Expectations and Economic Choices:

Essays on Projection Bias, Expectation-Based Reference Points, and the Emergence of Extreme Political Systems

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2012

vorgelegt von

Lukas Buchheim

Referent:	Professor Dr. Uwe Sunde
Korreferent:	Professor Dr. Klaus M. Schmidt
Promotionsabschlussberatung:	15. Mai 2013

Datum der mündlichen Prüfung: 08. Mai 2013

Namen der Berichterstatter: Uwe Sunde, Klaus M. Schmidt, Florian Englmaier *für meinen Vater*

Acknowledgements

I would not have made the choice to begin working on a dissertation if I had not expected to receive support from numerous people. At the time I have started, however, I did not know who these people were going to be. Now I know, and this is the right place to thank at least some of them.

Foremost, I would like to thank my advisors Uwe Sunde and Klaus M. Schmidt whose doors were always open when I needed advice and guidance. I have learned a lot from conversations and discussions with them. Sven Rady provided help and guidance to the same extent during the first two and a half years of work on this dissertation before my research interests started to diverge from his, for which I am the only one to blame. For serving as the oral examiner and especially for helpful advice regarding my professional development I would like to thank Florian Englmaier.

During the last couple of years I had the pleasure to start joint projects with – in chronological order – Robert Ulbricht, Thomas Kolaska, and Martin Watzinger (the projects with Robert and Thomas are part of this dissertation). It was (and is) great fun to conduct research with all three of them; and more importantly, all of them have become good friends over the years of joint work. Thank you, Robert, Thomas, and Martin! Additionally, I would like to thank my colleagues Manuela Beckstein, René Cyranek, Clemens König, Johannes Maier, Michal Masika, Ludwig Reßner, Caspar Siegert, Rudi Stracke, Sebastian Strasser, Ludwig Straub, Piers Trepper, and Larissa Zierow for creating a stimulating and cooperative atmosphere at work making the university a place (almost) like home.

Working on the dissertation has, at times, been difficult despite all the help I have received. For keeping up my morale during all times and for providing support throughout the years big thanks go to Thomas Meschede, my family, and Verena Risse. I do not have the right words to express my gratitude to them – especially Verena – for being there for me; for this reason, I have to leave it at that. My dissertation is dedicated to my late father.

Lukas Buchheim, October 2013

Contents

List of Tables

List of Figures

Preface

This dissertation combines three contributions to the fields of behavioral and political economics. Each contribution corresponds to one chapter; all chapters are self-contained and can be read independently. Despite their diverse topics, all chapters have a common focus: the interdependence of expectations and economic choices.[1]

There are several ways how expectations and economic choices can be related to each other. The obvious way concerns situations in which economic agents are making forward-looking decisions under uncertainty. In such situations, an agent's expectation regarding the consequences of her decisions will determine her preferences over available alternatives and, hence, her choice behavior. The standard economic approach for analyzing these kinds of situations is expected utility theory; it rests, among others, on the following two implicit assumptions. First, it is assumed that agents predict their utility from all possible consequences of their actions correctly; in other words, agents know their future preferences. Second, agents always choose the option with the highest expected utility, which is calculated by weighting the utility of each consequence of an alternative with its expected probability. Apart from this, expectations do not affect preferences over alternatives.

Two recent theories from the field of behavioral economics dispense with these two assumptions: Contrary to the first assumption, Loewenstein et al. (2003) argue that agents' perceptions of their state-dependent future preferences are not necessarily correct and instead biased towards their preferences at today's state of the world. Contrary to the second assumption, Kőszegi and Rabin (2007) show how probabilistic expectations may affect preferences apart from weighting the utility of each possible consequence. They argue that expectations serve as reference points that determine whether a consequence is evaluated favorably as a gain or unfavorably as a loss. Whether or not these theories describe behavior accurately ultimately needs to be studied empirically. This is the aim

[1] Throughout this preface, we will use the term "expectations" in the broad sense of describing an agent's general perception of situations involving uncertainty. In particular, we will make no distinction between expectations and beliefs. In the main chapters of this dissertation, however, the terms "expectations" and "beliefs" will be used in their usual, more narrow, sense.

of the first two chapters of this dissertation.

A second aspect of how choices and expectations are related is that the behavior of one agent may affect the expectations of others. In particular, if information is asymmetrically distributed, uninformed agents can learn from observing choice behavior of the well-informed. This may create incentives for the well-informed parties to disclose or conceal their information in order to influence the expectations – and, hence, actions – of the lesser informed agents. In Chapter 3 we analyze theoretically, how such signaling considerations may affect political transitions and the emergence of extreme political systems.

The first chapter, which is based on joint work with Thomas Kolaska, evaluates whether individual expectations regarding their future preferences are, on average, correct. While this assumption is generally maintained in standard economic theory, Loewenstein et al. (2003) argue that individuals may actually have difficulties predicting their future utility accurately. In particular, individuals tend to underestimate the degree to which their future preferences will change due to changes in the state of the world. This bias – labeled *projection bias* by Loewenstein et al. – can be illustrated by a simple example: Individuals tend to overweight utility from food when hungry, which is why shopping on an empty stomach should be avoided.

We test for projection bias by studying advance sales of an outdoor movie theater. In this context, projection bias predicts that good weather on the purchase-date leads customers to overvalue their utility from visiting the theater on the movie-date. Hence, the number of advance sales for the theater should be positively related to good weather on the purchase-date.

The decision problem of whether or not to buy tickets in advance has a number of characteristics which make it a suitable testing ground for projection bias. First, utility from visiting an open air theater is clearly weather- and, hence, state-dependent. Second, the variation in the state "weather" is independent of the individual decision to buy tickets such that the chain of causality is clear. Third, we demonstrate that weather is highly variable at the location of the theater such that the current state (current weather) does not provide useful information regarding the realization of future states (future weather). Finally, the stochastic process of weather should be among the stochastic processes best understood by individuals, as it is part of daily experience. This and the simplicity of the decision problem should minimize the propensity of decisions being biased at all.

We contribute to the literature by demonstrating that projection bias affects choices in an economically significant way even in this simple decision problem. According to our estimates, a one standard deviation change in the daily sunshine duration (a salient indicator of weather) leads to a change in ticket sales between 10 and 25 percent on average, irrespective of how many days (or even weeks) in advance tickets are sold.

We rule out a number of alternative explanations for this result. We demonstrate that neither the information content of current weather, weather-related shifts in the number of customers, nor weather-related capacity constraints of the movie theater can account for all regularities we find in the data. Only projection bias is consistent with all our findings.

Similar to the first chapter, the second is devoted to testing a behavioral theory for decision making under risk. The theory tested here – Kőszegi and Rabin's (2006; 2007) model of reference dependent preferences – is an extension of the theory of *loss aversion,* an essential part of one of the oldest and most influential behavioral models in economics: Kahnemann and Tversky's (1979) prospect theory.

Loss aversion describes the idea that individuals evaluate consequences of their choices not only by the outcome itself, but also assess these outcomes in comparison to a reference point. If the outcome is better than the reference point, they perceive this as a gain; on the other hand, if the outcome is worse than the reference point, individuals experience a loss. It is generally maintained that the experience of losses is a stronger feeling than the experience of gains such that individuals dislike losses of a certain size more than they like gains of the same size.

Since gains and losses emanate from comparing outcomes with reference points, the question of what constitutes the reference point is central for any theoretical application of loss aversion. The contribution of Kőszegi and Rabin is to provide a theory of how reference points are formed endogenously. Their central ideas are summarized in the following quote: "A person's reference point is the *probabilistic belief* she held *in the recent past* about outcomes" (Kőszegi and Rabin, 2006, p. 1134; emphases added). It is worth elaborating on the implications of this quote because this allows us to derive the testable prediction of Kőszegi and Rabin's theory that is at the heart of our work.

First, equating the reference point with *probabilistic beliefs* about outcomes or, in other words, expectations, implies that individuals feel gains and losses with respect to all possible outcomes they expect. As an example, consider an individual endowed with a lottery that pays 27 Euro or 7 Euro with equal probability. When the reference point is given by the payments of this lottery, she will evaluate the unfavorable outcome of 7 Euro both with respect to itself (yielding neither gain nor loss) and with respect to the favorable outcome of 27 Euro (yielding a loss of 20 Euro). Of course, the exact reverse reasoning holds for the evaluation of the favorable outcome, which provides zero gain-loss utility when compared to itself and a gain of 20 Euro when compared to the payment of 7 Euro. Since losses loom larger than gains, expecting to face a risky lottery and incorporating this expectation in the reference point thus necessarily leads to negative expected gain-loss utility. In contrast, expecting a safe outcome of, say, 15 Euro entails gain-loss utility of zero, since expectations are always exactly met.

Second, only expectations held *in the recent past* (as opposed to current expectations) are assumed to be relevant for the reference point; in other words, Kőszegi and Rabin postulate that reference points adjust to new information only with a lag. Note that for the evaluation of the lottery and the safe outcome in the above example it is important whether or not the reference point adapts. If there is only a short time-span between receiving information about the lottery and the resolution of uncertainty, the reference point will not adapt to new information. Then, the payments of the lottery are compared to the expectations previously held. The same holds for the safe outcome such that both alternatives in the above example are compared to the same referent.

Clearly, an individual's choices alter the distribution over future outcomes and thus her expectations. By our above arguments, preferences for a lottery depend on whether or not the reference point adapts to the new expectations (following from choice) before the outcome of the lottery is revealed. In particular, the Kőszegi-Rabin model predicts that individuals should be more likely to prefer a safe payment over a lottery in situations in which the referent adapts compared to situations in which the referent does not adapt. As we have argued above, this is because the expectation of a lottery always leads to negative expected gain-loss utility.

In Chapter 2 we ask whether individual choice behavior is consistent with this prediction. We analyze choice behavior of subjects in an experiment in which they had to decide between a risky lottery and a safe payment. In a between subject design, we vary the length of the lag between the time at which subjects make their choices and the time at which the uncertainty for the lottery is resolved. In one treatment, individuals learn the outcome of the lottery immediately after their choices; hence, reference points have only very little time to adapt. In a second treatment, the outcome of the lottery is revealed not until 24 hours after choices have been made, providing ample time for the reference point to absorb new information.[2]

In contrast to the predictions of the Kőszegi-Rabin model, we find that the time structure of the decision problem does not affect individual preferences for the lottery: the share of

[2]In a third treatment, we vary the timing of information about available alternatives relative to the timing of uncertainty resolution. Kőszegi and Rabin predict that this affects choice behavior in a specific way compared to the first two treatments discussed above. We refer to Chapter 2 for a detailed description and discussion of this treatment.

subjects choosing the lottery over the safe payment remains constant across treatments. Nevertheless, our findings could be consistent with a theory of expectations-based reference points in which referents adapt almost immediately to new expectations. Such a theory may be appropriate in experimental contexts in which the "attachment" of subjects to previously expected outcomes is potentially weaker than in real-life situations.

While the first two chapters provide empirical tests of behavioral theories for individual decision making under uncertainty, Chapter 3 – which is based on joint work with Robert Ulbricht – studies the dynamics of political systems in a theoretical framework. Within this framework, the interdependence of agents' expectations and policy choices crucially drives political dynamics, as will become clear below.

Chapter 3 is a contribution to the economic literature on political transitions (see, for example, the seminal contributions of Acemoglu and Robinson 2000b, 2001). While the previous literature has focused mainly on single transition events, our work places the dynamic process that describes the evolution of political systems at the core of the analysis. More specifically, we ask the following questions: First, which types of political systems arise endogenously when transitions occur peacefully via reforms or violently via revolts? Second and related, through which transition mechanisms – reforms or revolts – are democratic and autocratic political systems established? And third, how stable are different political systems over time?

Providing answers to these questions requires a model in which the political system that emerges from the transition mechanisms of reforms and revolts is endogenous to the choices of political actors. For revolts, we assume that each agent who is excluded from political power individually decides whether or not to participate in the risky undertaking of revolting. If the revolt is successful, those agents who participated form the new regime. Regimes after revolts are thus determined by the individual decisions of political outsiders. The prospects of a revolt – and hence the incentives to support it – thereby depend, on the one hand, on the fraction of political outsiders supporting it, and, on the other hand, on the internal stability of the regime currently in place. The latter is private knowledge of the ruling elite such that they are better informed than political outsiders about the likelihood of a revolt to succeed.

To alleviate the potential threat of a revolt, members of the regime may reform the political system by enfranchising an arbitrary fraction of political outsiders. A reform thus reduces the mass of outsiders in the population and thereby the number of those who potentially join a revolt. Since the scope of reforms is not fixed, the political system emerging from reforms is *a priori* unspecified.

Our first finding from the model is that reforms always lead to democratic political systems. At the heart of this finding is the signaling value of reforms. Because internally weak regimes have, *ceteris paribus*, a higher incentive to conduct reforms than internally strong regimes, political outsiders expect the regime to be weak when reforms are observed in equilibrium. If the regime is weak, however, the incentives for political outsiders to join the revolt are large. Hence, reforms need to enfranchise a large fraction of outsiders in order to offset the increasing degree of coordination among outsiders (and to effectively reduce revolutionary pressure).

In addition, we show that successful revolts always lead to autocracies. Our model thus predicts that democracies are always established via reforms. This finding is in line with findings from political science, according to which members of the old elites play an important role in the establishment of democracies (e.g. Rustow, 1970). We further demonstrate that democratic political systems are stable because they lack a meaningful opposition of outsiders; in contrast, autocracies are instable due to being frequently overthrown by revolts or succeeded by democracies after reforms. Nevertheless, given that revolts in general result in autocracies similar to their predecessors, autocratic political systems persist over time. Our model thus predicts a long-run distribution of political systems with mass concentrated on extreme political systems, which mirrors the empirically observed distribution.

Chapter 1

Projection Bias under Risk*

1.1 Introduction

In many economic decision problems, the utility from choice materializes in the future such that individuals need to predict their future utility in order to make informed decisions. While standard economic models assume that individuals predict their utility correctly, Loewenstein et al. (2003) argue that individuals make systematic errors. Specifically, people tend to underestimate the extent to which changes in the state of the world alter utility. Hence, predicted future utility (at unknown future states) is biased towards utility at today's state. Loewenstein et al. call this error *projection bias*.

There is accumulating evidence that projection bias affects economic decisions like house, car, and apparel purchases (Conlin et al., 2007; Busse et al., 2012) or college choice (Simonsohn, 2010). For example, in parallel work Busse et al. demonstrate that sales of 4-wheel drive vehicles increase by 6 percent after a snowstorm, that is at times when the weather-related utility from owning a 4-wheel drive is very high. All these papers have in common, however, that the weather-related dimension of utility, which serves as the testing ground for projection bias, is most likely not of primary importance for decision makers. If this is the case and individuals devote only limited attention to predicting the weather-related dimension of utility for available alternatives, they may be more prone to make errors in that dimension.[1] Therefore, it remains an open question whether people are able to overcome projection bias when their attention is drawn to the state-dependent nature of utility.

In this study, we test for projection bias in a situation where state-dependent utility is

* This chapter is based on joint work with Thomas Kolaska.

[1] See Schwartzstein (2012) for theoretical and Hanna et al. (2012) for empirical evidence on how limited attention may lead to errors.

expected to be at the center of decision makers' attention: We study online advance sales for an outdoor movie theater. In this context, projection bias predicts that good weather on the purchase-date leads customers to overvalue their utility from visiting the theater on the movie-date. Hence, the number of advance sales for the theater should increase if purchase-date weather is good.

The state-dependent nature of utility in this setting is salient for a number of reasons. First, the presence of risk when buying tickets in advance highlights the possibility of facing a different state of the world in the future. Customers face risk when deciding whether or not to buy tickets today for an outdoor movie night in the future because only movie-date weather – as opposed to purchase-date weather – affects utility. This risk is obvious because tickets are only valid for one particular show and are non-refundable (tickets are a perishable good). Additionally, the ticketing website points out the risk in a clear way by stating: "The show is going to take place regardless of weather conditions. (...) You have to pay for your tickets even if you do not collect them."[2] We demonstrate that there is considerable risk because weather at the location of the theater is highly variable.

In addition to risk, two further characteristics of the decision problem are expected to de-bias potential customers. First, the weather-related dimension of utility is a very important component of total utility derived from the movie night – in a survey, the majority of customers states that weather is at least as important as the movie shown.[3] Since weather is important to customers, they should devote a considerable amount of attention to predicting weather-related utility correctly. Second, when considering to purchase tickets few days in advance, customers can condition their decision on reliable, unbiased, and free information provided by weather forecasts to overcome projection bias.

Contrary to our conjecture that potential customers are by and large de-biased, we find that variations in purchase-date weather explain variations in advance sales to a large degree, controlling for the weather forecast. Across different time horizons – with the number of days tickets bought in advance ranging from one day up to three weeks – a one standard deviation increase in sunshine duration leads to an increase in sales between 10 and 25 percent on average. Our findings are robust to considering different subsets of customers. Notably, the results do not change when we consider the behavior of customers with prior bad experiences (defined as rainfall during a previous show they purchased tickets for). The dependence of ticket orders on current weather is thus prevalent for customers who had the possibility to learn from previous mistakes.

[2] Authors' translation from https://www.didax.de/kms/index.php [4 October 2012].

[3] We conducted a survey at the theater on a total of 13 nights, interviewing 443 customers. For details, see Section 1.3.3.

We rule out a number of alternative explanations for this finding. First, we show that purchase-date weather has at most negligible predictive power for movie-date weather, ruling out the possibility that current weather is an informative signal for future weather.

Second, we investigate whether the positive effect of purchase-date weather on aggregate sales merely reflects an increase in the number of potential customers who consider visiting the theater as an attractive leisure activity without affecting individual decisions directly. This may be the case, for example, if good current weather reminds people of the possibility to visit the theater. We use a strategy similar to Conlin et al. (2007) to distinguish between the latter explanation and projection bias by looking at the decision to collect the tickets that have been purchased in advance. If, on the one hand, projection bias affects individual purchase decisions, utility of customers is upward biased at times of good purchase-date weather. Then, tickets are mistakenly purchased with a higher likelihood. We therefore expect that the likelihood that customers let their tickets expire increases with better purchase-date weather if projection bias affects decisions. If, on the other hand, individual decisions are unbiased and purchase-date weather solely affects the aggregate number of potential customers, there should be no effect on tickets collected. We find a negative effect of purchase-date weather on the probability that tickets are collected, providing further evidence for projection bias.

Third, we argue that weather-related market interactions cannot explain why sales depend on purchase-date weather. In particular, there may be a "precautionary" rationale for purchasing tickets at times of good weather as good weather may increase the perceived probability for the theater to sell out in advance. However, this seems unlikely to be the sole explanation for our findings for two reasons. On the one hand, sales well in advance of the movie-date – when the probability for the theater to sell out is essentially zero – are also weather-dependent. On the other hand, we show that hourly variations in weather explain hourly changes in ticket sales. This is in line with projection bias, but does not fit an explanation based on "precautionary" purchasing motives because the perceived probability of the theater to sell out is unlikely to vary with hourly changes in weather.

By showing that projection bias affects individual decisions even in situations in which the state-dependence of utility is particularly salient, our study complements the emerging literature on projection bias in economics discussed above. In addition to this literature in economics, there is a number of papers in psychology providing evidence for projection bias. This literature deals mostly with how current visceral states – for example hunger or sexual arousal – affect decision making.[4] See Loewenstein and Schkade (1999) for an overview.

[4]See for example Loewenstein (1996), Loewenstein et al. (1997), Read and van Leeuwen (1998), van Boven and Loewenstein (2003), and Nordgren et al. (2007).

Furthermore, it is important to note that projection bias is observationally equivalent to agents holding subjective beliefs that assess the current state of the world to be more likely in the future (this has been pointed out by DellaVigna, 2009).[5] There is some evidence for agents holding these types of beliefs, which Fuster et al. (2010) call “extrapolation bias”. For example, several papers in behavioral finance find that individuals tend to choose assets with high current returns more frequently even if current returns do not predict future ones (Benartzi, 2001; Kaustia and Knüpfer, 2008; Barber et al., 2009; Choi et al., 2009). Similar evidence comes from the literature on heterogeneous expectations (see Hommes, 2011, for an overview of the literature). For example, Chavas (2000) estimates that 47 percent of cattle producers use the current price as proxy for future prices when planning future supply, despite large fluctuations in price over time (widely known as “hog cycle”).

The remainder of the chapter is structured as follows. In the next section we develop a simple model and derive predictions regarding how current weather may affect advance sales and the subsequent decision of customers whether or not to visit the theater. In Section 1.3 we describe the data in greater detail. Section 1.4 discusses our main empirical findings. In Section 1.5 we evaluate alternative explanations for our findings as well as their robustness. Section 1.6 concludes.

1.2 A Simple Model and Hypotheses

To fix ideas, this section provides a simple model of individual purchase decisions as well as aggregate purchasing behavior. The model nests rational behavior as well as projection bias and the “reminder-effect” of weather, where the latter two are models of how the current state – weather on the purchase-date – may affect individual choices and total sales. From the general model, we derive testable predictions to distinguish between rational behavior and the two potential explanations for weather-dependent individual decisions and sales.

1.2.1 Individual Purchase Decisions and Aggregate Sales

Individual Decisions Survey results indicate that weather is an important determinant of the utility derived from an outdoor movie night.[6] Overall, 81 percent of respondents state that dry weather is “very important” or “important” for having a good night at

[5] Recall that projection bias is a mistake in predicting utility at unknown future states. The beliefs regarding the likelihood of each state are assumed to be correct.

[6] See Section 1.3.3 for a description of the survey.

the movie; comfortable temperatures are of importance for 66 percent. In our model, therefore, each customer derives weather-related utility $u(w_\tau)$ when watching a movie on date τ given weather conditions $w_\tau \in \mathbb{R}$.[7] The utility function $u(\cdot)$ is assumed to be increasing, twice differentiable, and concave on the real line. When not visiting the theater, individuals receive utility $u(\eta)$ from a heterogeneous outside option $\eta \in \mathbb{R}$, which is distributed within the population according to the distribution $G(\cdot)$.

On the purchase-date $t < \tau$, an individual decides whether or not to buy a ticket for the movie-date at costs c (in utility terms). On the purchase-date, the realization of weather on the movie-date is uncertain. We denote the distribution of w_τ at t by $H(\cdot)$, which is known to potential customers. We assume that $H(\cdot)$ belongs to the location family of distributions with location parameter f_t and is independent of actual weather w_t (we will justify this assumption empirically in Section 1.5.1).[8] The parameter f_t denotes the weather forecast at t for the movie-date τ, which is available to individuals free of charge. The forecast predicts expected weather on the movie-date and contains all relevant information regarding movie-date weather at t: $E[w_\tau | \, f_t] = E[w_\tau | \, f_t, w_t] = f_t$, where E is the expectations operator with respect to $H(\cdot)$.

To incorporate projection bias in our model, we adopt the formulation of "simple projection bias" (Loewenstein et al., 2003) and assume that the current state – in our case current weather – receives weight $\alpha \in [0, 1]$ in an agent's expected utility function. Clearly, the case $\alpha = 0$ represents fully rational behavior. The case $\alpha > 0$ captures that individuals cannot fully assess the extent to which a change in the state of the world will alter their utility and thus unconsciously anchor their utility on the current state.[9]

Expected utility from purchasing a ticket on the purchase-date for an individual with outside option η is then given by

$$v^B(f_t, w_t, \eta) = (1 - H(\eta)) \Big((1-\alpha) \, E[u(w_\tau) | \, w_\tau \geq \eta, f_t] + \alpha u(w_t) \Big) + H(\eta) \, u(\eta) - c. \quad (1.1)$$

A customer who owns a ticket will only visit the theater if movie-date weather exceeds the outside option ($w_\tau \geq \eta$). In this case, captured by the first term of (1.1), she receives weather-related (expected) utility from visiting the theater, which may have excessive

[7]For simplicity, the model abstracts from potential explanations for ticket orders different from weather such as the popularity of a movie. In the setting we are analyzing, these factors are orthogonal to purchase-date weather such that omitting them in this analysis does not alter the empirical implication of the model. We control for popularity of the movie in one of the robustness checks in Section 1.5.3.

[8]In practice, $H(\cdot)$ would depend on the forecast horizon $\tau - t$ and the season of the year as well. Considering these factors does not change the analysis. To ease notation, we therefore omit them here.

[9]As mentioned in the introduction, this interpretation is equivalent to individuals holding beliefs about the distribution of future states which are unconditionally biased towards the current state.

weight on the current state. If movie-date weather turns out to be unexpectedly bad ($w_\tau < \eta$), she will let her ticket expire and choose the outside option instead. In either case, she has to bear the ticket costs c.

Clearly, an individual with outside option $\bar{\eta}$ will be indifferent between buying and not buying a ticket on the purchase-date iff

$$F^P(f_t, w_t, \bar{\eta}) \equiv v^B(f_t, w_t, \bar{\eta}) - u(\bar{\eta}) = 0 \tag{1.2}$$

A natural candidate for optimal choice behavior is that all individuals with low outside options $\eta \leq \bar{\eta}$ buy tickets on the purchase-date and all individuals with high outside options $\eta > \bar{\eta}$ do not. Lemma 1.1 below states that optimal choices can indeed be completely described by a unique $\bar{\eta}$ satisfying (1.2). Before stating the lemma, however, we need to assume sufficient conditions for a unique fixed point to exist.

Assumption 1.1.

(i) For all f_t there exists an η satisfying $F^R(f_t, \eta) \equiv (1 - H(\eta))\Big(E[u(w_\tau)|\, w_\tau \geq \eta, f_t] - u(\eta)\Big) - c = 0$.

(ii) The hazard rate of $H(\cdot)$, $h(w)/(1 - H(w))$, is weakly increasing.

Assumption 1.1 (i) ensures that there is at least one potential customer, who, given the optimal use of information, would be indifferent between buying a ticket on the purchase-date and not buying a ticket at all. This ensures existence of a fixed point of (1.2). Assumption 1.1 (ii) is the monotone hazard rate assumption, which provides a sufficient condition for uniqueness of the fixed point and holds for a variety of frequently used distributions like the normal and uniform distributions. Given this, we can state the following lemma:

Lemma 1.1. *Suppose Assumption 1.1 holds. Then, for each (f_t, w_t) a unique $\bar{\eta}$ satisfying (1.2) exists.*

All proofs are relegated to Appendix A.1. A direct implication of the above lemma is that there is always a positive probability, $G(\bar{\eta}) \in (0, 1)$, that some customer will buy a ticket on the purchase-date.

Aggregate Sales Given the individual propensity to buy a ticket, expected aggregate sales depend on the total number of potential customers. Here, we incorporate the idea in our model that good weather makes the choice option "outdoor movie theater" more salient and thus enlarges the customer base. One possible interpretation is that customers face cognitive restrictions regarding the number of choice options they can consider at a given time. For this reason, they consider a choice option only if it "comes to mind", which is supposed to be positively related to its attractiveness at the current state.[10]

If the number of potential customers is weather-dependent, ticket sales may be driven by weather even if individual decisions to buy tickets are fully rational. To allow for this explanation in our model, we assume that the number of potential customers $n(w_t)$ is increasing in purchase-date weather w_t. The expected total number of sales on purchase-date t is thus given by $y(f_t, w_t) = n(w_t)\, G(\bar{\eta}(f_t, w_t))$. If customers are fully rational – that is, they have all choice options in mind at all times –, $n(\cdot)$ is independent of w_t.[11]

1.2.2 Hypotheses

Our empirical analysis in Section 1.4 is guided by testable predictions derived from the model. Our first hypothesis deals with the effect of purchase-date weather on sales.

Hypothesis 1.1. *If customers are rational ($\alpha = 0$ and $\partial n(w_t)/\partial w_t = 0$) advance sales are independent of purchase-date weather. Otherwise, sales increase when purchase-date weather is good.*

If we reject the implications of rational behavior in our data – if variations in purchase-date weather explain variations in advance sales –, our model assumes that this effect can be explained by projection bias or a reminder-effect of current weather. For this to be the case, we expect customers to be unaware of the limitations underlying their choices – otherwise, they could adopt strategies to arrive at optimal choices nevertheless. This conjecture provides a plausibility test for our model: If customers are indeed unaware of the impact on weather on their purchase decisions, sales should be affected by purchase-date weather regardless of customers' past experiences or the time horizon between purchase-date and movie-date.

Furthermore, we derive testable predictions to disentangle whether the current state w_t affects individual decisions via projection bias or whether purchase-date weather solely affects the total number of potential customers. Since we do not observe individuals who abstain from buying a ticket, we answer this question by examining the individual decision

[10] Another possible interpretation for a weather-dependent customer base is that good weather at the purchase-date facilitates the coordination of larger groups.

[11] A third possible explanation for a positive relation between good weather on the purchase-date and sales is that customers expect the theater to be sold out with higher probability when purchase-date weather is good. We discuss this potential explanation theoretically after Hypothesis 1.2 and empirically in Section 1.5.2.

to collect paid-for tickets on the movie-date. Our model predicts that individuals buy a ticket if their outside option is worse than $\bar{\eta}$ and to collect it if movie-date weather is sufficiently nice ($w_\tau > \eta$). The probability that a customer collects her ticket is therefore given by

$$\Pr(\text{collect} \mid \text{ buy}) = 1 - \frac{\Pr(w_\tau < \eta < \bar{\eta})}{\Pr(\eta < \bar{\eta})} = \begin{cases} G(w_\tau)/G(\bar{\eta}) & \text{if } w_\tau < \bar{\eta} \\ 1 & \text{else.} \end{cases} \tag{1.3}$$

If individual purchase decisions are affected by current weather, good weather increases the expected utility of buying tickets in advance and thus leads to a higher $\bar{\eta}$. Since the realization of movie-date weather w_τ is independent of purchase-date weather w_t, the likelihood that a customer prefers her outside option therefore increases if purchase-date weather was nice. In contrast, if current weather has no effect on individual decisions (but only on the aggregate number of customers), the likelihood of ticket collection is expected to be independent of purchase-date weather. The following hypothesis summarizes this argument.

Hypothesis 1.2. *If customers are rational (*$\alpha = 0$ *and* $\partial n(w_t)/\partial w_t = 0$*) or if current weather increases the pool of potential customers (*$\partial n(w_t)/\partial w_t > 0$*), the probability that tickets are collected is independent of purchase-date weather. Otherwise – if individual decisions are affected by projection bias (*$\alpha > 0$*) and if movie-date weather is worse than expected – the probability that tickets are collected decreases when purchase-date weather was good.*

Before we continue, it is important to point out a few assumptions upon which our model and hypotheses rest. First, as noted above, we assume that purchase-date weather has no information value for movie-date weather. Otherwise, our results could be explained by customers taking current weather as informative signal. We show in Section 1.5.1 that purchase-date weather is indeed not informative. Nevertheless, individuals could perceive current weather to be informative for the future. As discussed in the introduction, we cannot rule out this explanation if individuals perceive purchase-date weather to be informative regardless of the time span between purchase-date and show. However, it is natural to assume that the perceived information content of current weather is declining in the time horizon one is trying to predict. Then, we would expect that the effect of purchase-date weather on sales becomes weaker with increasingly long horizons. In Section 1.5.2 we will see that this is not the case.

Finally, to keep the model simple, we have abstracted from the fact that potential customers essentially face a dynamic problem when they decide on which date they would like to buy their tickets. Clearly, the timing of buying tickets can be affected by purchase-date weather, for example if the latter affects the perceived probability that the theater may sell out. We discuss this potential alternative explanation in more detail in Section 1.5.2.

1.3 Data

Our data comes from four different sources. An outdoor movie theater located in Munich, Germany, provided us the record of their online advance tickets sales platform. The Meteorological Institute of the University of Munich shared their detailed data on the weather conditions in Munich with us; the local weather forecast was collected from the archives of the newspaper "Süddeutsche Zeitung", a high quality newspaper located in Munich. Finally, we conducted a survey among visitors of the theater at 13 different nights of the 2011 season.

1.3.1 Weather and Forecast

We collect data on weather and weather forecasts for the months June to August of the years 2004 to 2011, which are the times at which the theater screens and for which we have sales data (for details, see below).

The Meteorological Institute of the University of Munich provides us hourly measures for precipitation (measured in 1/100 mm), temperature (measured in degrees Celsius) as well as the average sunshine duration (in percent) between 8 am and 7 pm in Munich.[12] Most statistical inferences uses daily averages (24 hours) of these three weather variables.

We hand-collect the weather forecast from the archives of the daily newspaper "Süddeutsche Zeitung", which is published every day except Sundays and public holidays.[13] It provides a regional forecast for each day, one to four days in advance, for the South of Bavaria including Munich. The forecast comprises of forecasted maximum and minimum temperature (in degrees Celsius) and one of the following weather symbols: sunny, partly sunny, shower, rain, and scattered thunderstorms.[14]

The weather in Munich is highly variable, especially during the summer months, when

[12] The latter restriction ensures that the changing times of dusk and dawn do not confound our measure of sunshine duration.

[13] Weather forecasts take up a lot of memory capacity which is why they are not stored by any German weather firm.

[14] There are in total 12 observations of the symbol overcast, which we group with "shower" to simplify the exposition of results. Undoing this grouping does not lead to any significant changes throughout.

Table 1.1: Summary Statistics: Weather and Forecast

	Weather		
	All day	Evening	SD within day
Avg. Sunshine Duration	54.32	47.86	18.73
	(34.99)	(38.06)	(13.96)
Avg. Temperature	19.01	19.15	3.01
	(3.60)	(3.99)	(1.28)
Avg. Rainfall	11.23	20.27	28.38
	(24.81)	(63.16)	(62.83)
	Forecast		
	Minimum	Maximum	
Forecasted Temperature	12.66	23.56	
	(2.75)	(4.02)	

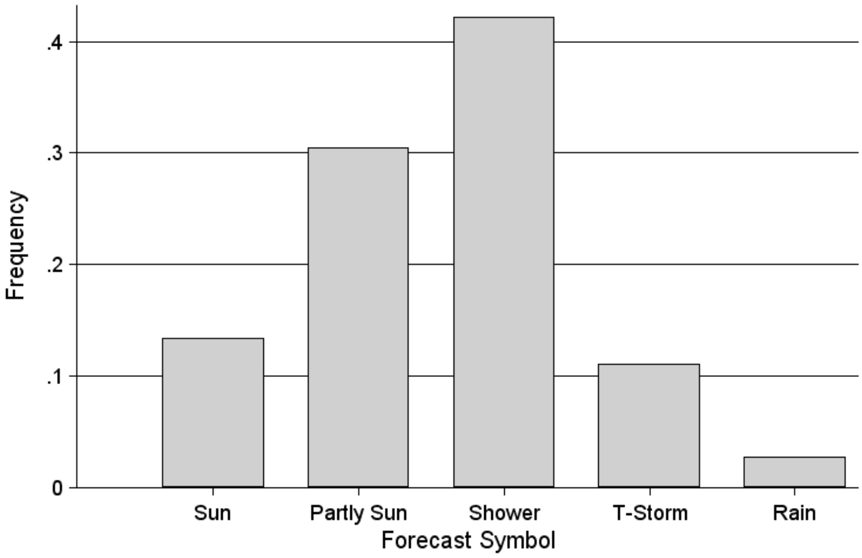
Notes: We report the means of variables; their standard deviations are in parentheses. Sunshine duration is mesured in percent per hour, temperature is measured in degrees Celsius, and rainfall is reported in 1/100 mm per hour. In the column "SD within day" we report the average of the variables' standard deviations within a single day.

there are daily shows at the movie theater. This is mostly due to the proximity of the Alps, which leads to frequent and often unexpected rainstorms. These tend to occur especially in the evening hours. For this reason, there is high monthly precipitation in the summer months when total precipitation is on average 123 mm per month (for comparison: London 51 mm, New York City 92 mm, and Berlin 61 mm). Long periods of stable good weather are the exception; rather, there are frequent shifts in weather patterns every few days as reflected by the mean number of 12.4 rain days per month (days with at least 1 mm of rain) during the summer months (for comparison: London 10.5 days, New York City 8 days, and Berlin 8.7 days).[15]

The weather varies within as well as across days. This can be seen in Table 1.1 where the summary statistics of average daily weather are depicted in the first two columns. Standard deviations of sunshine duration as well as rainfall are high compared to their respective means; the coefficient of variation for sunshine duration is 1.55 and for rainfall 0.45. Furthermore, it is noteworthy that rainfall in the evening hours is considerably higher than during the day reflecting the higher likelihood of rainstorms at these times. The third column of Table 1.1 provides information about the variation of weather within days by depicting the mean of within-day standard deviations of the respective weather variable. Note that both sunshine duration and precipitation exhibit high within-day variation. The within-day variation for temperature is not very informative, as there is a

[15]Sources of long term monthly averages: World Meteorological Organization http://worldweather.wmo.int/ [4 October 2012].

Figure 1.1: Distribution of Forecasted Weather (Symbols)
This figure plots the distribution of forecast symbols pooling over forecast horizons (one to four days in advance).



cyclical pattern of temperature over the course of each day. (Keep in mind that sunshine duration is only measured between 8 am and 7 pm such that darkness does not contribute to the within-day variation of sunshine duration.)

Regarding the forecast, note first that average forecasted temperatures are in a similar range as average temperatures (Table 1.1), which is what we expect. The distribution of weather symbols for all forecast horizons – as shown in Figure 1.1 – again reflects the high variations in local weather across days.[16] Note furthermore that the forecast frequently predicts scattered thunderstorms and showers, which indicates rather unstable weather conditions within days as well.

1.3.2 Ticket Sales

The data on advance ticket sales were provided by “Kino, Mond und Sterne” [Movies, Moon, and Stars], one of four outdoor movie theaters in Munich. The theater usually screens daily during the months of June, July, and August, and shows the movie regardless of weather conditions. The latter fact is important for our study because it implies that tickets bought in advance are non-refundable. A consumer who buys a ticket for this theater in advance thus bears the full weather risk.[17] Customers are expected to be aware of this risk, as it is explicitly mentioned prominently on the ticketing website.

[16]The distributions of symbols separately by each forecast horizon (one to four days in advance) do not differ substantially.

[17]None of the seats are covered. See Appendix A.3 for a picture of the theater.

The theater has a total of 1,300 seats available, tickets for which are sold at the box office and various advance ticket sales locations. The majority of advance tickets are sold online where tickets for a particular show are available until 6 pm on the day of the screening. Our goal is to explain these advance sales such that our main data set comprises of all online ticket orders for the theater between 2004 and 2011. This amounts to a total of 20,999 orders.[18] For each order, the system records the number of tickets bought, the exact date of the transaction and a unique alphanumeric customer ID, which allows us to track repeat customers.

Additionally to the data on online sales for the years 2004 to 2011, we have data on the total number of visitors of the theater – including box office sales – for the years 2009 to 2011. This allows us to assess the importance of advance online sales, which amount to 24 percent of the total number of tickets during this period. More than half (almost 60 percent) of online tickets are sold on the day of the show. Our main analysis focuses on sales between one and four days before the show, on which the weather forecast for the movie-date is available. Within this period, 30 percent of online tickets are sold, with percentages declining between one and four days in advance. The remaining 10 percent of online tickets are sold five days or earlier before a show.

Our main variable of interest is aggregate ticket sales on a daily base. More precisely, one observation is the sum of ticket orders on a single day for a specific show. If no tickets are sold on a day at most 23 days before the show, we add an observation with aggregate orders of zero. This results in at least 24 observations for every single movie shown, one for each day between 0 and 23 days out. We construct additional aggregates of ticket sales for robustness checks. For example, we count orders of repeat customers (identified by their unique customer ID) who have bought tickets more than once since 2004. Another noteworthy variable are ticket orders by repeat customers who had previously bought tickets for a show during which it was raining.

For the years 2009 – 2011 we additionally know for each order whether tickets were in fact collected at the evening of the show. Of the total of 4,102 orders, the vast majority (88 percent) of tickets was collected on the movie-date.

The summary statistics for ticket sales are presented in Table 1.2, organized according to how early in advance tickets were sold. The average number of ticket orders decreases from 7 one day in advance to 1 four days in advance, representing the declining pattern of orders. The number of tickets sold per order remains stable at about 2.6, independent of the time horizon. About half of the ticket orders are placed by repeat customers, who have bought tickets online more than once.

[18] Ticket prices have been stable at about 7.40 Dollar (5.70 Euro) each during the entire period.

Table 1.2: Summary Statistics: Ticket Orders

	Day of show	1 day out	2 days out	3 days out	4 days out
Avg. Orders	24.74	7.18	2.78	1.36	0.89
	(33.97)	(10.82)	(4.30)	(2.24)	(1.37)
Tickets per Order	2.46	2.55	2.55	2.63	2.58
	(0.69)	(0.92)	(0.88)	(1.24)	(1.24)

Notes: We report the means of daily ticket orders bought on the day of the show as well as one to four days in advance. For the same five days we additionally provide average numbers of tickets per order. Standard deviations in parentheses.

1.3.3 Survey

During the 2011 season, we conducted a survey among visitors of the cinema. As many visitors spend some time in the theater before the movie starts, the willingness to participate in the survey was high. Overall we received 443 questionnaires for 13 different days with considerable variance in weather conditions (and accordingly varying number of questionnaires obtained per day). This amounts to more than 10 percent of the audience on these days on average. Of all surveyed customers, 25 percent bought their ticket online (compared to 24 percent of all customers in the years 2009 – 2011) and 7 percent purchased it one to four days in advance (compared to 8 percent of all customers between 2009 and 2011). Throughout, we use the survey to provide supporting evidence for our arguments. That being said, none of our main results depends on data from the survey.

1.4 Empirical Analysis

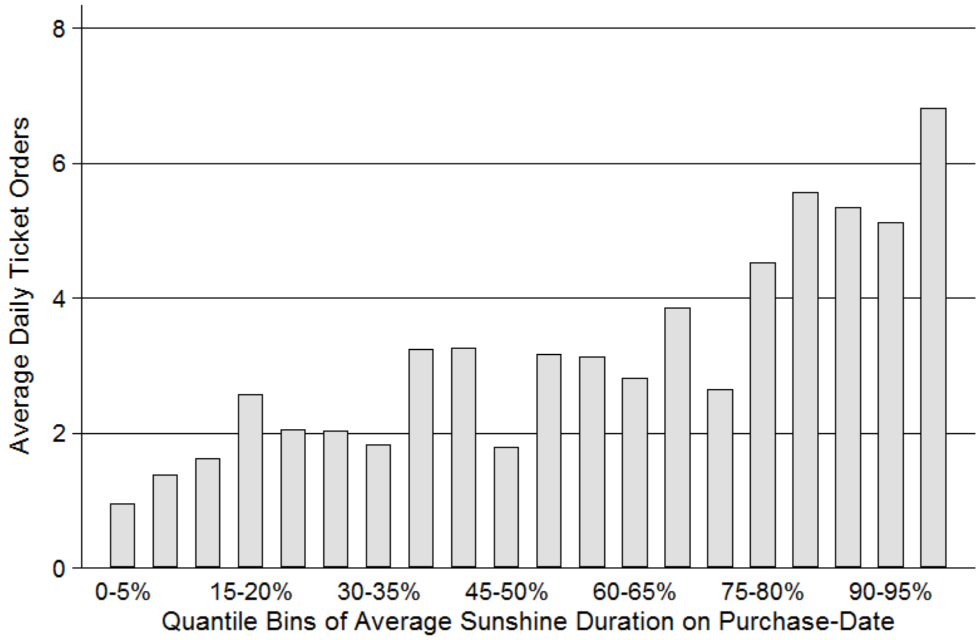
In this section we test the hypotheses derived in Section 1.2. We first show that weather on the purchase-date explains variation in ticket orders for various model specifications, rejecting rational behavior from Hypothesis 1.1. Furthermore, we show that good weather on the purchase-date decreases the likelihood that the purchased tickets are collected on the movie-date, providing evidence for projection bias (Hypothesis 1.2).

1.4.1 Purchase-Date Weather and Ticket Orders

Figure 1.2 illustrates the effect of weather on ticket orders by comparing the number of orders across different weather conditions on the purchase-date. For the sample of ticket orders one to four days ahead of the movie-date – which we use in the empirical analysis below – we group ticket orders into bins based on five percent quantiles of purchase-date

Figure 1.2: Purchase-Date Weather and Ticket Orders

This figure plots the average number of daily ticket orders (between one and four days in advance) for bins based on five percent quantiles of purchase-date sunshine duration. Bins are sorted from dates with shortest sunshine duration (to the left of the horizontal axis) to days with longest sunshine duration (to the right of the horizontal axis).



sunshine duration and plot, for each of these bins, the average number of ticket orders per day. Consistent with projection bias, Figure 1.2 shows that the average number of daily ticket orders strongly rises parallel to an increase in sunshine duration from the left to the right of the horizontal axis.

An obvious concern with the graphical analysis above is that other factors which explain ticket sales – like the weather forecast – may possibly be correlated with purchase-date weather. To address this concern, we estimate the effect of purchase-date weather on the number of daily ticket orders in a number of regressions. In all of these regressions, we include average sunshine duration as well as average precipitation on the purchase-date t as explanatory variables (collected in the weather vector $\mathbf{W_t}$).[19] In addition, we control for the weather forecast at t for the movie-date τ by adding the forecasted maximum and minimum temperatures, as well as separate dummy variables for each forecast symbol as independent variables; these variables are collected in the forecast vector $\mathbf{F_{\tau t}}$. Because the forecast is only available for a horizon Δ of up to four days, we limit the sample to ticket orders between one and four days ahead of the show.

[19]We omit average temperature from our analysis, since it is highly correlated with sunshine duration ($\rho = 0.6$), which makes the analysis of the respective coefficients difficult. We chose to keep sunshine duration for its greater salience compared to temperature. However, our results are not qualitatively affected by this choice. For further details see the discussion in Section 1.5.3.

For the first empirical model we organize the data in a panel structure with the movie-date τ as unit of observation and advance sales one to four days in advance as observations over time. Within this structure the model is

$$y_{\tau t} = \mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{W}} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{D}'_{\tau\mathbf{t}}\beta_{\mathbf{D}} + \upsilon_{\tau t} \tag{1.4}$$

where $\mathbf{D}_{\tau\mathbf{t}}$ includes dummy variables for each time difference between purchases and show. We assume that the error term $\upsilon_{\tau t}$ is iid between different shows τ but may be arbitrarily correlated between advance sales for the same show. To control for unobserved heterogeneity possibly correlated with our regressors, we estimate (1.4) as a fixed effects model.

For the second econometric model we organize our data as cross sections separately for each purchase-date being $\Delta \in \{1, 2, 3, 4\}$ days ahead of the movie-date. This gives us less power due to limiting observations, but allows us to exploit cross sectional variation and to include a set of controls $\mathbf{X}_{\tau\mathbf{t}}$ which are for most observations time invariant between t and τ. Specifically, we control for the day of the week of the show τ, average sunshine duration and precipitation of the past two weeks before t, as well as dummy variables for year and month. For each Δ, we estimate the following model:

$$y_{\tau t} = \mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{W}} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{X}'_{\tau\mathbf{t}}\beta_{\mathbf{X}} + \varepsilon_{\tau t}. \tag{1.5}$$

Since $\Delta = (\tau - t)$ is fixed, there is a single observation for each movie night τ. Imposing the identifying assumption from above – that errors $\varepsilon_{\tau t}$ are uncorrelated across movie-dates τ – we can estimate (1.5) by OLS.

Table 1.3 displays the estimation results for the two models. Similar to the graphical analysis in Figure 1.2, the results provide strong support for Hypothesis 1.1: the effect of sunshine duration on aggregated ticket sales is positive and significant throughout. In the fixed effects model (1.4) and the cross sections (1.5) for one and three days in advance, average rainfall has furthermore a negative effect on ticket sales, which is significant at least at the ten percent level. Moreover, as predicted by the theoretical model, we find significant effects of the weather forecast on sales at least for one to three days out. This is especially true for forecasted temperature. Forecast symbols seem to have an effect one and two days out, only. In the fixed effect model (1.4) we find no statistically significant effect of these symbols at all, which may mostly be explained by their limited within variance.

In order to interpret the estimated parameters of the variables of interest and to compare their impact across different advance sales horizons Δ, we calculate the statistic $m(x) =$

Table 1.3: Effect of Purchase-Date Weather on Ticket Orders

	Daily Ticket Orders				
	Fixed Effects	1 day out	2 days out	3 days out	4 days out
Avg. Sun	0.023***	0.027**	0.016***	0.0061*	0.0046**
	(0.0054)	(0.014)	(0.0054)	(0.0032)	(0.0021)
Avg. Rain	−0.014**	−0.028**	0.013	−0.0059*	0.0020
	(0.0063)	(0.014)	(0.012)	(0.0036)	(0.0033)
Forecasted Maxtemp.	0.19**	0.68***	0.26***	0.15***	0.053*
	(0.072)	(0.17)	(0.084)	(0.041)	(0.027)
Forecasted Mintemp.	−0.014	0.61**	0.12	−0.0063	0.019
	(0.090)	(0.24)	(0.12)	(0.056)	(0.030)
Symbol Partly Sunny	0.30	−4.05*	−2.31**	0.32	−0.55**
	(0.79)	(2.19)	(0.94)	(0.40)	(0.24)
Symbol Shower	−0.21	−6.92***	−3.30***	−0.17	−0.46*
	(0.74)	(2.23)	(0.90)	(0.37)	(0.24)
Symbol Rain	−1.13	−8.09***	−3.19***	0.49	−0.61
	(1.02)	(2.62)	(1.14)	(0.59)	(0.50)
Symbol T-Storm	−0.53	−10.7***	−1.95*	0.34	−0.12
	(0.81)	(2.32)	(1.04)	(0.47)	(0.36)
2 Days Out	−4.44***				
	(0.38)				
3 Days Out	−5.94***				
	(0.47)				
4 Days Out	−6.43***				
	(0.49)				
Time-invariant Controls	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1635	413	411	406	405
Adjusted R^2	0.282	0.350	0.306	0.205	0.181

Notes: We report the coefficients and robust standard errors of OLS regressions of total daily ticket orders on purchase-date weather, forecast, and control variables. In the first column, the sample consists of all daily ticket orders between one and four days before the show (show-date fixed effects and horizon dummies included). In the remaining columns, the sample is split according to the number of days tickets are purchased in advance. "Avg. Sun" is the average sunshine duration in percent on the purchase-date between 8 am and 7 pm; "Avg. Rain" denotes average rainfall on the purchase-date in 1/100 mm. Forecasted temperatures are from the forecast at the purchase-date for the movie-date and measured in degrees Celsius. The variable "Symbol Partly Sunny" takes the value 1 if the forecast for the movie date is partly sunny on the purchase-date and 0 otherwise. Other symbol-variables are defined accordingly; the baseline forecast symbol is sunny.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

$\beta(x)\, s(x)/\bar{y}_\Delta$ for the estimated models. The nominator of $m(x)$ is the product of the coefficient $\beta(x)$ of an independent variable x and its standard deviation $s(x)$, which gives us the impact of one standard deviation change of x on the number of advance sales. To compare this effect across different advance sales horizons Δ, we normalize it by the mean of the respective number of advance sales $\bar{y}_\Delta$. Thus, the statistic $m(x)$ denotes the impact of a one standard deviation change of x on sales as percentage of mean sales for a given empirical specification.

In the estimated model, one standard deviation change of actual sunshine duration leads to a change in sales between 10 and 25 percent of the mean. In comparison, one standard deviation of the forecasted temperature has an effect on sales between 10 and 40 percent of the mean – the effects of these two determinants of sales are thus of comparable size. This leads to our first result.

Result 1.1. *Purchase-date weather has a statistically and economically significant effect on aggregate ticket orders.*

We further investigate the conjecture from Section 1.2.2 that the effect of weather on sales is independent of customers' past experiences to address the concern that the results are driven by inexperienced customers.[20] To this end, we estimate the fixed effects model (1.4) replacing total advance sales on a given day by sales to three different subsets of repeat customers as the dependent variable.

The results of this exercise are reported in Table A.1 in Appendix A.2.[21] Average sunshine duration has a positive and highly significant effect on sales to the sub-population of repeat customers who had bought tickets at least once between 2004 and 2011. The same holds for the subset of customers with multiple visits per season. Ticket orders by customers who had previously bought tickets for a show during which it was raining can also be explained by variations in sunshine duration on the purchase-date. The economic significance as measured by statistic $m(\cdot)$ is in the same range for both forecast and purchase-date weather as the estimates for the complete set of customers. These results suggest that decisions of experienced customers are influenced by current weather even if weather at previous visits turned out to be bad.

[20] We investigate the additional conjecture that the effect of weather is independent of the time horizon between purchase-date and movie-date in Section 1.5.2

[21] Although we only report the results from the fixed effect model (1.4), similar results are obtained from model (1.5). The results can be obtained from the authors on request.

1.4.2 Purchase-Date Weather and Ticket Collection

So far, our analysis has focused on explaining aggregate purchase behavior. Using aggregate data, we cannot distinguish whether the current state – purchase-date weather – affects the number of potential customers (through reminding them of their choice options) or whether it affects individual decision making directly (through projection bias). We assess to what extent purchase-date weather alters individual behavior by exploiting information on the individual decision whether or not to collect paid-for tickets on the movie-date.

According to the model in Section 1.2, the likelihood that tickets are collected decreases with good weather on the purchase-date if individual decisions are directly affected. In contrast, if good weather at purchase merely increases the number of potential customers, the decision to actually visit the theater is expected to be independent of purchase-date weather.

Let $\psi_{it\tau} = 1$ denote the decision of a customer i to collect a ticket, which had been purchased on the purchase-date t, at the movie-date τ. (The decision to let the ticket expire is denoted by $\psi_{it\tau} = 0$.) The likelihood that a customer collects her ticket on the movie-date is estimated using the probit model

$$Pr(\psi_{it\tau} = 1) = \mathbf{\Phi}(\mathbf{W}'_{\mathbf{t}}\beta_{\mathbf{Wt}} + \mathbf{W}'_{\tau}\beta_{\mathbf{W}\tau} + \mathbf{F}'_{\tau\mathbf{t}}\beta_{\mathbf{F}} + \mathbf{X}'_{\tau\mathbf{t}}\beta_{\mathbf{X}}). \tag{1.6}$$

Since the model predicts that individual collection decisions depend on the movie-date weather, we include movie-date weather $\mathbf{W}_{\tau}$ on the right hand side of (1.6) additional to purchase-date weather $\mathbf{W_t}$, forecast $\mathbf{F}_{\tau\mathbf{t}}$, and controls $\mathbf{X}_{\tau\mathbf{t}}$ as defined in Section 1.4.1. Since sunshine duration ceases to be a salient indicator for actual weather at night, we add a dummy variable indicating whether tickets were purchased later than 8 pm to the vector $\mathbf{X}_{\tau\mathbf{t}}$. In order to assess the robustness of the estimates, we re-estimate (1.6) without controls $\mathbf{X}_{\tau\mathbf{t}}$.

The first two columns of Table 1.4 report the estimated coefficients from model (1.6) for all customers who had purchased tickets one to four days in advance in the years 2009 to 2011. Extended sunshine duration on the purchase-date tends to reduce the likelihood that tickets purchased in advance are actually collected at the box office. However, the estimated coefficients are not significantly different from zero at any common level. We conjecture that this is due to limited variance of the dependent variable: 93 percent of all customers collect their ticket. In fact, the model predicts that the probability for advance tickets to be collected equals one if the realized weather at the movie turns out to be at least as good as the outside option of the marginal customer. In other words, our model

Table 1.4: Effect of Purchase-Date Weather on Ticket Collection

	Tickets Collected			
	(1) Full Sample	(2) Full Sample	(3) Restricted Sample	(4) Restricted Sample
Avg. Sun	−0.0020	−0.0019	−0.0049**	−0.0049*
	(0.0016)	(0.0017)	(0.0025)	(0.0026)
Avg. Rain	0.00080	0.00062	0.0024	0.0000020
	(0.0023)	(0.0025)	(0.0034)	(0.0037)
2 Days Out	−0.28***	−0.30***	−0.28**	−0.39***
	(0.10)	(0.11)	(0.14)	(0.15)
3 Days Out	−0.084	−0.13	0.057	0.0036
	(0.14)	(0.15)	(0.18)	(0.20)
4 Days Out	−0.52***	−0.61***	−0.40*	−0.46*
	(0.16)	(0.16)	(0.23)	(0.24)
Sun before Film	0.0067***	0.0065***	0.012**	0.016***
	(0.0015)	(0.0017)	(0.0047)	(0.0046)
Rain Film	−0.0011**	−0.00099*	−0.00098*	−0.0018**
	(0.00050)	(0.00056)	(0.00059)	(0.00076)
Temp. Film	0.15***	0.15***	0.17***	0.16***
	(0.020)	(0.021)	(0.025)	(0.028)
Forecasted Maxtemp.	−0.082***	−0.086***	−0.047*	−0.057*
	(0.023)	(0.026)	(0.028)	(0.032)
Forecasted Mintemp.	−0.024	−0.018	−0.065*	−0.061
	(0.028)	(0.029)	(0.036)	(0.039)
Symbol Partly Sunny	0.10	0.094	0.25	0.28
	(0.13)	(0.14)	(0.20)	(0.24)
Symbol Shower	0.0059	−0.012	0.10	0.12
	(0.14)	(0.15)	(0.23)	(0.25)
Symbol T-Storm	0.21	0.22	0.34	0.58**
	(0.17)	(0.18)	(0.24)	(0.27)
No. of Tickets	0.097**	0.10**	0.046	0.051
	(0.043)	(0.044)	(0.043)	(0.044)
Evening	0.19*	0.20**	0.11	0.13
	(0.097)	(0.097)	(0.14)	(0.14)
Controls	*No*	*Yes*	*No*	*Yes*
Observations	2620	2620	874	874
Pseudo R^2	0.220	0.227	0.200	0.218

Notes: We report the coefficients and robust standard errors of Probit regressions explaining individual decisions to collect advance tickets. The first two columns refer to the full sample of all advance sales for the seasons 2009 – 2011. Column (3) and (4) provide estimates for a restricted sample considering only shows during which the weather was worse than expected (see the text for details). Variables indicating purchase-date weather ("Avg. Sun" and "Avg. Rain") and variables of the purchase-date weather forecast ("Forecasted Max(Min)temp.", "Symbol ...") are defined as in Table 1.3. The baseline forecast is sunny. We control for movie-date weather by "Sun before Film" (average sunshine duration between 5 pm and 7 pm in percent), "Rain Film" (rainfall in 1/100 mm between 7 pm and 11 pm), and "Temp. Film" (temperature between 7 pm and 11 pm), as well as for whether the ticket was bought at darkness on the purchase-date (Evening = 1), the number of tickets ordered ("No. of Tickets"), and the number of days tickets were bought in advance ("x Days Out").
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

predicts that customers should decide to let their tickets expire only in cases in which movie-date weather is a negative surprise. Including all customers in our analysis should therefore bias the coefficient for purchase-date sunshine duration upwards.

To test this conjecture, we estimate (1.6) for a sample of movie-dates at which realized weather was a negative surprise. For all movie-dates in the restricted sample, expected sunshine duration $\hat{S}_t$ as predicted by model (1.7) in Section 1.5.1 is greater than realized sunshine duration shortly before the movie starts. As predicted by the theoretical model, almost all customers (97 percent) dropped from the sample chose to pick up their tickets; in contrast, 15 percent of the remaining customers let their ticket expire.

The results from re-estimating (1.6) for this sample are depicted in columns (3) and (4) of Table 1.4. It becomes apparent that the magnitude of the (negative) effect of purchase-date sunshine duration on the likelihood of tickets to be collected doubles at the mean of the sample. This confirms our conjecture that the coefficients in columns one and two are upward biased.[22] Furthermore, the coefficients become statistically significant at the five and ten percent level, respectively. This leads to our second result.

Result 1.2. *Customers are less likely to collect their tickets on the movie-date if they experienced good weather on the purchase-date, providing evidence for projection bias (Hypothesis 1.2).*

1.5 Alternative Explanations and Robustness

In the previous section we have shown that projection bias can account for both, weather-dependent sales and decisions to collect purchased tickets. However, there may exist alternative explanations that could explain these findings as well. In this section, we discuss plausible alternative explanations and the robustness of our results to different empirical specifications.

1.5.1 Is Current Weather Informative for Future Weather?

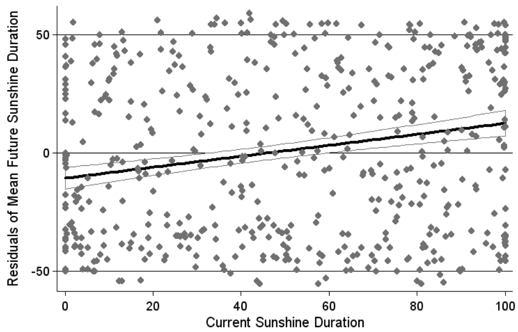
An immediate concern of our analysis so far is that individuals use current weather to update their beliefs about future weather conditions. There are two reasons why this could be optimal. First, current weather may be informative by itself such that looking up the weather forecast is unnecessary. Second, current weather may enhance the prediction of

[22] The estimates from the restricted sample are potentially still upward biased if customers are subject to the sunk cost fallacy (Arkes and Blumer, 1985).

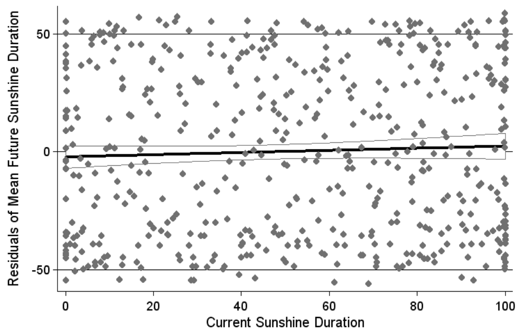
Figure 1.3: Predictive Power of Current for Future Sunshine Duration

This figure provides a scatterplot of current sunshine duration against residuals of a regression of future sunshine duration (1 to 4 days ahead, respectively) on month and year dummy variables. The black solid line depicts the estimates of a regression of residuals on current weather, the grey lines depict the 95 percent confidence interval.

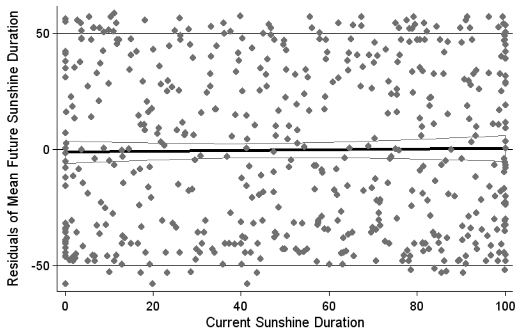
(a) 1 Day Ahead



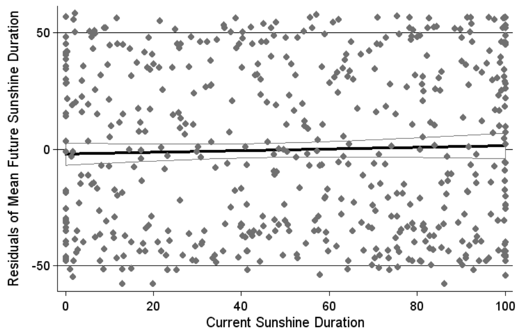
(b) 2 Days Ahead



(c) 3 Days Ahead



(d) 4 Days Ahead



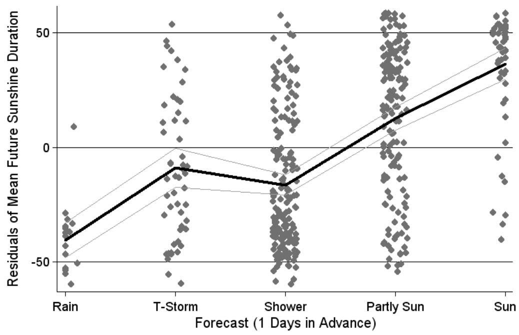
future weather, even given the weather forecast. This may be the case, for example, if the forecast cannot take regional factors into account sufficiently well.

We argue that the information content of current weather for future weather is, in general, limited if not nil due to large day to day fluctuations of local weather in Munich. In Figure 1.3, we plot average sunshine duration one to four days ahead (purged for seasonal effects by year and month dummies) against current sunshine duration (for the summer months). It turns out that tomorrow's sunshine hours are at best slightly positively related to today's sunshine duration. Furthermore, today's weather has no explanatory power for weather two or more days ahead.

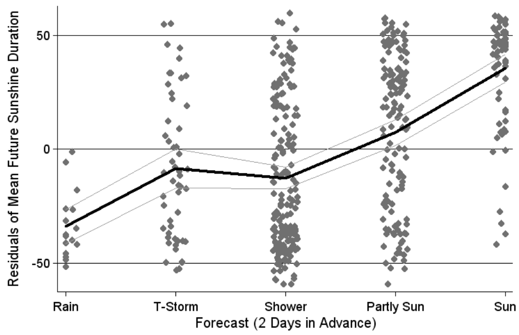
In contrast, the weather forecast is able to explain future sunshine duration well. Figure 1.4 again plots average sunshine duration purged for seasonal effects as above but this time against the forecast as given by forecast symbols. Evidently, there is a clear positive relationship between symbols indicating good future weather and realized sunshine duration. Especially, the symbols "rain", "partly sunny", and "sunny" seem to predict the

Figure 1.4: Predictive Power of the Weather Forecast for Future Sunshine Duration
This figure provides a scatterplot of current forecast symbols against residuals of a regression of future sunshine duration (1 to 4 days ahead, respectively) on month and year dummy variables. The black solid line connects the means of future sunshine conditional on the forecast, the grey lines connect the 95 percent confidence intervals of the conditional means.

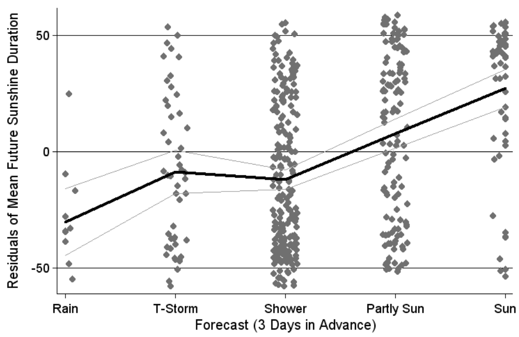
(a) 1 Day Ahead



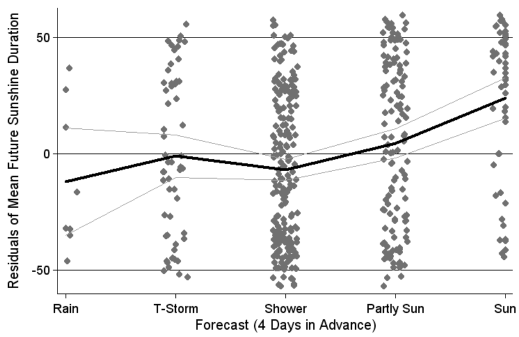
(b) 2 Days Ahead



(c) 3 Days Ahead



(d) 4 Days Ahead



weather quite well, even as far as four days into the future.

In order to reassure that the predictive power of current weather – even when not controlling for the forecast – is low, we complement the graphical analysis above with empirical estimates. In particular, we forecast evening sunshine duration S_h (in percent) at some date $h >$ t with the following model

$$S_h = \mathbf{W_t'}\gamma_{\mathbf{W}} + \mathbf{F_{ht}'}\gamma_{\mathbf{F}} + \mathbf{V_t'}\gamma_{\mathbf{V}} + \xi_{ht} \tag{1.7}$$

where controls $\mathbf{V_t}$ include average sunshine duration and precipitation of the past two weeks before t as well as year and month dummies. Current weather $\mathbf{W_t}$ and forecast indicators $\mathbf{F_{ht}}$ are defined as above. We estimate model (1.7) with and without including the forecast $\mathbf{F_{ht}}$; the results including the forecast are displayed in Table A.2, the results without forecast in Table A.3 in Appendix A.2. Confirming the graphical results, current weather does not help to assess future weather except for one day ahead where the coefficients of current sunshine duration are statistically significant but small (a one percent increase in sunshine duration today leads to an increase in sunshine duration tomorrow of at most 0.21 percentage points). In contrast, the predictive power of the forecast is sizable since adding it to the model leads to a roughly threefold increase in variance explained. To further appreciate the predictive power of weather symbols, note that if the forecast symbol for four days in advance is "shower" instead of "sunny", evening sunshine duration decreases by 70 percent of one standard deviation.[23]

Given this, the question arises whether customers appreciate the predictive power of the forecast. Our survey results indicate that this is indeed the case as customers report to consult the weather forecast frequently and appreciate its reliability. From all respondents, 84 percent consult the weather forecast at least every other day or when they are planning weather-related activities. Regarding forecast reliability, 85 (86) percent state that the forecast for tomorrow (two days ahead) will be correct in at least 80 (60) percent of cases.

Overall, given the above results, it seems unlikely that customers knowingly base their predictions of future weather-dependent utility on actual weather – especially since the vast majority of customers are locals, who should be expected to know the regional weather conditions well.

1.5.2 Probability of Ticket Availability

Another concern is that the ticket sales are driven by capacity constraints. The theater has 1,300 seats such that higher ticket sales at any given point in time lead to a higher risk that the movie may sell out. Thus, if customers believe that the likelihood that tickets will be available on the movie-date decreases with good purchase-date weather, they have a higher incentive to buy on the purchase-date. Such a "precautionary" motive for buying tickets early at times of good weather would shift purchase decisions to earlier dates with good weather, which is a potential explanation for our results.

In fact, the movie theater in question has been sold out in 13 percent of evenings over the entire time span of our analysis, but has, so far, never been sold out in advance. In general, customers seem to understand that they are always able to buy tickets online: 88 percent of customers state that it is "unlikely" or "very unlikely" that all tickets for tomorrow's screening will be sold out in advance.[24]

Some customers may nevertheless perceive the probability of ticket availability to depend

[23]Here, we assess the predictive power of current weather and the weather forecast for the main weather indicator of interest – sunshine duration – only. Repeating the exercise for precipitation and temperature gives similar results.

[24]The likelihood that the theater sells out, however, is empirically unrelated to weather (and the forecast) prior to the movie-date.

Table 1.5: Effect of Purchase-Date Weather on Early Ticket Orders

	Daily Ticket Orders			
	5 - 11 Days Out	9 - 15 Days Out	13 - 19 Days Out	17 - 23 Days Out
Avg. Sun	0.0012***	0.00091***	0.00041	0.00058**
	(0.00043)	(0.00034)	(0.00026)	(0.00023)
Avg. Rain	−0.00046	0.00036	−0.00014	−0.00022
	(0.00043)	(0.00034)	(0.00029)	(0.00021)
Horizon Indicators	*Yes*	*Yes*	*Yes*	*Yes*
Observations	3472	3472	3462	3437
Adjusted R^2	0.035	0.010	0.004	0.002

Notes: Coefficients and robust standard errors are reported for OLS regressions of daily ticket orders on purchase-date sunshine duration (in percent of time), purchase-date rainfall (in 1/100 mm), and horizon indicators (dummy variables for the number of days between purchase-date and movie-date). Fixed effects for the show are included. One observation is the number of sales for a particular show per day. The column headings indicate how many days in advance tickets are purchased in the sample used.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

on current weather few days before the show. However, deferring the purchase decision to a later date should be perceived to be risk-less for particularly early purchase-dates, for instance five days in advance and earlier. Thus, if customers' concerns that the theater may sell out were the sole explanation for the effect of current weather on sales, particularly early ticket orders should be unaffected by purchase-date weather. In contrast, if our results can be explained by projection bias (or a reminder-effect of good weather), we expect to find an effect on early orders as well.

To analyze this prediction, we estimate (a variant of) the fixed effects model (1.4) for separate sets of advance sales which are defined by how many days in advance tickets were sold. More precisely, we estimate the effect of weather on ticket orders between 5 and 11 days in advance. Since weather forecasts for this time horizon are lacking, we cannot include them in these regressions. We repeat this exercise for time spans between 9 and 15, 13 and 19, and 17 and 23 days in advance.[25]

Table 1.5 reports the results. It becomes apparent that the effect of average sunshine duration on sales is significantly greater than zero at least at the five percent level for most estimated models. The notable exception are the results with sales between 13 and 17 days in advance; here, the coefficient of sunshine duration is rather small.[26] Regarding

[25] Obviously, the choice of beginning and end days of these time-spans is arbitrary. However, our qualitative results do not depend on the exact location of the time spans as long as they are sufficiently long (greater than four days) to allow for enough within variation for early sales.

[26] We observe fairly small coefficients for all intervals which include the time horizon of exactly 16 days. Excluding observations for this time horizon leads to significant coefficients throughout, which suggests that this time horizon is an outlier for which we have no plausible explanation.

Table 1.6: Effect of Hourly Changes in Weather on Changes in Ticket Orders

	Hourly Ticket Orders		
	Morning & Afternoon	Morning	Afternoon
Diff. Sun per h.	0.00038**	0.00043	0.00043*
	(0.00018)	(0.00029)	(0.00024)
Diff. Rain per h.	0.000019	−0.000028	0.000027
	(0.000038)	(0.000042)	(0.000048)
Hour Indicators	*Yes*	*Yes*	*Yes*
Observations	22454	8984	11225
Adjusted R^2	0.001	0.001	0.000

Notes: We report the coefficients and robust standard errors of OLS regressions of the first difference of hourly ticket orders on the first difference of hourly purchase-date sunshine duration (in percent of time), the difference in purchase-date rainfall (in 1/100 mm), and hour indicator variables. Column 1 reports coefficients for all orders between 8 am and 8 pm. The two remaining columns split the dataset into orders before and after 2 pm.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the economic interpretation of the estimates, one standard deviation change in weather explains variations in sales between 10 and 17 percent of (quite low) mean sales for the respective periods and is therefore in a similar range as for all our previous estimates in Section 1.4.1.

Note that these results can also be seen as a robustness check for the concern that information content of current weather drives the results. As early as three weeks before the movie-date, (perceived) information content of purchase-date weather for movie-date weather should be nil.

Another instance in which changes in weather can naturally be assumed to have little impact on the probability of ticket availability are variations in weather from one hour to the next. Again, we only expect to find an effect of hourly changes in weather on changes in ticket orders if the state of the world by itself – and not its effect on results of market interactions – affects choice behavior.

To test this prediction, we regress, for a given movie-date, the first difference of ticket orders per hour on the first difference of sunshine duration and precipitation. In addition, hour dummies are included as independent variables to control for different sales volumes over the course of the day. We restrict the sample to hours with potentially positive sunshine duration (8 am to 8 pm) as well as to hours in the morning (8 am to 2 pm) and afternoon (2 pm to 8pm) between one and four days ahead of the movie-date.

In this analysis, the effect of weather on ticket orders is identified through variations in weather within a given day. Given the low within-day variation in sales and thus low mean differences as dependent variable, the estimated coefficients are rather small (see Table 1.6). Still, hourly changes in weather have a statistically significant effect on changes in ticket orders. The estimates seem to be mainly driven by sales in the afternoon when most tickets are ordered and therefore hourly variation in sales is highest.

In light of the evidence that both, very early sales as well as hourly changes in sales are affected by current weather, we conclude that current weather explains ticket orders even when the probability of ticket availability is independent of purchase-date weather. Summarizing our above arguments, projection bias is the only explanation that can simultaneously account for all our empirical findings.

1.5.3 Robustness

We examine the robustness of our empirical results along various dimensions.[27] First, we check whether the results depend on estimating linear models in the form of (1.4) and (1.5). We therefore repeat the entire analysis using count data models (poisson and negative binomial regressions). None of the results are altered by estimating either of these models.

Second, we evaluate whether the main results in Table 1.3, for which we have most statistical power, are driven by specific subgroups of customers, movie genres, or the weekday of the show. To this end, we repeat the analysis for the subgroup of orders conditioning on the timing of the order (morning, afternoon, and evening), the number of tickets ordered (one, two, three, and more than three), the age of the buyer (below and above 30 years), movie genre (drama, comedy, and action/adventure), and the weekday of sale and movie respectively (weekend, weekday). The number of total orders across these subgroups varies considerably – for example, more than 50 percent of ticket orders comprise of two tickets, while less than 10 percent of customers order three tickets – which is accompanied by different degrees of statistical power for the analysis.

Nevertheless, the estimated effect of sunshine duration and (to a lesser extent) rainfall on total orders continues to be statistically and economically significant at least for the fixed effect model (1.4). The only exception from this rule are ticket orders on weekends (Friday to Sunday), for which the number of observations is lowest (N = 291). Additionally, the cross-section models (1.5) yield meaningful effects for current weather on sales for the majority of specifications.

Third, we check whether our results are driven by the selection of average sunshine dura-

[27] Due to the large number of robustness checks, we do not include the detailed results here. The estimation results for all analyses mentioned in this subsection are available from the authors on request.

tion as the relevant weather variable. To do so, we substitute average sunshine duration by average temperature in all regressions with ticket orders as the dependent variable. By and large, this substitution leaves the results unchanged. Similarly, we test whether our conclusions are sensitive to the choice of the independent variable. Instead of using aggregated ticket orders as the quantity to be explained, we could have also used the overall number of tickets sold as independent variable. These two measures are highly correlated, such that it is not surprising that this modification does not lead to different conclusions.

A fourth possible concern could be that ticket sales depend on recent rather than current weather. For example, customers may be more inclined to buy tickets if weather was good for a couple of days, possibly indicating a stable high pressure weather system. We account for this by including one period lagged weather variables in our analysis of models (1.4) and (1.5). In these empirical models, current weather persists to have explanatory power, contrary to weather one day earlier.

Finally, we examine the robustness of all empirical results to the inclusion and exclusion of various control variables. First, we attempt to proxy for the probability of ticket availability directly by (a) including a dummy indicating whether the theater turned out to sell out for the particular show of interest and (b) including the number of ticket orders until the purchase-date in all variants of cross-section models (1.5). Second, we control for the popularity of the movie by including either the number of theaters in which the movie was shown on the opening weekend in Germany, or movie gross in Germany on the opening weekend (or both) as independent variables in all variants of model (1.5).[28] Neither of these modifications changes the results in any meaningful way. The same holds true for excluding controls $\mathbf{D}_{\tau\mathbf{t}}$ or $\mathbf{X}_{\tau\mathbf{t}}$ in all the models we estimate.

1.6 Conclusion

There is a growing literature which shows that projection bias impedes the ability of individuals to consistently predict future utility. Predicted utility at unknown future states of the world tends to be biased towards utility at today's state of the world.

This study presents evidence for projection bias in a simple decision problem (purchasing

[28]The choice of the popularity indicator is somewhat arbitrary; for example, we could have chosen the total gross of the movie shown as well. We opted for opening weekend measures to avoid measurement error due to the total time the particular movie has been screened in theaters. The concern for measurement error arises because the outdoor movie theater in question shows recent films as well as classics. All data for this analysis have been retrieved from the database http://www.boxofficemojo.com [October 2011].

advance tickets for an outdoor movie theater), in which the transient nature of today's state is obvious due to risk and explicitly pointed out to decision makers. The extent to which decisions are biased should therefore be minimal. The availability of unbiased and precise forecasts regarding future states, on which individuals may condition their decisions, should further reduce the extent to which projection bias affects choices. Nevertheless, we find that the current state – current weather – influences choices to a large extent, which suggests that de-biasing decision makers may turn out to be challenging.

Put in a broader context, our result that projection bias is present even in very simple decision problems points towards the possibility that it may have important aggregate implications This may especially be the case when decisions of individuals observing the same state of the world are biased in the same direction. It has been shown recently that such correlated errors can be amplified through feedback effects in markets, leading to potentially large fluctuations (Hassan and Mertens, 2011). To evaluate this hypothesis, further research is needed to answer the question whether projection bias affects choices over alternatives whose utility depends on an endogenous state of the world (like consumption and savings decisions depending on the state of the economy).

Finally, studying projection bias under risk highlights the need to understand how exactly individuals mispredict future utility. Do they indeed undervalue the extent to which utility varies with the state (and hold correct beliefs regarding future outcomes), the standard interpretation of projection bias? Or are their beliefs regarding the likelihood of future states biased towards the current state (and the predictions of state-dependent utilities correct)? Answering these questions is certainly important for finding ways to help individuals to predict future utility accurately.

Chapter 2

Do Lagged Expectations Determine Reference Points? — A Test of Kőszegi and Rabin's Equilibrium Concepts

2.1 Introduction

Loss aversion, according to which individuals dislike losses more than they like equal-sized gains, has been found to be an important factor in explaining individual decisions under uncertainty. Gains and losses are defined relative to a reference point such that the specification of the reference point is a key part of any application of loss aversion to explaining choice behavior. While in many applications the identity of the reference point represented a degree of freedom for the researcher, Kőszegi and Rabin (2006, 2007, 2009, henceforth KR) provide modeling discipline by introducing a theory of endogenous reference points. In a series of influential papers, they argue that "a person's reference point is the probabilistic belief she held in the recent past about outcomes" (KR 2006, p. 1134).

Defining the reference point as *recent probabilistic belief* has two important implications. First, if the reference point is given by *probabilistic beliefs*, it is possibly stochastic.[1] This means that individuals will compare each realized outcome as gain or loss compared to the entire expected distribution of outcomes. To give an example, consider an individual

[1]The term "reference distribution" is probably more appopriate for describing a stochastic reference point. However, we follow KR (and the subsequent literature) and use the term reference point regardless of wheter the reference point is given by a degenerate or a non-degenerate distribution.

endowed with a lottery paying 7 Euro or 27 Euro with equal probability, and let the reference point be given by the distribution of outcomes implied by this lottery. According to KR, the individual will compare the unfavorable outcome of 7 Euro with equal probability both to itself (yielding neither gain nor loss) and to the favorable outcome of 27 Euro (yielding a loss of 20 Euro). The exact reverse holds for gain-loss utility attached to the favorable outcome, which provides a gain of 20 Euro when evaluated in comparison to the unfavorable outcome and zero gain-loss utility when compared to itself. Because losses loom larger than gains, expecting to face a risky lottery and incorporating this expectation in the reference point thus necessarily leads to negative expected gain-loss utility. In contrast, being endowed with and expecting a safe payment yields expected gain-loss utility of zero.

Second, equating the reference point with *recent beliefs* means that there is a lag between the time at which an individual receives new information about the distribution of outcomes and the time at which this new information is incorporated into the reference point. Clearly, whether or not the reference point adapts is important for how outcomes are evaluated. If the reference point does not adapt to the expected distribution of outcomes, all outcomes in the above examples are compared to the expectations previously held. In this case, the lottery and the safe payment are evaluated relative to the same distribution over outcomes such that both alternatives yield similar gain-loss utility.

To the extent that individual choices alter the expected distribution over outcomes, reference points may be endogenous to choice. Whether or not the reference point is endogenous depends on the time structure of the decision problem: According to KR, there needs to be sufficient time between individual choice and the resolution of uncertainty for the reference point to adapt.

Kőszegi and Rabin argue that decision makers should anticipate whether the reference point is affected by their choice. To account for this, KR introduce two conditions for optimal choice behavior with choice-endogenous reference points. First, when there is sufficient time for the reference point to adapt to the consequences induced by choice, the condition of "Choice-Acclimating Personal Equilibrium" (CPE) applies. CPE requires that *choices* are optimal given their distribution over outcomes as the reference point. Second, when the reference point does not adapt to choices (because uncertainty is resolved directly after choosing), but the choice set is known well in advance, KR argue that the reference point is given by *choice plans*. Because individuals cannot commit to plans, the condition of "Preferred Personal Equilibrium" (PPE) requires that it is optimal for individuals to stick to their plans given these plans as the reference point. Finally, while CPE and PPE correspond to situations with choice-endogenous reference points, we label situations in which there is insufficient time for the reference point to adapt "Surprise situations". In these situations, the reference point is exogenously given by previously held beliefs.

This study tests KR's equilibrium conditions for the first time. Since the applicability of these conditions depends on the time structure of the decision problem, a testable prediction of KR's theory is that individual preferences for a risky lottery compared to a safe payment depend on the relative timing of information about the choice set, decision making, and the resolution of uncertainty. Specifically, individuals are predicted to take risks with higher likelihood in a Surprise than in a CPE or PPE situation because expecting to face a risky lottery leads to negative expected gain-loss utility.

We conduct a controlled lab experiment to test whether the time structure of the decision problem affects risk preferences as predicted by KR. We ask subjects to choose between a safe payment and a lottery in three treatments. In the first treatment – corresponding to a Surprise situation – the outcome of the lottery becomes known immediately after the choice set is introduced and choices are made. In the second treatment – corresponding to a CPE situation – uncertainty is resolved with a lag of 24 hours such that the reference point can adapt to choices. In the third treatment – corresponding to a PPE situation – we inform subjects about the choice set that they will encounter 24 hours later, elicit their choices after that time, and resolve uncertainty immediately. In this situation, we expect the reference point to adapt to choice plans, but not to factual choice. To avoid confounds, subjects have to visit the laboratory on two consecutive days and are paid after the second date in all treatments.

Note that the PPE-treatment allows us to distinguish KR's theory from other theories which predict that individuals prefer lotteries with early resolution of uncertainty (as in the Surprise-treatment) to lotteries with delayed resolution (as in the CPE-treatment). Drèze and Modigliani (1972) and Spence and Zeckhauser (1972) argue that there is an option value in receiving information early because consumption plans can be optimally adjusted.[2] Moreover, Wu (1999) and Caplin and Leahy (2001) introduce the possibility that individuals experience disutility from anxiety while waiting for the revelation of risk. However, none of these theories relates observable risk preferences to the timing of information about available choice options as inherent in PPE situations.

Summarizing the results of the experiment, risk preferences are statistically identical across treatments. Compared to the Surprise situation, neither the possibility of planning nor the delayed resolution of risk affects the willingness of subjects to choose the lottery. This finding is robust to controlling for individual risk preferences. Furthermore,

[2]See Kreps and Porteus (1978) and Epstein and Zin (1989) for axiomatic treatments on how preferences depend on the temporal resolution of uncertainty.

the theory presented in Section 2.2 suggests that the treatment effect is present only for individuals with risk preferences within a specific range. In contrast to this prediction, we do not find evidence for heterogeneous treatment effects. Our results are thus consistent with expected utility theory, but not with theories that predict a relationship between preferences for risk and the temporal resolution of uncertainty. The results can be reconciled with the KR-model, however, if referents adapt immediately to information in our experimental setting.

This is the first study that seeks to test predictions of KR's equilibrium conditions empirically. Testing whether referents adapt to choices is important in order to evaluate the recent theoretical literature that applies KR's idea of stochastic reference points to various settings.[3] Often, these models differ with respect to whether they require consistency between choices and reference points. For example, in the branch of behavioral industrial organization that analyzes optimal contracts for loss averse consumers, some models assume that consumption plans are a PPE. This assumption requires that the (stochastic) reference points regarding price and quality of planned consumption are consistent with the plans given these referents (Heidhues and Kőszegi, 2008; Karle and Peitz, 2010). In contrast, the approach of Carbajal and Ely (2012) introduces stochastic referents regarding the quality of the object consumed, but does not (necessarily) require that these referents are consistent with the optimal consumption plans they induce.[4]

There is a growing literature that tests implications of the KR-models in the lab. The majority of these papers is devoted to providing evidence for expectation-based reference points (Abeler et al., 2011; Ericson and Fuster, 2011; Gill and Prowse, 2012).[5] Relatedly, Sprenger (2010) finds empirical support for stochastic reference points, confirming a key prediction of KR.[6] Yet, these papers do not tackle the question of how fast the referent adapts to new information and choices, which is at the heart of our work. Rather, by deriving their predictions using the concepts of CPE or PPE these papers have implicitly assumed that the referent adjusts to changes in beliefs within time spans of few minutes (within a single session in the lab).[7] However, their findings that the referent adapts to

[3]In the following, we use the terms "reference point" and "referent" interchangeably.

[4]Similar work includes Herweg and Schmidt (2012) who study optimal contracts with the possibility of renegotiating these contracts. In their model, renegotiation comes as a surprise such that the contract serves as the referent. Furthermore, the studies by Herweg et al. (2010) and Daido and Murooka (2012) on optimal incentive contracts in principal-agent relationships with loss averse agents rely on CPE as equilibrium concept as well as work by Eisenhuth (2012) on auction design with loss averse buyers.

[5]Note that while an endowment effect based on expectations has been found by Ericson and Fuster (2011), this finding could not be reproduced in similar settings by either Heffetz and List (2011) or Smith (2008).

[6]These results are strengthened by the studies of Mas (2006), Card and Dahl (2011), Crawford and Meng (2011), and Pope and Schweitzer (2011) all of which provide evidence for expectation-based referents from field data.

[7]The exception to this rule is the work by Sprenger (2010) who assumes that the referent adjusts

expectations as predicted is indicative for more or less immediate adaption of the referent to new information, consistent with our results.

The evidence from a few studies, which explicitly tests whether referents adapt with a lag to new information, remains inconclusive regarding the speed of adaptation. In a lab experiment, Song (2012) varies whether subjects learn a fixed payment 24 hours or 5 minutes before the experiment and finds no effect on behavior. He thus concludes that the referent adapts immediately.[8] In contrast, Post et al. (2008) find that individual risk aversion in a TV game show with large stakes depends on past expectations and interpret this as evidence for lagged adaptation of reference points. Similarly, Matthey and Dwenger (2008) argue that it takes time for the referent to adjust to expectations, as original information provided long before their experiment influences risk taking while new information provided during the experiment does not. Our work adds to this literature by being the first that explicitly tests the implications of KR's equilibrium concepts.

Finally, this study contributes to the small literature that evaluates whether the timing of uncertainty resolution matters for behavior, and extends it by including the possibility for subjects to make plans well before actually making a choice. In the existing experimental studies, individuals have to take multiple choices such that the reference points predicted by KR become very complex. Therefore, the result of this work cannot be related to the predictions of KR tested in our study, in which subjects only make one decision. Nevertheless, the literature does not find a clear pattern of how the temporal structure of uncertainty resolution affects preferences. Van Winden et al. (2011) find that immediate resolution of risk increases the willingness to invest in a risky project with high probability of a fairly low good outcome but does not change investments into a (mirrored) risky project with the same expected value but a low probability of a fairly high good outcome. Our result that the timing of uncertainty resolution does not matter for behavior is consistent with their results, as our probabilities and payoffs lie in the middle of their range. All of these results can further be reconciled with findings from Gaudecker et al. (2011). Their structural estimations suggest that choices of the median subject from a representative sample are independent of the temporal structure of uncertainty resolution, even though these preferences are heterogeneously distributed within the population. This is also consistent with the mixed results found when subjects can express their preferences for the temporal resolution of uncertainty directly (Chew and Ho, 1994; Ahlbrecht and

quickly to the choice option repeatedly presented to subjects. Thus, in his context, referents adapt to information, but not necessarily to the consequences induced by the option chosen.

[8] Somewhat relatedly, Zimmermann (2012) tests the prediction of KR (2009) according to which individuals should prefer to receive information about outcomes clumped rather than piecewise since new information may lead to feelings of losses with respect to previously held beliefs. Zimmermann finds that individuals do not avoid piecewise information and interprets this as evidence that subjects do not expect losses from changes in expectations, which could be the case if referents adapt immediately.

Weber, 1997; Arai, 1997; Noussair and Wu, 2006; Eliaz and Schotter, 2007; Kocher et al., 2009).

The remainder of the chapter proceeds as follows. In Section 2.2 we derive the hypothesis on how risk preferences are related to the relative timing of 'information regarding the choice set', 'individual choice', and 'uncertainty resolution' from a theoretical framework based on KR (2007). In Section 2.3 we describe the experiment and discuss its results in Section 2.4. Section 2.5 concludes.

2.2 Theoretical Framework and Hypotheses

In this section, we derive the theoretical predictions which guide our experimental design. As we are primarily interested in how the time span between learning the choice set, making choices, and uncertainty resolution affects individual risk preferences in the KR model, we derive the implications of this model first. We then compare KR's predictions with predictions from other models.

2.2.1 Kőszegi-Rabin Preferences and Equilibrium Concepts

According to the KR model, an individual i evaluates a sure consumption outcome c with respect to a certain reference point r according to

$$u_i(c|r) = m_i(c) + \mu(m_i(c) - m_i(r)).$$

The term $m_i(c)$ denotes individual consumption utility from c, while the function μ denotes gain-loss utility as a function of the difference in consumption utility between outcome c and referent r. Following KR, we assume that gain-loss utility μ is piecewise linear with $\mu(g) = \eta g$ for $g \geq 0$ and $\mu(g) = \lambda\eta g$ for $g < 0$. The parameter $\eta \geq 0$ measures the utility weight of gain-loss utility relative to consumption utility; the case $\eta = 0$ nests standard utility depending only on outcomes. The parameter $\lambda > 1$ denotes the degree of loss aversion. Regarding individual consumption utility, we assume that m_i is strictly increasing and concave. Furthermore, we incorporate heterogeneity in the model by assuming that individuals differ with respect to risk aversion over consumption utility as represented by the curvature of m_i. As will become clear below, the hypothesized treatment effects depend on the presence of some heterogeneity within the population, either in the parameters λ or η, or in the curvature of m_i. We assume heterogeneity in the curvature of m_i because, first, there is substantial evidence for heterogeneity in risk attitudes (see, for example, Holt and Laury, 2002, Bruhin et al., 2010, Dohmen et al., 2011, or Gaudecker et al., 2011), and second, because it allows us to derive an additional testable hypothesis on the basis of our data.

Given these primitives, the first important innovation of KR is that the referent r may be stochastic and drawn from a distribution G such that each factual outcome may be compared to multiple expected outcomes. Hence, with stochastic outcomes c drawn from a distribution F, the general formulation of (expected) reference dependent utility becomes

$$U_i(F|G) = \int \int u_i(c|r)dG(r)dF(c). \tag{2.1}$$

To ease the following exposition and to cover the situation encountered by subjects in the experiment, we consider a situation in which the individual has a choice between two alternatives. One alternative pays a safe amount x for sure, while the other alternative is given by a lottery F that pays $\tilde{y}$ or $\tilde{z}$ with equal probability. The expected value of F is greater than x $(1/2(\tilde{y} + \tilde{z}) > x)$ and payoffs satisfy $\tilde{y} > x > \tilde{z}$. We further normalize $m_i(x) = x$ for every individual and define $y_i = m_i(\tilde{y}) - x$ and $z_i = x - m_i(\tilde{z})$ as the difference in consumption utility between the safe option and the favorable and unfavorable outcome of the lottery, respectively. As risk aversion in consumption utility – given by the curvature of m_i – translates into risk aversion in KR-preferences, the value of the ratio y_i/z_i (which is decreasing in risk aversion in consumption utility) determines the individually optimal choice between x and F.

KR's second important innovation is that referents are given by lagged beliefs about outcomes. This entails the notion that the reference distribution does not adjust immediately to new information such that expected gain-loss utility – and hence, optimal choice – crucially depends on the time span between receiving information regarding the distribution of outcomes and the resolution of uncertainty. In the following, we closely follow KR (2007) in defining properties of optimal choice in three situations that differ with respect to the timing of information, choice making, and the resolution of uncertainty.

Consider first a situation in which an individual faces the choice between x and F unexpectedly and uncertainty is resolved directly after choice. Since there is only a short time between learning the choice set and uncertainty resolution, the previously held reference beliefs G do not adapt. Hence, outcomes are evaluated with respect to G such that individual choice is optimal if it satisfies the following property.

Definition 2.1. For some expected distribution of outcomes G, selecting x from $\{x, F\}$ is a *Surprise Equilibrium (SurpriseE)* if $U_i(x|G) \geq U_i(F|G)$. Otherwise, F is the Surprise Equilibrium.

Next, think of a situation in which the individual knows the choice set she will face well ahead, but will make her choice in the future. Uncertainty is resolved shortly after making a choice. The individual's reference point will then be unaffected by her decision at the time she learns the outcome of her choice. However, the referent will be endogenous to the choice plans the individual makes after learning the choice set. Then, her plan and the reference beliefs thus induced are only consistent if she is willing to execute her plan given her (reference) beliefs. In other words, an individual should only plan to choose x if this choice is optimal given x as reference point (and equivalently for F).

Definition 2.2. If selecting x from $\{x, F\}$ is a *Preferred Personal Equilibrium (PPE)*, then $U_i(x|x) \geq U_i(F|x)$. If selecting F from $\{x, F\}$ is a PPE, then $U_i(F|F) \geq U_i(x|F)$. When $U_i(x|x) \geq U_i(F|x)$ and $U_i(F|F) \geq U_i(x|F)$ hold simultaneously, then x is the PPE if $U_i(x|x) \geq U_i(F|F)$; otherwise, F is the PPE.

Finally, when individuals commit to their decisions well before the outcome is realized, the decision itself determines the beliefs with respect to which the outcome will be evaluated as gain or loss. In such situations, the optimal choice is therefore given by the alternative that maximizes expected reference-dependent utility given itself as the referent.

Definition 2.3. Selecting x from $\{x, F\}$ is a *Choice-Acclimating Personal Equilibrium (CPE)* if $U_i(x|x) \geq U_i(F|F)$. Selecting F is a CPE if $U_i(F|F) \geq U_i(x|x)$.

Because these three equilibrium conditions imply different requirements for optimal choice, individual decisions may vary in a predictive way among different time structures. We will evaluate next how preferences for risk are affected. For conciseness, we will thereby refer to the situations described above by their corresponding equilibrium concepts whenever no confusion arises.

2.2.2 Hypotheses

We begin by discussing the perception of the lottery F in a PPE situation compared to a Surprise situation. In PPE, planning to choose F entails the expectation of gains and losses of size $y_i + z_i$ while planning to choose x does yield gain-loss utility of zero. Thus, for an individual that selects F in PPE, the gain of F in consumption utility compared to x needs to compensate for the negative gain-loss utility. In contrast, in a Surprise situation both F and x are evaluated with respect to the same exogenous reference lottery G (which may be degenerate). Hence, expecting G implies that gain-loss utility after choosing either F or x is in a similar range such that choosing F over x does not lead to additional expectations of losses.

The difference between a PPE and CPE situation is that in a CPE situation individuals can commit to an alternative which ex post may not be optimal given itself as referent, but in a PPE situation they cannot. Since expecting risk decreases utility, there may be cases in which an individual would prefer the safe outcome ex post in a PPE situation but cannot commit to it ex ante. In such cases, the CPE is x but the PPE is F. Furthermore, since agents are more risk averse in PPE than in Surprise situations, every individual choosing x in a Surprise situation chooses x in a CPE situation.

In the following proposition, we use these arguments to link individually optimal choices in different situations to risk aversion in consumption utility as measured by y_i/z_i. In doing so, we need to ensure that at least some individuals are indeed willing to choose the risky lottery in CPE and PPE situations, in which losses are expected for sure. Hence, losses must not loom too large in comparison to gains and consumption utility. For our parametrization, it turns out that choosing F can only be supported as PPE or CPE if $\eta(\lambda - 1) < 2$.[9]

Proposition 2.1. *Assume $\eta(\lambda-1) < 2$. Then, there exist $\underline{a}, a,$ and $\bar{a}$ with $1 < \underline{a} < a < \bar{a}$ such that for individual i endowed with the choice set $\{x, F\}$*

(i) F is the SurpriseE, PPE, and CPE if $\bar{a} \leq y_i/z_i$,

(ii) F is the SurpriseE and PPE, and x is the CPE if $a \leq y_i/z_i < \bar{a}$,

(iii) F is the SurpriseE, and x is PPE and CPE if $\underline{a} \leq y_i/z_i < a$, and

(iv) x is the SurpriseE, PPE, and CPE if $y_i/z_i < \underline{a}$.

The proof is relegated to Appendix B.1. Proposition 2.1 contains two testable hypotheses. The first is that the share of individuals choosing the safe option x should differ between Surprise, PPE, and CPE situations if there is a positive mass of individuals whose consumption utility satisfies $y_i/z_i \in (\underline{a}, a)$ and $y_i/z_i \in (a, \bar{a})$.

Hypothesis 2.1. *The share of individuals choosing the safe option x from $\{x, F\}$ is highest in CPE situations and lowest in Surprise situations. In PPE situations, there is an intermediate share of safe choices.*

Furthermore, the time structure of the decision problem should only affect individuals whose consumption utility satisfies $\underline{a} < y_i/z_i < \bar{a}$.

Hypothesis 2.2. *The effect from Hypothesis 2.1 is only present for individuals with individual risk aversion over consumption utility in an intermediate range.*

[9] Note that this holds for the frequently used parameter values $\eta = 1$ and $\lambda \in (1, 3)$.

Before describing the experiment designed to test Hypotheses 2.1 and 2.2 in Section 2.3, we briefly discuss whether theories other than KR can account for the hypotheses derived above.

2.2.3 Kőszegi and Rabin's Hypotheses in Relation to Other Models

We evaluate whether the models of disappointment aversion (Bell, 1985; Loomes and Sugden, 1986), preferences for temporal resolution of uncertainty (starting with Kreps and Porteus, 1978), or models of anticipatory emotions (Wu, 1999; Caplin and Leahy, 2001) can account for Hypothesis 2.1 or Hypothesis 2.2.

Disappointment Aversion The models of disappointment aversion (DA) of Bell (1985) and Loomes and Sugden (1986) are closely related to KR in that they also represent theories of expectation-based reference points. However, in their models the reference point of a lottery is given by the certainty equivalent of the lottery and is therefore not stochastic. Hence, expected reference-dependent utility is given by $U_i^{DA}(F|G) = \int m_i(c) + \mu(m_i(c) - E_G[m_i(r)])dF(c)$, where E_G denotes the expectation operator with respect to the distribution G.

Since the models of DA assume that the chosen option is a CPE (that is, it maximizes U_i^{DA} given itself as the referent), the models of disappointment aversion do not make predictions on how the time structure of the decision problems affects behavior. In order to compare KR's results with the ones of DA nevertheless, we impose KR's equilibrium requirements of SurpriseE, PPE, and CPE on optimal choices under DA preferences.[10]

Now, consider how a referent that is based on the certainty equivalent of recent beliefs over outcomes affects utility.[11] We analyze a Surprise situation first. Here, gain-loss utility of the lottery F and the safe option x depend on the value of expected consumption utility, $EM_i(G) := E_G[m_i(r)]$, relative to the safe payoff x and the payoffs of F, $x+y_i$ and $x-z_i$.[12] If $EM_i(G)$ is close to x, the individual will appear more risk averse since it can avoid the feeling of losses by choosing x. In contrast, if $EM_i(G)$ is close to one of the payoffs of F, gains and losses are of similar size regardless of choosing F or x; the individual will then appear to be risk neutral. For this reason, the share of individuals predicted to choose the safe option by the DA model in a Surprise situation may be larger or smaller than in

[10] See Appendix B.2 for exact definitions.

[11] For proofs of all claims in this and the next paragraph, see Appendix B.2.

[12] Note that by the definition of the certainty equivalent, $EM_i(G)$ is the consumption utility provided by the reference point in the DA models.

a PPE or CPE situation.

A similar reasoning applies when comparing PPE and CPE situations. Choosing F is only optimal when its expected consumption utility $EM_i(F)$ is larger than x. However, expecting to choose F is only consistent with ex-post optimal behavior in a PPE situation if $EM_i(F)$ is distant enough to x, as otherwise choosing x is tempting to avoid the expectation of relatively large losses from F. Thus, committing to F may be beneficial in CPE situations but not consistent in PPE situations such that the DA model predicts a lower share of safe options chosen in CPE than PPE situations, the exact opposite of the KR model.

We conclude that the (modified) DA model cannot account for any of the predictions from Hypothesis 2.1. Similar to Hypothesis 2.2, the DA model predicts that the timing of the decision problem should only alter choices for individuals whose risk aversion over consumption utility is within a certain range.

Preferences for Temporal Resolution of Uncertainty Individuals may have direct preferences for the temporal structure of uncertainty resolution (see Kreps and Porteus, 1978, for an axiomatic treatment). Thus, individuals may directly prefer early resolution to late resolution, for example due to the option value of forming optimal consumption plans. Thus, such preferences could explain the first part of Hypothesis 2.1 according to which the share of individuals choosing the safe option is higher in Surprise than in PPE or CPE situations.

However, such preferences do in general not entail preferences over the timing of making choices. Hence, in these utility representations, making choices and choice plans is equivalent as long as the distribution over outcomes and the timing of uncertainty resolution are unaffected. Thus, preferences for temporal resolution of uncertainty would predict no difference in choice behavior between PPE and CPE situations, contrary to Hypothesis 2.1. Furthermore, preferences for the timing of uncertainty resolution and risk preferences are typically not linked such that these preferences cannot account for Hypothesis 2.2.

Anticipatory Emotions The models of Wu (1999) and Caplin and Leahy (2001) incorporate anticipatory emotions evoked by uncertainty of future outcomes in utility representations.[13] As such, they are special cases of preferences for the time structure of uncertainty resolution and, hence, can account for a higher willingness to take risks in Surprise situations compared to PPE or CPE situations but not for differences in risk

[13]Wu incorporates anticipatory emotions by means of probability weighting depending on the time of uncertainty resolution. Caplin and Leahy directly assume that anxiety is decreasing in the expected value of future outcomes and increasing in their variance.

taking between PPE and CPE situations. Since they are special cases of preferences for the temporal structure of uncertainty resolution, models based on anticipatory emotion cannot account for Hypothesis 2.2.

2.3 Experimental Procedure

The experiment is designed to capture the key elements from the model discussed above. For achieving this goal, two aspects are of particular importance.

First, we need as much control as possible over choice-induced reference points. For this reason, we let subjects make one choice in one decision problem only. Given this, the calibration of the decision problem becomes important: if most subjects preferred the same option, we were not to find any effects. Using a subject pool of undergraduate students of the University of Munich, we therefore tested several parametrizations of decision problems in which subjects could choose between a safe payment and a lottery. From the decision problems tested, the one in which choices were most evenly distributed was implemented in the experiment. This was the case for the choice between a safe option paying 15 Euro for sure and a risky option paying either 7 Euro or 27 Euro with equal probability.[14]

A second important aspect of the experimental design is the specification of the time lag needed for the referent to adapt to expectations. As pointed out in the introduction, existing empirical results do not provide clear guidance regarding the appropriate length of the lag. For the experiment, we set the lag equal to 24 hours. One full day is a lag of meaningful length providing subjects some time to ponder about the decision; the referent should have at least partially adapted to expectations after that time. Furthermore, a lag of 24 hours is of practical importance, because at both dates the experiment is conducted at the same time of the day, which avoids possible time of the day confounds.

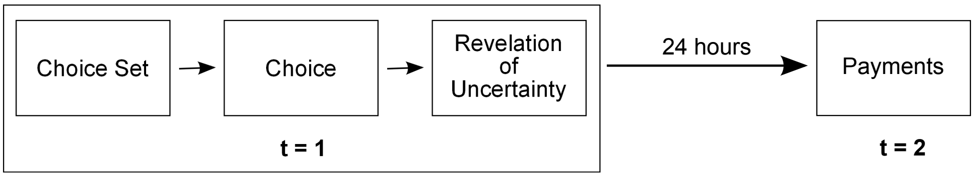
Treatments Given these considerations, we used a between-subject design to evaluate whether risk preferences vary with the time structure of the decision problem as predicted by KR. Across all treatments, the experiment started with a first set of instructions on the first day and ended with paying the subjects on the second day. All subjects thus had to visit the laboratory on two consecutive days, which was known in advance. Furthermore, for all treatments the ordering of tasks subjects faced was identical.[15]

[14] At the time of the experiment, 15 Euro were approximately equal to 20 Dollar; 7 and 27 Euro corresponded to about 9.5 and 39 Dollar, respectively.

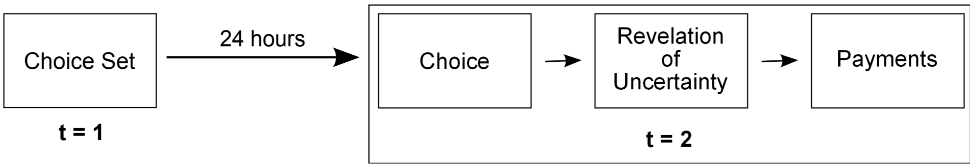
[15] There is one exception to this rule: across treatments, all subjects were asked the parameters of the lottery at the beginning of the second day's session (at a different stage in the ordering of tasks). See

Figure 2.1: Treatments

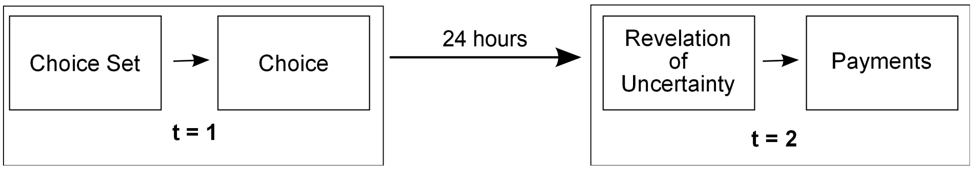
(a) Treatment Surprise



(b) Treatment PPE



(c) Treatment CPE



Three treatments were designed to capture the time structure of Surprise, PPE, and CPE situations (see Figure 2.1). In treatment *Surprise*, subjects learned that they could choose between the safe and the risky option on the first screen of the experiment. Directly thereafter, subjects had to make their choice, and uncertainty was resolved by the role of a die as follows: A set of lucky numbers – either $\{1, 2, 3\}$ or $\{4, 5, 6\}$ – was assigned to each subject who had chosen the risky option. Then, one randomly chosen subject rolled a die and announced the number which had come up. If the number was in the set of a subject's lucky numbers, the subject received 27 Euro; otherwise, she received 7 Euro.

In treatment *PPE*, subjects learned their decision problem at the first date as well as that they would be able to make their choice on the next day. As in treatment Surprise, uncertainty was resolved by the role of a die directly after choices were made on the second day. Finally, in treatment *CPE* subjects could make their choice on the first day of the experiment knowing that uncertainty would not be resolved until the next day.

Procedural Details Across treatments, the experiment started with instructions being read aloud.[16] In the instructions, the task was described as a decision problem between a safe payment called "Option A" and a lottery called "Option B" using other payments than those encountered by subjects during the experiment in order to avoid possible

below for details.

[16]See Appendix B.3 for an English translation of the original German instructions.

Table 2.1: Task for Elicitation of Risk Preferences

	Option A		Option B	
Row	with prob. 1/2	with prob. 1/2	with prob. 1/2	with prob. 1/2
1	18.00 €	16.00 €	24.00 €	6.00 €
2	18.00 €	16.00 €	20.00 €	10.00 €
3	18.00 €	16.00 €	16.00 €	14.00 €
4	22.00 €	12.00 €	16.00 €	14.00 €
5	23.50 €	10.50 €	16.00 €	14.00 €
6	25.00 €	9.00 €	16.00 €	14.00 €
7	26.50 €	7.50 €	16.00 €	14.00 €
8	28.00 €	6.00 €	16.00 €	14.00 €
9	29.50 €	4.50 €	16.00 €	14.00 €
10	31.00 €	3.00 €	16.00 €	14.00 €

Notes: Each row represents a decision problem between Option A and Option B. See the text for details.

reference point adaptation.[17] Furthermore, the time structure of the decision problem was emphasized.

In addition to providing information about the decision problem, subjects were informed that they would be asked about details of the decision problem they were going to face on the second day of the experiment. It was made clear that they could earn extra money if details were recalled correctly. We included this information in the instructions to provide incentives for subjects to memorize the choice problem and to have some control over beliefs held on the second day of the experiment. Upon arrival on the second date, subjects then had to recall all possible payments (15 Euro for the safe option and 7 Euro or 27 Euro for the risky option) as well as the probabilities of the outcomes of the risky option. They could earn up to 2 Euro if all details were recalled correctly.

On the second day (after the revelation of uncertainty in treatments PPE and CPE), we further elicited subject's risk preferences using a method suggested by Maier and Rüger (2010).[18] As in the standard risk elicitation task of Holt and Laury (2002), subjects had to choose between two lotteries (again called Option A and Option B, respectively) in a price list format displayed in Table 2.1. The main difference between this approach and the one of Holt and Laury is that the riskiness of the gambles in the Holt and Laury task is measured by the coefficient of risk aversion based on expected utility theory while the riskiness of gambles in the Maier and Rüger task is based on the concept of

[17]Note that we need to avoid the adaptation of the reference point to the "true" choice set in treatment Surprise. Furthermore, it follows directly from Proposition 2.1 that whether or not the reference point adapts to the alternatives presented in the instructions in treatment Surprise is irrelevant for the hypotheses tested.

[18]We provide a translation of the instructions for this part of the experiment at the end of Appendix B.3.

stochastic dominance. The Maier and Rüger risk elicitation task does therefore not rely on a specific model of individual risk attitudes. This is important here, as it is our goal to test predictions derived from the KR model, which differs substantially from the standard expected utility framework.

To see how Maier and Rüger measure risk preferences, note first that the expected values of both options as well as the probabilities of the good and bad outcomes remain constant across the decision problems in Table 2.1. Because Option A has a higher expected payment, risk neutral subjects would always choose it. Furthermore, in the first two rows Option A is weakly less risky than Option B, and in row three Option A first order stochastically dominates Option B. Therefore, there may be risk loving subjects who switch from B to A in either the first or the second row and choose A thereafter. Finally, from row four onward, Option A becomes strictly more risky in the sense that each later row is second order stochastically dominated by earlier rows. Hence, very risk averse subjects would switch from A to B early after the third row, and less risk averse subjects switch in later rows.

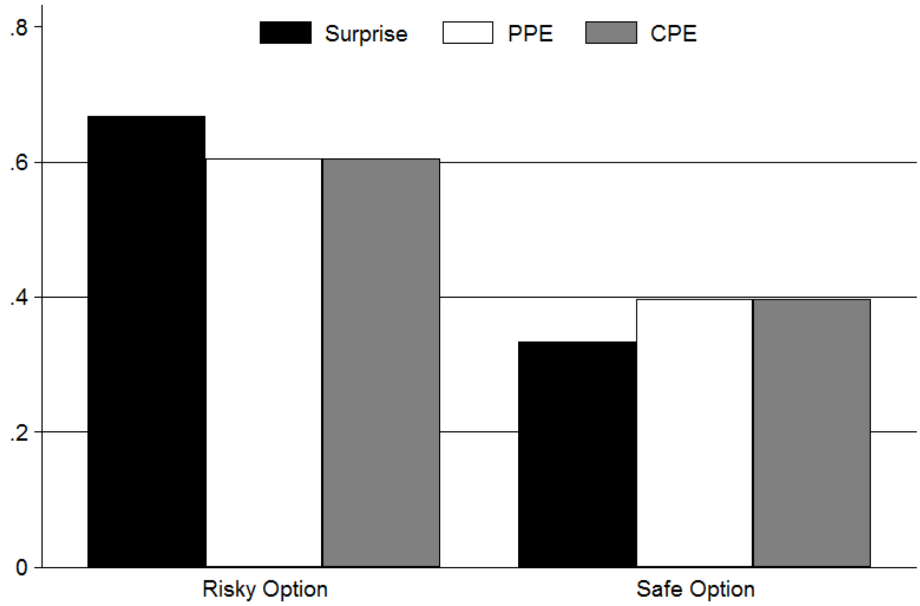
In order to give subjects incentives to report their preferences truthfully, one decision of two randomly selected subjects in each session was paid out (most sessions were conducted with 24 subjects). In a short questionnaire at the end of the experiment, subjects furthermore provided a measure of self-assessed risk preferences as well as few personal details like their age, gender, and subject of study.

The experiment was conducted using z-tree (Fischbacher, 2007) at MELESSA, the social science laboratory of the University of Munich. Subjects were invited using ORSEE (Greiner, 2004); in total, 141 subjects participated in the experiment (45 subjects in treatment Surprise and 48 subjects each in treatments PPE and CPE). Subjects average total earnings from the experiment were 19.90 Euro including a show-up fee of 4 Euro.

2.4 Results

Consider, first, overall choice behavior in the experiment. In Figure 2.2 we plot the fractions of subjects choosing the risky lottery and the safe payment across treatments. In treatment Surprise, 33 percent of subjects preferred the safe option, which is slightly less than in treatments PPE and CPE (40 percent). However, even though the safe option is chosen more frequently in CPE and PPE than in Surprise as predicted by KR, these differences are by no means statistically significant. Using a Wilcoxon rank-sum test, we cannot reject the null-hypothesis that choice behavior in Surprise is identical to choice behavior in PPE or CPE (p-value 0.48). Furthermore, the finding that behavior in

Figure 2.2: Fraction of Risky and Safe Options Chosen by Treatments



treatments PPE and CPE is identical in the experiment contrasts the prediction of the KR-model that the safe option is chosen more frequently in PPE than in CPE.

We further explore whether the reason why we do not find a treatment effect in the analysis of summary statistics is that individual heterogeneity is not accounted for by experimental randomization. Using a Probit model, we estimate the likelihood that a subject chooses the safe option; the results of this exercise are reported in Table 2.2. Here, we control for gender and age in all specifications. We add individual risk preferences as further controls (columns (2) and (4)) and estimate separate treatment effects for subjects who recalled the decision problem correctly on the second day of the experiment (columns (3) and (4)).[19] Risk preferences are controlled for by including dummy variables for the number of times a subject had chosen Option A (with higher expected value) over Option B in the risk elicitation task.[20]

In line with the previous findings, the treatment effects of CPE or PPE as compared to Surprise are insignificant in all model specifications (all p-values are larger than 0.3). Regarding the predicted treatment differences between PPE and CPE, note that in case the coefficient of CPE is larger than the one of PPE as predicted, the difference between

[19] 82 percent of subjects recalled the decision problem correctly.

[20] All results are robust to using other indicators of risk preferences such as the number of the first row a subject switched from one option to the other. We chose the number of Option A choices as a measure for risk preferences because it allows us to deal in a consistent manner with subjects who repeatedly switched between A and B. This is important, since 30 percent of subjects displayed such inconsistent choice behavior. Excluding these subjects from the analysis does not alter results. See below for a more extensive discussion.

Table 2.2: Probit Estimates of Treatment Effect on Choice of the Safe Payment

	Choice of Safe Option = 1			
	(1)	(2)	(3)	(4)
PPE	0.058	0.058	0.090	0.13
	(0.10)	(0.11)	(0.12)	(0.12)
× recall wrong			−0.22	−0.18
			(0.27)	(0.26)
CPE	0.070	0.060	0.058	0.049
	(0.10)	(0.11)	(0.11)	(0.12)
× recall wrong			−0.055	0.044
			(0.37)	(0.36)
Recall wrong			0.19	−0.0092
			(0.36)	(0.30)
Risk preferences	*No*	*Yes*	*No*	*Yes*
Gender and age	*Yes*	*Yes*	*Yes*	*Yes*
Observations	141	141	141	141

Notes: Marginal effects and standard errors of a Probit regression of subjects' choices of the safe option on treatment dummies (CPE, PPE) and controls are reported. Coefficients denote the treatment effects of PPE and CPE compared to Surprise for all subjects (first two columns) and for subjects who recalled the choice set correctly on the second day of the experiment (last two columns) at the sample mean. "Recall wrong" equals 1 if the subject recalled the choice set correctly, and zero otherwise. We control for risk preferences in columns (2) and (4) using dummy variables for the number of Option A choices in the risk elicitation task.
Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the coefficients is rather small and by no means statistically significant.

Result 2.1. *Across treatments, subjects are equally likely to choose the safe option.*

Our results are thus contrary to Hypothesis 2.1, which states that individuals are most likely to prefer the safe option in CPE and least likely to prefer it in Surprise.

As suggested by Hypothesis 2.2 we may not find treatment effects for the population of all subjects because they are only there for specific subgroups. Therefore, we investigate next whether there are heterogeneous treatments effects depending on subject's individual risk attitudes over consumption utility. To this end, we classify subjects as either risk neutral, moderately risk averse, or highly risk averse according to their choices in the risk elicitation task and evaluate the treatment effects for each class of risk preferences separately.

Clearly, the method used to measure risk preferences determines how subjects are classified. To assess whether results are sensitive to the measure used, we use three different methods of classification and compare the results obtained with each method. Table 2.3 shows how many subjects are categorized as risk neutral as well as moderately or highly

Table 2.3: Number of Subjects by Risk Preferences for Different Classifications

	# Option A	1st Switch	1st Switch & Consistent
Risk Neutral	27	30	18
Moderately Risk Averse	83	66	55
Highly Risk Averse	31	45	26

Notes: See the text on how subjects are classified as moderately or highly risk averse or risk neutral using three different methods of classification.

risk averse for each of the method used, which we will describe in the following.

First, we measure a subject's risk preferences by the total number of times she prefers Option A over Option B. Since Option A has a higher expected payoff but is also – for most decision rows – more risky, a higher count of Option A choices indicates less risk aversion. Accordingly, we classify subjects with less than five choices of Option A as highly risk averse (yielding 31 subjects in that category). Subjects who chose Option A between five and seven times are classified as having a moderate degree of risk aversion (83 subjects), and subjects with more than seven Option A choices are considered to be approximately risk neutral (27 subjects). Note that the latter category also includes risk seeking subjects, who choose Option B in the first one or two decision rows and Option A in the remaining eight rows.

The second method classifies subjects according to the row in which they first switch from Option A to B or vice versa. The mapping from the switch point to the class of risk preferences is such that this method classifies consistent subjects – that is, subjects who switch only once – in the same way as the method based on the count of Option A choices. Hence, subjects with a switch point in row four or five are classified as highly risk averse (31 subjects), and subjects who switched first in rows six, seven, or eight are classified as moderately risk averse (66 subjects). The group of approximately risk neutral subjects (30 subjects) comprises of those who either switched in row nine or never, and of those who switched between rows one and three (remember that switching in the first three rows indicates risk seeking behavior).

Third, we consider only subjects who showed consistent behavior by switching at most once. For these subjects, both methods described above are equivalent and yield groups of 18 risk neutral, 55 moderately risk averse, and 26 highly risk averse subjects.

Before we proceed, it is worth pointing out that the results presented in the following are insensitive to the particular cutoff values used for assigning subjects to classes of risk preferences. The cutoff values used here merely ensure that the sizes of the resulting groups are as balanced as possible and do not become too small when considering only consistent subjects.

Table 2.4: Heterogeneous Treatment Effects on Choice of the Safe Payment

	Choice of Safe Option = 1					
Risk Measure	# Option A		1st Switch		Consistent	
Model	Probit	OLS	Probit	OLS	Probit	OLS
Risk Neutral						
PPE	−0.26	−0.21	0.88***	0.15		−0.039
	(0.17)	(0.18)	(0.025)	(0.099)		(0.048)
CPE	0.062	0.054	0.83***	0.40**	0.78***	0.36
	(0.29)	(0.26)	(0.029)	(0.19)	(0.041)	(0.25)
Moderate Risk Aversion						
Mod. Risk A.	−0.007	−0.007	0.99***	0.32***	0.95***	0.24**
	(0.21)	(0.18)	(0.005)	(0.10)	(0.033)	(0.11)
PPE	0.14	0.13	0.068	0.074	0.094	0.090
	(0.15)	(0.14)	(0.16)	(0.17)	(0.19)	(0.19)
CPE	0.13	0.12	0.057	0.059	0.075	0.068
	(0.13)	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)
High Risk Aversion						
High Risk A.	0.35	0.35	1.00***	0.59***	0.99***	0.55***
	(0.25)	(0.24)	(0.002)	(0.14)	(0.007)	(0.21)
PPE	0.080	0.070	0.024	0.026	0.20	0.18
	(0.23)	(0.23)	(0.15)	(0.18)	(0.25)	(0.24)
CPE	−0.11	−0.13	−0.020	−0.025	−0.13	−0.15
	(0.20)	(0.25)	(0.17)	(0.21)	(0.20)	(0.28)
Gender and age	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	141	141	141	141	89	99

Notes: This table reports the estimated treatment effects and robust standard errors of CPE and PPE as compared to Surprise for each group of subjects classified either as risk neutral, moderately risk averse, or highly risk averse. The baseline group consists of risk neutral subjects; "Mod. Risk A." and "High Risk A." take the value 1, if a subject is classified as belonging to this group. In the first two columns, risk preferences are characterized by the number of choices of Option A in the risk elicitation task. In the remaining columns, risk preferences are characterized by the row in which individuals switched first; in the last two columns, only subjects with consistent switching behavior are considered (see the text for details). For Probit models, we report the size of treatment effects at the sample mean. Standard errors are adjusted accordingly.
Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4 reports estimated treatment effects – both from a Probit and a Linear Probability (OLS) Model – of CPE and PPE compared to Surprise separately for each class of risk preferences and classification method. Irrespective of how risk preferences are measured, these treatment effects are small and insignificant for risk averse subjects. The estimates for these groups suggest that the likelihood of choosing the safe option in PPE and CPE is at most ten percent larger than in Surprise, similar to the results from the pooled sample above. Note furthermore that the treatment effect of CPE is never significantly larger than the one of PPE, contrary to KR's prediction.

In contrast, we find moderate support for KR's prediction that the safe option is chosen more frequently in CPE and PPE than in Surprise for subjects classified as risk neutral by their first switch point. The OLS estimates suggest that under PPE (CPE) – as compared to Surprise – the safe option is 15 (40) percent more likely to be chosen. Yet, these results are not robust to considering only those subjects who behave consistently in the risk elicitation task; for them, the OLS model finds no statistically significant treatment differences. In contrast to the OLS results, the treatment effects estimated using a Probit model are always statistically significant but seem to be upward biased (note that we report marginal effects in Table 2.4). They predict the safe option to be chosen almost with certainty, even though only 37 percent of subjects classified as risk neutral chose it under CPE.

Result 2.2. *Overall, there is little evidence supporting Hypothesis 2.2, which predicts that treatment effects should be heterogeneous and depend on subjects' risk attitudes. Although there are significant differences in choice behavior across treatments for risk neutral subjects when classified by the first switch point, these results are not robust to considering alternative measures of risk preferences.*

2.5 Conclusion

In this study, we examined whether the time structure of a decision problem alters the distribution of individual risk preferences. Standard expected utility predicts that individual decisions should be independent of the length of lags between the date on which the individual learns a decision problem, the date on which she takes a decision, and the date on which uncertainty is revealed. In contrast, Kőszegi and Rabin (2007) argue that reference points adapt to new information with a lag such that a given choice option will be evaluated differently depending on whether or not there is sufficient time for the referent to adapt before the resolution of uncertainty. In our experimental data, however, we find no support for their predictions that the time structure of decision problems affects behavior.

Of course, our results could be in line with the model of Kőszegi and Rabin (2007) if we have misspecified the time needed for the referents to adapt. Since there is no theoretical guidance of whether the referent adapts within a few minutes or only after a number of days (or longer), this is a challenge we share with other approaches studying the adjustment of reference points to new information (see, for example, Zimmermann, 2012). In our context, we were not to find any treatment effect if subjects perceived all treatments equivalently, either because the factual time span needed for referents to adjust is very short (a few minutes) or quite long (longer than a day). Given that most tests of expectations as reference points (e.g. Abeler et al., 2011; Gill and Prowse, 2012) implicitly assume that referents adapt within a single experimental session, it seems unlikely that referents in our application only adapt after more than a day. Hence, our results can be interpreted as evidence for a very short lag between new information and reference adjustments.

As such short lags of adjustment are basically meaningless in real situations, the question arises whether the speed of adjustment is context-dependent. Many studies, which use loss aversion to explain individual behavior in field settings, require that it takes a significant time for reference points to adjust.[21]It is thus a question for future work whether there are certain factors – such as the attachment of individuals to certain outcomes – which determine the speed of reference point adjustments.

[21]Consider, for example, the loss aversion based explanations for the well known disposition effect in finance, according to which private investors are reluctant to realize paper losses from their assets they own but quick to realize paper gains (Odean, 1998; Barberis and Xiong, 2009). Here, the purchase price of the assets is typically taken as the reference point, even though there is typically a lot of information regarding the actual asset value.

Chapter 3

Emergence and Persistence of Extreme Political Systems — Transition Dynamics in an Unrestricted Polity Space[*]

3.1 Introduction

There is a growing economics literature exploring causes and circumstances of political transitions. This recent literature describes how political reforms and revolts can be supported within a rational-agents framework and how political systems are changed by these transition mechanisms. Even though many aspects of political transitions are dynamic in nature, these studies have largely abstracted from dynamic issues and focused on isolated transition events. This study takes a step towards filling this gap, by placing the dynamic process that describes the evolution of political systems at the core of the analysis. To this end, we endogenize outcomes of political transitions to a continuum of *a priori* attainable political systems and ensure the co-existence of reforms and revolts along the equilibrium path, allowing us to focus on the endogenously arising dynamic properties of political transitions.

More specifically, we introduce a dynamic framework where the space of political systems ranges continuously from single-man dictatorships to full-scale democracies. Actual polities are determined endogenously and result from political transitions that either can be initiated from within a regime (i.e., reforms) or can be enforced from outside (revolts); the likelihood of transitions is thereby determined endogenously. Within this framework, we

[*] This chapter is based on joint work with Robert Ulbricht.

address the following key questions. Which types of political systems arise from reforms, and which arise from revolts? Similarly, through which of these transition mechanisms are particular systems such as democracies most likely to emerge? And how frequently is either type of transition observed depending on the political system in place?

Model overview Our modeling approach aims to resemble the key mechanisms behind political transitions explored in the literature, but generalizes them in order to ensure the co-existence of reforms and revolts and to endogenize their outcomes.

To endogenize political systems that emerge after revolts, we dispense with the simplifying approach of a representative "political outsider". Instead we consider an economy in which agents that are excluded from political power are heterogeneously adapted to the current regime. As is standard in the literature, political outsiders can attempt to acquire political power by supporting a subversive attempt against the regime. Prospects of subverting depend on, first, an unobserved ability of the regime to withstand such an attempt and, second, the total mass of outsiders supporting it. For deciding whether or not to support a revolt, agents weight these prospects against their individual adaptation utility to the current regime. As a consequence, a coordination game similar to the literature on global games endogenously determines the regime type after a successful revolt.[1]

Reforms are modeled as in the seminal paper by Acemoglu and Robinson (2000b) in that members of the current regime ("political insiders") may conduct preemptive reforms in order to alleviate the threat from a revolt. However, we generalize their original approach by permitting political insiders to enfranchise an arbitrary fraction of the population, allowing for a continuum of *a priori* unspecified political systems to emerge from these reforms.

Finally, we assume that while insiders are perfectly informed about their ability to withstand a revolt, outsiders are strictly less informed about the prospects of subverting. As a consequence, conducting reforms will be endogenously associated with being intrinsically weak, which in equilibrium helps outsiders to coordinate their actions. This effectively increases the costs of reforming and provides an incentive for weak regimes to take tough stance rather than to negotiate on moderate reforms. Because, in equilibrium, excessive repression translates into a substantial risk to be overthrown, asymmetric information, crucially, ensures the co-existence of reforms and revolts along the equilibrium path and allows us to jointly analyze these two transition mechanisms in our model.

[1]Although outsiders in our model share the same amount of information, we use heterogeneous opportunity costs to ensure that subverting and not subverting is always a dominant strategy for some outsiders in our model. Iterated elimination of (interim) dominated strategies then gives rise to a unique outcome of this coordination game. This is essentially the same mechanism that determines equilibria in global games.

Results Our first set of findings characterizes the political systems that endogenously arise in equilibrium. We show that while revolts result in autocracies where a minority of the population forms the ruling class, political reforms enfranchise the majority of the population and establish democratic political systems. Intermediate types of political regimes, by contrast, do not arise along the equilibrium path, so that political systems tend to be extreme.

Furthermore, this first set of results implies that democracies are only established from within regimes, giving theoretical support to a long-standing view in political science according to which members of former autocracies are key actors in the establishment of democracies (Rustow, 1970; O'Donnell and Schmitter, 1973; Huntington, 1991). Or, as Karl (1990, p. 8) puts it: “no stable political democracy [in South America] has resulted from regime transitions in which mass actors have gained control, even momentarily, over traditional ruling classes”.

Our second set of results concerns the stability and persistence of political systems. From our analysis it follows that democratic regimes are intrinsically stable, characterized by long episodes without political change. In contrast, autocracies are subject to frequent regime changes—either via revolts or reforms. This is in line with the empirical literature on regime stability, which observes that democratic political systems are significantly more stable than autocratic ones (Przeworski, 2000; Gates et al., 2006; Magaloni and Kricheli, 2010).[2]

Nevertheless, our findings suggest that despite their instability, autocratic systems are persistent over time. This is because even though single autocratic regimes are relatively short lived, political change is frequently initiated by a small group of insurgents, resulting in autocracies very similar to their predecessors. Interestingly, this reasoning further implies that revolts tend to be serially correlated over time as they go along with a selection into politically instable regimes, leading to periods of political instability.

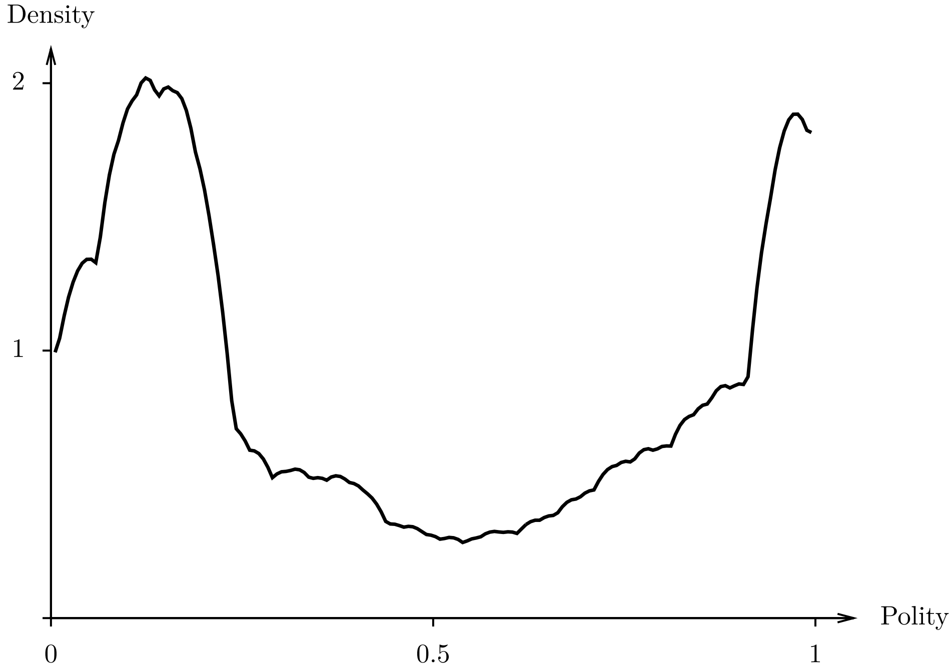
In combination, our results imply that the long-run distribution of political systems is double hump-shaped with mass concentrated on extreme political systems. Our model thus provides a foundation to the empirically observed distribution of political systems since World War I, plotted in Figure 3.1.[3] Taking a look at the underlying dataset (for

[2]From these results it follows that the mode of transition—peaceful reforms or violent revolts—is important for the characteristics of the resulting regimes. For transitions to democracy, a similar point has been highlighted by Cervellati et al. (2007, 2011), who show that consensual transitions foster civil liberties and property rights provision in contrast to violent transitions.

[3]The underlying data is taken from the Polity IV Project (for details, see Section 3.5). It has been disputed whether intermediate scored regimes on this index should nevertheless be classified as either democratic or autocratic due to nonlinearities in the index (Cheibub et al., 2010). Note, however, that this is to say that different measurements would only lead to more mass on the extremes, not altering the basic conclusion drawn for our purposes.

Figure 3.1: Distribution of Political Systems since World War I

Political systems range from extremely autocratic (0) to extremely democratic (1). Units of observation are country-days.



details, see Section 3.5), we also find similar support for the findings outlined above.

Related literature So far, the literature on political transitions has primarily focused on developing arguments for why autocratic regimes may conduct democratic reforms. Bourguignon and Verdier (2000), Lizzeri and Persico (2004), and Llavador and Oxoby (2005) argue that reforms are reflective of situations where autocratic decision makers are better off in a democratized political system than under the status quo. A number of other studies are based on the idea of preemptive reforms introduced by Acemoglu and Robinson (2000b) (e.g., Conley and Temini, 2001; Boix, 2003). These papers share with ours the basic logic behind reforms; i.e., autocratic regimes may use political reforms to credibly commit to redistribution and to reduce revolutionary pressure.[4]

In contrast to these papers, the emphasis of this analysis is on the dynamics of political transitions, including but not restricted to democratization. In this respect, this study relates more closely to Acemoglu and Robinson (2001) and Acemoglu et al. (2010), who consider settings where preemptive reforms co-exist with coups along the equilibrium path, and to Ellis and Fender (2011), who consider preemptive reforms that co-exist with mass revolutions. In particular, Ellis and Fender choose a similar approach in studying how autocracies may strategically manipulate the degree of subversive coordination in

[4]See Aidt and Jensen (2012) and Przeworski (2009) for empirical studies suggesting that subversive threats are indeed the driving force behind democratization.

the presence of asymmetric information. In their model, outsiders sequentially choose whether or not to support a subversive attempt, which succeeds only if it is unanimously supported. They find that asymmetric information provides an incentive to refrain from stabilizing reforms despite the presence of revolutionary pressure (see also Acemoglu and Robinson, 2000a; and for information manipulation in global games, see Angeletos et al., 2006 and Edmond, 2011).

However, all of these papers have in common that they exogenously restrict the set of political systems that result from transitions. In contrast, our approach of an unrestricted polity space leaves the outcomes of reforms and revolts unspecified. This is central to our analysis, allowing us to endogenously derive the properties of these transition mechanisms and to analyze their implications for the stability and persistence of political systems.

We also relate to Justman and Gradstein (1999), Jack and Lagunoff (2006), and Gradstein (2007), who study the incentives of political regimes to conduct democratic reforms in frameworks, in which—as in our approach—continuous extensions of the franchise are possible. Similar to the literature discussed above, these authors provide conditions under which (possibly gradual) extensions of the franchise are to be expected. In contrast to our work, however, they do not allow for political change to be initiated from political outsiders (via revolts), preventing them from analyzing transition dynamics in the generality that follows from the interplay between reforms and revolts, which is at the core of our contribution.

Outline The remainder of the chapter is organized as follows. Section 3.2 introduces the model economy. In Section 3.3, we characterize the equilibrium and illustrate the strategic considerations determining political transitions. The law of motion of the dynamic economy and our main predictions are derived in Section 3.4. In Section 3.5, we present some empirical evidence, and Section 3.6 concludes.

3.2 The Model

We consider an infinite horizon economy with a continuum of two-period lived agents. Each generation has a mass equal to 1. At time t, fraction λ_t of the population has the power to implement political decisions, whereas the remaining agents are excluded from political power. We refer to these two groups as (political) "insiders" and "outsiders".

When born, the distribution of political power among the young is inherited from their parent generation; that is, λ_t agents are born as insiders, while $1 - \lambda_t$ agents are born as outsiders. However, agents who are born as outsiders can attempt to overthrow the current regime and thereby acquire political power. To this end, outsiders choose individually and simultaneously whether or not to participate in a revolt.[5] Because we will assume that all political change takes effect at the beginning of the next period, only young outsiders have an interest in participating in a revolt. Accordingly, we denote young outsider i's choice by $\phi_{it} \in \{0, 1\}$ and use the aggregated mass of supporters, $s_t = \int \phi_{it}\, di$, to refer to the size of the resulting revolt.

The probability that a revolt is successful is given by

$$p(\theta_t, s_t) = \theta_t h(s_t), \tag{3.1}$$

where $\theta_t \in \Theta$ is a random state of the world that reflects the vulnerability of the current regime or their ability to put down a revolt, and h is an increasing and twice differentiable function, $h : [0, 1] \to [0, 1]$, with $h(0) = 0$. That is, the threat of a revolt to the current regime is increasing in the mass of its supporters and in the vulnerability of the regime. When a revolt has no supporters ($s_t = 0$) or the regime is not vulnerable ($\theta_t = 0$), it fails with certainty.

The purpose of θ_t in our model is to introduce asymmetric information between insiders and outsiders that, as will become clear below, explains the prevalence of revolts along the equilibrium path. Formally we have that the state θ_t is uniformly distributed on $\Theta = [0, 1]$, is i.i.d. from one period to the next, and is revealed to insiders at the beginning of each period. Outsiders only know the prior distribution of θ_t.

After they learn θ_t, insiders may try to alleviate the threat of revolt by conducting reforms. We follow Acemoglu and Robinson (2000b) by modeling these reforms as an extension of the franchise to outsiders, which is effective in credibly preventing them from supporting a revolt.[6] However, since our model is aimed at endogenizing the polity λ_t, we generalize this mechanism by allowing insiders to continuously extend the regime by any fraction, $x_t - \lambda_t$, of young outsiders, where $x_t \in [\lambda_t, 1]$ is the reformed political system.[7] Because preferences of insiders will be perfectly aligned, there is no need to specify the decision making process leading to x_t in detail.

Given the (aggregated) policy choices s_t and x_t, and conditional on the outcome of a

[5]For notational convenience, we abstract from the possibility of insiders participating in a revolt. In Appendix C.1.1, however, we show that this is without loss of generality, since it is never optimal for insiders to support a revolt against fellow members of the regime.

[6]As argued in Footnote 5 and shown in Appendix C.1.1, it is indeed individually rational for enfranchised outsiders to not support a revolt.

[7]Note that by assuming $x_t \in [\lambda_t, 1]$, we are ruling out reforms that withdraw political power once it has been granted. This is in line with the idea that granting someone the status of an insider is a credible and irreversible commitment in the logic of Acemoglu and Robinson (2000b).

revolt, the political system evolves as follows:

$$\lambda_{t+1} = \begin{cases} s_t & \text{if the regime is overthrown, and} \\ x_t & \text{otherwise.} \end{cases} \tag{3.2}$$

When a revolt fails (indicated by $\eta_t = 0$), reforms take effect and the old regime stays in power. The resulting political system in $t+1$ is then given by x_t. In the complementary case, when a revolt succeeds ($\eta_t = 1$), those who have participated will form the new regime. Accordingly, after a successful revolt, the fraction of insiders at $t+1$ is equal to s_t. Note that this specification prevents non-revolting outsiders from reaping the benefits from overthrowing a regime; there are no gains from free-riding in our model.[8]

To complete the description of our model, we still have to specify how payoffs are distributed across the different groups of agents at t. As for outsiders, we assume that they receive a constant per period payoff of γ_{it} which is privately assigned to each agent at birth and is drawn from a uniform distribution on $[0,1]$. We interpret this heterogeneity of outsiders as different degrees of economical or ideological adaptation to a regime, determining their propensity to revolt.

In contrast, insiders enjoy per period payoffs $u(\lambda_t)$, where u is twice differentiable, $u' < 0$, and $u(1)$ is normalized to unity. One should think of $u(\cdot)$ as a reduced form function that captures the various benefits of having political power (e.g., from extracting a common resource stock, implementing preferred policies, etc.). One important feature of u is that it is decreasing in the current regime size and, hence, extending the regime is costly for insiders (e.g., because resources have to be shared, or preferences about policies become less aligned). Another thing to note is that $u(\lambda_t) \geq \gamma_{it}$ for all λ_t and γ_{it}; that is, being part of the regime is always desirable. In the case of full democracy ($\lambda_t = 1$) all citizens are insiders and enjoy utility normalized to the one of a best-adapted outsider (i.e., $u(1) = 1$).

To simplify the analysis, we assume that members of an overthrown regime and participants in a failed revolt are worst-adapted to the new regime. Formally, $\gamma_{it} = 0$, resulting in zero payoff.

For the upcoming analysis it will be convenient to define the expected utility of agents

[8] The theoretical possibility for free-riding arises since we depart from the common assumption of treating the opposition as a single player in order to endogenize the political system resulting from a revolt. However, as long as there are some private benefits that provide incentives for outsiders to support a revolt, the working of this model is unaffected by (moderate) incentives to free-ride. Entirely abstracting from the collective action problem is merely a model simplification.

that are born at time t, which is given as follows:

$$V^I(\theta_t, \lambda_t, s_t, x_t) = u(\lambda_t) + [1 - p(\theta_t, s_t)] \times u(x_t), \tag{3.3}$$

$$V^O(\theta_t, \gamma_{it}, s_t, \phi_{it}) = \gamma_{it} + \phi_{it}\, p(\theta_t, s_t) \times u(s_t) + (1 - \phi_{it}) \times \gamma_{it}, \tag{3.4}$$

where superscript I and O denote agents that are born as insiders and outsiders, respectively. In both equations, the first term corresponds to the first period payoff (unaffected by the policy choices of the young agent's generation), while the other terms correspond to second period payoffs. (Since agents do not face an intertemporal tradeoff, we do not need to define a discount rate here).

The timing of events within one period can be summarized as follows:

1. The state of the world θ_t is revealed to insiders.
2. Insiders may extend political power to a fraction $x_t \in [\lambda_t, 1]$ of the population.
3. Outsiders individually and simultaneously decide whether or not to participate in a revolt.
4. Transitions according to (3.1) and (3.2) take place, period $t+1$ starts with the birth of a new generation, and payoffs determined by λ_{t+1} are realized.

In what follows, we characterize the set of perfect Bayesian equilibria that satisfy the trembling-hand criterion (due to Selten, 1975); that is, perfect Bayesian equilibria that are the limit of some sequence of perturbed games in which strategy profiles are constrained to embody "small" mistakes.[9] To increase the predictive power of our model, we thereby limit attention to equilibria that are consistent with the D1 criterion introduced by Cho and Kreps (1987), a standard refinement for signaling games. The D1 criterion restricts outsiders to believe that whenever they observe a reform x' that is not conducted in equilibrium, the reform has been implemented by a regime with vulnerability θ', for which a deviation to x' would be most attractive.[10]

[9] Here, the concept of trembling-hand perfection rules out "instable" equilibria, in which $s_t = 0$, but iteratively best-responding to a (perceived) second-order perturbation of s_t would lead to a different equilibrium with a first-order change in s_t. For details see the proof of Proposition 3.1. Except for these instabilities, the set of trembling-hand perfect equilibria coincides with the set of perfect Bayesian equilibria in our model. An alternative approach to rule out these instabilities would be to restrict attention to equilibria which are the limit to a sequence of economies with a finite number of outsiders, where each agent's decision has non-zero weight on s_t.

[10] Formally, let $\bar{V}^I(\theta', \lambda_t)$ be the insiders' payoff in a candidate equilibrium when the regime has a vulnerability θ'. Then the D1 criterion restricts beliefs to the state θ' that maximizes $D_{\theta',x'} = \{\hat{\theta} : V^I(\theta', \lambda_t, s(\hat{\theta}, x'), x') \geq \bar{V}^I(\theta', \lambda_t)\}$, where $s(\hat{\theta}, x')$ is the mass of outsiders supporting a revolt, given the beliefs $\hat{\theta}$ and reform x'. $D_{\theta',x'}$ is maximal here, if there is no θ'', such that $D_{\theta',x'}$ is a proper subset

Anticipating our results, we simplify our notation as follows. First, outsiders' beliefs regarding the regime's vulnerability will be uniquely determined in our setup. We therefore denote the commonly held belief by $\hat{\theta}_t$, dropping the index i. Second, there are no non-degenerate mixed strategy equilibria in our game. Accordingly, we restrict the notation in the main text to pure strategies and introduce mixed strategies only to define the perturbations required by trembling-hand perfection.

This leads to the following definition of equilibrium for our economy.

Definition. Given a history $\delta = \{\lambda_0\} \cup \{\{\phi_{i\tau} : i \in [0,1]\}, \theta_\tau, x_\tau, \eta_\tau\}_{\tau=0}^{t-1}$, an equilibrium in this economy consists of policy mappings $x_\delta : (\theta_t, \lambda_t) \mapsto x_t$ and $\{(\phi_{i\delta} : (\hat{\theta}_t, x_t) \mapsto \phi_{it}) : i \in [0,1]\}$, and beliefs $\hat{\theta}_\delta(\lambda_t, x_t) \mapsto \hat{\theta}_t$, such that for all possible histories δ:

a. Reforms x_δ maximize insider's utility (3.3), given states (θ_t, λ_t), beliefs $\hat{\theta}_\delta$, and perturbed policy mappings $\{\omega_{i\delta}^k : i \in [0,1]\}$ for all values of k;

b. Each outsider's policy choice $\phi_{i\delta}$ maximizes (3.4), given perturbed policy mappings σ_δ^k, $\{\omega_{j\delta}^k : j \in [0,1] \setminus i\}$, and corresponding beliefs $\hat{\theta}_\delta^k$ for all values of k;

c. Beliefs $\hat{\theta}_\delta = \lim_{k\to\infty} \hat{\theta}_\delta^k(x_t)$, where $\hat{\theta}_\delta^k$ are obtained using Bayes rule given σ_δ^k; and $\hat{\theta}_\delta$ satisfies the D1 criterion;

d. States (λ_t, η_t) are consistent with (3.1) and (3.2);

e. The perturbed policy mappings $\{\{\omega_{i\delta}^k : i \in [0,1]\}, \sigma_\delta^k\}_{k=0}^\infty$ are sequences of completely mixed strategy profiles converging to profiles that place all mass on $\{\phi_{i\delta} : i \in [0,1]\}$ and x_δ, respectively.

3.3 Political Equilibrium

In this section, we characterize the political equilibrium in the model economy. Our analysis will be simplified considerably by the overlapping generations structure of our model, which gives rise to a sequence of "generation games" between young insiders and young outsiders. Since the distribution of political power at time t captures all payoff-relevant information of the history up to t, the only link between generations is λ_t. We can therefore characterize the set of equilibria in our model by characterizing the equilibria of the generation games as a function of λ_t. All other elements of the history up to

of $D_{\theta'',x'}$. That is, beliefs are attributed to the state in which a deviation to x' is attractive for the largest set of possible inferences about the regime's vulnerability (implying that the regime gains most by deviating).

time t may affect the equilibrium at t only by selecting between multiple equilibria of the generation game.

The generation game consists of two stages that determine the polity at $t+1$. First, outsiders have to choose whether or not to support a revolt. Because the likelihood that a revolt succeeds depends on the total mass of its supporters, outsiders face a coordination problem in their decision to revolt. Second, prior to this coordination problem, insiders decide on the degree to which political power is extended to outsiders. On the one hand this will decrease revolutionary pressure along the extensive margin by contracting the pool of potential insurgents. However, extending the regime may also contain information about the regime's vulnerability. As a result, reforms may increase revolutionary pressure along the intensive margin by increasing coordination among outsiders who are not subject to reforms. Insiders' policy choices will therefore be governed by signaling considerations.

We proceed by backward induction in solving for the equilibrium, beginning with the outsiders' coordination problem.

3.3.1 Stage 2: Coordination Among Outsiders

Consider the outsiders' coordination problem at time t. For any given belief, $(\hat{\theta}_t, \hat{s}_t) \in \Theta \times [0,1]$, individual rationality requires all outsiders to choose a ϕ_{it} that maximizes their expected utility $\mathbb{E}_t\{V^O(\cdot)\}$.[11] At time t, outsider i with adaptation utility γ_{it} will therefore participate in a revolt if and only if

$$\gamma_{it} \leq p(\hat{\theta}_t, \hat{s}_t)\, u(\hat{s}_t) \equiv \bar{\gamma}(\hat{s}_t). \tag{3.5}$$

Here $\bar{\gamma}(\hat{s}_t)$ is the expected benefit of participating in a revolt that is supported by a mass of $\hat{s}_t$ outsiders. Since $\bar{\gamma}(\hat{s}_t)$ is independent of γ_{it}, it follows that in any equilibrium the set of outsiders who support a revolt at t is given by the agents who are least adapted to the current regime. Suppose for the time being that $\bar{\gamma}(\hat{s}_t) \leq 1$. Then, $\bar{\gamma}(\hat{s}_t)$ defines the fraction of young outsiders that participates in a revolt, and, therefore, the size of a revolt, s_t, that would follow from $\bar{\gamma}(\hat{s}_t)$ is given by

$$f(\hat{s}_t) \equiv (1 - x_t)\, \bar{\gamma}(\hat{s}_t). \tag{3.6}$$

Further note that in any equilibrium it must hold that $s_t = \hat{s}_t$. Therefore, as long as

[11] Note that by our specification of p, V^O is linear in θ_t, and thus $\mathbb{E}_t\{V^O(\theta_t, \cdot)\} = V^O(\hat{\theta}_t, \cdot)$, where $\hat{\theta}_t \equiv \mathbb{E}_t\{\theta_t\}$. That is, the expected value of θ_t, given the posterior distribution of θ_t (outsiders' beliefs), is a sufficient statistic for computing V^O. Henceforth we define $\hat{\theta}_t$ accordingly, disregarding any higher moments of outsiders' beliefs.

$\bar{\gamma}(\hat{s}_t) \leq 1$, the share of outsiders that support a revolt at t has to be a fixed point to (3.6). To guarantee that this is always the case and to further ensure that a well-behaved fixed point exists, we impose the following assumption.

Assumption 3.1. *For $\psi(s) \equiv h(s) \cdot u(s)$,*

a. $\psi' \geq 0$ *and* $\psi'' \leq 0$*;*

b. $\lim_{s \to 0} \psi'(s) = \infty$.

Intuitively, Assumption 3.1 states that participating in a revolt becomes more attractive if the total share of supporters grows. This requires that the positive effect of an additional supporter on the success probability outweighs the negative effect of being in a slightly larger regime after a successful revolt. Put differently, 3.1 states that the participation choices of outsiders are strategic complements. To ensure existence, we further require that the strategic complementarity is sufficiently strong when a revolt is smallest, and is decreasing as it grows larger.

Using 3.1, the above discussion leads to the following proposition.

Proposition 3.1. *In any equilibrium, the mass of outsiders supporting a revolt at time t is uniquely characterized by a time-invariant function, $s : (\hat{\theta}_t, x_t) \mapsto s_t$, which satisfies $s(0, \cdot) = s(\cdot, 1) = 0$, increases in $\hat{\theta}_t$, and decreases in x_t.*

All formal proofs are relegated to Appendix C. Proposition 3.1 establishes the already discussed tradeoff of conducting reforms: On the one hand, reforms reduce support for a revolt along the extensive margin. In the limit, as regimes reform to a full-scaled democracy, any subversive threat is completely dissolved. On the other hand, if reforms signal that the regime is vulnerable, they may backfire by increasing support along the intensive margin.

3.3.2 Stage 1: Policy Choices of Insiders

We now turn to the insiders' decision problem. Since more vulnerable regimes have higher incentives to reform than less vulnerable ones, conducting reforms will shift beliefs towards being vulnerable and, therefore, indeed stipulate coordination among outsiders who are unaffected by reforms. This generates the tradeoff established in Proposition 3.1, which is the main driving force behind the following result.

Proposition 3.2. *In any equilibrium, policy choices of insiders and beliefs of outsiders are uniquely characterized by time-invariant functions $x : (\theta_t, \lambda_t) \mapsto x_t$ and $\hat{\theta} : (\lambda_t, x_t) \mapsto \hat{\theta}_t$,* *such that*

$$x(\theta_t, \lambda_t) = \begin{cases} \lambda_t & \text{if } \theta_t < \bar{\theta}(\lambda_t) \\ \xi(\theta_t) & \text{if } \theta_t \geq \bar{\theta}(\lambda_t), \end{cases}$$

and

$$\hat{\theta}(\lambda_t, x_t) = \begin{cases} \bar{\theta}(\lambda_t)/2 & \text{if } x_t = \lambda_t \\ \bar{\theta}(\lambda_t) & \text{if } \lambda_t < x_t < \xi(\bar{\theta}(\lambda_t)) \\ \xi^{-1}(x_t) & \text{if } \xi(\bar{\theta}(\lambda_t)) \leq x_t \leq \xi(1) \\ 1 & \text{if } x_t > \xi(1), \end{cases}$$

where ξ is a unique increasing function with $\xi(\theta_t) > \lambda_t + \mu$, and $\bar{\theta}(\lambda_t) > 0$ for all λ_t and some $\mu > 0$.

Proposition 3.2 defines insiders' policy choices for generation t as a function of (θ_t, λ_t). Because the logic behind these choices is the same for all values of λ_t, we can discuss the underlying intuition keeping λ_t fixed. Accordingly, in Figure 3.2 we plot reform choices (left panel) and the implied probability to be overthrown (right panel), sliced along a given λ_t plane. It can be seen that whenever a regime is less vulnerable than $\bar{\theta}(\lambda_t)$, insiders prefer to not conduct any reforms (i.e., $x_t = \lambda_t$), leading to a substantial threat for regimes with θ_t close to $\bar{\theta}(\lambda_t)$. Only if $\theta_t \geq \bar{\theta}(\lambda_t)$, reforms will be conducted ($x_t = \xi(\theta_t)$), which in equilibrium effectively mitigate the threat to be overthrown, ruling out marginal reforms where $\xi(\theta_t) \to \lambda_t$.

To see why marginal reforms are not effective in reducing revolutionary pressure consider Figure 3.3. Here we plot equilibrium beliefs (left panel) and the corresponding mass of insurgents (right panel) as functions of x_t. If the political system is left unchanged by insiders, outsiders only learn the average state $\bar{\theta}(\lambda_t)/2$ of all regimes that pool on $x_t = \lambda_t$ in equilibrium. On the other hand, every extension of the regime—how small it may be—leads to a non-marginal change in outsiders' beliefs from $\hat{\theta}_t = \bar{\theta}(\lambda_t)/2$ to $\hat{\theta}_t \geq \bar{\theta}(\lambda_t)$ and, hence, results in a non-marginal increase in revolutionary pressure along the intensive margin. It follows that there exists some $\tilde{x}(\lambda_t)$, such that for all $x_t < \tilde{x}(\lambda_t)$ the increase of pressure along the intensive margin dominates the decrease along the extensive margin. Thus, reforms smaller than $\tilde{x}(\lambda_t)$ will backfire and *increase* the mass of insurgents (as seen in the right panel of Figure 3.3), explaining why effective reforms have to be non-marginal.

Furthermore, optimality of reforms requires that the benefit of reducing pressure compensates for insiders' disliking of sharing power. Because $\tilde{x}(\lambda_t) - \lambda_t > 0$, it follows that $u(\tilde{x}(\lambda_t)) - u(\lambda_t) < 0$. Moreover, any reform marginally increasing the regime beyond

Figure 3.2: Equilibrium Reforms and Implied Probability to be Overthrown

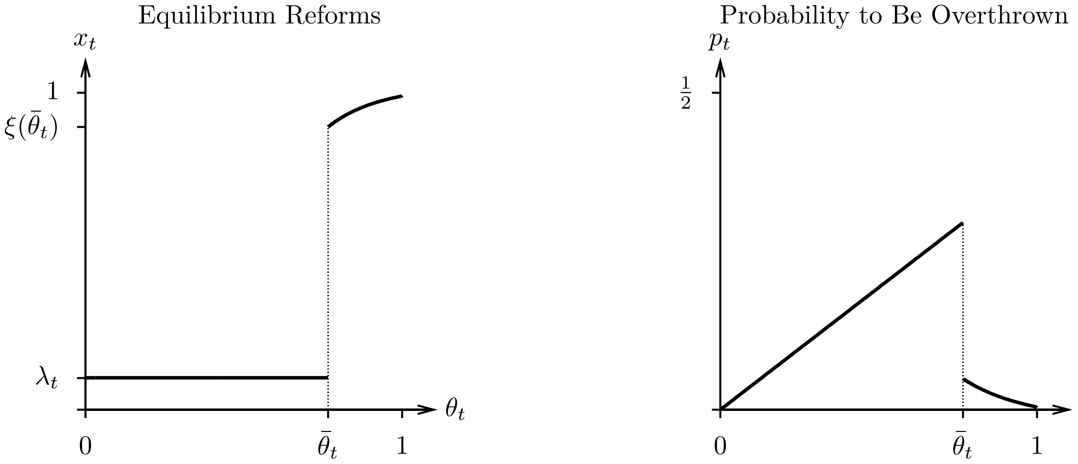
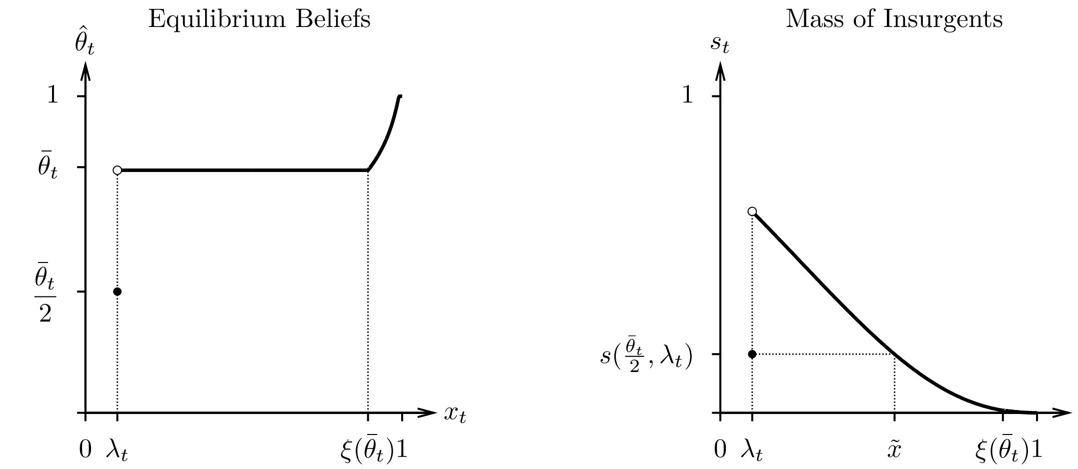


Figure 3.3: Equilibrium Beliefs and Implied Mass of Insurgents



$\tilde{x}(\lambda_t)$ leads only to a marginal increase in the likelihood to stay in power. Hence, there exists a non-empty interval, given by $[\tilde{x}(\lambda_t), \xi(\bar{\theta}(\lambda_t))]$, in which reforms are effective, yet insiders prefer to gamble for their political survival in order to hold on to the benefits of not sharing power in case they survive. This explains the substantial threat for regimes with θ_t close to $\bar{\theta}(\lambda_t)$, as seen in the right panel of Figure 3.2.[12]

3.3.3 Existence and Uniqueness of Equilibrium

Propositions 1 and 2 uniquely pin down the policy choices in every state, which in return determine the evolution of polities. We conclude that there is no scope for multiple equilibria in our model economy; if there exists an equilibrium it is unique. Verifying the

[12]More precisely, gambling for survival increases the likelihood to be overthrown in two ways. First, since on the margin it is more vulnerable regimes that join the pool at $x_t = \lambda_t$, these regimes obviously face a high threat by not conducting reforms. Second, since these regimes also shift the pooling belief towards more vulnerable, the threat further increases for regimes of all vulnerabilities in the pool.

existence then permits us to reach the following result.

Proposition 3.3. *There exists an equilibrium, in which for all histories δ, policy mappings x_δ and $\{\phi_{i\delta} : i \in [0,1]\}$, as well as beliefs $\hat{\theta}_\delta$ correspond to the time-invariant mappings given by Propositions 3.1 and 3.2. Furthermore, for any given initial political system λ_0, this equilibrium is unique.*

3.4 Transition Dynamics

In the preceding section, we have established that in the unique equilibrium, policy mappings are time-invariant, implying that (λ_t, θ_t) is a sufficient statistic for characterizing the transition dynamics of the political system from time t to $t+1$. Integrating out θ_t, political systems in our equilibrium follow a Markov process where the probability that $\lambda_{t+1} \in \Lambda$ is given by

$$\begin{aligned} Q(\lambda_t, \Lambda) = \rho^S(\lambda_t) \times Q^S(\lambda_t, \Lambda) + \rho^R(\lambda_t) \times Q^R(\lambda_t, \Lambda) \\ + \{1 - \rho^I(\lambda_t) - \rho^R(\lambda_t)\} \times \mathbb{1}_{\lambda_t \in \Lambda}. \end{aligned} \tag{3.7}$$

Here ρ^S and ρ^R denote the probabilities that a transition occurs via subversive attempts (i.e., revolts) and reforms, respectively; Q^S and Q^R are conditional transition functions; and $\mathbb{1}$ is an indicator function equal to unity whenever $\lambda_t \in \Lambda$.[13] Accordingly, the first term in (3.7) defines the probability that state $\lambda_{t+1} \in \Lambda$ emerges through a revolt, the second term defines the probability that $\lambda_{t+1} \in \Lambda$ emerges from a reform, and the third term refers to the event of no transition. Decomposing the law of motion into these conditional channels, we are now ready to state our main predictions.

3.4.1 Political Systems after Transition

By (3.7), political systems that arise after transitions are summarized by Q^S and Q^R. Our first result states that polities that emerge after reforms differ fundamentally from

[13] Formally, we have that

$$\begin{aligned} \rho^S(\lambda_t) &= \textstyle\int_0^1 \grave{p}(\theta)\, d\theta \\ \rho^R(\lambda_t) &= \textstyle\int_{\bar{\theta}(\lambda_t)}^1 \{1 - \grave{p}(\theta)\}\, d\theta \\ Q^S(\lambda_t, \Lambda) &= \{\rho^S(\lambda_t)\}^{-1} \textstyle\int_{\theta:\grave{s}(\theta)\in\Lambda} \grave{p}(\theta)\, d\theta \\ Q^R(\lambda_t, \Lambda) &= \{\rho^R(\lambda_t)\}^{-1} \textstyle\int_{\theta:\grave{x}(\theta)\in\Lambda\setminus\lambda_t} \{1 - \grave{p}(\theta)\}\, d\theta, \end{aligned}$$

where $\grave{x}(\theta) \equiv x(\lambda_t, \theta)$, $\grave{s}(\theta) \equiv s(\hat{\theta}(\lambda_t, \grave{x}(\theta)), \grave{x}(\theta))$, and $\grave{p}(\theta) \equiv p(\theta, \grave{s}(\theta))$.

those that emerge from revolts. The following proposition states the formal result.

Proposition 3.4. *For all states λ_t,*

$$Q^R(\lambda_t, (\tfrac{1}{2}, 1]) = 1 \qquad \textit{and} \qquad Q^S(\lambda_t, (0, \tfrac{1}{2})) = 1;$$

i.e., reforms lead to majority regimes with $\lambda_{t+1} > \frac{1}{2}$ and revolts lead to minority regimes with $\lambda_{t+1} < \frac{1}{2}$.

The first part of Proposition 3.4 states that any reform leads to a democratic polity, in which the majority of citizens holds political power. The intuition for this result mirrors the one for Proposition 3.2. Because conducting reforms will be associated with being intrinsically weak, coordination is increased along the intensive margin. For the benefits along the extensive margin to justify these costs, reforms therefore have to be far-reaching, leading to the enfranchising of the majority of the population.

In contrast, the second part of Proposition 3.4 establishes that successful revolts always lead to minority regimes, in which a small elite rules over a majority of political outsiders. Underlying this result is that in equilibrium subversive attempts are conducted by only a small group of insurgents. Mass revolutions on the other hand are off-equilibrium. To see what drives the result, first note that rationality of reforms implies that revolts are largest when regimes abstain from reforms and choose to repress the population. However, because abstaining from reforms is optimal, both, in times when regimes are strong and when they hide their weakness through taking tough stance, uncertainty about a regime's weakness is largest from the perspective of outsiders exactly when a regime abstains from reforms. Accordingly, prospects of revolting are only moderate and only those with large gains from winning political power (i.e., outsiders who are least adapted to the current regime) will find it rational to take the risk of revolting.

An interesting implication of Proposition 3.4 is that democratic regimes arise if and only if it is optimal for the regime to enfranchise former political outsiders. The commonly made assumption in the previous literature that democracies are established by means of reforms conducted by the elites is thus an endogenous outcome of our model. The other channel through which democracies hypothetically could be established are mass revolutions. Their severe threat, however, is always mitigated by rational regimes, such that mass revolts are events off the equilibrium path. This observation gives support to a long-standing view in political science, according to which members of former autocracies are key actors in the establishment of democracies, which is based on, e.g., the observation of Karl (1990, p. 8) that no stable South American democracy has been the result of mass revolutions (see also Rustow, 1970; O'Donnell and Schmitter, 1973; Huntington, 1991).

Finally, note that from Proposition 3.4 it follows that there is a (possibly quite large) open interval $\bar{\Lambda}$ around 1/2, such that $Q(\lambda_t, \bar{\Lambda}) = 0$ for all λ_t. That is, there is a range of intermediate regimes that are completely off the equilibrium path, suggesting a long-run distribution with mass only on the extremes. In a parametric example below, we will see that this is indeed the case.

3.4.2 Probabilities of Transition

The next proposition describes how the likelihood of either type of political transition depends on the political system λ_t.

Proposition 3.5. *For all $\lambda_t > \bar{\lambda}$, $\partial\rho^S/\partial\lambda_t < 0$ and $\partial\rho^R/\partial\lambda_t \leq 0$; and for all $\lambda_t < \underline{\lambda}$, $\partial\rho^S/\partial\lambda_t < 0$ and $\partial\rho^R/\partial\lambda_t > 0$ if $\lim_{\lambda\to 0} \partial u/\partial\lambda < \underline{u}$, and $\partial\rho^S/\partial\lambda_t > 0$ and $\partial\rho^R/\partial\lambda_t < 0$ if $\lim_{\lambda\to 0} \partial u/\partial\lambda > \bar{u}$, and some $(\bar{\lambda}, \underline{\lambda}, \bar{u}, \underline{u}) \in [0,1)^2 \times \mathbb{R}^2_-$, whereas $\bar{\lambda} \geq \underline{\lambda} > 0$ if $\bar{\theta}(0) < 1$.*

From Proposition 3.5 it follows that as regimes become more democratic, they eventually become more stable. This is generally true for polities, in which no reforms are conducted; and further holds for sufficiently democratic regimes ($\lambda_t > \bar{\lambda}$). For autocratic polities, in contrast, properties of the likelihood of political change depend on the exact specification of u. Still, Proposition 3.5 suggests that ρ^R and ρ^S are hump-shaped when marginal reforms for autocratic regimes are very costly or rather cheap, respectively. Otherwise, the likelihood for either type of transition tends to be decreasing as the political system becomes more democratic.

3.4.3 A Parametric Example

To illustrate the dynamics implied by Propositions 3.4 and 3.5 and to further study the implications of the model in the long-run, we now introduce a parametrized version of our model economy. We choose the following functional forms,

$$h(s_t) = s_t^{\alpha}$$

and

$$u(\lambda_t) = -\exp(\beta_1 \lambda_t) + \beta_0.$$

Here one may think of β_0 as a common resource stock or some other type of private benefits, which decline at a exponential rate β_1 as power is shared with more agents. To pin down the free parameters, we further assume that $\psi'(1) = 0$; i.e., the strategic effect of an additional outsider supporting a revolt becomes negligible when revolts are supported by the full population. Together with our assumptions on u and h, this pins down α and β_0 in terms of β_1, which is restricted to approximately satisfy $\beta_1 \in (0, 0.56)$.[14]

Intuitively, β_1 measures the costs of enfranchising political outsiders. In practice, these costs are expected to be high if members of the regime have access to a large pool of resources, or if there is a large degree of economic and political inequality.[15] Thus, when β_1 is close to its upper bound, extending the franchise is costly and the incentives to gamble for survival are strong. Consequently, for large β_1, one should expect to observe revolts frequently in equilibrium. On the other hand, if β_1 is low, conducting reforms is cheap and one should expect political insiders to quickly reform to a fully integrated society.

To give an overview of the transition dynamics, Figure 3.4 displays a simulated time series of the model economy for different values of β_1 and for 500 periods each. For each time path, we plot the political system, λ_t, at time t and indicate the dates where transitions occur via revolts (marked by Δ) and reforms (marked by $\times$). It can be seen that low costs of reforms in Setting 1 ($\beta_1 = 0.35$) result in immediate democratic reforms and the absence of successful subversive attempts. As the costs of reforms are increasing in Setting 2 ($\beta_1 = 0.40$) and Setting 3 ($\beta_1 = 0.45$), successful revolts become more frequent and are followed by periods of frequent regime changes, where autocracies succeed each other. In contrast, democratic reforms give rise to long periods of political stability.

Polarization Although Figure 3.4 is the result of a random simulation, it captures many essential transition dynamics that arise in our model. First, in line with Proposition 3.4, it can be seen that transitions lead to a polarization of regimes; i.e., revolts lead to autocratic regimes, whereas reforms result in fairly inclusive democracies. A more complete picture is provided by Figure 3.5, which displays the distribution of political systems that emerge from each transition mechanism for $\beta_1 = 0.4$.[16] From the left panel, it becomes apparent that approximately two different types of autocracies emerge after revolts: dictatorships, corresponding to regimes that emerge after revolts against democracies, and autocracies which emerge after succeeding other autocracies. From the right panel of Figure 3.5, it becomes apparent that reforms lead to democratic political systems where political power

[14] The implied values for the other two parameters are $\alpha = \beta_1 \exp(\beta_1)$ and $\beta_0 = \exp(\beta_1) + 1$, restricting $\beta_1 \in (0, \exp(-\beta_1)) \approx (0, 0.56)$.

[15] In particular, note that $u(\lambda) = \exp(\beta_1) - \exp(\beta_1 \lambda) + 1$ is increasing in β_1 for all λ, so that also the inequality between insiders and the average outsider, $\int (u(\lambda) - \gamma)\, d\gamma$, is increasing in β_1 for all λ.

[16] For computing the distributions, originating polities are weighted by their long-run distribution Ψ; e.g., the distribution of polities after reforms is given by $pdf(\lambda_{t+1}) = \int_0^1 Q^R(\lambda_t, \lambda_{t+1})\, d\Psi(\lambda_t)$. While the long-run distribution itself varies considerably with β_1 (see also Figure 3.8), the conditional distributions displayed in Figure 3.5 remain largely unaffected by changes in β_1.

Figure 3.4: Simulated Time Series of the Model Economy

Reforms are marked by “×”, successful revolts are marked by “△”.

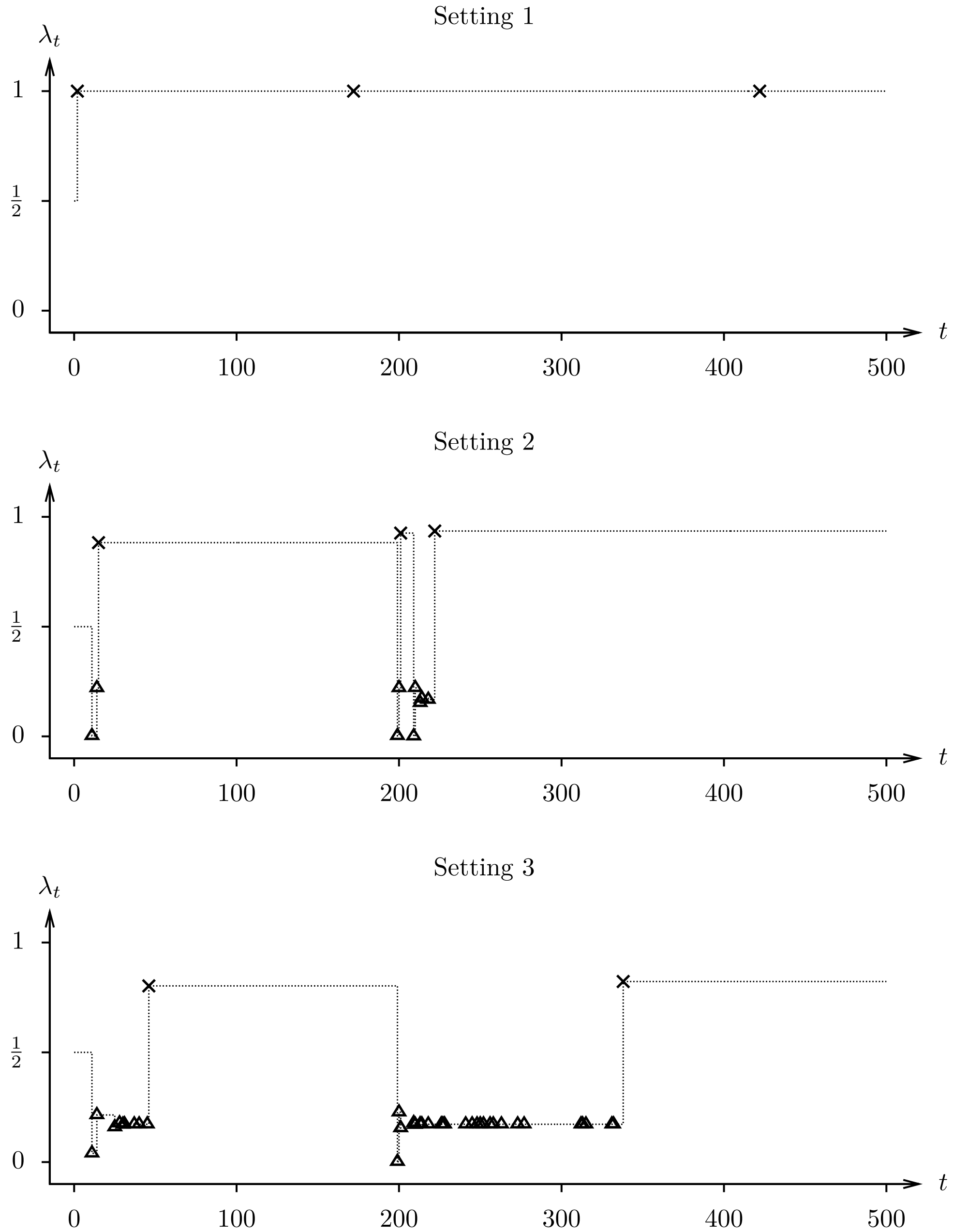


Figure 3.5: Distribution of Political Systems after Revolts and Reforms

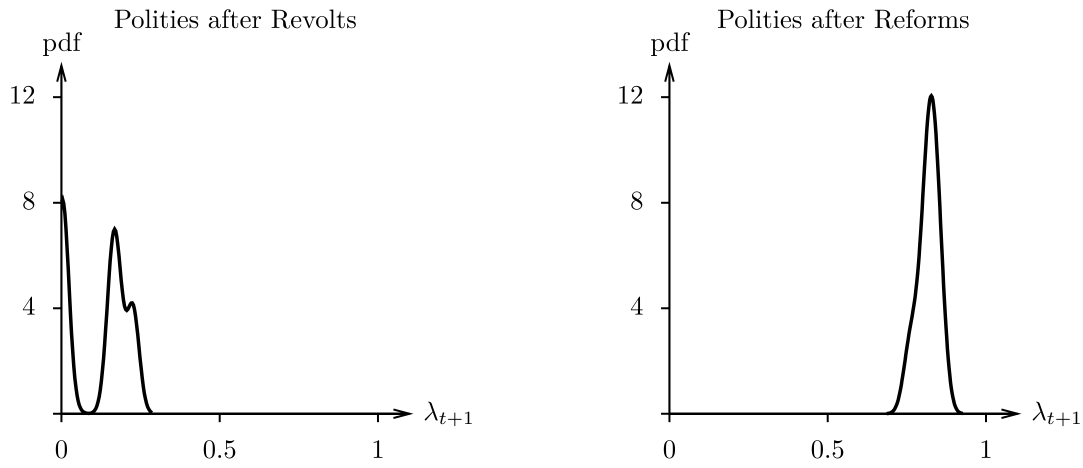
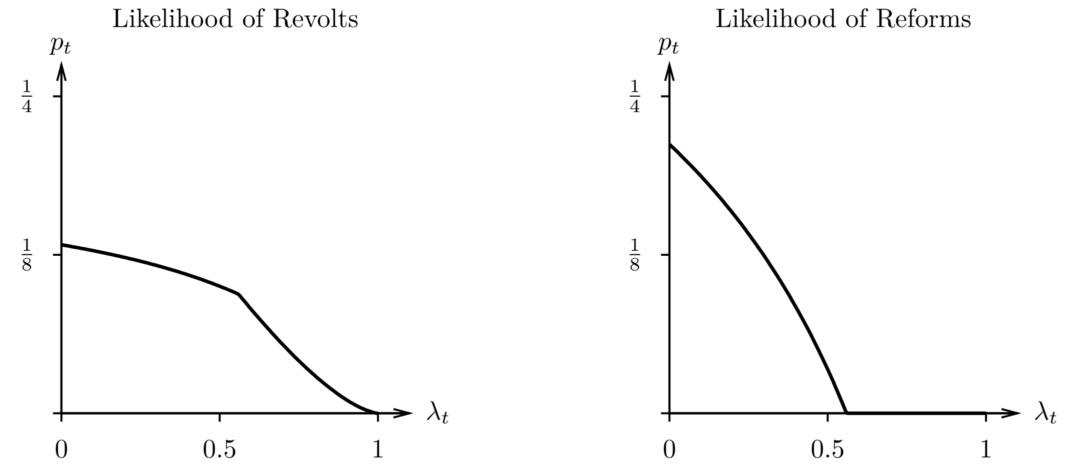


Figure 3.6: Likelihood of Revolts and Reforms



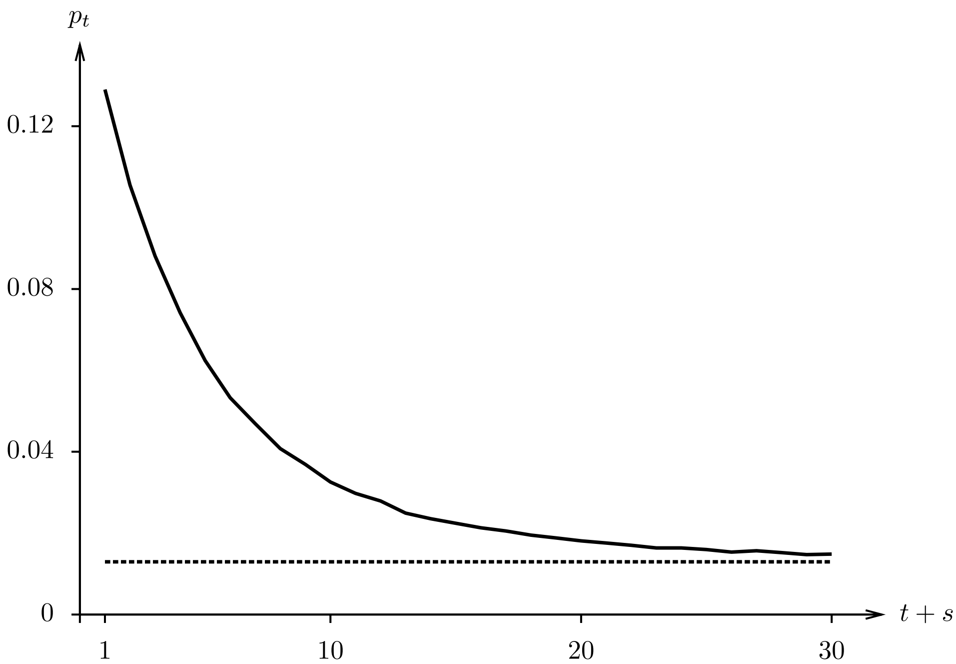
is shared among the majority of the population. Furthermore, it can be seen that a large set of polities around 1/2 is neither emerging from reforms, nor from revolts.

Stability The second observation that can be drawn from the simulations in Figure 3.4 concerns the stability of political regimes. In line with Proposition 3.5, it is evident that democracies are characterized by long episodes without political change. In contrast, autocracies are subject to frequent regime changes. The underlying transition probabilities are depicted in Figure 3.6. Here we plot the likelihood of political transitions via revolts (left panel) and reforms (right panel) as a function of λ_t. It can be seen that both relations are decreasing in λ_t, such that autocracies are more likely than democracies to experience transitions of either type.

Turbulent and peaceful times Another interesting observation suggested by the simulations in Figure 3.4 is that revolts tend to be serially correlated over time. Underlying

Figure 3.7: Serial Correlation of Revolts

Likelihood of a Successful revolt at time $t + s$ conditional on a revolt s periods before (solid) and unconditional likelihood (dashed).



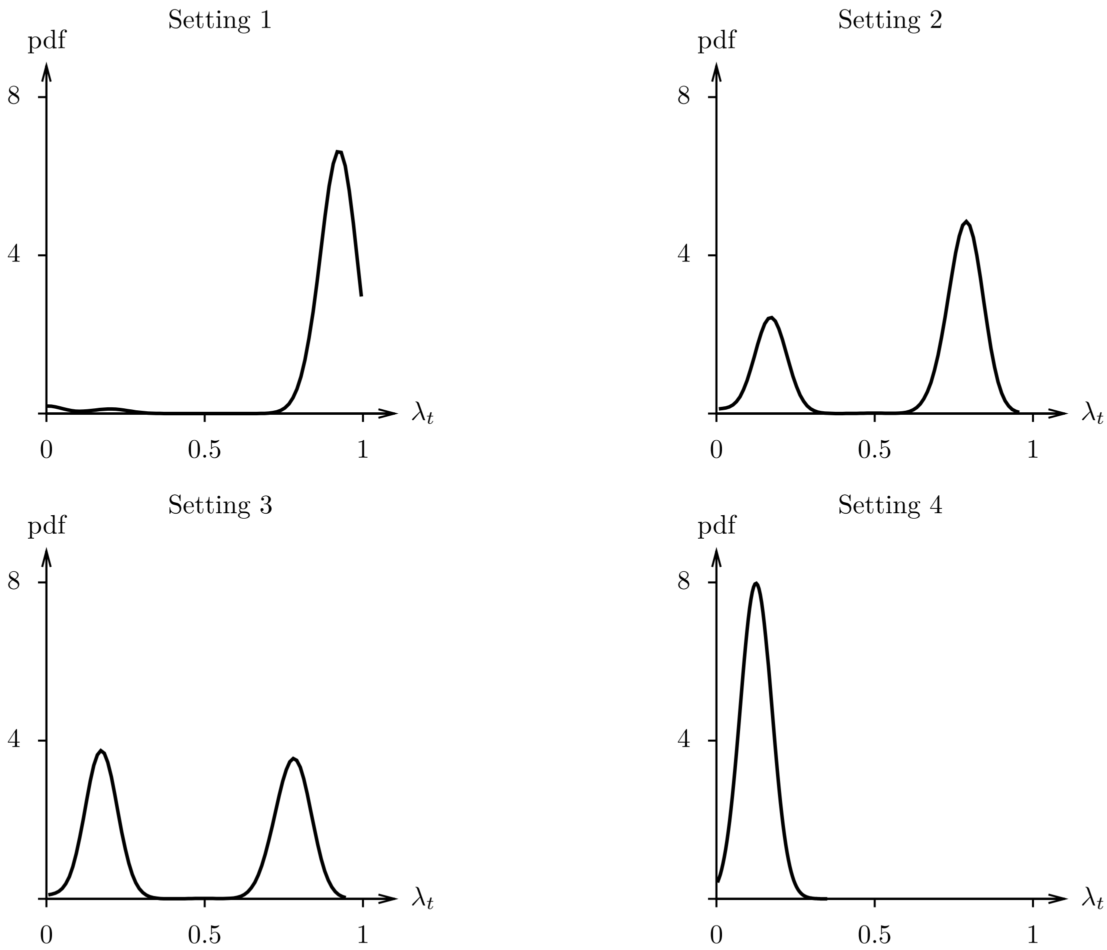
this observation is a statistical selection into autocratic regimes after successful revolts, seen in Figure 3.5. Because succeeding autocracies are frequently overthrown themselves, seen in Figure 3.6, the serial correlation follows. A direct assessment of this effect is provided in Figure 3.7, which plots the likelihood of a revolt at time $t + s$ conditional on a successful revolt at time t (represented by the downward sloping solid line).

The converse is true for reforms, which by Propositions 3.4 and 3.5 lead to democratic regimes, for which further political change is unlikely. Our model predicts, therefore, that via selection into particular polities, revolts lead to “turbulent” times, while reforms lead to “peaceful” periods.

Persistence A side effect of the considerations in the preceding paragraph is that despite their instability, autocratic systems are persistent over time. That is, while individual autocracies are relatively short-lived, they are frequently overthrown by small groups of insurgents, resulting in autocracies very similar to their predecessors. Settings 2 and 3 of our simulations in Figure 3.4 illustrate this implication further.

Long-run distribution Taken together, polarization to extreme regimes and the persistence of these suggests that the long-run distribution of polities is polarized as well. In Figure 3.8, we plot the invariant distribution of polities for different values of β_1. It can

Figure 3.8: Invariant Distribution of Political Systems



be seen that the distributions are double hump-shaped, with most mass concentrated on extreme political systems. Whether polities are mostly democratic or autocratic depends on the costs of reform as given by β_1. For low values of these costs (Settings 1 and 2), reforms are commonly used to mitigate most subversive threats, revolts are unlikely, and mass is mainly concentrated on democratic systems. If the costs of conducting reforms are high (Settings 3 and 4), less reforms are conducted, revolts are more frequent, and most mass is concentrated on autocratic political systems.

3.5 A Look at the Data

Our model predicts a number of properties about political transitions that are in principle accessible to an empirical investigation. In this section, we take an exploratory look at data that combines information on political transitions and political systems to evaluate the model's predictions. While we are able to demonstrate that our predictions are consistent with descriptive statistics from the data, we make no claims of capturing causal relations, which would be beyond the scope of this exercise.

3.5.1 Data Construction

As a measure for the model's polity variable, we use the *polity* variable, scaled to $[0,1]$, from the Polity IV Project (Marshall and Jaggers, 2002), which ranks political regimes on a 21 point scale between autocratic and democratic. In order to examine the model's predictions, we combine this dataset with data on political transitions.

To classify successful revolts, we use the Archigos Dataset of Political Leaders (Goemans et al., 2009). The dataset is available for the time period between 1919 and 2004 such that we limit attention to political systems and transition in these years. We record a successful revolt if a leader is irregularly removed from office due to domestic popular protest, rebel groups, or military actors (defined by Archigos' *exitcodes* 2, 4 and 6), and if at the same time the leader's successor takes office in irregular manner (defined by an *entrycode* 1). Furthermore, we take a revolt to be causal for a change in the political system if a polity change is recorded in the Polity IV database within a two week window of the revolt.

Finally, we use the dataset on the Chronology of Constitutional Events from the Comparative Constitution Project (Elkins et al., 2010) to classify reforms. We define reforms by a constitutional change (*evnttype* equal to *new*, *reinstated*, or *amendment*) accompanied by a positive change in the political system (as indicated by the variable *durable* from the Polity IV Project) which is not matched to a revolt or another irregular regime change from the Achigos Dataset.

The resulting dataset is a daily panel on the country level, which covers 175 countries and records 251 revolts and 97 reforms.

3.5.2 Empirical Properties of Political Systems and Transitions

Overview Table 3.1 summarizes the resulting dataset. Panel A displays average polities and annualized empirical likelihoods for a transition of either type. It can be seen that on average, revolts are observed with a frequency of 2.8 percent per year and country, and reforms are observed with a frequency of 1.1 percent. On average, this corresponds to a transition every 25 years per country.

The mean polity is given by 0.49—almost exactly the midpoint of the polity scale. As can be seen in the second column, however, the standard deviation of polities is quite large. The reason for this becomes clear in light of Figure 3.1, which displays the distribution

Table 3.1: Descriptive Statistics

	Mean	Standard Deviation	Observations
	A. Regimes		
Polities	0.493	0.376	3 289 400
Annual likelihood of a revolt			
Unconditional	0.028		3 289 400
If polity ≤ 0.25	0.030		1 452 533
If polity ≥ 0.75	0.012		1 238 720
Annual likelihood of a reform			
Unconditional	0.011		3 289 400
If polity ≤ 0.25	0.018		1 452 533
If polity ≥ 0.75	0.001		1 238 720
	B. Transitions		
Resulting polities			
After revolts	0.316	0.235	251
After reforms	0.672	0.242	97

Notes. Units of observation in Panel A are country-days. Units of observation in Panel B are transitions.

of political systems in our dataset: Only a minority of regimes are located in the middle of the polity scale. Instead, in line with our predictions, most mass is concentrated on extreme political systems. More precisely, 44 percent of all regimes are rather autocratic with a polity index of 0.25 and below, while 38 percent of all regimes are rather democratic with an index value of 0.75 and above.

Our model identifies two reasons for why the distribution of polities is extreme: Polarization via the transition mechanism and persistence of extreme polities.

Polarization To examine whether regimes are polarized via political transitions, consider Panel B of Table 3.1, which displays the mean polity index for regimes emerging after revolts and reforms, respectively. As predicted by Proposition 3.4, revolts on average lead to autocratic regimes with a polity index of 0.32, while reforms lead to rather democratic political systems with a mean polity index equal to 0.67.

Further insight can be gained from the conditional distribution of polities emerging after either type of transition. Figure 3.9 displays these distributions. From the left panel it is obvious that indeed the majority of political systems that emerge after revolts is autocratic. In contrast, the evidence about political reforms is less clear. On the one hand, the right panel of Figure 3.9 suggests that the majority of systems that are established through reforms are democratic. On the other hand, it also can be seen that, in contrast to the model's predictions, a significant number of reforms lead to regimes that are less

Figure 3.9: Empirical Distribution of Political Systems after Revolts and Reforms

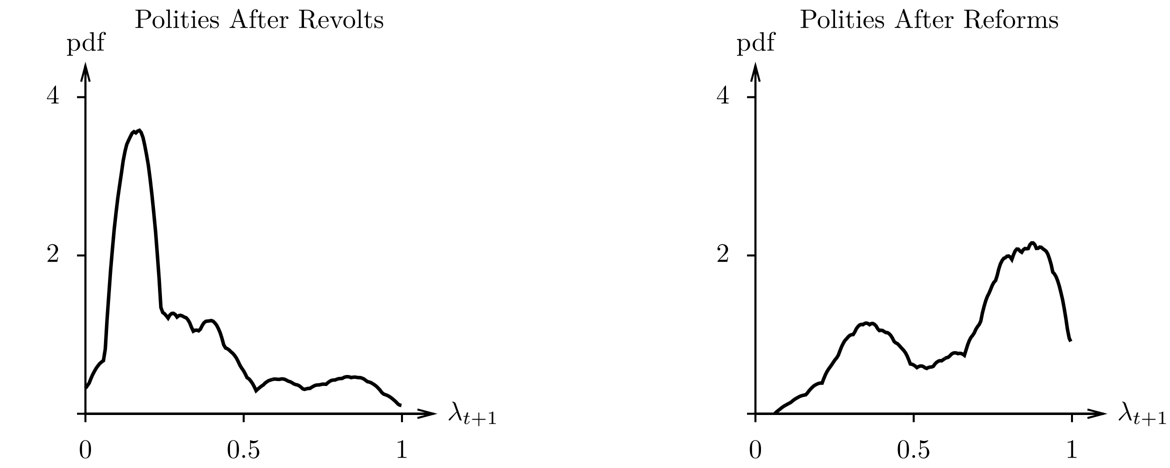
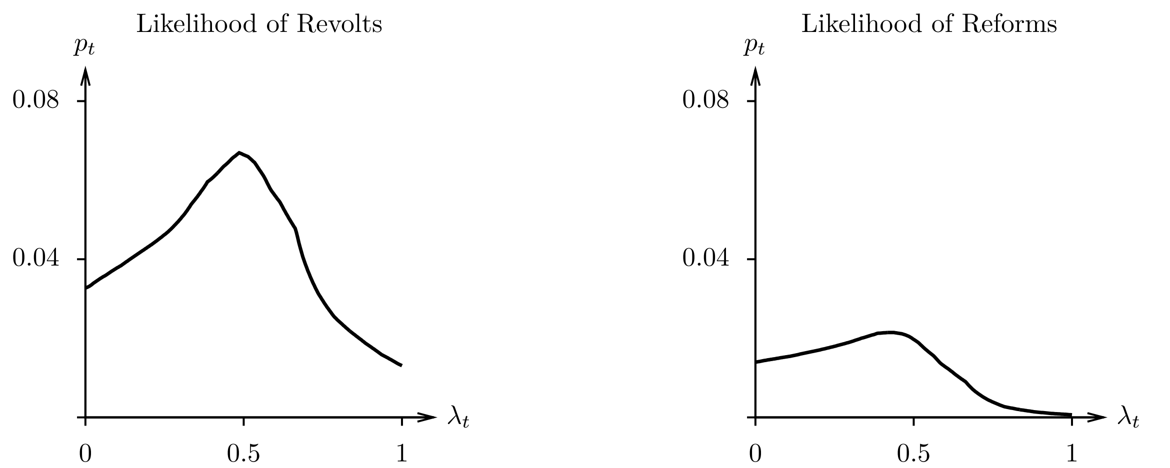


Figure 3.10: Empirical Likelihood of Revolts and Reforms

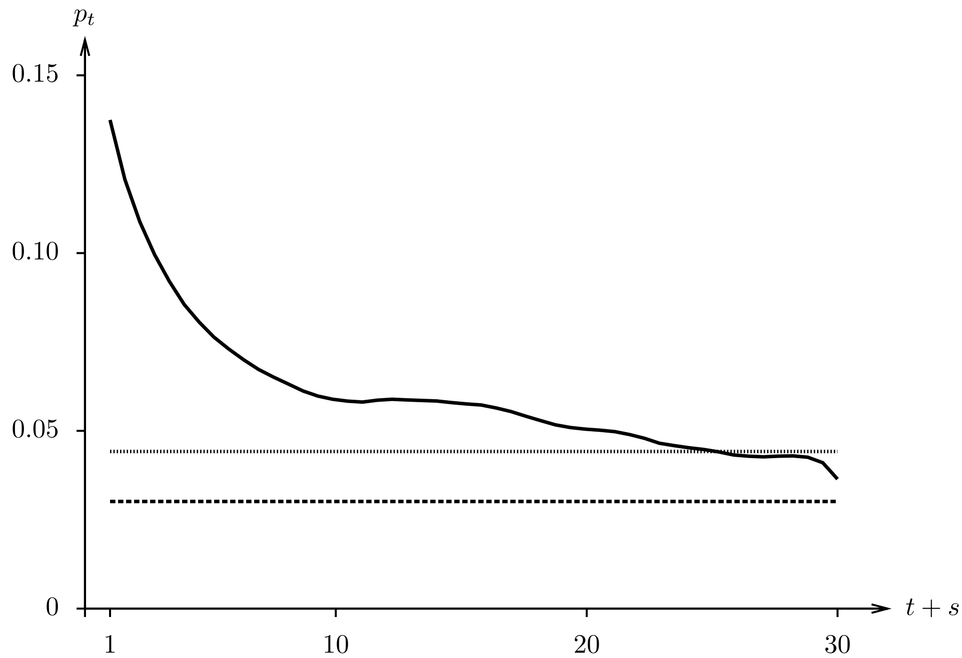


democratic.

However, while some reforms are less democratic than predicted, Figure 3.9 still suggests that the majority of democratic regimes are established via reforms, consistent with Proposition 3.4.

Stability and persistence To examine the stability of political systems, consider Figure 3.10. Here we plot the empirical likelihood functions for revolts and reforms, derived from a local polynomial estimation. Both likelihoods are hump-shaped in the polity index, with regimes in the middle of the scale being most likely to be overthrown. Nevertheless, as can be seen in Panel A of Table 3.1, autocracies with a polity index of 0.25 or below are more than twice as likely to fall to a revolt than democratic regimes with an index value of 0.75 and above. Moreover, autocratic regimes are about 18 times more likely to conduct reforms than democracies. Overall, autocracies survive for an average

Figure 3.11: Empirical Serial Correlation of Revolts
Empirical likelihood of a revolt at date $t+s$ conditional on a revolt s years before (solid) and unconditional likelihood for all countries (dashed) and countries with at least one transition (dotted).



of about 21 years, while democracies survive for an average of about 79 years. Hence, while in contrast to Proposition 3.5 full-scale democracies face a nonzero probability to be overthrown, they are nevertheless considerably more stable than all other regime types, confirming the qualitative predictions made by the model.

According to our model, even though autocracies are more instable than democracies, a serial correlation between revolts results in a persistence of autocratic political systems. The descriptive statistics reported above already suggest that the statistical selection mechanism underlying the persistence in our model might also be at work in the data. That is, we have seen that revolts are likely to result in autocracies, which are themselves likely to be overthrown again (see the left panels of Figures 3.9 and 3.10). As can be seen in Figure 3.11, the suggested correlation is indeed present in the data. The solid line in Figure 3.11 reflects the likelihood of observing a revolt at date $t+s$ conditional on that there was a successful revolt s years before. This likelihood is considerably larger than the unconditional likelihood of revolts across all countries (dashed line) and also compared to the unconditional likelihood in countries with at least one observed transition (dotted line). Compared to the latter benchmark, the difference is statistically significant at the 5 percent level for $s \leq 15$.

Summary In summary, the moments and correlations predicted by our model are consistent with the corresponding empirical moments and correlation. As predicted by the model, transitions lead to a polarization of political regimes, giving rise to autocracies after revolts and democracies after political reforms. While democracies are found to be empirically stable, autocracies are found to be short-lived. Yet, consistent with the model, a statistical selection gives rise to autocorrelation of successful subversions, explaining persistence of autocracies in the long-run. Consistently, as predicted by the model, the overall empirical distribution has mass mainly concentrated on extreme political systems.

3.6 Concluding Remarks

This is the first study which analyzes the dynamic properties of political transitions in a general framework allowing for endogenous outcomes of reforms and revolts. Our results suggest that transitions to democracy occur peacefully via reforms under participation of the former ruling elites. In contrast, violent transitions are driven by a small groups of insurgents and thus always lead to autocratic political regimes. Furthermore, democratic political systems face only a small opposition and are, hence, inherently stable, while autocratic regimes are short-lived due to the significant threat of revolts and the resulting strong incentives to conduct reforms.

These predictions are derived from a model in which the threat of revolt is the driving force of political change. We enrich the pioneering work of Acemoglu and Robinson (2000b) by allowing for arbitrary reforms conducted by the elite and endogenous formation of revolts through coordination of outsiders. While the predictions from this model fit descriptive statistics on political transitions quite well, our work points out promising avenues for future research. In particular, one simplifying assumption of our model is that the vulnerability of the incumbent regime is independently drawn anew in each period. Relaxing this assumption by allowing for serial correlation of the incumbents' strength would allow outsiders to learn about the prospects of revolting over time. A model, in which such endogenous learning is possible, could thus foster our understanding of the dynamic processes which ultimately lead to transition events (revolts or reforms).

Another interesting question regards the existence of mass movements. While from Figure 3.9 one can see that the majority of regimes that emerge after successful revolts is indeed autocratic in the data, there is also a nonzero mass of democratic regimes emerging from revolts. In our model mass revolutions and, hence, violent transitions to democracy are events off the equilibrium path. Therefore, only strategic mistakes could trigger mass revolts within our framework. For example, the elite may erroneously signal weakness by making small concessions, or outsiders may rally because of a commonly held belief that the regime is weak (for example due to information cascades as in Kuran, 1989 or

Lohmann, 1994). While it seems plausible that costly mass revolutions are the result of strategic mistakes, there thus remains the challenge to find a rational explanation for the emergence of mass revolutions when the regime has the power to counteract them via reforms.

Appendix A

Appendix to Chapter 1

A.1 Proofs

A.1.1 Proof of Lemma 1.1

We begin the proof by making two observations. First, it is easy to see that $F^R(f_t, \eta)$ is strictly decreasing and convex in η due to strict concavity of $u(\cdot)$. Hence, the fixed point $\tilde{\eta}$ satisfying $F^R(f_t, \tilde{\eta}) = 0$, which exists by assumption 1.1 (i) is unique. Second, note that for all $\eta > w_t$ we have that

$$F^P(f_t, w_t, \eta) < (1 - H(\eta))\Big((1-\alpha)\, E[u(w_\tau)|\, w_\tau \geq \eta, f_t] + \alpha u(\eta) - u(\eta)\Big) - c \leq F^R(f_t, \eta),$$

since $E[u(w_\tau)|\, w_\tau \geq \eta, f_t] \geq u(\eta)$. Thus, for large η, $F^R(\cdot)$ is an upper bound for $F^P(\cdot)$, where the former is smaller than zero for $\eta > \tilde{\eta}$.

Now, note that there always exists a fixed point $\bar{\eta}$ for which $F^P(f_t, w_t, \bar{\eta}) = 0$. To see this, consider the cases $\alpha = 0$ and $\alpha = 1$. For $\alpha = 0$, $F^P(f_t, w_t, \eta_t) = F^R(f_t, \eta_t)$ and by assumption 1.1 (i) there always exists η' satisfying $F^R(f_t, \eta') = 0$. For $\alpha = 1$, $F^P(f_t, w_t, \eta) = (1 - H(\eta))\, (u(w_t) - u(\eta)) - c$, which equals zero for some $\eta'' < w_t$ due to the strict concavity of $u(\cdot)$. Hence, some $\bar{\eta} \in [\eta', \eta'']$ satisfies $F^P(f_t, w_t, \bar{\eta}) = 0$ for $\alpha \in (0, 1)$.

Finally, we need to show that $\bar{\eta}$ is unique. Because $F^R(\cdot)$ is an upper bound for $F^P(\cdot)$ for large η, it is sufficient to show that $F^P(\cdot)$ is quasi-convex in η. The first and second derivatives of $F^P(\cdot)$ with respect to η are

$$\frac{\partial F^P(f_t, w_t, \eta)}{\partial \eta} = \alpha h(\eta)\, (u(\eta) - u(w_t)) - (1 - H(\eta))\, u'(\eta)$$

and

$$\frac{\partial^2 F^P(f_t, w_t, \eta)}{\partial \eta^2} = \alpha h'(\eta)\left(u(\eta) - u(w_t)\right) + (1+\alpha)h(\eta)u'(\eta) - \left(1 - H(\eta)\right)u''(\eta).$$

Hence, if there is some $\hat{\eta}$ with $\partial F^P(f_t, w_t, \hat{\eta})/\partial \eta = 0$, we have

$$\frac{\partial^2 F^P(f_t, w_t, \hat{\eta})}{\partial \eta^2} = \frac{\left(1 - H(\hat{\eta})\right)h'(\hat{\eta}) + h(\hat{\eta})^2}{h(\hat{\eta})}u'(\hat{\eta}) + \alpha h(\hat{\eta})u'(\hat{\eta}) - \left(1 - H(\hat{\eta})\right)u''(\hat{\eta}) > 0,$$

since $\left(1 - H(\hat{\eta})\right)h'(\hat{\eta}) + h(\hat{\eta})^2 \geq 0$ by assumption 1.1 (ii). This completes the proof. $\square$

A.1.2 Proof of Hypothesis 1.1

Clearly, $\partial y/\partial w_t > 0$ if either $\partial n(w_t)/\partial w_t > 0$ or $\partial \bar{\eta}/\partial w_t > 0$. The second follows directly from application of the implicit function theorem on (1.2), iff $\alpha > 0$. $\square$

A.1.3 Proof of Hypothesis 1.2

Obvious, following the same argument as in the proof of Hypothesis 1.1. $\square$

A.2 Tables

Table A.1: Effect of Purchase-Date Weather on Ticket Orders for Repeat Customers

	Daily Ticket Orders		
	Fixed Effects Multiple Orders	Fixed Effects Multiple Orders per Season	Fixed Effects Multiple Orders with Bad Experience
Avg. Sun	0.012***	0.0080***	0.0037***
	(0.0028)	(0.0018)	(0.00089)
Avg. Rain	−0.0074***	−0.0031*	−0.00018
	(0.0028)	(0.0017)	(0.00080)
Forecast	*Yes*	*Yes*	*Yes*
Horizon Indicators	*Yes*	*Yes*	*Yes*
Observations	1635	1635	1635
Adjusted R^2	0.260	0.241	0.132

Notes: Coefficients and robust standard errors are reported for OLS regressions of daily ticket orders on purchase-date weather (sunshine duration in percent of time and rainfall in 1/100 mm), forecast of movie-date weather (forecasted temperatures and indicator variables for symbols), and horizon indicators (dummy variables for the number of days between purchase-date and movie-date). Fixed effects for the show are included. In the first column, the sample is restricted to sales to customers who have previously ordered tickets at least once between 2004 and 2011. In the second column, the sample is restricted to customers with multiple orders per season and in the final column to customers who had previously experienced rainfall during a show they had tickets for.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Predictive Power of Current Weather and Forecast

	Evening Sunshine Duration			
	1 Day Out	2 Days Out	3 Days Out	4 Days Out
Avg. Sun	0.032	0.023	−0.0027	0.077
	(0.051)	(0.050)	(0.050)	(0.052)
Avg. Rain	0.076	0.012	−0.040	−0.027
	(0.068)	(0.071)	(0.074)	(0.076)
Forecasted Maxtemp.	1.27**	0.91	1.39**	0.95
	(0.62)	(0.71)	(0.68)	(0.70)
Forecasted Mintemp.	−0.38	−0.026	0.097	0.13
	(0.79)	(0.93)	(0.93)	(0.85)
Symbol Partly Sunny	−20.6***	−27.0***	−16.3***	−19.8***
	(4.86)	(4.98)	(5.41)	(5.62)
Symbol Shower	−47.9***	−45.4***	−31.7***	−27.4***
	(5.10)	(5.36)	(5.42)	(5.58)
Symbol Rain	−43.2***	−43.0***	−36.3***	−25.2***
	(5.98)	(6.21)	(6.64)	(6.70)
Symbol T-Storm	−63.7***	−59.1***	−47.1***	−31.9**
	(9.04)	(10.1)	(12.0)	(14.0)
MA Rain 14 days	−0.11	0.27	0.29	0.11
	(0.27)	(0.28)	(0.29)	(0.29)
MA Sun 14 days	0.19**	0.23**	0.31***	0.25**
	(0.093)	(0.098)	(0.10)	(0.10)
Year and Month Indicators	Yes	Yes	Yes	Yes
Observations	470	469	470	471
Adjusted R^2	0.369	0.302	0.246	0.159

Notes: We report the coefficients and robust standard errors of OLS regressions estimating expected sunshine duration (in percent) in the evening between 5 pm and 7 pm based on information one to four days in advance (indicated in the column heading). The information comprises of current average sunshine duration, current rainfall (in 1/100 mm), the current weather forecast for the respective time horizon, a moving average of sunshine duration and rainfall of the past fortnight as well as year and month dummies.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Predictive Power of Current Weather without Forecast

	Evening Sunshine Duration			
	1 day out	2 days out	3 days out	4 days out
Avg. Sun	0.16***	0.043	0.024	0.052
	(0.049)	(0.049)	(0.050)	(0.050)
Avg. Rain	−0.064	−0.051	−0.071	−0.0047
	(0.065)	(0.066)	(0.066)	(0.066)
MA Sun 14 days	0.29***	0.35***	0.33***	0.27***
	(0.100)	(0.10)	(0.10)	(0.10)
MA Rain 14 days	−0.058	0.14	−0.00052	−0.12
	(0.28)	(0.29)	(0.29)	(0.29)
Year and Month Indicators	Yes	Yes	Yes	Yes
Observations	572	570	568	566
Adjusted R^2	0.130	0.099	0.088	0.082

Notes: We report the coefficients and robust standard errors of predictions of evening sunshine duration. The estimated models are identical to Table 8 with the exception that the variables of the weather forecast are excluded from the dependent variables.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Picture

Figure A.1: Location of the Theater

This picture shows the location of the theater. Visitors are sitting in the amphitheater on different rows on flaggings or on wooden boards (the area at the bottom left corner of the picture). The screen is on the left of this picture (not shown).

Appendix B

Appendix to Chapter 2

B.1 Proof of Proposition 2.1

We consider an individual i with KR-preferences as described by (2.1) who faces the choice between a safe option with consumption utility normalized to x and a lottery F providing consumption utility $y_i + x$ and $x - z_i$ with equal probability. KR (2007) show in their Proposition 4 that an individual who prefers F in a PPE situation never prefers x in a Surprise situation, and in Proposition 8 that an individual who prefers x in a PPE situation never prefers F in a CPE situation.

Lemma B.1. *(KR 2007) If F is a PPE in the choice set $\{x, F\}$, then for any lottery G, $U_i(F|G) > U_i(x|G)$. Furthermore, if x is the PPE from choice set $\{x, F\}$, then F is not a CPE.*

The proof is a simple application of Propositions 4 and 8 from KR (2007). □

Given Lemma B.1, it is left to show that $a, \underline{a}$, and $\bar{a}$ satisfying the conditions given in Proposition 2.1 exist. Consider first case (ii) that F is the SurpriseE and PPE and x is the CPE. Using Lemma B.1 this implies

$$U_i(F|x) \geq U_i(x|x) > U_i(F|F). \tag{B.1}$$

The first inequality can be written as $x + 1/2\,(y_i - z_i) + 1/2\,\eta(y_i - \lambda z_i) \geq x$, or $\frac{y_i}{z_i} \geq \frac{1+\lambda\eta}{1+\eta} =: a$. The second inequality of (B.1) requires $x > x + 1/2\,(y_i - z_i) + 1/4\,\eta(1-\lambda)(y_i + z_i)$, or $\frac{y_i}{z_i} > \frac{1+\lambda\eta - 1/2\eta(1+\lambda)}{1+\eta-1/2\eta(1+\lambda)} =: \bar{a}$. Note that $\bar{a} > 0$ due to $\eta(\lambda - 1) < 2$ and that $\bar{a} > a$ due to $\lambda > 1$. Thus, x is the CPE and F is the PPE and SurpriseE if $a \leq y_i/z_i < \bar{a}$, showing (ii) of Proposition 2.1. Note that our arguments so far also imply case (i): F is SurpriseE, PPE, and CPE if $\bar{a} \leq y_i/z_i$.

Now, consider case (iii) requiring that F is the SurpriseE given reference distribution G and x is PPE and CPE. That x is the PPE implies $U_i(x|x) > U_i(F|x)$ which is equivalent to $y_i/z_i < a$ by the arguments above. Then, x is the CPE by Lemma B.1. For F to be the SurpriseE it has to hold that $U_i(F|G) \geq U_i(x|G)$. This inequality can be written as

$$y_i\Big(1+\eta\Big(G(x+y_i)+\lambda[1-G(x+y_i)]\Big)\Big)+\eta(\lambda-1)\int_x^{x+y_i}(r-x)dG(r) \geq$$
$$z_i\Big(1+\eta\Big(G(x-z_i)+\lambda[1-G(x-z_i)]\Big)\Big)+\eta(\lambda-1)\int_{x-z_i}^{x}(r-x)dG(r).$$

Integrating by parts and collecting terms gives

$$(y_i-z_i)(1+\eta\lambda) \geq \eta(\lambda-1)\left(\int_x^{x+y_i}G(r)dr-\int_{x-z_i}^{x}G(r)dr\right). \tag{B.2}$$

Note that an upper bound of the right hand side (RHS) of (B.2) is given by $\eta(\lambda-1)\left(y_iG(x+y_i)-z_iG(x-z_i)\right)$ such that a sufficient condition for (B.2) to hold is $\frac{y_i}{z_i} \geq \frac{1+\eta+\eta(\lambda-1)(1-G(x-z_i))}{1+\eta+\eta(\lambda-1)(1-G(x+y_i))}$. Since the lower bound of the RHS of (B.2) is 0, a necessary condition for (B.2) is $\frac{y_i}{z_i} \geq 1$. Thus, there is some $\underline{a} \in [1, \frac{1+\eta+\eta(\lambda-1)(1-G(x-z_i))}{1+\eta+\eta(\lambda-1)(1-G(x+y_i))}]$ such that (B.2) holds with equality for $y_i/z_i = \underline{a}$. Furthermore, note that $\underline{a} \leq \frac{1+\eta+\eta(\lambda-1)(1-G(x-z_i))}{1+\eta+\eta(\lambda-1)(1-G(x+y_i))} < a$ such that there exists some y_i/z_i satisfying $\underline{a} \leq y_i/z_i < a$. In this case, F is the SurpriseE and x is the PPE and CPE proving case (iii). Case (iv) follows because for $y_i/z_i < \underline{a}$, x is the SurpriseE and, by Lemma B.1, also the PPE and CPE. This completes the proof. □

B.2 Optimal Choice in the DA-Model

In this section we show how optimal choices predicted by the DA-model differ across Surprise, PPE, and CPE situations. For the presence of treatment differences, however, we need to apply the equilibrium concepts of SurpriseE, PPE, and CPE on preferences described by the DA-model first. Before stating the proposition to be shown, we thus start with a few definitions.

In the DA-model, the reference point is given by the certainty equivalent of consumption utility given beliefs about future outcomes r as described by the distribution $G(\cdot)$. We therefore define $EM_i(G) := E_G[m_i(r)]$. As in the main text, the DA-utility of a lottery described by the distribution $F(\cdot)$ over outcomes c given reference beliefs distributed according to $G(\cdot)$ is then given by

$$U_i^{DA} = \int m_i(c) + \mu(m_i(c) - EM_i(G))dF(c).$$

Given this, we apply the concepts of SurpriseE, PPE, and CPE to the DA model yielding the following definitions.

Definition B.1. For some expected distribution of outcomes G, selecting x from $\{x, F\}$ is a *DA Surprise Equilibrium (DA-SurpriseE)* if $U_i^{DA}(x|G) \geq U_i^{DA}(F|G)$. Otherwise, F is the DA surprise equilibrium. If selecting x from $\{x, F\}$ is a *DA Preferred Personal Equilibrium (DA-PPE)*, then $U_i^{DA}(x|x) \geq U_i^{DA}(F|x)$. If selecting F from $\{x, F\}$ is a DA-PPE, then $U_i^{DA}(F|F) \geq U_i^{DA}(x|F)$. When $U_i^{DA}(x|x) \geq U_i^{DA}(F|x)$ and $U_i^{DA}(F|F) \geq U_i^{DA}(x|F)$ hold simultaneously, then x is the DA-PPE if $U_i^{DA}(x|x) \geq U_i^{DA}(F|F)$; otherwise, F is the DA-PPE. Finally, selecting x from $\{x, F\}$ is a *DA Choice-Acclimating Personal Equilibrium (DA-CPE)* if $U_i(x|x) \geq U_i(F|F)$. Selecting F is a DA-CPE if $U_i(F|F) \geq U_i(x|x)$.

Given these definitions, we can state the following proposition.

Proposition B.1. *Assume $\eta(\lambda - 1) < 2$. Then, there exist $\underline{b}$, b, and $\bar{b}$ with $1 < \underline{b} < b < \bar{b}$ such that for individual i endowed with the choice set $\{x, F\}$*

(i) F is the DA-CPE if $y_i/z_i \geq \underline{b}$ and x is the DA-CPE otherwise,

(ii) F is the DA-PPE if $y_i/z_i \geq b$ and x is the DA-PPE otherwise,

(iii) F is the DA-SurpriseE if $y_i/z_i \geq \bar{b}$ and x is the DA-SurpriseE for $y_i/z_i \leq 1$. For $1 < y_i/z_i < \bar{b}$, the DA-SurpriseE depends on the location of $EM_i(G)$.

Consider first case (i), the optimal choice under DA-CPE. If F is the DA-CPE this implies $U_i^{DA}(F|F) \geq U_i^{DA}(x|x)$, which, after some algebra, implies $\frac{y_i}{z_i} \geq \frac{2+\eta(\lambda-1)}{2+\eta(\lambda+1)} =: \underline{b}$.

Given this, consider the consistency requirement for F to be a DA-PPE, $U_i^{DA}(F|x) \geq U_i^{DA}(x|x)$. Since choosing F given x as the referent can only be optimal if it has higher expected consumption utility ($EM_i(F) > x$), the latter requirement can be written as $EM_i(F) + 1/2\eta\Big(y_i - EM_i(F) + \lambda(z_i - EM_i(F)]\Big) \geq x_i + \eta\lambda(x_i - EM_i(F))$, with $EM_i(F) = x_i + 1/2(y_i - z_i)$. This is equivalent to $\frac{y_i}{z_i} \geq \frac{2+\eta(3\lambda-1)}{2+\eta(\lambda+1)} =: b$. Clearly, $b > \underline{b}$ such that if F is a DA-CPE it also is a DA-PPE.

The consistency requirement for x to be a DA-PPE is given by $U_i^{DA}(x|x) \geq U_i^{DA}(F|x)$. Since $y_i + x > x > x - z_i$, this implies that expecting x may be consistent if $x \geq EM_i(F) + 1/2\eta(y_i - \lambda z_i)$ which is equivalent to $\frac{y_i}{z_i} \leq \frac{1+\eta\lambda}{1+\eta}$. Since $\frac{1+\eta\lambda}{1+\eta} > b > \underline{b}$, x is the DA-PPE if $y_i/z_i < b$. This completes the proof of (ii).

Whether F or x are the DA-SurpriseE for individual i given beliefs over outcome G depend on the value of $EM_i(G)$. We have to distinguish the cases (a) $EM_i(G) > x + y_i$ or $EM_i(G) < x - z_i$, (b) $x > EM_i(G) \geq x - z_i$, and (c) $y_i + x \geq EM_i(G) \geq x$.

Consider case (a) first and assume that $EM_i(G) > x + y_i$, such that the individual will always experience losses regardless of choosing F or x. Then, F is a SurpriseE if $U_i^{DA}(F|G) \geq U_i^{DA}(x|G)$, which – after some algebra – can be shown to be equivalent to $y_i/z_i \geq 1$. This also holds for the case $EM_i(G) < x - z_i$, in which individuals always experience gains.

Now, consider case (b), $x > EM_i(G) \geq x - z_i$. In this case, the individual will experience losses only if she chooses F and the low outcome $x - z_i$ is realized. The condition for F to be the SurpriseE can be written as

$$y_i(1+\eta) \geq z_i(1+\eta\lambda) - \eta(\lambda-1)(x - EM_i(G)). \tag{B.3}$$

Since $\eta(\lambda-1)(x-EM_i(G)) \geq 0$ and $(x - EM_i(G)) \in [0, z_i]$, (B.3) holds with equality for some $\frac{y_i}{z_i} \in [1, \frac{1+\eta\lambda}{1+\eta}]$ depending on the value of $EM_i(G)$. For case (c), a similar argument establishes that the indifferent individual is characterized by $\frac{y_i}{z_i} \in [1, \frac{1+\eta\lambda}{1+\eta}]$ as well. With $\bar{b} := \frac{1+\eta\lambda}{1+\eta}$ and summarizing cases (a), (b), and (c) we thus conclude that F is the SurpriseE for all $\frac{y_i}{z_i} \geq \bar{b}$ and x is the SurpriseE for all $\frac{y_i}{z_i} \leq 1$. For all values $\frac{y_i}{z_i} \in (1, \bar{b})$, it depends on the value of $EM_i(G)$ whether F or x is the SurpriseE. □

B.3 Instructions

Below, we provide English translations of the instructions for the experiment; we thereby replicate the typeset as closely as possible. In passages, in which treatments differ, this is indicated by square brackets [. . .].[1]

Welcome to the experiment and thank you for your participation!

From now on, please do not talk to other participants of the experiment and switch off your mobile phones.

This experiment serves the investigation of economic decision making. You can earn money, which will be paid to you in cash after the experiment in private. **The entire experiment is split to take place on two dates.** The second date of this experiment is tomorrow at the same time. **If you are not going to have time tomorrow, please raise your hand now.** During the experiment, you and all other participants will be asked to make decisions.

[1] The experiment by Pahlke et al. (2012) was conducted at MELESSA shortly before the experiment presented here. Since their decision problem was similar to ours, we benefited from being able to re-use parts of their instructions.

On each date, the experiment lasts approximately **45 minutes**. Please raise your hand in case you have any questions during the experiment. The experimenter will then come to you and answer your question in private. In the interest of clarity, we use male terms only in the instructions.

Payment

You receive 4 Euro for arriving in time in addition to your earnings from the experiment. This premium for arriving in time as well as your entire earnings from your decisions during the experiment **will be paid tomorrow after the second date of the experiment. Thus, you will receive no payment today.**

Support

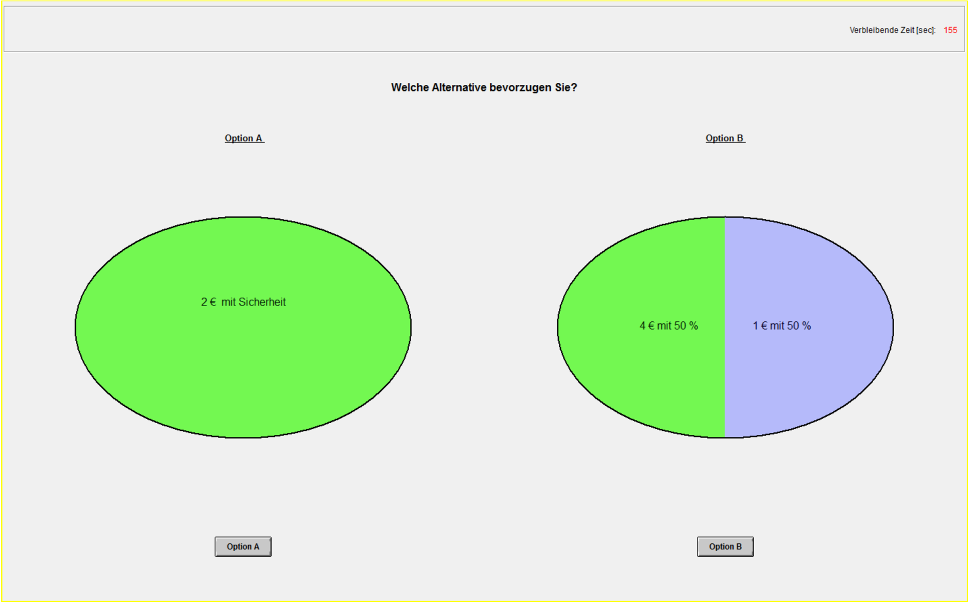
Please type your decisions into the computer. While making your decisions, there is a clock counting down in the right upper corner of the computer screen. This clock serves as a guide for how much time it should take for you to make your decisions. Of course, you are allowed to exceed the time. Only information screens, on which you are not asked to make a decision, will be skipped automatically when the clock has counted down.

Lottery decision making

You do not interact with other participants of the experiment at any point during the experiment. Your final payment is determined exclusively by your own decisions and according to the rules explained in the following. Other participants do not learn your decisions or how much you have earned at any point during or after the experiment. Nor will you learn the decisions of other participants and how much they have earned.

Task

In the experiment, you have to **decide <u>once</u> between two options**. On your screen, the decision problem will look like the following diagram.



In the above example, you have the choice between an **Option A**, which yields **2 Euro with certainty**, and an **Option B**, which yields **4 Euro with a probability of 50 %** and **1 Euro with the probability of 50 %**. An alternative like Option A in the above example is called a **"safe payment"** since it is paid with probability 100 %. An alternative like Option B in the above example is called a **"lottery"** since either the first amount is paid with probability of 50 % or the second amount is paid with probability of 50 %.

[PPE: In the experiment, you will have to make **one single decision** between a safe payment and a lottery like in the above example **not until tomorrow**. **However, you will learn the decision problem already today** and you will be asked to state the decision you are likely to make tomorrow. Yet, **your (tentative) decision today** comes **without any commitment**. Only **your decision tomorrow** will **determine your payment from this part of the experiment.**]

[Surprise & CPE: In the experiment, you will have to make **one single decision today** between a safe payment and a lottery like in the above example.] Note, however, that the **payments** of the alternatives you face **in the experiment are considerably larger** than in the above example. **Think about your decision carefully, as it will determine your earnings from the experiment to a large degree.** [PPE: You can also think about your decision over the course of the day.] Note as well that tomorrow you will be **asked** about **details of both options (payments and probabilities)**. Tomorrow, you will be able to earn additional money with each correct answer.

Determination of payments

[Surprise & CPE: If you have chosen the **safe payment**,] [PPE: If you choose the **safe payment** tomorrow,] you will receive the amount of the safe payment at the end of tomorrows' date.

[Surprise & CPE: In case you have chosen the **lottery**] [PPE: In case you choose the

lottery tomorrow], numbers ranging from 1 to 6 are assigned to each payment. Since each payment in the lottery you will be able to choose has a probability of 50 % to get paid, the lottery numbers 1, 2, and 3 are assigned to one amount and the numbers 4, 5, and 6 to the other amount. The computer determines randomly which amount gets assigned to the low numbers and which amount gets assigned to the large numbers. You will learn the assignment of the lottery numbers [Surprise: **directly after your decision**] [PPE: **tomorrow directly after your decision**] [CPE: **not until tomorrow**]. (If you have chosen the safe payment, lottery numbers are not assigned as you will receive the safe payment for sure.) [Surprise: **Directly thereafter**] [PPE & CPE: **When you have learned the assignment of the lottery numbers tomorrow,**] **a randomly chosen participant throws a die under the supervision of the experimenter.** At the end of tomorrows date, all participants who have chosen the lottery will receive the amount, which has been assigned to the number rolled. Thus, if you choose the lottery [PPE: tomorrow], you will learn your payment [Surprise: today directly after your decision.] [PPE: tomorrow directly after your decision.] [CPE: not until tomorrow.]

*Example: [Surprise & CPE: You] [PPE: Tomorrow, you] choose Option B, which pays **4 Euro** with probability 50 % and **1 Euro** with probability 50 %. You learn [Surprise: **directly after your choice**] [PPE: **tomorrow directly after your choice**] [CPE: **tomorrow**] that the computer has assigned **lottery numbers 4, 5, and 6** to the amount of **4 Euros** and **lottery numbers 1, 2, and 3** to the amount of **1 Euro**. You thus have a 50 % probability to receive 4 Euro and a 50 % probability to receive 1 Euro. [Surprise: **Thereafter**,] [PPE & CPE: **When you have learned the assignment of lottery numbers tomorrow,**] a randomly chosen participant roles a die once and determines your payment. If a 2 is rolled, your earnings are 1 Euro. If, on the other hand, a 6 is rolled you receive 4 Euro.*

End of the first set of instructions.

Below, we provide an English translation of the instructions for the risk elicitation task following Maier and Rüger (2010). These instructions were handed out (and read aloud) on the second date of the experiment and after the resolution of uncertainty for the main task in treatments CPE and PPE. Until these instructions were handed out, subjects did not know that they would be facing an additional choice task.

Decisions between lotteries

In this part of the experiment, you have to make a **total of 10 decisions** between two lotteries. The decision problems will be displayed in a table with the following form:

	Option A			Option B		
Zeile	A1: Wahrscheinlichkeit 50% (Würfel 1, 2 oder 3)	A2: Wahrscheinlichkeit 50% (Würfel 4, 5 oder 6)	Ihre Wahl	B1: Wahrscheinlichkeit 50% (Würfel 1, 2 oder 3)	B2: Wahrscheinlichkeit 50% (Würfel 4, 5 oder 6)	Zeile
1.	18.0 €	16.0 €	A ○ ○ B	24.0 €	6.0 €	1.
2.	18.0 €	16.0 €	A ○ ○ B	20.0 €	10.0 €	2.
3.	18.0 €	16.0 €	A ○ ○ B	16.0 €	14.0 €	3.
4.	22.0 €	12.0 €	A ○ ○ B	16.0 €	14.0 €	4.
5.	23.5 €	10.5 €	A ○ ○ B	16.0 €	14.0 €	5.
6.	25.0 €	9.0 €	A ○ ○ B	16.0 €	14.0 €	6.
7.	26.5 €	7.5 €	A ○ ○ B	16.0 €	14.0 €	7.
8.	28.0 €	6.0 €	A ○ ○ B	16.0 €	14.0 €	8.
9.	29.5 €	4.5 €	A ○ ○ B	16.0 €	14.0 €	9.
10.	31.0 €	3.0 €	A ○ ○ B	16.0 €	14.0 €	10.

In the above table, **each row represents a decision problem** between an Option A and an Option B. For example, in the first row of the table above, you may choose Option A, which pays out 18 Euro with a probability of 50 % and 16 Euro with a probability of 50 %. Option B in the first row, on the other hand, pays out 24 Euro with a probability 50 % and 6 Euro with a probability of 50 %. Note that the **probabilities** are **identical** across decision problems. With 50 % probability the high amount of each option will be paid out and with probability 50 % the low amount will be paid out. However, the **amounts** that each option pays **vary** across decision problems. Option A becomes (from top to bottom of the table) increasingly risky since the spread between high and low amounts increases, while Option B becomes rather safer (or does not change at all) since the spread between high and low amounts decreases. In each decision problem – in each row of the table – you have to make a decision between Option A and Option B.

Determination of payments

When all participants have made their decisions, **the computer determines randomly <u>two</u> participants** for each of whom **one decision will be paid out.** (The probability to be chosen by the computer is identical for all participants and completely independent of your choices.) For the two participants chosen, the computer randomly chooses one of the decision problems that will determine the payment. (Each decision problem – that is each row of the table – is chosen to be payoff relevant with the same probability.) The **payment from the option** that the participant has chosen in the relevant decision problem will then be **determined by the participant by the role of a die.** The **high amount of each option** will be paid out if a **1, 2, or 3 is rolled.** The **low amount of each option** will be paid out if a **4, 5, or 6 is rolled.** If you are not chosen by the computer for being paid out, you will not receive a payment from this part of the experiment.

Example: A participant chooses Option A in the third row of the table and makes other decisions in the remaining 9 rows of the table. He is chosen by the computer as one of two participants whose decision is paid out. For the participant chosen, the computer determines randomly the decision problem in row 3 as payoff relevant. Thereafter, the participants roles a die. ***If, for example, a 3 comes up****, he receives* ***18 Euro*** *(the high amount of Option A).* ***If, one the other hand, a 5 comes up****, he receives* ***16 Euro*** *(the low amount of Option A).*

End of the second set of instructions.

Appendix C

Appendix to Chapter 3

C.1 Proofs

C.1.1 Insiders never subvert, outsiders always join the regime

Insiders' choice set includes $x_t \in [\lambda_t, 1]$. It thus holds that $(1 - p(\cdot, x_t))u(x_t) \geq (1 - p(\cdot, 1))u(1) = u(1) \geq \psi(1) \geq \psi(s_t) \geq \hat{\theta}_t \psi(s_t)$, where the first inequality follows from revealed preferences, the second inequality follows from $h(\cdot) \in [0, 1]$, the third inequality follows from ψ increasing, and the last inequality follows from $\theta_t \in [0, 1]$. Hence, it is not attractive for any individual insider to support a revolt against his own regime. As for outsiders we need to differentiate two cases. First, outsiders that are targeted by a reform and would otherwise support a revolt prefer to join the regime using exactly the same argument as above. Second, outsiders that are targeted by a reform and would otherwise not support a revolt prefer to join the regime since again by revealed preferences it holds that $(1 - p(\cdot, x_t))u(x_t) \geq (1 - p(\cdot, 1))u(1) = u(1) \geq \gamma_{it}$ for all i and t. □

C.1.2 Proof of Proposition 3.1

We first establish that any solution to the outsiders' coordination problem is a fixed point to equation (3.6). From our discussion in the main body of the paper it is clear that this is the case if and only if $\bar{\gamma}(\hat{s}_t) \leq 1$ for all $\hat{s}_t$. From Assumption 3.1 it follows that $\bar{\gamma}$ is increasing in $\hat{s}_t$, and therefore $\bar{\gamma}(\hat{s}_t) \leq 1$ holds if $\bar{\gamma}(1) = p(\hat{\theta}_t, 1)\, u(1) \leq 1$. Since $u(1) = 1$ and $p(\cdot) \in [0, 1]$ this is indeed the case.

Hence, consider any fixed point to (3.6). Since $f(0) = 0$ for all $(\hat{\theta}_t, x_t) \in \Theta \times [0, 1]$, there always exists a fixed point at $\hat{s}_t = 0$. Whether or not $\hat{s}_t = 0$ is consistent with the concept of trembling-hand perfection, and whether or not other fixed points exist, depends on the values of $\hat{\theta}_t$ and x_t. We have to distinguish two cases.

First, if $\hat{\theta}_t = 0$ or $x_t = 1$, then $f(\hat{s}_t) = 0$ for all $\hat{s}_t$, and therefore $\hat{s}_t = 0$ is obviously the only fixed point to (3.6). To establish that $\hat{s}_t = 0$ is also trembling-hand perfect, it suffices to show that for all i, $\phi_{it} = 0$ is a best response to some sequence of totally mixed strategy profiles $\{\omega_{jt}^k : j \in [0,1] \setminus i\}_{k=0}^{\infty}$ that converges to the equilibrium profile where all i play $\phi_{it} = 0$ with probability 1. Since for $\hat{\theta}_t = 0$ and $x_t = 1$ playing $\phi_{it} = 0$ is a (weakly) dominant strategy, this is trivially true.

Second, consider the case where $\hat{\theta}_t \neq 0$ and $x_t \neq 1$. In this case the fixed point at $\hat{s}_t = 0$ is not trembling-hand perfect. To see this let $z^k = \min_i\{\omega_{it}^k(1)\}$ denote the minimum probability with which any agent i plays $\phi_{it} = 0$ in the kth element of sequence ω_{it}^k. The requirement of trembling-hand perfection that $\{\omega_{it}^k\}$ is totally mixed for all i and k implies that $z^k > 0$ for all k. Hence, $s_t^k = (1-x_t)\int_i \omega_{it}^k(1)\, di \geq (1-x_t)\, z^k > 0$. However, from $h(0) = 0$ in combination with Assumption 3.1(b) it follows that for any $s_t^k > 0$, $\bar{\gamma}(s_t^k) = \hat{\theta}_t \psi(s_t^k) > 0$ and, hence, a strictly positive fraction of outsiders strictly prefers to choose $\phi_{it} = 1$ in response to $\{\omega_{jt}^k : j \in [0,1]\}$. We conclude that $\hat{s}_t = 0$ can not be supported in any trembling-hand perfect equilibrium if $\hat{\theta}_t \neq 0$ and $x_t \neq 1$.

Having ruled out $\hat{s}_t = 0$ as a solution to the coordination problem for $\hat{\theta}_t \neq 0$ and $x_t \neq 1$, we now show that there is a unique $\hat{s}_t > 0$ solving (3.6) for $\hat{\theta}_t \neq 0$ and $x_t \neq 1$, which is also consistent with the concept of trembling-hand perfection. From $\bar{\gamma} \in [0,1]$ it follows that f is bounded by its support, $[0, 1-x_t]$. Moreover, by Assumption 3.1 we have that $\lim_{\hat{s}\to 0} \psi'(\hat{s}) = \infty$, implying that $\lim_{\hat{s}\to 0} f'(\hat{s}) = \infty$. Hence, there exists a $\tilde{s} > 0$, such that $f(\tilde{s}) > \tilde{s}$. Together with continuity of ψ (and thus of f), it follows that there exists a strictly positive fixed point to (3.6), which by concavity of ψ (and thus of f) is unique on $(0, 1]$.

Let $s_t^* = f(s_t^*)$ denote this fixed point. It remains to be shown that s_t^* is consistent with the concept of trembling-hand perfection. To show this, consider the following sequences $\omega_{it}^k(1) = 1 - \varepsilon^k$ for all $i \in \{j : \gamma_{jt} \leq \bar{\gamma}(s_t^*)\}$ and $\omega_{it}^k(1) = \frac{\bar{\gamma}(s_t^*)}{1-\bar{\gamma}(s_t^*)}\varepsilon^k$ for all $i \in \{j : \gamma_{jt} > \bar{\gamma}(s_t^*)\}$, with some $\{\varepsilon^k\}_{k=0}^{\infty}$ such that $\lim_{k\to\infty} \varepsilon^k = 0$. Then, by construction,

$$s_t^k = (1-x_t)\left((1-\varepsilon^k)\,\bar{\gamma}(s_t^*) + \frac{\bar{\gamma}(s_t^*)}{1-\bar{\gamma}(s_t^*)}\,\varepsilon^k(1-\bar{\gamma}(s_t^*))\right) = (1-x_t)\,\bar{\gamma}(s_t^*) = f(s_t^*),$$

and hence $\{\phi_{it} : i \in [0,1]\}$ being mutually best responses implies that $\{\phi_{it} : i \in [0,1]\}$ are best responses to $\{\omega_{it}^k : i \in [0,1]\}$ for all values of k.

The above arguments establish that s_t is uniquely determined by a (time-invariant) function $s : (\hat{\theta}_t, x_t) \to s_t$. It remains to be shown that $\partial s/\partial\hat{\theta}_t \geq 0$ and $\partial s/\partial x_t \leq 0$. Given that s_t is a fixed point to (3.6), we have that

$$\pi(s_t, x_t) \equiv s_t - (1 - x_t)\,\hat{\theta}_t\,\psi(s_t) = 0.$$

Implicit differentiation implies that

$$\frac{\partial s_t}{\partial x_t} = -\hat{\theta}_t\,\psi(s_t) \times \left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1}$$

and

$$\frac{\partial s_t}{\partial \hat{\theta}_t} = (1 - x_t)\,\psi(s_t) \times \left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1},$$

where

$$\frac{\partial \pi_t}{\partial s_t} = -(1 - x_t)\,\frac{\partial \bar{\gamma}}{\partial s_t} + 1.$$

Since ψ is bounded by $\psi(1) = 1$, (3.6) implies that $\lim_{\hat{\theta}_t \to 0} s_t^* = \lim_{x_t \to 1} s_t^* = 0$, and therefore the case where $\hat{\theta}_t = 0$ or $x_t = 1$ is a limiting case of $\hat{\theta} \neq 0$ and $x_t \neq 1$. From the implicit function theorem it then follows that s is differentiable on its whole support. Moreover, the previous arguments imply that $f(\tilde{s}) > \tilde{s}$ for all $\tilde{s} < s_t^*$ and $f(\tilde{s}) < \tilde{s}$ for all $\tilde{s} > s_t^*$, implying that $f'(s_t^*) < 1$ or, equivalently, $\partial\bar{\gamma}/\partial s_t < (1 - x_t)^{-1}$ at s_t^*. Thus $\partial \pi_t/\partial s_t > 0$ for all $(\hat{\theta}_t, x_t) \in \Theta \times [0,1]$, which yields the desired results.

Finally, while we focused on pure strategies when proving the results above, it is easy to see that the proposition generalizes to mixed strategies. By the law of large numbers, any mixed strategy equilibrium beliefs about s are of zero variance and, hence, the arguments above apply, implying that all outsiders, except a zero mass i with $\gamma_i = \bar{\gamma}(s_t^*)$, strictly prefer $\phi_i = 0$ or $\phi_i = 1$. We conclude that there is no scope for (nondegenerate) mixed best responses. □

C.1.3 Proof of Proposition 3.2

The proof proceeds by a series of lemmas. To simplify notation, in what follows we drop λ_t as an argument of x and $\hat{\theta}$ where no confusion arises. Furthermore, we use $\tilde{V}^I(\theta_t, \hat{\theta}_t, x_t) = (1 - \theta_t h(s_t))\,u(x_t)$ to denote insider's indirect utility (up to a constant $u(\lambda_t)$), as follows from $s_t = s(\hat{\theta}_t, x_t)$ given Proposition 3.1.

Lemma C.1. *x is weakly increasing in θ_t.*

Proof. Suppose to the contrary that $x(\theta'') < x(\theta')$ for $\theta' < \theta''$. Let $x' \equiv x(\theta')$, $x'' \equiv x(\theta'')$, $u' \equiv u(x')$, $u'' \equiv u(x'')$, $h' \equiv h(s(\hat{\theta}(x'), x'))$, and $h'' \equiv h(s(\hat{\theta}(x''), x''))$. Optimality of x' then requires that $\tilde{V}^I(\theta', \hat{\theta}(x''), x'') \leq \tilde{V}^I(\theta', \hat{\theta}(x'), x')$, implying $u'h' - u''h'' \leq (u' - u'')/\theta' < (u' - u'')/\theta''$, where the last inequality follows from $\theta' < \theta''$ and $u' < u''$. Hence, $\tilde{V}^I(\theta', \hat{\theta}(x''), x'') \leq \tilde{V}^I(\theta', \hat{\theta}(x'), x')$ implies that $\tilde{V}^I(\theta'', \hat{\theta}(x''), x'') < \tilde{V}^I(\theta'', \hat{\theta}(x'), x')$, contradicting optimality of x'' for θ''.

Lemma C.2. *Suppose x is discontinuous at θ', and define $x^- \equiv \lim_{\varepsilon \uparrow 0} x(\theta' + \varepsilon)$ and $x^+ \equiv \lim_{\varepsilon \downarrow 0} x(\theta' + \varepsilon)$. Then for any $x' \in (x^-, x^+)$, the only beliefs consistent with the D1 criterion are $\hat{\theta}(x') = \theta'$.*

Proof. Let $\theta'' > \theta'$, and let $x'' \equiv x(\theta'')$. Optimality of x'' then requires that $\tilde{V}^I(\theta'', \hat{\theta}(x''), x'') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+)$ and, thus for any $\tilde{\theta}$,

$$\begin{gathered}\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x''), x'') \quad \text{implies that} \\ \tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+)\,.\end{gathered}$$

Moreover, arguing as in the proof of Lemma C.1,

$$\begin{gathered}\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+) \quad \text{implies that} \\ \tilde{V}^I(\theta', \tilde{\theta}, x') > \tilde{V}^I(\theta', \hat{\theta}(x^+), x^+)\,.\end{gathered}$$

Hence, if $\tilde{V}^I(\theta'', \tilde{\theta}, x') \geq \tilde{V}^I(\theta'', \hat{\theta}(x^+), x^+) = \bar{V}^I(\theta'')$, then $\tilde{V}^I(\theta', \tilde{\theta}, x') > \tilde{V}^I(\theta', \hat{\theta}(x^+), x^+) = \bar{V}^I(\theta')$. Therefore, $D_{\theta'', x'}$ is a proper subset of $D_{\theta', x'}$ if $\theta'' > \theta'$. (For the definition of $D_{\theta, x}$, see Footnote 10.) A similar argument establishes that $D_{\theta'', x'}$ is a proper subset of $D_{\theta', x'}$ if $\theta'' < \theta'$ and, thus, the D1 criterion requires that $\hat{\theta}(x') = \theta'$ for all $x' \in (x^-, x^+)$.

Lemma C.3. *There exists $\bar{\theta}(\lambda_t) > 0$, such that $x(\theta_t, \lambda_t) = \lambda_t$ for all $\theta_t < \bar{\theta}(\lambda_t)$. Moreover, $x(\theta'') > x(\theta') > \lambda_t + \mu$ for all $\theta'' > \theta' \geq \bar{\theta}(\lambda_t)$ and some $\mu > 0$.*

Proof. First, consider the existence of a connected pool at $x_t = \lambda_t$. Because for $\theta_t = 0$, $x_t = \lambda_t$ dominates all $x_t > \lambda_t$, we have that $x(0) = \lambda_t$. It follows that there exists a pool at $x_t = \lambda_t$, because otherwise $\hat{\theta}(\lambda_t) = 0$ and, therefore, $p(\cdot, s(\hat{\theta}(\lambda_t), \lambda_t)) = 0$, contradicting optimality of $x(\theta) > \lambda_t$ for all $\theta > 0$. Moreover, by Lemma C.1, x is increasing, implying that any pool must be connected. This proves the first part of the claim.

Now consider $x(\theta'') > x(\theta')$ for all $\theta'' > \theta' \geq \bar{\theta}(\lambda_t)$ and suppose to the contrary that $x(\theta'') \leq x(\theta')$ for some $\theta'' > \theta'$. Since x is increasing, it follows that $x(\theta) = x^+$ for all $\theta \in [\theta', \theta'']$ and some $x^+ > \lambda_t$. W.l.o.g. assume that θ' is the lowest state in this pool.

Then Bayesian updating implies that $\theta^+ \equiv \hat{\theta}(x^+) \geq (\theta' + \theta'')/2 > \theta'$ and, therefore, $\tilde{V}^I(\theta', \theta^-, x^+) > \tilde{V}^I(\theta', \theta^+, x^+)$ for all $\theta^- \leq \theta'$. Hence, because θ' prefers x^+ over $x(\theta^-)$, it must be that $x(\theta^-) \neq x^+$ for all $\theta^- \leq \theta'$ and, hence, $x(\theta^-) < x^+$ by Lemma C.1. Accordingly, let $x^- \equiv \max_{\theta^- \leq \theta'} x(\theta^-)$. Then from continuity of $\tilde{V}^I$ and $\theta^+ > \theta'$ it follows that there exists an off-equilibrium reform $x' \in (x^-, x^+)$ with $\tilde{V}^I(\theta', \theta', x') > \tilde{V}^I(\theta', \theta^+, x^+)$. Hence, to prevent θ' from choosing x' it must be that $\hat{\theta}(x') > \theta'$. However, from Lemma C.2 we have that $\hat{\theta}(x') = \theta'$, a contradiction.

Finally, to see why there must be a jump-discontinuity at $\bar{\theta}(\lambda_t)$ note that $\tilde{V}^I(\bar{\theta}(\lambda_t), \bar{\theta}(\lambda_t)/2, \lambda_t) = \tilde{V}^I(\bar{\theta}(\lambda_t), \bar{\theta}(\lambda_t), x(\bar{\theta}(\lambda_t)))$; otherwise, there necessarily exists a θ in the neighborhood of $\bar{\theta}(\lambda_t)$ with a profitable deviation to either λ_t or $x(\bar{\theta}(\lambda_t))$. From the continuity of $\tilde{V}^I$ and the non-marginal change in beliefs from $\bar{\theta}(\lambda_t)/2$ to $\bar{\theta}(\lambda_t)$ it follows that $x(\bar{\theta}(\lambda_t)) > \lambda_t + \mu$ for all λ_t and some $\mu > 0$.

Lemma C.4. *x is continuous and differentiable in θ_t on $[\bar{\theta}(\lambda_t), 1]$.*

Proof. Consider continuity first and suppose to the contrary that x has a discontinuity at $\theta' \in (\bar{\theta}(\lambda_t), 1)$. By Lemma C.1, x is monotonically increasing in θ_t. Hence, because x is defined on an interval, it follows that for any discontinuity θ', $x^- \equiv \lim_{\varepsilon \uparrow 0} x(\theta')$ and $x^+ \equiv \lim_{\varepsilon \downarrow 0} x(\theta')$ exist, and that x is differentiable on $(\theta' - \varepsilon, \theta')$ and $(\theta', \theta' + \varepsilon)$ for some $\varepsilon > 0$. Moreover, from Lemmas C.2 and C.3 it follows that in equilibrium $\hat{\theta}(x') = \theta'$ for all $x' \in [x^-, x^+]$. Hence, $\tilde{V}^I(\theta', \theta', x^-) = \tilde{V}^I(\theta', \theta', x^+)$, since otherwise there necessarily exists a θ in the neighborhood of θ' with a profitable deviation to either x^- or x^+. Accordingly, optimality of $x(\theta')$ requires $\tilde{V}^I(\theta', \theta', x') \leq \tilde{V}^I(\theta', \theta', x^-)$ and, thus, $\tilde{V}^I(\theta', \theta', x^-)$ must be weakly decreasing in x. Therefore, $\partial \tilde{V}^I / \partial \hat{\theta}_t < 0$ and $\lim_{\varepsilon' \downarrow 0} \partial \hat{\theta}(x^- - \varepsilon')/\partial x_t > 0$ (following from Lemma C.3) imply that $\lim_{\varepsilon' \downarrow 0} \partial \tilde{V}^I(\theta', \hat{\theta}(x^- - \varepsilon'), x^- - \varepsilon')/\partial x_t < 0$. Hence, a profitable deviation to $x^- - \varepsilon'$ exists for some $\varepsilon' > 0$, contradicting optimality of $x(\theta')$.

We establish differentiability by applying the proof strategy for Proposition 2 in Mailath (1987). Let $g(\theta, \hat{\theta}, x) \equiv \tilde{V}^I(\theta, \hat{\theta}, x) - \tilde{V}^I(\theta, \theta', x(\theta'))$, for a given $\theta' > \bar{\theta}(\lambda_t)$, and let $\theta'' > \theta'$. Then, optimality of $x(\theta')$ implies $g(\theta', \theta'', x(\theta'')) \leq 0$, and optimality of $x(\theta'')$ implies that $g(\theta'', \theta'', x(\theta'')) \geq 0$. Letting $a = (\alpha\theta' + (1-\alpha)\theta'', \theta'', x(\theta''))$, for some $\alpha \in [0,1]$ this implies

$$0 \geq g(\theta', \theta'', x(\theta'')) \geq -g_\theta(\theta', \theta'', x(\theta''))(\theta'' - \theta') - \tfrac{1}{2} g_{\theta\theta}(a)(\theta'' - \theta')^2,$$

where the second inequality follows from first-order Taylor expanding $g(\theta'', \theta'', x(\theta''))$ around $(\theta', \theta'', x(\theta''))$ and rearranging the expanded terms using the latter optimality condition. Expanding further $g(\theta', \theta'', x(\theta''))$ around $(\theta', \theta', x(\theta'))$, using the mean value theorem on $g_\theta(\theta', \theta'', x(\theta''))$, and noting that $g(\theta', \theta', x(\theta')) = g_\theta(\theta', \theta', x(\theta')) = 0$, these inequalities can be written as

$$\begin{aligned}
0 &\geq g_{\hat{\theta}}(\theta', \theta', x(\theta')) + \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'} \times [g_x(\theta', \theta', x(\theta')) \\
&\quad + \tfrac{1}{2} g_{xx}(b(\beta))(x(\theta'') - x(\theta')) + g_{\hat{\theta}x}(b(\beta))(\theta'' - \theta')] + \tfrac{1}{2} g_{\hat{\theta}\hat{\theta}}(b(\beta))(\theta'' - \theta') \\
&\geq -[g_{\theta\hat{\theta}}(b(\beta')) + \tfrac{1}{2} g_{\theta\theta}(a)](\theta'' - \theta') - g_{\theta x}(b(\beta'))(x(\theta'') - x(\theta')),
\end{aligned}$$

for $b(\beta) = (\theta', \beta\theta' + (1-\beta)\theta'', \beta x(\theta') + (1-\beta)x(\theta''))$ and some $\beta, \beta' \in [0,1]$. Because $\tilde{V}^I$ is twice differentiable, all the derivatives of g are finite. Moreover, continuity of x implies that $x(\theta'') \to x(\theta')$ as $\theta'' \to \theta'$ and, therefore, for $\theta'' \to \theta'$,

$$0 \geq g_{\hat{\theta}}(\theta', \theta', x(\theta')) + \lim_{\theta'' \to \theta'} \frac{x(\theta'') - x(\theta')}{\theta'' - \theta'} \, g_x(\theta', \theta', x(\theta')) \geq 0.$$

By Lemma C.3, x and, hence, $\hat{\theta}$ are strictly increasing for all $\theta \geq \bar{\theta}(\lambda_t)$. Arguing similarly as we did to show continuity, optimality of x, therefore, requires that $g_x = \partial\tilde{V}^I/\partial x_t \neq 0$ and, hence, the limit of $(x(\theta'') - x(\theta'))/(\theta'' - \theta')$ is well defined, yielding

$$\frac{dx}{d\theta_t} = -\frac{\partial\tilde{V}^I/\partial\hat{\theta}_t}{\partial\tilde{V}^I/\partial x_t}. \tag{C.1}$$

Lemma C.5. *$x(\theta_t, \lambda_t) = \xi(\theta_t)$ for all $\theta_t > \bar{\theta}(\lambda_t)$, where ξ is unique and $\partial\xi/\partial\theta_t > 0$.*

Proof. From Lemma C.4 we have that ξ is differentiable, and by Lemma C.3, $\partial\xi/\partial\theta_t > 0$. We thus only need to show that ξ is unique. By the proof to Lemma C.4, $dx/d\theta_t$ is pinned down by the partial differential equation (C.1), which must hold for all $x_t \geq x(\bar{\theta}(\lambda_t))$. Moreover, whenever $\bar{\theta}(\lambda_t) < 1$, in equilibrium $\hat{\theta}(x(1)) = 1$ and, therefore, it obviously must hold that $x(1, \lambda_t) = \arg\max_{x_t} \tilde{V}^I(1, 1, x_t)$, providing a boundary condition for (C.1). Because $\tilde{V}^I$ is independent of λ_t, it follows that $x(\theta_t, \lambda_t)$ is uniquely characterized by a function, i.e., $\xi : \theta_t \mapsto x_t$, for all $\theta_t \geq \bar{\theta}(\lambda_t)$.

Lemma C.6. *$\bar{\theta}(\lambda_t)$ is unique.*

Proof. Suppose to the contrary that $\bar{\theta}(\lambda_t)$ is not unique. Then there exist $\bar{\theta}'' > \bar{\theta}'$, defining two distinct equilibria for a given λ_t. By Lemma C.5, there is a unique $\xi(\theta)$ characterizing reforms outside the pool for both equilibria. Optimality for type $\theta \in (\bar{\theta}', \bar{\theta}'')$ then requires $\tilde{V}^I(\theta, \theta, \xi(\theta)) \geq \tilde{V}^I(\theta, \bar{\theta}'/2, \lambda_t)$ in the equilibrium defined by $\bar{\theta}'$, and $\tilde{V}^I(\theta, \theta, \xi(\theta)) \leq \tilde{V}^I(\theta, \bar{\theta}''/2, \lambda_t)$ in the equilibrium defined by $\bar{\theta}''$. However, $\tilde{V}^I(\theta, \bar{\theta}'/2, \lambda_t) > \tilde{V}^I(\theta, \bar{\theta}''/2, \lambda_t)$, a contradiction.

This establishes uniqueness of $x(\theta_t, \lambda_t)$, with all properties given by Lemmas C.3 and C.5, and the corresponding beliefs $\hat{\theta}(\lambda_t, x_t)$ following from Lemma C.2 and Bayesian updating. Again, for the purpose of clarity we have established this proposition by focusing on pure strategy equilibria. In the following we outline how the proof generalizes to mixed strategy equilibria; a detailed version of these steps can be attained from the authors on request.

Replicating the proof of Lemma C.1, it is trivial to show that if $\tilde{V}^I(\theta', \hat{\theta}(x'), x') = \tilde{V}^I(\theta', \hat{\theta}(x''), x'')$, then $\tilde{V}^I(\theta'', \hat{\theta}(x'), x') < \tilde{V}^I(\theta'', \hat{\theta}(x''), x'')$ for all $\theta' < \theta''$ and $x' < x''$. It follows that (i) supports, $\mathcal{X}(\theta)$, are non-overlapping, and (ii) $\min \mathcal{X}(\theta'') \geq \max \mathcal{X}(\theta')$. Moreover, noting that $\tilde{x}(\theta) \equiv \max \mathcal{X}(\theta)$ has a jump-discontinuity if and only if type θ mixes in a nondegenerate way, (ii) further implies that there can be only finitely many types that mix on the closed interval $[0, 1]$. Lemmas C.2, C.3, and C.4 then apply with minor changes, ruling out any jumps of $\tilde{x}$ on $[\bar{\theta}(\lambda_t), 1]$. This leads to the conclusion that at most a mass zero of types (i.e., $\theta_t = \bar{\theta}(\lambda_t)$) could possibly mix in any equilibrium (with no impact on $\hat{\theta}$) and, thus, there is no need to consider any nondegenerate mixed strategies. $\square$

C.1.4 Proof of Proposition 3.3

From the discussion in the main body of the paper it is clear that the equilibrium is uniquely pinned down by the time-invariant mappings given by Propositions 3.1 and 3.2 if it exists. We are thus left to show existence, which requires us to verify that the equilibrium mappings are consistent with the D1 and trembling-hand criterion. The first is a direct implication from the proof of Proposition 3.2 where we apply Lemma C.2 to restrict off-equilibrium beliefs, such that $\hat{\theta}$ is necessarily consistent with the D1 criterion.

To show consistency with the concept of trembling-hand perfection, we need to show that $\{\phi_i : i \in [0, 1]\}$ and x are best responses to a sequence of completely mixed strategy profiles $\{\{\omega_i^k : i \in [0, 1]\}, \sigma^k\}_{k=0}^{\infty}$ that converge to a profile that places all mass on $\{\phi_i : i \in [0, 1]\}$ and x, respectively.

Accordingly, for $\phi_i(\hat{\theta}^k(\cdot, x_t), x_t)$ to be a best-response to x_t and the perturbed strategy profile $\{\omega_i^k : i \in [0, 1]\}$ for the marginal outsider i with $\gamma_i = \bar{\gamma}(s_t)$, we need that $\hat{\theta}^k(\cdot, x_t)\, \psi(s_t^k(x_t)) = \bar{\gamma}(s_t)$, requiring any change in beliefs along the perturbation path to be offset by trembles of outsiders $j \neq i$. Because for $x \in [\xi(1), 1]$, $\hat{\theta}(\cdot, x) = 1$ can never be sustained in a completely mixed equilibrium with a continuum of types, this implies that we need to adjust for $\hat{\theta}^k(\cdot, x) < \hat{\theta}(\cdot, x)$ by introducing asymmetric trembles, leading to $s^k(x) > s(\hat{\theta}(\cdot, x), x)$. Hence, let $s^k(x(1)) = s(x(1)) + \varepsilon_k$ for some $\{\varepsilon^k\}_{k=0}^{\infty}$ such that $\lim_{k\to\infty} \varepsilon^k = 0$ and $\varepsilon^k \in (0, \bar{\varepsilon})$ for all k.

A necessary (and for $\theta \in (\bar{\theta}(\cdot), 1)$ sufficient) condition for $x \in [\xi(\bar{\theta}(\cdot)), \xi(1)]$ to be optimal against s^k is that $s^k(x)$ satisfies the inverse differential equation (C.1) for $x(\cdot,\theta)$ fixed,

$$\frac{ds^k}{dx} = -\frac{\partial V^I/\partial x}{\partial V^I/\partial s}\bigg|_{s=s^k}, \tag{C.2}$$

which in combination with $s^k(x(1))$ pins down $s^k(x)$ for all $x \in [\xi(\bar{\theta}), \xi(1)]$. Note that $s^k(x(1)) > s(\cdot, x(1))$ implies that $s^k(x) > s(\cdot, x)$ for all $x \in [\xi(\bar{\theta}), \xi(1)]$ since the indifference condition (C.1) is unique. Moreover, since optimality of x requires that $\bar{\theta}$ is necessarily indifferent between λ_t and $\xi(\bar{\theta})$, $s^k(\xi(\bar{\theta}))$ pins down $s^k(\lambda_t) > s(\cdot, \lambda_t)$.

For off-equilibrium $x \in (\lambda, \xi(\bar{\theta})) \cup (\xi(1), 1]$ we are free to assign any $s^k(x)$ that (1) assures optimality of x, and (2) converges to $s(\cdot, x)$. As to (1), we can for instance set $s^k(x) = s(\bar{\theta}, x) + s^k(\xi(\bar{\theta})) - s(\cdot, \xi(\bar{\theta}))$ for $x \in (\lambda, \xi(\bar{\theta}))$ (which is continuous around $\xi(\bar{\theta})$ and has slope $ds(\bar{\theta}, x)/dx \geq ds^k(\xi(\bar{\theta}))/dx$, so that by (C.2) no type has an incentive to deviate), and $s^k(x) = s(\cdot, x) + \varepsilon^k f^k(x)$ for $x \in (\xi(1), 1]$ with some $f^k : [\xi(1), 1] \to \mathbb{R}_+$ such that $df^k(\xi(1))/dx = \{ds^k(\xi(1))/dx - ds(\cdot, \xi(1))/dx\}/\varepsilon^k$ and f^k sufficiently convex for V^I to be concave on $[\xi(1), 1]$, so that $\xi(1)$ is the global optimum for $\theta = 1$.

Note that these definitions imply that $s^k(x) \downarrow s(\hat{\theta}(\cdot, x), x)$ for all x and, hence, $\hat{\theta}^k(\cdot, x) \uparrow \hat{\theta}(\cdot, x)$ for all x as implied by the indifference condition of the marginal outsider, $\hat{\theta}^k(x) = \bar{\gamma}(s(\cdot, x))/\psi(s^k(x)) \in (0, \hat{\theta}(\cdot, x))$. By construction, these sequences assure optimality of $\{\phi_i : i \in [0,1]\}$ and x along the perturbation path. To conclude the proof it therefore suffices to show the existence of $\{\{\omega_i^k : i \in [0,1]\}, \sigma^k\}_{k=0}^{\infty}$ yielding $\{s^k, \hat{\theta}^k\}_{k=0}^{\infty}$.

Consider $\{s^k\}_{k=0}^{\infty}$ first. Define $\tilde{\varepsilon}$ such that $\max_x s^k(x) < 1 - \lambda$ for $\varepsilon^k = \tilde{\varepsilon}$ and suppose that $\bar{\varepsilon} \leq \tilde{\varepsilon}$.[1] Then any s^k can be sustained by setting

$$\omega_i^k(1)(x) = \begin{cases} 1 - \varepsilon^k & \text{for all } i : \gamma_i \leq \bar{\gamma}(s(\hat{\theta}(\cdot, x), x)) \\ c^k(x)\varepsilon^k & \text{for all } i : \gamma_i > \bar{\gamma}(s(\hat{\theta}(\cdot, x), x)), \end{cases}$$

with $c^k(x) = \{s^k(x) - (1 - \varepsilon^k)s(\cdot, x)\}/\{(1 - x)(1 - \bar{\gamma}(x))\varepsilon^k\}$. Note that ω_i^k is completely mixed if $\bar{\varepsilon} < 1$ and $\varepsilon^k c^k(x) \in (0,1) \iff c^k(x) \in (0, 1/\varepsilon^k) \iff s^k(x) + \varepsilon^k s(\cdot, x) < 1 - x$. From $s^k(x) > s(\cdot, x)$ we have that $c^k(x) > 0$ and because $s^k \to s$, using the same arguments as in Footnote 1, there exists some $\hat{\varepsilon}$ such that $c^k(x) < 1/\varepsilon^k$ holds for all $\bar{\varepsilon} \leq \hat{\varepsilon}$.

Finally, consider $\{\hat{\theta}^k\}_{k=0}^{\infty}$. It is straightforward to verify by Bayes rule that any $\hat{\theta}^k$ with

[1] To see that $\tilde{\varepsilon}$ exists, note that $s(\hat{\theta}(\cdot, x), x) < 1 - x \leq 1 - \lambda_t$ since otherwise $\bar{\gamma}_t = 1$, which requires $\bar{\theta}_t = 1$ and $s_t = 1$, contradicting that s is strictly decreasing in x. Convergence of s^k to s then implies that one can always find some $\tilde{\varepsilon}$ that is sufficiently small.

$\hat{\theta}^k(x) > 0$ for all x can be sustained by setting

$$\sigma^k(x)(\theta,\cdot) = \begin{cases} \varepsilon^k & \text{if } \theta > \hat{\theta}^k(x) \text{ and } (x > \lambda_t \text{ or } \theta > \bar{\theta}_t) \\ d^k(x)\varepsilon^k & \text{if } \theta < \hat{\theta}^k(x) \text{ and } x > \lambda_t \\ 1 - R^k(\theta) & \text{if } \theta \geq \bar{\theta}(\lambda_t) \text{ and } x = \xi(\theta) \\ T^k & \text{if } \theta \leq \hat{\theta}^k(\lambda_t) \text{ and } x = \lambda_t \\ Z^k & \text{if } \theta \in (\hat{\theta}^k(\lambda_t), \bar{\theta}(\lambda_t)) \text{ and } x = \lambda_t, \end{cases}$$

with $d^k(x) = (1 - \hat{\theta}^k(x))^2/\hat{\theta}^k(x)^2$, $R^k(\theta) = \int_{\theta > \hat{\theta}(x)} \varepsilon^k \, dx + \int_{\theta < \hat{\theta}(x)} d^k(x)\varepsilon^k \, dx$, $T^k = \inf_{\theta < \bar{\theta}(\lambda_t)}(1 - R^k(\theta))$, and $Z^k = \{T^k\hat{\theta}^k(x)^2 + \varepsilon^k[2(1 - \bar{\theta}(\lambda_t))\hat{\theta}^k(\lambda_t) - 1 + \bar{\theta}(\lambda_t)^2]\}/\{\bar{\theta}(\lambda_t) - \hat{\theta}^k(\lambda_t)\}^2$. With a slight abuse of notation, in the definition of σ^k, R^k , T^k and Z^k denote probabilities, while ε^k are understood to be probability densities. Note that σ^k is completely mixed if $T^k, R^k(\theta) \in (0,1)$ and $Z^k \in (0, R^k(\theta))$ for all θ. This is obviously true for some $\breve{\varepsilon}$, such that $\bar{\varepsilon} < \breve{\varepsilon}$. Finally, note that the above definition is incomplete in the sense that $R^k(\theta) + T^k < 1$ or $R^k(\theta) + Z^k < 1$ for some types $\theta < \bar{\theta}(\lambda_t)$. In these cases the remaining probability mass can be distributed (almost) arbitrary over atoms on $(\lambda_t, 1]$ without impact on the resulting beliefs.[2]

We conclude the proof by setting $\bar{\varepsilon} = \min\{1, \tilde{\varepsilon}, \hat{\varepsilon}, \breve{\varepsilon}\}$. □

C.1.5 Proof of Proposition 3.4

Consider $Q^R(\lambda_t, (\frac{1}{2}, 1]) = 1$ first. By Proposition 2, for any reform $x_t > \lambda_t$, $x_t = \xi(\theta_t)$, with ξ increasing. To show the claim, it thus suffices to show that $\tilde{x} \equiv \xi(\tilde{\theta}) > 1/2$ for $\tilde{\theta} = \min_\lambda \bar{\theta}(\lambda)$. Also, define $\tilde{\lambda} = \arg\min_\lambda \bar{\theta}(\lambda)$. Then, optimality of $\tilde{x}$ implies $s^* \equiv s(\tilde{\theta}/2, \tilde{\lambda}) > s(\tilde{\theta}, \tilde{x}) \equiv s^{**}$. Using (3.6),

$$s^* = (\tilde{\theta}/2)(1 - \tilde{\lambda})\,\psi(s^*) \equiv w^*\psi(s^*), \tag{C.3}$$

$$s^{**} = \tilde{\theta}(1 - \tilde{x})\,\psi(s^{**}) \equiv w^{**}\psi(s^{**}). \tag{C.4}$$

Note that, in analogue to the proof of Proposition 3.1, for a general $w_t \equiv \hat{\theta}_t(1 - x_t)$ it holds that

$$\frac{\partial s_t}{\partial w_t} = -\psi(s_t)\left(\frac{\partial \pi_t}{\partial s_t}\right)^{-1} > 0.$$

[2] For instance, each type θ could place the remaining probability mass on an atom at $x = \lambda_t + \theta(1 - \lambda_t)/\bar{\theta}(\lambda_t)$.

Hence, $s^* > s^{**}$ implies $w^* > w^{**}$, or $(\tilde{\theta}/2)(1-\tilde{\lambda}) > \tilde{\theta}(1-\tilde{x})$. Rearranging, then proves the claim,

$$\tilde{x} > 1 - \frac{1-\tilde{\lambda}}{2} \geq \frac{1}{2}.$$

Now consider $Q^S(\lambda_t, (0, \frac{1}{2})) = 1$. Again, optimality of x_t implies that $s(\hat{\theta}(\lambda_t, x), x)$ is decreasing in x. Hence, for all λ_t,

$$s(\hat{\theta}(\lambda_t, x_t), x_t) \leq s(\bar{\theta}(\lambda_t)/2, \lambda_t) \leq s(1/2, 0),$$

where the last inequality follows since s is increasing in its first and decreasing in its second argument. Hence, it suffices to show that $s(1/2, 0) < 1/2$.

Let $s^* \equiv s(1,0) \leq 1$ and let $s^{**} \equiv s(1/2, 0)$. From (3.6), $s^* = \psi(s^*)$ and $s^{**} = \psi(s^{**})/2$. Moreover, by Proposition 3.1, $s^* > s^{**}$. Hence, since ψ is strictly increasing,

$$s^{**} = \frac{\psi(s^{**})}{2} = \frac{\psi(\psi(s^{**})/2)}{2} < \frac{\psi(\psi(s^*)/2)}{2} = \frac{\psi(s^*/2)}{2} < \frac{\psi(s^*)}{2} = \frac{s^*}{2} \leq \frac{1}{2}.$$

□

C.1.6 Proof of Proposition 3.5

From Footnote 13,

$$\rho^S(\lambda_t) = \int_0^{\bar{\theta}(\lambda_t)} \theta h\left(s\left(\bar{\theta}(\lambda_t)/2, \lambda_t\right)\right) d\theta + \int_{\bar{\theta}(\lambda_t)}^1 \theta h\left(s\left(\theta, x(\theta)\right)\right) d\theta, \tag{C.5}$$

and

$$\rho^R(\lambda_t) = \int_{\bar{\theta}(\lambda_t)}^1 \left(1 - \theta h\left(s\left(\theta, x(\theta)\right)\right)\right) d\theta. \tag{C.6}$$

Also, note that $\bar{\theta}(\lambda_t) \in (0, 1]$ is implicitly defined as the solution to

$$F(\bar{\theta}, \lambda_t) \equiv \tilde{V}^I(\bar{\theta}, \bar{\theta}/2, \lambda_t) - \tilde{V}^I(\bar{\theta}, \bar{\theta}, \xi(\bar{\theta})) = 0, \tag{C.7}$$

if an interior solution exists. Otherwise, for λ_t there is a corner solution $\bar{\theta}(\lambda_t) = 1$, which implies $\tilde{V}^I(1, 1/2, \lambda_t) > \tilde{V}^I(1, 1, \xi(1))$.

First, consider $\lambda_t > \bar{\lambda}$. Suppose that there exists $\bar{\lambda}$, such that for all $\lambda_t \in (\bar{\lambda}, 1]$, $\bar{\theta}(\lambda_t)$ is a corner solution. Then clearly for all $\lambda_t > \bar{\lambda}$, $\partial\bar{\theta}(\lambda_t)/\partial\lambda_t = 0$, such that $\partial\rho^S(\lambda_t)/\partial\lambda_t =$

$\partial h(s(1/2, \lambda_t))/\partial \lambda_t < 0$, by Proposition 3.1. Furthermore, $\partial \rho^R(\lambda_t)/\partial \lambda_t = 0$. Otherwise, if there exists no $\bar{\lambda}$, such that for all $\lambda_t \in (\bar{\lambda}, 1]$, $\bar{\theta}(\lambda_t)$ is a corner solution, then there necessarily exists a λ^*, such that for $\lambda_t \in (\lambda^*, 1]$, $\bar{\theta}(\lambda_t)$ is an interior solution. But then, because $\rho^S(1) = \rho^R(1) = 0$, continuity of ρ^S and ρ^R implies that $\partial \rho^S(\lambda_t)/\partial \lambda_t < 0$ and $\partial \rho^R(\lambda_t)/\partial \lambda_t < 0$ for all $\lambda_t > \bar{\lambda}$ and some $\bar{\lambda} < 1$.

Now consider $\lambda_t < \underline{\lambda}$ and $\bar{\theta}(0) < 1$. Then, F differentiable implies that $\bar{\theta}(\lambda_t)$ has an interior solution and is differentiable for all $\lambda_t \in [0, \lambda^*)$ for some $\lambda^* > 0$. Implicit differentiation of F, substituting for $x'(\bar{\theta})$ from (C.1), and using $F(\bar{\theta}, \lambda_t) = 0$ yields

$$\frac{\partial \bar{\theta}(\lambda)}{\partial \lambda} = \frac{-\bar{\theta} h_1^p s_2^p u^p + (1-p^p) u_1^p}{\frac{\bar{\theta}}{2} h_1^p s_1^p u^p + \frac{u^p - u^s}{\bar{\theta}}}, \tag{C.8}$$

where subscript i denotes the derivative with respect to the ith argument, and superscripts p and s denote that the function is evaluated at the pooling or separating values, respectively (where $\hat{\theta}^p = \frac{\bar{\theta}}{2}$, $x^p = \lambda$ and $\hat{\theta}^s = \bar{\theta}, x^s = x(\bar{\theta})$).

Using this, the signs of $\partial \rho^S/\partial \lambda_t$ and $\partial \rho^I/\partial \lambda_t$ are given by

$$\begin{aligned}\operatorname{sign}\left\{\frac{\partial \rho^S(\lambda_t)}{\partial \lambda_t}\right\} = \operatorname{sign}\Bigg\{ & u^P \left(\frac{(p^P - p^S)(1-2p^S)}{1-p^S}\right) + \\ & + (1-p^P) u_1^P \left((1-\lambda_t) - \frac{2(p^P - p^S)}{\bar{\theta} h_1^P s_2^P}\right)\Bigg\}\end{aligned} \tag{C.9}$$

and

$$\operatorname{sign}\left\{\frac{\partial \rho^R(\lambda_t)}{\partial \lambda_t}\right\} = \operatorname{sign}\left\{-\frac{\partial \bar{\theta}(\lambda_t)}{\partial \lambda_t}(1-p^S)\right\}, \tag{C.10}$$

where we have used that $(1-p^P)u^P = (1-p^S)u^S$ from (C.7) and $s_1^P/(-s_2^P) = 2(1-\lambda_t)/\bar{\theta}$ by the proof of Proposition 3.1.

Evaluated at $\lambda_t = 0$, all terms except u_1 in (C.9) are strictly positive.[3] Thus, $\partial \rho^S(0)/\partial \lambda_t$ is weakly positive if and only if for $\lambda_t = 0$ it holds that

$$u_1^P \geq -u^P \left(\frac{(p^P - p^S)(1-2p^S)}{1-p^S}\right) \left[(1-p^P)\left((1-\lambda_t) - \frac{2(p^P - p^S)}{\bar{\theta} h_1^P s_2^P}\right)\right]^{-1}. \tag{C.11}$$

[3]Note that $p^S = \bar{\theta} h(s^S) < 1/2$ for $\lambda_t = 0$ is not obvious. To see that this is indeed the case, assume to the contrary $p^S > 1/2$ implying $p^P = \bar{\theta}/2\, h(s^P) > 1/2$. By Proposition 3.4, $s^P = \bar{\theta}/2\, h(s^P) u(s^P) = p^P/2\, u(s^P) < 1/2$ and hence $u(s^P) < 1/p^P < 2$ by $p^P > 1/2$. Furthermore, optimality of $\bar{\xi} \equiv \xi(\bar{\theta})$ requires $(1-p^S)u(\bar{\xi}) \geq 1$, since an indirect utility of 1 is always attainable by setting $x = 1$. This implies $u(\bar{\xi}) \geq 2$ by $p^S > 1/2$. Thus, $p^S > 1/2$ implies $u(s^P) < 2 \leq u(\bar{\xi})$ for $\lambda_t = 0$. However, by Proposition 3.4, $s^P < 1/2 < \bar{\xi}$ such that $u(s^P) > u(\bar{\xi})$, a contradiction.

Likewise, note that the sign of $\partial\rho^R/\partial\lambda_t$ is the opposite sign of $\partial\bar{\theta}(\lambda_t)/\partial\lambda_t$. Hence, because all terms except u_1 in (C.8) are strictly positive, $\partial\rho^R/\partial\lambda_t$ is weakly negative if and only if

$$u_1^P \geq \bar{\theta} h_1^p s_2^p u^p (1 - p^P)^{-1}. \tag{C.12}$$

Let u' and u'' be the values of the right hand sides of (C.11) and (C.12) when evaluated at $\lambda_t = 0$. Then, from our discussion above it follows, that $\partial\rho^S(0)/\partial\lambda_t > 0$ and $\partial\rho^R(0)/\partial\lambda_t < 0$ if $u_1(0) > \bar{u} \equiv \max\{u', u''\}$. The converse—that is, $\partial\rho^S(0)/\partial\lambda_t < 0$ and $\partial\rho^R(0)/\partial\lambda_t > 0$—holds true, if $u_1(0) < \underline{u} \equiv \min\{u', u''\}$. Differentiability of ρ^S and ρ^R around 0 thus establishes the claim for all $\lambda_t \in [0, \underline{\lambda}]$ for some $\underline{\lambda} > 0$. $\square$

Bibliography

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Acemoglu, D. and J. A. Robinson (2000a): "Democratization or Repression?" *European Economic Review*, 44, 683–693.

——— (2000b): "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective," *Quarterly Journal of Economics*, 115, 1167–1199.

——— (2001): "A Theory of Political Transitions," *American Economic Review*, 91, 938–963.

Acemoglu, D., D. Ticchi, and A. Vindigni (2010): "A Theory of Military Dictatorships," *American Economic Journal: Macroeconomics*, 2, 1–42.

Ahlbrecht, M. and M. Weber (1997): "Preference for Gradual Resolution of Uncertainty," *Theory and Decision*, 43, 167–185.

Aidt, T. S. and P. S. Jensen (2012): "Workers of the World, Unite! Franchise Extensions and the Threat of Revolution in Europe, 1820-1938," *Cambridge Working Paper in Economics.*

Angeletos, G.-M., C. Hellwig, and A. Pavan (2006): "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, 114, 452–484.

Arai, D. (1997): "Temporal Resolution of Uncertainty in Risky Choices," *Acta Psychologica*, 96, 15–26.

Arkes, H. R. and C. Blumer (1985): "The Psychology of Sunk Cost," *Organizational Behavior and Human Decision Processes*, 35, 124–140.

Barber, B. M., T. Odean, and N. Zhu (2009): "Systematic Noise," *Journal of Financial Markets*, 12, 547–569.

Barberis, N. and W. Xiong (2009): "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation," *Journal of Finance*, 64, 751–784.

Bell, D. E. (1985): "Disappointment in Decision Making Under Uncertainty," *Operations Research*, 33, 1–27.

Benartzi, S. (2001): "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," *Journal of Finance*, 56, 1747–1764.

Boix, C. (2003): *Democracy and Redistribution*, Cambridge and UK: Cambridge University Press.

Bourguignon, F. and T. Verdier (2000): "Oligarchy, Democracy, Inequality and Growth," *Journal of Development Economics*, 62, 285–313.

Bruhin, A., H. Fehr-Duda, and T. Epper (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78, 1375–1412.

Busse, M. R., D. G. Pope, J. C. Pope, and J. Silva-Risso (2012): "Projection Bias in the Car and Housing Markets," *NBER Working Paper*, 18212.

Caplin, A. and J. Leahy (2001): "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55–79.

Carbajal, J. C. and J. C. Ely (2012): "Optimal Contracts for Loss Averse Consumers," *Working Paper*.

Card, D. and G. B. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 126, 103–143.

Cervellati, M., P. Fortunato, and U. Sunde (2007): "Consensual and Conflictual Democratization, Rule of Law, and Development," *CEPR Discussion Paper*, 6328.

——— (2011): "Democratization and Civil Liberties: The Role of Violence during the Transition," *IZA Discussion Paper*, 5555.

Chavas, J.-P. (2000): "On Information and Market Dynamics: The Case of the U.S. Beef Market," *Journal of Economic Dynamics and Control*, 24, 833–853.

Cheibub, J. A., J. Gandhi, and J. R. Vreeland (2010): "Democracy and Dictatorship Revisited," *Public Choice*, 143, 67–101.

Chew, S. H. and J. L. Ho (1994): "Hope: An Empirical Study of Attitude Toward the Timing of Uncertainty Resolution," *Journal of Risk and Uncertainty*, 8, 267–288.

Cho, I.-K. and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2009): "Reinforcement Learning and Savings Behavior," *Journal of Finance*, 64, 2515–2534.

Conley, J. P. and A. Temini (2001): "Endogenous Enfranchisement when Groups' Preferences Conflict," *Journal of Political Economy*, 79–102.

Conlin, M., T. O'Donoghue, and T. J. Vogelsang (2007): "Projection Bias in Catalog Orders," *American Economic Review*, 97, 1217–1249.

Crawford, V. P. and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

Daido, K. and T. Murooka (2012): "Team Incentives and Reference-Dependent Preferences," *Working Paper*.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 9, 522–550.

Drèze, J. H. and F. Modigliani (1972): "Consumption Decisions under Uncertainty," *Journal of Economic Theory*, 5, 308–335.

Edmond, C. (2011): "Information Manipulation, Coordination, and Regime Change," *NBER Working Paper*, 17395.

Eisenhuth, R. (2012): "Reference Dependent Mechanism Design," *Working Paper*.

Eliaz, K. and A. Schotter (2007): "Experimental Testing of Intrinsic Preferences for Non-lnstrumental Information," *American Economic Review*, 97, 166–169.

Elkins, Z., T. Ginsburg, and J. Melton (2010): "Chronology of Constitutional Events, Version 1.1," .

Ellis, C. J. and J. Fender (2011): "Information Cascades and Revolutionary Regime Transitions," *Economic Journal*, 121, 763–792.

Epstein, L. G. and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Ericson, K. M. M. and A. Fuster (2011): "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Quarterly Journal of Economics*, 126, 1879–1907.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10, 171–178.

Fuster, A., D. Laibson, and B. Mendel (2010): "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 24, 67–84.

Gates, S., H. Hegre, M. P. Jones, and H. Strand (2006): "Institutional Inconsistency and Political Instability: Polity Duration, 1800–2000," *American Journal of Political Science*, 50, 893–908.

Gaudecker, H.-M. v., A. van Soest, and E. Wengström (2011): "Heterogeneity in Risky Choice Behavior in a Broad Population," *American Economic Review*, 101, 664–694.

Gill, D. and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102, 469–503.

Goemans, H. E., K. S. Gleditsch, and G. Chiozza (2009): "Introducing Archigos: A Dataset of Political Leaders," *Journal of Peace Research*, 46, 269–283.

Gradstein, M. (2007): "Inequality, Democracy and the Protection of Property Rights," *Economic Journal*, 117, 252–269.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen 2003*, ed. by K. Kremer and V. Macho, Göttingen: Ges. für Wiss. Datenverarbeitung, 79–93.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2012): "Learning Through Noticing: Theory and Experimental Evidence in Farming," *NBER Working Paper*, 18401.

Hassan, T. A. and T. M. Mertens (2011): "The Social Cost of Near-Rational Investment," *NBER Working Paper*, 17027.

Heffetz, O. and J. A. List (2011): “Is the Endowment Effect a Reference Effect?” *NBER Working Paper*, 16715.

Heidhues, P. and B. Kőszegi (2008): “Competition and Price Variation When Consumers Are Loss Averse,” *American Economic Review*, 98, 1245–1268.

Herweg, F., D. Müller, and P. Weinschenk (2010): “Binary Payment Schemes: Moral Hazard and Loss Aversion,” *American Economic Review*, 100, 2451–2477.

Herweg, F. and K. M. Schmidt (2012): “A Theory of Ex Post Inefficient Renegotiation,” *Working Paper*.

Holt, C. A. and S. K. Laury (2002): “Risk Aversion and Incentive Effects,” *American Economic Review*, 92, 1644–1655.

Hommes, C. (2011): “The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab,” *Journal of Economic Dynamics and Control*, 35, 1–24.

Huntington, S. P. (1991): *The Third Wave: Democratization in the Late Twentieth Century*, Norman (OK): University of Oklahoma Press.

Jack, W. and R. Lagunoff (2006): “Dynamic Enfranchisement,” *Journal of Public Economics*, 90, 551–572.

Justman, M. and M. Gradstein (1999): “The Industrial Revolution, Political Transition, and the Subsequent Decline in Inequality in 19th-Century Britain,” *Explorations in Economic History*, 36, 109–127.

Kahneman, D. and A. Tversky (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 263–291.

Karl, T. L. (1990): “Dilemmas of Democratization in Latin America,” *Comparative Politics*, 23, 1–21.

Karle, H. and M. Peitz (2010): “Pricing and Information Disclosure in Markets with Loss-Averse Consumers,” *CEPR Discussion Paper*, 7785.

Kaustia, M. and S. Knüpfer (2008): “Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions,” *Journal of Finance*, 63, 2679–2702.

Kőszegi, B. and M. Rabin (2006): “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–936.

Kocher, M., M. Krawczyk, and F. van Winden (2009): "Let Me Dream On! Anticipatory Emotions and Preferences for Timing in Lotteries," *Working Paper*.

Kreps, D. M. and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Kuran, T. (1989): "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution," *Public Choice*, 61, 41–74.

Lizzeri, A. and N. Persico (2004): "Why did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'," *Quarterly Journal of Economics*, 119, 707–765.

Llavador, H. and R. J. Oxoby (2005): "Partisan Competition, Growth, and the Franchise," *Quarterly Journal of Economics*, 120, 1155–1189.

Loewenstein, G. (1996): "Out of Control: Visceral Influences on Behavior," *Organizational Behavior and Human Decision Processes*, 65, 272–292.

Loewenstein, G., D. Nagin, and R. Paternoster (1997): "The Effect of Sexual Arousal on Expectations of Sexual Forcefulness," *Journal of Research in Crime and Delinquency*, 34, 443–473.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 118, 1209–1248.

Loewenstein, G. and D. Schkade (1999): "Wouldn't It Be Nice? Predicting Future Feelings," in *Well-Being: The Foundations of Hedonic Psychology*, ed. by D. Kahneman, E. Diener, and N. Schwarz, New York: Russell Sage Foundation.

Lohmann, S. (1994): "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91," *World Politics*, 47, 42–101.

Loomes, G. and R. Sugden (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies*, 53, 271.

Magaloni, B. and R. Kricheli (2010): "Political Order and One-Party Rule," *Annual Review of Political Science*, 13, 123–143.

Maier, J. and M. Rüger (2010): "Measuring Risk Aversion Model-Independently," *Working Paper*.

Mailath, G. J. (1987): "Incentive Compatibility in Signaling Games with a Continuum of Types," *Econometrica*, 55, 1349–1365.

Marshall, M. and K. Jaggers (2002): "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2002," .

Mas, A. (2006): "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 121, 783–821.

Matthey, A. and N. Dwenger (2008): "Don't Aim Too High: The Potential Costs of High Aspirations," *Working Paper*.

Nordgren, L. F., J. van der Pligt, and F. van Harreveld (2007): "Evaluating Eve: Visceral States Influence the Evaluation of Impulsive Behavior," *Journal of Personality and Social Psychology*, 93, 75–84.

Noussair, C. and P. Wu (2006): "Risk Tolerance in the Present and the Future: an Experimental Study," *Managerial and Decision Economics*, 27, 401–412.

Odean, T. (1998): "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.

O'Donnell, G. and P. C. Schmitter (1973): *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies*, Baltimore (MD): John Hopkins University Press.

Pahlke, J., S. Strasser, and F. M. Vieider (2012): "Risk-taking for Others under Accountability," *Economics Letters*, 114, 102–105.

Pope, D. G. and M. E. Schweitzer (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 101, 129–157.

Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008): "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98, 38–71.

Przeworski, A. (2000): *Democracy and Development: Political Institutions and Material Well-Being in the World, 1950-1990*, Cambridge (UK): Cambridge University Press.

——— (2009): "Conquered or Granted? A History of Suffrage Extensions," *British Journal of Political Science*, 39, 291.

READ, D. AND B. VAN LEEUWEN (1998): "Predicting Hunger: The Effects of Appetite and Delay on Choice," *Organizational Behavior and Human Decision Processes*, 76, 189–205.

RUSTOW, D. A. (1970): "Transitions to Democracy: Toward a Dynamic Model," *Comparative Politics*, 2, 337–363.

SCHWARTZSTEIN, J. (2012): "Selective Attention and Learning," *Working Paper*.

SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

SIMONSOHN, U. (2010): "Weather to Go to College," *Economic Journal*, 120, 270–280.

SMITH, A. (2008): "Lagged Beliefs and Reference-Dependent Utility," *Working Paper*.

SONG, C. (2012): "An Experiment on Reference Points and Expectations," *unpublished*.

SPENCE, M. AND R. ZECKHAUSER (1972): "The Effect of the Timing of Consumption Decisions and the Resolution of Lotteries on the Choice of Lotteries," *Econometrica*, 40, 401–403.

SPRENGER, C. (2010): "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points," *Working Paper*.

VAN BOVEN, L. AND G. LOEWENSTEIN (2003): "Social Projection of Transient Drive States," *Personality and Social Psychology Bulletin*, 29, 1159–1168.

VAN WINDEN, F., M. KRAWCZYK, AND A. HOPFENSITZ (2011): "Investment, Resolution of Risk, and the Role of Affect," *Journal of Economic Psychology*, 32, 918–939.

WU, G. (1999): "Anxiety and Decision Making with Delayed Resolution of Uncertainty," *Theory and Decision*, 46, 159–199.

ZIMMERMANN, F. (2012): "Clumped or Piecewise? Evidence on Preferences for Information," *Working Paper*.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

14. Dezember 2012

Lukas Buchheim