New Methods for New Data: Developing Techniques for Analysing Alternative Social Science Data

Thesis Submitted To

University of Oxford

in partial fulfilment of the requirements
of the award of the degree

DPhil in Sociology

Jonathan A. Mellon
Nuffield College
Department of Sociology
University of Oxford

Trinity 2013

Thesis Submitted To

University of Oxford

in partial fulfilment of the requirements
of the award of the degree

DPhil in Sociology

by

Jonathan A. Mellon

Nuffield College
Department of Sociology
University of Oxford
Trinity 2013

Abstract: In social science research, methodology matters. Seemingly arcane issues of measurement are not trivial but can have profound influences on how we view the social world. This thesis seeks to fill methodological gaps that have opened up as new research techniques become available. This thesis consists of five chapters covering three projects. The first project analyses whether Google Trends data can be used as proxies for issue salience. A methodology for validating Google Trends data against survey data is developed, with searches validated in the United States, the United Kingdom and Spain. The dangers of not testing the validity and reliability of Internet derived data is discussed.

The second project examines the translation equivalence of survey instruments into different languages by using corpus linguistic on Internet corpora in 10 languages. Using these corpora, dissimilarity measures are constructed, allowing clusters of meaning to be identified quantitatively and examined qualitatively. This method is illustrated using a postmaterialism item from the World Values Survey. This item is found to suffer from important translation problems, as the translations of the word *ideas* have clearly non-equivalent meanings across languages. These differences predict the item's correlations with other variables within the World Values Survey.

The third project looks at the role of information processing within the US State Department's foreign policy analysis using social network analysis (SNA). First, the predictors of information flow are identified, showing diplomats focus most strongly on trade. Next, a new type of relational event model is developed to account for specific communication features communication within this bureaucracy: a relational event conversation history actor initiated network model (RECHAIN). Using this model, we find the importance of trade is set by the Bureaus, with Embassies adopting this agenda against their preferences. Finally, we examine the evidence behind different mechanisms of communication agenda setting.

Acknowledgements

I would like to thank my supervisor Steve Fisher for his careful and insightful feedback throughout my thesis and my examiners Chris Wlezien and Geoffrey Evans who gave many useful suggestions for taking this work further. I am also thankful for the invaluable feedback I received from Tom Snijders, Nan Dirk De Graaf and Catherine de Vries during my transfer and confirmation of status.

Countless conversations with friends and colleagues at Nuffield College and the Oxford University Department of Sociology have helped to shape the ideas and direction of this thesis. The many discussions with my co-author James Hollway have been particularly useful in shaping this work. Red Owl Analytics gave me a great opportunity to learn about relational event modelling and Josh Lospinoso and Per Block were infinitely patient in answering hundreds of my questions about social network analysis.

I would like to thank the Economic and Social Research Council for funding my research and the Department of Sociology for nominating me for this studentship.

My parents have always been an invaluable source of support and advice.

Finally, I am hugely grateful to my wonderful fiancée Greer, whose constant support (and proof reading) helped me to finally finish this thesis.

Contents

1 Internet search indices and Issue Salience **8**
- 1.1 Previous Work . . . 9
  - 1.1.1 Google Trends in Political Science . . . 10
  - 1.1.2 The Different Uses of MIP . . . 12
- 1.2 Data . . . 14
  - 1.2.1 Google Internet Search Data . . . 14
  - 1.2.2 Representativeness of Internet Users . . . 15
  - 1.2.3 Gallup MIP Data . . . 16
  - 1.2.4 Differences between MIP and Search Data . . . 16
- 1.3 Methodology . . . 17
  - 1.3.1 Face Validity . . . 17
  - 1.3.2 Content Validity . . . 18
  - 1.3.3 Criterion Validity . . . 18
  - 1.3.4 Time series complications . . . 20
  - 1.3.5 Prediction of weekly issue salience series . . . 22
  - 1.3.6 Out-of-sample testing . . . 23
- 1.4 Results . . . 24
  - 1.4.1 Face Validity . . . 24
  - 1.4.2 Content Validity . . . 24
  - 1.4.3 Time series complications . . . 25
  - 1.4.4 Criterion Validity . . . 26
  - 1.4.5 Predicting weekly issue salience series . . . 27
  - 1.4.6 Out-of-sample reliability tests . . . 30
- 1.5 Comparison of Internet Search Measures with MIP Measures . . . 32
- 1.6 Conclusion . . . 34

2 Where and When Can We use Google Trends to Measure Issue Salience **38**
- 2.1 Introduction . . . 38
- 2.2 Previous Approaches to Using Google Trends Data . . . 41
- 2.3 Validating Google Data Cross-Nationally . . . 42
- 2.4 Results . . . 45
  - 2.4.1 For which issues do Google Trends provides good proxies? . . . 48
- 2.5 Why do *these* search terms work? . . . 52

3 Validating Survey Translation Using Corpus Linguistics **55**
- 3.1 Introduction . . . 55
- 3.2 Meaning and Equivalence . . . 56
  - 3.2.1 Checking for the Confounding Error Component's Equivalence . . . 58
- 3.3 Previous Approaches to Assessing Equivalence . . . 59

3.4 Assessing Meaning Using Corpus Linguistics . . . 61
3.5 Comparing Chinese and English Associations for a Postmaterialism Item . . 63
3.5.1 Translate and Clean . . . 64
3.5.2 Differences in the Meaning of *ideas* Between Chinese and English . . 65
3.6 Comparing the Meaning of *ideas* Translations Across 10 Languages . . . 69
3.6.1 Construct Dissimilarities Matrix . . . 69
3.6.2 Cluster Analysis . . . 70
3.6.3 Summary of Linguistic Findings for the Meaning of *ideas* in the World Values Survey Postmaterialism Question . . . 74
3.6.4 Evaluation of Other Words in the Postmaterialism Question . . . 74
3.6.5 Summary of Method . . . 74
3.7 Survey Analysis of Associations Between the Question and Other Concepts 76
3.7.1 Hypotheses . . . 77
3.7.2 Method . . . 78
3.7.3 Results . . . 79
3.7.4 Discussion of Results . . . 81
3.8 Discussion . . . 81
3.9 Conclusion . . . 83

4 What structures the flow of information between different Embassies in the US State Department? 85
4.1 Introduction . . . 85
4.2 Background . . . 87
4.2.1 State Department . . . 87
4.2.2 Information processing . . . 90
4.3 Potential factors that could induce communication flows . . . 93
4.3.1 Geographic variables . . . 93
4.3.2 International trade . . . 94
4.3.3 Militarized interstate disputes . . . 95
4.3.4 Arms trades . . . 96
4.3.5 International Organizations . . . 96
4.3.6 Alliances . . . 97
4.3.7 Migration Flows . . . 98
4.4 Communication in the State Department . . . 99
4.4.1 The Central Foreign Policy Files . . . 99
4.4.2 The structure of a diplomatic cable . . . 99
4.4.3 Actors (A, B, D, E and F) . . . 101
4.4.4 Operationalizing the communication structure . . . 101
4.5 Multiple Regression Quadratic Assignment Procedure (MRQAP) . . . 104
4.5.1 Further comparisons . . . 106
4.6 Results . . . 106
4.6.1 Predictors of inter-Embassy communication . . . 107
4.6.2 Qualitative triangulation . . . 110
4.6.3 Trade . . . 111
4.6.4 Shared IGO memberships . . . 112
4.6.5 Comparison of Bureau and Embassy agendas . . . 113
4.6.6 Longitudinal model . . . 114
4.7 Discussion . . . 116

5 Speaking Diplomatically: Examining Agenda Setting Using Relational Event Modelling **119**
5.1 Introduction . . . 119
5.2 Agendas in the State Department . . . 122
5.2.1 Bottom up . . . 122
5.2.2 Defining the agenda . . . 125
5.3 Communication in the State Department . . . 128
5.3.1 Senders and Recipients (A, B, C, D, E and F) . . . 129
5.3.2 References (K) . . . 129
5.3.3 Times (C) . . . 131
5.4 The structure of communication . . . 131
5.4.1 How many messages are sent? . . . 131
5.4.2 Conversation chains . . . 132
5.4.3 How persistent are the recipient sets between cables? . . . 134
5.5 Relational Event Models . . . 137
5.5.1 Limitations of current Relational Event Models . . . 139
5.6 The RECHAIN Model . . . 140
5.6.1 Model Specifications . . . 142
5.6.2 Modelling all stages of communication . . . 145
5.6.3 Summary . . . 147
5.7 Mechanisms of communication agenda setting . . . 147
5.7.1 Which actors focus on trade ties? . . . 147
5.7.2 Recipient adoption . . . 149
5.7.3 Recipient inheritance . . . 154
5.7.4 Additional control effects . . . 156
5.7.5 Implications of different mechanisms . . . 157
5.7.6 Modelling approach . . . 158
5.8 Results . . . 158
5.8.1 Preferences for communicating along trade ties . . . 158
5.8.2 Recipient adoption . . . 163
5.8.3 Recipient inheritance . . . 164
5.8.4 Summary . . . 165
5.9 Discussion . . . 166

6 Conclusions **169**
6.1 Links between all the projects . . . 174
6.1.1 Computation . . . 174
6.2 Computational Challenges in RECHAIN . . . 176
6.2.1 Validation and triangulation . . . 178
6.2.2 Factors affecting the adoption of methodological innovations . . . 180
6.3 Links between specific chapters . . . 181
6.3.1 Linking the meaning of words to comparative measurement . . . 181
6.3.2 Agendas . . . 182
6.4 Limitations . . . 182
6.5 Future research . . . 184
6.5.1 A change in methodology . . . 184
6.5.2 Wider validation of surveys . . . 184
6.5.3 Communication models . . . 184
6.5.4 Thermostatic RECHAIN Models . . . 186

6.5.5 Tracing influence in other political institutions 186

A Appendices for Internet Search Indices and Issue Salience 205
1.1 Content Validity Results 205
1.2 Glossary 205

B Appendices for Validating Survey Translation Using Corpus survey translation 208
2.1 Computational Details 208
2.2 Derivation of Relationship between Confounding Error Component and Individual Attributes 209
2.2.1 Error Component Definitions 209
2.2.2 Item Equivalence Conditions 210
2.3 Number of Observations used in Logistic Regressions 212
2.4 Linguistic Analysis of Other Words in postmaterialism Question 212
2.4.1 A Stable Economy 212
2.4.2 The Fight Against Crime 213
2.4.3 Progress Toward a Less Impersonal and More Humane Society . . . 214
2.4.4 Summary of Linguistic Analysis 217
2.5 Human translation of collocations 217

C Appendices for What structures the flow of information between different Embassies in the US State Department? 220
3.1 Software used in this paper 220
3.2 Missing data 220

D Appendices for Speaking Diplomatically: Examining Agenda Setting Using Relational Event Modelling 223
4.1 Timestamp parsing and imputation 223
4.2 Difficulties with extracting references 226
4.2.1 Biases introduced by these heuristics 227
4.3 Modelling citation choice 228
4.3.1 Citation risk sets 228
4.4 Modelling Alter Removal 229
4.5 Possible reasons for fewer recipients over time 230

Introduction

In social science research, methodology matters. Seemingly arcane issues of measurement are not trivial but can have a profound influence on how we view the social world. In a world where Google searches are being viewed as an insight into epidemiology and public opinion, social network analysis is being used to analyse vast quantities of metadata and global surveys are more feasible than ever, there is a need to ensure that the methodological rigour keeps up with the pace of empirical analysis. This thesis is motivated by these challenges. It seeks to fill some of the methodological gaps that have opened up as new research techniques have become available. In the course of this methodological research, I also tackle several important substantive issues: the role of postmaterialism in various countries, the provision of information in the US State Department and the agenda setting power of different parts of this bureaucracy. Thus, the thesis shows the immediate relevance of these methodological innovations to answering critical questions within social science.

Traditional routes of data collection and collation have involved either direct contact with the source of the data through surveys or have been derived from administrative records. In the past two decades, online data resources including collections of text written in many languages, internet search records and digitized historical archives have allowed questions to be asked that would have been impossible to answer using traditional survey methods. Researchers cannot realistically survey everyone on their views of what is important on a weekly basis or record billions of words of natural language in multiple countries on a rolling basis but data derived from the Internet offers the prospect of doing just this.

However, using these data sources requires more than simply applying existing techniques to be used effectively. In considering how best to utilise such datasets, it is essential to consider the specific limitations of each data type and the potential biases it could

introduce.

This is a paper based dissertation on the theme of providing new methods for new data. This differs from the traditional book thesis in that multiple topics are explored within the theme of improving research methodology using non-traditional data sources. The thesis does not follow a single narrative through the chapters. However, this format best allows for multiple contributions to the field of research methodology on emerging data sources.

The thesis consists of five chapters grouped into three projects:

- The first project (chapters 1 and 2) examines the Google Trends service that provides aggregated measurements of the number of searches for a given search term over time.
- The second project (chapter 3) uses the Leeds Collection of Internet Corpora to test the translation quality of questions on comparative surveys.
- The third project uses a recently released digitized archive of 300,000 diplomatic cables available through the National Archives and Records Administration to develop and examine a theory of State Department communications and a new relational event model for analysing communication data.

Chapters 4 and 5 are both part of a co-authored project with James Hollway of Oxford's Department of Politics and International Relations. However, chapter 4 is entirely written by Jonathan Mellon. Chapter 5 was conceived jointly between James Hollway and Jonathan Mellon. The RECHAIN model was designed and implemented by Jonathan Mellon and all the analysis and writing has been undertaken by Jonathan Mellon who is first author on the paper. Although these papers are part of a wider project on foreign policy formulation, the two chapters presented here are both independent pieces of work that are intended to address a specific aspect of foreign policy information provision. Because of this, these chapters and the descriptions of them in the introduction and conclusion are written using the plural pronouns "we" and "us" rather than "I" and "me". James Hollway is not currently planning to use any material from this project in his dissertation.

Structure of the thesis

This section describes the content of each chapter of the thesis in further detail, including the main research questions guiding each chapter.

The first two chapters of this thesis discuss the Google Trends service and the circumstances in which it is valid to use these data for measuring public opinion. Chapter 1 examines whether Google Trends data (tracking aggregate searches over time) can be used as a proxy for public opinion data, particularly issue salience. Chapter 1 addresses three research questions: 1) what procedures can be used to validate Google Trends procedures against existing sources of public opinion data in the United States? 2) Are any Google Trends time series valid proxies for public issue salience in the United States? 3) How valid and reliable are the Google Trends data that are able to be used as proxies? To answer these questions, the chapter outlines a methodology for validating aggregate behavioural data against aggregate measures of public opinion. I find that Google Trends data can be used as a proxy for the salience of several issues. However, many other plausible searches do not show validity, casting doubt on some research using these data without validation. Chapter 1 has been published in the Journal of Elections, Public Opinion and Parties.

Chapter 2 extends this methodology to two more countries: Spain and the United Kingdom. It focuses on two research questions. 1) Can Google Trends be used to measure issue salience in country contexts outside of the United States? 2) Are the same search terms valid for measuring issue salience in different country and language contexts? The case of the United Kingdom is particularly important given that chapter 1 evaluated data on the United States and English is the predominant language in both countries. As a result, this comparison provides some evidence towards answering the question of whether Google Trends validation is portable to countries with the same primary language. In fact, while there are search terms that are valid for measuring issue salience in the United Kingdom and Spain, these terms are not predictably the same as those that worked in the United States. Chapter 2 has been published in *PS: Political Science and Politics*. It is therefore written in American English and according to the APSA guidelines.

Chapter 3 makes use of new sources of natural language data available online to assess survey translation. A large body of work in social science relies on comparative surveys.

The conclusions reached in these studies rest on the quality of these surveys and the way they are translated into other languages. A substantial literature has looked into the problems of survey equivalence in terms of different relationships between the item or scale and the underlying variable across countries. However, there has been less work on detecting possible confounding biases in measurement where different additional variables are correlated with the indicator across different countries.

Due to the vast quantities of text that are now available on the Internet, it is possible to get a picture of language usage across many countries and languages. One such repository is the Leeds Collection of Internet Corpora that covers 10 languages with corpora that include around 100 million words each. These data have many possible uses but this chapter focuses on improving traditional survey analysis.

Chapter 3 asks three research questions. 1) Can corpus linguistics data derived from web corpora be used to assess translation equivalence? 2) Are there translation problems in existing survey questions (drawing on the specific example of the postmaterialism items from the World Values Survey)? 3) Do these non-equivalent translations lead to predictable correlations with other questions on the survey?

The corpora are used to assess the linguistic associations between key words in a question and other concepts. By using a corpus for each language, we can compare whether some languages include associations that are lacking in other languages. If this is the case, then there is the potential for confounding variables to be introduced into the measurement of these variables. I find that a widely used measure of postmaterialism in the World Values Survey has confounding linguistic associations in Polish, Russian, Arabic and Chinese. These linguistic differences predict a different correlation structure with other items on the survey that relate specifically to the linguistic differences discovered using the corpus method. The method in chapter 3 will be released as an R package along with the chapter when it is published as an article.

Chapters 4 and 5 return to the issue of agenda measurement in a different context: the US State Department. The agenda of governments is an area of study in which non-survey methods have been used for a long time. For instance content analysis of political speeches goes back decades (Jennings, & John 2009; Moen 1988; Gilberg et al. 1980). However, this analysis has usually focused on political leaders rather than the bureaucrats who back

them up. This focus persists despite a substantial line of research (see Workman, Jones, & Jochim (2009) for an overview of this literature) that argues that bureaucracies' use of information is a key component of the policy process.

In chapter 4, I make use of a newly released archive of diplomatic cables from the US State Department, covering the years 1973 and 1974, to look at the changing issue agendas of different actors in the State Department. Chapter 4 addresses five research questions. 1) Which issues structure the flow of information to Embassies in the US State Department? 2) Does the relative weighting of these issues differ between the Embassies in the field and the Bureaus based in Washington DC? 3) Is the issue agenda of the State Department stable over time? 4) Is communication structured around multiple issues at a time within the US State Department? 5) Does the content of diplomatic cables sent across a dyad (a link between a pair of actors) match up with the issues that the network models predict are important across that dyad?

I argue that Embassies will tend to distribute information based on who they consider to be relevant to their interests. It is possible to see which aspects of countries' relationships to each other individual Embassies consider most important by examining which other Embassies they transfer information to. Therefore, the hypothesis would be that if an Embassy is focused primarily on security issues, it will transfer information based on the host country's alliances and arms trading. By contrast, if the Embassy is concerned with international trade, then it will tend to focus on communicating with its host country's trading partners. I find that Embassies and Bureaus transfer information on the basis of multiple issues, but they focus most heavily on trade links and joint Inter-Governmental Organisation (IGO) memberships between their host countries. Particularly striking is the similarity in the relative weighting of different issues by Bureaus and Embassies indicating that they share their communication agenda.

Chapter 5 continues the examination of agendas in the State Department. We (the paper is co-authored with James Hollway) ask whether the importance of trade in the structure of information provision is introduced by Bureaus or Embassies.

Chapter 5 is based around three research questions. 1) What features are needed in a social network model that can accurately account for the features of the US State Department's communication network? 2) Do Embassies or Bureaus set the conversation agenda

in terms of introducing trade into the structure of information provision? 3) What mechanisms of communication agenda setting are present within the US State Department's cables?

To answer these questions we turn to relational event models. These models take an individual message as the unit of analysis and empirically fit a utility function that describes how an actor weighs up different aspects of the potential recipient. We argue that existing relational event models are inappropriate for this kind of dataset, as they assume conditional independence of events. However, we show that communications in the State Department are actually structured into ongoing conversations. This means that the decisions of the sender of a message are not independent of those that were made in the previous message of the conversation. We show that communication decisions take the previous message in a conversation as the starting point and then make decisions to change the conversation from that starting point. To account for this structure, we define a new relational event model that we term a relational event conversation history actor initiated network model (RECHAIN). This models decisions as modifying the recipient sets of a previous conversation.

Using this model, we show that only Bureaus directly react to the trade links of potential recipients. By contrast, the Embassies are somewhat less likely to add recipients whose host countries have strong trade links with the Embassy's own host country, compared with those who have weak trade links. Instead the Embassies end up communicating with trade partners due to conversational inertia (when they continue existing conversations started by Bureaus) and by reacting to endogenous communications such as in their tendency to send messages to Embassies that have previously provided information to them.

I conclude by discussing the findings of the thesis and highlighting the substantive and methodological links between the different chapters.

Chapter 1

Internet search indices and Issue Salience

Measures of issue salience are used widely in political science, particularly in agenda-setting research (Soroka 2002);[1] and research on issue voting. However, these studies have often been restricted by the limitations of available data. Almost all research on issue salience in the public uses the "most important problem" question (MIP)[2] or close variants of it. In the United States data for these questions are only available monthly, at irregular intervals and usually only at the national level. In most other countries the situation is worse, with only sporadic survey data obtainable.

Using data from Internet searches to measure issue salience could potentially solve these problems, as weekly search data are available for free from Google between 2004 and 2010 for almost every country, state and city in the world – and for any possible search term. The data from Google are available as indices that represent the number of Google searches that include a given search term in each period (aggregated either daily for a 90 day period or weekly for up to eight years).

Search indices are measures of the frequency of Internet searches that include a given word or set of words, referred to as "search terms". If the search term "jobs" is entered into Google Trends, it will return a search index that measures how many searches for the word "jobs" there were over time (including searches for the word "jobs" and ones for "government

[1] For an overview of the agenda-setting literature see McCombs (2005).

[2] The full question is "what do you think is the most important problem facing this country today?"

jobs", "secretary jobs", etc.) Some of these search indices could plausibly reflect the salience of an issue among the public.[3]

The potential value of search data has started to be recognised by social scientists (Granka 2010; Graefe, & Armstrong 2012; Boulton et al. 2011; Reilly, Richey, & Taylor 2012; Ellis, Swearingen, & Ripberger 2011; Oltra 2011; Reis, & Brownstein 2010; Weeks, & Southwell 2010; Dube, & Kaplan 2012; Kahn, & Kotchen 2010; Pelc 2013) and the media (for instance Weiner (2012); Silver (2012)) to make a variety of claims about public opinion. However, these claims are all based on the assumption that the trends in search data track those in the wider population. This article aims to provide a method by which researchers can test whether the Internet search data they use are valid proxies for the phenomena they are interested in by comparing the search data with existing survey data.

To assess when Internet data can be used, a process is outlined that tests the reliability and validity of each candidate search index as measures of a particular issue's salience. Content validity is tested by looking at whether the Internet searches in each index actually refer to what they are assumed to measure, i.e. the jobs search index is measuring the salience of employment issues and not Steve Jobs (the former CEO of Apple Inc.). Criterion validity is assessed by quantitatively comparing search indices with the existing MIP measures of issue salience. Although this paper focuses on issue salience, the approach could be equally applicable to aggregate measurements of other attitudes such as approval ratings and other Internet data where representativeness is in question, such as Twitter mentions.

This paper finds that four of the seven issues that were salient in the US between 2004 and 2010 (according to Gallup's MIP surveys across this period) are able to be proxied using search indices: immigration, terrorism, fuel prices and the economy. Additionally, the importance of the validation process is shown by a number of plausible looking search indices were not valid proxies of the issues' salience (such as "Iraq War" and "economy").

1.1 Previous Work

This section discusses some of the previous work in political science using Internet search data or the MIP question. Issue salience has been understood in a number of ways.

[3] A glossary of some of the technical terms is provided in Appendix 1.2.

Firstly, it has been seen as the importance an individual attaches to an issue, i.e. the weight placed on an issue when making political decisions such as voting (Wlezien 2005). Secondly, it has been viewed as how prominent an issue is in the mind of an individual, that is, how accessible the topic is when the individual is prompted to make a political decision (Wlezien 2005). Both conceptualisations fit well with search data, in that perceived issue importance is likely to have a strong effect on decisions about information seeking, and an issue must reach a certain level of prominence in a person's mind before they will search for information about it. So far, MIP has been used widely as a measure of issue salience and its properties are well understood (Soroka 2002). Therefore, to assess the validity of an Internet search index with regards to issue salience, it makes sense to corroborate it against the better understood MIP data.

1.1.1 Google Trends in Political Science

Despite being a relatively new data source, many political scientists have already begun to make use of Google Trends as a measure of issue salience or attention (Granka 2010; Graefe, & Armstrong 2012; Boulton et al. 2011; Reilly, Richey, & Taylor 2012; Ellis, Swearingen, & Ripberger 2011; Oltra 2011; Reis, & Brownstein 2010; Weeks, & Southwell 2010; Dube, & Kaplan 2012; Kahn, & Kotchen 2010).[4] These have been used to address problems from constituent reactions to WTO trade disputes (Pelc 2013) to the context of the Occupy movement (Dube, & Kaplan 2012).

However, these studies generally do not address the question of whether the search indices are representative of the public agenda. For instance, in agenda setting studies it is plausible that a small subset of the population who search using a certain term may be more or less responsive to media coverage than the public as a whole. So, findings of agenda setting are entirely consistent with the alternative explanation that a small politically interested minority follow each development of a political story obsessively and react strongly to media coverage, while the vast majority of the population pay no attention, or hear about the story through different mechanisms. A good proxy variable must be shown to closely track the theoretical variable of interest (in this case aggregate issue salience in

[4]There are many more studies that make brief use of search indices to illustrate increased public interest in the author's topic and a large number of working papers in progress.

the public).

A few papers try to address this problem. Ripberger claims to validate Internet search indices on the basis that some closely match the coverage of issues in the *New York Times* (Ripberger 2011). However, correlations with media coverage are insufficient to validate a search index as a measure of an issue's salience to the public: the *New York Times* is not a representative measure of issue salience in the general population (although correlations with media coverage can provide an additional line of evidence through convergent validity).

Two other articles looked at search data's relationship to survey measures of issue salience. A recent research note (Scheitle 2011) briefly presented correlations between MIP and Internet search data to illustrate a call for more research using search data but without examining any methodological concerns regarding time series or content validity. Scharkow and Vogelgesang attempted to validate search engine queries as a measure of issue salience (Scharkow, & Vogelgesang 2011). Survey responses relating to a controversial 2005 German politician, Paul Kirchof, were compared to the search volumes of "Paul Kirchof". The authors found a weak relationship between the two series ($R^2 = 0.24$), concluding that "search engine queries are a somewhat different but still related measure of the public agenda". However, the evidence for the relationship between the series is weak: due to the small sample size (500 respondents per time point, with a maximum of 3% mentioning Paul Kirchof), variations observed in the time series cannot be distinguished from sampling error.[5]

The above discussion shows that there are problems with drawing conclusions from previous work using Internet search data. This paper seeks to provide a method for assessing when a search index can be used as a proxy for public opinion and when this claim is unjustified. It is not necessary that the people searching for a term are perfectly representative of the population, but the dynamics of their interest needs to closely track that of the whole population.[6]

[5] An additional problem is that the authors use only 34 data points across the time period, and these do not form a continuous time series because no surveys were conducted on weekends. This means that they are unable to perform any tests for stationarity. These tests are particularly necessary in light of the low fit between the series. In the presence of non-stationarity, both R^2 values and levels of significance are likely to be greatly inflated (Granger, & Newbold 1974) and spurious relationships are likely to be found.

[6] Conceptualising public opinion as the average of opinions across a population is not the only possible approach. Price points out that that public opinion has been discussed in a variety of ways. Price argues that survey research shows that "the bulk of the general population is both uninterested and uninformed on most matters that could be construed as public affairs" (Price 1992). Rather than looking at either

1.1.2 The Different Uses of MIP

MIP measures (and the closely related "most important issue" question)[7] are widely used within political science. There are three main areas of political science research where these issue salience measures are used: research on issue voting, agenda setting and public policy.[8]

The valence voting literature has used MIP in a number of ways. An early example was Repass who found that party identification was a weaker predictor of vote choice in the 1960 and 1964 US presidential elections if the voter said that the other party (the one they didn't identify with) was more competent on the issues they identified as most important(RePass 1971). More recently, Clarke et al. found that American voters who viewed the Republican party as best able to handle the 'most important issue' were more likely to vote for the Republican House candidates in the 2010 midterm elections even after accounting for party identification, political attitudes and demographic characteristics(Clarke et al. 2012).

It should also be noted that the valence approach is one of several approaches to issue voting. For instance another strand has focused primarily on positional issues (those issues where there is ideological divison over the correct course of action). Indeed Clarke et al. also found support for the effect of positional issues such as the Obamacare reform and environmental protection on voting in the 2010 House elections in the same article.

There have been more mixed findings when using the MIP question to predict vote

the general public (the whole population) or the voting public (the subset who vote at least occasionally), Price highlights the attentive public (those who pay attention to politics) and the active public (those who actually participate in politics) as potentially more important groups in terms of policy making. Google Trends data could potentially provide measures for all of these publics. For instance we might use search terms that would reflect increasing levels of political sophistication in order to distinguish between the interest of the different publics (e.g. 'jobs' versus 'unemployment' versus 'output gap'). For this article I focus just on the general public for two reasons. First, the existing literature on issue salience overwhelmingly uses the general public so creating measures that track the general public will be most relevant to researchers. Second, the available aggregate issue salience data overwhelmingly focus on the general public. Although it is plausible that Google Trends could track the other forms of public opinion, it is no less necessary to validate Internet data when looking at the active public than for examining the general public. The approach in this paper would be equally applicable to validating against data measuring public opinion conceptualised in the other ways discussed by Price.

[7] Although there are differences on some issues (Jennings, & Wlezien 2011), the 'most important problem' and 'most important issue' questions are generally believed to tap into the same underlying construct.

[8] One additional area of research on issue salience that has started to be investigated is issue salience congruence (Reher 2012) where the fit between the public's issue agenda and that of political parties is examined. Aggregate issue salience data could help to look at short and long term effects of the public agenda on the issue priorities of parties. Previous studies that have looked at the fit between issue salience in manifestos and the public have tended to focus on a single time point without considering a longitudinal dimension (Shephard 2007).

directly (rather than with the question of which party is considered most competent). Niemi and Bartels found that the issue salience measures in the 1980 National Election Study did not improve the explanatory power of models when they were used to weight an individual's evaluation of candidates on different issues(Niemi, & Bartels 1985). However, some research has reported positive findings including Bélanger and Meguid who found that issue salience mediates the effect of issue ownership on vote choice(Bélanger, & Meguid 2008).

Internet search data are less relevant to issue voting studies such as the ones cited here, as these studies usually use individual level measures whereas Google data is aggregated. However, one recent paper by Bartle and Laycock shows the relevance of aggregate issue salience data to vote choice. Bartle and Laycock argue that a voter's choice of salient issue does not seem to reflect the weight she puts on it in her voting decision (Bartle, & Laycock 2012). Indeed, they found that voters' evaluations of political parties' economic competence has less influence on vote choice among voters who chose the economy as the most important issue. However, they found that the overall distribution of answers on the 'most important issue' question closely reflected the weight voters put on comparative evaluations of party competence on each issue. In other words, aggregate issue salience results are better measures of the weight individuals put on issue evaluations than their own answers to that question. Google data could potentially provide a useful measure of the relative weight different issue evaluations have on vote intentions over time although, we would still lack good longitudinal measures of comparative competency evaluations.

Issue salience measures have been heavily used in the agenda setting literature. Beginning with McCombs and Shaw's Chapel Hill Study (McCombs, & Shaw 1972), MIP measures have been used in more than 100 papers on the agenda-setting effect of the media according to Soroka (Soroka 2002). Some of these (including McCombs and Shaw) use individual level measures but many others compare aggregate MIP measurements with time series of media coverage of these issues. Most of these studies have looked at long term relationships using surveys conducted monthly or even less frequently (see Brosius, & Kepplinger (1992) for one exception). However, experimental evidence suggests that people's responses to political messages are relatively short lived (Chong, & Druckman 2010) so studies of the effects would ideally use much more frequent measurements of issue

salience. Weekly Internet search data could provide a much more practical alternative to conducting weekly MIP surveys.

Finally, longitudinal MIP data are also used to examine the interplay between public policy and issue salience (Jennings 2009; Jones, Larsen-Price, & Wilkerson 2009; Soroka, & Wlezien 2009). For instance Jones found that legislatures pay more attention to public opinion on an issue when it is salient (or at least to the preferences of the public that is paying attention) (Jones 1994). These studies also tend to look at the long term relationship but more frequent issue salience measurements could help isolate the mechanisms that link policy change and issue salience. For instance, the extent to which people respond to actual or expected policy change.

1.2 Data

This section outlines the properties of Gallup MIP data and Google Internet search data. It then discusses data issues that arise when comparing the two. The data are for the US between January 2004 and March 2010. The Google data are aggregated weekly across the time period whilst the Gallup data are available once a month at irregular intervals. The US was chosen due to having the highest number of Google searches in the world and its centrality in the agenda-setting literature. This paper attempts to measure the following issues: macroeconomics, Iraq, healthcare, dissatisfaction with government, immigration, terrorism and fuel prices, which were the MIP issues with sufficient numbers of respondents choosing them over the time period to allow analysis.

1.2.1 Google Internet Search Data

Search indices were accessed through the Google Trends website (Google 2004)[9] which allows a user to retrieve search indices aggregated either weekly across any period, or daily for up to 90 days at a time. This paper uses the weekly data to allow the validation of time series over the full time period which allows the analysis of longer term trends.[10]

[9] Until September 2012, Google released its data through two sites: Google Trends and Google Insights for Search. These were then merged into a single Google Trends site that takes most of its relevant functionality from Google Insights for Search. There may be small differences in functionality over time and this paper discusses the functionality that was available in 2011 when the data were downloaded.

[10] The daily search data might also be valuable in some studies, particularly if a researcher is interested in changes in issue salience surrounding particular political events such as elections or crises. From a

The Google data are not given in absolute volumes but are indexed to the highest observed search volume, which is set to 100. Consequently, it is not possible to ascertain the frequency of searches that took place at any given time, but only how the searches have changed over time.

Because Google data are indexed on an arbitrary scale, this paper converts valid search indices onto the same scale as the MIP survey responses (i.e. the percentage of respondents choosing that issue) by using the search index to predict the proportion of respondents who would cite an issue as important.

1.2.2 Representativeness of Internet Users

A potential problem with Internet data is that users are not fully representative of the population as a whole demographically and, as such, some groups are underrepresented in the population of potential searchers. Analysis of a 2010 Pew Internet use poll (Pew 2010) shows that younger age, higher education and higher income predict increased probability of using the Internet. To test whether these differences could bias the Google search data, the individual level Gallup MIP data was reweighted to reflect the demographics of Internet users.[11] The MIP time series resulting from the reweighted data was correlated against the original MIP series for each issue. The correlations between the series are extremely high ($R^2 > 0.9$ in all but one case), indicating that the demographic differences between Internet users and non-Internet users are unlikely to introduce significant biases.[12] Search behaviour *among* Internet users[13] has also been shown to vary according to demographic and psychological variables (Ford, Miller, & Moss 2001), but it is unlikely that these other

methodological viewpoint validating short run data such as this would run into several difficulties. First, we would require many waves of an MIP survey conducted within the 90 day period to compare the search data to. Second, there are substantial day of the week patterns within searches that would need to be adjusted for. Finally, due to the 90 day restriction, there would be insufficient data to find and correct seasonality. These problems would all be solvable with daily search data over a longer time period but make the use of the current daily data difficult. I initially considered concatenating the 90 day data to create longer run daily data but this quickly runs into difficulties with the rounding in the daily data. Google could hypothetically release daily data over longer periods or at even lower levels of aggregation but these data are not currently available publicly.

[11] Weights were created from predicted probabilities of Internet usage derived from the Pew survey using age group, sex, education and income level as independent variables. These weights were then combined with the Gallup survey weights to derive the time series for each issue. Only 30 of the 75 Gallup surveys across the time period in this study had individual level data available through iPoll.

[12] The one exception to this were MIP responses for healthcare issues: the two series only correlated with an R^2 of 0.79.

[13] Other research on Internet searching suggests that many people search iteratively, so will make multiple searches using some of the same search terms in order to hone their results(Du, & Spink 2011).

effects would be larger than the effect of demographic characteristics, so they should not seriously bias the results. Even given this analysis, the people searching with some search terms are likely to be much less typical of the public than Internet users in general. Hence, it is necessary to demonstrate the correspondence between each search index and data that is representative of issue salience in the wider population.

1.2.3 Gallup MIP Data

For this paper, summaries of Gallup's MIP responses were collated between 2004 and 2010 for all seventy-five months (Gallup 2010). The open-ended survey responses were coded in line with the Public Policy Agendas (PPA) schema (Baumgartner et al. 2006), with the exception of terrorism, which is included as a single category. The PPA category of macroeconomics includes the economy in general, unemployment, taxes, budget deficits and inflation.

This resulted in seven substantive MIP series, as only issues mentioned by at least 10% of respondents in at least one survey were included. As such, this paper attempts to find Internet search measures for the following issues: macroeconomics, Iraq, healthcare, dissatisfaction with government, immigration, terrorism and fuel prices.

1.2.4 Differences between MIP and Search Data

Although Gallup's MIP data are released once a month, the fieldwork is generally conducted for three days so one survey is an aggregate of just three days of public opinion. From this view, the Gallup series is a time series of three day windows with 90% of observations missing. That is, any variation in public opinion that occurs outside of the three day window each month is not captured at all.

For the validation of Google searches against Gallup's MIP, a week of Google searches is chosen that corresponds most closely to each Gallup poll's timing. If the days on which Gallup interviewed respondents do not fit into a single week of Google searches, the week with the most interview days is selected as corresponding to that Gallup survey.

Another difference between Gallup's MIP and Google's search indices is that Gallup respondents are at least partially restricted in the number of responses they can give to the MIP question. Consequently, the responses in each category are correlated negatively

with responses in other categories (Wlezien 2005). By contrast, Google searches are not structurally constrained by searches for other issues,[14] which may lead to lower correlations between search indices and MIP series. It is preferable to have unconstrained salience measures for testing theories that refer to the level of importance the public gives to an issue, so this paper will not adjust the models used to create constrained series.

1.3 Methodology

This article has two main aims: 1) to assess whether any Internet search indices can function as measures of issue salience and 2) to generate a set of issue salience measures based upon these indices. To fulfil these aims it is important that the measures satisfy two of the key requirements for measurement: reliability and validity. The validity of each search index is assessed in several steps: face validity, content validity, seasonal adjustment and criterion validity. At each of these steps, the search index is excluded if it lacks validity at that stage, reducing the number of search indices throughout the process. This full process is shown in Figure 1.1.

For the remaining search indices an OLS regression model is fitted to the MIP and search index for an issue. This model is used to predict MIP values for every week across the period, for valid search indices. In addition to these validity checks, the reliability of the generated issue salience measures is estimated through the use of out-of-sample prediction.

1.3.1 Face Validity

This paper begins by selecting a number of indices with face validity for each potential issue. This list is not intended to be exhaustive but just a reasonable selection of terms with face validity. For instance, the words "immigration" and "immigrant" have face validity for measuring concern about immigration. This contrasts with the approach of some previous papers that looked at all possible search indices (around 50 million), selecting only those that fitted statistically with the dependent variable of influenza outbreaks (Ginsberg et al. 2008). That approach risks finding significant relationships simply by chance due to the

[14] Empirically, concern over one issue may genuinely decrease concern about others due to cognitive capacity and prioritisation in the public (as well as the limited amount of time any individual can spend searching on the Internet). However, the relationship is likely to be more complex than simply dividing a permanently fixed quantity of importance between issues.

high number of tests being performed.

1.3.2 Content Validity

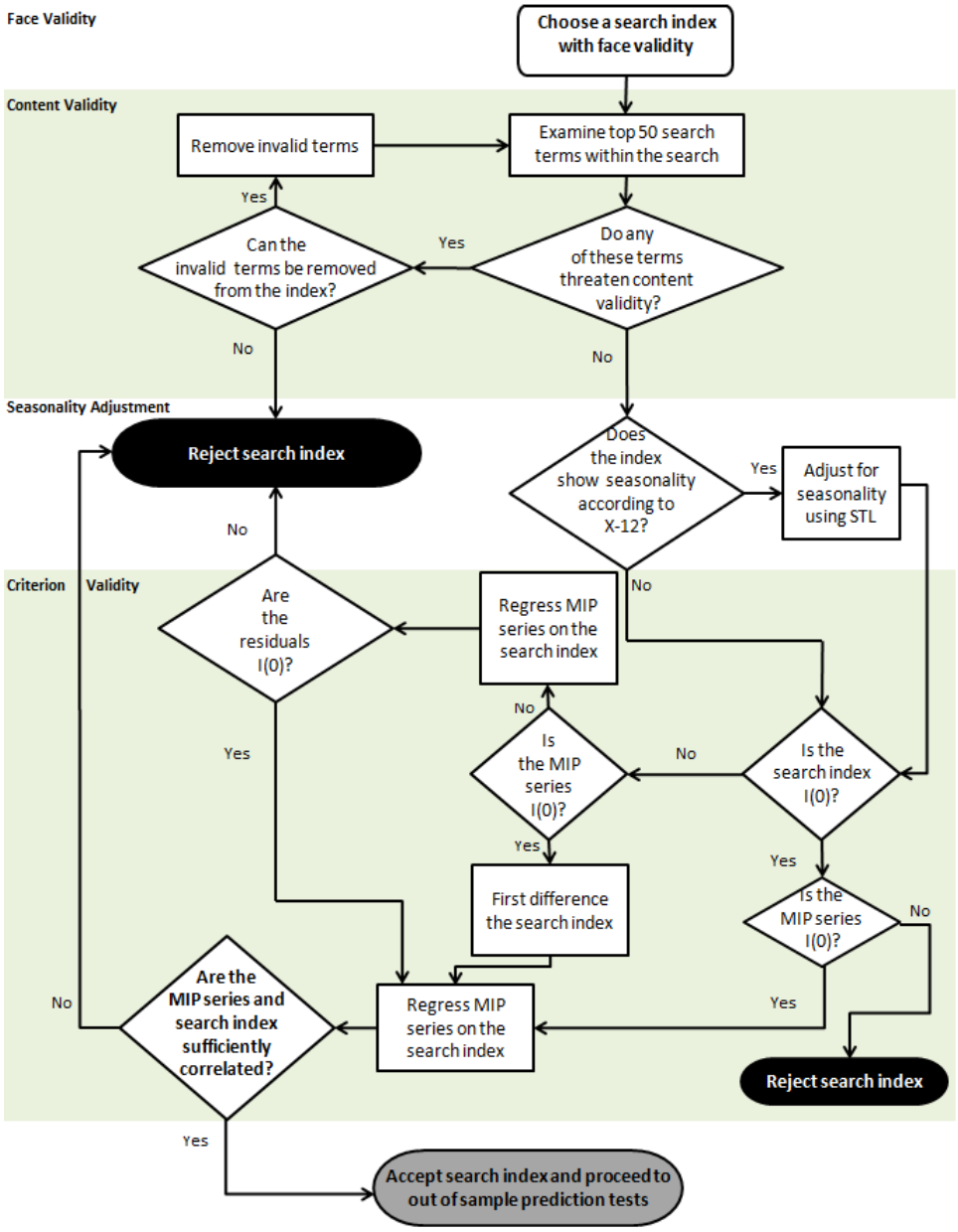
Carmines and Zeller define content validity as the extent to which "an empirical measurement reflects a specific domain of content" (Carmines, & Zeller 1979). In this study content validity is the extent to which a search index contains search terms that reflect the issue. That is, searches that someone who is concerned about a given issue would make.

Content validity is assessed by examining the top fifty search terms in each index for each year, which allows for the identification of terms not plausibly related to an issue's salience which make a substantial proportion of the search index. For instance, the search index of Iraq might include searches for "Iraq music", which is not in the domain of searches made by someone concerned about the issue of the Iraq conflict. These confounding searches can then be removed from the search index by adding "-[confounding term]" to the query on the Google Trends website. In the example, the query would read "Iraq –music". However, only a limited number of terms can be removed, so an index would be rejected if there were more confounding terms than this example, as shown in the content validity section of Figure 1.1.

1.3.3 Criterion Validity

Criterion validity describes how closely a measure corresponds to an existing measure of the concept under study (Carmines, & Zeller 1979). Since MIP is well established as a measure of issue salience, it makes sense to 1) use it as a criterion to compare new search index measures to and 2) distinguish between those indices that can measure issue salience and those that cannot. Criterion validity is assessed by an OLS regression using MIP as a dependent and the relevant search index as the independent variable, with Newey-West standard errors that are robust to autocorrelation and heteroskedasticity (Wooldridge 2009; Newey, & West 1987). Criterion validity is best regarded as a continuous property rather than a binary one (Carmines, & Zeller 1979), so the weight that can be placed on the results of criterion validity tests will depend on the strength of association found between the MIP series and Google searches (i.e. R^2 values showing the percentage of variance explained).

Figure 1.1: Flowchart showing the process involved in selecting search indices that are valid for measuring issue salience.



1.3.4 Time series complications

Before the regressions can be performed, it is necessary to take account of several time series complications. Determining criterion validity is considerably more complex when using time series data, as seasonal patterns and non-stationarity can generate misleading results. This subsection discusses how to deal with these concerns.

Many of the search indices include a substantial seasonal component not present in the MIP series. It is not possible to definitively say why MIP series show less seasonality than the Internet searches. However, one possibility is that respondents to the Gallup surveys discount seasonal trends in the obtrusiveness of a problem when responding to pollsters; for instance, Internet searches for "jobs" dip in December before rebounding in January. This wouldn't reflect a stronger economy every December but simply reduced activity in the labour market over the holiday period. As people are aware of this fact, they will discount this reduction when reporting their concerns about the economy to the pollster. Consequently, it is necessary to adjust the search index to remove the seasonal component.

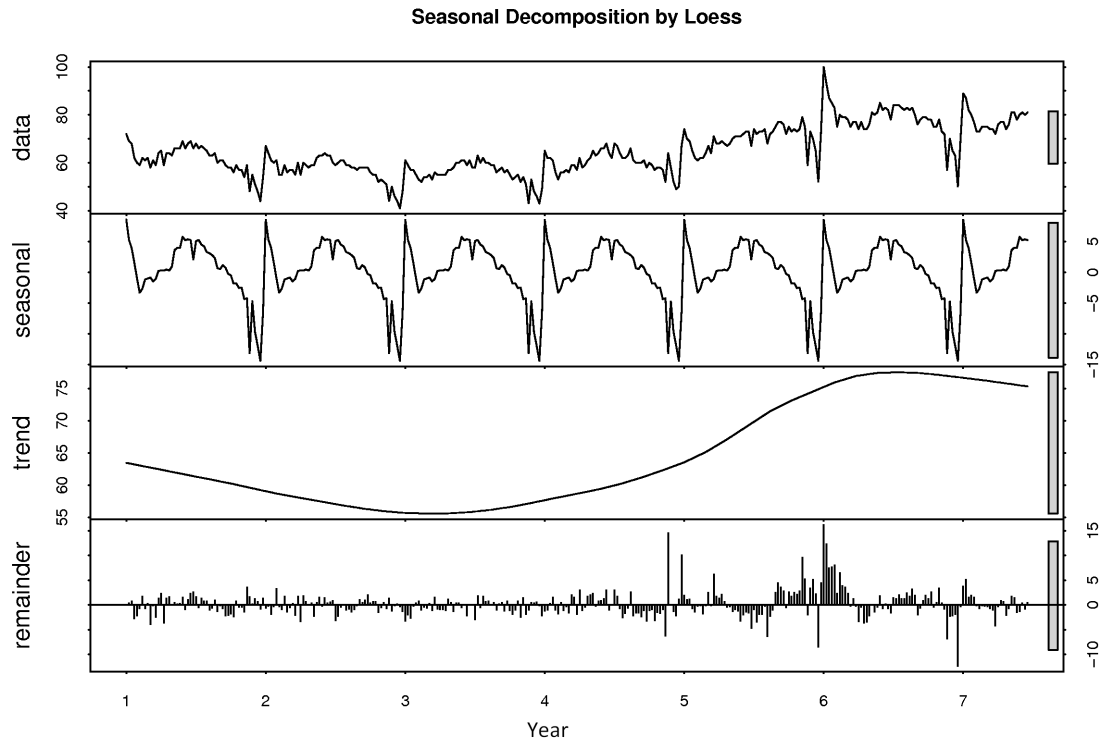
To decide which MIP and search index time series to adjust for seasonality, the X-12 seasonality diagnostic tests were run on each series, as suggested in the US census guidelines for seasonal adjustment (McDonald-Johnson et al. 2010). Seasonal adjustment was performed using Seasonal Trend Decomposition by LOESS (STL) (Cleveland et al. 1990) , which separates a time series into trend, seasonal and remainder components (Cleveland et al. 1990), with 52-week periodicity. To explain how this works, an example is shown for the "jobs" search index in Figure 1.2 . The initial series (labelled data) clearly displays a yearly pattern with dips in searches for jobs around December, which is confirmed by the large seasonal component (seasonal) identified. Once the seasonal component is removed, a smooth trend component and the residual variance (remainder) are returned and recombined to create a seasonally adjusted series.

Non-stationary time series are prone to spurious correlation (Granger, & Newbold 1974), which will compromise criterion validity tests. The MIP series and Google Trends series are first tested for stationarity using Augmented Dickey-Fuller tests attempting to reject the null hypothesis of non-stationarity.

Where two series are different orders of integration[15] it is not possible to estimate

[15]Order of integration is expressed as I(x) where x is the number of first differences required to obtain a

Figure 1.2: Seasonal Trend Decomposition by LOESS of searches for "jobs" on Google showing the original data and its seasonal, trend and remainder components. The axes are all expressed on the same arbitrary scale.



accurately the relationship between them.

The different combinations of orders of integration and the responses to them are shown in the criterion validity section of the flowchart in Figure 1.1 (note that a glossary of some terms used can be found in Appendix 1.2).

If the search index is I(1) and the MIP series I(0), the search index is first differenced to obtain a stationary series which is used in the regression.[16]

However, when the MIP series is I(1) and the search index is I(0), the search index is rejected because taking the first difference would mean predicting the change rather than the level of the issue salience, which is not substantively interesting (this paper is concerned with generating time series of *levels* of issue salience rather than generating time series of the changes in issue salience since the previous month).[17] If both the search index and MIP series are I(1) and cointegrated, then OLS estimates of coefficients are consistent and unbiased (Wooldridge 2009). The series are cointegrated if their residuals form a stationary sequence. To test for cointegration, OLS is run and its residuals tested for cointegration using adjusted critical values (Davidson, & MacKinnon 1993) to assess whether the null hypothesis of non-stationarity can be rejected for the residuals. Series that are cointegrated can be correlated consistently and are included in the next stage.

Once these potential complications are accounted for, the OLS regressions are run between the MIP series and search index. These use Newey-West standard errors which account for the effect of autocorrelation on the standard errors of the estimates.

1.3.5 Prediction of weekly issue salience series

The second aim of this article is to generate weekly issue salience measures for use in empirical investigation, for those search indices that are accepted after following the process outlined in the flowchart in Figure 1.1. The monthly MIP series for each issue is regressed on the relevant search index (that has been adjusted for seasonality, if appropriate). This

stationary sequence. For a wider discussion of stationarity and orders of integration see Wooldridge (2009).

[16] By taking the first difference we create a sequence of the differences between each observation and the previous one. If we have a non-stationary time series X, its mean and variance will change over time. However, the mean and variance of the first difference of X will not necessarily do so. This means that we can get a stationary sequence from a non-stationary series by differencing X one or more times (until the series is stationary). The number of times we have to difference X to obtain a stationary time series is called X's order of integration or $I(d)$, where d is the number of differences.

[17] Note that this would also mean using the weekly data to predict a time series of monthly changes.

model is then used to predict MIP responses for each week between January 2004 and March 2010 for that issue, which produces a weekly time series measured on the same scale as MIP responses, allowing comparisons to assess which issues are most salient at different times.

1.3.6 Out-of-sample testing

The utility of the time series generated in the previous section is dependent upon how well they extrapolate to time points without MIP data. Two different kinds of extrapolation would be valuable when using the instruments generated from search data – firstly, extrapolation to more detailed time points within the time period under study (weekly data) and secondly, it would be useful to predict issue salience values outside of this period in future, without having to fit new models between MIP and search indices. The existing MIP data are split into training and test sets, with a model fitted to the training set, which is then used to predict the test set. Root mean square errors are calculated for the training and test samples to assess the degree to which the error increases when the model is fitted to new data. Apart from the reduced data to which they are fitted, the models are specified in the same way as when predicting weekly values in the previous section.

Two separate versions of these tests are performed to reflect the two kinds of extrapolation performed in this paper. As no weekly MIP data are available, the issue of extrapolating to finer time intervals is addressed by arbitrarily defining the training sample as all even years and the test sample as odd years. This split was chosen in order to ensure the independence of the training and test samples, which could be undermined by autocorrelation if alternate months or a random sample were used. This means the test measures the accuracy of predicted values within the time period. Since the model is fitted to less data than the actual model used to generate the weekly series, this test should produce an estimate of the lower bound of weekly issue salience measure reliability.

To address the issue of extrapolation outside of the time period, the training sample is defined as the years 2004-2006 and the test sample as 2007-2010. This test looks at whether the relationship changes systematically over time. This second out-of-sample test is not performed on time series which are trend stationary, as they need to include a linear time trend in the regression to be calculated accurately and this would potentially lead to

unrealistic predictions, such as negative salience.

1.4 Results

This section details the outcomes of each section of the methodology outlined in this article when applied to Google search data and Gallup's MIP series in the US between 2004 and 2010. The search indices are measures of how many times a search term (such as the word "immigration") is searched for over time. Each Gallup series measures what proportion of respondents chose a particular issue (for instance 'the economy') as the most important problem facing the country at the time of each survey. In this paper, a search index is only included in each subsequent stage if it passed the requirements of the previous one.

1.4.1 Face Validity

Search indices with face validity were chosen for each of the seven issues that could be tracked across the complete period of interest, using MIP: macroeconomics, Iraq, healthcare, dissatisfaction with government, immigration, terrorism and fuel prices. Some potential search indices such as "health reform", used for concern about healthcare, could not be selected because they lacked sufficient search volume to track weekly frequencies across the full period. As 'macroeconomics' is a composite issue of several MIP responses, search indices such as "unemployment", "fiscal" and "taxes" were chosen because they corresponded to the components. In total, twenty search indices with face validity were selected to measure the seven issues that are important across this time period, as shown in Table 1.1 in Appendix 1.1.

1.4.2 Content Validity

Each of the twenty search indices was examined for threats to content validity. A good example of the content validity methodology is the jobs search index employed to measure the salience of "macroeconomics". The bulk of the searches comprised various types of jobs ("catering jobs", "government jobs", etc.), although two invalid terms were found in the fifty top searches – "blow jobs" and "Steve Jobs" (the former CEO of Apple Inc.) – which were then excluded from the search index. Table 1.1 in Appendix 1.1 summarizes the results of

the content validity checks. For each search index (shown in the second column) the table indicates the relevant issue it is intended to measure (first column), which terms within the index threatened content validity (third column) and finally whether the index was deemed to be content valid (fourth column). Of twenty search indices, six were rejected for lacking content validity and three initially included problematic search terms that were able to be removed. After the content validity checks, fourteen search indices covering all seven issues remained for further testing. Note that the full content validity tables are included as supplemental material.

1.4.3 Time series complications

Before assessing the criterion validity of the remaining search indices, it is necessary to account for the time series properties of the data, namely seasonality and non-stationarity. Eight of the fourteen remaining search indices showed evidence of seasonality and were adjusted, but none of the MIP series exhibited seasonality.

Non-stationary time series can be spuriously correlated, which has to be accounted for before assessing criterion validity. Five indices could not be used in OLS regressions because of differing orders of integration.[18] Although these search indices are rejected here, it is worth repeating that the ADF tests used to identify order of integration are not powerful, so it is possible that some series are wrongly rejected. As longer time series of Google searches become available, these tests should be able to more accurately assess which series are stationary and which are non-stationary.

Another four search indices and their relevant MIP series were I(1). These were tested for cointegration with the relevant MIP series. Only one of these, "jobs", was cointegrated, so the other three search indices were rejected.

[18] If two time series really are different orders of integration then they are not good proxies for each other as one has a probability distribution that varies over time and one does not. Hence we cannot rely on the relationship between them staying stable over time. As I note in the methodology, we can difference a non-stationary sequence to obtain a stationary series but this does not make sense in the context of MIP series as we want to predict its value not its first difference. Some readers may consider the stationarity tests unduly demanding. Some valid search indices may have been excluded, as augmented Dickey-Fuller tests have low statistical power (Hansen 1995), so often fail to reject non-stationarity even when the series is actually stationary. This paper is intended to provide a set of valid issue salience time series that can be used in empirical research with a high level of confidence. However, if a lower confidence level is acceptable, these criteria can be relaxed to generate more series for empirical use by choosing a lower significance cut-off for the tests

Table 1.1: Criterion validity testing by OLS regression analysis

Independent Variable (Search Index)	Dependent variable (MIP series)	Coefficient	Newey-West Std. Error	R^2
Jobs	Macroeconomics	1.726*	0.132	0.75
Illegal immigration	Immigration	0.284*	0.038	0.67
Terrorism and Al Qaeda (time trend)	Terrorism	0.305*	0.051	0.76
Fuel	Fuel Prices	0.522*	0.063	0.72
Gas	Fuel Prices	0.824*	0.218	0.49
Terrorism (time trend)	Terrorism	0.029	0.037	0.67

* $p < 0.01$ n=75 for all regressions

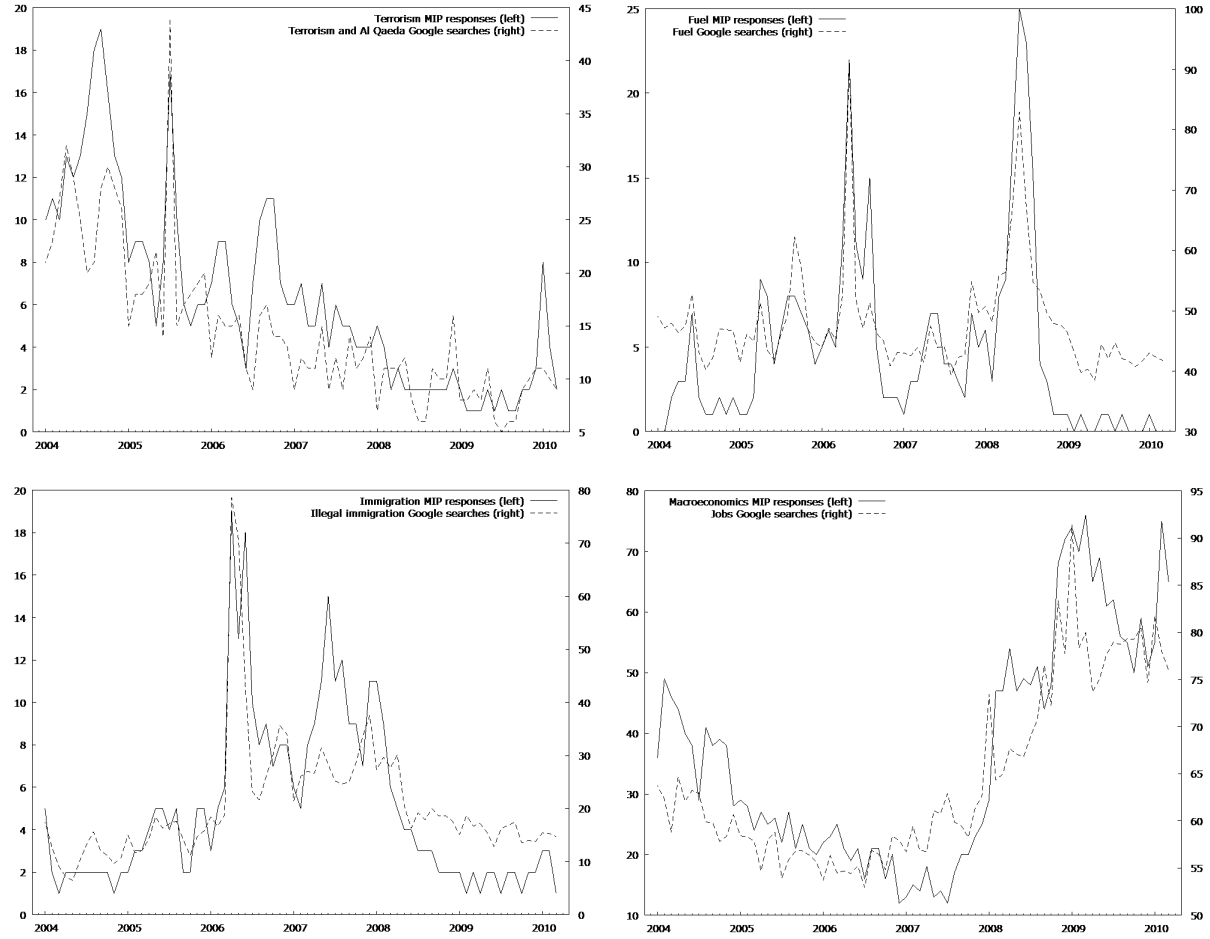
1.4.4 Criterion Validity

The remaining six search indices were tested for criterion validity by regressing each MIP series on the relevant search index using OLS with Newey-West standard errors. The results are shown in Table 1.1, with coefficients and standard errors shown for each search index predicting the corresponding MIP series. The "terrorism" regressions include time terms because the "terrorism" MIP series is trend stationary. Five search indices were found to predict the relevant MIP series significantly, indicating high levels of association, but searches for "fuel" were selected over those for "gas" as a measure of fuel price concern because the association was substantially higher at an R^2 of 0.72 compared to 0.49.[19] The "terrorism" search index was not significantly related to the MIP series, but the search index combining searches for "terrorism + al Qaeda" showed a strong fit, which left four issues with search indices that were significantly and substantively related to the relevant MIP series. The R^2 values ranged from 0.67 to 0.75, demonstrating a very strong association with the MIP series. These covered four issues: macroeconomics, immigration, terrorism and fuel prices.

The four MIP series and corresponding search indices that survived the validity tests are shown in Figure 1.3 for the weeks where a Gallup survey was available. The left axis shows the proportion of respondents mentioning the particular issue as the MIP that month and the right axis corresponds to the arbitrary scale for that search index. These graphs confirm that the Internet search indices are visually similar to the equivalent MIP time

[19] A separate model was fitted with both "gas" and "fuel" search indices predicting the fuel MIP series. However the "fuel" search index was no longer significant in this model and the overall fit only slightly increased (adjusted R^2 from 0.711 to 0.718).

Figure 1.3: Frequency of Internet searches and MIP responses for terrorism, fuel, immigration and macroeconomics for each week where there is an MIP survey.



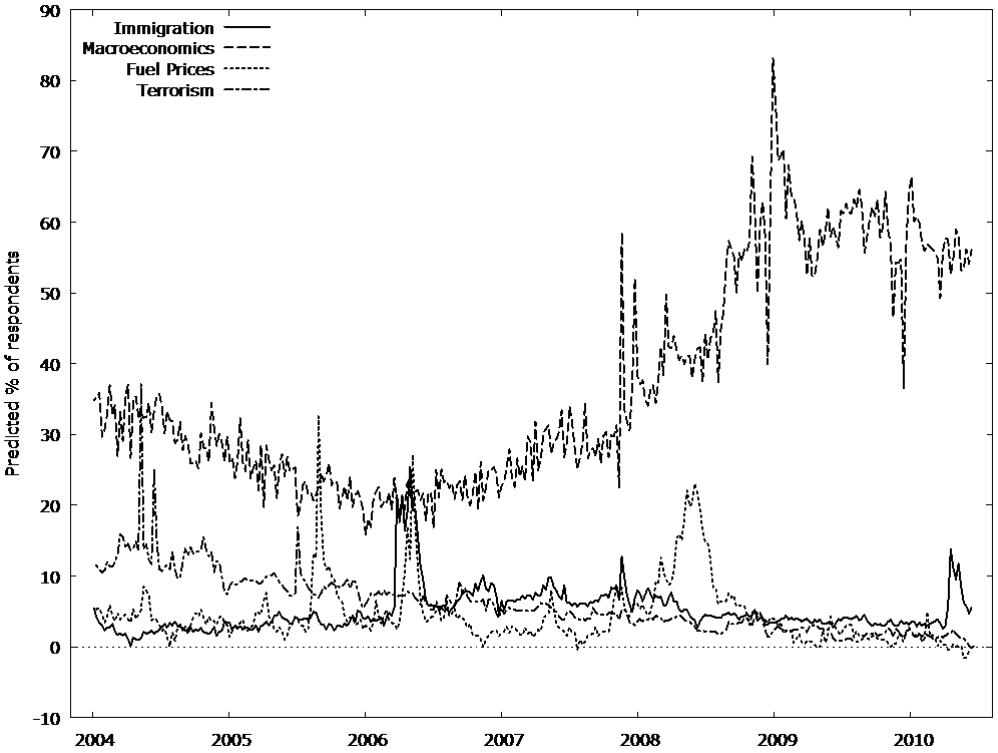
series, with similar trends and peaks visible in both time series.

1.4.5 Predicting weekly issue salience series

To create weekly series, OLS models were fitted between the monthly MIP series and the equivalent search index week, and then used to predict MIP scores for every week across the time period. The final four predicted issue salience series for immigration, macroeconomics, fuel prices and terrorism are shown in Figure 1.4.

These time series allow a comparison of each issue's salience over time and against other issues. The chart shows terrorism's salience gradually declined over time, whilst macroeconomic problems increased in salience as the economy worsened. By contrast, the salience of immigration and fuel prices is characterised by short periods of greatly increased

Figure 1.4: Predicted percentage of weekly MIP responses for fuel prices compared to monthly MIP percentage of responses.



salience (each issue briefly overtakes macroeconomics in importance), but generally they elicit only low levels of importance.

A comparison of the weekly predictions and monthly survey responses is shown for fuel prices in Figure 1.5. The predicted fuel series looks immediately suspect, with a large peak in the salience in September 2005, whilst the MIP series gives no indication of such a peak. However, an explanation becomes clear when examining gasoline prices around that time, as for one week the average gasoline price was greater than $3 per gallon for the first time ever, but it fell sharply the week after (US Energy Information Administration 2012). Since Gallup did not survey in that week, it missed the spike in salience accompanying the fuel price rise.

This reinforces the point made in Section 1.2.4 that the Gallup series are not monthly data in the sense of representing each month. They are a snapshot of 3 days within each month. If the survey happens not to be conducted in the week that something important, such as a fuel price spike, occurred, then there may be no reflection of this in the data.

Figure 1.5: Predicted percentage of weekly MIP responses for fuel prices compared to monthly MIP percentage of responses.

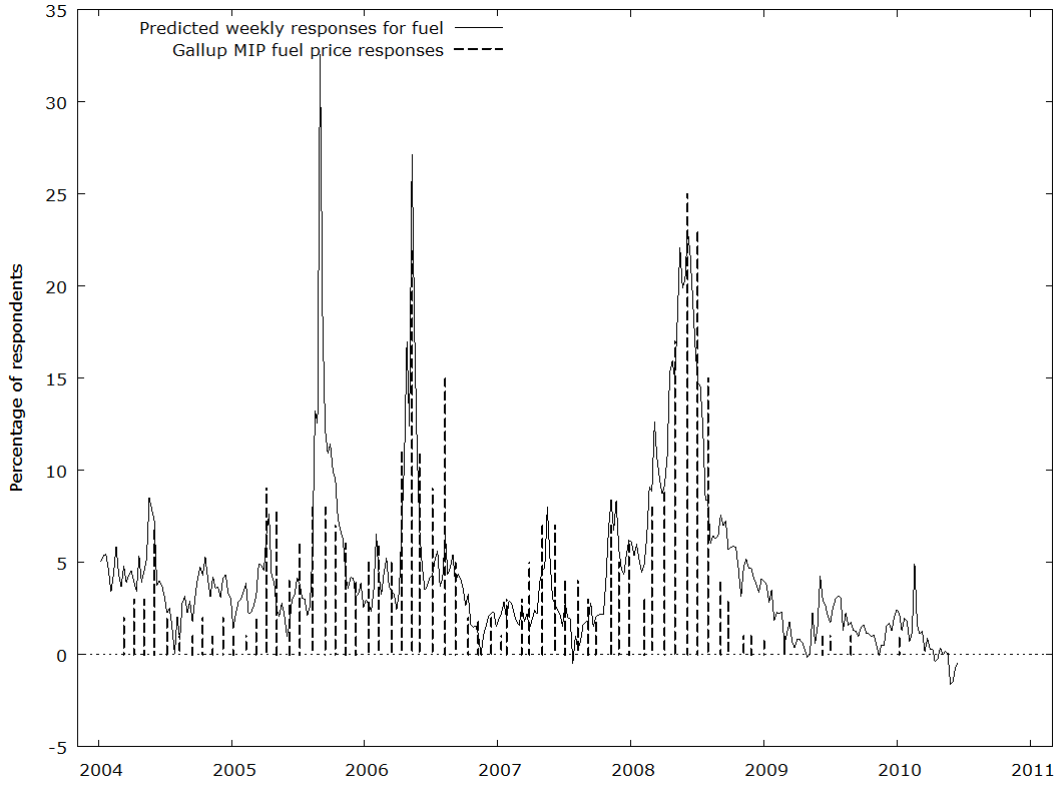


Table 1.2: Even/odd year out-of-sample reliability test results

Independent variable (search index)	Dependent variable (MIP series)	RMSE in-sample	RMSE out-of-sample	Mean MIP series
Jobs	Macroeconomics	8.12	11.54	36.7
Illegal immigration	Immigration	2.13	2.67	5
Terrorism and Al Qaeda	Terrorism	2.82	2.31	6.3
Fuel	Fuel Prices	3.42	2.33	4.7

1.4.6 Out-of-sample reliability tests

The generated weekly series were checked for face validity, content validity and criterion validity, but it still needs to be determined how reliable they will be outside the data to which they were fitted. To achieve this end, an out-of-sample test is performed by regressing the search indices on monthly MIP data for even years and testing the resulting model in predicting monthly MIP data for odd years. The results for this out-of-sample test are shown in Table 1.2.

Overall, the "jobs", "illegal immigration", "terrorism + al Qaeda" and "fuel" search indices show good reliability in predicting MIP series within the 2004-2010 time period, suggesting the weekly measures of issue salience will be reliable within this period. The root mean square error (RMSE) of the models changed between the training sample and test samples. In the case of the "jobs" search index predicting the macroeconomics MIP series, the RMSE increased but is still reasonable given that 37% of respondents choose a macroeconomic problem on average across the time period. The "terrorism", "illegal immigration" and "fuel" search indices' RMSEs did not substantially change in the test sample, suggesting a high degree of reliability outside the data the model was fitted to.

The second out-of-sample test fits a model to the years 2004-2006 and tests the predictions against data from 2007-2010 to assess how well the models are likely to perform outside the time period to which they were fitted. The terrorism weekly series is not tested in this section, as it was trend stationary. The results for these reliability tests are shown in Table 1.3.

This second set of out-of-sample tests broadly agrees with the first. The "jobs" search index RMSE greatly increases out-of-sample, which suggests that it will not be highly reliable when extrapolated beyond the period to which the model was fitted. The "illegal immigration" and "fuel" search index RMSEs increase moderately but are still broadly

Table 1.3: Contiguous out-of-sample reliability test results

Independent variable (search index)	Dependent variable (MIP series)	RMSE in-sample	RMSE out-of-sample	Mean MIP series
Jobs	Macroeconomics	6.54	14.7	36.7
Illegal immigration	Immigration	1.99	2.8	5
Fuel	Fuel Prices	3.01	3.15	4.7

reliable. These results suggest that the "illegal immigration" and "fuel" search indices would be reliable measures in the future, even without fitting new models to more MIP data, while the "jobs" search index would be unreliable if extrapolated beyond the period to which it was fitted.

The number of search indices narrowed throughout this paper as follows:

1. Face validity – the method was started with twenty search indices with face validity.

2. Content validity – six indices were removed for lacking content validity, leaving fourteen indices.

3. Seasonality – eight out of fourteen indices showed seasonality and were adjusted accordingly.

4. Criterion validity – eight search indices were rejected because they could not be tested appropriately using time series methods. One further series was rejected because it did not demonstrate criterion validity when tested against MIP data. This left five indices covering four issues.

At both the content validity stage and criterion validity stage a high proportion of the search indices were removed. The reduction from twenty indices to five shows that simply assuming validity would have been misleading and have led to erroneous conclusions: if researchers use invalid measurements in research, they are not answering the questions they believe they are answering. Importantly, however, only one search index out of six was rejected at the criterion validity testing stage, suggesting that the content validity analysis was successful in removing many time series that were not suitable measures of issue salience. The method also led to the successful measurement of four of the seven issues.

1.5 Comparison of Internet Search Measures with MIP Measures

The four newly generated issue salience measures are available for use in future empirical work,[20] so it is useful to compare their properties to those of traditional MIP measures. The new procedures use MIP as the standard against which to measure, so it is not possible to test directly whether these Internet search measures can surpass MIP as gauges of issue salience.

The primary advantage of Internet search measures is their greater availability both in terms of frequency. The Google measures are aggregated at a weekly level compared to the MIP measures which only allow us to observe a single three day window each month (when the fieldwork was conducted).

Another disadvantage of MIP is that only allow one problem to be reported by the respondent, so when one issue becomes important in the population it will reduce MIP responses for other issues, even if they are also important to respondents. This introduces an artifactual, negative correlation between different issues and thus can lead to problems with interpreting analyses that use MIP, as responses for one issue will be correlated negatively with any variable that is correlated with any other issue. Search indices do not constrain each other directly, so these artificial correlations are not introduced. The difference is apparent when examining the different issue salience time series, in that the MIP series for macroeconomics and immigration correlate strongly ($r = -0.62$), whereas the equivalent series generated from search indices correlate far less strongly ($r = -0.21$).

Existing MIP measures also violate the time series assumption of uniform intervals between measurements. The time between surveys ranges between three and five weeks, which means that time series models will not be specified correctly because they will not calculate the effects of lags consistently. However, MIP series are still subject to the problems experienced with time series (autocorrelation and non-stationarity), so models that ignore these aspects of the data will draw erroneous conclusions. By contrast, the search indices consist of evenly spaced weekly data, so they can be modelled accurately using time series techniques.

[20] The four issue salience series are available at [URL anonymized].

The major limitation of Internet search measures of issue salience is that they will not necessarily capture the importance of all major issues in a time period. This contrasts with MIP measures which will capture the salience of any issue a respondent believes is important. MIP, then, provides a snapshot of what issues the public considers important at a point in time, while Internet search data is limited by our ability to find suitable terms so may not provide a complete picture. This paper was able to measure the salience of fuel prices, macroeconomics, terrorism and immigration but was unable to capture the importance of dissatisfaction with government, healthcare and the Iraq conflict. The salience of the healthcare issue could not be captured using search data for two reasons. The MIP responses for healthcare broadly follow political debates on healthcare reform, whereas searches for "healthcare" are mainly focused on searching for health insurance information and are stable over time. The search term "healthcare reform" was initially considered, as it would probably be driven by political concern over healthcare, but the search volume was too low to track it throughout the whole time period.

The issue of the Iraq conflict had several plausible search indices including "Iraq" and "Iraq war", but the "Iraq" search index was problematic because a large number of searches were for graphic videos of an American contractor being beheaded in Iraq and pictures of the abuse from Abu Ghraib. These types of searches were therefore not plausible as a measure of concern over the Iraq war, especially as an extremely high proportion of the Iraq searches in the year where the beheading incident took place were for these videos. It is unlikely that most of the concern for the Iraq war was related to that one incident that year.

The third issue, dissatisfaction with government,[21] was difficult to measure because it was not obvious which search terms reflect a generalized concern about government. There were high levels of MIP responses for dissatisfaction with government under both George W. Bush and Barack Obama, but it seems unlikely that the respondents would use similar terms for concern about each administration. The keyword "big government" was examined and appeared conceivable (although it is generally more related to conservative concerns over government), but the time series included a structural break in September

[21] Note that this issue describes those MIP responses which talk about dissatisfaction with government rather than explicit government satisfaction questions.

2009 following the opening of a conservative political website at *www.biggovernment.com* (Quigley 2009).

Internet search measures still possess the useful properties of MIP. Since search data are used to predict MIP values, the resultant time series can be used to measure the relative importance of issues over time as well as the dynamics of a single issue's salience. MIP is plausible as a measure of issue salience and is derived from a broadly representative sample of the population. As such, the close fit of Internet search measures to MIP indicates that they are also tracking issue salience in the population.

1.6 Conclusion

This paper set out to assess whether Internet search data could provide measures of issue salience, which would be desirable because they are available more frequently and in more countries than existing MIP measures. To achieve this aim it was necessary to 1) create a process for assessing search data validity and 2) establish whether this process was able to identify search indices that were valid measures of issue salience within the US between 2004 and 2010. Out of the seven issues that were salient over the time period, four issues (immigration, macroeconomics, fuel prices and terrorism) were able to be measured using search data. This conclusion was reached after showing that the search indices were generated from qualitatively plausible search terms (content validity) and that they closely tracked the most widely used issue salience measure (criterion validity). The correlations between the Internet search measure and the MIP measure were greater than 0.8 for all four issues. In addition, the search measures were shown, by using out-of-sample testing, to be reliable outside of the time period to which they were fitted. These findings mean that the strength of the relationships between search indices and MIP measures is of a similar magnitude to that found between the two most popular measures of issue salience, namely MIP and MII (Jennings, & Wlezien 2011). Indeed, the correlation of 0.86 between "jobs" searches and macroeconomic responses is considerably higher than the 0.44 correlation found between MIP and MII responses on the economy (Jennings, & Wlezien 2011).[22]

[22]It should be noted that the comparison of MIP and MII was over a considerably longer period (1977-2001) possibly making the composition of the economy category more heterogeneous over the twenty-four year period compared to the six year period studied here. Jennings and Wlezien also point to possible differences in coding criteria for these categories between Gallup and Ipsos-MORI. Nevertheless, it is

These findings have a number of implications for future research. As a result of the greater frequency of data, it is possible to test more complex hypotheses in agenda-setting. For instance, it should be possible to test whether it takes extended periods of media coverage to change public perceptions or if there is a reciprocal relationship between media coverage and salience. Establishing any of these relationships requires frequent observations of issue salience and time series techniques, both of which assume regular intervals between observations. Furthermore, search data has frequent and regular observations, whereas most MIP series do not. More work will be needed to test how well Google Trends data work in other countries and languages. It remains to be seen whether the same kinds of issues are easily measured across countries and whether similar terms work across borders. Countries with lower and changing Internet usage may be harder to validate Google data in as might smaller countries where there will be a lower total volume of searches.

The tests employed here have some limitations. The ideal content validity test would survey Internet users directly about what their intentions were when searching with a particular term. However, this would be impractical (it would require that we take a sample of people who use a particular search term) and expensive. The ideal criterion validity test would validate the Internet data against surveys conducted continuously and aggregated at the same weekly level. Finally, an ideal test would validate against a measure without the problems that have often been cited for MIP (see Wlezien (2005) for some important examples). However, if the available data fulfilled these requirements it would obviate the need for using Google Trends data in the first place. Given the available validation data, the method employed here is a good substitute for these ideal tests. If better survey data (either in terms of availability or conceptual strength) on issue becomes available, then the validation approach here can be replicated against this benchmark.

The results also highlight the importance of testing the validity of Internet search measures. The salience of four out of seven issues was able to be measured, but many potential search measures were rejected because they either had problematic time series properties or because an examination of the search terms within them showed that they were unlikely to reflect the importance of an issue to the person searching. Picking a

striking that the relationship between MII and MIP would be weaker than that between MIP and Internet searches.

plausible sounding search term and examining its trends is insufficient to make claims about the general public. For example, Graefe, & Armstrong (2012) attempted to measure the relative salience of the Iraq war and economic problems using searches for "Iraq" and "economy", respectively. However both of these were shown to lack content validity. The Iraq searches are dominated by searches for beheading videos and the economy searches largely refer to fuel economy.[23]

Where validity can be established for search indices they offer many advantages for a researcher. However, there are several limitations that need to be considered when using these measures. The first limitation is that it may not always be possible to measure all the important issues in the time period. In this study, citizens' concern over three issues that were mentioned by a substantial number of Gallup respondents could not be measured with internet data: dissatisfaction with government, healthcare and the Iraq conflict. It could be the case that because these issues are very broad in scope, people have difficulty translating their concerns into specific Google search terms. However, a more plausible explanation is that these issues generate a wide variety of searches that are difficult for researchers to identify. For example, the importance of healthcare (as measured through MIP) broadly follows the political conversation at the federal level (with peaks during presidential elections and the push to pass the Patient Protection and Affordable Care Act). However, the public discourse on healthcare changed substantially from 2004 to 2010 with terms such as 'death panels' and 'Obamacare' being used around the Affordable Care Act but not in earlier debates about healthcare. The difficulty in measuring the importance of healthcare is probably not because people are struggling to search for information on it, but because the search terms they use are very time specific.

The second limitation is that content validity could become threatened in future by new, confounding search terms, which might invalidate some measures but would be detected easily through repeating the content validity tests to look for new confounding terms.

Finally, although search data is more widely available than survey data, it is still restricted by the need for MIP survey data against which to validate the new measures. Nonetheless, this validation could be achieved with fewer MIP data points than the seventy-

[23]These search terms were also tested for criterion validity in order to check whether Graefe and Armstrong's findings are likely to be affected. Unsurprisingly, given the content validity results, both of these series fail criterion validity tests as well.

five available in this study. This limitation will become less important over time, as even countries where only a few MIP surveys are conducted each year will have enough data for validity testing, as the time period jointly covered by internet search data and surveys increases. Indeed, the benefits are potentially greater for these areas, as researchers may be able to move from a biannual time series to a weekly one.

Using Internet search data might seem complex due to the many ways in which its validity must be assessed. Nevertheless, not using Internet search data limits the scope of problems that can feasibly be studied, as the most important problem can change quickly and processes of agenda-setting may vary across countries. The current situation of infrequent data originating from a small group of countries limits causal inferences and the generalizability of any results. Measures of issue salience derived from Internet searches will allow for a more detailed and wider examination of agenda setting.

Chapter 2

Where and When Can We use Google Trends to Measure Issue Salience

2.1 Introduction

Google search data have several major advantages over traditional survey data. First, the high costs of running frequent surveys mean that most survey questions are asked only occasionally making comparisons over time difficult. By contrast, Google Trends provides information on search trends measured weekly. Second, there are many countries where surveys are only conducted sporadically, whereas Google search data are available anywhere in the world where sufficient numbers of people use its search engine. The Google Trends website allows researchers to download data for almost all countries at no cost, and to download time series of any search term's popularity over time (provided enough people have searched for it). For these reasons, Google Trends is an attractive data source for social scientists.

Despite these benefits, researchers using search data face a challenge because the meaning of a Google Trends series is ambiguous: an Internet search is not an expressive act, as answering survey responses are, but a practical one intended to achieve a goal. The aim of answering a survey question is to express one's view, whereas the aim of a search is to find

some information. Intuitively, a greater proportion of survey respondents answering that the economy is the most important problem facing the country reflects that more people hold this view.[1] A greater number of people searching for the word "economy" on Google might reflect students reviewing for economics exams, drivers looking for information on fuel economy, or the public's concern with macroeconomic conditions (or some combination of these). Without validating Google Trends against such expressive data, we cannot be sure that these aggregates reflect the underlying attitudes we are interested in. Past research has validated particular search terms against US survey data but there has been little work examining whether these validations hold outside the US context. This article expands the validation approach to two more cases: the United Kingdom and Spain and explores whether similar search terms can effectively capture public opinion trends in the three cases (the United States, United Kingdom, and Spain).

This article argues both that there is a huge potential for using search data to capture trends in public opinion at a fine-grained level and that there is a huge potential to draw false conclusions if the data are not validated. This argument is based on the success in finding good proxies for the salience of five issues in the United Kingdom and two in Spain using Internet search data, on the one hand, and the high proportion of initially plausible search terms that had to be discarded, on the other.

Previous research on Internet data have identified several problems with using Internet data to make claims about the general population. The first problem is representativeness. In most countries, strong differences in Internet usage exists between demographic groups particularly in age and socioeconomic status (Fuchs 2009). Previous research on Google Trends in the United States suggests that this problem probably does not bias results as much as might be feared (Mellon 2014), although it is an open question whether results in lower-usage countries would be as robust (for a discussion of some of these problems see the article by Koehler-Derrick (2013)).

A second potential problem is that the meaning of search data is not as self-evident as with survey data (Mellon 2014). When a representative sample of the population is asked "what is the most important problem facing your country?", the results will show us

[1] There are also issues with how respondents construct their answers to survey questions but the results of a survey at least reflect the proportion of a population who are willing to express a particular opinion, even if there is a further question of what underlying attitude this answer actually represents.

what proportion of people chose a particular issue as most important at that time. The interpretation is complicated by how respondents actually understand "importance" and "problem" but the results have a clear interpretation. Google Trends are constructed by counting the number of searches that include a particular word or phrase. Consequently, a high number of searches for "Iraq" might reflect the political concern over the conflict but it is also possible that Iraq's soccer team is playing important matches that week. To avoid these problems, we need to check the actual searches reflected in each Google Trends (and hopefully remove the offending terms using a Boolean search).

The third potential problem is that, even if a set of searches are plausibly related to the issue we are trying to track, the patterns in searches may not track general concern about the issue. Any number of factors may affect the probability of an individual translating concern about an issue into Google searches: strength of partisanship, political interest, political knowledge, or strength of interest all could play a role in the link between individual concern and search behavior. These biases were probably observable in the 2012 Republican primaries where Google searches for Ron Paul outstripped those of any other candidate despite the fact that he failed to win a single primary contest (see figure 2.1 (Google 2012)). There may be scientific value in analyzing trends for candidates,[2] but the link between them is clearly more complex than a one-to-one relationship with poll results.

Here is an analogy to the problem of unvalidated search data: we are handed a large set of surveys tracking interest in a particular issue over time, but we are not told what population had been sampled or how it was weighted. Without knowing these two factors, we consider if it could be a national representative survey, a survey of *New York Times* readers, or a survey oversampling people with exceptionally high interest in politics. Each of these populations are potentially interesting but we would want to know which of them we are looking at and would be skeptical of research that did not make the distinction.

Returning to Google data, suppose we want to look at whether public interest in an issue predicts policy announcements from government officials. We might correlate searches about that issue with the announcements and find that they have some predictive power. Depending on whether the searches track concern in the general public or the political

[2] If the searches reflect enthusiasm for the candidate, searches could even potentially be a predictor of the potential pool of campaign activists in an area.

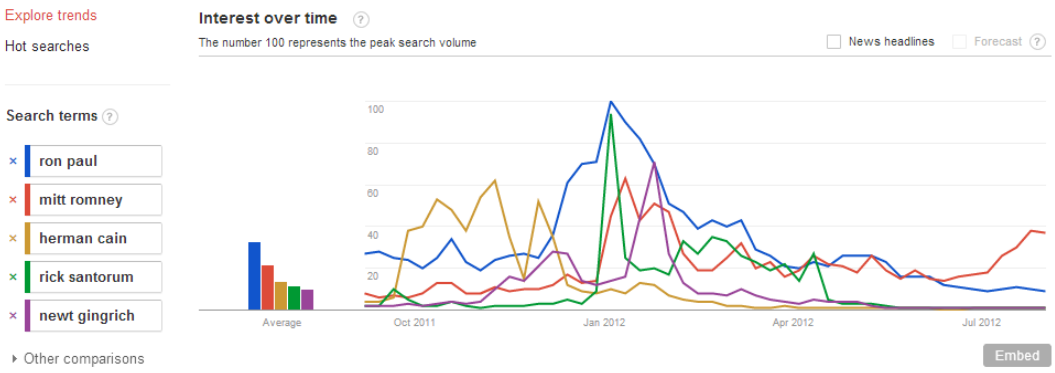


Figure 2.1: The Google Trends of the 2012 Republican Primary Candidates

elite, we should draw different substantive conclusions about the ability of the public to influence policy makers. Unless we test which population the searches are tracking, we cannot use the results to answer our question.

2.2 Previous Approaches to Using Google Trends Data

Previous research has tried to overcome these ambiguities by comparing Google Trends' data against time series of survey results from pollsters such as Gallup (Mellon 2014; Scheitle 2011). These articles focused on measuring issue salience using Google Trends. This approach led to the successful measurement of the salience of several issues in the United States with high correlations between Google Trends and the salience of several issues (Mellon 2014). Weekly issue salience series could then be generated by predicting the proportion of respondents picking an issue based on the level of Google searches.

However, the results also demonstrated the danger of not validating the Google data: many plausible Google Trends either turned out not to refer to the issue they were trying to capture or did not correlate with the Gallup trends. For instance, searches for "Iraq War" did not work well as a proxy for concern over the conflict because a high proportion of the searches were for videos of an American contractor being beheaded (Mellon 2014), which seems to reflect macabre curiosity more than political concern. Other plausible terms simply did not correlate strongly with the concern of issues measured by Gallup.

This validation approach has obvious drawbacks as it requires a source of survey data to use for validation. This excludes the use of Google data in many of the contexts in which it would be most useful, including countries where it is hard to conduct survey

research and where there is insufficient survey data to use for validation purposes. This means that Internet data are likely to be most useful in the countries where we already have relatively abundant data. An alternative justification for a search term's validity is if some issues can be consistently measured by particular search terms across a variety of contexts. As an example, in the United States, the Google Trends for "jobs" strongly tracks the salience of the issue of "unemployment" (Mellon 2014). If this finding were replicated across many countries we could reasonably use searches for "jobs," or the local equivalent, to track unemployment concern, even where we do not have long time series of survey data to use for validation.

2.3 Validating Google Data Cross-Nationally

This article applies the validation approach to two more countries: the United Kingdom and Spain. This approach has two purposes: first, to discover which search terms can be used as issue salience proxies in the two countries to allow a more fine- grained analysis of temporal patterns in issue salience and second, to explore any regularities in the terms that are valid across countries (such as the "jobs" example in the preceding text). If these regularities exist, we could potentially use search data in contexts without surveys to use for validation. If the regularities are not clear enough to use searches without validation, the analysis helps researchers decide whether search data is likely to be useful by indicating which issues are likely to be measurable and which initial search terms they should try.

In this article, the analysis process closely follows the validation process used in previous work on Google Trends (Mellon 2014). The search terms are tested against survey measurements of issue salience to assess their value as proxies. Each candidate Google Trends series is sequentially tested in four stages: content validity, stationarity, and significance in predicting the relevant issue salience time series (derived from survey results) both with, and without, a time trend. Trends are only retained at each stage if they pass the previous stage.

The two countries in this article, Spain and the United Kingdom, provide an interesting contrast in their patterns of Internet usage.[3] A 2011 Eurobarometer survey showed that

[3]A point of similarity between the cases is the market share of Google, which dominates in both markets with 92% of searches in the United Kingdom and 97% in Spain between August 2008 and July

74% of households in the United Kingdom had Internet access, whereas slightly more than half of Spanish households (52%) had access (Eurobarometer 2012).

The levels of penetration were much lower, however, at the start of the period, 2005, with 50% of UK households and just 28% of Spanish households having Internet access (Eurobarometer 2006).

These cases also provide an interesting test of the level of Internet usage required to use Internet search data (i.e., whether Google Trends data can track any issues when Internet penetration is low).[4]

Survey data from each country are compared with the Google Trends results: in the United Kingdom, a collection of Ipsos Mori polls asking "what do you see as the most important issues facing Britain today?" (Ipsos Mori 2012), and in Spain, Centro de Investigaciones Sociológicas surveys asking for up to three major problems facing Spain *Centro de Investigaciones Sociológicas.*Time series are created using the survey results for each salient issue in this period, showing what proportion of respondents choose that issue each month. The Google data are available from 2004 to the present but are of poor quality in the first year both in Spain and the United Kingdom, so 2004 is excluded, and the time series run from 2005 to 2012.

For each issue, search terms are chosen that might plausibly track concern about the issue. This task is considerably more difficult for some issues than others. It is relatively easy to think of terms that might relate to environmental issues ("global warming" and "climate change") but much more difficult to come up with plausible terms for concern over the "political class" in Spain. To generate Spanish search terms, I asked native speakers which words would be used by someone who was concerned about each issue. Many of these possible terms were discarded as there were insufficient numbers of searches for Google to track the volumes weekly. Several Spanish terms such as "inmigración ilegal" and "inseguridad ciudadana" (for public security) could not be measured weekly.

The first check on the Google search series determines whether the searches correspond to the issue. This content-validity check looks at the top searches that include the search

2012 according to StatCounter (2012)

[4]Ideally I would examine the effect of Internet penetration directly. However the increase in Internet penetration was almost perfectly linear across this time period so would be indistinguishable from any other time trend. Future research comparing periods of increases in Internet usage with steadier levels of penetration once the market is saturated should be better able to disentangle these effects.

term of interest. This strategy avoids cases such as the "Iraq War" example in the preceding text, where most searches were for politically less-relevant terms. When there are irrelevant searches, we can try to remove them by using a "-" sign in the search string (e.g., "Iraq War –beheading"). If nearly all searches including a word are irrelevant, then the search term is rejected as this indicates the word is primarily being used in a different way (as happened with the Iraq example).

Before the comparison can be made with the survey data time-series properties that can obscure relationships (seasonality) or lead to spurious correlations (nonstationarity) must be accounted for (Mellon 2014).[5] Google Trends that showed seasonal variation[6] were adjusted using Seasonal Trend Decomposition by Loess (STL) (Cleveland et al. 1990). Nonstationary time series (that do not have the same generative mechanism over time) often show spurious correlations with one another. Consequently, several tests need to be performed before the correlations between the series can be tested.[7].

Finally, the remaining searches are correlated against the time series of most important problem responses for the appropriate issue. This correlation is done in two stages: without and with a time trend. First, it is important that Google Trends has predictive power in a bivariate relationship. Because we cannot guarantee that any linear trend will continue outside of the time period, the search series should be able to stand in as a proxy by itself. However, it is also important that the strength of the relationship between the two time series is not simply due to each series' correlation with time. For trends that show a significant association, the strength of the relationship (measured by the R^2 from an ordinary least squares regression between the two time series) can be used to assess how strong a proxy the Google Trends series will be for measuring that issue's salience.

[5] See Wooldridge (2009) for an overview of potential time series problems.

[6] Seasonality is detected using the X-12 seasonality diagnostics(Findley et al. 1998).

[7] Time series are accepted either if they are both integrated order zero or if either is integrated order one and the combination passes the Phillips-Ouliaris cointegration test. These criteria are the same as in Mellon (2014) but the KPSS test is used for stationarity tests instead of the Augmented Dickey-Fuller tests as the null hypothesis of nonstationarity is more reasonable for these data. The method also differs slightly in allowing series to be tested that have different orders of integration (provided they have a cointegrating vector). This is a point where approaches differ but it makes relatively little substantive difference as most time-series of different orders of integration have low correlations as they reflect different generative processes. For a fuller discussion of the problems non-stationarity causes for inference see Granger, & Newbold (1974) and Wooldridge (2009)

2.4 Results

The results of this process are shown in table 2.1. The first column (search term) shows the keyword used to generate the Google Trends search, and the second column shows the issue it is intended to measure. Black circles show whether a Google Trends series passes the relevant testing stage (or if it needed to be adjusted for seasonality) and white circles indicate that it failed the test. Finally, the goodness of fit ($R^2 = 0.56$) is shown between the issue salience series and Google Trends series. After plausible and trackable terms were chosen, content validity was checked by looking at terms terms used in each Google Trend. Several problematic terms emerged. In the United Kingdom, results for the search term "jobs" included searches for "Steve Jobs," as well as ones related to finding employment.

A similar problem occurred for Spain when looking at searches for the Basque separatist group "ETA.". The word "eta" means "and" in Basque. As a result, a Basque nongovernmental organization supporting children with cancer, *Pirritx, Porrotx eta Marmimotots*, appears high in the searches.

In both of these cases, the confounding term can be removed using the Google Trends interface by excluding searches including the terms "Steve" and "Pirritx," respectively. In other cases, too large a proportion of the searches are confounding to use the trend. In Spain, Google Trends for "corrupción" mainly consist of searches for "corrupción en Miami," the Spanish name of the television show *Miami Vice*. Similarly, searches for 'EU,' intended to measure the British concern over the European Union instead mainly involve searches for the EU *World of Warcraft* server. In total, four Spanish and eight British search terms are excluded in the content validity stage.

The next stage checked for time-series problems. None of the issue salience time series examined showed evidence of seasonality, but several of the Google Trends showed considerable seasonal variation. Searches for 'jobs' in the United Kingdom, shown in figure 2.2 show particularly strong seasonal shifts. Panel a shows the unadjusted Google Trends data. Panel b shows the seasonal component extracted using STL. Panel c shows the smoothed trend and panel d shows the remainder component. The trend and remainder components are recombined to create a seasonally adjusted series. Thirteen of the British trends showed seasonality and were adjusted. Surprisingly, none of the Spanish terms

United Kingdom

Search Term	Issue	Content Valid	Seasonally Adjusted	Stationarity Tests+	Significant*	Significant (including time)*	R^2
job seekers	Economic Issues	●	●	●	●	●	0.84
job seekers allowance	Economic Issues	●	●	●	●	●	0.82
fuel prices	Fuel prices	●	○	●	●	●	0.67
education	Education	●	●	●	●	●	0.61
diesel price	Fuel prices	●	○	●	●	●	0.58
fuel	Fuel prices	●	○	●	●	●	0.57
petrol prices	Fuel prices	●	○	●	●	●	0.53
pensions	Pensions	●	○	●	●	●	0.41
petrol	Fuel prices	●	○	●	●	●	0.41
afghanistan war	Defence Issues	●	○	●	●	●	0.28
global warming	Environment	●	●	●	●	●	0.27
taxes	Taxes	●	●	●	●	●	0.21
illegal immigrants	Immigration	●	○	●	●	●	0.06
report crime	Crime	●	○	●	●	●	0.04
iraq war	Defence Issues	●	●	●	●	○	
ofsted	Education	●	●	●	●	○	
tax cuts	Taxes	●	○	●	●	○	
eu regulation	European Union	●	○	●	○		
crime	Crime	●	○	●	○		
british troops	Defence Issues	●	○	●	○		
housing list	Housing	●	○	●	○		
council housing	Housing	●	●	●	○		
inflation	Inflation	●	●	●	○		
winter fuel allowance	Pensions	●	○	●	○		
pollution	Environment	●	●	●	○		
illegal immigrant	Immigration	●	○	●	○		
illegal immigration	Immigration	●	○	●	○		
tax	Tax	●	●	●	○		
unemployment	Economic Issues	●	●	●	○		
uk economy	Economic Issues	●	●	●	○		
knife crime	Crime	●	○	○			
eu	European Union	○					
crimes	Crime	○					
criminals	Crime	○					
environment	Environment	○					
litter	Environment	○					
diesel	Fuel prices	○					
economy	Economic Issues	○					
Number of Trends	**39**	**31**	**13**	**30**	**17**	**14**	

Spain

Search Term	Issue	Content Valid	Seasonally Adjusted	Stationarity Tests	Significant*	Significant (including time)*	R^2
paro	Economic Issues	●	○	●	●	●	0.83
desempleo	Economic Issues	●	○	●	●	●	0.8
trabajo	Economic Issues	●	○	●	●	●	0.53
trabajos	Economic Issues	●	○	●	●	●	0.51
inmigrante	immigration	●	○	●	●	●	0.47
inmigración	immigration	●	○	●	●	○	
vivienda	housing	●	○	●	●	○	
precio de vivienda	housing	●	○	●	●	○	
delincuencia	insecurity	●	○	●	●	○	
ETA	terrorism	●	○	●	○		
vivienda social	housing	●	○	●	○		
educación	education	●	○	○			
seguridad	insecurity	○					
salud	health	○					
sudaca	immigration	○					
Number of Trends	**14**	**12**	**0**	**11**	**9**	**5**	

+ Missing months in the UK MIP series were imputed for the stationarity and criterion validitty tests by taking the mean of the two adjacent observations.

* Significance tested at the 10% level using Newey-West robust standard errors to account for autocorrelation and heteroskedasticity

Table 2.1: Google Trends Issue Salience Validity Results for the United Kingdom and Spain

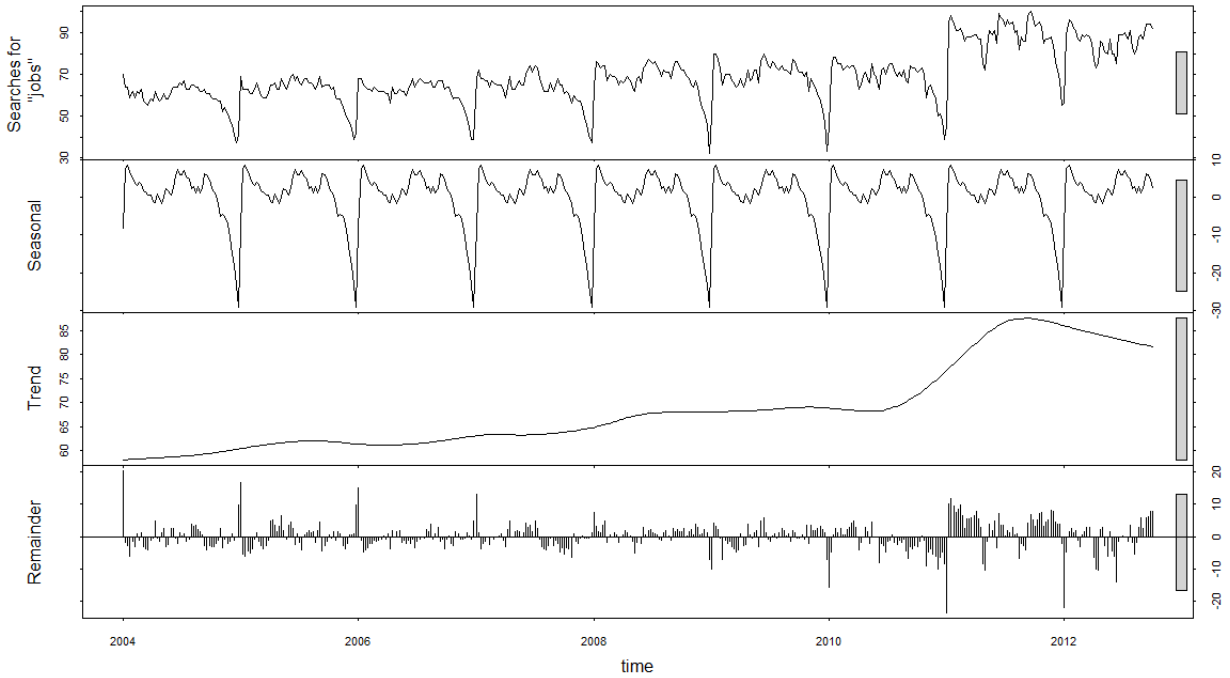


Figure 2.2: Seasonal Decomposition of Google Searches for 'jobs' in the United Kingdom, showing the original series seasonal, trend, and remainder components.

showed significant seasonal variation. One Spanish and two British trends were excluded after the stationarity tests.

After the time-series problems have been accounted for, it is possible to interpret the association between the series can be interpreted. Thirteen British terms were excluded in the bivariate test, excluding the remaining terms for the issue of housing. One of the terms excluded was "illegal immigration" (a term that showed a strong relationship with immigration concern in the United States). Four additional terms were no longer significant after a time trend was included in the regression.

Two Spanish terms were not significant in the bivariate test, and an additional four were excluded after including a time trend. These exclusions include searches for "ETA," which were not a significant predictor of the salience of terrorism.

In total, 25 out of 39 UK search trends were excluded in these steps, leaving 14 search trends measuring nine issues. Ten of the 15 initial Spanish search trends were excluded, leaving five terms covering two issues. This result partially reflects the stringent nature of

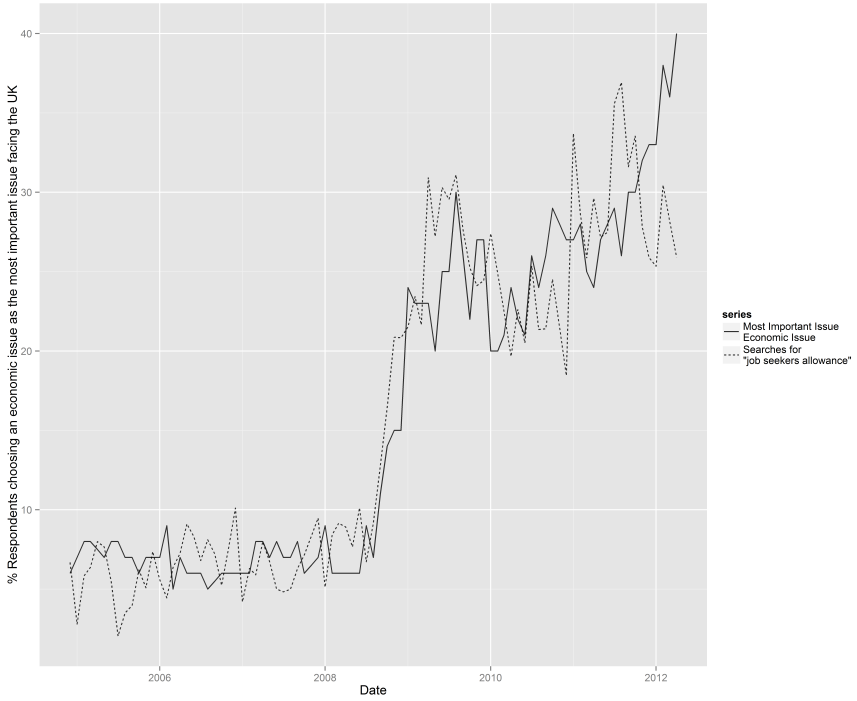


Figure 2.3: UK searches for 'job seekers allowance' (rescaled) compared to the percentage of respondents citing an economic issue as the most important issue.

the tests— longer time periods may show that some of the rejected terms were valid.[8] Nevertheless, the high level of attrition should concern researchers using search data without validation.

2.4.1 For which issues do Google Trends provides good proxies?

Although the remaining search terms in the final column of table 2.1 all show significant relationships with the relevant issue, their utility as issue salience proxies is dependent on the strength of this relationship.

In both Spain and the United Kingdom, economic issues were the most easily tracked using Google searches. Two terms, "job seekers" and "job seekers allowance" (the British term for unemployment benefits), were strong predictors of economic concern, with the latter searches explaining 86% of the variance in concern. The comparison of Google Trends for "job seekers allowance" and the Ipsos Mori time series for concern over economic issues is plotted in figure 2.3.

[8] In particular, the searches rejected after including a time trend in the regression may turn out to be valid once the trend flattens out. Nevertheless, it seems reasonable to require a proxy to function significantly better than a linear time trend.

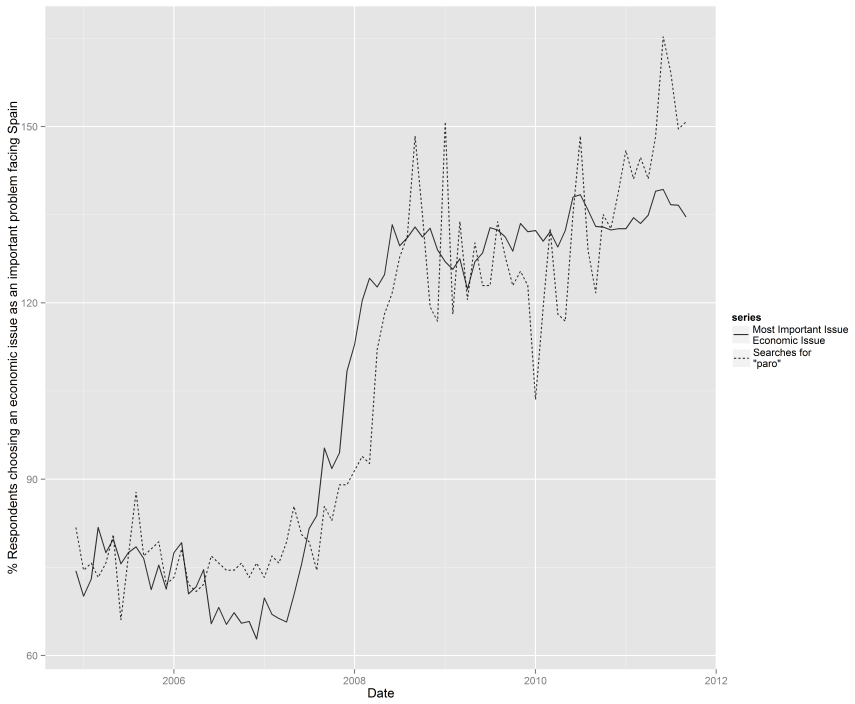


Figure 2.4: Searches for 'paro' in Spain (rescaled) compared to the percentage of respondents citing economic issues as an important problem.

In Spain, the results were similarly strong for economic issues. Four different search terms were related to economic concern. The strongest relationship was "paro" (unemployment) that explained 83% of the variance in concern about economic issues. The comparison between the two time series is shown in figure 2.4. An alternative translation for unemployment, 'desempleo', explained a similar proportion of variance ($R^2 = 0.80$). Searches for 'trabajo' (work) and 'trabajos' (jobs) each also explained more than half the variance.

In addition to economic issues, several other search terms showed substantial relationships to issue salience series. Several UK search trends were strong predictors for the public's concern about fuel prices. The term "fuel prices" was most effective, explaining 67% of the variance in concern. However, four other terms were also strong predictors with each explaining between 40% and 60% of the variance. The comparison using searches for 'fuel prices' is shown in figure 2.5. A combination search that combined searches for any of these five terms was also tested. However, the combined index performed considerably less well than the best of the individual terms with a goodness of fit $R^2 = 0.56$.

In the United Kingdom, searches for "education" also showed a strong relationship with

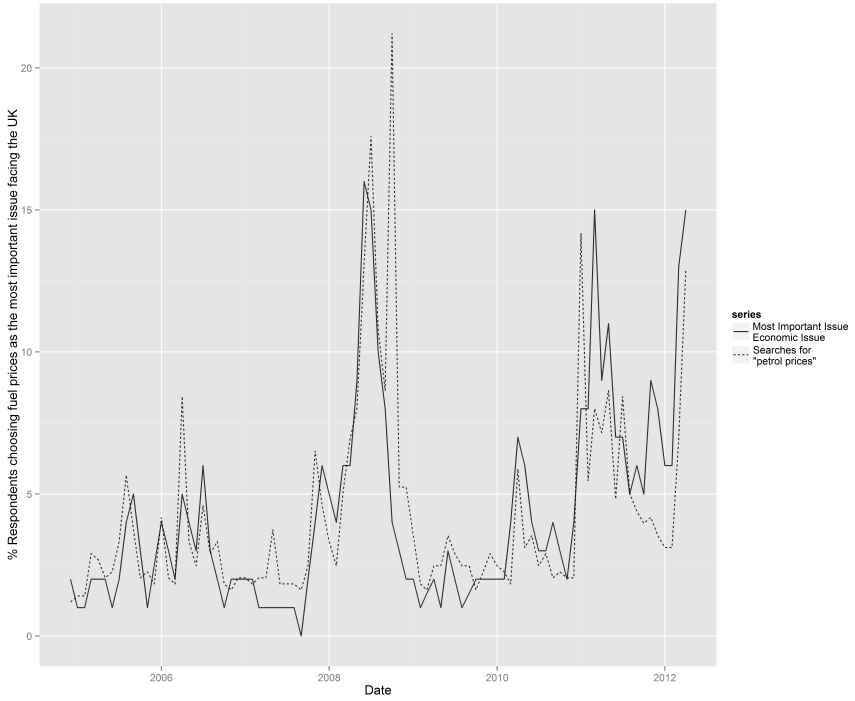


Figure 2.5: UK searches for 'fuel prices' (rescaled) compared to the percentage of respondents citing fuel prices as the most important issue.

concern about that issue, explaining 61% of the variance.

Another two Google Trends, 'pensions' and 'global warming' showed strong relationships with the survey results for each issue (R^2 of 0.41 and 0.27, respectively) with the relevant issue but probably not strong enough to justify its use as a proxy in empirical work.

Finally, four terms, 'Afghanistan war', 'taxes', 'pollution', and 'illegal immigrants' and 'report crime' are significant but the percentage of variance explained is low.[9]

Only one noneconomic issue was able to be tracked in Spain. Searches for "inmigrante" explained close to half the variance in concern about immigration ($R^2 = 0.47$). A visual comparison of the time series (see figure 2.6) shows that the 'inmigrante' series is particularly noisy at the start of the period, possibly because of a relatively low level of overall searches in Spain. It is possible, then, that the underlying relationship between these series

[9]Oddly, the coefficient for 'Afghanistan War' was actually negative when compared to defence issues, although the coefficient is positive when a time trend is included. For these reasons it would be inadvisable to use Afghanistan War as a proxy for overall concern with defence issues. Unfortunately, *Ipsos Mori* do not break out concern over Afghanistan and Iraq separately as it could potentially be a stronger predictor there. A combination search trend using 'Iraq War' or 'Afghanistan War' searches had a coefficient in the right direction but was not significant when including a time trend.

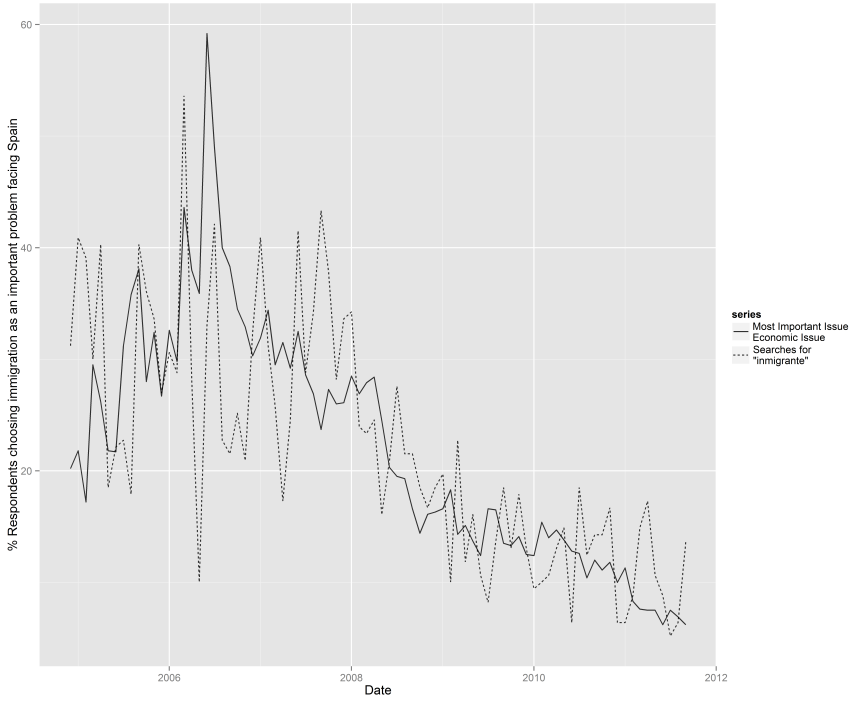


Figure 2.6: Searches for 'inmigrante' in Spain (rescaled) compared to the percentage of respondents citing immigration as an important problem.

may actually be stronger, but a longer time period is needed before this can be tested. Also, note that some of the rejected Spanish terms may be valid (particularly those rejected after including a time trend), but the tests conducted did not reject the claim that any relationship might be spurious. More issues are likely to be captured as the time period increases; many terms were rejected because of low search frequency in the first few years.

These results show progress toward the first goal of finer-grained issue salience measurement in the United Kingdom and Spain. Weekly time series of concern about five issues– economic issues, fuel prices, education, the environment, and pensions– can be generated for the United Kingdom, although the accuracy of these proxies varies considerably. Similarly, for Spain, weekly time series of concern about the economy and immigration now can be generated. But what about the second goal, to look for generalizations about which search terms are likely to be valid?

2.5 Why do *these* search terms work?

In the trends examined here, issue importance links to the frequency of searches for a particular term in two ways. First, people can be directly affected by a problem and increasingly search for practical information. Second, people can be interested in the issue and seek further information. Overall, searching for practical information seems to be the more common motivation behind the valid Google searches in this article. Perhaps practical searches are more related to issue salience or that it is easier to identify useful practical search terms than ones that reflect the searcher's interest without being obscured by irrelevant factors. Although there are similar motivations behind searches in each country (and previous results in the United States), the similarities are not sufficient to identify terms that would work in other countries without surveys to use for validation.

In the United Kingdom and Spain, economic issues are clearly the most easily tracked using Google Trends as shown in previous work on the United States (Mellon 2014). In each of these cases, the best predictors of concern with the economy are terms that relate to people taking actions as a result of economic conditions (signing up for unemployment benefits and looking for jobs) rather than terms related to general interest in the economy as an issue (for instance, searches for "UK economy" are not significantly related to economic concern). The success of practically motivated searches fits with previous successful uses of Google data in the medical literature. For instance, Google's Flu Trends (Ginsberg et al. 2008) rely on practically motivated searches for flu symptoms and cures. In contrast, the specific aspects of the economic system in each country mean that the precise terms that work still vary. In both Spain and the United States, the word for "jobs" showed a substantial relationship with economic concern. This suggests that concern with economic issues might be generally captured with these searches. In the United States "jobs" searches were by far the strongest predictor of concern about unemployment. However, in Spain much more variance is explained with "paro" than "trabajos" and "jobs" is not a significant predictor in the United Kingdom after controlling for a linear time trend. All of these results make sense, but picking a term without the aid of survey data would likely lead to a much poorer proxy.

There are also similarities in the other trackable issues. In the United States and

the United Kingdom, the salience of fuel prices can be tracked successfully using searches related to price searches: "fuel prices" and "diesel prices," respectively. This result seems to reflect a similar mechanism to the searches for jobs tracking unemployment concern: when fuel prices are of greater concern, more people will look for information about the price of fuel.

Note that "petrol prices" were a substantially weaker proxy than "fuel prices" or "diesel prices" in the United Kingdom because petrol is the more common fuel in Britain (Evans 2012). Unfortunately it was not possible to test this issue in Spain as the surveys do not categorise concern over fuel prices separately.

Immigration concern could be tracked well in both Spain and the United States. In both cases, the search terms that worked reflected an intention to find information about the issue rather than practical steps to deal with it. One term significantly predicted concern over immigration in the United Kingdom, but the degree of explanatory power was trivial, despite the issue's importance on the agenda during this time. One reason for this result may be that an important term in the British discourse on immigration, "asylum seekers," , was rejected at the content- validity stage as it included too many searches looking for information related to asylum charities such as "asylum aid" and "refugee council." This example reinforces the point that Google Trends can be driven by multiple motivations, especially if terminology is used by both proponents and opponents of a contentious issue.

This discussion shows that, for economic trends, some similarities emerge between the terms that work across countries. However, finding the best term to capture economic concern still requires validation against survey data. The success of some Google Trends in Spain also suggests that even relatively low levels of Internet penetration can generate search data that closely match survey measures for some issues.

The differences in terms used in each country's discourse around immigration, the environment, and terrorism mean that different searches reflect the concern with these issues. We still need to validate Google Trends against existing data to make reasonable inferences about public opinion. Although many of the tens of thousands of possible Google Trends, undoubtedly track interesting social phenomena, many more will track nothing at all. Without a strong argument that a suggested Google Trend is in the former, rather than latter, category it is brave to conclude much from the patterns in Internet searches.

One other possible explanation for which search terms work is the level of Internet penetration in the two countries that may affect the trackable searches. Given that fewer than 30% of Spanish households had access to Internet at the start of the study, many searches had to be rejected prior to any testing because of insufficient volume.

Although relatively few countries have sufficient survey data to validate against at present, these data will increase over time. Google provides data from as far back as 2004, so even surveys conducted twice a year will give a sample of 18 to test the Google Trends against, and as the period covered by Google increases, search data will be testable in more countries. Rather than replacing the traditional survey, using Internet data may require researchers to do more survey work, particularly in hard to reach areas. Paradoxically researchers may only be able to unlock the potential of the Internet data being generated today, in retrospect, when they have enough data generated with conventional survey tools to test the validity of search data.

Chapter 3

Validating Survey Translation Using Corpus Linguistics

3.1 Introduction

Social science research has paid increasing attention to the problems of the non-equivalence of survey questions asked across countries, cultures and languages (Heath, Fisher, & Smith 2005). In order to address these problems, recent literature has developed new statistical methods for detecting and correcting differential item functioning (where the relationship between the latent variable of interest and the indicators of it differ across contexts) in existing surveys (Stegmueller 2011; Pérez 2011; Nye et al. 2008). More systematic standards for pre-testing and translating surveys have also been developed (Harkness 2003; Harkness, Villar, & Edwards 2010; Willis et al. 2010). However, there has been little academic work on detecting the confounding variables introduced by survey translation in existing surveys. As these confounding variables can potentially undermine the validity of empirical research, this has left a gap, where we know little about the quality of many of the longest running and most widely used cross-national surveys.

Survey translation has been recognized as a potential cause of bias since the first multinational surveys were developed (Verba 1971). However, it is difficult for researchers to assess whether translation problems have an impact on their results, unless they are fluent in all the languages used within the survey. Without a way to directly assess the quality of translation, researchers may not know whether translation idiosyncracies constitute a

major source of bias or are simply a minor inconvenience.

One form of translation bias occurs when an important word in a question has different meanings across languages. Even if these differences are relatively subtle, the concepts that are tapped into in different languages could lead to confounding variables being introduced in some versions of the survey, but not others. This biases both population averages and estimates of individual level relationships involving the variable of interest. Even in those surveys translated using best practice, the closest word in the target language to the original word in the source language may still not be equivalent, as a word may have connotations and associations in one language, that it lacks in another.

One way to detect these differing meanings is to use corpus linguistics. Corpus linguistics takes very large collections of text from a language (between 80 and 300 million words for each of ten languages in this paper), then returns the words most closely associated with the word of interest. By comparing these associations *across* languages, it is possible to see whether the semantic associations of a word differ systematically when it is translated.

This paper outlines a method to assess the extent to which the associations of a translated keyword differ across languages and the content of these differences. To illustrate the method, it compares the associations of the word *ideas* and its translations, as taken from a postmaterialism question in the World Values Survey (World Values Survey 2011). I first compare the associations of *ideas* qualitatively for Chinese and English.

The method is then expanded to provide a quantitative comparison of the meaning of the word *ideas* in ten languages using cluster analysis. The word *ideas* is found to have systematically different associations in Western European languages, compared with Chinese, Arabic, Polish, and Russian. The different meanings of the word *ideas* in these other languages are shown to predict the relationships of the postmaterialism item with other variables in the World Values Survey.

3.2 Meaning and Equivalence

Cross-national surveys rely on achieving equivalence of meaning within their questions across different languages and cultural contexts (Heath, Fisher, & Smith 2005). Although there are many definitions of equivalence(Johnson 1998), it is useful for this paper to define

it in terms of measurement error.

For the case of a single country, a respondent, i, gives a response, $\hat{\theta}_i$, to an item, that measures a trait θ_i, with an error component ε_i, as shown in Equation 3.1.

$$\hat{\theta}_i = \theta_i + \varepsilon_i \tag{3.1}$$

The trait corresponds to the construct we are interested in, happiness, or postmaterialism, for example. Following Bollen, we can decompose the measurement error into four further orthogonal components (Bollen 2002), as seen in Equation 3.2. These components are defined by their relationships with the trait, θ_i, and the vector of all other attributes of an individual, i: X_i. X_i includes both measured and unmeasured attributes. 1) The random component, ε_{ri} has a mean of zero and is uncorrelated with θ_i and X_i. Accidentally ticking the wrong boxes on a survey would be an example of this kind of error. 2) The specification component, ε_{si}, is correlated with θ_i and allows for non-linear relationships between the trait and measurement of it. The specification component would be important if we find that higher-income individuals are more likely to understate their income. 3) The uniform component, ε_{ui}, is a constant systematic bias that is independent of any individual characteristics or the trait of interest. An example of a uniform error would be if there is a systematic tendency for respondents to overstate their height. 4) The confounding variable component, ε_{ci} is correlated with X_i. For instance, confounding measurement error is introduced if happier people overstate their true health.

$$\varepsilon_i = \varepsilon_{ri} + \varepsilon_{si} + \varepsilon_{ui} + \varepsilon_{ci} \tag{3.2}$$

Moving beyond the case of a single country, we can define non-equivalence as one or more error components differing systematically across countries, k.[1] Within a country, ε_{uik} and ε_{sik} are not problematic for variables measured on an arbitrary scale, such as happiness or health, as the uniform bias affects all respondents' measured scores equally. However, these biases become important if they vary across surveys or countries, as the scales become non-equivalent. Response bias affecting the equivalence of ε_{uik} and ε_{sik} across countries affects population means derived from the instrument, but doesn't strongly

[1] This section refers to k as a country, but it could also refer to a language or cultural group.

impact the instrument's correlations with other variables, provided the trait's variance is still measured. This difference in uniform and specification bias is often referred to as differential item functioning (DIF) and is common in translation. For instance, Bjornskov notes that the translation of happiness questions can introduce this kind of bias: "the Danish translation 'lykkelig' is a stronger concept than the parallel notion of 'happy' in English and would arguably require more to achieve" (Bjornskov 2010).

The confounding error component, ε_{cik}, is problematic even if it is equivalent across countries, however it causes additional problems with comparison when it is not equivalent. The confounding component is non-equivalent if the correlation between the error and individual attributes varies across different countries. Bjornskov also provides a potential example of this kind of bias from the happiness literature: "the Russian and French translations of 'happy' both mean happy and lucky, implying that the word can have multiple meanings that make the scores less comparable across countries" (Bjornskov 2010). Measured differences in happiness between English and Russian speaking countries could be driven by how lucky the Russian speaking respondents perceive themselves to be.[2]

Problems with specification and uniform equivalence have been studied at length (see Section 3.3). This paper focuses on the problem of non-equivalence of confounding errors, which have been studied in less depth.

3.2.1 Checking for the Confounding Error Component's Equivalence

This paper assesses non-equivalence resulting from differences in confounding bias by looking at differences in ε_{cik} across languages. Although we can't observe the confounding error component's relationship with individual attributes directly, we can observe its relationship to the measured values, $\hat{\theta}_{ik}$ which depends on the covariance of X_{ik} with both θ_{ik} and ε_{ik}. It is not possible to test all of X_{ik} for its relationship with the error. Firstly, some attributes of a respondent are not measurable. Secondly, in any real dataset there would be a high risk of Type I errors, if all measured variables were tested as confounding variables of measurement. As a result we only consider a small subset of X_{ik}, denoted as Z_{ik}, that we have good a-priori reasons, such as mistranslation, for suspecting as a confounder. We

[2]Formal definitions of each error component and the conditions under which it can be considered equivalent are shown in Appendix 2.2.

make the assumption that the true relationship between θ_{ik} and Z_{ik} is constant across countries. This is justified, provided we only test a small number of potential confounders and don't have theoretical reasons to suspect a systematically different relationship between the trait and the confounder in different countries. Hence, any observed difference in the relationship between $\hat{\theta}_{ik}$ and Z_{ik} can be attributed to ε_{cik} not being equivalent across k.

3.3 Previous Approaches to Assessing Equivalence

Considerable attention has been paid to identifying and correcting DIF where there is a difference in the uniform or systematic response bias across groups. DIF can sometimes be corrected using anchoring vignettes (King et al. 2004). For instance if the word *healthy* had a stronger meaning in Chinese than in English we could correct for this by measuring the respondents' ratings of situations such as:

> Bob has difficulty walking more than 100 metres and sometimes misses work due to this. Using this scale from 1 - 10 where 1 is not healthy at all and 10 is completely healthy, how healthy would you say Bob is?

If the word *healthy* implies a better physical condition in English than Chinese, we would find that Chinese respondents rate Bob's health better than English respondents do. After measuring this difference, it is possible to rescale the responses about one's own health to an equivalent scale between countries (King et al. 2004). Importantly this rescaling is neutral regarding the cause of differences in uniform and systematic response bias. It works whether the bias is caused by differences in expectations of good health or a difference in the strength of the meaning of the words.

For constructs that are measured using multiple indicators, several statistical approaches have been proposed for achieving equivalence. These include multi-group confirmatory factor analysis (Nye et al. 2008), multiple indicators multiple causes (Pérez 2011) and item response theory latent variable models (Stegmueller 2011).

These approaches all aim at addressing DIF. However, none of the approaches can directly detect or correct for confounding response bias. Instead they either make zero

confounding response bias an assumption (King et al. 2004; Stegmueller 2011) or point to high inter-item correlations as evidence against its presence (Pérez 2011). These responses are understandable, as detection of confounding measurement error requires some a-priori reason to suspect a particular confounding variable in the measurement. Testing all available variables would risk finding many false positives due to multiple testing or many false negatives if an adjustment was made for multiple testing. Insofar as the assumption of no confounding response bias is false, the measurement models previously suggested are biased due to misspecification and incorrectly estimate DIF.

A commonly used assessment tool is back translation, where the translated text is re-translated (by a different translator) into the original source language so the original document can be compared to it. Harkness and Schoua-Glusberg describe back translation as a "primitive" tool that catches some errors but misses many others, as it only indirectly accesses problems in the target language (Harkness 2003). In particular, back translation cannot identify situations where a translation is as close as possible to the source but still non-equivalent. For instance, with the Russian *happy* example there may not be a better Russian translation of *happy*, so back-translation may not find a problem with the translation, even though it is bringing in the construct of luck as well.

Harkness, Villar, and Edwards describe team translations as current best practice for achieving equivalence (Harkness, Villar, & Edwards 2010). Willis et al. outline one such approach: the Translation, Review, Adjudication, Pretesting and Documentation model (TRAPD) (Willis et al. 2010). Several translators produce drafts, which a larger team then reviews. Once a draft has been agreed, pretesting through in-depth interviews with members of the target population identifies comprehensibility problems and errors in translation.

Current translation techniques such as TRAPD aim at improving the translation of a survey prior to using it in the field. However, this is not useful for researchers who use existing multinational survey data translated using other approaches. Currently, researchers must simply hope that a survey adequately conveys the intended meaning across different languages, or else avoid using any survey not translated using best practices (which are far from universally agreed upon). The latter approach is overly wasteful as this would exclude much of the social science data currently available; many questions probably convey

meaning across different languages adequately, even without implementing the gold standard. However, a researcher has no way of knowing which questions they can use without making false inferences.

Most research implicitly follows the former option, using whichever data are available, regardless of the methods used to translate the survey. This shows that there is a need for easily accessible tools that can assess translation problems in existing surveys. One possible method for detecting translation problems is corpus linguistics.

3.4 Assessing Meaning Using Corpus Linguistics

Corpus linguistics looks at the structure and meaning of language by examining large bodies of naturally generated text from a language. Corpus linguistics has been used for a wide variety of purposes.[3] Within linguistics it has been used to study almost every aspect of language usage from grammar, to vocabulary change (Baker 2011; Michel et al. 2011) to turn taking in conversation (Baldry, & Thibault 2009)[4]. Lexicographers routinely use corpus linguistics to identify the meanings of words when constructing dictionaries and assess the relative prevalence of different meanings (Atkins, & Rundell 2008). Corpus linguistics has also found uses in a number of applied settings. For example, O'Keeffe and Walsh used corpus linguistics to give feedback on instructors engaged in small group teaching (O'Keeffe, & Walsh 2012)and, in the legal realm, forensic linguistics is used to help ascertain whether a suspect wrote a particular document, based on its similarity to a corpus of their other writings (Solan, & Tiersma 2004).

This paper utilises a technique developed within corpus linguistics: collocation. Two words are collocations of each other if they are found together more often than expected by chance. A number of statistical techniques can then assess the strength and significance of the association between the two words (Dunning 1993; Thanopoulos, Fakotakis, & Kokkinakis 2002).

The context a word is found in gives insight into the way the word is used and its meaning. This approach has a strong justification in terms of how vocabulary is acquired. Although the precise method by which children acquire the meaning of words is disputed

[3] For an overview of corpus construction, analysis and applied research see O'Keeffe, & McCarthy (2010).
[4] It should be noted that specifically collected corpora are often required to address these questions.

(Bloom 2002), it is generally agreed that children infer the meaning of a word from the context it is used in (Bloom 2002). A large part of this context is derived from other words found near the word being learnt (Bloom 2002; Chaffin, Morris, & Seely 2001). Thus, if a word in one language is surrounded by a very different set of collocations to those surrounding a word in a second language, it is implausible that they mean the same thing, as the mechanisms of language acquisition would lead people acquiring the vocabulary to infer a different meaning.

For the purposes of this paper, a pair of words are collocated if they are found within 10 words of each other. This span is wider than that used in many corpus studies. There are two reasons for this choice. First, the collocations need to be comparable across languages and a shorter span makes the collocations more dependent on the syntax of a language. For instance sentence structure varies according to the language being used. A short span would bias towards finding certain parts-of-speech in the collocations. As an example, in German the verb often appears at the end of a sentence.

A longer span reduces the effect of syntax by including all of the sentence, in most cases. The second reason is that when subjects are faced with novel words, they reread and scan a wide section of text to try and determine the word's meaning (Chaffin, Morris, & Seely 2001). This suggests that words found within the wider span are likely to be relevant to the meaning of a word.

In this paper, the Leeds Collection of Internet Corpora (Sharoff 2011) is used to retrieve the top 100[5] collocations in each language, listed in order of the strength of association between the collocation and keyword as calculated by a log-likelihood ratio statistic.[6] This statistic compares whether the two words appear together more often than would be expected by the two words' frequencies in the whole corpus. This article uses corpora in 10 languages: English, German, French, Spanish, Portuguese, Russian, Polish, Chinese, Arabic, and Italian. Each corpus is collected using the same method(Sharoff 2011) and has between 80 million and 300 million words derived from between 35,000 and 40,000

[5]It might be useful to retrieve more collocations than this but this is currently the maximum allowed through the corpus interface.

[6]Alternatively, the interface can order collocations using t statistics or mutual information scores. However, mutual information scores are highly distorted by sparse matrices, listing very rare words as highly associated (Manning, Schütze, & MITCogNet 1999) and t statistics are generated using extremely unrealistic assumptions about word distributions (Manning, Schütze, & MITCogNet 1999).

documents (Sharoff 2011).

Corpora are used in translation in two ways: by translators to look at examples of typical language in a particular domain (O'Keeffe, & McCarthy 2010) and in machine translation where supervised learning models try and make inferences about the most likely translation of a text (Xuan, Li, & Tang 2012). Whereas machine translation asks the question "given these sentences in a source language, what are the most likely equivalent sentences in a target language?", this paper asks the computationally simpler question "is this particular word in the source language equivalent in meaning to a particular word in the target language?" The decision space is vastly greater for the former compared to the latter making them quite different problems to address.

In this paper, collocations are used to compare the meanings of translated words in survey questions. This is meant to identify the confounding variable component of measurement error in the survey, as discussed in Section 3.2.1.

3.5 Comparing Chinese and English Associations for a Postmaterialism Item

In the most simple case, corpus linguistics can be used to examine the differences in the meaning of a survey question across a pair of languages by comparing the collocations of important words used in each language's version of the survey question. A question from the World Values Survey on postmaterialism is used as an example. The comparison of Chinese and English collocations is shown to demonstrate the type of differences that can be picked up by this method. Postmaterialist values "emphasize self-expression and the quality of life" (Inglehart, & Welzel 2005). Inglehart argues that individuals are more likely to adopt these types of values if they do not have to worry about survival, so postmaterialism should increase as societies develop and become wealthier (Inglehart, & Welzel 2005). The survey question reads:

> In your opinion, which one of these is most important?
>
> And what would be the next most important?:
>
> - A stable economy

- Progress toward a less impersonal and more humane society
- Progress toward a society in which ideas count more than money
- The fight against crime

The first and last options are considered materialist responses and the second and third are postmaterialist. A number of words are important for a respondent's interpretation of the question. In this section I look at the word *ideas* from the third option (several other words are considered in appendix 2.4 where no additional translation problems are found). *Ideas* is an abstract word which needs to be interpreted by a respondent, in order for her to assess its value relative to *money*. As a result, it is crucial that a translation with equivalent meaning is used. There is little additional context for the word in the question, so the respondent's understanding of *ideas* are most likely driven by their pre-existing associations with the word.

Table 3.1 shows the translations of *ideas* used in the World Values Survey (column 3) along with the total number of words in each corpus (column 2), the number of times *ideas* appears in the corpus (column 4) and the percentage of the total corpus that *ideas* makes up (column 5).

Language	Corpus Size	'Ideas' Translation	'Ideas' Frequency	'Ideas' Percentage of corpus
English	181,376,006	Ideas	58,846	0.0324%
Chinese	281,660,631	精神	52,270	0.0186%
Arabic	165,674,718	الأفكار	7,938	0.0048%
French	263,727,981	Idées	82,268	0.0312%
German	255,595,925	Ideen	41,311	0.0162%
Italian	155,569,389	Idee	42,635	0.0274%
Polish	86,605,024	Ideały	2,530	0.0029%
Portuguese	129,920,567	Idéias	20,625	0.0159%
Russian	198,509,029	идеалы	5,754	0.0029%
Spanish	145,572,631	Ideas	52,766	0.0362%

Table 3.1: Details of Each Language's Corpus and the Translation of *ideas* Used for that Language

3.5.1 Translate and Clean

In order to compare the collocations of the keyword in different languages, it is necessary to translate them into a common language. This paper uses Microsoft's Bing Translator

online service (Microsoft 2011)[7] to translate each list of collocations into English,[8] allowing a direct comparison.

To check whether using machine translation of collocations affects the results, they are compared to results generated by human translation. Although there are minor differences in a few translations, the same overall patterns are observed in the linguistic analysis. This is not greatly surprising: trained and experienced humans translators can translate better than software due to their understanding of the context. However, collocation lists do not provide any context for the humans to use, so the results are comparable for different methods of translation. For the full comparison of the human and machine translated results see Appendix 2.5.

The lists are cleaned of semantically non-important function words such as pronouns and conjunctions, as these do not provide information about meaning, and are more related to the grammar of a language which would bias the comparison.

3.5.2 Differences in the Meaning of *ideas* Between Chinese and English

The collocations of *ideas* for Chinese and English are shown in Table 3.2. The number of times each collocation appears with *ideas* is shown in the first column for each language. Chinese collocations translated into the same English word are summed to represent a single collocation. The English collocations seem to fit reasonably well with how *ideas* are usually understood by native English speakers. The list mainly focuses on expression (e.g. *express*, *exchange*, and *share*) and generation (e.g. *creative*, *develop*, and *concept*). The Chinese list of collocations has a very different set of themes. The two main themes are teamwork/team spirit (*carry forward*, *dedication*, and *team*) and community (*culture*, *community*, and *national*). From this comparison it seems very unlikely that *ideas* and the Chinese translation of this word, as it is used in the World Values Survey, have equivalent

[7] Bing Translator includes an Application Programming Interface that allows it to be automatically called by R scripts which can greatly speed up the process compared to entering the lists manually. This step is likely to introduce some errors into the process as Bing Translate sometimes makes incorrect translations. However, the errors it introduces are likely to make the two lists look less similar (i.e. two equivalent words are not translated equivalently) rather than more so. By contrast, we can be reasonably confident that collocations that appear on two lists are correctly translated, as the probability of mistranslating a word into one that happens to be on a specific list is low. The substantive interpretation of the collocations works by examining collocations that appear on multiple lists so false positives are much more problematic than false negatives.

[8] English is chosen as the common language primarily for convenience purposes but also because it is the most commonly translated language, resulting in higher quality translations.

meanings.

Table 3.2 : Collocation frequencies and percentages for translations of *ideas* in English and Chinese

	English		Chinese		
	# Collocations	Collocation %	# Collocations	Collocation %	% Point Difference
about	2249	6.5%	0	0.0%	6.5%
all	1311	3.8%	3188	4.3%	0.5%
also	734	2.1%	5741	7.7%	5.6%
any	919	2.7%	0	0.0%	2.7%
book	760	2.2%	0	0.0%	2.2%
can	1613	4.7%	1428	1.9%	2.8%
come	777	2.3%	0	0.0%	2.3%
concept	432	1.3%	0	0.0%	1.3%
creative	303	0.9%	0	0.0%	0.9%
develop	574	1.7%	0	0.0%	1.7%
different	458	1.3%	0	0.0%	1.3%
exchange	459	1.3%	0	0.0%	1.3%
express	358	1.0%	0	0.0%	1.0%
find	615	1.8%	0	0.0%	1.8%
get	1022	3.0%	0	0.0%	3.0%
give	644	1.9%	1342	1.8%	0.1%
good	1308	3.8%	1991	2.7%	1.1%
great	666	1.9%	0	0.0%	1.9%
help	688	2.0%	0	0.0%	2.0%
here	553	1.6%	0	0.0%	1.6%
idea	1577	4.6%	0	0.0%	4.6%
information	651	1.9%	0	0.0%	1.9%
like	632	1.8%	0	0.0%	1.8%
make	1002	2.9%	1329	1.8%	1.1%
many	953	2.8%	0	0.0%	2.8%
more	1318	3.8%	1522	2.1%	1.8%
new	2201	6.4%	1200	1.6%	4.8%
one	1027	3.0%	0	0.0%	3.0%
out	882	2.6%	0	0.0%	2.6%
own	746	2.2%	2294	3.1%	0.9%
people	1079	3.1%	5050	6.8%	3.7%
present	481	1.4%	0	0.0%	1.4%
share	672	2.0%	0	0.0%	2.0%
student	750	2.2%	1123	1.5%	0.7%

Continued on next page

Table 3.2 – *Continued from previous page*

	English		Chinese		
	# Collocations	Collocation %	# Collocations	Collocation %	% Point Difference
think	709	2.1%	0	0.0%	2.1%
thought	374	1.1%	1008	1.4%	0.3%
time	684	2.0%	727	1.0%	1.0%
use	1163	3.4%	0	0.0%	3.4%
work	1112	3.2%	1694	2.3%	0.9%
again	0	0.0%	1501	2.0%	2.0%
body	0	0.0%	890	1.2%	1.2%
carry forward	0	0.0%	503	0.7%	0.7%
community	0	0.0%	1106	1.5%	1.5%
culture	0	0.0%	1814	2.4%	2.4%
dedication	0	0.0%	529	0.7%	0.7%
flesh	0	0.0%	791	1.1%	1.1%
humanities	0	0.0%	685	0.9%	0.9%
implementing	0	0.0%	488	0.7%	0.7%
innovation	0	0.0%	808	1.1%	1.1%
land	0	0.0%	2118	2.9%	2.9%
large	0	0.0%	1932	2.6%	2.6%
learning	0	0.0%	1557	2.1%	2.1%
let	0	0.0%	1614	2.2%	2.2%
life	0	0.0%	1797	2.4%	2.4%
material	0	0.0%	1415	1.9%	1.9%
most	0	0.0%	1275	1.7%	1.7%
national	0	0.0%	1333	1.8%	1.8%
need	0	0.0%	974	1.3%	1.3%
professional	0	0.0%	605	0.8%	0.8%
pursuit	0	0.0%	715	1.0%	1.0%
reflect	0	0.0%	858	1.2%	1.2%
said	0	0.0%	2260	3.0%	3.0%
science	0	0.0%	870	1.2%	1.2%
species	0	0.0%	5486	7.4%	7.4%
spirit	0	0.0%	3882	5.2%	5.2%
state	0	0.0%	1161	1.6%	1.6%
team	0	0.0%	753	1.0%	1.0%
training	0	0.0%	693	0.9%	0.9%

Continued on next page

Table 3.2 – *Continued from previous page*

	English		Chinese		
	# Collocations	Collocation %	# Collocations	Collocation %	% Point Difference
while	0	0.0%	2372	3.2%	3.2%
year	0	0.0%	1714	2.3%	2.3%

Table 3.2: Collocation frequencies and percentages for translations of *ideas* in English and Chinese

3.6 Comparing the Meaning of *ideas* Translations Across 10 Languages

Visually comparing two lists of collocations can be instructive for understanding the differences between two languages. However, when looking at several languages, it becomes difficult to understand the overall structure of similarities between all the languages. For 10 languages, such a comparison involves simultaneously considering 45 two-way comparisons. However, this method of comparing translations using corpus linguistics can be generalized beyond two languages. This section expands the method to compare the meaning of *ideas* across 10 languages, using cluster analysis to highlight possible cases of non-equivalent meaning. The cluster structure is interpreted semantically by looking at which collocations are shared between languages in a cluster.

3.6.1 Construct Dissimilarities Matrix

In order to analyse translation differences we first need to define difference in this context. Table 3.2 shows the collocation lists for English and Chinese. The first column for each language shows the number of times the collocation was found together with *ideas*. For instance, *new* was found 1,200 times in the Chinese Corpus and 2,201 times in English. The second column shows the percentage of the list that the collocation makes up. In this case *new* makes up 1.4% of the listed collocations of *ideas*. The final column shows the percentage point difference between the two words for each collocation: *new* was found 3.9 percentage points more frequently in the English collocation list than in the Chinese list.

The final difference score is created by summing these percentage point differences and dividing the total by 2 to give a score between one (no collocations in common) and zero (the two languages have exactly the same proportions of each collocation). In this case the percentage point differences sum to 142% giving a final dissimilarity score of 0.71.

The final difference score for two languages (a and b) is summarized in equation 3.3 where f_{xi} is the joint frequency percentage for the ith collocation on the combined list of collocations for language x, and n is the number of collocations on the combined list.

$$\frac{\sum_{i=1}^{n} |f_{ai} - f_{bi}|}{2} \tag{3.3}$$

This equation is applied to each pair of languages, and the resulting scores are entered into a dissimilarity matrix. The dissimilarities matrix for *ideas* is shown in Table 3.3.[9] This paper uses the ISO 639-1 language codes to identify languages in figures and tables. The relevant codes are English (en), German (de), French (fr), Spanish (es), Portuguese (pt), Russian (ru), Polish (pl), Chinese (zh), Arabic (ar), and Italian (it).

	ar	zh	en	fr	de	it	pl	pt	ru	es
ar	0									
zh	0.92	0								
en	0.80	0.71	0							
fr	0.87	0.84	0.70	0						
de	0.81	0.74	0.54	0.70	0					
it	0.84	0.84	0.67	0.56	0.63	0				
pl	0.88	0.99	0.90	0.94	0.89	0.92	0			
pt	0.78	0.82	0.59	0.61	0.70	0.52	0.87	0		
ru	0.86	0.86	0.81	0.77	0.79	0.81	0.73	0.83	0	
es	0.77	0.84	0.65	0.56	0.67	0.44	0.88	0.39	0.78	0

Table 3.3: Dissimilarities matrix for *ideas* translations

3.6.2 Cluster Analysis

A hierarchical clustering algorithm, DIANA, is used to analyse the dissimilarity matrix (Kaufman, & Rousseeuw 1990) to group translations with shared meanings together. All languages are initially considered as a single cluster. The language with the highest average

[9] To look at whether the translation engine affected the dissimilarities matrix, the *ideas* matrix generated by Bing Translate was correlated against that translated using Google Translate. They correlated at r = 0.96 after excluding zeroes which would have further increased the correlation showing that the translation method used is not greatly skewing the findings.

dissimilarity to all other languages is removed from the cluster, and is used as the basis for a new cluster. Each language in the initial group is then examined to see whether it is closer, on average, to the new cluster or the original. If the language is closer to the new group, it is moved into the new cluster. This process is repeated until no language in the initial group is closer to the new group than its own. The process of splitting is applied iteratively to each of the new clusters, until all clusters consist of a single language.

The cluster analysis of the translations is intended to provide a guide to where a researcher should look for potentially problematic translations, but shouldn't be seen as evidence of translation problems by itself.[10] The substantive importance of the clustering structure can be tested by looking at the collocations shared within a cluster. If the collocations shared in one cluster are qualitatively different to those in another cluster, then that division is probably meaningful in dividing the meaning of the keywords across languages.

The result of applying divisive clustering analysis to *ideas* is shown in Figure 3.1. The clustering structure shows Arabic, Chinese, Polish, and Russian are all split into their own clusters early on, although Polish and Russian are more similar. Within the Western European languages, the Germanic languages of English and German are split off first, with the Romance languages more closely clustered.

The overall picture from using these techniques on the *ideas* translations is as follows. The Western European translations bring in concepts of innovation and sharing. Polish and Russian translations bring in primarily political and moral concepts. The Chinese translation is associated with community and dedication to a goal as well as some association with mental wellness. The Arabic translation seems to be mainly negative with some religious (or anti-religious) connotations.

This conclusion is based on the analysis of the collocations shared between languages in each cluster. Firstly, we examine the clustering within the Western European languages. Columns 1 and 2 of Table 3.4 show the collocations shared within English and German, compared to those shared within French, Spanish, Portuguese, and Italian. Although the lists contain quite a few different words, the concepts they are related to are similar.

[10]The results in this paper do not seem to be sensitive to the form of hierarchical clustering used, so only those generated using the DIANA algorithm are shown.

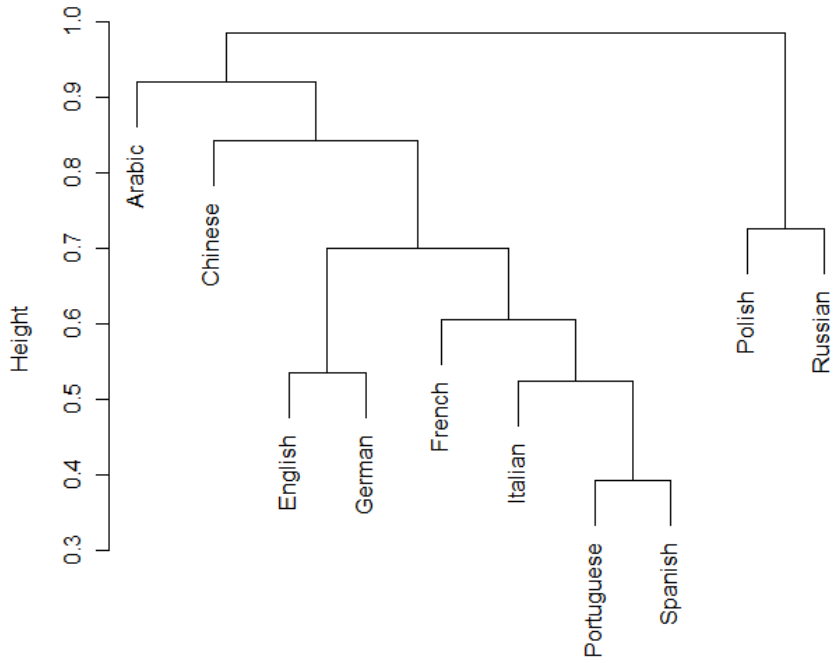


Figure 3.1: Hierarchical clustering dendogram of *ideas* dissimilarities

Both lists include many words related to creation and development (e.g. *creative*, *develop*, *express*), as well as some related to transmission (e.g. *give*, *exchange*). The split seems substantially unimportant and the two lists do not support the hypothesis that the word *ideas* means something very different between the two sets of languages.

A considerably different picture emerges when comparing the Polish and Russian cluster to the Western European cluster as a whole, shown in columns 3 and 4 of Table 3.4. While the Western European languages follow the pattern previously seen, the Polish and Russian collocations more closely relate to ideology and morality (e.g. *moral*, *justice*, *freedom*, *believe*, *Christian*).

(1) French, Spanish, Portuguese, and Italian		(2) English and German		(3) Polish and Russian		(4) English, German, Spanish, Italian, Portuguese, French	
Collocations	# languages shared in	Collocations	# languages shared in	Collocations	# languages shared in	Collocations	# languages shared in
all	4/4	about	2/2	another	2/2	all	6/6
also	4/4	all	2/2	believe	2/2	also	6/6
duty	4/4	also	2/2	can	2/2	give	6/6
express	4/4	can	2/2	christian	2/2	idea	6/6
give	4/4	come	2/2	freedom	2/2	more	6/6
idea	4/4	concept	2/2	great	2/2	new	6/6
more	4/4	creative	2/2	ideal	2/2	develop	5/6
new	4/4	develop	2/2	justice	2/2	express	5/6
power	4/4	find	2/2	man	2/2	own	5/6
say	4/4	give	2/2	moral	2/2	about	4/6
without	4/4	here	2/2	only	2/2	come	4/6
world	4/4	idea	2/2	political	2/2	concept	4/6
between	3/4	make	2/2	same	2/2	duty	4/6
develop	3/4	more	2/2	sublime	2/2	exchange	4/6
exchange	3/4	new	2/2	value	2/2	lot	4/6
go	3/4	one	2/2	what	2/2	other	4/6
lot	3/4	other	2/2	world	2/2	power	4/6
own	3/4	own	2/2			say	4/6
person	3/4	time	2/2			share	4/6
politician	3/4	what	2/2			time	4/6
share	3/4					what	4/6
thing	3/4					without	4/6
						world	4/6

Table 3.4: Collocations shared within potential clusters of languages for translations of *ideas*

Table 3.5 shows the collocations of *ideas* for Arabic. Three themes emerge from the Arabic list. There are several negative words such as *suspicious*, *sneaky*, and *misconceptions*. There are also several religious terms such as *Al-Isra* and *Miraj* (two parts of the night journey taken by the prophet Mohammed) and political terms such as *liberalism* and *leader*. These themes lead to a tentative conclusion that the Arabic translation of *ideas* is used to describe contentious ideological or religious concepts.

3.6.3 Summary of Linguistic Findings for the Meaning of *ideas* in the World Values Survey Postmaterialism Question

In light of this discussion, the important findings of the linguistic analysis are:

- In the six Western European languages, the word *ideas* reflects notions of innovation and sharing concepts;
- In Arabic, the translation of *ideas* refers more to anti-religious feelings and subversive concepts;
- In Chinese, the translation of *ideas* is related to community, dedication, and team spirit;
- The Polish and Russian, the *ideas* translations both bring in political and religious concepts.

3.6.4 Evaluation of Other Words in the Postmaterialism Question

The word *ideas* is not the only important one for the interpretation of this postmaterialism question. The following words are also checked for differences in meaning in the question: *economy*, *progress*, *impersonal*, *humane*, *fight*, and *crime*. None show a clear difference in meaning (although there are some cases where the word is not found frequently enough within the corpus to generate a stable list of collocations). For the full analysis of the other words see Appendix 2.4.

3.6.5 Summary of Method

The method in this paper can be summarized in the following steps:

Collocation	Joint Frequency with *ideas*
about	1203
Al-Isra	90
any	194
because	133
between	456
can	229
during	180
each	554
even	210
group	122
idea	242
ideas	687
including	144
leader	91
liberalism	110
life	120
lot	133
may	445
mean	197
miraj	90
misconceptions	77
model	90
more	146
new	210
non	228
nor	300
occasion	104
outlooks	90
political	112
salutes	90
scandalous	90
sneaky	90
society	127
subtract	80
suspicious	90
thinker	91
thought	100
views	154
what	817
world	146

Table 3.5: Collocations of *ideas* in Arabic from the Leeds Collection of Internet Corpora

1. A keyword within the survey question of interest is identified along with the equivalent keyword in translations of the survey question.

2. A list of words most strongly associated with a keyword (its collocations) are retrieved in each language from the Leeds Collection of Internet Corpora. These show how the keyword is used in each language.

3. The collocations are then translated into a common language using translation software (Bing Translate in this paper) (Microsoft 2011).

4. The lists of collocations are stripped of semantically unimportant words to avoid biases introduced by varying grammatical structures in different languages.

5. A dissimilarity score is calculated between each pair of languages based on how different their lists of collocations are. The scores are entered into a dissimilarity matrix.

6. The dissimilarity matrix is analyzed using cluster analysis to find which groups of languages may share a similar meaning of the keyword.

7. The clusters are interpreted semantically by looking at the collocations shared between languages within a cluster. This analysis is used to understand where potentially important differences in question meaning lie.

3.7 Survey Analysis of Associations Between the Question and Other Concepts

The linguistic analysis of the translations in this paper suggest the word *ideas* could tap into different meanings across languages. Therefore, respondents choosing the *ideas* response category may do so for different reasons depending on which language they are asked the question in. To demonstrate the potential problems that this could cause in empirical analysis, this section looks at some of the associations found between the postmaterialism item and other variables in the World Values Survey and European Values Study, across the different clusters of languages found in Section 3.6.

3.7.1 Hypotheses

The linguistic analysis, summarised in Section 3.6.3, leads to the prediction of several possible variables that could bias the measurement of postmaterialist values, i.e. the variables that are correlated with the confounding error component depending on the language (see Section 3.2).

The Chinese translation of *ideas* was found to be related to community and dedication while the Russian/Polish translations were related to political ideology. Both of these are likely to generate a positive relationship between choosing the *ideas* item in the postmaterialism question and nationalist feelings, compared to the Western European translations. To test these hypotheses, two questions from the World Values Survey are used that measure nationalist attitudes. The first asks "How proud are you to be [Your nationality]?", with the responses coded on a 4 point scale from "Not at all proud" to "Very proud" and the second asks "Of course, we all hope that there will not be another war, but if it were to come to that, would you be willing to fight for your country?" with the respondent answering yes or no.

The linguistic associations of *ideas* with community and dedication in Chinese leads to the prediction of a positive correlation between choosing the *ideas* item and traditional values. To test this hypothesis I use a question asking how similar a hypothetical person is to the respondent. The person is described as follows: "Tradition is important to this person; to follow the customs handed down by one's religion or family." The responses are measured on a six point Likert scale from "Very much like me" to "Not at all like me".

The Russian, Polish and Arabic translations of *ideas* were linked to religious concepts, with Russian/Polish referring to religious terms and Arabic referring to anti-religious terms. From this, it is predicted that choosing the *ideas* item would show a negative relationship with religious belief for Arabic respondents and a positive relationship for Russian/Polish respondents. Religious belief is measured using the question "How important is God in your life?", measured on a 10-point scale.

3.7.2 Method

The *ideas* response category is transformed to a binary variable coded as 1 if the respondent chooses the *ideas* option as either their first or second choice. Respondents were classified according to which of the four language clusters they fall into (Western European, Polish and Russian, Chinese, or Arabic). Respondents who speak other languages, or who are minority language speakers in their country, are excluded from the analysis.[11]

Logit models are fitted with the *ideas* option as the dependent variable. Each potential confounding variable is interacted with a dummy for language cluster to look for the presence of significantly different effects within each language cluster. The Western European group is used as the reference group in each model. A fixed effect is included for each survey so that country and time specific changes do not affect the results. The models also include controls for age, education[12] and gender.[13]

The models[14] are specified as follows:

$$logit(ideas) = \beta_0 + \beta_1 Confounder + \beta_2 Chinese * Confounder +$$
$$\beta_3 Arabic * Confounder + \beta_4 Polish/Russian * Confounder$$

Because the questions asked in the World Values Survey vary over time, the regression models have different numbers of observations depending on the independent variables used in the analysis. Table 2.1 in Appendix 2.3 shows the number of observations available for each of the independent variables used by language cluster.

[11] The survey language is not recorded in older World Values Surveys and European Values Studies. For these surveys, language is imputed if more than 90% of respondents in that country answer in a single language in years where language is recorded. For instance, Canadian surveys are not included for earlier years as they do not record whether they were conducted in French or English.

[12] These are entered as dummies for each age of completed education with <12 and >21 years categories.

[13] It is arguable whether demographic controls should be included in the models, as the hypotheses posit that both the confounding variable and the *ideas* item tap into the same latent variable. If a demographic variable was correlated with both the confounder and dependent variable, it is most likely because the demographic variable directly affects the latent variable. As a result, including controls in the regression may make the association less clear by reducing the variance in the underlying attitude. However, the same results are found when the demographic controls were excluded.

[14] Demographic controls and survey fixed effects not shown for clarity.

3.7.3 Results

The results for the models are shown in Table 3.6. The demographic controls are not shown.

Nationalism

For both of the nationalism questions, the interactions with Polish/Russian language cluster and Chinese language cluster are positive, showing that there is a positive relationship between nationalist concepts and the *ideas* items in these surveys compared to Western European countries. In addition, both nationalism questions also show a positive interaction with the Arabic language. This relationship was not predicted *a-priori* but may make sense in the Arab political context as secular nationalist movements are often contrasted with Islamic ones (Tibi 1997).

Traditional Values

As expected, the interaction between the importance of tradition and Chinese language is strongly positive, as can be seen in Table 3.6.

Religiosity

The Russian/Polish cluster shows a significantly positive relationship with *ideas* compared to the Western European cluster, as predicted. The confidence interval on the Arabic coefficient is too large to meaningfully test the item's relationship with religiosity. This reflects a low variance in the importance of God among Arabic speakers: over 90% of Arabic speaking respondents give a score of 10 (Very important).[15]

[15] If the Russian and Polish languages are analysed separately, the confidence intervals on the Polish interaction also become very large, so the coefficient is driven primarily by variation in the importance of God among Russian respondents. Over 50% of Polish speaking respondents respond with the highest score of importance.

	(1) Importance of tradition	(2) Willingness to fight for country	(3) Importance of God	(4) Proud of nationality
Confounder: Importance of tradition	-0.0442***			
	(0.0131)			
Confounder: Willingness to fight for country		-0.169***		
		(0.0493)		
Confounder: How important is god in your life			-0.0294**	
			(0.00914)	
Confounder: How proud of nationality				-0.213***
				(0.0321)
Interaction of confounder with Chinese	0.121***	0.330***	-0.00896	0.390***
	(0.0125)	(0.0474)	(0.00860)	(0.0318)
Interaction of confounder with Polish/Russian	-0.0280	0.175*	0.0516***	0.351***
	(0.0256)	(0.0765)	(0.0106)	(0.0552)
Interaction of confounder with Arabic	-0.0158	0.252*	0.00607	0.289**
	(0.0445)	(0.113)	(0.0368)	(0.0936)
Constant	-0.868***	-0.774***	-0.695***	-0.115
	(0.139)	(0.0672)	(0.102)	(0.121)
Observations	24964	68647	84204	83919

Standard errors in parentheses. Fixed effects for each survey, education categories, age and gender are included in all models but not shown.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6: Logit models predicting the probability of picking *ideas* option first or second using potential confounding attitudes

3.7.4 Discussion of Results

The survey analysis is largely consistent with the hypotheses derived from the linguistic analysis in Section 3.7.1. The two exceptions to this are in Arabic, where there is an unexpected positive finding for nationalism and no significant difference for religiosity. Nevertheless, these findings provide strong support for the claim that the translations introduce a confounding bias into the measurement of the postmaterialism item.

As Section 3.2.1 shows, this interpretation relies on the assumption that the true relationship between postmaterialism and each of the suspected confounders does not differ across the language clusters. It is always relatively easy to come up with *post-hoc* explanations for any finding. For example, it is conceivable that in the former-communist countries (Poland and Russia), nationalism is related to a rejection of communism, which is encouraged by postmaterialist values. It is never possible to completely dismiss these explanations but it seems unlikely that several important empirical relationships happen to fit well with observed differences in linguistic associations.

3.8 Discussion

One important question is whether the findings in this paper match the understandings of fluent speakers of the languages. To test this, I spoke to experts in each of the languages where the corpus method indicated potential equivalence problems. Russian and Polish speakers both confirmed that the words (идеалы and Ideały) were mistranslated and the words used in the survey would be better translated into the word *ideals* in English. The Chinese speaker also said that she felt the word 精神 was not the most appropriate translation for the English word *ideas*. She said that translation of 精神 is actually closer to the English word *spirit*, although there is no exact translation. The Arabic speaker said that the nuance and varied use of both the word الأفكار in Arabic and the word *ideas* in English made it difficult to say whether the words are directly equivalent. However, he did confirm that this Arabic word is the most appropriate translation for the English word *ideas*. Overall, these discussions reinforce the conclusions reached using the corpus method and indicate that the findings do reflect genuine translation problems.

To ensure that the choice of corpus does not drive the results, the analysis in this paper

is replicated, replacing the English and Russian Internet corpora with the British National Corpus (BNC Consortium 2007) and Russian National Corpus (Russian Academy of Sciences 2007), respectively. The results are not substantively different: the same clusters of meaning appear as when using Internet Corpora alone.[16]

A potential issue with the method is over-interpretation of collocation lists. It is easy to give an overly specific explanation of a word's meaning based on a list of 20 words. This is particularly problematic for small differences between collocation lists. As a result, it is probably advisable to interpret the lists in terms of a few broad categories rather than trying to divine complex definitions.

In this case study it is possible to test the hypotheses about how the confounding error component differed across countries. This is because the World Values Survey has a wide range of items to choose from. In practical settings this is not always the case. A survey may have many fewer potential questions to test for confounders with. Consequently, the linguistic analysis may need to be used without additional survey evidence to confirm the presence of confounding bias.

The technique may also be more difficult to apply to questions which give more context to the word of interest. The word *ideas* is given little context to help a respondent work out what it is referring to, but other questions give considerably more information that may change the importance of small differences in meaning across languages. It may be possible to measure the effect of context by looking at the collocations of phrases as well as individual words. However, this comes up against sample size problems if the phrasing is uncommon. Again, this concern would also be worth noting when constructing a new survey.

The Leeds Collection of Internet Corpora is a valuable resource because it allows this method to be used by researchers without the considerable effort involved in collecting corpora themselves. However, there are some limitations to using this data source, as the interface does not allow direct access to the corpus data. Collecting corpora directly would allow a wider variety of statistical analyses to be performed and a more flexible definition of

[16]These corpora are accessed through the Leeds Corpora Website. It is not clear which year's version of each corpus is used on the website but it appears to have been developed in 2007, so this is the year given for the citations.

collocation to be used.[17] For instance, an interesting alternative would be to include the full sentence (or one either side) as the collocation span instead of the 20 word span. This may further increase the relevance of the collocations to the word's meaning. Custom corpora would also allow the inclusion of more than the 100 most associated collocations which may lead to some biases, particularly given the exclusion of many collocations because they are function words.

Future work could extend the method extended by adding corpora covering more languages and regional dialects. The only requirement is that its speakers have a sufficiently large internet presence to generate a corpus with enough variety of documents. Internet corpora have been created in languages from Kiswahili (De Pauw, De Schryver, & Wagacha 2009) to Farsi (Pendar, & Sharoff 2009), so it will be possible to expand this method beyond those languages covered by the Leeds Collection of Internet Corpora.

3.9 Conclusion

This paper presents a method for detecting differing meanings of translations of important words using corpus linguistics. When these different meanings appear in a survey question they can lead to confounding variables in the measurement of the latent variable of interest, because of the way the respondents interpret what the question is asking. The meanings of a keyword are identified by looking at the most closely associated words with the keyword. By comparing these associations, clusters of translations are identified that shared similar meanings.

The case study in this paper demonstrates the ability of corpus linguistic methods to detect translation problems in important words in survey questions. In this case, the translation problems introduce substantial confounding variables in some groups of countries, as shown by the analysis of the postmaterialism item and its correlates. These correlates are predicted on the basis of differing associations of the translation of *ideas* used in the World Values Survey question. These confounding variables indicate that the confounding

[17] Ideally pre-constructed corpora would be available to download but it is not currently legal to distribute corpora due to copyright restrictions on the works contained within them. Corpora only using open source texts are available but these reflect a highly skewed sample of a language. Out-of-copyright works can also be used, but these are always many decades old and won't reflect modern language usage. For an overview of the methods used to construct Internet Corpora and examples of their usage see Hundt, Nesselhauf, & Biewer (2007).

measurement error component is non-equivalent for the postmaterialism item across different languages. This could undermine analysis using this item, leading to false inferences or masking the true relationships with other variables of interest. For instance, the attitudes driving the Chinese 'postmaterialist' responses appear to be very different to those affecting Western Europeans.

The method in this paper opens up several lines of future research. Most importantly, the translation of existing surveys should be examined and any potential implications for empirical results in social science tested. In particular, items that attempt to measure abstract concepts should be assessed for translation problems. An example of this is the literature on trust which attempts to measure a fairly elusive concept with a variety of questions.

This tool can help researchers using cross-national data to avoid confounding error and have greater confidence in the validity of the data they rely on. The process in this paper will be released as an R package that automatically implements all the steps described in Section 3.6.5, for a set of translations entered by a user. This provides a quick low cost tool for assessing translation validity before data analysis.

Another extension would be to collect larger corpora as low frequency words are currently difficult to examine using this method. The Google Ngrams project (Michel et al. 2011) could be one source of data in the future but is currently limited to seven languages and can only provide collocations one word to the left or right of a keyword, which would bias the dissimilarity scores in favour of grammatically similar languages.

The potential to systematically identify the presence of confounding bias opens up new possibilities for measurement models. It may be possible to create models which can correct for both differential item functioning and differential confounding biases and move considerably closer to measurement equivalence in survey research.

Chapter 4

What structures the flow of information between different Embassies in the US State Department?

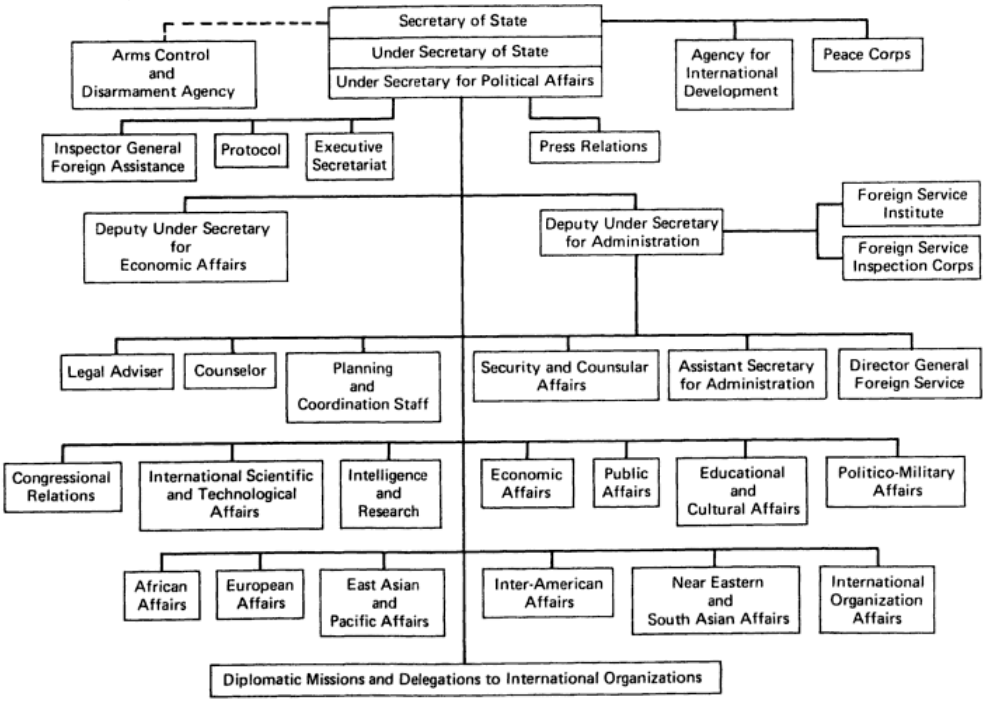
4.1 Introduction

The US State Department acts simultaneously as a representative of the US government abroad and as an information gathering operation on foreign policy. Although scholars have argued that the State Department's influence decreased throughout the 20th century, as more decision making moved into the White House and Department of Defense, it still accounts for a large proportion of US government staff stationed overseas and its political reporting is consumed heavily throughout the US foreign policy community. However, there is little academic literature on how the State Department operates internally to produce and refine this information.

This paper examines the question of how information is distributed throughout the global network of US Embassies. We look at the predictors of the magnitude of communication between pairs of Embassies in order to understand which exogenous factors are prioritised in communication. These exogenous factors include economic ties, military

disputes and migration flows between the host countries of the Embassies. For instance, international trade links the economic fates of two countries by transmitting economic shocks in one country to another. Alliances can have a direct impact of leading one country to war (so understanding the behaviour of a first country's partners may help one understand the behaviour of the first country) but they also reflect more subtle security cooperation between two countries. As sites of collective action, international organizations can strongly affect the behaviour of their members. There is also evidence that they play a role in upholding and transmitting international norms (Greenhill 2010). These types of linkages therefore potentially make information in one country relevant to an Embassy in a second country. We rely on a dataset of 300,000 diplomatic cables sent within the State Department in 1973 and 1974 to test how these various factors are weighted against each other by diplomatic offices when choosing which other actors to communicate with.

The paper first reviews the small amount of existing literature on the State Department as well as the relevant insights from the information processing literature that provides a framework for understanding how bureaucracies monitor the policy environment. We argue that exogenous variables relating to the relationships between countries are likely to be key influences on communication flows. Since the meat of the State Department's reporting is on the relationships between other countries, it is natural that it would focus on factors that affect these relationships. However, it remains to be seen which of these factors will be considered important by actors in the State Department and whether this communication agenda is stable or transitory. We then describe the State Department's communication structure and how we operationalize the communication agenda using the available archival data. In addition, we outline the factors that we use to predict communication flows and explain their potential relevance. This is followed by a discussion of the Multiple Regression Quadratic Assignment Procedure model that we use to analyse the diplomatic cables. We find that the State Department focuses on multiple factors simultaneously, but that the relative weights show substantial change over time. Importantly, these factors are almost identically weighted in the communications between Embassies and those originating from Bureaus in Washington DC, suggesting a consistent agenda between the centre and periphery of the Department.



SOURCE: *United States Government Organization Manual, 1970/71.*

Figure 4.1: Department of State Org Chart 1970/1971

4.2 Background

This section first discusses the existing literature on the State Department in general and the particular facts of the 1973-1974 time period under study. We then discuss the information processing literature and how it might be applied to the State Department given its organizational and political context. We conclude this section by drawing a set of research questions about how the State Department uses information.

4.2.1 State Department

The Department of State is split into several bureaus covering regional and policy areas. Examples include the Bureau of Economic Affairs and the Bureau of Near Eastern Affairs. In addition, the State Department incorporates the foreign service consisting of embassies, consulates and missions based in countries across the world. Figure 4.1 shows the organization chart in 1970/1971. The chart is largely similar to that in 1973 and 1974 with only some minor changes.

Communication between the Diplomatic Missions and delegations, and the Bureaus in

Washington DC are conducted via diplomatic cable.[1] While the underlying technology of this communication has changed substantially since 1973 (when the current format was established (Ewing 2005)), the format and drafting of cables has remained remarkably similar across this time. Cables are drafted by various foreign service officers within the Embassy and go through a multilayered approval process within the mission before being sent to Bureaus in Washington DC and other Embassies in the field (Department of State 2012a).

Academic analyses of the State Department

Relatively little academic literature (particularly in the last two decades) has examined the State Department as an entity as opposed to examining the larger foreign policy bureaucracy of which it is part (see Choi, James, & Kim (2013) and Rosati, & DeWitt (2012) for relevant reviews).

Much of the literature has argued that the State Department has declined in influence during the 20th century (Hook 2003; Sampas 1987; Garrison 1999; Rubin 1987). Explanations of this decline have included the loss of direct access to the president with the establishment of the National Security Council, the mistrust of the State Department from several presidents and the competition with other departments including the Department of Defense and the various intelligence services.

One recent article by Choi, James, & Kim (2013) looks at State Department's human rights reports as predictors of US military intervention. They find that the reports have predictive power for military interventions, outstripping that of media coverage of human rights violations. These same reports are used by Poe, Carey, & Vazquez (2001) to examine whether US international interests skew the human rights reporting in the State Department (compared with Amnesty International's reporting on these countries). They find that there are some discrepancies between the two reports. Drezner (2000) also examined the Human Rights Bureau and argued that it faced challenges in its analysis due to having a mission that conflicted with the rest of the department.

A number of earlier articles examined internal aspects of the State Department with

[1] This is still true today although emails and other communication methods have also become prominent (Rosati, & Scott 2010).

Scott (1969) commenting negatively on the Department's culture and Warwick (1979) analysing how organizational changes in the State Department adversely affected the morale and culture of the Foreign Service Officers.

The State Department between 1973 and 1974

In addition to these general articles, scholars have examined the 1973-1974 period in the State Department in more detail. The 1973-1974 period covers the culmination of a series of institutional conflicts between the State Department and the National Security Council (NSC). William P. Rogers was Nixon's original Secretary of State (1969-1973) and was reportedly chosen for his relative lack of experience in foreign policy (Garrison 1999). Nixon distrusted the State Department and chose instead to direct foreign policy through his National Security Advisor, Henry Kissinger, who controlled access to the President (Garrison 1999) through his leadership of the NSC.

Kissinger reportedly ran foreign policy almost exclusively, with the State Department only able to independently work on "trade and aid (though not security assistance to major countries); cultural, scientific, and information programs; UN and other international organization matters; and low-priority regions" (Rubin 1987).

However, Rubin (1987) reports that the State Department regained some influence after Rogers resigned and Kissinger took over the Secretary of State role (1973-1977) while retaining his National Security Advisor post. The balance of power reportedly shifted further still toward the State Department after Ford forced Kissinger to give up his National Security Advisor role (Kennedy 1996). However, this happpened after the time period included in this study.

Garrison (1999) argues that this domination of foreign policy by Kissinger was made possible by his decision to vet the information presented to the National Security Council in advance. This allowed Kissinger to control the agenda despite Rogers having greater access to information through his control of the State Department.

Kissinger's leadership at the State Department is said to have been considerably more centralized than Rogers', with Rubin (1987) reporting that one FSO saying that "you might as well turn the bottom six floors into a warehouse" while Kissinger led the Department. This centralization reportedly came at a cost, with strong results in areas where Kissinger

paid close attention and missed opportunities in those he was spending less time on (Rubin 1987).

Taken together, the period here provides an interesting backdrop to examine information provision in the State Department. It provides a chance to examine what were reportedly two very different leadership styles from Kissinger and Rogers.

4.2.2 Information processing

This section discusses how the State Department's reporting can be conceptualised within the literature on information processing in political institutions. Information processing is defined as the "collecting, interpreting and prioritizing signals from the environment" (Jones, & Baumgartner 2005). Information processing can be split into two components: information prioritisation (the decision about which information to analyse and pass on) and information supply (which information is gathered and sourced) (Jones, & Baumgartner 2005; Workman, Jones, & Jochim 2009).

Jones, & Baumgartner (2005) explain the process of information prioritisation. Given finite cognitive capacities, there is a limit to the number of issues that a political institution can consider at once and how much attention can be given to each. Consequently, some form of prioritisation must take place. Jones, & Baumgartner (2005) argue this leads to the neglect of some issues and a gradual drifting apart of the policy and reality on these issues, with the current policy on these neglected issues becoming increasingly inappropriate. When these issues do become salient again (often brought on by crisis), the tendency is to disproportionately process these issues leading to the rapid change in policy. As a result, policy tends to be characterised by punctuated equilibrium, with long periods of stability followed by rapid change.

Initially, issue prioritisation was the sole theme of the information processing literature, with the concept of information supply largely relegated to a secondary role. The argument used was that the policy making environment is characterised by 'information oversupply' rather than scarcity. According to Jones, & Baumgartner (2005), the main difficulty for political institutions is not finding and generating new information on an issue, but deciding which information to pay attention to, out of the very large amount available.

More recently though, scholars have argued that information supply can also play a

substantial role in information processing. In particular, the role of bureaucracies in sourcing relevant information has been discussed. Information supply can also potentially be subject to the same mechanisms of disproportionate attention as information prioritisation (Workman, Jones, & Jochim 2009). If an issue is particularly prominent, then there will be an incentive for various political actors to supply more information on the issue. Think tanks will produce reports, pollsters will survey the public on the issue and bureaucracies will direct more resources towards understanding and gathering information on the problem. In turn, this influx of new information may induce further political attention. This process constitutes a positive feedback loop that can exacerbate the tendency to disproportionately process information.[2]

Bureaucracies are able to overcome some of the limits of political attention due to their ability to process issues in parallel (Workman, Jones, & Jochim 2009). Parallel processing of issues means that different parts of a bureaucracy work on different problems (or aspects of problems) simultaneously. This gives bureaucracies the ability to source information on many issues at once and consequently have a more complete picture of the policy environment. Consequently, the whole of the federal bureaucracy can potentially be processing many more issues than Congress or the president would be able to consider at one time.

The dangers of not coordinating this processing is shown empirically by May, Workman, & Jones (2007) who analyse the case of the US Department of Homeland Security that was set up in the wake of the events of 9/11. The department became very heavily focused on terrorism preparedness despite its mandate covering other issues including natural disasters. Decision making was heavily centralised at the top of the department meaning that there was little scope for delegated authority. Since the Department heads were focused on terrorism, they paid relatively little attention to coordinating natural disaster preparation. Because decisions were centralised they could not parallel process on these issues and the natural disaster preparation became inconsistent and disorganized. This poor coordination culminated in a delayed and inappropriate response to hurricane Katrina. This example

[2]It is worth noting that this positive feedback loop is a largely short term process. In the long term, the negative feedback mechanisms of thermostatic politics discussed by Soroka2010 will tend to predominate and push policy back in line with public preferences. However, this error correction may well be sudden due to the disproportionate processing and supplying of information in the policy environment.

shows that the effectiveness of information processing in an organization is closely related to how it coordinates its actions in different parts of the organization.

Returning to the State Department, foreign policy formulation and information sourcing are also tasks that would need to be run in parallel to be effective. However, they are also tasks that require coordination between parts of the bureaucracy. The information demanded from the State Department is varied, but an important component is reporting on the intentions and interests of other states (Smith 2011). Since the interests of countries – especially on the world stage – are related to each other, reporting on the interests and concerns of one state requires understanding the actions and intentions of other relevant states. Hence, a significant portion of Embassies' information processing will require coordination with other Embassies to be effective.

Given the coordination requirements of foreign policy, we argue that the attention of the State Department should be reflected in information sharing between offices. If a factor is being heavily processed by an Embassy, then part of this processing will involve informing other Embassies of this issue. Since this information will only be shared with Embassies that have a relevant reason for interest in this issue, we would expect to see heavier communication traffic between Embassies when that factor is being heavily processed.

As a concrete example, if Embassies are heavily processing information about international trade, we would expect to see more communication between countries with strong trade ties. By contrast, if trade became less of an important issue, we might see less communication across these ties, as coordination of trade policy became less of a priority.

These are concepts we will focus on in our study based on the communications of the State Department in 1973 and 1974. As we discussed in the previous section, this period covers two quite different Secretaries of State: William P. Rogers and Henry Kissinger. Historical accounts of Kissinger's State Department suggest that he ran the organization in a very centralised fashion (Rubin 1987). It will therefore be interesting to see to what extent the State Department is able to process multiple issues at a given time or whether the centralized processing necessitates sequential processing of issues. By contrast, most accounts of Rogers' time as Secretary suggest that he was quite dependent on the Department for advice and direction (Garrison 1999). We might therefore expect to see a

wider range of issues processed but a looser link between the priorities of the Bureaus in Washington DC and those of the Embassies .

4.3 Potential factors that could induce communication flows

We focus on three research questions in this chapter.

- Firstly, we are interested in the most general question of what factors the State Department structures its communication around. In other words, what is the agenda of the State Department?
- Secondly, we consider the longitudinal aspect of the agenda. We ask whether it is processed primarily sequentially (with different issues taking precedence at different times) or in parallel (with different issues processed simultaneously)? We also consider how stable this agenda is over time. However, the relatively short time period precludes a full answer to this question.
- Finally, we ask whether the agenda of the Bureaus in Washington DC differs from that of the Embassies in the field?

Given these research questions, we need to consider what factors may induce communication flows. Intensity of communication flows may indicate which factors are being processed by embassies (because they will communicate with other embassies that are relevant on the salient issue). Potential explanations for communication flows include organizational needs for communication, or mechanisms based on findings from the International Relations and historical literature. Hence, this section contributes to our theory about the linkages between communication flows within information gathering bureaucracies and the interests and relations of the countries these bureaucracies are monitoring.

4.3.1 Geographic variables

As the descriptive plots of the State Department communication networks show in the next section, geography is an important factor in structuring communication. As Stinnett et al. (2002) point out, most conflicts and interactions on the world stage are localised, with just a few 'great powers' extending influence at greater distances. To some extent geographic

variables are likely to work indirectly through other factors that will be discussed here such as conflict, trade or migration. However, the geographic variables are useful variables to control for the correlation that each of the other variables has with proximity.

To capture geographic effects, we consider contiguity between states. We rely on the Correlates of War contiguity dataset for this purpose (Stinnett et al. 2002). There are a number of versions of contiguity we can consider and we include two in this model. First, we include the narrowest contiguity variable of sharing a land border (or a border across a river) and second, we include the widest definition of contiguity that includes borders across sea of up to 400 miles. Since we are looking only cross-sectionally at this point, we take just the contiguity data from 1974.

4.3.2 International trade

Bilateral trade is a key way in which two states become interdependent (Barbieri, Keshk, & Pollins 2009), with economic changes in one country capable of having effects on its trade partners. An economic boom in one country can strengthen its trade partners by providing a growing market for their exports. For example, trade ties have been shown to promote contagion of currency crises (Glick, & Rose 1999) and recessions (Ferrantino, & Larson 2009). Previous research has found a relationship between diplomatic representation between countries and the strength of their trade ties, suggesting that the countries themselves rate information from their trade partners as important to their interests.

Trade also occupies a significant portion of the attention of states in the international arena: with trade treaties and disputes being important events between countries. Finally, a variety of indirect effects of trade have been discussed including the promotion or prevention of terrorism and the diffusion of labour and human rights norms.

All of these effects of trade mean Embassies tasked with understanding their host country will likely need information about developments in their host country's trade partners.

International trade was one of the main areas that Rubin (1987) claims the State Department focused on while Rogers was Secretary of State. The leadership change between Rogers and Kissinger may therefore coincide with a reduced focus on trade, if the literature's claims about the balance of power between the NSC and State Department (Rubin 1987; Garrison 1999) are correct.

To measure trade, we turn to Correlates of War and their dyadic trade dataset. The values for trade between each pair of countries are derived from a number of sources, particularly the IMF's *Direction of Trade Statistics* data supplemented with data from reported national trade figures where there are missing data-points in the IMF's data. There are several caveats that are covered in the codebook (Barbieri, Keshk, & Pollins 2009) but the data seem to broadly capture the magnitudes of trade flows between various States. We use the values for 1974 and impute zero trade for dyads that are not reported or that have missing values.

4.3.3 Militarized interstate disputes

One of the most rudimentary interactions between states is war and its precursor disputes. Information transfer between Embassies in warring states clearly has value. If Embassies are trying to understand whether a dispute is likely to escalate, they need to understand the intentions, interests and perceptions of their host country but also those of the other countries involved in the dispute.

To test this hypothesis, we use the concept of militarized interstate disputes, defined by Jones, Bremer, & Singer (1996) as:

> cases of conflict in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state. Disputes are composed of incidents that range in intensity from threats to use force to actual combat.

This is operationalised using the Militarized Interstate Dispute (MID) data collection from the Correlates of War Project (Ghosn, Palmer, & Bremer 2004). We recode this dataset into a binary matrix that identifies whether two States are co-involved in any dispute in 1973 or 1974. We make the necessary but overly strong assumption that all parties to any dispute are involved in a dispute with each other.[3]

[3]The correlates of war website states that "Dyadic MID data is only available from the Correlates of War project from 1993-2001, because incident data (showing direct dyadic interaction) has only been systematically coded since 1993. For the pre-1993 period, either assumptions must be made in coding dyadic interaction (critically, that all states on opposite sides interacted), or additional research must be performed to identify actual interactions. In addition, coding to a dyadic form must carefully check dates

While this assumption would be problematic if we were trying to predict conflict, information transfer should be relevant both among collaborators in a war and antagonists. In future studies, with access to dyadic data on conflicts (as is available for later time periods), it would be useful to separate out the collaborator and antagonist effects on communication.

4.3.4 Arms trades

Arms trades capture an important component of countries' relationships because they imply a more active relationship than a formal alliance. Kapstein (2002) argues that arms trades are an important part of maintaining an alliance, both in terms of diplomatic links and the interoperability of two countries' armed forces. These deals also potentially provide information about the future threats that a country expects to face in future (i.e. if they are buying anti-aircraft equipment it may imply a different strategic view than if they are buying a shipment of small arms).

We use data on arms trades from the Stockholm International Peace Research Institute which tracks arms transfers between states (Stockholm International Peace Research 2012). Because the data do not give a precise value to each trade between states, and many transfers are active over extended periods of time, we code arms trades as a binary variable tracing whether there are any ongoing arms trades between two states in 1973 or 1974.

4.3.5 International Organizations

Intergovernmental organizations (IGOs) have important effects on the interests and behaviour of their member states. Scholars have studied several aspects of the role IGOs play in international politics. Some of these purported roles include transmitting international norms (Greenhill 2010), sanctioning norm violators (Lebovic, & Voeten 2009) and even helping states construct their understanding of their interests on the international stage (Finnemore 1996).

Given that IGOs have substantial effects on their members, understanding a country's

of state involvement to verify that states were opponents at the same time."(Ghosn, Palmer, & Bremer 2004) We take the first option here. This also means that allies who fight the same war will be coded as being involved in a dispute together. Since we are predicting salience rather than antagonism this is not unreasonable as it seems likely that there will be relevant information to transfer between offices in two allies or enemies. However, it is important to remember this definition when interpreting the results.

interests and likely actions may include understanding its relationship to IGOs. Since IGOs act on the basis of their members decisions, information from one member state will be useful for understanding the interests of all other member states.

IGOs also play a role in facilitating collective action between states (Russett, & Sullivan 2009; Abbott, & Snidal 1998; Nielson, & Tierney 2003), meaning that a diplomat's host country may take actions due to their membership of an IGO. As a result, the intentions of other member states will be useful for understanding the behaviour of the diplomat's host country.

We measure IGO comemberships using the Correlates of War IGO membership dataset (Pevehouse, Nordstrom, & Warnke 2004). We define the value of joint IGO membership between a set of States as the number of IGOs that they are both members of in 1974.

In addition to comembership, there is likely to be additional information needed by the State Department offices in countries hosting an IGO. Since diplomats in the same city as the headquarters of an IGO will be the ones directly monitoring and interacting with the IGO, they are also likely to be the most in need of information about the member states. This also applies in the other direction, where diplomats may seek information about an IGO from the State Department's offices in the country the IGO is headquartered in. On this view, offices in IGO headquarters are likely to act as central sources of information about these IGOs. This is particularly likely for those IGOs where the State Department has a dedicated delegation or office to that IGO such as NATO, the UN or the European community.

International organizations are one of the areas that Rubin (1987) reports that the State Department under William Rogers was relegated to dealing with. Consequently, we may see a greater focus on IGOs at the start of the time period before Kissinger becomes Secretary of State.

4.3.6 Alliances

Alliance between countries could potentially induce communication flows for similar reasons to IGOs (indeed some IGOs, such as NATO are also alliances). We focus on defense pacts (where one or both of the states in a dyad are required to come to the defense of the other state if it is attacked). While there are reasons to think that any type of alliance could

have an effect on communication, the defense pact is the most clear in terms of its meaning and obligations.

Because a defense pact has the potential to bind the actions of a state, information about the actions of one participant of an alliance is directly relevant to whether the second member will be involved in a conflict.

In addition to the formal aspect of defense pacts, they are also indicators of defense and strategic cooperation between states. Consequently, information about the actions and intentions of one ally is likely to be useful to understanding those of another ally. Allies also share information with each other so there may be directly applicable information about a diplomat's host country available from its allies.

Once again, the Correlates of War dataset provides a useful source of information on these agreements (Gibler 2013). We code a defense pact as present if either country in a dyad has a defense obligation to the other.

4.3.7 Migration Flows

The final issue we consider is migration. Given the impact of immigration (and to a lesser extent emigration) on domestic politics (Ceobanu, & Escandell 2010), migration is an issue that most states pay attention to. Since migration involves flows of people between countries, it is also an issue that links countries. If the salience of migration is high, then we would expect Embassies based in two countries with a high level of bilateral migration to share information.

Again, if the State Department under William Rogers is relatively marginalized on 'high politics', then migration may be an issue that we would expect to see a greater focus on at the start of the time period.

In this paper, we derive data on immigration from the World Bank's global bilateral migration database (Ozden et al. 2011). The immigration figures capture the number of people in each State who were born in each other State. The closest available data to this period is derived from 1970: three to four years earlier than this study. The data look at the immigration stock rather than immigrant flows but should nonetheless provide some indication of the movement of people between pairs of countries over the long term.

4.4 Communication in the State Department

This section describes how we measure the flow of information in the State Department using archival records of diplomatic cables.

4.4.1 The Central Foreign Policy Files

This chapter uses a relational event dataset created from Department of State internal diplomatic cables as recorded in the Central Foreign Policy Files since 1973 and stored in the National Archives and Records Administration (Burr 2006). We present the sources, formatting, and descriptives of the dataset. The relational event dataset comprises State Department electronic text records for declassified cables. These were retrieved from a private archive that had compiled them for 1973-1974 (BACM Research 2012). In the 22 months recovered (March 1973 to December 1974), nearly 300,000 cables were sent.

In addition to the full text cables, the Central Foreign Policy Files also include withdrawal cards for cables that cannot be released for various reasons.[4] In total there are an additional 82,756 withdrawn cables in this period. However, the withdrawal cards do include the metadata from the cables about where they orginate and are sent to. Cables on topics of less historical import (for instance 'Library Services' and 'Visas') are not permanently archived and so are not included in our dataset (Langbart et al. 2007).

4.4.2 The structure of a diplomatic cable

Cables can vary in length from one paragraph to several pages and may concern almost any topic of interest to foreign policy. All cables are structured similarly. Figure 5.2 shows a sample cable (a short cable was deliberately chosen for presentation reasons) with some key components highlighted: date and time (A), senders and recipients (B/C/D), tags (E), references (F), authors (G), and classification (H). We explain the parsing of the senders and recipients below.

[4]These were not included in the private archive so we wrote a text scraper in R to retrieve the records from the National Archives and Records Administration website.

PAGE 01 TEL AV 07501 260859Z

13

A ACTION SS-25

B INFO OCT-01 ISO-00 /026 W
---------- 064965

C R 251339Z SEP 73

D FM AMEMBASSY TEL AVIV

E TO SECSTATE WASHDC 9885

F INFO AMEMBASSY MOSCOW

G S E C R E T TEL AVIV 7501

H EXDIS

E O 11652: GDS

I TAGS: PFOR, MASS, UR, US, IS

J SUBJECT: TERRORIST POSSESSION OF SOVIET ORIGIN
GROUND-TO-AIR MISSILES

K REF: STATE 189495

1. I NOTE THAT DEPUTY SECRETARY INFORMED ISRAELI AMBASSADOR DINITZ SEP 10 THAT WE HAD APPROACHED SOVIETS SEVERAL TIMES REGARDING DANGERS OF STRELA MISSILE GETTING INTO HANDS OF TERRORISTS (STATE 181236). SINCE UNDER SECRETARY PORTER HAS NOW TAKEN UP THIS MATTER AGAIN WITH SOVIETS (REFTEL, I BELIEVE WE SHOULD INFORM ISRAELIS OF THIS ACTION IN INTERESTS OF REASSURING THEM OF OUR STRONG CONCERN AND, HOPEFULLY, OF REINFORCING THOSE ISRAELI LEADERS WHO WOULD COUNSEL RESTRAINT RE ANTI-TERRORIST MEASURES WHICH MAY BE UNDER CONSIDERATION. THIS NOTIFICATION, OF COURSE, COULD BE DONE EITHER HERE OR IN WASHINGTON.

2. PLEASE ADVISE.

L KEATING

SECRET

<< END OF DOCUMENT >>

Figure 4.2: Example of a cable sent from the American Embassy in Tel Aviv

4.4.3 Actors (A, B, D, E and F)

The sender of the message is defined in the FM line (D). In this case, the cable was sent by the American Embassy in Tel Aviv. If the message is sent from the State Department in Washington DC, the FM line will read SECSTATE WASHDC. However, this does not necessarily indicate that the Secretary wrote or even saw the message (Department of State 2012b). If the message was sent from the State Department, then the ACTION field (A) would be replaced with an ORIGIN field showing where in the State Department the message originated.

There are a number of fields in the cable that define its distribution list. Firstly, we have the ACTION line (A) that refers to one of the offices within the State Department. In this case the cable is sent to SS which is the Office of the Secretary of State. Additional offices within the State Department are specified on the INFO line below (B). Offices outside of the State Department are specified in the TO and INFO lines (E and F). The TO line here specifies SECSTATE WASHDC which is put on every message sent to the State Department (as with the FM line). In this case, the message is actually sent to the Secretary but this is specified in the ACTION line rather than the TO line which is just a formality. The INFO field in this message shows that the cable is being distributed laterally to the American Embassy in Moscow as well as to the State Department.

4.4.4 Operationalizing the communication structure

The communication network is defined by the number of messages passing between two nodes. So if State Department offices in Belgium send 121 messages to State Department offices in the Netherlands, then 121 is the weight assigned to the tie between the two offices. This measure assumes that it is the communication happening between the offices in the Netherlands and Belgium that is important, even if some of these communications also include other recipients.[5]

Figure 4.3 shows the communication patterns of the Embassies with each other based on this measure. In order to make the picture clearer, we only include ties where more than 25 messages are sent across the dyad and exclude isolated nodes. However, this is

[5]One alternative would be to make each event have equal weight with multiple recipients getting only a fractional edge weight.

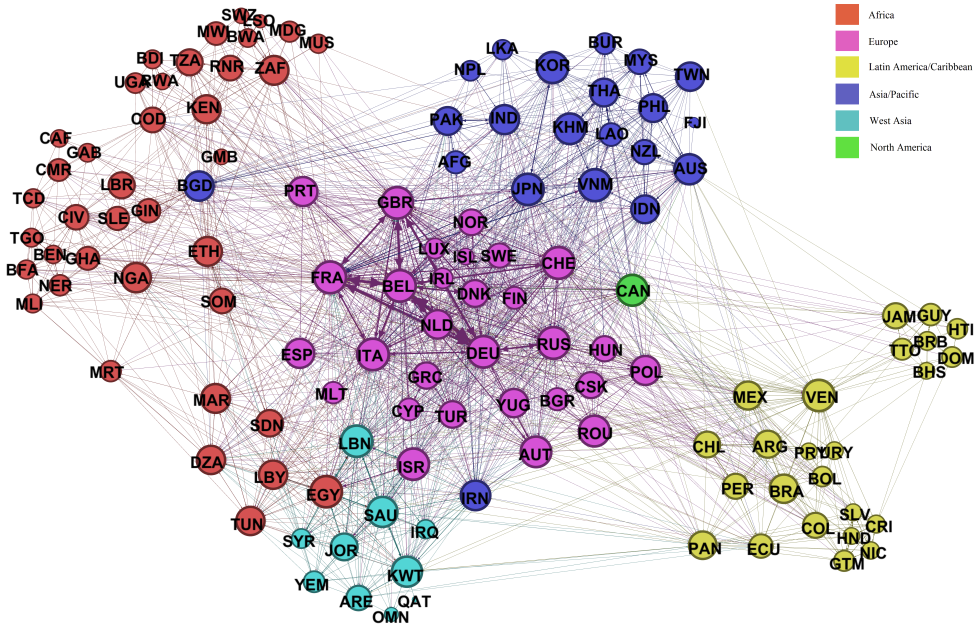


Figure 4.3: Network plot of inter-Embassy communications in the State Department

purely for visual purposes and these ties and nodes are included in our models. The nodes are sized by the total number of messages sent or received and are coloured by the region they are based in. More messages sent between embassies are indicated by thicker lines (called edges) between the country nodes.[6]

The graph shows that there is clustering of diplomatic communications within regions but that there is a substantial level of communication between regions. It should also be noted that the network is considerably more dense when the weaker ties are included as well, with much more cross-regional communication and a less clear regional structure. Overall, the plot shows that there is substantial communication between State Department offices and that this structure is not random.

In addition to modelling which factors structure inter-Embassy communication, we are also interested in looking at whether similar factors structure the communication of Bureaus to multiple Embassies. In other words, what factors lead a Bureau to include two Embassies together in a message. This is also potentially a coordinating activity, with the same information being sent to multiple subordinates.

[6]The nodes are arranged using the ForceAtlas 2 algorithm in gephi using LinLog and prevent overlap modes, an edge weight influence parameter of 0.3, scaling of 0.2 and gravity of 2.0. These parameters are chosen purely to show the general structure of the network.

Because State Department cables can have multiple recipients, we can measure which Embassies are included as recipients together by Bureaus. There is a question about the best way to weight the co-sending relationships. In a message sent to three recipients, there are three combinations of those recipients. However, in a message sent to ten recipients, there are 45 combinations of those recipients. It seems implausible that we should weight a message sent to ten recipients 15 times more importantly than one with three recipients. If we simply weight all of these edges equally, then we will be measuring the tendency of a recipient to be included in a message including many other recipients. Instead, we weight the co-sending so that each event contributes a total weight equal to the number of recipients in the message.

As an example, in the message with 10 recipients each of the 45 combinations receives a weight of 0.22 whereas in the three recipient message, each combination of those recipients receives a weight of one. This approach is consistent with our measurement of inter-embassy communication. In the case of inter-Embassy communication, a message with 5 recipients would count as one tie from the sender to each of the five recipients equalling a total of five ties.

The plot of co-sending by Bureaus is shown in figure 4.4. The overall pattern is reasonably similar to that of the Embassies' communications. Once again, there is a clear clustering by region. The plot has been simplified to aid interpretation so we should be cautious about reading in any substantive conclusions based on the comparison. Nevertheless, the plots show that co-sending does appear to be structured and hence may be able to tell us about the differences in the communication agendas of Embassies and Bureaus.

There are a number of simplifications made in the analysis for this paper. First, we are ignoring the multiple DOS offices that exist within some countries, such as Embassies, Consulates and Missions and aggregating them as single nodes. This is necessary for comparing the communication patterns with international variables between countries. However, it does leave out communications between offices within a country and simplifies the organizational structure of the DoS.

The second simplification is that we focus purely on the relationships between offices in the field. There is undoubtedly an important set of factors influencing which Embassies contact which Bureaus, but this is beyond the scope of the current paper.

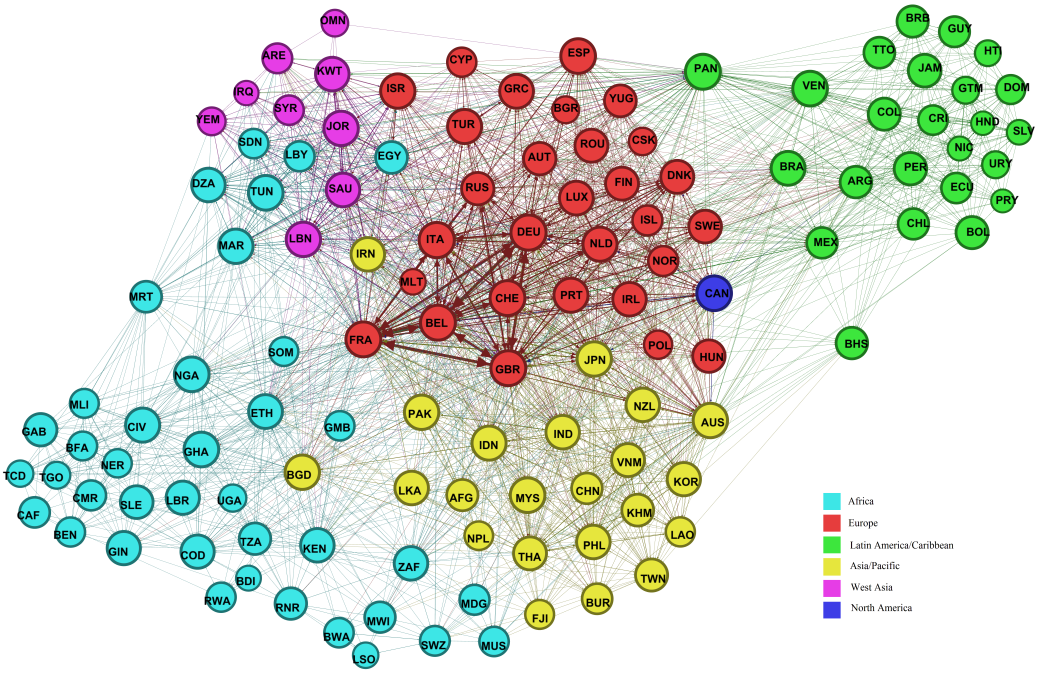


Figure 4.4: Network plot of recipients included together in messages from Bureaus in the State Department

4.5 Multiple Regression Quadratic Assignment Procedure (MRQAP)

This section discusses the MRQAP model that we use to look at which factors predict communication between Embassies. Since we are analysing relational data, we have to account for the correlation within rows and columns of an adjacency matrix. Consequently, we cannot simply use OLS regression.

The analyses here use the Multiple Regression Quadratic Assignment Procedure (MRQAP)[7] method (Dekker, Krackhardt, & Snijders 2007). At its heart, this method is an ordinary least squares (OLS) linear regression with the number of messages sent from each node to the other (the adjacency matrix) as the dependent variable, and other equally sized matrices as the independent variables.

The main analysis fits a series of MRQAP models to predict the level of communications between DoS offices in each pair of countries using the variables outlined in the previous section. We fit a second set of models using the number of times a set of embassies are

[7]We use the netlm command from the SNA package to fit the MRQAP models (Butts 2013).

included together as recipients in a message from a Bureau, in order to look at whether different types of information are distributed centrally and laterally in the DoS.

The linear model is shown in equation 4.1, where Y and X are nn matrices and β is a scalar parameter. Z is an nnq array of all the other independent variables.

$$Y = \beta_1 X_1 + ... + \beta_q X_q \tag{4.1}$$

While the parameters are fitted using OLS, the standard errors calculated would be highly biased because there is usually autocorrelation between observations in the same rows and columns of an adjacency matrix. The MRQAP approach calculates p values through a permutation test that uses the dataset to assess how often the estimated parameter would have been greater than the fitted parameter in these permutations. Because there is dependency within rows and columns the permutations simultaneously permute the labels for rows and columns. This maintains the same dependency structure within individuals in the dataset but still gives us an estimate of the natural variance of the data.

The simplest version of this permutation just permutes the dependent variable, so that each count of messages sent from one node to another is assigned randomly to a set of independent variables. However, Dekker, Krackhardt, & Snijders (2007) show that this significance test will sometimes not reject the null hypothesis at the correct rate, particularly when the independent variables are correlated.

Instead, they use a permutation test for each matrix (hereafter referred to just as X) conditional on the array of (q - 1) other matrices, referred to as Z. They suggest using a residual permutation approach: in particular their double semi-partialling approach (DSP). Dekker, Krackhardt, & Snijders (2007) show that using t-statistics with the DSP method most reliably approximates the correct type I error rate, while retaining good power. The t-statistic is derived from the partial correlation coefficient for each parameter which in turn is derived from the beta coefficient in the permuted regression model in equation 4.2 (see Dekker, Krackhardt, & Snijders (2007) for the full derivation of this test).

Z is first regressed on X and then the residuals from this model, $\hat{\epsilon}_{XZ}$ are permuted (the permuted residuals are notated as $\pi(\hat{\epsilon}_{XZ})$) and regressed on Y along with Z, as shown in equation 4.2.

$$Y = \beta\hat{\epsilon}_{XZ} + Z\gamma + E \tag{4.2}$$

The coefficient β on X is then used to define the reference distribution for the t-statistic and a p-value is calculated based on how often the permuted t-statistic is greater than the observed statistic.

Importantly, the results from a QAP model can be read like an OLS table with the parameters telling us how a one unit change in the independent variable changes the dependent variable, and the p value telling us how often we would expect to see a parameter of this magnitude under the null hypothesis.

4.5.1 Further comparisons

In addition to modelling the communications of the Embassies to each other, we also model the number of messages that Embassies are included in together by Bureaus in Washington DC. This approach allows for a comparison of the factors structuring the communication of Bureaus and Embassies.

Since the magnitude of the co-sending and communication variables are not necessarily at all directly comparable, we standardize both variables with a mean of zero and standard deviation of one.

In addition, to allow a comparison across coefficients we standardize the independent variables to also have a mean of zero and standard deviation of one. The one exception to this is binary variables which we leave untransformed.

The final comparison is of the agendas of both Bureaus and Embassies over time. To do this, we simply fit separate MRQAP models to subsets of the dataset. Once again, we standardize all the variables to allow for a comparison between time periods, predictors and the centre and periphery.

4.6 Results

This section discusses the results of the MRQAP models aimed at answering our research questions. We begin by describing the results of the inter-Embassy communication model. We then validate the results of this model by qualitatively coding the content of cables

sent across dyads predicted to contain information on a particular topic. After this, we compare the predictors of the communication agenda among Bureaus and Embassies in the State Department. Finally, we examine how each of these agendas changes over time by fitting a series of models to 35 day subsets of the data.

4.6.1 Predictors of inter-Embassy communication

Table 4.1 shows the MRQAP results predicting the level of communication flows across dyads of US Embassies.

	International Trade		**Migration**		**Security**		**International Org.**	
	Estimate	p	Estimate	p	Estimate	p	Estimate	p
Intercept	26.343	0.22	25.877	0.16	13.928	0.26	-59.182	0
Contiguity	125.799	0.01	125.459	0.03	9.187	0.7	3.927	0.84
Land border	188.037	0.01	154.61	0.01	94.313	0.08	85.432	0.01
Trade	0.568	0	0.562	0	0.526	0	0.441	0
Migration			0.126	0.01	0.133	0	0.11	0
Dispute					374.904	0	380.854	0
Alliance					182.23	0	117.924	0
Arms trade					61.975	0	-22.479	0.1
IGO							3.184	0
IGO HQ							42.31	0
R Squared	0.44		0.45		0.49		0.56	

Table 4.1: Communication QAP Models. n = 18090. Number of countries =135

The first model looks at how geographic factors and international trade influence communication flows between Embassies in the State Department. The bilateral trade variable has a large effect. We also control for whether the two countries share a land border or are contiguous at 400 miles. Taking into account these three variables, we already explain 44% of the variance in communication between US embassies. A large bulk of this variance is explained by the bilateral trade between the countries. A bivariate model with only trade already explains 43% of variance in communication.

Model 2 adds in historical migration (i.e. the number of people born in each country living in the other). This has a substantial impact with each 7932 additional migrants living in the other country associated with an additional message sent between the two embassies.

Migration marginally improves the model fit but probably does not explain that much variation because the migration matrix was fairly sparse in the 1970s when the migration stock was substantially lower than today (United Nations 2002). However, migration was still a major issue in some countries' relationships at this time such as that between India and Pakistan, or Poland and Germany.

Model 3 adds in the security related variables of arms trades, militarized interstate disputes and alliances. All three variables predict large increases in communications across dyads. In particular, the effect of disputes leads to an average of 375 additional messages sent across a dyad. Alliances similarly predict a substantially higher flow of information. The inclusion of these variables greatly reduces the strength of the geographic variables with a shared sea border only predicting 94 additional messages between US Embassies in the neighbouring countries and no longer significant at the 5% level.

The fourth model adds in factors related to IGOs: co-membership and headquarters. Both variables have strong effects. Each additional joint IGO membership predicts 3 more messages across the dyad. This makes for a substantial number of messages when we consider that the mean number of joint IGO memberships is 20 with a standard deviation of 8.7.

The IGO headquarters variable also plays a large role with Embassies sending on average 42 more messages across a dyad if the IGO that one host country is a member of has headquarters in the other Embassy's host country. This explains the centrality of

countries such as Belgium, Luxembourg and Austria in the communications of the State Department, as each of these host the headquarters of several IGOs.

The IGO variables substantially increase the fit of the model with 56% of the variance explained by the final model. The high fit of the model shows us that the variables considered here account for a substantial portion of the Embassies' agenda.

Overall, the models here show that Embassies can process information on multiple issues. Despite the claims that US foreign policy was focused primarily around grand soviet strategy and security issues in this time period (Siniver 2008), the State Department in this time period transmits far more information along dyads with high international trade and joint membership in international organizations.

4.6.2 Qualitative triangulation

The MRQAP results tell us which variables can explain variance in communication patterns. However, such correlations always need to be interpreted cautiously because of the possibility of omitted variable bias and other misspecification.

Communication datasets have an advantage over other datasets in assessing the plausibility of mechanisms: we can read the messages that are communicated across a particular dyad.

Trade and international organizations are the variables that most increase the fit of the model, so we focus on whether the cables sent across dyads with high levels of trade or high levels of IGO co-membership actually make mention of trade or IGOs.

Since it is unlikely that a communication model will perfectly explain all observations, we don't qualitatively examine cables from all dyads but only from those that are well predicted by the model. More precisely, the approach is as follows:

1. Define a set, of high (top quartile)/low (lowest quartile)/zero (variable = 0) conditions for a subset of the independent variables;

2. Sample dyads that fulfil each of the conditions in and are in the lowest quartile of residuals from the MRQAP model (so that we are sampling dyads that are well explained by the model);

3. We then read these cables and code for whether they make mention of the factor that was used to select the dyad.

4.6.3 Trade

The first set of cables we consider are sampled from high trade and low residual dyads. We read and coded 50 of these cables based on whether they mentioned trade. Three cables were missing and could not be coded. Of the remaining 47 cables, 17 explicitly discuss trade (36%) and an additional three discuss the provision of aid between two countries. Seven more cables are about general international meetings between countries, or nuclear non-proliferation. Both of these could potentially have economic and trade elements, but trade was not discussed explicitly in these cables so we coded them as not being related to trade. Although these messages about visits might seem like relatively trivial communications, interviews with diplomats suggest that the itinerary and talking points of trips are important times for Embassies to set the agenda. By saying that "We recommend you meet with these people and you try to do this" diplomats can try and influence the visitors perception of issues (Ewing 2002).

Some quotes from these cables show examples of these trade discussions.[8] For instance in cable 1973CARACA10937 the Embassy reports that:

> urging other producer countries in Latin America, notably Ecuador, to adopt nationalistic stance toward oil companies has long been element of Minmines Perez Lasalvia's policy, and was one of factors that lead him to promote creation of Latin American energy organization (OLADE) [. . .] another Perez motivation was to encourage other oil exporting nations, such as Ecuador, and potential exporters to keep their prices high enough not to undercut competitiveness of Venezuelan oil

In cable 1974BUDAPE52, sent from Hungary, the Embassy discusses the effect of exchange rates on trade between Hungary and other European countries:

[8] Note that we edit the quotes for length (as indicated) and by correcting obvious spelling errors. In addition we change the text to lower case from its original all caps. The full text of the cables and the coding are available from the authors.

> Normal conclusion that this move will have trade effect of discouraging Hungarian exports to west (especially West Germany) while encouraging western imports would not be accurate [. . .] Since prices are determined basically by western market conditions, forint/hard currency exchange rate (through associated foreign trade multiplier) is not a causal factor in pricing, nor will it have a direct effect on foreign traders.

The cable 1974ISLAMA1634, the Islamabad Embassy discusses Pakistan's drugs policy in the context of the international market for opium:

> USG and world recognition that [. . .] there is actually a substantial world shortage of licit opium obviously suggests to Pakistan the idea that it need not hasten to ban the poppy. On the contrary, GOP now seems to feel it can move to become a legal exporter of opium.

In cable 1973BAGHDA517, the Embassy discusses the strategic implications of US investment in Iraq:

> Brown and Root 122 million dollar contract for INOC deep sea terminal is breakthrough for U.S. on commercial front, and has political significance as well. Unlike Boeing a/c sale, this contract will result in influx of Americans and continuing relationship between GOI and U.S.

In each of these cases, diplomats discuss the implications of trade. The proportion of cables discussing trade is consistent with trade being an important concern of the State Department when considering these dyads.

4.6.4 Shared IGO memberships

A similar pattern is observed with IGO co-memberships. We sampled 50 cables from dyads with a high number of co-memberships in IGOs and low residuals in the MRQAP model. One cable was missing so we discarded it. Of the remaining 49 cables, 17 referred to at least one IGO (35%). One additional cable referred to the work of an non-governmental organization acting in several African countries and several more referred to international

meetings and treaties. However, we aren't able to tie these back to a specific IGO so we coded them as not referring to an IGO. The most commonly mentioned IGO is the European Community (EC). Discussion of an EC Member State's foreign policy is repeatedly discussed in the context of the country's relationship with the EC.

For instance, in cable 1973PARIS20534 on French reactions to recent exchange rate changes the Embassy writes:

> At French insistence EC is also on record that object of MTN was trade expansion and not reestablishment of balance of payments equilibrium for any given country.

Other messages report on specific meetings involving the members of an organization. For instance, the cable 1974VIENNA7903, directed to the Embassies in the various OPEC member states, discusses a dinner (hosted by the Austria foreign minister) attended by OPEC ministers:

> Bauer said OPEC ministers were cautious in their discussions of substantive issues during dinner. His own personal impression was that no major change in oil price structure expected, and that he does not anticipate 'unpleasant' surprises for balance of this year.

Again, the content of these messages provides strong support that the MRQAP models are correctly picking up on the importance of IGO memberships to Embassies when they choose recipients. The sample size used in both of these cases is relatively small but this analysis is still an important additional line of evidence supporting our interpretation of the aggregate MRQAP models.

These qualitative results give some prima facie support for our claim that information flows can represent the agenda of an organization.

4.6.5 Comparison of Bureau and Embassy agendas

The next question we are interested in is whether the agenda of the Embassies and Bureaus differs. To do this, we look at whether the predictors of Bureau co-sending are the same as those that predict inter-Embassy communication. As we described in section 4.5.1, we make

	Embassy		Bureau	
	Estimate	p	Estimate	p
Intercept	-0.034	0.27	-0.033	0.27
Contiguity	0.016	0.81	0.231	0.08
Land border	0.356	0.06	0.216	0.12
Trade	0.509	0	0.474	0
Migration	0.071	0	0.094	0
Dispute	1.587	0	1.288	0
Alliance	0.491	0	0.41	0
Arms trade	-0.094	0.06	-0.003	0.94
IGO	0.115	0	0.131	0
IGO HQ	0.248	0	0.265	0
R Squared	0.56		0.55	

Table 4.2: Comparison of foreign service communications and DoS co-sending (all variables standardized except for binary predictors)

these types of communication comparable by standardizing the communication matrices. Consequently, the coefficients in table 4.2 should be read as the effect that increasing the independent variable by one standard deviation has on the dependent variable in terms of how many standard deviations it changes by. We standardize all the continuous variables but leave binary variables untransformed. This means the coefficients on the binary variables should be read as reflecting what effect the variable has on communication in terms of standard deviations.

The fully comparable results show that there is very high similarity between the factors structuring the Bureaus' co-sending behaviour and the communication between embassies in the foreign service. At the 5% level, the same factors are significant predictors for both the Bureaus' co-sending behaviour and inter-Embassy communication. There is however, a difference in the size of the contiguity and arms trade variables between the two countries.

Despite these differences, the overall picture is one of remarkable consistency in the significance and magnitude of factors predicting the attention the Bureaus and foreign service pay to different international relationships.

4.6.6 Longitudinal model

So far we have seen that the State Department does communicate around multiple issues. However, it is still possible that these factors are actually being processed sequentially

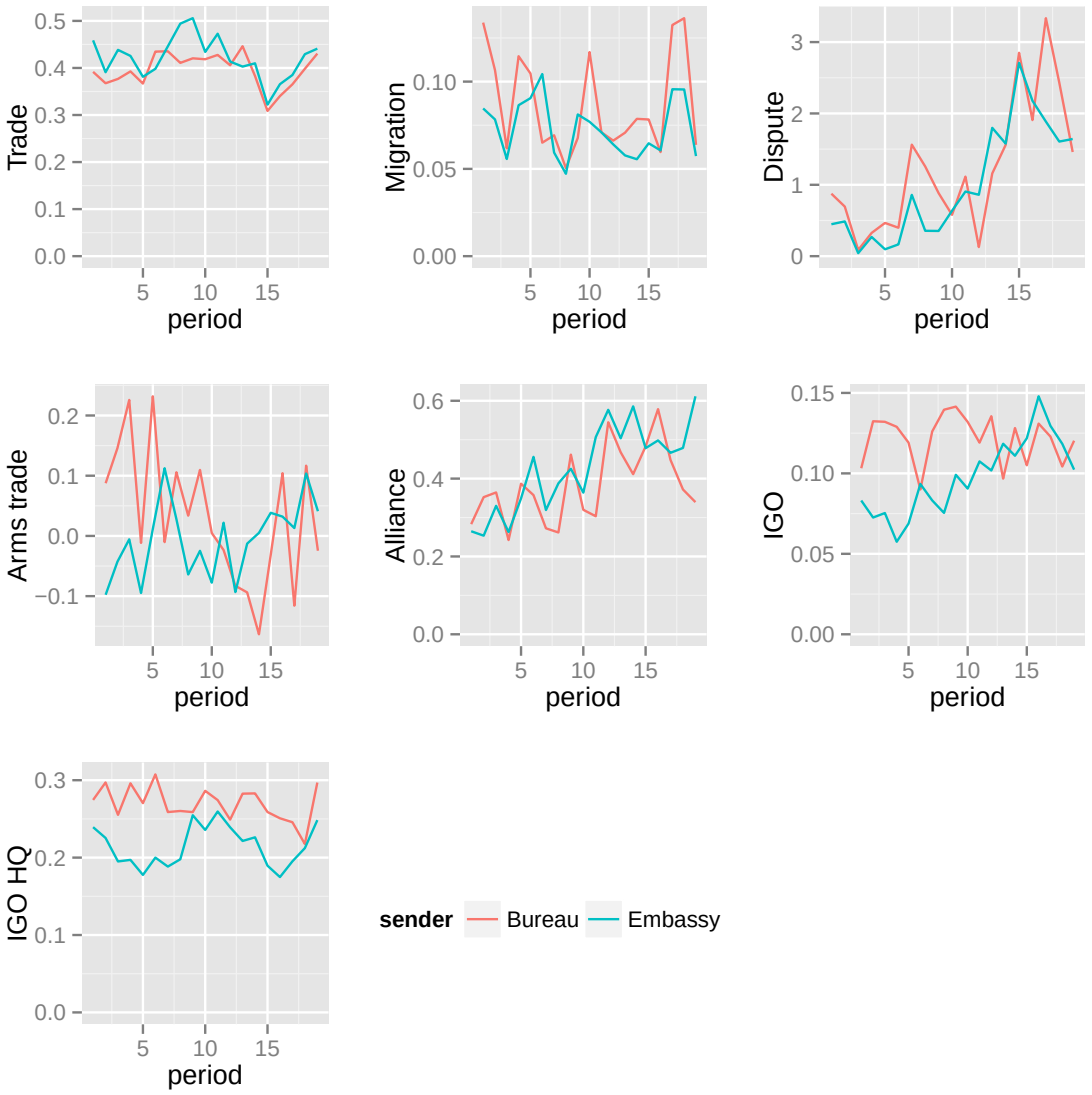


Figure 4.5: Trends in QAP parameter values over time for Bureaus and Embassies

rather being processed in parallel. To discover which type of processing is occurring, we look at whether the State Department focuses on the same issues over time in its communication patterns and whether multiple issues are important predictors at any given time.

Figure 4.5 shows the trends in the fitted parameters over nineteen 5 week periods. It shows lines both for the Embassies' communications and the co-sending of messages to Embassies by Bureaus. Because the variables in the MRQAP models were standardized, the size as well as changes in the coefficients are comparable.

These results provide further evidence for the tight link between the agendas of the

Embassies and Bureaus in the State Department. The one exception to this agreement is arms trades, where the level of attention sometimes varies substantially between Bureaus and Embassies. However, this may be due to the low number of dyads with arms trading relationships leading to noisy estimates of the parameter.

The plot also shows that there does not appear to be crowding out between different issues. Most of the issue attention is stable with only minor variation over time. One exception is military disputes. The effect of a military dispute between two countries increases communication between the US Embassies by one standard deviation at the start of the study. However this effect nearly triples by the 15th period. This is consistent with the claims of historians (Rubin 1987) that the State Department was able to focus more on high politics including military disputes under Kissinger than it had been able to under Rogers. However, the timing is not exactly right, as Kissinger took over the Secretary of State role around the sixth period considerably earlier than the spike in the salience of military disputes which occurs around the 12th period.

4.7 Discussion

We began this paper by asking how information is distributed in the State Department. After examining the information flows between Embassies and those from Bureaus to Embassies, we see that multiple issues structure this information flow. This supports the claims in the information processing literature that bureaucratization allows parallel processing on multiple issues. Among other issues, the State Department simultaneously considers the interactions of countries in IGOs, the evolving trade ties of their host countries and possible changes in alliance dynamics.

Interestingly, even at the height of the Cold War and the drawdown of the Vietnam War when one might expect security to be an overarching concern, the bulk of communication is directed along trade ties and memberships of international organizations. Although military disputes do prompt substantial communication, their relative rarity means that most communication concerns the low politics of trade and international organizations.

The issue agenda is broadly stable, but the importance of military disputes increases rapidly in 1974. There are a number of possible explanations for this change. This could

reflect the reportedly greater relevance of the State Department once Kissinger became Secretary of State (Rubin 1987). Alternatively, it may reflect the lower level of inter-departmental competition on security issues following the end of the Vietnam War. Because we use one overall measure of whether disputes occurred in a dyad at any point in 1973 or 1974, we cannot completely dismiss the possibility that the change in salience reflects events. However, this seems relatively unlikely as the two most important wars in this period, Yom Kippur and Vietnam, reduced in intensity in 1974.

The stability of most items on the agenda suggests that most factors remain similarly salient over time. A longer time period will be needed before we can definitively answer the question of whether the Department's attention conforms to the archetypal punctuated equilibrium model of information processing.

The stability of this agenda also casts doubt on some previous claims that the State Department under Kissinger paid attention only to a narrow range of issues (Rubin 1987). The changes here suggest that the Department has an equally varied agenda over time. However, it could also have been possible that, following May, Workman, & Jones (2007), the State Department hierarchy focused on a narrow range of issues while leaving the collection of information on other issues largely unfocused. However, this story is not fully consistent with the data. If a focus on a narrow set of issues led to less coordination on other issues, then we would expect to see greater agenda divergence under Kissinger. However, the agreement between the Bureaus and Embassies remains tightly coupled throughout the study.

Future research could benefit from integrating event data on conflicts into communication models in order to examine how the State Department reacts to conflict on a more disaggregated scale. However, the most comprehensive sources of this type of data start later than the time period covered in this paper (Uppsala Conflict Data Program 2013; Leetaru, & Schrodt 2013). Nonetheless, as the available State Department communications data increases, it would be useful to make this comparison. More data may be available either through further declassifications of historical data which are expected soon (personal correspondence with NARA staff, 2013) or possibly through the use of other sources of data. This might plausibly include leaked cables such as those from WikiLeaks.[9]

[9]There are some drawbacks to using the WikiLeaks data compared with the archival cables. It is

One of the most important findings of the chapter is that there is a very high level of agreement in the communication agendas of the Bureaus in Washington DC and the Embassies located across the world. This analysis cannot yet explain the reasons for this agenda agreement. There are several possible mechanisms that could lead to this level of agreement including a strong shared culture in the Department, active agenda setting by either the Embassies or Bureaus, or a more subtle process of status quo bias, once groups of Embassies begin working together. Future research should aim to disentangle the mechanisms driving this agreement.

Finally, this paper also makes a methodological contribution on agenda measurement. We show that the aggregate predictors of information flow match up to the actual content of the messages sent. This suggests that information sharing is an important element of the issue agenda within organizations. This could have value beyond the State Department. For instance, we could look at communication between branches of domestic bureaucracies, to look at which aspects of their agenda they are processing and sharing information on. Future research on bureaucracies could benefit from examining the agreement and divergence of the agenda of leadership and rank and file members of an organization.

uncertain how complete the WikiLeaks cable dataset is and modern sources report that the number of emails sent in the State Department outnumbers the numbers of cables by several orders of magnitude (Rosati, & Scott 2010). This means that the cable channel will be more biased than in the archival data.

Chapter 5

Speaking Diplomatically: Examining Agenda Setting Using Relational Event Modelling

> one fact ... is clear to anyone with experience in government: the springs of policy bubble up; they do not trickle down.

Dean Acheson, Secretary of State (1949-1953) (Hammond 1986)

> Increasingly, options sent up reflected decisions already made at the top, meaning that analysis and intelligence were being shaped to meet the preconceptions of powerful but ill-informed policymakers-perhaps the greatest single internal cause of recent U.S. foreign policy disasters.

Barry Rubin on Nixon and Kissinger's control of US foreign policy (Rubin 1987, p148)

5.1 Introduction

Who has influence over US foreign policy? Answers have included the media (Soroka 2003), the public (Soroka 2003; Wlezien 1996), business leaders (Jacobs, & Page 2005) and communities of experts (Jacobs, & Page 2005; Haas 1992; Adler 1992). Other scholars have pointed to the importance of federal bureaucracies tasked with formulating options

for foreign policy and sourcing relevant information (Kissinger 1966; Hammond 1986; Choi, James, & Kim 2013).

One line of research on foreign policy decision making has focused on the role of the president's advisors (and their staff) in setting the agenda by formulating possible policy options (Majeski, & Sylvan 1999; Garrison 1999). These advisors' proposals rely on information gathered by various foreign policy bureaucracies. So, in order to understand influence in this setting, we are faced with a parallel question of who directs and leads on the collection of this information?

One important source for foreign policy information is the US State Department. The network of Embassies, Consulates and other diplomatic missions send hundreds of thousands of messages a year reporting and analysing developments on the world stage. While the State Department is not the only source of foreign policy information, it is well placed to report on many international issues.

However, we know relatively little about how information is sourced in the State Department. There are differing views about this topic. These views are best summarized by the two quotes at the top of this paper. The Acheson quote expresses the view that the building blocks of State Department policy come fundamentally from the bottom up. By contrast, the alternative view expressed above by Rubin (1987), is that the leadership in the executive branch can —at least in this case —control the flow of information and the sourcing process.

A key test between these views is whether the Embassies and missions are generally the source of the State Department's agenda or instead act in accordance with the agenda defined by the Bureaus in Washington DC.

When analyzing this question, it is important to make the distinction between the influence of the State Department and the influence of the Secretary of State. It is possible that the Secretary of State could have little bargaining power with the presidents' other foreign policy advisors but that the flow of information out from the Embassies still set the agenda for the debates within the White House.

In order to answer these questions, it is first necessary to measure the issue agenda within the State Department (i.e. which issues it focuses on). As Pump (2011) points out, operationalizing the agenda has been a challenge in policy research on agenda setting,

due to the difficulty of assessing which sources give an accurate picture of an actor's prioritisation of different issues. While there are numerous aspects of the State Department agenda, we focus on the structure of information sharing between Embassies as this tells us which other offices they are collaborating with and hence which ties are considered as important. Since processing foreign policy issues requires collaboration between Embassies, an office's choice of recipients tells us which ties are being utilised and therefore which issues are processed efficiently by these offices.

This means that we are fundamentally interested in the question of which actors determine which Embassies work together and share information. Is this decided endogenously by the Embassies themselves? Or is the structure of communication and collaboration issued from the Bureaus in Washington DC?

In this paper, we use a dataset of approximately 300,000 diplomatic cables sent between Embassies and Bureaus in the State Department during 1973 and 1974. To disentangle the influence of Embassies and Bureaus on each other, we focus not just on the overall flow of information between each office, but also on the sequence of communication events that constitute this flow of information. An emerging class of social network analysis models, relational event models, look at exactly this type of question. These models take the unit of analysis to be each individual communication event and statistically model the series of choices that an actor makes during that communication, dependent on what other events have happened recently. For instance, we can see whether factors related to the messages an Embassy has received make the Embassy more likely to subsequently contact particular other Embassies.

However, we show that existing relational event models have not taken into account the dependency between communication events, namely that messages are part of longer conversations. To address this shortcoming, we define a new Relational Event Conversation History Actor Initiated Network model (RECHAIN), that sees each message in a conversation as taking the previous message as its starting point. Under this model, an actor deciding to send a message first decides which (if any) conversations she wants to continue. Once these are chosen she decides whether to remove existing recipients or add additional ones.

In applying this model to the State Department's communications, we reach a num-

ber of findings. We first show that our assumption that there is dependence between the recipients of messages in a conversation does characterise communications in the State Department. Consequently, actors who start conversations have substantial communication agenda setting power. We show that the Bureaus' previous messages are cited disproportionately often compared to how often they are sent. This suggests that Bureaus are able to set the agenda through starting conversations.

Surprisingly , we also find that only Bureaus choose recipients on the basis of the trade ties between host countries. Embassies do not pay attention to trade links in their choices of recipients. However, since both Bureaus and Embassies communications are strongly correlated with trade ties at the aggregate level, we conclude that the Bureaus must be setting the conversation agenda in terms of introducing trade into the structure of information provision.

Overall, within the realm of the conversation agenda around trade, there is more evidence for policy trickling down in the State Department. If there is bubbling up, it appears to be happening elsewhere in the policy process.

5.2 Agendas in the State Department

This section provides more detail about what mechanisms within the State Department could encourage a top down or bottom up mode of information processing. We then discuss the relationship between agendas and communication based on how agendas structure the decisions that are made within an organization.

As we indicated above, there are arguments given both for and against the State Department's agenda being generated from the top or bottom. We discuss arguments both empirical and theoretical for why we might expect each type of processing. It should be noted that we do not dismiss *a-priori* the possibility that the influence could be bidirectional with a iterative process to reach agreement.

5.2.1 Bottom up

While bureaucracies are often associated with strict hierarchies and rigid vertical flows of communication (Cremer 1980; Gajduschek 2003), the reality of communication is clearly

going to be more varied.

Incentives to reduce risk. another reason why we might expect to see some agenda setting by the Embassies is actually referenced in the quote from Rubin (1987). If the information that bureaucrats is determined from above, the bureaucrats will no longer focus on novel signal in the policy environment. One of the functions of the State Department is to allow a faster change of the agenda when facts on the ground change. This argument is less compelling in this time period given the claims that this was what in fact happened. However, it at least suggests that hierarchical information processing may not be a sustainable organizational style over the long term.

There is evidence from historical accounts that even Kissinger recognised the role of Embassies in risk reduction. Kissinger wrote in his memoirs that the unexpected election in Chile of the left-wing Allende was "the kind of unacceptable risk that policymakers are hired to avoid" (Rubin 1987, p161).

Information advantages. The risk reduction function of the State Department also potentially creates a principal agent problem. Since the Embassies have more direct access to the situation "on the ground" in the host country, they have informational advantages over Bureaus in Washington. Because they can decide which information is relevant to report back to Washington, the reporting may be skewed towards their own preferences. Because too tight a control over Embassy reporting risks ignoring emerging issues, the Embassy may be allowed a more freedom in reporting its version of events.

Cost reduction. A related principal agent problem relates to the additional costs required for a Bureau to do its own separate analysis of an issue. On this view, the additional work required to modify or reject an Embassy's analysis will tend to mean its views are accepted unless they differ substantially from the preferences of the Bureau. Halperin, & Clapp (2006, p. 108) sums this up well arguing that "on routine matters all Presidents prefer and press for agreed recommendations from their principal advisers. This saves them the trouble of having to get into the issue to make a decision." The same basic logic applies at all levels, of bureaucracy (although with different levels of detail), with a more efficient bureaucracy making decisions at all levels and only pushing decisions higher where there

is a conflict in preferences over action.

Epistemic communities. An epistemic community is "a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain" (Haas 1992). The State Department's foreign service could potentially fit this definition. Epistemic communities gain their influence through internal debate and presenting a uniform front to the outside world.

This claim is not unprecedented. Indeed Davis Cross (2008) argues that the European diplomatic corps has sufficient expertise and strength to constitute an epistemic community that can act separately from member-state preferences.

However, it may be stretching the definition to claim that Embassies are differentiated enough from Bureaus to allow such a community to emerge. While the State Department as a whole might be able to act as an epistemic community, the Embassies do not have a separate internal channel of debate between them but must send every message back to Washington as well as laterally between Embassies.

Coalition Building. The final influence that could encourage bottom up influence is the need for coalition building within the Department. Warwick (1979) reports "beneath the formal allocation of authority in State is an intricate web of conflicts, rivalries and alliances." In such an environment, inter-Bureau disagreements could make alliances with some Embassies valuable even at the expense of ceding some agenda control.

Top Down

Despite the multiple factors potentially pushing for more bottom up processing of information in the State Department, there are several countervailing forces that may encourage greater top down control of information supply.

Information/processing advantages. Although Embassies have greater direct access to information, Bureaus in Washington DC have more dedicated analysts whose job it is to synthesize information from multiple sources. The Bureaus are also more directly under the control of the Secretary of State and other high level officials, so might be expected to provide instruction from above.

This suggests that Embassies might be willing to accept a division of labour where they try and provide the highest quality analysis and the Bureaus draw wider links strategically between different areas.

Hierarchy. Another factor that might encourage top down processing is the reportedly highly hierarchical structure of the Department (Rosati, & Scott 2010), with an explicit ordering of individuals by level and an implicit status ordering between different offices and career tracks.

This hierarchy potentially makes top down processing more likely due to clear lines of authority (or at least ordering of authority). It may also encourage diplomats to try and impress the hierarchy to help their careers by moving up in the organization. Smith2011 reports that the State Department has far too much information and FSOs are constantly competing to gain attention from the hierarchy. This is backed up by several interviews with diplomats where they outline the attention they paid to how their information would be received in Washington (Kennedy 1997; Kennedy 2007).

Empirical reasons related to Kissinger. Historical accounts of this time period discuss Kissinger's attempts to control information flow to the National Security Council and the President (Rubin 1987; Siniver 2008; Garrison 1999). Allowing an independent voice for Embassies could potentially be another threat to his authority (first as National Security Advisor and then as Secretary of State).

Siniver (2008) gives an example of this behaviour: "Kissinger employed various tactics to mould U.S. foreign policy to his and Nixon's liking. One of them concerned the entanglement of the bureaucracy in what Landau described as 'a web of useless projects and studies, cleverly shifting an important locus of advisory power from the Cabinet departments to his own office."' (Siniver 2008, p.56)

5.2.2 Defining the agenda

This section discusses how an organization's agenda is linked to its communication patterns. Agendas within organizations have often been measured in terms of the content of documents produced by the organization. For instance, there are many examples of

coding politicians' agendas by using content analysis on their speeches (John, & Jennings 2010; Wood, & Peake 1998; Jennings, & John 2009). Content is an important aspect of an agenda. However, we believe that where the information is distributed within an organization is also an important aspect of the agenda. These two approaches are complimentary rather than in tension. indeed as we showed the in previous chapter, the communication agenda appears to line up with the content of the communications.

Hammond (1986) argues that the structure of an organization constitutes its agenda. This is because the structure of the organization determines which options will be compared, by whom and in what order. Hammond derives a series of formal models showing that the institutional arrangements (of who reports to whom and how policy responsibilities are carved up) can have a strong impact on which policy choices. This makes the choice of institutional jurisdictions and structure of reporting relationships key to who receives which proposals. Consequently Hammond's model is using the flow of communication as the explanation of the link between organizational structure and the agenda of a bureaucracy.

However, this is too strong an assumption about the effect of formal structure on communication. There is certainly a relationship between the two. However, as figure 5.1 shows, the structure of communications in an organization is actually more complicated than the assumption of reporting between subordinate and director would suggest. The figure shows the network of communications between different offices in the State Department. We only show ties where at least 100 messages were sent from one office to another (a lower cut-off than 100 makes the plot entirely unreadable). Embassies and Consulates are coloured blue, while Bureaus are coloured red.

While it is not possible to draw any strong inferences from the plot, it is not possible to see a clear hierarchical structure with well delineated lines of authority and boundaries between units. Instead there is a dense network of communications without clear boundaries. This is not to say that there is no effect of organization on communication structure, just that the assumption of a completely deterministic relationship between the two is not sustainable at least for the State Department.

However, the wider point of Hammond's theory is still correct, that the reporting of information and proposals constitutes an agenda. However, if the structure of this agenda is

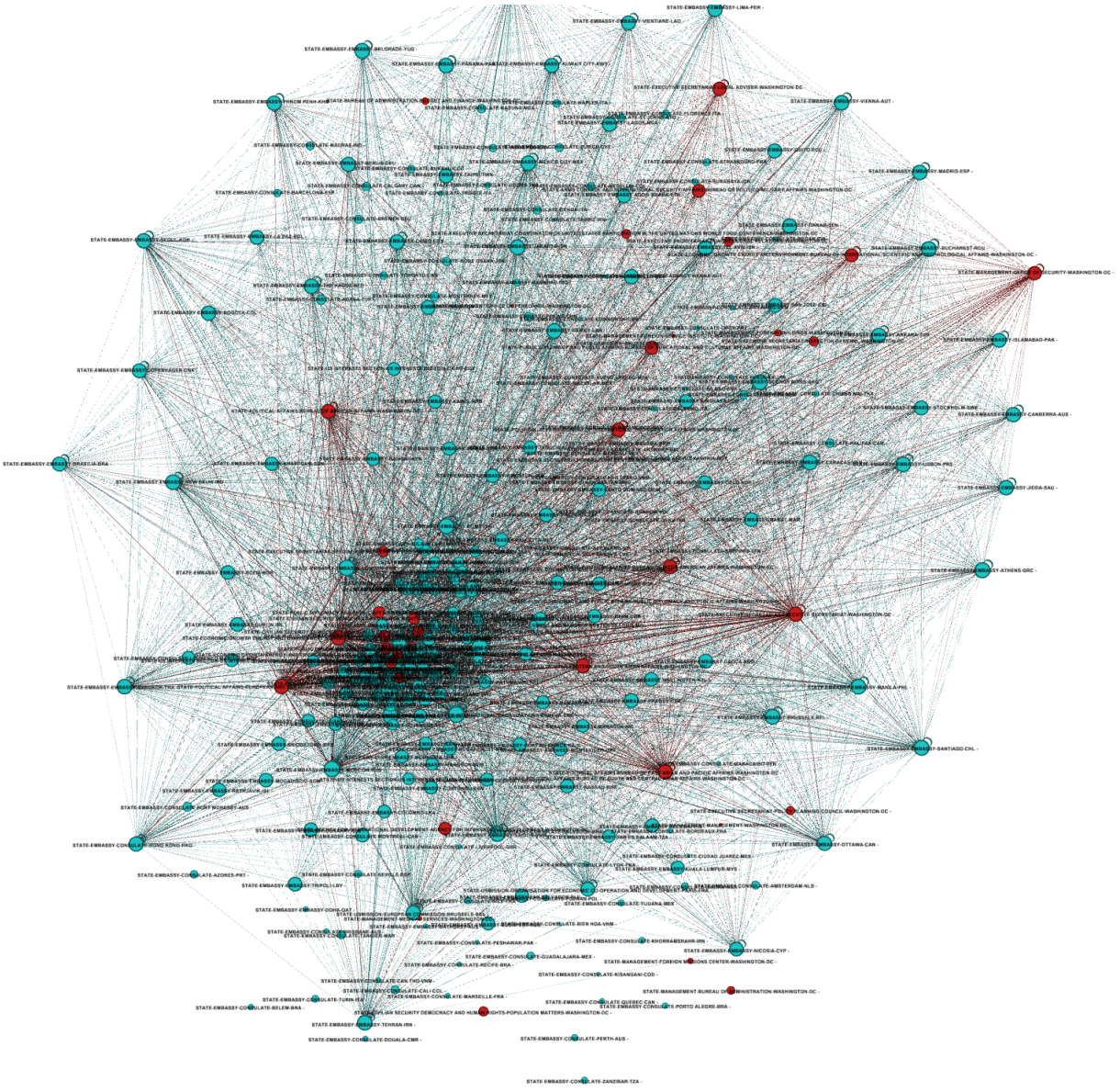


Figure 5.1: Force-directed network plot of State Department communication flows including Embassies and Bureaus

not fully determined by the formal structure of the organization but by the communication choices of actors, then there is the potential for agenda setting from different levels of the bureaucracy.

Importantly, the choice of communication partners changes how efficiently information on different offices can be transmitted through an organization. If there are high levels of communication between Embassies in countries with strong trade ties, then discussion of trade is more likely to be transmitted across the network.

This theory is given extra credence by the results of the previous paper which showed that factors that predicted the strength of ties between State Department offices in different countries are related to the content of those messages. That paper's results showed that State Department communications were structured around multiple issues. However, by far the strongest predictor were levels of trade between host countries, which explained nearly half of variance in communication strength between countries. Given the very strong cross-sectional importance of trade in State Department communications, we focus on which actors (Bureaus or Embassies) introduce this variable into the communications network.

5.3 Communication in the State Department

The diplomatic cable is one of the key ways in which information and policy information is transferred within the State Department. This is particularly true in the time period under study here before the use of classified email communications (Rosati, & Scott 2010). When telegrams were first introduced in the 1850s their use was restricted to the Secretary of State and Ambassadors based in the capitals of great powers (as these were the only countries connected to the telegram network) (Nickles 2003). By 1973, there are no such restrictions on which actors are able to communicate within the State Department as we saw in figure 5.1.

A typical (if slightly shorter than average) cable is shown in figure 5.2. There are a number of important pieces of information to note on the cable. In this article, we make use of the fields related to senders and recipients (A, B, C, D, E and F), references (K) and date (C). We briefly describe the information that is taken from each of these fields

and refer the reader to the relevant appendices for a much more in-depth description of the parsing process and issues involved in extracting this information.

5.3.1 Senders and Recipients (A, B, C, D, E and F)

The first fields we consider define where the cable is sent from and where it is distributed to. In this case the sender, the US Embassy in Kuwait, is indicated in the FM field (D). The recipients in Washington DC are defined in the ACTION (A) and INFO (B) fields. In this case the main recipients is PM (Bureau of Political-Military Affairs). The message is also cc'ed to a number of other recipients including OCT (Office of the Deputy Secretary for Communications)and NEA (the Bureau of Near Eastern Affairs) and CIAE (the Central Intelligence Agency).

Other embassies and some recipients outside of the State Department (such as the Defense Department) are included in the TO (E) and INFO (F) fields. All messages from Embassies are formally sent to the recipient SECSTATE WASHDC. However, this is just a formality and merely indicates that the message will go to the State Department (Department of State 2012a).

5.3.2 References (K)

The second important field for this paper is the REF (K) field. This field defines which other messages are being referred to in the current message. This is more than just an offhand reference as State Department procedure requires the Embassy receiving a message to manually retrieve the cited cables from their paper based archive in order to refer back to the information (Lotus Development Corporation 1996). As we show later in this section, these references have an important role in structuring communications into conversations that tend to keep the same set of participants over time.

The references field is not always well formatted by diplomats and a total of 29% of references could not be matched correctly to another cable in the dataset. We explain more details of the matching process and the biases that may be introduced in appendix 4.2.

CONFIDENTIAL

PAGE 01 KUWAIT 00845 051203Z

43
A ACTION PM-07

B INFO OCT-01 NEA-10 ISO-00 INR-10 EB-11 AID-20 IGA-02

CIAE-00 NSAE-00 RSC-01 SSO-00 INRE-00 MC-02 DRC-01

OMB-01 /066 W
---------- 118944
C O P 051144Z MAR 74
D FM AMEMBASSY KUWAIT
E TO SECDEF WASHDC//ISA(SA)/DSAA//IMMEDIATE
SECSTATE WASHDC IMMEDIATE 8940
F INFO USCINCEUR VAIHINGEN GERMANY//PEIORITY
CNO WASHDC//OP-63//IMMEDIATE
CSAF WASHDC//IMMEDIATE

G C O N F I D E N T I A L KUWAIT 0845

H E.O. 11652: GDS
I TAGS: MASS KU
J SUBJ: FMS KUWAIT: JOINT AIRCRAFT EFFORT

K REF: KUWAIT 0826 (DTG 031259Z MAR 74)

1. REFTEL REQUESTED A7E AND F5E PILOTS JOIN KUWAIT ACFT PRESENTATION. PLEASE HOLD PILOTS AND CONTRACTOR SUPPORT PENDING FURTHER DISCUSSION WITH KUWAIT MOD. INFORM LTV AND NORTHRUP. WILL ADVISE.
STOLTZFUS

CONFIDENTIAL

NNN

Figure 5.2: Example of a cable sent from the American Embassy in Kuwait

	Receiver			
	Bureas	Embassies	All	Number of events
sender				
Bureas	584824	452665	132159	63176
Embassies	2902200	2298466	603734	223107
All	3487024	2751131	735893	286283

Table 5.1: Breakdown of communications received by each type of recipient (columns) from each type of recipient (rows)

5.3.3 Times (C)

The final field we make use of is the time the cable was sent (C). This field is important because we are focusing on the process of influence so understanding the ordering of events is vital to assessing influence. Because the cables have timestamps down to the nearest minute, we can get a highly detailed picture of the ordering of communications between different offices. There are a number of issues in parsing these dates (including missing timestamps on some cables) that are necessary to deal with. For details on these problems and the imputation procedures used to correct them see appendix 4.1.

5.4 The structure of communication

This section describes the main structures in the State Department cable data. We first outline some very basic descriptive information about the communications network in terms of number of messages, recipients and their distributions. We then focus on the conversation structures revealed by references between cables and show how these references influence future recipient choices.

The network plot in figure 5.1 tells us that there is complexity in the network of communications in the State Department and that communication does not just follow hierarchical lines. However, this complexity means that the plot does not tell us much more than that.

5.4.1 How many messages are sent?

Table 5.1 shows some basic descriptive statistics about the State Department cables. The first three columns show the number of messages that are received by Bureaus, Embassies

and in Department as a whole. The rows show which senders these messages are received from. The fourth column shows the total number of events sent by each type of sender. There are many more messages received than sent within the network because almost all cables have multiple recipients. In this chapter, we are primarily concerned the distribution of information to Embassies, so are focused on the messages and recipient choices in the second column. This is not to deny the importance information distribution to Bureaus. However, the mechanisms that influence this distribution are likely to be different from those influencing distribution to Embassies. Although these messages are beyond the scope of this paper, future research should analyse the determinants of this type of information distribution. Importantly these results show that Embassies receive a majority of their information from other Embassies with only a minority coming from Bureaus in Washington DC.

5.4.2 Conversation chains

An important feature of the State Department cables is the references field, as this allows cables to continue discussion or projects over time. By treating the messages as nodes and the references between them as edges (ties), we create a network of cables that represents the path of the conversations over time.

We look at a sample of these conversation structures in figure 5.3 (each of these structures is a weakly connected component i.e. it's all the cables that can be connected together with either an incoming or outgoing tie). Each circle is a message and the lines are citations between messages. The highest circles are the messages sent the earliest with more recent messages further down. The circles are sized according to how many of the messages in the conversation can be traced back to them. This means that conversations that weave together many chains will have multiple large circles at the top that are unconnected. More linear conversations will have a single large node at the top with most other messages in a conversation continuing from that one point.

The plot suggests that there are quite different conversation structures throughout: some are deep conversations, some carry out a linear chain (at the start, end, or middle), some span out in a star formation, some weave older citations back in and some thread otherwise separate conversations together.

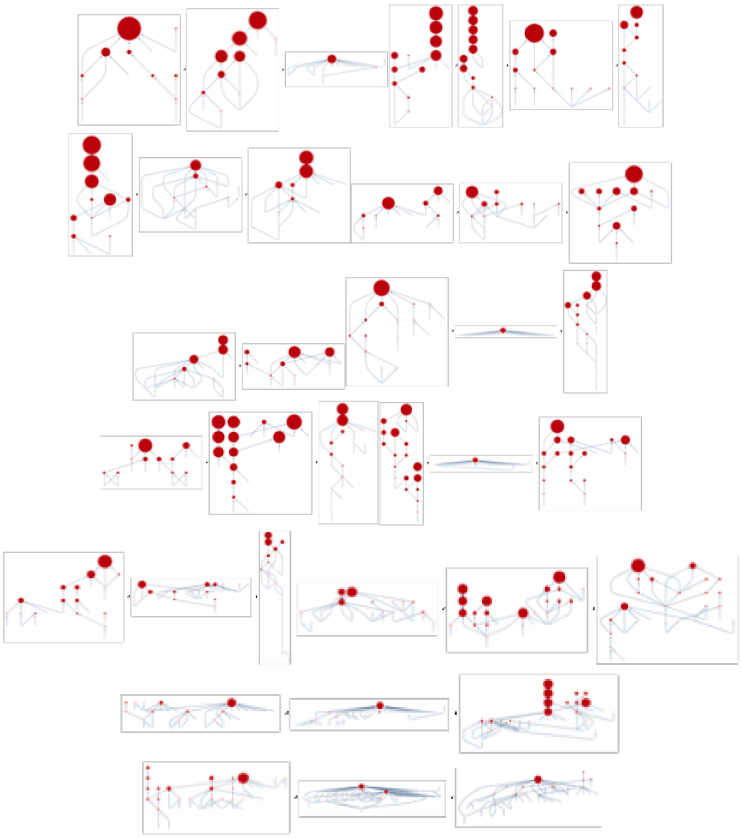


Figure 5.3: A sample of component structures in the citations graph

The largest single thread of conversation that can be traced back to a single message consists of 377 messages. However, most chains are much shorter with the mean number of additional messages being just 2.6. Given that most chains do not get very far but some get very large, there is the potential for a disproportionate influence of highly cited messages.

5.4.3 How persistent are the recipient sets between cables?

The next question is to what extent these conversations tend to keep the same set of participants over time. If we consider a new cable that cites an older cable, the sender of the new cable could keep the recipient set exactly the same as the previous cable (not including herself). Alternatively she could add or remove some recipients to change the audience.

Figure 5.4 shows what percentage of the older cable's recipients are included in the new cables recipient set.[1] The results show that by far the most common proportion of inheritance is 100%. On average, 63% of recipients from the older cable are kept in the new cable. To put this in perspective, if sender chose recipients completely at random, a recipient from a previous set would only have a 1.1% probably of being included in a subsequent set.[2] Even considering the fact that there is additional structure, a 63% rate of inclusion is extremely high.

These results show the importance of the conversation structures we saw above. If there is a high rate of inheritance between one cable and the next, then starting a conversation potentially gives a degree of control over the agenda in the rest of that conversation. Subsequent messages do make changes, but most do not.

Although we see that recipients are inherited between messages, we still need to understand how large a proportion of all recipients this account for before we can assess the mechanism's impact on information flow.

Table 5.2 shows the number and proportion of recipients that are included, added and inherited by Embassies and Bureaus. The table only includes Embassies (as defined

[1] This is only calculated for cables where at least one cable is cited where there is a valid recipient in the set.

[2] This calculation assumes a uniform probability of choosing each of the 285 possible recipients and then applies these probabilities to the actual recipient set sizes observed in the data.

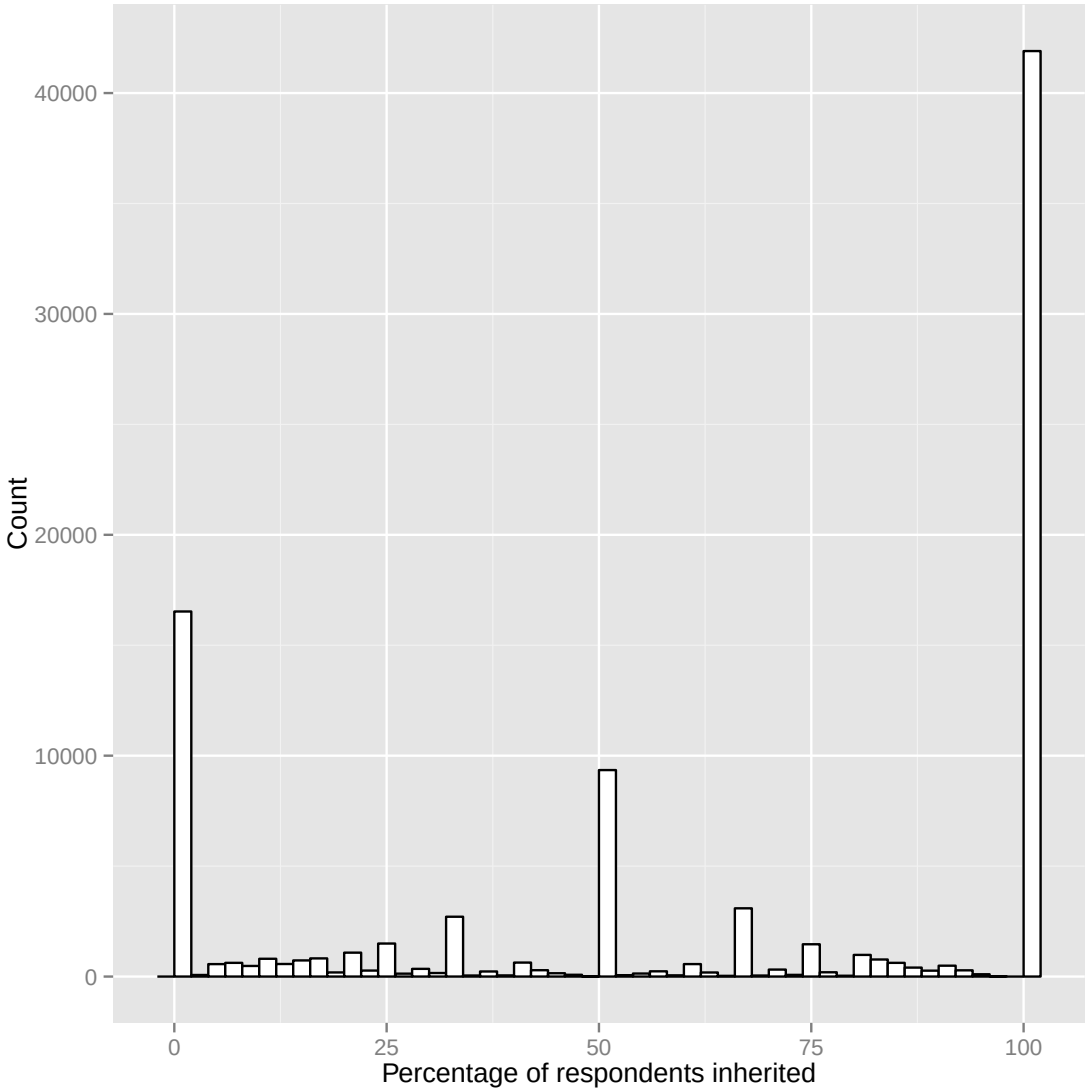


Figure 5.4: Histogram showing the proportion of respondents that are kept from cited cables.

earlier) as recipients and doesn't include recipients in Washington DC or outside of the State Department. There is a similar split between adding and inheriting recipients for the Bureaus and Embassies. In both cases around 30% of recipients are inherited from previous messages in a conversation and 70% of recipients are added by the sender of the message. This confirms that a substantial part of the information flow around the State Department is determined by previous messages rather than actively by the current sender.

	Bureaus	Embassies
Total recipients included	132159	603734
Number of recipients added	94333	423542
Percentage of recipients added	71	70
Number of recipients inherited	37826	180192
Percentage of recipients inherited	29	30

Table 5.2: Breakdown of how many cable recipients are chosen and how many are inherited

Given that recipient inheritance plays a role in the recipients of a message, it is worth looking at which messages are being cited. If Embassies are simply citing their own messages, then recipient inheritance is simply reinforcing their own previous decisions. Table 5.3 shows the breakdown of messages sent by Embassies and Bureaus. This shows that 78% of messages in the State Department are sent by Embassies, with 22% sent by the Bureaus in Washington DC. Despite the low initial percentage of messages sent by Bureaus, a considerably higher percentage of citations from Embassies are made to messages from written by the Bureaus in Washington DC (37%).

There is a substantial split between the self citing behaviour of Embassies and Bureaus. 21% of citations by Bureaus are to previous messages written by themselves. By contrast, 45% of messages cited by Embassies originated at that Embassy.

Particularly notable is that only 5% of citations by Bureaus are to messages written by other Bureaus. By contrast, 18% of citations by Embassies cite messages written by other Embassies.

These results are an underestimate of the true proportion of inherited recipients, due to limitations of this version of the dataset. In total, 29% of references could not be matched correctly. Assuming that the missing references include a similar number of recipients as those which are present, approximately 42% of recipients will be inherited in total. We hope to make improvements to the parsing process to capture and match more of these

references. However, some of these references simply refer to cables that are either top secret classification (which we may be able to retrieve from the paper archive at some stage) or messages with only temporary TAGS neither of which are present in the current dataset.

	Messages sent	Messages cited by Embassies	Messages cited by Bureaus
Embassy Count	223107	64699	21850
Embassy Percentage	78	63	74
Bureau Count	63176	37599	7796
Bureau Percentage	22	37	26
Self count	-	45662	6254
Self percentage	-	45	21

Table 5.3: Senders of all messages, those cited by Embassies and those cited by Bureaus

5.5 Relational Event Models

In this section we briefly discuss what modelling communication agenda setting requires. We argue that relational event models are the most appropriate type of model to analyse this type of agenda setting.[3] We then give further details about the alternative relational event models and describe the additional features our model requires to correctly capture the structure of the diplomatic communications data.

As we discussed in section 5.2.2, we are interested in the communication agenda of the State Department, i.e. who shares information with whom. This agenda setting question is both relational (it involves understanding the influence of actors on each other) and longitudinal (by its very nature agenda setting is a process that takes place over time).

Until recently, the way to analyse longitudinal relational data was to aggregate communication patterns across time periods and model the change between these periods (for instance, with Siena models (Snijders, Bunt, & Steglich 2010)). However, a new set of methods know as relational event models have been developed that take the communication event itself as the unit of analysis.

Most of the current work on relational event models builds on the framework outlined by

[3]This overview is meant just to highlight the key features of the alternative approaches. For an in-depth discussion of the alternative approaches and their comparison with relational event models see the book by Stadtfeld (2012) which has a chapter reviewing these models.

Butts (2008). The key insight of that paper is that we can model each communication (or relational event) in a multinomial logit model, where there is a utility function that depends on the attributes of the possible choices. This utility function can then be estimated using maximum likelihood estimation, where the parameters in the utility function maximise the likelihood of the observed choices.

Butts (2008) defines a dyadic model where the utility function describes the likelihood of a message being sent between a dyad: meaning it models a choice between $n \cdot n$ possible dyads in each event. A later extension by Stadtfeld (2012) divides up this choice two stages: ego (the sender of the message) selection and alter (the recipient of the message) selection, conditional on the chosen ego: $P(event) = P(ego) \cdot P(alter|\text{chosen ego})$

This means that the ego selection is between n options and alter selection is between $n-1$ options (because ego can't choose to communicate with themselves. This is considerably simpler computationally and also changes the model to focus on an actor's perspective.

To give a clear example of alter selection in this model, let's imagine that we have just four offices in our system: London, Paris, Moscow and St Petersburg. We will leave ego selection for now and focus on alter selection. Let's suppose that the Moscow Embassy is making a choice of which other Embassy to contact. In our model Moscow makes this decision based on just two variables in its utility calculation: whether the other office is in the same country and whether the Embassy is based in a US ally. These are given weights of 0.5 and 1, respectively.

For the moment let's just consider the choice of whether Moscow will contact London (and at this point Moscow can choose exactly one recipient). The total utility is given by a linear combination of one or more variables related to the potential choice $0.5 \cdot \text{US ally} + 1 \cdot$ same country. In this case London is a US ally, but is not in the same country as Moscow so the total utility is 1. Now that we have the utility we still need a way of turning this into likelihoods. The transformation simply takes the exponent of London's utility and divides it by the sum of the exponents of the utility of all possible alters.

$P(\text{Moscow contacts London}) = \frac{exp(1)}{exp(0.5)+exp(0.5)+exp(1)} = 0.45$

Under this utility function, Moscow has a 27% likelihood of contacting London or Paris and a 45% likelihood of contacting St Petersburg. In an empirical model, we fit the parameters of the utility function based on the observed choices. We use maximum

likelihood estimation to choose the parameters that maximise the likelihood of observing the event choices that were actually made in the data.

Importantly, these relational event models also endogenous variables based on past relational events. So we could for instance look at an actor' level of reciprocity by including a variable measuring how many messages ego had received from each possible alter in the last week. A higher parameter on this variable would indicate a greater tendency of actors to contact offices who have contacted them in the recent past.

These models can be increased in complexity and are flexible in the types of attributes that can be included as predictors of communication. Further work has extended these models to cover events with multiple recipients. Given that recipient sets could potentially be large and the decision space of potential sets increases exponentially with size, it is not computationally feasible to apply a multinomial model directly to the sets. Instead work by Lospinoso et al. (2013) models alter selection as a sequence of decisions by the current ego to add further alters into a recipient set. Each subsequent addition is made conditional on the alters that have already been added to the set. This makes it possible to model whether senders tend to contact sets of alters whose members are similar to the other alters who are contacted at the same time. Applying this concept in our model, we look at whether Bureaus tend to contact sets of Embassies which have strong trade links with each other.

5.5.1 Limitations of current Relational Event Models

Relational event models provide a useful tool for modelling communications. However, these models make the assumption that events are independent from each other, conditional on the parameters in the model and the data.

The descriptive analysis in section 5.4.3 shows that this is not a valid assumption in the State Department's communications. Rather than being independent, the messages are structured within existing conversations, with some chains of communication involving hundreds of messages. This paper's analysis also shows that there is a high level of inheritance of recipients between these messages. On average, 63% of recipients in a cited message are included in the subsequent message. These findings mean that a standard relational event model's assumptions will be violated and the model will be misspecified.

A hypothetical example of why this could make the analysis of agenda setting in the State Department difficult is as follows. Suppose we model the decision of Embassies to add recipients to cables based on trade ties. The parameter would measure the tendency for ego to contact alters with strong trade ties to ego's host country. In addition, let's also suppose that the true value of this parameter for Embassies is zero i.e. that Embassies don't have a preference for alters with strong trade ties.

When Embassies cite messages by Bureaus and inherit the recipients from those messages, they will end up contacting Embassies in countries with which their host country has strong trade links. However, this situation would not actually reflect an active preference among Embassies for recipients with trade links and hence it would not reflect the Embassies' own preferences over the agenda.

Given the problems with using an existing relational event model and the importance of citations in conversation practices in the State Department, we derive a new model that addresses the limitations of existing relational event models we identified above. This section outlines the model in detail, while the next section discusses our operationalization of variables.

5.6 The RECHAIN Model

This section presents how a Relational Event Conversation-History Actor-Initiated Network (RECHAIN) model is adapted from models developed by Lospinoso et al. (2013) and the work of Butts (2008). The Reveal model decomposed the probability of a relational event as the product of the probabilities of its ego, mode, topic and recipient set (Lospinoso et al. 2013). These choices are modelled on the basis of the event history and other covariates of the actors in the network.

Our model changes the probability decomposition to reflect the specific facts of the Department of State's communications. First, we observe only one mode of communication (diplomatic cables) so modelling this aspect of communication is unnecessary. Second, we introduce citation choice into the model in order to reflect the importance of citations in themselves and for determining the topic and audience of a conversation. Third, we model the choice of recipients from the starting point of those included in the cited cables rather

than from an empty set. This means that we model both the addition and removal of recipients.

For now, we condition on the time and ego of the event but ego selection can be modelled in exactly the same way as for the Reveal model.

The full selection process for RECHAIN proceeds as follows:

1. *ego selection*: multinomial selection model of who the next sender will be conditional on the time of the current event. i.e. given that we know an event happens at time t, how likely is it that London, Paris or Moscow etc. will be the sender of that event?

2. *citation selection*: sequential multinomial models of which previous cables the sender of the current message will choose, conditional on the time and sender of the current message. i.e. given that we know that Paris is the sender of the message at time t, how likely is it that each of cables 1973PARIS5, 1973LONDON55, 1973STATE67 or 1973PARIS6 etc will be included in the citation set? If there are multiple citations the product of the probabilities is taken with each subsequent citation conditional on those already included. So if a cable cites 1973PARIS5 and 1973LONDON55, then the probability of the recipient set is decomposed as
$P(1973PARIS5) \cdot P(1973LONDON55|1973PARIS5)$.

3. *alter removal*: sequential multinomial models of which recipients that are removed from the recipient set; this is conditional on the time, sender, and citations of the current message. i.e. how likely is it that each of the initial inherited recipients London, Melbourne and Halifax will be removed from the recipient set, conditional on the time the message was sent, the sender of the message, and the citations included in the message? Once again, this is modelled sequentially, so the probability of the observed removals is the product of each removal that is made. So if ego remvoes London and Melbourne the probability of these removals is
$P(removinglondon) \cdot P(\text{removing Melbourne}|\text{removing London})$.

4. *alter addition*: sequential multinomial models of which recipients are added into the recipient set by the sender; this is conditional on the time, sender, citations, and removed recipients of the current message i.e. how likely is it that the sender will

add each of Moscow, Mexico City and Berlin as recipients etc in the message. Again this selection is sequential so that each actor is added into the recipient set one at a time with a separate selection for each one.

The model outlined above, reflects the full decision process for sending an event. In the wider StateNet project we will focus on all of these choices, but this paper focuses on one part of the model: alter addition. Because we have defined what this decision is conditional on, the results of the addition model are not affected by whether we model the other stages at the same time. While the RECHAIN model decomposes the event as above, any stage can be modelled

All the selection stages are important to understanding agenda setting. However, the most important stage is the alter addition as this is the stage where actors are introduced into recipient sets. Subsequently, these decisions are compounded and curtailed by the citation selection and alter removal. However, alter addition defines the initial set of possibilities for these other stages. Because of this, we focus on analysing the alter addition decision for Bureaus and Embassies.

5.6.1 Model Specifications

This subsection goes into more details about the precise specifications of our model. We define an event as set of a timestamp τ , an ego i, a citation set of zero or more citations to other events C and a recipient set of one or more recipients J. We follow the shorthand of Lospinoso et al. (2013) by indicating the event that one of these applies to with a function that takes the event and returns the relevant component of the event so $i(e_a)$ is the ego of event a and $J(e_a)$ is the recipient set for event a. We will refer to the current event as e with no subscript. The event history, $\mathcal{E}$, is defined as all events and covariates that have taken place up until the time of e. This means that future information cannot affect the value of effect statistics.

There are many possible ways to decompose the probability of an event but we model

the following decomposition:

$$\begin{aligned} f(e|\mathcal{E},\theta) = f(i(e),\tau(e)|\mathcal{E},\theta)\cdot \\ f(C(e)|i(e),\tau(e),\mathcal{E},\theta)\cdot \\ f(J(e)|C(e),i(e),\tau(e),\mathcal{E},\theta) \end{aligned} \tag{5.1}$$

For the citation and alter densities, the size of the potential choices is too large to model directly. As Lospinoso et al. (2013) note, the outcome space for these types of distributions can be huge even for relatively small numbers of objects. For instance, with just 12 potential actors to select from, there are 32,767 unique recipient sets. Instead we follow the approach of Lospinoso et al. (2013) and model these sets as an ordered set where ego adds citations/actors to each set sequentially.

We explain the citation choosing process in much greater detail in appendix 4.3

Alter Selection

We depart from the approch of Lospinoso et al. (2013) when modelling the recipient sets. The Reveal model sees the probability of the recipient set in an event as being the product of the probabilities of each alter being added to the recipient set(conditional on those already included), so that the set moves from empty to the final set. However, the discussion in section 5.4.3 shows that this decomposition may be inappropriate for diplomatic cables as the citations lead to recipient inheritance.

Consequently, RECHAIN deviates from modelling selection as starting from an empty recipient set with elements added sequentially (as we do for citations), but instead takes the set of unique elements that were present in the recipients sets of the cited events and sequentially removes, then adds elements.

The initial set $J(e)_{initial} = \{j \in J(e_n)|e_n \in C(e), j \neq i(e_n), j \neq i(e_n)\}$ is defined as all recipients that were included in the citations excluding the currently selected ego and the egos of the cited events. We then consider the recipient set probability density,$f(J(e)|T(e),C(e),i(e),\tau(e),\mathcal{E},\theta)$, to be the combined likelihood of the set of additions and removals of recipients necessary to move from $J(e)_{initial}$ to the final observed recipient set $J(e)$. Presumably additions to the recipient set and removals are

not chosen according to the same criteria (it is likely that different -perhaps opposite- factors drive the decision to add or remove a recipient). Therefore we break down the change set further into a removal set, $J(e)_{removal} = J(e)_{initial} \setminus J(e)$ and an addition set $J(e)_{addition} = J(e) \setminus J(e)_{initial}$.

We provide a much more detailed account of the alter removal process in appendix 4.4.

Modelling Alter Addition

After alters have been removed from the set, the ego can then choose to add additional recipients.

There are many possible ways to calculate an alter risk set[4] but the approach here requires a minor adjustment. Let f(e)be an arbitrary function that returns a risk set for an event, the addition risk set is the relative complement of $R_{removal}$ in $f(e)$. $R_{removal} = J(e)_{removal} \setminus f(e)$. [5]

If no citations were selected in the current event or there are no valid recipients in the cited cables' recipient sets, then the model starts from an empty set and the alter selection process is equivalent to Lospinoso et al.'s (2013) model.

This paper considers a relatively constrained set of actors for Embassies and Bureaus to choose between. Consequently, rather than defining a risk set dynamically, we use the full set of 300 State Department actors who are not based within the wider Washington DC area.[6]

Modelling the decision not to add further citations/alters

There was a similar issue with including the option to stop adding citations or recipients. This will be added in future versions. For now, it is important to interpret the parameters as comparing the choice to add one recipient or another rather than between adding a recipient or no recipients at all.

[4] These will be covered in detail in the Lospinoso et al. paper.

[5] Note, that this formulation does not allow the reversal of choices by ego. Once an actor is removed, she cannot be re-added to the risk set. This is for two reasons: first, it would be difficult to model such a decision in the current framework because we do not observe reversed decisions and second, there are good reasons to think that ego's decision to exclude someone who was previously part of a conversation is different from ego changing her mind about whether to add someone.

[6] Note that this means we include the USUN delegation in New York and other similar missions as possible actors to choose.

In this iteration of the addition model we do not include the option of not choosing to add further recipients into the model. The reason for this is that the average number of recipients changes substantially across the course of the 22 months in this study. This heterogeneity means that any other variable that varies with time may be confounded by secular changes in recipient set size. We discuss some possible explanations for this change in appendix 4.5.

5.6.2 Modelling all stages of communication

Although our decomposition of an event's probability has selection stages for ego, citations, removals and additions, we do not use all of these stages in this chapter. As we are interested in why information is distributed as it is, we focus on the process of audience change rather than conversation choice.

This means that we model the removal and addition decisions. The citation stage will also be important for understanding information distribution. We leave this stage for future research because there is substantial descriptive background work needed on the structure of conversations before we can begin inferential modelling of this choice. For a brief overview of the citations structures and the challenges in modelling citation choice see Appendix 4.3.

Ego selection is also an important stage for understanding information provision in general. An ego selection model can help to understand the dynamics of requesting or coaxing information out interlocutors. For instance, we could begin to look at actors' ability to elicit activity from other actor depending on their relative positions in the State Department hierarchy.

While these are important questions for future work, they are not as directly relevant for answering why a particular actor provides information to another. Instead they are more general questions about the tendency of actors to provide information at all.

Selection ordering

The current paper consider the alters to have been added in the order they are listed on cables. This is an important assumption when considering order variant effects. Order variant effects are ones that change their value depending on the actors that are already

in the recipient set (remember that we add alters sequentially and begin with the set of inherited recipients).

As an example we could have an effect looking at how similar a candidate alter is to those alters already in the recipient set. For instance, perhaps Bureaus are likely to add Embassies based in countries with similar levels of GDP. We would then define an effect that assessed how similar each candidate Embassy is to the Embassies already in the recipient set (in terms of GDP of host country).[7]

Decays

For the time being we use an exponential decay function with a half life of two weeks. This means that an event occurring zero seconds before the current event will be given a weight of one in our model whereas an event occurring 14 days previously will only be given a weight of 0.5.

This decay time is chosen in order to minimise the effect of time-zone and weekday effects (see figure 4.1 for an example of these in appendix 4.1), while capturing relationships beyond the very short term conversation dynamics.

For effects that depend on past events these decayed values are usually summed to produce the statistic that relates to choosing a particular option. So for the habit effect we model an embassy's decision to contact a particular alter based upon the sum of the decayed times that they previously contacted each alter. The decay function, $\delta\big(E, \tau(e)\big)$, takes an event set E and the current time, and returns the sum of the decayed differences for that set. Exponentially decayed using a chosen decay parameter, λ.

$$\delta\big(E, \tau(e)\big) = \sum_{e_n \in E} exp(-\lambda(\tau(e) - \tau(e_n)\big) \tag{5.2}$$

[7] An alternative approach to using the observed ordering of alters is discussed by Lospinoso et al. (2013). They take the ordering of recipients to be unobserved so include an additional step off expectation maximization in the estimation procedure. In short, this requires using the expected score and information matrix in parameter estimation. The expectations are approximated by MCMC sampling of sequences of alters from the estimated parameters. This extra step makes model estimation far more computationally intensive. In addition, the expectation maximisation approach will only be superior if the order variant effects are more informative about the ordering of recipients than the stated ordering in the data.

5.6.3 Summary

In total, then we have outlined (and implemented) RECHAIN, a model that sequentially models an event in terms of ego selection, citation selection, alter removal and alter addition, all from the actor's perspective. For our purposes we focus on alter addition, while acknowledging that the other stages are potentially useful.

5.7 Mechanisms of communication agenda setting

Now that we have outlined the RECHAIN model and narrowed our (current) focus to the process of alter addition, we set out the questions we hope to answer with our analysis. Our primary question is whose preferences over trade become reflected in the overall structure of information provision. In other words, with regards to trade whose preferences set the communication agenda.

The secondary question is then which mechanisms of agenda setting link the agendas of the Bureaus and Embassies in the State Department. This section expands on how we will analyse the communications data to analyse these questions. We can ask what mechanisms of agenda setting might be observable within the communication patterns. To begin with, we have the question of whether there is actually agreement over the communication agenda:

- H0: There is a persistent gap in the communication agendas of the Embassies and Bureaus in the State Department regarding the importance of trade links

We can largely reject H0 due to the evidence in the previous chapter that Embassies and Bureaus put a very similar weight on different issues including international trade. Hence, we need to find mechanisms that explain this agreement. It is possible that their agendas are divergent in terms of the actual content of these messages. However, in terms of the information distribution patterns there is no such gap.

5.7.1 Which actors focus on trade ties?

The next question is whether Embassies or Bureaus (or both) are the ones introducing trade ties into the communication patterns. Given that we know the trade is important

to the overall communication structure, it must be reflected in the utility function of at least one of the Bureaus or Embassies. We focus only on trade at this point, in order to focus on a single issue and its relationship to the endogenous mechanisms, rather than simultaneously considering the interactions between the exogenous predictors. This still captures a high proportion of the agenda, as trade ties explain more cross-sectional variance than all the other exogenous variables combined (R^2 of 0.43 for the bivariate trade model compared to 0.36 for the model using all other predictors except trade).

- H1: Embassies focus on trade more heavily than Bureaus in their recipient choices
- H2: Bureaus focus on trade more heavily than Embassies in their recipient choices
- H3: There is agreement on the importance of trade in the communication agendas of Embassies and Bureaus, with both putting a similar weight on trade when choosing recipients

If H1 or H2 is the case, then there is agenda setting to be explained, as we need to explain how the overall information provision is the same between Bureaus, while their utility functions put different weights on trade.

To test these hypotheses, we define a two variables that capture how sensitive Bureaus and Embassies are to trade ties.

Trade

The trade effect captures the tendency of an Embassy to contact another Embassy dependent on how much trade their host countries have with each other. This effect is crucial to understanding whether Embassies actively make the choice to share information about the issue of international trade.

The trade effect is simply defined as the level of imports to the ego's host country from the candidate recipient's host country. So the Paris Embassy would have a value of this effect when considering whether to contact the Embassy in London because France imports a lot of goods from the United Kingdom. Importantly, the variable returns zero for alters based in the same country. We define $Y(i, j, variable)$ as the value of a dyadic *variable*

from i to j. So $Y(Delhi, Beijing, imports)$ is the value of the goods India imports from China. This means that the trade effect is defined as follows in equation 5.3.

$$Y(i, j_a, imports) \tag{5.3}$$

Exports to recipient set

The exports to recipient set effect captures the tendency of an actor to add further recipients based on their level of trade with recipients who are already in the set. If a Bureau has added the American Embassies in Venezuela and Brazil into the recipient set, it might be more likely to add Argentina as well if Argentina has high levels of trade with Venezuela and Brazil. We define this as the mean level of exports that the candidate alter's host country has with each of the host countries of the Embassies already included in the recipient set of the message as defined in equation 5.4. This effect is key to understanding whether Bureaus take account of trade relations in their communication behaviour. If the Bureaus add on the basis of trade and Embassies inherit the Bureaus' recipient choices, then we would see Embassies communicating along trade ties even if Embassies do not add recipients on the basis of trade ties.

$$\begin{aligned} if(J_a(e) = \varnothing) &: 0 \\ else &: \\ \frac{\sum_{k:k \in J_a(e)} Y(k, j_a, trade)}{|J_a(e)|} & \end{aligned} \tag{5.4}$$

5.7.2 Recipient adoption

The first mechanism that could explain agenda setting is recipient adoption. This mechanism posits that one set of actors (A) reacts directly to the behaviour or preferences of the other actors rather (B) than weighing up the value of recipients to A directly. To put this another way, this theory would say that A's communication utility function contains a term for B's past communication behaviour. Because of this A will end up communicating similarly to B, not because A has similar preferences to B but because A has a preference for communicating similarly to B. While we refer to this as one mechanism, there

are many possible psychological mechanisms that could account for this type of behaviour. The most direct mechanism would be that of persuasion (Hovland, Janis, & Kelley 1953; Perloff 2010) where an actor changes their opinions/preferences based on having changed their beliefs about the decision through argument. Another mechanism would be rational imitation (Hedstrom 1998) where one actor follows the lead of another actor because they believe that the other actor is acting rationally so their decisions are likely to be beneficial to adopt. Hedstrom (1998) gives the example of cars on a motorway slowing down when they see other cars slow down (because they believe that the car in front may have encountered an obstacle). However, this form of imitation leads to the phenomena of cars speeding up and slowing down in unison as they follow each other's lead. A similar process could be present in communications where an Embassy or Bureau follows the behaviour of other actors because they assume that these decisions are based on a rational calculation.

In both cases, these mechanisms will involve one actor changing their behaviour in response to the actions of a second actors so that the first actor's actions are more in line with those of the second actor. Although these mechanisms are importantly different, they are likely to be observationally identical within the current framework.

As we discussed in section 5.2, there are good reasons a-priori to suspect that this agenda could be set from above or below.

- H4: Bureaus imitate the past communication behaviour of Embassies
- H5: Embassies imitate the past communication behaviour of Bureaus

We define a number of effects that can help capture this kind of influence. For modelling Bureaus' recipient adoption preferences we consider geometric mean triadic closure, recipient set triadic closure and indegree popularity. For modelling Embassies recipient adoption preferences we consider geometric mean triadic closure, introduction by bureau and indegree popularity. We explain each of these effects in detail below.

Triadic closure (geometric mean)

The final standard effect is triadic closure. Triadic closure is a commonly used effect in many social network models. It derives from balance theory that posits that if i has a relationship with k and k has a relationship with j, the social situation is out of balance

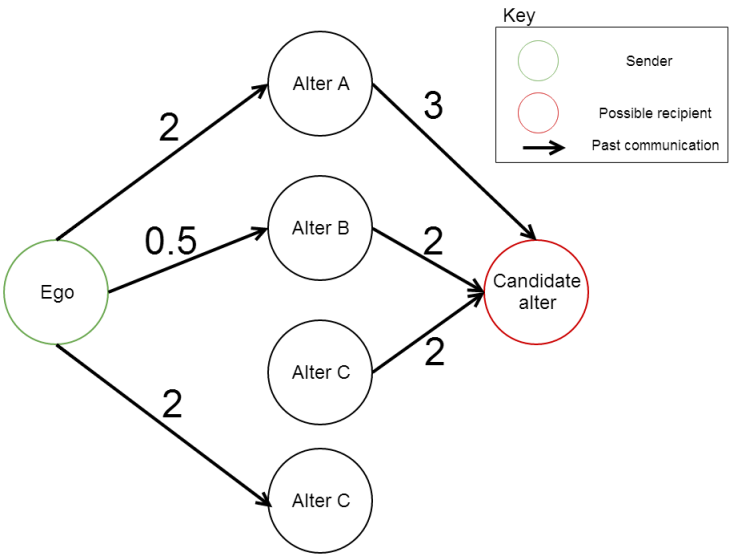


Figure 5.5: Diagram showing the past communications counted by the triadic closure effect. Only complete pathes between ego and the candidate alter count towards the value of the statistic.

unless i also has a relationship with j i.e. there is a tendency towards being friends with the friends of my friends. Since we are dealing with an event history rather than a simple binary network, we need to decide how to define the ties from i to k and k to j as well as how to combine these. In this version we decay the value of the ties to measure the strength of the link. We then follow Opsahl, Agneessens, & Skvoretz (2010) in taking the geometric mean of these numbers by taking their product and then the square root of the two sums. Since there is more than one possible actor k we sum these values for all possible actors. Equation 5.5 shows the formalization of this effect.

$$\sum_{k:k\in actors} \sqrt{\delta\big(\{e_a|i=i(e_a),k\in J(e_a)\}\big)\cdot\delta\big(\{e_b|k=i(e_b),j_a\in J(e_b)\}\big)} \tag{5.5}$$

An example of this effect is shown in figure 5.5. To aid in interpretation we have given the ties weighted values. The first thing that we notice is that alters C and D are not relevant to the statistic as there is no tie from ego to C and no tie from D to the candidate alter. The triadic value for the path from ego to the alter via A is $\sqrt{2\cdot 3}=2.45$ and the values for the path via B is $\sqrt{0.5\cdot 2}=1$. This means the total value for the triad statistic between ego and the candidate alter is 3.45.

Indegree popularity effect

The first standard behavioural effect is indegree popularity. This effect simply takes the value of the incoming ties to a candidate alter. This effect captures the tendency of people to contact those who have recently received many messages. As with most of the effects, we decay these messages to reflect the greater importance of recent behaviour for explaining future behaviour. Equation 5.6 shows the type of past events that indegree popularity is counting (those in which the candidate alter was in the recipient set).

$$\delta\big(\{e|j_a \in J(e)\}\big) \tag{5.6}$$

Recipient set triadic closure

The recipient set triadic closure variable captures the tendency for a Bureau to add recipients into the set who have been contacted a lot by those who are already in the set. In other words, whether the Bureau pays attention to the communication patterns of the Embassies its contacting. This effect is analagous to the introduction effect for the Embassies. It is defined as the average level of decayed communication from each of the recipients already added to the set to the candidate alter as explained in equation 5.7. An example of this is given in figure 5.6 which shows an example of past communications from the already chosen alters to the candidate alter. In this case they have weights of 1 and 5 to the candidate making the value of the recipient set triadic closure 3 to the candidate alter.

$$\begin{aligned} if(J_a(e) = \varnothing) &: 0 \\ else &: \\ \frac{\sum_{k:k\in J_a(e)} \delta\big(\{e|k = i(e), j_a \in J(e)\}\big)}{|J_a(e)|} & \end{aligned} \tag{5.7}$$

Introduction by Bureau Effect

The introduction by Bureau effect is included in the model predicting the sending behaviour of Embassies. It looks at whether Embassies send more messages to other Embassies if they were included together in messages from Bureaus. Importantly, this effect is included

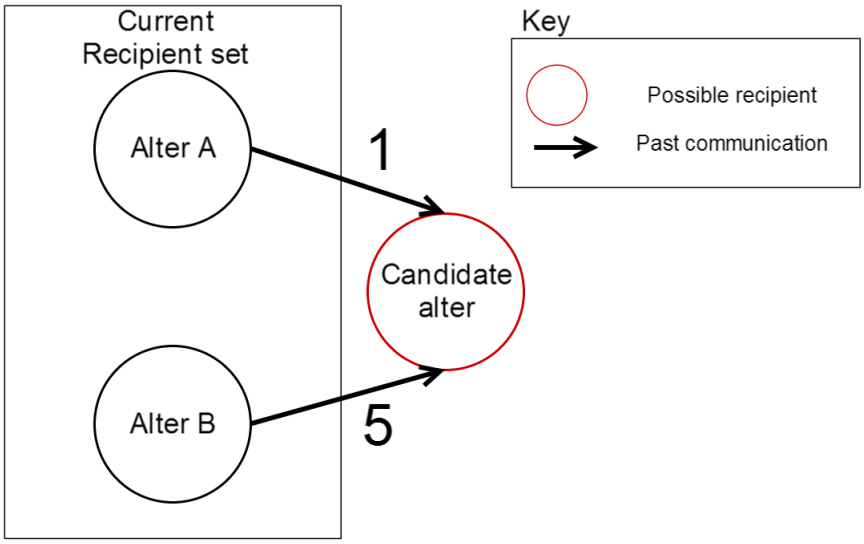


Figure 5.6: Diagram showing the ties counted by the recipient set triadic closure effect.

in the alter addition stage so it is counting an additional adoption of recipients on top of the simple inheritance effect. If this effect is important then it would suggest that Bureaus influence the sending choices of Embassies directly rather than passively.

The introduction by bureau effect measures how many messages ego was included in together with the candidate alter when the message was sent by a bureau. The more intuitive formulation of this is "who does my boss think I should be talking to?". This effect is designed to capture a direct influence of bureaus on the beliefs of embassies about which other Embassies they should direct information to. We define a new function $B(k)$ where $B(k) = 1$ if the actor k is a bureau and 0 otherwise. Consequently the introduction by bureau effect is defined as follows in equation 5.8.

$$\delta\big(\{e|B(i(e)) = 1, i \in J(e), j_a \in J(e)\}\big) \tag{5.8}$$

For now we only consider introduction by Bureaus. This is because we are trying to look at the interplay between the sending of information between Embassies and sending of information to multiple Embassies by Bureaus. These are the two networks shown to be closely related in the previous chapter. Future work should expand the number of potential influences further.

5.7.3 Recipient inheritance

As we showed in section 5.4.3, recipient inheritance accounts for a substantial proportion of all communication in the dataset. Given the stickiness of recipient decisions within conversations, the original initiator of a conversation potentially has agenda setting power over the conversation. We examine this by looking at which actors appear at the beginning of chains of messages.

A citation chain starts with an initial message that cites no other messages. This chain can then be continued as other cables cite the original message or subsequent messages in the chain. Importantly, all messages in a chain can be traced back to the original message. Chains can be woven together when a cable cites multiple existing chains. In this case, there will not necessarily be a single cable that all messages can be traced back to.

Some chains end in a cable that has outgoing but missing citations, often because the cable that is cited was not included in the dataset. In these cases we exclude the chain because it would introduce a selection bias to include those cables that cite missing cables.

As well as looking just at the number of chains started we also look at how long these chains are and estimate the total number of recipients that are able to be traced to a cable initiated by a Bureau or an Embassy.

- H6: cables written by Embassies initiate conversations with many more messages that those written by Bureaus
- H7: cables written by Bureaus initiate conversations with many more messages that those written by Embassies
- H8: cables written by Bureaus and Embassies initiate conversations with similar numbers of total messages

In addition to the direct impact of inheriting recipients on communication structure, there could also be indirect effects. A possible mechanism for this would be self-perception theory in social psychology, where individuals infer their attitudes or beliefs from their own actions (Bem 1967). This mechanism tends to be stronger where individuals would have had little cost to changing their behaviour. Given the relative simplicity of adding or removing alters from a message, this condition would seem to hold for recipient inheritance.

The self-perception theory has long been heavily contested (Fazio, Zanna, & Cooper 1977; Zanna, & Cooper 1974). However, there are reasons to think that it may be more plausible when the actors are offices rather than individuals. Since multiple people write messages from each Embassy, they do not directly observe which recipients on a message were actively added and which were inherited. Consequently, when a different person reads over a message sent from their office they may internalize all the recipients as being relevant rather than just the ones that were actively added by the drafter. In this way, actors would learn which other embassies to add to messages based on who they had contacted in the past (regardless of whether they had added those actors themselves). This mechanism sees the conversation structure as feeding back into social structure.

- H9: Bureaus are more likely to choose to add recipients who they inherited in the past
- H10: Embassies are more likely to choose to add recipients who they inherited in the past

To examine this mechanism, we define two new effects.

Recipient Habit Effect

The first of these effects is the habit effect. This simply captures whether the actor contacts the same others over time. It is defined as the decayed values of past events where ego sent a message to the candidate alter as shown in equation 5.9. A higher weight on this effect would mean that actors have a stronger tendency to repeatedly contact the same alters. Importantly, for the current hypotheses, this includes both recipients that ego added actively and those she inherited.

$$\delta\big(\{e|j_a = i(e), i \in J(e)\}\big) \tag{5.9}$$

Addition Habit Effect

The recipient habit effect defined in equation 5.9 captures the tendency to contact the same others regardless of ego's agency in adding these alters previously. This is important

because recipient inheritance means that many of the actors that ego has contacted in the past were not added actively but simply inherited. The addition habit effect captures the tendency of ego to add the same actors as he has *added* previously. Importantly, this is a different set of people than those who he *contacted* previously.

As discussed, self-perception theory would suggest that actors might attribute past choices of recipients to themselves due to a tendency to attribute agency to past actions, even where those actions were not entirely free choices. In this way we might expect to see recipient habit to predict future communications even after accounting for those recipients that ego actually chose to add with the addition habit effect.

The addition habit effect is defined the same way as the habit effect but only counts recipients if they were part of ego's recipient set in the past event as equation 5.10 shows.

$$\delta\big(\{e|i = i(e), j_a \in J_{addition}(e)\}\big) \tag{5.10}$$

5.7.4 Additional control effects

This final set of effects are not directly important to the hypotheses under study in this chapter. However, they are important to include because they capture potentially important influences on the sending behaviour of actors in the State Department.

Alter response Effect

Another standard effect in network models is reciprocity. This just captures ego's tendency to form ties with actors who have made ties to ego. We simply use the decayed values of cables that the candidate alter sent to ego as in equation 5.11. The graph also gives another view of the event we are trying to capture. This effect is important as a control but may also be related to if actors' pay attention to the actions of other actors. However, we should avoid overinterpreting this effect as it will be confounded with any unincluded covariate between countries that is also reciprocal. Since we are only including trade in this model we avoid giving a substantive interpretation to the alter response effect.

$$\delta\big(\{e|j_a \in J(e)\}\big) \tag{5.11}$$

Alter same country

The same country variable is defined as 1 if the ego and candidate alter are based in the same country and is zero otherwise $X(i, country) = X(j_a, country)$. This is an important control variable for two reasons. First, there is significant coordination between State Department offices in the same country, with consulates under direct control of the embassies in the country. Second, we need to control for the fact that there is no covariate value for trade between State Department offices based in the same country. Some of the highest communication is within country, which we give a zero value for trade. The fit line (in an OLS model) would be negative despite the clear positive relationship within countries that have a defined value for trade.

By defining the trade effect as zero between offices in the same country and including a dummy variable indicating that the ego and alter are in the same country, the trade variable is actually an interaction effect between "ego and alter are not from the same country" and "trade between the host countries of ego and alter".

5.7.5 Implications of different mechanisms

There are different implications for understanding the State Department depending on which of these mechanisms is at work. Firstly, it is important to know whether Embassies or Bureaus are primarily responsible for setting the structure of information sharing. If the embassies are responsible, then information provision can be seen as an emergent activity that is independent of the views of Washington DC. It would also suggest that the State Department is likely to be resistant to changes of priority coming in from above.

By contrast, if the Bureaus are responsible, then the State Department can be seen as a more tightly controlled bureaucracy that can be more easily bent towards the needs and interests of its leadership. However, it would raise the possibility of an echo chamber where information is sourced and provided that largely agrees with the status quo instead of challenging it.

The mechanism by which this agenda setting happens is also important. Conversation structures are a side effect of a particular organizational and technological setup (cables with multiple recipients and the ability to cross-reference between messages in a standard-

ized way). If these are largely responsible for the structure of information provision, then it implies a strong role for different technologies within bureaucracies. By contrast, if the agenda is set through influence, then this is a social process that should be less dependent on particular technologies or organizational forms.

It should be noted that it is very possible that more than one of these mechanisms will be at play simultaneously.

5.7.6 Modelling approach

This section has discussed the RECHAIN model and outlined the addition effects we will use to distinguish between the different mechanisms of agenda agreement. We will run a RECHAIN addition model on a 16,500 event sample of the cables sent by Bureaus in the State Department and a second RECHAIN addition model on a 16,500 event sample of the cables sent by Embassies in the State Department.[8]

5.8 Results

This section presents the results testing the above hypotheses on the communication behaviour of Embassies and Bureaus. We present a RECHAIN addition model for the Bureau's sending behaviour and a second RECHAIN addition model for Embassies' communication behaviour. Finally, we also present additional descriptive information about which actors tend to start communication chains.

5.8.1 Preferences for communicating along trade ties

Table 5.4 shows the RECHAIN addition model for Bureaus. Importantly, we see that the Bureaus pay attention to the trade ties between Embassies when adding them into recipient sets. A candidate Embassy is more likely to be added into a recipient set if the candidate Embassy has strong trade ties with those Embassies that are already included. This means that at least some of the importance of trade ties in structuring information flows is driven by the Bureaus' choice of which recipients to include together.

[8] An important extension to this paper will be to run the analysis presented here on the full dataset of 286,283 cables rather than the percent sample used here. See the conclusion for an explanation of some of the reasons why this is difficult.

Table 5.5 shows the Embassy RECHAIN addition model. Embassies are actually slightly less likely to add recipients if the recipient's host country has high trade relations with the sender's host country.

Taking these results together with the importance of trade cross-sectionally, the positive preference for trade ties among Bureaus and a slight negative preference among Embassies implies that Bureaus must be the actors who make trade an important factor in information provision.

	Parameter	Std..Error	Explanation
Habit (recipients)	−0.03762	0.00317	Does ego provide information to the same recipients over time?
Habit (addition)	0.08948	0.00370	Does ego provide information recipients who they have added previously?
Response	0.01577	0.00030	Does ego add Embassies who have provided information to ego recently?
Indegree Popularity	−0.00114	0.00009	Does ego add Embassies who have received many recent messages?
Triadic Closure (generalized mean)	0.00598	0.00014	Does ego contact 'friends of friends'?
Exports to recipient set	0.06846	0.00882	Does ego add groups of Embassies whose countries trade with each other?
Recipient set triadic closure	0.01718	0.00037	Does ego add groups of Embassies that contact each other a lot?

n: 16500 (events sampled), 11040 (events modelled) , 23818 (additions modelled)

Table 5.4: RECHAIN model of factors influencing Bureaus' choice of which Embassies to contact

These results allow us to reject H1 and H3, that Embassies are setting the agenda in terms of trade or that there is no difference between the two. It should be noted that this finding does not rule out the possibility that Embassies do have an agenda setting effect on Bureaus, just that on the issue of trade the Bureaus' agenda trumps the Embassies'.[9]

[9]This should not necessarily be interpreted as Embassies actively avoiding communication with economically related offices. The overall picture from the previous chapter showed that trade flows explain nearly half of the variation in communications between embassies. Instead, this finding merely shows that Embassies do not add recipients with trade ties conditional on the recipients that are already included in the recipient set because of inheritance. Nonetheless, this finding shows that Bureaus are responsible for the structuring of information flow around trade.

	Parameter	Std..Error	Explanation
Habit (recipients)	−0.02170	0.00388	Does ego provide information to the same recipients over time?
Habit (addition)	0.20284	0.00557	Does ego provide information recipients who they have added previously?
Response	0.02756	0.00104	Does ego add Embassies who have provided information to ego recently?
Indegree Popularity	0.00203	0.00007	Does ego add Embassies who have received many recent messages?
Triadic Closure (generalized mean)	0.00466	0.00015	Does ego contact 'friends of friends'?
Introduction by Bureau	−0.03007	0.00131	Does ego add Embassies that a Bureau introduced them to?
Alter trade	−0.01740	0.00625	Does ego add Embassies whose host trades with ego's host?
Same country	1.94153	0.02791	Does ego add recipients based in the same country as ego?

n: 16500 (events sampled), 7540 (events modelled) , 23850 (additions modelled)

Table 5.5: RECHAIN model of factors influencing embassys' choice of which Embassies to contact

5.8.2 Recipient adoption

The second set of hypotheses refer to the tendency for one set of actors to directly react to the preferences of the other. Looking at the Bureau results in table 5.4, we see that Bureaus do have significant reactions to the communication behaviour of other actors. The recipient set triadic closure effect is positive which gives some support to the influence mechanism, where Bureaus pay attention to the communication choices of Embassies when deciding which others to include in a recipient set.

However, indegree popularity comes out negative, meaning that Bureaus are actually less likely to add Embassies as recipients if the Embassy received many cables recently. This unexpected result should probably be interpreted together with the triadic closure effect. The triadic closure effect is essentially indegree popularity weighted by the Bureau's communication with the other actors. Bureaus are still more likely to contact popular Embassies but only if those Embassies are popular among the Bureau's contacts. This could potentially reflect the specialised areas of focus of different Bureaus, with a Bureau taking a particular area of expertise.

Embassies' communication patterns also show some indications of recipient adoption. The Embassies show a strong triadic closure effect meaning an Embassy is more likely to communicate with alters who are contacted by the Embassy's closest contacts. Unlike the Bureaus, Embassies also show a positive indegree popularity effect. This means that Embassies are more likely to contact Embassies that have received many incoming message recently (a bandwagon effect). This may reflect the more generalized work of Embassies compared with Bureaus who specialise on particular areas. However, an Embassy (A) doesn't appear to be influenced by Bureaus more strongly than other offices. In fact, there is a significantly negative parameter on choosing to contact Embassies that a Bureau introduces Embassy A to. This suggests there could be a degree of pushback from the Embassies in this way.

The introduction by Bureau effect comes out negative showing that an Embassy is not more likely (and is actually less likely) to add a candidate alter if the Embassy has been included in a bureau's message with the candidate alter. This of course is already accounting for the effect of inheritance and is also conditional on the triadic closure effect.

	Bureau		Embassy	
	Number	%	Number	%
Chains	11229	31	24859	69
Spawned recipients	39754	39	62992	61

Table 5.6: Number of chains and recipients traceable back to messages from Bureaus or Embassies

Nevertheless, there does not seem to be a direct influence of Bureaus' behaviour on the preferences of Embassies.

5.8.3 Recipient inheritance

The third set of hypotheses refer to the mechanism of recipient inheritance. We already determined that recipient inheritance is a widespread phenomena in section 5.4.3 that showed that a minimum of 30% of recipients are inherited rather than added and that the percentage is probably closer to 42%. We now know that recipients added by Bureaus tend to have trade links to each other, while those added by Embassies do not have a tendency to have strong trade ties. Consequently, recipients inherited from the choices of Bureaus will tend to increase the overall importance of trade ties in the network whereas recipient choices inherited from Embassies will tend to decrease the overall importance.

Table 5.6 shows the total number of messages sent by Bureaus that can be traced back to a cable started by an Embassy or Bureau as well as the total number of chains that can be traced back. Bureaus start longer chains than Embassies on average, as they account for 31% of chains but 39% of all recipients in these chains. However, this still means that more than six out of ten recipients in a chain can be traced back to

Overall, the picture is that Bureaus are overrepresented in inheritance compared to the number of messages they send. However, there is still a high level of inheritance from messages written by both Embassies and Bureaus. Inheritance amplifies the voice of the Bureaus but it isn't sufficient for them to dominate the agenda.

Internalisation

There is no evidence of internalisation of inherited recipients in either the Embassy or Bureau communications. In both cases the two habit effects are in opposite directions with

addition habit having a strong effect but a substantial negative effect for recipients. This means that Bureaus and Embassies tend to add the same recipients over time. However, neither of them internalize recipients which they merely inherited but did not actively add. This result shows that the mechanisms of self-perception theory do not seem to play a role in the choices of Bureaus or Embassies adding recipients and we therefore reject H9 and H10.

5.8.4 Summary

Taking all of these results together, the key finding is that Embassies do not react directly to the trade ties their host country has to other Embassies' host countries. Nevertheless, the cross-sectional pattern is that Embassies communicate strongly along trade ties. This means that Embassies must include trade ties through some other mechanism than a direct preference for these types of ties. Since Bureaus do take trade ties into account directly, we can conclude that the importance of trade ties is put on the agenda by Bureaus.

The results are somewhat less clear in terms of the precise mechanism of agenda setting. Given the proportion of recipients which are inherited, this appears to be one major mechanism at play.

We can definitely reject the internalisation of inherited recipients mechanism, as the effects are strongly and consistently in the opposite direction than predicted by this mechanism.

There is mixed evidence with regards to the recipient adoption mechanisms, where an actor reacts directly to other actors' behaviour. Bureaus have a strong effect of recipient set triadic closure and triadic closure suggesting they react to the choices that other offices make including those of Embassies. Embassies also have a positive triadic closure effect meaning they react to the choices of other offices who they have strong ties to. Embassies however, appear to react against aedding recipients included by Bureaus (outside of the context of recipient inheritance).

These results do not completely disentangle the effect of recipient inheritance and recipient adoption. Since some of these ties that actors react to are the result of recipient inheritance, these two mechanisms may well reinforce each other. Further analysis will be needed to specify models that can fully distinguish between these behaviours.

5.9 Discussion

We started this article with two conflicting views about how agendas are set within the State Department. One view sees the agenda as emerging from the bottom of the bureaucracy with those closest to the changing policy environment best placed to change the priorities of the bureaucracy. The opposite view saw the lower level bureaucracy as searching for information determined by the hierarchy.

Our results, while suggesting a degree of reciprocal influence, weigh much more heavily towards the top down view of influence within the State Department. The strongest support for this view comes simply from the fact that US Embassies' overall communication patterns are strongly structured by trade (explaining nearly half of all variation) but that the RECHAIN model of their actual decisions shows that they do not directly react to these trade links but instead appear to inherit these trade related recipients from Bureaus.

We outlined and tested a number of alternative mechanisms of agenda setting in the State Department. As we have noted, there is little evidence of the shared values mechanism, as Embassies do not choose on the basis of trade ties. There is some evidence for social influence through a tendency of offices to contact actors who their close contacts communicate with. This form of triadic closure is common in many networks and may be one of the mechanisms of agenda setting here.

There is the strongest evidence for the mechanism of recipient inheritance. This mechanism makes Embassies transfer information passively by simply continuing with the terms of the existing conversation rather than starting anew. This view is reminiscent of Kissinger's view of the power of the status quo in bureaucracies, describing "the curious phenomenon that decisions taken with enormous doubt and perhaps with a close division become practically sacrosanct once adopted." (Kissinger 1966).

What are the implications for our understanding of foreign policy making in light of these conclusions? The first point to make is that minds do not have to be changed for the agenda to be set. A first mover advantage and the power of inertia is sufficient to set the terms of the conversation.

Secondly, these results suggest that it is possible for the hierarchy of a bureaucracy to change the agenda of the organization. This contrasts with a common view of foreign

policy bureaucracy generally (Sylvan, & Majeski 2003, p. 9) and the State Department in particular (Kennan 1958; Rubin 1987, p. 96) as fundamentally inertial. The results here suggest that the bureaucracy can be very responsive to changing priorities, at least at the level of Embassies' information provision. It is of course possible that the inertia comes in at the level of Bureaus rather than at the lowest levels of the bureaucracy. Future research should try to trace the agenda further up the hierarchy both within the State Department and outside of it.

At the moment we have only examined the recipient addition process within the RECHAIN model, while analysing citation selection and alter removal only descriptively. Both of these are potentially further sites of agenda change.

The removal model will allow us to examine when one actor rejects a second actor's choice of the communication agenda, by removing the second actor's chosen recipients. In particular, it will be interesting to look in more detail at possible strategic uses of recipient addition and removal, where information is withheld or contributed as part of an attempt to build a coalition around an actor's preferred policy. Similarly, citation choice may be another way of setting the agenda by choosing cables that have recipients sets of Embassies who share trade links.

Once the various selection stages are modelled, it will be possible to simulate communication streams based on these parameters. These simulations will allow us to test whether the mechanisms we have identified here are sufficient to generate the cross-sectional structure of information provision that we identified in the previous chapter.

It will also be possible to tweak the parameters of the simulation to look at the sensitivity of this information structure to different behaviours. For instance, does a lower level of recipient inheritance allow longer term agenda disparities between Bureaus and Embassies?

Importantly, the RECHAIN model has potential applications outside of the State Department. It can potentially be used to examine any communication events that form part of larger conversations. These could include memos and emails, where the reply-all option may play a similar role to recipient inheritance. Even non-communication relational events such as violence between countries (grouped into longer running disputes) or trades between firms (grouped into business relationships) might be able to be modelled using

variants of RECHAIN.

Possibly the most interesting possibility of this research will be to look at how these results generalize outside of the 1973-1974 State Department. Historians cite Kissinger's interactions with the State Department as unique and the State Department has gone through many reorganizations of Bureaus over time. As more data are released, we can begin to examine how patterns of influence change (or stay the same) under different leaders and different institutional structures.

Chapter 6

Conclusions

This thesis covers three projects; evaluating the use of Google Trends as a measure of issue salience, using corpus linguistics to assess survey translation and using social network analysis to examine the agenda of bureaucrats in the US State Department. In this chapter, I discuss the key results from each of the papers. I then discuss the links between chapters in more detail and discuss the further research topics that this thesis suggests.

Chapters 1 and 2 cover the first project on using Google Trends to measure issue salience. Chapter 1 outlines the potential benefits of Google Trends data on Internet searches to measure issue salience. These benefits include: its availability at a weekly level; its geographic coverage across many countries, sub-regions and even cities; and its breadth in capturing all searches on Google. However, the chapter argues that the inherent self-selection in Google searches means that there is no set population that is being sampled. Consequently, the value of Google Trends data rests crucially on whether it consistently tracks existing measures of issue salience. Chapter 1 asked three research questions. 1) what procedures can be used to validate Google Trends data against existing sources of public opinion data? 2) Are any Google Trends time series valid proxies for public issue salience in the United States? 3) How valid and reliable are the Google Trends data that are able to be used as proxies? This chapter outlines a replicable method for validating Google Trends data against public opinion sources drawing on methods from econometrics. The results in this chapter show that in the United States, some searches show a good fit with traditional time series of issue salience using the proportion of respondents choosing an issue as "most important". Nonetheless many other searches do not show a good fit.

Chapter 2 asks whether the Google Trends measures validated in chapter 1 are portable across contexts. 1) Can Google Trends be used to measure issue salience in countries outside of the United States, including the United Kingdom and Spain? 2) Are the same search terms valid for measuring issue salience in different country and language contexts?

The results of this chapter show that Google Trends can be used validly outside of the United States. However, while several search series are shown to be valid in Spain and the United Kingdom, there does not appear to be a consistent pattern in which searches work across different countries. Even in the United Kingdom and United States, where the same language is used, different search terms are valid in these countries. Consequently, I conclude that validity assessments of Google Trends data are not portable across contexts, meaning that a new validation is needed for each situation.

This project shows that new data sources based on monitoring behaviour on the Internet can potentially be used for social science research. However, just because the data is ubiquitous, doesn't mean that it's actually interpretable everywhere. Paradoxically, the availability of large data sources based on Internet behaviour makes survey research more useful because the surveys can be used as a benchmark to make the aggregate behavioural data interpretable.

Chapter 3 is drawn from the second project on survey translation. Within the survey methodology literature, substantial attention has been paid to the problem of differential item functioning in surveys (where respondents in one country give systematically higher responses to a question than respondents in another country, even when the underlying attitude doesn't differ). However, current studies still assume that question responses are only related directly to the variable of interest and that no other variables are introduced in the measurement process. There has been little analysis of confounding bias where the indicator is correlated with other underlying variables, particularly where this varies across contexts. Chapter 3 asks three research questions. 1) Can corpus linguistics data derived from web corpora be used to assess translation equivalence? 2) Are there translation problems in existing survey questions (drawing on the specific example of the postmaterialism items from the World Values Survey)? 3) Do these non-equivalent translations lead to predictable correlations with other questions on the survey?

The chapter makes use of a set of corpora based on web pages in 10 languages. These

corpora are used to analyse how key terms in comparative survey questions are used in different languages and in particular, whether they have the same set of associations across different languages. Using this method, I find that a commonly used question on postmaterialism in the World Values Survey has serious confounding biases in Polish, Russian, Chinese and Arabic. These linguistic associations predict confounding correlations with other items in the survey.

Chapters 4 and 5 are part of the third project on agenda setting and communication in the US State Department. To explore these concepts, this project uses an electronic database of around 300,000 diplomatic cables sent between Embassies and Bureaus in the State Department between 1973 and 1974.

Chapter 4 addresses five research questions. 1) Which issues structure the flow of information to Embassies in the US State Department? 2) Does the relative weighting of these issues differ between the Embassies in the field and the Bureaus based in Washington DC? 3) Is the issue agenda of the State Department stable over time? 4) Is communication structured around multiple issues at a time within the US State Department? 5) Does the content of diplomatic cables sent across a dyad (a link between a pair of actors) match up with the issues that the network models predict are important across that dyad?

The chapter examines these questions by looking at the predictors of the volume of messages sent between offices. The logic is that bureaucrats will prioritise sending information to other offices that they believe are most relevant to their job and that this relevance will differ depending on what issues are salient. Therefore, if they consider understanding trade relations, for example, as a particularly important issue then they will provide more information along the lines of trade links. However, if they consider understanding security relations as highly important, then they may focus on supplying information to Embassies in countries that their host country trades arms with, is allied to or is at war with.

Based on this analysis, the State Department appears to focus on several issues. However, the most important factors to communication are the level of trade between two host countries and the number of IGOs in which both hosts are members. Embassies and Bureaus focus on almost exactly the same set of factors when distributing information to Embassies. This suggests that the agenda of the State Department does not vary systematically between the Bureaus in Washington DC and the Embassies in the field. The

communication agenda is also relatively stable over time and involves processing of information on multiple issues. Finally, we use the network models to predict which dyads are likely communicate about which topics. We conduct a small scale qualitative examination of cables sent across these dyads and found that they reflected the expected content.

The fifth chapter asks whether influence over the communication agenda in the State Department comes from above or below. After reviewing historical accounts of the State Department, it is clear that there is support for both possibilities. Chapter 5 is based around three research questions. 1) What features are needed in a social network model that can accurately account for the features of the US State Department's communication network? 2) Do Embassies or Bureaus set the conversation agenda in terms of introducing trade into the structure of information provision? 3) What mechanisms of communication agenda setting are present within the US State Department's cables?

In order to distinguish the two directions of influence, we move from looking cross-sectionally at the provision of information in the State Department to using relational event models that take an individual communication as the unit of analysis. This method allows the modelling of the impact of exogenous factors (such as trade) as well as endogenous factors (such as past communications) on an actor's communication choices. We note that existing relational event models do not take account of the dependence between communication events (e.g. diplomatic cables or emails). However, in the real world communication events tend to be structured within ongoing conversations. This means that the assumption of conditional independence of events is clearly violated within existing relational event models. To deal with this dependency, we define a new relational event conversation history actor initiated network model (RECHAIN) that takes the previous message in a conversation as a starting point and asks whether the sender will add or remove recipients from the conversation. This means that in our analysis we avoid conflating the choices made by other actors with the choices made by the current sender.

By using this model, we show that only Bureaus choose recipients on the basis of trade links. Embassies actually have a slight negative preference for recipients whose host countries have strong trade links with their host country. However, given that the cross-sectional analysis in chapter 4 shows a strong effect of trade on structuring communications (nearly half of the variation in the number of messages sent across dyads is explained by

trade), we argue that this is strong prima facie evidence that Bureaus' preferences on trade relationships structure communications in the State Department, leading Embassies to communicate on the basis of these factors too.

In this chapter, we also try to examine which mechanisms of agenda setting are taking place within the State Department. The inheritance mechanism (of continuing a conversation with the same recipients as the previous message) clearly plays a role (as it accounts for 30% of all recipients contacted). There is less evidence that Embassies are directly influenced by the choices of recipients made by Bureaus. However, there is some evidence that Bureaus are influenced directly by Embassies in their choice of recipients as the RECHAIN model suggests that they send messages to groups of Embassies that often contact each other.

The overall picture of communications in the State Department is that the system is set up so that Bureaus tend to set the agenda. However, most of the mechanisms we observe are passive rather than active forms of influence. Indeed, from the results of this model we cannot even be sure that the Bureaus are aware that their preferences are structuring the provision of information in the State Department.

In the course of the five chapters of this thesis, a number of themes have emerged. Figure 6.1 shows the overlap in the three projects. All three projects are based around the idea of using new methods and new data to tackle important questions within social science. Consequently, all five chapters are relevant to the following methodological issues in social science research: measurement, reliability, triangulation and validation, cost reduction and the increasing power of computation.

In addition, there are specific thematic links between some of the chapters. Clearly, chapters 1 and 2 are strongly linked through their ongoing development of research using Google Trends. Similarly, chapters 4 and 5 focus on the same dataset of diplomatic cables and share a concern over understanding the agenda within the US State Department. There are also strong links between the different projects.

The first links are between the work on Google Trends in chapters 1 and 2 and the work on survey translation in chapter 3. Both projects share a focus on measuring public opinion; the meaning of words used in measurement and the issues with transferring measurement across different contexts. Similarly, chapters 1 and 2 share a substantive focus with chapters

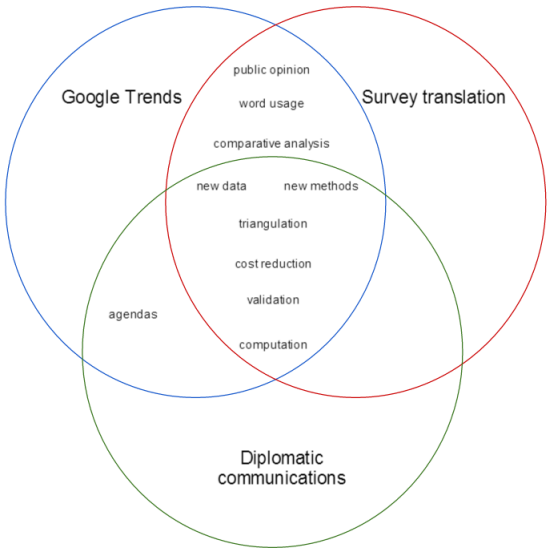


Figure 6.1: Venn diagram showing the linkages between the three projects in this thesis

4 and 5 - measuring agendas.

6.1 Links between all the projects

6.1.1 Computation

Big data has become a very fashionable buzzword within social science. *Political Analysis* released a virtual issue on 'big data' in 2013 and there have been numerous calls for integrating big data into social science research(Monroe 2013). Without engaging in a full history of the term 'big data', the term has generally been used to refer to two things; first, datasets that are sized and structured in such a way that it is difficult to analyse them with traditional tools and second, data collected incidentally, usually as a by-product of an individual's interaction with an electronic system.

As many have pointed out, social science has actually used big data in the first sense for a long time(Boyd, & Crawford 2012). For instance, censuses covering hundreds of millions of records are far larger than many datasets now termed big data. Nevertheless, there has been a genuine increase in the availability of datasets in the last decade. The second aspect of big data is relatively new and has mostly originated in the past two decades.

All three projects in this thesis touch upon both of these understandings of big data. The Google Trends data is an example of incidentally collected data. Thus, the problems I discuss in the first two chapters reflect this method of data collection. Google search data

is unproblematically relevant to some questions such as "how many Internet searches are there likely to be for this product in March?" However, moving from questions of prediction to explanation is more difficult as it requires inferring from non-representative data to a population of interest (generally the adult population).

The findings of these two thesis chapters show the limits of big data. A huge sample size and coverage cannot make a sample more representative. Ignoring this fact puts researchers back in the realm of the classic literary digest poll, mistaking sample size for validity.[1]

The second point to note about Google Trends is that it is an example of big data that has been converted into small data due to the fact that Google searches are compiled into aggregate level statistics before they appear online. While the number of searches flowing through Google's system is vast, the data that a researcher receives is small. For a given search term, the amount of data the researcher will receive is a single datapoint per week, or around 468 datapoints over 9 years (the service goes back as far as 2004). Consequently, the appropriate techniques for analysing Google Trends data are the same as those used for other time series. Chapters 1 and 2 draw on research from the econometrics literature to interpret these data.

Chapter 3 also makes use of big data in the sense of collecting incidentally created data. The large (100 million words+) corpora of different languages offer researchers a unique capability to compare natural language usage across different contexts. In addition, Internet corpora are orders of magnitude cheaper to construct than traditional corpora (Burnard 2000).[2]

There are still representativeness issues with corpora collected from the Internet. However, for the questions examined in chapter 3, the corpora are likely to be sufficiently representative of language usage. This is because the method involves looking for differences in the meaning of a small set of words between countries. Even if the sample of a language, in one of these corpora, is not perfectly representative of its usage, it is still mutually intelligible with all the other variants. Thus, the meaning of a word should be

[1] The Gallup/Literary Digest story is somewhat more complicated than many accounts have allowed for, with some accounts putting more emphasis on non-response bias than representativeness. See the following exchange on pollster.com for more details (Blumenthal 2010).

[2] I had intended to collect my own corpora for use directly. Unfortunately the search APIs needed to generate them have been commercialised and funding will be needed to construct new corpora. Even with the new pricing policy on search APIs, the cost would only come out to around $100 per corpus. See http://datamarket.azure.com/dataset/bing/search for pricing details.

broadly similar across these contexts.

Comparing the findings in this chapter with the project on Google searches, in general, the assumption that language usage on the Internet is representative of the broad differences in meanings of words is more easily defensible than the claim that changes in the volumes of a Google search term automatically represent public interest in that issue (the meaning of a word varies less across people speaking the same language than the salience of an issue does).

6.2 Computational Challenges in RECHAIN

The diplomatic communications project covered discussed in chapters 4 and 5 makes use of big data in both senses of the term. It uses communications data created as a by-product of the US State Department's communications system. It is also of a size and structure that lead to computational difficulties in running some analyses. It is perhaps surprising that the dataset in this paper that best fits the definition of big data is created from records created in 1973 and 1974. This illustrates that big data tools can apply to diverse research contexts.

Higher computational capabilities have also made some analyses available that would not have previously been feasible. The RECHAIN model in chapter 5 is an example of these computational capabilities. Within the RECHAIN model, each decision to add a recipient involves evaluating the sender's communication utility to send to each of 280 possible recipients. With 9 variables in the utility function, this means calculating 2,520 statistics for every choice. Each of these variables in turn captures such relations as "the candidate has contacted the current sender recently" or "the candidate is heavily embedded in the current sender's local network". To calculate these types of relations it is necessary to count patterns in the past events at the time of every event.

The upshot of this modelling approach is that the 15,000 event/25,000 decision model of Embassy communications involves 63,000,000 statistic calculations, each of which requires searching through several thousand events. Even with many optimisations and running on three computers the model takes around two weeks to run. Running the model over the full dataset (rather than the 5% sample used here) would take many months to run. Without

modern computers, it would be impossible to run this analysis at all. I wrote the program entirely in R which is a relatively slow language (despite recent improvements in R 3.0). It also scales poorly, as it requires all objects to be in memory at all times. I also wrote a less memory intensive database solution in couchDB but this slows the performance further.

The time required for running long models – such as RECHAIN – changes the analysis workflow. In many ways, it encourages good practices of generating many descriptive statistics prior to running inferential models and also spending a substantial amount of time considering which variables to include in the model. However, the long computation times reduces the flexibility to run diagnostic tests and alternative specifications of a model. For instance, testing different values of the decay parameter would have been a useful sensitivity check in chapter 5. However, this was simply not practical to run on a realistic timescale.

The fact that a RECHAIN model can be run at all, speaks to a major advantage of relational event models over other social network analysis. The other two major forms of social network analysis, exponential random graph models and stochastic actor oriented models, both rely heavily on simulation of networks to estimate parameters. This means that it is impossible to use sampling because every modelling step relies on previously simulated ties.

By contrast, relational event models take the observed events to be fully constitutive of the network. This means that all events can be modelled independently of one another, as the modelling for each event relies only on the observed data and not on anything calculated for any other event. This means that a 25,000 choice sample could be drawn from a dataset of 300,000 events (as in this case) or 3 billion events and not take substantially longer to run. Each of the 25,000 choices modelled can use information from the 3 billion events (although only a tiny subset are likely to be relevant), but they are only used as context and only a tiny fraction need be modelled. Provided that we are happy with the level of information from 25,000 cases, models can be fitted that represent the tendencies of communication patterns in any sized dataset. To my knowledge, this thesis is the first academic work to make use of sampling in a relational event model rather than modelling the entire population of events.[3]

[3]Stadtfeld (2012) uses subsets of events in stable periods but takes smaller slices of time rather than

6.2.1 Validation and triangulation

Although the techniques described in this paper are different from the standard survey/content analysis techniques often used in the past, they are intended as complementary methods of analysing these problems.

A repeated theme of this thesis is the need for the validation of novel data sources. The need for validation is demonstrated directly in chapters 1 and 3. Chapter 1 shows that most plausible Google search terms are not strongly correlated with existing measures of the public agenda. It is possible that they are tapping into another aspect of issue salience. However, the burden of proof is surely on those who want to make the claim that these searches measure a novel form of issue salience. The conclusion of chapter 1 is that most published research using Google Trends is probably not using valid measures.

A similarly clear example of the need for validation is shown in chapter 3, where a key term in a commonly used question in the World Values Survey's postmaterialism index is shown to have different meanings across several languages. In particular, the translations of the word ideas appear to have a substantial impact on the correlations of the postmaterialism variable with other values in the countries where the meaning differs.

This leads to odd trends over time, as shown in figure 6.2. The standard postmaterialism argument claims that as societies get richer they tend to become more postmaterialist in outlook, but the results for China show the opposite result. Nearly 40% of respondents picked the mistranslated postmaterialism option in a survey question comparing different values in 1990 compared with only around 15% preferring the option in 2007. This change would be largely inexplicable without the information that the 'ideas' item correlates positively with traditional values in China, rather than the strong negative correlation observed in almost every other country. This difference is most likely attributable to the translation problems discussed.

In theory, a control mechanism against these measurement issues should be the mere presence of odd empirical results such as these. However, there are enough theories about genuine differences between societies that these results can be interpreted as genuine differences. For instance, Zhu, & Rosen (1993) argued that post-materialist values have a

sampling events directly.

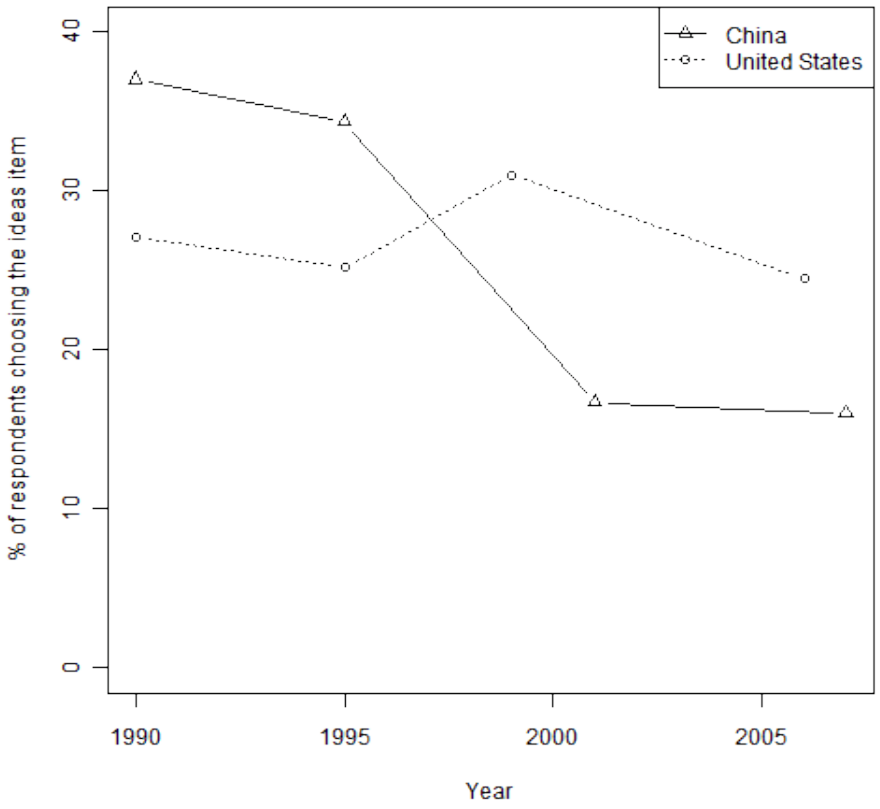


Figure 6.2: The trends in the proportion of respondents choosing the 'ideas' item in different waves of the World Values Survey in the United States and China

negative relationship with protest in China –the opposite relationship to that in Western countries. While Zhu, & Rosen (1993) use a different post-materialism item, this example shows that different relationships between variables in different contexts can always be given a substantive as well as measurement interpretation. Given that there are generally greater professional benefits within social science to making a surprising substantive claim than to making a measurement/methodology one, such errors appear to be able to persist unchecked for a long time.

Problems with the translation of large cross-national surveys have the potential to undermine a large quantity of research. Because of the high cost of running these surveys, the datasets generated tend to be used heavily by social scientists. A document on the World Values Survey website lists 3,443 'recent' articles that cite the World Values Survey (World Values Survey 2008).

6.2.2 Factors affecting the adoption of methodological innovations

New data sources are valuable not only because they allow us to address questions that couldn't have been addressed previously, but in some cases they also allow us to address questions at a lower cost to researchers. Where there is a valid public opinion time series to validate against, Google Trends data can potentially provide very frequent measurement of issue salience at essentially zero cost, compared with the high cost of conducting surveys at that frequency.

Similarly, using corpus linguistics to validate existing survey translations puts a diagnostic tool within realistic reach of survey users as well as those who collect the data. It is of course possible that someone analysing the data could have used a team of translators to test the validity of surveys they were using. However, it is notable that these translation issues do not seem to have been reported in the academic literature in the 32 years of the question's use (admittedly several of these translations are more recent). It is possible that some of these translation issues have been referenced briefly in a footnote of a paper at some point, but there is no simple source for a user of the question to consult. Since writing this chapter, several academics have told me that they know of concerns with the translation of the World Values Survey questions, but this seems to be an example of 'tacit knowledge' (Gerholm 1990; Gertler 2003; Cowan, David, & Foray 2000) rather than based

on published studies.

Cost and convenience are genuine concerns for the adoption of better methodology in social science and contributing to these will improve the practice of the disciplines.

The network methods in chapters 4 and 5 also hold the possibility of reducing the cost of reading through archives. This is not because the methods replace reading, but because they can potentially structure the reading of documents. If a researcher wants to answer a specific set of questions based on an archive of correspondence, then understanding the structure of communication will help to pinpoint the documents most likely to be relevant to their questions. This thesis demonstrates a simple application of this approach in chapter 4, where the predictors of communication patterns are used to find cables on specific topics.

However, more sophisticated searches should also be possible using these techniques. If there is a particular relational behaviour that we believe is important in a case, we would sample "likely events" i.e. those events where the model predicts a high likelihood of the chosen recipient and where the independent variable of interest is also high. This should allow behavioural sampling. Future research should expand our understanding of the link between network behaviour and the content of messages.

6.3 Links between specific chapters

6.3.1 Linking the meaning of words to comparative measurement

The first two projects (consisting of chapters 1, 2 and 3) share a focus on measuring public opinion, particularly across countries. One of the key issues both projects face is how the meaning of words in different countries affects comparative analysis.

Chapter 3 shows how even a question from a widely used survey can include various confounding biases due to differences in the meaning of key terms across countries. Chapter 2 shows that there are comparable problems with the portability of Google Trends variables across countries. This is despite the fact that Google Trends data is often billed as a more universal data source than traditional surveys. Chapter 2 shows that even across countries that speak the same language (the United Kingdom and the United States), the same search terms do not work consistently. As I discuss in chapter 2, these differences appear to partially stem from the different phrases used in political discourse in different

countries. These findings largely undermine the most enthusiastic claims made for Internet search data. Internet data's availability across time periods and countries does not help us understand anything about the world unless we have a reliable method for making inferences from searches to populations.

6.3.2 Agendas

This thesis uses both the Google Trends data and the diplomatic communications data to measure issue agendas. Both projects measure agendas based on measurements of behaviour rather than attitudes. In the case of Google Trends, the act of seeking information online is taken be a possible indicator of a belief about the importance of an issue. In the diplomatic cables, the distribution of information across Embassies is hypothesised to indicate the priorities of those distributing this information.

In both cases, it is important to check that the content of the behaviour is actually what we assumed. For Google Trends this means that the searches for a term should actually reflect the content we expect. The classic example is that we need to be sure that searches for 'jobs' refer to broader topic of employment rather than, for instance, to Steve Jobs.

Similarly, the coefficients for the importance of different issues within the State Department's communications could reflect the importance of these issues or other omitted variables. To make sure that the cables reflect concern with the issues indicated by the statistical models, we read the text of a sample of the cables predicted to be substantial, due to strong trade ties or strong IGO links. In both cases, a high proportion of the cable traffic referred to these issues.

6.4 Limitations

As is always the case in research, there are a number of limitations to the work in this thesis. The limitations are discussed in each chapter but it is worth considering some of the most important ones here.

As discussed above, the Google Trends data has a number of limitations in its usage, including the portability of measures and the need for data to validate against.

In addition, the method requires the researcher to define the initial search terms to test.

Since there seems to be a link between terms used in political discourse and the search terms found to be valid, an in-depth understanding of the specific political system may be necessary to find suitable Google search terms. This reduces the ability of researchers to find good search terms to analyse, unless they have a strong grounding in the political discourse of a country they are researching.

Additionally, there is no guarantee that any variable will necessarily have a suitable search term (or terms) that can effectively act as a proxy for the variable. For instance, the Iraq war was a highly salient issue in the US, but there are no search terms I could find that were reasonable proxies of public perceptions of the importance of this issue.

The corpus method also has limitations. Its application is limited to words that appear regularly enough within a corpus to reliably assess what other words (collocations) they appear with. However, if this limitation is present, then it is worth considering whether the term used in a survey is well known enough that all respondents will be expected to understand it. Another limitation of this approach is that the method can only indicate if there are non-equivalencies and what concepts they might be related to. At present, the method is not designed to correct these non-equivalences in analysis or to suggest better translations.

The diplomatic data is also limited in several ways. First, there are a significant number of cables missing from our dataset because some cables are still classified as top secret (which probably accounts for a relatively small proportion of the total but are likely to be of greater significance) or unimportant (which account for a very high volume but are probably not reducing the substantive content of the cable communication infrastructure). There is some additional missing information in certain fields of the cables, particularly the references field that is often misformatted. Future research in this project will attempt to reduce any data lost in this way.

The second limitation is that we only observe one channel of communication between diplomats. The dataset does not include phone calls or face-to-face meetings. However, we do have the records of all messages that diplomats wanted to commit to paper (at least in terms of information sent to other parts of the department). We also only currently considering the internal communications of the US State Department and are not considering where the preferences of Bureaus and Embassies originate.

Finally, we do not currently consider all aspects of the cable sender's decision process. However, we have fully outlined and implemented all aspects of the RECHAIN model including ego selection, citation selection and alter removal. These stages of selection will provide the opportunity to test further mechanisms of agenda setting.

6.5 Future research

6.5.1 A change in methodology

As I discussed above, adoption of methodology is partially a function of cost and convenience. Therefore, I am working on developing the processes from the first three chapters of the thesis into free R packages that will streamline the process of checking survey questions or search indices for other researchers.

6.5.2 Wider validation of surveys

The goal with the R corpus translation validation package will be to simplify the process so that an applied researcher could run these checks on secondary data without a large time investment. It is also worth noting that a validation need only be run once for each survey item. Therefore it would also be useful to set up a system for tracking translation quality ratings for different survey questions. This would allow secondary researchers to consult a source where any known translation problems with a survey question would be reported (as well as checks that did not indicate a problem).

6.5.3 Communication models

Although diplomatic cable communications have a number of idiosyncratic features (e.g. the possibility of replying/citing multiple previous messages at once and the requirement that a Bureau is always included as a recipient), the model outlined here should be applicable to other communications data that allows for multiple recipients, such as emails or memos sent within many organizations. For example, in the Enron email dataset (the largest open-source email dataset) 62% of messages are part of email chains (Klimt, & Yang 2004). Although we should be careful about inferring too much from this one dataset, it suggests that email communication is likely to follow the same general pattern of behaviour

as the cables and hence that the RECHAIN model could be very useful for analysing email data.

A key extension of this line of research should be linking the internal behaviour of the State Department with that of other political actors. We have found that the flow of information in the State Department is largely directed by the Bureaus. This means we have traced the source of the communication priorities as far upstream as the bureaucrats based in Foggy Bottom and largely excluded an endogenous source of influence among the network of diplomats themselves. However, we do not yet know how these priorities are formulated within the State Department. Are these priorities influenced by other political actors, the media or the intelligence services? It would also be useful to see how information transmission outside the department affects broader political agendas. For instance, the NSA and CIA heavily consume cable traffic even before it has been interpreted by analysts in Bureaus, so the information sent by Embassies may have direct effects on the agendas of other actors.

To look at the interplay between external agendas and the priorities within the State Department, we will probably need a longer time period than the 22 months used in the current study. NARA have recently released additional cables covering 1973 to 1976, although it is not in an easily accessible format. They also plan to eventually release cables through 1988, but the process of declassification means that this will probably take some time.[4]

As part of the analysis of wider foreign policy agendas, the archives of National Security Study Memorandum and National Security Study Decisions will be useful. These are a set of matched documents detailing aspects of foreign policy that the President wants information on with the decision eventually taken on that area of policy. Tracing the route of these documents and decisions through the communications of the State Department could allow us to better analyse the link between communications, bureaucracy and policy formation.

[4]Documents are meant to be declassified automatically after 25 years but this is still subject to manual checking for sensitive details and personal information which is a time consuming process that can only be undertaken by staff with security clearances.

6.5.4 Thermostatic RECHAIN Models

The current RECHAIN models assume that each office has a value they place on the utility of contacting offices with certain trade ties to their own office. However, it is also possible that these apparent preferences for trade ties are really preferences for a certain level of communication about trade within the State Department. This would be directly analogous to the thermostatic model of policy (Wlezien 1996), where actors prefer a certain level of policy and demand more or less of it depending on whether the current level of policy exceeds their preferred level or not. Hence, political pressure creates negative feedback that tends to push policy back to equilibrium over the long term. Relative preferences can potentially be modelled within RECHAIN by interacting the sender's preference for communicating across trade ties with the amount of recent communication traffic along trade ties that the sender has received. In other words, a sender would then potentially send more trade messages when the volume is lower than their preferred level and send fewer messages about trade when the current volume is above their preferred level.

6.5.5 Tracing influence in other political institutions

The State Department is not the only institution where we could potentially look for agenda setting and political influence using these techniques. Clearly we are limited in the institutions we can examine due to data constraints.[5] However, there are several existing sources of data that would allow further hypotheses about political influence within political institutions to be examined.

First, 2.4 million emails within the Syrian government have been publicly released, covering the period 2006 to 2012. Whilst there will be many things to learn from these data, it would be particularly interesting to look at where power is located within this authoritarian regime. Are there communities of dissent within the government? Are different parts of the bureaucracy played off against each other? How centralised is decision making? What informal groups within the government influence this decision making? Do interests within the bureaucracy organise to set the agenda?

[5] Hopefully the emails from the George W. Bush White House will eventually be made public, and this would be an interesting archive to examine using these methods (see http://www.georgewbushlibrary.smu.edu/Research/Electronic-Records.aspx).

The second applied case extends the relational event models to cover both communications and voting behaviour. I have constructed a dataset of the letters between delegates to the Continental Congress (1774-1789) and their voting behaviour (both in the Congress and in the Constitutional Convention, for those that attended). These could potentially be used to examine whether scholars' success in explaining voting behaviour according to the delegates' economic interests (see for instance, Mcguire, & Ohsfeldt (1984)) is primarily because the delegates voted rationally or because delegates were more open to persuasion by delegates with similar class backgrounds.

All of the papers in this thesis show the benefit to periodically reassessing whether existing lines of research could be improved and extended using new techniques and data resources. When these resources are changing as quickly as they are presently, new links between different areas of research are feasible for the first time. Whether it is drawing links between foreign policy bureaucracies and social network analysis; issue salience and internet search behaviour or corpus linguistics and survey methodology; there are benefits to combining insights and methods from multiple sources.

In addition to the substantive topics explored in the thesis, these methodological advances are potentially useful for scholars in many fields. This thesis should help future research on public agendas, research using cross-national survey data or research that uses communication data. Since many different areas of social science are concerned with these areas, the methods developed in this thesis should have wide applicability.

Bibliography

Abbott, Kenneth W., & Duncan Snidal (1998). "Why States Act through Formal International Organizations". *Journal of Conflict Resolution* 42.1, pp. 3–32.

Adler, Emanuel (1992). "The emergence of cooperation: national epistemic communities and the international evolution of the idea of nuclear arms control". *International Organization* 46.01, pp. 101–145.

Arel-Bundock, Vincent (2013). *countrycode: Convert country names and country codes.*

Atkins, Sue, & Michael Rundell (2008). *The Oxford guide to practical lexicography.* Oxford: Oxford University Press.

BACM Research (2012). *Department of State Electronic Telegrams: 1973 - 1974.* URL: `http://www.paperlessarchives.com/dos\_telegrams.html` (accessed on 07/30/2013).

BNC Consortium (2007). *The British National Corpus.* URL: `http://www.natcorp.ox.ac.uk/` (accessed on 10/01/2012).

Baker, Paul (2011). "Times May Change, But We Will Always Have Money: Diachronic Variation in Recent British English". *Journal of English Linguistics* 39.1, pp. 65–88.

Baldry, Anthony, & Paul Thibault (2009). "Multimodal turn-taking". In: *International Symposium on Multimodal Approaches to Communication.* Vol. 41, p. 12.

Barbieri, Katherine, Omar M. G. Keshk, & Brian M. Pollins (2009). "Trading Data Evaluating our Assumptions and Coding Rules". *Conflict Management and Peace Science* 26.5, pp. 471–491.

Bartle, John, & Samantha Laycock (2012). "Telling more than they can know? Does the most important issue really reveal what is most important to voters?" *Electoral Studies* 31.4, pp. 679–688.

Baumgartner, Frank, Bryan Jones, Scott Adler, & John Wilkerson (2006). *Public Policy Agendas Project Topic Codebook*. URL: http://www.policyagendas.org/page/topic-codebook (accessed on 10/20/2012).

Bem, Daryl J. (1967). "Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena". *Psychological Review* 74.3, pp. 183–200.

Bjornskov, Christian (2010). "How Comparable are the Gallup World Poll Life Satisfaction Data?" *Journal of Happiness Studies* 1.1, pp. 41–60.

Bélanger, Éric, & Bonnie M. Meguid (2008). "Issue salience, issue ownership, and issue-based vote choice". *Electoral Studies* 27.3, pp. 477–491.

Bloom, Paul (2002). *How children learn the meanings of words*. The MIT Press.

Blumenthal, Mark (2010). *Dr. George Gallup and the Literary Digest Poll*. URL: http://www.pollster.com/blogs/dr_george_gallup_and_the_liter.php?nr=1 (accessed on 09/30/2013).

Bollen, Kenneth A (2002). "Latent variables in psychology and the social sciences". *Annual Review of Psychology* 53.1, pp. 605–634.

Boulton, Andrew, Lomme Devriendt, Stan Brunn, Ben Derudder, & Frank Witlox (2011). "City networks in cyberspace and time : using Google hyperlinks to measure global economic and environmental crises". In: *ICTs for Mobile and Ubiquitous Urban Infrastructures*. Ed. by Firmino, Rodrigo J., Duarte, Fabio, & Ultramari, Clovis. Curitiba: IGI Global, pp. 67–87.

Boyd, Danah, & Kate Crawford (2012). "Critical Questions for Big Data". *Information, Communication & Society* 15.5, pp. 662–679.

Brosius, HansBernd, & Hans Mathias Kepplinger (1992). "Linear and nonlinear models of agendasetting in television". *Journal of Broadcasting & Electronic Media* 36.1, pp. 5–23.

Burnard, Lou (2000). "Where did we Go Wrong? A Retrospective Look at the British National Corpus". In: *Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Ed. by Kettemann, Bernhard, & Marko, Georg. Rodopi, pp. 51–70.

Burr, William (2006). *U.S. National Archives Web Site Uploads Thousands of Diplomatic Cables*. URL: http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB188/ (accessed on 07/30/2013).

Butts, Carter T. (2008). "A Relational Event Framework for Social Action". *Sociological Methodology*, pp. 155–200.

— (2013). *sna: Tools for Social Network Analysis.*

Carmines, Edward G., & Richard A. Zeller (1979). *Reliability and Validity Assessment.* London: SAGE.

Carrier, Norman H (1959). "A note on the measurement of digital preference in age recordings". *Journal of the Institute of Actuaries [JIA]* 85, pp. 71–85.

Ceobanu, Alin M., & Xavier Escandell (2010). "Comparative Analyses of Public Attitudes Toward Immigrants and Immigration Using Multinational Survey Data: A Review of Theories and Research". *Annual Review of Sociology* 36.1, pp. 309–328.

Chaffin, Roger, Robin K Morris, & Rachel E Seely (2001). "Learning new word meanings from context: A study of eye movements." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1, pp. 225–235.

Choi, Seung-Whan, Patrick James, & Youngwan Kim (2013). "U.S. Military Intervention: Mass Media versus Expert Influence". *via personal correspondence.* Working Paper.

Chong, Dennis, & James N. Druckman (2010). "Dynamic Public Opinion: Communication Effects over Time". *American Political Science Review* 104.04, pp. 663–680.

Clarke, Harold D., Allan Kornberg, Thomas J. Scotto, & Marianne C. Stewart (2012). "Political Choices in Hard Times: Voting in the 2010 U.S. House Elections". *Journal of Elections, Public Opinion & Parties* 22.2, pp. 139–165.

Cleveland, Robert, William Cleveland, Jean Mcrae, & Irma Terpenning (1990). "STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)". *Journal of Official Statistics* 6, pp. 3–73.

Cowan, Robin, Paul David, & Dominique Foray (2000). "The explicit economics of knowledge codification and tacitness". *Industrial and Corporate Change* 9.2, pp. 211–253.

Cremer, Jacques (1980). "A partial theory of the optimal organization of a bureaucracy". *The Bell Journal of Economics* 11.2, pp. 683–693.

Davidson, Russell, & James G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Davis Cross, Mai'a K. (2008). "A European Epistemic Community of Diplomats". In: *The Diplomatic Corps as an Institution of International Society*. Ed. by Sharp, Paul, & Wiseman, Geoffrey. Vol. 11. Basingstoke: Palgrave Macmillan, pp. 221–245.

De Pauw, Guy, Filles-Maurice De Schryver, & Peter W Wagacha (2009). "A corpus-based survey of four electronic Swahili English bilingual dictionaries". *Lexikos* 19.1, pp. 340–352.

Dekker, David, David Krackhardt, & Tom A. B. Snijders (2007). "Sensitivity of MRQAP tests to collinearity and autocorrelation conditions". *Psychometrika* 72.4, pp. 563–581.

Department of State (2012a). *5 FAH-1 H-210 How to Use Telegrams*. URL: `http://www.state.gov/documents/organization/89319.pdf` (accessed on 08/12/2013).

— (2012b). "Preparing Telegrams in the Department". *US Department of State Foreign Affairs Manual* 5.1, pp. 1–18.

Drezner, Daniel W. (2000). "Ideas, Bureaucratic Politics, and the Crafting of Foreign Policy". *American Journal of Political Science* 44.4, pp. 733–749.

Du, Jia Tina, & Amanda Spink (2011). "Toward a web search model: Integrating multitasking, cognitive coordination, and cognitive shifts". *Journal of the American Society for Information Science and Technology* 62.8, 1446–1472.

Dube, Arindrajit, & Ethan Kaplan (2012). "Occupy Wall Street and the Political Economy of Inequality". *The Economists' Voice* 9.3, pp. 1–7.

Dunning, Ted (1993). "Accurate methods for the statistics of surprise and coincidence". *Computational linguistics* 19.1, pp. 61–74.

Ellis, Curtis, Colin Swearingen, & Joseph T. Ripberger (2011). "Examining the Impact of Public Attention on Fundraising in U.S. Senate Elections". In: *SSRN eLibrary*.

Eurobarometer (2006). *E-communications Household Survey 2005*. Special 249.

— (2012). *E-communications Household Survey 2011*. Special 381.

Evans, Warren (2012). "Petroleum". In: *Digest of United Kingdom Energy Statistics 2012*. Ed. by MacLeay, Iain, Harris, Kevin, & Annut, Anwar. Digest of United Kingdom Energy Statistics (DUKES). London: Department of Energy and Climate Change.

Ewing, Raymond (2002). *Interview with Mr. Robert Hopper*. URL: `http://memory.loc.gov/cgi-bin/query/D?mfdip:82:./temp/~ammem\_pVRl::`.

Ewing, Raymond (2005). *Interview with Ambassador Denis Lamb*. URL: http://memory.loc.gov/cgi-bin/query/D?mfdip:2:./temp/~ammem_dlym::.

Fazio, Russell H., Mark P. Zanna, & Joel Cooper (1977). "Dissonance and self-perception: An integrative view of each theory's proper domain of application". *Journal of Experimental Social Psychology* 13.5, pp. 464–479.

Ferrantino, Michael J., & Aimee Larson (2009). "Transmission of the global recession through US trade". In: *The Great Trade Collapse: Causes, Consequences and Prospects*. CEPR, pp. 171–180.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, & Bor-Chung Chen (1998). "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program". *Journal of Business & Economic Statistics* 16.2, 127–152.

Finnemore, Martha (1996). *National Interests in International Society*. Ithaca: Cornell University Press.

Ford, Nigel, David Miller, & Nicola Moss (2001). "The role of individual differences in Internet searching: An empirical study". *Journal of the American Society for Information Science and Technology* 52.12, 1049–1066.

Fox Translate (2012). *Translation Services*. URL: http://www.foxtranslate.com.

Fuchs, Christian (2009). "The Role of Income Inequality in a Multivariate Cross-National Analysis of the Digital Divide". *Social Science Computer Review* 27.1, 41–58.

Gajduschek, Gyorgy (2003). "Bureaucracy: Is It Efficient? Is It Not? Is That The Question?: Uncertainty Reduction: An Ignored Element of Bureaucratic Rationality". *Administration & Society* 34.6, pp. 700–723.

Gallup (2010). "US National Polls 2004-2010". In: *The Roper Center, University of Connecticut*.

Garrison, Jean A. (1999). *Games Advisors Play: Foreign Policy in the Nixon and Carter Administrations*. College Station: Texas A&M University Press.

Gerholm, Tomas (1990). "On Tacit Knowledge in Academia". *European Journal of Education* 25.3, pp. 263–271.

Gertler, Meric S. (2003). "Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there)". *Journal of Economic Geography* 3.1, pp. 75–99.

Ghosn, Faten, Glenn Palmer, & Stuart A. Bremer (2004). "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description". *Conflict Management and Peace Science* 21.2, pp. 133–154.

Gibler, Douglas M (2013). "Release Notes for Version 4 : Correlates of War Formal Interstate Alliance Dataset , 1816-2012 Data Files Included with Release". *Release notes* 4, pp. 1–5.

Gilberg, Sheldon, Chaim Eyal, Maxwell McCombs, & David Nicholas (1980). "The State of the Union Address and the Press Agenda". *Journalism & Mass Communication Quarterly* 57.4, pp. 584–588.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, & Larry Brilliant (2008). "Detecting influenza epidemics using search engine query data". *Nature* 457.7232, pp. 1012–1014.

Glick, Reuven, & Andrew K. Rose (1999). "Contagion and trade". *Journal of International Money and Finance* 18.4, pp. 603–617.

Google (2004). *Google Trends*. http://www.google.com/trends. URL: `http://www.google.com/trends` (accessed on 10/19/2012).

— (2012). *Google Trends - Web Search Interest: ron paul, mitt romney, herman cain, rick santorum, newt gingrich - Worldwide, Sep 2011 - Jul 2012.*

Graefe, Andreas, & J. Scott Armstrong (2012). "Predicting elections from the most important issue: A test of the take-the-best heuristic". *Journal of Behavioral Decision Making* 25.1, 41–48.

Granger, Clive W. J., & Paul Newbold (1974). "Spurious regressions in econometrics". *Journal of Econometrics* 2.2, pp. 111–120.

Granka, Laura A. (2010). "Measuring Agenda Setting with Online Search Traffic: Influences of Online and Traditional Media". In: *SSRN eLibrary*.

Greenhill, Brian (2010). "Norm Transmission in Networks of Intergovermental Organizations". PhD thesis. University of Washington.

Haas, Peter M. (1992). "Introduction: Epistemic Communities and International Policy Coordination". *International Organization* 46.1, pp. 1–35.

Halperin, Morton H., & Priscilla Clapp (2006). *Bureaucratic Politics and Foreign Policy*. Washington, D.C.: Brookings Institution Press.

Hammond, Thomas H (1986). "Agenda Control, Organizational Structure, and Bureaucratic Politics". *American Journal of Political Science* 30.2, pp. 379–420.

Harkness, Janet (2003). "Questionnaire translation". *Cross-cultural survey methods* 325, p. 35.

Harkness, Janet, Ana Villar, & Brad Edwards (2010). "Translation, Adaptation, and Design". In: *Survey Methods in Multinatuional, Multiregional, and Multicultural Contexts*. Ed. by Harkness, Janet, Braun, Michael, Edwards, Brad, Johnson, Timothy, Lyberg, Lars, Mohler, Peter, Pennell, Beth-Ellen, & Smith, Tom. Hoboken: John Wiley and Sons.

Harrell Jr., Frank E, & Contributions from many other users. (2010). *Hmisc: Harrell Miscellaneous*.

Heath, Anthony, Stephen Fisher, & Shawna Smith (2005). "The globalization of public opinion research". *Annual Review Political Science* 8, pp. 297–333.

Hedstrom, Peter (1998). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.

Hook, Steven W. (2003). "Domestic Obstacles to International Affairs: The State Department Under Fire at Home". *PS: Political Science & Politics* 36.1, pp. 23–29.

Hovland, Carl I., Irving L. Janis, & Harold H. Kelley (1953). *Communication and persuasion; psychological studies of opinion change*. New Haven: Yale University Press.

Hundt, Marianne, Nadja Nesselhauf, & Carolin Biewer (2007). *Corpus linguistics and the web*. Vol. 59. Amsterdam: Rodopi.

Inglehart, Ronald, & Christian Welzel (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge: Cambridge University Press.

Ipsos Mori (2012). *Ipsos MORI Issues Index Archive*. URL: `http://www.ipsos-mori.com/researchpublications/researcharchive/2420/Issues-Index-Archive.aspx?view=wide` (accessed on 10/06/2012).

Jacobs, Lawrence R., & Benjamin I. Page (2005). "Who Influences U.S. Foreign Policy?" *American Political Science Review* 99.1, pp. 107–123.

Jennings, Will (2009). "The Public Thermostat, Political Responsiveness and Error-Correction: Border Control and Asylum in Britain, 1994–2007". *British Journal of Political Science* 39.04, pp. 847–870.

Jennings, Will, & Peter John (2009). "The Dynamics of Political Attention : Public Opinion in the United and the Queen ' s Speech Kingdom of Manchester". *American Journal of Political Science* 53.4, pp. 838–854.

Jennings, Will, & Christopher Wlezien (2011). "Distinguishing Between Most Important Problems and Issues?" *Public Opinion Quarterly* 75.3, pp. 545–555.

John, Peter, & Will Jennings (2010). "Punctuations and Turning Points in British Politics: The Policy Agenda of the Queen's Speech, 1940–2005". *British Journal of Political Science* 40.03, pp. 561–586.

Johnson, Timothy P (1998). "Approaches to equivalence in cross-cultural and cross-national survey research". *ZUMA Nachrichten Spezial* 3, pp. 1–40.

Jones, Bryan D. (1994). *Reconceiving Decision-Making in Democratic Politics: Attention, Choice, and Public Policy*. Chicago: University of Chicago Press.

Jones, Bryan D., & Frank R. Baumgartner (2005). *The Politics of Attention: How Government Prioritizes Problems*. Chicago: University of Chicago Press.

Jones, Bryan D., Heather Larsen-Price, & John Wilkerson (2009). "Representation and American Governing Institutions". *The Journal of Politics* 71.01, pp. 277–290.

Jones, Daniel M., Stuart A. Bremer, & J. David Singer (1996). "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns". *Conflict Management and Peace Science* 15.2, pp. 163–213.

Kahn, Matthew E., & Matthew J. Kotchen (2010). *Environmental Concern and the Business Cycle: The Chilling Effect of Recession*. Working Paper 16241. National Bureau of Economic Research.

Kapstein, Ethan B (2002). "Allies and Armaments". *Survival* 44.2, pp. 141–155.

Kaufman, Leonard, & Peter J Rousseeuw (1990). *Finding groups in data: an introduction to cluster analysis*. Vol. 39. Wiley Online Library.

Kennan, George F. (1958). "America's Administrative to Its World Response Problems". *Daedalus* 87.2, pp. 5–24.

Kennedy, Charles Stuart (1996). *Interview with Richard H. Solomon*. URL: `http://memory.loc.gov/cgi-bin/query/D?mfdip:49:./temp/~ammem\_BRFZ::`.

— (1997). *Interview with Frank C. Carlucci III*. URL: `http://memory.loc.gov/cgi-bin/query/D?mfdip:9:./temp/~ammem\_cc7t::`.

Kennedy, Charles Stuart (2007). *Interview with Mr. Lawrence Cohen*. URL: http://memory.loc.gov/cgi-bin/query/D?mfdip:4:./temp/~ammem_aMNn::.

King, Gary, Christopher J L Murray, Joshua A Salomon, & Ajay Tandon (2004). "Enhancing the validity and cross-cultural comparability of measurement in survey research". *American Political Science Review* 98.01, pp. 191–207.

Kissinger, Henry A . (1966). "Domestic Structure and Foreign Policy". *Daedalus* 95.2, pp. 503–529.

Klimt, Bryan, & Yiming Yang (2004). "The Enron Corpus: A New Dataset for Email Classification Research". In: *Machine Learning: ECML 2004*. Ed. by Boulicaut, Jean-François, Esposito, Floriana, Giannotti, Fosca, & Pedreschi, Dino. Vol. 3201. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 217–226.

Koehler-Derrick, Gabriel (2013). "Quantifying Anecdotes: Google Search Data and Political Developments in Egypt". *PS: Political Science & Politics* 46.02, pp. 291–298.

Lang, Duncan Temple (2010). *RCurl: General network client interface for R*.

— (2011). *XML: Tools for parsing and generating XML within R and S-Plus*.

Langbart, David, William Fischer, Lisa Roberson, & The National Weights and Measures Laboratory (2007). *Appraisal of records covered by N1-59-07-3-P*. Tech. rep. College Park: National Archives.

Lebovic, James H., & Eric Voeten (2009). "The Cost of Shame: International Organizations and Foreign Aid in the Punishing of Human Rights Violators". *Journal of Peace Research* 46.1, pp. 79–97.

Leetaru, Kalev, & Philip A Schrodt (2013). "GDELT : Global Data on Events, Location and Tone, 1979-2012". *International Studies Association Annual Meeting, San Francisco*. International Studies Association, pp. 1–49.

Lospinoso, Josh A., Brian Godsey, Per Block, & Jonathan Mellon (2013). "Reveal : Relational EVEnt AnaLysis". *Internal Red Owl Paper*.

Lotus Development Corporation (1996). *2005 Computerworld Honors Program: CABLEXPRESS*. URL: http://www.cwhonors.org/search/his_4a_detail.asp?id=2469.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, & Mia Hubert (2005). "Cluster Analysis Basics and Extensions".

Majeski, Stephen, & David Sylvan (1999). "How foreign policy recommendations are put together: A computational model with empirical applications". *International Interactions* 25.4, pp. 301–332.

Manning, Christopher D, Hinrich Schütze, & MITCogNet (1999). *Foundations of statistical natural language processing*. Vol. 999. Boston: MIT Press.

May, Peter J., Samuel Workman, & Brian D. Jones (2007). "Organizing Attention: Responses of the Bureaucracy to Agenda Disruption". *Journal of Public Administration Research and Theory* 18.4, pp. 517–541.

McCombs, Maxwell E., & Donald L. Shaw (1972). "The Agenda-Setting Function of Mass Media". *Public Opinion Quarterly* 36.2, pp. 176–187.

McCombs, Maxwell (2005). "A Look at Agenda-setting: past, present and future". *Journalism Studies* 6.4, pp. 543–557.

McDonald-Johnson, Kathleen M., Brian Monsell, Ryan Fescina, Roxanne Feldpausch, Catherine C. Harvil Hood, & Monica Wroblewski (2010). *Seasonal Adjustment Diagnostics*. Tech. rep., pp. 1–5.

Mcguire, Robert A, & Robert L Ohsfeldt (1984). "Economic Interests and the American Constitution: A Quantitative Rehabilitation of Charles A. Beard". *The Journal of Economic History* 44.2, pp. 509–519.

Mellon, Jonathan (2014). "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience". *Journal of Elections, Public Opinion & Parties* 24.1, pp. 45–72.

Michel, Jean-Baptiste, Yuan K Shen, Aviva P Aiden, Adrian Veres, Matthew K Gray, Joseph P Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, & Others (2011). "Quantitative analysis of culture using millions of digitized books". *Science* 331.6014, p. 176.

Microsoft (2011). *Bing Translator API*. URL: `http://msdn.microsoft.com/en-us/library/dd576287.aspx`.

Moen, Mathew C (1988). "The political agenda of Ronald Reagan: A content analysis of the State of the Union messages". *Presidential studies quarterly* 18.4, pp. 775–785.

Monroe, Burt (2013). *Virtual Issue: Big Data in Political Science*. URL: http://www.oxfordjournals.org/our_journals/polana/virtualissue4.html (accessed on 09/30/2013).

MyGengo (2012). *Translation Services*. URL: www.mygengo.com.

Newey, Whitney K., & Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". *Econometrica* 55.3, pp. 703–708.

Nickles, David Paul (2003). *Under the Wire: How the Telegraph Changed Diplomacy*. Boston: Harvard University Press.

Nielson, Daniel L., & Michael J. Tierney (2003). "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform". *International Organization* 57.02, pp. 241–276.

Niemi, Richard G., & Larry M. Bartels (1985). "New Measures of Issue Salience: An Evaluation". *The Journal of Politics* 47.04, pp. 1212–1220.

Nye, Christopher D, Brent W Roberts, Gerard Saucier, & Xinyue Zhou (2008). "Testing the measurement equivalence of personality adjective items across cultures". *Journal of Research in Personality* 42.6, pp. 1524–1536.

O'Keeffe, Anne, & Michael McCarthy (2010). *The Routledge handbook of corpus linguistics*. 1st. Routledge handbooks in applied linguistics. London: Routledge.

O'Keeffe, Anne, & Steve Walsh (2012). "Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education". *Corpus Linguistics and Linguistic Theory* 8.1, pp. 159–181.

Oltra, Christian (2011). "Stakeholder perceptions of biofuels from microalgae". *Energy Policy* 39.3, pp. 1774–1781.

Opsahl, Tore, Filip Agneessens, & John Skvoretz (2010). "Node centrality in weighted networks: Generalizing degree and shortest paths". *Social Networks* 32.3, pp. 245–251.

Ozden, Çağlar, Christopher R. Parsons, Maurice Schiff, & Terrie L. Walmsley (2011). "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000". *The World Bank Economic Review* 25.1, pp. 12–56.

Paradis, Emmanuel, Julien Claude, & Korbinian Strimmer (2004). "{APE}: analyses of phylogenetics and evolution in {R} language". *Bioinformatics* 20, pp. 289–290.

Pelc, Krzysztof J. (2013). "Googling the WTO: What Search-Engine Data Tell Us About the Political Economy of Institutions". *International Organization* 67.03, pp. 629–655.

Pendar, Nick, & Serge Sharoff (2009). "Supervised Lexical Acquisition for Persian from a Web Corpus". *Computational Approaches to Arabic Script-based Languages*, p. 106.

Pérez, Efrén O (2011). "The Origins and Implications of Language Effects in Multilingual Surveys: A MIMIC Approach with Application to Latino Political Attitudes". *Political Analysis* 19.4, pp. 434–454.

Perloff, Richard M. (2010). *The Dynamics of Persuasion: Communication and Attitudes in the Twenty-First Century*. New York: Routledge.

Pevehouse, Jon, Timothy Nordstrom, & Kevin Warnke (2004). "The Correlates of War 2 International Governmental Organizations Data Version 2.0". *Conflict Management and Peace Science* 21.2, pp. 101–119.

Pew (2010). *Pew December 2010 Survey Dataset*. http://pewinternet.org/Shared-Content/Data-Sets/2010/December-2010–Social-Side-of-the-Internet.aspx. URL: `http://pewinternet.org/Shared-Content/Data-Sets/2010/December-2010--Social-Side-of-the-Internet.aspx` (accessed on 10/20/2012).

Poe, Steven C., Sabine C. Carey, & Tanya C. Vazquez (2001). "How are These Pictures Different? A Quantitative Comparison of the US State Department and Amnesty International Human Rights Reports, 1976-1995". *Human Rights Quarterly* 23.3, pp. 650–677.

Price, Vincent (1992). *Public Opinion*. Newbury Park: SAGE.

Pump, Barry (2011). "Beyond Metaphors : New Research on Agendas in the Policy Process". *Policy Studies Journal* 39.1, pp. 1–12.

Quigley, Robert (2009). *Why Big Government Could Make Andrew Breitbart the Right's Huffington*. URL: `http://www.mediaite.com/online/could-andrew-breitbart-become-the-rights-arianna-huffington/`.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

RePass, David E. (1971). "Issue Salience and Party Choice". *The American Political Science Review* 65.2, pp. 389–400.

Reher, Stefanie (2012). "Substantive Representation and Satisfaction with Democracy in the 2009 German Federal Elections: The Case for Issue-Salience Congruence". *Elections, Public Opinion and Parties Conference 2012, Oxford*.

Reilly, Shauna, Sean Richey, & J. Benjamin Taylor (2012). "Using Google Search Data for State Politics Research An Empirical Validity Test Using Roll-Off Data". *State Politics & Policy Quarterly* 12.2, pp. 146–159.

Reis, Ben, & John Brownstein (2010). "Measuring the impact of health policies using Internet search patterns: the case of abortion". *BMC Public Health* 10.1, p. 514.

Ripberger, Joseph T. (2011). "Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness". *Policy Studies Journal* 39.2, 239–259.

Rosati, Jerel A., & Scott DeWitt (2012). "The Department of State". In: *Routledge Handbook of American Foreign Policy*. Ed. by Hook, Steven W., & Jones, Christopher M. London: Routledge, pp. 178–192.

Rosati, Jerel A., & James M. Scott (2010). *The Politics of United States Foreign Policy*. Boston: Cengage Learning.

Rubin, Barry (1987). *Secrets of State*. New York: Oxford University Press.

Russett, Bruce M., & John D. Sullivan (2009). "Collective Goods and International Organization". *International Organization* 25.4, pp. 845–865.

Russian Academy of Sciences (2007). *The Russian National Corpus*. URL: `http://ruscorpora.ru` (accessed on 04/01/2010).

Sampas, Dorothy M. (1987). "The Role of the Department of State in the Foreign Policy Decision-Making Process". *Strategic Studies Project Report*.

Scharkow, Michael, & Jens Vogelgesang (2011). "Measuring the Public Agenda using Search Engine Queries". *International Journal of Public Opinion Research* 23.1, pp. 104–113.

Scheitle, Christopher P. (2011). "Google's Insights for Search: A Note Evaluating the Use of Search Engine Data in Social Research". *Social Science Quarterly* 92.1, 285–295.

Scott, Andrew M (1969). "Department of State: Formal Organization and Informal Culture". *International Studies Quarterly* 13.1, pp. 1–18.

Sharoff, Serge (2011). *Leeds Collection of Internet Corpora*. URL: `http://corpus.leeds.ac.uk/internet.html` (accessed on 09/30/2012).

Shephard, Mark (2007). "Multiple Audiences, Multiple Messages? An Exploration of the Dynamics between the Party, the Candidates and the Various Constituencies". *Journal of Elections, Public Opinion & Parties* 17.2, pp. 181–201.

Silver, Nate (2012). *Maine Caucus Offers an Opportunity for Paul*. URL: `http://fivethirtyeight.blogs.nytimes.com/2012/02/11/maine-caucus-offers-an-opportunity-for-paul/` (accessed on 10/18/2012).

Siniver, Asaf (2008). *Nixon, Kissinger, and U.S. Foreign Policy Making: The Machinery of Crisis*. Cambridge: Cambridge University Press.

Smith, Raymond F. (2011). *The Craft of Political Analysis for Diplomats*. Washington, D.C.: Potomac Books, Inc.

Snijders, Tom A.B., Gerhard G. van de Bunt, & Christian E.G. Steglich (2010). "Introduction to stochastic actor-based models for network dynamics". *Social Networks* 32.1, pp. 44–60.

Solan, Lawrence M, & Peter M Tiersma (2004). "Author identification in American courts". *Applied linguistics* 25.4, pp. 448–465.

Soroka, Stuart N. (2002). "Number of Responses and the Most Important Problem". *Nuffield College Politics Working Paper Series*.

— (2003). "Media, Public Opinion, and Foreign Policy". *The Harvard International Journal of Press/Politics* 8.1, pp. 27–48.

Soroka, Stuart N., & Christopher Wlezien (2009). *Degrees of Democracy: Politics, Public Opinion, and Policy*. 1st ed. Cambridge: Cambridge University Press.

Stadtfeld, Christoph (2012). *Events in Social Networks*. Karlsruhe: KIT Scientific Publishing.

StatCounter (2012). *StatCounter Global Stats*. URL: `http://gs.statcounter.com` (accessed on 11/11/2012).

Stata Corporation (1999). *Stata statistical software*. Stata Corp.

Stegmueller, Daniel (2011). "Apples and Oranges? The Problem of Equivalence in Comparative Research". *Political Analysis* 19.4, pp. 471–487.

Stinnett, Douglas M., Jaroslav Tir, Paul F. Diehl, Phillip Schafer, & Charles Gochman (2002). "The Correlates of War (Cow) Project Direct Contiguity Data, Version 3.0". *Conflict Management and Peace Science* 19.2, pp. 59–67.

Stockholm International Peace Research (2012). *Sipri: Trade Registers*. URL: `http://portal.sipri.org/publications/pages/transfer/trade-register` (accessed on 07/30/2013).

Sylvan, David, & Stephen Majeski (2003). "An Agent-Based Model of the Acquisition of U.S. Client States". *Annual Convention of the International Studis Association, Portland*.

Thanopoulos, Aristomenis, Nikos Fakotakis, & George Kokkinakis (2002). "Comparative evaluation of collocation extraction metrics". In: *3rd International Conference on Language Resources and Evaluation*. Vol. 2. Citeseer, pp. 620–625.

Tibi, Bassam (1997). *Arab nationalism: between Islam and the nation-state*. Houndmills: Palgrave Macmillan.

US Energy Information Administration (2012). *Gasoline and Diesel Fuel Update*. URL: `http://www.eia.gov/petroleum/gasdiesel/` (accessed on 10/20/2012).

United Nations (2002). *International Migration Report 2002*. Tech. rep. New York.

Uppsala Conflict Data Program (2013). *UCDP Conflict Encyclopedia*. Tech. rep. Uppsala: Uppsala University.

Verba, Sidney (1971). "Cross-national survey research: The problem of credibility". *Comparative methods in Sociology: Essays on trends and applications*. Ed. by Vallier, Ivan, pp. 309–356.

Warnes, Gregory R., Ben Bolker, & Thomas Lumley (2013). *gtools: Various R programming tools*.

Warnes, Gregory R, with contributions from Ben Bolker, Gregor Gorjanc, Gabor Grothendieck, Ales Korosec, Thomas Lumley, Don MacQueen, Arni Magnusson, Jim Rogers, & Others (2011). *gdata: Various R programming tools for data manipulation*.

Warwick, Donald P. (1979). *A Theory of Public Bureaucracy: Politics, Personality, and Organization in the State Department*. Boston: Harvard University Press.

Weeks, Brian, & Brian Southwell (2010). "The Symbiosis of News Coverage and Aggregate Online Search Behavior: Obama, Rumors, and Presidential Politics". *Mass Communication and Society* 13.4, pp. 341–360.

Weiner, Rachel (2012). *Google shows how debate hurt Rick Santorum*. URL: `http://www.washingtonpost.com/blogs/the-fix/post/google-shows-how-debate-hurt-rick-santorum/2012/02/23/gIQAZYoDVR_blog.html` (accessed on 10/18/2012).

Wickham, Hadley (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

— (2011). "The Split-Apply-Combine Strategy for Data Analysis". *Journal of Statistical Software* 40.1, pp. 1–29.

Willis, Gordon, Martha Kudela, Kerry Levin, Alicia Norberg, Debra Stark, Barbara Forsyth, Pat Brick, David Berrigan, Frances Thompson, Deidre Lawrence, & Anne Hartman (2010). "Evaluation of Multistep Survey Translation Process". In: *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Ed. by Harkness, Janet, Braun, Michael, Edwards, Brad, Johnson, Timothy, Lyberg, Lars, Mohler, Peter, Pennell, Beth-Ellen, & Smith, Tom. Hoboken: John Wiley and Sons.

Wlezien, Christopher (1996). "Dynamics of Representation: The Case of US Spending on Defence". *British Journal of Political Science* 26.1, pp. 81–103.

— (2005). "On the salience of political issues: The problem with 'most important problem'". *Electoral Studies* 24.4, pp. 555–579.

Wood, B Dan, & Jeffrey S Peake (1998). "The Dynamics of Foreign Policy Agenda Setting". *The American Political Science Review* 92.1, pp. 173–184.

Wooldridge, Jeffrey M. (2009). *Introductory Econometrics: A Modern Approach*. Cengage Learning.

Workman, Samuel, Bryan D. Jones, & Ashley E. Jochim (2009). "Information Processing and Policy Dynamics". *Policy Studies Journal* 37.1, pp. 75–92.

World Values Survey (2008). "Recent publications with references to World Values Survey". *World Values Survey Website*.

— (2011). *World Values Survey Website*. URL: `http://www.worldvaluessurvey.org/`.

Xuan, H. W., W. Li, & G. Y. Tang (2012). "An Advanced Review of Hybrid Machine Translation (HMT)". *Procedia Engineering* 29, pp. 3017–3022.

Zanna, Mark P., & Joel Cooper (1974). "Dissonance and the pill: an attribution approach to studying the arousal properties of dissonance." *Journal of personality and social psychology* 29.5, pp. 703–709.

Zhu, Jian-Hua, & Stanley Rosen (1993). "From Discontent To Protest: Individual-Level Causes of the 1989 Pro-Democracy Movement in China". *International Journal of Public Opinion Research* 5.3, pp. 234–249.

Appendix A

Appendices for Internet Search Indices and Issue Salience

1.1 Content Validity Results

1.2 Glossary

- MIP – questions that ask respondents their opinion on the most important problem facing our country. There are a number of minor variants but the wording used in these surveys is: "what do you think is the most important problem facing this country today?" I(0) and I(1) - These refer to the order of integration of a time series. I(1) means that a time series is integrated order 1. The order refers to the number of times that the series must be differenced in order to return a stationary time series.

- Search engine – A search engine is a website used to find webpages by entering relevant words (or search terms) into a query. In this paper the Google search engine is used.

- Search term – the word or phrase used in an internet search on a search engine. For instance, typing "illegal immigration" into Google means that "illegal immigration" is a search term.

- Search index – the time series tracking the frequency in usage of a search term. The

Table 1.1: Content validity problems for each search index

Issue	Search Index	Problematic searches	Passes content validity check
Macroeconomics	Economy	Fuel economy, foreign countries	No
Macroeconomics	US economy	Greece, China	Yes
Macroeconomics	Jobs	Steve, blow	Yes
Macroeconomics	Unemployment	None	Yes
Macroeconomics	Tax burden	None	Yes
Macroeconomics	Taxes	Terms related to financial planning	No
Macroeconomics	Fiscal	Spanish and Portuguese terms related to finance	No
Macroeconomics	Budget Deficit	None	Yes
Iraq	Iraq	Beheading, Abu Ghraib	No
Iraq	Iraq War	None	Yes
Iraq	Iraq Casualties	Afghanistan, Vietnam	Yes
Healthcare/Hospital	Healthcare	None	Yes
Government	Government	Jobs	No
Government	Big government	None	Yes
Immigration	Illegal immigration	None	Yes
Immigration	Immigration	Visas, green cards, etc.	No
Terrorism	Terrorism	None	Yes
Terrorism	Terrorism or Al Qaeda	None	Yes
Fuel Prices	Fuel	None	Yes
Fuel Prices	Gas	None	Yes

search index for "illegal immigration" would be a measure of how many times the search term "illegal immigration" is used on Google each week.

- STL – Seasonal trend decomposition by LOESS. This is a technique for decomposing a time series into seasonal, trend and remainder components. The seasonal component captures the part of the series that varies seasonally (for instance if searches were higher in summer and lower in winter). The trend component is a LOESS smoothed curve through the data. Finally, the remainder component captures the variation not captured in the other 2 components. This means that, by definition, the $data = trend + seasonal + remainder$. This paper removes the seasonality by just recombining the trend and remainder components so that: $adjusted = trend + remainder$.

- X-12 – is a set of seasonality detection tools developed by the US Census Bureau. In this paper it is used to decide which series to adjust using STL.

Appendix B

Appendices for Validating Survey Translation Using Corpus survey translation

2.1 Computational Details

The analysis of the meanings of words is performed in R (R Development Core Team 2004). The functions for retrieving the collocations from the Leeds Collection of Internet Corpora uses the RCurl (Lang 2010) and XML (Lang 2011) packages. Manipulation of the text and data structures is handled using functions from the base package (R Development Core Team 2011), Harrel's miscellaneous package (Harrell Jr. & users. 2010), and the gdata package (Warnes et al. 2011).

The divisive hierarchical DIANA clustering algorithm is implemented in the cluster package (Maechler et al. 2005). Additional checks on the dissimilarities data are performed using the Analyses of Phylogenetics and Evolution package (Paradis, Claude, & Strimmer 2004).

The analysis of the World Values Survey is performed using Stata (Stata Corporation 1999).

2.2 Derivation of Relationship between Confounding Error Component and Individual Attributes

2.2.1 Error Component Definitions

The random error component is uncorrelated with either the trait: or any other individual attributes, measured or unmeasured, X_{ik}, as shown in Equation B.1. However, the magnitude of random error may vary according to X_{ik} or θ_{ik}. For instance, lower educated respondents may be more likely to give random answers to a question.

$$Cov(\varepsilon_{rik}, X_{ik} + \theta_{ik}) = 0 \tag{B.1}$$

The confounding variable component of the error is that part of the measurement error that correlates with other attributes of the respondent, as expressed in Equation B.2.

$$|Cov(\varepsilon_{cik}, X_{ik})| > 0 \tag{B.2}$$

Response biases that affect the confounding variable component can be problematic even within a single country, as confounding bias means that the variable is measuring something other than it is meant to.

The uniform error component, ε_{uik},is a constant systematic bias that is independent of any individual characteristics or the trait, as shown in Equation B.3.

$$\varepsilon_{uik} = a_k \tag{B.3}$$

The specification error component, ε_{sik}, is correlated with the trait: $|Cov(\varepsilon_{sik}, \theta_{ik})| > 0$. This is to allow for biases such as digit preference, where respondents disproportionately report ages ending in zero or five (Carrier 1959) or tendencies for the happiest people to under-report their true happiness. The specification error component is uncorrelated with all other individual attributes, as shown in Equation B.4.

$$Cov(\varepsilon_{sik}, X_{ik}) = 0 \tag{B.4}$$

Response bias generally includes the uniform and confounding components, but these

should be analysed separately, as they are problematic in different situations and call for different responses.[1]

2.2.2 Item Equivalence Conditions

Item equivalence can be defined as the situation where for any individual with true value θ'_i and a vector of all other attributes, X'_i, the probability density function of their measured scores, $pdf(\hat{\theta}'_i)$, is the same regardless of country.[2] Combining this with the definitions of the error components in Section 2.1 gives the following four conditions for equivalence in Equations B.5-B.8.

Firstly, the variance of the random error component should be constant, b, across all countries after accounting for the effect of covariates on the magnitude of ε_{rik}, as shown in Equation B.5.

$$\forall k : var(\varepsilon_{rik}) | X_{ik} = a \tag{B.5}$$

Secondly, the covariance between the confounding error component and all properties of an individual respondent must be constant, a, across countries k, as shown in Equation B.6.

$$\forall k : Cov(\varepsilon_{cik}, X_{ik}) = b \tag{B.6}$$

Thirdly, the uniform component should be constant across all countries and individuals, as shown in Equation B.7.

$$\forall k : \varepsilon_{uik} = c \tag{B.7}$$

Finally, covariance of the specification component with the trait should be constant across all countries and individuals, following Equation B.8.

[1] The random and specification components can also be affected by survey design and response errors but they are outside the scope of the paper.

[2] This is a relatively weak form of equivalence that merely requires that the item does not affect measurement differences across countries. A stronger form of equivalence would require that given θ'_i, $pdf(\hat{\theta}'_i)$ is the same regardless of k or X_i. This would require the same item equivalence conditions as the weak form but also that ε_{rik}'s magnitude is independent of X_{ik} and that ε_{cik} is zero. The effects of ε_{rik} and ε_{cik} are possible to account for within statistical models, if they are constant across countries. Hence the weaker form of equivalence is also important.

$$\forall k : Cov(\varepsilon_{sik}, \theta_{ik}) = d \tag{B.8}$$

To see whether there is a violation of confounding error component's equivalence, we look at a small subset of X_{ik}, Z_{ik}, and its effect on the errors. Since Z_{ik} is a subset of X_{ik}, shown by Equation B.9, the relationships between X_{ik} and the error components also hold for Z_{ik}.[3]

$$Z_{ik} \subset X_{ik} \tag{B.9}$$

Consequently, the covariance of the trait and the potential confounding variable can be expressed in terms of the error components in Equation B.10.

$$Cov(\hat{\theta}_{ik}, Z_{ik}) = Cov(\theta_{ik} + \varepsilon_{rik} + \varepsilon_{cik} + \varepsilon_{uik}\varepsilon_{sik}, Z_{ik}) \tag{B.10}$$

Given the covariance of the random, uniform and specification components with X_{ik} from Equations B.1, B.3, and B.4 respectively, this can be simplified to Equation B.11.

$$Cov(\hat{\theta}_{ik}, Z_{ik}) = Cov(\theta_{ik} + \varepsilon_{cik}, Z_{ik}) \tag{B.11}$$

In addition, we assume that the true relationship is equivalent across countries in Equation B.12.

$$\forall k : Cov(\theta_{ik}, Z_{ik} = e) \tag{B.12}$$

Combining this assumption with the relationship between the measured score and confounder derived in Equation B.11, it is simple to show that the confounding error component's equivalence condition from Equation B.6, that the covariance of ε_{cik} and X_{ik} is constant across countries, cannot hold.

Language cluster	Tradition important	Fight for your country	Proud of nationality	Importance of God
Western Europe	21,808	71,036	81,681	84,936
Chinese	1,664	4,701	4,933	2,514
Russian/Polish	3,835	13,072	15,036	13,023
Arabic	5,345	7,175	14,184	14,223

Table 2.1: Number of observations used in each logit analysis by language cluster

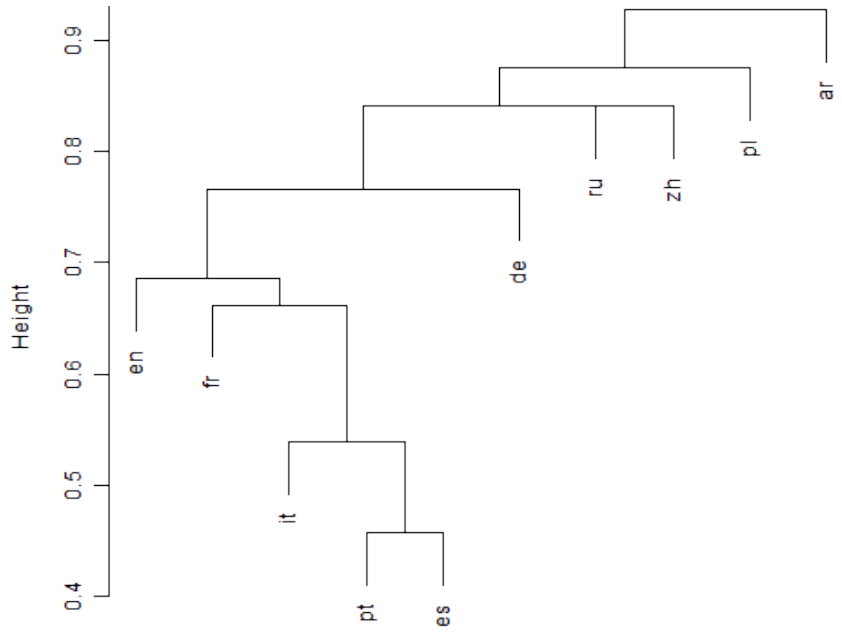


Figure 2.1: Hierarchical clustering dendogram of *economy* dissimilarities

2.3 Number of Observations used in Logistic Regressions

2.4 Linguistic Analysis of Other Words in postmaterialism Question

Before the effects of the differing meanings of *ideas* can be predicted, it is necessary to look at the translation of other important words in the same question.

2.4.1 A Stable Economy

The word *economy* shows little evidence of clustering, as can be seen in Figure 2.1. The languages break off at regular intervals without clear groups of languages. The collocations in the possible cluster of English, French, Italian, Portuguese and Spanish shows no

[3] Z_{ik} is referred to as a single variable here but the result also applies to a vector of confounding variables.

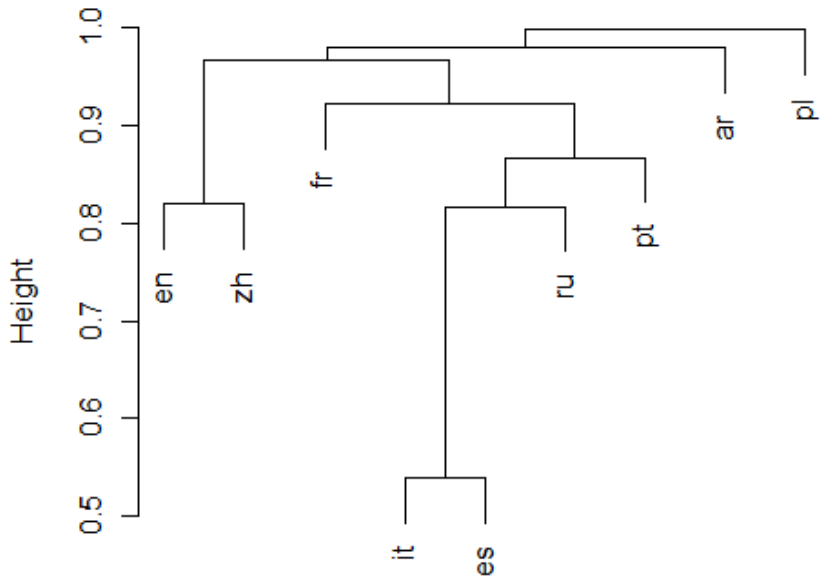


Figure 2.2: Hierarchical clustering dendogram of *fight* dissimilarities

difference in the types of collocations from the rest of the languages.

2.4.2 The Fight Against Crime

The translations of *fight* shows differences in meaning across languages and some evidence of clustering in Figure 2.2, with Italian and Spanish being particularly close, as are English and Chinese. Examination of the shared collocations reveals that, in Chinese and English, a metaphor with violence is prominent, whereas all the other languages use terms that are more general ways of reducing something. In other words, the option "the fight against crime" refers to the same policy stance in every language (reducing crime), but the metaphor of combat is absent from the majority of translations. Although the exact meanings of the word *fight* may differ, it is not clear that it should greatly affect responses to the question, as it is clear that the same type of policies are being referred to in all languages.

The word *crime* shows no important differences between languages. All languages show associations with unlawful behaviour and responses to it, such as police and prison.

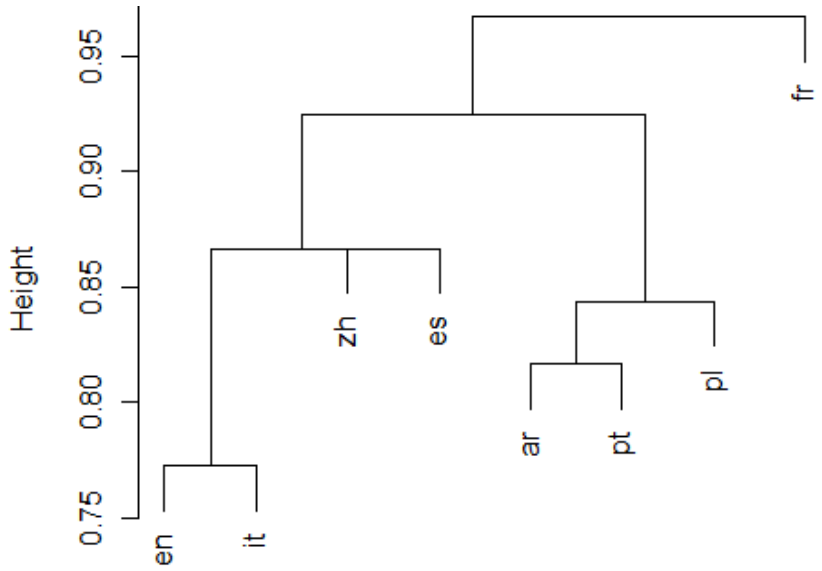


Figure 2.3: Hierarchical clustering dendogram of *progress* dissimilarities

2.4.3 Progress Toward a Less Impersonal and More Humane Society

The other postmaterialist option in the question has several words that are important to the interpretation of the question: *progress*, *impersonal*, and *humane*.

The word *progress* shows some evidence of clustering in the dendogram in Figure 2.3. However, this clustering is shown to be semantically unimportant when we look at the collocations shared within each cluster in Table 2.2. The shared collocations, between the cluster of English and Chinese, do not differ greatly in meaning from those shared between the cluster of French, Italian, Portuguese, and Spanish. In both cases, the primary meaning is related to development and movement. This meaning is also reflected in the shared collocations of the less clustered languages of Arabic and Polish, which also share many of the same collocations as the other two clusters. In this case, the clustering that is found does not appear to be reflecting something that is problematic for the survey question.

It is not possible to run the analysis on the translations of *humane* in German, Polish, and Russian, because it doesn't appear often enough in their respective corpora. The DIANA analysis shows two clusters of meaning. The first consists of Arabic and English. The second cluster contains French, Portuguese, Italian and Spanish.

It is not completely clear whether there is a difference in meaning between the clusters from the lists of shared collocations alone. The second cluster includes a wide range of

English and Italian		Polish, Arabic and Portuguese	
Collocation	Number of languages shared in	Collocation	Number of languages shared in
all	2/2	each	3/3
also	2/2	great	3/3
continue	2/2	human	3/3
country	2/2	social	3/3
development	2/2	society	3/3
further	2/2	world	3/3
great	2/2	about	2/3
human	2/2	also	2/3
more	2/2	another	2/3
new	2/2	any	2/3
other	2/2	between	2/3
social	2/2	century	2/3
towards	2/2	civilization	2/3
		development	2/3
		economic	2/3
		even	2/3
		evolution	2/3
		form	2/3
		growth	2/3
		humanity	2/3
		life	2/3
		man	2/3
		may	2/3
		nation	2/3
		new	2/3
		political	2/3
		progress	2/3
		scientific	2/3
		technological	2/3
		what	2/3
		work	2/3

Table 2.2: Collocations shared within *progress* clusters

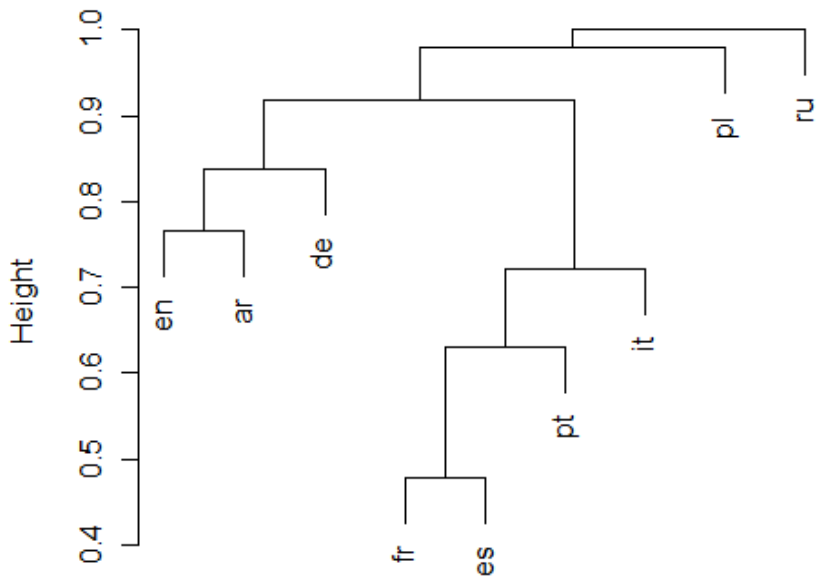


Figure 2.4: Hierarchical clustering dendogram of *humane* dissimilarities

quite philosophical words such as *existence*, *natural*, and *conciousness*. Conversely, Arabic and English both include much more practical words on their lists, such as *treatment* and *organization*. Arabic's list of words mostly refer to humanitarian situations, while English's most commonly refer to animal cruelty, with some mention of societal issues such as *euthanasia*. The main difference between the clusters is the generality of the term *humane* which seems to be much more specific in English and Arabic.

Although this difference in meaning is present, it is not clear what the effect of it should be. With the word *impersonal* in the question as well, it is unlikely that English speaking respondents would think primarily of animal cruelty in answering the question.

No language has more than 800 instances of *impersonal* in its corpus, which is not sufficient to create a reliable dissimilarity matrix. A qualitative comparison of the collocations do not suggest that there are widely differing meanings across languages. The main themes that appear are 'a lack of emotion' and 'objectivity'. These themes highlight a potential ambiguity in the question, as respondents may answer differently depending on which meaning is salient. However, the addition of the word *humane* in the question probably reduces this risk, as it signals a negative, connotation of *impersonal*.

2.4.4 Summary of Linguistic Analysis

The analysis of the keywords within the question do not clearly show any differences between the languages that would be problematic for interpretation, with the exception of *ideas*. However the difficulty in running the full analysis due to low frequency of the words within the corpora suggest that the question may be difficult to interpret in some languages, as the vocabulary is likely to be unfamiliar to some respondents. In particular, it is worrying that the postmaterialist responses use more complicated vocabulary than the materialist options, as this could introduce a bias towards these responses among less educated respondents. Since the major thesis of postmaterialism is that economic development increases postmaterialist attitudes (Inglehart, & Welzel 2005), this bias could be problematic for these findings. A plausible alternative interpretation would be that development increases education, which leads to a greater proportion of respondents who are familiar with the vocabulary used in the postmaterialist options in the survey.

2.5 Human translation of collocations

In order to check that machine translation of the collocation lists do not lead to biased results, the analysis in this paper is replicated using human translators. Translators for Spanish, French, Chinese, Portuguese, Russian, German and Italian are recruited through the myGengo website (MyGengo 2012). Polish and Arabic translators are recruited through the Fox Translate website (Fox Translate 2012). All translators are given the following set of instructions:

> This is a lists of words associated with particular topics generated from collections of texts in your language. The symbols and question marks can just be copied over unless they have a specifically different meaning between the languages. Function words should be translated into the English equivalent if there is one. If there is no equivalent English term, leave a question mark (?) instead. Where there are multiple possible translations, pick the one you consider most likely and just leave that.
>
> This translation is being used to check translations that are provided without

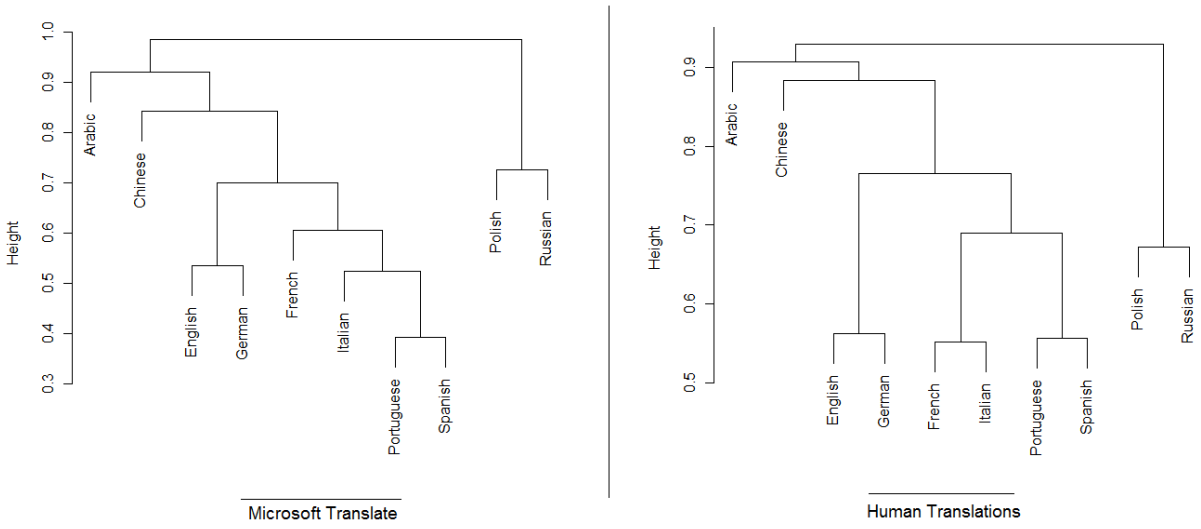


Figure 2.5: Comparison of divisive hierarchical clustering of *ideas* translations using human and machine (Microsoft) translations of collocations

> additional context to see whether they are stable (to compare to a project involving machine translation). The most likely translation for each term is all that is needed. I'm not looking for dictionary entries, just the closest translation.
>
> Thank you for your time.

Once the translation lists are obtained[4] the same methodology used for the machine translation is followed for generating and analysing a dissimilarity matrix.

Human translation of the collocations does not lead to noticeably different results from those found with the machine translation. The dendogram for *ideas* generated from human translations is shown in Figure 2.5, along with that generated by the machine translation in this paper. The cluster structure does not differ substantively depending on the translation method used. The collocations shared in the Western European and Polish/Russian cluster also follow the same themes as were found when using machine translation as shown in Table 2.3.

[4]Some translators gave several possible translations of each word. In these cases the first mentioned word is chosen for use in the analysis. A table containing the original collocations in each language and the human, Google and Bing translations are available from the author.

Polish and Russian		English, German, Spanish, Italian, Portuguese, French	
Collocations	Number of Languages Shared In	Collocations	Number of Languages Shared In
all	2/2	give	6/6
believe	2/2	idea	6/6
christian	2/2	more	6/6
common	2/2	new	6/6
freedom	2/2	thought	6/6
good	2/2	develop	5/6
great	2/2	express	5/6
ideal	2/2	good	5/6
life	2/2	own	5/6
moral	2/2	all	4/6
only	2/2	come	4/6
other	2/2	concept	4/6
people	2/2	how	4/6
political	2/2	much	4/6
value	2/2	say	4/6
world	2/2	share	4/6
		time	4/6
		what	4/6
		without	4/6
		world	4/6

Table 2.3: Collocations shared within Russian and Polish compared to those shared between Western European language for *ideas* using human translated collocations

Appendix C

Appendices for What structures the flow of information between different Embassies in the US State Department?

3.1 Software used in this paper

The analysis in this paper was performed in R (R Core Team 2013) and made use of the SNA (Butts 2013), ggplot2 (Wickham 2009), countrycode (Arel-Bundock 2013), gtools (Warnes, Bolker, & Lumley 2013) and plyr (Wickham 2011) packages.

3.2 Missing data

Although our dataset contains a large number of diplomatic cables, there are several classes of cables that are not included. These are broadly in two groups: very low level administrative cables and the most sensitive 'top secret' cables. We were unable to find figures for the numbers of cables in each of these categories.

The disposition authority document for the diplomatic cables comprises 220 pages of discussion of which TAGS indicate cables with historical value. For each subject TAG they review 100-200 messages and categorize those messages according to the level of research

TAGS	Field	Title
AART	Administration	Art-in-Embassies Program
AAUD	Administration	Audits
ABLD	Administration	Buildings
ABUD	Administration	Budget Services and Financial Systems
ACOM	Administration	Departmental Communications
AFIN	Administration	Financial Services
AFSI	Administration	Foreign Service Institute
AFSP	Administration	Post Administration
ALIB	Administration	Library Services
ALOW	Administration	Allowances
AMED	Administration	Medical Services
APER	Administration	Personnel
APUB	Administration	Publishing, Printing, Distribution, and Library Services
AREC	Administration	Commissary and Recreation
AREG	Administration	Regulations and Directives
ASAF	Administration	Safety
ASCH	Administration	Overseas Schools
ASUP	Administration	Supplies and Equipment
ATRN	Administration	Transportation
AWRD	Administration	Awards
BLIB	Business	Services Commercial Libraries
BPUB	Business	Services Business-Commercial Publications and Libraries
CPAS	Consular Affairs	Passports and Citizenship
CVIS	Consular Affairs	Visas
OEXC	Operations	Educational and Cultural Exchange Operations
OSCI	Operations	Science Grants
OTRA	Operations	Travel and Visits

Table 3.1: Table of temporary subject TAGS reproduced from FAQ Central Foreign policy File, 1973-1976 Appendix II p.13

potential they have: no research potential, low, medium or substantial. They then assessed whether a substantial proportion of messages with research potential would be lost (bearing in mind that a message with a temporary TAG will still be kept if it was also TAGsed [sic] with a permanent subject TAG). Correspondence with the NARA curators suggests that around 25% of cables were excluded because they had only temporary subject TAGS. The results of this appraisal were then discussed and amended with input from Department of State (Langbart et al. 2007).

Given the lengthy process directed by archivists familiar with the State Department cables, we believe that this process of removing irrelevant cables is likely to overall improve the quality of the dataset by focusing it more strongly on politically relevant communications.

The pre-exclusion of irrelevant cables has advantages and disadvantages.

Advantages:

- The temporary cables are also unlikely to have been written or even seen by politically important actors within an embassy such as the ambassadors or senior diplomats.
- It also avoids creating too much network structure through purely administrative functions such as processing visa applications.
- Focuses attention on those cables that are interesting from a historical/political perspective

While these advantages are substantial for most applications, there are some potential applications that could have made use of these removed cables.

Disadvantages:

- Even purely administrative communication may have some purpose and set the stage for more important communications later
- Some interesting cables could have been erroneously tagged with temporary labels

Overall, the situation is probably preferable to having all cables written between these embassies without any indication of their political importance but less useful than having this indication.

Appendix D

Appendices for Speaking Diplomatically: Examining Agenda Setting Using Relational Event Modelling

4.1 Timestamp parsing and imputation

The draft time and date of the cable is given in Coordinated Universal Time (UTC) including hours and minutes (C). This is important as it gives us the temporal granularity key for relational event analysis, where ordering of communication matters.

Times were directly extracted for 293,370 cables (73.2%) for the period 1973-1974. However, 106,160 archived and withdrawn cables had date but not time information available, which meant 26.5% lacked a full date and time.[1] Given the importance of communication ordering in relational event analysis, we imputed time of day for these cables based on message ordering implied by the cable's ID number and the standard sending patterns of the sending embassy. A full explanation of this procedure and the results of it are explained in appendix 4.1.

Although we are missing timestamp data for some cables, we have complete data on

[1] An additional 1134 cables (0.2%) held no date or time information and were thus dropped from the dataset.

Cables	Date and Time	Short date only (normal cable)	Short date only (withdrawn cable)	No information	Total
Number	293370	23404	82756	1134	400664
Percentage	73.2	5.8	20.7	0.2	100

Table 4.1: Summary statistics for the parsing of timestamp data for the diplomatic cables

message ID. Each diplomatic cable has a unique ID formatted: [year][sender][number]. For example, 1973STATE224422 is a cable sent from the Secretary of State in 1973 with number 224422. The number at the end appears to run incrementally over time within each sender. We can thus infer that cable 1973STATE3001 is almost certainly sent after cable 1973STATE3000 but not necessarily sent after 1973OTTAWA3000. This means that, for most of the missing times, we know what date it was sent on and which messages it was sent before and after. For cables sent from busy locations (such as the London Embassy or the State Department in Washington), using the ordering method often narrows down the sending time to a range of a few minutes.

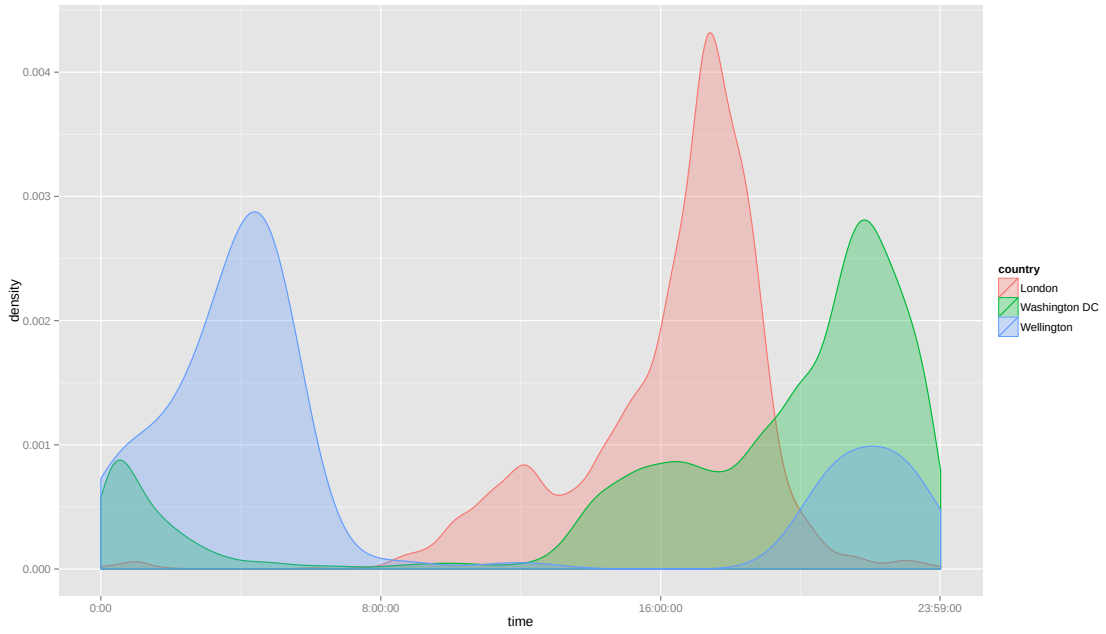


Figure 4.1: Distributions of State Department cable timestamps from various locations

Although embassies send cables at all times of day, the vast majority of cables are sent during working hours. We can immediately see this by comparing the distribution of sending times for London, Washington DC and Wellington.

The three locations have clearly different patterns of cable sending that reflect the work

Timestamp	
22:51:00	
22:52:00	Lower bound
missing time	
missing time	
22:55:00	Upper bound
22:56:00	
22:56:00	

Table 4.2: Example of a stretch of timestamps with two missing times to be imputed

days. We can take advantage of this information when imputing dates by drawing times from the empirical distribution of timestamps from that embassy.[2]

The final imputation procedure combines both these heuristics to impute the best guess at the time a cable is sent. The method starts with a contiguous block of one or more missing timestamps from an embassy. It then takes the first complete timestamp sent from that embassy on the same day prior to the missing values and uses this as the lower bound for the imputed times. The equivalent is then done for the upper bound. The following table shows an example of a short stretch of timestamps for the State Department.

There are some limitations to this procedure however. Where the upper and lower bounds produce too few times to sample from, such as in table 4.2 (where there are fewer than (5 times the number of missing cables) times to sample from), we draw instead from a uniform distribution between the two bounds to avoid creating too many tied times.[3] Where there are no upper or lower bounds, for example because a missing cable is the first or last sent from the embassy that day, the bound is replaced by the start or end of that day. If just one cable is sent on a particular day but without a timestamp, it will be imputed from the unrestricted time distribution at that embassy.[4]

Following imputation, only 1,134 cables (0.3%) lacked timestamp information. We test

[2]One alternative approach to sampling would have been to simply use the expected value of the sending time. This would have the advantage of making the imputation deterministic. However, we decided to sample as this keeps the imputed time distributions much closer to the true ones, whereas imputing the mean would reduce the dispersion.

[3]A small number of IDs do not line up with time ordering. It is not completely clear why these cables are not time ordered. When a missing time lies between two times in the wrong order, we reverse the bounds so that we still assume that the time lies between the two times on that date. Fortunately this only occurs in a small number of cases.

[4]A small number of sending entities had no non-missing time values to sample from. In this case another entity in the same country and time zone was used to impute times from. One example of this were cables from NATO Brussels which were all missing. Times were imputed using cables sent from the Brussels Embassy instead.

the accuracy of the imputation process by running the imputation process over a sample of 10,000 known times and calculating the root mean squared error (in minutes) of the prediction. The differences are shown in figure 4.1 below showing that the imputation gives a strong improvement over random imputation.

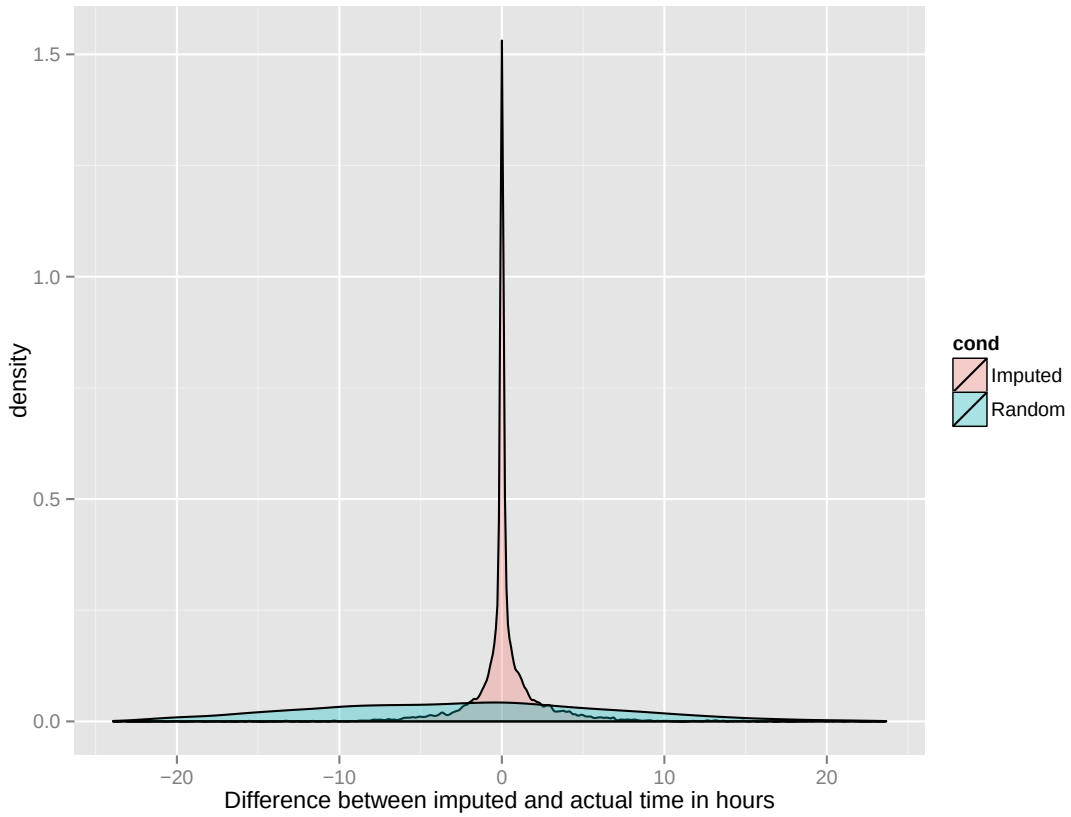


Figure 4.2: Comparison of errors when imputing timestamps randomly or using our imputation method

4.2 Difficulties with extracting references

The reference field in the cables can be formatted in a number of ways. One example is below:

```
REFERENCES: ( A) CAN THO 0141, ( B) CAN THO 0182, ( C) CAN THO 0149
```

Each unique cable id referenced is listed but these references do not include years. The cable ids restart within the embassy each year which means that we cannot necessarily distinguish between `CAN THO 0141` sent in 1972 and `CAN THO 0141` sent in 1974.

We use a number of heuristics to try and match up references with the cable IDs. The first heuristic is simply that a cable cannot cite cables in the future. [5]

The second heuristic we use is that historically important cables are more likely to be cited. That is, those cables that made it into our dataset (because they had politically relevant TAGS), are more likely to be cited by other cables in our dataset than those we don't observe. This generalisation is supported by the data which show that we manage to make matches for 71% of citations despite most cables not being present. In other words the historically important cables again seem to constitute their own separate discourse.

The third heuristic we used was to favour recent possible citations over those sent longer ago. There is some qualitative evidence behind this heuristic. A summary paper describing the cableXpress cable system (the successor to the system used at the time of the StateNet dataset) describes how the recipient of a cable must retrieve any cables referenced in it from a paper filing system (Lotus Development Corporation 1996). This kind of retrieval is likely to mean greater difficulty in citing or retrieving older cables.

4.2.1 Biases introduced by these heuristics

These heuristics lead to an overall sensible structure that makes sense when we examine chains of citations qualitatively. However, some network statistics are likely to be biased when calculated using these measures.

The second heuristic, assuming that cables in the dataset are more likely to be cited, is helpful in making connections but may introduce some essentially random connections into the citation graph. This means that conversations may end up appearing more connected than is actually the case (large components become increasingly likely to be connected if random ties are made).

It was first necessary to identify these conversations within the dataset. A full, directed network of cables of citing or cited nodes and citations as the ties binding them was first produced by producing an edge list from the cable citations. While withdrawn cables held some other metadata, this did not extend to outgoing references. However, this citation network does include references from the full-text declassified cables to withdrawn cables.

[5] While this seems like a fairly ironclad assumption, it appears that there are some cases where two cables are prepared together by an embassy and cross-cite each other. These references will not be correctly captured in this iteration of the dataset.

This means that there may be a bias towards withdrawn cables 'fathering' conversations. We exclude cable chains from some of the analyses if their initial cable has outgoing references for this reason. We indicate where this is the case when describing the analysis.

4.3 Modelling citation choice

For the citation set we consider it an ordered set:$C(e) = e_1, e_2, ..., e_(|C(e)|)$. The probability density function for citation sets is decomposed as follows, where $C_*(a)$ is the set of events already added to the citation set:

$$f(C(e)|...) = \prod_{a=1}^{|C(e)|} f(e_a|C_*(a), ...) \tag{D.1}$$

Each candidate element e_a in the decomposition is modelled as:

$$f(e_a|C_*(a), ...) = \frac{exp(\beta^T S_C(e_a, .))}{\sum_{e^* \in Rcitation(i(e), C_*(a), ...)} exp(\beta^T S_C(e^*, .))} \tag{D.2}$$

where $R_{citation}(i(e), C_*(a), ...)$ is a risk set calculation function that defines the possible events that ego is choosing between and where $\beta \subset \theta$ and $S_c(e_a, .)$ is a statistic.

4.3.1 Citation risk sets

To calculate the probability of a discrete choice, we must first define the options that an actor is choosing between (Lospinoso et al. 2013; Stadtfeld 2012). We refer to this as a risk set. For the citation choice, we need to define a set of possible messages an actor could cite. Since there are several hundred thousand messages in the dataset, it is both psychologically unrealistic (no one has the mental resources to consider hundreds of thousands of options) and computationally infeasible to calculate. The first restriction we put on the risk set is that the actor must have had the possibility of seeing the message before they can consider citing it. There are four ways in which we capture this:

1. The actor sent the cable;
2. The actor received the cable;
3. The actor cited the cable;

4. The cable was cited in a message the actor received.

These criteria alone would lead to huge risk sets and flat likelihoods so we put two further restrictions on the risk set. First, the citation must have been recruited within a certain number of days and second, only the most recent x messages meeting one of these criteria are included.

For this paper we chose a maximum risk set size of 300 and a sliding window of 35 days. 71% of cables that were cited were in the risk sets. This should be considered against the fact that only 91% of cables would be included in the risk sets if the windows were set to their maximums. [6]

Finally, we exclude those recipients who have already been added to the recipient set, $\{e_1...e_{a-1}\}$, from the risk set.

Citations outside the risk set

For now we do not model citations that are not in the risk set. This causes a bias in the probability model but shouldn't be large enough to affect the substantive conclusions. However, we will correct this in future iterations by explicitly modelling the possibility of choosing "someone else". We did not include this option in the current model because the large risk set sizes lead to technical issues when including this type of actor and there are additional questions about what factors should be included to understand choosing someone else.

4.4 Modelling Alter Removal

Following Lospinoso et al.'s (2013) approach to decomposing the alter set, we consider the removal set as an ordered set of actors that are added (into the removal set) sequentially from a risk set. This means the removal set probability density is decomposed as follows:

$$f(J(e)_{removal}|...) = \prod_{a=1}^{|J(e)_{removal}|} f(j_a|J_*(a),...) \tag{D.3}$$

[6] This is likely to be due to some unobserved citations in the withdrawn cables. We intend to continue investigating this to see what is leading to the missing 9%. It is also possible that diplomats are occasionally made aware of some cables through another channel. However, we suspect that this is a data issue.

, where $J_*(a)$ is the set of recipients currently in the recipient set (i.e. the initial set minus the recipients already removed).

Each candidate alter for removal j_a in the decomposition is modelled as:

$$f(j_a|J_*(a),...) = \frac{exp(\beta^T S_{Jremoval}(j_a,.))}{\sum_{j^* \in R_{removal}(i(e),J_*(a),...)} exp(\beta^T S_{Jremoval}(j^*,.))} \tag{D.4}$$

where $\beta^T \subset \theta$ and $S_{Jremoval}(j,.)$ is a statistic. The removal risk set, $R_{removal}$, is simply the initial set $J(e)_{initial}$ excluding $i(e)$ and including the special j_{close} actor that represents the choice to stop removing recipients from the recipient set. Finally, the risk set excludes any recipients already removed from the recipient set.

4.5 Possible reasons for fewer recipients over time

Across the course of this study, we observe a pattern of higher volume of messages with fewer recipients per message. This could potentially suggest greater centralization of the communication flow through the Bureaus. There are many potential explanations for why embassies may include fewer recipients over time while also increasing the volume of messages sent.

The first is that the State Department's leadership chagnes in our study From William Rogers to henry Kissinger. The shift in numbers of recipients appears to line up approximately with Kissinger taking over the job of Secretary of State.

Another factor is that the telecommunications system used here essentially starts at the beginning of our study period (although the previous system has many analagous features). Consequently, it's possible that the system took some time to find an optimal distribution approach. This explanation seems somewhat unlikely as the shift is actually fairly sharp.

Although cables are the primary means of political reporting (especially in this period), there are other lines of communication including telephone and face-to-face meetings. This seems unlikely to be the cause of behaviour change, however, as there is no reason to think that these methods would suddenly become more appealing.

Many of these potential explanations are topics worthy of further study in their own right but are beyond the scope of this chapter. Instead, we take the pragmatic approach of restricting our analysis to the part of the system we have a theoretical reason for studying

and can examine in full detail rather than introducing a whole new set of confounding factors into the model that we don't understand well.