ESSAYS ON SOCIAL INFLUENCE

BY

DAVID JAMES HENDRY

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Political Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor James H. Kuklinski, Chair
Professor Wendy K. Tam Cho
Professor Jeffery J. Mondak
Assistant Professor Milan Svolik
Assistant Professor Cara J. Wong
Abstract

This dissertation explores various aspects of social influence processes in political behavior research. Specifically, three separate essays explore the ways in which communication, social conformity, cultural transmission, and other interpersonal influence processes operate to generate similar public behaviors among groups of interconnected individuals. Broadly, the dissertation project develops theoretically a set of general principles about the cognitive processes, social dynamics, and institutional structures that shape the development, transmission, and transition of social norms. The theoretical arguments are used to inform a set of experimental research designs that provide empirical evidence for some of the short-term processes at work, as well as a set of simulations that are informative about the more long-term processes.

After a brief introduction in Chapter 1, Chapter 2 presents a simulation model of the cognitive tradeoff between stating a truly held internal attitude and avoiding social deviancy, as well as the dynamics of interpersonal influence in forced interaction settings. The key motivation behind the focus on these specific factors is the idea that for most people, many, if not most, day-to-day social interactions compel them to interact with people who they likely would not otherwise select to interact with absent common membership in some institution or organization. It is in these settings with unfamiliar others that the tradeoff between the expressive satisfaction of stating what one truly believes and the social satisfaction of exhibiting the group majority behavior is likely to be most obvious. One takeaway point from this essay is that when individuals place relatively equal weight on the values of internal consistency and social conformity, the behavior and overall satisfaction levels of the society at large become much less predictable in the short term, and the average overall satisfaction of the population takes its lowest value. That is, populations that place significant weight
on either conformity or consistency experience more satisfaction on average.

Chapter 3 presents a series of laboratory experiments on social influence. In addition to dealing with non-self-selected group interactions and conformity pressures, as in Chapter 2, several other ideas drove the design employed. One goal was to study very short-term micro-level processes of social interaction, and the benefit of the experimental setting is the ability to measure, rather than merely assume, the relative weight that individuals place on internal consistency, as opposed to social conformity. But the largest goal was to take a set of general ideas from the empirical research on social conformity, and apply them specifically to the question of individual expression of equality between groups. In a prediction that I call the *Social Dissonance Hypothesis*, I suggest that individuals come to the laboratory both with their own internal attitudes and some general notion about the range of attitudes that are socially acceptable in the broader population. When asked to give their views in isolation, some individuals are willing to express an attitude that runs counter to what is socially acceptable in the broader population. And among these people, there is a very strong tendency to change responses when they face a unanimously opposed majority, and they are compelled to express their attitudes publicly.

Chapter 4 develops a second simulation model in which the intention is to move away from forced interaction settings. In this model, the key moving parts are the abilities of individuals to self-select into social relationships, the assumption that individuals can influence each others’ types through interaction over time, and that individuals are characterized by traits in more than one dimension. Self selection into social relationships is obviously ubiquitous in reality, and the idea that individuals can influence types is meant to capture the intuitive sense that friends tend to become more alike through more interactions. The assumption that individuals are multidimensional captures the idea that social relationships might be formed on the basis of one dimension, but later other information may come to light that can make those social relationships more or less costly. The model shows that types that were relatively successful tended to be imitated more, and therefore ended up dominating the population over time. No type tended to have an advantage in general, but for any given simulation, an agent’s type was consequential for the ultimate utility experienced. A final chapter offers a very brief conclusion.
To Mom, Dad, Amy, and Sunhee, with the utmost love, respect, and thanks.
Acknowledgments

My successful completion of this dissertation and graduate school would not have been possible without the help of a large number of family, friends, and colleagues. The people that deserve thanks are too numerous to completely cover here. But a select few deserve specific mention.

First and foremost, I owe the largest debt of gratitude to my family. To my sister, Amy, my first teacher, this is the seed that you planted bearing fruit. To Mom and Dad, I could never possibly thank you for everything that you have done to get me to where I am today. Your love and support in everything that I pursued was one of the main reasons that I could stick with it until the end. I’m sorry that it has taken so long.

After beginning graduate school at Florida State University, a happy accident led me to the University of Illinois. I am extremely grateful to Jeff Mondak for encouraging me to transfer. Soon after arriving at Illinois, a research assistantship with Jeff, as well as a course with Scott Althaus, led to my first publications. For these, I will be forever indebted to Jeff and Scott. Following that, a series of research positions allowed me to have the flexibility to more thoughtfully develop my dissertation topic and pick up a variety of technical skills along the way. For hiring me in various research capacities, I am exceedingly thankful to Cara Wong, Gisela Sin, Brian Gaines, Bob Rich, and, especially, Jim Kuklinski.

After developing a dissertation topic, I sought the services of Wendy Cho, Jim Kuklinski, Jeff Mondak, Milan Svolik, and Cara Wong to serve on my committee. All obliged, and all offered me thoughtful counsel whenever I sought it. Special thanks in this regard are due to Wendy Cho, who encouraged and helped me to seek out a variety of outside opportunities to better develop my dissertation and other projects. But the utmost thanks are due to Jim Kuklinski. As a teacher, no one pushed me to work harder; as a mentor, no one was more
generous with his or her time; as a co-author, no one helped me to produce more thoughtful scholarship. With Jim, I have built a working relationship that I hope will carry on long into the future. And I know that I have gained a friend. I could never repay Jim for what he has done for me, and my work will be forever influenced by him.

I was also very fortunate to have been surrounded by a group of fellow graduate students who have been, and continue to be, great colleagues and friends. Meals, work conversations, help of various kinds, and all-around good company came with or from Andy Bloeser, Jason Coronel, Matthew Hayes, Aya Kachi, Dongsuk Kim, Dona-Gene Mitchell, and, especially, Sergio Wals, one of the most generous people I’ve ever had the pleasure of meeting.

Additional thanks are due to the Department of Psychology at the University of Illinois. The experiments performed as part of this dissertation required laboratory facilities that are not normally available to members of the Department of Political Science. After taking a course with Mike Regenwetter, I told him of my proposed research plan and asked whether he knew of any facilities on campus that would fit my needs. Mike then generously vouched for me and began the process of securing permission to use the Psychology labs. For getting the ball rolling, I am extremely grateful to Mike, and for securing the final permission, I am extremely grateful to Dave Irwin.

And most importantly, I would like to thank my wife, Sunhee Park, for her enduring love and encouragement throughout this entire process. Words cannot say what her support has meant to me. We started this process together, and we are now moving on to bigger and better things as partners. My life would be incomplete without her.

To anyone who was inadvertently left off of this list, please accept my apology.
# Table of Contents

Chapter 1  Introduction ................................................................. 1

Chapter 2  Norm-Development in Non-Homophilic Group Interactions: Model and Simulations .................................................. 4
  2.1 The Norm of Racial Equality in Public Speech .......................... 6
  2.2 Micro-level Cultural Transmission: Small Groups and Social Conformity ................................................. 7
    2.2.1 Theoretical and Methodological Framework ..................... 9
  2.3 The Model ..................................................................... 10
  2.4 Example Model Dynamics .................................................. 13
  2.5 Numerical Experiments ...................................................... 16
    2.5.1 Results .................................................................. 17
  2.6 Discussion ..................................................................... 28
  2.7 Conclusion ..................................................................... 32

Chapter 3  Laboratory Evidence on Social Influence ....................... 33
  3.1 Social Influence ................................................................. 35
    3.1.1 General Notions of Social Influence ............................. 35
    3.1.2 Social Influence Through Interpersonal Interactions .......... 36
    3.1.3 Social Conformity ..................................................... 37
  3.2 Development of Hypotheses ............................................... 39
  3.3 The Social Pressure Experiments ......................................... 44
    3.3.1 The Behavioral Task .................................................. 44
    3.3.2 Participants ............................................................. 45
    3.3.3 The Looming Social Pressure Experiment ..................... 48
    3.3.4 The Real-Time Social Pressure Experiment ................... 68

Chapter 4  Homophilic Agent Interactions and Network Development: Model and Simulations ............................................. 91
  4.1 Introduction ..................................................................... 91
  4.2 Strategic Model of Network Formation ................................ 93
    4.2.1 Simultaneous Link-Announcement Game with Heterogeneous Agents ......................................................... 94
    4.2.2 Illustration: Four Agents and Four Types ..................... 95
  4.3 Numerical Experiments ...................................................... 99
  4.4 Discussion and Conclusion ............................................... 110
Chapter 1

Introduction

In the contemporary world, norms of individual equality are a staple of democratic political culture. Following long histories of political, economic, and social systems based on gender, racial, religious, and other forms of hierarchy, principles of equal protection of citizens’ rights are now codified into law in virtually all modern participatory democracies. Across democratic societies, citizens consistently exhibit high levels of support for the idea that all people should be treated as political and social equals, without regard to personal characteristics such as gender, race, religion, or socioeconomic status. Based on public opinion measures, support for principles of equality seem to be a pervasive feature of modern democratic polities.

To be sure, though, general popular acceptance of egalitarian norms has its limits. Even genuine support for abstract notions of equality tend to mask a patchwork of compromises that individuals are willing to tolerate under particular circumstances. Additionally, though various forms of inequality persist in modern democracies, citizenries often show very little support for government action to address them. Even more important, however, the expressed levels of support for certain egalitarian principles may not be genuine at all. Because citizens in democratic societies generally carry the belief that egalitarian principles are widely held, publicly expressing attitudes of an inegalitarian nature may carry the possibility of embarrassment, or even social sanction. In fact, this threat may have been reinforced through repeated social interactions over time. In the end, among a possibly substantial subset of a given populace, self reports of adherence to egalitarian principles may reflect conformity to social rules that govern appropriate behavior rather than truly held beliefs.

Though it is plausible to believe that this sort of social desirability bias would serve to inflate expressed levels of public support for egalitarian principles, little is known about how
the desire to give the appropriate social response operates at the level of interpersonal interactions. That is, within a broader social climate that evokes a strong norm of egalitarianism, we know little about how everyday interpersonal communications serve to perpetuate both the norm of acceptable behavior and variation in individuals’ actual attitudes. Empirical studies of social conformity with respect to general, arbitrary norms established through social interactions have a long history in the social sciences (Asch 1951, 1952, 1956; Deutsch and Gerard 1955; Sherif 1935, 1936). However, with respect to political and social attitudes, empirical examinations of conformity at the interpersonal level are virtually nonexistent (but see Verhulst and Levitan 2009). Therefore, though we have strong reasons to believe that expressed public support for egalitarian principles on surveys are inflated, we know little about how this norm structures, and is structured by, interactions between individuals.

The current study examines various aspects of the relationship between interpersonal influences and the levels of egalitarianism that individuals are willing to express in public and private settings. A strong theme throughout will be the maintenance of a distinction between individuals’ true attitudes, and the attitudes that those same individuals are willing to express publicly. Using formal theory, computational simulations, and a set of laboratory experiments, I relate individual-level cognitive processes coupled with interpersonal interactions to large-scale societal changes. In so doing, I attempt to shed light on general processes of norm development, transition, and change.

The dissertation proceeds as three freestanding essays on social influence in the political context. Though some portions may contribute to understanding of social influence processes in more general contexts, a strong theme bringing the essays together is the relationship of the social influence process under examination to the study of the development, transition, and transmission of egalitarian and inegalitarian norms. In the following chapter, I present a formal model of a particular social influence process in non-self-selected (or non-homophilic) group settings. I then translate the model into a series of computational simulations in order to derive important implications. In the subsequent chapter, I provide a brief review of existing literature on social influence processes in order to derive a set of micro-level empirical hypotheses and test them in a series of laboratory experiments. Following the experiments, another chapter offers a formal model of actor behavior when self-selection of social partners
(or homophilic interaction) is possible, and again uses computational simulations to derive some important results. A final chapter offers concluding remarks.
In the twentieth century, the United States underwent a massive cultural shift. Norms of public discourse with respect to racial equality were completely reversed. The idea that the races could and should be considered social and political equals went from being a view rarely communicated outside of a small number of intimate social groups to being the only acceptable mode of public expression regarding race. This transformation has been deep, wide, and pervasive, effecting large-scale change in discourse across virtually all contexts of public communication. In most white social circles—in casual friendships, in school, in the workplace, in politics—whereas open expression of racial egalitarianism had once been a cause for social censure, eventually any statements of an inegalitarian nature would be met with strong disapproval, and could potentially cause irreparable damage to a reputation.

The scope of this cultural change is difficult to understate. The body politic of the United States in the twentieth century experienced a complete displacement of a prevailing social norm that had helped to define and constrain both mass and elite rhetoric since the founding of the nation. From establishment, through the Civil War, through Reconstruction, and into the twentieth century, public agreement with the idea of the inherent superiority of whites over African Americans was a necessary condition for entry into national, state, and most local politics. In particular, for the vast majority of white elected officials of the Southern United States, at least a nominal defense of racial separation was necessary until the 1960s, and, in some cases, beyond. The necessity of this stance for political survival and career advancement was a direct consequence of the then-prevailing social norm of racial inequality in public speech. In rapid historical fashion, however, the idea of racial equality, once publicly held by only an extremely small minority of the population, grew in the latter half of the twentieth century to be the prevailing social norm, thereby displacing the norm of racial
inequality.

To claim that this shift in American culture has occurred is relatively uncontroversial. Observers of American politics and history will surely note the dramatic conversion of the United States from a country with a legalized system of black slavery to today’s legal, social, and political environment that promotes values such as racial tolerance, inclusion, and diversity, and in which being labeled a racist or bigot is likely to make one a social outcast. But given the status quo of American thought on matters of racial differences (and particularly of white racial attitudes) in the mid-twentieth century, the complete turnaround by the century’s end certainly does not appear to be a priori self-evident. In fact, given the level of expressed white resistance to the idea of blacks as social and political equals, and the widespread willingness of elites to exploit and take on leadership roles in that resistance for electoral gains, the eventual entrenchment of the norm of racial equality is a puzzling phenomenon. The rapidity with which the change occurred is even more profound. And the implications for American politics have been far-reaching, permeating social interactions between and among elites and masses, changing the nature of political campaigns nationwide, altering the cleavage structure of major-party political competition, and helping to redefine the coalitions of electoral support for the major political parties.

How did this process of norm displacement occur and, just as importantly, how did it occur so quickly? In this chapter, I address an important piece of the answer to these question by situating the process of the displacement of norms of public speech within a broader story of cultural evolution and focusing on the theoretical impact of forced social interactions on the motivations of individuals. In particular, I explore individual agents’ sometimes competing desires to publicly express views that are consistent with their own internal preferences on the one hand, and to maintain social standing within their networks of social relationships on the other. Under a system of homophilic selection into networks of relationships within a population that is relatively homogeneous with respect to internal preferences, desires for internal consistency and social acceptance are unlikely to be in tension. However, when institutional mechanisms construct non-homophilic networks of social interactions, these two desires may frequently be opposed. When individuals are forced to interact with non-familiar others, and when exogenous events lead individuals to be rela-
tively uncertain about the distribution of preferences in the population, it is reasonable to expect that publicly expressed opinions may not necessarily comport with privately held values. Individual idiosyncracies in estimating the distribution of preferences along with the desire to maintain social status may establish conditions conducive to the rapid transition from one norm of public expression to another. And importantly, the evolution of norms of public expression can occur independently of any changes in internal values.

The chapter proceeds as follows. In the following section I briefly discuss the historical context of the norm change that is the basis of my theoretical perspective in order to provide a direct link between the substantive phenomenon and the more abstract model presented herein. Following this, I situate the individual-level mechanisms at play in the theory within the broader social scientific literature on cultural evolution and social pressures to develop insights into the individual-level mechanisms underlying key aspects of the dynamics of norms of public speech. Then, I present a simplified model that formalizes these individual-level mechanisms and some of the features of social structure within small-scale institutions that are thought to be important components to norms of public discourse on sensitive social issues. After this, I use the model to perform numerical experiments that demonstrate how the proposed individual-level mechanisms lead to population-level outcomes as a function of key parameters. Following the numerical experiments, I offer concluding remarks.

2.1 The Norm of Racial Equality in Public Speech

As stated previously, the transition from a norm of racial inequality to a norm of racial equality in public speech has permeated all levels of society, having failed to penetrate only a small minority of closed social groups. In the most general terms, the process of change unfolded over a period of several hundreds of years, operating at a relatively slow overall rate for much of U.S. history, and finally achieving a rapid pace in the latter half of the twentieth century (see, e.g., Mendelberg 2001, chs. 2 and 3 for a more detailed historical development). Accepting this description as an accurate characterization of the macro-level process of norm transition does not preclude the existence of significant variation in the pace of micro-level change. In fact, the society-wide displacement of the norm of racial inequality by the norm of
racial equality was marked by different rates of piecemeal progression and digression across social groupings and geographic locations. Over time, the complex relationship between elite leadership, public opinion, and mass political action gradually advanced the social and political rights of African Americans and established an ever-changing set of legal codes and institutions that structure interracial interactions.

With respect to changes in institutions, of particular significance was the patchwork of public policies aimed at desegregating previously racially segregated institutions. Policies to combat racial discrimination in various aspects of society along with the eventual vigorous enforcement of judicial orders to desegregate schools led to a system of interracial contacts that had previously been unimaginable. In the modeling exercise that follows, while I do not specifically model interracial interactions, the model is intended to capture aspects of interactions with unfamiliar others. In a changing environment in which it is known that interracial contacts are occurring at an increasing rate, the consequences of making racist or inequitable statements even in an intraracial context could be severe. It is this newfound uncertainty that I contend was among the root causes of movement away from a norm of racial inequality to a norm of racial equality. No work has focused on the specific impact that changes in the degree and character of interracial contacts has had on the social acceptability of speech espousing racial inequality. In survey research on racial attitudes, the potential social desirability biases of direct measures are widely known and often mentioned, but a coherent theory or body of empirical evidence on the origin of such biases is absent.

2.2 Micro-level Cultural Transmission: Small Groups and Social Conformity

To develop a model of the micro-level process of norm development and evolution that has general appeal, and that can be readily applied to the case of the United States, I begin with the assumption that conformist transmission (also referred to as frequency-dependent bias) is adaptive under a wide range of conditions. That is, rather than develop a specific model of the genetic development of conformist transmission in humans, I appeal to the literature with respect to its origins and dynamics (Boyd and Richerson 1985, 2005; Henrich 2001;
Henrich and Boyd 1998, 2001) and begin at the starting point of conformist transmission as a pervasive human phenomenon. In brief, conformist transmission is the tendency of individuals in a population to acquire a particular cultural variant due to the relative frequency of that variant among the individuals’ cultural “parents” (Boyd and Richerson 1985, p. 10). Boyd and Richerson (1985, 2005) and Henrich and Boyd (2001) note that conformist transmission is not identical to the concept of social conformity as it has typically been examined in social psychology (e.g., Asch 1951; Sherif 1935), and the difference is rooted in the motivations of individual agents. Specifically, conformist transmission results in the adoption of cultural traits that are popular because individuals use popularity as an indirect measure of a trait’s value. Social conformity, on the other hand, results in the adoption of traits that are popular because individuals consciously try to avoid appearing deviant (Henrich and Boyd 2001). The former represents an internalization of a cultural variant, while the latter can be a simple change in observable behavior unaccompanied by internal changes.

If we consider norms of public speech to be cultural variants, the distinction between the concepts of conformist transmission and social conformity becomes murky. It is plausible to think that changes in public speech with respect to a social referent can and do occur without always being accompanied by an associated change in internal attitudes with respect to that same referent. Using the characterization of Henrich and Boyd (2001), such a process would suggest that patterns of public speech at the micro level could be the result of social conformity. At the macro level, however, large-scale changes in norms of public speech with respect to a social referent are the result of a broad internalization of what does and what does not constitute deviant behavior. Notice that I am not claiming that changes in norms of public speech must be accompanied by large-scale shifts in attitudes. Rather, I am explicitly maintaining a distinction between public speech and internal attitudes. Further, though it is clear that public speech and internal attitudes with respect to the same social referents must be related in some ways, it is also reasonable to expect that certain environmental conditions may cause the two to evolve on different—and sometimes even opposing—trajectories. Specifying those conditions theoretically and deriving logical implications are some of the goals of the present chapter.

The dynamics of norms of public speech at the level of the small group, as they have
been presented here, can be summarized as follows. Norms of public discourse are developed and maintained through a process of social learning and cultural transmission. Over successive instances of publicly expressing one’s opinion with respect to a particular social referent (e.g., race), individuals use the frequency of a publicly expressed value as an indirect measure of its appropriateness in social settings. Positive social reinforcement of the expressed view acts to strengthen the enculturation process. Additionally, individuals identify unacceptable public pronouncements by either witnessing or directly experiencing the social censure of unpopular expression. When disapproval of a particular expressed value is sufficiently frequent or intense, individuals internalize the unacceptability of that value as a social fact (Durkheim [1895] 1965). In this way, positive feedback and social censure act as reinforcement mechanisms that condition the attainment of particular cultural variants (Bandura 1977). However, internalizing the norms of appropriate social behavior does not necessarily indicate the internalization of congruent internal attitudes. Critically, norms of public expression can and often do run counter to privately held individual preferences. Public compliance with the expressed opinions of others with whom one interacts is considered to be a type of social conformity that results from a cultural evolutionary process of conformist transmission.

2.2.1 Theoretical and Methodological Framework

The transition from a norm of public rhetoric that compelled individuals to express a view of the inherent superiority of whites over other races to the present norm of public rhetoric that compels individuals to express opinions consistent with full racial equality represents a truly profound change. To provide a rigorous explanation of the social processes that brought about this change, I do not simply reconstruct and describe the specific historical details. Rather, my purpose is to supply a more general description of the essential features of social interactions that could plausibly have brought about such a change (see, e.g., Carnap 1962; Ullmann-Margalit 1977). That is, given a set of reasonable assumptions and expectations about human behavior, the nature of social interactions, and the consequences of policy changes, I intend to show that the transition in norms of rhetoric was in fact a plausible
With respect to the dynamics and social effects of interpersonal communication within institutions, I argue that a variety of institutional contexts, including schools and workplaces, serve to structure interpersonal interactions between individuals who may not necessarily select into direct social relationships outside of those settings. That is, membership in settings like schools and workplaces offers opportunities, and sometimes compels individuals, to communicate with a wider and more diverse group of peers than is likely to be present in, for instance, familial or friendship groups (Granovetter 1986; Huckfeldt and Sprague 1995; Mutz and Mondak 2006). As in other institutional contexts that structure social interactions, individuals within forced-interaction settings develop a shared understanding of the types of public communication that are socially acceptable with respect to a broad spectrum of social referents (Mead 1934), including race and racial equality. Through repeated interpersonal interactions accompanied by positive and negative social reinforcement (e.g., Bandura 1977), individuals internalize norms of socially acceptable public speech with respect to race that may or may not be in line with their actual racial attitudes. When called upon to make public pronouncements about race or racial equality, if behavioral norms and true attitudes are in tension, individuals must then balance preferences for stating what they truly believe with preferences for not appearing deviant from their broader social group.

2.3 The Model

I now present a model of the public expression of attitudes with respect to sensitive social issues that accounts for the features of non-homophilic social interactions structured by institutions and the preferences of individuals to conform to the behaviors of others with whom they are socially connected. The purpose is to capture the types of intimate interactions that are forced upon individuals within institutional and organizational settings such as schools and workplaces. Following Akerlof (1980) and Jones (1984), I represent the social status of agents by the utility that they derive from behaving in accordance with others. But as in Bernheim (1994) and Kuran (1995), preferences for conformity must be balanced by
preferences to exhibit behavior that is consistent with internal values.\(^1\) The model will intentionally take on an abstract quality, but the intention is to provide a reasonable explanation for an important aspect of the dynamics of rhetorical norms described above.

Formally, let \(N = \{1, \ldots, n\}\) be a set of agents represented as the nodes of a graph, \(g \in G(N)\), where \(G(N)\) is the set of all undirected graphs on \(N\). Further, let \(N_i(g) \subset N\) be agent \(i\)'s neighborhood of direct connections. Agent \(i \in N\) is characterized by a fixed scalar \(v_i \in [0, 1]\) representing preferences over a behavior, \(b_i \in \{0, 1\}\), that is observable for all \(j \in N_i(g)\). In the vein of the previous discussion, we may think of agents as having more or less well defined preferences for racial equality or racial inequality (or egalitarianism or inegalitarianism with respect to some other sensitive social issue). If \(b_i = 1\) indicates a public pronouncement of agent \(i\) in favor of, say, egalitarianism, and \(b_i = 0\) in favor of inegalitarianism, then we may think of \(v_i > 0.5\) as representing an internal preference for egalitarianism and \(v_i < 0.5\) as representing an internal preference for inegalitarianism.

Each agent in the population is compelled to select some publicly observable behavior \(b \in B \equiv \{0, 1\}\). Agents have intrinsic preferences over \(B\) represented by the payoff function \(f(b_i, v_i)\), where \(f(\cdot)\) achieves a maximum when \(v_i = b_i\) and is assumed to be symmetric. In addition, agents have extrinsic preferences over \(B\) represented by the payoff function \(h(b_i, \phi_i, g)\), where \(h(\cdot)\) achieves a maximum at \(b_i = \phi_i\) and is assumed to be symmetric, and where \(\phi_i\) is a summary representation of agent \(i\)'s estimate of the behaviors of others in her neighborhood of direct connections\(^2\) (to be explained below). Given the assumptions imposed on \(f(\cdot)\) and \(h(\cdot)\), a convenient representation is to let \(f(b_i, v_i) = -(b_i - v_i)^2\) and \(h(b_i, \phi_i, g) = -(b_i - \phi_i)^2\). Agent \(i\) chooses \(b_i\) to solve \(\text{arg} \max_{b_i \in B} U(b_i, v_i, \phi_i, g)\), where

\[
U(b_i, v_i, \phi_i, g) = \alpha f(b_i, v_i) + (1 - \alpha) h(b_i, \phi_i, g)
= -\alpha (b_i - v_i)^2 - (1 - \alpha) (b_i - \phi_i)^2,
\]

\(^{1}\)In other words, the balance is between conformity with others and consistency between internal values and expressed behavior. For a similar take on a different phenomenon, Bednar et al. (2010) and Page, Sander, and Schneider-Mizell (2007) use the same terminology of conformity and consistency to develop models of the balance between conformity with others and internal consistency across a variety of issue dimensions.

\(^{2}\)The term “neighborhood of direct connections” can be substituted with the term “group.” Given the choice of notation, I will use the language of graph/social network theory.
where $\alpha \in [0, 1]$ is a relative weighting parameter determining the value that agents place on publicly expressing their internal preferences as compared to the value that they place on not appearing deviant from their immediate social group. In other words, $\alpha$ is the relative weight given to cognitive consonance as opposed to social consonance. Additionally, the additive form of $U(\cdot)$ shows that it is simply a convex combination of weighted intrinsic and extrinsic payoffs.

For a specified number of time steps, agents are ordered according to their social status as determined by their extrinsic payoff in the previous period, and in this order each agent chooses an observable behavior. Allowing social status to determine the order in which agents publicly express an opinion on a sensitive social issue can be thought to capture concepts such as boldness or confidence. Let $S \equiv \sigma(N) = (s_1, \ldots, s_n)$ be a permutation of $N$ that lists agents in descending order according to their status. For example, suppose in a population of three agents that agent 3 has the highest status, agent 1 the second highest, and agent 2 the lowest. We would then have $S \equiv (s_1, s_2, s_3) = (3, 1, 2)$. In a given time step, after each agent $s_j$ chooses $b_{s_j}$ her action is observable by all other agents in her neighborhood of direct connections, denoted $N_{s_j}(g)$. Therefore, any agent $s_k$ such that $j < k$ and $s_j \in N_{s_k}(g)$ will be able to incorporate the observation of $b_{s_j}$ into her estimate of the distribution of behaviors in the neighborhood, $\phi_{s_k}$. Formally, let $m_i$ be the total number of agents in the neighborhood of direct connections for agent $i$. Then, at time step $t$, agent $i = s_k$ develops an estimate of the distribution of behaviors in her neighborhood according to the following rule:

$$
\phi_{s_k,t} = \frac{1}{m_{s_k} - 1} \left[ \left( \sum_{j < k : s_j \in N_{s_k}(g)} b_{s_j,t} \right) + \#\{j > k : s_j \in N_{s_j}(g)\} p_{t-1} \right],
$$

(2.2)

where $p_{s_k,t-1}$ is the mean of the observable behaviors among other agents in the population in the previous time step. Specifically,

$$
p_{t-1} = \frac{1}{n} \sum_{j \in N} b_{j,t-1}.
$$
Substantively, the form of $\phi_{sk}$ shows that agents are establishing an estimate of the distribution of behaviors in their neighborhood of direct connections by observing others who have gone before them in the current round, and combining this with the information from the entire population in the previous round.

After all agents in $N$ choose a behavior to solve $\arg \max_{b_i \in B} U (b_i, v_i, \phi_i, g)$, each receives an actual payoff of $U (b_i, v_i, p, g)$. That is, the extrinsic payoff actually received is based on the average behavior of other agents rather than the mere estimate employed to choose behavior. Agents’ extrinsic payoffs in time $t$ then determine the order in which agents choose behavior in time $t + 1$, and so on. The operation of the model dynamics can be summarized as follows:

1. Define $N = \{1, \ldots, n\}$ and choose $g \in G (N)$ and a group size, $k$;
2. For each $i$ in 1 through $n$, assign $v_i \in [0, 1]$, and an initial social status, $q_0 \in [0, 1]$;
3. Define $S = \sigma (N) = (s_1, \ldots, s_n)$ so that agents are placed in descending order according to their social status;
4. Agents $s_1$ through $s_n$ calculate $\phi_{sj}$ and choose $b_{sj}$ to solve $\arg \max_{b \in B} U (b_{sj}, v_{sj}, \phi_{sj}, g)$;
5. Agents receive payoff $U (b_i, v_i, p_i, g)$ and $h (b_i, p_i, g)$ becomes agent $i$’s current status;
6. Repeat steps 3 through 5.

### 2.4 Example Model Dynamics

To unpack the black-box aspect of the above description, I now present an illustration of the one-period dynamics of the model for a contrived example population. Figure 2.1 presents an example population of 8 agents, with a neighborhood or group size of 4. In this example, agents 1 through 4 form one group and agents 5 through 8 form a second group. Agent preferences, $v_i$, and social status in the previous period, $h_{i,t-1}$, are presented above or below the respective nodes.

From the graphical presentation in Figure 2.1, the network structure of the population becomes clear. Groups are actually fully connected island cliques embedded in the larger population. That is, within groups, all members are connected to all other members, and there is no overlapping membership between the several groups. Additionally, note that the subscripts indicate that agents’ social statuses are indexed by time, while their preferences
are not. Though it is plausible to think that individual preferences are subject to change over time, the simple formulation examined here is intended to capture behavioral dynamics over a relatively brief span of time. Hence, preferences are assumed to be static. Social status, on the other hand, changes each round as a function of each agent’s behavior in relation to the behavior of their neighborhood of direct connections and that of the broader population.

In this particular example, agents begin with a social status determined by the behaviors exhibited in the previous period. Given the values of $h_1$ through $h_8$, the permutation that places agents in descending order according to social status is $S \equiv \sigma (N) = (3, 8, 5, 6, 1, 2, 4, 7)$. Further, assume that in the previous round, 3 out of the 8 agents exhibited behavior 1, so that $p = 0.375$. And for simplicity, assume that agents place equal weight on expressing their true preference and on not appearing deviant from their social group, so that $\alpha = 0.5$. Given that agent 3 is compelled to be the first mover, she does not have any information about the behavior in her neighborhood of direct connections in the current period. She does, however, know the distribution of behaviors in the broader...
population in the previous time period. Therefore, armed with a weak preference for behavior 0 \((v_3 = 0.409)\), and the knowledge that 5 out of the eight agents exhibited behavior 0 in the previous period, there is no inherent conflict between internal and social preferences, and after forming the estimate of \(\phi_3\) agent 3 chooses behavior 0. After agent 3, agent 8 is compelled to exhibit a behavior. Agent 8 is the second mover in the population, but the first mover in her particular neighborhood of direct connections. Like agent 3, agent 8 uses only her internal preference and the population distribution of behaviors in the previous period to make her decision. In this case, however, a conflict arises because although agent 8 has a relatively strong preference for behavior 1 \((v_8 = 0.892)\), she knows that a majority of agents in the previous period chose behavior 0. Agent 8 forms her estimate of neighborhood connections, \(\phi_8\), according to equation \((2.2)\) and then solves \(\arg\max_{b_i \in \mathcal{B}} U(b_i, v_i, \phi_i, g)\) according to equation \((2.1)\), finding that behavior 1 is the optimal choice. Agent 5 is then compelled to exhibit a behavior, with the ability to account for the behavior of agent 8 in the current period; then agent 6, with the ability to account for the behaviors of agents 5 and 8; then agent 1, with the ability to account for the behavior of agent 3; then agent 2, with the ability to account for the behaviors of agents 1 and 3; followed by agents 4 and 7, each with the ability to account for the behaviors of all of their fellow group members.

In each period of the model, each agent \(i\), in order according to her social status, simply estimates \(\phi_i\) and then uses that estimate to choose \(b_i\) to optimize equation \((2.1)\). Then, after each agent exhibits a behavior, they earn an actual extrinsic utility that is based on their own behavior in relation to the behaviors of the population at large. Extrinsic utility then becomes agents’ social status, and they begin the next iteration in the order determined by that social status. Table \(2.1\) summarizes the one-period dynamics for the population illustrated in Figure \(2.1\).

As in the previous period with which this example began, we again see that three agents exhibit behavior 1 and five exhibit behavior 0. Further, if the behaviors of individual agents are matched to their internal preferences shown in Figure \(2.1\), we see that agents 2 and 4 exhibit behavior 0 even though they have an intrinsic preference for behavior 1. At least two features of the setup lead to this outcome. First, agents 2 and 4 have lower social status than their fellow group member, agent 3, who has a preference for behavior 0. Second, agents 2
Table 2.1: One-Period Dynamics for Example Population

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
<th>Step 6</th>
<th>Step 7</th>
<th>Step 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
<td>(\phi_i, b_i)</td>
</tr>
<tr>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
<td>0</td>
<td>0.30</td>
<td>0</td>
<td>0.30</td>
<td>0</td>
<td>0.30</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
<td>0.53</td>
<td>1</td>
<td>0.53</td>
<td>1</td>
<td>0.53</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.77</td>
<td>0</td>
<td>0.77</td>
<td>0</td>
<td>0.77</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.67</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>0.30</td>
<td>1</td>
<td>0.30</td>
<td>1</td>
<td>0.30</td>
<td>1</td>
<td>0.30</td>
</tr>
</tbody>
</table>

and 4 know that the majority behavior in the previous period was behavior 0. Combining both of these facts, when agent 2 acts she knows that one of her fellow group members exhibited behavior 0 in the current period and that the majority of the population at large exhibited behavior 0 in the previous period, and therefore herself earns greater estimated utility for exhibiting behavior 0 because of the desire to not appear deviant. Then when agent 3 acts, she knows that two of her fellow group members exhibited behavior 0 in the current period and, again, that the majority of the population exhibited behavior 0 in the previous period. Agent 3, too, estimates that her utility would be higher by choosing behavior 0. Just like the problem of cooperation in prisoner’s dilemma-type settings, if agents 2 and 4 could coordinate their behavior, they could together alter the majority behavior to be in line with their own preferences. But the uncertainty about others’ intrinsic preferences and the desire to not appear deviant lead them to falsify those preferences by acting in an opposed fashion.

### 2.5 Numerical Experiments

A set of numerical experiments were constructed to explore variation in some of the key features of individual-level preferences and social context that are expected to impact the distribution of behaviors in the population over time. With a chosen population size, \(n\), and group size, \(k\), graphs were constructed so that each agent was assigned to a group of
size $k$ or less.$^3$ The resulting networks take the form of those presented in Figure 2.1. In these numerical experiments, the purpose of random assignment to groups is to capture non-homophilic interactions, and the purpose of random assignment to island-cliques is to capture an important feature of schools, workplaces, and other institutional and organizational social settings that structure interpersonal interactions. That is, institutions composed of many actors structure interactions around tasks that must be performed, and those tasks will call for some interactions between individuals to be very intimate and others to range from casual to non-existent. The particular network structures used to investigate the social dynamics employ the simplification that groups of individuals are intimately connected and there is no cross-group membership. This captures certain important features of situations like work groups, which structure short-term interpersonal relationships between individuals embedded within the same institutions.

For all of the experiments described herein, initial agent populations were constructed by drawing the vector $(v_1, \ldots, v_n)$, where $v_i \sim Unif(0, 1)$. As stated previously, this vector of characteristics remained static. In addition, each population under consideration took on assumed values of $n$, $k$, and $\alpha$, which were varied across simulations (to be explained below). To begin the iterative process, the initial social status of agents was determined by agent preferences, $v_i$ (and therefore was also randomly assigned). And from this starting point, the behaviors and social statuses of the agents were allowed to develop endogenously. The model was then allowed to iterate for a specified number of time periods and the behavior was examined.

### 2.5.1 Results

For the simulations that follow, the population size, $n$, was allowed to vary between 50 and 200 (taking values of 50, 60, 70, 80, 90, 100, and 200), the group size, $k$, was allowed to vary between 3 and 10 (increasing in increments of 1), and the weight placed on internal consistency, $\alpha$, was allowed to vary between 0.1 and 0.9 (increasing in increments of 0.1).

---

$^3$Agents in a given population were assigned to groups of size less than $k$ only if the ratio of $n$ to $k$ was not integer valued. For instance, suppose $n = 50$ and $k = 4$. In this case, agents would be assigned to 12 groups of size 4, and the 2 remaining agents would be assigned to a final group of size 2.
For each initial set of parameters, 500 simulations were run for 100 iterations. To take an initial look at the system dynamics over time, Figure 2.2 presents an illustration that holds \( n \) constant at 70, \( k \) constant at 5, and allows \( \alpha \) to vary over its range.\(^4\) In each panel, the thin lines represent the average behavior of the population at each iteration for each of the 500 simulated runs of the model. The thick black line is the average at each iteration across the 500 simulations.

The pattern exhibited for this choice of \( n \) and \( k \) is indicative of the general pattern that emerges for any given choice of \( n \) and \( k \). That is, for any value of \( \alpha \) less than 0.5, the system almost always quickly settles into an equilibrium in which all of the agents exhibit one or the other of the behaviors. Substantively, this makes perfect intuitive sense and is not at all surprising. Because \( \alpha \) captures the weight that agents place on internal consistency, that is, exhibiting a behavior that is consistent with their true preferences, low values of \( \alpha \) should be associated with greater levels of conformity and preference falsification. Since internal

\(^4\)The results for \( \alpha = 0.1 \) are omitted for presentation purposes, but suffice to say that the results are very similar to those for \( \alpha = 0.2 \).
preferences are random draws across the range of possible preferences, about half of the agents should hold a preference for behavior 1 and about half should hold a preference for behavior 0, on average and over the long run. Therefore, equilibria in which all agents exhibit one behavior are the outcomes with the highest levels of conformity. As $\alpha$ increases from 0.2 to 0.4, the number of iterations required for the system to fall into equilibrium increases slightly, but all simulated populations eventually do fall into an all-0 or all-1 steady state. Further, the location of the mean trend line indicates that there does not seem to be any inherent bias in the system leading to more choices of behavior 0 or behavior 1.

As we move to the situation in which agents place equal weight on their desires to express their own internal preferences and their desires to not appear deviant from the group, that is, when $\alpha = 0.5$, the general behavior of the system becomes more difficult to assess. A substantial proportion of the simulated runs fall into equilibria in which all agents exhibit behavior 0 or all agents exhibit behavior 1. But at least two characteristics are clearly different from the system dynamics for $\alpha \leq 0.4$. First, when simulated populations do fall into an all-0 or all-1 equilibrium, there is no apparent pattern to when this happens. For $\alpha \leq 0.4$ all simulated populations fall into one or other of these patterns prior to the twentieth iteration of the model. For $\alpha = 0.5$, however, populations are continually moving into an all-0 or all-1 equilibrium throughout the 100 iterations. Additionally, whereas for $\alpha \leq 0.4$ all simulation runs quickly fall into one of these steady states, for $\alpha = 0.5$ a substantial proportion of the simulations remain in flux through the first 100 iterations. It is possible many of these simulated populations would reach an all-0 or all-1 equilibrium if the system had been allowed to iterate beyond 100 periods. But given the behavior exhibited, it is also likely that many would continue to fluctuate indefinitely.

For $\alpha \geq 0.6$, the picture changes yet again. For these parameter values, no simulations of the 500 reach an all-0 or all-1 equilibrium. To be sure, some simulated populations do reach equilibria with a mix of behaviors in the population. But the conformity pressures are now reduced to the point that lower proportions of agents are willing to falsify their preferences in order to conform to the behaviors of others. As $\alpha$ increases from 0.6 to 0.9, the range of values of average behavior tighten further and further around the mean of 0.5. Again, the substantive interpretation of what is happening is quite intuitive. The pressure to conform
to a view that runs counter to one’s preferences is a direct function of the strength with which they hold the desire to maintain internal consistency. Values of $\alpha$ greater than 0.5 mean that agents value expressing a view that is consistent with internal preferences more than they value consistency with proximate others who can witness their behavior.

As stated previously, the pattern of results in Figure 2.2 is not particular to the chosen values of $n$ and $k$. With little exception, this is the pattern that holds for any values of $n$ and $k$ that could be chosen. But for $\alpha$ values greater than 0.4, it becomes difficult to ascertain just how many steady states are achieved among the 500 simulations. Close examination of the plots reveals a great deal of fluctuation across the first 100 iterations for many of the simulations. To get a better sense of how many populations do in fact achieve a steady state for a given set of parameter values, Figure 2.3 presents the proportion of the 500 simulations for each set of starting parameter values that reach equilibrium prior to the 90th iteration.\(^5\) That is, to be included in the proportions presented here, a population had to be in a steady state for at least 10 periods.

As shown previously, for $\alpha = 0.4$, 100% of the simulated populations fall into equilibria. Additionally, for $\alpha = 0.5$, a very high proportion of the populations eventually move into equilibrium. And with few exceptions, there appears to be an emergent pattern in which the proportion of equilibria reached is an increasing function of group size. That is, when agents place equal weight on expressing their true preference and not appearing deviant from their social group, it tends to be the case that the greater the number of individuals in one’s immediate social group, the greater is the propensity to fall into a state of clear, predictable behavior from which agents do not deviate. This is not something that necessarily would have been predicted ahead of time, but it makes intuitive sense upon reflection. Essentially, the larger the proportion of the broader population that the individual agent has more intimate interactions with, the more likely they are to begin to mirror the behavior of one another and become highly predictable.

When $\alpha$ increases beyond 0.5, yet another interesting pattern is revealed. Specifically, moving from $\alpha = 0.5$ to $\alpha = 0.6$, the proportion of simulated populations that reaches

\(^5\)Results for values of $\alpha \leq 0.3$ are not presented because they are identical to the results for $\alpha = 0.4$. Results for values of $n = 200$ are excluded due to space considerations.
Figure 2.3: Proportions of Simulations that Reach a Steady-State Equilibrium

Figure continued on next page.
Note: For a given set of parameter values, bar heights represent the proportion of simulations out of 500 that reach a constant population equilibrium behavior prior to the 90th iteration of the system.
equilibrium drops off dramatically, from almost 0 when \( n = 50 \) or \( n = 60 \) to exactly 0 when \( n \geq 70 \). Increasing to \( \alpha = 0.7 \), the proportion generally decreases even further or remains steady. But when \( \alpha \) increases to 0.8, the proportion of simulated populations reaching equilibrium begins to increase. And when \( \alpha \) moves from 0.8 to 0.9, the increase in the proportion is quite dramatic. That is, as agents move from holding only a slight preference for conformity over consistency to an overwhelming preference for consistency, the behavior of the population at large becomes more and more predictable. This is also something that was not predicted prior to observing the simulation results, but that can be explained easily upon further reflection. The high values of \( \alpha \) indicate a stronger preference for expressing views that are consistent with one’s internal preferences. Therefore, equilibria at high levels of \( \alpha \) are likely due to a relatively large number of groups composed of like-minded individuals expressing their true preferences. This idea is further supported by the relationship between the population size and the number of equilibria achieved, while holding \( k \) and \( \alpha \) constant. To see this, scanning down the panels of Figure 2.3, we see that for, say, \( \alpha = 0.9 \) and any given value of \( k \), the number of simulated populations that reach equilibrium decreases as \( n \) increases. If the above explanation is correct, through random sampling it should be expected that agents with similar values of \( v_i \) will occasionally end up assigned to groups together at relatively high rates in the population. When this occurs, public compliance between group members is not a result of a strong drive for not appearing deviant, but it nonetheless mirrors what would be expected when that drive is strong. As the population size increases, sampling will still lead to some groups composed of like-minded agents, but the rate at which this occurs will decrease on average. And this is exactly what is borne out by the results. As \( n \) increases from 100 to 200 (not presented), the number of simulated populations that reach equilibria falls to nearly 0 for all values of \( \alpha \geq 0.6 \).

The claim being made to explain the pattern of equilibrium behavior across simulations is the intuitive one that steady states achieved with values of \( \alpha \leq 0.5 \) will tend to be composed of a relatively large proportion of individual agents who are falsifying their preferences and simply complying with an established group and/or population behavior because they have a strong desire to not appear deviant. For \( \alpha \geq 0.6 \), however, equilibria will largely be composed of groups of like-minded individual agents exhibiting behaviors that are generally
in line with their internal preferences. To further buttress this idea, Figure 2.4 presents a summary of the distributions of the average proportion of agents who falsify their preferences across simulations when $n = 50$. That is, for $n = 50$, and for any values of $k$ and $\alpha$, the average number of agents who falsify their preferences over 100 iterations of the model were calculated, and these averages themselves were averaged over the 500 simulations run for each set of parameter values. The panels in Figure 2.4 display boxplots of those average proportions. The population size of $n = 50$ was chosen merely for illustrative purposes, but other population values not displayed lead to very similar results, with the major difference being that the range of the average proportion tends to decrease as population size increases.

The pattern displayed in the boxplots is exactly what would be expected. For $\alpha \leq 0.5$ we find the highest levels of individuals who falsify their preferences when they exhibit a behavior, and this level remains relatively constant between $\alpha = 0.1$ and $\alpha = 0.5$. At $\alpha = 0.5$, the range of the proportion of falsifiers begins to increase sharply, but the maximum value rarely goes above 0.5 and never goes above 0.6. At $\alpha = 0.7$, the range remains large, but
the densest portion of the distribution decreases dramatically. The median and the range of the distribution continue to decrease as $\alpha$ increases to 0.9. Taken by itself, Figure 2.4 presents the intuitive and expected result that when agents place more emphasis on internal consistency, they are less likely to falsify their behavior. But taken together with the results displayed in Figure 2.3, the clear result is that when $\alpha > 0.5$, as the proportion of agents falsifying their behavior decreases to near 0, the number of simulated populations that reach an equilibrium increases substantially. This indicates that when agents place more weight on the intrinsic utility of exhibiting behavior consistent with internal preferences, the proportion of groups exhibiting similar behavior in any given population tends to increase, while the number of people who falsify what they prefer simultaneously decreases.

The results presented thus far summarize the types of behavior that we can expect to play out on average over a range of different circumstances. Intuitively, when agents care less about expressing their true preferences, more agents tend to falsify their preferences in order to conform to the behaviors of others. And when this conformity occurs in larger numbers, more of the simulated populations carry on the conformity from period to period in steady-state equilibria. Decisions of how to behave are based partially on the estimated utility generated by pressure to conform to those in one’s group who have already chosen a behavior in the current round, as well as an estimate of how the remaining members of the group will behave based on the distribution of behaviors in the broader population in the previous iteration. But after all agents have chosen a behavior, the actual utility that they receive is based on actual behaviors exhibited, rather than estimates.

To summarize the utility gains that agents receive from under various circumstances, I once again employ an example population value and examine a summary of the distributions across values of $k$ and $\alpha$. I do so because, again, the same relationship seems to hold across values of $n$. Figure 2.5 presents summary distributions of the average utility earned across the 500 simulations for the experiments in which $n$ was set equal to 100. These utility calculations are again the average of the average utility for each simulated population across the 100 iterations. Interestingly, little variation in utility is exhibited for different values of $k$. This indicates that the intimacy of the group interactions has little explanatory power for the amount of benefit that agents receive when putting their behavior on public display.
Interestingly, for any given group size, the greatest level of average utility is generally received when agents place the lowest amount of weight on the internal consistency portion of their utility calculations. As that weight increases from $\alpha = 0.1$ to $\alpha = 0.6$, the variability of average utilities increases while their central tendency decreases. This is not unexpected, given the underlying assumptions of the model. Across the simulated populations, internal preferences are randomly assigned and remain static over time, whereas extrinsic utilities, or social statuses, are allowed to develop endogenously subject to initial conditions. It is reasonable to think that as agents place more and more weight on the portion of their utility calculation that is a function of unchanging parameter values, they will also have less and less control over utility maximization at the margins.

When the weight increases from $\alpha = 0.6$ to $\alpha = 0.9$, however, the average utility of the simulated populations begins to increase slightly. That is, when agents place less and less weight on the utility of conforming to the group behavior, away from the situation of equal weight on intrinsic and extrinsic preferences, the population seems to benefit more on average. This increase is only slight, but it is also somewhat unexpected. The likely
explanation for this result is that when the weights placed on intrinsic and extrinsic utility are closer to parity, agents will more often find themselves in circumstances in which they see the benefit in changing their own responses from the previous period, and they will therefore more often be attempting to achieve marginal gains in social status from iteration to iteration. Because all agents will find themselves in the same situation, there will simply be more fluctuation in population behavior from period to period, and therefore greater uncertainty in the estimates of the population behavior on which they partially construct their decisions. This explanation is further buttressed by the results presented in Figure 2.3, where I previously noted that the number of simulated populations that fall into a steady-state equilibria is often trivial at mid-range values of $\alpha$.

Further, we can also examine the question of how well agents who falsify their preferences fare in these simulated dynamics. To that end, Figure 2.6 presents the summary distributions of average utilities for the same population size as presented in Figure 2.5 ($n = 100$), but only for those agents in the population who falsified their preferences. Given that agents could potentially change their behavior from period to period, taking an average for each simulated population over the 100 iterations does not make sense in this case. The summary distributions are therefore taken at iteration number 50. And, again, suffice to say the choice of $n$ size on which to focus did not change the general statements that can be made about the system dynamics as $k$ and $\alpha$ vary.

For preferences falsifiers, Figure 2.6 shows that the same general pattern exhibited across values of $k$ holds as with the population at large, but it does so at a lower level of utility. In other words, as expected, falsifiers tend to earn less utility on average than non-falsifiers. For $\alpha = 0.1$, the difference in central tendency between falsifiers and the population at large is only slight, but as $\alpha$ increases to 0.2 and greater, the differences in utility become far more substantial. However, when focusing only on preference falsifiers, it is still the case that when $\alpha$ increases from 0.6 to 0.9, the central tendency of the utility also increases slightly. And here a similar explanation holds. Given the low proportions of simulated populations

---

6Summary distributions were examined at a variety of chosen iterations, and there was little variation in results. The exception to this general statement is that there tended to be greater variability exhibited at earlier iterations.
falling into equilibria at values of $\alpha$ around 0.6 and 0.7, the fluctuations in behavior from iteration to iteration mean that a steady and substantial proportion of the population are falsifying their preferences from period to period, each in reaction to the fluctuations in the population at large. This, in turn, creates greater uncertainty and poorer estimates of the behavior of others, and hence lower extrinsic utility on average. What holds for the larger population also holds for the preference falsifiers specifically. As agents place greater weight on intrinsic utility, behaviors become more predictable on average, and agents are better able to predict the social status they will obtain by exhibiting a certain behavior, and hence increase their social status at the margins.

### 2.6 Discussion

The model and numerical experiments are admittedly very abstract, but the intention is to capture an important, if stylized, set of characteristics of cognitive and social dynamics that
govern interpersonal communication in non-homophilic social settings. Among the most visible social changes in recent history is the shift in norms regarding the social acceptability of racist or bigoted public speech. This is a change that has permeated both elites and masses, as well as members of all social classes. Whereas in a previous era, norms of socially acceptable public speech dictated that whites in most settings should publicly espouse the superiority of whites over other races, by the last decade of the twentieth century this norm had completely reversed, and remains so today. In the current social and political climate, espousing any view other than complete racial equality is a dangerous endeavor that can lead to a permanently tarnished reputation. Regardless of what individuals may truly believe, expressions consistent with the ideas that one race is inherently superior to another, or that the traditional white hierarchy should be maintained, are now wholly unacceptable in almost all white social circles. It is clear that expressions in favor of racial inequality may occur and remain acceptable within self-selected groups of like-minded individuals, but in the types of interpersonal interactions that make up the bulk of day-to-day interpersonal communications for most individuals, they simply are not. The historical trajectory of this change is difficult to summarize in a concise manner, having been subject to wide geographic and temporal variability, but it can safely be claimed that the pace of change in the latter half of the twentieth century was particularly brisk. Coinciding with that brisk change was a monumental shift in public policies with respect to institutions that structure social interactions.

In this chapter, I have attempted to provide a formal account of some of the underlying individual and structural mechanisms that could have plausibly brought about such changes. The model presented above provides a simplified representation of the types of social interactions that I claim were largely responsible for structuring the witnessed changes in norms of rhetoric, that is, everyday interactions within organizational settings. The key feature of those interactions is that they are non-homophilic. When combined with individual preferences for conformity, non-homophilic interactions are expected to carry the potential for individuals forced to express an opinion on a sensitive social issue to potentially put forth a view that is dissonant with internal preferences. If such a phenomenon takes place on a wide scale, the general implication is that norms of public discourse can shift in such a way
that a substantial proportion of individuals that make up a given population consistently express opinions that differ from what they truly believe.

I attempted to capture these substantive dynamics by considering the notion of a dichotomous choice of behavior, where the dichotomy can be thought of substantively as the choice of exhibiting a behavior that is consistent with racial equality or consistent with racial inequality. In the model, agents are compelled to choose one behavior or the other, and they have a desire to both behave in a way that is consistent with their own internal values, but also to not appear deviant from their immediate social groups. Because individuals in the model act in sequence, and because networks of direct social connections were randomly assigned, the two desires often found themselves in tension. And this tension, I contend, is an ubiquitous feature of population-level social interactions. In other words, in the real world, when behavior with respect to some issue takes on a dichotomous-choice nature, yet underlying preferences are expected to be nuanced and variable, it is clear that widespread falsification of internal preferences may result. In the case of sensitive social issues, this is no doubt expected.

Many of the results of the numerical experiments were quite intuitive and predictable ahead of time. For instance, the results bore out the idea that as the weight that agents place on the utility from internal consistency increases, the preponderance of preference falsification decreases. Beyond a certain level, no simulated populations moved into equilibria in which all of the agents exhibited one behavior or the other. In other words, when agents placed greater value on exhibiting behavior consistent with internal preferences, there remained a substantial mix of behaviors in the population. Relating these results to the substantive situation of the change in rhetorical norms regarding race, if the model presented here holds any relation to the real world, it must be the case that the desire to not appear deviant is in fact quite strong. In other words, as discussed previously, the United States in the twenty-first century is a society in which the race question has been settled in the court of expressed public opinion. That is, to the extent that public behavior with respect to race can be boiled down to a dichotomous choice between a preference for equality and a preference for inequality, behavior is indeed skewed far to the side of equality. It may be the case that high values of $\alpha$ may be realistic for some individuals, but not for broad swaths of
the population at large.

Other results were not necessarily predictable ahead of time, but make sense upon further reflection. For instance, when equal weight is placed on intrinsic and extrinsic preferences (i.e., when \( \alpha = 0.5 \)), there is a general pattern in which the proportion of simulated populations that achieve equilibrium increases as the sizes of the groups to which they are assigned increases. Substantively, larger group sizes mean that a larger proportion of the overall population is in an intimate social setting together in which behaviors directly influence one another in any given current period. And the larger the public, the more uncertain are predictions about the behavior and preferences of others from which a given agent derives utility. Relating this back to the substantive example, the more proximate others who are able to witness an agent’s behavior, the more uncertain agents will tend to be in estimating the extent of preferences for racial (in)equality in the network of direct connections. In these instances, more individuals will tend to look to the previous behavior of the population at large in deciding what behavior to exhibit. From there, it is straightforward to surmise that this will more often lead to equilibrium behavior in one direction or the other, depending largely on the initial distribution of \( v_i \) in the population. In the simulations, both behaviors were given an equal chance to develop, on average and over time. But for the real-world example of racial rhetoric, it is likely the case that over time, social interactions outside of the organizational settings in which individuals experienced the bulk of their non-homophilic social interactions would have led the starting values in the population eventually to be biased toward the behavior representing racial equality.

Additionally, the average utilities of both the population at large as well as the falsifiers in the population suggest that, on average, agents gain more utility by being on either end of the internal consistency weight scale. In other words, high and low values of \( \alpha \) lead to greater average utility levels than mid-range values. As discussed previously, higher average utility levels are associated with higher proportions of steady-state behaviors. At both ends of the \( \alpha \) spectrum, relatively more equilibria are achieved across simulated populations, but the reasons differ. When agents place little weight on being internally consistent, steady behavioral states occur because they simply conform to the behaviors of first movers. When agents place a great deal of weight on being internally consistent, steady behavioral states
occur because like-minded individuals are randomly assigned to groups together at relatively high rates. In either case, agents receive a relatively high level of utility on average because they either do not care what others think, or they care, but they simply happen to get assigned to proximate others with similar values. Given the substantive situation of interest here, it is clear that taking averages and assigning the same value of $\alpha$ for all agents is an oversimplification. In the real world, it is highly likely that substantial proportions of the population exhibited high, mid-range, and low values for internal consistency. Those with high values are likely to have been the uncompromising first movers whose behaviors led to wider-scale change.

2.7 Conclusion

In the end, the model presented in this chapter is only a start as to capturing the most important processes underlying norm development, transition, and transmission. Much further work is required to bring together the appropriate set of micro-level cognitive elements and features of social interactions that will be able to generally describe the change in norms with respect to racial equality in the United States.
Scholarly thought on the interpersonal underpinnings of social influence processes is buttressed by compelling theoretical accounts and anecdotal evidence, but little in the way of systematic empirical demonstrations. At least one reason for the dearth of empirical research is that tapping the differences between true and expressed attitudes is exceedingly difficult in practice. In the pursuit of measuring internal attitudes, we are faced with external expressions of attitudes; in the pursuit of detecting social pressure effects, we are typically without observations of individuals embedded within social settings.

Take as one example the phenomenon of social desirability bias in the measurement of racial attitudes. If one examines, for instance, white attitudes toward blacks over large swaths of time in the United States, a clear trend will emerge in which expressed negative attitudes exhibit a marked decrease (see, e.g., Schuman et al. 1998). In any thoughtful account, interpretations of this trend will be accompanied by statements of caution about changing social mores that govern the social acceptability of expressions of negative racial attitudes. The standard, plausible, interpretation is that likely many respondents to any given poll or survey are masking their true attitudes and simply giving the researcher the perceived socially acceptable response. That is, the social customs of a particular time and place dictate a certain level of perceived social acceptability of negative racial attitudes of whites toward blacks, that level has decreased over time, and individuals whose true attitudes are discordant with that level face perceived pressure to mask their views.

That interpretation will not be criticized here. It is, in fact, highly plausible and quite compelling. But it is useful to note at least one phenomenon that it implies, and about which we know very little empirically. That is, whether or not the distribution of racial attitudes changed significantly over time, the mass public nonetheless internalized the change in norms
of socially acceptable public expression regarding race. But exactly how individuals came to perceive a changing social climate, as well as their own place in that change, remains empirically elusive.

Social desirability bias in the case of white racial attitudes is just one example of a general phenomenon of norm change and internalization. Similar transformations have been witnessed with respect to gender, various religious and ethnic groups, homosexuals, and so on. With respect to each of these cases, we can reasonably surmise that there is a general understanding among the populations of most institutionalized democracies that expressions of egalitarianism constitute the appropriate social response in most circumstances. That is, norms governing appropriate public discourse are such that in order to avoid appearing deviant from one’s peers, expressions of equality between groups are a safer bet than any justifications or preferences for inequality. This implies that individuals perceive that a belief in egalitarianism is the norm in the broader population. And the development of that perception must have come through a process of learning and internalization over time.

When internally held attitudes and norms of appropriate discourse are in tension, public expressions of attitudes may not be what they seem (e.g., Noelle-Neumann 1984). And therein lies the rub for empirically oriented social scientists. Those interested in obtaining valid measures of attitudes on sensitive social issues through self reports cannot be sure that respondents are telling the truth. And those interested in the real-world dynamics of the social influence processes themselves will have difficulty observing them because their primary manifestation is likely to be in the form of everyday social interactions. In either case, without a careful research design, these phenomena will continue to elude empirical examination.

The present chapter examines the dynamics of social pressure effects in a laboratory setting. For now, we will leave aside questions of whether the opinions that individuals express on sensitive social matters represent their true, internally held, attitudes, and rather focus attention on the dynamics of social influence. Specifically, examinations of social pressure effects will be with respect to baseline attitudinal measures on sensitive social matters collected in an isolated setting. Isolation is not expected to eliminate social desirability bias stemming from researcher effects, but it is expected to provide a contrast with which to
compare more direct forms of social pressure exerted in peer-group settings. The chapter proceeds as follows. In the next section, I present a summary of social scientific research on social influence. In the subsequent section, I use the ideas from previous research to draw a set of empirical hypotheses that can be tested in a laboratory setting. Following this discussion, I describe the designs and results of two experimental studies that allow for testing of the hypotheses. Finally, I offer concluding remarks.

### 3.1 Social Influence

Scholars have long been familiar with the idea that when it comes to salient social issues, individuals tend to be aware of whether there are norms of appropriate behavior that govern the range of thought deemed socially acceptable (e.g., Crowne and Marlowe 1960; Kuran 1995; Mendelberg 2001; Noelle-Neumann 1984). It has also been widely theorized that through interpersonal interactions, common knowledge of socially acceptable behavior can exert social pressure on individuals to act according to behavioral norms (Festinger 1954; Kelley 1952; Sherif 1936), even when such behavior is incongruent with preferences (Festinger 1957; Kuran 1989, 1991, 1995; MacKuen 1990). Everyday social interactions, therefore, act as a type of social verification system that drives individuals in the same society toward a shared understanding of the range of behaviors that are socially acceptable (Hardin and Higgins 1996), if not actual perceptions of right and wrong.

#### 3.1.1 General Notions of Social Influence

Individuals' social surroundings can substantially affect how they come to understand the political and social world (Hardin and Higgins 1996), and the mechanisms by which general social influence can occur vary greatly. One line of research that has received a great deal of attention by social scientists concerns the manner in which social context structures attitudes and behaviors. Beginning with the basic notion that individual behavior is developed through individual-environment interactions (e.g., Tingsen 1937), it is widely recognized that an important part of that environment is the social context. Over long spans of in-
dividuals’ lives, they will find themselves embedded within a set of institutions, networks, social groups, and so on, that for most practical analytical purposes can be considered static (Granovetter 1985). The major mechanism by which context imparts social influence is by limiting the number and structuring the types of opportunities for social interaction (Eulau 1986; Huckfeldt and Sprague 1987). If we can safely assume that a significant proportion of political talk that an individual engages in takes place in casual interactions between individuals who just happen to be proximate (MacKuen 1990; Verhulst and Levitan 2009), the impact of social context in the development of true attitudes, perceptions of the attitudes of others, and the social consequences of norm deviation would seem to be profound.

In addition to the immediate social context, individuals are also embedded within the broader society. Mass media, public opinion, and salient social groups operate beyond the bounds of individuals’ interpersonal environments. Exposure to mainstream media provides individuals with information about the lines being drawn on salient political topics, the range of acceptable points of elite debate on those topics, as well as measures and interpretations of broad public opinion. Further, individual perceptions about the aggregate distribution of preferences have been found to influence both the willingness to express an opinion (Noelle-Neumann 1974, 1984) and, when willing, the direction of preferences expressed (Mutz 1998).

The cumulative impact of these exposures should be to instill in individuals a sense of societal norms of appropriate social behavior. To actors within the current social climate of the United States, at least one social rule that should be apparent is the notion of equality between individuals, without respect to race, ethnicity, religion, gender, and so on. Regardless of one’s true preferences, it is clear that publicly espousing an inegalitarian view carries with it greater risk of social costs than espousing an egalitarian view.

3.1.2 Social Influence Through Interpersonal Interactions

The media provide information about society at large and the social context structures our exposure to others, but the bulk of social influence takes place through direct interpersonal interactions. Within political science, studies of interpersonal influence have tended to focus on the direct effects of information filtering by opinion leaders (Katz 1957; Katz and
Lazarsfeld 1955; Lazarsfeld, Berelson, and Gaudet 1948), and information exchanges between peers (e.g., Ahn, Huckfeldt, and Ryan 2010; Barabas 2004; Huckfeldt and Sprague 1987; Mutz 2002) in altering attitudes and behaviors. Political scientists have paid less attention to normative explanations, however (but see Verhulst and Levitan 2009). Without any exchange of factual information, people still tend to draw on the attitudes and behaviors of others as guideposts for the ranges of attitudes and behaviors that are socially appropriate (Festinger 1954; Kelley 1952). In certain situations, individuals may have a sense that there is truly a correct and incorrect attitude with respect to a given social referent, and that the attitudes of one’s peers serve as a benchmark against which to evaluate the appropriateness of their opinions (Burt 1987; Festinger 1954; Kelley 1952; Levitan and Visser 2008; Visser and Mirabile 2004). In certain other situations, proximate others may serve as standards for appropriate social behavior without regard to factual accuracy (Deutsch and Gerard 1955; Kelley 1952).

3.1.3 Social Conformity

The particular type of social influence most relevant for the present study is social conformity, that is, the tendency of individuals to publicly comply with a known or perceived group opinion. Outside of political science, social conformity has been a well studied phenomenon. Following Sherif (1935) and Asch (1951), a long line of research in social psychology and related fields has demonstrated that social conformity is in fact empirically verifiable in the laboratory. Sherif (1935, 1936) employed an optical illusion known as the autokinetic effect (Adams 1912), in which a stationary pinpoint of light is projected onto the wall of a darkened room, and to almost all human observers, the light appears to move. Employing confederates, Sherif demonstrated that experimental subjects were quite susceptible to the power of suggestion with respect to judgments about how far the light had traveled. That is, when subjects were placed in a group setting and asked to publicly express their estimation of how far the light had traveled after confederates had already done so, the judgments of subjects tended to fall in line with those of the confederates, even when the expressed judgments of confederates were wildly different than baselines established in a control condition.
The Sherif studies demonstrated that when humans are asked to express judgments about relatively ambiguous phenomena, there is a strong tendency to look to the majority opinion as a rule of thumb for the appropriate response.

But what about relatively unambiguous phenomena? Working in the tradition of Sherif (1935, 1936), Asch (1951, 1952) examined social conformity in situations in which the judgment task had clear right and wrong answers. Specifically, subjects in the Asch studies were asked to express which among a set of drawn lines was identical in length to another drawn line. Line lengths were purposely drawn so that subjects would easily be able to differentiate their lengths, and subjects making the judgment in an isolated control condition rarely expressed the wrong answer. Again, subjects were placed in a group setting and asked to express their judgment out loud, this time following a group of confederates who expressed the wrong answer. Though a majority of subjects gave the correct response in this group condition, a surprisingly large proportion of subjects gave the same wrong answer as the group of confederates. Given little evidence that the subjects examined by Asch (1951, 1952) were unsure about the correct answer, the results point to a strong tendency for humans in public settings to conform to expectations about the appropriate social response.

While the results from the research programs spawned by Asch and Sherif are extremely useful as the largest and most prominent collection of empirical evidence on small-group conformity, it is not clear whether the type of conformity demonstrated has a direct connection to norms of public speech with respect to salient social referents (including attitudes about the equality of salient social groups), which are the focus of the present study. In particular, the typical decision-making task under examination in these studies is highly arbitrary, and cannot reasonably be expected to carry any social significance for the subject outside of the laboratory. In one way, this can be seen as a strength of the experimental designs in that the researcher is able to eliminate subjects’ preconceived notions about particular referents. In another way, however, if we consider those preconceptions to be worthy of study in their own right, and if we consider the clear qualitative difference between expressed judgments on arbitrary tasks and expressions of opinion on salient—and sometimes sensitive—attitudinal measures, existing work on social conformity leaves open the question of how small-group pressures might operate on norms of public speech regarding egalitarianism.


3.2 Development of Hypotheses

The lines of research spawned by Asch and Sherif offer a jumping-off point from which to develop a set of expectations about the impact of social pressure on individual expressions of political attitudes. As discussed previously, the takeaway point from these lines of research is that in ambiguous and unambiguous decision-making settings, experimental subjects exhibit a strong tendency to look to the group majority as an indicator of the appropriate social response. Individuals may not internalize the perceived group opinion as their own, but avoidance of appearing deviant provides a strong incentive to publicly comply, nonetheless.

Again, however, though researchers in social psychology and related fields have tended to couch the phenomena that they examine in terms of general descriptions of social norms, decision-making, opinion formation, and so on, it is not completely clear that the types of behavioral tasks that they examine provide a close analogue to expressed attitudes on sensitive social issues, which are the focus of the present study. On the one hand, attitudes, unlike lines drawn on index cards, are by their nature matters of opinion without a clear right and wrong answer. If individuals, no matter how strongly they hold their own opinions, are able to see that certain political questions are coupled with a range of potential responses across which reasonable people might disagree, then the impact of social pressure to conform to some perceived group opinion may be less than in unambiguous decision-making settings. On the other hand, the lack of a clearly correct answer, as with the autokinetic effect, may lead individuals to be even more susceptible to the power of suggestion by immediate peers. Because previous research does not offer clear guidance in this respect, I will begin the process of investigation by situating the expression of political attitudes on sensitive social issues within the general framework of decision-making tasks examined by previous research on social conformity. This decision leads to the following:

**General Social Conformity Hypothesis 1:** Individuals who are able to learn the distribution of attitudes in their immediate peer group will exhibit a tendency to publicly comply with the group opinion.

First note what *General Social Conformity Hypothesis 1* is not saying. Specifically, there is no inherent assumption that individuals who express attitudes in the absence of specific
knowledge about the opinions of their peers are free from all social pressure. It is in fact expected that individuals carry with them some notion of social expectations, and that being forced to express an attitude to a researcher at all is often enough to invoke conformity pressures. Knowledge of the distribution of attitudes among a proximate group of peers, however, is expected to be stronger, and to manifest itself through dissipation or reversal of biases exhibited in isolated response settings. That is, if a clear picture of the distribution of expressed attitudes in the population can be established among individuals compelled to express their attitudes in isolation, General Social Conformity Hypothesis 1 states that groups of individuals can predictably be lured away from that population distribution if they are compelled to give their attitudes publicly in the face of a group of peers whose expressed attitudes are markedly different than those of the population.

While General Social Conformity Hypothesis 1 offers a prediction about what to expect when comparing groups of individuals who express attitudes in isolation to groups of individuals who express attitudes in public, it does not make a claim about individuals changing the responses that they would otherwise give in another circumstance. Even if the expectation is supported, we do not know whether specific individuals in the group of people compelled to give their responses publicly would have given a different response if they happened to have given their response in isolation. But if we continue to think about the expression of political attitudes on sensitive social issues as just one example of the types of decision-making tasks and norms of behavior examined by Asch and Sherif and their followers, it would seem to be implied that individuals could be drawn away from their own true belief in the face of social pressure to provide a different response. That is, if the same individual were compelled to provide an attitude expression both in isolation and in the face of a proximate and opposed group of peers, the group responses may provide a form of social pressure that leads individuals to alter their responses between the public and private settings. This discussion leads to the following:

*General Social Conformity Hypothesis 2: When asked to first express an attitude*
privately, and then publicly in the face of a unanimously opposed majority, individuals will tend to alter their privately expressed attitude in order to publicly comply with the unanimous group attitude.

We can imagine a situation in which individuals are compelled to give their opinion on a social issue, and where they have some sense that this opinion will be kept private. Then later, those same individuals are compelled to express their opinions on the same issue in a group situation in which (1) there is no sense of privacy, and (2) the fellow group members have unanimously given an opinion that is opposed to what the individual stated in private. In this situation, General Social Conformity Hypothesis 2 predicts that the subject will change the attitude that they previously expressed in order to conform to the group opinion.

General Social Conformity Hypothesis 1 and General Social Conformity Hypothesis 2 are drawn directly from previous social conformity research. But as stated above, it is not clear that political attitudes operate in the same way as arbitrary decision-making tasks. As has been discussed, the inherently subjective nature of attitudes means that most people will understand that there is no clear right or wrong answer. On most sensitive social issues, however, there is a sense of what is and what is not appropriate to say publicly. So while individuals could genuinely feel that reasonable people might disagree, they also carry the looming feeling that publicly appearing deviant from the dominant opinion in the population may carry with it social costs.

It has been argued convincingly elsewhere that when people hold conflicting cognitions, they experience cognitive dissonance, that is, a sense of psychological discomfort that leads them to try to reduce the conflict (Festinger 1957). Analogous to the idea of conflicting cognitions, individuals may also experience psychological discomfort from holding internal attitudes that are in conflict with the known appropriate social response. Borrowing the terminology of cognitive dissonance theory, I offer the following hypothesis:

The Social Dissonance Hypothesis: When asked to first express an attitude privately, and then publicly in the face of a unanimously opposed majority, individuals will be more likely to alter their privately expressed attitude in order to publicly
comply with the unanimous group attitude if the attitude that they expressed in private runs counter to norms of social acceptability.

The Social Dissonance Hypothesis is not opposed to the General Social Conformity Hypothesis 1 or General Social Conformity Hypothesis 2; rather, the Social Dissonance Hypothesis simply predicts a more nuanced set of circumstances leading to conformity than the latter two. Revisiting the hypothetical scenario put forth above in which an individual is compelled to express an attitude privately and then publicly in the face of unanimous opposition, imagine further that the particular attitude being expressed is on a sensitive social issue, and that individuals have an accurate sense of the range of attitudes that are deemed socially appropriate with respect to that issue. The Social Dissonance Hypothesis predicts that if the individual expressed an attitude in private that falls outside of the range of social acceptability, she will be under more pressure to conform to the unanimously opposed majority than if she expressed an attitude that falls in line with what is deemed socially appropriate. Assuming that individuals in the United States recognize attitudes expressing inegalitarian viewpoints as generally socially unacceptable, those who express an inegalitarian opinion in private are more likely than those who express an egalitarian opinion in private to submit to interpersonal pressures to express the opposite in public, according to the Social Dissonance Hypothesis.

Even if we do find that under either general or limited circumstances, individuals have some systematic tendency to conform to the opposed opinion of a group of peers, public compliance in and of itself does not indicate whether the change in responses is due to internalization of a new attitude (e.g., through learning or updating) or preference falsification to avoid appearing deviant from the group. On the one hand, Sherif (1936) and Verhulst and Levitan (2009) demonstrated that even brief interactions can lead to continued influence weeks, or even months, later. Also, cognitive dissonance theory would predict that the very act of providing a particular response publicly would lead an individual to be more likely to internalize that attitude (Festinger 1957). Based on these considerations I propose the
The Attitude Internalization Hypothesis: When asked to first express an attitude privately, then publicly in the face of a unanimously opposed majority, and then again privately, individuals who altered their responses between the first private setting and the group setting will maintain the attitude expressed in the group setting in the second private setting.

However, and as has been discussed, it is also plausible to think that expressions of attitudes on sensitive social issues do not operate in exactly the same manner. For instance, in a large body of work, Kuran (1987, 1989, 1990, 1991, 1995), has used formal logic, anecdotal accounts, and qualitative analysis of political events to argue convincingly that individual public compliance with a belief that is incongruent with internal attitudes is a theoretically plausible phenomenon. At the individual level, one can think of agents earning a type of expressive benefit by stating publicly their true preferences with respect to social referents and incurring a type of expressive cost by stating publicly a view that goes against the majority opinion (Kuran 1995). The primary mechanisms leading to these costs and benefits are negative and positive reinforcement by other group members. Under very general circumstances, it can be shown that norms of public compliance that run counter to some or all of a group’s members’ internal attitudes can persist if the costs of going against the group are sufficiently high. Based on this discussion, I offer the following:

The Preference Falsification Hypothesis: When asked to first express an attitude privately, then publicly in the face of a unanimously opposed majority, and then again privately, individuals who altered their responses between the first private setting and the group setting will alter them again in the second private setting in the direction of the original response.

Predictions about whether any attitude changes witnessed as a result of social pressure are real attitude changes or artificial public compliance are expressed as competing hypotheses because the previous literature offers competing expectations.
3.3 The Social Pressure Experiments

I now present a set of experiments that were employed to test the hypotheses put forth in the previous section. The studies took place from September 2011 through April 2012 and used as subjects undergraduates from the University of Illinois at Urbana-Champaign. The design objective was to situate each study within the tradition of previous research on social conformity from social psychology and related fields, but to push the envelope into territory not yet explored by previous researchers—that is, public expression of political attitudes as it relates to egalitarian norms (for the only known example of a similar design with respect to political attitudes more generally, see Verhulst and Levitan 2009). In one instance using a between-subjects design, and in another instance using a within-subjects design, either a control group or a pre-treatment measurement of treated subjects was used to establish a baseline measurement of the types of responses to potentially sensitive social issues that subjects were willing to give in isolation. These were then used in fictitious group settings to construct treatments intended to generate social pressure in the opposite direction of the baseline measurements.

3.3.1 The Behavioral Task

To relate the experimental designs to the preceding discussion, a behavioral task had to be chosen that would tap subjects’ levels of adherence to egalitarian norms. In the United States, as in many other multicultural societies, it is without controversy to claim that in most public social settings, norms of appropriate social behavior dictate compliance with a norm of individual and group equality, without respect to race, ethnicity, religion, gender, and so forth. That is, individuals’ true attitudes with respect to, say, religious equality, are expected to vary—sometimes greatly—in reality, but the range of appropriate public expression of religious attitudes lies wholly on the equality side of the scale. With respect to specific identifiable groups, however, tolerance of dissent from the norm of equality can
vary at any given time given contemporaneous social circumstances. Therefore, in order to examine differences between subjects who are willing to express an inegalitarian attitude and those who are not, a target group had to be chosen that would elicit a relatively high level of variation in subject responses. Based on a variety of polling data, as well as a current social climate in which the threat of violence by Islamist groups looms large in the Western public conscience, expressed attitudes about Muslims and Islam by non-Muslims was chosen as the behavioral task of interest for examining social pressure.\(^1\)

Table 3.1 presents the questionnaire items that will be used as behavioral measures to test the various hypotheses put forth above. For each question, a determination was made as to what constituted the egalitarian and inegalitarian responses, and these decisions are indicated in the latter columns of the table. The number of response options available to subjects on the questionnaire is also indicated. Further, each question is given an identification number (first column). In much of the discussion that follows, it will be useful and efficient to refer to questionnaire items by their identification number rather than their full question wording.

### 3.3.2 Participants

Subjects were recruited from undergraduate political science courses at the University of Illinois at Urbana-Champaign in exchange for extra credit. Subjects were recruited under the auspices of taking part in a study on “group and individual decision making.” Summary statistics for the participants included in the analyses presented here are presented in Appendix B.

\(^1\)The incentive for subjects to participate in the study was extra course credit for one of their political science courses, and therefore, in the name of fairness, all students in a given course were offered the opportunity to participate. That is, Muslim students were not excluded from the recruitment pool. However, a survey question asked students for their religious affiliations. In all analyses that follow, self-identified Muslims are excluded.
Table 3.1: Summary of Questions About Muslims and Islam Used to Examine the Effect of Social Pressure

<table>
<thead>
<tr>
<th>ID</th>
<th>Question Wording</th>
<th>Response Scale</th>
<th>Egalitarian Response</th>
<th>Inegalitarian Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Would you be nervous if a Muslim man were on the same flight as you?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q2</td>
<td>Would you mind living in a neighborhood where half of your neighbors are Muslim?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q3</td>
<td>Do you favor special security checks at airports for Muslims?</td>
<td>4-point</td>
<td>Str. oppose/ Oppose</td>
<td>Str. favor/ Favor</td>
</tr>
<tr>
<td>Q4</td>
<td>Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?</td>
<td>4-point</td>
<td>A little/ Not at all</td>
<td>Mod. amount/ A lot</td>
</tr>
<tr>
<td>Q5</td>
<td>In your best estimation, what proportion of Muslims in the world are anti-American?</td>
<td>3-point</td>
<td>Some or Just a few</td>
<td>Most or Almost all/About half</td>
</tr>
<tr>
<td>Q6</td>
<td>Do people have the right to keep Muslim organizations out of their neighborhoods?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q7</td>
<td>Would you mind if you yourself or your children were to have a Muslim teacher at school?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q8</td>
<td>Do you think that most, many, only some, or very few Muslims living in the United States today support the use of violence by Muslims against non-Muslims?</td>
<td>4-point</td>
<td>Only Some/ Very Few</td>
<td>Most/ Many</td>
</tr>
<tr>
<td>Q9</td>
<td>How worried are you, if at all, about radicals within the U.S. Muslim community?</td>
<td>3-point</td>
<td>Not worried at all</td>
<td>Very worried/ A little worried</td>
</tr>
<tr>
<td>Q10</td>
<td>Would you mind if your children were to attend a predominantly Muslim school?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q11</td>
<td>In general, do you think that Islam is more likely than other religions to encourage violence among its believers?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 3.1: Summary of Questions About Muslims and Islam Used to Examine the Effect of Social Pressure (continued from previous page)

<table>
<thead>
<tr>
<th>ID</th>
<th>Question Wording</th>
<th>Response Scale</th>
<th>Egalitarian Response</th>
<th>Inegalitarian Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q12</td>
<td>Do you think that most, many, only some, or very few Muslims living in the United States today are peaceable and do not condone violence?</td>
<td>4-point</td>
<td>Most/Many</td>
<td>Only Some/Very Few</td>
</tr>
<tr>
<td>Q13</td>
<td>Would you mind having a Muslim supervisor at work?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q14</td>
<td>Would you be nervous if a Muslim woman were on the same flight as you?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q15</td>
<td>Would you mind if a close relative who is male married a Muslim?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q16</td>
<td>Would you mind if a close relative who is female married a Muslim?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q17</td>
<td>Do you think that most, many, only some, or very few Muslims living in the United States today support the goals of Al Qaeda and the Islamic fundamentalists?</td>
<td>4-point</td>
<td>Only Some/Very Few</td>
<td>Most/Many</td>
</tr>
<tr>
<td>Q18</td>
<td>Do you favor laws that prohibit discrimination against Muslims?</td>
<td>4-point</td>
<td>Str. favor/Favor</td>
<td>Str. oppose/Oppose</td>
</tr>
<tr>
<td>Q19</td>
<td>Based on the events of September 11th, 2001, would you favor or oppose the U.S. government tightening immigration laws to restrict the number of immigrants from Arab or Muslim countries entering the United States?</td>
<td>4-point</td>
<td>Str. favor/Favor</td>
<td>Str. oppose/Oppose</td>
</tr>
<tr>
<td>Q20</td>
<td>Should someone who says that terrorist attacks against America are the fault of how America behaves in the world be allowed to teach in a public school?</td>
<td>2-point</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note: ID numbers are used to refer to particular questions in the text. Combined response options in the egalitarian and inegalitarian categories are separated by slashes.
3.3.3 The Looming Social Pressure Experiment

In the first study, a control group of subjects came to the laboratory and answered a questionnaire that tapped, among other items, attitudes about Islam and Muslims. These control measures were then used to construct a treatment intended to generate social pressure in the opposite direction of that established in the baseline setting (this is similar to the method used by Sherif 1935, 1936; see also Jacobs and Campbell 1961). This social pressure treatment was administered via a contrived group setting in which subjects in the treatment group were provided information about the distribution of opinion among a group of other participants whom they would meet in a face-to-face setting at a later time.

Procedure for Control Group

Up to 20 subjects at a time gathered in a waiting room adjoined to a set of individual, private, enclosed computer terminals. Subjects arrived individually, typically over a period of 10 to 20 minutes. After filling out preliminary paperwork, subjects were told that they would be participants in a broader research study on group and individual decision making, and that today, they would simply take part in an individual survey. Subjects were then assigned to a computer terminal, and followed a set of instructions as the software led them through a questionnaire.

To simply gather a distribution of responses to the questionnaire, it was not strictly necessary to have subjects gather together in a common waiting room or to go through the computer-based survey in an enclosed, private computer terminal. But these steps were taken to attempt to eliminate any confounding effects of the experimental context itself. Given that the primary measures of interest in the study are subject responses to questions tapping sensitive social issues, it seems reasonable to expect that certain features of the laboratory setting may come to bear on the on the extent to which individuals feel confident about the anonymity of their responses. Specifically, it is likely that the social nature of
the way the subjects gather and the privacy of the physical setting in which they answer
the questionnaire could each impact the degree of general social desirability bias exhibited
in the survey responses. The goal of the control setting was not to eliminate biases in
responses stemming from social desirability effects, but rather to ensure that such biases can
be excluded as a competing explanation for any differences witnessed between the treatment
and control groups.

Procedure for Treatment Group

Overview After being given information about the supposed distribution of attitudes on
several public opinion items among a subset of other study participants, subjects answered
a computer-based questionnaire tapping those same attitudinal measures, among others. A
ruse was employed to convince subjects that their own responses to the questionnaire would
eventually be made public to that same group of other participants, following which time
they would engage in a face-to-face task with them. Because the manipulation relied on
a deception, following the questionnaire subjects were queried as to whether the deception
was convincing.

Preliminaries Up to 10 subjects at a time gathered in the same waiting room that was
used in the control condition. Again, subjects arrived individually, typically over a period of
10 to 20 minutes. As subjects filled out preliminary paperwork, the researcher informed them
that another group of subjects was taking part in a study in a nearby room, and that the
progress of that group would have to be monitored. This other group of participants did not
in fact exist, but to complete the ruse, the researcher periodically left the room to supposedly
check on them. Specifically, the researcher would apologize and leave the room for a period
of one to two minutes. Upon returning to the room, the researcher would go to a computer
in the waiting room and look through a spreadsheet full of seemingly random numbers in full
view of the participants. The purpose of each of these steps was to make the deception more
convincing. This process was then repeated every several minutes until the agreed-upon start time of the study had passed and each subject had experienced the researcher leaving the room and returning at least once. At that point, the researcher addressed the entire group of subjects, instructing them to file into their assigned computer terminals, close the door, and await further individual instructions while the researcher confirmed the status of the supposed other group’s progress. Specific instructions are provided in Appendix A.

**Individual Instructions 1** As the subjects filed into their assigned computer terminals and closed the door, the researcher left the room for a period of one to two minutes to supposedly check on the other group of participants. Upon returning, the researcher walked into each computer terminal and gave a set of supposedly individualized instructions. In fact, however, the instructions recited to each subject were identical. Specifically, the researcher walked into each computer terminal carrying a file folder containing four sheets of paper with printed material on them. The researcher informed subjects that part of the purpose of the study was to analyze group decision-making situations when people are allowed to learn about their fellow group members prior to engaging in a face-to-face task. To that end, the subjects were told that they were assigned to a group composed of four people from the other room, and that the printed material in the file folder represented the responses of these four people to six randomly chosen questions from a questionnaire that they had answered. To eliminate possible suspicion about the deception at this point in the study, the researcher informed subjects that the participants today were assigned two different roles, and that the difference between the people who gathered with the subject and the people in the other room was that the latter did not know about the group component prior to answering the questionnaire, and that this difference would be used to test something. The researcher instructed subjects on how to read the printouts of others’ responses, and asked them to look them over for a few minutes and await further instructions. The printed materials given to subjects are presented in Figure 3.1 and the specific instructions recited are presented in
Appendix A. The researcher then left the private computer terminal, closed the door, and in succession gave the same set of instructions to the remaining subjects.

**Individual Instructions 2** After each subject went through the procedure described above, the researcher returned and informed them that they would now go through the same questionnaire answered by their fellow group members. They were also informed that following the questionnaire, they would have a brief break before the group decision-making task, at which time the researcher would make a similar printout about them and present it to their fellow group members before the initial face-to-face meeting. The group decision-making task, they were told, would be very brief, and then their participation in the study would be over. Then at their individual computer terminals, subjects answered the questionnaire, received a debriefing about the deception employed in the study, and then answered questions about whether the deception was successful.

**Results**

Comparing the printed materials in Figure 3.1 to Table 3.1, we see that four of the six questions of interest were used to generate the social pressure treatment. For the question of whether subjects would be nervous if a Muslim man were on the same flight as them (Q1), the supposed distribution of opinion among the four group members was unanimously “Yes”; for the question of whether subjects would worry about having a Muslim as a neighbor (Q4), the distribution was two answering “A lot” and two answering “A moderate amount”; for the question of what proportion of Muslims in the world are anti-American (Q5), the distribution was one answering “Most/Almost all” and three answering “About half”; and for the question of whether people have the right to keep Muslim organizations out of their neighborhoods (Q6), the distribution was the distribution was unanimously “Yes.” All of the responses of future group members, therefore, fall on the inegalitarian side of the scale (see Table 3.1). In the discussion that follows, these questions will be referred to as direct
| Do people have the right to keep Muslim organizations out of their neighborhoods? | Yes | No | X |
| Do you approve or disapprove of the way that President Obama is handling the situation in Afghanistan? | Strongly Approve | Strongly Disapprove | X |
| Do you favor or oppose the idea that homosexuals should be allowed to serve openly in the United States military? | Strongly Favor | Strongly Oppose | X |
| Would you be nervous if a Muslim man were on the same flight as you? | Yes | No | X |
| Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor? | A lot | A moderate amount | A little | Not at all | X |
| In your best estimation, what proportion of Muslims in the world are anti-American? | Most/Almost All | Some/Just a Few | X |

---

| Do you have the right to keep Muslim organizations out of their neighborhoods? | Yes | No | X |
| Do you approve or disapprove of the way that President Obama is handling the situation in Afghanistan? | Strongly Approve | Strongly Disapprove | X |
| Do you favor or oppose the idea that homosexuals should be allowed to serve openly in the United States military? | Strongly Favor | Strongly Oppose | X |
| Would you be nervous if a Muslim man were on the same flight as you? | Yes | No | X |
| Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor? | A lot | A moderate amount | A little | Not at all | X |
| In your best estimation, what proportion of Muslims in the world are anti-American? | Most/Almost All | Some/Just a Few | X |

---

Figure continued on next page.
### SURVEY RESULTS SEPTEMBER 28, 2011 SESSION 2 PARTICIPANT # 6

<table>
<thead>
<tr>
<th>Question</th>
<th>Participant # 6</th>
<th>Participant # 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do people have the right to keep Muslim organizations out of their neighborhoods?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Do you approve or disapprove of the way that President Obama is handling the situation in Afghanistan?</td>
<td>Strongly Approve</td>
<td>X</td>
</tr>
<tr>
<td>Do you favor or oppose the idea that homosexuals should be allowed to serve openly in the United States military?</td>
<td>Strongly Favor</td>
<td>X</td>
</tr>
<tr>
<td>Would you be nervous if a Muslim man were on the same flight as you?</td>
<td>Yes</td>
<td>X</td>
</tr>
<tr>
<td>Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?</td>
<td>A lot</td>
<td>A little Not at all</td>
</tr>
<tr>
<td>In your best estimation, what proportion of Muslims in the world are anti-American?</td>
<td>Most/Aprmost</td>
<td>Some/Just a Few</td>
</tr>
</tbody>
</table>

Note: Each panel is one page of a printout given to experimental subjects that supposedly represented the distribution of opinion among fellow group members across six randomly chosen questions. Materials given to subjects were identical across and within study days, with the exception that the date on the top of each page of the printout always reflected the current date. Materials were contained in a file folder with “# 11” written on the outside (see instructions in Appendix A for details).
questions, given that they are the questions for which direct social pressure was exerted. The remaining questions (Q2-Q3 and Q7-Q18) will be referred to as indirect questions.

To get an initial sense of the distribution of responses in the treatment and control groups, Figure 3.2 presents barplots of the questions drawn directly from the materials given to subjects. For each question with more than two response options, the egalitarian and inegalitarian response categories are each collapsed. The first point to note is that the deliberate construction of a treatment to exert social pressure in the opposite direction of the distribution of opinion established in a control setting necessarily dictated that the treatment distribution would be skewed toward inegalitarianism. Comparing the heights of the dark bars within each panel, it is clear that subjects in isolation show an overwhelming tendency to choose the egalitarian response option on these particular questions. Given the large body of research on Americans’ racial attitudes, this result is not at all surprising. For question Q4, the difference between the set of control subjects who gave the egalitarian and inegalitarian responses was about 93 percentage points.

The second thing to notice is that the percentage point difference between the subset of subjects giving the egalitarian and inegalitarian responses was dissipated across the board in the treatment group. For each question, when informed that the attitudes of their fellow group members were unanimously on the inegalitarian side of the response scale, and under the expectation that their own responses would be made public to a group of peers, the percentage of subjects willing to give the inegalitarian response increased by about 18 percentage points for Q1, by about 1 percentage point for Q4, by about 13 percentage points for Q5, and by about 0.04 percentage points for Q6. The directional effects clearly exhibit initial support for General Social Conformity Hypothesis 1, but we cannot be confident yet that these differences are significant.

Focusing just on these direct questions, Table 3.2 presents $\chi^2$ difference of proportion tests for the proportion of subjects giving the inegalitarian response between the treatment and control groups. Also presented are Fisher’s exact tests of independence for the contingency
Figure 3.2: Distribution of Egalitarian and Inegalitarian Responses in the Looming Social Pressure Experiment for Questions Drawn Directly From Materials

Q1: Would you be nervous if a Muslim man were on the same flight as you?

Q4: Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?

Q5: In your best estimation, what proportion of Muslims in the world are anti-American?

Q6: Do people have the right to keep Muslim organizations out of their neighborhoods?

Note: For response choices identified as egalitarian and inegalitarian, refer to Table 3.1. Combined response options in the egalitarian and inegalitarian categories are separated by slashes. Bar heights are percentages within categories. Numbers above bars are the actual number of respondents in each category.
Table 3.2: Treatment-Control Differences in the Looming Social Pressure Experiment for Questions Drawn Directly From Materials

<table>
<thead>
<tr>
<th>ID</th>
<th>Inegalitarian Control Proportion</th>
<th>Inegalitarian Treatment Proportion</th>
<th>$\chi^2(1)$</th>
<th>Asymptotic p-value</th>
<th>Fisher’s Exact p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.182</td>
<td>0.358</td>
<td>5.893</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>Q4</td>
<td>0.034</td>
<td>0.113</td>
<td>3.347</td>
<td>0.034</td>
<td>0.039</td>
</tr>
<tr>
<td>Q5</td>
<td>0.324</td>
<td>0.453</td>
<td>2.272</td>
<td>0.066</td>
<td>0.067</td>
</tr>
<tr>
<td>Q6</td>
<td>0.151</td>
<td>0.189</td>
<td>0.182</td>
<td>0.335</td>
<td>0.328</td>
</tr>
</tbody>
</table>

Note: Second and third columns are the proportions of respondents in the control and treatment conditions, respectively, that gave the response identified as inegalitarian (see Table 3.1). Null hypothesis for Fisher’s exact test is that the number of people who gave the inegalitarian response is unrelated to whether they were in the treatment or control group. Alternative hypothesis is that the number of people in the treatment group who gave the inegalitarian response is greater than the number in the control group. Asymptotic $p$-value is based on the $\chi^2(1)$ approximation of the hypergeometric null distribution (Fisher 1922) using Yates’ continuity correction (Yates 1934).

table in which subjects are classified according to whether they were in the treatment or control group and whether they gave the inegalitarian response. For both tests, the null hypothesis of no relationship was put in contrast to the alternative that the proportion or number in the treatment group was greater than the proportion or number in the control group. The $\chi^2 p$-values and the exact $p$-values match quite closely, and they tell a story that comports with the illustration in in Figure 3.2.

The two questions that seem to elicit the largest social pressure effect tap a general sense of concern among the subjects about being in close proximity to Muslims. Upon reflection, this sense of concern forms a qualitative similarity between Q1 and Q4 that also provides a qualitative difference with the other questions. When asked about the level of nervousness or worry about being on a flight with a Muslim or having a Muslim as a neighbor, subjects are more susceptible to the power of suggestion from their peers than with respect to other types of questions. For both of these questions, the difference in proportion of subjects giving the inegalitarian response between the treatment and control groups achieves a conventional level of statistical significance, and the 17 percentage point difference is particularly pronounced in the case of Q1. At least one plausible interpretation of these results is that a large proportion of non-Muslims carry some positive level of concern about being in close proximity to Muslims, but they typically allow that concern to remain latent if asked. Upon learning
that a group of non-Muslims peers are willing to express openly their concern in these situations, those individuals receive a cue that it is permitted, if not necessary, to allow those concerns to come out of latency to the forefront. Notice that this interpretation is not stating that individuals are being compelled to express an attitude that they do not believe. Rather, it is simply stating that in the absence of a defined group opinion, the overwhelming tendency is for individuals to keep those concerns on the backburner. And it is reasonable to believe that at least one reason for this is the perception of an appropriate social response.

Q5, tapping subjects’ perceptions of the pervasiveness of Muslim anti-Americanism, is qualitatively different from Q1 and Q4, but the result is similar. The difference in proportion giving the inegalitarian response between treatment and control groups achieves marginal statistical significance in this case. And because this question is closer to a statement of factual knowledge than questions tapping personal concern, it is difficult to provide the same interpretation. To the extent that individuals encounter this question having not considered it before, it is perhaps the case that the unanimously inegalitarian responses of fellow group members provide information about the correct answer to this question. To the extent that individuals do have well formed attitudes with respect to this question, movement away from these attitudes must necessarily represent public compliance with views that they do not truly hold. But because it is likely that on-the-fly opinion formation and alteration of true attitudes are both occurring, teasing out the degree of simple learning and the degree of preference falsification is simply not possible with the data at hand. What we can say is that there is strong evidence of a public compliance effect, regardless of the underlying mechanism producing it.

For Q6, the differences in proportions giving the inegalitarian response between the treatment and control groups does not come close to achieving a conventional level of statistical significance. Here the reason behind the result could plausibly reside in the fact that this particular question may be tapping less of a personal attitude about Muslims and more of a personal attitude about the civil rights of minority groups in general. Regardless of individ-
Table 3.3: Treatment-Control Differences in the Looming Social Pressure Experiment for Questions Not Drawn Directly From Materials

<table>
<thead>
<tr>
<th>ID</th>
<th>Control Inegalitarian Proportion</th>
<th>Treatment Inegalitarian Proportion</th>
<th>(\chi^2(1))</th>
<th>Asymptotic p-value</th>
<th>Fisher’s Exact p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>0.426</td>
<td>0.453</td>
<td>0.033</td>
<td>0.428</td>
<td>0.427</td>
</tr>
<tr>
<td>Q3</td>
<td>0.223</td>
<td>0.264</td>
<td>0.175</td>
<td>0.338</td>
<td>0.333</td>
</tr>
<tr>
<td>Q7</td>
<td>0.082</td>
<td>0.113</td>
<td>0.167</td>
<td>0.341</td>
<td>0.584</td>
</tr>
<tr>
<td>Q8</td>
<td>0.061</td>
<td>0.152</td>
<td>2.758</td>
<td>0.048</td>
<td>0.054</td>
</tr>
<tr>
<td>Q9</td>
<td>0.709</td>
<td>0.830</td>
<td>2.370</td>
<td>0.062</td>
<td>0.059</td>
</tr>
<tr>
<td>Q10</td>
<td>0.581</td>
<td>0.623</td>
<td>0.134</td>
<td>0.357</td>
<td>0.359</td>
</tr>
<tr>
<td>Q11</td>
<td>0.333</td>
<td>0.434</td>
<td>1.303</td>
<td>0.127</td>
<td>0.127</td>
</tr>
<tr>
<td>Q12</td>
<td>0.054</td>
<td>0.019</td>
<td>0.468</td>
<td>0.247</td>
<td>0.259</td>
</tr>
<tr>
<td>Q13</td>
<td>0.095</td>
<td>0.057</td>
<td>0.320</td>
<td>0.714</td>
<td>0.876</td>
</tr>
<tr>
<td>Q14</td>
<td>0.047</td>
<td>0.075</td>
<td>0.178</td>
<td>0.337</td>
<td>0.322</td>
</tr>
<tr>
<td>Q15</td>
<td>0.182</td>
<td>0.154</td>
<td>0.065</td>
<td>0.600</td>
<td>0.747</td>
</tr>
<tr>
<td>Q16</td>
<td>0.320</td>
<td>0.358</td>
<td>0.118</td>
<td>0.365</td>
<td>0.362</td>
</tr>
<tr>
<td>Q17</td>
<td>0.047</td>
<td>0.038</td>
<td>0.000</td>
<td>0.500</td>
<td>0.607</td>
</tr>
<tr>
<td>Q18</td>
<td>0.524</td>
<td>0.415</td>
<td>1.433</td>
<td>0.884</td>
<td>0.546</td>
</tr>
</tbody>
</table>

Note: Second and third columns are the proportions of respondents in the control and treatment conditions, respectively, that gave the response identified as inegalitarian (see Table 3.1). Null hypothesis is that the number of people who give the inegalitarian response is unrelated to whether they were in the treatment or control group. Alternative hypothesis is that the number of people in the treatment group who give the inegalitarian response is greater than the number in the control group. Asymptotic p-value is based on the \(\chi^2(1)\) approximation of the hypergeometric null distribution (Fisher 1922) using Yates’ continuity correction (Yates 1934).

As a point of comparison to the questions for which direct social pressure was exerted in the treatment, Table 3.3 presents \(\chi^2\) difference of proportion tests and Fisher’s exact tests for the indirect questions, that is, the remaining questions on the survey tapping attitudes about Islam and Muslims for which no information about the distribution of peer responses was given. And the big takeaway point from this table is that very little in the way of differences between treatment and control groups is exhibited. Differences between treatment and control achieve marginal statistical significance for Q8 and Q9, which tap individuals’ personal views of Muslims and Islam, the overwhelming majority is likely to recognize the rights of all groups to operate within the bounds of the law. In hindsight, the lack of a social pressure effect with respect to this question is not surprising, and in fact perhaps should have been expected.
subjects’ attitudes about militant Islam within the United States. Another question, Q11, also measures attitudes about radical Islam, and the differences between treatment and control in this case are approaching marginal significance. It is plausible to think that individuals who encounter a unanimous opinion from their peers on an issue to which they have not given much thought might use that information to make inferences about the distribution of opinion on related questions. Then public compliance to a perceived group attitude may operate on inferred group attitudes just as well as realized group attitudes. But if this is the case, then it is not clear why other questions tapping opinions about militant Islam in the U.S. (e.g., Q12 and Q17) do not show a similar effect. In fact, for no questions other than Q8, Q9, and Q11 do differences between treatment and control even come close to being statistically significant. The fact that each of these questions seems qualitatively to measure a similar subcategory of attitudes about Islam and Muslims is promising evidence for the case that something systematic may be left uncovered. But the overall picture is that a form of indirect social pressure, in which subjects use available information to make inferences about social pressure in other areas, is not occurring in general.

The analyses thus far have not accounted for individual-level characteristics beyond assignment to treatment or control groups. To control for characteristics of the individual, a series of logistic regressions were used to model the individual-level probability of providing an inegalitarian response. For each model in Table 3.4, the unit of analysis is the individual subject, and the binary dependent variable is 1 if the subject gave the inegalitarian response to the question identified in each column heading, and 0 otherwise. Dummy variables are used to capture race (1 = white), gender (1 = female), religion (1 = Christian), and partisan identification (1 = Republican), and a count variable is used to capture education (1 = first year through 4 = fourth year or greater). Interestingly, at this level of analysis, when controlling for race, gender, religion, partisan identification, and education, only Q1, the question tapping nervousness about being on a flight with a Muslim man, exhibits a treatment effect that achieves a conventional level of statistical significance. However, the
Table 3.4: Logistic Regression Models of Inegalitarian Responses on Questions Used for Direct Social Pressure

<table>
<thead>
<tr>
<th></th>
<th>Question 1 Estimate (s.e.)</th>
<th>Question 4 Estimate (s.e.)</th>
<th>Question 5 Estimate (s.e.)</th>
<th>Question 6 Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.982* (0.385)</td>
<td>1.063 (0.660)</td>
<td>0.585 (0.362)</td>
<td>0.388 (0.456)</td>
</tr>
<tr>
<td>White</td>
<td>0.353 (0.465)</td>
<td>-0.405 (0.753)</td>
<td>-0.339 (0.397)</td>
<td>0.448 (0.597)</td>
</tr>
<tr>
<td>Female</td>
<td>0.162 (0.375)</td>
<td>0.325 (0.650)</td>
<td>0.159 (0.340)</td>
<td>0.441 (0.425)</td>
</tr>
<tr>
<td>Christian</td>
<td>-0.111 (0.424)</td>
<td>0.686 (0.850)</td>
<td>0.550 (0.389)</td>
<td>0.519 (0.553)</td>
</tr>
<tr>
<td>Republican</td>
<td>0.725† (0.421)</td>
<td>0.298 (0.766)</td>
<td>1.309*** (0.389)</td>
<td>1.227** (0.446)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.225 (0.198)</td>
<td>-0.570 (0.466)</td>
<td>-0.422* (0.182)</td>
<td>-0.003 (0.213)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.624** (0.611)</td>
<td>-2.767* (1.169)</td>
<td>-0.570 (0.526)</td>
<td>-3.080*** (0.794)</td>
</tr>
<tr>
<td>N</td>
<td>197</td>
<td>197</td>
<td>197</td>
<td>196</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-98.187</td>
<td>-38.739</td>
<td>-115.031</td>
<td>-79.645</td>
</tr>
</tbody>
</table>

Note: Cell entries are logistic regression coefficients; standard errors are in parentheses. Binary dependent variable is 1 if the respondent gave the inegalitarian response, and 0 otherwise.

***p < 0.001; **p < 0.01; *p < 0.05; †p < 0.10

Estimated treatment effects for Q4 and Q5 come very close to a 90% confidence level (for Q4, p ≈ 0.107, and for Q5, p ≈ 0.108). As expected from the tests of independence and differences of proportions between treatment and control groups, the treatment effect for Q6 does not come close to achieving statistical significance. Among the subject covariate profiles the only substantial predictor of the likelihood of giving an inegalitarian response to these questions is partisan identification. Specifically, self-identified Republicans are significantly more likely to express nervousness about being on a flight with a Muslim man, to estimate that most or about half of Muslims in the world are anti-American, and to state that people have the right to keep Muslim organizations out of their neighborhoods. Additionally, a greater number of years in school is associated with a lower likelihood of estimating that most or about half of Muslims in the world are anti-American.
Figure 3.3: Predicted Probabilities from Logistic Regression of Inegalitarian Responses on Questions Used for Direct Social Pressure

Note: Predicted probabilities are for treatment and control groups for each model in Table 3.4. Calculations are for a baseline subject is a white, female, Christian, Republican, college freshman. Actual probabilities are presented above the bars.

If we can momentarily put aside the lack of statistical significance for Q6, further interpretation is available through the calculation of predicted probabilities. Figure 3.3 provides a graphical representation of the predicted probability of giving the inegalitarian response for members of the treatment and control groups based on the estimates in Table 3.4. For each of these calculations, all other independent variables are held constant at 1, and therefore the baseline hypothetical subject under consideration is a white, female, Christian, Republican, college freshman. As expected, the most dramatic treatment-control probability difference is for Q1. A baseline subject in the control group has a probability of only 0.327 of openly expressing nervousness about being on the same flight as a Muslim man, while for a baseline subject in the treatment group this probability increases to 0.565 (an increase of roughly 0.24, from below the 0.5 threshold to above it). For Q4 and Q5, this increase is roughly 0.12, about half that of Q1. Interestingly, for Q5, tapping the estimation of anti-Americanism among Muslims, the baseline subject already exhibits a relatively high likelihood of providing the inegalitarian response (due largely to the strong impact of Republican party identifica-
Table 3.5: Logistic Regression Models of Inegalitarian Responses, Pooling Across Direct and Indirect Social Pressure Categories

<table>
<thead>
<tr>
<th></th>
<th>Indirect Questions</th>
<th>Direct Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (s.e.)</td>
<td>Estimate (s.e.)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.242 (0.166)</td>
<td>0.724** (0.275)</td>
</tr>
<tr>
<td>White</td>
<td>-0.123 (0.176)</td>
<td>0.055 (0.318)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.083 (0.156)</td>
<td>0.243 (0.262)</td>
</tr>
<tr>
<td>Christian</td>
<td>0.366* (0.170)</td>
<td>0.322 (0.307)</td>
</tr>
<tr>
<td>Republican</td>
<td>0.653*** (0.174)</td>
<td>1.071*** (0.289)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.106 (0.077)</td>
<td>-0.280* (0.139)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.303*** (0.237)</td>
<td>-2.060*** (0.430)</td>
</tr>
</tbody>
</table>

Observations 2745 787
Subjects 197 197
Var(αᵢ) 0.582 0.582
Log-likelihood -1522.737 -362.458

Note: Cell entries are logistic regression coefficients, with standard errors in parentheses. Binary dependent variable is 1 if the respondent gave the inegalitarian response, and 0 otherwise. All models include subject-level random intercepts, αᵢ, where i indexes subjects. The unit of observation is the subject-question.

***p < 0.001; **p < 0.01; *p < 0.05; †p < 0.10

tion, but the treatment has the impact of increasing this likelihood even further. For Q4, both treatment and control subjects exhibit a relatively low likelihood of openly providing the inegalitarian response.

It may also be the case, however, that focusing on subject-level analyses for individual questions is somehow masking a more general process that might be revealed by pooling across responses to questions. To begin to examine the impact of combining responses across questions, Table 3.5 presents a set of logistic regressions where, again, the binary dependent variable is 1 if the subject changed her response, and 0 otherwise. But in this case, the unit of analysis is the subject response to a particular question (i.e., the subject-question), so that each row of the data matrix is a unique subject-question combination. To account for the hierarchical nature of the data, subject-level intercepts are assumed to be drawn from a normal distribution and are allowed to vary randomly. The first model pools only across indirect questions and the second pools only across direct questions.

Because the directional relationships in the raw data for each question of interest here all point in the direction of support for General Social Conformity Hypothesis 1, it is plausible
to think that through borrowing strength across indirect questions and direct questions, the estimated relationship for the indirect questions may become more substantial, and that those for the direct questions may become less substantial (due to the lack of significance for the treatment control differences with respect to Q6). But neither of these are the case, and the relationships borne out by pooling are consistent with the previous discussions. First and foremost, the treatment effect in the case of the direct questions is to substantially increase the probability of expressing the inegalitarian response, while in the case of indirect questions the treatment has no estimated effect. In the case of the indirect questions, the only significant predictors of providing the inegalitarian response are subject self identification as Christian and Republican. Both the condition of being a Christian and the condition of being a Republican are expected to increase the likelihood of providing an inegalitarian response to an indirect question. Moving to the case of direct questions, the impact of partisan identification as Republican operates in the same direction and becomes even more substantively meaningful. The impact of being a Christian, however, dissipates, while the number of years in school becomes significantly associated with a decrease in the likelihood of giving inegalitarian responses.

Once again, we can calculate predicted probabilities to help with interpretation of the findings. Figure 3.4 provides a graphical representation of the fixed effects estimates from the model in Table 3.5, and the form of the characteristic subject is again assumed by setting the values of all covariates to 1. We find that the estimated increase in the probability of providing an inegalitarian response when moving from control to treatment is substantial (about 0.18). The condition of being treated with information about an inegalitarian response distribution among peers has the effect of moving the likelihood of providing an inegalitarian response from below to above the 0.5 threshold.

Combining the data into unique subject-question observations is not the only means of

---

2Estimates for $\alpha_i$ are set aside for the time being. Addition of the subject-level random intercepts can be considered as deviations away from the global constant presented in Table 3.5.
pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, along with least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, along with least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.

In terms of estimated directional effects and statistical significance, the negative binomial and OLS specifications mirror one another, and the results tell a familiar story given the pooling information across questions, however. Using the subject as the unit of analysis, we may also consider counts and proportions of inegalitarian responses given during the course of the study. To that end, Table 3.6 presents a series of negative binomial regressions of the number of inegalitarian responses given by subjects, alongside a series of ordinary least squares regressions of the proportion of responses that are inegalitarian. Data are presented for all questions of interest considered together, and for direct and indirect questions considered separately.
Table 3.6: Negative Binomial and Ordinary Least Squares Regressions of Counts and Proportions of Inegalitarian Responses

<table>
<thead>
<tr>
<th></th>
<th>All Questions</th>
<th>Indirect Questions</th>
<th>Direct Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative Binomial</td>
<td>Negative Binomial</td>
<td>Negative Binomial</td>
</tr>
<tr>
<td></td>
<td>Estimate (s.e.)</td>
<td>Estimate (s.e.)</td>
<td>Estimate (s.e.)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.301* (0.126)</td>
<td>0.208† (0.121)</td>
<td>0.514** (0.181)</td>
</tr>
<tr>
<td>White</td>
<td>-0.042 (0.135)</td>
<td>-0.073 (0.129)</td>
<td>0.148 (0.177)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.093 (0.116)</td>
<td>-0.092 (0.110)</td>
<td>0.148 (0.177)</td>
</tr>
<tr>
<td>Christian</td>
<td>0.270* (0.125)</td>
<td>0.252* (0.121)</td>
<td>0.301 (0.221)</td>
</tr>
<tr>
<td>Republican</td>
<td>0.498*** (0.125)</td>
<td>0.443*** (0.119)</td>
<td>0.687*** (0.189)</td>
</tr>
<tr>
<td>Education</td>
<td>-0.097 (0.059)</td>
<td>-0.095† (0.056)</td>
<td>-0.177† (0.096)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.330*** (0.180)</td>
<td>1.226*** (0.171)</td>
<td>-0.650* (0.307)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>3.690 (0.747)</td>
<td>5.268 (1.399)</td>
<td>10.308 (14.429)</td>
</tr>
<tr>
<td>$N$</td>
<td>184</td>
<td>184</td>
<td>185</td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>-447.536</td>
<td>-415.153</td>
<td>-221.884</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.168</td>
<td>0.144</td>
<td>0.148</td>
</tr>
</tbody>
</table>

Note: Cell entries in the negative binomial columns are negative binomial regression coefficients, where the dependent variable is the number of egalitarian responses given. $\theta$ is the estimated conditional variance of the response variable. Cell entries in the OLS columns are ordinary least squares coefficients. Standard errors are in parentheses. The unit of observation is the subject.

***p < 0.001; **p < 0.01; *p < 0.05; †p < 0.10
previously discussed findings. Across the board, self-identified Republicans are associated with a greater number or proportion of inegalitarian responses than non-Republicans. The number of years in school is generally marginally or close to marginally significantly related to a lower number or proportion of inegalitarian responses. And for the analyses in which all questions are pooled and indirect questions are pooled, self-identified Christians are associated with higher numbers or proportions of inegalitarian responses than non-Christians. But for our purposes here, the most noteworthy finding is that those in the treatment group are associated with higher numbers and greater proportions of inegalitarian responses no matter how the data are broken down. Holding all other factors constant, the treatment effect is smallest in the case of indirect questions only, and largest in the case of direct questions. The OLS models tell us that for indirect questions, a movement from control to treatment is expected to increase the proportion of inegalitarian responses by about 0.06, holding all other factors constant. For direct questions, a move from control to treatment is expected to increase the proportion of inegalitarian responses by about 0.108.

For the negative binomial specifications, a similar story emerges, but here it will once again prove useful to calculate predicted probabilities. Figure 3.5 presents predicted probabilities for each of the negative binomial specifications in Table 3.6, and for which the hypothetical subject under consideration is again assumed by setting all of the covariates equal to one. For each of the ways of subsetting the data, members of the control group have a higher probability of espousing a low number of inegalitarian responses. For instance, out of the four direct questions, the members of the control group have a higher probability of expressing no or only one inegalitarian response. The members of the treatment group, however, have a higher probability of expressing two, three, or four inegalitarian responses. In the model for indirect questions only, out of the fourteen possible responses, members of the control group have a higher probability of expressing zero through five inegalitarian responses, while members of the treatment group have a higher probability of giving seven through 14. In the model that uses all questions combined, out of 18 possible responses, the probability of
expressing zero through seven inegalitarian responses is higher for members of the control group than in the treatment group, whereas the probability of expressing eight through 18 inegalitarian responses is higher for members of the treatment group. In addition to there being a generally higher probability of expressing a relatively large number of inegalitarian responses in the treatment group, for each model it is also true that the peak probability occurs at a lower number of inegalitarian responses in the control group than in the treatment group.

Discussion

The results of the Looming Social Pressure study have been examined in a variety of ways, all of which point to at least weak support for General Social Conformity Hypothesis 1. As we know from a plethora of previous public opinion research, when Americans are asked to express their attitudes on possibly sensitive social issues, the attitude distribution is
typically skewed toward whatever happens to be the so-called politically correct response. In the case of questions tapping a sense of equality between racial, ethnic, religious, gender, or other groups, the politically correct response in modern times is certainly in favor of egalitarianism under most circumstances. This result is exactly what was seen with respect to the control group. However, when individuals are able to learn about the distribution of opinions in a peer group that they will later have to meet for an unspecified face-to-face decision-making task, and when that distribution is skewed in favor of inegalitarianism, the results presented here suggest that they can be pulled away from that general egalitarian position. These data do not allow us to say whether individuals in the treatment group are actually generally in favor of a more egalitarian position and some significant number are falsifying their preferences to publicly comply with the group opinion; or whether members of the control group are actually generally in favor of a more inegalitarian position, and some significant number are altering their responses because of typical social desirability effects; or whether a significant proportion of subjects simply do not have a well formed attitude with respect to many of the questions in the survey, and they are looking to whatever information is available to them to form an opinion on the fly. Without knowing the source of the tendency toward public compliance, however, the results are clear that a significant amount of public compliance is in fact occurring.

The next study attempts to tease out more of the issues with respect to whether and what kind of individuals are publicly complying by expressing an attitude that they would not otherwise express. I turn to that study now.

### 3.3.4 The Real-Time Social Pressure Experiment

To test the hypotheses put forth regarding the propensity of individuals to alter their expressed attitudes in the face of social pressure requires comparisons of the behavior of individuals with respect to the same task, but under different social circumstances. That is,
testing requires a within-subjects design that records measurements on subjects at multiple time points. To develop such a design, it is useful to again revisit the general framework put forth by Asch and Sherif, but make some important alterations and extensions.

Scholars in a variety of fields have argued over how to interpret the results from the Asch and Sherif research programs, with some asserting that subjects are publicly complying with the group while privately maintaining a different viewpoint (e.g., Kiesler and Kiesler 1969), and others contending that subjects are unsure about the correct answer and are using the views of fellow group members to learn the appropriate response (e.g., Thelen, Dollinger, and Kirkland 1979). Boyd and Richerson (1985), on the other hand, argue convincingly that one should leverage the difference in the level of ambiguity between the two decision-making tasks and that doing so may lead to different conclusions about the phenomenon exhibited by subjects. Specifically, since the judgment situation under study in the Asch experiments has an answer that is clearly correct, while the judgment situation from the Sherif experiments does not, it is plausible to conjecture that the subjects who conform to the view of the confederates in the Asch experiments are exhibiting mere public compliance, while subjects who conform in the Sherif experiments are exhibiting a form of social learning. But this has not yet been tested in the extant literature. In the Real-Time Social Pressure Experiment, I provide a design that allows one to explicitly differentiate between these explanations empirically.

In the present study, I employ a design that takes a baseline attitude measurement of subjects in a private, isolated condition. I then use the responses given in private to simulate social pressure in a public group setting to determine if there is in fact a propensity for individuals to succumb to social pressure. Following that, I measure subjects once more to obtain some evidence on whether the response offered in the public setting was internalized.
Procedure

Overview At a computer terminal, subjects provided responses to survey questions first in isolation, then in a simulated group composed of other participants simultaneously taking part in the study, then again in isolation. To simulate the experience of participating in a computer-networked group interaction, subject photographs (headshots) were used in conjunction with original software\textsuperscript{4} to mimic certain aspects of social networking websites, with which a large proportion of subjects are expected to be familiar. Because the experiment relied on a deception, following their participation subjects were questioned as to whether the deception was convincing.

Preliminaries Up to 10 subjects at a time gathered in the same waiting room adjoined to the set of individual, private, enclosed computer terminals described in the Looming Social Pressure Experiment. After filling out preliminary paperwork, subjects were escorted one at a time into a terminal in which the computer was equipped with a webcam. At this terminal, subjects used a computer interface to enter identifying information and upload a headshot of themselves to a server. Depending on the number of subjects attending a given session, the preliminary picture upload portion of the study took approximately 5 to 15 minutes to complete, during which time the subjects not uploading a photo at any given moment were allowed to converse with one another. After the final photo was taken, all subjects were simultaneously read a set of instructions outlining the structure of the experiment, specifically noting that they would first answer a set of survey questions as individuals, then be assigned to a group of other participants and perform some tasks in groups, and then perform more tasks as individuals. Specific instructions are provided in Appendix A.

\textsuperscript{4}Software to implement the design was written by the author in consultation with a professional developer. Given the dynamic requirements of the experimental design, and to ensure cross-platform compatibility for future development, the use of web forms was chosen as the most straightforward implementation method. Page content, layout, actual dynamics, simulated dynamics, and database queries were handled using HTML, CSS, PHP, JavaScript, and MySQL, respectively. More detailed information about software development is available from the author upon request.
**Individual Pre Period** Following the instructions, subjects took seats at enclosed computer terminals where they encountered a set of instructions on how to begin the individual questionnaire. After first selecting and confirming the identifying information provided during the picture upload process, subjects answered a series of survey questions in typical computer-interface fashion. As before, the primary items of interest for the purposes of hypothesis testing are questions that tap non-Muslim subjects’ attitudes with respect to Muslims and Islam. To assist in masking the intended purpose of the study, in between questions of primary interest subjects were asked questionnaire items that tapped other political attitudes, as well as demographic and personality characteristics. All questionnaire items measure concepts worthy of study in their own right in a typical mass survey, but the primary purpose of the non-Muslim, non-Islam questions in this study was to serve as distraction between repeated tapping of the primary concept of interest.

**Group Period** Following the initial individual-level survey items, subjects reached a waiting screen where they were told that their group was being formed for the group period. After a brief delay, each subject was taken to a screen where their own picture was shown alongside two other participants taking part in the study at the same time. Subjects were told that this collection of participants would comprise their group. Additionally, at this time subjects were instructed that the group period of the study would consist of answering questions similar to those asked in the Individual Pre Period, the difference being that all group members would be able to witness all other group members’ responses. Additionally, subjects were told that the order in which group members would answer each question would be randomly chosen. As subjects navigated through the questions during the group round, the order in which group members would answer each question was made known to the subjects by displaying group members’ pictures in order from left to right, with group members’ responses displayed below their pictures.

In reality, the groups were fictitious constructions. Pictures of other participants taking
part in the experiment and pre-programmed dynamic responses actually played the role of confederates. That is, for any given subject, group members’ supposed responses were either predetermined or a function of the subject’s previous responses in Individual Pre Period. Specifically, intermixed within a set of distraction items that subjects answered in the first or second position in the group, and for which they faced a mixture of agreement and disagreement from their fellow group members, subjects encountered a series of questions about Muslims and Islam that they had previously been asked in the initial round of questioning. For these items (the questions of primary interest), subjects were asked to answer last (or, in one case, in the middle position) in the face of a group whose responses were unanimously in diametric opposition to the response given by the subject in the initial round of questioning. For example, suppose that in the Individual Pre Period a subject responded to a question by choosing the egalitarian (inegalitarian) response. Then, in the group round, they would be asked to respond in the face of a unanimous majority choosing the inegalitarian (egalitarian) response. In this way, the responses that subjects supply in isolation provide a baseline measure of attitudes from which to gauge their propensities to deviate from that baseline when confronted with an opposed majority group attitude. This feature of the design parallels the basic approach employed by Asch (1951, 1952, 1956) and Sherif (1935, 1936) with respect to arbitrary decision-making tasks, but applies it to the measure of political attitudes that should be expected to carry some social significance for the subjects outside of the laboratory.

Following a series of questions about Islam and Muslims in which subjects were compelled to answer after other group members while facing a unanimously opposed majority, subjects then encountered a set of similar questions in which they were compelled to answer in the first group position. Similar to the design executed by Jacobs and Campbell (1961), the repeated instances of answering in the last group position allows the subject the opportunity to learn about the supposed distribution of opinion in the group. Then the instances that follow in which the subject answers in the first group position allow for hypothesis testing
Table 3.7: Summary of Questions Used to Examine the Effect of Social Pressure

<table>
<thead>
<tr>
<th>ID</th>
<th>Question Wording in Group Round</th>
<th>Subject Position</th>
<th>Egalitarian Response</th>
<th>Inegalitarian Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Would you be nervous if a Muslim man were on the same flight as you?</td>
<td>Last</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q2</td>
<td>Would you mind living in a neighborhood where half of your neighbors are Muslim?</td>
<td>Last</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Q3</td>
<td>Do you favor special security checks at airports for Muslims?</td>
<td>Middle</td>
<td>Str. oppose/Oppose</td>
<td>Str. favor/Favor</td>
</tr>
<tr>
<td>Q4</td>
<td>Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?</td>
<td>Last</td>
<td>A little/Not at all</td>
<td>Mod. amount/A lot</td>
</tr>
</tbody>
</table>

Note: Questions appear in the order that subjects encountered them in the group round, separated by a series of distraction items. Group dynamics in the experiment are such that if a subject gave the egalitarian (inegalitarian) response in the Individual Pre Period, peers unanimously gave the inegalitarian (egalitarian) response in the Group Period. See text for details. ID numbers are used to refer to particular questions in the text.

about the propensity of individuals to perpetuate learned group opinion in further group interactions. The questions of interest used to examine social pressure effects, as well as the group dynamics with respect to these questions, are presented in Table 3.7. Question identification numbers, as well as question wording and response options in Table 3.7 correspond exactly to those in Table 3.1.

**Individual Post Period** Following the Group Period, subjects were again asked to answer a set of questionnaire items in isolation under the condition of anonymity. Other than the specific set of distraction questions asked, this period directly mirrored the Individual Pre Period. The purpose of this second isolation period was to allow for hypothesis testing about the nature of the responses given across the study (importantly, whether responses given in the group setting were carried over into the post-group setting). Survey questions were followed by a debriefing that informed subjects of the deception employed in the study, and then a set of questions that probed for suspicion about the deception.
Experimental Conditions Each subject went through the above procedure. In addition, two factors were varied that made three slightly different experimental conditions.

1. Expectation of a Face-to-Face Meeting:

In one of the experimental conditions, subjects were given instructions that informed them explicitly that they would not engage in any face-to-face meeting with their fellow group members following the computer portion of the study. In two other conditions, subjects were given instructions that included a ruse that they would have a face-to-face meeting with the same group following the computer-based interactions. These instructions are provided in Appendix A.

2. Priming of Negative Muslim Attitudes:

In one of the experimental conditions, subjects were asked a battery of questions tapping attitudes about the anti-terrorism policies of the United States government.

The three experimental conditions are provided in Table 3.8. Ultimately, the differences between the experimental conditions will not be leveraged substantially for the purposes of hypothesis testing. The intention of telling one subset of participants that they would not engage in a face-to-face meeting with their computer-networked group while telling another subset that they would was to lend further evidence for the relationship posited in General Social Conformity Hypothesis 1. This will be discussed briefly in what follows. In a simple comparison of proportions, the difference in propensity to change responses between conditions only worked according to the hypothesis for one of the four survey questions that were the focus of the analysis. The intention of giving one subset of subjects a battery of terrorism policy questions prior to the group portion and not the other subset was to test whether this would lead to a greater number of changes in responses between the Individual Pre Period and the Group Period. Ultimately, no such differences between conditions were found. In the discussion that follows, breaking the subjects down into their experimental
Table 3.8: Experimental Conditions That Result by Varying Two Factors

<table>
<thead>
<tr>
<th>Condition</th>
<th>Expectation of a Face-to-Face Meeting</th>
<th>Inclusion of Terrorism Policy Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Condition 2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Condition 3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

conditions is primarily used to categorize them according to whether they went through exactly the same protocol.

Results

To get an initial sense of subject responses on the questions of primary interest for the Real-Time Social Pressure Experiment, Figure 3.6 presents the raw data on expressed attitudes for the questions in Table 3.7. Data are separated both by experimental condition (including a category that pools across conditions) and whether subjects gave the egalitarian response or the inegalitarian response in the Individual Pre Period. Panel labels in the figure indicate both the condition and the initial response given by subjects. Individual data points are jittered for presentation purposes.

From simple visualization, the most striking feature of these data is that subjects overwhelmingly tend to give egalitarian responses to the questions. On each question, and across conditions, a far greater proportion of respondents initially gave the egalitarian response than otherwise. This is not surprising given the results previously presented. There is, however, interesting variation. On Q1, for example, only about 14% of subjects gave the inegalitarian response in private. On Q2 over 31% of subjects did so; on Q3, about 19% of subjects did so; and on Q4 only 2% of subjects did so. As expected, the vast majority of subjects give the egalitarian response, regardless of the question, and regardless of the experimental condition. But somewhat surprisingly, the question that elicited the greatest proportion of inegalitarian responses was the question tapping concern about living in a
Figure 3.6: Individual Responses Across Individual Pre, Group, and Individual Post Settings in the Real-Time Social Pressure Experiment

**Q1:** Would you be nervous if a Muslim man were on the same flight as you?

**Q2:** Would you mind living in a neighborhood where half of your neighbors are Muslim?

**Q3:** Do you favor special security checks at airports for Muslims?

**Q4:** Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?

Note: Subject responses for each condition experimental condition, and all conditions pooled, are separated according to whether subjects provided the egalitarian response or the inegalitarian response in the Individual Pre Period. Panel labels indicate the experimental condition and the initial response. On egalitarian and inegalitarian response choices, consult text and Table 3.7.
predominantly Muslim neighborhood, while the question that elicited the lowest proportion tapped concern about one Muslim neighbor. This suggests that non-Muslims who encounter these hypothetical circumstances generally do not feel compelled to express concern about Muslims living in proximity to them until the number of Muslims reaches some threshold. Beyond this point, such people feel comfortable enough to express their concern, against the norm of egalitarian rhetoric.

A second obvious pattern that emerges from these data speaks directly to the question of whether individuals tend to change their responses in the face of a unanimously opposed majority. For Q1, Q2, and Q3 at least, the figure indicates that a large percentage of subjects who initially provided the inegalitarian response publicly complied with the egalitarian group response when compelled to follow their fellow group members. On the flipside, a very small percentage of subjects who initially provided the egalitarian response in isolation ended up changing to express the inegalitarian response in the face of an inegalitarian group majority. For Q4, a small percentage of subjects who initially provided the egalitarian response publicly complied with the unanimous group response, but not enough subjects initially provided an inegalitarian response to make a meaningful comparison.

While Figure 3.6 is suggestive, Table 3.9 presents formal statistical tests of the propensity of subjects to conform to the group response based on whether they gave the egalitarian or the inegalitarian response in the initial private setting. These analyses pool subjects across experimental conditions. Looking at the raw proportions who change between groupings of initial responses, the proportions of initial inegalitarians who changed their responses in the group round are uniformly larger than initial egalitarians. The $p$-values indicate that for all questions under consideration, the proportions who change responses between the Individual Pre Period and the Group Period among those who initially give an egalitarian response are significantly different than among those who initially gave an inegalitarian response.

The results in Table 3.9 do not control for any subject-level characteristics, however. To do so, Tables 3.10 and 3.11 present multivariate logistic regression models of the likelihood of
Table 3.9: Tests of Difference in Proportions Between Egalitarian and Inegalitarian Initial Responses

<table>
<thead>
<tr>
<th>Proportion Who Change Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Pre → Group</td>
</tr>
<tr>
<td>Egalitarian</td>
</tr>
<tr>
<td>Q1</td>
</tr>
<tr>
<td>Q2</td>
</tr>
<tr>
<td>Q3</td>
</tr>
<tr>
<td>Q4</td>
</tr>
</tbody>
</table>

Note: Proportions are the subset of subjects who changed responses between the *Individual Pre Period* and the *Group Period*. For responses identified as egalitarian and inegalitarian, refer to Table 3.7. Asymptotic p-values are based on the $\chi^2(1)$ approximation of the hypergeometric null distribution (Fisher 1922) using Yates’ continuity correction (Yates 1934).

changing responses between the *Individual Pre Period* and the *Group Period* using self-report measures from the questionnaire. Just as in the analyses in the Looming Social Pressure Experiment, dummy variables are used to capture race, gender, and religion, while a four-level variable is used to capture the number of years in school. In addition, these models control for the size of the group that gathered in the common waiting area for the study on a given date. The idea here is that because the hypothesized mechanisms of subjects altering or masking their attitudes have to do with considerations about appearing deviant from the broader population or the immediate social group, the number of subjects in the group could plausibly prime some sense of sociality among the subjects and therefore induce different behavior based on the size of the group. The models in Table 3.10 pool over responses to all four of the questions of interest; that is, the unit of observation is the unique subject-question combination. To control for unobserved subject characteristics, these models also include subject-level random coefficients.

A brief scan of Table 3.10 reveals that the characteristics of individuals do not seem to add very much explanatory value beyond what was already known. Whether the subject initially provided the egalitarian response is a highly significant predictor of a lower probability of changing responses in the face of group pressure. But little else of substance can be found among the other covariates. While it is useful to control for demographic variables, the fact that they do not seem to be systematically related to the probability of changing responses
Table 3.10: Logistic Regression Models of Changing Responses Between Individual and Group Rounds, Pooling Across Questions

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td></td>
<td>(0.412)</td>
<td>(1.048)</td>
<td>(0.993)</td>
<td>(0.510)</td>
</tr>
<tr>
<td>White</td>
<td>-0.470</td>
<td>-0.900</td>
<td>0.152</td>
<td>-0.518</td>
</tr>
<tr>
<td></td>
<td>(0.494)</td>
<td>(1.126)</td>
<td>(1.230)</td>
<td>(0.544)</td>
</tr>
<tr>
<td>Female</td>
<td>0.180</td>
<td>0.088</td>
<td>2.415†</td>
<td>-0.516</td>
</tr>
<tr>
<td></td>
<td>(0.461)</td>
<td>(1.019)</td>
<td>(1.345)</td>
<td>(0.531)</td>
</tr>
<tr>
<td>Christian</td>
<td>-0.546</td>
<td>-2.455†</td>
<td>-0.449</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>(0.524)</td>
<td>(1.437)</td>
<td>(1.116)</td>
<td>(0.598)</td>
</tr>
<tr>
<td>Republican</td>
<td>0.815</td>
<td>0.831</td>
<td>1.467</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.550)</td>
<td>(1.552)</td>
<td>(1.252)</td>
<td>(0.595)</td>
</tr>
<tr>
<td>Education</td>
<td>0.326</td>
<td>1.038†</td>
<td>0.603</td>
<td>-0.262</td>
</tr>
<tr>
<td></td>
<td>(0.234)</td>
<td>(0.589)</td>
<td>(0.477)</td>
<td>(0.285)</td>
</tr>
<tr>
<td>Group Size</td>
<td>0.177</td>
<td>0.122</td>
<td>-0.417</td>
<td>0.624**</td>
</tr>
<tr>
<td></td>
<td>(0.190)</td>
<td>(0.387)</td>
<td>(0.463)</td>
<td>(0.237)</td>
</tr>
<tr>
<td>Condition 2</td>
<td>0.484</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.672)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition 3</td>
<td>-0.185</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.721)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.159</td>
<td>-2.161</td>
<td>-0.746</td>
<td>-2.646</td>
</tr>
<tr>
<td></td>
<td>(1.583)</td>
<td>(3.444)</td>
<td>(3.258)</td>
<td>(1.491)</td>
</tr>
<tr>
<td>Observations</td>
<td>600</td>
<td>144</td>
<td>152</td>
<td>304</td>
</tr>
<tr>
<td>Subjects</td>
<td>150</td>
<td>36</td>
<td>38</td>
<td>76</td>
</tr>
<tr>
<td>Var(αᵢ)</td>
<td>2.092</td>
<td>2.105</td>
<td>2.731</td>
<td>0.283</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-158.625</td>
<td>-32.927</td>
<td>-40.935</td>
<td>-72.904</td>
</tr>
</tbody>
</table>

Note: Cell entries are logistic regression coefficients, with standard errors in parentheses. The binary dependent variable is 1 if subject changed responses to a given question between the first individual round and the group round, and 0 otherwise. All models include subject-level random intercepts, αᵢ, where i indexes subjects. The unit of observation is the subject-question. ***p < 0.001; **p < 0.01; *p < 0.05; †p < 0.10

is not all that surprising. It is not clear a priori why gender, religion, or party identification would be related to the probability of changing responses. We might have surmised that education would lead to a lower probability of changing responses due to simple need for response consistency, or that group size would lead to a higher probability of changing responses due to the inducement of a greater sense of social pressure. But neither of these relationships are borne out by the data at hand.
The impact of the initial response, however, is substantial. To ease interpretation, Figure 3.7 presents predicted probabilities for the individual condition models from Table 3.10. The plot illustrates the estimated effect of moving from an inegalitarian response to an egalitarian response while holding all of the demographic and education variables equal to one, and holding group size at its mean value. The effect is very slight within condition one, largely due to significant covariation with religion and education. In conditions two and three, the estimated effect of providing an inegalitarian initial response is quite substantial. In the case of condition two, the probability of changing responses between the initial private period and the group period is only about 0.1. But if a subject initially gave the inegalitarian response to a question, that probability increases to about 0.75, an increase of roughly 5

Note that the coefficient estimates for Education and, especially, Christian, are substantially different in the model for condition one as opposed to the same coefficient estimates for all other cases. Examining the summary statistics in Appendix B, there is no obvious difference between conditions that would be expected to lead to this anomaly. Without further investigation, the most plausible assessment is that a small number of subjects giving the initial inegalitarian response along with a few anomalous behaviors within religion and education categories contributed to this difference. The direction and substantive significance of the coefficient estimate for giving the initial egalitarian response, however, is quite consistent across specifications.
0.65. That is, the probability goes from around a 10% chance of changing responses between the *Individual Pre Period* and the *Group Period* when a subject initially falls in line with the socially desirable response in the broader population to an almost 75% chance when a subject initially goes against egalitarian norms in private. In condition three, the jump is not quite as dramatic, but the relationship is nonetheless quite strong. In this case, the probability of changing response between the initial private setting and the real-time social pressure setting increases from 0.032 if a subject initially gives the egalitarian response to just over 0.5 if a subject initially gives the inegalitarian response (an increase of about 0.47). Momentarily setting aside the anomaly of condition one, the results lead to the conclusion that when subjects initially express the inegalitarian attitude in private they typically have a greater than 50% chance of changing their responses in the face of group pressure to do so.

The results from the difference of proportion tests of the single variable and the logistic regressions presented thus far show strong initial support for *Social Dissonance Hypothesis*, and little reason to believe *General Social Conformity Hypothesis 2*. All signs thus far point to the idea that if the subject initially gives a response that runs counter to known norms in the broader population, she will be more susceptible to group pressures to provide the opposite response. Subjects who express a view in agreement with the broader population norm show little susceptibility to group social pressure to publicly comply with a group norm that runs counter to the broader population norm.

Table 3.11 presents a set of similar models, but estimates each question separately, here pooling across experimental conditions. Similar to the above analysis that pooled responses across questions, the covariate profile in the multivariate specification offers little in the way of explanatory power beyond the main variable of interest. The effect of the initial response, however, remains substantial. To be sure, there is significant variation across questions in the actual substantive impact of initially giving an egalitarian response. Holding all other factors constant, the impact of the initial choice on the probability of changing is most substantial
<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>Egalitarian Initial Response</td>
<td>-3.548***</td>
<td>-2.541***</td>
<td>-4.329***</td>
</tr>
<tr>
<td></td>
<td>(0.743)</td>
<td>(0.569)</td>
<td>(0.908)</td>
</tr>
<tr>
<td>White</td>
<td>-0.270</td>
<td>-0.883†</td>
<td>-0.260</td>
</tr>
<tr>
<td></td>
<td>(0.697)</td>
<td>(0.519)</td>
<td>(0.772)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.101</td>
<td>-0.086</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>(0.669)</td>
<td>(0.497)</td>
<td>(0.727)</td>
</tr>
<tr>
<td>Christian</td>
<td>-0.393</td>
<td>-0.758</td>
<td>-0.372</td>
</tr>
<tr>
<td></td>
<td>(0.756)</td>
<td>(0.553)</td>
<td>(0.810)</td>
</tr>
<tr>
<td>Republican</td>
<td>0.385</td>
<td>0.498</td>
<td>-0.453</td>
</tr>
<tr>
<td></td>
<td>(0.823)</td>
<td>(0.586)</td>
<td>(0.866)</td>
</tr>
<tr>
<td>Education</td>
<td>0.088</td>
<td>0.361</td>
<td>0.706†</td>
</tr>
<tr>
<td></td>
<td>(0.320)</td>
<td>(0.248)</td>
<td>(0.400)</td>
</tr>
<tr>
<td>Group Size</td>
<td>0.495†</td>
<td>-0.057</td>
<td>0.242</td>
</tr>
<tr>
<td></td>
<td>(0.272)</td>
<td>(0.209)</td>
<td>(0.312)</td>
</tr>
<tr>
<td>Condition 2</td>
<td>0.507</td>
<td>0.505</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>(0.915)</td>
<td>(0.696)</td>
<td>(1.181)</td>
</tr>
<tr>
<td>Condition 3</td>
<td>-0.198</td>
<td>-0.991</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(1.065)</td>
<td>(0.779)</td>
<td>(1.195)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.888</td>
<td>0.554</td>
<td>-3.767</td>
</tr>
<tr>
<td></td>
<td>(2.136)</td>
<td>(1.837)</td>
<td>(2.822)</td>
</tr>
<tr>
<td>Subjects</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-35.731</td>
<td>-56.767</td>
<td>-30.374</td>
</tr>
</tbody>
</table>

Note: Cell entries are logistic regression coefficients, with standard errors in parentheses. The binary dependent variable is 1 if subject changed responses to a given question between the first individual round and the group round, and 0 otherwise. The unit of observation is the subject.

**p < 0.001; *p < 0.01; *p < 0.05; †p < 0.10

for Q3, followed by Q1, and then Q4 and Q2. In the cases of Q1, Q2, and Q3, this relationship achieves the 99% level of statistical significance, while for Q4 the relationship achieves a 90% level. The marginal significance of the effect for Q4 is somewhat expected given the results already discussed.

Again, the substantive impact of the estimated effects can be assessed further by examining predicted probabilities. Figure 3.8 presents predictions from each model in Table 3.11 for moving from an egalitarian to an inegalitarian initial response, using the same representative
assumed values for the covariate profile as discussed previously, and separating subjects by condition. A similar pattern emerges across conditions, with some important variation. For instance, the moderately large difference between the estimated effect of an initial egalitarian response between condition one and conditions two and three for Q1 and Q4, and the substantially large difference for Q3, offers some further insight into the differences between condition one and the others exhibited in the preceding analyses. In condition one, with no expectation of a later face-to-face meeting, subjects seem to have simply had a generally lower baseline probability of changing responses in the face of a unanimously opposed group attitude distribution. This lends further support to General Social Conformity Hypothesis. In moving from an egalitarian initial response to an inegalitarian initial response for Q1, subjects exhibit a substantial increase in the probability of changing responses between the initial setting and the group pressure setting that is fairly consistent across experimental conditions (differences in probability of about 0.47 in condition one, 0.51 in condition two, and 0.641 in condition three). For conditions two and three, the same increase in probability for Q3 is even more dramatic (about 0.79 and 0.76, respectively). For Q2, this increase in probability is more modest, and for Q4 even more modest still. Further, in conditions two and three, the representative hypothetical subject who initially expresses the inegalitarian attitude to Q1, Q3, or Q4 changes responses in the face of group pressure with a probability approaching certainty.

In short, the results in Table 3.11 and Figure 3.8 generally offer significant credence to the relationship posited in the Social Dissonance Hypothesis. Whether examining subject behavior with respect to specific questions, all questions pooled together, or across or within conditions, the subject choice to initially provide the inegalitarian response, that is, the response that is clearly not the socially desirable one, turns out to be highly consequential in attempting to predict whether the subject will alter that expressed attitude when encountering unanimously opposed group pressure. Subjects who initially expressed the egalitarian attitudinal response, in agreement with the socially desirable response in the broader pop-
Figure 3.8: Predicted Probabilities from Logistic Regressions Between Individual and Group Rounds, Individual Questions

Note: Predicted probabilities are based on the second, third, and fourth models in Table 3.11. Calculations are for a baseline subject is a white, female, Christian, Republican, college freshman, and group size is held at its mean value.
ulation, are overwhelmingly likely to maintain that attitude even when they are compelled to express it publicly in front of an audience who unanimously expresses an opposing view. On the other hand, subjects who initially expressed the inegalitarian response, that is, the socially unacceptable pronouncement in the broader population, are overwhelmingly likely to alter their expressed attitude in the face of opposing group pressure.

At this point, it would be useful to further explore the mechanism underlying the propensity of subjects to alter their responses between the Individual Pre Period and the Group Period. That is, the Individual Post Period allows us to assess whether any changes in responses between the first two periods are evidence of learning and internalizing the response given under group pressure, or whether they are evidence of preference falsification. However, because, as we have seen, altering responses under group pressure is so overwhelmingly associated with an initial inegalitarian response, and because so few subjects actually gave an inegalitarian response in the initial private setting, we are left with a very small sample of subjects with whom to examine the relationship between group behavior and the following private behavior. Therefore, the discussion that follows will necessarily only be suggestive and preliminary.

Table 3.12 presents preliminary evidence speaking to whether the subjects who change their expressed preferences are merely publicly complying by falsifying their preferences. N sizes listed are the number of subjects who publicly complied with the expressed group response. As stated, these numbers are incredibly small for formal analysis. Simply examining the raw proportions of subjects who go back to their original response after having changed under group pressure, the strongest evidence for preference falsification lies with Q4, while the strongest evidence for learning or on-the-fly preference formation lies with Q3. Interestingly, for all of the questions that tap subjects’ general level of concern or worry about being in proximity to Muslims (Q1, Q2, and Q4), over half of those who altered their responses between the Individual Pre Period and the Group Period went back to their original responses in the Individual Post Period. For Q3, which taps subjects’ preferences
Table 3.12: Proportions of Subjects Who Publicly Comply by Falsifying Their Preferences

<table>
<thead>
<tr>
<th>Question</th>
<th>Combined Initial Responses</th>
<th>Inegalitarian Initial Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion Who Change</td>
<td>Proportion Who Change</td>
</tr>
<tr>
<td></td>
<td>N Group → Individual Post</td>
<td>N Group → Individual Post</td>
</tr>
<tr>
<td>Q1</td>
<td>7</td>
<td>0.571</td>
</tr>
<tr>
<td>Q2</td>
<td>14</td>
<td>0.571</td>
</tr>
<tr>
<td>Q3</td>
<td>18</td>
<td>0.333</td>
</tr>
<tr>
<td>Q4</td>
<td>10</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.583</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: Data presented are for subjects who altered their initial responses to publicly comply with the group opinion. Data are separated by pooling across egalitarian and inegalitarian initial responses and by focusing only on the subjects who provided the inegalitarian response in the initial period. N sizes are the total number in each category who changed their responses between the Individual Pre and Group periods. Proportions are the proportions in each category who change their responses again between the Group and Individual Post periods.

for special security checks at airports for Muslim travelers, only a minority of subjects who changed under group pressure changed back when asked again in private. Speculating for a moment, it may be the case that subjects see security checks as more of a policy issue than a personal feeling, and policy issues are more subject to learning and internalization than personal feelings. This is something that can be explored further in future iterations.

I can also examine the differences between those who initially express egalitarian and inegalitarian attitudes to the four questions. For Q1, all of the subjects who exhibited evidence of preference falsification initially gave the inegalitarian response to the question. For Q2, seven out of the eight subjects who changed back to their original preference initially expressed an inegalitarian attitude. In the case of Q3, the policy question, however, the number of initial egalitarians who change their responses is the same as the number of initial inegalitarians who change. And for Q4, interestingly, among the 10 total subjects who changed their responses between the Individual Pre Period and the Group Period, six changed back to their original responses in the Individual Post Period, and all of those six initially gave the egalitarian response to the question.

At this point, little can be said with respect to preference falsification vs. learning and internalization because of the relatively small number of respondents who change responses under group pressure. More data collection will be required to determine whether the patterns evident now will hold in reasonable sample sizes. We cannot claim support for either
the *Attitude Internalization Hypothesis* or the *Preference Falsification Hypothesis*. At this point, differentiating between these two competing explanations will remain inconclusive.

Finally, and briefly, after encountering Q1, Q2, Q3, and Q4 in the *Group Period*, and being compelled to answer in the face of group opposition to what had been expressed in private, subjects encountered a further set of questions about Islam and Muslims for which they found themselves in the first group position. This dynamic was included in the design to determine whether, upon learning the distribution of responses across a variety of related questions, the subjects would still feel compelled to alter their responses, and whether the underlying mechanism would be similar to the mechanism evident in the situations in which subjects answered first. That is, the basic question was whether, after some conditioning, the tendency to publicly comply with an expected group behavior would operate in the same way as the tendency to comply with a known group behavior.

Figure 3.9 presents the raw data for the questions in which subjects answered first in the *Group Period*. It is clear that in each case, some small minority of subjects do in fact alter their original responses when they are under the impression that others will witness their behavior. But unlike the questions in which the subject answered last in the face of an opposed group, no clear pattern emerges here with respect to the whether the subject initially provided the egalitarian or inegalitarian response. This was confirmed by performing similar statistical analyses to those presented previously. These raw data are simply offered in the name of completeness.

**Discussion**

A variety of methods have been used to examine the results of the Real-Time Social Pressure Experiment, all of which point to strong support for the *Social Dissonance Hypothesis*. As we saw, and as is consistent with a large body of previous research, when a pool of subjects are presented with a set of attitudinal questions about sensitive social issues, tapping concepts such as religious equality, the vast majority express attitudes in favor of the egalitarian point
Figure 3.9: Individual Responses Across Individual Pre, Group, and Individual Post Settings in the Real-Time Social Pressure Experiment, where Subject is in First Group Position

**Q10:** Would you mind if your children were to attend a predominantly Muslim school?

**Q19:** Based on the events of September 11th, 2001, would you favor or oppose the U.S. government tightening immigration laws to restrict the number of immigrants from Arab or Muslim countries entering the United States?

**Q20:** Should someone who says that terrorist attacks against the America are the fault of how America behaves in the world be allowed to teach in a public school?

---

Note: Subject responses for each condition experimental condition, and all conditions pooled, are separated according to whether subjects provided the egalitarian response or the inegalitarian response in the Individual Pre Period. Panel labels indicate the experimental condition and the initial response. On egalitarian and inegalitarian response choices, consult text and Table 3.7.
of view. This result was clearly borne out in the Individual Pre Period. However, in the Group Period, interesting systematic variation became evident based on whether subjects had initially provided the egalitarian response. Specifically, the minority of subjects who initially provided the inegalitarian response were shown to be significantly more likely to alter their responses when later asked to answer the same question in the face of a unanimously opposed group opinion. In other words, on any given question tapping non-Muslim subjects’ attitudes with respect to Islam and Muslims, a large majority of subjects expressed support for an egalitarian view. These subjects were later put into a situation in which they faced a group opinion on the same question that unanimously supported an inegalitarian view, and were asked to publicly express their attitude on this question again. In this case, the overwhelmingly dominant tendency was for those subjects to resist social pressure to publicly comply and stick with the response that they originally gave privately. However, there was also a minority of subjects who were willing to express an inegalitarian view when they were initially asked a given question about Islam and Muslims. Then, when they were put into a situation in which they were compelled to state their opinion on that same question publicly in front of a group of peers showcasing a unanimous egalitarian response, the vast majority of these subjects altered their initial response in order to publicly comply with the group. This relationship is quite strong, and carries across specific questions and slight variations in the experimental protocol.

This relationship is exactly what was posited in the Social Dissonance Hypothesis. Borrowing the language of cognitive dissonance theory (Festinger 1957), social dissonance refers not to internal inconsistencies within the individual, but rather to inconsistencies between an opinion held by the subject and a known appropriate social response. Specifically, it is expected that on matters of egalitarianism with respect to various groups in societies, individuals who are willing to express an inegalitarian view in private will almost surely be aware that the attitude that they have expressed runs counter to the expected appropriate social response in the broader population. Because individuals have a preference for not
appearing deviant from the broader population, the dissonance between the inegalitarian view expressed and the assumed social view is expected to be a salient thought in the minds of subjects when they encounter that same question again a short time later. If the immediate social group then confronts the subject with a unanimous view that falls in line with that of the broader population, the subject’s preference for not appearing deviant will have a tendency to overwhelm any preferences for stating what one truly believes, or even for consistency with a view that they expressed only a short time prior. This preference is then likely to lead to public compliance with the expressed group opinion. This expectation was strongly supported by the results of the Real-Time Social Pressure Experiment.

With respect to the specific form that the public compliance takes, the jury is still out. The data presented here remain inconclusive on whether the changes in responses represent preference falsification or attitude internalization. While three out of the four questions examined here saw a majority of subjects who changed their responses between the Individual Pre Period and the Group Period then go back to their original response in the Individual Post Period (on the face exhibiting evidence in favor of preference falsification), there are not yet enough subjects who changed their responses in the first place to say with any confidence whether this a general phenomenon.
Chapter 4

Homophilic Agent Interactions and Network Development: Model and Simulations

4.1 Introduction

Recent empirical research in political science has demonstrated that social networks are both ubiquitous and highly consequential in a wide range of social and political settings (e.g., Fowler 2006; Fowler and Smirnov 2007; Franzese and Hays 2008; Gibson 2001; McClurg 2006; Siegel 2009). It would therefore stand to reason that the decision-making processes by which individuals choose to form, maintain, and break social ties is an important area of inquiry for political scientists. Studies on random network formation have a long pedigree in physics and mathematics (e.g., Erdős and Rényi 1959; Rapaport 1957). More recently, scholars have begun to take note of the specific features of social networks, and are increasingly aware that the choices of individuals, and not simply random chance, play a crucial role in how social networks form and change over time. A fairly large body of social science literature is developing around the idea that individuals tend to form social relationships that are mutually beneficial in some respect, and to sever relationships that are not beneficial (see, e.g., Bloch and Dutta 2009; Bloch and Jackson 2007; Caroyal and Roux 2009; Goyal and Joshi 2003; Hojman and Szeid 2008; Jackson 2008; Jackson and Wolinsky 1996; Joshi 2008), but surprisingly little work has been done on strategic network formation in political science. Scholars in political science have long been interested in questions such as alliance formation (Walt 1987; Waltz 1979; Reiter 1994) and the formation of political coalitions (Adrian and Press 1968; Grofman 1982; Laver and Shepsle 1990), processes that explicitly involve the selection of partners to achieve political goals. But these areas of political science have not
taken on these questions from the perspective of network formation.

Given the nature of the questions addressed by scholars of strategic network formation, and the growing size of the body of work, the failure of political scientists to explicitly address this literature is somewhat puzzling. I begin the process of bridging the gap between political scientists and scholars who focus on the strategic factors influencing partner selection in social network formation and dynamics. Specifically, in this chapter I present a simple game theoretic model of network formation in which players earn utility by forming links with other players. The basic framework of this model has been examined extensively elsewhere, but I include the added feature of players with heterogeneous types across multiple dimensions. Previous research on strategic network formation has assumed (explicitly or implicitly) that actors play network formation games in one-dimensional issue space. Even in the examination of non-strategic influences such as homophily—the tendency of actors to form links with similar others—researchers are typically interested in only one dimension of similarity. For example, previous studies employing network concepts have focused on single dimensions such as federalism (Broschek 2010), democratization (Aleman 2009; Lee 2009), trade and financial markets (Belleflamme and Bloch 2004; Fagiolo, Reyes, and Schiavo 2010; Goyal and Joshi 2003; Mauleon, Song, and Vannetelbosch 2010; Saggi and Yildiz 2010; Weber, Davis, and Lounsbury 2009), public expenditures (Jones and Breunig 2007), the environment (Cao and Prakash 2010), and anti-smoking (Shipan and Volden 2008). This simplifying assumption is limited in being able to explain network formation in a real-world environment, where actors have multiple identifiable characteristics, e.g., language, religion, sex, race, political, social, and economic ideology, and so on (Breton et al. 1995; Brown 2000; Kellas 1998; Smith 1998). If similarity can be defined on multiple dimensions, strategic actors may have incentives to form relationships with others as similar to themselves as possible. The study of preferences and behaviors over multidimensional space has a long history in political science, but has been examined very little in specifically network-analytic settings.

In this chapter, I model these more complicated situations by introducing the notion
of player types in two-dimensional space. After deriving some insight from this model by examining its properties in a very simple setting, I then use the outcome of the link formation game as the initial state of a network whose structure is allowed to change over time as a function of the utility-maximizing behavior of agents. I show that with random initial conditions, the networks that emerge can have very different forms and can lead to rather large differences in both individual and social utility for the agents.

The chapter proceeds as follows. In the next section I present the model of link formation. I then apply the model to a simple case with a small number of agents. In the section that follows, I present a numerical experiment of network change over time and examine the results. I follow this with some concluding remarks.

### 4.2 Strategic Model of Network Formation

In this section, I consider a model of the formation of network links in which agents receive positive utility by forming links with others similar to themselves, and negative utility by forming relationships with dissimilar others, where similarity can be defined in multiple dimensions. This setup is motivated by social and political situations in which agents may have objectives over multiple issue dimensions, and they feel compelled to form social relationships in order to achieve them. Let us say that preferences over multiple dimensions define agents’ types. With sufficient variation in the distribution of types in the population, clearly situations will arise in which agents agree with others on some dimensions but not others. Additionally, as many scholars have pointed out previously, actors may be limited in their abilities to gather information and to process the information gathered. In the setting examined here, without perfect information about the distribution of types, situations may arise in which agents are indirectly connected with individuals with whom they are opposite on all dimensions.
4.2.1 Simultaneous Link-Announcement Game with Heterogeneous Agents

Following Myerson (1977), and using the terminology and notation of Jackson (2008), I define the simultaneous link-announcement game as the game in which all players simultaneously and independently announce the set of players with whom they wish to form links. Links are formed when any two players announce that they want to be linked with one another. For the set of players, \( N = \{1, \ldots, n\} \), the strategy space for player \( i \in N \) is the power set of \( N \setminus \{i\} \), denoted \( S_i \). If \( s_i \in S_i \) is the strategy played by player \( i \) and \( s_j \in S_j \) is the strategy played by player \( j \), link \( ij \) forms if and only if \( i \in s_j \) and \( j \in s_i \). Further, define \( G(N) \) as the set of all undirected networks on \( N \), define \( g \in G(N) \), and let \( s = \{s_1, \ldots, s_n\} \) be the profile of strategies played. Then the network that forms in the simultaneous link-announcement game is \( g(s) = \{ij| i \in s_j \text{ and } j \in s_i\} \).

Players in the model are endowed with a two-dimensional vector of types, \( \Theta = (\theta_1, \theta_2) \), where \( \theta_k \in \{0, 1\} \), and \( k = 1, 2 \) indexes dimensions. Player \( i \)'s payoff for strategy profile \( s \) is then given by the following distance-based utility function (Bloch and Jackson 2007):

\[
u_i(s) = \sum_{j\neq i: j \in N_i^{n-1}(g)} b(\ell_{ij}(g))\phi_{ij} - d_i(g)c_m \tag{4.1}\]

where \( N_i^{n-1}(g) = N_i(g) \cup \left( \bigcup_{j \in N_i(g)} N_j^{n-2}(g) \right) \) represents all nodes that can be reached from \( i \) by walks (i.e., sequences of links connecting sequences of nodes) of length no more than \( n - 1 \), \( \ell_{ij}(g) \) is the shortest path between \( i \) and \( j \) in \( g \), and \( b : \{1, 2, \ldots, n-1\} \rightarrow \mathbb{R}^+ \) is the net benefit that a player receives from direct and indirect connections. Additionally,

\[
\phi_{ij} = \begin{cases} 
2a & \text{if } \theta_1^i = \theta_1^j, \theta_2^i = \theta_2^j; \\
\theta_1^i \neq \theta_1^j, \theta_2^i \neq \theta_1^j, \theta_2^i \neq \theta_2^j; \\
-2a & \text{if } \theta_1^i \neq \theta_1^j, \theta_2^i \neq \theta_2^j \\
\theta_1^i \neq \theta_1^j, \theta_2^i \neq \theta_2^j. 
\end{cases}
\]
for \( a > 0 \), where superscripts index players. The first term of the utility function therefore ensures that players receive the highest utility for forming direct connections with other players of the same type.

In the second part of the utility function, \( d_i(g) \) is the degree of player \( i \) in network \( g \), that is, the number of agent \( i \)'s direct links, and \( c_m \) is the cost of maintaining a link. This cost is intended to capture the intuitive idea that maintaining large networks of connections can be costly.

### 4.2.2 Illustration: Four Agents and Four Types

To illustrate some of the properties of the simultaneous link-announcement game with heterogeneous agents, consider the simple case of a four-agent society in which each of the four possible types are represented. Specifically, let \( N = \{1, 2, 3, 4\} \), and let \( \Theta_1 = \{1, 1\} \), \( \Theta_2 = \{0, 1\} \), \( \Theta_3 = \{1, 0\} \), and \( \Theta_4 = \{0, 0\} \). For any given player, I refer to any other player who differs on both dimensions of \( \Theta_i \) as an opposite type, who differs on exactly one dimension of \( \Theta_i \) as a one-type match, and who is the same on both dimensions of \( \Theta_i \) as a same type. For simplicity, let \( a = 1 \), let \( c_m = 0 \) (forming links is costless), and let \( b(x) = \frac{n-x}{n-1} \), so that \( b(1) = 1 \), \( b(2) = 2/3 \), and \( b(3) = 1/3 \). This simplification allows for trivial calculations while retaining a sensible notion of distance-based utility (Bloch and Jackson 2007).

In the four-agent case, there exist six possible dyads, and therefore \( \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 64 \) possible networks that could arise from any given strategy profile. Figure 4.1 presents these possibilities. It is important to note that the simultaneous link-announcement game requires that each player involved in a link announces that she wants to form a link with the other. The fully connected network, therefore, can only arise from one strategy profile: the profile in which each player announces a desire to form a link with every other player. Each of the other possible network structures, however, can arise from multiple strategy profiles. So for instance, the second network in the first row of Figure 4.1 can arise
from a strategy profile in which neither player 3 nor player 4 announces a desire to link with the other, or the strategy profiles in which only one of these players announces a desire to link with the other. For more sparsely connected networks, the number of corresponding strategy profiles becomes very large.
Given the nature of players’ utility calculations, not all of the networks in Figure 4.1 are sensible. In particular, forming a link with an opposite-type player leads to negative utility. It is then immediately clear that any strategy that involves announcing a desire to form a link with an opposite type is weakly dominated, and regardless of the solution concept employed should not be predicted to occur. We may therefore speculate that any networks involving connections between players 1 and 4 and between players 2 and 3 are unlikely to occur. Using the concept of Nash stability, this is in fact the case. Following Jackson (2008), I define a network \( g \) to be Nash stable if it results from a pure strategy Nash equilibrium of the simultaneous link-announcement game, where player \( i \)'s payoff as a function of the strategy profile is given in Equation (4.1). The full set of Nash stable networks in the four-player simultaneous link-announcement game with four types represented are presented in Figure 4.2.

At least one problem with deriving predictions from the simultaneous link-announcement game is that it has many Nash equilibria (Jackson 2008). In particular, Figure 4.2 shows that one-quarter of the 64 possible outcomes of the game are Nash stable networks. To refine the concept of stability in mutual-consent relationships, Jackson and Wolinsky (1996) introduced the notion of pairwise stability to account for the idea that if any two players want to form a link, then such a link should be an expected outcome even in non-cooperative settings. Introducing the notation \((g - ij)\) to mean the network that results by removing link \( ij \) from network \( g \) and \((g + ij)\) to mean the network that results by adding link \( ij \) to network \( g \), Jackson and Wolinsky (1996) define a network \( g \) as pairwise stable if:

(i) \( \forall ij \in g, u_i(g) \geq u_i(g - ij) \) and \( u_j(g - ij) \), and

(ii) \( \forall ij \notin g, \) if \( u_i(g + ij) > u_i(g) \), then \( u_j(g + ij) < u_j(g) \)

A network that is both Nash stable and pairwise stable is referred to as pairwise Nash stable. Pairwise stable networks can sometimes be overconnected, in that some player might benefit
Figure 4.2: Nash Stable and Pairwise Nash Stable Outcomes of the Simultaneous Link-announcement Game with Four Actors and Four Types Represented.

Note: Panels are the subset of networks from Figure 4.1 that are Nash stable or pairwise Nash stable.

by deleting multiple links at once (Jackson and Wolinsky 1996; Jackson 2008), but it has been argued elsewhere that pairwise Nash stability is a reasonable notion of network stability in a large variety of settings (e.g., Bloch and Jackson 2007; Calvó-Armengol and Ilkiliç 2009; Goyal and Joshi 2003; Belleflamme and Bloch 2004). In this simple representation of the simultaneous link-announcement game, the only pairwise stable network is the network in which each player is connected to both of her one-type matches in the population. This is pictured in the bottom right panel of Figure 4.2.

The unique pairwise Nash stable network in this simple setting seems reasonable on the
face. Each player's direct links provide positive utility. Each player is also indirectly connected to her opposite type, but indirect links are sufficiently discounted. It must be kept in mind, however, that the results are specific to the fact that there are only four players and that all four types are represented. The example can provide intuition about situations with larger populations in which types are equally distributed across agents, and forming links is costless. In these situations, we will find the unique pairwise Nash stable network to be the one in which all agents are connected to all of their same-type and one-type matches, and no direct links exist between opposite types. However, when types are not equally distributed, when link formation is costly, and when agents are not able to form links with any other agent in the society, expectations about the networks that should arise become quite difficult to describe analytically. To examine settings such as these I employ a series of numerical experiments.

4.3 Numerical Experiments

In the numerical experiments that follow, I consider an extension of the simultaneous link-announcement game presented in the previous section. Specifically, agents are initially distributed across geographic space and are able to observe and form links only with other agents within a specified distance band. After the initial formation of links according to the simultaneous link-announcement game, I allow for the possibility that agents’ types are influenced by the other agents to whom they are connected (Page 2007), and that this influence is a function of the utility received by others and of the strength of agents’ types. As the relationships persist over time, more successful partners are more likely to be influential than less successful partners. This influence can be thought of in terms of explicit imitation of success (e.g., Schlag 1998; Young 1998), or simply as a more subconscious process of success having greater social influence than failure. Based on changes in types, the potential utility of links, and costs for forming and severing links, agents make decisions in each time period.
about whether to reform their current set of links. As in the initial formation model, links
in each period result from a mutual desire by the actors involved, and the resulting network
confers utility on the actors based on their system of direct and indirect connections. Be-
cause agents are able only to observe other agents within a specified geographic distance, the
formation of direct links is characterized by a great deal of uncertainty about the indirect
connections that may result. Indirect connections with opposite types are therefore likely,
and they impose negative externalities on those indirect connections.

Formally, agents are characterized by a two-dimensional vector, \( Z_i = (x_i, y_i) \), representing
their location in two-dimensional geographic space, a two-dimensional vector of latent types,
\( \Sigma_i = (\sigma_{1i}, \sigma_{2i}) \), and a two-dimensional vector of observed types, \( \Theta_i = (\theta_{1i}, \theta_{2i}) \). In the
initial stage of the model, \( x_i, y_i, \sigma_{1i}, \sigma_{2i} \sim U[0, 1] \), and latent types lead to observed types
\( \theta_{1i}, \theta_{2i} \in \{0, 1\} \), such that if \( \sigma_{ki} \geq 0.5 \) then \( \theta_{ki} = 1 \), and if \( \sigma_{ki} < 0.5 \) then \( \theta_{ki} = 0 \), \( k = 1, 2 \).
Latent types are intended to capture the strength of agents’ types, such that latent type
values close to 0.5 represent weakly held observed types, and values close to 0 or 1 represent
strongly held types.

Agents have the ability to form links with a subset of other agents. Each agent’s potential
partners are those within a given distance, \( l \). Agents are located on a torus to correct for
the possibility of edge effects. Practically, this means that agents who would appear to be
on opposite edges of a two-dimensional grid are actually close to one another in geographic
distance.

I define \( p_i \) as an actor-specific parameter that determines the relative weight given to
temporal autoregression (the influence of an agent’s own state in the previous period), and
the relative weight given to network influences (the influence of an agent’s neighborhood
of direct connections), and is based on the utilities in each actor’s neighborhood of direct
connections, where utilities are again calculated according to Equation (4.1). Specifically,
\( p_i = \frac{u_i(g) - u_{N_i}^{\text{min}}(g)}{u_{N_i}^{\text{max}}(g) - u_{N_i}^{\text{min}}(g)} \), where \( u_i(g) \) is actor \( i \)’s utility, and \( u_{N_i}^{\text{min}}(g) \) and \( u_{N_i}^{\text{max}}(g) \) are the minimum
and maximum utilities, respectively, among actor \( i \)’s neighborhood of direct connections.
Therefore, if agent $i$ has the highest utility in her neighborhood, $p_i = 1$, and if she has the lowest, $p_i = 0.$

Let $t \in \{1, 2, \ldots \}$ denote time periods. Types are then updated according to

$$
\sigma_{ki,t+1} = p_{i,t} \sigma_{ki,t} + (1 - p_{i,t}) \sum_{j \neq i} \delta_{ij} \sigma_{kj,t} \frac{m - 1}{m - 1}
$$

for $k = 1, 2$, where $m$ is number of members of actor $i$’s direct-link neighborhood, and $\delta_{ij} = 1$ if $j$ is directly linked to $i$, and 0 otherwise. Agents with relatively higher utilities are therefore expected to remain relatively similar to their own latent type in the previous period, whereas agents with relatively lower utilities are expected to be more heavily influenced by the latent types in their neighborhood.

After types are updated in each period, new utilities are calculated, and actors make decisions about whether to reform their network connections. Actors observe the portion of their realized utility that is derived from each actor to which they are directly connected. Using this information, actors determine whether they wish to maintain their current connections. If the utility derived from a particular link is greater in magnitude than the cost of severing that link, then an actor wishes to maintain that relationship. For potential connections to whom the actor is not currently connected, actors make expected utility calculations just as in the initial network formation stage. With the information about current connections and the projections about potential new connections, all actors announce a new profile of potential links. If a connection appears in two actors’ profiles, a link between the two is formed or maintained. Otherwise, the link does not form or is severed.

**Results**

Using the model description in the previous section, I ran a series of simulations varying the costs of forming and severing links to derive conclusions from a wide variety of settings. Specifically, I focus on four basic conditions: (1) low-cost link formation, low-cost link sev-
Figure 4.3: Individual-level Utility Trajectories for Low-cost Link Formation and Low-cost Link Severance

Note: Lines are individual utility levels at each iteration.

Of particular interest are questions of whether players become locked into various levels of utility after some initial set of moves, and whether initial conditions can be used to predict the types of outcomes that are possible. As an illustration, Figure 4.3, presents individual-level utility trajectories for the five simulation runs of low-cost link formation and low-cost link severance. A brief examination of the figures indicates that a variety of outcomes are in fact possible depending upon the initial configuration of actors across geography and across types. In particular, we see that in each simulation run, actors become broken into particular regimes of utility levels. In three of the five simulations, a set of individuals breaks away from the pack, performing significantly better than the other agents in the society. There is always a small group of players that never moves away from a utility of zero–agents that fail
to form links for the entire time period analyzed. Then a series of agent blocs form early
and remain stable around several levels of positive utility. In the three simulations in which
a clear outlier group of high utility emerges, the groups of agents at the intermediate levels
of utility top off at roughly between one-third and one-half of the utility level of the outlier
group. Agents receiving negative utility seems to be a rare phenomenon, and looks to be
self-correcting. That is, agents who receive negative utility tend to quickly and universally
move back into the realm of non-negative utility.

To illustrate that these general conclusions are not specific to the situation of low-cost
link formation and low-cost link severance, Figure 4.4 presents the individual-level utility
trajectories for high-cost link formation and high-cost link severance. The general description
of the results for this condition is quite similar to that of the previous condition. We once
again see that various regimes of intermediate levels of positive utility form, and on occasion
we observe a break-out group that performs significantly better than the rest of the society.
Actual levels of utility are quite similar to the previous condition, if slightly higher on average. The third panel of Figure 4.4 demonstrates an anomaly for this particular set of simulations. In particular, at time period 8, the group of agents with the highest level of utility, along with a subset of an intermediate group, are able to jump to a coalescence around a higher level of utility. Across the 20 simulations performed for this analysis, this was the latest that the ultimate high level of utility was reached, and represents one of the only times that such a jump occurred after a group had seemingly been in a steady state for several periods.

The individual-level trajectories for low-cost link formation, high-cost link severance, and high-cost link formation, low-cost link severance are not presented here due to space constraints. But suffice to say that the general description of the results do not change significantly depending upon the condition. The highest individual-levels of utility are experienced when severing links carries high costs, regardless of whether the formation of links carries a high or low cost. This would seem to suggest that the myopia of the agents, when combined with low-cost of link severance, leads to agents severing relationships predicted to be costly in the short-term that could potentially carry greater benefits in the long-term. Additionally, agents do not become trapped in negative levels of utility because though the severance of links carries a cost, it never carries a greater cost than the negative utility experienced. Agents are almost always able to break free from costly relationships in a matter of one period. Overall, the most striking and consistent feature of the sets of individual-level trajectories is that but for a few individuals, agents quickly lock into a particular level of utility resulting from a set of mutually beneficial network connections. Short-term considerations and structural constraints insure that agents would be wise not to break their existing links, but they also find themselves unable to form links that might continue to move the utility trajectories of themselves and their groups to a higher level.

The visualization of the utility trajectories brings up interesting questions about whether the random initial conditions lock players into states from which they are unable to break out. In the model, there is nothing particularly advantageous for an agent in being one
Table 4.1: ANOVA Models of Individual Utility in the Final Period, Conditioning on Initial Types

<table>
<thead>
<tr>
<th>Formation Cost</th>
<th>Severance Cost</th>
<th>Simulation Number</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>1</td>
<td>4.183</td>
<td>0.008</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>2</td>
<td>6.830</td>
<td>0.000</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>3</td>
<td>4.861</td>
<td>0.003</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>4</td>
<td>2.545</td>
<td>0.061</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>5</td>
<td>2.466</td>
<td>0.067</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>1</td>
<td>5.873</td>
<td>0.001</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>2</td>
<td>1.414</td>
<td>0.244</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>3</td>
<td>3.656</td>
<td>0.015</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>4</td>
<td>8.641</td>
<td>0.000</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>5</td>
<td>7.289</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>1</td>
<td>2.355</td>
<td>0.077</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>2</td>
<td>7.419</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>3</td>
<td>6.850</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>4</td>
<td>7.683</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>5</td>
<td>1.852</td>
<td>0.143</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>1</td>
<td>1.956</td>
<td>0.126</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>2</td>
<td>2.793</td>
<td>0.044</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>3</td>
<td>2.594</td>
<td>0.057</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>4</td>
<td>4.150</td>
<td>0.008</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>5</td>
<td>3.753</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Note: F-statistics are derived from a one-way analysis of variance, conditioning on types in the initial period. Response variable is utility in the final period. p-values are based on $F(3, 96)$.

type or another, except as this relates to the subpopulation of other agents who are in close proximity. If the initial distribution of types across geographic space does not confer benefits upon one group or another, then we should not expect agents’ types in the initial stage to be significant predictors of agents’ utilities in the final stage. As a first step toward examining this relationship, Table 4.1 presents the results of hypothesis tests from one-way analysis of variance models for each simulation, in which the response variable is utility in the final period, and the predictor is initial type. Sums of squares are omitted for presentational purposes. The p-values indicate that except for a few exceptional cases, initial type is at least a marginally statistically significant predictor of the ultimate utilities. In short, far more
often than not, the mean values of final utilities for the four different initial types differ by more than should be expected by chance. These results indicate not that any particular group has an advantage in general, but rather that for any given set of initial conditions, an agent’s type is consequential for the long-run level of utility that can be expected.

Looking more closely at the individual-level dynamics of the model, we find that in any given simulation, the type updating rule based on the relative utilities in an agent’s neighborhood leads to the organization of agents into one dominant group. This is a universal result across the series of simulations employed here. After beginning with a relatively balanced distribution of agents across types, agents form connections, utilities are received and observed, and types are updated. The continuation of this process leads to the relatively successful types in the society having a great deal of influence over the types of their neighbors, eventually having a sort of cascading effect.

Figure 4.5 illustrates this process for one simulation run of the low-cost formation, low-cost severance condition. In each row, the first panel indicates the time period, the second panel displays the utilities received from the network of links in that time period, and the third panel displays agents’ degree centrality (a count of the agents’ numbers of direct links). Each of the latter two panels are broken down by agent type. Examining the change from $t = 0$ to $t = 1$, the number of type changes is immediately dramatic. Types (1,0) and (0,1) in particular lose a significant portion of their original members, and the number of Type (0,0) agents grows substantially. By $t = 5$, the change is even more dramatic, and by the final period, no agents are left of Type (0,1). Examining the upper left and lower left panels of the utility calculations at $t = 1$ gives a hint to how this may have occurred. Type (0,1) agents would typically desire to form connections with Type (1,1) agents and Type (0,0) agents (in addition to their own type matches). The utility calculations after the initial period place Type (0,0) agents at the highest utility levels, on average, whereas Type (1,1) agents reside at the lowest utility levels. Type (0,1) agents with connections to both other types would then be more heavily influenced by the Type (0,0) agents. This process continues until Type
Figure 4.5: Utilities and Centrality by Agent Type Over Time, Low-cost Formation, Low-cost Severance Condition.

$t = 0$

$t = 1$

$t = 5$

$t = 10$

Note: Rows indicate individual utilities and degree centrality at specified time step (iteration) for each type.
agents are completely absent.

In this example, as in other simulations not presented, the final dominance of one type does not indicate that the members of that type perform universally better than members of the minority types. Members of the dominant group do receive higher utility on average, but they also become stuck in low-utility trajectories.

The degree centrality panels of Figure 4.5 indicate that in the first several time periods, the relationship between the number of links and the level of utility is not immediately apparent. The relationship only begins to emerge as the dominant type emerges. Because by the final period, the dominant type makes up an overwhelming proportion of the population, the relationship between the number of connections and utility received remains unclear. And it is important to keep in mind at this point that this example is representative of the outcomes that obtained across simulations.

Given that random initial conditions seem to be consequential for individual utility trajectories, it stands to reason that they may also be consequential for the social welfare obtained by the society as a whole. In the model employed here, agents are not dividing utility across a fixed pie, but rather returns are allowed to increase indefinitely until all types are the same, and the network is fully connected subject to the imposed geographic constraints. Given a random array of initial agents, therefore, the set of possible outcomes for a society would appear to be wide ranging. This is, in fact, what is found here. Figure 4.6 presents the trajectories of social utility for each of the simulations discussed. I operationalize social utility as simply the sum total of utility across agents at each time period. For each combination of costs of link formation and link severance, the total of five social utility trajectories are presented. The figure illustrates that even for a very small number of simulations across the various conditions, a wide variety of social outcomes obtain. The highest stable trajectory of social utility occurred in the low-cost formation, low-cost severance condition, and the lowest stable trajectory occurred in the high-cost formation, low-cost severance condition. But the relationship between the costs of formation and severance and social utility remain

108
unclear. The relatively large vertical separation between trajectories in each panel indicates that if initial conditions are randomly given, the social outcomes that can be expected will be very difficult to predict.
The simulation results suggest a clear avenue for future inquiry: studying the specific sets of initial conditions that can be expected to lead to particular individual and social outcomes. The analyses of model results presented here only scratch the surface of the possible questions that can be addressed with this type of model. Because initial conditions are randomly assigned in these simulations, the wide variety of outcomes in any one condition suggest that the particular initial values are likely to be consequential.

4.4 Discussion and Conclusion

Random network formation has a long pedigree in mathematics and the physical sciences, and the amount of scholarship on models that account for individual choices in social network formation are growing rapidly. However, though political science has long been concerned with the use of social network concepts to explain various aspects of political behavior, the discipline is a relative latecomer to the specific study of strategic network formation. Models of the strategic formation of social and political relationships would seem to be a natural fit for political scientists interested in such substantive areas as the formation of political coalitions, the formation of military alliances, and the formation of trade relationships, particularly given the popularity and importance of these topics. In this chapter, I have attempted to take a small step toward a greater understanding of strategic network formation in areas of interest to political scientists.

Specifically, the chapter is concerned with situations in which actors have incentives to form relationships with similar others, but in which similarity is defined over multiple dimensions. The study of models of decision and choice in multi-dimensional issue space has generated a great deal of work in political science. But such questions have been studied very little in specifically network-related terms. Likewise, studies of strategic network formation have given a great deal of attention to the formation of social ties between individuals with similar characteristics and goals, but little work in this area has explicitly considered sim-
ilarity in multiple dimensions. By bridging these two areas of scholarly inquiry, both have
great potential to benefit.

I began with a simple model of network formation via a non-cooperative game of si-
multaneous announcement of links by players. The simultaneous link formation game was
developed elsewhere (Myerson 1977), and has been studied extensively (see, e.g., Jackson
2008). But I added the feature of agents who are heterogeneous over a two-dimensional
space of types. A simple example demonstrated that under a very simple configuration, a
large set of equilibrium predictions obtain using the naïve solution concept of Nash stability,
but a unique prediction obtains under the refinement of pairwise Nash stability (Jackson
and Wolinsky 1996). I stress, however, that the existence of a unique pairwise Nash stable
network depends upon the simplicity of the example. Situations characterized by larger
numbers of players and varying distributions of types lead to a multiplicity of predictions
about the stable networks that can be expected to emerge.

Using the simultaneous link-announcement game as a basis, I then employed numerical
simulations to examine the dynamics of network change when agents are myopic and have
a limited ability to form ties with others, when the forming and severing of links is costly,
when the initial distribution of types in two-dimensional space is randomly assigned, and
when agents are allowed to influence and be influenced by others’ types as a function of
success. A wide array of results obtain at both the individual and societal levels, making
general conclusions about the relationships between various conditions difficult. Across
simulations, individuals and the society as a whole tended to lock into particular levels of
utility fairly early in the process and remain there for the duration. Additionally, there was
a strong tendency for the influence of neighbors’ types as a function of success to lead to one
type overwhelmingly dominating the final population of agents. The variation in possible
outcomes in terms of both individual and social utility suggest that the randomly assigned
initial conditions may be highly consequential.

This naturally will lead to future inquiry about the specific initial conditions that lead to
particular types of outcomes. From this relatively basic model, it is possible to examine a very large set of possible initial circumstances. The representation presented here, therefore, has opened the door to a fruitful line of future research.
Chapter 5
Conclusion

In the contemporary industrialized world, norms of appropriate social behavior dictate that on questions of race, ethnicity, gender, religion, and so on, support for egalitarian principles is the only acceptable mode of mainstream social and political behavior. And not surprisingly, citizens, especially those with greater awareness about and knowledge of the political system, express high levels of support for egalitarian principles across issue domains. But it remains a constant struggle for empirical researchers to determine the truth and falsehoods underlying those responses.

The main source of the trouble is the understanding that humans have a strong desire to not appear to be deviant from those around them, especially when those around them are able to witness their behavior. It is that desire that leads to the thought that in a public setting, when asked to publicly express a view on a sensitive social issue like racial equality, some substantial portion of citizens are likely to falsify their true preferences in favor of the actual or perceived distribution of preferences in the group. To the extent that many individuals feel this way, widespread preference falsification might be expected to ensue.

In three essays, this dissertation explored various aspects of the relationship between interpersonal interactions and the behaviors that individuals are willing to exhibit in public settings. The first essay did so by examining behavior in situations in which actors were compelled to exhibit some behavior in non-self-selected settings. The model was quite abstract, and could potentially be applied to a variety of different settings in which individuals might experience some level of tension between intrinsic and extrinsic preferences. But the model was interpreted in the context of changing norms of racial rhetoric witnessed in the
United States in the latter part of the twentieth century. In the second essay, I explored a series of laboratory experiments on social influence and group pressure to conform. These studies showed general support for what I termed the *Social Dissonance Hypothesis*, which essentially states that individuals will be more likely to alter the behavior that they previously exhibited in private when that behavior runs counter to prevailing social norms. More generally, the experiments demonstrated an asymmetric effect of social pressure in which those subjects who gave inegalitarian responses were more likely than subjects who initially gave egalitarian responses to change their behavior in the face of social pressure to do so.

The third essay moved away from situations in which individuals were exogenously assigned to networks of relationships and examined a setting in which individuals make decisions to form relationships and networked populations are allowed to develop endogenously. Using random initial parameter values in this setting leads to a wide variety of possible outcomes at both the individual and societal levels.

These essays provide the foundation for future work on social influence. Future research will have to more closely integrate the experimental and simulation portions in order to provide empirical support to a wide variety of long, over-time processes.
Appendix A

Experimental Instructions

Here I present the instructions read to participants and the laboratory materials used in the experiments described in Chapter 3.

A.1 The Looming Social Pressure Experiment

After all participants in a given session had gathered in a common waiting room and filled out a consent form to participate, the researcher recited to the group the following set of instructions:

Today you are participating in a study on group and individual decision making. The first part of the session will take place on these computer terminals. So you all can each choose an open terminal. You should go into the room and shut the door. I am going to come around to you one-by-one and give you instructions. It may take a few minutes for me to get to you. You may read or use the computer while you wait for me. But you will find that there is a web browser open and minimized on your computer. Don’t touch that until I tell you to do so. So go ahead and choose a computer. What you are going to do has to be coordinated with the other group of participants, so I have to go make sure that everything is ready in the other room.

As the participants filed into the computer terminals and closed the door, the researcher left the room. After a period of one to two minutes, the researcher returned, walked into each terminal, one-by-one, and recited the following set of instructions:

Part of what we are interested in here today is what happens in group decision-making situations when people are able to learn a little bit about one another before they get together in groups to do something.

So there is a group of people in another room who have filled out a questionnaire with a series of public opinion questions. I have selected six of the questions at random, and
I am going to give you a computer printout of four people’s answers. You have been assigned to be in a group with these four people for something that happens later.

The people participating in the experiment today have been assigned different roles. For the purposes of this study, the only difference between the people in this room and the people in the other room is that the people in the other room filled out the questionnaire without knowing anything about the group part of the study. You have a different role in that you will be allowed to learn a little bit about your fellow group members before you do anything for the study.

[Researcher points to the top page of a stack of computer printouts (refer to Figure 3.1).] On the printout, you’ll see a survey question along with the set of responses, and then you’ll see an X marked next to the other person’s answers. Other people are identified only by a participant number in the upper right corner of the printout. And for this study, you are participant number 11.

So I will leave these printouts of the other people’s answers with you for a couple of minutes. Just look them over and I’ll be right back after I give these instructions to a few other people. [Researcher hands stack of printouts to subject, leaves the room, and closes the door.]

The researcher then left the room to give the instructions to another subject. After all subjects were given the above set of instructions, the researcher began again with the first subject, walked into each terminal, and recited the following set of instructions:

So now you are going to fill out the same questionnaire that your fellow group members already completed. When you’re finished, there will be a brief break for you, and at that time I am going to give your fellow group members a similar printout about you. After that, you will go through a group task that takes just a few minutes.

The questionnaire has about 60 questions on it, and it generally takes people about 25 to 30 minutes to complete. You’ll answer this, and then there will be a short break before the group task. And the software should provide all of the instructions that you need as you navigate through the questionnaire. Just read everything very carefully and go ahead and get started.

A.2 The Real-Time Social Pressure Experiment

Two experimental conditions were administered via instructions given to the group of subjects prior to the beginning of the computer portion of the study. In one condition, subjects were told that they would meet in a face-to-face setting following the computer portion of the study (Condition 2), and in another they were told that they would not engage in any
Welcome.

Today you are going to be participants in a broader research study on individual and group decision making. What you do today will be used to compare across a variety of situations in which individuals make decisions in individual settings, in group settings over a computer network, and in face-to-face group settings.

Today, you will not be involved in any face-to-face tasks. That is, you will engage in some tasks as individuals, and then some more tasks in groups over the computer network, and then you’ll be finished. Today, you will do some of all three of these. That is, you will engage in some tasks as individuals, some more tasks in groups over the computer network, and then some more tasks in face-to-face groups.

The study today will take place on these computer terminals. The first part of the study today will take place on these computer terminals. You will first go through some questions as an individual. Then following the individual questioning, you will be assigned to a group of other people participating in the study today, and go through some similar questions as a group. The software will assign people to groups based on when people finish the individual portion, so you may have to wait a few minutes as people finish the first part. Then after the group portion finishes, you’ll go through a second short round of questioning as an individual. Then the software will instruct you that you are finished and that you can come out into the common area. Then you will be finished with the computer portion of the study, and the software will instruct you to come out in the common area and meet your computer-networked group face to face. At that point you will go through another very brief set of tasks as a group. And then you’ll be finished. In total, this generally takes people about 40-50 minutes to go through.

Does anyone have any questions? [Q&A]

So to reiterate, you’re going to sit at a computer terminal, then go through a set of individual tasks, a set of computer-networked group tasks, then another set of individual tasks. Then the software will instruct you that you are finished. Then you’ll come out of the computer terminal and meet the same group face to face, and go through a brief set of tasks together.

So go ahead and choose a computer terminal. You will find instructions already on the screen. Just read everything very carefully, and the software should lead you through the whole process on the computer and let you know when to come out. Go ahead and begin and just read everything very carefully. And I am going to come around and shut your doors.
Appendix B

Summary Statistics From Experiments

The following tables present summary statistics for the subjects who participated in the experiments described in Chapter 3.
Table B.1: Summary Frequency Distributions of Categorical Variables for Subjects in the Looming Social Pressure Experiment

<table>
<thead>
<tr>
<th>Race:</th>
<th>White</th>
<th>Black</th>
<th>Asian</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>112</td>
<td>75.68</td>
<td>8</td>
<td>5.41</td>
<td>19</td>
</tr>
<tr>
<td>Treatment</td>
<td>40</td>
<td>75.47</td>
<td>3</td>
<td>5.66</td>
<td>5</td>
</tr>
<tr>
<td>Gender:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>40</td>
<td>27.21</td>
<td>107</td>
<td>72.79</td>
<td>147</td>
</tr>
<tr>
<td>Male</td>
<td>25</td>
<td>47.17</td>
<td>28</td>
<td>52.83</td>
<td>53</td>
</tr>
<tr>
<td>Religion:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chris.</td>
<td>99</td>
<td>67.81</td>
<td>47</td>
<td>32.19</td>
<td>146</td>
</tr>
<tr>
<td>Non-Christ.</td>
<td>38</td>
<td>71.70</td>
<td>15</td>
<td>28.30</td>
<td>53</td>
</tr>
<tr>
<td>Party ID:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dem.</td>
<td>56</td>
<td>38.10</td>
<td>26</td>
<td>17.69</td>
<td>41</td>
</tr>
<tr>
<td>Ind.</td>
<td>29</td>
<td>54.72</td>
<td>8</td>
<td>15.09</td>
<td>9</td>
</tr>
<tr>
<td>Rep.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oth.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Yr.</td>
<td>60</td>
<td>40.54</td>
<td>48</td>
<td>32.43</td>
<td>27</td>
</tr>
<tr>
<td>2nd Yr.</td>
<td>36</td>
<td>67.93</td>
<td>10</td>
<td>18.87</td>
<td>4</td>
</tr>
<tr>
<td>3rd Yr.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th Yr.+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All measures are self reports. For each category of each variable, the first column gives the frequency, and the second column gives the percentage.
Table B.2: Summary Frequency Distributions of Categorical Variables for Subjects in the Real-Time Social Pressure Experiment

<table>
<thead>
<tr>
<th>Race:</th>
<th>White</th>
<th>Black</th>
<th>Asian</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. 1</td>
<td>26 72.22</td>
<td>2 5.56</td>
<td>6 16.67</td>
<td>2 5.56</td>
<td>36 100</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>29 76.32</td>
<td>1 2.63</td>
<td>7 18.42</td>
<td>1 2.63</td>
<td>38 100</td>
</tr>
<tr>
<td>Cond. 3</td>
<td>49 64.47</td>
<td>7 9.21</td>
<td>11 14.47</td>
<td>9 11.84</td>
<td>76 100</td>
</tr>
<tr>
<td>Gender:</td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 1</td>
<td>19 52.78</td>
<td>17 47.22</td>
<td>36 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 2</td>
<td>21 55.26</td>
<td>17 44.74</td>
<td>38 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 3</td>
<td>36 47.37</td>
<td>40 52.63</td>
<td>76 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religion:</td>
<td>Chris.</td>
<td>Non-Christ.</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 1</td>
<td>22 61.11</td>
<td>14 38.89</td>
<td>36 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 2</td>
<td>18 47.37</td>
<td>20 52.63</td>
<td>38 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 3</td>
<td>43 56.58</td>
<td>33 43.42</td>
<td>76 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cond. 1</td>
<td>16 44.44</td>
<td>8 22.22</td>
<td>8 22.22</td>
<td>4 11.11</td>
<td>36 100</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>25 65.79</td>
<td>5 13.16</td>
<td>7 18.42</td>
<td>1 2.63</td>
<td>38 100</td>
</tr>
<tr>
<td>Cond. 3</td>
<td>32 42.11</td>
<td>9 11.84</td>
<td>26 34.21</td>
<td>9 11.84</td>
<td>76 100</td>
</tr>
<tr>
<td>Ideology:</td>
<td>Lib.</td>
<td>Mod.</td>
<td>Con.</td>
<td>Oth.</td>
<td>Total</td>
</tr>
<tr>
<td>Cond. 1</td>
<td>15 41.67</td>
<td>14 38.89</td>
<td>7 19.44</td>
<td>0 0.00</td>
<td>36 100</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>21 55.26</td>
<td>12 31.58</td>
<td>5 13.16</td>
<td>0 0.00</td>
<td>38 100</td>
</tr>
<tr>
<td>Cond. 3</td>
<td>34 44.74</td>
<td>16 21.05</td>
<td>26 34.21</td>
<td>0 0.00</td>
<td>76 100</td>
</tr>
<tr>
<td>Education:</td>
<td>1st Yr.</td>
<td>2nd Yr.</td>
<td>3rd Yr.</td>
<td>4th Yr.+</td>
<td>Total</td>
</tr>
<tr>
<td>Cond. 1</td>
<td>11 30.56</td>
<td>13 36.11</td>
<td>6 16.67</td>
<td>6 16.67</td>
<td>36 100</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>8 21.05</td>
<td>12 31.58</td>
<td>7 18.42</td>
<td>11 28.95</td>
<td>38 100</td>
</tr>
<tr>
<td>Cond. 3</td>
<td>4 5.26</td>
<td>15 19.74</td>
<td>30 39.47</td>
<td>27 35.53</td>
<td>76 100</td>
</tr>
</tbody>
</table>

Note: All measures are self reports. For each category of each variable, the first column gives the frequency, and the second column gives the percentage.

Table B.3: Summary Distributions of Group Size Variable for Subjects in the Real-Time Social Pressure Experiment

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Med.</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Mean</th>
<th>S.D.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. 1</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7.39</td>
<td>1.54</td>
<td>36</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>6.87</td>
<td>1.19</td>
<td>38</td>
</tr>
<tr>
<td>Cond. 3</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>5.67</td>
<td>1.11</td>
<td>76</td>
</tr>
</tbody>
</table>

Note: Group size is the number of subjects who showed up on a given study date and waited together in the common waiting room. See text in Chapter 3.
Appendix C

Experimental Results

To test hypotheses about individual propensities to give the egalitarian and inegalitarian responses under various circumstances, many of the analyses in Chapter 3 combined response options for questions with more than two choices. This Appendix illustrates graphically the raw uncombined data from the Looming Social Pressure Experiment for questions with more than two response options. These are presented in Figure C.1.
Q3: Do you favor special security checks at airports for Muslims?

Q4: Would you worry a lot, a moderate amount, a little, or not at all about having a Muslim as a neighbor?

Q5: In your best estimation, what proportion of Muslims in the world are anti-American?

Q8: Do you think that most, many, only some, or very few Muslims living in the United States today support the use of violence by Muslims against non-Muslims?

Figure continued on next page.
Q9: How worried are you, if at all, about radicals within the U.S. Muslim community?

Q12: Do you think that most, many, only some, or very few Muslims living in the United States today are peaceable and do not condone violence?

Q17: Do you think that most, many, only some, or very few Muslims living in the United States today support the goals of Al Qaeda and the fundamentalists?

Q18: Do you favor laws that prohibit discrimination against Muslims?

Note: Bar heights are percentages within categories. Numbers above bars are the actual number of respondents in each category.


