

What can Quantitative and Formal Models Teach us About Transitional Justice

Milena Ang
The University of Chicago

Monika Nalepa
The University of Chicago

The authors are grateful to Douglas Lawrence, Jens Meierhenrich and Alex Hinton for feedback on this chapter and to the participants of the Workshop on Transitional Justice, Amherst MA. All mistakes are the authors' responsibility.

1 Introduction

The study of Transitional Justice has been undertaken by a variety of disciplines, and political science is no exception. Transitional Justice is approached by all four traditional subfields of this discipline, with a sizable representation in normative Political Theory,¹ International Relations,² and slightly lower numbers of Comparativists³ and Americanists⁴ studying how new democracies reckon with former autocrats and human rights' abusers. Yet political scientists using advanced quantitative or formal methods have at best token representation among scholars of transitional justice.⁵ Few studies on Transitional Justice leverage formal models and quantitative evidence to analyze Transitional Justice policies, exacerbating the skepticism of how complex and heterogeneous phenomena that make up transitional justice can be formally modeled, and then measured and analyzed with statistical tools.

Why are cutting edge methods designed by political scientists for political science research largely absent from transitional justice research? Part of it might stem from the belief that complex and heterogeneous social and political processes, like the ones that make up transitional justice, are not well suited for numerical abstraction. As a result, scholars are skeptical that transitional justice can be formally modeled, and then measured and analyzed with statistical tools.

This chapter calls this skepticism into question. We argue that research on transitional justice benefits greatly when applying formal and quantitative models. We begin by arguing that formal models are particularly well suited for answering normative questions

¹The work of Colleen Murphy (2017) is but one excellent example. Others include Elster (2004) and Tucker (2015).

²See especially Binningsbo, Loyle, Gates & Elster (2012), Sikkink (2011), and Prorok (2017).

³See especially (Payne 2007, Lupu & Peisakhin 2017, Rozenas, Schutte & Zhukov 2017)

⁴Most of these works are devoted to the topic of reparations for slavery (Elster 2006, Aukerman 2002)

⁵Among the few exceptions are Dancy, Kim & Wiebelhaus-Brahm (2010) who present an analyze an extensive database of truth commissions, (Aguilar, Balcells & Cebolla-Boado 2011) who analyze Spanish survey data in the aftermath of the country's passage of the Memory Law reversing Spain's decades' old transitional justice policy of letting bygones be bygones, and (Binningsbo et al. 2012), who collected and analyzed transitional justice data on 326 "peace periods" following civil wars.

and for understanding causal relationships between empirical phenomena. We finalize our discussion by clarifying that formal models need not, and in fact cannot, be tested, and that their value to knowledge is not thwarted by this. We then advocate for quantitative models, arguing that these approaches can complement existing scholarship on transitional justice in at least two ways: by providing case selection criteria that minimizes bias and strengthen the probative value of empirical evidence, and by quantifying and aggregating evidence from various cases, enabling generalizability of the claims made. These benefits are possible precisely because numbers abstract from the details and complexities of the ‘real world,’ and not in spite of them.

Throughout this chapter, we rely on lustration policies to present our arguments. Lustration is a transitional justice mechanism that screens political elites for secret participation or collaboration with the ancien régime. For example, the Polish lustration law of 1997 required all persons holding or running for public office to declare in advance whether or not they had collaborated with the secret authoritarian police prior to the transition. If a person admitted to collaboration, such admission was put on the ballot, and voters could decide whether to cast their vote on a former collaborator. Otherwise, denials of collaboration were verified by a special division of the Institute for National Remembrance against information assembled in the archives of the former secret political police. Proven collaborators who lied on their declarations were banned from running for office for 10 years.⁶ A host of scholars have made the argument about the “strategic” uses and misuses of transitional justice (Nalepa 2010*a*, Appel & Loyle 2012, Leebaw 2008). Although the arguments we set forth here are applicable to other transitional justice mechanisms as well, we believe that lustration, in particular, stands to gain from a formal and quantitative

⁶Although this is the most cited example of lustration (Kamiński & Nalepa 2014, Nalepa 2010*a*, Nalepa 2012, Letki 2002, Williams, Fowler & Szczerbiak 2005), it is hardly typical, because it allows two types of collaborators to escape direct sanctions: (1) the collaborator who admits he worked as a secret collaborator. (See (Nalepa 2008) for a discussion of whether a positive declaration is, indeed, not a sanction) and (2) the collaborator who failed to own up to his past, but was not uncovered, because evidence against him had been destroyed. A more typical lustration law carries with it an explicit sanction for anyone who is proven to have worked for the secret police as an informer (as in Hungary) or who fails to provide evidence of his or her innocence (as in the Czech Republic).

approach.

This chapter proceeds as follows. The next section provides an overview of the contribution of formal models to transitional justice, using existing models as illustrative examples. In this section, we also offer, as a slightly more detailed illustration, our original model of lustration blackmail. Section 3 is devoted to quantitative models of transitional justice—what is gained and what is lost when using them. This section illustrates the use of quantitative methods with a statistical model complementing the explanation provided with the formal model from section 2. Section 4 talks about the limitations associated with using either the formal or the quantitative approach alone, but then it also problematizes the association between formal and quantitative models and criticizes the approach which has come to be known as the Empirical Implications of Theoretical Models (EITM) research program. Section 5 concludes.

2 What do Formal Models tell us about transitional justice

In their influential book “A Model Discipline,” Clarke & Primo (2012) introduced to political scientists the semantic conception of formal theory originating in the work of Alfred Tarski (1944). This idea explains the relationship between theory to the real world as follows: a theory is best thought of as a collection of models, which through theoretical hypotheses are connected to a system of features of the real world. The theoretical hypotheses specify in which respects and to what degree each model is similar to each feature of the real world system.

We illustrate this conception in figure 1 below. The models are represented with M_1, \dots, M_5 in the left hand side set, whereas the features of the real world are represented by F_1, \dots, F_6 in the set on the right hand side. Theoretical hypotheses (TH_1, \dots, TH_5) connect some formal models with some features of the real world. For instance, M_1 represents feature F_4 , as stipulated by TH_1 . More than one model can represent the same feature of the real world.

Figure 1: The Semantic Conception of Formal Theory

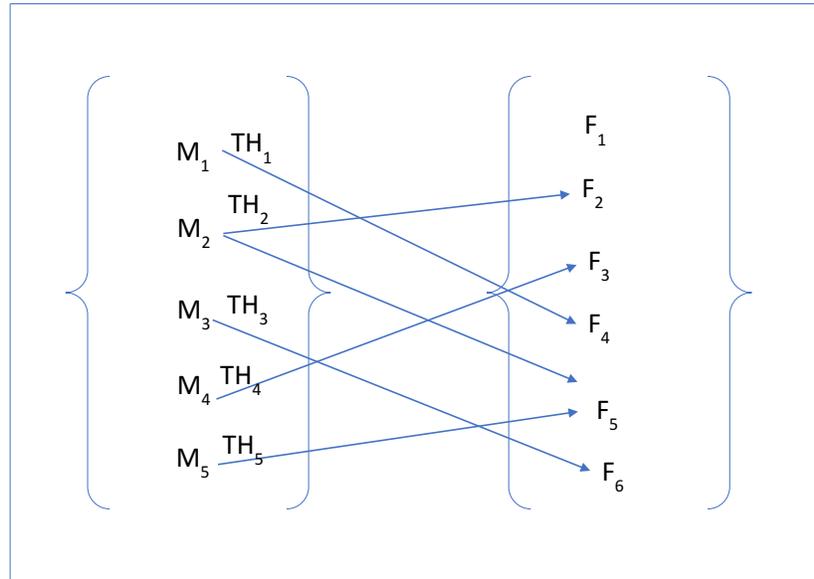


Figure based on Clark and Primo (2012).

For instance, both M_2 and M_5 represent F_5 and there can be features that are not represented at all, such as F_1 . Moreover, some, especially rich models can represent multiple features of the real world, as M_2 represents both F_2 and F_5 .

In this semantic representation, multiple features of the real world remain unrepresented any particular model, a feature that is a key advantage of formal models. Put differently, models abstract and simplify reality, and in doing so they cannot include all features of reality. Models, at a practical level, ought to be thought of as maps of some geographic terrain. Maps cannot be true in the sense of capturing every detail of the terrain; maps record some important features at the expense of others, and provide some valuable information while missing or simplifying features of that terrain. Just as a map, a model can be useful for a particular purpose or situation but not for another, and just as a map it is its usefulness that should be the key to a model's evaluation.

A formal model can be useful for at least three reasons. First, a model can be "foundational," shifting the way scholars think about a certain phenomenon, and paving the

way for the creation of new models extending the original one. Monika Nalepa (2010*b*)'s model of transitions to democracy with and without transitional justice is a good example of such a foundational model. Previously to its publication, the literature expected all anti-authoritarian dissidents to be proponents of transitional justice. The model departs from this notion, and instead allows both former autocrats and the former opposition to have a negative or positive stake in carrying out transitional justice. In doing so, the proposed model allows for the possibility of anti-authoritarian dissidents opposing transitional justice.

Second, "organizational" formal models can be successful if they "collect under one organizational framework seemingly disparate empirical generalizations" (Clarke & Primo 2012). An example of this type of model is Kamiński & Nalepa's (2014) model of strategic preemption. In it, the authors explain the often puzzling implementation of lustration in six formerly communist countries using a workhorse agenda setter model adopted from Romer & Rosenthal (1978), a classical example of a foundational model in its own right. Kamiński & Nalepa's (2014) model allows them to predict not only the implementation, but also amendments to transitional justice legislation making lustration more or less restrictive.⁷

A third kind of model is the "exploratory model," which investigates the mechanism underlying a phenomenon of interest. The model from Ang & Nalepa (2017) is an example of an exploratory model, as it examines the logical soundness of the forward-looking argument for lustration. According to this argument, revealing whether politicians collaborated with the secret police of an authoritarian regime prevents the blackmailing of these politicians: after all, if such information is not made public, agents of the former regime can threaten to divulge this information lest they succumb to the blackmailer's demands. Ang & Nalepa (2017)'s model forces us to reevaluate the forward-looking argument in fa-

⁷An example of a organizational model from American Politics is Krehbiel's (2010) which builds on the model presented by Romer & Rosenthal (1978), but additionally explains the occurrence of gridlock over many decades of US history

vor of lustration: it shows that this alleged effect is highly contingent on modifying factors, such as how easy it is to make empty threats, and on how much the voters are willing to penalize former collaborators-turned-politicians. To illustrate, we present a narrative of this model, and a summary of the main idea behind its solution, and the formal model and solution can be found in the appendix to this chapter.

The model features the interaction between a former agent of the authoritarian regime who may or may not have access to secret police files, and a politician who does not know if evidence of his collaboration with the *ancién* regime survived the transition or not.⁸ The agent can be one of two ‘types,’ depending on whether they have evidence of collaboration or not. In this model, information is distributed asymmetrically, with the agent knowing whether he has the evidence or not (which “type” he is), and the politician not knowing this, and therefore this is an example of a model of incomplete information.

The model assumes that the informed party—the agent—moves first, deciding whether to place a policy demand on the politician, who then has a chance to give in or reject this policy demand. Since the politician might learn something about the agent’s type from his action, this is a model with ‘updating,’ referred to also as a dynamic model of incomplete information. Additionally, the informed party moving first is referred to as the ‘sender of the signal,’ while the uninformed party, moving second, is referred to as the ‘receiver of the signal.’ For this reason, such models are frequently referred to as “signaling models.” These models have specific categories of equilibria, called “separating,” “pooling,” and “semi-separating.” In the first category, each type of the informed party chooses a different action (in this sense, the types “separate”). In the second type of equilibrium, both types “pool” on the same action. Finally, in the third category, of semi-separating equilibria, at least one of the types uses a mixed strategy, meaning, he sometimes pools

⁸Given that secret police apparatuses in many contexts falsified evidence of collaboration, even politicians who did not knowingly collaborate with the the former authoritarian secret police can be uncertain about existing evidence of their collaboration. For reasons why such falsification would be beneficial see (Nalepa 2008).

with the other type and sometimes separates from the other type.⁹

The first observation, we make in our model, is that the forward looking argument is only consistent in the case of the separating equilibrium. There, blackmail with secret information occurs only when evidence survived the transition. Hence the more stringent the lustration law, the less blackmail will take place. Thus, the key test of the logical consistency of the forward looking argument for lustration is whether separating equilibria are the only kind that exist. Unfortunately, we do find, for plausible parameter values the two other equilibria—the pooling equilibrium and the semi-pooling. Under both of these equilibria, blackmail—and the accompanying extraction of policy concessions by the blackmailing agent—takes place even with lustration policies in place. Pooling equilibria occur more frequently when the cost of to politician of having his “skeletons in the closet” exposed is greater. If we were to venture for an empirical interpretation of when that happens, it could be the immediate aftermath of the transition from authoritarianism, when the salience of who was and who was not a collaborator of the ancien regime is at its highest. This could lead us to the implication that lustration implemented shortly following the transition works less well than when it is implemented later after the transition.

The second finding from the analysis of our model is that the chances of a separating equilibrium decrease with the distance between the newly elected politician and the former secret police agent. This means that the more divergent the preferences of the secret police officer and the politician, the less likely the pooling equilibrium. This is most clearly seen in Figure 5 in the Appendix, which graphs (as $a - \sqrt{a^2 - c}$) the cutoff between the separating and other equilibria as a function of a , the parameter measuring the divergence in preferences between the politician and the officer.¹⁰ What this implies is that lustration will work better in less polarized societies. The third and most critical finding of our model is about how the occurrence of blackmail with “skeletons in the closet” re-

⁹We note here that our explanation of signaling models assumes discrete types, although signaling models also allow continuous types.

¹⁰The graph shows this for three different values of the third parameter in the model, c , which represents media scrutiny of false accusations of collaboration.

sponds to increasing lustration severity. A theoretically robust finding is that blackmail decreases in response to lustration severity.¹¹

We conclude this section with a word of caution on formal modeling and parsimony. The model discussed above assumes that unconstrained by “skeletons in their closet” politicians would be implementing voters’ preferences. Obviously, this does not mean that we believe that politicians universally implement what the voters want them to. Politicians may misrepresent voters for a variety of reasons. Yet in our modeling enterprise, we are interested only in the added effect that lustration has on that benchmark. In order to capture that additional and distilled effect that transitional justice may have on democratic representation we start with the baseline of implementing what the voters want and then explore the departure from that baseline that skeletons in the closet cause.

3 What Can Quantitative Models tell us about transitional justice

Similarly to formal models, large-N approaches to empirical research exploit the advantages of abstraction from the ‘real world.’¹² Since quantitative scholars have made many arguments in support of large-N studies,¹³ which are impossible to summarize here, in this section we will limit our discussion to advantages that, we believe, are most likely to impact the study of transitional justice. We focus on (1) the driving principles of case selection, and (2) the aggregation of information. The first allow researchers of complex social phenomena, like transitional justice, to reach adequate causal conclusions based on evidence. The second refers to the capacity of numerical approaches to consider experiences

¹¹By theoretically robust, we mean that regardless of assumptions about the functional form of players’ preferences, this relationship holds.

¹²Large-N studies rely on large numbers of units of observation, and on statistical evidence to prove or illustrate an argument. In contrast, small-N analysis rely on one or a few cases, and the evidence presented is mostly qualitative.

¹³In Political Science, King, Keohane & Verba (1994) famously argued that the logic of large-N studies can, and should, be applied to small-N analysis.

from many countries (or many units of observation). As we will show, this aggregation allows for the creation and testing of generalizable theories. Just as in the previous section, we will illustrate our claims using lustration. We conclude this section by addressing some of the potential imitations of quantitative analysis.

3.1 Epistemologies and case selection

Proponents of quantitative approaches often emphasize the importance of case selection to arrive to appropriate conclusions, meaning conclusions that are both warranted by the observed evidence and that can potentially explain unobserved cases. Case selection, or the criteria for determining which cases will be included in a specific study, can fall short of achieving these two goals, leading to what is often referred to as ‘selection bias.’¹⁴

Take for example, Horne (2017)’s study of the relationship between the nature of transitional justice mechanisms and societal and political trust. After observing that in countries with a large network of covert collaborators, revealing the notoriety with which citizens were spying on one another might cause interpersonal trust to drop, she shows that a wide and compulsory lustration procedure results in real bureaucratic turnover, increasing political trust in governmental institutions. Thus, although lustration causes societal trust to drop, it causes political trust to increase.¹⁵ Case selection is crucial because it enables her to disaggregate the two types of trust in a single country, but it is unclear whether such trust is causally connected to the mechanisms of transitional justice that are being implemented. This also limits the scope of Horne’s argument: since her evidence is based on places where lustration was implemented (countries in Eastern Europe). As a result of this case selection, we do not know, and we cannot know, if similar processes and/or outcomes also occur in places where lustration was not implemented.

¹⁴Selection bias in the statistical sense refers to the non-randomized selection of individuals, which leads to a sample that is likely not representative of the population of interest.

¹⁵Prior to Horne’s research most of the literature assumed that transitional justice is important for trust-building, but nobody disaggregated trust in this way.

A related issue, and one that is much more prevalent in studies of transitional justice, is what is often referred to as 'selection on the dependent variable.' This refers to the criteria of including cases because they present the outcome of interest. This widely adopted practice follows the logic of studying an outcome by observing instances where it happens, and it follows from the logic of agreement set forth by Mill in *A system of logic* (Mill 1884): if everything but one characteristic is different between two units that exhibit the same outcome, then such characteristic must be the cause of said outcome.

One example of this research strategy is the work of Williams et al. (2005), where the authors wonder why lustration was adopted in the Czech Republic, Hungary and Poland, three cases with quite different histories. The authors argue that lustration is the result of a political process, one that results in the adoption of a lustration bill if its defenders are able to convince a plural body of legislators that lustration is necessary for democracy. The convincing ability of defenders, of course, varies from case to case, which results in differences in the promptness with which lustration laws are passed and implemented. Although the authors identify the political and contested process of adoption of lustration laws, something that we have argued above is crucial, all three cases exhibit the outcome of interest, passage of a lustration law. Their selection on the dependent variable precludes an appropriate comparison. In order to illustrate a theory of the causes of lustration the authors should have identified a case where lustration failed to pass.

However, even if they had found such a case, Mill's logic is not flawless, as it necessitates some pretty strong assumptions to be inferentially valid. Most importantly, as Nichols (1986) notes, Mill himself clarified that these methods are inadequate to analyze complex phenomena with multiple causes, and scholars have noted that this approach is limited when dealing with probabilistic events, that is, events that are more or less likely to happen given a certain set of conditions. These assumptions are unlikely to be met when studying social phenomena.¹⁶

¹⁶See Nichols (1986) for a discussion on the use of Mill's methods particularly in the social sciences, and see Lieberson (1991) for a specifically quantitative criticism of the use of Mill's methods.

The importance of allowing multiple causes was partially developed in the section 2 above. Quality of representation post-transition is a function of many different inputs, one of which is implementation and quality of lustration. The point of the model was to show that lustration indeed affects quality of representation, not that it is the only element. Therefore, an approach that allows for multiple causal paths is the best approach: Rather than providing competing explanations for high quality of representation, our theoretical model, and the empirical evidence that supports it demonstrates the effect of lustration *independently and additionally* to other determinants of quality of representation. In short, quantitative modes of reasoning are well equipped to deal with arguments aimed at establishing causality when the phenomenon in question is as complicated as lustration.¹⁷

3.2 Aggregation and generalization

Measurement is complex: the researcher must define the phenomena of interest, list the particular observable or recordable attributes of such a definition, and determine how to translate those attributes into numbers. These decisions are of particular importance, as some features of the real world will no doubt be lost this way. In this section, we exemplify some of the trade-offs that researchers often face when conducting large-N analysis. We acknowledge the relevance of concept definition, particularly when it comes to quantitative approaches, but in this section we focus mostly on the translation from observable attributes to numeric characteristics.

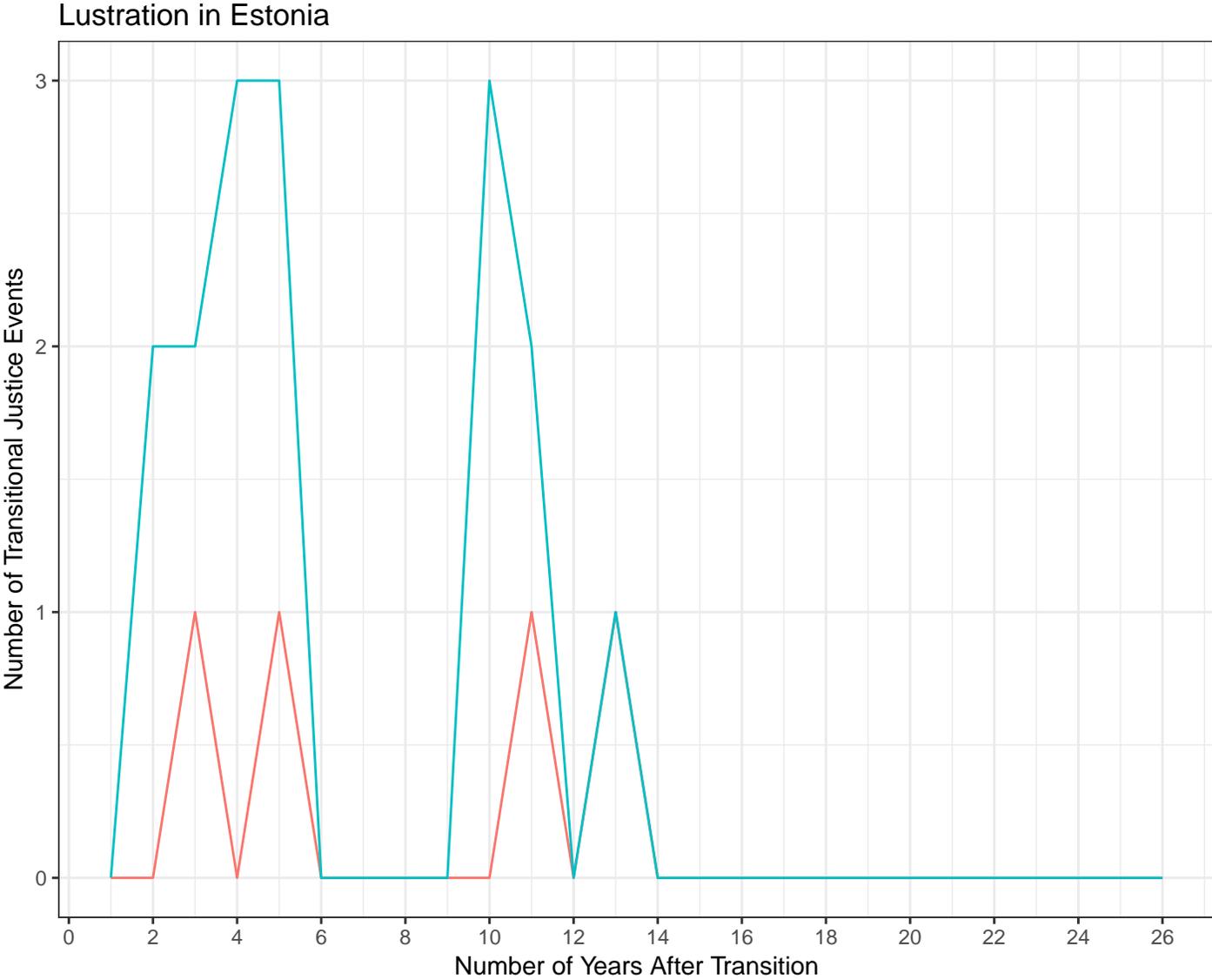
As an example, consider two databases that record transitional justice mechanisms since World War II: Olsen & Reiter (2010) focus on countries that have transitioned from autocracy to democracy, and Binningsbo et al. (2012) focus on post-conflict societies. After collecting the appropriate data, both teams of authors coded the presence or absence of a specific transitional justice mechanism, such as the existence of amnesty, or a purge.

¹⁷Other authors have proposed alternative ways of establishing and evaluating evidence of causality. See particularly Slater & Ziblatt's (2013) discussion on case studies and causality, and Pierson's (2003) insights into causal processes in social sciences.

These researchers' decisions to code the relevant phenomena as the presence or absence of a transitional justice mechanism allows them to aggregate the data and translate it into simple numerical data. Both databases provide relevant information to understand the adoption of transitional justice mechanisms in different contexts. Of course, this aggregation obscures other relevant parts of transitional justice that might be relevant. Specifically, by coding only the presence or absence of a transitional justice mechanism during the entire post-transition era, researchers might be reducing a sometimes very complex process of shepherding a transitional justice proposal through the legislative process to a dummy variable.

Large-N approaches allow for improvement on preexisting measures and for the incorporation of alternative dimensions of interest. For example, Ang & Nalepa (2017) acknowledge the political and contested nature of the implementation of transitional justice, in particular of lustration. This point is illustrated in figure 2, which summarizes the consequences of such reductionism applied to measuring lustration. It contains data on regressive and progressive lustration events from Estonia. Capturing the implementation of lustration by a simple 0 or 1 is definitely unfair to the turbulent process Estonia had with transitional justice. The highly complex and sometimes lengthy processes associated with the adoption of lustration have in the Olsen & Reiter (2010) and Binningsbo et al. (2012) datasets been straight-jacketed into a dummy variable. Figure 2 illustrates at least two additional sources of variation: variation in contention and temporal variation in the events. To address the first one, the authors define two kinds of events: progressive and regressive. A progressive TJ event is the submission of a TJ proposal to the floor of the legislature, the passage of such legislation, the upholding of such legislation as constitutional by a supreme court, or the overturning of a presidential veto against such legislation. In the case of truth commissions, the publication of the commission's report(s) and the extension of the commission's mandate are also considered to be progressive TJ events. Regressive transitional justice events are defined, in contrast, as the voting down,

Figure 2: Progressive and Regressive Lustration Events in Estonia since Transition in 1990



vetoing, or striking down by the supreme court of a transitional justice proposal or law. Similarly, expanding the set of persons targeted by TJ or broadening the set of “offenses” (where “offense” is defined in light of the TJ procedure in question) to include more past or present positions constitutes a progressive transitional justice event, whereas attempts to narrow the set of targets or “offenses” are coded as regressive TJ events. The guiding principle in determining if an event is regressive or progressive is whether it advances the TJ event forward or backward.

Second, the figure shows a temporal variation in the implementation of lustration laws. Acknowledging that not all events occur immediately in the aftermath of transition, and that the lapse of time between one event and the other are indicative of a process, rather than a characteristic, allows us to numerically capture this temporal variation. The measure proposed by Bates, Cinar & Nalepa (2018) which employs event data like the one in figure 2, offers yet another way of using time to evaluate lustration severity.

In short, quantification allows scholars to simplify complex characteristics by assigning them a numerical value. Although some information is, no doubt, lost in measurement, the decision to focus on some common features of lustrations allows researchers to collect data on many more phenomena and does not force them to restrict their attention to only a small part of the world. This stands in contrast with the practice of studying only one region where a phenomenon in question is prevalent.¹⁸

4 EITM:the best of both worlds?

Our final substantive section is devoted to the Empirical Implications of Theoretical Models (EITM), a program launched by the National Science Foundation in 2003 for clarifying the role of formal models in quantitative analysis and vice versa—the role of statistical mod-

¹⁸Of course, not all quantitative studies gather data from multiple countries, but a limitation of Horne (2017)’s work is that despite providing a rich description of lustration trajectories for as many as 12 countries, all of her observations come from postcommunist Europe.

els in quantitative work.

According to one interpretation in the EITM tradition, scholars after finding the equilibrium or equilibria of a game, would find how the equilibrium outcome responds to one or more of the parameters of the model. After operationalizing these parameters with empirical variables one can as it were, “test the model’s outcomes.” This process involves using comparative statics on the game’s equilibrium to develop hypotheses, and then, operationalizing parameters of the model with variables that the scholar has collected for a sufficient number of observations to yield significant results.

Writing about testing formal theory has a unfavorable reputation among political scientists (Johnson 2014).¹⁹ In clarke2012model work, the idea of literally testing formal models is ridiculed. The authors point out that a model cannot be tested simply because by virtue of being a system of deductive propositions, it cannot be false. Following from this, the authors are also skeptical about testing formal models with data, precisely because empirical models are still models. Data by itself cannot “test” anything. It needs first to be modeled, simplified. And just like formal models, empirical models rely on a series of assumptions, though these assumptions pertain to the data generating process. Compared to a formal model, an empirical model is no more objective. It is therefore not clear why data would be more suitable for testing the formal model rather than the other way around.

Other scholars approach the idea of “testing a formal model” with caution for practical reasons.²⁰ Bracketing for a moment whether models can be tested in the first place, these scholars believe they should never be tested by the same scholars who constructed them because it is clear that the incentives of the person attempting to “test” the model are biased in favor of finding supportive evidence. Due to this confirmation bias, they advocate,

¹⁹Many of the reflections contained in this section have been inspired by the panel on Formal Modeling and Qualitative Case Studies that Monika Nalepa co-organized with Peter Lorentzen at the Midwest Political Science Association Meeting in April of 2017. Participants included aside from the organizers, Andrew Little, Scott Gehlbach, Heins Geomans, Benjamin Lessing, Anne Meng and John Patty.

²⁰To quote Andrew Little, they will “reach for their wallet”.

a theoretical model should not be tested by its author.

Yet even Clarke & Primo (2012) are far from arguing that empirical models should appear unaccompanied by formal models. In the end, they agree that both kinds of models may appear side by side as a parts of a political science *explanation*. Through these lens, the findings from section 2 and ?? can be interpreted as two independent contributions—one to the theory of personnel transitional justice and one to the practices of data gathering. Both, we believe, shed light on how we should be thinking about lustration, purges, and truth commissions.

There is also a very productive strand of criticism aimed at EITM for privileging large-n analysis over qualitative case studies when it comes to testing models. According to this critique, quantitative tests ignore the theoretically rich material that a formal models has to offer. Formal models, as it were, produce very complex kinds of implications because an equilibrium, regardless of the solution concept one chooses to employ, is not actually an outcome (or representation of the state of the world), but an entire *strategy profile*. A strategy profile, in turn, is a collection of strategies—one for each player of the game—containing a complete plan of action, even for contingencies that may never arise. Thus, a strategy profile, for any dynamic game, will contain rationally consistent (in game theoretic terminology, “best”) responses “off the equilibrium path.” A thorough “test” of the model should not ignore these off-the equilibrium path responses. But limiting data collection only to variables that operationalize the equilibrium outcome is guilty of exactly that kind of omission.

Qualitative researchers may be better equipped to perform such tests thanks to process tracing and analytic narratives (?). Only qualitative data can unveil what is ultimately unobservable, namely the reason certain actors chose the actions they did (Goemans & Spaniel 2016, Bates 2007, Lorentzen, Fravel & Paine 2017). A semi-structured elite interview with a bureaucrat or politician can reveal he is doing x in order to avoid y, offering the researcher insight into off-the-equilibrium path behavior (Gailmard & Patty 2012). Histor-

ical memoirs and documents in archival repositories can similarly disclose the reasoning behind choosing certain actions and not others(?). While in principle, there is no obstacle to collecting such data on a larger set than just a handful of cases, such research would place enormous resource demands on scholars.

This observation just brings home the message that behind every large dataset that is constructed in political science out of observational data, is a series of (hopefully) more or less consistent qualitative judgement calls, which take place during coding. Typically, these judgement calls remain undisclosed. Hence even though quantitative tests rest on qualitative assumptions, they may be much less transparent about how they classify cases. Within transitional justice some dataset projects do this better than others. The Post-conflict Justice database (Gates, Binningsbo & Lie 2007), for instance, offers a complete set of raw data that led to every coding decision. Similarly, the Personnel Transitional Justice Dataset (Bates et al. 2018) supplies a website detailing how data was collected from primary sources and a list of primary and secondary sources that were consulted to assemble the raw data. The dataset website—[www.https://ipekcinar.shinyapps.io/personnal_transitional_justice_dataset/](https://ipekcinar.shinyapps.io/personnal_transitional_justice_dataset/)—contains a Frequently Asked Questions document describing how the research team resolved coding issues when collecting data. The qualitative information gathered to code each event could well be used also in a qualitative research design.

In light of the above criticisms, why not simply settle on an EITM that employs case studies?

We believe that this amounts to throwing out the baby with the bath water. First, because all the biases associated with “testing” theories performed by the authors of these theories are just compounded in the qualitative approach, where the threshold of confirmation is much lower. When all one needs is a handful of documents corroborating a theory, the researcher is tempted to rest his case sooner. Second, since “historians disagree with each other all the time” and “we [as political scientists] are not even trained as his-

torians”²¹ accumulating the handful of anecdotes confirming a theory seems to present a much lower hurdle than a large-n statistical test. Indeed, this last concern is corroborated by recent work exposing the lack of uniform standards guiding the process of combining case studies with formal modeling (Lorentzen et al. 2017).

4.1 Connecting formal and quantitative worlds

Keeping all these caveats in mind, in evaluating the model discussed in section 2 we turn to quantitative analysis. To clearly outline our supportive evidence, we present the insights from the formal model in the form of two hypotheses:

1. As the severity of lustration increases, the quality of representation will also increase.
2. As preferences of the former autocrats and the new democratic diverge, the country’s quality of representation will improve.

If the formal model allowed us to derive implications of blackmail, an action that by its nature is hard to observe, the use of quantitative approaches allows us to analyze empirically if these implications bear out. Specifically, we argue that if politicians are indeed subject to blackmail, we can observe the aggregate effects of such blackmail *even if observing specific blackmailing is impossible*. This is so because, on average, politicians that are blackmailed will be less capable of consistently delivering the policies offered to voters. In turn, this inconsistent behavior will be observable at the party level: in places where information can be easily leveraged, politicians will have a hard time building robust political parties that offer policy proposals and implement them.²²

Therefore, the best feasible empirical probe for our hypothesis relies on party-level measures of quality of representation. Specifically, we use the Democratic Accountability and Linkages Project (DALP), a survey of country experts survey for measures of our dependent variable: degree of programaticness of the party system. Specifically we build on

²¹Scott Gehlbach, personal communication at MPSA Panel, April 2017

²²As discussed before, this is not the only reason why countries have poor quality of representation.

(Kitschelt & Freeze 2010) proposed measure of partisan consistency, a variable that measures how consistent each party is and how well have they been able to situate their policy proposals in the relevant policy space.²³ This variable ranges from 0 to 1, where 0 is the least programmatic party and 1 is the most programmatic.

Our second hypothesis is that as the preferences of former autocrats and politicians under democracy diverge, the quality of representation will improve. Since preferences of former autocrats are also unobservable, we also exploit the advantages of aggregate measures. Specifically, we argue that ideological placements of former autocrats can be estimated using the ideological placement of successor parties, or parties that ruled during the authoritarian era and were able to linger in the post-transition political landscape. Thus, we use DALPs data to measure the placement of parties on a left-right continuum and to calculate the ideological distance between successor parties and democratic parties.

In addition, there are a few parameters of the model that ought to be accounted for in the statistical model. For example, we have to account for the cost that politicians pay when compromising information about them is revealed, and for the cost that the agent pays when he is bluffing (suggesting he is in possession of the evidence when in fact he has no such evidence). We account for the former with the number of years lapsed since the democratic transition. The rationale here is that revealing this information is more damaging shortly following the transition rather than further along. We account for the latter by using a measure of press freedom: we argue that countries with free presses will be better at unearthing the truth, and exposing the agent's bluff. Finally, although not in the model, we also include a variable measuring the status of the opposition under the former authoritarian regime. After all, it could be that authoritarian regimes that allow the opposition to exist in their legislatures, were less likely to implement lustration and were also less likely to build high quality parties after the transition.

²³Our variable builds on Kitschelt et. al.'s *cosalpo₄*, which measures party programmaticness with cohesion, saliency, and polarization. However, to avoid correlation with one of our independent variables—divergence of preferences between former autocrats and new democratic politicians—we eliminated the polarization component from this measure.

Table 1: Explaining quality of representation (country intercepts not shown)

	Model 1	Model 2	Model 3	Model 4	Model 5
<i>distance_party</i>	-0.093** (0.043)	-0.089** (0.043)	-0.091** (0.043)	-0.094** (0.043)	-0.090** (0.043)
<i>severity</i>		0.217*** (0.078)	0.209** (0.093)	0.687** (0.339)	0.206** (0.095)
<i>n_year</i>			-0.055 (0.187)	0.125 (0.221)	-0.072 (0.200)
<i>press_freedom</i>			0.283 (0.301)	0.375 (0.301)	0.299 (0.313)
<i>n_year * severity</i>				-0.985 (0.672)	
<i>opp_status</i>					-0.016 (0.060)
Constant	0.393*** (0.038)	0.309*** (0.047)	0.113 (0.215)	-0.022 (0.230)	0.118 (0.219)
Observations	158	158	158	158	158
Log Likelihood	89.776	91.727	91.028	92.608	89.155
Akaike Inf. Crit.	-171.552	-173.453	-168.056	-169.216	-162.311
Bayesian Inf. Crit.	-159.302	-158.140	-146.618	-144.716	-137.810

Note:

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

Our data on programmaticness is at the party level and hence our unit of analysis is a party in a country. In light of this “nested” structure of our data, the appropriate statistical model to use is a hierarchical model, known also as a random effects model. We present the two variables of interest, which correspond to our hypotheses, in the top two rows of our table. In order to assess their robustness, we begin with presenting just the bivariate regression between programmaticness and party distance. The effect of party distance on programmaticness is negative, which is consistent with our model, since increasing the distance, changes the cutoff between the pooling and separating equilibrium. This negative effect remains significant after adding other covariates, including lustration severity, our second factor of interest in predicting the effect of lustration. Lustration too appears to improve programmaticness. As our model suggests, increasing lustration severity will increase democratic representation. This effect is clear from the effect of π , the probability that evidence survived the transitional justice process in all three equilibria. This is the case even after accounting for the remaining covariates responsible for representing features of the formal model in the empirical set up. None of these remaining variables are significant. The lesson learned from this empirical exercise is that for both theoretical and empirical reasons we should expect two phenomena:

1. Polarization between the former autocrats and the incoming democratic politicians will jeopardize the lustration process, making it work not as the proponents of the forward looking argument predict it will. Empirically, this impacts programmatic representation negatively.
2. Increasing the severity of lustration, both theoretically leads to the better quality representation and empirically is associated with more programmatic parties

The next section, we discuss the way the two exercises in modeling fit together. Before we do so, however, we list some limitations of quantitative research alone.

4.2 Limitations

Large-n analysis that uses observational data presents a set of challenges, which have become recognized during the so-called “identification revolution” in the social sciences. The essence of this movement was to point out that making causal claims on the basis of correlations discovered in observational data is unwarranted. For the purposes of this specific chapter, we point out that such a fallacy is particularly dangerous when studying associations between transitional justice as the independent variable and the quality of democratic representation as the outcome variable. The same factors that lead states to engage in transitional justice may also be pushing them to greater democratic representation. Thus any relationship we find between transitional justice and the quality of democratic representation could be spurious. How can this problem of endogeneity be resolved? One way, which is proposed by Ang, Bates & Nalepa (2018) is to leverage the time series nature of the transitional justice data in a difference in difference design. What this design allows researchers to do is to construct counterfactuals of countries that adopted transitional justice by exploiting similarity in trends with countries that did not adopt transitional justice at the same point in time. Implementing this design does require, however, the time consuming collection of time series data both for transitional justice and quality of democracy indicators. And it is problematic as to whether in the case of transitional justice event data—whether we divide it into regressive and progressive or other types of events—it makes sense or whether democracy indicators have enough over time variation to be meaningfully used (Pennings, Keman & Woldendorp 2016, Wright 2017).

5 Conclusion

This chapter has attempted to clarify certain misconceptions about the uses and misuses of formal and quantitative modeling, particularly in their application to transitional justice research. There is no doubt that these phenomena exhibit a great degree of complexity.

No formal model can capture the richness of a mechanism that holds past perpetrators and their collaborators accountable. Similarly, no two transitional justice processes are exactly alike. So no statistical aggregation can do justice to the unique way in which states deal with their pasts. Yet, in the parsimony and reductionism of modeling lies simultaneously its strength. Like maps, formal models abstract and offer practical directions on how to anticipate unintended consequences of certain transitional justice mechanisms. Formalization allows us to examine the consistency of theoretical arguments, as the model of Ang & Nalepa (2017) does. Quantitative models offer the ability to generalize and discipline our natural tendencies to extrapolate beyond what is warranted.

Together, formal and quantitative models allow us to make inferences that are impossible to make using case studies and normative theorizing alone. In this chapter we have shown such advantages using the case of lustration, and the possibility that its improper implementation will decrease the quality of democracy by enabling blackmail. Of course, the methods discussed here are by no means a substitute for qualitatively intense case studies, nor will they ever displace such research. Instead, we propose to think of these multiple methods as existing in conversation. But it is very important that modeling become more integrated into the transitional justice mainstream if the research on transitional justice is to become recognized as part of political science.

A Formal Model

The game starts with a move of Nature, which determines with probability $\pi \in (0, 1)$ that the officer has evidence against the politician, and with probability $1 - \pi$ that he does not. π in this model represents the exogenously given severity of lustration, as the more severe lustration is, the less likely it is that unearthed evidence of collaboration indeed remains in the hands of the officer. The *Politician*, P , knows the value of π , but not its specific realization. In the second stage of the game, the *Officer*, (O) decides whether or not to

make a policy demand towards the *Politician*. We represent this decision by one of two exogenously given actions {demand, no demand}. If no demand is made, the game ends and the Politician implements the policy corresponding to his ideal point.

For simplicity, we assume that the politician's ideal point is given by $p = 0$ and the officer's ideal point is represented by $a > 0$. Thus, a captures the ideological distance between the politician and the officer. P observes whether or not a demand has been made, but not whether evidence against him exists (the realization of π). In the third stage, P decides whether to make a policy concession in response to the officer's demand, and how big this concession should be. We model this concession as proposing a policy $x \in [0, a]$. If evidence against the politician does not exist, the game ends and the officer pays the cost of bluffing, $0 < c < a^2$.²⁴ The bluffing cost, c , captures the cost of obtaining and disclosing some plausible information about the politician's past. This is costlier for the officer in a world of independent media, where such disclosures are verified by journalists and, if false, exposed.

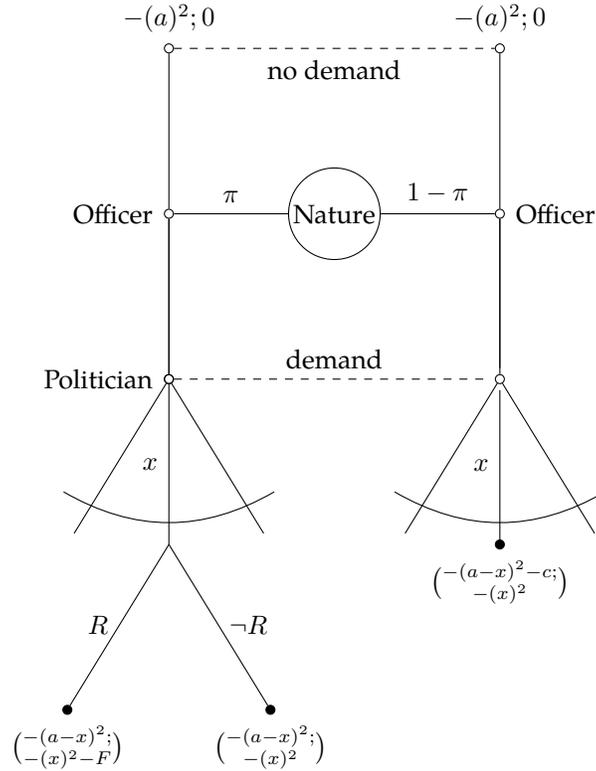
If evidence does exist, the officer decides whether or not to reveal the evidence (R or $\neg R$). The revelation is interpreted as the officer exercising his threat to reveal compromising evidence against the Politician. In the event that evidence is revealed, the politician pays the cost of being fired, F .

A.1 Strategies and beliefs

A strategy for the officer is a triple $s_O = (z^E, r, z^{\neg E})$, where $z^E \in \{\text{Demand, No Demand}\}$ denotes O 's action when he is in possession of evidence against the politician P and $z^{\neg E} \in \{\text{Demand, No Demand}\}$ denotes O 's action when he is not in possession of such evidence. Suppose that $r \in \{R, \neg R\}$ denotes the action taken by O following P 's counteroffer x . This part of O 's strategy can best be represented as a rejection region $R = \{x : x \in [0, a] \rightarrow O \text{ Reveals}\}$. A strategy for the politician is a proposal $x \in [0, a]$

²⁴The second restriction is purely technical and merely simplifies equilibrium calculation.

Figure 3: Lustration Blackmail Game



that P makes in the event that O makes a policy demand. Note that O can simply ignore the demand if he chooses $x = 0$.²⁵

A.2 Preferences

The payoffs are a quadratic function of the Euclidean distance between the players' respective ideal points and implemented policy as well as the two types of costs characterized above: (1) the cost to the politician of being fired as a result of revealing skeletons in his closet; and (2) the cost of bluffing incurred by the officer if he makes an empty threat. The utilities, along with the entire game tree, are presented in Figure 3 above.

²⁵Since P does not observe whether evidence exists or not, but only observes whether a demand was placed or not, we only need a single action to describe his strategy.

A.3 Equilibrium Analysis of the Formal Model

We first will verify the plausibility of the pure separating equilibrium in which all agents equipped with evidence choose a different strategy from all agents not equipped with such evidence.

A.3.1 Pure separating equilibrium

Since it is never rational for O not to make a demand when he has evidence against the Politician,²⁶ the only possible separating equilibrium is: (D, R, ND, x^*) . The requirement for this to be a Bayesian equilibrium is that the posterior beliefs are $Pr(E|ND) = 0, Pr(\sim E|ND) = 1, Pr(E|D) = 1, Pr(\sim E|D) = 0$. The following four steps will lead to uncovering the conditions for this equilibrium.

1. First, suppose x^* is the proposal accepted in equilibrium. P knows that if he proposes $x < x^*$, he will his skeletons will be exposed with certainty. Therefore, on the one hand, if he is going make a proposal, in equilibrium, it must be accepted, because if it were to be rejected P should propose $x = 0$. On the other hand, if a demand is placed and P makes a counteroffer $x \neq 0$, P 's utility must be greater from having his proposal accepted than from having it rejected, i.e.,

$EU_P(x^*|D, ND) \geq EU_P(0|D, ND)$, which is equivalent to $-(x^*)^2 \geq -0 - F$. The last expression simplifies to

$$x^* \leq \sqrt{F}$$

2. Second, note, that the lowest x^* that P is willing to accept is $x^* = \sqrt{F}$. Therefore, O 's optimal rejection region is $R = [0, \sqrt{F}$
3. Third, note that to prevent O from bluffing (and making demands when there is

²⁶Note that P cannot make a counterproposal x unless O has made a demand. Thus, the worst O can do when making a demand is $-|a - d|$, which would be his payoff if P 's counterproposal were $x = a$, that is if P made no concession at all. But without making any demand O is guaranteed to receive $-(a)^2$ and no more.

no evidence), it has to be the case that that $EU_O(D|x^*, \sim E) \leq EU_A(ND|x^*, \sim E)$, which is equivalent to $-(a - x^*)^2 - c \leq -|(a)^2|$. This last expression simplifies to $x^{*2} + 2ax^* - c \leq 0$,

4. Fourth, since the expression is on the RHS is a quadratic expression, we have to solve for the inequality by applying the quadratic formula to $x^{*2} + 2ax^* - c = 0$ to identify $x_1^* = a - \sqrt{a^2 - c}$ and $x_2^* = a + \sqrt{a^2 - c}$. Because the counterproposal must be in $[0, a]$, x_2^* is eliminated and x_1^* is not, by virtue of our assumption from above, $c < a^2$.
5. This leads to $x^* \leq a - \sqrt{a^2 - c}$ which after substituting for x^* to

$$F \leq (a - \sqrt{a^2 - c})^2$$

We conclude that a pure separating equilibrium exists only when the cost of having skeletons revealed (F) relative to the cost of bluffing c is quite low. In this pure separating equilibrium, blackmail is effective with probability π and the average democratic misrepresentation is proportional to the cost of having skeletons in the closet exposed. In the next two sections, we show how this departure from programmatic representation compares with the effectiveness of blackmail under the pure pooling and hybrid equilibria.

A.3.2 Pure pooling equilibrium

In pooling equilibria, officers with and without evidence will choose the same action, implying that the Politician cannot update his prior beliefs to posterior beliefs by conditioning on the agent's action. In the analysis above, we established that the Officer will never refuse to place a demand when evidence is present. Thus, the only possibility of a pooling equilibrium in this game is $(D, R, D; x'')$ with accompanying beliefs: $Pr(E|D) = \pi, Pr(\sim E|D) = (1 - \pi)^{27}$. In this equilibrium, the Officer always places a

²⁷Note that paths that involve the agent not making a demand are off the equilibrium path and we do not have to specify the beliefs there

demand and the the Politician always offers the same counterproposal, x'' .

1. For such an equilibrium to hold, the dissident has to prefer to have his proposal accepted to having skeletons in the closet revealed (in which case, he would simply propose his ideal point, 0). Thus, it must me the case that $EU_D(x'|D, D) \geq EU_D(0|D, D)$, which is equivalent to $-(x')^2 \geq \pi(-0-F) + (1-\pi)*0$.²⁸ The last expression simplifies to :

$$x' \leq \sqrt{\pi F}$$

2. Since $x' = \sqrt{\pi F}$ is the highest proposal the Politician will accept, given his beliefs, the Officer's optimal rejection region is $[0, \sqrt{\pi F}]$.
3. To ensure the Officer always has an incentive to place a demand, it has to be the case that $EU_A(D|x', \sim E) \geq EU_A(ND|x', \sim E)$, which is equivalent to $-(a - x')^2 - c \geq -(a)^2$. This last expression simplifies to $x' \geq a - \sqrt{a^2 - c}$ (using the solution to the quadratic formula from above, except with the sign on the quadratic coefficient flipped).
4. Substituting for x' , we arrive at:

$$F \geq \frac{(a - \sqrt{a^2 - c})^2}{\pi}.$$

Summing up, a pure pooling equilibrium exists only when the cost of having skeletons revealed (F) relative to the cost of bluffing c is quite high. Equivalently, we can also state:

$$\pi \geq \frac{(a - \sqrt{a^2 - c})^2}{F}.$$

²⁸Note that since the Politician cannot tell which type—with or without evidence—he is facing any better than he could before the Officer took an action, his expected utility from making a proposal outside of the acceptance region is weighted by his priors about the probability that evidence exists.

The second expression shows that the pooling equilibrium is more likely for higher values of π , which correspond to systems with less severe lustration. In this pure pooling equilibrium, blackmail is always effective (takes place with probability 1). The distortion it causes relative to the Politician's ideal point and is $\sqrt{\pi F}$. It is proportional to the cost of firing and the extent to which evidence exists. Our final subsection of the equilibrium analysis looks at the effectiveness of blackmail with secret police files under the semi-pooling (hybrid) equilibrium.

A.3.3 Semi-separating equilibrium

In addition to the pure separating equilibrium discussed above, we also derive the conditions (and verify their plausibility) of a semi-pooling or (or hybrid) equilibrium. In this equilibrium, the Officer plays a mixed strategy. He always makes a demand when evidence is present, but he also with some probability λ makes a demand if evidence does not exist (and with probability $1 - \lambda$ does not make a demand). Consequently, any hybrid equilibrium must fit the format $(D, R, \lambda; x'')$.²⁹ Note the beliefs consistent with this semi-pooling equilibrium are: $Pr(E|ND) = 0, Pr(\sim E|D) = \frac{\lambda(1-\pi)}{\pi+\lambda(1-\pi)}, Pr(\sim E|ND) = 1 - \lambda, Pr(E|D) = \frac{\pi}{\pi+\lambda(1-\pi)}$. The calculation of conditions for which this equilibrium obtains proceeds in six steps:

1. First, to find the equilibrium value of λ^* , we calculate the expected utility of the Politician from responding x'' to the Officer's demand ($EU_D(x''|D, \lambda)$) and set it equal to the expected utility of the dissident's choosing his ideal point, 0, which is outside of the Officer's acceptance region, $EU_D(0|D, \lambda)$ This yields the equality $\pi(-(x'')^2) + (1 - \pi)\lambda(-(x'')^2) + 0 = -\pi F$, leading to:

$$\lambda = \frac{-\pi F + \pi(x'')^2}{(1 - \pi)[-(x'')^2]} = \frac{\pi(F - (x'')^2)}{(1 - \pi)(x'')^2} \quad (1)$$

²⁹This means that the Politician makes a demand with probability 1 if evidence exists and with probability λ if evidence does not exist.

2. To ensure that $0 < \lambda < 1$ and is a probability we need:

$$-(x')^2 < F \quad (2)$$

and

$$\frac{\pi F - \pi(x'')^2}{(1 - \pi)(x'')^2} < 1 \quad (3)$$

, which simplifies to $F < \frac{(x'')^2}{\pi}$.

3. Next, to pin down x'' , we make use of the fact that when evidence does not exist, O must be indifferent between placing a demand and not placing one, i.e.: $EU_O(D|x'', \sim E) = EU_A(ND|x'', \sim E)$, which reduces to: $-(a)^2 = -(a - x'')^2 - c$. This last quadratic equality is solved again using the familiar quadratic formula, where we obtain just $x''_1 = a + \sqrt{a^2 - c}$, which does not satisfy the constraint on the counter-proposal $0 < x < a$ and $x''_2 = a - \sqrt{a^2 - c}$, which does.

4. Finally, substituting x' into equation 1, we arrive at:

$$\lambda = \frac{\pi(F - (a - \sqrt{a^2 - c}))}{(1 - \pi)c} \quad (4)$$

5. And to ensure the λ is a probability, we will need (by substituting x'' into 3:

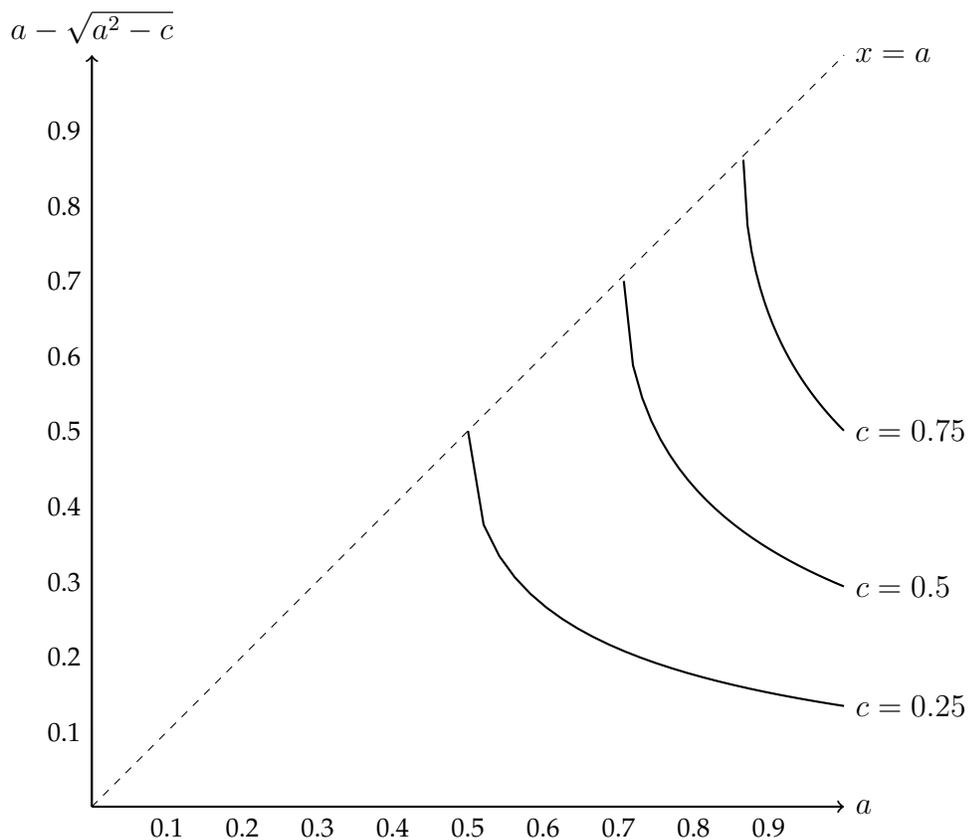
$$(a - \sqrt{a^2 - c})^2 < F < (a - \sqrt{a^2 - c})^2 \pi \quad (5)$$

6. Given the condition above, note that the game only has a semi pooling equilibrium if the set of F 's satisfying condition defined in 5 is non-empty.

Finally, just as in the case of the separating and semi-pooling equilibrium, we can calculate the expected departure from perfect representation in this equilibrium. It will be given by:

$$\pi(a - \sqrt{a^2 - c}) + (1 - \pi) \frac{\pi(F - (a - \sqrt{a^2 - c})^2)}{(1 - \pi)(a - \sqrt{a^2 - c})^2}, \text{ which simplifies to } \pi(a - \sqrt{a^2 - c}) + \frac{F}{(a - \sqrt{a^2 - c})^2} - 1)$$

Figure 4: Constraints on c and a ensuring that the counterproposal x falls between 0 and a



The two figures below can aid our interpretation of these results. First, Figure ?? shows how the requirement that $c < a^2$ affects the equilibrium counterproposals in the the semi-pooling equilibrium and the cutoffs between the separating and pooling equilibrium. This cutoff is relevant for determining for which parameter values lustration works “as it ought to” according to the normative forward looking argument. We see that as a increases, the cutoff between the separating equilibrium and pooling equilibrium, making the separating equilibrium less common.

References

- Aguilar, P., Balcells, L. & Cebolla-Boado, H. (2011), 'Determinants of attitudes toward transitional justice: An empirical analysis of the Spanish case', *Comparative Political Studies* 44(10), 1397–1430.
- Ang, M., Bates, G. & Nalepa, M. (2018), 'Identifying the effect of personnel transitional justice on the quality of democratic representation'.
- Ang, M. & Nalepa, M. (2017), 'Secret authoritarian legacies, transitional justice, and the quality of representation', *Unpublished Manuscript*.
- Appel, B. J. & Loyle, C. E. (2012), 'The economic benefits of justice: Post-conflict justice and foreign direct investment', *Journal of Peace Research* 49(5), 685–699.
- Aukerman, M. J. (2002), 'Extraordinary evil, ordinary crime: A framework for understanding transitional justice', *Harv. Hum. Rts. J.* 15, 39.
- Bates, G., Cinar, I. & Nalepa, M. (2018), 'Taking time seriously: Introducing a new transitional justice dataset: 1946-2016'.
- Bates, R. H. (2007), From case studies to social science: A strategy for political research, in 'The Oxford Handbook of Comparative Politics'.
- Binningsbo, H. M., Loyle, C. E., Gates, S. & Elster, J. (2012), 'Armed conflict and post-conflict justice, 1946–2006: A dataset', *Journal of Peace Research* 49(5), 731–740.
- Clarke, K. A. & Primo, D. M. (2012), *A model discipline: Political science and the logic of representations*, Oxford University Press.
- Dancy, G., Kim, H. & Wiebelhaus-Brahm, E. (2010), 'The turn to truth: Trends in truth commission experimentation', *Journal of Human Rights* 9(1), 45–64.

- Elster, J. (2004), *Closing the books: Transitional justice in historical perspective*, Cambridge University Press.
- Elster, J. (2006), 'Retribution and reparation in the transition to democracy'.
- Gailmard, S. & Patty, J. W. (2012), *Learning while governing: Expertise and accountability in the executive branch*, University of Chicago Press.
- Gates, S., Binningsbo, H. M. & Lie, T. G. (2007), *Post-conflict justice and sustainable peace*, The World Bank.
- Goemans, H. & Spaniel, W. (2016), 'Multimethod research: A case for formal theory', *Security Studies* 25(1), 25–33.
- Horne, C. M. (2017), *Building Trust and Democracy: Transitional Justice in Post-Communist Countries*, Oxford University Press.
- Johnson, J. (2014), 'Models among the political theorists', *American Journal of Political Science* 58(3), 547–560.
- Kamiński, M. M. & Nalepa, M. (2014), 'A model of strategic preemption: why do post-communists hurt themselves?', *Decyzje* (21), 31–66.
- King, G., Keohane, R. O. & Verba, S. (1994), *Designing social inquiry: Scientific inference in qualitative research*, Princeton university press.
- Kitschelt, H. & Freeze, K. (2010), Programmatic party system structuration: Developing and comparing cross-national and cross-party measures with a new global data set, in 'APSA Annual Meeting, Washington, DC'.
- Krehbiel, K. (2010), *Pivotal politics: A theory of US lawmaking*, University of Chicago Press.
- Leebaw, B. A. (2008), 'The irreconcilable goals of transitional justice', *Hum. Rts. Q.* 30, 95.

- Letki, N. (2002), 'Lustration and democratisation in east-central europe', *Europe-Asia Studies* **54**(4), 529–552.
- Lieberson, S. (1991), 'Small n's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases', *Social forces* **70**(2), 307–320.
- Lorentzen, P., Fravel, M. T. & Paine, J. (2017), 'Qualitative investigation of theoretical models: the value of process tracing', *Journal of Theoretical Politics* **29**(3), 467–491.
- Lupu, N. & Peisakhin, L. (2017), 'The legacy of political violence across generations', *American Journal of Political Science* **61**(4), 836–851.
- Mill, J. S. (1884), *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*, Vol. 1, Longmans, green, and Company.
- Murphy, C. (2017), *The conceptual foundations of transitional justice*, Cambridge University Press.
- Nalepa, M. (2008), 'To punish the guilty and protect the innocent: Comparing truth revelation procedures', *Journal of Theoretical Politics* **20**(2), 221–245.
- Nalepa, M. (2010a), 'Captured commitments: an analytic narrative of transitions with transitional justice', *World Politics* **62**(2), 341–380.
- Nalepa, M. (2010b), *Skeletons in the closet: Transitional justice in post-communist Europe*, Cambridge University Press.
- Nalepa, M. (2012), 'Tolerating mistakes: How do popular perceptions of procedural fairness affect demand for transitional justice?', *Journal of Conflict Resolution* **56**(3), 490–515.
- Nichols, E. (1986), 'Skocpol on revolution: comparative analysis vs. historical conjuncture', *Comparative Social Research* **9**(1), 163–186.

- Olsen, Tricia D., P. L. & Reiter, A. (2010), 'Transitional justice in balance: Comparing processes, weighing efficacy'.
- Payne, L. A. (2007), *Unsettling accounts: Neither truth nor reconciliation in confessions of state violence*, Duke University Press.
- Pennings, P., Keman, H. & Woldendorp, J. (2016), 'Relating theory and concepts to measurements: bridging the gap', *Handbook of Research Methods and Applications in Political Science* p. 54.
- Pierson, P. (2003), 'Big, slow-moving, and... invisible', *Comparative historical analysis in the social sciences* pp. 177–207.
- Prorok, A. K. (2017), 'The (in) compatibility of peace and justice? the international criminal court and civil conflict termination', *International organization* **71**(2), 213–243.
- Romer, T. & Rosenthal, H. (1978), 'Political resource allocation, controlled agendas, and the status quo', *Public choice* **33**(4), 27–43.
- Rozenas, A., Schutte, S. & Zhukov, Y. (2017), 'The political legacy of violence: The long-term impact of stalin's repression in ukraine', *The Journal of Politics* **79**(4), 1147–1161.
- Sikkink, K. (2011), *The Justice Cascade: How Human Rights Prosecutions Are Changing World Politics (The Norton Series in World Politics)*, WW Norton & Company.
- Slater, D. & Ziblatt, D. (2013), 'The enduring indispensability of the controlled comparison', *Comparative Political Studies* **46**(10), 1301–1327.
- Tarski, A. (1944), 'The semantic conception of truth: and the foundations of semantics', *Philosophy and phenomenological research* **4**(3), 341–376.
- Tucker, A. (2015), *The Legacies of Totalitarianism: A Theoretical Framework*, Cambridge University Press.

Williams, K., Fowler, B. & Szczerbiak, A. (2005), 'Explaining lustration in central europe: a post-communist politics approach', *Democratization* **12**(1), 22–43.

Wright, J. (2017), 'The latent characteristics that structure autocratic rule'.