The Missing Link in Macro-Quantitative Political Science Research: The Case of Electoral System Choice

Abstract: This article critically discusses the quantitative analysis of political reforms that are the result of non-recurrent collective decision-making in presence of strategic interactions using the example of the debate on electoral system choice. An expert survey among historical experts specialized in political, economic, and social history in the period 1890-1939 shows that country experts remain unconvinced by the currently dominant explanations of electoral system choice in political science. This negative assessment cannot be explained by a general reluctance on the part of the historical experts to endorse quantitative approaches, generalization, and comparison. We present an alternative explanation for this negative assessment. We develop our argument in three steps. Firstly, we argue that political decisions are made in a historical context. In particular, these decisions are made in presence of multiple, non-independent issues on the political agenda. Secondly, in such a context macro-political actors (e.g. parties, governments, interest groups) are likely to suffer from collective decision-making problems, thereby making collective behavior virtually impossible to predict. Thirdly, while we might assume the absence of collective decision-making problems, such an assumption is only useful iff the theoretical predictions are tested. This is, however, for several reasons rarely done. Macro-quantitative approaches are thus often unsuitable for the causal analysis of political reforms.

Patrick Emmenegger, University of Southern Denmark, emm@sam.sdu.dk
Klaus Petersen, University of Southern Denmark, klaus.petersen@hist.sdu.dk

Word count: 9’523 (main text and footnotes)

Acknowledgements: We thank all country experts for participating in the expert survey. In addition, we thank the journal editors and the researchers who helped us locate the country experts. We thank Aurélien Buffat for translating the questionnaire into French and Agner Holmbjerg Schibler, Emil Damkjær Herløv Hansen, and Steffen Gantzel Pedersen for helping us create the web-based expert survey. An earlier version of this paper has been presented at the seminar on comparative methodology (Odense, May 2011) and the 18th annual conference of the Council for European Studies (Barcelona, June 2011). We thank all participants and in particular John Gerring, Gary Goertz, Markus Kreuzer, Jon Kvist, Caroline de la Porte, and Sven Steinmo for their helpful comments. Last but not least, we thank André Blais, Carles Boix, and Torben Iversen for revising our summaries of their arguments.
1 Introduction

The choice of electoral systems in the period 1890 to 1939 is one of the most important topics in contemporary comparative political science research. Electoral institutions are core institutions of political systems. Sartori (1968: 273, quoted in Lijphart 1994: 139) calls the electoral system “the most specific manipulative instrument of politics”. Yet we know little about what causes cross-national variation in electoral laws, despite a recent surge in the literature on electoral system choice with important contributions by some of the most prominent scholars in the field (see Table 1). Among recent contributions, all of them briefly discussed in part two of this paper, Blais et al. (2005) and Cusack et al. (2007, 2010) stress political consensus with regard to electoral system reform, while Boix (1999, 2010) and Alesina & Glaeser (2004) emphasize political conflict. At the same time, Boix (1999, 2010) and Cusack et al. (2007, 2010) stress how established political groups attempted to secure their position in the political system – either to safeguard themselves against a socialist threat (Boix 1999, 2010) or to protect co-specific assets (Cusack et al. 2007, 2010) – while Alesina & Glaeser (2004) and Blais et al. (2005) emphasize processes of democratization – either coming from within the country (Alesina & Glaeser 2004) or coming from outside the country (Blais et al. 2005). No matter how you look at this debate, this is not a very satisfactory situation.

[Table 1]

In a discussion of Boix (1999) and Cusack et al. (2007), Kreuzer (2010) believes to have found the reason for this situation. He argues, “political scientists commonly draw on history but often do not read actual historians carefully” (Kreuzer 2010: 369). Moreover, he argues, “it would be beneficial to first do the more nuts-and-bolts work of using historical knowledge to improve the quantitative study of institutional origins” (Kreuzer 2010: 385). He makes three empirical contributions: Firstly, he revisits their data collection. Secondly, he replicates their statistical analyses (using the new data) to test the robustness of the findings. Thirdly, he tests more observable implications of the causal mechanisms put forward by these authors. Kreuzer (2010: 383) concludes with a rather negative assessment of Cusack et al. (2007), while he argues that “Boix’s closer dialogue with historical knowledge is vindicated by the
greater robustness of his findings”, even though he also notes that only “nine of the 24 cases (36%) match all five of the causal links in Boix’s overall argument” (Kreuzer 2010: 380).

We applaud Kreuzer (2010) for his efforts, but we believe his critique only scratches the surface. In this paper, we argue that macro-quantitative political science research suffers from a more fundamental problem. Taking Kreuzer’s (2010: 370) call to take not only history, but also the work of historians seriously as a starting point, we conducted an expert survey among historical experts specialized in political, economic, and labor history in the period 1890-1939 from 20 countries. Among others, we asked the participants to rate the explanations of Boix (1999, 2010), Blais et al. (2005), and Cusack et al. (2007, 2010) on a scale from 1 to 10. The results of this survey are reported in the third part of this paper. Overall, the experts’ assessment of the three explanations is rather negative. In addition, we show that this negative assessment cannot be explained by a general reluctance on the part of historical experts to endorse quantitative approaches, generalization, and comparison. Rather, the negative assessment seems to be the result of other factors.

In the fourth part, we provide an explanation for the negative assessment. We develop our argument in three steps: Firstly, borrowing from existing critiques of macro-quantitative political science research (e.g. Capoccia and Ziblatt 2010), we stress the context-dependence of political reforms. In particular, we stress the violation of the causal homogeneity assumption, emphasize the lack of independence of the cases due to diffusion processes between nation-states, and document the presence of multiple issues on the political agenda.

Secondly, we argue that given strategic behavior on the part of political actors and in presence of multiple issues on the political agenda, we cannot treat collective behavior (of macro-political actors such as parties and interest groups) as if it simply were aggregate individual behavior. Rather, we have to acknowledge that collective decisions cannot easily be predicted because of social choice problems. This is what we refer to as the ‘missing link’. By ignoring the possibility of social choice problems, macro-quantitative political science research makes implausible assumptions about collective decision-making.

Thirdly, while we might assume the absence of strategic behavior and thus the absence of social choice problems, such an assumption is only useful iff the theoretical predictions are tested. This is, however, rarely done in macro-quantitative political science research. On the one hand, advocates of a given theory are typically too familiar with their cases to adequately

---

1 Cusack et al. (2010) forcefully reject Kreuzer’s (2010) assessment, arguing that all but one of Kreuzer’s historical claims are incorrect. Boix’s (2010) response is more positive. However, he contends that only three (instead of ten) cases disconfirm his theory (Boix 2010: 405).
test their macro-models. Therefore, theories are rarely tested using *new* data.\(^2\) On the other hand, empirical tests based on observational data are never fully conclusive and therefore typically rejected by the advocates of the tested theory. We provide examples for both of these explanations.

Let us quickly add what we are *not* arguing: Firstly, our argument applies only to cases in which the macro-phenomenon is the result of (non-recurrent) collective decision-making in presence of strategic interactions. This is typically the case for decision-making in macro-political units such as parties and governments (e.g. political reforms), but not necessarily the case for aggregate behavior in markets (e.g. demand for goods). In addition, our argument does not apply to individual-level outcomes (e.g. political decisions by individuals). We focus on non-recurrent collective decision (such as electoral system choice) because in case of recurrent collective decision-making (such as decisions on public social expenditure) it is possible that due to the large number of decisions the average result approximates the expected value. This is of course only the case if we assume that there is no systematic bias that affects the average over the long run.

Secondly, we are not rejecting statistics and quantitative approaches per se. Statistics is a powerful way of organizing and analyzing data. We simply argue that quantitative approaches, as commonly used, are sometimes not useful for causal analysis in macro-comparative political science research. This is particularly the case if the dependent variable is the result of a (non-recurrent) collective decision-making process in presence of multiple issues on the political agenda and strategic behavior on the part of political actors.

Thirdly, we are not arguing that historical events are essentially unique and thus not comparable. Rather, we argue that context-dependence has to be taken into account, but context-dependence does not make comparison impossible. We therefore conclude this paper with a discussion of possible alternative research strategies that would allow for the macro-comparative analysis of electoral system choice.

### 2 Four and a half arguments about electoral system choice

This section briefly presents the four and a half arguments on electoral system choice in the period 1865 and 1939 (see Figures 1 to 5). We use these starting and end points because Thomas Hare proposed the first system of proportional representation (PR) in 1864 and

\(^2\) In fact, in macro-quantitative political science research it is often impossible to test theories using new data because the set of cases under analysis is identical with the universe of cases.
because the onset of World War 2 with the accompanying collapse of several democracies using PR challenged the previous consensus on the democratic virtues of PR (Blais et al. 2005: 182, 186). The first country to use a proportional representation system was Belgium in 1899. Thereafter, several other countries adopted PR, while some countries kept using majoritarian representation systems (MR).

The debate on the determinants of electoral system choice started with Boix’s (1999) prize-winning “Setting the Rules of the Game”, where he, based on Rokkan (1970), identifies a ‘socialist threat’ as the main determinant of the adoption of PR. Given that MR rewards strong parties and punishes weak ones, the emergence of a new (strong) party could endanger the position of the established ones. More concretely, electorally strong socialist parties endangered the established parties if the established parties were fragmented (and thus weakened) and the extension of suffrage to lower social classes (who are expected to vote for the socialists) was likely to boost the vote share of the socialists. In such a situation, the established parties, if endangered, benefitted from a move to PR and, given their still powerful position, were able to enact PR unilaterally.

[Figure 1]

Boix (2010) refined his argument in his response to Kreuzer (2010). He now distinguishes between segmented electoral arenas (those where the support of a particular party is highly concentrated in a particular geographic area or social sector) and competitive electoral arenas (those where several parties contend for the vote of a least some fraction of the electorate). In segmented electoral arenas, the established parties favored PR if the new entrant threatened their electoral hegemony (mainly urban parties). Otherwise they did not support PR. In contrast, in more competitive electoral arenas, the position of the established parties was shaped by the extent to which they were dominant in the electoral arena vis-à-vis the other established parties conditional on the entry of third parties. The party that expected to become the focal point around which non-socialist voters would eventually rally had little incentive to support PR. By contrast, those established parties that could not expect to become the dominant non-socialist party preferred PR.

[Figure 2]
According to Alesina & Glaeser (2004), the proportional representation movement began in the middle of the nineteenth century and was subsequently taken up by the mobilized, revolutionary left. Given that MR stymied their electoral chances, the political left demanded proportional representation. Using strikes, street protests, the threat of violence or simply their increasing political power, the left was able to force reform against the will of the established parties, especially parties representing the old elite. In some countries, the left took advantage of the weakness of the right during national crises such as in the aftermath of World War 1. Thus, PR was implemented either by the left after the left came to power, or by the right after the right was forced to do so by a mobilized, revolutionary left. In countries in which the left lacked the necessary power resources, the right successfully defended MR electoral systems.

Blais et al. (2005) highlight two factors facilitating the shift to PR. Firstly, the spread of democratic ideas and the general perception that PR was the fairest electoral system increased the pressure on countries to democratize and to adopt PR. Secondly, the presence of a majority run-off electoral system (in contrast to a plurality electoral system) led to a higher number of parties and therefore a more fragmented party system. In these fragmented party systems, opposition to the introduction of PR was weaker because of the regular occurrence of coalition governments and the greater uncertainty as to the optimal strategies for winning elections. In addition, the presence of smaller, electorally disadvantaged parties meant that some parties strongly favored the adoption of PR. As a consequence, PR was adopted in many countries without much debate.

According to Cusack et al. (2007, 2010), PR was adopted in countries with traditions of cooperation and negotiated decision-making. These traditions of cooperation encouraged the production of co-specific assets, i.e. investments by both companies and workers and where return on investment was only possible in presence of cooperation between these diverse actors. By the end of the nineteenth century, industrialization turned local workers’
organizations into a national movement and increased the role of the national level in regulatory policy-making. This created a collective action problem because MR did not allow for the proportional representation of all relevant social and economic interests at the national level. Consequently, PR was (consensually) adopted in order to restore a negotiation-based political system in which national parties represented all relevant social and economic interests.³

[Figure 5]

The arguments stress different factors: traditions of economic cooperation, the power of the left, the spread of democratic ideas, and the strategic behavior of partisan actors. Alesina & Glaeser (2004) and Boix (1999, 2010) stress the role of the political left, Blais et al. (2005) and Boix (1999, 2010) the role of fragmented party systems, and Alesina & Glaeser (2004) and Blais et al. (2005) the role of democracy (either in the form of democratic ideas or the emancipation of the labor movement). The argument by Cusack et al. (2007, 2010) is more unique because it emphasizes economic interests and historical legacies. All authors claim that the available empirical evidence supports their argument.

In the next section, we present the results of an expert survey among historical experts specialized in the areas of political, economic, and social history in the period 1890-1939. Among others, we asked the participants to evaluate the arguments by Blais et al. (2005), Boix (1999, 2010), and Cusack et al. (2007, 2010). We did not consider the argument by Alesina & Glaeser (2004) because the argument is not as developed as the other arguments. In particular, it cannot explain why PR was sometimes introduced by the political right against the will of the political left and sometimes by the political left against the will of the political right. The argument is thus underspecified.⁴

3 Political science seen through historians’ looking glass

³ According to Cusack et al. (2010: 397), the argument pertains to the choice of electoral systems once the transition to democracy had been completed because before then calls for PR were often tied to calls for democracy, which were resisted by the political right.
⁴ We would like to emphasize that the political left is likely to have played a more active role than it is acknowledged in the other three arguments. Using an indicator of the position of the political left on the issue of electoral system choice (based on data from the experts survey) and the effective electoral threshold as an indicator of the proportionality of electoral systems (Boix 1999, 2010; Cusack et al. 2007, 2010), we find a bivariate correlation of r=0.83 (N=20). Thus, we tend to find PR where the political left supported PR (cf. Rodden 2009).
This section presents the results of our expert survey. We have compiled the list of experts by contacting leading national and international history journals and by asking them to nominate experts to participate in the survey. Subsequently, we have contacted these historical experts and asked them to participate in an online-survey. The survey consisted of four parts. In the first part, we asked for the participants’ academic background, countries and periods of expertise, and the extent of collaboration with social scientists (defined as political scientists and sociologists). Subsequently, we asked them to assess on a very general level the usage of history by social scientists. The third part of the survey was introduced by a short description of the debate on electoral system change and the summaries (max. 400 words) of the arguments by Blais et al. (2005), Boix (1999, 2010), and Cusack et al. (2007, 2010). In case of the arguments by Boix (1999, 2010) and Cusack et al. (2007, 2010) we incorporated their recent revisions.\(^5\) The participants in the survey were then asked to rate the three arguments from 1 (does not explain at all the presence or absence of electoral system change) to 10 (fully explains the presence or absence of electoral system change). In the last part of the survey, the participants were asked a second battery of general questions, centering on issues such as the need for generalizations and the usage of quantitative approaches. Throughout the survey, participants had the opportunity to complement their answers with written comments (see Appendix for a detailed description of the survey including the summaries of the three arguments).

Figure 6 displays for each country the average evaluation of the best performing explanation (the average evaluation across all three explanations is shown in the Appendix). Three lessons can be drawn from Figure 6. Firstly, overall, the evaluation is rather negative. Only 7 out of 20 cases score above the point of maximum ambiguity (5.5). In two cases (Canada and Ireland), all explanations received an average score of 1. Secondly, historical experts consider the argument by Cusack et al. (2007, 2010) to be the most convincing. In 8 out of 20 cases (Australia, Denmark, France, Germany, Great Britain, Italy, Netherlands, and New Zealand), the argument by Cusack et al. (2007, 2010) has received the highest average score. In contrast, historical experts consider in only 3 cases (Japan, Norway, and Sweden) and 2 cases (Finland and Spain), respectively, Boix (1999, 2010) or Blais et al. (2005) to be the most convincing. In the remaining seven cases, at least two explanations performed equally. Finally and somewhat surprisingly, ‘pure historians’ have evaluated the three arguments more positively than historical experts who declared to have some social science background. The

\(^5\) We are very grateful to André Blais, Carles Boix, and Torben Iversen for revising our summaries of their arguments.
latter gave the three arguments an average score of 4.11. In contrast, ‘pure historians’ scored them on average 4.72. This difference is significantly different from zero at the 10 percent level.\textsuperscript{6} Thus, the negative evaluation is unlikely to be caused by a lack of familiarity with social science research.

[Figure 6]

How can we explain this result? At first sight, we might dismiss these results by pointing to disciplinary differences, in particular the (often claimed) general reluctance on the part of historical experts to endorse quantitative approaches, generalization, and comparison (see Elman and Elman 2001 for a critical discussion). However, as Table 2 shows, these three factors are unlikely to explain the negative evaluation. After evaluating the three arguments, the survey participants generally expressed high levels of support for generalizations and comparisons, and even expressed some support for quantitative approaches. Thus, the negative assessment seems to be the result of other factors.

[Table 2]

As Table 3 shows, two other factors seem to play a more important role. Firstly, most historical experts are doubtful that we can always give clear answers to questions about historical facts. This is important because in macro-quantitative political science research, we are generally dependent on being able to assign scores for every case being analyzed. The level of consensus among political groups on the issue of electoral system choice is one of the key observable implications of the arguments put forward (see Table 1). Using this example, we have asked the participants whether we are generally able to give clear answers to questions about issues such as the level of consensus among political parties. Somewhat surprisingly, 58 percent of the survey participants answered that it is only sometimes possible (see Table 3). Only about 36 percent answered that it is always or almost always possible to clear answers to such questions. This sheds a different light on the debate between Cusack et

\textsuperscript{6} This difference increases when we remove those respondents from the sample that consider themselves to be social scientists first and historians second (4.72 vs. 3.82, t-value of -2.42).
al. (2010, 2010b) and Kreuzer (2010, 2010b) on the coding of indicators. Some cases might simply defy a straightforward coding.

Secondly, historical experts are critical of variable-oriented medium-N approaches to study historical developments (see Table 3). Variable-oriented medium-N approaches typically make generalizations across the whole sample (or the whole universe of cases) even though they cannot explain every case. This means that even for countries for which the explanation is not correct, the argument should be considered an important part of the specific national explanation. Almost 70 percent of the participants think that this is never or almost never a reasonable way to study historical developments. Given the otherwise rather positive attitudes towards generalization, comparison, and quantitative approaches, it is quite likely that it is the focus on variables instead of cases that makes historical experts reject this kind of approaches. In the comment sections, several experts expressed their irritation how an explanation could be considered a general explanation if it cannot explain every case.

[Table 3]

This view is also confirmed by the survey results displayed in Table 4. These questions were asked before the empirical example was introduced. As is to be expected, historical experts are rather critical of social scientists’ historical knowledge and think that social science research does not pay sufficient attention to the historical dimension of political and social developments. Potentially more damaging, they believe that social scientists are generally more interested in defending their theories than in analyzing actual historical developments and that social scientists often draw facts from historically grounded accounts and then use them for timeless propositions (see also Thies 2002). For all four questions, the means are significantly different from the point of maximum ambiguity (5.5) at the 1 percent level. Interestingly, historical experts with some social science background are again more critical of social science research than ‘pure historians’.

[Table 4]

---

7 The difference between ‘pure historians’ and historians with some social science background is significantly different from zero at the 5 percent level in case of the evaluation of social scientists’ historical knowledge and the attention social science research pays to the historical dimension of political and social developments.
Thus, ‘taking historians seriously’ (and asking them to evaluate our work) is a painful experience for political scientists. Overall, the evaluation is rather negative. What is particularly bewildering is the fact that the evaluation by historical experts with social science background is more negative than the evaluation by what we refer to as ‘pure historians’. Therefore, the negative evaluation cannot be dismissed by pointing to the lack of familiarity of historical experts with social science research. Neither can it be dismissed by pointing to disciplinary differences such as the (seemingly) general reluctance on the part of historical experts to endorse quantitative approaches, generalization, and comparison. Rather, the problem seems to lie elsewhere. Historical experts are doubtful of the possibility to provide clear answers to questions about historical facts (which we need in order to make indicators), they stress the context-dependence of historical events, and they reject variable-oriented approaches (that make generalizations across the whole sample even though they cannot explain all cases).

4 The missing link in macro-quantitative political science research

In this section we present an explanation for this negative assessment. The section is structured in two parts. In the first part, we stress the context-dependence of political reforms. In particular, we stress the violation of the causal homogeneity assumption, emphasize the lack of independence of the cases due to diffusion processes between nation-states, and document the presence of multiple issues on the political agenda. The second part then explains why the context-dependence of political reforms makes the methodological approach chosen by Blais et al. (2005), Boix (1999, 2010), and Cusack et al. (2007, 2010) inappropriate for analyzing electoral system choice in the period 1890 to 1939. More precisely, we argue that in presence of multiple issues on the political agenda, non-recurrent decisions by collective actors (such as parties, governments, and interest groups) cannot easily be predicted because of social choice problems. This is what we refer to as the ‘missing link’. Furthermore, while we might assume the absence of social choice problems, such an assumption is only useful iff the theoretical predictions are tested. However, for several reasons these predictions are rarely tested.

4.1 Electoral system choice in the period 1890 to 1939: The historical context

Following Sartori (2009: 16), comparability requires homogeneity, that is, comparisons can only be applied to things that belong to the same ‘genus’, the same class. Once differences-in-
kind are identified, differences-in-degree within the same class of units can be meaningfully compared. Put differently, once we have separated apples and pears, we can start comparing the pears and the apples among each other. The crucial point here is that the classification of units into different classes of units is logically prior to the comparison of some properties within the sub-species. This is point is not new, but as Sartori (2009: 16) argues, these “taxonomical requisites of comparability are currently neglected if not disowned”.

This is also the case for electoral system choice in the period 1890-1939. Free and competitive elections are a core part of every definition of democracy. In democracies, the government is determined by free and competitive elections, while in non-democracies elections do not play the same role. Thus, it is an uncontroversial claim to argue that electoral system reform has a different significance in democracies than in non-democracies. In a similar vein, it is a completely different thing to introduce proportional representation (PR) during a process of democratization, given that PR was considered more democratic (Blais et al. 2005), than in a country that has been democratic for almost 100 years. In 1899, when Belgium was the first country to introduce PR, the United States of America had already been democratic for 100 years.

Figure 7 displays the level of democracy, measured as the difference between the 10-point democracy index and the 10-point autocracy index in the Polity IV dataset (Marshall et al. 2010), of three countries considered in all studies under investigation. All three countries adopted PR: Switzerland and the Netherlands in 1918, Italy in 1919. However, these countries made the transition to PR in completely different contexts. Switzerland was a stable democracy with universal male suffrage, democratic since 1848, and not involved in the First World War. The Netherlands made the transition to PR during a process of democratization (from -2 to +10 according to the Polity IV dataset), which also led to the introduction of universal male suffrage. Unlike the Netherlands and Switzerland, Italy was involved in the First World War, suffering many casualties. In addition, Italy was, despite universal male suffrage, far from being democratic. Overall, the political situation in postwar Italy could be described as messy (in 1922, Mussolini became Italian Prime Minister).

[Figure 7]

However, despite these differences-in-kind, Boix (1999), Blais et al. (2005), and Cusack et al. (2007, 2010) treat these cases as if they belong to the same ‘genus’, that is, the same class.
Thus, Boix (1999) and Cusack et al. (2007) compare the average effective threshold in Italy in the period 1919-1923 (average level of democracy = -1.6) to the average effective threshold in the United States of America in the period 1919-1939 (average level of democracy = +10). In a similar vein, Blais et al. (2005) treat the 1925 reform in Japan (level of democracy = +1) the same way as they treat the 1918 reform in Switzerland (level of democracy = +10). In addition, they all ignore that the democratic system of several of these countries collapsed only few years after the introduction of PR. For instance, Boix (1999) uses the period 1931-1936 in Spain as a case, ignoring that Spain did not return to democracy before 1978. In sum, all these authors assume causal comparability despite differences-in-kind.

The historical experts participating in the expert survey have been equally irritated by the lack of attention to the political context. For instance, Figure 6 shows that none of the historical experts covering Canada and Ireland considers any of the three arguments helpful. One Irish survey participant considered it impossible to evaluate the three arguments given that the electoral system had been imposed by the United Kingdom. With regard to Canada, one participant wrote, “there are unique historical facts that explain the strong support for the first past the post system and the rejection of PR”. Finland is another very exceptional case. Becoming independent only in 1917, Finland immediately introduced universal suffrage and PR because it had no room for disputes since “no party wanted to give the Russian government the chance of intervention”.

We could list more examples, but they would not change the main message: To be truly convincing, Blais et al. (2005), Boix (1999, 2010), and Cusack et al. (2007, 2010) would need to show that all cases are cases of the same class of units. Put differently, the authors need to show that all these cases are indeed causally comparable. Figure 7 shows that this is unlikely to be the case. Admittedly, emphasis on context can be pushed too hard, but comparing electoral system change in stable democracies and in autocracies is unlikely to satisfy the assumption of causal homogeneity that underlies macro-quantitative political science research.

Figure 7 allows for a second observation. All three countries adopted PR around the same time, at the end of World War 1. This is in fact a general pattern: Belgium introduced PR in 1899, Finland (after becoming independent) in 1907, and neighboring Sweden in 1909, which was followed by Denmark (1915). Between 1918 and 1922, Austria, Germany, Italy, Ireland, the Netherlands, Norway, and Switzerland adopted PR. Greece and Japan followed in 1925.
and 1932 respectively (Colomer 2005). Given this temporal pattern, we are unlikely to be able to treat these countries as independent cases (Galton’s problem).

Both within political science and historical research there has been a growing interest in the diffusion of ideas and policy learning across national boundaries (Rose 1993; Werner & Zimmermann 2002; Iriye & Saunier 2009). Behind the different concepts of diffusion and transnational history lays the basic conclusion that nation-states are not closed containers but are very closely interrelated.\(^8\) In a recent study on the spread of democracy in the late part of the 20\(^{th}\) century, Elkins & Simmons (2005) have analyzed diffusion mechanisms such as ‘learning’ and ‘adaptation to altered conditions’. Rodgers (1998) documents that the first decades of the 20\(^{th}\) century were also characterized by intense transnational interaction and internationalism. Experts, civil servants, political parties, and organizations such as the US-based Proportional Representation League (1893), the France-based Ligue pour la Représentation Proportionnelle (1901), the British Electoral Reform (1883) and Proportional Representation (1905) societies as well as similar movements in other countries formed a network were ideas, arguments, and experiences could travel.

There can be no doubt that the successful introduction of PR in pioneering countries such as Switzerland (on the cantonal level) and Belgium influenced other countries. As early as 1895, we find long and detailed discussions of Switzerland’s “very valuable experiments in the organization of democracy” (Wuarin 1895: 1), soon to be followed by intense discussions of ‘the Belgian solution’ (Mahaim 1900). Williams (1914: 121) argued in 1913 that the “Swiss example was infectious rather than contagious”, while Humphrey (1911, appendix IV) argued that Finland followed the Belgian model even though “the Finnish system shows boldness, originality, and, it must be added, no little complexity of procedure.” Decisions taken in one country legitimized demands for PR in other countries and these demands grew ever stronger as the majority of European countries had opted for PR. In this context, it is important to note that only Blais et al. (2005) have taken the lack of independence of cases into account. In contrast, Boix (1999, 2010) and Cusack et al. (2007, 2010) treat their cases as if they were completely independent from each other.

Finally, and in the context of this article most importantly, all of these authors fail to take sufficiently into account the fact that electoral system choice was but one among many issues on the political agenda. Blais et al. (2005) argue, the spread of democratic ideas increased the pressure on European governments to adopt PR, and Boix (1999) correctly underlines the

---

\(^8\) This is also a well documented in more recent studies on the conceptual history of ‘democracy’ and ‘Liberalism’ (cf. Rosanvallon 1995; Leonhard 2001; Dunn 2005; Nevers 2011).
importance of considering the question of universal male suffrage (even though it is not part of his empirical analysis), but one could also add items such as protection of political minorities, secret ballots, boundaries of constituencies, republicanism, female suffrage, one- or two-chamber systems, and ticket methods. In this context, it is important to note that the adoption of PR generally took place in the context of intensive democratization (e.g. Austria, Denmark, Germany, Greece, the Netherlands, and Sweden) or in the process of independence (e.g. Finland, Ireland, and Norway). In contrast, the adoption of PR was a rare feature among stable democracies (Switzerland and to a certain extent Belgium). In general, stable democracies opted to keep MR (Australia, Canada, France, New Zealand, United Kingdom, and United States of America).

As a result, the decision to adopt PR was often part of comprehensive reforms that included multiple issues (cf. Dodd 1910, 1911). Given the complexity of these reform packages, the effect of proposed reforms were often impossible to predict. In addition, PR was often not the main concern of party leaders. In the intense French discussions around 1900, PR was second to the choice of ticket methods (Garner 1913) and for socialist parties the expansion of male suffrage was generally the most important issue (Boix 1999). For instance, in Sweden, the introduction of PR was the result of a political compromise that traded universal male suffrage (main demand of the Socialists and Liberals) for the introduction of PR (main demand of the Conservatives) (Särälvik 2002: 236). In Switzerland, the introduction of PR was one of the core demands in the 1918 general strike. Other core demands included universal female suffrage, the introduction of a public old-age pension scheme, and shorter working hours. Switzerland introduced PR at the federal level in 1918, however, not by means of a parliamentary discussion, but by means of a constitutional amendment following a popular initiative (Lutz 2004). If not for this very unique political institution that lets voting-age (male) citizens propose and decide on constitutional amendments, Switzerland would have been likely to keep MR for at least another couple of years.

In sum, we argue that PR has been adopted in very diverse contexts (autocracy, democracy, democratization, independence), thereby violating the assumption of causal homogeneity, that the decision to adopt was influenced by events outside the nation-state, thereby violating the assumption of independence, and that electoral system choice was not the only issue on the political agenda, thereby allowing for strategic maneuvering on the part of political actors. As we will argue in the next section, in presence of multiple political issues and strategic

---

interactions, social choice problems are virtually unavoidable. If political parties allow for some sort of collective decision-making, i.e. if we assume that there is no ‘dictator’, decisions by collective political actors become unpredictable.

4.2 Why ‘taking historians seriously’ undermines macro-quantitative political science research

Macro-quantitative political science research has been subjected to increased criticism for several decades (e.g. Lieberson 1985; Achen 1986; Ragin 1987; Tilly 1997; Hall 2003; Brady 2004; Ebbinghaus 2005; Kittel & Winner 2005; Shalev 2007; Przeworski 2008). Many of these points are well taken, but as Scruggs (2007: 310) writes in his response to Shalev (2007), many of them are also addressed in basic econometrics textbooks. In many cases then, the problem is rather one of correct application of data analysis techniques rather than the technique itself. We agree with this diagnosis, but we also note that econometrics textbooks rarely use macro-comparative political science examples, and in particular no examples typical of the kind of research this article is concerned with.10

In this section, we argue that ‘taking historians seriously’ undermines macro-quantitative political science research. The reason for this is not to be found in econometrics textbooks, but rather in our assumptions about how the world works. Macro-quantitative political science research makes simplifying assumptions that are most certainly wrong. This could be dismissed, following Friedman (1953), by arguing that theories can make correct predictions even if assumptions are wrong. However, theories in macro-comparative research are rarely tested. This is likely to be true in our working example (electoral system choice), but also quite likely to apply to this kind of research more generally.

How do macro-phenomena like industrialization or socialist threats lead to a political reaction? These questions are typically answered using Coleman’s (1990) ‘bathtub’ (see Figure 8). For instance, using an example from Boix (1999), we could argue that the ‘socialist threat’ made (some) members of the Belgian Liberal Party worry about the future of their party (macro-micro). These members then reacted to this threat by demanding a change of strategy (micro-micro). Finally, the new preferences of party members were aggregated through a complex intra-party decision-making process into the party’s position on electoral system change (micro-macro).

10 This is also true for textbooks specialized in political science (cf. Morgan and Winship 2007).
In macro-quantitative political science research, these micro-foundations are hardly ever ‘tested’. None of the papers discussed in section 2 does so. Rather, in macro-quantitative political science research, we typically compare the variable claimed to influence individual behavior (e.g. the level of the socialist threat) and the collective decision (e.g. the party’s position on electoral system change).11 As a consequence, in macro-quantitative political science research we typically compare macro-phenomena, while we make assumptions about the micro-foundations that produce the relationship (macro-macro) we observe.12

Each of the three transitions of Coleman’s (1990) ‘bathtub’ can be critically discussed, but we focus our attention on the last micro-macro transition because it is here that most problems for macro-quantitative political science research arise, and it is here that macro-quantitative political science research becomes what Cartwright (2002: 142) and Kittel (2006: 662) have called a “crazy methodology”. It is very well known from social choice theory that in presence of strategic behavior of multiple groups and several issues on the agenda (which was almost always the case, see previous section), the outcomes of collective decision-making processes are virtually impossible to predict (Arrow 1963; Nurmi 1999; Mueller 2003). This is because “most of the general situations either lack an analytical equilibrium solution, or face an infinite variety of them” (Kittel 2006: 660). Only simple alterations of the decision-making process suffice to change the outcome, and it is quite possible that a rather small minority prevails over the majority in a collective decision making process (Nurmi 1999: 70-73). However, the relationship linking micro behavior and macro behavior is rarely discussed.

The famous Ostrogorski paradox could for instance explain why the Belgian Catholic party used its absolute majority in parliament to introduce PR in 1899. Belgium was the first country to introduce PR on the national level and inspired many other countries to do likewise (Ahmed 2010). 65 Catholics and 5 Liberals supported the 1899 law introducing PR, 35 Catholics, 7 Liberals, and 21 Socialists opposed it (Mahaim 1900: 397). This result is

---

11 ‘Testing’ the micro-foundations in macro-quantitative political science research would be virtually impossible given time and space restrictions as well as data availability.
12 Historians are rather critical of these macro-macro comparisons. As argued by Roberts (1996: 16), “historians do not explain the occurrence of complex historical events by subsuming them under covering laws.” Rather, “they explain the occurrence by tracing the sequence of events that brought them about”. Of course, this methodological stance makes it impossible for historians to compare more than a handful of nation-states in the framework of one publication.
remarkable for several reasons, among them the fact that it was mainly the Liberals who were facing the ‘socialist threat’ and not the Catholics, or the fact that the opposing parties benefitted from the reform, while the Catholics went on to suffer heavy losses in terms of parliamentary representation. Boix (2010: 411) ignores these anomalies and considers Belgium a case confirming his theory because it was mostly the urban Catholics (facing some socialist threat) who supported PR, while the Catholic party’s rural wing (facing no socialist threat) opposed PR.\textsuperscript{13} What Boix (2010) does not mention is the fact that these vulnerable urban Catholics were a clear minority in the party (Ahmed 2010: 1082). The Ostrogorski paradox can explain why the minority could prevail over the majority in presence of multiple on the political agenda (see section 4.1), but the example also shows how dangerous it is to make simplifying assumptions about intra-party collective decision-making. If Boix (1999, 2010) were really correct, the Catholic party, led by its rural majority, would have let the Liberal party perish instead of deliberately saving it (Ahmed 2010: 1079).\textsuperscript{14} Collective decision-making problems make the prediction of aggregate behavior (such as political reforms or party positions) extremely difficult. However, aggregate behavior is not always subject to similar collective decision-making problems (cf. Kittel 2006). Economists, for instance, are mostly interested individuals’ reaction to incentives. Based on assumptions about a representative agent (homo economicus), they expect individuals to react independently of each other in a typical (representative) way. As a consequence, economists face fewer problems when analyzing aggregate economic behavior. Quite the contrary, economists benefit from ‘the wisdom of the masses’ or, to put it differently, the ‘law of large numbers’. The situation is completely different in politics though. Here we typically assume strategic behavior and conflict. Political organizations (such as governments, parties, interest groups) are typically divided with every camp trying to prevail by the strategic use of power and institutions. In such a situation, the outcomes of collective decision-making processes are no longer predictable (except in very unique circumstances). Assuming a representative agent

\textsuperscript{13} We are focusing on Boix (1999, 2010) here because Boix is the one only making clear statements about the actors enforcing these changes.

\textsuperscript{14} The reality was in fact much more complicated including interventions by King Leopold, violent street protests that led the Catholics to withdraw their first proposal (which was more beneficial for them), the role of the previous electoral system, the plan of the Catholic party to use the Liberals as a buffer against the Socialists, and the role of the geographical distribution of votes, which gave the Catholics an incentive to support PR. We thank Markus Kreuzer for informing us about the complexities of the Belgian case. The simplified story, however, suffices to illustrate our argument.
despite far-ranging collective decision-making problems makes for a very implausible assumption (Kittel 2006: 661).\textsuperscript{15}

Assumptions, however, do not have to be correct. Quite the contrary, assumptions are typically wrong because the purpose of theory is to explain much by little. As argued by Friedman (1953: 15), “the relevant question to ask about the ‘assumptions’ of a theory is not whether they are descriptively ‘realistic,’ for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions”. We agree with Friedman (1953) that assumptions need to simplify.\textsuperscript{16} From that point of view, assuming a representative agent despite collective decision-making problems is not per se incorrect, in particular if the model’s predictions cannot be falsified using empirical data. But what if we cannot test the predictions?

Testing predictions is of fundamental importance in the neopositivist methodology (methodology here, following Sartori (2009: 13), understood as the “concern with the logical structure and procedure of scientific inquiry”). As famously argued by Popper (1979), we can only make progress in the realm of knowledge by accepting that there is no “absolute certainty in the whole field of our knowledge” (Popper 1979: 77) and that “\textit{all theories are hypotheses; all may be overthrown}” (Popper 1979: 29, emphasis in the original). Put differently, falsificationism is the neopositivist methodology’s main strategy for dealing with what is sometimes referred to as the ‘Cartesian problem’: “how does a world-independent mind gain reliable knowledge of a mind-independent world?” (Jackson 2011: 46). The neopositivist strategy is to test hypotheses against the mind-independent world (Jackson 2011: 71).

However, for two main reasons, macro-quantitative political science research struggles to test hypotheses. Firstly, studies relying on observational data always face the ‘fundamental problem of causal inference’ because causal statements are ultimately based on the comparison of something that did occur (the ‘factual’) and something that did not occur (the

\textsuperscript{15} Kittel (2006) discusses these differences in detail. He identifies two crucial factors that separate aggregate market behavior from decision-making by collective political actors. Firstly, textbook market consumers react to price changes independent of the reaction of other consumers (independence). In contrast, decision-making by political parties is characterized by strategic bargaining between different political factions. Secondly, textbook market consumers are expected to react in an identical way to price changes (identity). The ‘law of large numbers’ then identifies the representative agent. In contrast, decision-making by political parties is characterized by multiple factions with different interests and constituencies. For instance, representatives of rural interests are unlikely to react to political issues in the same way as representatives of urban interests within the same party. Consequently, we are unable to identify a representative agent.

\textsuperscript{16} Whether assumptions have to be plausible is a contested issue in the literature.
“counterfactual”). Since we cannot per definition observe the counterfactual – the outcome if the hypothesized cause would not have occurred – we never know a causal inference for certain (Lieberson 1985; King et al. 1994; Brady 2004; Przeworski 2008). Thus, in non-experimental settings, falsifications are never really definitive. This is also the case in our working example. Both Boix (2010) and Cusack et al. (2010) reject the alleged falsifications of their theories by Cusack et al. (2007) and Kreuzer (2010) respectively.\textsuperscript{17}

The inability of statistical tests to irrevocably falsify hypotheses using observational data is widely acknowledged. However, statistics in macro-comparative political science research can still be helpful in identifying “unobvious regularities” (Hoover 2002: 173). If a theory predicts unobvious regularities and we generally find these regularities in the data, the theory might in fact have high explanatory power. It is here where the second reason why we rarely test hypotheses in macro-quantitative political science research enters the scene: the observed regularities are rarely unobvious. Quite often, and most certainly in our working example, the regularities are in fact quite obvious. Take, for instance, Cusack et al. (2007): They ‘test’ their hypothesis by regressing the effective electoral threshold on an indicator of ‘preindustrial coordination’ and control variables. Preindustrial correlation is virtually perfectly correlated with the country classification used in the varieties of capitalism literature (Estevez-Abe et al. 2001; Hall and Soskice 2001; Iversen 2005). This is not surprising given that the varieties of capitalism literature generally claims that preindustrial coordination caused the different varieties of capitalism to emerge (Iversen 2005; Martin and Swank 2008; Iversen and Soskice 2009). In this context, it is interesting to read that Cusack et al. (2007: 373) are aware of the fact that “as Gourevitch (2003) and Gourevitch and Shinn (2005) have pointed out, current varieties of capitalism are almost perfectly correlated with electoral systems”. As a result, we can no longer speak of an ‘unobvious regularity’. The correlation was known from the beginning.\textsuperscript{18}

We should not call these quantitative analyses ‘tests of predictions’ because researchers are typically quite familiar with the cases they are analyzing.\textsuperscript{19} Researchers use the very same

\textsuperscript{17} For students of philosophy of science, this is of course not surprising (see Jackson 2011: 52-59).

\textsuperscript{18} The same is true for Boix (1999, 2010) and Blais et al. (2005). Boix (2010) explains the adoption of PR ex post (see discussion of Belgium and Switzerland above). The two main independent variables used by Blais et al. (2005) are essentially dummy variables for Europe (spread of democracy) and Anglo-Saxon countries (plurality system) because only European countries could score a high value on the spread of democracy variable (number of democracies in your world region), while all Anglo-Saxon countries, but only few other countries had a plurality voting system at that time. Boix (1999) comes closest to a real test of theoretical predictions, but he focuses on one observable implication only (see Kreuzer 2010). In addition, Cusack et al. (2007) have rejected Boix’s (1999) empirical analysis.

\textsuperscript{19} This familiarity with the cases in macro-comparative political science research is one of the main reasons why Ragin (1987) suggests abandoning variable-oriented research in favor of case-oriented research.
cases they used to develop the theory to test the theory. Thereby they ignore standard advice given in methodology textbooks (cf. King et al. 1994; Geddes 2003) that we are to test theories using new data or at least new observable implications (assuming that there are new observable implications the authors are not yet familiar with). Rather they seem to rely on what Skocpol and Somers (1980: 176) coined the ‘parallel demonstration of theory’, which does not test theories, but repeatedly demonstrates the usefulness of a theory by displaying its ability to convincingly order evidence.  

In sum, macro-quantitative political science research relies on assumptions about collective decision-making that are extremely implausible. Theoretical arguments based on these assumptions are rarely tested because of statistical problems in presence of observational data and the fact that often the same data is used to both develop and ‘test’ these theories. In such a situation, macro-quantitative political science research is essentially providing a summary of some data, which is, as Cartwright (2002: 142) candidly notes, the original meaning of the word ‘statistics’.

5 Possible research strategies for the macro-comparative analysis of electoral system choice

The debate on electoral system choice is a perfect example to illustrate our point about macro-quantitative political science research. Firstly, the number of cases covered and the statistical techniques used are quite typical for this kind of research. In addition, the event captured by the dependent variable is generally non-recurrent. Secondly, the authors involved are among the leading researchers in comparative politics. Finally, the theoretical arguments put forward by these authors are in fact very convincing. Boix (1999, 2010) has systematized the classical argument by Rokkan (1970). He is right to focus on strategic behavior of political agents. There is strong evidence that established parties were indeed dealing with a ‘socialist threat’. In a similar vein, Blais et al. (2005) make an important contribution by highlighting the pressure for democratizing the political system and by providing a convincing explanation why some countries were characterized by more fragmented party systems already before adopting PR. Finally, Cusack et al. (2007, 2010) develop a completely new, innovative

\[\text{20} \] From this point of view, it would be more advisable to test the discussed theories against data on the level of consensus among the major political actors. In our expert survey, we asked the participants whether most political actors agreed on adopting/not adopting proportional representation at the climax of the debate about adoption of PR in the period 1890-1939. The results are displayed in Figure A2 in the Appendix. Figure A2 shows that there was widespread consensus in countries such as Finland, Ireland and Norway (part of independence process), but also the Netherlands, Switzerland, and the USA. On the other hand, there was widespread conflict in Canada, Greece, France, Austria, Italy, and Germany.
argument why some countries had a stronger propensity to adopt PR. Reading the theoretical parts of these articles is extremely informative. In our opinion, they all make important contributions.

It is their empirical parts we are quarreling with. In a nutshell, we argue that in order to make their empirical approach work, the authors have to make a range of implausible simplifying assumptions (causal homogeneity, no collective decision-making problems, no international diffusion, outcome of collective decision-making process is independent of other political issues). However, the theoretical predictions based on these assumptions are not tested because of statistical problems in presence of observational data and the familiarity of authors with the data prior to the empirical test. In such a situation, macro-quantitative political science research is essentially describing data, while the assumed micro-foundations are highly questionable because they rely on unlikely scenarios. In a nutshell, ‘taking historians seriously’ (and many social scientists for that matter) undermines macro-quantitative political science research.

However, there is no reason to abandon the macro-comparative enterprise. We argue that we should engage in macro-comparative studies that are more historically sensitive (Capoccia and Ziblatt 2010). The goal of research cannot be to explain just a little bit more than alternative explanations. Rather, the goal must be to get it right. This should not be misunderstood as a historicist argument that context is everything or as a plea for excessive particularism. Being contextually sensitive does not mean that everything can be argued (Berlin 1960; Evans 1997). In a similar vein, rejecting macro-quantitative political science research is not synonymous with advocating postmodernism, as Boix (2010: 405) seems to imply.\footnote{Boix’s (2010: 405) argument that “historical narratives are so rich in content … that one can find almost all kinds of causal stories in them” is essentially a postmodernist rejection of most historical research. Interestingly, though, Boix adopts a more positivist perspective for the remainder of his article.} Context can be addressed in a systematic way. It is all about striking the right balance.

We therefore end with a series of possible research strategies for the analysis of electoral system choice in the period 1890 to 1939. We distinguish between medium-N and small-N research designs. With regard to medium-N research designs, we emphasize that research designs must allow for equifinality because the causal homogeneity is unlikely to hold. Broadly speaking, countries seem to cluster into countries that adopted PR in the process of independence (Finland, Ireland, and to a lesser extent Norway), countries that adopted PR in the process of democratization (Austria, Denmark, Germany, Greece, the Netherlands, Spain,
and Sweden), countries with more or less stable democracies at the time of adoption (Belgium and Switzerland) and countries with defective democracies at the time of adoption (Italy and Japan). In addition, the research design must take account of international diffusion processes.

Following Penadés (2008), we suggest focusing on actors’ preferences prior to the decision rather than their actual choices in the decision-making process. Preferences prior to the decision are more likely to reflect the true preferences of the political actors. However, scholars should not underestimate the extent to which preferences can change over time. For instance, the French electoral system was reformed seven times during the period 1870 to 1940, with French political parties adapting their position on electoral system choice very frequently (Carstairs 1980; Alexander 2004). In addition, researchers should be careful to read history ‘forward’ and not ‘backward’ (Ahmed 2010; Capoccia and Ziblatt 2010), i.e. researchers should avoid a ‘hindsight bias’. As Lebow (2010: 8) notes, social scientists tend to upgrade the probability of events once they have occurred, while they consider the future to be highly contingent. This view ignores that crucial decisions in the past have typically been made in situations of high uncertainty (Capoccia and Ziblatt 2010). “Once we know what has happened, it is difficult to recall how unsure we used to be about the future” (Tetlock et al. 2006: 3). Lebow (2010: 10) adds, “if major historical developments are so inevitable, … the underlying conditions responsible for these events should have been apparent at the time to scholars and policymakers alike, making them … to some degree predictable” (see also Evans 1997).

Alternatively, researchers might want to consider changing the level of analysis. Federal political systems such as Germany, Switzerland, and the United States of America allow for the analysis of PR adoption at the substate level (Wuarin 1895; Hoag & Hallett 1926; Barber 1995; Amy 1996). The Swiss case is particularly interesting. Not only was the Swiss canton of Ticino the first political unit to introduce PR in 1891 (followed by Geneva and Neuchatel in 1893 and Zug in 1894), the (then) 25 Swiss cantons (19 full cantons and 6 half cantons) are also characterized by a lot of variation. Some Swiss still use MR for the election of cantonal parliaments today. In addition, there is also considerable variation with regard to the cantonal rules for the electoral of cantonal governments and the representatives in the two federal parliamentary chambers. Alternatively, researchers can (quantitatively) analyze the voting behavior of citizens in the vote on the constitutional amendment that introduced PR in Switzerland in 1918, or analyze the campaigns of the different political actors (parties, interest groups, government) in the weeks prior to the vote. Changing the level of analysis has
the advantage of keeping numerous factors constant. However, it is questionable to what extent the results are translatable to other levels of nationals.

The final alternative is historical case studies, as advocated by Capoccia and Ziblatt (2010) and exemplified by Ahmed (2010) in her study of electoral system choice in Belgium and the United Kingdom. Historical case studies are probably best capable of dealing with issues such as causal heterogeneity, Galton’s problem, multiple issues on the political agenda, and strategic behavior on the part of the political actors. In contrast, there are question marks with regard to external validity and the ability of case studies to test theories.

Ultimately, then, researchers interested in electoral system choice can choose from a large menu of options. However, we advise against choosing a macro-quantitative research design to study electoral system choice.
Table 1: Recent contributions on electoral system choice

<table>
<thead>
<tr>
<th>Safe or challenge old system?</th>
<th>Consensus or conflict among major political actors?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe</td>
<td>Consensus: Cusack et al. (2007, 2010)</td>
</tr>
</tbody>
</table>

Table 2: Survey results on generalization, comparison, and quantitative approaches (percent)

<table>
<thead>
<tr>
<th>Does it make sense to look for explanations that can be generalized to more than one country?</th>
<th>How much can we learn about our country of expertise by comparing it with other countries?</th>
<th>Can a quantitative approach contribute to the explanation of historical developments?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always</td>
<td>Learn a lot</td>
<td>Always</td>
</tr>
<tr>
<td>Almost always</td>
<td>Quite a lot</td>
<td>Almost always</td>
</tr>
<tr>
<td>Sometimes</td>
<td>Occasionally</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Almost never</td>
<td>Rarely</td>
<td>Almost never</td>
</tr>
<tr>
<td>Never</td>
<td>Never</td>
<td>Never</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
</tr>
</tbody>
</table>

Note: The complete questions are: (1) “Some social scientists advocate a quantitative approach to historical developments. These scholars argue that researchers should not focus on countries but rather on ‘variables’. In particular, they argue that social scientists should develop indicators measuring the ‘spread of democracy’ in the neighboring countries (Blais et al.), the level of a ‘socialist threat’ to established parties (Boix) or the level of ‘coordination between different group interests’ (Cusack et al.) and test whether these variables are systematically associated with the outcome of interest (here electoral system choice). Can this quantitative approach contribute to the explanation of historical developments?” (2) “Does it make sense to look for explanations that can be generalized to more than one country (e.g. in the analysis of historical events or developments)?” (3) “How much can we learn about our country of expertise by comparing it with other countries and the development of our outcome of interest (e.g. electoral system change) in these countries?”
Table 3: Survey results on historical facts and medium-N approaches (percent)

<table>
<thead>
<tr>
<th></th>
<th>Are we able to give clear answers to questions about historical facts (such as consensus or conflict)?</th>
<th>Are (variable-oriented) medium-N approaches are reasonable way to study historical developments?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always</td>
<td>4.48</td>
<td>6.06</td>
</tr>
<tr>
<td>Almost always</td>
<td>31.34</td>
<td>12.12</td>
</tr>
<tr>
<td>Sometimes</td>
<td>58.21</td>
<td>12.12</td>
</tr>
<tr>
<td>Almost never</td>
<td>4.48</td>
<td>46.97</td>
</tr>
<tr>
<td>Never</td>
<td>1.49</td>
<td>22.73</td>
</tr>
<tr>
<td>N</td>
<td>67</td>
<td>66</td>
</tr>
</tbody>
</table>

Note: The complete questions are: (1) “Some of the social scientists (political scientists and sociologists) contributing to the debate on electoral system change in the period 1890-1939 disagree on historical facts. Some argue that there was conflict between parties on the issue of electoral system choice in a given country, some argue that there was none. Are we in general able to give clear answers to such questions for a given country?” (2) “The social scientists contributing to this debate work with approximately 20 countries in order to develop a general argument about electoral system choice. They argue that if their argument is correct for most of these 20 countries, then it should be considered a general explanation for all 20 countries. This means that even for countries for which the explanation is not correct, the argument should be considered an important part of the specific national explanation. Do you think that this is a reasonable way of studying historical developments?”

Table 4: Survey results on general perception of social scientists: Ranked from 1 (disagree) to 10 (agree)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social scientists’ historical knowledge is generally very good</td>
<td>112</td>
<td>4.16</td>
<td>1.84</td>
</tr>
<tr>
<td>In general, social science research pays sufficient attention to the historical dimension of political and social developments</td>
<td>113</td>
<td>3.79</td>
<td>1.66</td>
</tr>
<tr>
<td>Social scientists are generally more interested in defending their theories than in analyzing actual historical developments</td>
<td>111</td>
<td>6.72</td>
<td>2.21</td>
</tr>
<tr>
<td>Social scientists often draw facts from historically grounded accounts and then use them for timeless propositions</td>
<td>105</td>
<td>6.74</td>
<td>1.96</td>
</tr>
</tbody>
</table>
Figures 1 to 5: Elaboration models of the causal arguments

Figure 1: Boix (1999) & Socialist Threat

- Strong political left (electorally)
- Fragmented political right
- Extension of suffrage (to lower social classes)
- Ruling parties initiate move to PR
- Enactment of PR

Figure 2: Boix (2010) & Socialist Threat

- Competitive electoral systems
- Expansion of suffrage (to lower social classes)
- Newer political party (electorally)
- Some established parties lose socialist voters
- Vulnerable parties support PR
- Pro-PR parties have insufficient political power
- Enactment of PR

Figure 3: Alesina and Glaeser (2004) & Left Power

- Strong political left
- Political right decides to introduce PR when in power
- Political left decides to introduce PR when in power
- Enactment of PR

Figure 4: Blais et al. (2005) & Democratization

- Spread of democratic ideas
- Consensus to switch to PR
- Enactment of PR
- Majority run-off electoral system
- High number of parties

Figure 5: Cusack et al. (2007, 2010) & Co-specific Assets

- Industrialization, urbanization, nationalization, etc.
- Regulatory policies now at the national level
- Co-specific assets
- Consensus to switch to PR
- Enactment of PR

Note: Red arrows indicate the empirically tested relationship.
**Figure 6:** Evaluation of the three explanations: Best performing explanation

**Note:** The Figure shows the average score of the best performing explanation. The best performer, given there is a single-best performer, is listed in the x-axis labels following the country name together with the number of observations.
Figure 7: Level of democracy and transition to PR: Italy, the Netherlands, and Switzerland

Note: The level of democracy is measured as the difference between the 10-point democracy index and the 10-point autocracy index in the Polity IV dataset using data from the Polity IV dataset, version 2009.
**Figure 8: Coleman’s ‘bathtub’**

Macro level: Socialist threat  →  Support for PR

Micro level: Members worry about future of party  →  Members demand change of strategy

**Source:** Coleman (1990).
Appendix

**Figure A1:** Evaluation of the three explanations: Average performance

Note: The Figure shows the average score of all three explanations. The number of observations is listed in the x-axis labels following the country name.

**Figure A2:** Level of conflict at the climax of the debate on the adoption of PR (1 = widespread consensus, 3 = widespread conflict)

Note: The Figure shows the level of conflict among the major political parties at the climax of the debate on the adoption of PR in the period 1890-1939 (1 = widespread consensus, 3 = widespread conflict). The number of observations is listed in the x-axis labels following the country name.
Presentation of the expert survey

The expert survey was organized in two steps. Firstly, we contacted the editors-in-chief of a total of 130 leading national, regional, and international journals within the following sub-disciplines of history: political history, contemporary history, economic (and business) history, parliamentary history, social science history, social history, labor history, and general history (only leading journals). The editors-in-chief were contacted by a standard email presenting the project and asking the editors-in-chief to appoint five or more experts to participate in the survey. The email was adapted to the specific country or region (except for the international journals). A remainder was sent out four weeks later.

We received positive answers from 42 journals covering all countries except Belgium, Ireland, and New Zealand. In order to improve country coverage as well as increase the total number of experts, we directly contacted 29 researchers and asked them to help us locate the most relevant experts (which could include themselves). We received 23 positive answers (by Klas Åmark, University of Stockholm; Neil Atkinson, Ministry of Culture & Heritage New Zealand; Rita Baeten, OSE Brussels; Christoph Conrad, University of Geneva; Robert Engelbert, University of Saskatchewan; Mark Francis, University of Canterbury; Effi Gazi, University of Peloponnes; Brian Girvin, University of Glasgow; Tore Gronlie, University of Bergen; Ana Guillen, University of Oviedo; Per Haave, University of Oslo; Kees van Kersbergen, University of Aarhus; Marc Lazar, Sciences-po, Paris; Henry Milner, Université Montréal; Makoto Murai, Tokyo University; Herbert Obinger, University of Bremen; Michel Offerlé, ENS Paris; Bruno Palier, Sciences-po, Paris; Gerhard A. Ritter, Ludwig-Maximilians-Universität München; Dimitri Sotiropoulos, University of Athens; Pat Thane, Institute of Historical Research London; Mariano Torcal, Universitat Pompeu Fabra Barcelona; Bart Vanhercke, OSE Brussels; Noel Whiteside, University of Warwick). Together with the suggestions made by the journal editors-in-chief, we were able to compile a list of 300 experts.

In a second step, we sent these 300 experts per email an invitation to participate in a web-based survey (created with the help of surveyXact). In addition, we sent two remainders. The first invitation was sent on January 21, 2011. The deadline for participation was March 20, 2011. Both the invitation and the web-based survey were prepared in three languages (English, French, and German). The participants were assured that we would treat their answers anonymously. The survey consisted of four blocks, containing in total 21 questions. The first block covered biographical information on scientific background, including questions on years in research, countries and periods of expertise, disciplinary training, and experience with cross-disciplinary research (history/social sciences only). The second block consisted of questions addressing general perceptions of social science research (see Table 4). The third block was introduced by a short description of the debate on electoral system choice. Subsequently, the participants were shown summaries of the three arguments (see below). Finally, the participants were asked the following five questions for each of their countries of expertise (i.e. this sequences of questions was asked more than once if the participant selected more than one country of expertise):

Q12 Does the argument by Blais et al. (2005) regarding the spread of democracy and the role of previous electoral institutions explain the presence or absence of electoral system change in [selected country]? (only one answer possible)
Answer options ranged from 1 (does not explain at all the presence or absence of electoral system change) to 10 (fully explains the presence or absence of electoral system change) with the additional option of ‘I don’t know’.

Q13 Does the argument by Boix (1999, 2010) regarding the ‘socialist threat’ to established parties explain the presence or absence of electoral system change in [selected country]? (only one answer possible)
Answer options ranged from 1 (does not explain at all the presence or absence of electoral system change) to 10 (fully explains the presence or absence of electoral system change) with the additional option of ‘I don’t know’.

Q14 Does the argument by Cusack et al. (2007, 2010) regarding the level of coordination between different group interests (see below for a summary) explain the presence or absence of electoral system change in [selected country]? (only one answer possible)
Answer options ranged from 1 (does not explain at all the presence or absence of electoral system change) to 10 (fully explains the presence or absence of electoral system change) with the additional option of ‘I don’t know’.

Q15 At the climax of the debate about the adoption of proportional representation in the period 1890 to 1939, did most major political actors agree on adopting/not adopting proportional representation in [selected country]? (only one answer possible)
Answer options: (1) widespread consensus; (2) some major political actors disagreed with the majority opinion; (3) widespread conflict among the political actors; (4) I don’t know.
Q16 At the climax of the debate about the adoption of proportional representation in the period 1890 to 1939, did the political left generally support the adoption of proportional representation in [selected country]? (only one answer possible)
Answer options: (1) the political left supported the adoption of PR; (2) the political left mostly supported the adoption of PR; (3) the political left neither supported nor opposed the adoption of PR; (4) the political left mostly opposed the adoption of PR; (5) the political left opposed the adoption of PR; (6) I don’t know.
In each case, the participants had the opportunity to complement their answers with written comments.

The final block of the survey consisted of five questions on how to conduct research on historical developments. As in the previous block, the participants had the opportunity to complement their answers with written comments. The five questions are as follows:

Q17 Some of the social scientists (political scientists and sociologists) contributing to the debate on electoral system change in the period 1890-1939 disagree on historical facts. Some argue that there was conflict between parties on the issue of electoral system choice in a given country, some argue that there was none. Are we in general able to give clear answers to such questions for a given country? (only one answer possible)
Answer options: (1) always; (2) almost always; (3) sometimes yes, sometimes not; (4) almost never; (5) never.

Q18 The social scientists contributing to this debate work with approximately 20 countries in order to develop a general argument about electoral system choice. They argue that if their argument is correct for most of these 20 countries, then it should be considered a general explanation for all 20 countries. This means that even for countries for which the explanation is not correct, the argument should be considered an important part of the specific national explanation. Do you think that this is a reasonable way of studying historical developments? (only one answer possible)
Answer options: (1) always; (2) almost always; (3) sometimes yes, sometimes not; (4) almost never; (5) never.

Q19 Some social scientists advocate a quantitative approach to historical developments. These scholars argue that researchers should not focus on countries but rather on ‘variables’. In particular, they argue that social scientists should develop indicators measuring the ‘spread of democracy’ in the neighbouring countries (Blais et al.), the level of a ‘socialist threat’ to the established parties (Boix) or the level of ‘coordination between different group interests’ (Cusack et al.) and test whether these variables are systematically associated with the outcome of interest (here electoral system choice). Can this quantitative approach contribute to the explanation of historical developments? (only one answer possible)
Answer options: (1) always; (2) almost always; (3) sometimes yes, sometimes not; (4) almost never; (5) never.

Q20 Does it make sense to look for explanations that can be generalized to more than one country (e.g. in the analysis of historical events or developments)? (only one answer possible)
Answer options: (1) always; (2) almost always; (3) sometimes yes, sometimes not; (4) almost never; (5) never.

Q21 How much can we learn about our country of expertise by comparing it with other countries and the development of our outcome of interest (e.g. electoral system change) in these countries? (only one answer possible)
Answer options: (1) we can learn a lot; (2) we can learn quite a lot; (3) we can occasionally learn something; (4) we can rarely learn anything; (5) we can never learn anything.

We were unable to find a contact address for 20 of the 300 nominated experts, giving us a total of 280 experts with correct contact details. 146 experts did not accept our invitation to participate in the survey, giving us a response rate of 47.9 percent. The remaining 134 experts participated in the survey, but not all of them finalized the survey. We have a complete set of answers from 66 participants (23.6 percent). Due to multiple countries of expertise, we have 88 sets of answers in block 3 of our survey (evaluation of explanations). See Figure 6 for the distribution of answers to countries. As regards incomplete sets of answers, most participants stop at the beginning of block 3 of the survey (evaluation of explanations). Judging from the written feedback that we received from those that did not finalize the survey, the main problem seems to have been that the experts did not feel qualified enough to evaluate the three explanations. In total, we consider the number of complete sets of answers (66) and the number of evaluations of explanations (88) to be high given the very specialized topic and the amount of time being used for answering the complete survey.

Summaries of the three arguments:

Argument 1 (Blais et al. 2005):
Blais et al. highlight two factors facilitating the shift to PR: (1) the spread of democratic ideas and (2) the presence of a majority run-off electoral system (in contrast to a plurality electoral system).

In the period 1865 to 1938, a period of democratization in many countries, the idea that each individual should have one vote and that each vote should count the same gained enormous ground. “From that perspective democracy and PR appeared to dovetail each other. PR came to be regarded as the fairest system” (p. 182) and the most ‘democratic’ electoral system. As a consequence, PR was adopted in many countries without much debate.

Blais et al. then ask why politicians in some countries nonetheless reacted less favourably to, or were better positioned to resist the public demand for PR. They argue that the decision to adopt PR was dependent on the existing electoral institutions. Most importantly, they maintain that “politicians’ reactions and positions were contingent on whether the choice was between a plurality system and PR or a majority system and PR” (p. 184).

Compared to plurality systems, majority systems lead to a higher number of parties. The reason for this is strategic voting. In plurality systems, all voters converge on the two biggest parties because the party that gets the most votes (the plurality of votes) wins. In contrast, in majority systems, which involve two or more rounds of voting, a majority (not just a plurality) of the votes is needed. As a consequence, voters have a weaker incentive to vote for one of the two biggest parties. In majority systems, party systems are more fragmented and this often leads to coalition governments.

The adoption of PR is more likely in majority systems because fragmented party systems, the regular occurrence of coalition governments, and the greater uncertainty as to the optimal strategies for winning elections, due to the greater number of parties. “As a consequence, governing parties in these countries offered little resistance to the widespread push for PR that was taking place at the time” (p.190).

“Although majority systems allow legislative representation for more parties than plurality systems, it is nevertheless the case that majority systems lead to a high degree of disproportionality. The minor, electorally disadvantaged, coalition parties are therefore likely to favour the adoption of PR and to make a strong case for it on grounds of fairness.” (p. 185).

**Argument 2 (Boix 1999, 2010):**

Until the turn of the twentieth century, parliamentary seats were using plurality or majoritarian rules everywhere. Given the strategic or coordination properties of those rules (resulting in high entry barriers to new parties), the existing parties (mainly, conservative, liberal or Christian democratic) had strong incentives to maintain them.

However, the gradual decision of trade unions to shift their support from liberal or progressive parties to strictly social democratic parties pushed electoral reform and the adoption of proportional representation (PR) to the forefront of the political agenda of non-socialist parties (and many socialist parties as well) in all countries (except those where socialist parties were small or did not exist).

Non-socialist parties favored PR if they felt that the emergence of a social democratic party threatened their electoral viability in the future. That threat was as a function of: (1) the type of electoral market (segmented versus competitive) in which the old parties competed with each other; and (2) the extent to which the old parties shared voters with the new socialist party.

In segmented electoral arenas (those where the support of a particular party is highly concentrated in a particular geographic area or social sector), the old parties favored PR if the new entrant (generally a social democratic party) threatened their electoral hegemony: mainly urban parties. Otherwise they did not support PR.

In more competitive electoral arenas (those where several parties contend for the vote of at least some fraction of the electorate), the position of the established parties was shaped by the extent to which they were dominant in the electoral arena vis-à-vis the other old parties conditional on the entry of third parties. The party that expected to become the focal point around which non-socialist voters would eventually coordinate had little incentive to support PR. By contrast, those old parties that could not expect to become the dominant non-socialist party preferred PR.

**Argument 3 (Cusack et al. 2007, 2010):**
“Countries that chose proportional representation electoral systems in the early twentieth century were those [...] that had historically had relatively negotiated forms of political decision making. [...] In these societies, [...] the nineteenth-century state and policy-making systems emerged out of Ständestaat traditions. They comprised densely institutionalized local and regional economies. Within these subnational communities [...] local decision making involved consensus-based negotiation and bargaining so that different group interests (except those without possessions) could be effectively represented. This allowed the solution of collective action problems, as well as the safe creation of cospecific assets with local and regional economic networks” (2010, p. 395).

“In these countries, a majoritarian electoral system worked adequately at the national level through much of the nineteenth century. Constituencies were represented in national politics by local notables elected by plurality and often unopposed. With economic interests generally geographically defined, these provided for their more or less proportional representation” (2010, p. 395).

“By the end of the nineteenth century, however, industrialization, urbanization, and the growth of the working class had made this majoritarian system of national representation increasingly disproportional [...] At the same time, economic networks and regulatory legislation were becoming increasingly national to reflect the accelerating growth of industrialization” (2010, p. 406).

This presented major problems for representative political parties “closely connected with organized interest groups representing important economic sectors of society” (2010, p. 397). This included confessional parties, which served as “negotiating communities” for many different economic groups. Consequently, PR was adopted in order to restore the proportional representation of all important economic interests. “Interest-carrying parties needed to preserve their identity to be able to continue to represent their interests at the national level. The transition to PR was a means to restore a negotiation-based political system in which different economic interests were effectively represented by parties. To do this, there was no obvious alternative to PR” (2010, p. 395). For the very same reasons, PR “was supported across the party spectrum” (2010, p. 395).

It is important to add that this argument pertains to the choice of electoral systems once the transition to democracy had been completed. Before then, calls for PR were often tied to calls for democracy, which were of course resisted by the political right. “[O]ur argument is about the relationship between party organization and electoral competition once reasonably democratic conditions had been met in the early 1920s” (2010, p. 397).
References


145-183.


Skocpol, Theda and Somers, Margaret (1980). 'The Uses of Comparative History in Macrosocial Inquiry', *Comparative Studies in Society and History* 22(2): 174-197.


