Essays on Networks and Social Interaction

by

Dean Knox

B.S. Nuclear Engineering, University of Illinois at Urbana-Champaign

Submitted to the Department of Political Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Political Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Signature redacted

Author .....
Department of Political Science
11 May 2017

Certified by ..... Signature redacted
Teppei Yamamoto
Alfred Henry and Jean Morrison Hayes Career Development Associate Professor
of Political Science
Thesis Supervisor

Accepted by .... Signature redacted .....
Ben Ross Schneider
Ford International Professor of Political Science
Chair, Graduate Program Committee

Essays on Networks and Social Interaction

by

Dean Knox

Submitted to the Department of Political Science
on 11 May 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Political Science

Abstract

This dissertation presents new models and experimental designs for understanding network behavior and social interaction. The first paper develops a model for a new kind of data, "path data," that represents the sequential decisions made by actors navigating social, geographic, and other kinds of networks. The model is validated in a benchmark test, then used to measure sectarian influences in the ways that Sunni and Shia individuals navigate the streets of Baghdad in a smartphone-based field activity. The second paper uses a novel experimental design to examine the social network search patterns employed by both sects. Using smartphone and self-reported data, the paper shows that differing search strategies result in differential access to public services in Baghdad. The third paper presents a new model for measuring rhetorical style and other modes of speech in political deliberation. The model is validated in a benchmark test of conflictual speech in political debates.

Thesis Supervisor: Teppei Yamamoto
Title: Alfred Henry and Jean Morrison Hayes Career Development Associate Professor of Political Science

Acknowledgements

Over the course of graduate school I have accumulated far too many debts to repay. I owe more than I can express to my teachers and mentors: Teppei Yamamoto, Danny Hidalgo, Jens Hainmueller, and Chad Hazlett, for showing me the ways of political methodology; Adam Berinsky, for the secrets to success (among other things, "brighter lights and a harder chair"); and Fotini Christia, for everything I know about implementing projects in the field. For most of my good ideas, I am indebted to Christopher Lucas, Dan de Kadt, Connor Huff, Horacio Larreguy, and Ben Morse. For emotional support, I am thankful to Jolie Lee, George Howell Coffee, and Aeronaut Brewing Company. I gratefully acknowledge financial support from the MIT Department of Political Science, Capital One Financial Corporation under the 0% introductory APR program, and the National Science Foundation Graduate Research Fellowship under Grant No. 1122374.

Contents

List of Tables

List of Figures

Introduction

This dissertation leverages recent advances in measurement technology and computation to develop new models and experimental designs for understanding network behavior and social interaction. Where past work has relied on simplifying assumptions about human behavior—independence between observations, for example, or the reduction of human speech to a mere "bag of words"—computational advances have made it possible to build increasingly realistic models of behavior. At the same time, the fine-grained data needed for such models, including call, location, and audio data, is rapidly becoming accessible through smartphones and other forms of measurement. In this dissertation, I use a combination of new statistical methods and technology-intensive original data collection to examine a range of topics in political science. The first paper develops a model for an increasingly prevalent kind of data, "path data," that represents the sequential decisions made by actors navigating social, geographic, and other kinds of networks. The model is validated in a benchmark test, then used to measure sectarian influences in the ways that Sunni and Shia individuals navigate the streets of Baghdad in a smartphone-based field activity. The second paper uses a novel experimental design to examine the social network search patterns employed by both sects. Using smartphone and self-reported data, the paper shows that differing search strategies result in differential access to public services in Baghdad. The third paper presents a new model for measuring rhetorical style and other modes of speech in political deliberation. The model is validated in a benchmark test of conflictual speech in political debates.

In my first paper, I examine path data, a form of behavioral data that describes steps taken to get

from point A to B. This data offers researchers the opportunity to test theories about network navigation, such as whether individuals avoid out-group neighborhoods in their daily walking routes, or how voters search the Internet for political information. However, the sequential decision-making process in path data violates the underlying assumptions of existing models, which assume some form of conditional independence between observations. I propose a new random-path model (RPM) that explicitly captures this pathwise dependence, develop an estimation procedure, and demonstrate its properties. The RPM builds on a random-walk model by incorporating a realistic but difficult constraint, that actors avoid repetitive loops. The model is validated in an analysis of the U.S. Interstate planning process, where existing approaches fail to recover a known qualitative benchmark. Finally, the RPM is used to test two competing explanations of Baghdad's recent segregation. Using smartphone-based behavioral data on Sunni and Shia participants in a field activity, I show that this segregation was not solely driven by need-based sorting in response to political violence. Instead, walking paths reveal a taste-based aversion to out-group neighborhoods. These results suggest that societal preferences have shifted in a way that makes re-integration unlikely.

In my second paper, joint with Fotini Christia and Jaffar Al-Rikabi, I employ new network measurement techniques to study how individuals draw on their social contacts to gain access to public goods and services. Using a pool of over 300 participants from paired Sunni and Shia neighborhoods in the highly sectarian context of contemporary Iraq, we conduct a small-world network experiment in which participants are randomly assigned to obtain information about local government services in Sunni- or Shia-dominated target areas. We trace how participants draw on their social networks and show that segregated social networks and different patterns of network search result in differential levels of access to services between groups. Our results shed light on the widely acknowledged but undertheorized relationship between ethnic fractionalization and lower availability of public goods and services. Contrary to expectations, we find that the politically dominant majority Shia group is substantially less able to access public services than the minority Sunni group. They pursue an inefficient network search strategy that relies on lower-quality

contacts, and are less able to leverage their social ties into costly assistance. The minority group appears to have developed better strategies for obtaining resources to which it would otherwise be denied access.

My final paper presents a new model for studying political communication, including the way in which people engage in political deliberation. While most work in this area focuses on the textual component of communication, audio data contains information, such as the speaker's emotion, that is particularly important in group settings in which individuals seek to persuade one another. However, in part due to the relative difficulty of processing and analyzing audio data, research has neglected this dimension of communication. To resolve this, I develop a general approach for classifying speech into discrete researcher-specified "modes," which can represent emotional content, speakers, or any other set of labels with distinct audio profiles. This model is the first of its type used in political science and provides three useful innovations over existing methods in computer science: (1) it develops a hierarchical hidden Markov model (HMM) that nests audio HMMs within a higher-level HMM that learns the flow between modes of speech, modeling the ebb and flow of speech to borrow strength from nearby utterances; (2) it incorporates ridge-like regularization that allows utilization of many more features than existing approaches in the lower-level audio HMM; and (3) it employs a principled approach to uncertainty via Bayesian bootstrap. I validate the model with a benchmark study of televised debates.

Paper 1: A Model for Path Data with Application to Sectarian Movement in Baghdad

Introduction

Goals are rarely accomplished in one fell swoop. Instead, actors work one step at a time, making a series of smaller decisions that ultimately lead to the intended destination. This sequential process forms a *path*. As a general phenomenon, paths are ubiquitous, and social scientists have developed a variety of theories about them. For example, sociologists are interested in how people search their social networks through a chain of intermediaries and referrals (Milgram, 1967; Killworth and Bernard, 1978; Dodds et al., 2003a), and development economists seek to evaluate the impact of transportation infrastructure, such as highway and railroad routes, on long-term growth (Fogel, 1962, 1964; Aschauer, 1989; Banerjee et al., 2012). In these theories, paths feature as both dependent and independent variables—just like any other form of data. Yet paths are rarely measured and studied as such when evaluating path-related theories. In this, they differ from other data types, such as event counts, where a broad consensus has emerged that theories should be tested by collecting appropriate forms of data and using appropriate models to analyze them (King, 1989). The gap is largely due to the absence of statistical models for path data.

In this paper, I propose and demonstrate the properties of a new model for path outcomes. Paths are represented as movement on a network: A forward-looking decision-maker chooses from a limited set of options, or neighboring nodes, each opening a new and different set of possible next steps. For example, when building a road to connect two cities, a planner must go through an inner suburb of the first, then an adjacent outer suburb, and so on until eventually arriving at the destination. The random-path model (RPM) allows researchers to learn about preferences based on this movement. The intuition underlying the model is simple. If roads commonly go out of their way to avoid mountainous regions, then the "cost" of elevation is larger than deviating from the shortest route.

The RPM builds on the random-walk model, which assumes that each decision is independent. Random walks are powerful and well-studied models that can explain how long-term patterns emerge from a series of small decisions. They have found application across a wide range of fields, including the random collisions of atoms, foraging patterns of animals, and stock-market fluctuations. In the random-walk model, a walker myopically takes steps based on their short-term attractiveness, until eventually reaching the goal by sheer luck. As an example, consider an individual navigating between diagonal corners of a 5×5 street grid. Under this model, the walker has a 99.2% chance of wandering back to a previously visited intersection at some point before arriving at the destination. While the random walk might be appropriate model of human decision-making in some circumstances (as a case in point, this particular problem is often called the "drunkard's walk"), I argue that actors generally plan ahead and work more efficiently.

To better model purposeful decision-making, the RPM adds a conditioning stage before starting, in which all candidate routes that contain repetitive loops are discarded. It can be shown that this is equivalent to a forward- and backward-looking walker that navigates toward the goal and avoids previously visited areas. In the same street-grid example, under the RPM, the walker pushes onward to the destination rather than wandering aimlessly in circles. Moreover, under this simple scenario, a walker following the RPM takes a shortest path—walking eight blocks, e.g. along the

diagonal—nearly three-quarters of the time. (Additional covariates can be incorporated to account for the walker's sense of direction or familiarity with the area.) A model that incorporates some form of long-term planning is better-suited for most applications in political science.

However, this additional constraint poses a challenge in that it makes the resulting probability mass function intractable. I demonstrate that despite this challenge, the RPM can still be estimated. To this end, I develop and assess numerical algorithms for sampling random paths, evaluating a simulated RPM likelihood function, and efficiently implementing Metropolis-Hastings sampling from the posterior distribution. A permutation-based extension of the model can be used to estimate the causal effects of path-assigned treatments.

In the remainder of this paper, I first discuss examples of path data in political science. Section 3.3 formally defines the model, outlines the estimation procedure, and contrasts the random path model with existing approaches. Section 1.4 validates the RPM with a study of the U.S. Interstate Highway System, where official priorities are known from detailed qualitative planning documents. I show that RPM estimates correspond closely to this benchmark, whereas existing spatial models produce substantively and statistically differing results.

Finally, section 1.5 provides a motivating empirical application. In Baghdad, a long history of Sunni–Shia coexistence and integration was overturned by a wave of ethnic cleansing in 2006–2007. The long-run effects of this conflict depend on whether recent segregation was only driven by need-based sorting—to avoid violence (Morrison, 1993)—or whether it also led to taste-based sorting (Schelling, 1969, 1971). While these drivers of out-group aversion often go hand-in-hand, they have very different implications for Baghdad's future development: If residents only fled mixed neighborhoods to avoid ethnic violence, then game-theoretic models (Young, 1998; Zhang, 2004a,b) predict gradual post-conflict reintegration. If conflict led to newfound taste-based aversion, however, ethnic attitudes can persist far beyond the end of conflict and make segregation difficult to escape.

To test these competing hypotheses, I analyze behavioral data from a Baghdad field study

by Christia and Knox (nd). Participants' movement in a treasure-hunt-type activity reveals that taste-based aversion is a significant factor in participants' daily movement. These results suggest that recent conflict has shifted societal preferences in a way that makes reintegration unlikely. Section 3.5 concludes with limitations and areas for future work.

Literature Review

Path Data in Political Science

Political scientists have theorized about the causes and effects of paths in a variety of settings. In this section, I offer examples that include the geographic paths of connective infrastructure, the social paths traced by individuals as they search their social networks, and the aggregate-flow paths of people and goods. Other path-related theories appear in economics, urban planning, operations research, and engineering. While the measurement of path *data* is relatively new, it is growing increasingly common, offering new opportunities for research but also presenting new challenges for statistical analysis.

An illustrative example of a path in political science is the highway, which connects cities through a series of intermediate counties. Researchers are often interested in evaluating the role of various political factors in transportation spending, including institutions, pork-barrel spending, and ethnic patronage (e.g. Lee, 2000; Burgess et al., 2015). Other examples from connective infrastructure include the impact of patronage in electrical grid construction (Briggs, 2012) and governance on oil pipelines (Carmody, 2009). Researchers have increasingly recognized that standard models fail to account for the dependence between units that arises in these contexts. That is, whether a county is connected to the highway system depends not only on *whether* neighboring counties are also connected, as in standard spatial models, but also *which* neighbors are connected.[1]

[1]For example, consider an east-west highway on a square grid. For county i to be connected, two conditions are required: (1) Its neighbors to the east and west must be connected, so that i can serve as the missing link; and (2)

Given that highways are typically designed to connect major metropolitan areas, a rural legislator's success in securing transportation spending is perhaps as much about diverting the course of already-planned segments as it is fabricating entirely new projects. However, existing models do not permit principled testing of hypotheses about factors shaping the trajectories of paths. I argue that paths are a unique class of dependent variable and should be modeled accordingly, much as event count outcomes are commonly modeled with Poisson regression.

Social scientists are not only interested in modeling paths as a dependent variable, but also in evaluating the effects of path-assigned treatments. In economics, a long-standing debate revolves around whether connective infrastructure leads to economic development (Fogel, 1962, 1964; Aschauer, 1989; Banerjee et al., 2012; Casaburi et al., 2013). More recent work has also linked highway construction to popular support for the Nazi party (Voigtlaender and Voth, 2014) and urban–suburban political polarization (Nall, 2013, 2015); electrification to liberalizing attitudes in the Tennessee Valley (Caughey, 2012); and oil pipelines to local revenue sharing (Blair, 2016). This work has generally relied on context-specific information to address inferential challenges in these settings—for example, approximating an ideal experiment in which several alternative highways are proposed, but only some selected for construction. I discuss a way to take this intuition and generalize it, by modeling the path assignment process.

Beyond connective infrastructure, paths also appear widely in studies of social network search. As a concrete example, Habyarimana et al. (2007) showed that co-ethnic networks increase "findability" of strangers in their study of public goods provision. In their experiment, randomly selected Ugandan "runners" were given photographs of strangers, then asked to locate them within 24 hours—which they did with startling levels of success. The ability to locate and sanction free riders through social networks, often easier among co-ethnics, is in turn linked to public goods provision in diverse societies (Miguel and Gugerty, 2005; Eubank, 2016). Christia et al. (nd) examine

neighbors to the north and south *cannot* be connected—otherwise, the route would already be complete and i would be superfluous.

related questions in the context of Iraq, showing that minority Sunnis adapt to a Shia-dominated society by developing more efficient network search strategies to access public services. Network search is also important in the spread of political information, where the search patterns of citizens actively seeking knowledge can distort the information they ultimately receive (Huckfeldt and Sprague, 1987, 1995). (The passive transmission of political information, like the spread of rumors, can also be thought of as a flow path over a network (Converse, 1962; Zaller, 1989).) Finally, the formation of buyer-seller ties in imperfect markets often involves searching a social network for exchange partners (Kranton and Minehart, 2001).

Path outcomes are also of great interest in public policy, where they can represent aggregate flows of, e.g., refugee migration or smuggled drugs. Potential determinants of migration routes, such as welfare policy and border security, have been debated with increasing urgency since a sharp increase in European migration in 2015 (U.N. High Commissioner for Refugees, 2016). The spillover effects of policies that divert such flows is a common concern. For example, Dell (2015) considers a game-theoretic model of path-based spillover, in which crackdowns by Mexico's National Action Party (PAN) force drug traffickers to reroute through nearby municipalities, and shows that the model is consistent with rising drug arrests after PAN victories.

The random-path model allows researchers to learn about the preferences of an actor, such as the highway planners, information-seeking citizens, and drug traffickers discussed above, based on an observed path or collection of paths. It can test whether patronage is a significant factor in the trajectories of highways and electrical lines—that is, whether infrastructure routes deviate from an "optimal" route in order to visit certain areas—or whether drug traffickers tend to avoid states with harsher penalties, such as minimum-sentencing laws. Furthermore, the model allows researchers to simulate various quantities of interest in a statistically principled way: How many additional miles of highway were built because of patronage? If one state cracked down on trafficking, what volume of drug shipments would divert into its neighbors? Finally, in ongoing work, I show that RPM can be used as a model of treatment assignment to allowing inference about the causal effects

of path-assigned treatments (Rubin, 1991). This approach also permits the study of spillover effects (Bowers et al., 2013; Aronow and Samii, nd), such as whether highways lead to growth in nearby areas or whether they contribute to out-migration and decline, by comparing highway towns and nearby areas to places that were as likely to be connected.

Alternative Methods

Most analyses of paths, such as infrastructure, have used standard regression or matching methods that ignore spatial dependence entirely (Rephann and Isserman, 1994; Chandra and Thompson, 2000; Michaels, 2008; Donaldson, ming). Others allow for correlation within a cluster of units, such as counties in a state, but neglect the fact that bordering counties on opposite sides of a state line are *also* highly dependent (Baum-Snow, 2007; Baum-Snow et al., nd). At best, researchers have employed spatial error or spatial autocorrelation models that assume dependence is (1) isotropic, so that each unit is positively correlated with its circular neighborhood;[2] (2) decaying at a constant rate with distance; and (3) stationary, so that the same correlation structure is constant the entire space (Cohen and Paul, 2004; Del Bo and Florio, 2012).

Spatial models are well-suited for analyzing a variety of phenomena, such as policy diffusion (Elkins and Simmons, 2005). However, they are not intended for the analysis of path data, which violate every one of the underlying assumptions outlined above. As a concrete example, consider the naturalistic simulation in appendix A.1, where a road curves around a mountain range to connect cities on opposite sides. Path data has strong positive dependence in some directions (for a town to be connected, the road must approach it from the front and back) and negative in others (if the road detours around one side, it will not pass through the town). Roads can easily exhibit long-range dependence—all towns on one side of the mountain are positively correlated with each other, and they are negatively correlated with towns on the opposite side—the road can only choose one side, and it connects many areas on that side simultaneously. Moreover, the distribution over

[2]Some analyses relax this assumption to allow for elliptical correlation structures.

possible roads is tighter in a mountain pass, where fewer viable routes exist.

Generally speaking, analyses that ignore pathwise dependence between observations lead to results that are as much of an "exercise in self-deception" as those that ignore clustering (Cornfield, 1978). Moreover, all of the models describe above essentially treat local dependence as a nuisance. Their output cannot be interpreted in terms of useful path-related quantities of interest—for example, the change in road length caused by the mountain described above.

In contrast to spatial models, the RPM is a model of network formation. It starts with the graph of all neighboring nodes, such as counties, and selects a subset of contiguous nodes and edges to connect a starting point to an endpoint. Other families of network models that have been used in political science include exponential random graph models (ERGMs) and latent-space models (Cranmer et al., 2016). Broadly speaking, existing network models deal with dyadic relationships while accounting for the contextual influence of local network structure. For example, ERGMs have proven valuable in the study of international relations for their ability to account for the way allies' relationships affect whether two nations go to war (Cranmer and Desmarais, 2011). The goal of the RPM differs from these models in that it models a purposeful attempt to connect two nodes; it is influenced by network structure over a much longer range, is subject to more severe constraints, and generally addresses a different class of questions.

Model

In this section, I briefly discuss two interpretations of random walks, which are closely related to the random-path model. Walks are introduced as a sequence of dependent random steps. This view is then shown to be mathematically equivalent to an alternative view in which entire walks are drawn, all at once, from a discrete set of sequences. I exploit this equivalence to conveniently express random-path models in the second view, then discuss the implied relationship between random-walk models and RPMs in the first. I then outline the computational challenges in es-

timating RPMs and outline a procedure to recover the posterior distribution of the random-path parameters, given a set of observed paths. The method is placed in the context of the simulated likelihood method and a rapidly growing literature on approximate Bayesian computation.

Random Walks: A Review

Define a weighted, possibly directed graph G as a set of nodes (vertices) denoted $V \in \{1, \cdots, N\}$, such as counties, and a row-stochastic edge-weight matrix $E = [\varepsilon_{i,j}] = [\boldsymbol{\varepsilon}_{1,*}^\top, \cdots, \boldsymbol{\varepsilon}_{N,*}^\top]^\top$. For a walker at i, $\varepsilon_{i,j}$ represents the probability that the walker's next step is to j; it takes on positive values for adjacent j—those in i's neighborhood, $\mathcal{N}_i$, which is the "choice set" for a walker at i—and zero otherwise. Self-links, or $\varepsilon_{i,i}$, are set to zero by convention.

A random walk, $\Gamma \equiv (v_0, \cdots, v_K)$, is defined by a starting node v_0, the transition distributions $v_t \sim \text{Categorical}(\boldsymbol{\varepsilon}_{t-1,*})$, and a stopping rule.[3] For illustrative purposes, I assume that walks stop upon reaching a single predesignated terminus, v_K. The observed path is denoted $\gamma = (\gamma_0, \cdots, \gamma_k)$, and the specified conditions require that $v_0 = \gamma_0$ and $v_K = \gamma_k$. Note that the number of steps in the walk, K, is also a random variable, with realization k. (Alternative stopping rules, such as after a fixed number of steps or when any of a set of nodes are found, may be more appropriate in other applications, and the proposed distribution is easily adapted for these cases.)

The random walk is analogous to the negative binomial distribution in that it can be thought of as either a sequence of dependent categorical random variables, as presented above, or a probability distribution over an infinite discrete set whose elements are sequences of varying length. In either case, given fixed endpoints, the probability of a particular realization is

$$\Pr(\Gamma = \gamma \mid v_0 = \gamma_0, v_K = \gamma_k) = \prod_{t=0}^{k-1} \varepsilon_{\gamma_t, \gamma_{t+1}}$$

[3]This defines a walk in terms of a node sequence, which leaves the intervening edges $v_t v_{t+1}$ implicit. An equivalent definition is that a walk is a subgraph of G, $G_\Gamma = (V_\Gamma, E_\Gamma)$, in which $V_\Gamma \subseteq V$ and E_Γ is a sequence of edges, $(v_0 v_1, \cdots, v_{K-1} v_K)$, in which all elements satisfy $\varepsilon_{v_t, v_{t+1}} > 0$.

It is straightforward to model step probabilities, $\varepsilon_{i,j}$, as a function of M covariates. Let $\boldsymbol{X}$ be a $N \times N \times (M+1)$ tensor where the m-th slice is a matrix of dyadic covariates, such as distance. $\beta = [\beta_0, \beta_1, \cdots, \beta_M]^\top$ is a vector of coefficients, and $\boldsymbol{X}_m\beta^m = \left[\sum_m \beta_m X_{*,*,m}\right]$ is a $n \times n$ matrix of linear predictors. Assume edge weights can be written as

$$\varepsilon_{i,j} = \frac{\exp\left([\boldsymbol{X}_m\beta^m]_{i,j}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m\beta^m]_{i,j'}\right)},$$

so that rows of E are the multinomial logistic transformation, or softmax, of rows of $\boldsymbol{X}_m\beta^m$. Fix β_0 at $-\infty$ and let $X_{*,*,0} = [\mathbf{1}(j \notin \mathcal{N}_i)]$, so that $\varepsilon_{i,j} = 0$ for $j \notin \mathcal{N}_i$, and the probability mass function (PMF) is

$$f_{\text{walk}}(\gamma \mid \boldsymbol{X}, \beta) \equiv \Pr(\Gamma = \gamma \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta) = \prod_{t=0}^{k-1} \frac{\exp\left([\boldsymbol{X}_m\beta^m]_{\gamma_t,\gamma_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m\beta^m]_{\gamma_t,j'}\right)}$$

The random-walk model is well-understood and has been used in the transportation literature. Fosgerau et al. (2013) observe that these models allow loops, including infinite loops, although they report that these are rare in their particular case.

Random Path Distribution as Conditional Random Walk

The random walk, while analytically tractable, is a poor model for many social phenomena because it assumes that each step is independent. Unlike sequential decision-makers in political science, such as highway planners, random walkers are neither forward- or backward-looking. Under typical conditions, they are likely to revisit many nodes—as I describe in section 3.1, a random walker crossing a $N \times N$ street grid will go in circles with near certainty for $N \geq 5$. Moreover, this problem cannot be fixed by incorporating covariates into the step probabilities (such as distance or direction), because loops are a property of the entire sequence rather than any particular step.

This paper proposes an alternative, the conditional random-walk distribution $(\Gamma \mid \Gamma \in \mathscr{P})$, where $\mathscr{P}$ is the set of all possible paths from γ_0 to γ_k—i.e., all walks from start to terminus that contain no loops. Formally, $\mathscr{P} \equiv \{\psi : \Omega_\Gamma, |\{\psi\}| = |\psi|\}$, where Ω_Γ is the sample space of Γ and the latter condition specifies that all nodes in path ψ are unique. Thus, $\mathscr{P}$ excludes all walks that return to a previously visited node. Given that the observed walk γ is a path, so that it automatically satisfies $\gamma \in \mathscr{P}$, the random-path PMF is found by renormalizing:

$$\begin{aligned} f_{\text{path}}(\gamma \mid \boldsymbol{X}, \beta) &\equiv f_{\text{walk}}(\gamma \mid \boldsymbol{X}, \beta, \Gamma \in \mathscr{P}) \\ &= \frac{\Pr(\Gamma = \gamma, \Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)}{\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)} \\ &= \frac{\Pr(\Gamma = \gamma \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)}{\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)} \\ &= \frac{\prod_{t=0}^{k-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, \gamma_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, j'}\right)}}{\sum_{\psi \in \mathscr{P}} \prod_{t=0}^{|\psi|-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\psi_t, \psi_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\psi_t, j'}\right)}}. \end{aligned} \tag{1.1}$$

This random-path distribution has properties that make it well-suited for modeling common decision-making processes, such as the political science applications discussed in section 1.2.1. Recall that in the view of the random walks as a sequence of random variables, at each step, the walker is only "backward-looking" insofar as the previous step determines the current options. That is, in a random walk, $\Gamma_t \not\perp\!\!\!\perp \Gamma_{t-1}$, but $(\Gamma_t \mid \Gamma_{t-1}) \perp\!\!\!\perp \Gamma_{t-2}$. In the same view of random paths, the walker is "fully" backward-looking in that it will tend to avoid the vicinity of all previously visited nodes. The walker is also forward-looking in that it tends to avoid traps and other local optima with foresight, anticipatorily moving in directions that will take it to the destination faster.

Estimation

Equation 1.1 suggests a likelihood-based approach for inference on the random-path model. I begin by briefly discussing an algorithm for exactly calculating this likelihood. Because this approach

becomes intractable for moderately sized or dense graphs, I then develop an simulation-based approximation that converges to the exact method as the number of simulations tends to infinity. Finally, I briefly discuss computational issues for Bayesian inference on RPM models.

Simulated RPM Likelihood

A natural approach for inference on random paths is to use the likelihood $\mathscr{L}_{\text{path}}(\beta \mid \boldsymbol{X}, \gamma) \equiv f_{\text{path}}(\gamma \mid \boldsymbol{X}, \beta)$, where the full expression for the right-hand side is given in equation 1.1. The chief difficulty in doing so is that the denominator of equation 1.1 varies with β and involves summing over the (typically large) set of possible paths between the observed start- and endpoints, γ_0 and γ_k. For example, the maximum likelihood estimate of β are the parameters that maximize the ratio of (a) the unconditional (random-walk) probability of the observed path to (b) the totaled random-walk probabilities of every other path that could have been drawn. In appendix A.0.1, I describe an exact method for doing so. This procedure uses a recursive search to explicitly enumerate every possible path, then sums the random-walk probabilities of mutually exclusive paths. In principle, the Fisher information matrix can also be derived in this way, with confidence intervals calculated using the delta method.

However, in practice, explicitly enumerating and operating on all possible paths is computationally infeasible, even for moderately sized or dense graphs. For example, in complete graphs, where every node is connected to every other, the number of possible paths is given by $\sum_{k=0}^{N-2} \frac{(N-2)!}{k!}$. Even in a ten-node complete graph, 109,601 paths are possible. Building on the intuition behind the exact approach, I develop algorithm 1 to approximate the likelihood function to arbitrary precision. Algorithm 1 is based on a common numerical approach for summations over hard-to-enumerate domains, Monte Carlo integration. The approach developed here is analogous to the following procedure for approximating the integral $\int_a^b f(x)\, dx$: Randomly sample points on the uniform $[a,b]$ distribution, evaluate $f(x)$ at each point, average the results, and multiply by the size of the sampling space $(b-a)$.

To apply this to the RPM case, let Ψ be the uniform distribution over the set of all possible paths $\mathscr{P}$. The following is a directly analogous approach for estimating the denominator of equation 1.1, which is equivalent to $\sum_{\psi \in \mathscr{P}} f_{\text{walk}}(\psi)$. Repeatedly draw $\psi \sim \Psi$, evaluate the random-walk probability for each sampled element, and average across draws to estimate $\mathbb{E}_\Psi[f_{\text{walk}}(\Psi)] = \frac{1}{|\mathscr{P}|} \sum_{\psi \in \mathscr{P}} f_{\text{walk}}(\psi)$. Then, multiply by the total number of paths $|\mathscr{P}|$ to find an estimate of the denominator. Because the numerator is calculated exactly, the simulated likelihood (Lee, 1992) inherits the desirable property of converging to the exact likelihood in appendix A.0.1 (up to a multiplicative constant) as the number of or simulations tends to infinity. This can be seen by noting that depth-first search finds every path exactly once, and the number of times that the uniform distribution draws each path converges to $\frac{S}{|\mathscr{P}|}$ as S grows large.

There are two complications in this procedure. The lesser complication is that we need to know the value of $|\mathscr{P}|$ to correctly normalize. This is a #P-hard problem (Valiant, 1979), meaning that it can only be solved by listing every possible path and then counting them—precisely the computational issue that we were trying to sidestep in the first place.[4] Fortunately, $|\mathscr{P}|$ does not involve the RPM parameters, β, and so it can be absorbed into the normalizing constant of the RPM likelihood function. The RPM likelihood is then given by

$$\begin{aligned}
\mathscr{L}_{\text{path}}(\beta \mid \boldsymbol{X}, \gamma) &= \frac{\prod_{t=0}^{k-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, \gamma_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, j'}\right)}}{\sum_{\psi \in \mathscr{P}} \prod_{t=0}^{|\psi|-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\psi_t, \psi_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\psi_t, j'}\right)}} \\
&= \frac{\prod_{t=0}^{k-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, \gamma_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, j'}\right)}}{|\mathscr{P}| \cdot \mathbb{E}_\Psi \left[\prod_{t=0}^{|\Psi|-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\Psi_t, \Psi_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\Psi_t, j'}\right)} \right]} \\
&\propto \frac{\prod_{t=0}^{k-1} \frac{\exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, \gamma_{t+1}}\right)}{\sum_{j'} \exp\left([\boldsymbol{X}_m \beta^m]_{\gamma_t, j'}\right)}}{\mathbb{E}_\Psi[\Pr(\Gamma = \Psi \mid \nu_0 = \gamma_0, \nu_K = \gamma_k, \boldsymbol{X}, \beta)]}.
\end{aligned} \tag{1.2}$$

[4] Roberts and Kroese (2007) explore an approximation for large graphs, but its accuracy is unknown.

After eliminating this constant, equation 1.2 could in principle be approximated by Monte-Carlo sampling from Ψ as described above. This brings us to the second complication. Unfortunately, Ψ cannot be sampled—there is no known algorithm for uniformly sampling paths. To deal with this, I adapt a non-uniform distribution for importance sampling on $\mathscr{P}$, the loop-erased random walk (LERW). [5] The LERW begins with a pure random walk, then retraces its steps and removes loops that return to previously visited nodes (the procedure is given as part of algorithm 1). Wilson (1996) proved that the spanning tree—i.e., a subgraph that connects all nodes, such as a maze, that contains no cycles—produced by iteratively combining LERWs will be a uniform draw from the set of all spanning trees. In proposition 1, I use this property to construct an importance-sampling scheme on $\mathscr{P}$.

Proposition 1 (Simulated RPM Likelihood). *Define the unweighted version of G, $\tilde{G}$, and let L be a path-valued random variable, with distribution $f_{\text{LERW}}(\psi : \tilde{G}, v_0, v_K)$, that can be sampled by the loop-erased random walk on $\tilde{G}$ from v_0 to v_K. The denominator of equation 1.2 can be rewritten*

$$\begin{aligned}&\mathbb{E}_\Psi[\Pr(\Gamma = \Psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)] \\ &\quad = \sum_{\psi \in \mathscr{P}} \Pr(\Gamma = \psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)\, f_{\text{LERW}}(\psi : \tilde{G}, v_0, v_K)\, w(\psi),\end{aligned}$$

and its importance-sampling estimate is

$$\begin{aligned}&\hat{\mathbb{E}}_\Psi[\Pr(\Gamma = \Psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)] \\ &\quad = \sum_{s=1}^{S} \Pr(\Gamma = L_s \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)\, w(L_s),\end{aligned}$$

[5] An alternative approach for sampling non-uniformly from $\mathscr{P}$ is the self-avoiding walk (SAW). The SAW is a random walk that sets transition probabilities to zero for previously visited nodes, as in *Snake* (Gremlin, 1976). Properties of the SAW are largely unknown, so this approach is not considered. As part of their algorithm, Roberts and Kroese (2007) attempt to estimate and correct for the bias of SAWs toward shorter paths by simulation. They employ an ad-hoc method that increases the probability of long paths by down-weighting transition probabilities to the final node (i.e., avoiding termination) based on the number of steps that have been taken. However, their approach under-samples nodes that are distant from the target, convergence rates of the various correction factors are unknown, and the resulting distribution is poorly understood.

where S is the number of importance-sampling draws. The adjustment factor $w(\psi) \propto \frac{1}{\det L_{(-i,-j)}(\tilde{G}/\psi)}$ *is the ratio between the target uniform distribution and the LERW distribution. The above holds for any i and j in V. The term* $\tilde{G}/\psi$ *is the iterated edge contraction of* $\tilde{G}$ *along all edges in path* ψ, $L(\cdot)$ *is the Laplacian matrix of a graph, and* $M_{(-i,-j)}$ *is the* (i, j) *minor of a square matrix M.*

A proof is given in Appendix A.0.2. Briefly, Wilson (1996) implies that the probability that a LERW draws a particular path, ψ, will be proportional to the number of spanning trees that include ψ as a subgraph. Proposition 1 uses the deletion-contraction recurrence to exclude trees that cannot arise under ψ, then applies Kirchoff's matrix tree theorem to the contraction to count the number of such spanning trees. Proposition 1 immediately suggests a simulated-likelihood analogue of equation 1.2. This simulated likelihood forms the basis for statistical inference.

Discussion

The methodological challenges that arise in the RPM are closely related to those in fixed-effects logistic regression model and related models. In the fixed-effect logit setting, conditioning on a sufficient statistic induces a combinatorics problem in the denominator of the conditional likelihood. This problem, which also arises in censored survival data, is commonly addressed with various analytical approximations (Breslow, 1974; Efron, 1977). In the RPM setting, the no-loop conditioning leads to a similar problem, but the combinatorics in the denominator are sufficiently complex that no closed-form approximations are known.

The simulated likelihood approach taken here is a response to this problem. It uses Monte Carlo simulations to integrate over the outcome space, and thus approximate the denominator, by computational rather than analytical methods. The procedure differs from typical applications of the simulated likelihood method, which use Monte Carlo integration to marginalize nuisance variables such as random coefficients (Bhat, 2001). A related approach to outcome-space integration is used in approximate Bayesian computation, when exact calculation of the likelihood is not practical or possible (for recent developments, see Marin et al., 2012).

This simulated likelihood is computationally intensive for two reasons. First, for each of the S sampled paths is drawn, an expensive matrix determinant must be calculated to find the adjustment factor. Second, each time the simulated likelihood is evaluated at some point in the parameter space, algorithm 1 loops over and operates on all S paths. (In the applications explored here, values of S are on the order of 10^6 paths).

Neither the simulated likelihood nor the true likelihood function to which it converges are necessarily well-behaved, particularly when the network size is small. Thus, numerical optimization of the likelihood (with confidence intervals by the delta method with numerical Hessian) is inadvisable for short paths. When the parameter space is low-dimensional, the simulated likelihood can simply be evaluated on a fine grid. As the number of parameters increases, this approach rapidly becomes infeasible. In appendix A.0.3, I develop a procedure to estimate RPM by Metropolis–Hastings (MH) and discuss further approximations that can greatly speed computation. In appendix A.1, I use simulations to evaluate the proposed estimation procedure in various scenarios. Consistency is shown to depend not only on the number of paths, but also their length (and indirectly, network size).

Data:

starting node γ_0, teminus γ_k, covariates $\boldsymbol{X}$, parameters β
unweighted graph $\tilde{G}$, number of simulations S

Result:

$\psi_1, \cdots, \psi_s \in \mathscr{P}$
$w_1, \cdots, w_s$, inverse importance weights
$\hat{\mathbb{E}}_\Psi[\Pr(\Gamma = \Psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)]$,
estimated denominator of random-path likelihood function (equation 1.2)

Algorithm `ApproxPrPath`$(\gamma_0, \gamma_k, \boldsymbol{X}, \beta)$

for $s \in 1, \cdots, S$ **do**
  draw $\psi_s \sim$ `LERW`$(\tilde{G}, \gamma_0, \gamma_k)$
  weight by $w_s = \frac{1}{\det L_{(-i,-j)}(\tilde{G}/\psi_s)}$
end
estimate $\hat{\mathbb{E}}_\Psi[\Pr(\Gamma = \Psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)] =$
$\frac{1}{\sum_{s=1}^{S} w_s} \sum_{s=1}^{S} w_s \Pr(\Gamma = \psi_s \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)$
return $\hat{\mathbb{E}}_\Psi[\Pr(\Gamma = \Psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)]$

Procedure `LERW`$(\tilde{G}, \gamma_0, \gamma_k)$

initialize $\psi = (\gamma_0)$, $i = \gamma_0$
while $i \neq \gamma_k$ **do**
  sample j uniformly from $\mathscr{N}_i$
  step to $i = j$ and append to ψ
end
initialize $t = 0$
while $t < |\psi| - 1$ **do**
  set t' to maximum index satisfying $\psi_t = \psi_{t'}$
  if $t' > t$ **then**
    erase elements in loop $(\psi_{t+1}, \cdots, \psi_{t'})$ from ψ
  end
  t+=1
end
return ψ

Algorithm 1: Approximating the probability that a random walk from γ_0 to γ_k is a path, up to the unknown multiplicative scaling factor $|\mathscr{P}|$, by importance-sampled Monte Carlo integration. The approximation converges to the exact likelihood as the number of simulations, S, approaches infinity. The loop-erasure proceeds along an unweighted random walk, identifies points where the walk returns to a previously visited node, then erases the second visit and all intervening nodes.

U.S. Interstate Highways

I apply the RPM to the U.S. Interstate Highway System, often called the "greatest public works project in history." The interstate highways, estimated to cost over 500 billion inflation-adjusted dollars, provided the first comprehensive national road network for national defense and economic development. It offers an ideal empirical testing ground in that the planning process was highly transparent and explicitly stated decision criteria are publicly available. Interstate paths are analyzed with the RPM and alternative spatial models to assess whether these models accurately recover the decision-making process. The random-path model produces estimates that are consistent with planning criteria, whereas alternative spatial methods yield conflicting results.

Qualitative Benchmark

The Interstate Highway System evolved over several decades and numerous iterations. The 1921 Pershing map—an Army proposal of important routes for military logistics, including emergency mobilization—was one of the first comprehensive drafts. The counties to connect were identified in a report by the National Interregional Highway Committee (1944), along with a detailed discussion of the route selection process. At this level, planners were fairly insulated from political pressures: An independent committee devised and applied a rule-based system, using census data, to identify counties to connect in each region. The committee was composed of professional bureaucrats—planners, civil engineers, and administrators from the Bureau of Public Roads. Legislative influence was primarily through a formula that fixed the distribution of funds among states, rather than influencing specific routes. Thus, the technocratic priorities described in this document represent a qualitative benchmark that can be used to test whether the RPM correctly models the decision-making process. The locations of highways within counties, in terms of the specific tracts of land to condemn, were proposed by the U.S. Bureau of Public Roads (1955) but are not considered here.

The highway committee states that "the recommended interregional system conforms closely" with military priorities, with extensive additions. The report discusses the specifics of route selection in detail, with the vast majority of attention devoted to connecting major urban centers. Planners started with a list of cities that had a 1940 census population over 100,000;[6]; between these, "the primary purpose was to select routes... which would join the principal centers of population and industry... by lines as direct as practicable." Thus, the Interstate Highway System was first and foremost designed to connect population centers, while keeping the overall system at a manageable length—in the ultimately recommended plan, just under 34,000 miles. While it was considered desirable to connect counties with high manufacturing capacity, planners observed that in practice, this goal could be achieved by maximizing the urban population served.[7]

Rural population was also described as an important consideration, with the proposed highways passing "en route between these hubs, through or very close to the denser clusters of population in small towns and populous rural areas."[8] Agricultural production, on the other hand, was described as something of an afterthought, though the proposed system was shown to be adequate for the purposes of transporting farm products.

Just as importantly, planning documents describe a number of factors that were not important in route selection. Perhaps surprisingly, topography was an influence in "remarkably few places," and soil quality was not considered at all. Nor were interstates built along existing routes. Except for a few sections in the Northeast and Detroit, existing roads generally did not meet standards for lane width and arrangement, and "existing rights-of-way are grossly insufficient to permit such

[6]Virtually cities with more than 100,000 residents were directly served, with three exceptions that were "passed in close proximity." The report explains that these cities could not be directly served without negatively impacting much larger adjacent cities.

[7]Roughly two-thirds of variation in 1939 manufacturing can be explained by population alone. From the committee report: "While slight differences exist in the relative importance of cities when they are measured on the one hand by their populations and on the other by the values added by their manufactures, on the whole the similarity of the measures is marked... It is, therefore, concluded that the recommended system closely approximates the system of optimum extent from the standpoint of service to manufacturing industry."

[8]The report continues, "Indeed, the courses of the recommended routes are shown by this map to be in most instances the inevitable selections, if service of population is to be considered important in the choice."

widening." These statements help justify the relatively simple model specifications used here.

Data

Following the National Interregional Highway Committee, I define the county as the unit of analysis. Based on National Highway Planning Network shapefiles, I convert a total of 57 two-digit interstate highways into sequences of adjacent counties.[9] I condition on the endpoints of these highways, as well as intermediate cities with populations over 100,000.[10] For example, I assume that I-5 was built as a route to connect Seattle to Los Angeles, with mandatory stops in Portland and Sacramento, but that planners were otherwise free to choose the intermediate steps. The interstates are thus split into 136 highway segments with fixed endpoints, passing through an average of 11 counties each. The highway system can be represented as a network in which nodes are highway counties and edges are dyadic highway connections. This highway network is a subgraph of the overall U.S. county network (shown in figure 1-1), because highway counties are a subset of all counties, and county-dyads with a direct highway connection are a subset of all geographically adjacent counties.[11]

Highway construction is modeled as a collection of random paths. Edge weights for the move

[9]In the Interstate Highway System numbering scheme, long-range east-west (north-south) highways are given even (odd) two-digit codes, such as I-90 (I-95). Three-digit auxiliary highways that start with an odd number are spurs (e.g., I-391) and those that start with an even number are circumferential highways ("ring roads," e.g., I-495). Auxiliary highways are discarded.

[10]This is consistent with the planning process, in which routes were selected so that virtually all cities with populations over 100,000 were directly served. For computational reasons, two additional fixed waypoints were added—I assume that I-90 had to pass through Sioux Falls, SD and Buffalo, WY (junctions with I-29 and I-25, respectively). This was done to break up a highway segment of I-90 that was otherwise an extreme outlier in terms of length.

[11]Two counties, A and B, are considered adjacent if (1) they share a border or (2) their county seats can be connected by a line that barely clips a third county, C. In a hypothetical set of 10 mi. $\times$ 10 mi. counties forming a square grid, the latter criteria allows for diagonal connections between counties. The threshold for "barely clips" is arbitrarily defined as one-quarter of county C's characteristic length (the square root of its area). For example, in the square grid described above, a road can clip at most a triangular region of area 2.5 sq. mi. (out of C's total area of 100 sq. mi.) before it is considered an $A-C-B$ path, rather than an $A-B$ path.

from i to j are modeled by a softmax function with covariates based on planning documents:

$$\varepsilon_{i,j} = \frac{\exp\left(z_{i,j}\right)}{\sum_{j' \in \mathcal{N}_i} \exp\left(z_{i,j'}\right)}, \text{where}$$

$$z_{i,j} = \beta_{\text{dist}}\texttt{dist}_{i,j} + \beta_{\text{pop}}\texttt{pop}_j + \beta_{\text{urb}}\texttt{urb}_j + \beta_{\text{mil}}\texttt{mil}_j + \beta_{\text{ind}}\texttt{ind}_j + \beta_{\text{agp}}\texttt{agp}_j$$

and covariates are defined as follows (census data from Haines, nd):

- $\texttt{dist}_{i,j}$: Minimum road distance between county seats of i and j, miles.[12]
- $\texttt{pop}_j$: 1940 log census population of county j.
- $\texttt{urb}_{j,m}$: 1940 urban census population of county j. Discretized into 5 dummy variables with breakpoints at 2, 10, 25, and 50 thousand, following planning documents.
- $\texttt{mil}_j$: Military facilities in j, as proxied by log spending from 1940–1945 reported by the Civilian Production Administration (CPA).
- $\texttt{ind}_j$: Industrial capacity of county j, as proxied by CPA-reported log total value of industrial facility expansions from 1940–1945.[13]
- $\texttt{agp}_j$: 1939 log value of agricultural products sold and traded in county j.

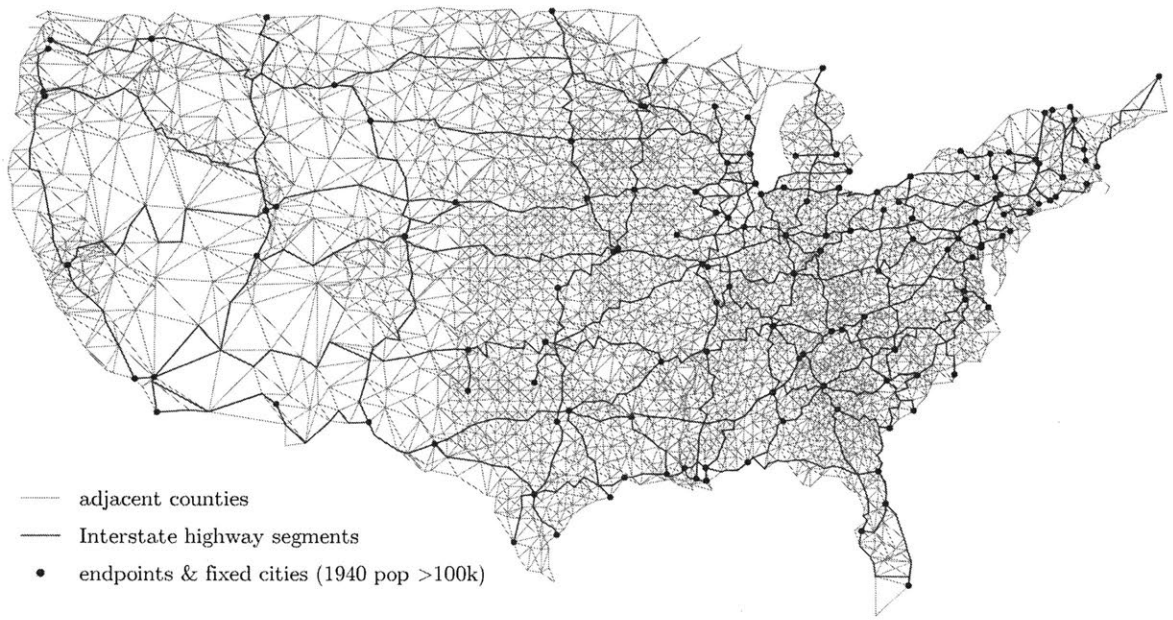
Results

Based on observed decisions by planners as they connect major cities through a series of intermediate counties, the random-path model estimates the priorities of the Interstate Highway System. For example, figure 1-2 shows a case in which Interstate 80 deviates from the shortest possible route between Cheyenne, WY and Omaha, NE. While such cases appear to suggest that planners are willing to trade off some additional distance in order to pass through small cities and military facil-

[12]For approximately "adjacent" counties that do not share a border, I first identify the points at which i's border is closest to j's border. The minimum road distance is then defined as the distance from i's seat to the point on its border, plus the minimum distance between i and j's border, plus the distance from the point on j's border to its county seat.

[13]Results are substantively and statistically indistinguishable when using 1939 log manufacturing value added, but this earlier data has substantial missingness in non-urban aras

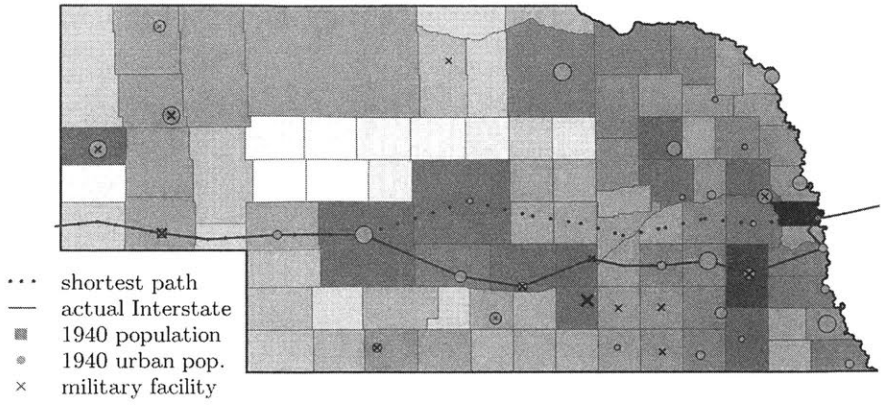
Figure 1-1: The graph of all adjacent counties is drawn in thin gray lines. The subgraph of adjacent counties connected by Interstate Highways is highlighted with thicker red lines. The subgraph is composed of paths between waypoints, plotted as blue circles.



ities, visual inspection alone cannot determine whether the relationship is statistically significant, nor assess the value placed on cities relative to military bases.

Because interstates are bidirectional but the RPM analyzes directed paths, I treat each segment as the equally weighted combination of both directions, e.g., the northbound and southbound parts. Parameter posterior distributions and convergence diagnostics are reported in appendix A.2 For interpretability, parameter estimates are converted to a probability scale in figure 1-3 according to the following scenario: Suppose a highway comes upon a mutually exclusive choice between two identical counties, j and j'. Holding the rest of the route fixed, the highway has an equal probability of passing through either. If j was changed so that it had some desirable property, such as a higher population, what would be the corresponding increase in the probability of highway construction through it?

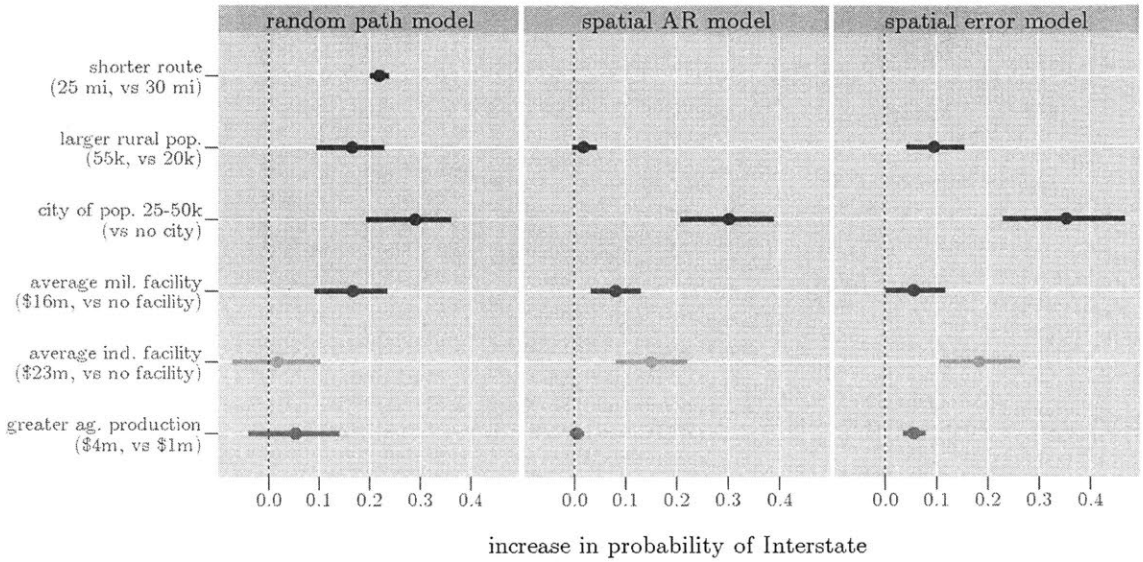
Figure 1-2: The I-80 segment between Cheyenne, WY and Omaha, NE (solid red line) deviates from the shortest route (dotted red line). Rather than minimizing distance, highway planners opted to connect small cities (yellow circles) and military facilities (black crosses).



RPM results suggest that planners preferred more direct routes. They reveal a preference for connecting counties with higher population, particularly when this population was concentrated in cities. The addition of an average manufacturing facility or a one-standard-deviation increase in agricultural production was found to be small and insignificant, whereas the addition of an average military facility was associated with a large increase in the chances of receiving an interstate. These model predictions agree well with the qualitative priorities outlined above (according to the National Interregional Highway Committee, routes were to be "as direct as practicable" and "close to the denser clusters of population," with "close proximity of... [military and naval] establishments to the recommended routes").

In contrast, alternative spatial models with identical specifications (except distance between counties, a dyadic covariate that cannot be incorporated) produce significantly different estimates that do not agree with documented interstate priorities. A standard probit model with state fixed-effects and a correction for spatially correlated errors incorrectly suggests that an average industrial facility is more important than an average military facility, by a large and statistically significant margin—a surprising and almost certainly incorrect result, given the planning process (and the fact

Figure 1-3: Predicted increase in probability of highway construction in one of two otherwise identical counties, holding all else fixed. Base distance of 25 miles is roughly the median distance between county seats connected by interstate highways. Various changes in county attributes (y-axis) are approximately one standard deviation in the attribute, except for military and industrial capabilities, which represent the value of an average facility. Points are posterior means; error bars are 95% posterior credible intervals.



that the system is named the "National System of Interstate and Defense Highways"). The spatial autoregressive probit model finds that if city size is held fixed, a one-standard-deviation increase in total population has no substantive effect on highway construction. This directly contradicts planning documents (from the National Interregional Highway Committee, 1944, "the recommended routes trace their courses along the country's most populous bands [of rural population]... the evidence of appropriate selection is marked").

The failure of alternative spatial models is due to two reasons. First, the path structure of highways is a violation of the underlying assumptions of these spatial models. Second, and more importantly, by nature they cannot account for the distance between sequential highway counties—perhaps the single most important factor in route choice. These alternative models do not capture how highways deviate from the shortest route to touch desirable counties. Instead, they essentially compare highway counties to non-highway counties, ignoring the fact that many rural areas were

never viable candidates for an interstate.

Navigating the Streets of Baghdad

In Baghdad, the relationship between majority Shia and minority Sunni Muslims has been one of peaceful coexistence for centuries (Tripp, 2000). This history of integration and intermarriage stands in stark contrast to a wave of ethnic cleansing in 2006–2007 which has dramatically reshaped the city's ethnic landscape (Baker et al., 2006). What are the long-run effects of this civil conflict—and the resulting counterinsurgency campaign—on Baghdad's political geography? The prospects for post-conflict reintegration hinge on whether these changes have been driven solely by *need-based* sorting (Morrison, 1993) or whether they have also been accompanied by the emergence of *taste-based* sorting (Schelling, 1969, 1971).

There is no dispute that need-based sorting—moving out of mixed neighborhoods to escape violence—was an important factor in Baghdad's sudden segregation. During the 2006–2007 conflict, sectarian militias drove many families out of their houses at gunpoint, and others were intimidated into moving preemptively. What is less clear is whether conflict fundamentally changed ethnic relations in Baghdad and led to taste-based sorting. If the conflict was simply a power struggle between armed factions, such as Ba'ath loyalists and the Mahdi Army, citizens' preferences might remain unchanged. On the other hand, conflict could have precipitated a shift in citizens' ethnic attitudes. In this case, even people who were unaffected by violence would move out of mixed-sect areas due to a newfound distaste for old neighbors.

This question is not merely of historical interest. Out-group aversion is an important parameter in game-theoretic models of segregation, and different values lead to very different predictions about the future trajectory of Baghdad or other cities segregated by violence (Young, 1998; Zhang, 2004a,b). If sorting was a need-based response to conflict, then segregation is not stable: After the conflict ends, the city will eventually return to its former integrated state. If there is a substantial

taste-based component, however, the current geographic division of Sunnis and Shia is likely to persist far beyond the end of conflict.

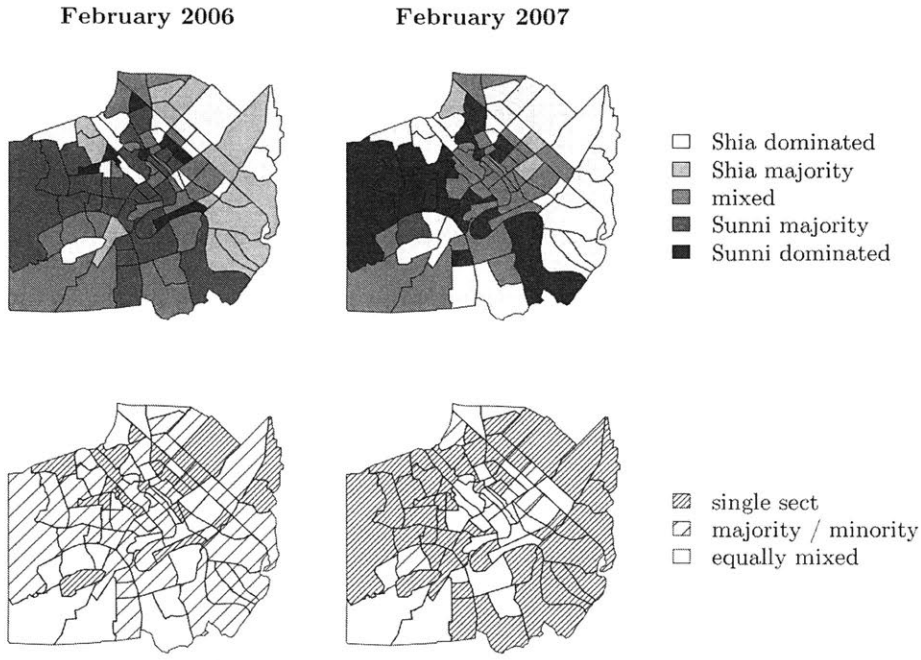
Using the RPM, I analyze behavioral data from a field activity by Christia and Knox, in which subjects participate in a "treasure hunt" in their own home neighborhoods. Results suggest that ethnic conflict over the past decade has led to the emergence of previously nonexistent taste-based sorting. In one scenario—a hypothetical one-kilometer walking task that can be completed in about 12 minutes—estimates indicate that Shia will go out of their way by 34% (4 minutes) to avoid Sunni areas, and Sunnis will go out of their way by 12% (1.5 minutes) to avoid Shia areas. These shifting societal preferences have implications not only for reintegration, but also for economic and political development as the Iraqi state attempts to rebuild after years of conflict.

Theory and Background

The geographic impact of conflict on Baghdad residents, shown in figure 1-4, is hard to overstate. Sunni and Shia once lived side-by-side in nearly every district of the city, but formerly mixed areas are now overwhelmingly dominated by one sect or the other. What drove the sudden segregation in 2006–2007? Clearly, need-based sorting in response to sectarian purging was an important factor. The effects of violence on migration are examined by Morrison (1993), who developed a model that incorporates preferences for safety in addition to economic considerations such as wage maximization. But did the aftershocks of this conflict also change sectarian attitudes and lead to taste-based sorting, where citizens moved due to a newfound aversion to their out-group neighbors? The answer to this question matters, because violence fades—recent deaths, though still substantial, have fallen below half the peak in late 2006 (Iraq Body Count, 2016). Ethnic attitudes, on the other hand, can persist for years if not decades.

The daily lives of Baghdadis can help shed light on these questions. In terms of daily movement, need-based sorting suggests that as citizens navigate their surroundings, their decisions are driven primarily by a desire to survive, earn a living, and procure food. Citizens will generally

Figure 1-4: Top panels show sectarian composition of Baghdad pre- and post-purges (adapted from International Medical Corps, 2007). Darker (lighter) districts have a higher proportion of Sunni (Shia). Bottom panels show the sectarian diversity of districts; denser hatch marks indicate less diverse areas. As a result of local sectarian cleansing, areas that were previously mixed with a Sunni majority (dark gray, slight hatching) tend to become all-Sunni (black, dense hatching), and those that were previously mixed with a Shia majority (light gray, slight hatching) become all-Shia (white, dense hatching).



act efficiently, taking direct routes when walking, except when facing the threat of violence. If Baghdad's segregation is predominantly a need-based response to this violence, then we may expect gradual desegregation after conflict ends: Citizens will stop avoiding out-group areas as they become safer, and families will move back into former homes or sort into new neighborhoods (Tiebout, 1956), with the city eventually shuffling back to an integrated state. Although the claim may seem implausible in the current climate, integration has proven to be surprisingly robust in the past. Baghdad did not segregate in response to Shia and Kurdish uprisings in 1991 or the brutal suppression that followed, even after a major uptick in sectarian tensions. To the extent that increased contact can slowly reduce ethnic animosity in post-conflict settings (Samii, 2013; Mironova and

Whitt, 2014; Hartman and Morse, nd), this scenario suggests the possibility of reconciliation.

Taste-based sorting, on the other hand, implies that citizens will pay costs to avoid the discomfort of out-group contact even when there is no threat to safety or economic rationale. For example, it predicts that they will walk far out of their way to avoid out-group areas. If true, this would be a new development in Baghdad's history of cosmopolitanism. Even after the fall of Saddam Hussein's dictatorship in 2003, when restrictions on movement were abolished, there was little to no change in Baghdad's ethnic composition—Sunni and Shia continued to live in close proximity. After the 2006–2007 purges, however, there are some reasons to suspect that this may be changing. For one, institutional changes have increasingly elevated the prominence of sectarian identity. Moreover, the counterinsurgency strategy of coalition forces has been one of divided political geography.[14] Coalition-built walls and armed checkpoints intended to protect single-sect neighborhoods have been criticized for "hardening the separation of Sunnis and Shias," with effects that could persist long after their ongoing dismantlement (Damluji, 2010). Research starting from Schelling (1969, 1971) shows that these conditions can lead to a rapid "tipping point," beyond which segregation is hard to escape. If political violence has contributed to or been accompanied by the rise of taste-based sorting, these agent-based and game-theoretic results suggest that society is locked in a long-term segregated state (Young, 1998; Zhang, 2004a,b).

Empirically testing these competing hypotheses is a difficult task. Surveys offer one proxy of sectarianism but are often subject to social desirability bias. For example, in a separate survey of religious Shia pilgrims, Christia et al. (nd) show that pilgrims claim to support Sunni–Shia interaction but still overwhelmingly favor co-ethnic neighbors in a conjoint experiment (Hainmueller et al., 2014).[15] While indirect survey methods can accurately measure sensitive attitudes and even certain kinds of past behavior, such as sensitive vote choices (Rosenfeld et al., 2016), they are

[14] As an example, the well-trafficked Bridge of the Imams, which served as a point of contact between Sunni Adhamiya and Shia Kadhimiya, was barricaded for years to stop armed conflict between the neighborhoods.

[15] Among these religious respondents, the only trait less desirable than Sunni faith was alcoholism, which is seen as a serious moral failing in Iraq.

unlikely to perform well in this particular context. This is because predictions about re-integration hinge on future behavior. Indirect methods that tap into sectarian attitudes do not directly address this question, since not all biased individuals will express their attitudes through costly and publicly visible behavior; moreover, survey questions that directly ask about future behavior in hypothetical scenarios are notoriously unreliable even for non-sensitive questions (Rogers and Aida, 2013).

Instead, I draw on behavioral data from Christia and Knox (nd) that isolates the taste-based channel through a field activity. We assign matched Sunni and Shia residents to a treasure hunt through mixed-sect areas in their home districts. By assigning participants to find the same target locations, we sidestep a fundamental problem in observational behavioral data: If people live in neighborhoods where all basic needs are met, it might appear that they avoid surrounding areas—including those populated by out-groups—but this would not indicate an unwillingness to move to those areas after violence dies down. In our field study, locations are carefully selected, with input from a mixed-sect team of local advisors and officials intimately familiar with the area, to eliminate any reasonable concerns that participants might have about their personal safety. This avoids a second confounder of taste-based preferences in observational data—that the threat of violence (generally unobservable, unless local knowledge is available) may be associated with out-group areas.

Sample and Design

To test these models of sorting, we recruited a group of University of Baghdad students from mixed-sect districts to participate in a field navigation activity—a treasure hunt. These recruits are not representative of Baghdad as a whole. Instead, they represent a subpopulation in which taste-based avoidance is "least likely" to be found (Eckstein, 1975; Gerring, 2007). If taste-based aversion exists even among well-educated students who attend a mixed-sect university and live in low-conflict, diverse areas, we may safely conclude that it is a widespread phenomenon.

We advertised around campus for students living in two districts, Ghazaliya and Jihad, chosen

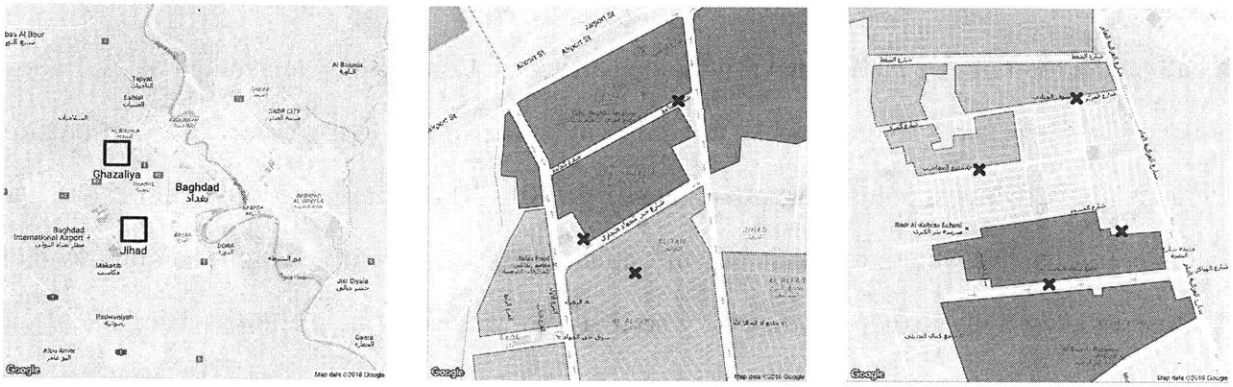
for their sectarian diversity and security (shown in figure 1-5). While both districts are mixed-sect, their sectarian landscapes differ in important ways—Ghazaliya tilts toward Sunnis and Jihad has a larger Shia population, though both districts contain substantial numbers of each sects. Central Ghazaliya has neighborhoods in which Sunni and Shia live side-by-side, but Sunnis (Shia) tend to spend their time on commercial streets in the Sunni-dominated (Shia-dominated) areas to the south (north). In contrast, neighborhoods in Jihad tend to be single-sect, but both groups frequent markets and cafes on the same major thoroughfares.

All potential recruits provided basic demographic information in an initial meeting, and 120 Sunni and Shia participants were chosen so that groups were comparable in terms of on gender, district, and household income sufficiency. The average participant was 21 years old. Our gender balance was skewed towards male participants (two-thirds), and 55% of participants reported that household income was sufficient to cover costs. By design, there were no significant differences between sects (see table 1.1).

The field navigation activity was embedded in a broader week-long smartphone study in which participants consented to our collection of behavioral data on social networks, traditional and social media consumption, and location. To incentivize participation, subjects were given a recent-model Android smartphone to use with their own SIM card for the duration of the study. They received a free one-month credit (covering the study period and three additional weeks) for free data, domestic calling and text messaging. In addition, subjects could earn up to 15,000 IQD in phone credit for completing the treasure hunt (about 13 USD, c.f. laborer day wages of 7–30 USD depending on skill, or civil service monthly wages of 500 USD). This was a substantial amount for participants and was seen as highly motivating. Smaller amounts of phone credit were offered for other, shorter experimental tasks. After the end of the study, over half of the participants won a contest that allowed them to keep their phone.

In each district, we chose centrally located targets that were near to both Sunni- and Shia-dominated areas. These were well-trafficked locations, such as markets, bus stations, schools, and

Figure 1-5: Left panel indicates location of Jihad and Ghazaliya in Baghdad. Center panel shows the treasure hunt area in Jihad, with targets (black $\times$) and Sunni (dark red) and Shia (light blue) residental areas. Neutral commercial streets and or mixed residential areas are not marked. Most Jihad participants walked between the western, northern, and southern targets in that order. The right panel shows the Ghazaliya playing field. Most Shia participants walked from the northern target to the western one, then finished at one of two southern targets; most Sunnis did the route in reverse order.



mosques. Locations were selected from a set of options that local advisors of both sects agreed would eliminate reasonable concerns about security. The vicinity of each treasure hunt was an area visited by people of both sects at least occasionally. Besides avoiding harm to participants, this also helped ensure that differences in walking routes were due to taste-based aversion, rather than need-based concerns about safety.

Each participant was assigned to a supervisor who provided treasure-hunt directions via an instant messaging app. Supervisors were instructed to initially provide a start location, but no further information. After a participant arrived and verified their location with a "selfie", they would receive the next location. A total of two additional locations were assigned, so that the shortest possible route would be about three kilometers long (30 minutes) and pass through residential neighborhoods of both sects. The resulting route consists of two paths: from start to midpoint, and from midpoint to endpoint. These locations are shown in figure 1-5, along with nearby sectarian neighborhoods.

In a follow-up survey, walkers reported their familiarity with the area and whether they stopped

to ask for directions in out-group areas, among other questions. To supplement our phone-based location data collection, walkers also self-reported their routes by drawing on a map of the treasure hunt area.

Nonresponse, Compliance, and Attrition

We gathered a total of 102 paths from 55 unique participants, containing an average of 21.1 correlated decisions per path.[16] Nonresponse was high, in large part due to extremely low participation by women (30 percentage points lower than men, $p = 0.01$). This pattern was predictable, given local gender norms. We were aware that treasure-hunt nonresponse would be higher among women, but chose to keep both genders in the study so that they would not be excluded from social media and other experimental modules. Other observable characteristics—sect, home district, age, or income sufficiency—are uncorrelated with nonresponse (participation and completion rates are given in table 1.2). Based on conversations with supervisors, subjects' participation in the treasure hunt appears to be largely driven by whether someone was willing to exercise for 30–60 minutes. However, we cannot rule out the possibility that subjects who are more averse to out-groups are also less likely to participate.[17]

Compliance with the task protocol was imperfect. A handful of early participants were accidentally informed of all targets at once, allowing them to walk the route in a different order than intended. In Ghazaliya, other participants visited a different grocery store than originally intended, due to ambiguous instructions. I address these deviations by assuming that they are independent of

[16]Five of these paths are instances where the participant returned to the starting point (thus adding a third leg) or revisited the treasure-hunt area at some point over the weekend and happened to walk a different route between the same locations. The latter is unsurprising since our locations are commonly visited by locals. I include these in the analysis because they provide additional information about walking patterns.

[17]One way this might happen is if highly averse people are also less mobile or active, perhaps due to fear. It might also arise if participants were aware of the sectarian nature of the task before starting, although steps were taken to prevent this. Walkers were asked not to discuss the task with other subjects, and they were unaware that all participants were assigned an identical set of targets in a neighborhood. In addition, in open-ended responses from debrief surveys, we saw no indication that subjects realized the treasure hunt was intended to send them to out-group areas. However, subjects were allowed to complete the treasure hunt at any time over a two-day period, and we cannot rule out the possibility that later participants might be aware that some targets are located in out-group areas.

out-group aversion, then conditioning on the start- and endpoints of each segment. If participants deliberately choose one store to avoid certain areas, the violation of this assumption will bias estimates toward zero.[18] Participants were free to withdraw at any time, and 13 failed to complete the treasure hunt despite succeeding in the first leg. This attrition was distributed evenly by participant's sect or home area. Anecdotally, it seemed to occur when participants learned that their next target was in an out-group area. This suggests that more-averse participants are more likely to withdraw, which will bias estimates toward zero. There is some evidence that lower-income participants have a lower attrition rate ($p = 0.085$), which could be due to the financial incentive for completion. In general, however, reactions to the task were quite positive. Typical responses in the debrief survey conveyed excitement ("an enjoyable new experience", "I felt adventurous", "I liked exploring") or discussions of the physical exertion ("tiring but entertaining", "good for health", "healthy exercise").

Table 1.1: Summary statistics for Sunni and Shia subjects, with p-values from t-test and chi-squared tests for continuous and binary variables, respectively.

	Sunni mean	Shia mean	p-value
Ghazaliya	0.72	0.62	0.36
Jihad	0.28	0.38	0.36
Age	20.9	21.3	0.21
Male	0.71	0.60	0.30
Income sufficient	0.57	0.53	0.87
N	62	58	

[18] If noncompliance is associated with higher levels of out-group aversion—for example, if more-averse subjects walk in the wrong order because it allows them to avoid out-group neighborhoods more easily—then conditioning on this decision will result in attenuation bias in estimates of out-group aversion.

Table 1.2: First and second columns describe response rates by subjects in various subgroups, with p-values for the difference based on a multivariate probit regression. Third and fourth columns describe completion (non-attrition) rates among subjects who started the treasure hunt.

	Participated in activity	p-value of difference	Completed both legs	p-value of difference
Sunni	0.48	0.78	0.80	0.43
Shia	0.43		0.72	
Female	0.27	0.00	0.91	0.36
Male	0.56		0.73	
Ghazaliya	0.51	0.15	0.76	0.36
Jihad	0.36		0.79	
Income sufficient	0.52	0.38	0.83	0.08
Income insufficient	0.41		0.65	

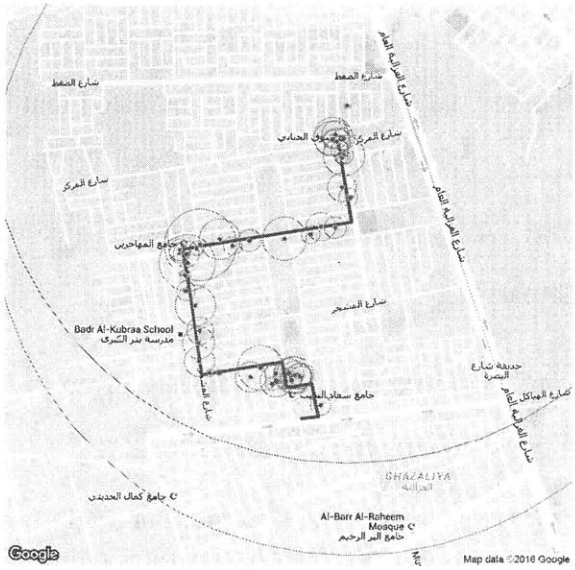
Data

Christia and Knox wrote a custom Android app to record the location of participants at one-minute intervals, as well as the accuracy of the location estimate. Because this app was active for the entire week-long study, we used a mix of GPS- and Wi-Fi-based measurement to compromise between accuracy and power consumption. GPS estimates are typically higher quality, with an accuracy within 10 meters in outdoor areas, but we find that Wi-Fi based location is generally sufficient for our purposes (often on the order of 10 to 100 meters) and that the far lower power requirements of Wi-Fi counterbalances its lower accuracy for long-term tracking. Walking routes are constructed from location data as shown in figure 1-6. When smartphone location data is unavailable due to technical issues, I use self-reported routes that participants drew on a map during debrief.[19]

I model treasure-hunt routes as random paths on a street network (modified from OpenStreetMap, 2016), where nodes are intersections and edges are street segments. Covariates are indexed as follows: i denotes the current node (the start of a step); j is the node to which the walker is moving (end of step); l represents a leg of the treasure hunt (e.g., from the mosque to the school); and k denotes an individual. Edgewise covariates that describe each street segment are coded based on

[19]When both sources are available, they are generally consistent.

Figure 1-6: Example of smartphone measurement of walking route. Blue points represent estimated locations. A thin circle is drawn around each point, with a radius corresponding to the estimated accuracy. A path that aligns these points with the street network is manually fit to the location data.



satellite imagery and input from our mixed-sect team of local advisors who were familiar with the area.[20] Additional variables are taken from the debrief survey. These covariates are given below.

- $\texttt{dist}_{i,j}$: Length of street segment (1 = 100 meters).
- $\texttt{direct}_{i,j,l}$: Directness of approach, or how much closer the $i \rightarrow j$ step brings a walker to the endpoint of leg l (in units of 100 meters).
- $\texttt{enclosed}_{i,j}$: Indicator for narrow enclosed streets in densely built residential areas. Coded from satellite imagery.
- $\texttt{thoroughfare}_{i,j}$: If street is a major thoroughfare, e.g. a major commercial avenue. Coded from local knowledge and satellite imagery.
- $\texttt{sunni\_walker}_k$ and $\texttt{shia\_walker}_k$: Participant sect (binary).

[20]Coders were given printed maps of the area, which they annotated by drawing borders around single-sect residential areas. Coders provided additional information about these areas, such as their safety level and professional composition. Major thoroughfares were traced and described, and major landmarks such as schools were marked. Annotated maps were then manually aligned and digitized.

- $\texttt{familiar}_{k,l}$: In debrief survey, whether participant indicated that they were familiar with the endpoint of leg l (binary).
- $\texttt{outgroup}_{i,j,k}$: Whether street passes through an area dominated by out-group residents (binary). Boundaries are based on local team's knowledge of neighborhoods.
- $\texttt{safety}_{k}$: In debrief survey, whether participant indicated that safety was a factor in their route choice (binary).
- $\texttt{familiar}_{k,l}$: In debrief survey, whether participant indicated that they were familiar with the endpoint of leg l (binary).

Examples of thoroughfares and enclosed neighborhoods are given in figure 1-7. Street networks are shown in figure 1-8 with these geographic covariates.

Figure 1-7: Annotated satellite imagery in Jihad, near starting point. Intersections are marked with dots, with lines depicting the connecting streets. Major thoroughfares (thick lines) cross near a bus station, at lower left. Streets that are fully enclosed by residential areas are drawn with dotted lines.

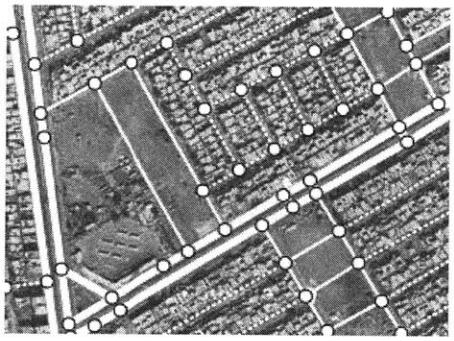
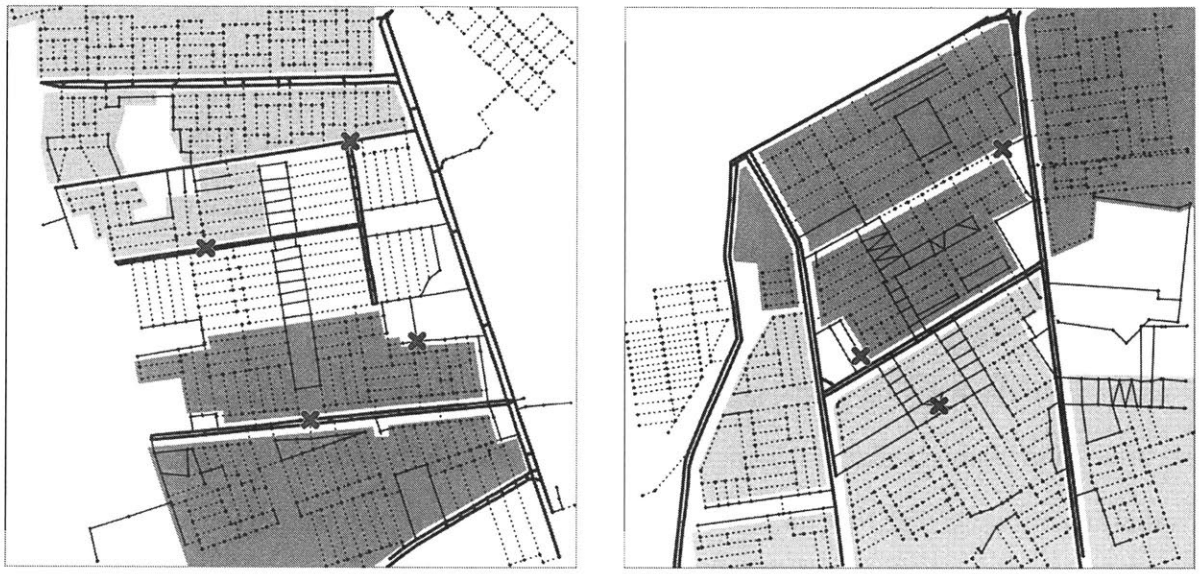


Figure 1-8: Street network in Jihad (left) and Ghazaliya (right). Targets are marked with a black $\times$. Sunni- and Shia-dominated residential areas are highlighted in dark red and light blue, respectively (neutral or mixed regions are left white). Major thoroughfares are indicated with solid thick lines, "open" streets in residential areas (e.g., bordering a park) are in solid thin lines, and "enclosed" streets in densely built residential areas are drawn in dotted thin lines.



Model and Results

I first describe the preferred model specification and describe the results from this model. I then test alternative explanations and robustness with other specifications. Estimates and 95% credible intervals for all models are given in figure 1-9. Finally, to interpret these results, I describe how the sectarian landscape affects Baghdadis' route choices in several simple but realistic scenarios.

Baseline Specification and Results

In the baseline specification, edge weights are modeled with the softmax function,

$$\varepsilon_{i,j,k,l} = \frac{\exp\left(z_{i,j,k,l}\right)}{\sum_{j' \in \mathcal{N}_i} \exp\left(z_{i,j',k,l}\right)}, \text{where}$$

$$z_{i,j,k,l} = \beta_{\text{dist}}\texttt{dist}_{i,j} + \beta_{\text{direct}}\texttt{direct}_{i,j,l} + \beta_{\text{enclosed}}\texttt{enclosed}_{i,j} + \beta_{\text{thoroughfare}}\texttt{thoroughfare}_{i,j} + \beta_{\text{SunniOG}}\texttt{sunni\_walker}_k \cdot \texttt{outgroup}_{i,j,k} + \beta_{\text{ShiaOG}}\texttt{shia\_walker}_k \cdot \texttt{outgroup}_{i,j,k}.$$

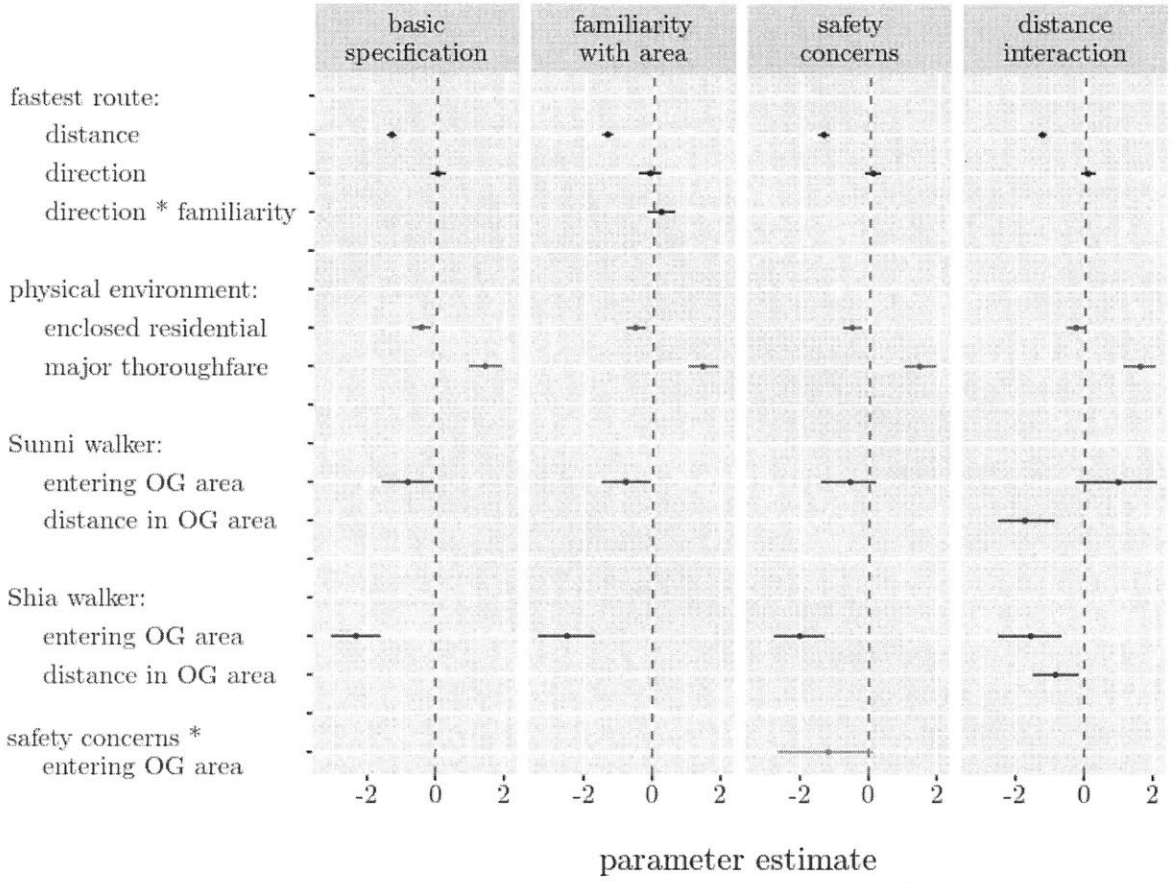
Note that the base terms for `sunni_walker` and `shia_walker` are omitted, because their coefficients are statistically unidentified. (From the perspective of a walker standing at any node i, all step options would have an identical constant added to the linear predictor, C; this can be rewritten as a multiplicative constant e^C on both the numerator and denominator, then canceled.) Three MCMC chains were run with 10,000 iterations per chain; convergence diagnostics are shown in appendix A.3.

Results from the baseline model show that minority Sunnis prefer to avoid entering out-group (Shia) areas. For Sunni walkers, the coefficient on "enter out-group area" is negative and statistically significant, indicating that these streets are less likely to be selected. However, Sunni aversion is relatively small compared to Shia aversion: Members of the Shia majority are significantly more reluctant to enter out-group (Sunni) areas. These results show that even after eliminating need-based reasons to avoid out-group areas, Baghdadis still exhibit sectarian taste-based aversion. Moreover, taste-based aversion is significant even among what is perhaps the best-integrated subpopulation in Baghdad, suggesting that it is likely stronger among the rest of society.

Other estimates are generally intuitive. Participants prefer shorter routes (negative coefficient on distance), avoid dense residential developments, and prefer to walk on major thoroughfares. A null result was found for the directness of route. This may be because directness and distance essentially measure the same concept—a walker who takes the shortest route to a destination is also moving in the correct direction—but in principle, including both terms leads to a more flexible specification because walkers may respond to long steps differently under certain circumstances, e.g. when they overshoot the destination. Findings for these covariates correspond well with anecdotal evidence from participants, who reported that thoroughfares such as commercial streets were "exciting," with more activity and shops, and that they "had no reason" to cross through

residential areas where they "did not belong." Coefficient estimates for the basic specification are interpreted in section 1.5.5, and alternative specifications are discussed in section 1.5.5.

Figure 1-9: Results from all model specifications. Points are posterior means; error bars are 95% posterior credible intervals. Distance is measured in units of 100 meters.



Interpreting Results

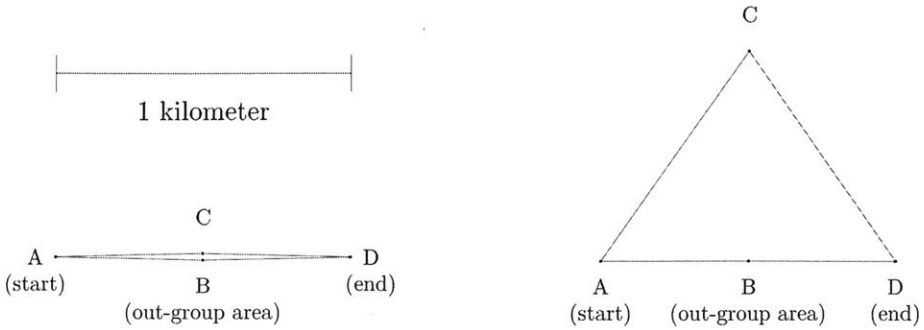
Coefficient estimates in figure 1-9 can be interpreted relative to each other. For example, the negative coefficient on `distance` is roughly the same size as the positive coefficient on `thoroughfare`, so the "cost" of walking 100 additional meters could be offset by the "benefit" of staying on a major thoroughfare. However, this sort of interpretation paints an incomplete picture of the iterative decision-making process that RPM is designed to model.

Instead, consider a hypothetical task in which a walker must decide whether or not to cross through an out-group area to get to a destination that is one kilometer away. Figure 1-10 presents two versions of this task. In the first scenario, both routes are equal in length. The corresponding estimates in the lower panel (leftmost error bars) show that both Sunni and Shia are significantly more likely to avoid the out-group route when it is costless (no additional distance required). Sunni will cross through Shia territory 31% of the time (significantly less than the 50% chance if sect were irrelevant), and Shia will choose to cross through Sunni areas only 9% of the time. In the second scenario, the out-group route is shorter, and the alternative is twice as long. In this case, the rightmost error bars show that both Sunni and Shia would almost certainly cut through out-group areas to save a kilometer of additional walking, as in the second scenario. Estimates in the middle show how these decisions change as the distance tradeoff shifts between these two extremes: Sunni become exactly indifferent between the route options when out-group avoidance "costs" 120 meters, or roughly a one-eighth increase in walking time (95% credible interval [10m, 240m]).[21] Shia are more reluctant, becoming indifferent at 340 meters—about a one-third increase in walking time ([240m, 460m]).

Figure 1-11 demonstrates the estimated behavior of participants in more complex scenarios, when many options are available, by simulating 1000 walking routes using point estimates of the RPM parameters. These plots show that in aggregate, members of the Sunni minority will tend to deflect slightly away from the most direct route to avoid Shia areas, but many individuals are willing to pass through. In contrast, Shia walkers stay just outside the border of the Sunni area, with almost no individuals cutting across. Members of both sects will go far out of their way to use a major thoroughfare, instead of walking through residential areas.

[21] For added distances in this range, the route choice probability is not significantly distinguishable from 0.5 at the 95% credible level.

Figure 1-10: Top panels depict scenarios in which a walker must choose between two potential routes. $A-B-D$ passes through an out-group area, while $A-C-D$ does not. In scenario 1, these routes are of equal length, so there is no incentive to pass through the out-group area. The corresponding estimates (leftmost error bars) show that under these conditions, both Sunni and Shia significantly avoid the out-group route. In scenario 2, $A-C-D$ is twice as long as $A-B-D$, so walking through the out-group area saves one kilometer. The rightmost error bars show that this provides a sufficient incentive for both Sunni and Shia to overcome their aversion. Estimates in the middle show results for various intermediate scenarios between these extremes.



(scenario 1) (scenario 2)

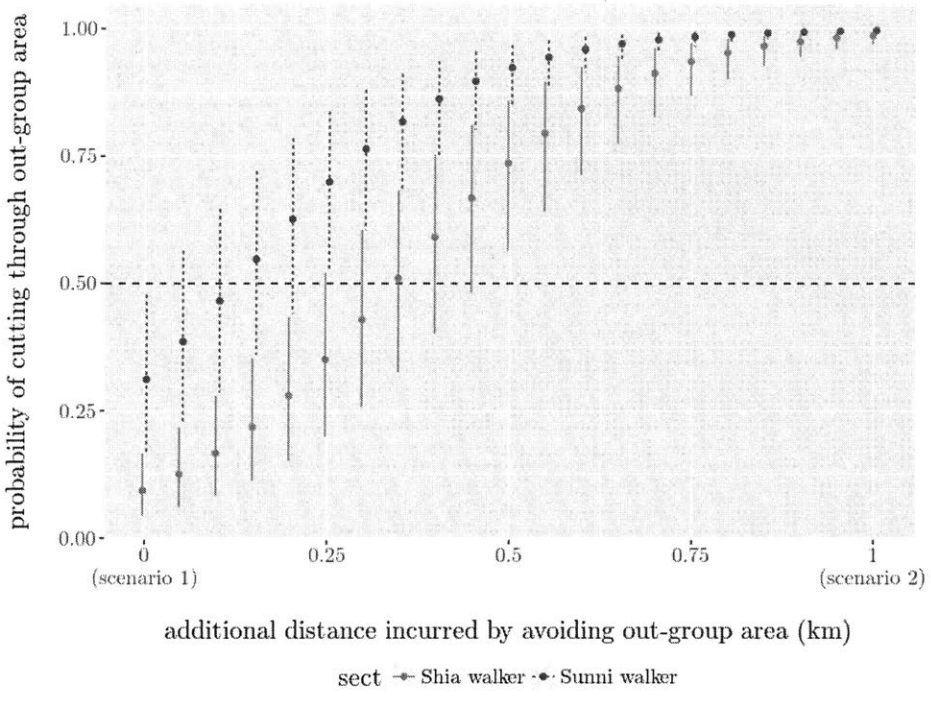
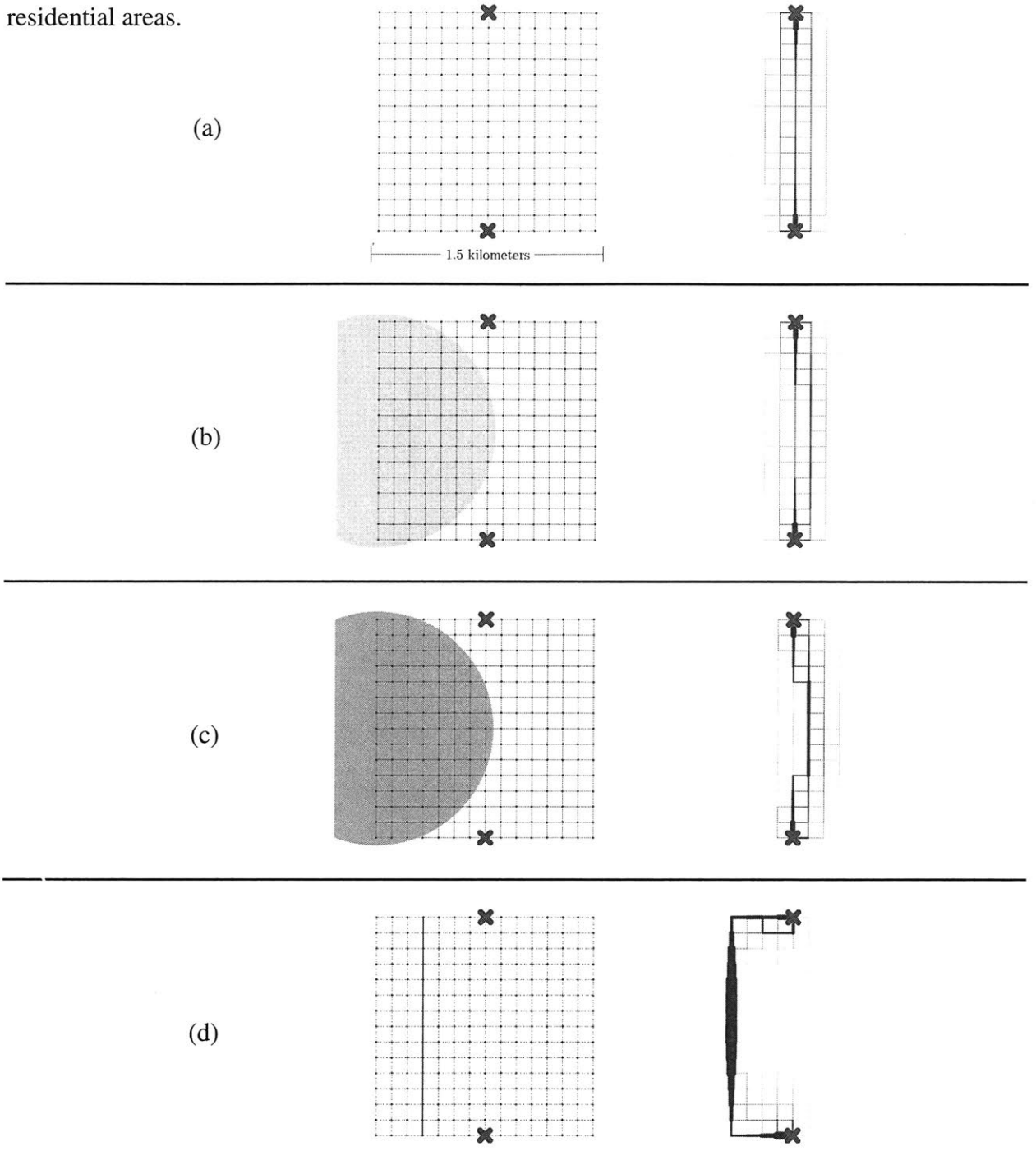


Figure 1-11: Estimated RPM distribution of walking routes in various scenarios. Each column depicts a hypothetical scenario (upper panel) and 1000 simulated walking routes (lower panel), using point estimates of RPM parameters from Baghdad data. In scenario (a), a walker crosses from bottom to top in a 15×15 square lattice; in the lower panel, thicker edges represent streets that are more likely to be used. In scenario (b), the walker is Sunni and the light blue region is Shia dominated. In scenario (c), the walker is Shia and the dark red region is Sunni dominated. In scenario (d), a vertical line is [illegible] :nclosed residential areas.



These results illustrate a strength of the RPM—that they illuminate how changes in short-term incentives can have broader implications for long-term behavior. Unlike existing alternatives, it is explicitly designed to model the path formation process. Thus, researchers can use the model as a tool to examine counterfactual quantities of interest, both by manipulating the walker's preferences (How much more efficient would people be if their aversion could be eliminated?) or altering the network context (How would the same person navigate a different neighborhood? How can urban planners design cities to encourage inter-group contact despite residents' preferences?).

Eliminating Alternative Explanations

In this section, I discuss alternative explanations for these findings and rule them out with additional model specifications. These alternative explanations are ruled out and results are shown to be robust.

One potential concern about the treasure hunt format is that participants might move differently in areas that they know better. If they happen to be more familiar with in-group areas, the difference might distort the estimate of out-group aversion. To address this issue, I add an interaction between `familiarity` with the target (as reported by participants in a debrief) and `direction` toward that target.[22] The basic idea is that people who know the location of their destination are more likely to walk directly toward it. The second panel of figure 1-9 shows results from this alternative specification. The estimate on this interaction is positive, as expected, but small and insignificant. Other results remain unchanged. Thus, there is no evidence that familiar subjects take more direct routes, and familiarity is not driving the other results. One possible explanation, discussed in section 1.5.5, is that people who are unfamiliar with the target are more likely to ask for directions.

I also consider the possibility that our design did not fully address participants' concerns about safety. In debrief surveys, 16% of participants said that safety was a consideration in their walking

[22]The base term, `familiarity` with the endpoint of a treasure-hunt leg, again drops out of the model because it is constant for all edges in a leg.

routes, despite efforts to choose a safe playing field. If participants avoided out-group areas due to a perceived chance of violence, then their walking patterns would not necessarily indicate taste-based aversion. To test whether this is the case, I interact `safety` concerns with an indicator for out-group areas ($\texttt{sunni\_walker} \cdot \texttt{shia\_area} + \texttt{shia\_walker} \cdot \texttt{sunni\_area}$). The third panel of figure 1-9 shows that, as expected, safety-conscious individuals are much less likely to enter out-group areas ($p = 0.074$). However, results among non-safety-conscious participants are broadly similar: Shia remain highly and significantly averse to Sunni areas, and Sunni are slightly averse to Shia areas (but statistical significance for Sunnis drops to $p = 0.146$).

Finally, I examine whether results change when considering the distance walked through out-group territory, rather than a binary indicator for merely entering it. In the fourth panel of figure 1-9, two estimates are reported for Sunni walkers—the "baseline" aversion, or the initial discomfort of setting foot into a Shia area, and the "marginal" aversion, or the additional discomfort of each additional 100 meters in Shia territory. The first is insignificant, but estimates show that Sunnis are increasingly reluctant to walk through Shia areas when the distance grows longer. When the two are added together to represent the total discomfort of walking 100 meters across Shia land (a typical distance between intersections), the resulting aversion is comparable to previous estimates and significant at the $p = 0.060$ level. For Shia walkers, both baseline and marginal aversion are significant and in the expected direction.

Discussion

These results strongly suggest that the segregation caused by sectarian conflict is likely to persist beyond the end of this conflict. Aversion is significant even among young and well-educated participants who were primed to explore unfamiliar terrain by the treasure-hunt nature of the task.[23] Among the general population, and during daily life, reluctance to enter out-group-dominated ar-

[23] In open-ended debrief responses, participants wrote, e.g., "entered new areas", "I saw new things", "it was fun to meet new friends."

eas is likely to be stronger. It almost certainly translates to an unwillingness to live in these areas. This alone is a sufficient condition for persistent segregation, even if people are still comfortable in mixed neighborhoods (Young, 1998, among others). The intuition behind this literature on "tipping-point" segregation is that when an equally mixed community begins tilting in one direction due to random migration, members of the smaller group will start to flee with increasing urgency. Moreover, segregation tends to be a one-way street, because individuals do not want to be the first to move back into an out-group community.

Heterogeneity in out-group aversion must be interpreted with care. Sectarian identities are not readily manipulable, and their "effects" are arguably undefined (Pearl, 2000; Woodward, 2003). Moreover, our Sunni and Shia participants are not a representative sample of Baghdad residents, and while they are comparable on observable variables, it is possible that unobserved individual- or neighborhood-level variables (e.g., subjects' past exposure to violence or the hostility of neighborhood residents) may be driving this heterogeneity. With these issues in mind, it is interesting to note that Shia participants were more uncomfortable in Sunni areas than vice versa, even though Shia groups have been in power for over a decade. More efficient behavior by Sunnis may reflect a necessary adaptation in a Shia-dominated society. Adaptation would accord with findings from Christia et al. (nd) in a separate study of young Baghdad cafe-goers, where Sunnis are shown to have developed more efficient network strategies for accessing public services: There, among other differences, Sunni were found to be more willing to seek Shia assistance when necessary.[24] Strong aversion among Shia is particularly troubling given their majority status and political power—if Shia are resistant to re-integration, it is unlikely to succeed. Zhang (2004a) develops a model showing that segregation remains stable even with one-sided aversion. This is because one group is willing to pay a premium for housing to avoid the other, and the non-averse group moves away

[24]Beyond geographic movement and access to public services, we also examine how information is disseminated in sectarian networks in ongoing work. A hypothesis of particular interest is whether Sunnis transmit information more efficiently among themselves, and if so, whether this is because of the tighter-knit structure of Sunni social networks or a behavioral adaptation.

in response to price incentives.

Out-group aversion has implications that extend beyond segregation, particularly given Shia control over a wide range of government resources. It also reduces the chances that gradual contact will lead to improved relations. By shaping the way that people roam around their own neighborhoods, aversion in walking routes may also lead to fewer interactions with out-group members. In the debrief survey, participants reported whether they asked for directions in each leg. The only significant predictor of "asked for directions" was whether the participant was familiar with their target. In particular, those walking through out-group areas were no less likely to stop and interact with locals—if anything, they were slightly more likely to ask for directions. One possible interpretation is that people are averse to the out-group as a whole, but they are still willing to interact with individual members when required, e.g. at university. If this is the case, walking patterns are reducing the chances for positive contact. However, no data is available on the sect of the person approached, and these results may also be confounded by individual heterogeneity.

Future Directions

In this paper, I describe a type of data that is common but underutilized in political science—path data. I show that while dependence between observations presents a statistical challenge, it does not prevent inference on factors shaping the trajectories of paths. The proposed random-path model allows social scientists to assess a wide range of previously untestable hypotheses.

However, much work remains to be done. An `R` package for random-path models is under development and will provide software tools for constructing and cleaning path data, estimating models in a familiar interface, conducting model diagnostics, and visualizing results. In future work, I plan to adapt common statistical procedures for the RPM case. This includes an approach to model selection using likelihood-based cross-validation (van der Laan et al., 2004) and sensitivity analyses to test robustness to omitted variables.

Finally, the RPM can be extended to estimate the effects of path-assigned treatments. In ongoing work, I use a model of the assignment process (Rubin, 1991) to simulate a distribution of possible treatment assignments. In the highway case, these can be thought of as proposed highway routes, among which one route is ultimately selected for construction. The resulting distribution over paths is well-suited for incorporation into Bayesian multilevel models or for drawing causal inferences. For example, this approach allows testing of sharp hypotheses of treatment effects in the Fisherian framework (Rosenbaum, 2002), without the need for untenable assumptions about spatial dependence in the outcome variable. Following Bowers et al. (2013), models of treatment assignment can also be used to evaluate hypotheses about interference between units—a particularly important question in connective infrastructure, where highways can either generate spillover growth in adjacent, unconnected communities or lead to out-migration and decline.

Paper 2: Experimental Evidence on Sect and Network Access to Services in Baghdad

(Joint work with Fotini Christia and Jaffar Al-Rikabi) In August 2015, while Iraqi forces were mired in a protracted fight against the Islamic State, tens of thousands of Iraqis took to the streets of Baghdad to protest against electricity blackouts. Carrying coffins draped in black shrouds inscribed with the words "electricity, water, services" and holding signs proclaiming their "death by sectarianism," Iraqis protested against the under-provision of public goods and services and the sectarian divisions between majority Shia and minority Sunnis that are seen as driving them. These rallies, which lasted six weeks and spread across eight Iraqi provinces, were also supported by the chief Shia religious authority in Iraq, Grand Ayatollah Sistani, who called for an end to the sectarian politics that have crippled Iraqi institutions since the US invasion and have brought public services to a halt. These calls, which were echoed as recently as 26 February 2016 at a massive Baghdad rally organized by prominent Shia Iraqi militia leader Moktada al-Sadr, reflect an empirical regularity in political economy: that fewer public goods and services are available to members of ethnically divided societies. Though segregated social networks are presumed to be a major driver behind this underprovision, their precise role remains largely under-theorized.

Much existing work has looked at how divisions such as ethnicity impact the availability of public goods. Research on this phenomenon at higher levels of aggregation, e.g. among local governments, is complicated by practical and ethical difficulties in randomizing the provision of

public goods and services. However, from the perspective of a citizen, the supply of these resources is only one part of the story. Our study helps fill this gap by showing that even when faced with the same task in the exact same local government office, members of different groups have significantly different abilities to access public goods and services, and that these gaps are in large part driven by the network strategies used to gain access. Specifically, we conducted a small world experiment in which participants were asked to draw on their social networks to obtain information about a local government service. After randomly assigning participants to in- or out-group neighborhoods as a way to measure the extent of access discrepancies due to sect, we also tracked their search behavior to understand the network dynamics driving such discrepancies.

We draw on the case of Iraq where the prominent societal cleavage is sect, and where Sunnis and Shiites control different official and unofficial resources. This network segregation implies that individuals of different sects have very different levels of access to services. While important and relevant, Iraq is also not exceptional. Sectarianism has become a salient identity divide in an array of strategically important areas of the Muslim world—such as Syria, Yemen, Pakistan, Lebanon, and Bahrain—and can arguably even capture majority-minority dynamics in ethnically divided societies more broadly. The intent is to use the case of Iraq to answer the broader social scientific question of how divided networks affect access to services and what the strategies are that people use to navigate a segregated society.

We find, contrary to our initial expectations, that the politically dominant majority Shia group is substantially less able to access public services than the minority Sunni group. Shia underperformance is still evident after controlling for a host of demographic characteristics in a variety of ways, including in the Shia target neighborhood where they should enjoy an in-group advantage. Using data from the search process, we present evidence that this gap is driven by two sets of mechanisms: (1) relative to the minority Sunni group, Shia participants pursue an inefficient network search strategy that relies on lower-quality contacts, and (2) Shia are less able to leverage their social ties into costly assistance from their contacts. One explanation that is consistent with

these results is that minority groups evolve better strategies for obtaining resources to which they would otherwise be denied access.[25]

Literature and Hypotheses

The lower levels of public goods in more fractionalized societies has been established in contexts across the developed and developing world and as such has been called one of the most robust findings in political economy (Alesina and Easterly, 1999; Banerjee et al., 2005; Miguel and Gugerty, 2005; Habyarimana et al., 2007, among others). Many have also looked at underlying mechanisms, including divergent preferences (Alesina and Easterly, such as) and the role of shared language, customs and norms, including social networks (termed "technologies of communication" by Habyarimana et al.). These results help explain why certain societies provide fewer public goods and services as a matter of policy. However, the mere availability of these resources is not sufficient for in- and out-group citizens to receive actual benefits. Our experiment examines whether citizens are able to access these services in a divided society and how they draw on their social networks to do so. Our results shed light on citizens' expectations of access, which likely plays an important role in societal support for providing these public goods. We are, however, not able to examine the dynamic interplay between these factors and how they influence future public goods provision.

Our focus on social networks is consistent with an increased interest in their general role in development—be it in how social networks affect the diffusion of micro-finance (Banerjee et al., 2013) or how they may affect the successful integration of winners of housing lotteries (Barnhardt et al., 2017). Our focus is specific to the context of public services, where we aim to assess whether differential access is due to differential use of within-group ties, lack of cross-group social ties, or whether cross-group ties exist but cannot be effectively leveraged for assistance. Existing works

[25]It is worth noting that Ba'athist rule most likely predates the political memory of our participants, who were typically 9 years old when the regime fell. Thus, this alternative explanation would require that political knowledge be passed down through a sort of group memory effect, or through their parents.

have theorized the societal and economic implications of a network's structure and how people navigate them (Rauch and Casella, 2001; Kleinberg, 2000; Granovetter, 2005, among others)). A rich literature on the "small world" phenomenon shows that seemingly distant members of society, such as a common citizen and the gatekeeper of a public service, are often connected by a small number (typically given as "six degrees") of intermediate contacts. Perhaps more surprising than the mere existence of these short paths, however, is the fact that across a variety of settings, individuals are often able to find these short paths even without knowing the global social network structure (Milgram, 1967; Dodds et al., 2003b). This well-established result highlights the important role of a search strategy, or the heuristics that people use to identify social resources—for example, reaching out to intermediate contacts based on geographic location or occupation (Killworth and Bernard, 1978). However, a common critique of small-world experiments (Schnettler, 2009) is that they tend to only examine successful searches—e.g. letter-forwarding chains that eventually reach the intended recipient–which limits the inferences that can be drawn (Geddes 1990). Our study collects data on search behavior for both successful and failed attempts to access public services, which helps identify search strategies that are associated with success. Here, we are interested in whether certain groups are more likely to approach high-quality contacts.

A second, closely related question is how, after identifying a contact, the searcher is able to secure costly assistance such as information or an introduction to yet another contact. A substantial literature on the sociology of job searches suggests a tension between the use of weak and strong ties: While strong ties facilitate the exchange of favors (for example Bian, 1997), they are also associated with tight-knit, homogeneous groups in which members have similar knowledge and gains to trade are thus limited. In contrast, weak ties are more likely to "bridge" the searcher to distant parts of the network where fresh information is more readily available (Granovetter, 1973).

Though these literatures raise interesting questions about the ways in which different groups may attempt to access public goods, they offer little in the way of clear theories about how people will navigate divided social networks. This is precisely the gap our study aims to cover. There

are, however, some suggestive hypotheses that are readily derived from existing works and which we used to identify what behaviors to look for and what measures to collect. Specifically, we anticipated the following outcomes:

H1: People belonging to the group in power are more likely to have better access to public goods and services.

The hypotheses below examine the network-related measures that may drive the differential mechanisms for success and failure.

H2: People will more readily turn to in-group ties than out-group ties for information or assistance. This is consistent with the role of strong ties in the literature that allows for higher levels of trust and reputational gains, as well as higher levels of enforcement and compliance.

H3: There will be higher levels of success among those who can effectively reach across groups when necessary. This draws from findings on "the strength of weak ties" and the role these contacts play in bridging disparate groups.

H4: We expect people to seek out contacts not merely based on their sectarian affiliation, but also using other heuristics to identify better-informed individuals. For instance, we expect people to reach out to older or more educated contacts, as these are more likely to have information about accessing public services.

Experimental Design

Intervention

The intent of our study is to draw inferences from actual behavior in a realistic task, rather than online behavior or stylized behavioral games as much of the existing literature. After carrying out qualitative fieldwork to identify a meaningful task in the Baghdad context that would allow us to measure access to public services, we settled on the act of obtaining a residency certificate from local council authorities. Such a certificate is required not only for access to services such as connecting a residence to electricity/water or enrolling a child at a local school, but also for job applications, voter registration, and moving, among other things.

In Baghdad—as in other areas where bureaucracies are less professionalized and political knowledge is low—citizens draw heavily on their social networks to access public services. Inefficiencies in service provision and overall sectarian politics largely prevent people from simply showing up at local government offices to request services. Customarily, a citizen would first identify and establish contact with a gatekeeper who can help in such a task, and then initiate an in-person meeting to proceed with securing the service. Though we want to understand how people navigate sect-dominated bureaucracies to access services, we did not want to pose an actual burden to local councils in Baghdad with unnecessary requests for residency certificates. Participants were instead only asked to find the address of the local council and the name and phone number of a gatekeeper—an official in that council that could provide the service in an assigned Shia or Sunni neighborhood.[26] The experiment thus mimics the initial process an Iraqi would have to go through to obtain such government services.

These local councils are neighborhood level councils that were founded by the Coalition Provisional Authority (Order 71) in 2004 to "achieve greater efficiency and economies of scale in

[26]To validate answers, we compared the information provided by participants against a government-provided list of location and council members in the target office.

coordinating the provision of public services, and to facilitate unified and coordinated administration of cities." Their tasks include issuing local identity documents and monitoring provision of services such as water, electricity, trash collection and sewage. The average size of a council is 5–6 members, with each member representing a specific neighborhood block (with an average of 25,000 residents per block). Their members, however, are appointed rather than elected and have largely been the same since their initial appointment in 2004. Though they were in principle meant to represent both sects, initial appointees were overwhelmingly Shia, not only as a reflection of the demographic realities of Baghdad but also in large part due to the requirement that members have no past affiliation with the Baath party.[27] The leadership of these councils has also been Shia-dominated as it is council members that select the council head and deputy head. Shia over-representation is not particular to local councils: more broadly, the Shia-majority composition of the current government is reflected across a wide range of government bodies.

Sample Selection

Our participants are a sample of 317 young men who were recruited from well-frequented cafes in two socioeconomically comparable neighborhoods in Baghdad, described below. While this is a convenience sample, we argue it is an important one as it captures the first generation of post-invasion young adults in the Iraqi capital. Due to local gender norms that largely preclude women from socializing publicly with men, we focused our sample on male participants only. Men tend to be the head of households and political decision makers in Iraqi society and as such constitute an interesting and meaningful population to study. We used identical recruitment strategies in both neighborhoods, recruiting participants from similarly sized cafes with comparable prices for narghile, a favorite pastime of Baghdadi men. Recruitment took place after local working hours, when men go out to cafes close to their homes to socialize.

Each participant received a T-shirt as compensation for participating, as well as up to the equiv-

[27] This suggests that they can be Shia-dominated even in neighborhoods that may have a majority of Sunni residents.

alent of US$50 dollars in phone credit if they completed the task. This is a non-negligible amount of money that greatly facilitated recruitment—a laborer in Baghdad makes anywhere between $7-$30 a day depending on the level of skill, and a civil servant receives an average monthly salary of $500.

After we described the task at hand to participants and they consented to taking part, we asked them to come to a school in each of the two recruiting neighborhoods where the study took place. Each participant completed a short survey before the start of the study and was also asked to write down the phone numbers of potential contacts he might need for assistance with the task at hand. We then took away the participants' personal phones and gave each one a study phone so as to ensure they all had as much phone credit as needed for calls/text messages and no access to internet data.[28] Each participant was then assigned to a separate classroom, and once all participants were in their rooms, an enumerator notified them one-by-one of the Baghdad neighborhood to which he had been assigned. Participants had up to 2 hours to complete the task.

Participating Neighborhoods

Because of the sensitive political situation in Baghdad, one cannot openly ask people their sectarian affiliation. We therefore recruited from two neighborhoods with distinct sectarian identities as a way to sort participants from both sects. The two participating neighborhoods are across from each other on Baghdad's Tigris river: Shia-majority Kadhimiya, on the west bank of the river and Sunni-majority Adhamiya on the east (figure B-1). The neighborhoods are connected by a bridge called "the Imams' bridge," aptly named as it connects two monumental mosques in these neighborhoods: Imam Kadhim's shrine, named after the 7th imam in the Twelver Shia tradition (in Kadhimiya); and Abu Hanifa's mosque, named after the founder of one of the four schools of Sunni Muslim thought (in Adhamiya). Because of the different religious tradition of these two

[28]The information on council locations and council member names and phone numbers is not available on the internet.

mosques, the two neighborhoods are almost exclusively homes for people of the respective sects. Apart from the distinct sectarian tradition, the two neighborhoods compare well on a range of socioeconomic characteristics such as population size, illiteracy levels, and labor participation, as well as access to water and sanitation (see Table B-1).[29]

In addition to relying on the neighborhood of recruitment for identifying sect, we further validated people's sect through their name, as Sunni and Shia tend to have distinct names (identified through first name, father's name, grandfather's name and tribal name). As expected, among those with identifiably sectarian names, over 98% of participants were of the dominant sect of the recruitment neighborhood. We also checked to see how many participants were local residents of the respective neighborhoods. Sunni participants are more likely to be local than Shia participants, which is consistent with the differential mobility patterns of the two groups reported below. Reported results hold when adjusting for participants' home location and we also conduct robustness tests in which we subset to local residents only.

We considered it too easy a task to assign participants to get information for a residency permit in their own neighborhood. Instead, we identified two distinct target neighborhoods, one majority Shia and another majority Sunni, the neighborhoods of Zayouna and Ghazaliya respectively (figure B-2). Asking participants to secure a residency certificate in a neighborhood other than their own would approximate the experience of an incoming resident trying to secure a residency certificate at his new neighborhood. Unlike Kadhimiya and Adhamiya, we do not have the same type of neighborhood level data for Zayouna and Ghazaliya. These two neighborhoods are known to be majority Shia and Sunni respectively, and were selected because they are roughly equidistant from the recruitment neighborhoods (a 20–25 minute drive). Similar distance from the recruitment neighborhoods was an important factor to ensure that participants were not more familiar with

[29]There is limited systematic data on the neighborhood level in Baghdad. Though we collected what was available from an array of local and international organizations, local data was available only for a subset of Baghdad's neighborhoods.

one neighborhood over the other due to proximity.[30] Based on qualitative accounts, these neighborhoods also compare well on the type of residents (professionals such as teachers, lawyers and doctors) and on real estate prices as provided by local agents (at an average of US$1900 a square meter in both neighborhoods). Despite representing neighborhoods with very different sectarian compositions, for the political reasons discussed above, both have a 5-member local council with three Shia members (including the council head and deputy head) and two Sunni members (see Table B-2).

Data

We had a total of 317 participants, of which 153 were recruited in Kadhimiya and 164 in Adhamiya. Half from each group were in turn randomly assigned to either the Shia target neighborhood (Zayouna) or the Sunni target neighborhood (Ghazaliya).

Measurement

We carried out a short pre-experiment socio-demographic survey where we asked participants about their age, education levels, employment, daily contacts and weekly mobility around Baghdad.[31] The participants were then asked to keep a log of all the calls they made while trying to complete the task. There they noted who they were calling; how they knew this person; age, education and occupation of the contact; what type of information they were hoping to secure; how long the call lasted; and if they got the help they were hoping for. The study phones we used were smartphones that allowed us to also back up the numbers called and messages sent. This was a way for us to compare the participants' logs against the actual calls, i.e. self-reported data against actual behavioral data. Despite a handful of isolated reports of participants across the two recruitment

[30] Self-reported familiarity during the debrief confirms that participants were equally familiar with both targets.

[31] We were not able to directly ask questions about their sect or political affiliation and civic engagement as this is considered sensitive information in the Baghdad context.

areas altering self-reported numbers or deleting call logs,the data appear to be largely consistent: 87% of self-reported phone numbers appeared in the call logs, and approximate self-reported call durations generally track the exact figures recorded in call logs.

After completing their calls, participants filled out a debrief survey responding to questions on prior experience securing a residency certificate; their levels of religiosity; and offering a qualitative description on how they decided whom to contact to try to secure the requisite information for the task at hand.

Participant Characteristics

Participants compare well in terms of age, years of education, marital status and prior experience securing a residency permit (figure B-3). But Sunni and Shia are not just labels–these are distinct populations with differences that come across in both demographic and behavioral measures, and there is no reason to expect balance across these groups. Looking at employment, it is clear that Sunni participants are notably more likely to be enrolled as students, whereas Shia participants are more likely to be employed in the manual, private and public sectors (figure B-4). Sunni and Shia also differ in terms of years of residency in Baghdad. Specifically, we see that the Sunnis have more new arrivals to the city as recent as in the past year (with 17% of the sample in the past five years), who are likely to have been displaced from the conflict against the Islamic State in Mosul and the Sunni triangle (figure B-5). Going beyond socio-economic measures, we see that Sunnis and Shia also differ in religious behavior. Though participants from both groups reported similar levels of mosque attendance and Friday prayer, Shia participants were considerably more likely to read the Quran and listen to recorded sermons outside mosque (figure B-6).

There are also notable differences in the pattern of mobility between the two groups: we find that Sunni participants are more insular on a variety of measures. When asked about the neighborhoods they visit in their daily lives (for, e.g. work, study, entertainment, socializing), relative to Sunni participants, Shia reported visiting a significantly wider range of neighborhoods, at a greater

average distance from home, with higher frequency (figures ?? and ??). When asked to name people with whom they interacted in their daily life, Sunni participants also reported a lower number of contacts, and these contacts tended to live significantly closer to the participant's home.

Results

We first report results by levels of success on the task by participant sect (Sunni or Shia) and by target assignment (co-sect or cross-sect). We then unpack two sets of mechanisms driving success. First, we examine participants' search strategies, or the type of contacts that they approach. Specifically, we look at whether participants reach out across sectarian lines, as well as whether they move horizontally within their network (i.e., contact peers) or vertically (reach out to older, more educated people). Then, we evaluate the effectiveness of their strategies, or the extent to which they are able to secure assistance from these contacts. We examine several dimensions of effectiveness: the amount of contact engagement (rate of contact pick-up, number of calls, call duration); whether contacts subsequently participate in the search process by drawing on their own ties; and whether contacts ultimately provide accurate information. We also reference qualitative results, where participants were encouraged to reflect upon the factors that influenced who they decided to reach out to and why.

Success in Access

Overall, 39% of participants were able to get correct information on the general vicinity of the assigned neighborhood council office, including 17% who found the exact location. For the next, more challenging step, identifying someone who could help secure a residency permit, 17% were able to secure a name. Only 2% (three participants from each sect) found the most difficult piece of information, a correct personal phone number.

Looking at these purely experimental differences in success between randomly assigned tar-

gets, we find that both Sunni and Shia participants were significantly more successful in the Sunni target, both at locating the council address and identifying a gatekeeper (by an average of 20–30 percentage points depending on outcome, at p-values of 0.01 or below for both groups).[32] This suggests that despite our efforts to find similar target neighborhoods—and despite participants reporting similar levels of familiarity with both neighborhoods—the Sunni target office was simply easier. This could be due to an array of idiosyncratic factors and it is not something at which we can get through the setup of our experimental study.

In our raw results, we also found unexpected patterns of heterogeneity in success: Shia participants performed drastically worse than Sunni participants in both target neighborhoods, including the Shia target in which Shia participants should have enjoyed a co-sect advantage. This contradicted our initial hypothesis that Shia, as members of the dominant group, would have greater access to services across the board. The disparity was particularly puzzling in light of the other Sunni handicaps discussed above, such as a larger proportion of internally displaced persons, greater insularity, and the fact that councils are led by Shia appointees even in Sunni neighborhoods.[33]

To rule out the possibility that the higher rate of Shia failure is driven by observable demographic differences between Sunni and Shia participants, we adjust for these differences by regression, matching, and reweighting by entropy balancing (Hainmueller 2012).[34] Regression results

[32]Nonparametric p-values calculated by block-bootstrapped rank sum (council address) and difference-in-means (gatekeeper name) test statistics. Parametric models with ordered probit (address) and logistic regression (name) yield essentially identical results.

[33]Given this variation between councils, to hold the difficulty of the task fixed, we would need to randomize over a larger set of Sunni and Shia target neighborhoods, rather than one of each as in our experiment. This is not something that our current experiment can explain. What we can get at is the differences, if any, in the way the two groups utilize their networks to get at the requisite information and this is our focus.

[34]In regression models, we adjust for the following demographic characteristics: age (linear and quadratic), time residing in Baghdad (linear and quadratic), religiosity (1st principal component of six measures), marital status, educational attainment and current enrollment status, and work sector and current employment status. We also adjust for the time of day at which participants start the experiment, as well as the distance of the assigned target from their home neighborhood. Matching of Sunni participants to Shia participants is done one-to-one with replacement by exact-matching on the assigned target and "local" status (living in the recruitment neighborhood); then minimizing the standardized Euclidean distance on age, student enrollment status, whether the participant was university-educated, time in Baghdad, and start time. Reweighting is done such that Sunni and Shia participants who are assigned to a particular target are identical on mean age, time in Baghdad, school enrollment status, university education, start time, and distance from their home neighborhood to the assigned target.

show that in fact, demographic characteristics only weakly predict access. The only significant predictor of success, whether substantively or statistically, is a university education: Relative to high-school educated participants, college-educated participants are roughly 25 percentage points more likely to find information about both council addresses and gatekeepers. Given the role of Iraqi universities as a place where people form broader social networks and interact with those from the other sect, this difference in performance is unsurprising. In contrast, factors such as living closer to the target, marital status, employment status, age, length of residency in Baghdad, or religiosity are not significantly associated with success, and point estimates on the differences associated with typical changes in these covariates[35] are associated with increased success of at most 10 percentage points (and usually much less).

In these models, the difference between Sunni and Shia success on locating the council address loses significance, but the sign of the difference generally remains the same–Shia perform worse (Table B-3). However, we find substantial gap on the most difficult task, identifying a gatekeeper. Here, success is more heavily dependent on personal relationships (as gatekeepers' names are less readily available and verifiable than the council location). Overall, Shia are roughly 10 percentage points less likely to succeed (Table B-4). The gap is not merely driven by the Sunni target neighborhood—if this were the case, the difference could be explained by an inability to work across sect lines. Rather, if anything, the Shia performance gap is more robust in the neighborhood that is dominated by fellow Shia.

Network Search Strategy

To explain these differences in access, we explore the underlying network search dynamics of the two groups further. In this section, we discuss the heuristics by which participants choose contacts to approach for assistance. To offer readily interpretable results in terms of magnitude, we present each outcome for the median player using the covariate adjusted model (generally the

[35] One standard deviation for continuous predictors, or a unit change for binary predictors.

most conservative model, except as noted) and varying sect or target assignment.

As expected, we find that participants rely heavily on geographic cues when deciding whom to call. Shia typically socialize with people that live an average of one kilometer from their home neighborhood, and less-mobile Sunnis socialize almost exclusively with neighbors in their daily lives. Both groups turn further afield when searching their network: The average contact shifts away from home (Table B-5) and toward the assigned target (Table B-6) by about one kilometer, and these results are highly robust and statistically significant. This behavior is consistent with a large literature on letter-forwarding experiments[36] (Killworth and Bernard, 1978), but because both Sunni and Shia employ similar geographic strategies, it does not help explain Shia underperformance.

Next, we highlight two weaker results. First, in terms of education, Sunni participants typically reach upwards to contacts that have completed high school and may be better informed or embedded in a broader set of social ties to draw upon. Shia participants behave similarly when assigned to the Shia target, but when assigned to the Sunni target they tend to approach significantly lower-educated contacts. However, this result is not robust to alternative specifications (Table B-7). Second, during their daily lives, about one-tenth of contacts mentioned by Sunni and Shia participants' have names that identifiably belong to the other sect—an unsurprisingly small proportion in this heavily sectarian context. When assigned to the co-sect target, where participants have little reason to reach across sect lines, we observe almost identical patterns. When assigned to the cross-sect target, on the other hand, both groups of participants seek out a significantly larger proportion of cross-sect search contacts, demonstrating that such relationships exist but that people usually choose not to draw on them (Table B-8). Sunnis participants are highly responsive in this regard, increasing their Shia contacts by 25 percentage points when assigned to the Shia target, versus

[36]Letter-forwarding tasks are the canonical small-world experiment, and several variants exist. In the most basic version, participants are given the name, location, and occupation of a target individual to who they are not directly connected. They then send a letter to an intermediate contact, who forwards the letter to yet another intermediate contact, until the letter ultimately reaches the designated recipient.

a corresponding cross-sect shift of only 10 percentage points for Shia participants. Differential responsiveness may be partly attributed to structural factors,[37] but when taken together with the education gap, we interpret these results as suggestive evidence that Shia may have a shallow pool of social ties in the cross-sect target neighborhood and as a result are forced to seek assistance from suboptimal contacts.

Coming into the experiment, a long tradition of sociological research on job searches raised questions about whether we should expect participants to draw on stronger or weaker social ties. A priori, there are countervailing theoretical arguments: Drawing on more dispersed, weaker ties allows a searcher to explore a wider and more varied space, but stronger ties are more likely to provide costly assistance. Using two proxies of tie strength—the participant's length of friendship with the contact and whether the contact was a family member—we find empirical support for an effect in both directions. Relative to daily life, we find that participants reach out to more recent friends (Table B-9). Participants typically socialize with friends they have known for nine years, but search through friends they have known for six (differences across participant sect are not significant). For family contacts, however, we observe the opposite effect: Both Sunni and Shia draw more on family ties when searching, relative to daily life (Table B-10). This shift was particularly large and robust among Sunni participants, for whom nearly 30% of search contacts were family members (an increase of 16 percentage points over reported family contacts in daily life). For Shia participants, who similarly drew on family members more during search (14%, increase of 8 percentage points), results are also significant in the regression model but are less robust. In addition, we find that Shia rely significantly less on family than otherwise identical Sunni participants, both in daily life and in search. We argue that the reluctance of young Shia to reach out to family members is related to the well-established finding that Shia society is more heavily hierarchical; this is bolstered by the fact that Shia participants' contacts are, on average,

[37] In particular, Shia are better-represented on both target councils, and the Sunni target neighborhood has a larger Shia minority (estimated to be 40%, relative to the 20% Sunni minority in the Shia target neighborhood). These factors may mean that Shia participants simply have less need to reach out across sect lines.

more than 3.5 years younger than those of an otherwise identical Sunni participant (Table B-11).[38] Additional evidence, discussed below, suggests that the resulting inefficiencies in information flow are likely to be a handicap for Shia search.

These differences in search strategy are not explained by different levels of effort. Both Sunnis and Shia spend roughly half an hour on the task and turn to an average of two contacts for the search, irrespective of whether they were assigned to a co-sect or cross-sect neighborhood.[39] Sunni participants make an average of 3 outbound calls irrespective of whether they are assigned to a co-sect or cross-sect area, and Shia also make 3 calls in the cross-sect area but 5 calls in the co-sect area. This could reflect a higher number of relevant individuals to contact, but in any case, indicates that the apparent lower quality of Shia contacts is not merely an artifact of lower effort.

We also collected qualitative information about the participant's search process at the end of the study when we asked participants to reflect on the primary factors that motivated their search strategies. While both groups mentioned turning to people who they considered to be well-connected, Sunnis were (1) nearly 40 percentage points more likely to mention the importance of trust as driving their search strategy and (2) roughly 15 percentage points more likely to mention selecting contacts based on how knowledgeable they were perceived to be, which they often described explicitly by referencing education or the epithet "resident of Baghdad," a local term roughly corresponding to "born and bred" and more sophisticated than people who are new arrivals from the periphery. In contrast, Shia participants were more likely to complain about the lack of local contacts as a hindrance in completing the task.

[38]Contacts of a typical Shia participant remain younger, by 2 years, than an otherwise identical Sunni participant even after controlling for whether the contact is a family member, but the difference becomes statistically insignificant.

[39]The small number of search contacts may suggest that participants expect their contacts to search further on their behalf, as discussed below. Shia also report roughly twice the number of contacts in their daily lives. Note that the number of daily contacts is a purely self-reported answer, while the number of contacts for the search is both self-reported in the contact log, and behaviorally validated via the smartphone log.

Search Effectiveness

While participants have complete agency over the type of contacts they choose to approach, or their search strategy, the assistance that they subsequently receive depends to a greater extent on variables beyond their control. For one, a contact may not have access to the desired information. Even when a contact knows or is able to procure the correct answer, the provision of accurate answers may be somewhat costly for a range of reasons, including the cognitive load of recalling an obscure piece of information; the opportunity costs of giving up potentially valuable knowledge; or, if the information is not immediately available, the effort and social capital expended through referral and indirect search. Perceived benefits depend on norms of reciprocity between participant and contact, including the expectation of repeated interaction; reputational benefits, which depend primarily on the number of shared friends; and any intrinsic value that the contact assigns to being helpful. For the most part, these factors are outside of the short-term control of an individual searching for service-related information, meaning that we can use various measures of the assistance received as proxies for the effectiveness of their chosen search strategy.

Anecdotal evidence, collected in an open-ended debrief interview, suggests that many participants reach out to contacts who help broaden their search, as opposed to merely providing information. This can take one of two forms: (1) a referral, in which the initial contact introduces the participant to a third party;[40] or (2) indirect search, in which the contact searches their own local social network on the participant's behalf.

To quantify indirect search, we turn to the phone call history, which records the start and end times of all calls to a particular number. Logistically speaking, an indirect search unfolds as follows: The participant calls a contact, asks for assistance, and the phone call is terminated.

[40] In the self-reported contact log, we attempted to collect data on the former: Participants were instructed to record, for every person approached, whether the contact was their own acquaintance or whether the contact was introduced by another "referrer." In addition, they were asked about their relationship with the referrer and the relationship between the referrer and the contact. Unfortunately, these questions appear to have been confusing for many participants—only eight referrals were reported, a far lower number than suggested by the qualitative interview.

The contact then makes a number of unobserved phone calls to obtain the access information. Finally, the answer is transmitted to the participant in a follow-up call or text message. We proxy for indirect search with a binary indicator for "recontact," or whether any follow-up call or text message occurred after some waiting period from the end of the first call. (This waiting period excludes, e.g., an immediate call back after losing reception. Results are reported for a waiting period of one minute but are substantively and statistically similar for alternative thresholds of up to five minutes.) This measure is a necessary condition for most indirect search, with some exceptions.[41] Under this definition, roughly 13% of friends approached by a typical participant had recontact. Among family members approached by Sunni participants, however, the recontact rate was a robust and statistically significant 10 percentage points higher (Table B-13)—a considerable boost, considering that recontact is the single largest predictor of whether a contact was described as having provided useful information (Table B-17). In contrast, Shia family members are not insignificantly more likely to recontact; if anything, they are slightly less likely to have done so. Thus, the family strategy appears to be a successful one for Sunni participants but not Shia, which is likely related to the fact that Sunni are more likely to draw on family contacts and Shia are not.

Next, we examine the extent to which participants receive and trust accurate answers from their contacts. As a first cut, we consider the proportion of submitted answers that are incorrect. While some participants may write down vague answers or answers about which they are uncertain, we argue that the open-ended nature of the task makes it less likely for participants to simply guess wildly about council addresses or gatekeepers—thus, the mere act of writing something down conveys a minimum level of faith in the information provided by contacts. Among participants who wrote down an answer, even after adjusting for demographic differences, we find that as the difficulty of the question increases, Shia increasingly provide incorrect answers at a higher rate than Sunni. On the location of the council office, where answers are public knowledge and are more eas-

[41] In one case described during a debrief, a participant's father posed the access question to a group of nearby friends, then reported the answer to his son.

ily verifiable, Sunni and Shia have nearly identical error rates in both target neighborhoods. When asked to identify a gatekeeper in the easier Sunni neighborhood, Shia have an error rate which is 14 percentage points higher but statistically indistinguishable from that of Sunni participants; in the more difficult Shia neighborhood, this gap is 17 percentage points higher and significant at $p = 0.02$, even though Shia participants enjoy a co-sect advantage (Table B-14). To confirm that these differences are not due to peculiarities about this measure of confidence (e.g., if Shia were more likely to guess at answers, leading to higher error rates), we use two additional proxies based on open-ended debrief questions. First, without any prompting, some participants asserted that their answer was correct at some point in the debrief process. Second, in another set of debrief questions, we asked participants what they found to be easy or difficult. While we initially intended this as a more general question about the search process, some participants indicated that they found the name question, the address question, or both to be "easy." We hand-code these statements[42] and find that among the 79 participants who provided an answer and claimed that the name was "correct," Shia were more likely to in fact be wrong ($p = 0.16$) and that the same pattern holds among the 48 respondents who claimed the name was "easy" ($p = 0.09$, $n = 48$).

The final measure of effectiveness, which we discuss last due to various measurement issues, is whether a contact was reported to provide useful assistance. This measure was also hand-coded from the self-reported contact log, in which participants were asked the paired pre- and post-contact questions, "How do you expect this person to help you?" and "What exact assistance did they offer you?" When participants indicated that a contact gave them either the council address or gatekeeper name, and the participant in fact later wrote down an answer to either question, we coded the contact as having provided an answer. As previously discussed, recontact is the single most important factor, being associated with roughly a 25-percentage-point increase in reported assistance. We also confirm that geography is indeed a good heuristic for selecting contacts: Relative to

[42]We offer two typical cases that were coded as indicators of "claimed correct." One participant wrote, "My friend... provided me the right answer." Another went further, indicating "I got the exact location."

contacts who lived the median distance from the assigned target (9 kilometers, or the same distance as the median participant), those who lived one standard deviation closer (5 kilometers from the target) were a significant 9 percentage points more likely to be reported as providing assistance. Finally, among Sunni participants, a 35-year-old contact was significantly more likely to provide assistance than the median 25-year-old contact (by 13 percentage points), but among Shia participants, older contacts were if anything slightly less likely to be of assistance. This further reinforces the notion that Shia hierarchy presents an impediment to information flow, and indirectly to Shia access to public services (Table B-16). To confirm that these results are not driven by reporting biases (e.g., a heightened tendency among one group to acknowledge assistance), we re-analyze with participant fixed effects and find results that are substantive and significantly identical, with the sole exception being that the coefficient on recontact becomes more noisily estimated among Sunni participants but remains consistent in sign and general magnitude. Results are also similar, with a slight weakening in statistical significance, when we only consider "validated" assistance by discarding reports of assistance from participants who did not receive at least partial credit on any question (Table B-17).

Discussion and Conclusions

We find no support for our original hypothesis, that participants belonging to the group in power—the Shia—will have better access. On the contrary, the minority group—the Sunnis—do better in identifying both easier information, the location of the local council, and harder information, the name of the local council member, irrespective of target assignment (be it co-sect or cross-sect). While some of these differences are driven by demographic differences between the two groups, particularly the greater number of Sunnis who are college-educated, the overall underperformance of Shia participants remains substantively and statistically significant even after adjusting for demographics in a variety of ways.

Ultimately, the Sunnis seem to pursue what should be a dominant strategy: They move vertically in their network, contacting older, more knowledgeable, and trusted people, to try to secure the requisite information. This is not surprising behavior. The paradox is that the Shia, the majority group and the group in power, do not pursue what one would expect to be the most promising strategy. Instead, Shia systematically employ a suboptimal "horizontal" strategy of approaching lower-quality peers that are younger and less well-educated. Sunnis also more readily reach across sectarian lines, contacting out-group individuals when assigned to the cross-sect neighborhood. In their qualitative responses they would emphasize the role of trust and strong ties as well as the contact's level of education.

In terms of effectiveness, Shia secure less assistance from their contacts than Sunnis, even after holding contact-level attributes fixed. In particular, the family members of Shia participants are less willing to expend social capital by searching on their behalf, and Shia participants are less able to obtain assistance from older contacts. Ultimately, Shia are too trusting in the low-quality information that they receive from their contacts.

Using mediation analysis (?), we find that the "strategy" and "effectiveness" classes of mechanisms are respectively more important in the Sunni and Shia targets with each accounting for roughly one-third of the gap in performance; the remaining direct contribution is likely due either to unobserved differences between Sunni and Shia or model misspecification.

We argue that there are two interrelated factors that are driving these results. First, the Shia are more hierarchical as a group, a statement long established in anthropological, historical and other recent work in political science (Patel, forthcoming). Hierarchical social ties in people pose an asymmetry in access to information, with people from the bottom not being able to tap into informational resources from more senior people. As such, younger Shia men are reluctant to contact older people for assistance and instead turn to co-sect peers. Second, they seem to be more insular as a group, as they are less likely to reach across group even when assigned to a cross-sect area. This doesn't necessarily mean that Shia have fewer cross-sect contacts. It could very well be

because all local councils are Shia dominated councils, and as such it is always a dominant strategy for a Shia to reach out to Shia contacts.[43] There is suggestive evidence, however, that points to the direction of lack of cross-sect ties amongst the Shia. We see for instance that the people they reach to when assigned to a cross-sect target are considerably less educated than the people they contact in the co-sect assignment. In addition, we see that those Shia who do reach cross-sect are more likely to get successful answers. Both these results are more suggestive of a lack of cross-sect ties rather than a choice not to contact them.

The insularity of the Shia, which may be driven by the comfort that comes with being the politically dominant majority group, may lead them to place too much trust in the information they obtain from these co-sect peers–information that ultimately proves of lower quality than that secured by their Sunni counterparts. This is reinforced in the patterns of false responses. This is a story of the dark side of social capital, also found in recent experimental findings by Levine et al. (2014), which suggest that ethnically homogeneous markets are more prone to suffer from financial bubbles than diverse ones as participants in the former are less likely to question the quality of information received.

Our findings are consistent with several possible explanations. First, it could be that as a minority under threat, the Sunnis have needed to work harder to get access to services offered by a Shia-dominated government and as such have evolved strategies to accomplish this more efficiently. Secondly, it could be some sort of legacy effect from the Saddam years, in that the Sunnis know how to work the system because they have the past experience of being the group in power. This would suggest deeply rooted behaviors of Sunnis knowing how to get things done. Relatedly, Sunni parents are likely to be better educated than the respective Shia parents, which may offer an advantage. Finally, research on network search demonstrates that networks can become difficult

[43] We cannot conclusively evaluate whether Shia are less likely to reach cross sect than Sunni (which they are by almost 20 percentage points) because they don't have cross-sect friends or because it just makes sense to ask Shia people for Shia council members. We would need to have a neighborhood with a Sunni dominated council in our experiment to be able to get at that and such a neighborhood does not exist.

to navigate when individuals are too widely connected (Kleinberg, 2000). This counterintuitive result stems from the fact that when social ties are purely local (e.g., in a grid network where every individual is only connected to immediate neighbors), searchers can use their knowledge of the general structure of the network to find optimal paths connecting them to a particular target, but an exceedingly large number of steps are required. At the other extreme, when social ties are widely dispersed, or essentially random, every individual may be connected to a target by a very small number of steps, but searchers are unable to identify these short paths because the network contains little structure to exploit—individuals are unable to identify the right direction in which to take a first step, because they have little information about where the friends of their friends are located. Baghdadi Shia, with their relatively high mobility and reputation for gregariousness, may fall into the latter category of overly dispersed and difficult-to-search social networks.

Our data does not allow us to conclusively identify one channel, especially as they could all be at play, but does offer some additional information on underlying mechanisms. We also cannot rule out the possibility that our cross-group comparisons are confounded by unobserved differences. For example, though we control for participant education or occupation in our analysis, we do not have data on parents' education or occupation. We can therefore not rule out that this could be driving part of the effect.

Paper 3: A Model for Classifying Mode of Speech in Political Deliberation

Introduction

Applications of text analysis in political science often examine corpora which were first spoken, then transcribed. To name but a few examples, in American and comparative politics, numerous articles study speech made by executives, legislators, and justices (Sigelman and Whissell, 2002a,b; Yu et al., 2008; Monroe et al., 2009; Quinn et al., 2010; Black et al., 2011; Proksch and Slapin, 2012; Eggers and Spirling, 2014; Kaufman et al., ND). Though methodologically diverse, this research shares in common an exclusive focus on the words alone. However, human speech contains information beyond simply the spoken text. The rhetoric and tone of human speech conveys information that moderates the textual content (El Ayadi et al., 2011a), and without appropriate methods to analyze the audio signal accompanying the text transcript, researchers risk overlooking important insights into the content of political speech. Moreover, studies of spoken speech span a range of fields, from speech made by elected officials (Proksch and Slapin, 2010) to deliberations and statements about foreign policy (Stewart and Zhukov, 2009; Schub, 2015).

Despite the frequency with which social scientists analyze speech, there is little to no research in the social sciences that explicitly models the audio that preceded these textual transcriptions. However, political scientists nonetheless study aspects of speech like emotion (Black et al., 2011)

and rhetorical style (Sigelman and Whissell, 2002a,b), which depend on tone of speech as well as the words used (Scherer and Oshinsky, 1977; Murray and Arnott, 1993; Dellaert et al., 1996). And though methods for analyzing text as data have received a great deal of attention in political science in recent years (Laver et al., 2003; Benoit et al., 2009; Clark and Lauderdale, 2010; Hopkins and King, 2010; Grimmer and Stewart, 2013; Lauderdale and Clark, 2014; Roberts et al., 2014; Lucas et al., 2015), none permit the inclusion of the accompanying audio features, even though recent work demonstrates that audio features can predict political outcomes (Dietrich et al., 2016).

I attempt to fill this methodological gap by proposing the speaker-affect model (SAM) for classifying auditorily distinct "modes of speech." SAM is a hierachical hidden Markov model in which each mode of speech is modeled as a HMM that generates high-frequency, high-dimensional audio features. The transitions between these modes—for example, over the course of a speech or debate—are modeled by a higher-level HMM. SAM is the first model of its kind in political science and expands on approaches in computer science which use hidden Markov models for speech classification (Schuller et al., 2003; Nwe et al., 2003; Nogueiras et al., 2001). SAM develops existing approaches in other fields, primarily computer science, in three ways. First, SAM is a hierarchical model that learns the flow between modes of speech, modeling each mode as a HMM and the transitions between these modes as a higher-level HMM. This modeling choice lets users analyze speech dynamics, which are particularly interesting in cases of strategic interaction that are often of interest to social scientists. Second, it incorporates ridge-like regularization that allows utilization of many more features than existing approaches and contributes to improved performance. Third, it employs a principled approach to uncertainty by Bayesian bootstrap, which is also of particular interest to social scientists interested in more than just predictive accuracy.

The remainder of this paper is as follows. In Section 3.2, I introduce "audio as data" for a political science audience, describing how recorded sound can be quantified as high-dimensional time-series data. Section 3.3 develops the model and inference, which is validated in Section 3.4. Finally, Section 3.5 concludes.

Audio as Data

As Section 3.1 notes, the number of papers developing and applying methods for "text as data" has increased rapidly in recent years. However, little effort has been devoted to the analysis of other data signals that often accompany text. How can the accompanying audio be similarly treated as data? In this section, I outline the preprocessing stages of an audio analysis workflow. I begin with a description of raw audio, then explain how that signal is processed before it may be input into a model like SAM.

The Raw Audio Signal

The human speech signal is transmitted as compression waves through air. A microphone translates air pressure into an analog electrical signal, which is then converted to sequence of signed integers. This recording process involves sampling the analog signal at a fixed sampling rate and rounding to the nearest discrete value as determined by the audio bit depth, or the number of binary digits used to encode each sample value.

In order to statistically analyze audio as data, we must first cut (typically long) audio recordings into a sequence of shorter segments which may correspond to sentence-length utterances, speaker turns, or simply short clips cut at regular intervals for convenience. For the remainder of this paper, I refer to such segments as "utterances" and modes of speech as "emotions," although the model is general. For each of these utterances, we compute a series of *audio features*.

Raw Audio to Audio Features

A typical approach to audio feature extraction divides each utterance's raw audio signal into short windows of time (here, 25 milliseconds in length), and extracts various summary statistics from each window. To maximize the extracted information, rolling or overlapping windows are often

used (here, windows are spaced at 12.5-millisecond intervals). Hence, longer utterances will have a greater number of windows from which features are extracted. This presents a challenge for typical classifiers that can only accept a fixed-size feature vector for each utterance, such as logistic regression or support vector machines; when such models are used, time series information are collapsed, e.g. by taking the mean and standard deviation of each feature. However, in audio classification benchmarks, modeling approaches that retain the full audio feature time series—HMMs and convolutional neural nets—have generally been shown to outperform those that discard this information.

Numerous studies in the audio classification literature have sought to identify the most informative features for tasks such as speaker identification or emotion detection. The model developed in this paper is agnostic on this point. Because the regularization approach employed here allows an essentially unlimited number of dimensions to be used, I simply extract the full range of features that have been used in the literature.[44]

Audio features can be broadly divided into three categories. In each window, some features are extracted from the raw signal. For example, sound intensity (measured in decibels) is defined as the log root mean square of the raw samples. Next, features based on the audio frequency spectrum are extracted. The audio signal (assumed to be stationary within the short timespan of the window) is decomposed into components of various frequencies, and the power contributed by each component is estimated by discrete Fourier transform. The shape of the resulting power spectrum, particularly the location of its peaks, provides information about the shape of the speaker's vocal tract, e.g. tongue position. Some artifacts are introduced in this process, most notably by truncating the audio signal at the endpoints of the 25-millisecond frame and by the greater attenuation of high-frequency sounds as they travel through air. I ameliorate the former with a Hamming window that downweights audio samples toward the frame endpoints, and compensate for the latter using

[44]For excellent reviews of the literature, including a more thorough discussion of these features, see Ververidis and Kotropoulos (2006); El Ayadi et al. (2011b).

Table 3.1: Audio features extracted in each frame. In addition, we include interactions between (*i*) energy and zero-crossing rate, and (*ii*) Teager energy operator and fundamental frequency. We also use the first and second finite differences of all features.

Features from raw audio samples		
energy	1 feature / frame	sound intensity, in decibels: $\log_{10}\sqrt{\overline{x_i^2}}$
ZCR	1 feature / frame	zero-crossing rate of audio signal
TEO	1 feature / frame	Teager energy operator: $\log_{10}\overline{x_i^2 - x_{i-1}x_{i+1}}$
Spectral features		
F0	2 features / frame	fundamental, or lowest, frequency of speech signal (closely related to perceived pitch; tracked by two algorithms)
formants	6 features / frame	harmonic frequencies of speech signal, determined by shape of vocal tract (lowest three formants and their bandwidths)
MFCC	12 features / frame	Mel-frequency cepstral coefficients (based on discrete Fourier transform of audio signal, transformed and pooled to approximate human perception of sound intensity in 12 pitch ranges)
Voice quality		
jitter	2 features / frame	average absolute difference in F0
shimmer	2 features / frame	average absolute difference in energy

a pre-emphasis filter that boosts the higher-frequency components. Finally, I extract measures of voice quality, commonly used to diagnose pathological voice, based on the short-term consistency of pitch and intensity. Various interactions used in the emotion-detection literature are calculated, and the first and second finite differences of all features may also be taken.

Table 3.1 shows the full set of features that we extract for each frame. The table divides features into those calculated directly from the raw audio, spectral features, and those measuring voice quality. Spectral features are those based on the frequency spectrum (for example, energy in the lower portion of the spectrum), while voice quality describes features that measure vocal qualities like "raspiness" and "airiness." Note that some rows describe classes of features, like energy in each of 12 frequency ranges.

Contiguous frames are grouped together into segments of continuous speech, e.g. into sentence-length *utterances*. When timestamped transcripts are available, they may be used to segment the audio by sentence. If unavailable, timestamps can be recovered by a procedure called "forced

alignment," e.g. in the Sphinx software package. Two preprocessing techniques are also implemented in R package that implements SAM, under development. The first uses a framewise classifier to detect events of interest, such as interruptions or applause. This is trained on manually coded start/stop times for the events of interest, as well as a few seconds before and after each instance to serve as a baseline. SAM currently implements a linear support vector machine to classify individual audio frames. Framewise classifications are smoothed and thresholded to reduce false positives. This simple classifier is an effective and computationally efficient method for isolating short sounds with distinct audio profiles, such as an offstage voices. The second approach uses a rule-based system that recursively partitions continuous speech by cutting it at brief pauses until utterance lengths are close to some target length.

The Speaker-Affect Model

In this section, I introduce the speaker-affect model, or SAM. SAM is a hierarchical hidden Markov model (HMM), meaning that each "state" in SAM is itself another hidden Markov model. Within SAM, states are the user-defined labels, like "angry" and "neutral" or "male" and "female." Each of these states, by contrast, is modeled as an *unsupervised* HMM, learned during the training process. In the case of speech modes, this is useful because it permits each mode of speech to be defined by learned transitions between "sounds," which can be inferred from the user-supplied labels.

In the remainder of this section, I introduce notation, define the model, and overview inference.

Notation

I assume a model of discrete speech modes, as is common in the emotion detection literature. However, in classifying political speech I depart from traditional models of so-called "basic" emotions such as anger or fear (Ekman, 1992, 1999), which are posited to be universal across cultures and often involuntarily expressed. Because such emotions are rare in political speech, of model of

them is not especially useful. Across a range of applications, I find that actors of interest tend to be practiced communicators with a reasonable degree of practice and control over their speech. Political speakers generally employ more complex modes of speech, such as skepticism or sarcasm, in pursuit of context-specific goals such as persuasion or strategic signaling. To this end, I develop a method that can learn to distinguish between arbitrary modes of speech specified by subject-matter experts.

This model segments continuous speech into utterances (or other short segments), generally bracketed by pauses. The mode of speech is assumed to be constant during an utterance. This is the quantity that we wish to measure, and it is generally unobserved unless a human coder listens to and classifies the utterance. Naturally, the mode of speech is not independent across utterances: A calm utterance is usually followed by another calm utterance. On a more granular level, each utterance is composed of an unobserved sequence of sounds, such as vowels, sibilants, and plosives. These sounds then generate a continuous stream of observed audio features.

Indices:

- Utterance index $u \in \{1, \cdots, U\}$: continuous segment of audible speech by a single speaker, preceded and followed by a period of silence or a transition between speakers.
- Time index $t \in \{1, \cdots, T_u\}$: position of audio window or video frame within an utterance. Advances by increments of 12.5 milliseconds.

Latent states:

- $S_u \in \{1, \cdots, M\}$: latent mode of speech for utterance u, e.g. anger. Indexed by m.
- $R_{u,t} \in \{1, \cdots, K\}$: latent sound at time t (e.g., sibilant, plosive). Indexed by k. Note that the same index may take on different meanings depending on the mode of speech. For example, sibilants may appear in both angry and neutral speech, but exact auditory characteristics will differ by emotion, and the index corresponding to the subjective label of "sibilant" may not be the same for each emotion.

Features:

- $\boldsymbol{X}_{u,t}$: column vector of D audio features at time t, such as sound intensity (decibels) or position of mouth corners. All feature vectors in an utterance are collected in the $T_u \times D$ matrix, $\boldsymbol{X}_u$ (with $D = 189$ features in total: 27 audio features, optionally with first and second derivatives).

Model

I assume that the feature series is generated by a hierarchical HMM with two levels. First, a speaker's emotional state in utterance u is assumed to be drawn from a first-order HMM, i.e., drawn based on the emotional state in the previous utterance. The probability of transitioning from emotion m to m' is given by $\Delta_{m,m'}$, and all transition probabilities are collected in the emotion transition matrix $\boldsymbol{\Delta}$.

$$S_u \sim \text{Cat}(\boldsymbol{\Delta}_{S_{u-1},*})$$

Second, given that utterance u was spoken with emotion $S_u = m$, the sequence of sounds and expr(essions that comprise an utterance are assumed to be generated by the m-th emotion-specific first-order HMM. The probability of transitioning from sound/expression k to k' is given by $\Gamma^m_{k,k'}$, and transition probabilities are collected in sound/expression transition matrix $\boldsymbol{\Gamma}^m$. I use superscripts to index the properties of states and sounds; subscripts index the elements of a vector or matrix.

$$(R_{u,t} \mid S_u = m) \sim \text{Cat}(\boldsymbol{\Gamma}^m_{R_{u,t-1},*})$$

Finally, during a particular sound, the vector of features at each point in time is assumed to be

drawn from a multivariate Gaussian distribution.

$$(X_{u,t} \mid S_u = m, R_{u,t} = k) \sim N(\boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k})$$

For example, the hypothetical "sibilant" sound might have a low value for $\mu_{\text{intensity}}^{\text{anger, sibilant}}$, and the covariance matrix $\boldsymbol{\Sigma}^{\text{anger, sibilant}}$ might contain a small variance for intensity and positive covariance with zero-crossing rate (often an indicator of hissing sounds). Given that the speaker is forming an angry sibilant, the exact audio characteristics of successive audio frames are assumed to be independent draws from this distribution. Despite the seemingly implausible independence assumption, features may still exhibit substantial autocorrelation, because expressions tend to persist for multiple frames and different expressions often have very different mean values. However, model misspecification and unfounded independence assumptions often lead to overly confident predictions and over-concentrated posteriors, adversely affecting model performance; in section 3.3.5, I present an approach to inference that directly addresses these concerns.

The remainder of this section describes the model can be fit by expectation-maximization, how the resulting parameters are used to classify unlabeled data, and a bootstrap-based approach to statistical inference. Numerical issues, missing data, and other complications are discussed in Appendix C.

Training

To estimate the parameters of the model, we select a training set of utterances for human coding. Each coder receives a subset of the training utterances, in random order, and classifies it into one of the M basic emotional states. Thus, during the training stage, the emotion labels, S_u, are known and the emotion-specific distributions are estimated.[45]

[45] In practice, because the perception of emotion is subjective, discrepancies between coders are to be expected and the emotion labels are not known precisely. We address this in the following section by weighting the contribution of an utterance to the model for emotion m by the proportion of human coders who classified the utterance as emotion m.

The training procedure consists of the following steps:

1. Estimate the parameters of the emotion-specific HMM distributions
   (a) Estimate the parameters of the sound/expression Gaussian distributions $\boldsymbol{\mu}^{m,k}$ and $\boldsymbol{\Sigma}^{m,k}$
   (b) Estimate the sound/expression transition matrix $\boldsymbol{\Gamma}^{m}$
2. Estimate the emotion transition matrix $\boldsymbol{\Delta}$
3. Select the optimal number of sounds/expressions in each emotion, K, via cross-validation

Training a Single HMM with a Single Utterance of Known Emotion

In this section, we first present a simplified version of our model—specifically, we start with a single utterance, and we assume that the emotion of that utterance is perfectly known. Some parameters in this section are marked with an overline to prevent confusion with the actual versions used in the following section, where we relax these constraints and present the actual procedure used in the training stage. The introduction to HMMs in this section is adapted from Zucchini and MacDonald (2009). Our full model is developed the following section.

Suppose that the speaker's emotional state during utterance u is known to be $S_u = m$. Audio and visual features, $\boldsymbol{X}_u = [X_{u,1}, \cdots, X_{u,T_u}]^\top$, are also observed. However, the underlying sounds and expressions that generated these features are completely unobserved: neither their labels ($R_{u,t}$, the order in which sounds and expressions occurred) nor their contents ($\boldsymbol{\mu}^{m,k}$ and $\boldsymbol{\Sigma}^{m,k}$, the auditory and visual characteristics of each sound/expression) are known. We treat the sound/expression labels as missing data and estimate by the expectation–maximization algorithm (EM).

For simplicity of exposition, we begin with a single utterance u, with known emotional state $S_u = m$. At each time t, the feature vector $\boldsymbol{X}_{u,t}$ could have been generated by any of the K sounds/-expressions associated with emotion m, so there are K^{T_u} possible sequences of sounds/expressions by which the feature sequence, $\boldsymbol{X}_u$, could have been generated. The observed-data likelihood is the joint probability of all observed features is found by summing over every possible sequence of

sounds and expressions:

$$\begin{aligned}
&\mathscr{L}^m(\boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k}, \boldsymbol{\Gamma}^m \mid \boldsymbol{X}_u, S_u = m) \\
&\quad = \Pr(\boldsymbol{X}_{u,1} = \boldsymbol{x}_{u,1}, \cdots, \boldsymbol{X}_{u,T_u} = \boldsymbol{x}_{u,T_u} \mid \boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k}, \boldsymbol{\Gamma}^m) \\
&\quad = \delta^{m\top} \boldsymbol{P}(\boldsymbol{x}_{u,1}) \left[\prod_{t=2}^{T_u} \boldsymbol{\Gamma}^m \boldsymbol{P}(\boldsymbol{x}_{u,t}) \right] \mathbf{1},
\end{aligned} \tag{3.3}$$

where δ^m is a $1 \times K$ vector containing the initial distribution of sounds/expressions (assumed to be the stationary distribution, a unit row eigenvector of $\boldsymbol{\Gamma}^m$), the matrices $\boldsymbol{P}(\boldsymbol{x}_{u,t}) \equiv \text{diag}\left(\phi_D(\boldsymbol{x}_{u,t}; \boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k})\right)$ are $K \times K$ diagonal matrices in which the (k,k)-th element is the (D-variate Gaussian) probability of $x_{u,t}$ being generated by sound/expression k, and $\mathbf{1}$ is a column vector of ones. The parameters $\boldsymbol{\mu}^{m,k}$, $\boldsymbol{\Sigma}^{m,k}$, and $\boldsymbol{\Gamma}^m$ can in principle be found by directly maximizing this likelihood.

In practice, given the vast number of parameters to optimize over, we estimate using the Baum–Welch algorithm, a flavor of expectation–maximization for hidden Markov models. This procedure involves maximizing the complete-data likelihood, which differs from equation 3.3 in that it also incorporates the probability of the unobserved sounds/expressions.

$$\begin{aligned}
&\Pr(\boldsymbol{X}_{u,1} = \boldsymbol{x}_{u,1}, \cdots, \boldsymbol{X}_{u,T_u} = \boldsymbol{x}_{u,T_u}, R_{u,1} = r_{u,1}, \cdots, R_{u,T_u} = r_{u,T_u} \mid \boldsymbol{\mu}^{m,*}, \boldsymbol{\Sigma}^{m,*}, \boldsymbol{\Gamma}^m) \\
&\quad = \delta^{m\top}_{r_{u,1}} \, \phi_D(\boldsymbol{x}_{u,1}; \boldsymbol{\mu}^{m,r_{u,1}}, \boldsymbol{\Sigma}^{m,r_{u,1}}) \times \\
&\qquad \prod_{t=2}^{T_u} \Pr(R_{u,t} = r_{u,t} \mid R_{u,t-1} = r_{u,t-1}) \, \phi_D(\boldsymbol{X}_{u,t}; \boldsymbol{\mu}^{m,r_{u,t}}, \boldsymbol{\Sigma}^{m,r_{u,t}}) \\
&\quad = \prod_{k=1}^{K} \left(\delta^{m\top}_k \, \phi_D(\boldsymbol{x}_{u,1}; \boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k}) \right)^{1\{R_{u,1}=k\}} \times \\
&\qquad \prod_{t=2}^{T_u} \left(\prod_{k=1}^{K} \left(\prod_{k'=1}^{K} \left(\Gamma^m_{k,k'} \right)^{1\{R_{u,t}=k', R_{u,t-1}=k'\}} \phi_D(\boldsymbol{X}_{u,t}; \boldsymbol{\mu}^{m,k}, \boldsymbol{\Sigma}^{m,k})^{1\{R_{u,t}=k\}} \right) \right),
\end{aligned} \tag{3.4}$$

The algorithm uses the joint probability of (*i*) all feature vectors up until time t and (*ii*) the sound at t, given in equation 3.5. Together, these are referred to as the *forward probabilities*,

because values for all t are efficiently calculated in a single recursive forward pass through the feature vectors.

$$\begin{aligned}\boldsymbol{\alpha}_{u,t} &\equiv \Pr(\boldsymbol{X}_{u,1} = \boldsymbol{x}_{u,1}, \cdots, \boldsymbol{X}_{u,t} = \boldsymbol{x}_{u,t}, R_{u,t} = k) \\ &= \delta_u^\top \boldsymbol{P}(\boldsymbol{x}_{u,1}) \left(\prod_{t'=2}^{t} \boldsymbol{\Gamma}^m \boldsymbol{P}(x_{u,t'}) \right)\end{aligned} \tag{3.5}$$

The algorithm also relies on the conditional probability of (*i*) all feature vectors after t given (*ii*) the sound/expression at t (equation 3.6). These are similarly called the *backward probabilities* due to their calculation by backward recursion.

$$\begin{aligned}\boldsymbol{\beta}_{u,t} &\equiv \Pr(\boldsymbol{X}_{u,t+1} = \boldsymbol{x}_{u,t+1}, \cdots, \boldsymbol{X}_{u,T_u} = \boldsymbol{x}_{u,T_u} \mid R_{u,t} = k) \\ &= \left(\prod_{t'=t+1}^{T_u} \boldsymbol{\Gamma}^m \boldsymbol{P}(\boldsymbol{x}_{u,t'}) \right) \mathbf{1}\end{aligned} \tag{3.6}$$

E step

The E step involves substituting (*i*) the unobserved sound/expression labels, $\mathbf{1}\{R_{u,t} = k\}$, and (*ii*) the unobserved sound/expression transitions, $\mathbf{1}\{R_{u,t} = k\}$, with their respective expected values, conditional on the observed features $\boldsymbol{X}_u$ and the current estimates of $\boldsymbol{\mu}^{m,k}$, $\boldsymbol{\Sigma}^{m,k}$, and $\boldsymbol{\Gamma}^m$ (collectively referred to as $\boldsymbol{\Theta}$).

For (*i*), combining equations 3.3, 3.5 and 3.6 immediately yields the expected sound/expression label

$$\mathbb{E}\left[\mathbf{1}\{R_{u,t} = k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u = m\right] = \tilde{\alpha}_{u,t,k}\, \tilde{\beta}_{u,t,k} \;/\; \tilde{\mathscr{L}}^m, \tag{3.7}$$

where the tilde denotes the current approximation based on parameters from the previous M step, and $\tilde{\alpha}_{u,t,k}$ and $\tilde{\beta}_{u,t,k}$ are the k-th elements of $\tilde{\boldsymbol{\alpha}}_{u,t}$ and $\tilde{\boldsymbol{\beta}}_{u,t}$, respectively.

For (*ii*), after some manipulation, the expected sound/expression transitions can be expressed

as

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m] \\
&\quad = \Pr(R_{u,t}=k', R_{u,t-1}=k, \boldsymbol{X}_u \mid \tilde{\boldsymbol{\Theta}}) \;/\; \Pr(\boldsymbol{X}_u \mid \tilde{\boldsymbol{\Theta}}) \\
&\quad = \Pr(\boldsymbol{X}_{u,1}, \cdots, \boldsymbol{X}_{u,t-1}, R_{u,t-1}=k \mid \tilde{\boldsymbol{\Theta}}) \; \Pr(R_{u,t}=k' \mid R_{u,t-1}=k, \tilde{\boldsymbol{\Theta}}) \times \\
&\qquad \Pr(\boldsymbol{X}_{u,t} \mid R_{u,t}=k') \; \Pr(\boldsymbol{X}_{u,t+1}, \cdots, \boldsymbol{X}_{u,T_u} \mid R_{u,t}=k') \;/\; \Pr(\boldsymbol{X}_u \mid \tilde{\boldsymbol{\Theta}}) \\
&\quad = \tilde{\alpha}_{u,t-1,k} \, \tilde{\Gamma}^m_{k,k'} \, \phi_D(\boldsymbol{x}_{u,t}; \tilde{\boldsymbol{\mu}}^{m,k}, \tilde{\boldsymbol{\Sigma}}^{m,k}) \, \tilde{\beta}_{u,t,k'} \;/\; \tilde{\mathscr{L}}^m. \qquad (3.8)
\end{aligned}$$

M Step

After substituting equations 3.7 and 3.8 into the complete-data likelihood (equation 3.4), the M step involves two straightforward calculations.

First, the maximum likelihood update of the transition matrix $\boldsymbol{\Gamma}^m$ follows almost directly from equation 3.8:

$$\left(\Gamma^m_{k,k'} \mid S_u = m\right) = \frac{\sum_{t=2}^{T_u} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m\right]}{\sum_{t=2}^{T_u} \sum_{k'=1}^{K} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m\right]} \qquad (3.9)$$

Second, the optimal update of the k-th sound/expression distribution parameters found by fitting a Gaussian distribution to the feature vectors, with weights given by the expected value of the k-th label.

$$\left(\boldsymbol{\mu}^{m,k} \mid S_u = m\right) = \boldsymbol{X}_u^\top \overline{\boldsymbol{W}}_u^{m,k} \qquad (3.10)$$

$$\left(\boldsymbol{\Sigma}^{m,k} \mid S_u = m\right) = \boldsymbol{X}_u^\top \mathrm{diag}\left(\overline{\boldsymbol{W}}_u^{m,k}\right) \boldsymbol{X}_u - \boldsymbol{\mu}^{m,k} \boldsymbol{\mu}^{m,k\top} \qquad (3.11)$$

$$\text{where } \overline{\boldsymbol{W}}_u^{m,k} \equiv \frac{\left[\mathbb{E}\left[\mathbf{1}\{R_{u,1}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m\right], \cdots, \mathbb{E}\left[\mathbf{1}\{R_{u,T_u}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m\right]\right]^\top}{\sum_{t=1}^{T_u} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}, S_u=m\right]}$$

Training Multiple Emotion-specific HMMs with Multiple Utterances of Imperfectly Observed Emotion

In practice, due to the subjective nature of perceived emotion, the emotional labels attached to training utterances may contain inter-coder variation: some proportion of coders, $\hat{S}^{\text{train}}_{u,m}$, will classify an utterance as the m-th emotion. I treat $\hat{\boldsymbol{S}}^{\text{train}}_{u} = [\hat{S}^{\text{train}}_{u,1}, \cdots, \hat{S}^{\text{train}}_{u,M}]^{\top}$ as a probability vector over the speaker's emotional state during utterance u, which naturally leads to the use of $\hat{S}^{\text{train}}_{u,m}$ as a weighting factor in the estimation of the m-th emotion model. In the E step, I modify equation 3.7 to

$$\mathbb{E}\left[\mathbf{1}\{R_{u,t}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right] = \hat{S}^{\text{train}}_{u,m}\, \tilde{\alpha}_{u,t,k}\, \tilde{\beta}_{u,t,k} \,/\, \tilde{\mathscr{L}}^{m}, \tag{3.12}$$

and equation 3.8 to

$$\begin{aligned}&\mathbb{E}[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}] \\ &\quad = \hat{S}^{\text{train}}_{u,m}\, \tilde{\alpha}_{u,t-1,k}\, \tilde{\Gamma}^{m}_{k,k'}\, \phi_D(\boldsymbol{x}_{u,t}; \tilde{\boldsymbol{\mu}}^{m,k}, \tilde{\boldsymbol{\Sigma}}^{m,k})\, \tilde{\beta}_{u,t,k'} \,/\, \tilde{\mathscr{L}}^{m}.\end{aligned} \tag{3.13}$$

The multi-utterance extensions of the M step equations 3.9, 3.10, and 3.11 are

$$\Gamma^{m}_{k,k'} = \frac{\sum_{u=1}^{U_{\text{train}}} \sum_{t=2}^{T_u} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right]}{\sum_{u=1}^{U_{\text{train}}} \sum_{t=2}^{T_u} \sum_{k'=1}^{K} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k', R_{u,t-1}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right]} \tag{3.14}$$

$$\boldsymbol{\mu}^{m,k} = \sum_{u=1}^{U_{\text{train}}} \boldsymbol{X}_u^{\top} \boldsymbol{W}_u^{m,k} \tag{3.15}$$

$$\boldsymbol{\Sigma}^{m,k} = \sum_{u=1}^{U_{\text{train}}} \left(\boldsymbol{X}_u^{\top} \operatorname{diag}\left(\boldsymbol{W}_u^{m,k}\right) \boldsymbol{X}_u\right) - \boldsymbol{\mu}^{m,k} \boldsymbol{\mu}^{m,k\top} \tag{3.16}$$

where $\boldsymbol{W}_u^{m,k} \equiv \dfrac{\sum_{u=1}^{U_{\text{train}}} \left[\mathbb{E}\left[\mathbf{1}\{R_{u,1}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right], \cdots, \mathbb{E}\left[\mathbf{1}\{R_{u,T_u}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right]\right]^{\top}}{\sum_{u=1}^{U_{\text{train}}} \sum_{t=1}^{T_u} \mathbb{E}\left[\mathbf{1}\{R_{u,t}=k, S_u=m\} \mid \boldsymbol{X}_u, \tilde{\boldsymbol{\Theta}}\right]}$

Extrapolation

Section 3.3.3 outlined a procedure to learn how a speaker generates audio features while speaking with a particular emotion. This procedure is based on a training set of utterances with human-coded emotions, and its output is M separate, emotion-specific hidden Markov models, each with some estimated parameters $\hat{\boldsymbol{\mu}}^{m,k}$, $\hat{\boldsymbol{\Sigma}}^{m,k}$, and $\hat{\boldsymbol{\Gamma}}^{m}$. In this section, I describe how emotion-specific HMMs can be used to predict the emotions of utterances in previously unseen speech. This task can be approached in two ways: we can either (*i*) "decode" the most likely *local* emotion of the u-th utterance, or (*ii*) decode the most likely *sequence* of emotions that generated all U utterances *globally*. In our applications, the two approaches produce results that differ in insubstantial ways, if at all.

In either case, the general procedure to estimate the higher-level HMM broadly parallels the preceding sections. The main difference is that emotional labels are available for the training set, which eliminates the need for the E step. Instead of assuming a the stationary distribution, we simply estimate the frequency of each emotion, $\overline{\boldsymbol{S}} \equiv \frac{1}{U_{\text{train}}} \sum_{u=1}^{U_{\text{train}}} \hat{\boldsymbol{S}}_u^{\text{train}}$, using the human-coded training utterances. Using consecutive training utterances, we also estimate the emotional transition matrix $\hat{\boldsymbol{\Delta}}$, by

$$\hat{\Delta}_{m,m'} = \frac{\sum_{u=1}^{(U_{\text{train}}-1)} j\hat{S}_{u,m}^{\text{train}} \hat{S}_{u+1,m'}^{\text{train}} \mathbf{1}\{u \text{ immediately precedes } u+1\}}{\sum_{u=1}^{(U_{\text{train}}-1)} \hat{S}_{u,m}^{\text{train}} \mathbf{1}\{u \text{ immediately precedes } u+1\}} \tag{3.17}$$

As before, we calculate the that the feature sequence in utterance u, was generated by the estimated model for the m-th emotion: $\Pr\left(\boldsymbol{X}_u \mid \hat{\boldsymbol{\Theta}}, S_u = m\right)$. These state probabilities are collected

in the diagonal matrix $\boldsymbol{P}(\boldsymbol{X}_u)$. As before, define the total, forward, and backward probabilities

$$\begin{aligned}\mathscr{L} &= \Pr(\boldsymbol{X}_1 = \boldsymbol{x}_1, \cdots, \boldsymbol{X}_U = \boldsymbol{x}_U) \\ &= \overline{\boldsymbol{s}}^\top \boldsymbol{P}(\boldsymbol{x}_1) \left(\prod_{u'=2}^{U} \hat{\Delta} \boldsymbol{P}(\boldsymbol{x}_{u'}) \right) \mathbf{1} \end{aligned} \tag{3.18}$$

$$\begin{aligned}\boldsymbol{A}_u &= \Pr(\boldsymbol{X}_1 = \boldsymbol{x}_1, \cdots, \boldsymbol{X}_u = \boldsymbol{x}_u, S_u = m) \\ &= \overline{\boldsymbol{s}}^\top \boldsymbol{P}(\boldsymbol{x}_1) \left(\prod_{u'=2}^{u} \hat{\Delta} \boldsymbol{P}(\boldsymbol{x}_{u'}) \right) \end{aligned} \tag{3.19}$$

$$\begin{aligned}\boldsymbol{B}_u &= \Pr(\boldsymbol{X}_{u+1} = \boldsymbol{x}_{u+1}, \cdots, \boldsymbol{X}_U = \boldsymbol{x}_U \mid S_u = m) \\ &= \left(\prod_{u'=u+1}^{U} \hat{\Delta} \boldsymbol{P}(\boldsymbol{x}_{u'}) \right) \mathbf{1} \end{aligned} \tag{3.20}$$

The local probabilities of the emotional states are then

$$\left[\Pr\left(S_u = m \mid \boldsymbol{X}, \hat{\boldsymbol{\Theta}} \right) \right] = \boldsymbol{A}_u \boldsymbol{B}_u / \mathscr{L} \tag{3.21}$$

and the globally most likely sequence of emotional states can be found by the Viterbi algorithm, a dynamic programming approach that efficiently finds the sequence of local emotions that maximizes the probability of the observed features.

Inference

The posterior distribution for SAM can be approximated in a variety of ways. However, the most typical use of an audio classification model is to generate posterior probabilities for a large corpus of unclassified audio segments, and many applications expect or require that the output from a classification model be well-calibrated in the sense that if human coders were to determine the actual mode of speech for each observation, among those with a posterior probability of p in class m, some fraction p would actually lie in class m. Typical Bayesian approaches for the model

described here do not possess this property, in large part due to violations of the independence assumption between successive frames that belong to the same state and sound. For example, it is mechanically true that the sound intensity of a 25-millisecond audio window will be correlated with the next window's intensity, because adjacent windows overlap by 12.5 milliseconds; moreover, nearby but non-overlapping windows can be nonindependent due to unmodeled medium-term time-varying factors such as distance from a microphone. This model misspecification leads to an overly concentrated posterior, which can manifest as posterior class probabilities that are numerically indistinguishable from zero or one.

I employ an alternative approach for posterior inference, the Bayesian bootstrap, that places a Dirichlet(1) prior over each utterance's probability of inclusion in the sample. These inclusion probabilities correspond to the weight attached to each observation and equal $1/N$ in expectation. Inference proceeds by taking repeated Dirichlet draws from the prior, assigning the resulting vector as weights, and repeating the deterministic EM procedure described above. Repeated draws numerically integrate over the prior, resulting in a posterior distribution over estimated modes of speech and model parameters.

Application

In this section, I demonstrate the application of `SAM` with an application to a corpus of televised political debates from Kim et al. (2012). These debates were held in Switzerland from 2006–2008 and span a variety of topics. Kim et al. divided the videos into 30-second segments, and all segments were then scored on the level of conflict expressed by participants. Mechanical Turk coders rated segments on a five-point Likert scale for each of the questions in Table 3.2, and summed scores were averaged over ten coders. It is noteworthy that a number of these questions involve non-auditory cues.

The Kim et al. dataset is a well-suited for benchmarking because (1) it is fully labeled by human

Table 3.2: Questions used to code debate segments as "conflictual.", adapted from Kim et al. (2012) (Table 1). Those marked with an asterisk are reversed before summing, and those marked with a dagger depend on non-auditory cues.

Question
People wait for their turn before speaking *
The atmosphere is relaxed *†
People show mutual respect *†
One or more people talk fast
People argue
One or more people raise their voice
People interrupt one another
One or more people are aggressive
One or more people compete to talk
People are actively engaged
The ambience is tense †
One or more people fidget †
One or more people shake their heads and nod †
One or more people gesture with their hands †
One or more people frown †

coders and (2) each segment is assigned a continuous conflict score. I discretize the conflict scale by cutting it at the midpoint, but these continuous scores are retained to help identify hard cases that lie close to the threshold. As I show later, the bulk of misclassified segments is composed of these hard cases.

I use this benchmark corpus to demonstrate that SAM can achieve high levels of accuracy while requiring markedly less coding, both in terms of the number of segments coded and the coding effort per segment (binary classification vs continuous scoring). The corpus as analyzed here contains slightly more than eight hours of audio data—a range in which it is still feasible to manually classify. However, a political deliberation experiment with a reasonable sample size could easily generate data volumes that are larger by several orders of magnitude. I show that using SAM, less than one hour of training data is sufficient to achieve out-of-sample accuracy of 82%, versus human "ground truth" labels—a fairly high value given the use of non-auditory cues

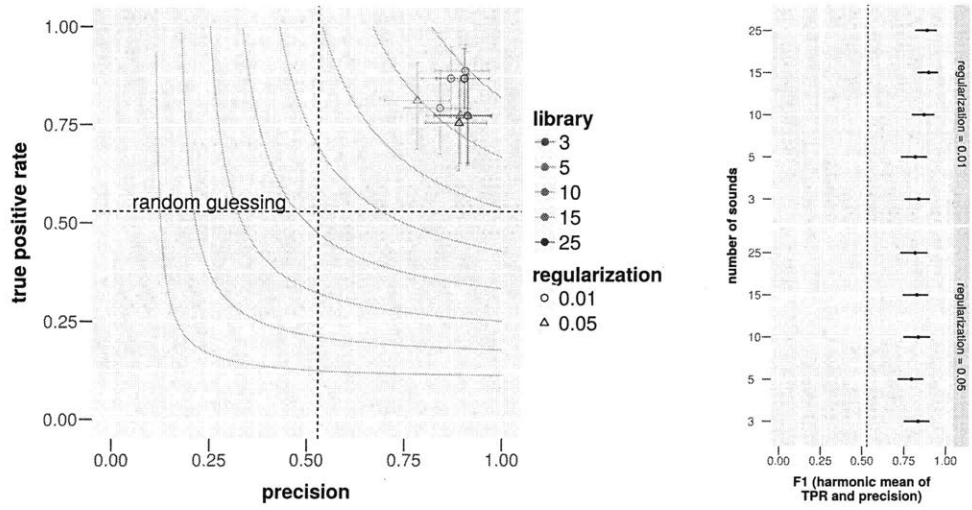
for classification, as well as the subjective nature of perceived "conflict." I further show that the resulting posterior class predictions are reasonably well-calibrated, opening the door for a range of new analyses and research designs.

The benchmark is intended to parallel the procedure that would be used to analyze such a political deliberation experiment. First, I divide the audio corpus into training (10% of corpus), validation (10%), and test (80%) sets. The former two represent the quantity of audio data that would require human coding in a realistic scenario (100 minutes), although this amount could be reduced by a factor of $V/(2V-2)$ if V-fold cross-validation is used (60 minutes of human-coded data for $V = 6$, increasing the computational cost by a factor of V).

In this benchmark, I evaluate eight SAM parameter combinations with the assumed number of sounds ranging from $K = 3$ to 25 and the regularization parameter taking on values of $\lambda = 0.01$ or 0.05. Figure 3-1 illustrates the performance of each model on the validation set. I discretize predictions by cutting posterior probabilities at 0.5, then compute various measures of binary classification performance. (Ongoing extensions include ROC-based selection and further calibration of the posterior probabilities through Platt scaling or isotonic regression.) Estimated accuracy in the validation set ranges from 0.75–0.91 depending on the model and whether positive (conflict) segments are pooled with negative (peaceful) segments or each class is considered separately. Model selection is conducted by selecting on the highest overall F_1 score, a commonly used binary classification performance metric that is a nonlinear function (the harmonic mean) of a classifier's precision and true positive rate. The selected model by this metric is a somewhat complex SAM that models 15 sounds with little regularization. Selection by highest out-of-sample likelihood would yield a model with 25 sounds and similarly low regularization, but this model is markedly slower to run and effectively indistinguishable in terms of classification performance.

Next, I evaluate the selected model's performance in a test set consisting of roughly 6.5 hours of debate. Estimated overall accuracy is 0.82 with a 95% binomial confidence interval of (0.80, 0.85). The upper panel of figure 3-2 shows that a greater proportion of negative (peaceful) segments are

Figure 3-1: Performance of various parameter combinations on validation set. Black dots and solid black errorbars indicate posterior means and 95% credible intervals. In the left panel, grey lines indicate contour lines of the F_1 metric, the harmonic mean of precision (x-axis) and true positive rate (y-axis), with higher values in the upper right. Dotted black lines denote the classification performance that would be achieved by randomly predicting labels according to the prevalence of each audio class.

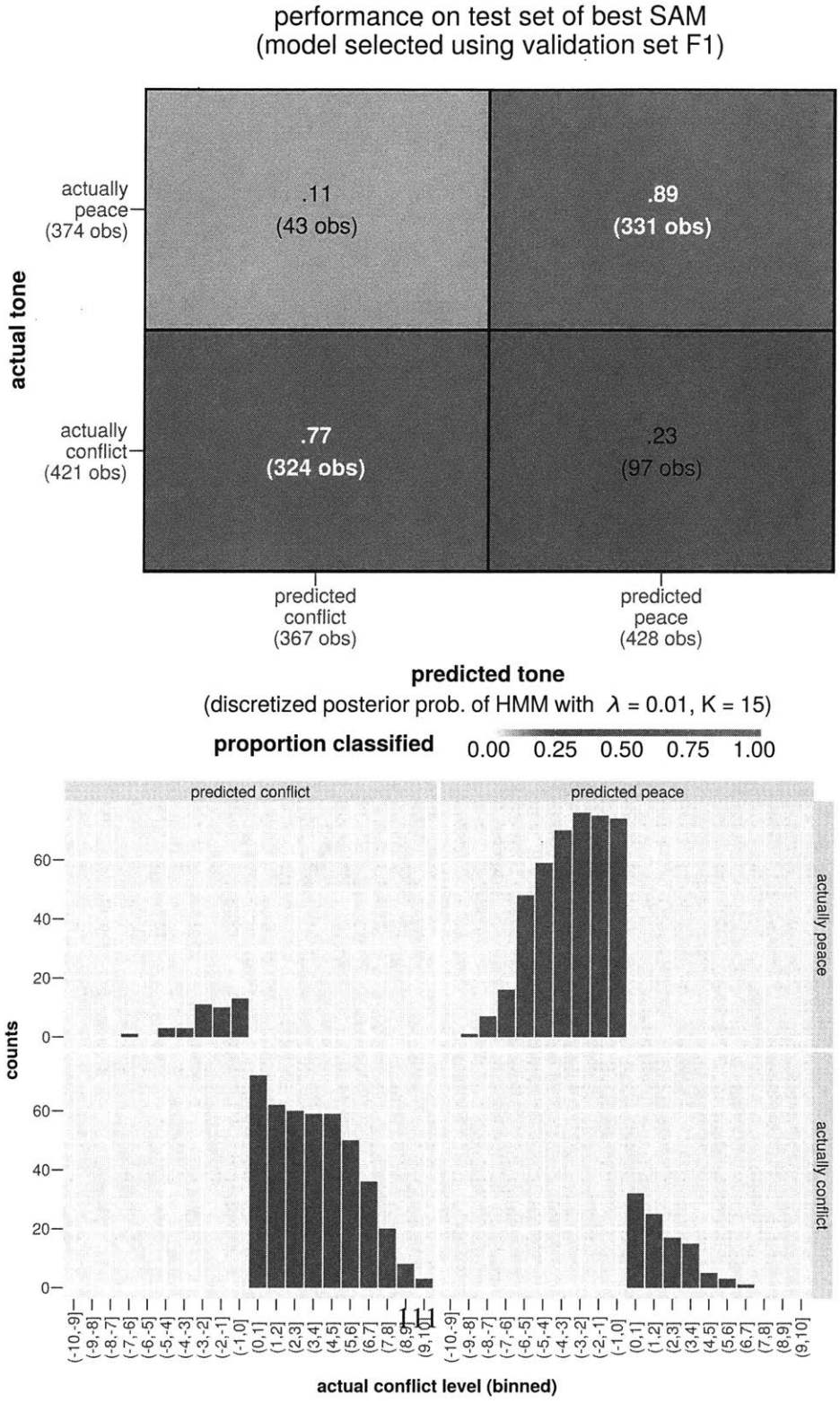


correctly classified (0.89, vs a true positive rate of 0.77). In the lower panel, the distribution of (non-dichotomized) human codings are plotted for each cell. These plots reveal that misclassified segments tend to lie close to the peace-conflict threshold, whereas segments that lie further from the threshold are more often classified correctly.

Together, these results show that some conflictual debate is auditorily difficult to distinguish from peaceful speech, partially because "conflict" is in fact a continuous spectrum and and partially because coding was influenced by nonverbal cues and text. This suggests that future work may improve performance by joint consideration of text, audio, and video.

SAM also recovers reasonable estimates for speech dynamics. In the test set, the probability that conflict debate continues from one 30-second segment into the next segment (self-transition) is 0.67. Thus, the expected duration of conflictual debate is 90 seconds. The posterior mean for this parameter is precisely 0.67 (0.57, 0.80). Starting from a peaceful segment, the probability sustained peace in the next segment is 0.60, so that peaceful debate can be expected to last roughly

Figure 3-2: Performance of selected model on test set. The upper plot shows that most peaceful segments are correctly classified as such. A greater proportion (0.23) of conflictual segments are erroneously classified as peaceful, in part due to nonverbal and textual cues that help human coders detect conflict but are indistinguishable to SAM. The lower plot illustrates the distribution of the underlying (non-dichotomized) scale in each cell, showing that misclassified segments are far more likely to lie close to the peace-conflict threshold.



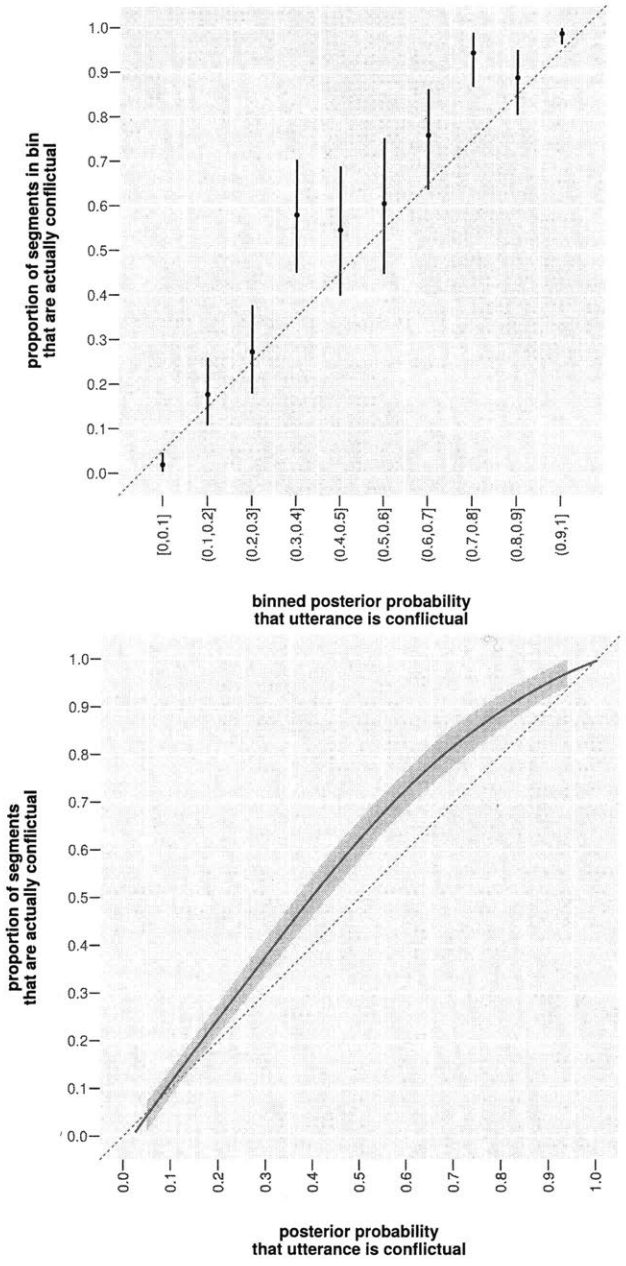
75 seconds. The posterior mean for peaceful self-transition is also 0.67 (0.52, 0.78).

Finally, Figure 3-3 shows that posterior class probabilities are moderately well-calibrated, though the selected model under-predicts conflictual debate. While there is little reason to expect that a Bayesian model should possess this frequentist property, it is often desirable from the perspective of interpretation and empirical application. Calibration is attained when among utterances in the test set that are assigned a particular posterior class probability p, the proportion that actually belong to that class is in fact p. Calibration can be improved by applying a monotonic transformation to the posterior probabilities, although this option is not explored here. These approaches typically run a logistic or isotonic regression in the validation set, where posterior probabilities are the independent variable and actual class membership is the outcome, then use the parameters learned in the validation set to transform model output for the test set.

Conclusion

In this paper, I introduced a new hierarchical hidden Markov model, the speaker-affect model, for classifying modes of speech using audio data. The model is shown to be suitable for use in measuring conflict in political deliberation, opening the door to new experimental designs that randomly assign debate formats or topics and use SAM to measure outcomes. The approach developed here has a broad range of possible substantive applications, from speech in parliamentary debates (Goplerud et al., 2016) to television news reporting on different political topics. With other interesting results on the importance of audio as data (Dietrich et al., 2016) accumulating, this model is a useful and general solution that improves on existing approaches and broadens the set of questions open to social scientists.

Figure 3-3: Performance of selected model on test set. The upper plot shows that most peaceful segments are correctly classified as such. A greater proportion (0.23) of conflictual segments are erroneously classified as peaceful, in part due to nonverbal and textual cues that help human coders detect conflict but are indistinguishable to SAM. The lower plot illustrates the distribution of the underlying (non-dichotomized) scale in each cell, showing that misclassified segments are far more likely to lie close to the peace-conflict threshold.



Appendix A

Supporting materials for Paper 1

Exact Enumeration of Paths

Proof of Proposition 1

This appendix is structured as follows. After introducing the necessary notation, I discuss some properties of the loop-erased random walk. I then outline a procedure that will be used in the proof. Finally, the proof is presented.

Notation

Where the notation in this appendix differs from the simplified exposition in the main text, a note is made.

Let $\tilde{G} = (V, \tilde{E})$ be an undirected, unweighted graph, where $\tilde{E}$ is set of edges (versus an edge-weight matrix in main text). The path $\psi = (V_\psi, E_\psi)$ is a connected subgraph of $\tilde{G}$ that contains no loops or branches (versus a node sequence in main text, where intervening edges were left implicit).

A subgraph of $\tilde{G}$ is a spanning tree if (*i*) it contains all vertices V, and (*ii*) every pair of vertices in V is connected by a single unique path on the subgraph. Denote the set of all spanning trees on

Data:
starting node γ_0, terminus γ_k, covariates $\boldsymbol{X}$, parameters β

Result:
$\mathscr{P} \equiv \{\psi : \Omega_\Gamma, |\{\psi\}| = |\psi|\}$, set of all paths from γ_0 to γ_k
$\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)$, probability that a random walk is a path

Algorithm `PrPath(`γ_0, γ_k, $t = |\gamma|$, $\boldsymbol{X}$, β`)`
initialize $\psi = (\gamma_0)$, $t = |\gamma| = 1$, $\mathscr{P} = \{\}$
populate $\mathscr{P}$ by `recursiveDFS(`γ, t, $\mathscr{N}_{\gamma_{t-1}}$`)`

initialize $\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta) = 0$
for $\psi \in \mathscr{P}$ **do**
  $\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)$ += $\Pr(\Gamma = \psi \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)$
end
return $\Pr(\Gamma \in \mathscr{P} \mid v_0 = \gamma_0, v_K = \gamma_k, \boldsymbol{X}, \beta)$

Procedure `recursiveDFS(`ψ, $t = |\psi|$, $\mathscr{N}_{\psi_{t-1}}$`)`
for $j \in \mathscr{N}_{\gamma_{t-1}}$ **do**
  if $j = \gamma_k$ **then**
    append j to ψ
    path to terminus found; append ψ to $\mathscr{P}$
  else if $j \in \psi$ **then**
    j already visited; proceed to next neighbor
  else
    append j to ψ
    continue search by `recursiveDFS(`ψ, $t+1$, $\mathscr{N}_j$`)`
  end
end
pop ψ_t from ψ

Algorithm 2: Calculating the probability that a random walk from γ_0 to γ_k is a path, using depth-first search (DFS) to exhaustively enumerate the set of all paths, $\mathscr{P}$. DFS starts at γ_0 and visits each neighbor in turn, expanding recursively as far as possible until the terminus γ_k is found or no new neighbors are available. The probability that a random walk is in $\mathscr{P}$ is then calculated by summing the probabilities of mutually exclusive events.

$\tilde{G}$ as $\mathscr{T}$, and let $\tilde{G}\tau(G)$ be the number of such trees. The path ψ is "on" a particular spanning tree $T = (V, E_T)$ if it is a subgraph of T; this holds if $E_\psi \subseteq E_T$, since necessarily $V_\psi \subseteq V$.

LERW Properties

Wilson's algorithm (Wilson, 1996) takes as input the graph $\tilde{G}$ and some ordering of its nodes $U = (u_1, \cdots, u_N)$, then returns a random sample from the set of possible spanning trees, $\mathscr{T}$. In brief, the algorithm starts from u_2 and performs a LERW until u_1 is reached, then marks all nodes and edges along the resulting path as visited. It then proceeds to the next node in U that has not been previously visited, performs a LERW until reaching any previously visited node, and again marks everything along that path as "previously visited." This process is iterated until all nodes have beeen visited. The resulting set of visited nodes and edges is a spanning tree on $\tilde{G}$; Wilson showed that for any choice of U, the procedure samples each element of $\mathscr{T}$ with equal probability. See also Lawler (1999)[pp. 211–212] for a more illuminating proof.

Corollary A.0.1. $\Pr\left(\mathit{LERW}(\tilde{\mathrm{G}}, \mathrm{u}_2, \mathrm{u}_1) = \psi\right)$ *is proportional to the number of spanning trees on* $\tilde{G}$ *that contain* ψ.

Proof. Let W be a spanning-tree-valued random variable whose probability mass is uniformly distributed over elements of $\mathscr{T}$. Wilson's algorithm is a procedure to sample W, in which $\mathtt{LERW}(\tilde{G}, u_2, u_1)$ is the first step. A spanning tree contains one unique path between u_2 and u_1. Therefore,

$$\Pr\left(W = T, \mathtt{LERW}(\tilde{G}, u_2, u_1) = \gamma\right) = \begin{cases} \Pr(W = T) & \text{if} \quad \gamma \text{ is on } T \\ 0 & \text{if} \quad \gamma \text{ is not on } T \end{cases}$$

It immediately follows that

$$\begin{aligned} \Pr\left(\mathtt{LERW}(\tilde{G}, u_2, u_1) = \psi\right) &= \sum_{T \in \mathscr{T}} \Pr\left(\mathtt{LERW}(\tilde{G}, u_2, u_1) = \psi \mid W = T\right) \Pr(W = T) \\ &= \sum_{T \in \mathscr{T}} \mathbf{1}\{\psi \text{ is a subgraph of } T\} \frac{1}{|\mathscr{T}|}. \end{aligned} \tag{A.1}$$

□

Deletion-Contraction Recurrence

I now outline a method that will be needed to count trees that contain a path. Consider an arbitrary edge, e, in $\tilde{G}$. The deletion-contraction recurrence (see, e.g. Bollobás, 1998, Theorem X.5.10, pp. 351–353) states that $\mathscr{T}$ can be divided into two disjoint sets: the set of spanning trees that do not use e, and the set of spanning trees that do. The former is in one-to-one correspondence with the set of spanning trees on the *deletion*, denoted $\tilde{G}-e$, formed by cutting e. The latter is similarly in one-to-one correspondence with the set of spanning trees on the *contraction*, $\tilde{G}/e$, formed by fusing the endpoints of e into a single node.[1] Thus, $\tau(\tilde{G}) = \tau(\tilde{G}-e) + \tau(\tilde{G}/e)$.

Proof

We are now ready to prove Proposition 1.

By recursive deletion-contraction, there is a bijection between (*i*) the set of spanning trees on $\tilde{G}$ that contain ψ as a subgraph and (*ii*) the set of spanning trees on the iterated contraction $\tilde{G}/e_{\psi,1}/\cdots/e_{\psi,K}$, where $e_{\psi,t}$ is the t-th edge in ψ. Kirchoff's matrix-tree theorem states that the number of spanning trees on a graph is given by the determinant of any minor of the graph's Laplacian matrix,

$$\tau(\tilde{G}) = \det L_{(-i,-j)}(\tilde{G}),$$

for any i and j, where the Laplacian, $L(\tilde{G}) = \tilde{D} - \tilde{A}$, is the diagonal degree matrix less the adjacency matrix. Substituting into equation A.1 yields

$$f_{\text{LERW}}(\psi) = \Pr(\texttt{LERW}(\tilde{G}, u_2, u_1) = \psi) = \frac{1}{\tau(\tilde{G})} \det L_{(-i,-j)}(\tilde{G}/e_{\psi,1}/\cdots/e_{\psi,K}),$$

for the LERW importance-sampling distribution, versus the target uniform distribution $f(\psi) = \frac{1}{|\mathscr{P}|}$. The corrective weight for importance sampling is the ratio of the latter relative to the former,

[1] Note that this procedure may result in a multigraph.

which is

$$\frac{\tau(\tilde{G})}{|\mathscr{P}| \det L_{(-i,-j)}(\tilde{G}/e_{\psi,1}/\cdots/e_{\psi,K})}$$

□

Details of RPM Estimation by MCMC

In this appendix, I discuss computational issues in estimation and provide an algorithm for sampling from the RPM posterior via MCMC. First, the sampling-reweighting procedure involves large numbers of simulated paths and expensive matrix determinants. Rather than repeating algorithm 1 in its entirety for each MH proposal, it is clearly advantageous to pre-compute a single batch of paths and their weights. This has the ancillary benefit of reducing noise in the MH acceptance ratio, as the simulated likelihood of both current and proposed parameters are estimated with the same path-set.

To evaluate the likelihood at any point in the parameter space, algorithm 1 must compute the unconditional (random-walk) probabilities of many paths. Because MCMC methods frequently revisit a relatively small, dense-probability region in the parameter space, a naïve implementation will spend considerable time repeatedly evaluating the likelihood at infinitesimally differing points. An alternative that considerably reduces running time, at the expense of initialization time and memory, is to pre-compute a finely gridded piecewise-constant approximation of the likelihood across a wide subspace. However, this contradicts the spirit of MCMC and is computationally infeasible for parameter spaces of moderate dimension. I implement a compromise by lazy evaluation of the likelihood over the parameter grid. In areas that are never sampled by MH, the computational cost is never incurred and memory usage is greatly decreased. After a cell is sampled by the MH proposal distribution, the likelihood is evaluated and cached for future use, or "memoized." Thus, chains will accelerate as they grow longer or more numerous, particularly when sampling the high-posterior-density region.

Data:

starting node γ_0, teminus γ_k, covariates $\boldsymbol{X}$

unweighted graph $\tilde{G}$, number of path simulations S

initial parameters $\beta^{(0)}$, gridded parameter space $\tilde{\mathscr{B}}$

number of Metropolis-Hastings samples R, proposal distribution $Q(\beta^*; \beta^{(t)})$

Result:

R correlated samples from posterior of parameters β

Algorithm `ChainMH`$(\gamma, \boldsymbol{X}, \beta^{(0)}, \tilde{\mathscr{B}}, Q)$

for $s \in 1, \cdots, S$ **do**

draw $\psi_s \sim \texttt{LERW}(\tilde{G}, \gamma_0, \gamma_k)$

calculate $w_s = \frac{1}{\det L_{(-i,-j)}(\tilde{G}/\psi_s)}$

end

set $\text{evaluated}_{\tilde{\beta}} = \texttt{FALSE}$ for all $\tilde{\beta} \in \tilde{B}$

for $r \in 0, \cdots, R$ **do**

draw proposed parameters $\beta^* \sim Q(\beta^*; \beta^{(r)})$

if parameter space is discretized **then**

calculate acceptance ratio $\alpha = \frac{\texttt{ApproxSimLikelihood}(\beta^*)}{\texttt{ApproxSimLikelihood}(\beta^{(r)})}$

else

calculate acceptance ratio $\alpha = \frac{\frac{\Pr(\Gamma=\gamma \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \beta^*)}{\sum_{l=1}^{s} w_s \Pr(\Gamma=\psi_s \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \beta^*)}}{\frac{\Pr\left(\Gamma=\gamma \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \beta^{(r)}\right)}{\sum_{l=1}^{s} w_s \Pr\left(\Gamma=\psi_s \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \beta^{(r)}\right)}}$

end

if $\alpha < 1$ and jump $\sim \text{Bern}(\alpha)$ **then**

set $\beta^{(r+1)} = \beta^*$

else

set $\beta^{(r+1)} = \beta^{(r)}$

end

end

return $\beta^{(0)}, \cdots, \beta^{(R)}$

Procedure `ApproxSimLikelihood`(β)

set $\tilde{\beta}$ to center of grid cell in $\tilde{B}$ containing β

if $\text{evaluated}_{\tilde{\beta}}$ **then**

return precomputed $\hat{\mathscr{L}}(\tilde{\beta} \mid \boldsymbol{X}, \gamma)$

else

set $\text{evaluated}_{\tilde{\beta}} = \texttt{TRUE}$

return and cache $\hat{\mathscr{L}}(\tilde{\beta} \mid \boldsymbol{X}, \gamma) = \frac{\Pr\left(\Gamma=\gamma \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \tilde{\beta}\right)}{\sum_{l=1}^{s} w_s \Pr\left(\Gamma=\psi_s \mid \nu_0=\gamma_0, \nu_K=\gamma_k, \boldsymbol{X}, \tilde{\beta}\right)}$

end

Algorithm 3: Implementing Metropolis-Hastings for a random-path model. Simulated likelihood calculations are memoized so that chains accelerate as they sample the highest-posterior-density region.

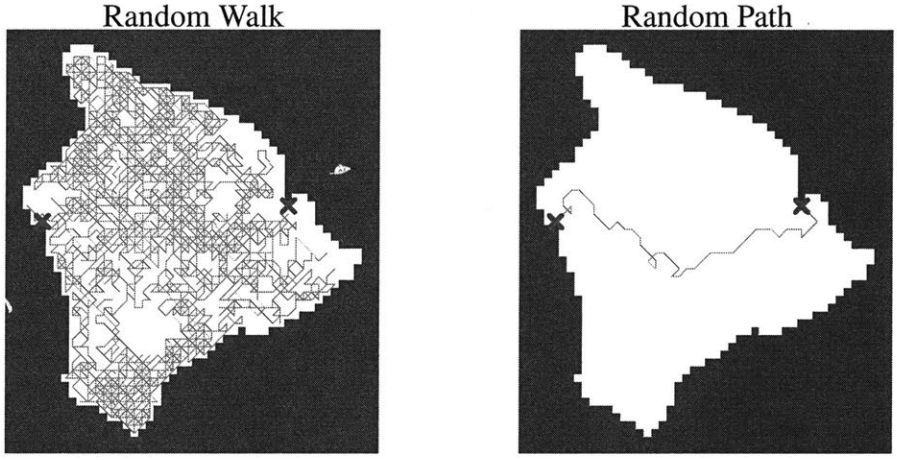
Simulation

In this section, I first demonstrate the properties of the random-path distribution with a naturalistic simulation. I then conduct a validation test in which a single path is drawn and its parameters are estimated by MCMC. This procedure is repeated at various sample sizes and graph resolutions in order to assess the consistency of the estimation procedure.

Simulation Distribution of Random Paths

The simulation ground is a virtual Hawai'i Island, rasterized into square cells of varying size. Each cell is connected to a tic-tac-toe board consisting of the 8 adjacent cells and excluding self-loops. I assume that a single road will be constructed from the western economic center, Kona, to the county seat in the east, Hilo. Figure A-1 depicts the difference between a typical random walk and path on the unweighted graph; it is perhaps unnecessary to point out that the random path bears a closer resemblance to actual Hawai'ian state highways.

Figure A-1: One draw each from the random walk and random path distributions, on a 50×50 grid, with all parameters set to zero.



One might reasonably expect a Hawai'ian road to avoid excessively mountainous regions, while

passing through as many villages as possible without deviating too far from a direct course. As a point of reference, actual state highways on the Big Island are roughly Θ-shaped, consisting of a circular coastal highway and Saddle Road, which cuts directly from Kona to Hilo. To capture this behavior, I include transformations of three covariates: (1) directness $\texttt{dir}_{ij}$, or how much closer the $i \to j$ step brings a walker to the target; (2) elevation $\texttt{elev}_j$; and (3) population "gravity."[2] The covariates are shown in figure A-2. For a walker at cell i, the unconditional (random-walk) probability of stepping to adjacent cell j is

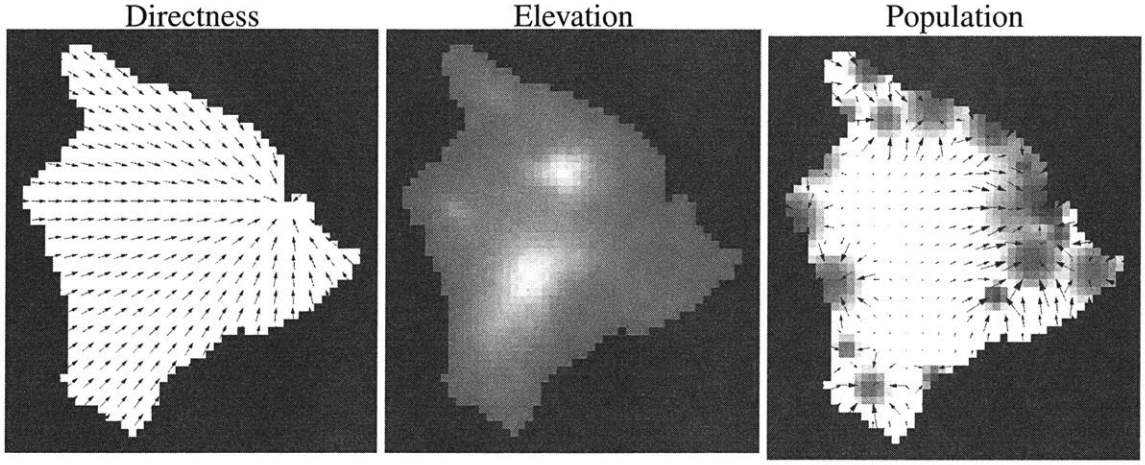
$$\frac{\exp\left(\beta_{\text{dir}} \cdot \texttt{dir}_{ij} + \beta_{\text{elev}}\texttt{elev}_j + \beta_{\text{pop}} \cdot \texttt{lpop}_{ij}\right)}{\sum_{j' \in N_i} \exp\left(\beta_{\text{dir}} \cdot \texttt{dir}_{ij'} + \beta_{\text{elev}}\texttt{elev}_{j'} + \beta_{\text{pop}} \cdot \texttt{pop}_{ij'}\right)}.$$

The random-path distribution is the conditional random-walk distribution, given that the walk does not contain cycles. The simulation distribution of a random-path model is the result of importance-sampling S paths, calculating the random-walk probability of each, then resampling from the S paths with probability proportional to random-walk probability times inverse-importance weights. The simulation distribution converges to the true RPM distribution as S increases; in the illustrations that follow, I use $S = 10^6$ and resample 10^2 paths.

In figure A-3, I show the result of increasing β_{dir}. The left panel in figure A-3 is a larger sample from the baseline distribution with all parameters set to zero (the same RPM that generated the right panel of figure A-1). The baseline distribution is the path-conditioned version of a random walk in which all adjacent cells are equally likely. After conditioning to walks that contain no cycles, the random-path distribution exhibits a strong baseline preference for shorter (more direct) paths. This is because the longer a walk continues, the more likely it is to double back on itself. In the

[2]Directness is calculated as the inner product of the step vector $(\texttt{location}_j - \texttt{location}_i)$ with a unit vector pointing from i to Hilo. Elevation is rasterized by averaging National Elevation Dataset values within j, then scaled and exponentiated to increase separation. Raster-cell population is generated to be consistent with 1940 census tract data (with Gaussian allocation of tract population around approximate coordinates of in-tract villages); each cell is assumed to generate a gravitational pull proportional to its log-population and the inverse squared distance, and $\texttt{pop}_{ij}$ is operationalized as the inner product of the step-vector $i \to j$ with the aggregate gravitational field at i.

Figure A-2: RPM covariates on a 50×50 grid. Direction toward target (left) indicated by arrows. Elevation (center) in terrain colors, green at sea level and white at $\sim 4,000$ m, around the peaks of Mauna Kea and Mauna Loa. Log-population (right) plotted in red, with higher density in more opaque regions. Arrows show the direction of population gravitational pull, with arrow size indicating force.



right panel, I show that this natural tendency can be reinforced by increasing β_{dir}; at higher values, the random path distribution becomes tighter and more focused. Figures A-4 and A-5 depict the effects of β_{elev} and β_{pop}, respectively.

Figure A-3: Higher values of β_{dir} (right) result in a tighter distribution with more direct paths than the baseline (left).

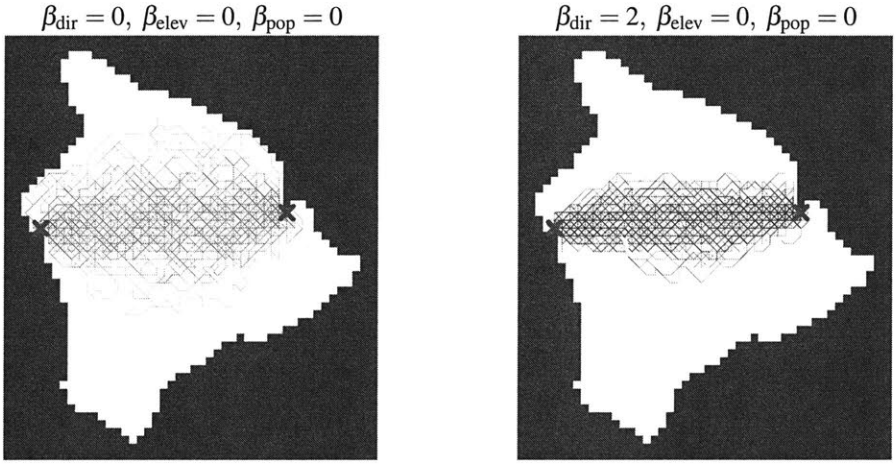
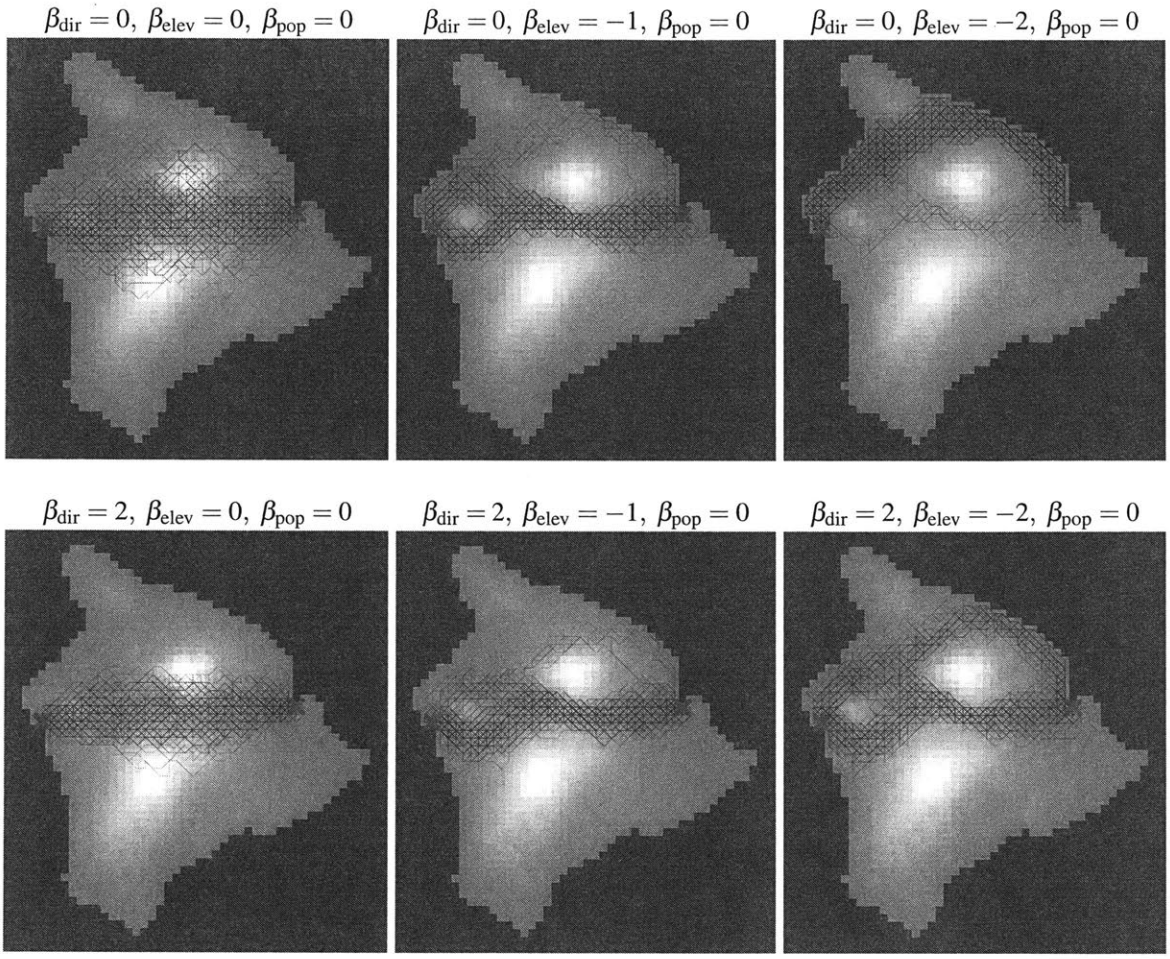


Figure A-4: Increasingly negative values of β_{elev} (moving right) result in distributions that avoid mountainous regions. However, this tendency can be partially overcome by higher values of β_{dir} (lower plots), which drive the path distribution over the saddle pass directly toward Hilo.

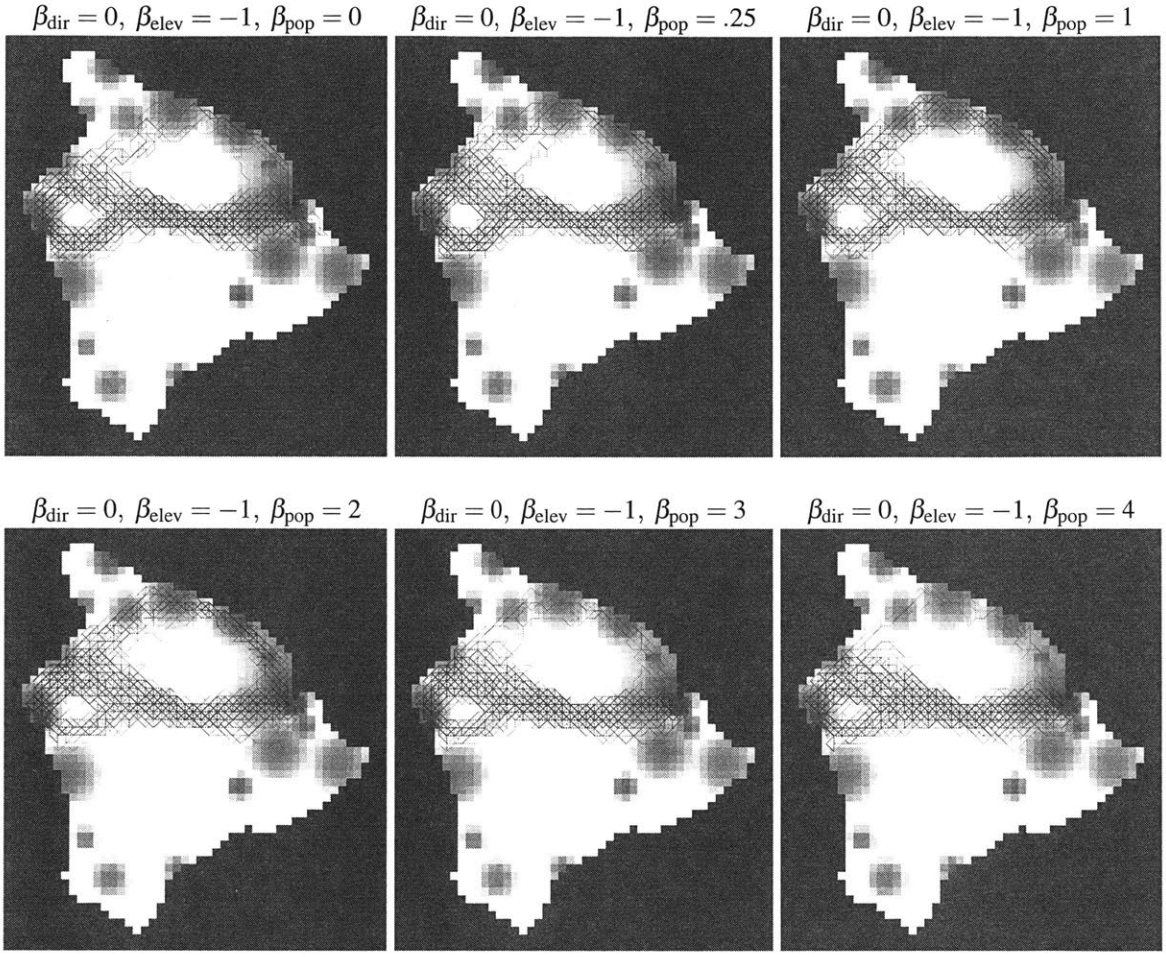


Validating the Estimation Procedure

Convergence

I first assess the MCMC convergence of the RPM posterior by randomly drawing a single path from RPM($\beta_{\text{dir}} = 0, \beta_{\text{elev}} = -1, \beta_{\text{pop}} = 0.5$). The true distribution was chosen such that with a single draw, equivalent to perusing a map, a reasonable human observer would consider $\hat{\beta}_{\text{elev}}$ to be negative and statistically significant and both $\hat{\beta}_{\text{dir}}$ and $\hat{\beta}_{\text{pop}}$ to be perhaps slightly positive but

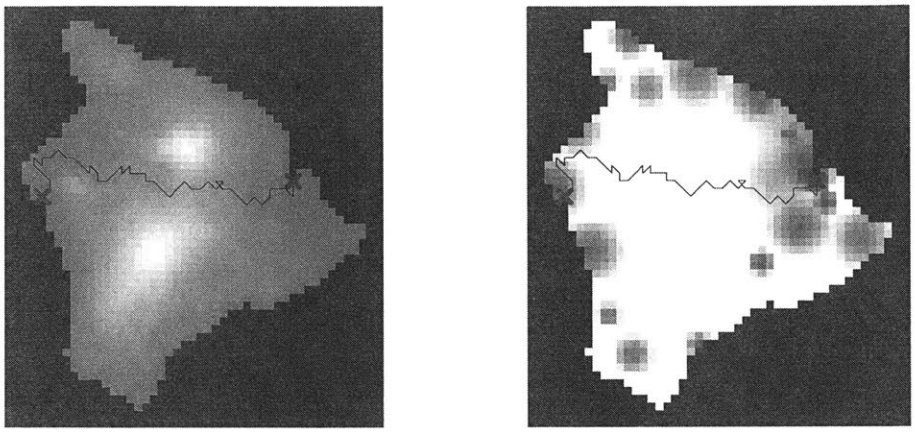
Figure A-5: Small increases in β_{pop} (upper row, moving right) make coastal paths more likely to visit small towns instead of passing by (esp. Waimea, on the northern peninsula), then begin to redirect paths away from the saddle pass and toward coastal population centers. At very large values, however, this effect reverses as paths are pulled directly over the pass by the strong gravitational pull of the large Hilo population (lower right).



indistinguishable from zero. In fact, the first sampled path (shown in figure A-6) captures this intent nicely. Starting in the west at Kona, the path tracks the city limits as it diverts around Hualalai, the volcano just outside the city, then traverses the saddle pass before exiting with a slight flourish. I examine the extent to which the RPM posterior reflects these patterns. The effective number of observations in a single path, after accounting for dependence, is somewhere in $[1, k]$.

I evaluate the mixing of MH-sampled MCMC and the resulting estimates. Chain length was

Figure A-6: A single draw from RPM($\beta_{\text{dir}} = 0, \beta_{\text{elev}} = -1, \beta_{\text{pop}} = 0.5$), plotted against elevation (left) and population (right).



5,000 draws and the reduction in effective posterior sample size due to autocorrelation was a factor of roughly 15, differing only slightly by parameter. This left an effective sample size of roughly 300–350 and sampling standard errors of parameter posterior means between 0.02 and 0.05—more than an order of magnitude smaller than estimated posterior standard deviations, and quite acceptable for present purposes. Chains for each parameter, posterior means, and 95% posterior credible intervals are shown in figure A-7; elevation was estimated to be negative and correctly signed, while all other parameter estimates were insignificant.

Consistency

Next, I examine the Bayesian consistency of the RPM estimation procedure. Specifically, I generate paths according to a true distribution, then evaluate whether the posterior means and variances of the distribution parameters go to the true parameter and zero, respectively, (*i*) as the number of paths increase, but approximate length of each path remains fixed; and (*ii*) as paths grow longer, but the number of paths remain fixed. To test (*i*), I examine the RPM posterior distribution given sample sizes of 1, 2, 4, 8, and 16 paths between fixed endpoints on the same graph. For (*ii*), I re-rasterize the Hawai'i simulation ground into 10×10, 20×20, and 40×40 square grids, then

compare the posterior on these graphs given a fixed number of sampled paths. Given computational constraints, I focus here on the elevation parameter only. The true parameter used below is $\beta_{\text{elev}} = -2$.

The true RPM distributions are shown in figure A-8 for each grid size. The procedure used is as follows: For the 10×10 grid, a single path was sampled and its posterior distribution was approximated by algorithm 3; this corresponds to the first horizontal line in the top-left panel of figure A-9. In total, 100 single paths were drawn on the 10×10 grid—results are shown in the top-left panel.

Next, paths were sampled from the true model, two at a time; the approximate posteriors for 100 sampled pairs are shown in the second panel in the top row. This was repeated with samples of 4, 8, and 16 paths. The entire process was repeated for the 20×20 and 40×40 grids (second and third row of panels).

In this simulation, results initially show that estimates are correctly signed but unmistakably biased toward zero for short paths. Variance goes to zero, but bias does not disappear as the number of short paths increase. This suggests that for small graphs, a bias-correction step, such as simulating paths from the posterior and re-estimating, may be necessary. As the graph grows larger and paths grow longer, this bias disappears and estimates converge toward the true parameter.

Figure A-7: RPM posterior for the path depicted in figure A-6, with sampled parameter values on the vertical axis and iterations on the horizontal. Parameter posterior means (dashed) and 95% marginal posterior credible intervals (dotted) are plotted horizontally over each chain. The true parameter is marked with a "×" on the vertical axis.

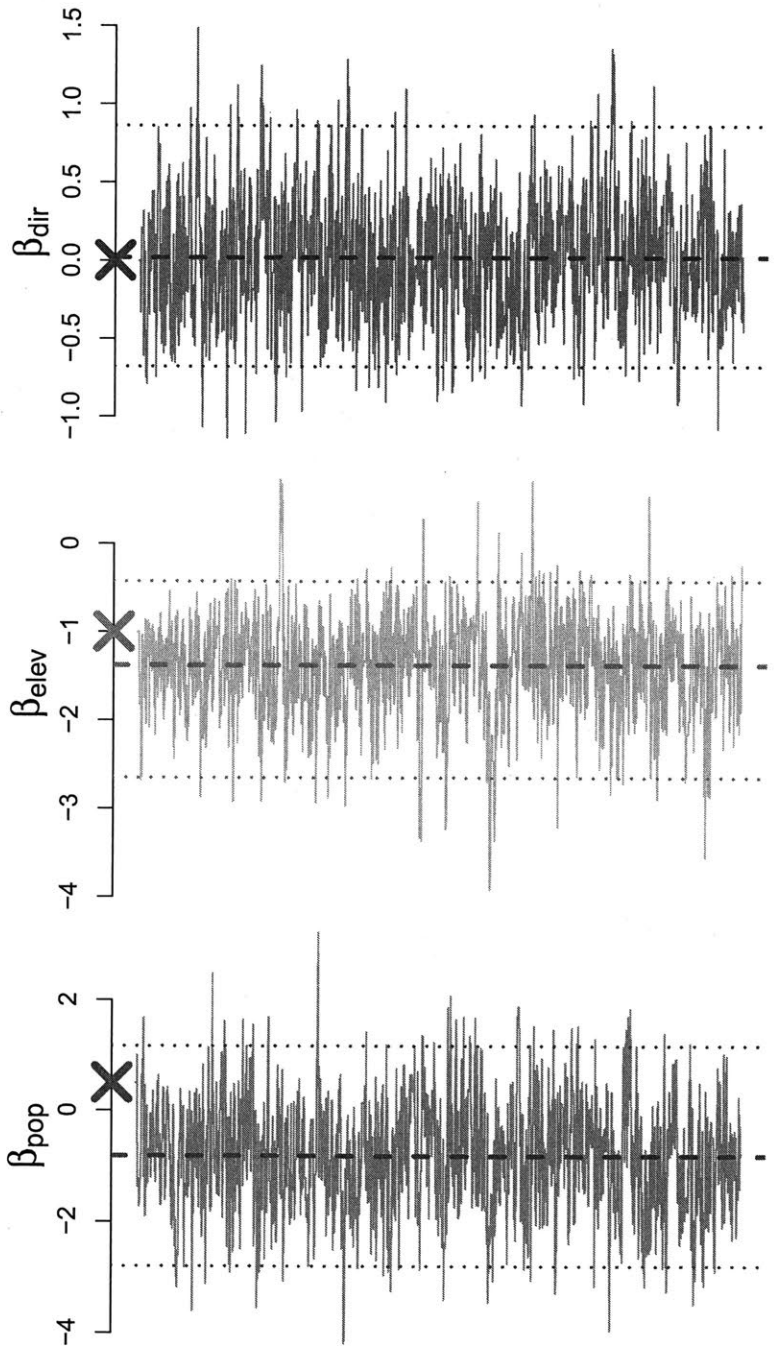


Figure A-8: 100 draws from RPM($\beta_{\text{elev}} = -2$) on 10×10, 20×20, and 40×40 square grids

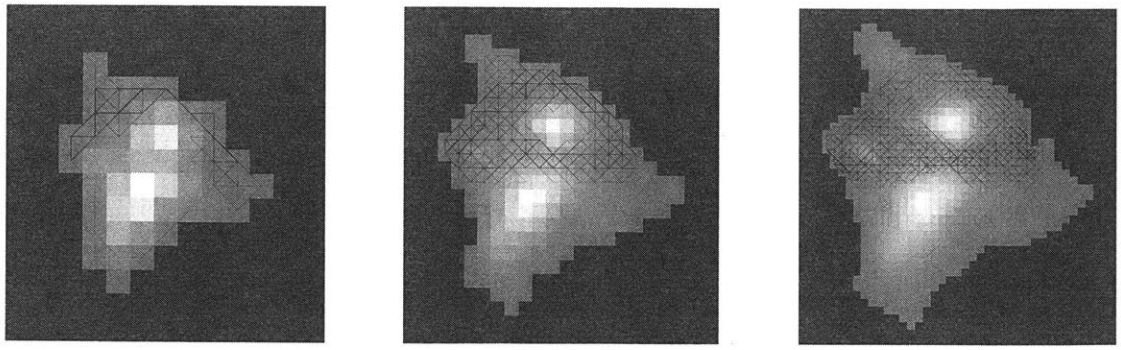
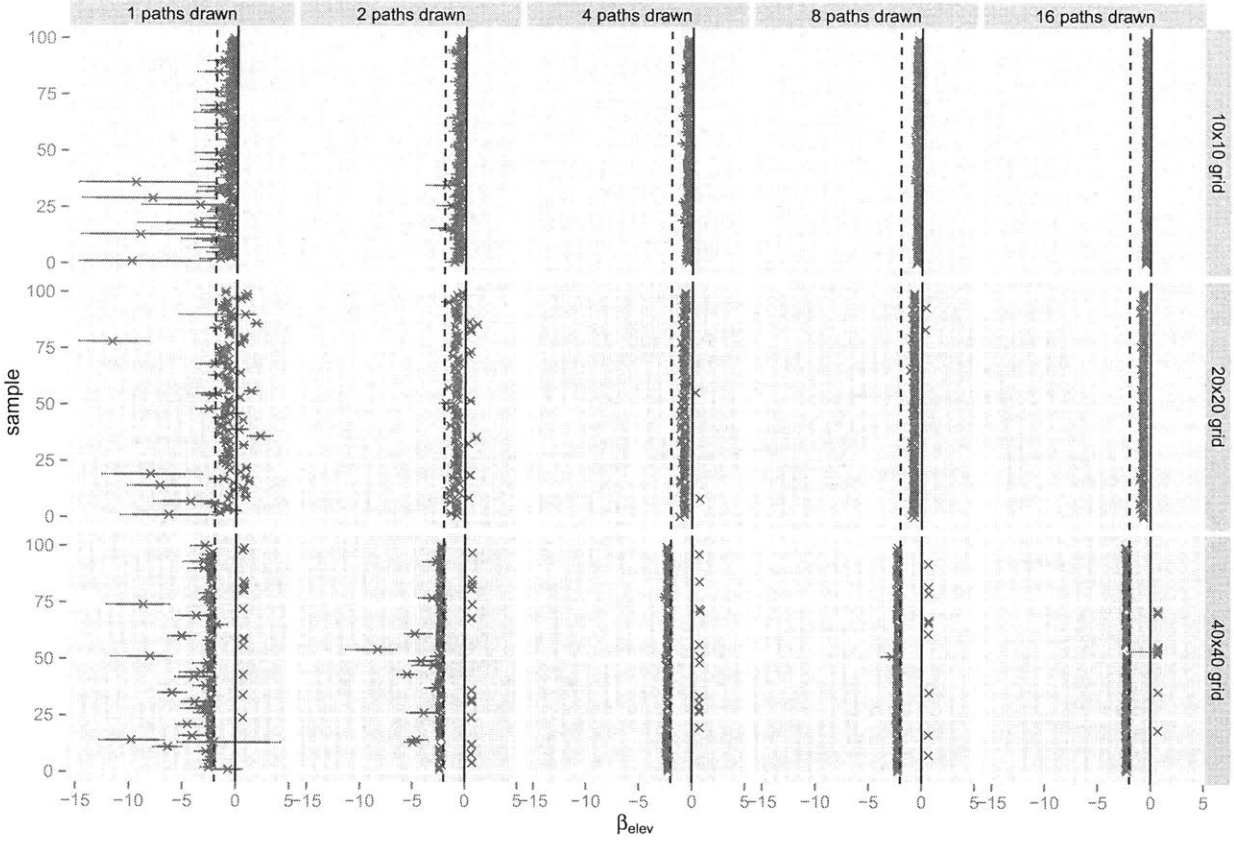


Figure A-9: For each combination of sample size and grid size, 100 samples were drawn. Posterior means are marked with "×", 95% credible intervals with thin horizontal green lines, and 80% intervals with thick green horizontal lines. The true parameter is shown with a vertical black dotted line. Results show that estimator variance converges to zero as sample size increases, but some bias remains when paths are short. This bias disappears as paths grow longer.



Convergence of U.S. Interstate Highways Estimates

The posterior of RPM parameters in the U.S. Interstate Highway application was simulated by MCMC. Five chains, of 10,000 samples each, were initialized at overdispersed locations. After a burn-in of 2,000 iterations, visual diagnostics show excellent mixing and low autocorrelation relative to chain length.

Figure A-10: After discarding the first 2,000 iterations of each chain as burn-in, marginal posterior densities of RPM parameters over the remaining 8,000 iterations are extremely similar. Separate colors and line types represent each chain. Vertical bars represent 2.5 and 97.5-th posterior percentiles.

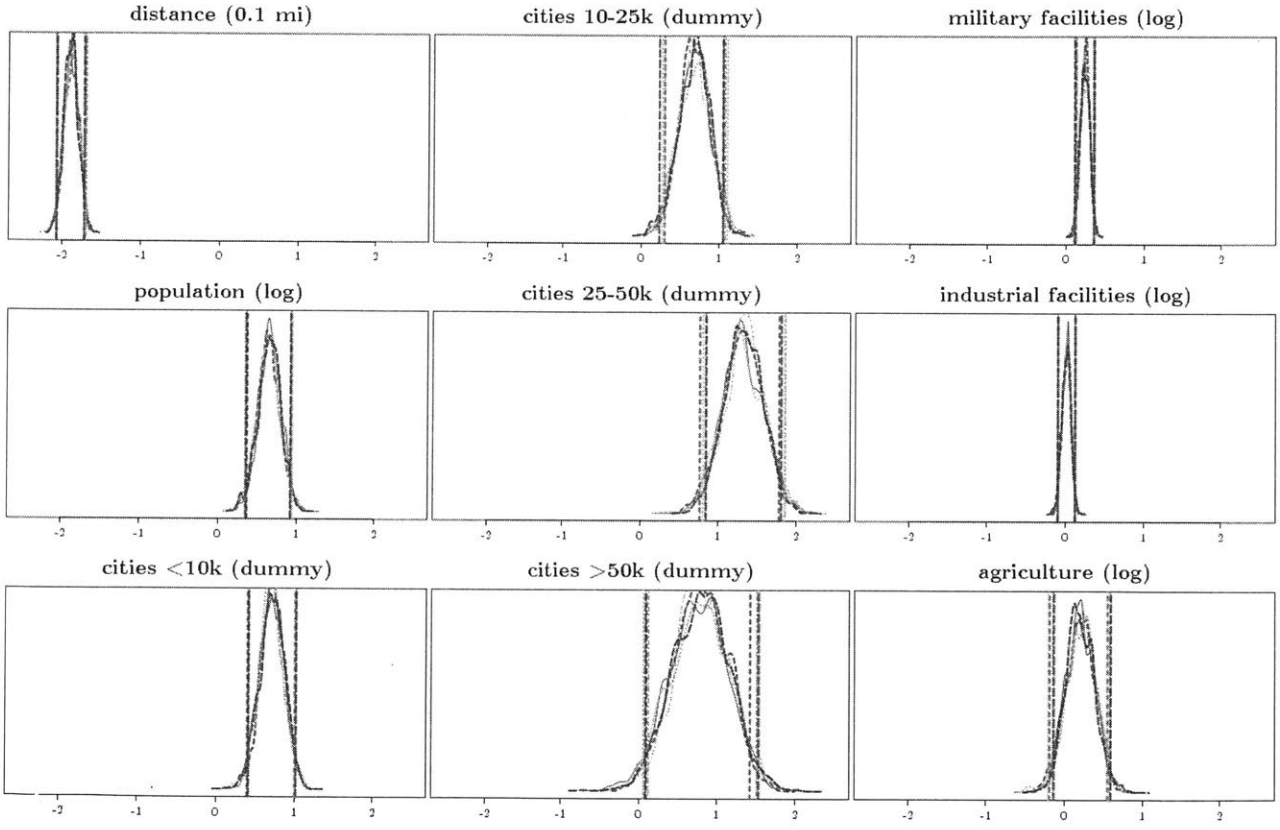


Figure A-11: Traces of five chains, denoted by color. Visual inspection suggests that a burn-in of 2,000 iterations is adequate and that autocorrelation is low relative to chain length.

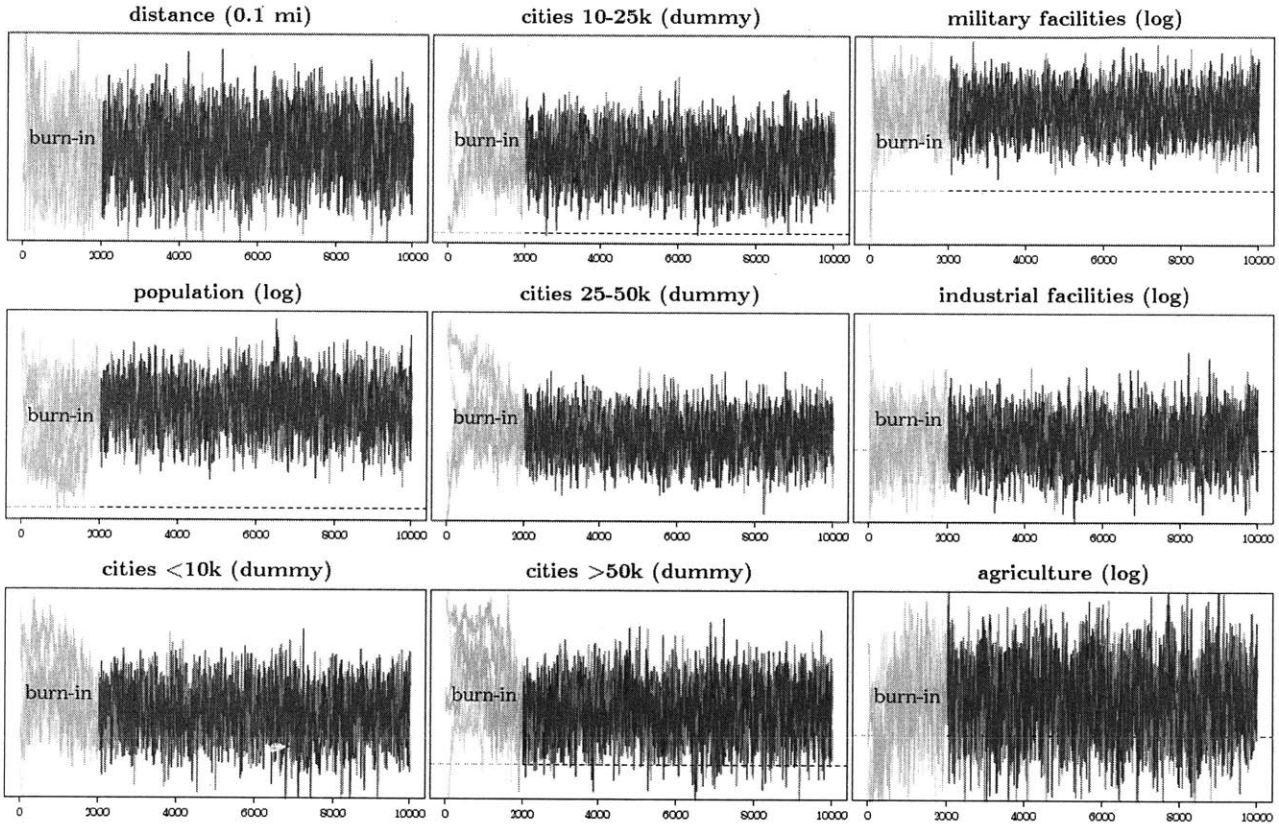
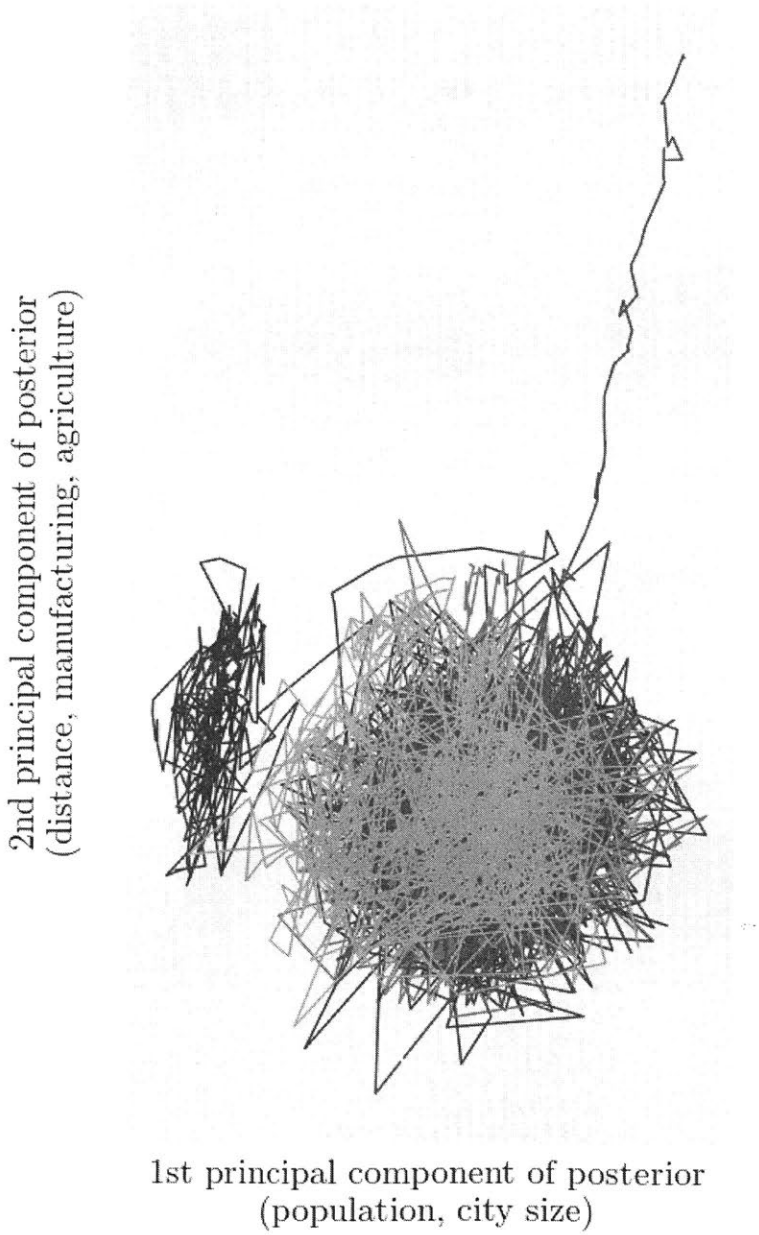


Figure A-12: Visualizing MCMC chains with the first two principal components of the posterior distribution. The plot shows that chains initialized at overdispersed starting positions converge to the same region in the parameter space, with excellent mixing. The first component roughly captures population-related covariates, and the second is a mix of the remaining covariates.



Convergence of Baghdad Sectarianism Estimates

The posterior of RPM parameters in the Baghdad walks application was simulated by MCMC. Three chains, of 10,000 samples each, were initialized at overdispersed locations. After a burn-in of 4,000 iterations, visual diagnostics show excellent mixing and low autocorrelation relative to chain length.

Figure A-13: After discarding the first 4,000 iterations of each chain as burn-in, marginal posterior densities of RPM parameters over the remaining 6,000 iterations are extremely similar. Separate colors represent each chain. Vertical bars represent 2.5 and 97.5-th posterior percentiles.

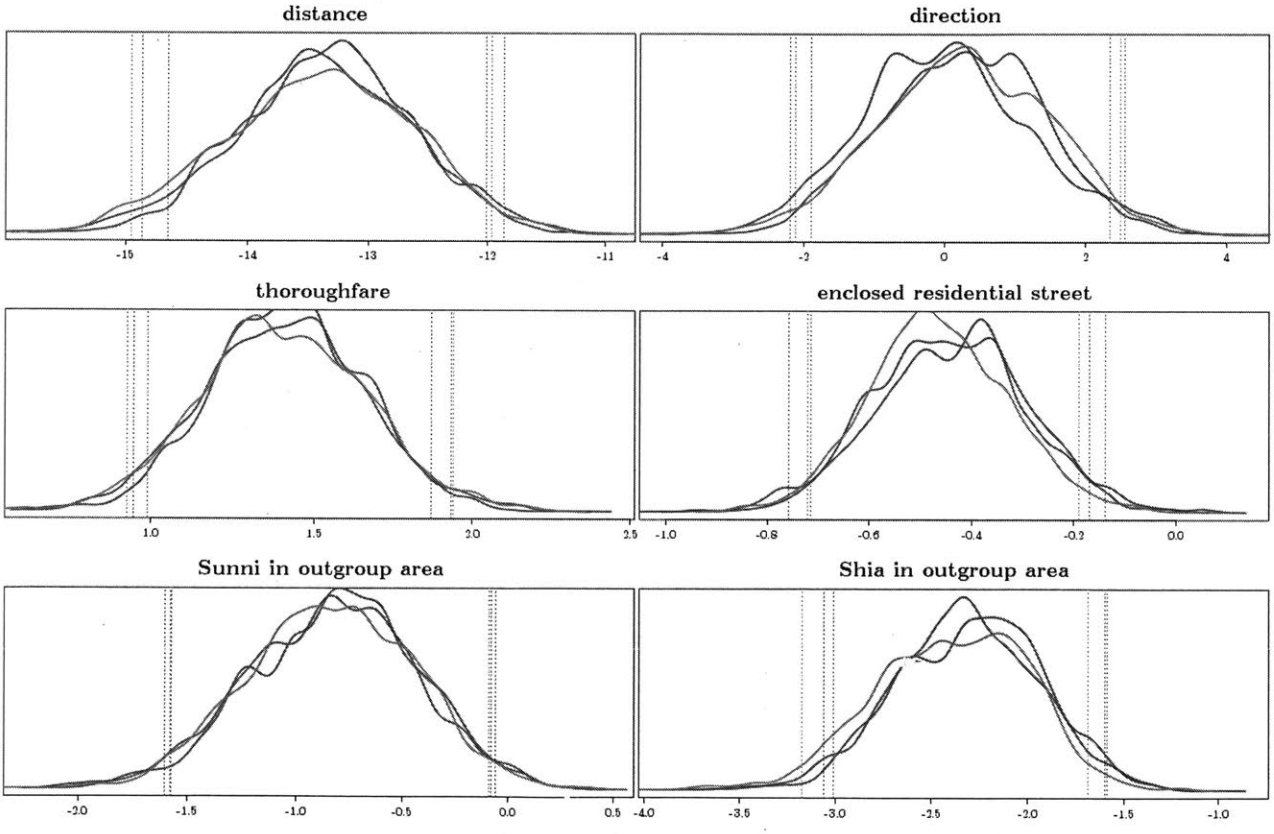


Figure A-14: Traces of three chains, denoted by color. Visual inspection suggests that a burn-in of 4,000 iterations is adequate and that autocorrelation is low relative to chain length.

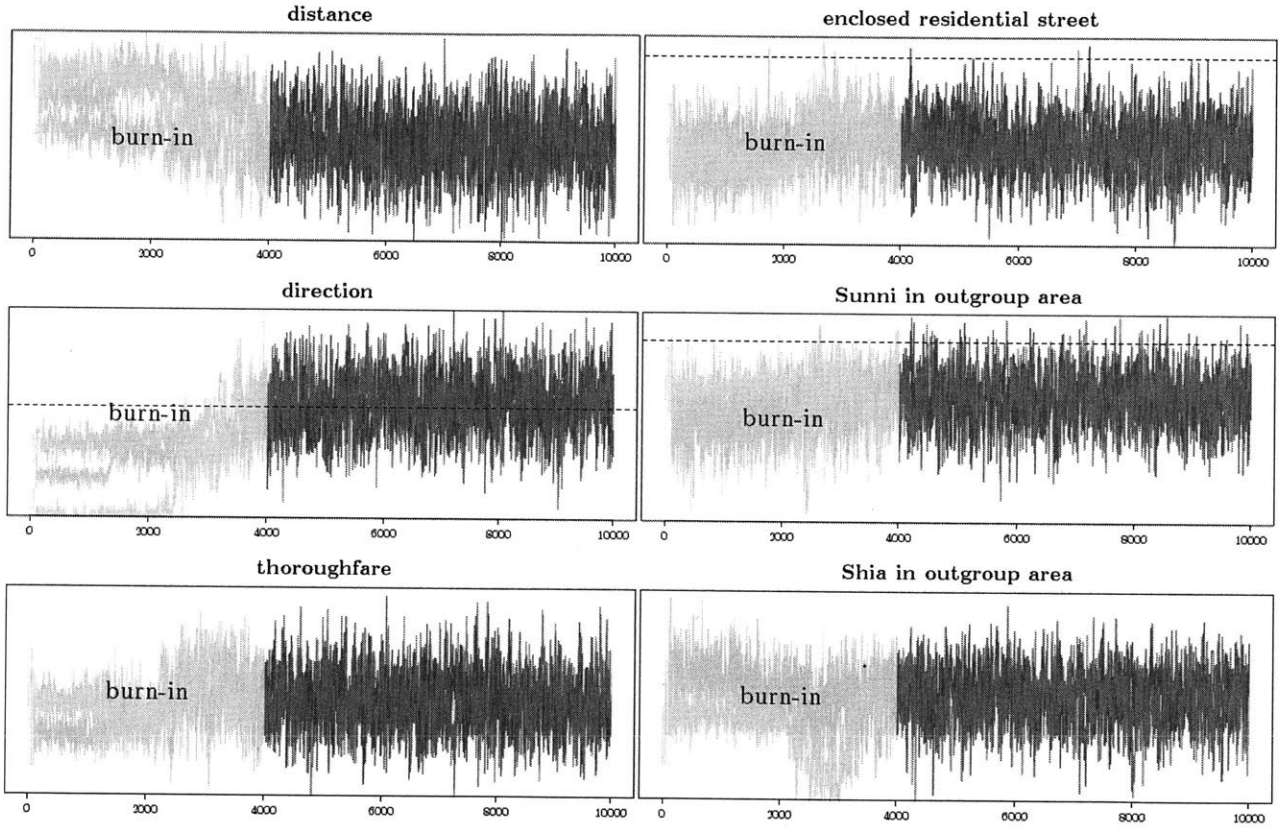
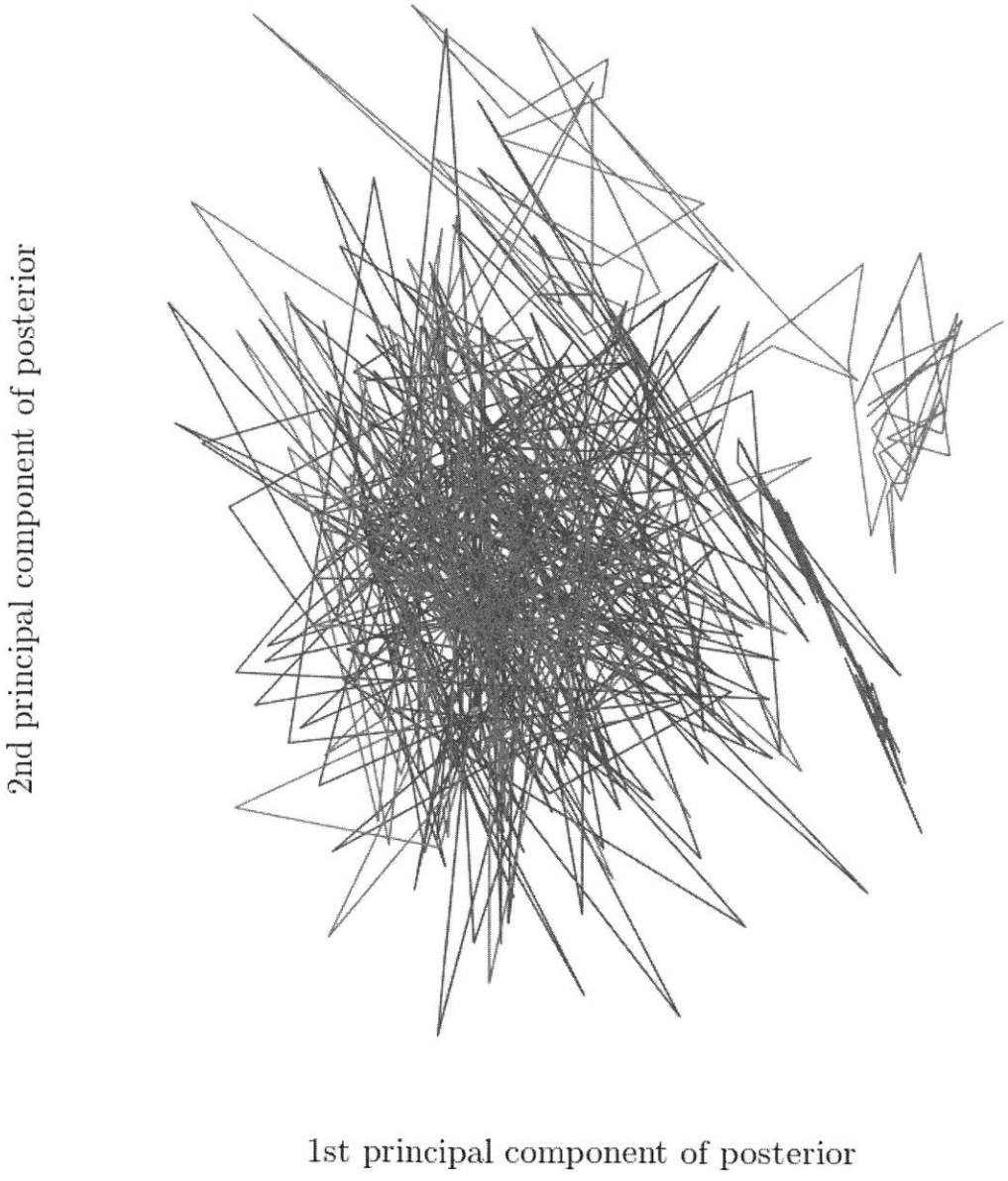


Figure A-15: Visualizing MCMC chains with the first two principal components of the posterior distribution. The plot shows that chains converge to the same region in the parameter space, with excellent mixing.

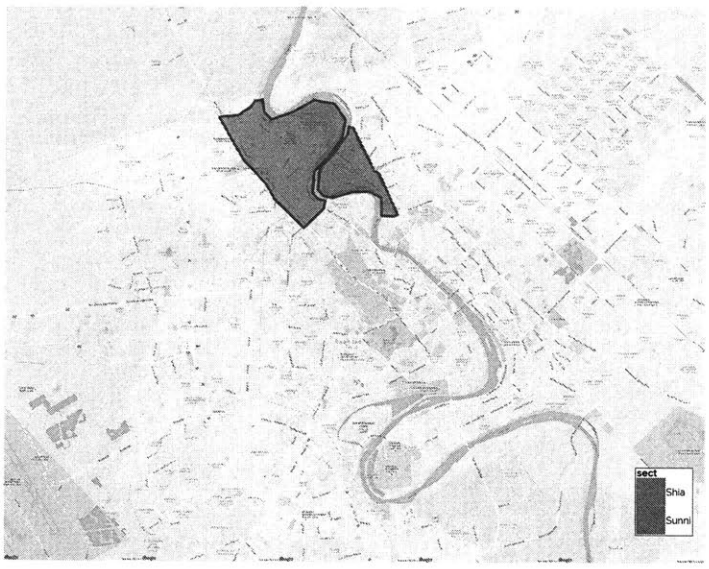


Appendix B

Supporting materials for Paper 2

Figures

Figure B-1: Recruitment neighborhoods. Shia Kadhimiya in blue, Sunni Adhamiya in red.



Tables

Figure B-2: Target areas, with triangle denoting council office location. Shia Zayouna in blue, Sunni Ghazaliya in red.

Figure B-3: Balance between Sunni and Shia participants on basic demographics.

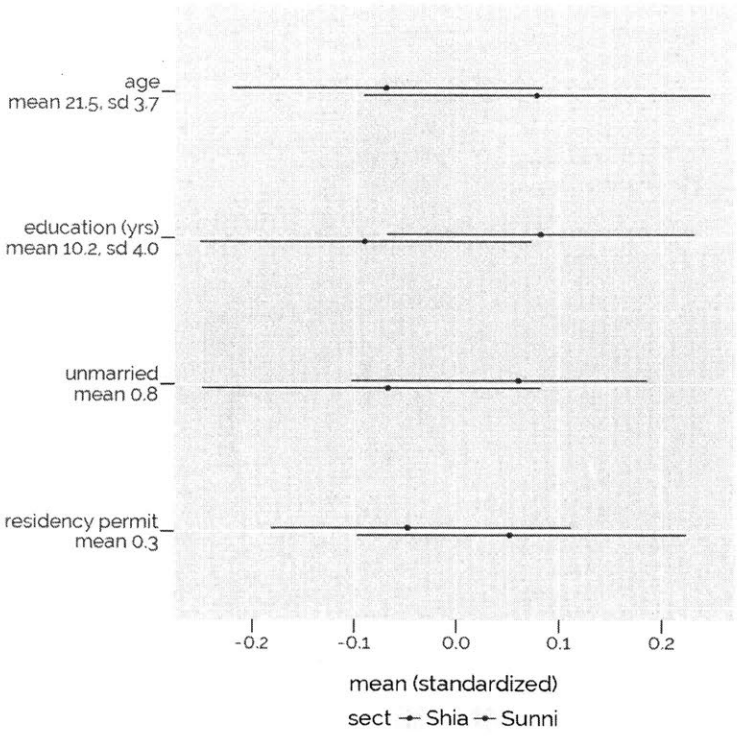


Figure B-4: Work sector of Sunni and Shia participants.

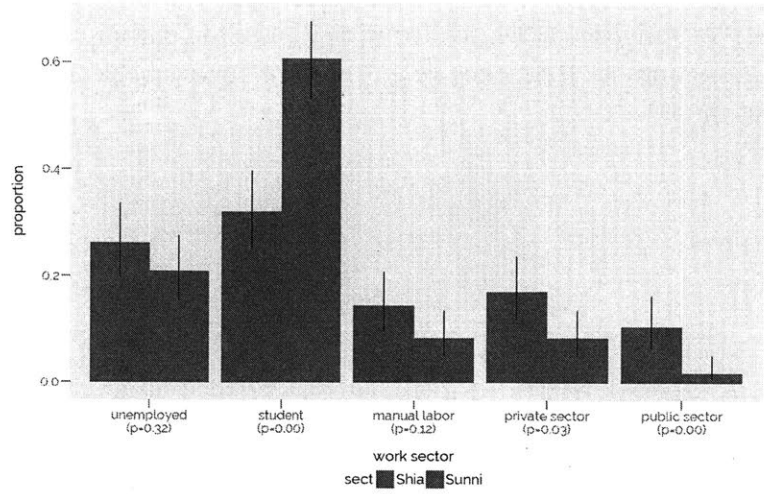


Figure B-5: Length of residency in Baghdad for Sunni and Shia participants. The large number of recent Sunni arrivals are likely to be internally displaced persons from conflict against the Islamic State.

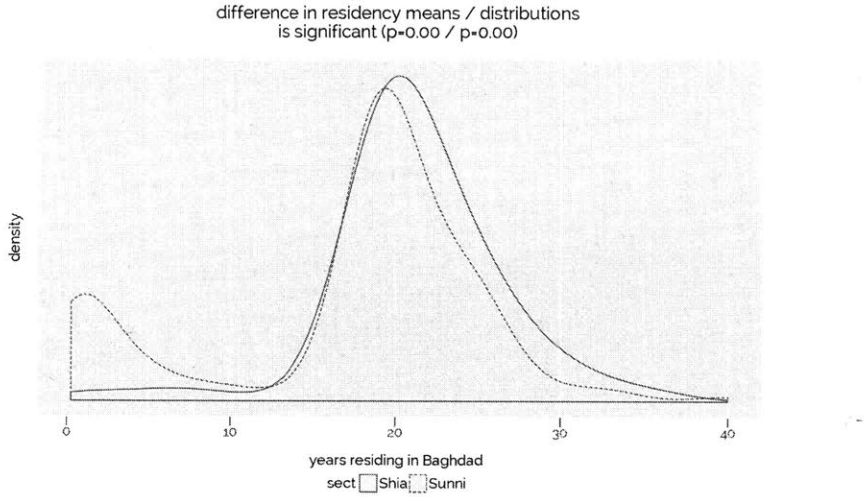


Table B.1: Recruitment neighborhoods

	Kadhimiya	Adhamiya	Baghdad typical (sd)
Population	842k	833k	800k (500k)
Illiteracy	7.0%	8.5%	10% (5pp)
Labor participation	85%	83%	83% (5pp)
Post-2006 IDPs	14%	12%	41% (9pp)
Water/sewer	100%	100%	90% (15pp)

Figure B-6: Frequency of religious activities for Sunni and Shia participants. Respondents chose from several categories of activity frequency (e.g., "at least once a week"), which were coded into approximate times per month.

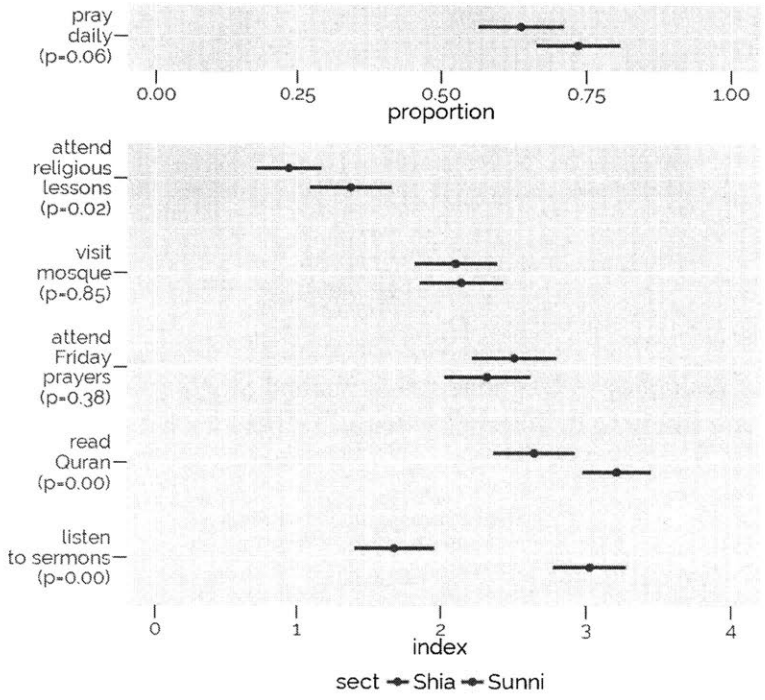


Table B.2: Assigned target neighborhoods

	Ghazaliya	Zayouna
Sect	~60% Sunni	~80% Shia
Residents	professionals	professionals
Real estate	US$1900/m^2	US$1900/m^2
Distance from Adhamiya/Kadhimiya	~15 km	~15 km
	20–25 minute drive	20–25 minute drive
Local council	5 members (3 Shia, 2 Sunni)	5 members (3 Shia, 2 Sunni)

Table B.3: Relative success in locating the assigned council office address

	Dependent variable: council address found			
	binary ($\geq$ partial)	ordered (none $<$ partial $<$ full)		
	OLS	*ordered probit*		
	raw	matched	reweighted	adjusted
	(1)	(2)	(3)	(4)
Shia participant in Shia target	-0.305***	-0.593	-0.612	-0.875**
(vs Shia participant in Sunni target)	(0.072)	(0.995)	(0.518)	(0.279)
Sunni participant in Shia target	-0.305***	-0.409	-0.591**	-0.572**
(vs Sunni participant in Sunni target)	(0.074)	(0.337)	(0.201)	(0.213)
Sunni participant in Sunni target	0.199**	0.354	0.381	-0.032
(vs Shia participant in Sunni target)	(0.080)	(0.684)	(0.291)	(0.292)
Sunni participant in Shia target	0.199**	0.538	0.402	0.270
(vs Shia participant in Shia target)	(0.067)	(0.827)	(0.430)	(0.318)
Experimental variables		matched	reweighted	adjusted
Demographic characteristics (main)		matched	reweighted	adjusted
Demographic characteristics (secondary)				adjusted
Observations	317	162	279	268

Note: cluster-robust standard errors by communities of players with shared contacts

*p<0.1; **p<0.05; ***p<0.01

Table B.4: Relative success in identifying a head, member, or employee from the assigned council

	Dependent variable: gatekeeper found (binary)			
	OLS raw	*OLS* matched	*OLS* reweighted	*logit* adjusted
	(1)	(2)	(3)	(4)
Shia participant in Shia target (vs Shia participant in Sunni target)	-0.158*** (0.046)	-0.323* (0.176)	-0.446** (0.174)	-3.943*** (0.971)
Sunni participant in Shia target (vs Sunni participant in Sunni target)	-0.244*** (0.067)	-0.242** (0.074)	-0.265*** (0.063)	-2.133*** (0.617)
Sunni participant in Sunni target (vs Shia participant in Sunni target)	0.207** (0.071)	0.042 (0.186)	-0.048 (0.182)	0.783 (0.756)
Sunni participant in Shia target (vs Shia participant in Shia target)	0.121** (0.040)	0.123** (0.047)	0.134*** (0.035)	2.593** (1.101)
Experimental variables		matched	reweighted	adjusted
Demographic characteristics (main)		matched	reweighted	adjusted
Demographic characteristics (secondary)				adjusted
Observations	317	162	279	268

Note: cluster-robust standard errors by communities of players with shared contacts

*p<0.1; **p<0.05; ***p<0.01

Table B.5: Distance from participant's home to contact's home

	Dependent variable: distance from participant's home to contact's home			
	OLS	*OLS*	*OLS*	*censored regression*
	raw	matched	fixed effects	adjusted
	(1)	(2)	(3)	(4)
Shia search contacts	1.419***	2.688***	1.517***	3.507***
(vs Shia daily contacts)	(0.408)	(0.461)	(0.325)	(0.839)
Sunni search contacts	2.342***	2.503***	2.150***	7.977***
(vs Sunni daily contacts)	(0.379)	(0.187)	(0.379)	(1.258)
Sunni daily contacts	-2.126***	-0.229		-6.666***
(vs Shia daily contacts)	(0.397)	(0.214)		(1.528)
Sunni search contacts	-1.203**	-0.414		-2.196
(vs Shia search contacts)	(0.490)	(0.485)		(1.361)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	1,169	1,169	1,159	1,023

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.6: Distance from contact's home to assigned target

	Dependent variable: distance from contact's home to assigned target			
	OLS raw	*OLS* matched	*OLS* fixed effects	*OLS* adjusted
	(1)	(2)	(3)	(4)
Shia search contacts	-0.864**	-1.682***	-0.961***	-0.946**
(vs Shia daily contacts)	(0.361)	(0.223)	(0.270)	(0.312)
Sunni search contacts	-0.849**	-1.032***	-1.134***	-1.126***
(vs Sunni daily contacts)	(0.286)	(0.142)	(0.315)	(0.303)
Sunni daily contacts	-1.624***	-1.775***		-0.225
(vs Shia daily contacts)	(0.432)	(0.320)		(0.328)
Sunni search contacts	-1.609***	-1.126***		-0.406
(vs Shia search contacts)	(0.412)	(0.290)		(0.469)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	1,185	1,185	1,177	1,023

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.7: Contact's education

	Dependent variable: contact education		
	ordered probit		
	raw	matched	adjusted
	(1)	(2)	(3)
Shia participant in cross-sect target	-0.167	0.683	-0.464**
(vs Shia participant in co-sect target)	(0.164)	(0.715)	(0.207)
Sunni participant in cross-sect target	-0.107	-0.086	-0.090
(vs Sunni participant in co-sect target)	(0.171)	(0.316)	(0.182)
Sunni participant in co-sect target	0.307*	0.444	0.147
(vs Shia participant in co-sect target)	(0.172)	(0.343)	(0.209)
Sunni participant in cross-sect target	0.367**	-0.325	0.521**
(vs Shia participant in cross-sect target)	(0.164)	(0.734)	(0.260)
Experimental variables		matched	adjusted
Demographic characteristics (main)		matched	adjusted
Demographic characteristics (secondary)			adjusted
Observations	472	468	407

Note: cluster-robust standard errors by participant *p<0.1; **p<0.05; ***p<0.01

Table B.8: Cross-sect contacts

	Dependent variable: contact is cross-sect		
	OLS raw	*OLS* matched	*logit* adjusted
	(1)	(2)	(3)
Shia participant in cross-sect target	0.143**	0.072	1.412*
(vs Shia participant in co-sect target)	(0.054)	(0.055)	(0.757)
Sunni participant in cross-sect target	0.279**	0.223**	1.536**
(vs Sunni participant in co-sect target)	(0.097)	(0.100)	(0.521)
Sunni participant in co-sect target	0.111**	0.146**	1.334*
(vs Shia participant in co-sect target)	(0.047)	(0.046)	(0.774)
Sunni participant in cross-sect target	0.247**	0.297**	1.458**
(vs Shia participant in cross-sect target)	(0.101)	(0.105)	(0.647)
Experimental variables		matched	adjusted
Demographic characteristics (main)		matched	adjusted
Demographic characteristics (secondary)			adjusted
Observations	391	192	338

Note: cluster-robust standard errors by participant *p<0.1; **p<0.05; ***p<0.01

Table B.9: Length of relationship with contact, excluding family (years)

	Dependent variable: length of relationship with contact, excluding family (years)			
	OLS raw	*OLS* matched	*OLS* fixed effects	*OLS* adjusted
	(1)	(2)	(3)	(4)
Shia search contacts	-3.245***	-3.512***	-2.926***	-3.244***
(vs Shia daily contacts)	(0.640)	(0.827)	(0.491)	(0.671)
Sunni search contacts	-1.569**	-1.727***	-1.412*	-1.687**
(vs Sunni daily contacts)	(0.501)	(0.247)	(0.562)	(0.538)
Sunni daily contacts	-2.217***	-1.611*		-1.205
(vs Shia daily contacts)	(0.643)	(0.978)		(0.839)
Sunni search contacts	-0.540	0.174		0.352
(vs Shia search contacts)	(0.647)	(0.505)		(0.797)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	1,037	1,037	1,023	896

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.10: Contact with family members

	Dependent variable: contact is family (binary)			
	OLS raw	*OLS* matched	*OLS* fixed effects	*logit* adjusted
	(1)	(2)	(3)	(4)
Shia search contacts	0.109***	-0.090**	0.129***	0.939**
(vs Shia daily contacts)	(0.030)	(0.044)	(0.026)	(0.341)
Sunni search contacts	0.118**	0.130***	0.169***	0.963***
(vs Sunni daily contacts)	(0.037)	(0.019)	(0.031)	(0.263)
Sunni daily contacts	0.071**	-0.007		0.927**
(vs Shia daily contacts)	(0.028)	(0.043)		(0.406)
Sunni search contacts	0.080**	0.212***		0.951**
(vs Shia search contacts)	(0.041)	(0.023)		(0.308)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	1,392	1,392	1,375	1,190

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.11: Contact's age

	Dependent variable: contact age		
	OLS raw	*OLS* matched	*OLS* adjusted
	(1)	(2)	(3)
Shia participant in cross-sect target	2.123	3.888	1.257
(vs Shia participant in co-sect target)	(1.316)	(2.574)	(1.547)
Sunni participant in cross-sect target	-1.973	-1.449	-1.447
(vs Sunni participant in co-sect target)	(1.497)	(1.651)	(1.532)
Sunni participant in co-sect target	3.873**	4.506**	4.972**
(vs Shia participant in co-sect target)	(1.503)	(1.991)	(1.831)
Sunni participant in cross-sect target	-0.223	-0.831	2.268
(vs Shia participant in cross-sect target)	(1.310)	(2.321)	(1.780)
Experimental variables		matched	adjusted
Demographic characteristics (main)		matched	adjusted
Demographic characteristics (secondary)			adjusted
Observations	488	264	422

Note: cluster-robust standard errors by participant *p<0.1; **p<0.05; ***p<0.01

Table B.12: Distance from contact's home to target (km)

	Dependent variable: distance from contact's home to target (km)		
	OLS raw	*OLS* matched	*OLS* adjusted
	(1)	(2)	(3)
Shia participant in cross-sect target (vs Shia participant in co-sect target)	-0.149 (0.681)	-3.498** (1.339)	-0.226 (0.715)
Sunni participant in cross-sect target (vs Sunni participant in co-sect target)	0.906* (0.521)	0.580 (0.566)	0.247 (0.603)
Sunni participant in co-sect target (vs Shia participant in co-sect target)	-2.106*** (0.540)	-3.210*** (0.688)	-0.578 (0.735)
Sunni participant in cross-sect target (vs Shia participant in cross-sect target)	-1.051 (0.666)	0.867 (1.281)	-0.105 (0.790)
Experimental variables		matched	adjusted
Demographic characteristics (main)		matched	adjusted
Demographic characteristics (secondary)			adjusted
Observations	456	237	394

Note: cluster-robust standard errors by participant *p<0.1; **p<0.05; ***p<0.01

Table B.13: Participant had repeated contact

	Dependent variable: participant had repeat call/text with contact with more than 1-minute gap			
	OLS raw	*OLS* matched	*OLS* fixed effects	*logit* adjusted
	(1)	(2)	(3)	(4)
Shia participant calling family	0.069	0.019	-0.099	-0.099
(vs Shia participant calling friend)	(0.082)	(0.094)	(0.137)	(0.601)
Sunni participant calling family	0.128*	0.103**	0.198*	0.845**
(vs Sunni participant calling friend)	(0.065)	(0.039)	(0.102)	(0.388)
Sunni participant calling family	0.039	0.035	-0.302	0.865
(vs Shia participant calling family)	(0.100)	(0.103)	(0.524)	(0.663)
Sunni participant calling friend	-0.020	-0.049	-0.599	-0.079
(vs Shia participant calling friend)	(0.045)	(0.039)	(0.506)	(0.380)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	453	453	215	353

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.14: Logit-scale differences in proportion incorrect, among participants who provided an answer.

	Dependent variable: answer incorrect	
	logit	*logit*
	Address	Gatekeeper
	(1)	(2)
Shia participant in Sunni target	-1.802**	-3.793**
(vs Shia participant in Shia target)	(0.731)	(1.166)
Sunni participant in Sunni target	-1.787***	-1.674**
(vs Sunni participant in Shia target)	(0.536)	(0.732)
Sunni participant in Sunni target	-0.052	-0.546
(vs Shia participant in Sunni target)	(0.810)	(0.826)
Sunni participant in Shia target	-0.067	-2.664**
(vs Shia participant in Shia target)	(0.711)	(1.056)
Experimental variables	adjusted	adjusted
Demographic characteristics (main)	adjusted	adjusted
Demographic characteristics (secondary)	adjusted	adjusted
Observations	181	146

Note: results among participants with completed answer. *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors by communities of players with shared contacts.

Table B.15: Summary statistics for confidence in answer

		Sunni participants		Shia participants		χ^2 test
Dependent variable	Confidence measure	Prop. correct	Count	Prop. correct	Count	p-value
council address	completed answer	0.65	125	0.55	82	0.20
council address	described as easy	0.71	51	0.73	44	1.00
council address	claimed answer correct	0.71	73	0.68	37	0.86
gatekeeper name	completed answer	0.42	100	0.21	68	0.01
gatekeeper name	described as easy	0.58	33	0.27	15	0.09
gatekeeper name	claimed answer correct	0.49	55	0.29	24	0.16

Table B.16: Reported assistance

	Dependent variable: participant reported assistance from contact			
	OLS raw	*OLS* matched	*OLS* fixed effects	*OLS* adjusted
	(1)	(2)	(3)	(4)
Participant calling contact 5km from target (vs 9km from target)	0.106***	0.228***	0.144***	0.092**
	(0.030)	(0.030)	(0.037)	(0.045)
Participant calling university-educated contact (vs high-school educated contact)	-0.047	0.108*	-0.063	0.068
	(0.073)	(0.057)	(0.077)	(0.101)
Shia participant calling cross-sect contact (vs co-sect contact)	-0.060	0.176*	-0.059	0.121
	(0.081)	(0.094)	(0.103)	(0.120)
Sunni participant calling cross-sect contact (vs co-sect contact)	0.036	0.024	0.004	0.014
	(0.096)	(0.059)	(0.112)	(0.186)
Shia participant calling 35-year-old contact (vs 25-year-old contact)	0.021	-0.065*	0.011	-0.013
	(0.035)	(0.039)	(0.041)	(0.043)
Sunni participant calling 35-year-old contact (vs 25-year-old contact)	0.082**	0.109***	0.095**	0.133**
	(0.041)	(0.023)	(0.042)	(0.064)
Shia participant receiving call-back (vs no recontact)	0.311***	0.307***	0.376***	0.234**
	(0.076)	(0.089)	(0.089)	(0.091)
Sunni participant receiving call-back (vs no recontact)	0.159	0.113*	0.182	0.269*
	(0.114)	(0.059)	(0.116)	(0.162)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	284	284	249	284

Note: cluster-robust standard errors by participant for all except participant fixed-effects model

*p<0.1; **p<0.05; ***p<0.01

Table B.17: Verified assistance

	Dependent variable: participant reported assistance from contact and participant had at least partially correct answer			
	OLS raw	*OLS* matched	*OLS* fixed effects	*OLS* adjusted
	(1)	(2)	(3)	(4)
Participant calling contact 5km from target (vs 9km from target)	0.116***	0.189***	0.145***	0.072*
	(0.030)	(0.032)	(0.039)	(0.038)
Participant calling university-educated contact (vs high-school educated contact)	-0.002	0.037	-0.022	0.030
	(0.057)	(0.045)	(0.062)	(0.084)
Shia participant calling cross-sect contact (vs co-sect contact)	-0.039	0.329**	-0.050	0.100
	(0.068)	(0.107)	(0.089)	(0.100)
Sunni participant calling cross-sect contact (vs co-sect contact)	0.020	0.008	-0.001	0.002
	(0.094)	(0.056)	(0.108)	(0.155)
Shia participant calling 35-year-old contact (vs 25-year-old contact)	0.014	-0.077*	-0.008	-0.004
	(0.026)	(0.040)	(0.029)	(0.036)
Sunni participant calling 35-year-old contact (vs 25-year-old contact)	0.026	0.049**	0.038	0.120**
	(0.034)	(0.018)	(0.037)	(0.054)
Shia participant receiving call-back (vs no recontact)	0.242***	0.281***	0.290***	0.154**
	(0.063)	(0.085)	(0.078)	(0.076)
Sunni participant receiving call-back (vs no recontact)	0.085	0.044	0.104	0.168
	(0.109)	(0.054)	(0.115)	(0.135)
Experimental variables		matched	fixed effects	adjusted
Demographic characteristics (main)		matched	fixed effects	adjusted
Demographic characteristics (secondary)			fixed effects	adjusted
Observations	284	284	249	284

Note: cluster-robust standard errors by participant for all except participant fixed-effects model *p<0.1; **p<0.05; ***p<0.01

Appendix C

Supporting materials for Paper 3

This appendix discusses various practical issues that arise in the estimation of the speaker-affect model.

Numerical Issues

HMMs are particularly vulnerable to numerical underflow in the calculation of forward, backward, and total probabilities; for a discussion, see Zucchini and MacDonald (2009).

Because we work with a large number of features, the M step involves fitting a high-dimensional Gaussian distribution for each sound/expression. When a sound/expression is relatively rare, this can occasionally lead to sample covariance matrices that are non-invertible. We address this by ridge-like regularization, or adding a small value to the diagonal of the sample covariance; the value of this regularization parameter is determined by cross-validation.

Like Gaussian mixture models, HMMs are sometimes subject to the "collapse" of smaller components. For example, a sound/expression distribution may move directly over a single data point, and its variance may go to zero; this results in a contribution to the likelihood that is infinite for one data point and zero for all others. We address this problem by detecting collapse by setting a

tolerance parameter for the determinant of the covariance matrix (roughly the "volume" of the state distribution) and randomly resetting the state to a randomly chosen position with a large, spherical covariance matrix. This requires standardizing all features before estimation, so that "volume" and the value of the tolerance parameter can be defined in a meaningful way.

Convergence

The likelihood functions of most HMMs can suffer from degeneracy, and the EM algorithm is not guaranteed to converge to a global maximum. Multiple EM chains can be run for each combination of model parameters, where the chain that produces the highest in-sample likelihood should be used for all downstream analysis; the variation in in-sample likelihood between chains may be used as a diagnostic.

When inference is conducted via Bayesian bootstrap, estimation for a single bootstrap reweighting can be treated as a mechanical procedure; because estimation uncertainty captured by a Dirichlet prior over each observation's inclusion in the training dataset. In this case, nonconvergence is merely a feature of the procedure that results in greater instability across bootstrap iterations.

Missing data

Features are frequently subject to partial missingness. For example, the fundamental frequency is notoriously difficult to estimate under certain conditions. Different tracking algorithms often produce widely varying results, and gender-based heuristics are generally needed to rule out implausible estimates. When this occurs, we treat the fundamental frequency as missing data; however, other features may still be available for the affected frames. Similarly, visual features may be obscured when a speaker turns away from the camera. We assume that missingness is at random, conditional on the remaining observed features. To calculate probabilities for these partially

missing frames, we integrate over all possible values that the missing data could have taken on.

Model selection

To select the regularization parameter and the number of sounds in an emotion, two standard approaches are considered. Both proceed by dividing the labeled data into a training set and a validation set, then selecting parameters that achieve optimal performance on the validation set in some sense. Multiple validation sets may be pooled as well, including though V-fold cross-validation, by taking the (weighted, if fold sizes are not perfectly balanced) average of separate performance metrics. The first approach selects parameters that optimize the total log-likelihood of the validation set(s) (van der Laan et al., 2004). This procedure possesses the "oracle" property in that it asymptotically selects the closest approximation, in terms of the Kullback–Leibler divergence, to the true data-generating process among the candidate models considered. In applications, this approach almost always selects the most complex model, i.e. the model with the largest number of states and lowest amount of regularization, most likely because even the most complex model considered is far less complex than actual human speech. The second approach is to select the model that achieves the best classification performance, e.g. as measured by the F_1 metric for binary classification, on the validation set. Across a range of applications, I find that returns to classification performance are generally declining in model complexity, and this approach to model selection tends to result in somewhat less complex models.

Bibliography

Alesina, A. and W. Easterly (1999). Public goods and ethnic divisions. *Quarterly Journal of Economics 114*(4), 1243–1284.

Aronow, P. M. and C. Samii (n.d.). Estimating average causal effects under interference between units. http://arxiv.org/pdf/1305.6156v1.pdf.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics 23*, 177–200.

Baker, III, J. A., L. H. Hamilton, L. S. Eagleburger, V. E. Jordan, Jr., E. Meese, III, S. D. O'Connor, L. E. Panetta, W. J. Perry, C. S. Robb, and A. K. Simpson (2006, December). The iraq study group report. report, U.S. Institute of Peace.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*, 363–370.

Banerjee, A., E. Duflo, and N. Qian (2012). On the road: Access to transportation infrastructure and economic growth in china. Technical Report 17897. http://www.nber.org/papers/w17897.

Banerjee, A., L. Iyer, and R. Somanathan (2005). History, social divisions, and public goods in rural india. *Journal of the European Economic Association 3*(2–3), 639–647.

Barnhardt, S., E. Field, and R. Pande (2017). Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india. *American Economic Journal: Applied Economics 9*(1), 1–32.

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics 122*, 778–805.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (n.d.). Roads, railroads and decentralization of chinese cities.

Benoit, K., M. Laver, and S. Mikhaylov (2009). Treating words as data with error: Uncertainty in text statements of policy positions. *American Journal of Political Science 53*(2), 495–513.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological 35*(7), 677–693.

Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in china. *American Sociological Review 62*(3), 366–385.

Black, R. C., S. A. Treul, T. R. Johnson, and J. Goldman (2011). Emotions, oral arguments, and supreme court decision making. *The Journal of Politics 73*(2), 572–581.

Blair, G. (2016). *On the Geography of Assets and Citizens: How Proximity to Oil Production Shapes Political Order*. Ph. D. thesis, Princeton University.

Bollobás, B. (1998). *Modern Graph Theory*. New York: Springer.

Bowers, J., M. Fredrickson, and C. Panagopoulos (2013). Reasoning about interference between units: a general framework. *Political Analysis 21*, 97–124.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics 30*(1), 89–99.

Briggs, R. C. (2012). Electrifying the base? aid and incumbent advantage in ghana. *The Journal of Modern African Studies 50*, 603–624.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. P. i Miquel (2015). The value of democracy: Evidence from road building in kenya. *American Economic Review 105*, 1817–1851.

Carmody, P. (2009). Cruciform sovereignty, matrix governance and the scramble for africa's oil: Insights from chad and sudan. *Political Geography 28*, 353–361.

Casaburi, L., R. Glennerster, and T. Suri (2013). Rural roads and intermediated trade: Regression discontinuity evidence from sierra leone. Technical report.

Caughey, D. (2012). *Congress, Public Opinion, and Representation in the One-Party South, 1930s–1960s*. Ph. D. thesis, University of California, Berkeley.

Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity? evidence from the rural interstate highway system. *Regional Science and Urban Economics 30*, 457–490.

Christia, F., E. Dekeyser, and D. Knox (n.d.). Gauging shia public opinion: A survey of iranian and iraqi religious pilgrims.

Christia, F. and D. Knox (n.d.). Geographic segregation in baghdad: Detecting out-group aversion in walking routes.

Christia, F., D. Knox, and J. Al-Rikabi (n.d.). Networks of sectarianism: Experimental evidence on access to services in baghdad.

Clark, T. S. and B. Lauderdale (2010). Locating supreme court opinions in doctrine space. *American Journal of Political Science 54*(4), 871–890.

Cohen, J. P. and C. J. M. Paul (2004). Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *Review of Economics and Statistics 86*, 551–560.

Converse, P. E. (1962). Information flow and the stability of partisan attitudes. *Public Opinion Quarterly 26*(4), 578–599.

Cornfield, J. (1978). Randomization by group: a formal analysis. *American Journal of Epidemiology 108*(2), 100–102.

Cranmer, S. J. and B. A. Desmarais (2011). Inferential network analysis with exponential random graph models. *Political Analysis 19*(1), 66–86.

Cranmer, S. J., P. Leifeld, S. D. McClurg, and M. Rolfe (2016). Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science Forthcoming*, 416–424.

Damluji, M. (2010). 'securing democracy in iraq': Sectarian politics and segregation in baghdad, 2003–2007. *Traditional Dwellings and Settlements Review 21*(2), 71–87.

Del Bo, C. F. and M. Florio (2012). Infrastructure and growth in a spatial framework: Evidence from the eu regions. *European Planning Studies 20*, 1393–1414.

Dell, M. (2015). Trafficking networks and the mexican drug war. *American Economic Review 105*(6), 1738–1779.

Dellaert, F., T. Polzin, and A. Waibel (1996). Recognizing emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, Volume 3, pp. 1970–1973. IEEE.

Dietrich, B. J., R. D. Enos, and M. Sen (2016). Emotional arousal predicts voting on the us supreme court. Technical report, Technical Report.

Dodds, P. S., R. Muhamad, and D. J. Watts (2003a). An experimental study of search in global social networks. *Science 301*, 827–829.

Dodds, P. S., R. Muhamad, and D. J. Watts (2003b). An experimental study of search in global social networks. *Science 301*, 827–829.

Donaldson, D. (forthcoming). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*.

Eckstein, H. H. (1975). Case studies and theory in political science. In F. J. Greenstein and N. W. Polsby (Eds.), *Handbook of Political Science*. Reading, MA: Addison-Wesley.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association 72*(359), 557–565.

Eggers, A. C. and A. Spirling (2014). Ministerial responsiveness in westminster systems: Institutional choices and house of commons debate, 1832–1915. *American Journal of Political Science 58*(4), 873–887.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion 6*, 169–200.

Ekman, P. (1999). Basic emotions. In T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion*. Chicester, England: Wiley.

El Ayadi, M., M. S. Kamel, and F. Karray (2011a). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition 44*(3), 572–587.

El Ayadi, M., M. S. Kamel, and F. Karray (2011b). Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recognition 44*, 572–587.

Elkins, Z. and B. Simmons (2005). On waves, clusters and diffusion: A conceptual framework. *American Academy of Political and Social Science 598*, 33–51.

Eubank, N. (2016). Social networks and the political salience of ethnicity.

Fogel, R. (1962). A quantitative approach to the study of railroads in american economic growth: A report of some preliminary findings. *Journal of Economic History 22*(2), 163–197.

Fogel, R. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins Press.

Fosgerau, M., E. Frejinger, and A. Karlstrom (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological 56*, 70–80.

Gerring, J. (2007). Is there a (viable) crucial-case method? Volume 40, pp. 231–253.

Goplerud, M., D. Knox, and C. Lucas (2016). The rhetoric of parliamentary debate. *Working Paper*.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Granovetter, M. S. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives 19*(1), 33–50.

Gremlin (1976). *Blockade*. San Diego: Gremlin Industries. Arcade game.

Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(4), 709–725.

Haines, M. R. (n.d.). Historical, demographic, economic, and social data: The united states, 1790–2002 (dataset). Technical Report 2896. http://www.icpsr.umich.edu/icpsrweb/RCMD/studies/2896.

Hainmueller, J., D. J. Hopkins, and T. Yamamoto (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis 22*(1), 1–30.

Hartman, A. C. and B. S. Morse (n.d.). Wartime violence, empathy, and altruism: Evidence from the ivoirian refugee crisis in liberia.

Hopkins, D. J. and G. King (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science 54*(1), 229–247.

Huckfeldt, R. and J. Sprague (1987). Networks in context: The social flow of political information. *American Political Science Review 81*(4), 1197–1216.

Huckfeldt, R. and J. Sprague (1995). *Citizens, politics and social communication: Information and influence in an election campaign*. Cambridge: Cambridge University Press.

International Medical Corps (2007). Iraqis on the move: Sectarian displacement in baghdad. Technical report.

Iraq Body Count (2016). Documented civilian deaths from violence. Database.

Kaufman, A., P. Kraft, and M. Sen (ND). Machine learning and supreme court forecasting: Improving on existing approaches.

Killworth, P. D. and H. R. Bernard (1978). The reversal small-world experiment. *Social Networks 1*, 159–192.

Kim, S., F. Valente, M. Filippone, and A. Vinciarelli (2012). Predicting the conflict level in television political debates: an approach based on crowdsourcing, nonverbal communication and gaussian processes. In *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 793–796.

King, G. (1989). Event count models for international relations: Generalizations and applications. *International Studies Quarterly 33*(2), 123–147.

Kleinberg, J. (2000). The small-world phenomenon: an algorithmic perspecive. In *Proceedings of the thirty-second Annual ACM Symposium on the Theory of Computing*. Association for Computing Machinery.

Kranton, R. E. and D. F. Minehart (2001). A theory of buyer-seller networks. *American Economic Review 91*(3), 485–508.

Lauderdale, B. E. and T. S. Clark (2014). Scaling politically meaningful dimensions using texts and votes. *American Journal of Political Science 58*(3), 754–771.

Laver, M., K. Benoit, and J. Garry (2003). Extracting policy positions from political texts using words as data. *American Political Science Review 97*(02), 311–331.

Lawler, G. F. (1999). Loop-erased random walk. In M. Bramson and R. Durrett (Eds.), *Perplexing Problems in Probability*. Boston: Birkhäuser.

Lee, F. E. (2000). Senate representation and coalition building in distributive politics. *American Political Science Review 94*, 59–72.

Lee, L.-F. (1992). On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory 8*, 518–552.

Levine, S. S., E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark (2014). Ethnic diversity deflates price bubbles. *Proceedings of the National Academy of Sciences 111*(52), 1524–18529.

Lucas, C., R. A. Nielsen, M. E. Roberts, B. M. Stewart, A. Storer, and D. Tingley (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*.

Marin, J.-M., P. Pudlo, C. Robert, and R. Ryder (2012). Approximate bayesian computational methods. *Statistics and Computing 22*(6), 1167–1180.

Michaels, G. (2008). The effect of trade on the demand for skill—evidence from the interstate highway system. *Review of Economics and Statistics 90*, 683–701.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of Public Economics 89*(11–12), 2325–2368.

Milgram, S. (1967). The small-world problem. *Psychology Today 1*(1), 61–67.

Mironova, V. and S. Whitt (2014). Ethnicity and altruism after violence: The contact hypothesis in kosovo. *Journal of Experimental Political Science 1*, 170–180.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2009). Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, mpn018.

Morrison, A. R. (1993). Violence or economics: What drives internal migration in guatemala? *Economic Development and Cultural Change, 41*(4), 817–831.

Murray, I. R. and J. L. Arnott (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America 93*(2), 1097–1108.

Nall, C. (2013). The road to division: How interstate highways caused geographic polarization.

Nall, C. (2015). The political consequences of spatial policies: How interstate highways facilitated geographic polarization. *Journal of Politics 77*, 394–406.

National Interregional Highway Committee, U. (1944). Interregional highways: Message from the president of the united states, transmitting a report of the national interregional highway committee, outlining and recommending a national system of interregional highways. Technical report, U.S. Government Printing Office.

Nogueiras, A., A. Moreno, A. Bonafonte, and J. B. Mariño (2001). Speech emotion recognition using hidden markov models. In *INTERSPEECH*, pp. 2679–2682.

Nwe, T. L., S. W. Foo, and L. C. De Silva (2003). Speech emotion recognition using hidden markov models. *Speech communication 41*(4), 603–623.

OpenStreetMap (2016). Openstreetmap data.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Proksch, S.-O. and J. B. Slapin (2010). Position taking in european parliament speeches. *British Journal of Political Science 40*(03), 587–611.

Proksch, S.-O. and J. B. Slapin (2012). Institutional foundations of legislative speech. *American Journal of Political Science 56*(3), 520–537.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science 54*(1), 209–228.

Rauch, J. E. and A. Casella (2001). *Networks and Markets*. New York: Russell Sage Foundation.

Rephann, T. and A. Isserman (1994). New highways as economic development tools: An evaluation using quasi-experimental matching methods. *Regional Science and Urban Economics 24*, 723–751.

Roberts, B. and D. P. Kroese (2007). Estimating the number of s–t paths in a graph. *Journal of Graph Algorithms and Applications 11*(1), 195–214.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science 58*(4), 1064–1082.

Rogers, T. and M. Aida (2013). Vote self-prediction hardly predicts who will vote, and is (misleadingly) unbiased. *American Politics Research 42*(3), 503–528.

Rosenbaum, P. R. (2002). *Observational Studies*. New York: Springer-Verlag.

Rosenfeld, B., K. Imai, and J. Shapiro (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science 60*(3), 783–802.

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics 47*(4), 1213—-1234.

Samii, C. (2013). Perils or promise of ethnic integration? evidence from a hard case in burundi. *American Political Science Review 107*(3), 558–573.

Schelling, T. C. (1969). Models of segregation. *American Economic Review 59*(2), 488—-493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*(2), 143–186.

Scherer, K. R. and J. S. Oshinsky (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and emotion 1*(4), 331–346.

Schnettler, S. (2009). A small world on feet of clay? a comparison of empirical small-world studies against best-practice criteria. *Social Networks 31*, 179–189.

Schub, R. (2015). Are you certain? leaders, overprecision, and war. Technical report, Working Paper (available at http://scholar. harvard. edu/schub/research).

Schuller, B., G. Rigoll, and M. Lang (2003). Hidden markov model-based speech emotion recognition. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, Volume 1, pp. I–401. IEEE.

Sigelman, L. and C. Whissell (2002a). " the great communicator" and" the great talker" on the radio: Projecting presidential personas. *Presidential Studies Quarterly*, 137–146.

Sigelman, L. and C. Whissell (2002b). Projecting presidential personas on the radio: An addendum on the bushes. *Presidential Studies Quarterly 32*(3), 572–576.

Stewart, B. M. and Y. M. Zhukov (2009). Use of force and civil–military relations in russia: an automated content analysis. *Small Wars & Insurgencies 20*(2), 319–343.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy 64*(5), 416–424.

Tripp, C. (2000). *A History of Iraq*. Cambridge: Cambridge University Press.

U.N. High Commissioner for Refugees (2016, March). Stabilizing the situation of refugees and migrants in europe. Proposals to the meeting of eu heads of state or government and turkey, United Nations.

U.S. Bureau of Public Roads (1955). General location of national system of interstate highways, including all additional routes at urban areas designated in september 1955. Technical report, U.S. Government Printing Office.

Valiant, L. G. (1979). The complexity of enumeration and reliability problems. *Siam Journal of Computing 8*(3), 410–421.

van der Laan, M. J., S. Dudoit, and S. Keles (2004). Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology 3*(1), 1–23.

van der Laan, M. J., S. Dudoit, S. Keles, et al. (2004). Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology 3*(1), 1036.

Ververidis, i. and C. Kotropoulos (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication 48*, 1162–1181.

Voigtlaender, N. and H.-J. Voth (2014). Highway to hitler. Technical Report 20150. http://www.nber.org/papers/w20150.

Wilson, D. B. (1996). Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing*. Association for Computing Machinery.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.

Yu, B., S. Kaufmann, and D. Diermeier (2008). Classifying party affiliation from political speech. *Journal of Information Technology & Politics 5*(1), 33–48.

Zaller, J. (1989). Bringing converse back in: Modeling information flow in political campaigns. *Political Analysis 1*(1), 181–234.

Zhang, J. (2004a). A dynamic model of residential segregation. *Journal of Mathematical Sociology 28*(3), 147–170.

Zhang, J. (2004b). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization 54*, 533–550.

Zucchini, W. and I. MacDonald (2009). *Hidden Markov Models for Time Series*. Boca Raton, FL: CRC Press.