Framing Policy Debates in the European Union

NSF Proposal submitted October 2010

Christine Mahoney, University of Virginia, PI

Frank R. Baumgartner, University of North Carolina at Chapel Hill, Co-PI

Requested amount: $386,885 for 24 months beginning on March 1, 2011.

This document contains the project summary, project description, and references.

Project summary

Intellectual Merit:

The proposed project is a large-scale study of framing by interest groups involved in consultations with the European Union. It proposes the use of new automated techniques to identify frames and assess their dimensionality in policy debates surrounding 120 issues. The investigators coordinate with a large team of scholars simultaneously conducting interviews and fieldwork associated with these same issues, thus contributing to a large and growing infrastructure for the study of policy processes and the roles of civil society organizations in the European Union. The larger project addresses issues of the democratic nature of debate and the relative impacts of nation-states, consumer and civil society organizations, and industries in shaping policy decisions in the world's largest new political system. The focus of this proposal is on the application of tools and the development of an infrastructure that will allow the analysis of the choice and effectiveness of arguments by interest groups seeking to affect policy outcomes. This focus on framing contributes to the larger collaboration but constitutes a coherent stand-alone project. All of the documents collected be made available on-line as a resource for other scholars of public policy to download and analyze with other software tools, to answer a broad range of theoretical questions.

The project seeks to make more systematic, quantitative, and rigorous a literature on framing which has often been highly qualitative. More importantly, it uses a framing approach to understand a number of substantively and conceptually important questions: what are the roots of policy stability? Are those roots based in shared policy paradigms that become predominant within professional communities, or are they due to institutional structures? How do models of policy negotiation and compromise work within uni- and multi-dimensional policy spaces? Are most issues debated in the policy process unidimensional or do they have multiple dimensions of active engagement by the stakeholders involved? Can one understand policy change with greater focus on argumentation and framing? Can one predict and understand movement in official positions on a policy debate with reference to the arguments put forward by advocates during the policy process? Do material resources, network structures, group type, or alliance with national governments affect the success of interest groups in promoting their preferred arguments within official policy statements?

Broader Impact:

The project seeks to enhance the study of framing and policy processes by initiating a large project focused on the European Union in collaboration with a network of European scholars conducting parallel studies of the same issues explored under this grant through automated content analysis techniques. The project enhances the study of policy processes through the development of new tools of automated text analysis, to generate important empirical findings about the dimensionality of debate across a sample of policy issues and the abilities of interest groups and government advocates to affect the frames used by government policymakers themselves, and to advance the discipline's understanding of the causes of policy stability and change. The project will train a number of undergraduate and graduate students and create a public web-based resource for the continued study of democratic decision making in the European Union.

Project Description

Introduction

We propose a large-scale study of framing by interest groups involved in public policy debates in the European Union in order to understand the impact of framing on lobbying success. We use new quantitative text analysis techniques to identify frames, assess the dimensional structure of a sample of policy debates, and determine the change in the distance between interest groups and the European institutions over the course of policy debates surrounding 120 issues. The European Commission publishes proposed regulations, seeks comments, and then issues final regulations. By measuring the substance and frames associated with the initial proposal, the comments received, and the final regulation, we can assess the degree to which movement from proposal to final regulation is associated with the comments received by interest groups, and which comments are most closely associated with any observed movement in the official position. By measuring the relative positions of the policymaking institutions and interest groups in a multidimensional space, we can assess quantitatively what has previously been done only with qualitative methods. And we do so over a large sample of issues.

We coordinate with a team of European scholars simultaneously conducting interviews and fieldwork associated with the same 120 issues that will be our focus. The larger project, INTEREURO (Comparative Research on Interest Group Politics in Europe), aims to develop a comprehensive theoretical and empirical understanding of the role of interest groups in the EU through synchronized data collection and sharing. Dozens of scholars and researchers will be collecting data at the EU-level and in six countries (Germany, the Netherlands, Slovenia, Spain, Sweden, and the UK; see: http://www.unc.edu/~fbaum/papers/esf_2010.htm). They will work collaboratively to examine interest group mobilization; organizational maintenance and professionalization; strategies for influencing political decision-making; multi-level lobbying efforts (that is, coordination of national-level and EU-level strategies); and the impact of groups on policy outcomes. This collaborative network received approval for funding from the European Science Foundation (ESF) in September 2010, and the current NSF proposal is designed to link closely with their efforts. Our focus here is on quantitative assessments of success in framing. Other collaborators will be conducting more qualitative assessments of different aspects of the lobbying process in individual countries and at the EU-level in Brussels. This large international collaboration seeks to create the infrastructure to support a new generation of studies of lobbying and the policy process in Europe, including a new web site with information about our collective efforts.

Our focus on framing at the EU-level contributes to the larger collaboration but constitutes a coherent stand-alone project as well. All of the documents we collect, process, and code will be made available on-line as resource for other scholars of public policy to download and analyze with other software tools, creating new infrastructure for the study of policy processes and comparative content-analysis techniques. The larger ESF-based project is also committed to maximum public availability of the source materials.

Unanswered Questions in the Study of Framing

Considerable ambiguity characterizes the literature on framing, as scholars in different disciplines or subfields of political science have used a variety of terms to refer to relatively

similar things and few studies have systematically looked at the effect of framing on policy outcomes. Various actors emphasize different points; Entman has described framing as "selecting and highlighting some features of reality while omitting others" (Entman (1991, 53). We follow Entman and define a frame as a specific aspect of a policy issue that is emphasized in a public policy debate. Policy issues are almost universally multi-dimensional, since complex policy proposals often have differential effects on different social and professional groups, different geographic regions, and can be justified in many ways. Baumgartner and colleagues, in their study of a random sample of 98 objects of lobbying in Washington DC, found not a single issue that could be considered along only a single dimension (2009). Baumgartner and Jones (2009 [1993]) describe "non-contradictory argumentation" in which proponents of environmental protection may for example discuss issues of loss of habitat whereas, in the same debate, business interests may focus on loss of jobs or tax revenues. Key to the discussion of framing is the idea that protagonists in policy debates often attempt to push collective attention more toward their preferred frames and away from those of rivals. Direct confrontation on a single dimension, for example statistical debates about *how many* jobs may be lost with a given new regulation, may be less common than indirect or non-contradictory debates where an argument about jobs is countered with an argument about habitat. Apples-and-oranges comparisons are typical. We aim to develop tools to make what has often been a very qualitative literature amenable to rigorous quantitative methods, allowing us to compare issues based on the frames and dimensional structure of debates and assess the effectiveness of interest groups' framing efforts in promoting particular policy outcomes.

Making use of new content analysis techniques, we will first identify the frames invoked on each issue as well as the dimensional structure of the policy debates. At the same time, we locate interest groups and the European institutions in the policy space and assess the movement of the EU institutions over time. If the distance between interest groups and the EU institutions in the policy space decreased over time, we consider the interest groups as successful. By controlling for other interest group and issue characteristics identified in the literature and by analyzing the frames employed by interest groups, we are thus able to assess whether their framing efforts can account for variation in lobbying success.[1]

Since Schattschneider (1960), scholars of public policy have known that participants in the policy process have incentives to describe the issues on which they deal in different ways in order to attract new participants in the process, or to justify their exclusion. As Baumgartner and Jones (2009 [1993]) described, a description of nuclear power that focuses on technical complexity and lack of broader social consequence (other than cheap and abundant electricity) served to justify the monopoly once enjoyed by nuclear physicists and others linked to the now defunct Atomic Energy Commission. These authors referred to 'policy images' and noted how that of nuclear power went from positive to negative in the 1960s, with huge public policy consequences. Since these early works, students of public policy have been keenly aware of the impact of framing on public policy.[2]

[1] Of course, we cannot assess whether the movement of the European Commission position was due to the arguments made by any particular interest groups. Rather than attempt to identify "influence" in this manner (which we believe to be impossible), we focus instead on what can indeed be measured: the frames that are used by different actors in the debate and the movement of the official position. Our collaborators in Europe will add additional qualitative assessments of the causes of the outcomes we observe.

[2] Framing studies are common in several other literatures within political science, including media-effects studies in the field of public opinion (see for example Berinski and Kinder 2006 or Gilliam and Iyengar 2000). Similarly, Lakoff (2004) focuses on how members of the public respond to different types of emotional stimuli. These

William Riker (1986, 1996) focused our attention on the ability of individual protagonists in the policy process to destabilize debates by focusing attention on a particular dimension. Since his important works, scholars have struggled with the determinants of success in such efforts. James Druckman (2001) was one of the first to study the limits to framing (e.g., the fact that many people may be strongly resistant to efforts by others to reframe a debate), but his study was at the mass level, not among participants in the policy process as we propose here. Baumgartner and colleagues (2009) have conducted the most extensive study to date of collective framing processes, covering a random sample of 98 objects of lobbying activity in Washington, DC. Their results suggest that few issues are reframed, at least in the short run. Following each of 98 issues over a four-year period, they found that fewer than five percent of the issues were significantly reframed. Much more common were stable frames understood by all members of the professional community of lobbyists and policymakers surrounding the issue. The individual lobbyists might seek to focus attention on one dimension over another, but none had the power unilaterally to redirect *collective* attention only to those dimensions favourable to their own position. Mahoney (2008) similarly found limits to the ability of lobbyists to reframe issues in her study comparing interest-group argumentation in the US and the EU. In contrast to members of the mass public, elite participants in a policy community have highly detailed understandings of the various elements of debate, even those with which they disagree. So it is not easy to introduce a "new" element of debate to a group of experts.

Few previous empirical studies of framing in public policy have addressed issues we will explore here: How many different dimensions of debate are present across a sample of issues? Do the issues with higher dimensional complexity differ from the simpler issues with regards to outcomes or the ease of reaching a policy solution? Can we associate shifts in the positions of policymaking institutions with the framing strategies employed by interest groups? Is success in framing associated with the material resources controlled by interest groups? Are there certain categories or types of frames that emerge as particularly powerful when it comes to moving policymakers? Do those who emphasize the same frames constitute homogenous coalitions with respect to group type and sector of the economy, or are these empirically defined coalitions more heterogeneous? Among those who disagree, do they engage on the same dimension of evaluation, or do they follow what Baumgartner and Jones (2009) called "noncontradictory argumentation"?

The range of important theoretical issues to be addressed here is matched by important methodological ones. The study of argumentation and framing at the elite level is being revolutionized by new developments in the systematic study of text as data (e.g. Laver et al. 2003, Purpura and Hillard 2006, Hillard, Purpura and Wilkerson 2008, Schonhardt-Bailey 2008, Slapin and Proksch 2008, Yu, Kaufmann and Diermeier 2008, Hopkins and King 2010, Quinn et al. 2010). We will apply a set of newly developed text analysis tools (e.g. Schonhardt-Bailey 2008) to an original database of over 7,000 documents which detail the political argumentation of thousands of advocates lobbying on 120 policy issues. Thus, we fit into a growing literature making use of the huge mass of publicly available documents associated with government decision-making.

---

literatures, focused on individual psychological or cognitive responses to public policy stimuli among members of the mass public, are related to but different from the focus within the field of public policy. This literature deals with elites, not members of the mass public, and is less interested in the individual-level cognitive response to frames than in the overall nature of the debate as reflected in publicly available documents. We follow the public policy tradition here.

Finally, we focus on the institutional context of the European Union. This allows comparison to previous NSF-funded work in the US (see Baumgartner et al. 2009, Mahoney 2008) and allows comparison of the effectiveness of civil society interest groups, business and trade associations, member states through the European Council, bureaucrats and agency officials through the European Commission, and elected officials in the European Parliament. Previous work on the US, for example, suggested that forty percent of all the "advocates" active in the policy process at the federal level were government officials, not outside interest groups, the normal focus of policymaking studies. Elected and appointed officials play important roles, and the EU setting allows us to study the roles of national governments as well as EU-level officials.[3] The availability of large stores of documentation through the EU consultation process makes possible a large-scale analysis of administrative decision-making in the complicated institutional setting of the EU.

Overall, then, our project seeks to address key issues in the study of framing, apply newly developed tools for automated coding of frames on a large scale for the first time, and create an infrastructure for research that other scholars can use for their own purposes. By expanding the study of framing from the US to the European context we are able to investigate whether framing success varies with contextual characteristics related to the political system and whether theories developed in American context apply to other political systems. In addition, we will systematically assess for the first time, the ability of advocates to move the debate closer to their ideal points by selective framing in their policy arguments.

Testing Theories of Framing & Frame Success Systematically

When does a frame successfully change policy? We test expectations about variation in framing success by looking at characteristics of the frame itself and the broader issue debate space. The first set of questions makes use of theories from the Resource Mobilization and Psychological Framing literatures, and the second makes use of theories of Frame Bridging and Frame Extension common in Sociology as well as Psychology and Economics-based Cognitive Limitation and Bargaining theories.

Is it the quality of the frame that moves policymakers to action? Or is it simply the quantity of communications so framed that moves policymakers? This should be a fundamental question in advocacy studies and, at base, is a resource mobilization question. Resource mobilization theories suggest that fired-up latent interest groups are not enough to change policy; groups of like minded individuals need to mobilize resources to organize and become active and effective in the policymaking process (McCarthy and Zald 1977). For each issue we will asses which actors are evoking which frames and whether those frames have independent effects on the positions of political institutions, controlling for group and coalition resources. Resource mobilization theories would suggest that those groups and coalitions of groups with more resources, and by extension the ability to magnify their frame through a barrage of communications, will be more likely to see their frame succeed in moving policymakers. In this perspective, it is not so much the argument that matters, but who is making it.

An alternative framework, which suggests there is power in argument, would allow for variation in framing success across actors and coalitions of actors with differing levels of resources. It is here that the extant literature provides only partial guidance and where we hope this project will provide the greatest insight. Frames may have different likelihoods of success

[3] The roles of each nation-state in the policy process is beyond the scope of what we expect to be able to do in this project, but it is a significant part of the larger European collaboration of which this project will form a part.

depending on whether they focus on gains or losses. Kahneman & Tversky's (1981) laboratory experiments suggest people are more likely to be risky, and thus policymakers would be willing to take a risk of policy change, if the issue is framed negatively (i.e. as a loss, or a pending crisis). However, research on philanthropy shows that there are cultural differences in responsiveness to negative frames, where Europeans respond to negative pleas for financial support of develop aid which evoke colonial guilt, Americans respond better to positive frames that evoke ideas of American philanthropy, generosity and equality of opportunity. Do negative frames always beat positive frames as Kahneman & Tversky's research would suggest? Does the effectiveness of negative frames vary depending on the issue – displaying strong returns on domestic issues but perhaps not on international issues? Lastly, is it the economy stupid? Do frames that focus on jobs and economic growth carry more weight with governing institutions that will be judged largely on the health of their economies, than frames emphasizing the impact on consumers, the environment, or public health? Or can frames that highlight how a set of policies result in bodily harm win, as Keck & Sikkink (1994) found in the case of human rights issues, even in the face of major economic interests?

With our data linking group type and framing strategies we will be able to test, on a large scale, Snow and Benford's (1986; 1988) concepts of "frame extension" and "frame bridging." We will better understand if corporations are always framing issues in economic terms, or thinking outside that box; if environmental groups are constrained to frame issues as environmental concerns, or if they have the flexibility to be creative and evoke economic, public health, and security dimensions as well. There are reasons to suspect that diffuse interests reliant on large citizen-based membership will be more restricted in their set of frames (that is, they will minimize their "frame extension" so as to not alienate core supporters), whereas trade and industry groups, who need worry less about mobilizing enthusiastic support from diffuse potential memberships, can be more strategic with respect to the frames they emphasize. We will investigate whether the range of frames used differs by group type.

We will also be able to study the extent to which actors are "frame bridging" and if that bridging increases the chances of a frame successfully moving policymakers. Snow and Benford (1986; 1988) suggest groups frame their positions so as to build bridges to other groups and therefore expand their supporting coalitions. Riker's focus on heresthetics and Baumgartner and Jones' idea of non-contradictory argumentation would suggest that each set of actors might focus only on that dimension most advantageous to it. Baumgartner et al. (2009) found, however, that the "sides" mobilized on the sample of issues they investigated in Washington DC were surprisingly heterogeneous and that all sides in the debate shared a common understanding of the underlying issues. This suggests that the sets of groups invoking particular dimensions of discussion may be more diverse. The likelihood of success of a frame may increase with the heterogeneity of the groups employing it. If a coalition of groups using the same frame is very homogenous (such as only automobile manufacturers), it is plausible to assume that this frame represents the opinion of only a small section of society. However, if different types of actors align by using the same frame (such as automobile and environmental groups), the European Institutions can be sure that this frame represents a broad section of society and therefore receives a lot of backing by the public. In any case, these are straightforward empirical questions given our research design.

Some issues may have only a few frames put forward by participants (or even only a single frame, to which virtually all actors subscribe), whereas others may have many. The degree of complexity in argumentation can be a measure of the dimensionality of the debate surrounding

an issue. We will study the structure of an issue debate by mapping the various frames employed by interest groups as well as the governing institutions and by identifying the underlying dimensions that structure the debate. Framing success may be expected to vary across issues according to the complexity of the structure of the debate. Cognitive limitations on the part of policy makers leads them, like all humans, to focus on just a few dimensions of an issue (see Simon 1984, Jones and Baumgartner 2005). Due to the cognitive demands placed on policymakers in a complex debate we would expect them to be more likely to take cues from large and diverse coalitions regardless of their frame. That is, the power of argumentation may be lost in complex issues where policymakers need to take mental short-cuts, equating numbers of supporters or diversity of supporters of a frame with a powerful argument.

Finally, bargaining theories suggest that if there are competing and unrelated dimensions of debate, space is open for compromise due to the possibility of issue linkages and package deals. However, if there is only a single dimension of conflict, bargaining is difficult since actors compete in a zero-sum environment and have nothing to exchange. Thus, we can assess the possible linkage between the number of dimensions of a policy debate and the likelihood of negotiated settlement. The literature on bargaining seems to stand in contrast with expectations from political science, which might suggest that a reduced (or unidimensional) political space leads to a simple solution: the median voter. We can test straightforwardly whether low-dimensional issues are more likely to see successful outcomes than high dimensional issues, and also whether the final position of the EU institutions is within an area predicted by median voter theorems, or whether the final EU documents reflect an official position outside the area of predicted compromise. Further, we can assess the number of relevant dimensions across our issues, an empirical point rarely studied.

For all these analyses, we will assess the initial and final policy locations of the European institutions (European Commission, Council, and the European Parliament). We will establish "closeness scores" for each advocate to the official position as reflected in official documents at the initial and final stages of the debate. Thus, we can establish the degree to which the different EU institutions take up the various frames in the debate and align with interest groups. In this manner, we will be able not only to assess the structure of conflict in a large sample of EU policy debates, but also the success of various interest groups in moving the location of the relevant EU institutions toward their policy position. Our project therefore seeks to enhance our theoretical knowledge of framing, lobbying, and policy outcomes. We hope not only to test a number of specific hypotheses following from the comments above, but more generally to provide some empirical framework for future studies that may be more focused on one or a few of these issues. Ours is a very broad project seeking to lay out the empirical terrain. We know remarkably little about how framing strategies sway policy debates – considering the centrality of argumentation in politics this is surprising and should be remedied; this proposed project moves us in that direction. The logistics of our project are quite unusual because of our ambition to gather and analyze large amounts of information, so we spend the rest of our proposal focusing on issues of data collection, analysis, and showing the feasibility of what we propose to do.

Logistics: Case Selection

Our project will focus on the same stratified random sample of 120 legislative proposals being studied in the larger collaborative project. This sample is drawn from the European Commission's Prelex database to create a list of legislative proposals introduced by the European Commission between 1 January 2007 and 31 December 2009. Limiting the list to proposals that

have been introduced in 2007 through 2009 will allow members of this larger team to know the final legislative outcome, collect the final legislation documentation and conduct interviews with policymakers and advocates. (Our focus here will be on analyzing the documents associated with these issues.) Legislative proposals will include 50 proposals for Regulations, and 40 proposals for Directives (two forms of binding laws), as well as 10 green and white books (official documents laying out proposed directions for future legislation); in this way the sample includes issues that are at different stages of the formal policy agenda. In this section of the sample we will exclude issues that did not generate public attention since there are many technical issues that do not motivate a public response. Operationally, we will exclude issues that are the object of fewer than five articles in *Agence Europe*, a widely used database of news coverage of EU affairs. With 100 issues chosen from among the relatively salient items on the EU docket, we will also select another 20 proposals with lower salience as a control. This sample of proposals that did not garner public attention will be stratified using the same proportions as for the publicly salient proposals: 10 for regulations, 8 for directives, and 2 green and white books. Thus we arrive at a sample of 120 issues, with varying levels of salience and at different stages in the policy making process, for which documentation about the positions of the EU institutions and the positions of interested advocates can be collected.

Document Collection & Identification of Interested Parties

For every issue in our random sample of 120 cases we will analyze three areas of public position taking at the supranational level:

1. Official Statements from the EU Institutions and Policymakers: This includes all white papers, green papers, proposals for Directives, proposals for Regulations, the Opinions from the Council and the EP during the legislative process, the final adopted legislation as well as the original policy if one existed as a measure of the status quo. These will allow us to track locations of EU institutions in the political space and to analyze any differences over time or from one EU institution to another.

2. Consultation Submissions: For those cases where public consultations are available, we will collect all submissions. The Commission's report of the Consultation will be included in the Official Statements database above. Submissions will come from interest groups (e.g., membership based organizations) but may also involve other private actors (e.g., individual corporations) as well as other actors (for example regional or national government institutions). All such submissions will be included in our analysis and our analysis of them will be as extensive as the dataset will allow (e.g., we will analyze differences among EU-level interest groups, national level groups, individual corporations, and public bodies at the lowest level of detail that our dataset allows).

3. Position Papers: In many cases a public consultation has not been held, interest groups have chosen to not comment in that forum, or groups have become involved later in the process; to capture the full range of interest group communications contributing to a debate we will go to websites of all active participants and download their press releases and position papers on the issue. Documents collected will be limited to position papers and press releases for comparability across groups. The list of all active participants on an issue will be constructed iteratively by triangulating information on groups mentioned in all three sources and from interviews conducted by the larger ESF research team.

The logistics of our data collection task depend on the number of documents found on each issue. As our sampling frame (for 100 of the 120 issues) excludes issues at the very lowest level of salience within the EU (e.g., purely technical regulations with little public discussion), we expect a substantial number of submissions for each issue. At the highest end of the scale, recent issues such as the Green Paper on the Review of the Consumer Aquis which lead to a new legislative proposal on consumer rights produced 307 consultation submissions. However, based on an earlier analysis of 56 consultations of the European Commission, the average number of consultation submissions can be expected to be about 60 (Klüver 2009). Our budget and staff plans are based on an estimate of approximately 7500 documents for the 120 issues to be analysed, including official documents (120 issues x 60 documents + a small margin).

Content Analysis

In order to analyze framing success, we will identify the frames brought up by interest groups as well as their policy positions and those of the European institutions, locate them in a policy space and examine the distance between the location of interest groups and the European institutions over the course of a legislative debate. If the distance between interest groups and the European institutions decreased over time, their lobbying activities were successful. In order to investigate the effect of framing on lobbying success, we control for alternative explanatory factors such as resources and organizational type. These variables will be gathered in cooperation with the ESF project.

Based on a review of recent literature on framing and text analysis and on a preliminary comparison of different quantitative text analysis techniques for one sample case, we came to the conclusion that a combination of cluster and correspondence analysis currently constitutes the most appropriate text analysis technique in order to study framing and dimensionality of legislative debates. Cheryl Schonhardt-Bailey has successfully used this technique to analyze framing and dimensionality of parliamentary speeches, speeches of presidential candidates and transcripts of the Federal Reserve's Federal Open Market Committee (Schonhardt-Bailey 2005, 2006, 2008; Bailey and Schonhardt-Bailey 2008). Two software packages are available with the necessary functions: ALCESTE (Image 2009) and T-LAB (Lancia 2009). Before we however start with the text analysis of all 120 issues, we will conduct a more elaborate comparison of T-Lab and Alceste with other text analysis techniques across ten policy issues in order to check the validity of the findings and to assess the quality of alternative content analysis approaches. Recent text analysis techniques that deserve further investigation are fully automated unsupervised learning techniques such as the topic model by Quinn et al (2009) or semi-automated supervised learning techniques such as the ReadMe package by Hopkins and King (2010). We will compare the results of these quantitative text analysis techniques with estimates gathered by traditional Hand-Coding in order to assess the validity of the results in a small-N study of ten policy issues before we ultimately decide upon the text analysis technique we will use in our project. However, based on our preliminary assessment of text analysis techniques, we are illustrating the use of T-LAB in order to identify frames and to assess the dimensionality of public policy debates on one sample case below. Analyzing framing and dimensionality with T-LAB involves three steps: Preparation of the text corpus, Cluster Analysis, and Correspondence Analysis.

A. Preparation of the text corpus

Preparation of documents includes: 1) transfer to digital format for any documents not already digitally readable (e.g. scanned faxes); 2) unification of British and American spellings and correction of spelling errors; and 3) removal of names of authors, non-content "stop words," unnecessary information such as contact information or repetition of the consultation questions. The resulting text files are merged into a single file with each original document tagged with identification variables such as name of the actor or actor type. All of these tasks can be done either automatically by drawing on a Python computer script or with undergraduate coders.

B. Cluster and Correspondence Analysis

T-LAB relies on co-occurrence analysis which is the statistical analysis of frequent word pairs in a text corpus. Using the presence or absence of words in each document, the program calculates an indicator matrix on which to base the classification process. This matrix contains documents in rows and the occurrence of words in each text in columns. Based on ascending hierarchical cluster analysis, T-LAB then identifies clusters of documents. The clusters can be interpreted as frames used by various actors (see also Miller 1997; Schonhardt-Bailey 2005, 2006, 2008).

In a second step, correspondence analysis is used to assess the dimensionality of these frames. Correspondence analysis allows spatial representation of the relation between the clusters whereby position estimates are contingent on correlations, thus distance reflects the degree of co-occurrence. T-LAB cross-tabulates document clusters and words in order to create a second matrix that can be used for factor correspondence analysis. Correspondence analysis provides a measure which indicates the amount of association explained by the dimensions. It aims to account for a maximum amount of association along the first dimension. The second dimension then seeks to account for a maximum amount of remaining association and so forth. The output of this analysis are coordinates for individual interest groups, the frames and the European institutions in the (potentially) multidimensional policy space (see figure 1 & table 4).

Illustration

In order to test the applicability of this content analysis technique, we analyzed the public policy debate concerning the legislative proposal on reduction of CO_2 emissions from cars, since our collaborator Heike Klüver (2009) had already used this issue for a comparison of Hand-Coding, Wordfish (Proksch and Slapin, 2008; Slapin and Proksch, 2008) and Wordscores (Laver et al., 2003). On 7 February 2007, the European Commission proposed a legislative framework to reduce CO_2 emissions from cars to 120g/km in 2012. The Commission called for improvements in vehicle technology that should account for an emission reduction to 130g/km, while efficiency improvements for tires and air conditioning systems as well as a greater use of bio fuels should contribute to further emissions cuts of 10g/km. Furthermore, the Commission suggested fiscal measures, improved consumer information and a code of good practice on car marketing to decrease the popularity of cars with high CO_2 emissions. From 7 February until 15 July 2007 the Commission then launched a public online consultation before adopting its final proposal in December 2007. A wide variety of interest groups took part in this consultation: Out of the twenty-three interest groups under analysis, six represent traditional automobile manufacturers, four are alternative industry associations (such as the biodiesel industry or manufactures of electric vehicles), six are environmental groups, two are media associations, one represents consumer interests, one is a trade union, one represents the tire industry, one represents security interests and one the interest of leasing companies.

Identification of frames

In a first step, a cluster analysis was conducted in order to identify the frames used in the analysis. Three document clusters could be identified (see table 1): The first and smallest cluster (12% of the documents) comprises texts using words such as "advertising," "press" and "media." Table 1 shows the list of typical words of this frame (cluster) which clearly indicates its focus on the impact of the legislative proposal on the advertising industry. The second cluster, which encompasses 28% of the documents, is marked by words such as "automotive," "segments" or "product." The table makes clear that this cluster comprises documents emphasizing the impact of the proposal on the automobile manufacturers. The third and largest cluster (60% of the documents) is represented by typical words such as "LPG," "biodiesel" and "natural." This cluster is actually comprised of two types of groups: Environmental groups as well as alternative industry groups. Further analysis using the keyword-in-context function of the open source text analysis program *Yoshikoder* (Lowe, 2009) reveals that both types of groups emphasize the negative effects of global warming on the environment. However, whereas environmental groups use this frame simply for the intrinsic value of environmental protection, alternative industry groups employ this frame in order to highlight the environmental superiority of their products.

Table 1: Most prominent words distinguishing clusters of actors in the CO_2 emissions debate

Rank according to Chi² Value	**Cluster 1: Press**	**Cluster 2: Industry**	**Cluster 3: Environment / Alternative Industry**
1	Advertising	target	LPG
2	press	political	energy
3	media	value	gas
4	promotional	function	fuel
5	print	approach	fuels
6	literature	Automotive	Biodiesel
7	Publishers	models	oil
8	survey	segments	fuelled
9	believe	reduction	duty
10	restrictions	product	Natural
11	marketing	complementary	light
12	information	system	Methane
13	claim	technologies	biogas
14	freedom	N1	biomethane
15	penalties	rental	diesel
No of texts	3	7	15
% of texts	12%	28%	60%

Note: The Table shows the most typical words per cluster according to their Chi^2 value. The analysis is based on comments from 23 interest groups submitted to the consultation process as well as the initial and final Commission documents.

In a second step, we compared a manual coding of group type to the clusters in which the program classified each actor. Drawing on information gathered on interest groups' websites, we coded them into five different categories: Traditional automobile industry; alternative industry; environmental; the European Commission itself; press / media; and other (this coding was done before the content analysis was conducted to ensure impartiality). Table 2 compares the clusters

from Table 1 with the group type in order to assess the validity of the measurement. Each row represents an interest group together with the cluster membership of the document it submitted to the consultation. The results show that the automated identification of clusters corresponds very strongly (though not perfectly) with a manual coding of group type. In those cases where the coding does not correspond, it is because the document associated with that organization had more in common with the groups in its statistically identified cluster than with our a priori assumption. Note as well that all groups are coded even those which were initially listed as "other" in the manual coding. The system also classifies the initial (Comm1) and final (Comm2) Commission document, showing its correspondence with each of the three clusters.

Table 2: A Comparison of Manual and Automated Classification of Interest Groups

Group			Cluster Membership Scores		
Name	Group type	Best Solution	Press	Industry	Environment
ADTS	Alt. Industry	Environment	0.11	0.31	0.58
AEGPL	Alt. Industry	Environment	0.10	0.19	0.72
EBB	Alt. Industry	Environment	0.13	0.23	0.64
ENGVA	Alt. Industry	Environment	0.09	0.19	0.73
COMM1	**Commission**	**Environment**	**0.22**	**0.36**	**0.42**
COMM2	**Commission**	**Industry**	**0.23**	**0.39**	**0.38**
FANC	Environ. Group	Environment	0.23	0.36	0.41
FOE	Environ. Group	Press	0.54	0.24	0.22
GREENPEACE	Environ. Group	Environment	0.23	0.35	0.43
RSPB	Environ. Group	Environment	0.25	0.35	0.41
TANDE	Environ. Group	Environment	0.27	0.31	0.43
WWF	Environ. Group	Environment	0.22	0.33	0.45
BEUC	Other	Industry	0.25	0.43	0.32
BVRLA	Other	Industry	0.19	0.54	0.27
ETRMA	Other	Environment	0.21	0.30	0.49
ETSC	Other	Environment	0.20	0.36	0.44
ETUC	Other	Industry	0.24	0.41	0.35
AAUK	Press	Press	0.68	0.16	0.16
FAEP	Press	Press	0.88	0.06	0.06
ACEA	Trad. Industry	Industry	0.18	0.56	0.26
JAMA	Trad. Industry	Industry	0.19	0.55	0.26
KAMA	Trad. Industry	Industry	0.19	0.53	0.28
RAI	Trad. Industry	Environment	0.22	0.36	0.43
SMMT	Trad. Industry	Industry	0.24	0.46	0.30
VDA	Trad. Industry	Industry	0.15	0.60	0.25

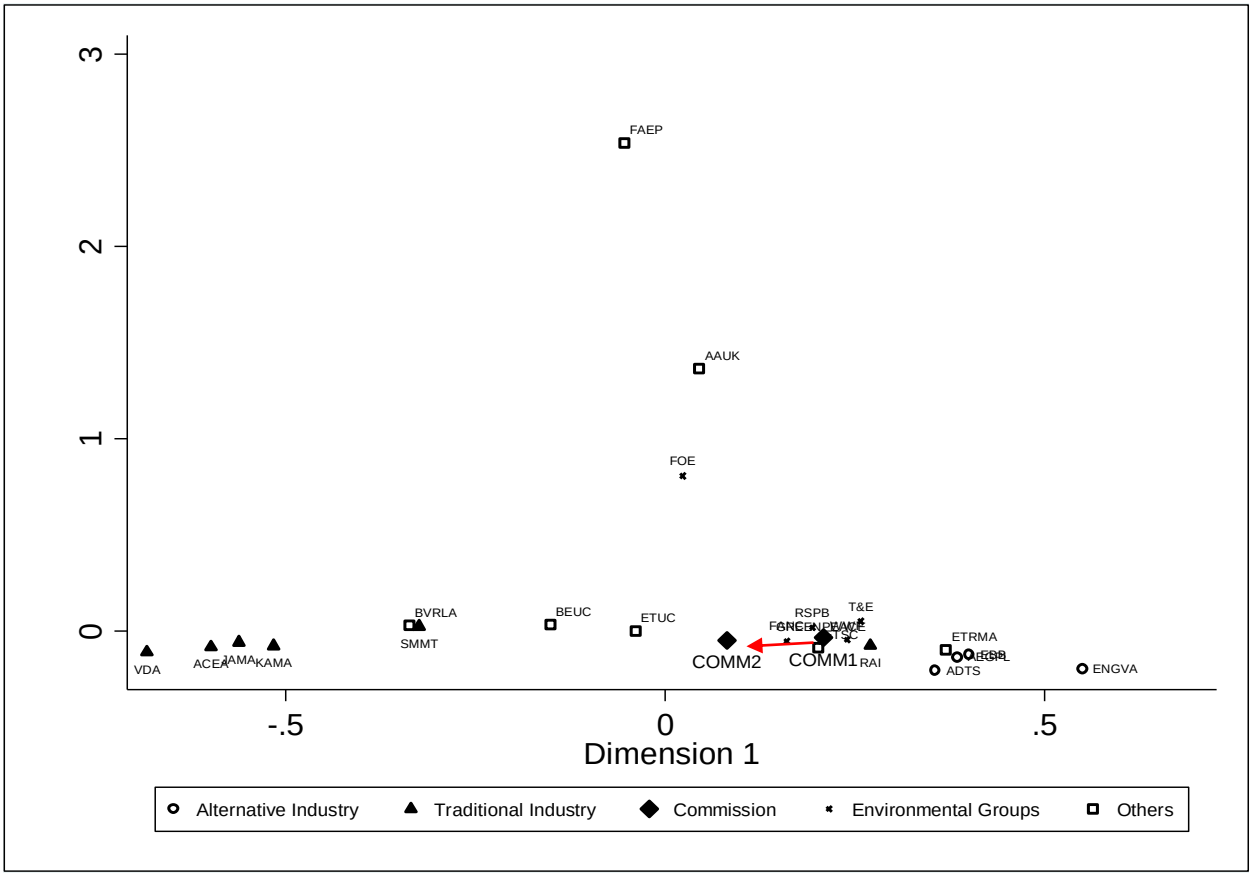
Note: Full names of the associations are available on request. Cluster scores represent the degree to which each document is a member of the various clusters as well as the best cluster solution according to these scores.

Examining dimensionality

Using the correspondence analysis procedure, the underlying dimensions of the frames are identified and the frames are mapped spatially. Correspondence analysis provides a measure which indicates the amount of association explained by the underlying dimensions. The correspondence analysis identifies a two-dimensional space in which the frames are located (see Figure 1). The first dimension accounts for 58 per cent of the association and the second dimension for 42 per cent. Whereas the "Environmental" and "Industry" frames mainly oppose each other on the first dimension and hardly differ in respect to the second dimension, the "Press" frame largely differs from the other two frames along the second dimension and is located more or less in the middle of the "Environment" and "Industry" frame on the first dimension. Hence, the debate surrounding the policy proposal on the reduction of CO_2 emissions

from cars is characterized by a two-dimensional space: "environmental control" and "advertising code of conduct."

Figure 1: A Two-Dimensional Issue-Space for the CO_2 Emissions Debate



	Eigenvalue	Percentage	Cumul. Percentage
Factor 1 (x-axis)	0.19	58.00	58.00
Factor 2 (y-axis)	0.14	42.00	100.00

Note: Dimension 1 is "environmental control;" Dimension 2, the "advertizing code of conduct." Factor scores below indicate the degree of association explained by these two dimensions. The initial (COMM1) and final (COMM2) positions of the European Commission are connected by a red arrow indicating the direction of movement.

As a validity check, we can compare these results with policy position estimates obtained by Hand-Coding and Wordfish (for details, see Klüver, forthcoming). Since Wordscores and Wordfish estimate policy positions only for a single dimension, we compare them separately to the two dimensions identified here (see table 3). The table shows strong correspondence for the first dimension, but little for the second one. These results suggest that the automated procedure accurately replicates the main dimensional structure but also allows additional cleavages not apparent to any technique assuming a single dimension.

Table 3: Correlation of T-LAB Coordinates with Wordfish and Hand-Coding

		T-LAB	Wordfish	Hand-Coding
	T-LAB	1.0		
Dimension 1	**Wordfish**	0.73***	1.0	
	Hand-Coding	0.76***	0.70***	1.0
	T-LAB	1.0		
Dimension 2	**Wordfish**	0.03	1.0	
	Hand-Coding	0.34*	0.70***	1.0

*p<0.1, **p<0.01, ***p<0.001. Source: Klüver (forthcoming)

Assessing the success of frames

Finally, as we have two measures of the location of the European Commission, we can assess the direction of any movement in the official position. In this example, we compare the initial location (t_0) in the February 2007 Communication and the final location (t_1) in the proposed Regulation in December of that same year, after the consultation materials described above had been submitted and reviewed. These are indicated in Figure 1 as COMM1 and COMM2, and a red arrow shows the direction of movement.

Using the coordinates in the multidimensional space (see table 4), one can compute measures of distance such as the Euclidean Distance between the Commission's positions and the frames (clusters) or even the single interest groups to assess in which direction the Commission moved from t_0 to t_1. The Euclidean Distance is given as

$$d(x,y) = \|x-y\|_2 = \sqrt{(x_1-y_1)^2+\cdots+(x_n-y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

Any individual actor can be assessed with regards to this analysis, or we can look more generally at the direction in which the Commission moved: Did it move toward the frame adopted by the greatest number of groups, the business groups, the government organizations, or what? In this example we calculate the distance between the Commission position and the three frames along the two dimensions of debate (see table 5). The analysis shows that the Press Frame is most distant from the Commission and this distance was maintained at the time of the second Commission position. The Environmental Frame is the closest to the Commission but the Commission increased its distance from 0.14 to 0.26 over time. The Industry Frame by contrast is closer to the Commission at time point t_1 than at t_0. In conclusion, the Commission moved over time towards the center of the political space, aligning more with Industry than in its initial position, but remaining closer to the position of the Environment Cluster than to the others even at the end of the process.

Table 4: Location of Interest Groups as well as the EC before and after Consultation

Actor	D1-Envi	D2-Media	Actor	D1-Envi	D2-Media
PRESS CLUSTER	0.01	1.77	ETSC	0.2	-0.09
INDUSTRY CLUSTER	-0.6	-0.08	ETUC	-0.04	0
ENVIRONMENT CLUST.	0.34	-0.08	FAEP	-0.05	2.54
COMMISSION. TIME$_0$	**0.21**	**-0.03**	FANC	0.16	-0.05
COMMISSION. TIME$_1$	**0.08**	**-0.05**	FOE	0.02	0.81
AAUK	0.04	1.36	GREENPEACE	0.2	-0.05
ACEA	-0.6	-0.08	JAMA	-0.56	-0.06
ADTS	0.36	-0.2	KAMA	-0.52	-0.08
AEGPL	0.38	-0.13	RAI	0.27	-0.08
BEUC	-0.15	0.03	RSPB	0.19	0.02
BVRLA	-0.34	0.03	SMMT	-0.32	0.02

EBB	0.4	-0.12	TANDE	0.26	0.05
ENGVA	0.55	-0.19	VDA	-0.68	-0.11
ETRMA	0.37	-0.1	WWF	0.24	-0.05

Table 5: Relative Distance of the EC Position from Major Actors in the Debate, t_0 and t_1

Frames	Distance at t_0	Distance at t_1	Success
Press	1.81	1.82	-0.01
Industry	0.81	0.68	+0.13
Environment	0.14	0.26	-0.12

We have illustrated one element of our project, suggesting ways in which we will be able to analyze the dimensional structure of debates, the positions of each actor, the frames being proposed by different actors, and movement of official bodies during the policy process. In the limited space available we have not discussed change in the position of the European Institutions over time, comparisons across the 120 issues in our sample (e.g., whether issues with low dimensional structures are more likely to see policy change than those with many unrelated frames simultaneously being discussed), and a variety of other factors to which we have alluded in the pages above. Concerning the movement of the European Institutions, we will have various official documents reflecting the official positions over the course of the legislative process such as the policy proposal of the European Commission, opinions of the European Parliament and the Council and the final adopted legislation. For most issues, we will also find official document preceding the policy proposal such as Communications, Green and White Papers as well as a status quo reflected in existing legislation in this policy area (Klüver's analysis of 56 issues on which consultations were available allowed an estimate of the status quo policy position for 75 per cent of the issues). We expect to be able to assess movement of the official position at least from the beginning to the end of the consultation process. However, we should note that we do not expect to be able to trace on a clear chronological basis all the dynamics of framing over time, for example the timing of when certain interest groups make certain claims. Across the sample of issues we do not expect sufficient documents at regular enough time units to support such a dynamic analysis, and our theoretical concerns do not require this.

Website Infrastructure

We have our own theoretical interests in this project but also hope to contribute to a transformation in how political scientists conduct research projects, contributing to ever-expanding databases that can be used by others for a variety of purposes. The PI and Co-PI have already contributed to the US policy agendas project (www.policyagendas.org) and to the lobbying and advocacy project (http://lobby.la.psu.edu), both supported by NSF, and both of which have led to large international collaborations based on the initial design (see http://www.comparativeagendas.org/ and http://sites.maxwell.syr.edu/ecpr/). The projects have also developed large secondary literatures based on use of the data. To this end, we take the creation of a user-friendly and highly assessable website seriously. We will develop a website where we will make all of the documentation publically available for other scholars; detail our data collection process; and our coding schemes so that interested scholars may analyze the original text documents that we collect; all the information will remain in the public domain. In addition, as part of the larger European Science Foundation collaboration, and newly established ECPR working group on interest groups co-chaired by the PI on this proposal, we are promoting the creation of a collaborative network of scholars who may replicate or create analogous data collection projects for the US, EU member states, and other countries.

Feasibility, Logistics, and Timeline

We estimate the data collection work here to be determined largely by the number of actors and documents involved in our sample of 120 issues. Based on an earlier analysis of 56 consultations, we base our calculations on 60 documents per issue. This leads to our request for undergraduate students to handle manual document preparation issues where necessary; graduate assistants at UVA to develop automated scraping and coding processes and to build the website, consulting with Heike Klüver who has detailed experience of the relevant software as well as EU-specific policy and institutional knowledge.

Time period	**Tasks completed**
Spring 2011	• Case selection complete through the associated ESF project.
Summer 2011	• Requested NSF funding begins, July 1, 2011 • Klüver will train UVA graduate & undergraduates on the collection of documents. • Password protected project website portal for sharing documents created.
	• Joint meeting of PI, co-PI and Klüver at UVA.
Fall 2011	• Document Collection, the establishment of a document database and the coding of documents according to information gathered on interest groups' website concerning group type, level of analysis and Member State. Students will be particularly trained in how to identify and search for relevant documents. • Text Processing which includes transfer to digital format for any documents not already digitally readable, unification of British and American spellings, correction of spelling errors, removal of unnecessary information (e.g. contact details) and merging the different documents into one single text file with each original document tagged with identification variables such as name of the actor or actor type. Students will be trained in using a Python Script to automatically edit the documents and in how to conduct the necessary manual steps. • Based on preliminary case example, document collection and text processing, we expect to finish 30 issues in 3 months.
Spring 2012	• Finish collecting documents on another 60 issues. • Graduate assistants will be trained in analyzing the processed texts using T-LAB and begin performing the text analysis. We estimate that they can finish the text analysis for 30 issues by the end of the semester.
	• Joint meeting of PI, co-PI and Klüver at UNC to assess progress.
Summer 2012	• Complete document collection for the last 60 issues. • Complete text processing for the last 60 issues. • Complete text analysis of another 60 issues.
Fall 2012	• Complete text analysis for the last 30 issues and begin analysis of data. • External website for sharing data with scholarly community will be established, including original and process documents, quantitative coding results, and background information on each case.
Spring 2013	• Data analysis finalized, project coordinators write up the results and present them at several conferences before submitting the papers to journals. • June 30, 2013 - funding expires

References

Bailey, Andrew and Cheryl Schonhardt-Bailey. 2008. Does deliberation matter in FOMC monetary policymaking? The Volcker Revolution of 1979. *Political Analysis*. 16(4): 404-427

Baumgartner, Frank R., Jeffrey M. Berry, Marie Hojnacki, David C. Kimball, and Beth L. Leech. 2009. *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago: University of Chicago Press.

Baumgartner, Frank R., Suzanna L. De Boef and Amber E. Boydstun. 2008. *The Decline of the Death Penalty and the Discovery of Innocence.* New York: Cambridge University Press.

Baumgartner, Frank R., and Bryan D. Jones. 2009. *Agendas and Instability in American Politics*. 2nd ed. Chicago: University of Chicago Press.

Baumgartner, Frank R., and Christine Mahoney. 2008. The Two Faces of Framing: Individual-Level Framing and Collective Issue-Definition in the EU. *European Union Politics* 9 (3): 435–49.

Berinski, Adam J., and Donald R. Kinder. 2006. Making Sense of Issues Through Media Frames: Understanding the Kosovo Crisis. *Journal of Politics* 68, 3: 640–56.

Druckman, James N. 2001. On the Limits of Framing Effects: Who Can Frame? *Journal of Politics* 63, 4: 1041–66.

Entman, Robert M. 1991. Framing US coverage of international news: Contrasts in narratives of the KAL and Iran air incidents. *Journal of Communication* 43(4): 51-58

Gilliam, Franklin D., Jr., and Shanto Iyengar. 2000. Prime Suspects: The Influence of Local Television News on the Viewing Public. *American Journal of Political Science* 44, 3: 560–573.

Image. 2009. ALCESTE. Logiciel d'Analyse de Données Textuelles. Version 4.7. http://www.image-zafar.com/. last accessed on 4 May 2009.

Klüver, Heike. forthcoming. Measuring interest group influence using quantitative text analysis. *European Union Politics.*

Klüver, Heike. 2009. Understanding policy-making in the European Union: Interest group influence on policy formulation. Paper presented at the 67th National Conference of the Midwest Political Science Association, Chicago. 2 - 5 April 2009.

Lancia, Franco. 2009. T-LAB Pro. Tools for Text Analysis. Version 6.1., http://www.tlab.it. last accessed on 4 May 2009.

Laver, Michael, Kenneth Benoit and John Garry. 2003. Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review* 97 (2): 311–331.

Lindblom, Charles E. 1977. *Politics and Markets*. New York: Basic Books.

Lowe, Will. 2009. Yoshikoder. Software for multilingual content analysis. Version 0.6.3., http://www.yoshikoder.org/. last accessed on 4 May 2009.

Mahoney, Christine. 2008. *Brussels vs. the Beltway: Advocacy in the United States and the European Union. Washington. DC: Georgetown University Press.*

Miller, M. Mark. 1997. Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review*. 15(4): 367-378β

Quinn, Kevin M. et al. 2008. How to Analyze Political Attention with Minimal Assumptions and Costs. Harvard University Department of Government, working paper, July 14.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2008. WORDFISH: Scaling Software for Estimating Political Positions from Texts. Version 1.2., URL (consulted Sept. 2008): http://www.wordfish.org

Riker, William H. 1986. *The Art of Political Manipulation*. New Haven: Yale University Press.
Riker, William H. 1996. *The Strategy of Rhetoric.* New Haven: Yale University Press.
Schattschneider, E. E. 1960. *The Semi-Sovereign People*. New York: Holt. Rinehart and Winston.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science* 52 (3): 705–722.
Schonhardt-Bailey, Cheryl. 2005. Measuring ideas more effectively: An analysis of Bush and Kerry's national security speeches. *PS: Political Science & Politics*. 38(4): 701-711.
Schonhardt-Bailey, Cheryl. 2006. *From the Corn Laws to Free Trade: Interests, Ideas, and Institutions in Historical Perspective*. Cambridge: MIT Press.
Schonhardt-Bailey, Cheryl. 2008. The Congressional Debate on Partial-Birth Abortion: Constitutional Gravitas and Moral Passion. *British Journal of Political Science*. 38 (3): 383–410.