Evolution, Bounded Rationality and Institutions

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2017

vorgelegt
von

Dipl. Math. Johannes Buckenmaier

aus

Balingen

Referent: Prof. Dr. Carlos Alós-Ferrer
Korreferent: Prof. Dr. Larbi Alaoui
Tag der Promotion: 06. November 2017

For Manuela

Table of Contents

List of Tables

List of Figures

Introduction and Summary

This dissertation consists of five self-contained research papers that cover theoretical work, simulation-based research, and experimental studies. My research interests are mainly focused on two interrelated areas within economics. One area is concerned with the economic consequences as well as the foundations of boundedly rational behavior. The other area is more specific and concerns the design of institutions and how they can be used to shape behavior and align incentives. The first chapter belongs to both areas, Chapters 2 and 3 cover topics from the former area, whereas the last two chapters contribute to the latter area. Chapter 1 concerns the role of trader matching with regard to the selection of market institutions by boundedly rational traders. Chapter 2 presents results on the stability of the Cournot-Nash and the Walrasian equilibrium under imitative behavior. Chapter 3 presents a model linking response times and iterative thinking and provides experimental evidence regarding the underlying processes of iterative thinking. Chapter 4 investigates the effects of a leniency mechanism on collusive bribery and tax evasion. Chapter 5 asks how the timing of punishment and the timing of the resolution of uncertainty affect deterrence of illicit behavior. In the remainder of this section I present a brief introduction for each chapter and summarize the main findings.

Chapter 1 is the result of joint work with Carlos Alós-Ferrer (University of Cologne) and has been published under the title "Trader Matching and the Selection of Market Institutions" in the *Journal of Mathematical Economics*. We analyze a stochastic dynamic learning model with boundedly rational traders who can choose among trading institutions with different matching characteristics. The framework allows for institutions featuring multiple prices (per good), thus violating the "law of one price." We find that centralized institutions are stochastically stable for a broad class of dynamics and behavioral rules, independently of which other institutions are available. However, some decentralized institutions featuring multiple prices can also survive in the long run, depending on specific characteristics of the underlying learning dynamics such as fast transitions or optimistic behavior. Work on this paper was shared among the authors as follows: Carlos Alós-Ferrer 50%, Johannes Buckenmaier 50%.

Chapter 2 is the result of joint work with Carlos Alós-Ferrer (University of Cologne) and has been published under the title "Cournot vs. Walras: A Reappraisal through Simulations" in the *Journal of Economic Dynamics and Control*. Best-reply behavior in Cournot oligopolies generally leads to Cournot-Nash equilibrium, but imitative behavior selects the Walrasian equilibrium as the unique stochastically stable state. Previous work (Alós-Ferrer, 2004) showed that in the presence of memory, imitative behavior leads to a non-trivial dynamics selecting all quantities between the Cournot and Walrasian outcomes. However, the scope of previous results was limited to specific assumptions on demand and cost functions, and did not provide information on the shape of the distribution of outcomes. We use computational simulations to address these limitations. We show that the selection result for non-trivial memory holds beyond the set of well-behaved Cournot games previously analyzed. Further, we find that, in Cournot games, the limit distribution of long-run outcomes is highly skewed towards the Walrasian quantity. Although longer memory increases the importance of the Cournot equilibrium, the competitive outcome remains the dominant prediction. Work on this paper was shared among the authors as follows: Carlos Alós-Ferrer 50%, Johannes Buckenmaier 50%.

Chapter 3, entitled "Cognitive Sophistication and Deliberation Times," is the result of joint work with Carlos Alós-Ferrer (University of Cologne). Cognitive capacities differ among individuals. Models of iterative thinking put forward heterogeneity in the depth of reasoning as a source of individual differences in behavior. So far there has been little direct evidence that sophistication (depth of reasoning) corresponds to cognitive effort. Choice data alone cannot provide such evidence, hence additional evidence is necessary. We argue that deliberation times can provide such evidence. We provide a simple model linking cognitive sophistication and deliberation times, taking into account stylized facts from the psychophysiological literature on response times. The key assumption is that deliberation time is a decreasing function of the hypothetical gain from conducting an additional step of reasoning. We then test the predictions in an experiment. We find longer deliberation times for choices commonly associated with more steps of reasoning in games where iterative thinking is salient, confirming the prediction of our model that deliberation time is increasing in cognitive sophistication. However, this relation breaks down when iterative thinking is not natural or when there is a conflict between alternative decision rules. Further, we find that larger incentives decrease the time required to perform a single step of reasoning, which, in line with our predictions, is consistent with a closeness-to-indifference effect. If the underlying processes are clearly identified, we observe a strong link between deliberation times and steps of reasoning supporting level-k thinking. Additionally, however, deliberation times also allow us to detect when other elements enter the picture, and hence are also helpful for further theory development. Work on this paper was shared among the authors as follows: Carlos Alós-Ferrer 50%, Johannes Buckenmaier 50%.

Chapter 4, entitled "Institutional History, Leniency and Collusive Tax Evasion," is the result of joint work with Eugen Dimant (University of Pennsylvania) and Luigi Mittone (University of Trento). We investigate the effects of an institutional mechanism, that incentivizes tax payers to blow the whistle through a leniency program, on collusive corruption and tax compliance. In our experiment, we nest collusive corruption within a tax evasion framework. We not only study how the presence of such a mechanism affects behavior, but also investigate the role of institutional changes, that is, the dynamic effect caused by the introduction and the removal of leniency. We find that in the presence of a leniency mechanism subjects collude less, accept less bribes and pay more taxes, while we find no evidence that it encourages bribe offers. Further, our results show that the introduction of the opportunity to blow the whistle decreases collusion, decreases the bribe acceptance rate, and increases the tax yield collected, while not encouraging bribe offers. In contrast, the removal of the institutional mechanism does not cause effects in the opposite direction, suggesting a positive spillover effect of leniency that persists even after the mechanism has been removed. Work on this paper was shared among authors as follows: Johannes Buckenmaier 33%, Eugen Dimant 33%, Luigi Mittone 33%.

Chapter 5, entitled "Timing, Uncertainty and Institutional Deterrence," is the result of joint work with Eugen Dimant (University of Pennsylvania), Ann-Christin Posten (University of Cologne) and Ulrich Schmidt (University of Kiel). Reducing criminal acts in society is a crucial duty of governments. Establishing punishment structures to attain this goal involves high costs. Typically, both theorists and practitioners resort to the adjustment of severity and/or certainty of punishment as effective deterrents of criminal behavior. One more cost effective, but scientifically understudied mechanism for effective deterrence is the swiftness or celerity of punishment. We carry out a controlled economic experiment to study the effectiveness of swiftness of punishment along the following two dimensions: the timing of punishment and the timing of the resolution of uncertainty (regarding the punishment). Our results indicate an inverted U-shaped relation between the delay of punishment, the delay of uncertainty resolution regarding the detection of deviant behavior, and any resulting deterrence. In fact, institutions that either reveal detection and impose punishment immediately or maintain uncertainty about the state of detection and impose punishment sufficiently late deter individuals at equal rates. Further, we find that the same institutional settings that are capable of reducing recidivism are also the ones deterring deviant behavior in the first place. Our results yield policy implications for designing effective institutions in mitigating misconduct and reducing recidivism. Work on this paper was shared among the authors as follows: Johannes Buckenmaier 25%, Eugen Dimant 25%, Ann-Christin Posten 25%, Ulrich Schmidt 25%.

Chapter 1
Trader Matching and the Selection of Market Institutions

1.1 Introduction

Market institutions come in many flavors. In many markets, institutions with different characteristics exist, even for the same good. Those can be formal, as e.g. specific Business-to-Business (B2B) or Business-to-Consumer platforms, middlemen agencies, or local markets for perishable products (fish and produce), or informal, as e.g. exchange arrangements, black markets, or the set of particular conventions surrounding real-estate and rental markets in certain countries (group-tenant vs. individual visits). The characteristics of such market institutions in turn influence market outcomes in terms of efficiency, surplus distribution and convergence to market-clearing outcomes. It is hence important to understand what promotes coordination on a specific institution.

In this work, we build upon the evolutionary approach to the selection of trading institutions, and in particular on Alós-Ferrer and Kirchsteiger (2010, 2015) and Alós-Ferrer et al. (2010). The essence of the approach is the study of long-run stability. Suppose a host of alternative institutions are present in a market, whatever their origin might be. Are there any particular institutions whose survival is more likely in the long run? To answer these questions, we analyze the selection and stability of market institutions when boundedly rational traders employ certain "rules of thumb" to decide at which competing institution to trade. The assumption of bounded rationality seems reasonable since, due to the complexity of the evaluation of institutional characteristics, rational learning is rather implausible. Following the evolutionary approach, we will concentrate on the long-run outcomes of the discrete-time, stochastic dynamical system which results when traders revise their institution choices over time on the basis of the behavioral rules.

We use the notion of stochastic stability in a dynamic learning framework with vanishing mistakes (Kandori et al., 1993; Young, 1993; Blume, 1993; Ellison, 2000) to determine which institutions survive in the long run.

For practical purposes, a market institution can be defined as a set of trading rules and conventions which determine the matching and price formation process, i.e. who trades with whom and at what price. As most of the existing literature, Alós-Ferrer and Kirchsteiger (2010, 2015), and Alós-Ferrer et al. (2010) focused on the price-formation part, analyzing market selection when institutions generate possibly biased prices (implying rationing) but feature a single price (per good) only. However, institutions also influence who trades with whom. This paper takes the next natural step and concentrates on the trader matching process. Specifically, we study whether traders learn to coordinate on centralized, market-clearing institutions or whether other institutions can survive in the long run, in a framework where institutions are solely characterized by a matching mechanism. Hence, we allow for violations of the law of one price, that is, we study the stability of general institutions including decentralized non-market clearing ones where a single good might be traded at different prices within a single institution. In order to isolate the effects of matching, however, we concentrate on the effects arising from differences in the matching mechanisms and abstract away from any other complications. In particular, and unlike in the works cited above, institutions will be characterized by market clearing (within each institution), excluding both rationing and price biases. Further, we exclude trader heterogeneity and consider a model with homogeneous buyers and sellers.

We hence identify each institution with a certain matching pattern for the traders who choose to use it. Examples include the "bazaar" where buyers and sellers are randomly matched, auction houses where each good is offered to a subgroup of buyers, and of course centralized markets. Our first result is that centralized institutions are always stable in the long run. This clear-cut result is conceptually in line with the stability of market-clearing institutions in Alós-Ferrer and Kirchsteiger (2010, 2015). It is a rather strong result, because it holds independently of the number and properties of other available institutions, of the characteristics of trader demand and supply, of the behavioral rules within the general class we consider, and of the exact specification of revision opportunities (and hence speed) in the dynamics.

Stochastic stability, however, only means long-run survival, and not necessarily the identification of a unique prediction. It turns out that other decentralized institutions can also survive in the long run. Unlike in the case of a centralized institution, we also show that their survival depends on, e.g., the characteristics of the behavioral rules and the specification of the dynamics. In keeping with our aim for generality, we ask ourselves whether general conditions can be identified without specifying concrete examples of behavioral rules, dynamics, and market characteristics. Our second main result identifies a general necessary condition for stochastic stability, which we term *matching-efficiency*. Informally speaking, a trading institution is matching-efficient if it leaves no unmatched trader when all or almost all traders have already coordinated on it. Although many institutions are matching-efficient, many others, as e.g. a bazaar defined merely by random matching, are not, and hence the condition does have cutting power.

Interestingly, under a strengthening of our assumptions on revision opportunities (requiring the dynamics to be fast enough), matching-efficiency fully characterizes the set of stochastically stable institutions. However, we also show that without this strengthening, matching-efficient institutions can fail to be stochastically stable in general. Hence, the take-home message is that, while full centralization ensures stochastic stability, other institutions might also survive, and a full characterization thereof for specific markets will require active market design, in the sense that institutions will need to be tailored to the specifics of trader behavior and other relevant market characteristics.

The article is structured as follows. Section 1.2 briefly reviews the related literature. Section 1.3 describes the elements of the model, i.e. the characteristics of market institutions, the behavioral assumptions underlying institution choice by (boundedly rational) traders, and the actual (discrete-time, stochastic) learning dynamics. Section 1.4 contains the results, starting with an analysis of the stochastic stability of centralized institutions and proceeding to the conditions under which decentralized institutions might be stochastically stable. Proofs are relegated to Appendix 1.A.

1.2 Related Literature

This article belongs to a line of research started in Alós-Ferrer and Kirchsteiger (2010), which studied the selection of alternative market institutions in a multi-good, general equilibrium setting, and continued in Alós-Ferrer and Kirchsteiger (2015), in a partial equilibrium buyers-sellers model. The main result of those works is that, even if alternative (biased) market institutions exist, market-clearing (unbiased) institutions are always stochastically stable. However, other, alternative institutions might also be stochastically stable and hence survive in the long run, giving rise to a multiplicity of institutions. Which other institutions survive depends on many factors, ranging from the elasticity of individual demands and the heterogeneity of the traders to the speed of the particular dynamics considered. Since this implies that the design of institutions becomes meaningful, Alós-Ferrer et al. (2010) considered fully rational market designers who actively design alternative market platforms, which are then chosen by boundedly rational traders.[1] The present contribution differs from Alós-Ferrer and Kirchsteiger (2010, 2015), and Alós-Ferrer et al. (2010) in that we allow for violations of the law of one price and study the effects of different trader matching within an institution, but we exclude the possibility of price biases and rationing.

The analysis here and in Alós-Ferrer and Kirchsteiger (2010) is also related to the literature on the stability properties of perfectly competitive behavior in learning models with boundedly rational agents. Those works (Alós-Ferrer and Ania, 2005; Mandel and Gintis, 2014) provide a learning-based foundation for perfectly competitive behavior when the market institution is fixed. In contrast, we do not consider the stability of outcomes by themselves, but rather the stability of market institutions which channel those outcomes. Hence, one of our aims is to examine the stability of

[1] Hence that work built a bridge to the "asymmetric rationality program" where rational firms are confronted with boundedly rational consumers (Ellison, 2006; Spiegler, 2006; Gabaix and Laibson, 2006). See also Shi (2015).

"Walrasian," centralized institutions allowing for full market clearing.

Conceptually, the current paper bridges the strand of the literature described above and the literature on evolutionary dynamics and surplus division. The latter asks how cooperative game solutions can be implemented via learning processes, either for bilateral trading (Nax and Pradelski, 2015; Klaus and Newton, 2016) or for general cooperative games modeling surplus sharing among more than two players (Agastya, 1999; Newton, 2012). Our work is complementary to those in that we examine the evolution of institutions as characterized by matching mechanisms, and, hence, the evolution of the matching process itself. However, there are also similarities in the results. The works just mentioned typically select outcomes within the core. Here and in Alós-Ferrer and Kirchsteiger (2010, 2015), we obtain that states where all traders coordinate on a single centralized institution are stable, a result which can be related to the core as no coalition of traders could improve (in the sense of increasing overall efficiency) by moving away.

1.3 The Model

There is a single homogeneous good to be traded by a finite population of traders consisting of buyers and sellers. We consider a buyers-sellers model with a fixed set B of $n \geq 2$ homogeneous buyers and a fixed set S of $m \geq 2$ homogeneous sellers, where traders' roles are fixed and predetermined. We view this setup as reasonably general while ensuring tractability.

Our model has three components, which will be discussed in three separate subsections below. First, we need to specify the characteristics of market institutions and how trade is conducted within an institution. Second, we will detail the behavioral assumptions underlying institution choice by (boundedly rational) traders. Third, we will describe the actual (discrete-time, stochastic) learning dynamics.

1.3.1 Matching and Institutions

Buyers and sellers can trade the good at different market institutions. We assume that there is a set of $N + 1$ different institutions $Z = \{z_0, \ldots, z_N\}$, and traders can choose at which institution they want to trade. For our purposes, the important part of a market institution is how the matching process is structured within it, and how the trading prices are determined; in other words, who can trade with whom and at what price. That is, we identify an institution with a trading rule that specifies the matching and price formation process. When modeling the matching process we rely on the following notion of a matching.

Definition 1. A **matching** for two (possibly empty) sets X and Y is

- a partition of X, $\{X_0, X_1, \ldots, X_\ell\}$, and
- a partition of Y, $\{Y_0, Y_1, \ldots, Y_\ell\}$,

such that $X_i \neq \emptyset \neq Y_i$ for all $i = 1, \ldots, \ell$. A matching is *non-trivial* if $\ell \geq 1$, or, equivalently, $X_0 \subsetneq X$ and $Y_0 \subsetneq Y$.

The interpretation is as follows. Given a set of buyers X and a set of sellers Y, all of them present at the same institution, a matching partitions all traders into matching groups or sub-markets (X_i, Y_i) for $i = 1, \ldots, \ell$, while possibly leaving a subset of buyers X_0 and a subset of sellers Y_0 unmatched. Unmatched traders do not trade at all. Buyers in X_i can potentially trade with sellers in Y_i, and vice versa, at a price to be determined by a specific price formation process that depends solely on X_i and Y_i. That is, within each matching group there will be a unique trading price, but the prices within a single institution can differ across matching groups.

The simplest market institutions could now be defined by assigning a fixed matching to each potential pair of sets (buyers and sellers). Market institutions, however, are rarely fully deterministic. Hence, an institution will rather be defined by a distribution over potential matchings, together with a specific price formation process that determines prices within matching groups. We formally define an institution as follows.

Definition 2. Given a set of buyers B and a set of sellers S, an **institution** z is characterized by a *matching function* M_z which, for any two subsets $B_z \subseteq B$ and $S_z \subseteq S$ specifies a probability distribution $M_z(B_z, S_z)$ over all matchings for B_z and S_z. An institution is *non-trivial* if *supp* $M_z(B_z, S_z)$ contains at least one non-trivial matching whenever $B_z \neq \emptyset$ and $S_z \neq \emptyset$.

For given subsets $B_z \subseteq B$ and $S_z \subseteq S$ of traders and sellers at an institution z, a realization of the distribution $M_z(B_z, S_z)$ is hence a matching $(B_i^z, S_i^z)_{i=0}^{\ell_z}$ in the support of $M_z(B_z, S_z)$.

We endow buyers with a common demand function d and sellers with a common supply function s. Demand and supply functions satisfy standard properties as captured by the following assumptions.[2]

M1. The demand function $d : \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \cup \{\infty\}$ is continuous and strictly decreasing in p, with $d(p) > 0$ for all $p \geq 0$ and $\lim_{p \longrightarrow \infty} d(p) = 0$.

M2. The supply function $s : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is continuous and (weakly) increasing in p, with $s(0) = 0$ and $s(p) > 0$ for all $p > 0$.

For any realized matching $(B_i^z, S_i^z)_{i=0}^{\ell_z}$, prices $p_1^z, \ldots, p_{l_z}^z$ at z are determined by local market clearing, i.e.

$$|B_i^z|\, d(p_i^z) = |S_i^z|\, s(p_i^z) \text{ for } i = 1, \ldots, \ell_z. \qquad \text{(LMC)}$$

Since all buyers are identical (and characterized by the demand function d) and all sellers are identical (and characterized by the supply function s), given the matching, condition (LMC) is enough to describe the results of trading. Under **M1** and **M2** there always exists a unique, strictly positive price p_i^z solving (LMC) for $i = 1, \ldots, \ell_z$, which yields demand and supply strictly above zero. Further, the price p_i^z only depends on the buyer-seller ratio $r_i^z = |B_i^z|/|S_i^z|$ within the respective matching group (B_i^z, S_i^z), and of course on the shape of the supply and demand functions, which we assume to be fixed. Given a ratio r, we denote the corresponding price by $p(r)$. Note also that $p(r)$ is strictly increasing in r.

Assume buyers $B_z \subseteq B$ and sellers $S_z \subseteq S$ want to trade at institution z. The matching function M_z determines all the matchings which occur with positive probability for B_z and S_z, namely the support of $M_z(B_z, S_z)$ (the

[2] **M1** and **M2** correspond to M1' and M2' in Alós-Ferrer and Kirchsteiger (2015).

exact probability is not important for our results). After matching groups have been realized trade takes place in the matching groups at the price specified by (LMC) at group level. Note that under (LMC) traders are not rationed.[3]

We will assume that $z_0 \in Z$ is always a centralized market-clearing institution as follows, i.e., an institution of this type is always available. This is an institution where all traders are matched in a single group (no trader remains unmatched) and hence trade at the same price. The (single) price at the centralized institution depends only on the number of buyers and sellers that wish to trade there.

Example 1. A **centralized institution**, denoted z_0, matches traders according to a fixed matching with $\ell_0 = 1$, $B_0^{z_0} = S_0^{z_0} = \emptyset$, $B_1^{z_0} = B_{z_0}$ and $S_1^{z_0} = S_{z_0}$ (whenever $B_{z_0} \neq \emptyset \neq S_{z_0}$). Thus z_0 always features a single price $p^0 = p(r_0)$ with $r_0 = |B_{z_0}|/|S_{z_0}|$.

Our definition of market institution, however, is rather general. It includes "classical," centralized market-clearing institutions as above, but also many others. The following is an example of a different institution that leaves some traders unmatched but still features a unique trading price.

Example 2. A **bazaar** is any institution z_B that matches buyers and sellers in pairs and leaves the remaining traders unmatched.

If the number of buyers and sellers at a bazaar are not identical, then this institution leaves some traders unmatched. In this symmetric setting, if trade occurs it does so in many sub-markets (each consisting of one buyer and one seller) but at the same single price given by $p(1)$. There are, however, also institutions that violate the law of one price, as the following, rather stylized, example shows.

Example 3. A **double one-to-many institution** is an institution z_D that always selects a single buyer and a single seller (provided at least two traders

[3]For example, Alós-Ferrer and Kirchsteiger (2015) study institutions characterized by a rationing parameter.

of each type are present) and matches all remaining sellers to the singled-out buyer and all remaining buyers to the singled-out seller. Hence the institution selects both a random monopolist and a random monopsonist. If there is only a single buyer (seller), it matches all sellers (buyers) to that single buyer (seller). This institution features at most two matching groups and hence at most two prices. Denote them by $\underline{p}_D \leq \overline{p}_D$. If there are at least 3 sellers and at least 3 buyers at z_D, then we have $\underline{p}_D < p(1) < \overline{p}_D$ and the institution fails the law of one price.

The double one-to-many institution is, in a sense, as far away from a centralized market-clearing institution as possible. Most of the time z_D will feature two prices, thus violating the law of one price. In sharp contrast to centralization, the spread between the realized prices will often be large.

1.3.2 Behavioral Assumptions

Traders select an institution where they want to trade from the set of feasible institutions $Z = \{z_0, z_1, \ldots, z_N\}$. The state of the learning dynamics is completely determined by the choices of traders. Recall that there are n buyers and m sellers, therefore the state space Ω is given by Z^{n+m}. We denote by $\omega(k) \in Z$ the institution chosen by trader k in state $\omega \in \Omega$. For a state $\omega \in \Omega$, we write $B_z(\omega) = \{i \in B \mid \omega(i) = z\}$ and $S_z(\omega) = \{j \in S \mid \omega(j) = z\}$ for the sets of buyers and sellers currently at z, respectively. The number of buyers and sellers at z is given by $n_z(\omega) = |B_z(\omega)|$ and $m_z(\omega) = |S_z(\omega)|$, respectively. We will also denote $n_z(\omega)$, $m_z(\omega)$ by n_z, m_z if no confusion can arise.

Given ω, the sets $B_z(\omega)$ and $S_z(\omega)$, $z \in Z$, determine the distribution of traders among the available institutions. For each institution z, a potential matching realization is an element $\gamma_z = (B_i^z, S_i^z)_{i=0}^{\ell_z} \in \mathit{supp}\, M_z(B_z, S_z)$. Let

$$\Gamma(\omega) = \{(\gamma_z)_{z \in Z} \mid \gamma_z \in \mathit{supp}\, M_z(B_z(\omega), S_z(\omega))\ \forall z \in Z\}$$

be the set of vectors of potential realizations with typical element $\gamma \in \Gamma(\omega)$. We then call $\overline{\Omega} = \{(\omega, \gamma) \mid \gamma \in \Gamma(\omega)\}$ the set of (potential) **state-matching pairs**.

For a given state-matching pair (ω, γ) we say that an institution z is **active** if γ_z is non-trivial, and **inactive** otherwise. Denote by $A(\omega, \gamma)$ the set of active institutions at (ω, γ). Further, for each institution $z \in A(\omega, \gamma)$, let $T_z(\omega, \gamma) = \{p_1^z, \ldots, p_{\ell_z}^z\}$ be the set of realized prices at z for (ω, γ).

We assume that traders observe the prices at all institutions. From the point of view of an individual trader k, the relevant market outcome at (ω, γ) is given by the pair $(p(k), q(k))$ containing the price at which he trades and the quantity he can trade. In our setting, however, if a trader trades at price p he can trade exactly the quantity he desires (that is, either $d(p)$ or $s(p)$, depending on his role). Thus, traders are never rationed. Hence, it is reasonable to base traders' behavior on the observation of prices only, since demand and supply given a price will always be fulfilled. Note that prices at a given institution are directly linked to the ratio of buyers to sellers within the institution's sub-markets through the local market-clearing condition (LMC).

Specifically, traders' behavior in our model is based on the following two main (and minimalistic) assumptions. First, traders prefer trade over no trade. Second, buyers prefer lower prices and sellers prefer higher prices. Of course, these assumptions could be obtained from first principles by postulating appropriate utility and profit functions compatible with the supply and demand functions, or alternatively by deriving decisions from consumer and producer surplus. We follow here Alós-Ferrer and Kirchsteiger (2010, 2015) and Alós-Ferrer et al. (2010) and base our behavioral model on these properties only.

We assume that agents look at observed, actually realized outcomes to (myopically) select an institution in the subsequent period. Traders' behavior is captured by discrete-time, stochastic **behavioral rules**, which are mappings specifying the probability of choosing each available institution given the previous market outcome. We will keep the approach as general as possible. In particular, the exact choice probabilities will not be important; rather, the key property of a behavioral rule will be which institutions can be selected with positive probability. Hence, it will be enough for our analysis to specify (families of) behavioral rules through the set $\mathcal{S}_k(\omega, \gamma) \subseteq Z$ of institutions which can be chosen with positive probability in the next period if the current state-matching pair is (ω, γ). That is, if (ω, γ) occurs at time t, trader k will choose some institution in $\mathcal{S}_k(\omega, \gamma)$ in $t+1$, each one with positive probability. A similar approach was adopted in Alós-Ferrer and Weidenholzer (2014).

An example of a behavioral rule is the Imitate-the-Best-Max (IBM) rule used in Alós-Ferrer and Kirchsteiger (2015). Essentially, for a trader k this rule is specified by defining $\mathcal{S}_k(\omega, \gamma)$ to be the set of active institutions that are evaluated best according to some evaluation function reflecting the behavioral fundamentals of the model (e.g., an indirect utility function). While rules of this type will be an example allowed in our analysis, we adopt a more general approach.

Our specification through the sets $\mathcal{S}_k(\omega, \gamma)$ already focuses on families of behavioral rules, since the exact choice probabilities might vary from rule to rule without affecting our results. We allow for even larger classes of behavioral rules and use an "axiomatic" approach. In other words, rather than focusing on a specific behavioral rule, we allow traders to use any behavioral rule satisfying two general assumptions.

Our first behavioral assumption is that traders prefer trade over no trade, hence traders never switch to inactive institutions if alternative active institutions are available. Recall that $A(\omega, \gamma)$ denotes the set of active institutions for a state-matching pair (ω, γ).

ACT. Consider an arbitrary trader k and a state-matching pair (ω, γ). If k is matched, or if k is unmatched but there is an active institution $z \neq \omega(k)$, then $\mathcal{S}_k(\omega, \gamma) \subseteq A(\omega, \gamma)$.

Matched traders are those currently active, i.e. matched and hence allowed to trade. In our setting, matched traders trade positive quantities, hence a switch to an inactive institution will never be beneficial for the trader (at least from his myopic perspective). Under **ACT** traders never switch to institutions that are inactive at the current state-matching pair as long as there is at least one alternative active institution available. For unmatched traders, **ACT** only applies if there is actually some active institution other than the one the trader is currently at. The reason is that, if a trader is unmatched but all other institutions are inactive, there is no clear (myopic) advantage to staying in the only active institution for the trader, and requiring to stay in the current institution would be unnecessarily restrictive.

For each trader k denote the price at which he trades by $p(k)$ (if he does trade); we abuse notation here by dropping the obvious dependence on (ω, γ). In the absence of rationing, any sensible model based on first principles will lead to the conclusion that buyers prefer low prices and sellers prefer high prices. Following this logic an institution z is "attractive" for a trader k if all the prices realized at z are weakly better for k than $p(k)$, the price at which he currently trades. In this case, a myopic trader will expect not to be worse off at z.

Definition 3. Consider an arbitrary trader k and an institution $z \neq \omega(k)$ with set of realized prices T_z at (ω, γ).

- For a matched trader k trading at price $p(k)$, we say z is **attractive for** k at (ω, γ) if $T_z \neq \emptyset$ and all $p \in T_z$ are weakly better than $p(k)$ for k (that is, $p(k) \leq p \; \forall \, p \in T_z$ if k is a seller, $p(k) \geq p \; \forall \, p \in T_z$ if k is a buyer).

- For an unmatched trader k, we say that z is **attractive for trader** k at (ω, γ) if z is active.

Our second main behavioral assumption states that if an institution z is attractive to some trader in the sense described above this trader will (at least with some positive probability) leave his current institution, for instance (but not necessarily) towards z.[4]

SELF. Consider an arbitrary trader k and a state-matching pair (ω, γ). If there exists an institution other than $\omega(k)$ that is attractive for k, then with positive probability k switches to some active institution $z \neq \omega(k)$, that is, $(\mathcal{S}_k(\omega, \gamma) \setminus \{\omega(k)\}) \cap A(\omega, \gamma) \neq \emptyset$. If k is unmatched and all

[4]Requiring the trader to move to z with positive probability whenever z is attractive would be unnecessarily restrictive. For instance, it would exclude all rules of the Imitate-the-Best-Max type, where the best institution according to some criterion is the only one selected even if there are several attractive ones.

institutions other than $\omega(k)$ are inactive, then $\mathcal{S}_k(\omega, \gamma) = Z$, i.e. k has positive probability of switching to every institution.

On the one hand, **SELF** is quite conservative in evaluating other institutions because it just requires a trader to leave his current institution with positive probability only if there exists another institution where all prices are "better" for the trader. It is, however, less conservative regarding the evaluation of the current institution because only the price at which the trader is actually trading is used for comparison purposes (in the sense of Definition 3). That is, the condition incorporates a form of **self bias** because the trader's actual decision might occasionally be triggered by the comparison of his own price (outcome) with the outcomes at another institution, neglecting outcomes of other traders at his own institution. Note, however, that **SELF** is just a sufficient condition for switching institutions with positive probability, but not a necessary one. That is, it only requires that traders do not stay with probability one if an attractive institution exists, but not that they always switch. The probability with which they leave can be very small. That is, **SELF** still allows for rationalistic rules where traders take all prices into account and perform complex computations to determine their next move, but with some small probability bolt into action if they see that their own price at the current institution is unsatisfactory.

For example, the following behavioral rule would fulfill both **ACT** and **SELF**. Compute the average price at each active institution, taking all prices into account, and move to the one with the best average (largest for sellers, smallest for buyers). In case the current institution is the best according to this criterion, but there exists an attractive institution (in the sense of Definition 3), switch there with a fixed (small) probability δ. If k is unmatched and all institutions other than $\omega(k)$ are inactive, then randomize uniformly among all institutions.

1.3.3 Learning Dynamics

We study a dynamic learning model where traders interact repeatedly in discrete time $t = 0, 1, 2, \ldots$. In each period traders observe all prices realized at all institutions in the previous period. Based on this information each trader k chooses an institution to trade at according to a behavioral rule $\mathcal{S}_k$ (possibly different across traders). Following traders' institutional choices, matchings and prices are determined at all institutions and demand and supply are realized. At the end of each period the proceeds of trade are consumed. Next period, demand and supply functions are reset, i.e. the game is played recurrently.

Agents are potentially allowed to revise their decisions (choice of institution) in any period, but might uphold their decisions (inertia) in some. The specification of how and when revision opportunities arise is an integral part of evolutionary dynamics. Results that are fragile with respect to minor variations regarding the specifications of the revision process could be criticized as lacking robustness (see, e.g., Alós-Ferrer and Netzer, 2015). We therefore refrain from imposing a specific form of how revision opportunities arise, but rather consider a general class of random revision processes that satisfy certain "minimal" assumptions, following Alós-Ferrer and Kirchsteiger (2015). Denote by $E(k,\omega)$ the event that agent k receives revision opportunity in state ω, and by $E^*(k,\omega)$ the event that agent k is the only agent of his type with revision opportunity at ω. We allow for any specification of revision opportunities that satisfy the following two assumptions. For every trader k and state ω,

D1. $Pr(E^*(k,\omega)) > 0$.

D2. either $Pr(E^*(k,\omega) \bigcap E^*(k',\omega)) > 0$ for any trader k' of the other type, or $Pr(E^*(k,\omega) \bigcap E(k',\omega)) = 0$ for all such k'.

The first condition ensures that in any state any trader has positive probability of being able to revise. Further, it requires that there is always a small probability that only one trader is allowed to revise. The second condition implies some form of independence of revision opportunities between buyers and sellers. Specifically, the assumption implies that there is no correlation in the presence of revision opportunities as there would be if, e.g., a pair formed by one buyer and one seller would always receive them together.

One could also consider stronger conditions on the dynamics, at the expense of generality. In particular, consider the following additional assumption on revision opportunities.

D3. $Pr\left(\bigcap_{k\in B} E(k,\omega)\right) > 0$ and $Pr\left(\bigcap_{k\in S} E(k,\omega)\right) > 0$.

Condition **D3** requires that there is always some positive probability that all buyers (respectively all sellers) revise simultaneously. Intuitively, this makes the dynamics relatively quick, since a whole market side might switch in a single period. Dynamics with independent inertia (meaning that in any period every agent has a positive, independent probability of not being able to adjust) satisfy **D3**, but dynamics with asynchronous learning (meaning that each period a single agent is randomly chosen and only that agent is allowed to revise) not. We will not assume **D3**, but we will return to this condition later for a particular result.

1.4 Analysis

1.4.1 Absorbing States

We consider a family of learning processes satisfying the assumptions laid out above, that is, a behavioral rule satisfying **ACT** and **SELF** together with a revision process satisfying **D1** and **D2**. Given two states $\omega, \omega' \in \Omega$, denote by $P^0(\omega, \omega')$ the probability of transition from ω to ω' in one period for a fixed learning process, which we will refer to as the **unperturbed dynamics**.[5] The transition matrix is given by $P^0 = [P^0(\omega, \omega')]_{\omega,\omega'\in\Omega}$. An *absorbing set* of the unperturbed dynamics is a minimal subset of states which, once entered, is never abandoned. An *absorbing state* is an element which forms a singleton absorbing set, i.e. $P^0(\omega, \omega) = 1$.

We first introduce some terminology. A state ω determines the sets of buyers B_z and sellers S_z at each institution $z \in Z$. Thus $(M_z(S_z, B_z))_{z\in Z}$ induces a probability distribution over vectors of realizations $\Gamma(\omega)$. In what

[5] Of course the actual transition probabilities depend on the specific behavioral rule, but in what follows we drop this dependence to increase readability.

follows we adopt the notation $\Pr(B_0^z = \emptyset \mid \omega)$ to denote the probability (conditional on the state ω) of the set of matchings of the form $\gamma = (\gamma_z)_{z \in Z}$ with $\gamma_z = (B_i^z, S_i^z)_{i=0}^{\ell_z}$ such that $B_0^z = \emptyset$. Other expressions of this type are defined analogously. Last, we say an institution z **matches all traders at** ω if $\Pr(B_0^z \cup S_0^z = \emptyset \mid \omega) = 1$, i.e. there is no state-matching pair (ω, γ) with unmatched traders at z (analogously, we will also speak of institutions "matching all buyers" or "all sellers").

A particular class of states will be of specific interest. Given an arbitrary institution z, the **monomorphic state** ω_z is the state where $B_z = B$ and $S_z = S$, i.e. all traders are at z. Monomorphic states are absorbing states, provided the institution manages not to leave traders unmatched when all traders choose that institution. The reason is simple. Since all traders are at z, all other institutions are inactive. If all traders at z are matched, by **ACT** they will stay at z.

Lemma 1. *Let Z be an arbitrary set of institutions. Assume* **ACT**. *For every institution z that matches all traders at ω_z, the monomorphic state ω_z is an absorbing state.*

This result already indicates that in general there will be a multiplicity of absorbing states, at least one per each institution which avoids unmatched traders in case of full coordination. Additionally, in general there might be non-singleton absorbing sets. Absorbing sets and states, however, are just an intermediate and not always necessary step of the analysis, as we are interested in the long-run stability of outcomes. To study the latter, it is not always necessary to characterize the former, especially if techniques along the lines of Ellison (2000) are used.

1.4.2 Stochastic Stability

Our analysis of the learning dynamics follows a stochastic stability approach using methods and concepts introduced by Kandori et al. (1993) and Young (1993). Detailed overviews can be found, e.g., in Samuelson (1997), Fudenberg and Levine (1998), Young (1998), and Sandholm (2010). In our context, we aim to analyze the stability of situations where traders coordinate on particular trading institutions. To this purpose, the dynamics is enriched with a perturbation in the form of experiments (or mistakes) in the following way. With an independent, small probability $\varepsilon > 0$, each agent, in each period, might discard the prescriptions of his behavioral rule and experiment (or make a mistake, or "mutate"). In that case, the trader simply picks an institution at random, independently of other considerations, with all institutions having positive probability.

The dynamics with experimentation is called the **perturbed dynamics**. Its transition matrix is denoted by P^ε. Since experiments make transitions between any two states possible, the perturbed process has a single absorbing set formed by the whole state space (i.e., the process is irreducible). There is a unique probability distribution over states $\mu_\varepsilon \in \Delta(\Omega)$ which, if taken as initial condition, would be reproduced in probabilistic terms after updating (more precisely, $\mu_\varepsilon P^\varepsilon = \mu_\varepsilon$). This μ_ε is called the **invariant distribution** of P^ε. For the perturbed dynamics P^ε the **limit invariant distribution** $\mu^* = \lim_{\varepsilon \to 0} \mu_\varepsilon$ exists and is an invariant distribution of the unperturbed dynamics P^0 (see e.g. Kandori et al., 1993; Young, 1993; Ellison, 2000).

The states in the support of μ^*, i.e. $\{\omega \in \Omega \mid \mu^*(\omega) > 0\}$ are the **stochastically stable states** or **long-run equilibria**. Standard results (see e.g. Ellison, 2000, Theorem 1) then imply that the set of stochastically stable states is a union of some absorbing sets of the original, unperturbed chain ($\varepsilon = 0$). In other words, stochastic stability selects among the absorbing sets of the unperturbed dynamics.[6]

To simplify terminology, we will say that a institution z is **stochastically stable** if ω_z is stochastically stable. Note, however, that the set of stochastically stable states might not be a singleton, for example if several market institutions are stochastically stable.

[6] In the following, whenever we say absorbing sets or states, we refer to the unperturbed dynamics. Since the perturbed dynamics is irreducible, no confusion should arise.

1.4.3 The Stability of Centralized Institutions

The centralized, market-clearing institution z_0 always matches all traders and features a unique price (recall Example 1). Our first result is that this institution is always stochastically stable, independently of how many and which other institutions are also available.

The proof of this result proceeds in two steps. Let ω_0 be the monomorphic state with full coordination at z_0. First, the following lemma, whose proof is in Appendix 1.A, shows that from any state (monomorphic or not) where z_0 is active, there is a dynamic pressure towards coordination on this institution. This results in a positive probability path towards the monomorphic state where all agents are at z_0.

Lemma 2. *Let Z be an arbitrary set of institutions with $z_0 \in Z$. Under* **D1**, **D2**, **ACT**, *and* **SELF**, *for any state-matching pair (ω, γ) where z_0 is active, there is a positive-probability path (of the unperturbed dynamics) leading to ω_0.*

Intuitively, Lemma 2 makes use of the fact that z_0 satisfies the law of one price featuring always a single price p only. For a given alternative institution z, p is then either larger than all prices at z, making z_0 attractive for all sellers, or smaller than all prices at z, making z_0 attractive for all buyers, or there exist prices at this institution that are both larger and smaller than p, making z_0 attractive for at least one buyer and one seller. This property can then be used iteratively to construct a path towards full-coordination on z_0 as stated in Lemma 2.

The last step relies on standard results from the stochastic stability literature (see Appendix 1.A). Essentially, the intuition is as follows. From any state, a few traders experimenting with z_0 suffice to make this institution active. In view of Lemma 2, it is hence easy to construct a path towards ω_0. This strong property allows us to complete the analysis without needing, for instance, a full characterization of the absorbing sets of the unperturbed dynamics, because they can all be easily destabilized independently of their particular characteristics (except for the singleton $\{\omega_0\}$). However, in order to destabilize ω_0 (which is absorbing by Lemma 1), it is necessary to have a large number of traders experimenting away, so that z_0 becomes inactive; else, by Lemma 2 again, the dynamics will lead back to ω_0 with positive probability and without further experiments.

Theorem 1. *Let Z be an arbitrary set of institutions with $z_0 \in Z$. Under* **D1***,* **D2***,* **ACT***, and* **SELF***, the centralized institution z_0 is stochastically stable.*

We conclude that centralized market clearing displays rather strong stability properties, in the sense that such institutions will survive in the long run independently of what other trading coordination opportunities are available. Indeed, Theorem 1 holds independently of how many other institutions are available and what their characteristics are. Further, the result holds independently of trader characteristics, as captured by demand and supply functions, and independently of the exact details of the behavioral rules and the specification of the dynamics as long as our basic conditions hold.

Note that there might very well be several centralized institutions available, and by Theorem 1 any such institution is then stochastically stable. In general, other institutions might also be stable, and their stability might depend on trader characteristics and the specification of the dynamics. The following subsections illustrate which and when other institutions are and are not stochastically stable.

1.4.4 Matching-Efficient Institutions

In the next step we seek to establish that many institutions are always stochastically *unstable* in our setting. In fact, a necessary condition for stochastic stability of an institution is that at or near full coordination on that institution, there should be no unmatched traders. This is captured by the following concept (recall that the number of buyers and sellers at z is given by $n_z(\omega) = |B_z(\omega)|$ and $m_z(\omega) = |S_z(\omega)|$, respectively).

Definition 4. An institution z is **matching-efficient** if the following three conditions hold.

- z matches all traders at ω_z;

- at every state ω with $(n_z(\omega), m_z(\omega)) = (n, m-1)$, z matches all buyers with probability one; and
- at every state ω with $(n_z(\omega), m_z(\omega)) = (n-1, m)$, z matches all sellers with probability one.

Proposition 1. *Let Z be an arbitrary set of institutions with $z_0 \in Z$. Assume* **D1**, **D2**, **ACT**, *and* **SELF**. *If an institution $z \in Z$ is stochastically stable, then it must be matching-efficient.*

To illustrate the intuition behind Proposition 1, consider an institution z that does not match all traders at ω_z, i.e. the first condition in Definition 4 fails. There is at least one unmatched trader $k \in S_z \cup B_z$ at ω_z. As all other institutions are inactive, by **SELF** there is positive probability that k switches to any other institution, for instance to z_0. We thus reach a state where a trader is at z_0. From this state, a single mutation (a trader of the other type switching to z_0) is enough for z_0 to become active. By Lemma 2, the process then drifts to ω_0 with positive probability. However, reaching ω_z from ω_0 requires destabilizing ω_0, which as commented above is not easy. It follows that ω_z is easier to leave than to reach. Standard results in stochastic stability (Ellison, 2000) are then enough to establish that ω_z is not stochastically stable.

Proposition 1 states that matching all traders at states where "almost all" traders are at a given institution is a necessary condition for stochastic stability. As a consequence, many institutions are unstable. The following example shows that the bazaar is one of them.

Example 4. Consider the bazaar of Example 2, z_B, and let ω_B be the corresponding monomorphic state. If $n \neq m$, there must be an unmatched trader at ω_B. If $n = m$, then there is no unmatched trader at ω_B, but at any state with $(n_z, m_z) = (n, m-1)$ a buyer is unmatched and at any state with $(n_z, m_z) = (n-1, m)$ a seller is unmatched. Therefore, the bazaar is not matching-efficient and can never be stochastically stable.

This is natural in our setting, because since we concentrate exclusively on matchings and not, say, price biases, the bazaar is just the institution which matches traders in pairs, all of them obtaining the same price. This aspect of the bazaar (matching groups as pairs) leads to instability. Actually, the argument above can be generalized to show that no institution with fixed group size can be stochastically stable.

Strikingly, under the additional assumption **D3** on revision opportunities, we obtain a full characterization of stochastically stable, non-trivial institutions, because the necessary condition to be matching-efficient becomes also sufficient.

Theorem 2. *Let Z be an arbitrary set of institutions with $z_0 \in Z$. Assume* **D1–D3***,* **ACT***, and* **SELF***. A non-trivial institution $z \in Z$ is stochastically stable if and only if it is matching-efficient.*

Since matching-efficiency is a relatively weak condition, Theorem 2 reveals that a large class of institutions is stochastically stable for quick dynamics. This should not be overinterpreted, though. The results of Theorem 2 depend heavily on the speed of the dynamics since a quick enough dynamics allows the process to "jump over" unstable states. As we will show below, this result fails for slower dynamics.

Further, as already commented above, matching-efficiency does exclude a relatively large number of potential institutions. For instance, consider one-sided institutions which try to implement price biases in favor of only one side. The only way to implement such an outcome is to leave traders unmatched, in particular in the case of full coordination.

1.4.5 Optimism, Pessimism, and Decentralized Institutions

To better understand the scope of the results, it is worth briefly exploring the behavioral assumptions, and in particular **SELF**. Remember that this assumption states that, if all the prices realized at an institution z are weakly better for a trader k than the price $p(k)$ at which that trader is currently trading, there is some positive probability that k leaves his current institution to some active one, for instance to z. One natural possibility yielding alternative assumptions is to capture more optimistic or pessimistic behavior. Consider the following possibility.

OPT. Consider an arbitrary trader k and a state-matching pair (ω, γ). If k is matched and there is an institution $z \neq \omega(k)$ that features some price that is weakly better than $p(k)$ (larger for $k \in S$, smaller for $k \in B$), then there is positive probability that k switches to some active institution other than $\omega(k)$. If k is unmatched, then k has positive probability of switching to every institution.

Under this (extreme) assumption, a trader can be interpreted as being optimistic because he focuses on the better prices at z, ignoring the ones worse than $p(k)$. For instance, if he actually switches to z, one interpretation is that after switching the trader believes that he will be able to achieve the best observed outcome at the new institution even if, in the previous period, it was only obtained by some traders there. Obviously, **OPT** implies **SELF** and the results derived above hold. In particular, in the presence of **D3**, the characterization of stochastically stable institutions identified in Proposition 2 remains unchanged.

For general dynamics (in the absence of **D3**), Theorem 1 shows that, given **ACT** and **SELF**, a centralized institution is stochastically stable independently of the details of the dynamics, of which other institutions are available and of what their specific characteristics are. If **SELF** is strengthened to **OPT**, it can be shown that all other matching-efficient but decentralized institutions are stochastically stable for all possible dynamics and alternative institutions. That is, a characterization as that in Theorem 2 holds.

Proposition 2. *Let Z be an arbitrary set of institutions with $z_0 \in Z$. Under* **D1***,* **D2***,* **ACT***, and* **OPT** *a non-trivial institution $z \in Z$ is stochastically stable if and only if it is matching-efficient.*

To gain some quick intuition, consider for example the double one-to-many institution z_D (recall Example 3). This institution usually features two prices which are highly asymmetric (a monopolistic price and a monopsonistic one). Hence, z_D is rarely attractive for conservative traders as they require both prices to be better than the one they trade at. However, for optimistic traders this asymmetry makes z_D always appealing for at least one market side. Thus **OPT** facilitates a transition towards the double one-to-many institution rendering it a long-run equilibrium for general learning dynamics.

The result above is instructive for the general research agenda, since it does not rely on restrictions on the dynamics but rather concentrates on a subclass of behavioral rules. For instance, as long as behavioral rules fulfilling **OPT** are considered reasonable, Proposition 2 shows that there is no reasonable strengthening of the current assumptions which would render centralized institutions *uniquely* stochastically stable.

Results as Proposition 2 are of course less robust than Theorem 1, as they hinge on more restrictive assumptions. One can conceive other behavioral rules or dynamics for which the alternative institutions fail to be stable in the long run. Consider, for instance, a pessimistic behavioral rule as follows.

PES. Consider an arbitrary trader k and a state-matching pair (ω, γ). If there exists an institution other than $\omega(k)$ that is attractive for k, then k switches to an institution $z \neq \omega(k)$ with positive probability if and only if z is attractive for k. If k is unmatched and all institutions other than $\omega(k)$ are inactive, then k has positive probability of switching to every institution.

Obviously, if a behavioral rule satisfies **PES**, it also fulfills **SELF**, but it must violate **OPT** (note that **PES** also implies **ACT**). A trader fulfilling this assumption can be interpreted as being overly cautious, since he will never switch to an institution where some realized price is worse than the one he is currently trading at. Since the double one-to-many institution is matching-efficient it follows from Theorem 2 that it is also stochastically stable under **PES**, provided the dynamics fulfills **D3**. However, under slower dynamics, this result is not true any more.

Proposition 3. *Let $Z = \{z_0, z_D\}$ and $n, m > 3$. Assume* **PES**. *Under asynchronous learning, the double one-to-many institution z_D is not stochastically stable.*

In general, the intuition is that for slow dynamics and pessimistic (or cautious) behavioral rules, the attractiveness of the double one-to-many institution vanishes and this institution fails to be stochastically stable. Again, this result is instructive. While for fast dynamics (fulfilling **D3**) a full, simple characterization of the class of stochastically stable institutions is feasible (as given by Theorem 2), the results above prove that for general dynamics such a simple characterization is impossible. Institutions as the double one-to-many example are stochastically stable for all dynamics and certain types of behavioral rules fulfilling **SELF**, but stop being stochastically stable for the same dynamics and other types of behavioral rules which, however, do fulfill **SELF**. Hence, there simply exists no characterization in the absence of assumptions on the speed of the dynamics beyond **D1–D2** and in the absence of stronger assumptions on the behavioral rules.

1.5 Conclusion

This contribution is a parsimonious step in the study of the selection of market institutions by boundedly rational traders. Our results have been obtained in a setting which is as general as possible in some dimensions (dynamics, trader behavior) but remains necessarily stylized in others. Accordingly, they pave the way for a number of possible extensions which are currently in our research agenda. First, the basic result can be used to study market design under asymmetric rationality as in Alós-Ferrer et al. (2010). Second, combining the results here with Alós-Ferrer and Kirchsteiger (2015) should allow to study more realistic institutions which combine restrictions on trader matching and price biases (rationing). Third, trader heterogeneity and multiple goods can be incorporated, either in buyer-seller models or in general equilibrium settings along the lines of Alós-Ferrer and Kirchsteiger (2010).

Appendix 1.A: Proofs

We start with some preliminary results. Given two absorbing sets A and B, denote by $c(A, B)$ the minimal number of mistakes required for a transition from A to B, called the transition cost from A to B. Note that any transition along a path that has positive probability under the unperturbed dynamics has a cost of zero; we refer to such paths as positive-probability paths. To show stochastic stability we use the following result, which is a straightforward adaptation of results in Ellison (2000, Theorem 3) (see also Alós-Ferrer and Kirchsteiger, 2010, Lemma 2).[7]

Lemma 3. *Let A be an absorbing set and define the Radius of A by*

$$R(A) = \min\{c(A, B) \mid B \text{ is an absorbing set }, B \neq A\}$$

and the Coradius of A by

$$CR(A) = \max\{c(B, A) \mid B \text{ is an absorbing set}, B \neq A\}$$

Then

(a) If $R(A) \geq CR(A)$, the states in A are stochastically stable.

(b) If $R(A) > CR(A)$, the only stochastically stable states are those in A.

(c) If the states in an absorbing set B are stochastically stable and $R(A) = c(B, A)$, the states in A are also stochastically stable.

(d) If B is an absorbing set with $c(B, A) < R(A)$, then B is not stochastically stable.

We say that an institution z is a *single-price institution* if the set of realized prices T_z is a singleton at any (ω, γ) such that z is active. We now prove a preliminary result for single-price institutions. Lemma 4 shows that if active, a single-price institution is always attractive for at least some traders.[8]

Lemma 4. *Consider a state-matching pair (ω, γ) where both a single-price institution z and another institution $z' \neq z$ are active. Then (at least) one of the following cases holds.*

[7] Ellison (2000) credits Evans (1993) with the introduction of the radius-coradius concept.

[8] This does not hold for an arbitrary decentralized institution. For example, suppose there are two active institutions z and z' with $\underline{p}_z < \underline{p}_{z'} < \overline{p}_z < \overline{p}_{z'}$. Then none of the cases from Lemma 4 applies. The crucial part of Lemma 4 is that, given a fixed, single-price institution z, it can be applied to z and *any* other institution z'.

(i) *z is attractive for all buyers at z' (and z' is attractive for all sellers at z),*

(ii) *z is attractive for all sellers at z' (and z' is attractive for all buyers at z), or*

(iii) *there exist a buyer i and seller j at z' such that z is attractive for both i and j.*

Proof of Lemma 4. Let z be a single-price institution. Fix a state-matching pair (ω, γ) with $z, z' \in A(\omega, \gamma)$ and let p be the (unique) price at z. Clearly $T_{z'} \neq \emptyset$ as z' is active. Further, p is larger than all prices in $T_{z'}$ (resulting in case (i)), or p is smaller than all prices in $T_{z'}$ (resulting in case (ii)), or p lies (weakly) between two prices at z', i.e. there exists $\underline{p}', \overline{p}' \in T_{z'}$ such that $\underline{p}' \leq p \leq \overline{p}'$. In the latter case, there is a buyer at z' trading at price $\underline{p}'$ that is (weakly) lower than p, and a seller at z' trading at price $\overline{p}'$ that is (weakly) higher than p, hence we are in case (iii). ∎

We now prove the results in Subsection 1.4.3. Note that a centralized institution is also a single-price institution, hence Lemma 4 applies. The proof of Lemma 2 makes use of Lemma 4 in an iterative fashion.

Proof of Lemma 2. Let $z_0 \in Z$. Fix a state-matching pair (ω, γ) where z_0 is active. If z_0 is the only active institution, then all traders currently not at z_0 are unmatched. Thus z_0 is attractive for all these unmatched traders. By **SELF**, given revision opportunity any such trader leaves his current institution with positive probability towards some active institution, i.e. towards z_0. On the other hand, by **ACT** no trader leaves z_0 as all are matched and there is no other active institution. As a consequence we reach a state where all traders are at z_0, which yields a positive probability path to ω_0 from any state where only z_0 is active.

Now suppose that another institution $z \neq z_0$ is active at (ω, γ). By Lemma 4, there are three cases to consider.

Case (i). z_0 is attractive for all $k \in B_z$ (and z is attractive for all $k \in S_{z_0}$).

Let $j \in B_z$ be a buyer. Since z_0 is attractive for j, by **SELF** there is positive probability that, given revision opportunity, j will leave z to some active institution $\hat{z} \neq z$. By **D1**, with positive probability j is the only trader of his type revising this period. Further, by **D2** either no seller gets revision opportunity, or there is positive probability that only j and some $i \in S_z$ get revision opportunity. Hence, with positive probability the process reaches a new state with strictly less traders at z.

Case (ii). z_0 is attractive for all $k \in S_z$ (and z is attractive for all $k \in B_{z_0}$).

This case is analogous to case (i).

Case (iii). There exist traders $j \in B_z$ and $i \in S_z$ such that z_0 is attractive for i and j.

By **SELF** there is positive probability that given revision opportunity i and j leave z to some active institution $\hat{z} \neq z$. Moreover, by **D1** and **D2** there is positive probability that either only i, or only j, or only i and j are allowed to revise this period. Hence, with positive probability the process reaches a new state with strictly less traders at z.

Our goal is to iteratively construct a positive-probability path from ω to a state ω' where z_0 is the only active institution. We have just shown that in all possible cases there is a positive-probability transition from ω to a new state ω_1 with strictly less traders at z, in which no empty institution, i.e. an institution with zero buyers and zero sellers, has become non-empty (as traders only leave z towards active institutions). If z is still active we can apply exactly the same reasoning to this new state ω_1 to construct a positive-probability path to a state ω_2 with strictly less traders at z than in ω_1. Continuing in this fashion, after a finite number of steps we eventually reach a state where either no buyer or no seller remains active at z, thus we have reached a state where z is inactive. In particular, all remaining traders at z are unmatched and by **SELF** leave z towards an active institution if given revision opportunity, hence we can reach a state ω_r where z is empty, hence inactive. In particular, the number of non-empty institutions is strictly smaller. If there is another active institution z' at (ω_r, γ_r), we can repeat the whole argument and thus obtain a positive-probability path to a state where also z' is empty. Proceeding iteratively in this fashion, we can finally construct a positive-probability path to a state ω' where z_0 is the only active institution. This completes the proof. ■

Proof of Theorem 1. Let $z_0 \in Z$. We first observe that $\{\omega_0\}$ is absorbing by Lemma 1 since z_0 always matches all traders, in particular at ω_0. From any given absorbing set two mutations (a buyer and a seller) suffice to reach a state where z_0 is active. By Lemma 2 the state ω_0 can then be reached from this state without further mutations. Hence $CR(\{\omega_0\}) \leq 2$.

On the other hand, by **ACT** the state ω_0 cannot be left with just one mutation. It follows that $R(\{\omega_0\}) \geq 2$. Hence applying Lemma 3(a), it follows that ω_0 is stochastically stable. ■

We now turn to the proofs of results in Subsections 1.4.4 and 1.4.5.

Proof of Proposition 1. Let $z_0 \in Z$. Consider a matching-inefficient institution z. One of the following conditions holds.

- z does not match all traders at ω_z.
- At some state with $(n_z, m_z) = (n, m-1)$, z leaves a buyer unmatched with positive probability.
- At some state with $(n_z, m_z) = (n-1, m)$, z leaves a seller unmatched with positive probability.

In the first case, there exists a state-matching pair (ω_z, γ) with $k \in S_z \cup B_z$ unmatched. Then every institution other than z is inactive, hence, by **SELF** k will switch to any institution, in particular z_0, with positive probability. Now one mutation from a trader $k' \in S_z \cup B_z$ of the other market side is sufficient to make z_0 active.

For the other two cases, a state ω with $(n_z, m_z) \in \{(n, m-1), (n-1, m)\}$ can be reached by one mutation towards z_0. Now there exists a state-matching pair (ω, γ) with an unmatched buyer $k \in B_z$, respectively an unmatched seller $k \in S_z$, and by **SELF** the unmatched buyer, respectively seller, switches to z_0 with positive probability so that z_0 becomes active.

In any case, one mutation from ω_z suffices to reach a state from which, by Lemma 2, ω_0 can be reached without further mutations. Hence $c(\omega_z, \omega_0) = 1 < 2 \leq R(\omega_0)$ and by Lemma 3(d) it follows that ω_z is not stochastically stable. ■

Proof of Theorem 2. Let $z_0 \in Z$. Any stochastically stable institution must be matching-efficient by Proposition 1. To see the converse, let $z \in Z$ be a non-trivial institution that is matching-efficient. Suppose we are in ω_0 and a single buyer j and seller i mutate switching to z so that it becomes active. In this new state ω there is only one buyer and one seller at z, hence they must be matched with positive probability by non-triviality of z and we have $T_z(\omega, \gamma) = \{p(1)\}$ for some (ω, γ).

Let $p_0(\omega)$ be the price at z_0 in state ω. If $p_0(\omega) \leq p(1)$ ($p_0(\omega) \geq p(1)$), then z is attractive for all $k \in S_{z_0}$ ($k \in B_{z_0}$) and $A(\omega, \gamma) \setminus \{z_0\} = \{z\}$, hence by **SELF** every member of the respective market side at z_0 will switch to z (the only active institution) with positive probability, given revision opportunity. By **D3**, with positive probability all members of the appropriate market side revise simultaneously, leading to a state with either $n_z = n$ or $m_z = m$; hence, z_0 becomes inactive. In particular, no other institution can become active and z is active by non-triviality, hence z is the only active institution. If after this transition there are still traders at z_0, again by **SELF** all remaining members of the other market side (they are all unmatched) follow with positive probability if given revision opportunity, which happens with positive probability by **D3**. We hence reach the state ω_z. Since ω_0 cannot be left with less than two mutations we obtain $c(\omega_0, \omega_z) = 2$.

On the other hand, by **ACT** ω_z cannot be left with less than two mutations as it is matching-efficient, while two mutations suffice for a transition towards ω_0 (as in the proof of Theorem 1), hence $R(\omega_z) = 2$. Thus ω_z is stochastically stable by Lemma 3(c). ■

Proof of Proposition 2. First note that, by Proposition 1, and since **OPT** implies **SELF**, every non-trivial, stochastically stable institution must be matching-efficient. Hence we only need to prove the converse.

By Theorem 1, z_0 is stochastically stable. Let $z \in Z$ be a matching-efficient, non-trivial institution, $z \neq z_0$. To show that z is also stochastically stable, by Lemma 3(c) it suffices to show that $c(\omega_0, \omega_z) = R(\omega_z)$. By Lemma 1, ω_z is an absorbing state because, by matching efficiency, it matches all traders at ω_z. By **ACT**, ω_z cannot be left with less than two mutations.

However, two mutations suffice to leave ω_z towards ω_0 (recall Lemma 2). It follows that $R(\omega_z) = 2$.

Next, we show that ω_z can be reached from ω_0 with two mutations. Starting at ω_0, two mutations (a buyer and a seller) are sufficient to reach a state ω where z is active. We now iteratively construct a path from ω to ω_z.

Consider a state ω' where z_0 and z are active, and all other institutions are empty. Let p_0 be the (unique) price at z_0 and consider a price $p \in T_z$. Then $p \leq p_0$ or $p_0 \leq p$, hence it is either weakly better for buyers at z_0 or weakly better for sellers at z_0. By **OPT** any member k of the respective market side at z_0 will switch to z (as it is the only active institution other than z_0) with positive probability, given revision opportunity. By **D1**, with positive probability k is the only trader of his type revising this period. Further, by **D2** either with positive probability k is the only trader with revision opportunity, or with positive probability only k and some k' of the other market side who is currently at z_0 receive revision opportunities. Hence, we can reach a new state with strictly more traders at z (at least k switches to z) where no other institution than z_0 and z is active (by **ACT** k' either switches to z or stays at z_0).

Applying this argument iteratively yields a positive-probability path from ω to a state where all buyers or all sellers are at z, hence z_0 is inactive. All remaining traders at z_0 (if there are any) are unmatched, hence switch to the only active institution z by **OPT**, given revision opportunity. We have thus constructed a positive-probability path leading to ω_z that requires only two mutations. This shows that $c(\omega_0, \omega_z) = 2$ and completes the proof. ■

Proof of Proposition 3. Let $Z = \{z_0, z_D\}$ and assume $n, m > 3$. Since only z_0 and z_D are available, $R(\omega_0) = c(\omega_0, \omega_D)$. Thus, by Lemma 3(d), it suffices to show that $c(\omega_D, \omega_0) < c(\omega_0, \omega_D)$. As in the proof of Theorem 1, $c(\omega_D, \omega_0) = 2$.

We have to show that a transition from ω_0 to ω_D requires at least three mutations. By contradiction, suppose there is a path from ω_0 to ω_D that requires at most two mutations. Under asynchronous learning in any transition along the path at most one trader can switch institutions at the same time. Along this path z_D has to become active, but by **ACT** no trader at z_0

can switch to z_D as long as it is inactive, which is the case at ω_0. Hence the path has to involve two mutations (a buyer and a seller) from ω_0 to the first state in the path where z_D is active. Note that, since $n, m > 3$ and only two mutations are possible, these transitions do not render z_0 inactive.

Hence, the path from ω_0 to ω_D cannot involve any further mutation. At some point along this path, z_0 has to become inactive, hence the path needs to contain a transition (without mutation) from a state ω where z_0 is active to a state ω' where z_0 is inactive. If z_D were inactive at ω, it would not be attractive for any trader at z_0 (as z_0 is active at ω), hence by **PES** the transition from ω to ω' could only occur via an additional mistake, a contradiction. Hence z_D must be active at ω.

The transition from ω to ω' involves a single trader switching from z_0 to z_D. Suppose this trader is a buyer (the case of a seller is symmetric). That is, $n_0(\omega) = 1$, $n_0(\omega') = 0$, and $m_0(\omega') = m_0(\omega) \geq 1$. By **PES**, z_D must be attractive at ω for the buyer switching from z_0 to z_D, that is, $p_0 \geq p$ for all $p \in T_{z_D}(\omega)$.

As z_D is active at ω it either features a single price $p_D = p(n-1)$ (for $m_D(\omega) = 1$), or it features two prices $\underline{p}_D \leq \overline{p}_D$ with $\overline{p}_D = p(n-2)$ (for $m_D(\omega) > 1$). Since $n \geq 4$ and $p(r)$ is strictly increasing in r, it follows that $p_D \geq p(2)$ and $\overline{p}_D \geq p(2)$. On the other hand, the single price p_0 at z_0 in state ω is given by $p_0 = p(\frac{1}{m_0(\omega)}) \leq p(1)$ as $m_0(\omega) \geq 1$. It follows that $p_0 \leq p(1) < p(2) \leq \min\{\overline{p}_D, p_D\}$, a contradiction to $p_0 \geq p$ for all $p \in T_{z_D}(\omega)$.

We have thus shown that a transition from ω to ω' requires at least three mutations. Therefore $c(\omega_0, \omega_D) \geq 3$, implying that z_D is not stochastically stable.

■

Chapter 2
Cournot vs. Walras: A Reappraisal through Simulations

2.1 Introduction

One of the main objectives of industrial organization is to determine to which extent do market outcomes deviate from the competitive, welfare-maximizing ideal. Both theoretical predictions and empirically estimated outcomes are measured against the perfectly competitive benchmark. The study of market outcomes, however, cannot be disentangled from individual behavior. Consider the case of quantity competition in oligopolistic markets (Cournot oligopolies). Perfectly competitive outcomes (Walrasian equilibria) obtain if individual behavior aims to maximize individual profits under the constraint that market prices are taken as given (which is a form of bounded rationality). In contrast, best-reply behavior underlies Cournot-Nash equilibria, where consumer welfare is lower and industry profits are larger than in the Walrasian case. Joint profit maximization leads to collusive outcomes, the mere suspicion of which might trigger market-regulation interventions.

Firm managers are often motivated by relative-performance concerns, that is, the comparison with the competition's outcomes. It is well-known that such concerns go hand-in-glove with imitative behavior, that is, mimicking the behavior of the best performers. In a seminal paper Vega-Redondo (1997) showed that imitative behavior in a noisy evolutionary Cournot oligopoly leads to the selection of the perfectly competitive Walrasian outcome where price equals marginal cost. Formally, in a discrete-time, finite-population stochastic dynamics where firms imitate best performers and make occasional mistakes, the long-run distribution of outcomes concentrates on the Walrasian outcome as the probability of mistakes vanishes: the selected outcome is called *stochastically stable* (for an introduction to stochastic stability models, see, e.g., Blume, 1993; Kandori et al., 1993; Young, 1993; Samuelson, 1997; Fudenberg and Levine, 1998). This striking result, which can be generalized to the class of aggregative games (Alós-Ferrer and Ania, 2005), is driven by the so-called *spite* effect (Hamilton, 1970; Schaffer, 1989): a deviation from the Cournot equilibrium to the Walrasian solution is detrimental in that it decreases one's payoff, but at the same time it decreases the payoffs of the other firms even more, hence making the deviating firm better off in relative terms. As a result imitation of highest profits quickly leads firms to choose the perfectly competitive quantity. This result is important, because it represents a counterpart to the concerns on reduced market-outcome competitiveness and points out that relative-performance concerns and (boundedly rational) imitative behavior might actually *increase* the competitiveness of those outcomes.

The analysis of Vega-Redondo (1997) exhibited a clear-cut but highly stylized result, which rests on some sharp behavioral assumptions. First and foremost among those is the length of memory. The selection of the Walrasian outcome under imitative behavior crucially depends on the assumption that previous (potentially more profitable) outcomes are immediately forgotten and as a consequence only relative payoffs matter. This is in stark contrast with the prediction of collusive outcomes arising when market competition is modeled as an infinitely repeated game, which hinges on the players' capability to condition on possibly distant past events (in order to sustain a subgame-perfect equilibrium in the infinitely repeated game). This limitation of Vega-Redondo (1997) was addressed in Alós-Ferrer (2004), which introduced bounded but possibly long memory into the dynamic model of Vega-Redondo (1997). Interestingly, even if imitative behavior is assumed, the assumption of non-negligible memory opens the door for intertemporal payoff comparisons and *better-reply* behavior. For, when firms can imitate whatever output level delivered the highest profits in memory, intertemporal comparisons allow to evaluate deviations from a given profile. For example, if a firm deviates away from the Cournot-Nash equilibrium, imitation will lead the firm to correct this "mistake", because the pre-deviation payoffs are larger than the post-deviation ones. This holds even if by a spite effect after deviation that firm is better off than the competition, because the remembered payoffs are even higher. Non-negligible memory hence creates a tension between actions leading to relative-payoff advantages and those leading to advantages in terms of absolute payoffs. The main result in Alós-Ferrer (2004) was that for well-behaved Cournot oligopolies and non-trivial memory (one round suffices) a non-trivial dynamics arises and all quantities between the Cournot and Walrasian outcomes are long-run equilibria. That is, all such states are stochastically stable, implying that the long-run distribution of outcomes does not concentrate on a single outcome and, as the probability of mistakes becomes small but positive, the dynamics is concentrated on outcomes between the Cournot and the Walrasian ones, but remains non-trivial. Of course, for the memoryless case convergence to the Walrasian outcome obtains (Alós-Ferrer and Shi, 2012 further showed that the latter result also holds if *some* firms are memoryless). This demonstrates the importance of memory with regard to outcome selection in Cournot oligopolies. Also, this result qualifies the original insight of Vega-Redondo (1997) and points to a richer (and possibly more realistic) dynamics, with the Walrasian and Cournot-Nash outcomes as stylized bounds of predicted market outcomes.

The main result of Alós-Ferrer (2004) has two practical limitations. First, the proof applies to a large but still specific class of Cournot oligopolies, as it relies on stronger assumptions on the structure of the underlying Cournot game compared to Vega-Redondo (1997). Essentially, the result is proven under the additional requirements that the inverse-demand function is strictly concave and the cost function is strictly convex. Second, the fact that all outcomes between the Walrasian and the Cournot-Nash ones are stochastically stable does *not* mean that they are all "alternative equilibria" in a classical sense. The reason is that stochastic stability refers to the limit as behavioral noise (the probability of mistakes) vanishes. The result has to be interpreted in terms of the dynamics for a positive but small noise level. The actual prediction is that the dynamics will quickly converge towards the interval of quantities between the Walrasian and Cournot-Nash outcomes, and then a rich, non-trivial dynamics within this interval is to be expected. How much time the system will spend at each output level compared to others is measured by the (limit) invariant distribution of the stochastic process.

Unfortunately, the stochastic stability techniques on which Vega-Redondo (1997), Alós-Ferrer (2004), and many other works are based do not allow for an estimation of that distribution, merely for an analysis of its support (the stochastically stable states). Hence, if one wishes to analyze how far apart from the Cournot (or Walras) prediction the system will be, one needs to determine the shape of the invariant distribution. Although this is analytically not feasible, a direct application of the Ergodic Theorem (e.g., Karlin and Taylor, 1975) shows that simulations can provide a sharp estimation of that shape.

The logic is simple. The Ergodic Theorem implies that, for almost all realizations of the dynamical system, in the long run the weights of individual outcomes in the invariant distribution correspond exactly to the percentage of time that the system spends in that outcome. Hence, the proportion of time spent in individual outcomes, averaged across (long enough) simulations, becomes a numerical estimate of the weights in the invariant distribution. As a consequence, extensive numerical simulations become an efficient tool for the systematic study of the characteristics of long-run predictions, and are particularly valuable when (as in Alós-Ferrer, 2004) the prediction is not a single outcome. For instance, the stochastic stability result quoted above cannot discriminate between outcomes where almost all weight is placed on or near the Walrasian quantity and outcomes where that weight is on or near the Cournot-Nash quantity. Systematic simulations then become invaluable to discriminate among such possibilities.

Our objective in this paper is twofold: On the one hand, we want to better understand the exact shape of the invariant distribution and hence whether the prediction is closer to the perfectly competitive outcome or rather to the classical Cournot-Nash outcome. On the other hand, we seek to investigate whether the main result in Alós-Ferrer (2004) holds beyond the limitations just described, that is, whether it extends to less well-behaved Cournot oligopolies. To those ends, we will rely on computational simulations to approximate the invariant distribution, systematically varying the specifications of the underlying Cournot oligopoly and the dynamical system.

This article is also linked to the literature studying convergence in Cournot oligopoly games where learning of boundedly rational agents is modeled through evolutionary algorithms (EA). An important factor determining the outcome the EA converges to is the type of learning it is based on. The literature mainly distinguishes between two types: individual learning exclusively from own past performance (i.e. in the absence of spite and imitation of others, when firms only learn through introspection) and social learning from own as well as others' past performance (i.e. imitation of others as considered in Vega-Redondo, 1997 and Alós-Ferrer, 2004). An early contribution in this line was provided by the simulations of Vriend (2000), who found convergence to the Cournot-Nash equilibrium when learning was individual (closer to myopic best-reply), but to the Walrasian outcome in the presence of social learning (closer to imitation). This is a natural observation in view of the fact that best-reply behavior underlies Nash equilibria and imitative behavior leads to Walrasian outcomes. Individual learning focuses on the own experience, while information about the performance of others is either not available or ignored. Hence there is no room for relative payoff comparisons and only absolute payoffs matter, which can, of course, drive behavior away from the competitive outcome. In a formal-analytical study relying on stochastic stability, Bergin and Bernhardt (2004) examined imitation dynamics and introspective dynamics in isolation and found that individual agents (which in a Cournot setting means firms, excluding consumer welfare) are worse off in a world of imitators than in a world where agents learn via introspection. In a similar framework Riechmann (2006) showed using EA simulations that even individual learning can lead to either Walras or Cournot, depending on the analytical sophistication of players and their degree of knowledge of the game. Individual learning mechanisms leading to the Cournot equilibrium require a sizable degree of knowledge and analytical sophistication, while simpler behavior might lead towards Walrasian outcomes.

The simulation results of Vriend (2000) were in contrast to Arifovic (1994), who found that both social and individual learning converge to the Walrasian outcome. However, this earlier work relies on an EA which is not based on best-reply or imitation (of actual past-performance) but rather takes elements from both. Firms behave as price-takers, but compute and compare hypothetical payoffs in a way similar to myopic best-reply considerations, hence the underlying behavior cannot be directly compared to the models described above. Arifovic and Maschek (2006) argued that Vriend's (2000) result of convergence to the Cournot outcome under individual learning hinged upon a very specific cost structure and particular EA implementation. Vallée and Yıldızoğlu (2009) analyzed the differences between the underlying mechanisms that lead to those diverging results. Using computational experiments, they confirmed that expectation-based learning (Arifovic, 1994) cannot converge to the Cournot-Nash equilibrium, while this is possible under the repetition-based learning model employed by Vriend (2000). The main reason for this difference is that the latter model belongs to a completely different family. For instance, trading strategies are tried out for 100 periods and the EA (actually, a genetic algorithm including recombination) acts only after each 100-periods block, on the basis of average profits. Hence, the EA acts only occasionally, which allows individuals to discover the decreasing relationship between market price and quantity. On the other hand, in the case of social learning the computational results in Vallée and Yıldızoğlu (2009) are in line with the theoretical prediction of Vega-Redondo (1997) driven by the spite effect.

Vallée and Yıldızoğlu (2013) use computational experiments to investigate the role of memory under both social and individual learning in the possible convergence to outcomes more collusive than the Cournot equilibrium. Their results for social learning are (up to a significant level of noise) in line with Alós-Ferrer (2004).[1] For specific types of individual learning mechanisms they find convergence towards quantities that are close to the collusive

[1] Vallée and Yıldızoğlu (2013) conduct simulations with high levels of noise and large probability of inertia (that is, infrequent strategy adjustments). Unsurprisingly, their results include distributions with long tails, which arise exclusively due to the levels of noise. Stochastic stability analyzes the limit as noise vanishes, modeled as the probability ε of behavioral mistakes. The key insight of the literature is that this limit differs from the heavily path-dependent behavior when there is no noise at all. Of course, for large ε the prediction is simply a high level of noise, from which little can be learned. Computational tests of stochastic-stability results need to concentrate on small but positive values of ε.

outcome.[2] However, although the social learning dynamic they study is, as ours, based on random experiments, imitation, and memory, the implementation is very different from the model studied here and in Alós-Ferrer (2004). First, agents are equipped with finite memory but only occasionally use it and thus agents act in a "memoryless" way most of the time. As a consequence, the results of Alós-Ferrer (2004) do not directly apply to their setting (a better comparison would be to Alós-Ferrer and Shi, 2012). Second, they consider imitation (of the best outcome in the previous period, i.e. with trivial memory) and the act of "using memory" (that is, imitation of the best outcome within the remembered time frame) as two distinct, independent events. Third, Vallée and Yıldızoğlu (2013) use an ad hoc notion of "convergence" which, regrettably, does not fully exploit the properties of the actual stochastic system. They simply compare the population distribution in the last (of many) periods in the simulation to the predicted theoretical limit distribution. For a single simulation, if the invariant distribution has a non-trivial support (multiplicity of stochastically stable states), this criterion produces essentially random results, especially if the simulation is too short. Averaging over simulations but relying only on the last period of each one is an appeal to the Fundamental Theorem of Markov chains (see, e.g., Karlin and Taylor, 1975), which states that the long-run probability of each individual state is numerically equal to its probability under the invariant distribution. Vallée and Yıldızoğlu (2013) average over only 500 runs of only 10, 000 periods each. It is unclear whether, with such a short length of the simulations, this relatively small number of simulations suffices to obtain an approximation through the Fundamental Theorem. The Ergodic Theorem (see also Section 2.3), however, provides a better approach, which relies on the whole simulation instead of just the last period. The key is that time averages, that is, the fraction of time spent at each state as time goes to infinity, converges to the invariant distribution, and hence empirical time

[2]This result requires "selective" memory, where agents can only remember the latest payoff associated with a given quantity and forget the payoffs associated with earlier, possibly more profitable instances where the same quantity was chosen. This dynamics is similar to trial and error learning as studied in Huck et al. (2004) (both analytically and through simulations).

averages approximate the invariant distribution.

At a conceptual level, the inconsistencies among the works quoted above arise because the different simulations they study differ in a relatively large number of dimensions (and convergence criteria), and it is not clear which of them are purely technical and which define qualitatively different classes of learning dynamics. In turn, this is made possible because those simulations are exploratory in nature (which is of course valuable) but are not, in general, conceived as a systematic test of the long-run outcomes of a well-defined class of dynamics. In contrast our research agenda starts with a clearly formulated theoretical framework (stochastic stability models where agents are endowed with behavioral rules) and builds upon existing analytical results, exploring their boundaries.

The article proceeds as follows. Section 2.2 introduces the discrete-time Cournot oligopoly and the imitation-based learning dynamic with memory. The main result of Alós-Ferrer (2004) in that framework is that if firms have positive memory then all quantities between the Walrasian and the Cournot ones are stochastically stable. Section 2.3 describes the simulation protocol and the simulation parameters used to generate our computational results. Section 2.4 contains the main results, starting with an analysis of the generality of the full support prediction for positive memory and proceeding to factors that influence the shape of the limit distribution, in particular the relative weights of the Cournot-Nash and Walrasian outcomes. Section 2.5 concludes.

2.2 Existing Theoretical Results

In this section we briefly review the discrete-time dynamic Cournot oligopoly studied in Vega-Redondo (1997), introduce the imitation-based learning dynamic with bounded memory studied in Alós-Ferrer (2004), state the main selection result with bounded memory obtained in that work, and discuss its limitations (which motivate the present work).

2.2.1 The Discrete Cournot Oligopoly

Consider an N-player, symmetric Cournot oligopoly with inverse-demand function $P : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and cost function $C : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$. Let P be twice-differentiable on $[0, q_{\max}]$ with $P' < 0$ and $P'' < 0$ in this interval (hence P is strictly decreasing and strictly concave). Further, assume that $P(0) = P_{\max} > 0$ and $P(q) = 0\ \forall q \geq q_{\max}$. Let C be twice differentiable with $C' > 0$ and $C'' > 0$ (hence C is strictly increasing and strictly convex), and assume $C'(0) < P_{\max}$. Call such a Cournot oligopoly *well-behaved.*

The Walrasian quantity q^W is defined by

$$P(Nq^W)q^W - C(q^W) \geq P(Nq^W)q - C(q)\ \forall q,$$

and it is unique under the assumptions given above. It is well-known that this quantity is optimal with respect to relative-payoff considerations (see, e.g., Alós-Ferrer and Ania, 2005). In contrast, the Cournot quantity q^C is optimal with respect to absolute payoffs, since it fulfills that

$$P(Nq^C)q^C - C(q^C) \geq P((N-1)q^C + q)q - C(q)\ \forall q.$$

Alós-Ferrer (2004) analyzes the (discrete-time) Cournot oligopoly described above within a dynamic context with memory. The model without memory was first studied in the seminal paper Vega-Redondo (1997). Firms play the Cournot game repeatedly in discrete time $t = 0, 1, 2 \ldots$. Each period firms observe quantities chosen and profits realized by all firms in that period. In addition, firms remember the outcomes, i.e. quantities and profits, from the last $K \geq 0$ periods in addition to the current one. Based on this information firms subsequently choose a quantity for the next period, taken from a finite grid $\Gamma = \{0, \delta, 2\delta, \ldots, \nu\delta\}$ with step size $\delta > 0$ where $\nu\delta = q_{\max}$. For concreteness, we assume that both q^W as well as q^C belong to Γ.

The state of the learning process for a given period is entirely determined by a vector $(q_1, \ldots, q_N) \in \Gamma^N$, therefore the state space for the model with memory $K \geq 0$ is given by $\Gamma^{N(K+1)}$. The dynamics is based on imitation of strategies that performed best within the remembered time frame and occasional experimentation (also called mutation). More precisely, for each period and every firm, there is a small probability $\varepsilon > 0$ that instead of imitating the firm experiments with a new quantity at random according to a probability distribution with full support on Γ. With the remaining probability $(1-\varepsilon)$ the firm imitates a quantity that yielded the highest profit in memory, i.e. within the last K periods (including the current period). Formally, let $q_i(t)$ be the quantity chosen by firm i in period t and $q_{-i}(t)$ the quantities of its $N-1$ competitors. Then the set of quantities chosen with positive probability in the next period for a state $\omega = (q(k))_{k=t-K}^{t} \in \Gamma^{N(K+1)}$ is given by

$$\begin{aligned} B_t^K(\omega) = &\{q_i(k) \mid i \in \{1,\dots,N\}, k \in \{t-K,\dots,t\} \text{ and } \Pi_i(k) \geq \Pi_j(k') \\ &\forall j = 1,\dots,N,\ \forall k' = t-K,\dots,t\} \end{aligned} \tag{2.1}$$

where $\Pi_i(t) = P(Q(t))q_i(t) - C(q_i(t))$ and $Q(t) = \sum_{i=1}^{N} q_i(t)$.

This behavioral rule is often referred to as Imitate-the-Best-Max (IBM) rule in this literature.[3] Imitation and experimentation are events that happen independently from each other across firms and time. The learning dynamics (with imitation and experimentation) defines a stationary Markov chain on the state space $\Gamma^{N(K+1)}$ that is analyzed using standard tools introduced by Blume (1993), Kandori et al. (1993) and Young (1993). One seeks to identify the *stochastically stable states*, which are those in the support of the limit invariant distribution of the process as the probability of experimentation vanishes. Given a quantity $q \in \Gamma$, denote by $\text{mon}(q)$ the quantity vector where all firms choose q. More generally, for $K \geq 0$ we denote $\text{mon}(q, K) = ((q,\dots,q),\dots,(q,\dots,q)) \in \Gamma^{N(K+1)}$. These states, where the same quantity has been produced by all firms as long as it is remembered, are called *monomorphic*. For imitation-based learning dynamics it is well-known that the absorbing states are exactly the monomorphic states, and as

[3] Another reasonable behavioral rule is the so-called Imitate-the-Best-Average (IBA) rule, however, this rule leads to completely different results. Bergin and Bernhardt (2009) have shown that IBA together with long enough memory leads to cooperative outcomes in the long-run, in particular in a Cournot oligopoly context the unique long-run prediction is the state where all firms choose the collusive quantity.

an easy consequence only those can be stochastically stable (see Alós-Ferrer, 2004 or Alós-Ferrer and Shi, 2012 for details).

2.2.2 Imitation with Bounded Memory

For the model without memory, i.e. $K = 0$, Vega-Redondo (1997) has shown that the unique stochastically stable state is the one where all firms choose the Walrasian quantity q^W, that is, the monomorphic state mon(q^W). The main result in Alós-Ferrer (2004) is that for positive memory ($K \geq 1$) every quantity q in the interval $[q^C, q^W]$ corresponds to a stochastically stable state, namely to the monomorphic state mon(q, K). Further, those states are the only stochastically stable states.

Theorem 1 (Alós-Ferrer, 2004)**.** *Consider a well-behaved N-player Cournot oligopoly with memory K on the grid Γ with step size δ ($q^C, q^W \in \Gamma$). For any $K \geq 1$, $N \geq 2$, and δ small enough, the set of stochastically stable states is* $\{\text{mon}(q, K) \mid q \in [q^C, q^W] \cap \Gamma\}$.

This result is in stark contrast to Vega-Redondo (1997) where the competitive equilibrium is the unique prediction. Allowing for positive memory, however, introduces another force driving selection that acts alongside the relative payoff comparisons that favor the Walrasian quantity. Memory increases the importance of absolute payoffs through intertemporal comparisons, which favor the Cournot quantity.

The selection result for the model with bounded memory has two limitations. First, stochastic stability techniques aim at identifying the stochastically stable states, but in case of multiplicity of stable states are generally not able to identify the shape of the limit distribution. With memory all quantities between the Walrasian and Cournot outcomes are stable under an imitation dynamics. However, this only implies that coordination on each of those outcomes will be observed a positive fraction of time in the long-run. Results based on stochastic stability techniques do not provide further information about the exact shape of the limit distribution, and as a consequence, we do not know whether those fractions of time are large or small.

Second, the result in Alós-Ferrer (2004) makes strong structural assumptions compared to Vega-Redondo (1997) (well-behaved oligopolies). These assumptions, however, mainly serve a technical purpose enabling the fairly involved construction on which the proof of the main result is based. The technical difficulty is that destabilizing the Walrasian quantity "downwards" is costly in terms of the number of mutations required. The proof proceeds through a series of transitions passing through quantities outside the stable region $[q^C, q^W]$ above q^W from where a transition to a quantity below q^W is comparably cheap because both forces, relative and absolute payoff improvements, are aligned. The proof, partially based on differential calculus, is not entirely intuitive and proceeds through a series of intermediate lemmata. These lemmata make intensive use of the assumptions on the structure of the game, specifically of the strict concavity of the inverse-demand function and the strict convexity of the cost function. Hence it is not feasible to generalize the proof so that it still applies with weaker assumptions (or at least we have failed to do so). Besides these technical considerations we hypothesize that the general logic driving the stability of the complete range of outcomes between q^C and q^W still holds in less well-behaved environments. In particular, we aim to examine oligopoly games with non-concave inverse-demand functions and cost functions that are not necessarily strictly convex.

In view of the overall motivation, the analysis of the invariant distribution translates into a series of more specific questions. What are the relative weights on the Walrasian and Cournot quantities, respectively? What is the exact shape of this distribution and how is it affected by the specific structure of the oligopoly, i.e. the specific demand and cost functions chosen? What are the effects of memory length on the limit distribution? The use of computational simulations allows us to shed light on these questions. Simulations deliver data on the complete distribution and are an effective means to explore the validity of the full support prediction beyond the structural assumptions made in Alós-Ferrer (2004). In the next section we detail the simulation protocol and the parameters used to tackle these questions.

2.3 The Simulations

The Ergodic Theorem from the theory of Markov chains allows the following interpretation of the limit invariant distribution of a stationary Markov chain: the fraction of time that the dynamic process spends at a specific state converges to the weight given to that state by the limit distribution. This has important consequences, and in particular opens the door for a computational approach to study the limit distribution through the use of simulations. Specifically, the Ergodic Theorem enables us to approximate the limit distribution of a stationary Markov chain, in our specific case the distribution over the monomorphic states mon(q) with $q \in \Gamma$, by time averages obtained via simulations. This is exactly the strategy pursued in this section. In the following we will give a detailed description of the simulations used to obtain approximations of the invariant distribution.

2.3.1 Simulation Protocol

In this section the Cournot oligopoly game described in Section 2.2 is transformed into a simple agent-based simulation protocol mimicking the evolutionary process underlying the theoretical results. The initial quantity $q_i(0) \in \Gamma$ chosen by agent i at time $t = 0$ is randomly drawn according to a uniform distribution on the grid Γ. In particular, we do not necessarily start with a monomorphic state. Denote by $Q(t) = \sum_{i=1}^{N} q_i(t)$ the aggregate quantity in period t. The payoff of player i in period t is then given by $\Pi_i(t) = P(Q(t))q_i(t) - C(q_i(t))$. This determines the set of "best-performing" strategies in the last $K + 1$ periods, $B_t^K(\omega)$, as defined in (2.1).

Agents will adopt strategies from B_t^K, randomly (and uniformly) picking one in case B_t^K is not a singleton. With probability $(1 - \varepsilon)$ agent i uses the so-determined strategy, however, with probability ε (independently drawn for each agent) she experiments with a new strategy chosen randomly according to the uniform distribution on Γ. This process is then repeated for a large number of periods. Since monomorphic states are the only absorbing sets of the unperturbed dynamics, the time spent in non-monomorphic states approaches zero as ε becomes small. Hence, we know that the system spends most of the time in monomorphic states (for small ε). For this reason, we only record the distribution over those and aggregate all non-monomorphic states in a residual, denoted res. Specifically, for each simulation (run) we record the fraction of time spent in each of the monomorphic states mon(q), $q \in \Gamma$, denoted by $F(q)$, as well as the fraction of time spent in non-monomorphic states, denoted by $F(\text{res})$. Formally, our main output is a probability distribution F over $\Gamma \cup \{\text{res}\}$. The restriction of F to Γ gives us the desired approximation of the limit invariant distribution. To further improve quality of the approximation and to guarantee robustness we recorded 150 repetitions (with different initial conditions) and then took the average distribution over those repetitions for our subsequent analysis.[4]

To facilitate later analysis and to ensure comparability of the results across the various configurations we determined the grid Γ endogenously in order to satisfy certain criteria: the scale (units) was normalized so that Γ would be contained in $[0, 1]$; Γ should contain the quantities of interest q^C and q^W; the cardinality of Γ should only depend on the step size δ and not on any other parameter; the smallest element in Γ should be "close to 0", meaning as close to zero as the discretization would allow.

We will conduct simulations for different grid sizes. Grid coarseness will be captured by the parameter M, which measures the number of steps in the grid from q^C to q^W. The step size, i.e. the increment within the discrete grid, is then given by $\delta = \frac{q^W - q^C}{M}$. That is, the step size is determined relative to the size of the interval $[q^C, q^W]$ as to ensure that the number of elements in the grid that lie within the interval $[q^C, q^W]$ is $M + 1$. We take the grid size to be a fixed multiple of $\left|q^W - q^C\right|$, specifically we set $\nu = 7M$ and hence $|\Gamma| = 7M + 1$. The position of the grid within $[0, 1]$ is specified as follows. For a given starting point $q_0 \in \mathbb{R}_+$ we set $\Gamma(q_0) = \{q_0, q_0 + \delta, \ldots, q_0 + \nu\delta\}$. We then determine the smallest element of the grid q_0 such the criteria above are met through an iterative procedure. Set $q_0 = q^C - \lfloor \frac{q^C}{\delta} \rfloor \delta > 0$. If $q^W \leq q_0 + 6M\delta$, set $\Gamma = \Gamma(q_0)$ and end the procedure. If not, set $q_1 = q_0 + M\delta$. If $q^W \leq q_1 + 6M\delta$, set $\Gamma = \Gamma(q_1)$ and end the procedure, else set $q_2 = q_1 + M\delta$.

[4]Due to the presence of mutations, the simulated dynamics is ergodic, that is, long-run predictions are independent of initial conditions.

Table 2.1: Inverse-demand functions used in the simulations.

Inverse-Demand Function	Properties
$P_1(Q) = \max\{1 - Q^2, 0\}$	strictly concave
$P_2(Q) = \max\{\frac{\log(2-Q)}{\log(2)}, 0\}$	strictly concave
$P_3(Q) = \max\{\frac{e}{e-1} \cdot (1 - e^{Q-1}), 0\}$	strictly concave
$P_4(Q) = \max\{1 - Q, 0\}$	linear
$P_5(Q) = (\frac{1}{Q+1})^2$	log-convex
$P_6(Q) = e^{-Q}$	convex, log-concave, isoelastic
$P_7(Q) = \max\{\frac{\log(\frac{4}{5}Q+\frac{1}{5})}{\log(\frac{1}{5})}, 0\}$	convex
$P_8(Q) = (24Q + 1)^{-\frac{1}{2}}$	convex, (α, β)-biconcave for $\alpha, \beta \leq \frac{1}{100}$
$P_9(Q) = \mathbf{1}_{Q\leq 1}(1 - Q^{\frac{3}{4}})^{\frac{4}{3}}$	convex, (α, β)-biconcave for $\alpha, \beta \leq \frac{3}{4}$
$P_{10}(Q) = \max\{\frac{1+(1-2Q)^3}{2}, 0\}$	non-concave, non-convex, S-shaped
$P_{11}(Q) = \max\{\frac{1-\arctan((2+\tan(1))Q-2)}{1-\arctan(-2)}, 0\}$	non-concave, non-convex, inverted S-shaped
$P_{12}(Q) = \frac{2^{(1-Q^2)}}{2}$	non-concave, non-convex, inverted S-shaped

Proceed iteratively generating $q_{k+1} = q_k + M\delta$ until $q^W \leq q_k + 6M\delta$ and then set $\Gamma = \Gamma(q_k)$. This procedure ensures that Γ satisfies all the criteria just mentioned except the last one. In particular, we always have $q^C, q^W \in \Gamma$, and that Γ exceeds q^W by at least $\left|q^W - q^C\right|$ so that neither q^W nor q^C is at the border of the grid. Due to the large variance in the distance $\left|q^W - q^C\right|$ across the different configurations, sometimes the grid could not start just above 0 because $\left|q^W - q^C\right|$ was relatively small. Hence, "close to 0" means "as close as possible without violating one of the other criteria". As we will mainly focus on values between the Walrasian and the Cournot quantities, we view this as an acceptable simplification.

2.3.2 The Simulation Parameters

Inverse-demand functions are taken from a set of twelve different functions designed to cover a wide spectrum of properties (e.g. linear, convex, concave, log-concave, log-convex, S-shaped). The specific inverse-demand functions used in the simulations and their properties are listed in Table 2.1 below. All demand functions are normalized so that $P_i(0) = 1$ and $q_{\max}(P_i) = \inf\{q \in$

Table 2.2: Cost functions used in the simulations.

Cost function	Properties
$C_1(q) = 0.25q$	linear
$C_2(q) = 0.5q$	linear
$C_3(q) = 0.75q$	linear
$C_4(q) = 0.25q^2$	convex
$C_5(q) = 0.5q^2$	convex
$C_6(q) = 0.75q^2$	convex

$\mathbb{R}_+ \mid P_i(q) = 0\} = 1$ if this latter quantity is finite[5] for $i = 1, \ldots, 12$. Note that only P_1, P_2 and P_3 satisfy the strict concavity assumption from Alós-Ferrer (2004), and hence only those are covered by Theorem 1. We employ a set of six cost functions, three linear ones and three quadratic ones. The specific cost functions used are shown in Table 2.2.

The set of demand functions was determined as follows. The functions P_1, P_2, and P_3 are strictly concave demand functions that are normalized versions of common examples in the literature (Anderson and Engers, 1992; Amir and Lambson, 2000). We then included P_4 as a very simple linear demand function. Function P_5 belongs to a class of log-convex functions used by Amir (1996, Example 3.3). Functions P_6 and P_7 were chosen as examples of convex demand functions with varying curvature. Functions P_8 and P_9 are examples of convex functions that are also biconcave (Ewerhart, 2014), that is, they become concave after simultaneous monotone transformations of price and quantity. Function P_{10} represents an example of an S-shaped function that is concave below and convex above a certain value. On the other hand, the functions P_{11} and P_{12} are examples of inverted S-shaped functions, i.e. their first part is convex, while they become concave above a certain threshold. The three last functions were included as they are neither always concave, nor always convex within the region of interest.

Each pair of inverse-demand and cost functions (P, C) defines a particular

[5]Demand functions P_5, P_6, P_8, and P_{12} approach 0 asymptotically, hence $q_{\max} = \infty$. For these four functions, the normalization was such that $P_i(1)$ was reasonably close to zero.

Table 2.3: Range of parameters used in the simulations.

Description	Name	Values
Inverse-Demand Function	P	$P_1, \ldots, P_{12}$
Cost Function	C	$C_1, C_2, C_3, C_4, C_5, C_6$
Number of Firms	N	$2, 3, 4, 5, 10, 15, 20$
Pr(Experimentation)	ε	$0.001, 0.005, 0.010$
Grid Coarseness	M	$10, 20, 30, 40, 50$
Memory Length	K	$1, 5, 10, 50$

Notes: The inverse demand functions are as in Table 2.1. The cost functions are as in Table 2.2.

Cournot oligopoly for a total of 72 different configurations. For each configuration we calculated the Cournot and Walras quantities.[6] Besides varying the inverse-demand functions P, the cost functions C, and the coarseness of the grid M, we also vary the number of firms N, probability of experimentation ε and the memory length K ($K = 0$ refers to the case without memory, see Vega-Redondo, 1997). Table 2.3 summarizes the parameters used in the simulations.

We therefore have a total of $30,240$ different parameter combinations. For each combination we ran 150 simulations for a total of $150 \times 30,240 = 4,536,000$ (about 4.5 million) simulations. Additionally, we run $150 \times 7,560$ simulations for the no-memory case $K = 0$ in order to make sure that the algorithms run correctly (of course, for all those convergence was as predicted in Vega-Redondo, 1997). We let each simulation run for a minimum of $100,000$ periods and up to a maximum of one million periods.[7]

When dealing with limit results, naturally, one has to deal with the question of how long is long enough (Ellison, 1993). As a first indicator we can compare the empirical distribution in the no-memory case $K = 0$ to the the-

[6] We used analytical solutions whenever feasible, and numerical solutions otherwise.

[7] We also implemented an endogenous stopping condition beyond a certain threshold to save computing time. For any period $t > 100,000$, the simulation stopped if the proportion of occurrences of monomorphic states lying between the Cournot and the Walrasian outcome exceeded 99%, i.e. whenever the condition $f_t([q^C, q^W])/(1 - f_t(\text{res})) > 0.99$ was fulfilled for some $t > 100,000$ where $f_t : \Gamma \cup \{\text{res}\} \longrightarrow [0, 1]$ is the relative frequency distribution up to period t. Our large-scale simulations nevertheless took a significant amount of time on a Super Computer (in total about 8.5 CPU years).

oretical distribution, which in this case places all weight on the Walrasian outcome (Vega-Redondo, 1997). Indeed, for $K = 0$ the process spends on average about 80% of the time at the Walrasian quantity for low levels of noise suggesting that the length of our simulations is sufficient to obtain a good approximation of the limit distribution. For positive memory, the results in Alós-Ferrer (2004) can be used (via a straightforward Radius-Coradius approach following Ellison, 2000) to show that the expected time of first arrival in the interval $[q^C, q^W]$ is of order ϵ^{-1}, which corresponds to (a constant times) 100, 200, and 1000 periods for low, medium and high noise, respectively. In our data the average time of first arrival in the interval was 381, 691, and 3475 periods for low, medium and high levels of noise, respectively, and hence well below the minimal simulation length of $100,000$ periods. For the averaged data (across the 150 repetitions per parameter combination) all monomorphic states within the interval were visited for each and every parameter combination with the coarsest grid ($M = 10$). With the finest grid ($M = 50$), still all monomorphic states within $[q^C, q^W]$ were visited for 90.3% of the parameter combinations.

In our analysis we will mainly use averages over the 150 repetitions for each parameter combination, hence we have a total of $30,240$ "observations" (data points) for our main variables of interest. The simulation was coded in C++ and run on the high performance computing cluster CHEOPS (Cologne High Efficient Operating Platform for Science) at the University of Cologne. The pseudo-code is given in Table 2.4.

2.4 Computational Results

In this section we present the results of our computational simulations. Specifically, we show that the main result of Alós-Ferrer (2004) holds beyond the assumptions considered in that paper. We then proceed to a more detailed analysis of the shape of the limit invariant distribution.

The objective of the simulations is to obtain estimates of the (limit) invariant distribution μ^* through the average time spent by the system at each monomorphic state. Formally, μ^* is a distribution over $\Gamma^{N(K+1)}$, i.e. over vectors of length $K+1$, with each vector entry being a profile of quantities across firms. However, in practice the computational analysis can be greatly simplified. Formally, the system evolves over states of length $K+1$ (which include what is available in memory), but of course each such system induces a dynamics on the state of strategy profiles $(q_1, \ldots, q_N)$. In the limit, since only monomorphic states can be observed a positive fraction of time, it is immediate that the system spends a proportion r of the time in the $K+1$-length state $\text{mon}(q, K)$ if and only if the induced system evolving over strategy profiles spends the same proportion r of the time in the profile $\text{mon}(q)$. Hence, computationally, for the purposes of the approximation it is enough to keep track of the proportion of time that the profiles $\text{mon}(q)$ are visited. In turn, the latter are one-to-one with quantities, that is, we can consider the estimated invariant distribution as a mapping $\mu^* : \Gamma \longrightarrow [0,1]$. Hence, our computational results below are referred to strategy profiles of the form $\text{mon}(q)$, indexed by the quantities q. This also allows us to avoid introducing artificial difficulties in the comparison across different values of K.

2.4.1 Cournot, Walras, or Both?

Our first objective is to investigate whether the full support prediction of Theorem 1 still holds when the assumptions of strict concavity of P and strict convexity of C are dropped. The theoretical prediction is that the invariant distribution $\mu^* : \Gamma \longrightarrow [0,1]$ has full support on $[q^C, q^W]$ and is identically zero outside of this interval (identifying each q with the corresponding monomorphic state). Formally, $\mu^*(q) > 0 \; \forall q \in [q^C, q^W]$ and $\mu^*([q^C, q^W]) = 1$. Recall that using the Ergodic Theorem we can approximate μ^* through the time averages obtained in our simulations. Specifically, for ε small enough the restriction of the relative frequency distribution $f : \Gamma \cup \{\text{res}\} \longrightarrow [0,1]$ to Γ converges to the theoretical limit distribution μ^*. Although our computational results are based on low experimentation probabilities, this limit result can of course not be obtained for non-vanishing ε as it is the case in our simulations, hence we will have $f([q^C, q^W]) < 1$.

Table 2.4: Pseudo-code of the N-player learning model.

Set parameters: P, C, N, ε, M, and K.
Generate discrete grid Γ (with cardinality $|\Gamma| = 7M + 1$): Determine step size as $\delta = \frac{q^W - q^C}{M}$. For $q \in \mathbb{R}_+$ set $\Gamma(q) = \{q, q+\delta, \dots, q+7M\delta\}$. Set $q_0 = q^C - \lfloor \frac{q^C}{\delta} \rfloor \delta$. For $k \geq 0$: If $q^W \leq q_k + 6M\delta$, set $\Gamma = \Gamma(q_k)$ and stop the procedure. Otherwise, set $q_{k+1} = q_k + M\delta$.
Period 0: An initial population profile $q(0) \in \Gamma^N$ is selected where each $q_i(0)$, $i = 1, \dots, N$, is randomly drawn from the uniform distribution over Γ.
Period t:
- Aggregate quantity is determined as $Q(t) = \sum_{i=1}^{N} q_i(t)$.
- Determine individual payoffs as $\Pi_i(t) = P(Q(t))q_i(t) - C(q_i(t))$.
- Determine set of best-performing strategies in the last $K + 1$ periods as

$$\begin{aligned} B_t^K = &\{q_i(k) \mid i \in \{1, \dots, N\}, k \in \{t-K, \dots, t\} \text{ and } \Pi_i(k) \geq \Pi_j(k') \\ &\forall j = 1, \dots, N, \ \forall k' = t-K, \dots, t\}. \end{aligned}$$

- For each $i = 1, \dots, N$ draw $\varepsilon_i \in B(1, \varepsilon)$ (Bernoulli distribution).
- If $\varepsilon_i = 0$, agent i randomly imitates one of the strategies in B_t^K (uniform), that is $q_i(t+1) \sim U(B_t^K)$.
- If $\varepsilon_i = 1$, agent i experiments and chooses a strategy at random (uniform) from the whole grid Γ, that is $q_i(t+1) \sim U(\Gamma)$.
- The population profile for next period is $q(t+1) = (q_i(t+1))_{i=1}^N$.
- If $t + 1 > 1{,}000{,}000$ or if $t > 100{,}000$ and $f_t([q^C, q^W])/(1 - f_t(\text{res})) > 0.99$ (where $f_t : \Gamma \cup \{\text{res}\} \longrightarrow [0, 1]$ is the relative frequency distribution over monomorphic states up to period t), then stop. Otherwise, increase the period counter t and proceed to next period.

Notes: N-player learning model with memory K, grid coarseness M, mutation probability ε, cost function C, and inverse-demand function P.

The fraction of time spent at monomorphic states with quantities within the interval $[q^C, q^W]$, formally defined as $\sum_{q \in [q^C, q^W] \cap \Gamma} f(q)$ and denoted by $f([q^C, q^W])$, will serve us as a measure of convergence towards the theoretical result. The closer the values of $f([q^C, q^W])$ are to one the more likely it is that the approximated distribution μ^* will indeed have full support between q^C and q^W. The left-hand part of Table 2.5 presents the average value of $f([q^C, q^W])$ across different memory lengths and noise levels. For the right-hand part we have calculated the average value of $f([q^C, q^W])$ for each of the

Table 2.5: Overview average and minimum fraction of time spent within the interval $[q^C, q^W]$.

	$\mathbf{mean}(f([q^C, q^W]))$			$\mathbf{min}_{(P,C)} f_{P,C}([q^C, q^W])$		
K	**low** ε	**med** ε	**high** ε	**low** ε	**med** ε	**high** ε
1	.97134	.90418	.82716	.96578	.89475	.81348
5	.97445	.91769	.84797	.96880	.90508	.82969
10	.97445	.91731	.84429	.96919	.90587	.82667
50	.97417	.90076	.78764	.96912	.87468	.67634

Notes: Left-hand panel shows the average fraction of time spent within the interval $[q^C, q^W]$ across all 30,240 simulation runs split by memory length K and experimentation probability ε. For the right-hand panel we computed for each (P, C) combination the average proportion of time spent within the interval $[q^C, q^W]$ across runs and report the minimum of those averages over all 72 (P, C) configurations. Low $\varepsilon = 0.001$, med $\varepsilon = 0.005$, high $\varepsilon = 0.010$.

72 (P, C) combinations, denoted by $\bar{f}_{P,C}([q^C, q^W])$, and we report the minimum of those averages, i.e. $\min_{(P,C)} \bar{f}_{P,C}([q^C, q^W])$, over the 72 (P, C) pairs for different memory lengths and different probabilities of experimentation.

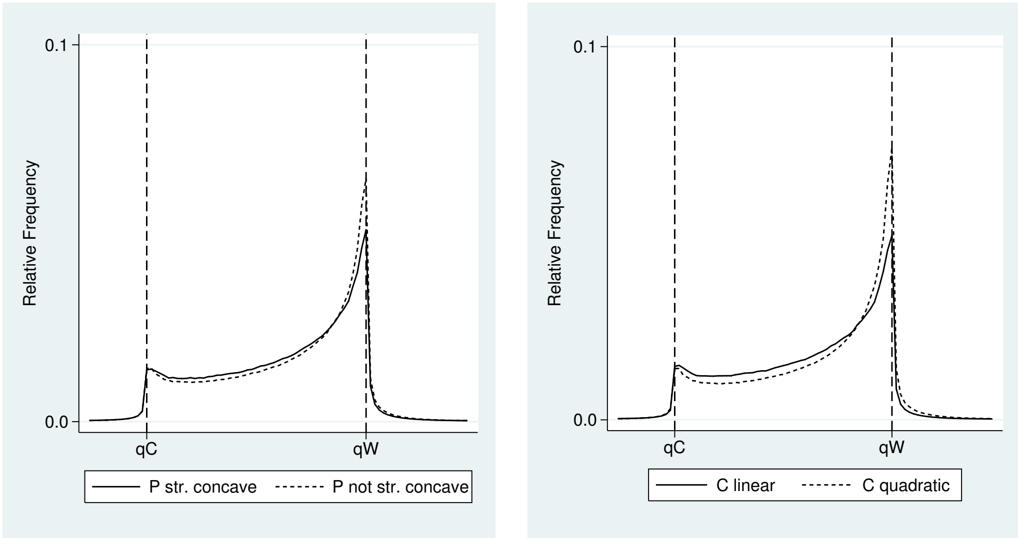
Our results clearly show that $f([q^C, q^W])$ approaches one as the probability of experimentation decreases. For high, medium, and low levels of ε we obtain average values of $f([q^C, q^W])$ that are always above 0.97, 0.90, and 0.78, respectively. For low ε the minimal value of $f([q^C, q^W])$ across all 72 combinations of inverse-demand and cost functions never drops below 0.96 independently of the length of memory. We interpret this as first evidence that the result of convergence towards full support on $[q^C, q^W]$ holds beyond the case of strictly concave inverse-demand and strictly convex cost.

2.4.2 The Shape of the Invariant Distribution

We want to have a closer look at the shape of the estimated invariant distribution, in particular over quantities within the main range of interest between the Cournot and the Walrasian outcomes. Of course, for the simulations the process will still spend at least some time at non-monomorphic states, that is $f(\text{res}) > 0$. Clearly a higher noise level leads to more experimentations, but the same is true for more firms (for a given level of ε) because each firm experiments with the same probability, hence the likelihood for a single experiment increases. Hence the weight placed on non-monomorphic states $f(\text{res})$ will be increasing in ϵ and N due to an increased likelihood of mutations. The consequence is a general level effect, that is, non-monomorphic states occur more often for larger values of ε and N. To control for this effect, we focus in our analysis on the restriction of the relative frequency distribution to Γ, which we denote by $F : \Gamma \longrightarrow [0, 1]$. That is, we consider the relative frequency of monomorphic states conditioning on the total number of monomorphic states. As a second step, in order to aggregate the relative frequency distributions, e.g. over different pairs of (P, C), we normalize F by identifying each quantity $q \in \Gamma$ with its position within the grid Γ relative to q^C for a fixed grid coarseness M. Specifically, for a given value of M, we consider the set of indices $\{0, \ldots, 2.5M + 1\}$ where the first $M/2$ indices are quantities below q^C, the last M are quantities above q^W, and the remaining $M + 1$ correspond to the quantities $q^C, q^C + \delta, q^C + 2\delta, \ldots, q^W$, where $\delta = \frac{q^W - q^C}{M}$. To accomplish this, for each quantity q in the interval $[q^C - \frac{q^W - q^C}{2}, q^W + (q^W - q^C)]$ we assign the index $\iota_M(q) = \frac{M}{2} + k$, where k is such that $q = q^C + k\frac{q^W - q^C}{M}$. Hence the quantities q^C and q^W always have the indices $M/2$ and $3M/2$, respectively. This allows us to compare the so-obtained normalized relative frequency distribution, denoted by $\hat{F}_M$ (or abusing notation simply by $\hat{F}$), across different values of P, C, δ, and N, although the quantities of q^C and q^W clearly vary with all four of those variables.

Of course, the exact shape of the distribution $\hat{F}$ and in particular $\hat{F}(q^C)$ and $\hat{F}(q^W)$ vary across P_1 to P_{12} and C_1 to C_6. All individual plots, however, share a number of general features which can be illustrated by averaged plots across appropriate subsets of simulations. Before we proceed to a formal analysis of the data, it is worth to look at these illustrative representations. Figures 2.1 to 2.3 plot the normalized relative frequency distribution $\hat{F}_M$ averaged over several subsamples split along the dimensions strictly concave or not strictly concave inverse-demand function (dummy; Fig. 2.1a), linear or quadratic cost function (dummy; Fig. 2.1b), memory (K; Fig. 2.2), and

Figure 2.1: Effect of strict concavity of P and strict convexity of C.



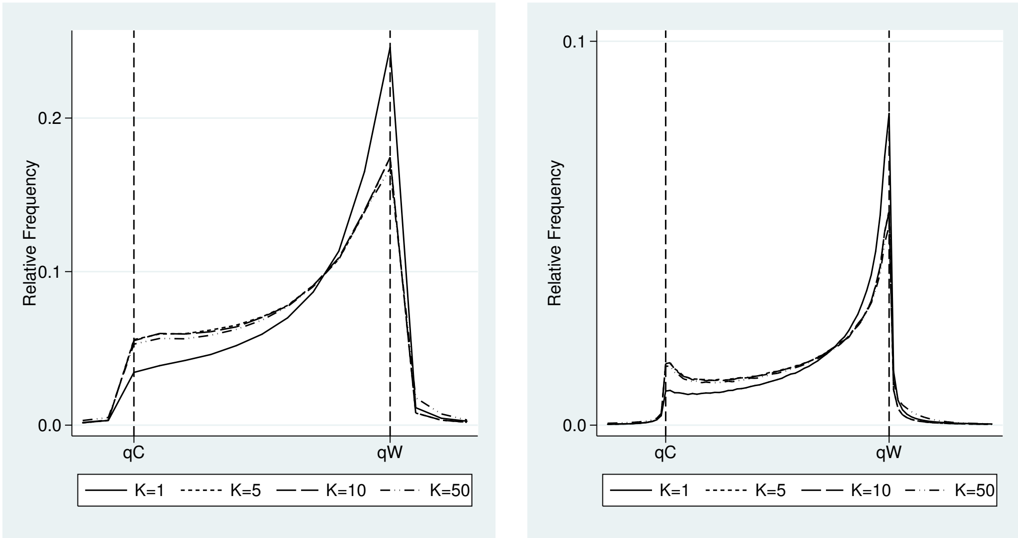
(a) Effect of P strictly concave. (b) Effect of C linear.

Notes: Normalized relative frequency distribution $\hat{F}$ centered around $[q^C, q^W]$. Subsamples are limited to the finest grid with the highest number of steps $M = 50$.

number of firms (N; Fig. 2.3). The normalized relative frequency distributions $\hat{F}_M$ look qualitatively very similar across the different grid coarseness values M (compare, e.g., the two panels of Fig. 2.2), hence in several of the figures we will present only the graphs for $M = 50$ (the regression analysis will of course rely on the whole data set).

The first observation is that the qualitative shape of the normalized relative frequency distribution is quite similar across the different subsamples (Fig. 2.1 to Fig. 2.3). In the unstable region outside the interval $[q^C, q^W]$ the value of the distribution $\hat{F}$ is close to zero, while it is well above zero for all values between the Cournot and the Walrasian quantities. In general the process spends the largest amount of time at the Walrasian quantity across all specifications, indicating that the Walrasian outcome is a robust long-term prediction in the sense that it is stable independently of the length of memory (as long as the latter is non-trivial), the number of firms, the specific inverse-demand function, and the specific cost function used in the underlying oligopoly game. The general shape is bimodal and generally convex,

Figure 2.2: Effect of memory.



(a) Effect of memory K for $M = 10$. (b) Effect of memory K for $M = 50$.

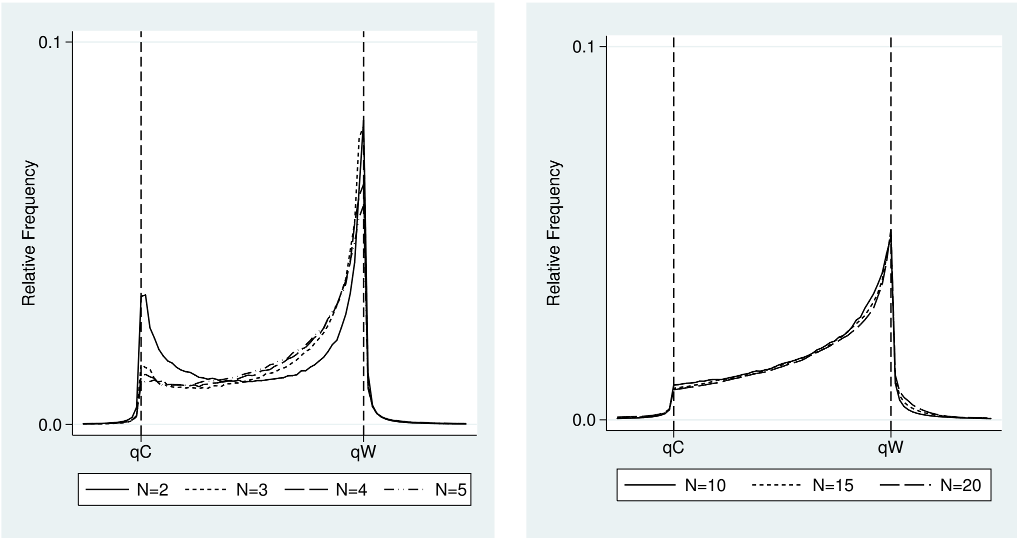
Notes: Normalized relative frequency distribution $\hat{F}$ centered around $[q^C, q^W]$. For the left figure the subsample is limited to $M = 10$, while for the right figure it is limited to $M = 50$.

with a large peak at the Walrasian quantity and another at or close to the Cournot quantity.

The second observation concerns the stability of the prediction of Theorem 1. Fig. 2.1a compares the frequency distributions of simulations with strictly concave inverse demand functions and those with not strictly concave ones. Fig. 2.1b compares the frequency distributions averaged across simulations with strictly convex cost functions and those with linear ones. These figures confirm that $\hat{F}$ approaches a full support on $[q^C, q^W]$. This obviously had to be the case for the combinations (P, C) covered by Theorem 1, but the result is also confirmed for non-concave inverse-demand and quadratic cost functions. That is, this general prediction seems unaffected by different specifications of P and C. This provides a further positive answer to the question of whether the result of Alós-Ferrer (2004) holds beyond the class of Cournot oligopolies considered in that work.

The third observation is that, even at this descriptive level, it can be readily observed that changes in the key variables do influence the relative

Figure 2.3: Effect of number of firms N.



Notes: Normalized relative frequency distribution $\hat{F}$ centered around $[q^C, q^W]$. Subsamples are limited to the finest grid with the highest number of steps $M = 50$.

weights on q^C and q^W, although the general shape of the distribution remains unaffected. As can be seen in Fig. 2.2, short memory ($K = 1$) clearly favors the Walrasian outcome, while longer memory shifts weight towards the Cournot quantity. This result confirms the intuition that longer memory makes intertemporal comparisons more important. Those comparisons essentially rely on absolute payoffs, hence longer memory benefits the Cournot quantity. This figure also includes different panels for $M = 10$ and $M = 50$ illustrating that the general characteristics of the distribution are not affected by the step size discretization.[8]

Another key variable is the number of firms, N. Fig. 2.3 plots the frequency distributions as a function of N and shows that there is also a clear effect. In the duopoly case, a much larger weight is placed on the Cournot quantity q^C compared to the settings with $N > 2$. That is, coordination on the Cournot equilibrium is more likely in the duopoly case, where strategic complexity is lowest. For more than two firms, however, there is no clear

[8]There is, of course, a purely mechanical "level effect:" for a smaller value of M, the individual values of $\mu^*(q)$ will be higher, since they capture weight which will be spread across several consecutive quantities for a larger M (finer grid).

Table 2.6: Overview average quantity and median quantity.

P	C	$m(\hat{F})$	med$(\hat{F})$	P	C	$m(\hat{F})$	med$(\hat{F})$
P_1	C_1	.675	.735	P_7	C_1	.675	.735
P_1	C_2	.676	.735	P_7	C_2	.678	.738
P_1	C_3	.676	.734	P_7	C_3	.680	.741
P_1	C_4	.687	.750	P_7	C_4	.685	.747
P_1	C_5	.689	.753	P_7	C_5	.693	.759
P_1	C_6	.687	.750	P_7	C_6	.705	.775
P_2	C_1	.672	.731	P_8	C_1	.671	.728
P_2	C_2	.672	.730	P_8	C_2	.671	.729
P_2	C_3	.674	.732	P_8	C_3	.676	.736
P_2	C_4	.686	.750	P_8	C_4	.847	.900
P_2	C_5	.681	.744	P_8	C_5	.854	.909
P_2	C_6	.694	.762	P_8	C_6	.852	.909
P_3	C_1	.673	.731	P_9	C_1	.674	.736
P_3	C_2	.673	.731	P_9	C_2	.678	.738
P_3	C_3	.670	.726	P_9	C_3	.679	.740
P_3	C_4	.687	.750	P_9	C_4	.687	.751
P_3	C_5	.680	.741	P_9	C_5	.708	.778
P_3	C_6	.693	.758	P_9	C_6	.706	.777
P_4	C_1	.686	.748	P_{10}	C_1	.673	.730
P_4	C_2	.684	.745	P_{10}	C_2	.675	.730
P_4	C_3	.688	.751	P_{10}	C_3	.677	.737
P_4	C_4	.681	.743	P_{10}	C_4	.675	.733
P_4	C_5	.701	.766	P_{10}	C_5	.684	.746
P_4	C_6	.697	.764	P_{10}	C_6	.692	.759
P_5	C_1	.670	.726	P_{11}	C_1	.670	.727
P_5	C_2	.674	.735	P_{11}	C_2	.673	.731
P_5	C_3	.681	.743	P_{11}	C_3	.676	.736
P_5	C_4	.759	.841	P_{11}	C_4	.677	.737
P_5	C_5	.772	.855	P_{11}	C_5	.692	.755
P_5	C_6	.778	.860	P_{11}	C_6	.693	.758
P_6	C_1	.669	.727	P_{12}	C_1	.671	.729
P_6	C_2	.675	.736	P_{12}	C_2	.684	.746
P_6	C_3	.682	.745	P_{12}	C_3	.673	.731
P_6	C_4	.729	.807	P_{12}	C_4	.695	.762
P_6	C_5	.743	.824	P_{12}	C_5	.706	.777
P_6	C_6	.758	.840	P_{12}	C_6	.721	.794

Notes: Average quantity $m(\hat{F})$ and median quantity med$(\hat{F})$ for the distribution $\hat{F}$ on the subsamples corresponding to the 72 combinations of P and C. Inverse-demand functions P are as in Table 2.1. Cost functions C are as in Table 2.2.

pattern. The intuitive reason for a larger weight on q^C in the duopoly case is that, if there are just two firms, a downward transition within the stable interval $[q^C, q^W]$ can readily be achieved by both firms simultaneously experimenting with a smaller quantity. Such a move yields a gain in terms of absolute payoff, and with only two firms there is no firm left at the original quantity, hence at least one of the firms is also better off in relative terms. For $N > 2$ a similar transition would require simultaneous mutations by at least three firms and is therefore very unlikely in comparison. Whereas a downward transition with only two mutations is also possible for $N > 2$, it is more complex than in the duopoly case requiring a series of transitions each involving simultaneous mutations by two firms.[9] Although both transitions occur with positive probability in the limit, our results show that this "difference in complexity" is reflected in the shape of the invariant distribution, as demonstrated by the larger weight on q^C for the duopoly.

We now shift to more analytical measures of the shape of the normalized relative frequency distribution $\hat{F}$, and of how it changes across the different combinations of (P, C). To this purpose, we construct two measures describing the main features of $\hat{F}$, which we will also use in the regression analysis below. First, we define the average quantity for a distribution $\hat{F}$ as

$$m(\hat{F}) = \frac{\sum_{q \in [q^C, q^W]} (\iota(q) - \iota(q^C)) \hat{F}(q)}{(\iota(q^W) - \iota(q^C)) \sum_{q \in [q^C, q^W]} \hat{F}(q)}$$

where $\iota(q)$ is the index of q within the grid Γ. The value of $m(\hat{F})$ indicates the position of the average quantity within $[q^C, q^W]$ weighted by its relative frequency according to $\hat{F}$. For higher values of $m(\hat{F})$ the distribution $\hat{F}$ puts more mass on quantities closer to q^W. For example $m(\hat{F}) = 0$ means $\hat{F}(q) = 0$ for all $q \in (q^C, q^W]$, i.e. full mass on q^C, while $m(\hat{F}) = 1$ would imply $\hat{F}(q) = 0$ for all $q \in [q^C, q^W)$, i.e. full mass on q^W. Second, we define

[9] Indeed the actual proof in Alós-Ferrer (2004) for the duopoly case was handled differently than the case $N > 2$.

the median quantity within $[q^C, q^W]$ for a distribution $\hat{F}$ as

$$\text{med}(\hat{F}) = \iota(\min\{q \in [q^C, q^W] \mid \sum_{q' \leq q} \hat{F}(q') \geq \frac{1}{2} \sum_{q \in [q^C, q^W]} \hat{F}(q)\}).$$

The median $\text{med}(\hat{F})$ is the index $\iota(q)$ of the smallest quantity $q \in [q^C, q^W]$ such that at least half of the mass of $\hat{F}$ within $[q^C, q^W]$ is on values weakly smaller than q. Table 2.6 summarizes the shape of the distribution $\hat{F}$ for the 72 possible combinations of (P, C) by means of these two measures. It illustrates that there is considerable variance with regard to the shape of the distribution as $m(\hat{F})$ varies from a minimum of 0.669 to a maximum of 0.853 over the different specifications of P and C. However, the most important observation is that both $m(\hat{F})$ and $\text{med}(\hat{F})$ are consistently above 0.6. This indicates that the distribution is highly skewed towards the Walrasian outcome, confirming the qualitative features observed in Fig. 2.1 to Fig. 2.3.

2.4.3 Regression Analysis

The shape of the relative frequency distribution is strongly influenced by the choice of a specific pair (P, C). However, to obtain a systematic relation describing the effect of a change from a pair (P, C) to another pair (P', C') one would need to order those along meaningful dimensions. In absence of those, it is more instructive to focus on variables that come with a clear-cut order and therefore can serve as an effective means to guide our intuition on how they influence the relative frequency distribution. We seek to investigate how the shape of the distribution $\hat{F}$ is affected by the following variables: memory length K, number of firms N, concavity of P, and linearity of C.

Table 2.7 shows the results of a linear regression on the measures $m(\hat{F})$ and $\text{med}(\hat{F})$ including the aforementioned dimensions as independent variables. As Fig. 2.3 suggests a special role of the duopoly we also include a dummy variable for the duopoly in the regression. The results indicate that both mean quantity $m(\hat{F})$ and median quantity $\text{med}(\hat{F})$ are significantly decreasing in the length of memory K. This result is in line with the prediction that longer memory K allows firms to focus more on intertemporal

Table 2.7: Linear regressions on average and median quantity.

	$m(\hat{F})$	$\text{med}(\hat{F})$
K	-0.00066^{***}	-0.00091^{***}
	(0.00002)	(0.00003)
Duopoly	-0.08491^{***}	-0.11758^{***}
	(0.00124)	(0.00164)
N	-0.00109^{***}	-0.00275^{***}
	(0.00007)	(0.00009)
P_{sconcave}	-0.02281^{***}	-0.02858^{***}
	(0.00091)	(0.00120)
C_{linear}	-0.04354^{***}	-0.05767^{***}
	(0.00079)	(0.00104)
ε	3.55312^{***}	2.64964^{***}
	(0.10738)	(0.14149)
Observations	30240	30240

Notes: Standard errors in parentheses. * $p < 0.001$, ** $p < 0.0001$, *** $p < 0.00001$. Linear regressions with dependent variables $m(\hat{F})$ and $\text{med}(\hat{F})$. Dummy variable Duopoly $= 1$ if $N = 2$. Dummy variable $P_{\text{sconcave}} = 1$ if P is strictly concave. Dummy variable $C_{\text{linear}} = 1$ if C is linear.

comparisons. Those rely more heavily on absolute payoffs, and hence the distribution of outcomes is shifted in the direction of the Cournot quantity, although most of the weight remains closer to the Walrasian one (recall Table 2.6).

The coefficients for the duopoly dummy as well as for N show a significant negative effect. For the duopoly the distribution places more weight on states closer to the Cournot quantity compared to situations with more than two firms, which confirms the observation in Fig. 2.3 of a larger weight on q^C for the duopoly. In contrast, beyond the duopoly case increasing the number of firms slightly shifts weight from Walras to Cournot. This effect, however, might simply be due to a higher level of noise. For a larger number of firms the probability of simultaneous mutations increases (for a given level of ε), and this results in an overall flatter distribution. As the weight on q^W is generally larger than that on q^C this "mechanical" flattening of the distribution can explain this weak negative trend beyond the duopoly.

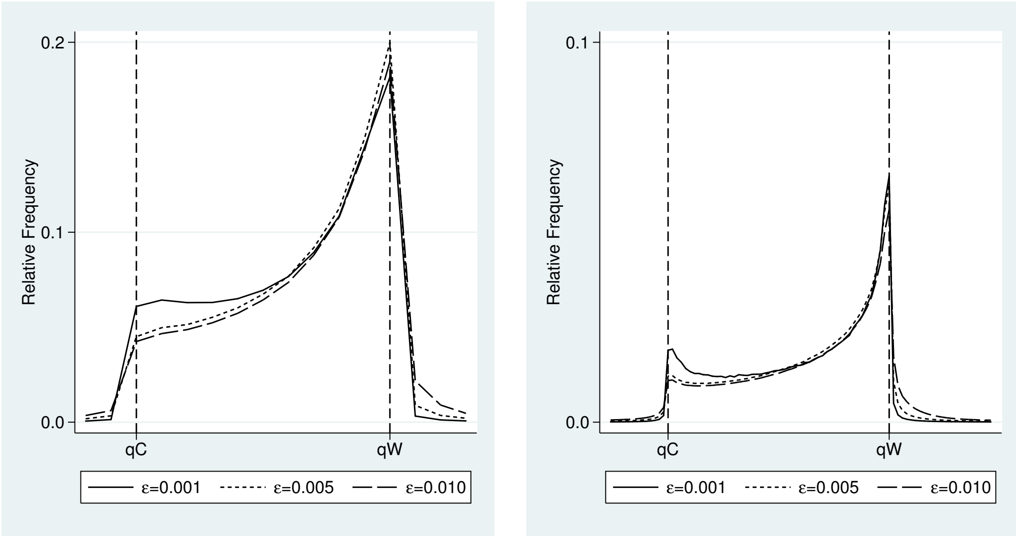
The dummies for strictly concave inverse demand functions and linear cost functions are significant and negative. This indicates that both classes of functions place more weight on states closer to the Cournot outcome and thus result in a distribution that is less skewed towards the Walrasian outcome. That is, although the prediction from Theorem 1 holds beyond the class of well-behaved oligopolies, it seems that, if inverse demand functions are strictly concave or costs are linear, states closer to the Cournot outcome will be observed more frequently.

Last, the regressions in Table 2.7 also control for the level of noise ($\varepsilon \in \{0.001, 0.005,\ 0.010\}$). The coefficients are significantly positive, which indicates that for vanishing noise, more weight is shifted in the direction of the Cournot quantity. Fig. 2.4 clarifies this effect. While the shape of the distribution remains qualitatively unchanged, the spike at q^C becomes larger as ε becomes smaller. However, this effect has to be understood in relative terms: across all our simulations, the weight close to the Walrasian quantity always remains larger than the weight close to the Cournot quantity.

We also seek to examine the effects of memory length or the number of firms in the market on convergence, as proxied by the percentage of time that the system spends in the interval $[q^C, q^W]$, and on the invariant distribution's weights on q^C and q^W. Since those quantities are frequencies, we turn to fractional logit regressions (see Papke and Wooldridge, 2008). Table 2.8 shows the results of fractional logit regressions. Dependent variables are the normalized relative frequencies of q^C and q^W, the mass strictly between q^C and q^W (excluding $\hat{F}(q^C)$ and $\hat{F}(q^W)$), and the mass within the interval $[q^C, q^W]$.

The fractional logit regressions confirm again that violating the main assumptions made in Alós-Ferrer (2004), namely strictly concave P or strictly convex C, does not negatively affect convergence towards full support on $[q^C, q^W]$ as captured by $\hat{F}([q^C, q^W])$. Allowing for non-concave inverse-demand functions does not result in a significant decrease on the average time the process spends on quantities between Cournot and Walras. For linear cost functions convergence even increases significantly compared to quadratic cost functions. Thus the prediction of Theorem 1 seems to hinge neither on the

Figure 2.4: Effect of noise.



(a) Effect of noise ε for $M = 10$. (b) Effect of noise ε for $M = 50$.

Notes: Normalized relative frequency distribution $\hat{F}$ centered around $[q^C, q^W]$. For the left figure the subsample is limited to $M = 10$, while for the right figure it is limited to $M = 50$.

strict concavity of P nor on the strict convexity of C. However, as we have already seen from Table 2.7 the characteristics of the inverse-demand function and the cost function affect the shape of the invariant distribution, it is less skewed towards q^W for strictly concave P and linear C.

For linear cost functions the distribution shifts more weight towards the Cournot outcome (see also Fig. 2.1b), while at the same time the weight on the Walrasian quantity decreases slightly. We seek to identify what drives this shift of weight from Walras to Cournot. To that end, consider a monomorphic state where all N firms produce the quantity $q^* < q^W$ and a single mutant firm experiments with a larger quantity, q, closer to the Walrasian quantity, that is, $q^* < q < q^W$. The chance of success of such an upwards deviation depends on whether the mutant fares better than the incumbents after the deviation has occurred. Denote the profit of a firm producing q when the total output in the market is Q by $\Pi(q|Q) = qP(Q) - C(q)$. Then, the post-deviation payoff of the mutant is $\Pi(q \mid q + (N-1)q^*)$, whereas the post-deviation payoff of the non-mutants is $\Pi(q^*|q + (N-1) + q^*)$. The

Table 2.8: Fractional logit regressions.

	$\hat{F}(q^C)$	$\hat{F}(q^W)$	$\hat{F}((q^C, q^W))$	$\hat{F}([q^C, q^W])$
K	0.00351^{***}	-0.00602^{***}	-0.00036	-0.00893^{***}
	(0.00021)	(0.00023)	(0.00017)	(0.00019)
Duopoly	0.85686^{***}	0.08372^{***}	-0.33407^{***}	-0.15500^{***}
	(0.01240)	(0.01302)	(0.00983)	(0.01164)
N	-0.02021^{***}	-0.01201^{***}	-0.02251^{***}	-0.08230^{***}
	(0.00078)	(0.00093)	(0.00058)	(0.00051)
P_{sconcave}	-0.00043	-0.21172^{***}	0.13788^{***}	0.02762^{***}
	(0.00986)	(0.00943)	(0.00676)	(0.00608)
C_{linear}	0.07979^{***}	-0.37862^{***}	0.28314^{***}	0.18333^{***}
	(0.00879)	(0.00882)	(0.00626)	(0.00612)
ε	-50.89090^{***}	-5.33708^{**}	-52.48244^{***}	-192.65921^{***}
	(1.25759)	(1.21091)	(0.88524)	(0.92550)
Obs.	30240	30240	30240	30240

Notes: Standard errors in parentheses. * $p < 0.001$, ** $p < 0.0001$, *** $p < 0.00001$. Dummy variable Duopoly = 1 if $N = 2$. Dummy variable $P_{\text{sconcave}} = 1$ if P is strictly concave. Dummy variable $C_{\text{linear}} = 1$ if C is linear. $\hat{F}((q^C, q^W)) = \sum_{q \in (q^C, q^W) \cap \Gamma} \hat{F}(q)$ is the mass that is strictly between q^C and q^W.

relative payoff advantage of an upward deviation to $q > q^*$ is thus $\Pi(q \mid q+(N-1)q^*) - \Pi(q^* \mid q+(N-1)q^*) = (q-q^*)P(q+(N-1)q^*) - (C(q)-C(q^*))$. In our simulations we used two types of cost functions, linear functions of the form $C(q) = aq$ and quadratic functions of the form $C(q) = aq^2$ with the same coefficients (recall Table 2.2). As an illustration, a pairwise comparison for a fixed coefficient a shows that a deviation upwards is more attractive in relative terms for quadratic costs if $a(q^2 - (q^*)^2) < a(q - q^*)$, which is equivalent to $q + q^* < 1$ (recall that we normalized $q_{\max}$ to 1). As a result a firm experimenting with a larger quantity $q > q^*$ makes larger profits than the incumbent firms for both types of cost functions, however, for $q + q^* < 1$ the increase in costs is larger for linear costs than for quadratic costs, and thus dampens the additional profits. Of course, for $q + q^* > 1$ the opposite holds, however, for our choices of P and C most values of q^W are below 1/2.[10] Moreover, a similar argument shows that an upward deviation

[10]For our parameter combinations the Walrasian quantity q^W is smaller than 1/2 for all

towards q^W is also more likely to generate an absolute payoff advantage compared to the previous period, which increases the likelihood that this deviation is successful. Therefore deviations upwards in the direction of q^W are more attractive in both relative and absolute terms if C is quadratic, which potentially explains the lower weight on q^W for linear cost functions.

We can also confirm our previous observation that the relative weight on q^C is increasing in memory length, while it has the opposite effect on q^W. That is, longer memory shifts some weight within the invariant distribution towards quantities closer to the Cournot one, although most of the weight remains closer to the Walrasian equilibrium.

Regarding the effect of the number of firms, the regressions show that the time spent at both q^C and q^W is decreasing in N. However, for $N > 2$ the interior mass does not increase significantly for higher N, but instead it even decreases. This indicates that beyond the duopoly the effect of N on the weight on the Cournot and Walras quantities is rather mechanical: as the number of firms increases, the probability of simultaneous, successful deviations to quantities outside the interval $[q^C, q^W]$ is higher, leading to a relative decrease of the weight on *all* states within the interval. However, for the duopoly we observe larger weights on both q^C and q^W while, at the same time, the interior mass $\hat{F}((q^C, q^W))$ decreases significantly compared to situations with more than two firms. This effect on the relative weight is especially pronounced for q^C, which is in line with our previous observation that for the duopoly the distribution shifts more weight towards the Cournot quantity.

Last, the regressions also show that a larger level of noise increases the weight on states outside the stable interval $[q^C, q^W]$ and decreases the weight at all monomorphic states within $[q^C, q^W]$. This is a trivial effect: the presence of higher noise increases the lower bound for the weight on every monomorphic state (for positive epsilon the invariant distribution is irreducible) resulting in a general flattening of the distribution. As remarked before, the predictions of stochastic stability refer to the vanishing-noise limit and have to be tested with small values of ε. For larger values, all one can

but four cases, and only slightly above 1/2 in those four.

learn is "noise in, noise out."

2.5 Discussion

This paper makes two main contributions. First, using computational simulations we show that the main prediction of Alós-Ferrer (2004) – full support of the limit invariant distribution on $[q^C, q^W]$ for $K > 0$ – holds beyond the set of well-behaved Cournot games used in that article. We provide evidence that this result is more general than previously shown. Specifically, our simulations suggest that the result holds for a wide spectrum of inverse-demand and cost functions which do not necessarily satisfy the assumptions of Alós-Ferrer (2004).

Second, we shed light on the exact shape of the limit distribution for imitative dynamics with memory in Cournot games, and how it is affected by different variables such as memory length and the number of firms. It turns out that the limit distribution is bimodal, with peaks at the Cournot and Walras extremes, but highly skewed towards the Walrasian quantity. This result is surprisingly robust across a large number of different specifications. Although longer memory increases the importance of the Cournot equilibrium the competitive outcome remains the dominant quantity. Throughout all specifications the process spends most time on the Walrasian quantity. Interestingly, the Cournot quantity is most attractive in a duopoly setting, while for more than three firms the exact number of competitors apparently has only minor influence on the relative weights on Cournot and Walras. Overall, the main message of our analysis is that, even when one considers more realistic behavior than in Vega-Redondo (1997) and less restrictive assumptions than in Alós-Ferrer (2004), relative-performance concerns and imitative behavior in quantity-setting markets will generally lead to increased competitiveness in market outcomes, even if the expected behavior is captured by a rich dynamics rather than a point prediction.

In summary, we have used simulations to numerically "prove" that the set of stochastically stable states derived in Alós-Ferrer (2004) is unchanged for more general Cournot oligopolies, while at the same time exploring a number of natural dimensions as the number of firms or the convexity of costs. Of course, simulations do not provide analytical results and are not meant to replace them. If analytical results are not feasible, however, simulations are clearly an efficient way to make progress and open new directions for theory development. In this sense, we hope that the work presented here can serve as an illustration of how computer simulations can serve as a useful tool for the analysis of dynamic, stochastic learning models where agents are endowed with behavioral rules.

Chapter 3
Cognitive Sophistication and Deliberation Times

3.1 Introduction

Cognitive capacities differ among individuals. Inter-individual differences in sophistication and cognitive effort have been put forward as an explanation for disparities in observed behavior, e.g. differences in the effectiveness of incentives and differences in cognitive ability. This gave rise to a rich theoretical literature that endows individuals with differing degrees of strategic sophistication or reasoning capability. Prominent models, such as the level-k model (Stahl, 1993; Nagel, 1995; Stahl and Wilson, 1995) and models of cognitive hierarchies (Ho et al., 1998) are build on the assumption that heterogeneity in depth of reasoning is the source of individual differences in behavior. These models have proven to perform well in describing observed behavior, however, there is also some recent behavioral evidence (Goeree et al., 2016) that is inconsistent with most models of iterative thinking. Reconciling behavior in their experiment with a model of iterative thinking, such as level-k, would require inordinate high levels compared to what is usually observed in the literature. This, at least, casts some doubt on whether these models can be understood as procedural, describing how decisions are arrived at, or whether they should rather be understood as purely descriptive, outcome-based models. Studying this question requires an individually measurable correlate of cognitive effort, we argue that deliberation times can be purposefully used as such a measure.

The use of response times is well-established in (cognitive) psychology as a tool to help understand decision processes in the human brain, that is, how decisions are made.[1] A well-studied stylized fact in that literature

[1]For a recent discussion of the benefits, challenges, and desiderata of response time analysis in experimental economics see Spiliopoulos and Ortmann (2017).

is that the human ability to discriminate between two stimuli is a function of the difference between the respective stimuli. With increasing difference the mean response time decreases, or in other words, decisions closer to indifference are found to be slower (Dashiell, 1937; Mosteller and Nogee, 1951; Moyer and Landauer, 1967). It is important to note that response times in that literature are usually very short (below one second), and hence have to be distinguished from deliberation times (or decision times) in more complex tasks, as the ones used to study iterative thinking. However, there is some recent evidence that this distance-to-indifference effect can also be found for longer decision times, to which we refer as deliberation times (Krajbich et al., 2014, 2015).

So far there has been little direct evidence that depth of reasoning (level of thinking) corresponds to cognitive effort. Most of the experimental literature has used observed choices to classify individuals in different cognitive-reasoning categories or types. This way, choice data is used to make inferences regarding the processes that lead to a particular choice. The problem with this approach is that the same choice is always attributed to the same level, although it might very well be the result of completely different decision rules. Choice data alone is not sufficient to distinguish such cases and, hence, additional data is necessary to make better inferences regarding the underlying processes. We argue that deliberation times can provide such evidence.

There is a growing literature employing other sources of evidence suggesting that individuals follow step-wise reasoning processes (e.g., Coricelli and Nagel, 2009; Brañas-Garza et al., 2012; Gill and Prowse, 2016). Those works show that reasoning in the beauty contest game (Nagel, 1995), which is the workhorse in the literature on iterative thinking, correlates with neural activity in areas of the brain associated with mentalizing (Theory of Mind network) and relate higher cognitive ability (as measured, e.g., by the Cognitive Reflection Test or the Raven test) with more steps of reasoning. Others have used click patterns recorded via MouseLab and eye-tracking to obtain information on search behavior, which is then used to make inferences regarding level-k reasoning (Costa-Gomes et al., 2001; Crawford and Costa-

Gomes, 2006; Polonio et al., 2015). Lindner and Sutter (2013) find that under time pressure behavior in the 11-20 game is, perhaps coincidentally, closer to the Nash equilibrium. Recently, Alaoui and Penta (2016b) have incorporated deliberation times into a model of endogenous depth of reasoning, with the key assumption being that each additional step of reasoning increases deliberation times. Gill and Prowse (2017) use deliberation times to measure strategic complexity in a repeated p-beauty contest game and show that “overthinking,” that is, thinking longer than usual, is detrimental to performance.

We provide a simple model linking cognitive sophistication and deliberation times, taking into account stylized facts from the psychophysiological literature on response times. In our model we view the total deliberation time of an observed choice as the sum of “decision times” for a chain of binary “hypothetical choices” as implicitly postulated in the literature on iterative thinking, and explicitly assumed in Alaoui and Penta (2016b). The key assumption of the model is then that the time required for each step is a decreasing function of the distance to indifference, as captured by the potential gain of conducting an additional step of reasoning. The model provides empirically testable predictions regarding the relation of deliberation times, choices and cognitive sophistication, as well as regarding the effects of incentives on both the level of cognitive sophistication, as revealed by choices, and the psychophysiological correlate embodied in deliberation times. We then test the predictions in an experiment using two different games commonly used to study iterative thinking: the beauty contest game (or guessing game) (Nagel, 1995), which is the workhorse in that literature, and the 11-20 money request game, recently introduced by Arad and Rubinstein (2012), in the graphical version of Goeree et al. (2016). In our experiment subjects play different variants of the 11-20 game, which allows us to vary the payoff structure, hence the incentives, without affecting the underlying best-reply structure. Using deliberation times as a proxy for cognitive effort in these strategic situations, we then argue that process data, such as deliberation times, can provide additional evidence that the underlying decision processes are indeed based on some form of step-wise reasoning.

In the beauty contest game we find longer deliberation times for choices commonly associated with more steps of reasoning, confirming the prediction of our model that deliberation time is increasing in cognitive effort. In three variants of the 11-20 game, we show that deliberation times are longer for higher-level choices, at least in situations where the payoff structure is such that following level-k type of reasoning is salient enough. However, when iterative thinking is not natural or when a conflict with alternative decision rules (e.g., based on the salience of high payoffs) is likely, we find no systematic relation between higher-level choices and deliberation times. Rather, in this situation we find overall longer deliberation times suggesting a conflict between competing decision rules. That is, features besides the best-reply structure matter as well, but deliberation times can serve as a tool to identify such situations. The second prediction of our model relates changes in the incentives to deliberation times and predicts shorter deliberation times for a given number of steps when the incentives are increased. This captures the stylized fact that "decisions" closer to indifference require more deliberation. In one treatment we find overall shorter deliberation times when incentives are increased, while at the same we observe more higher-level choices. This might sound counterintuitive at first, since intuitively more steps of reasoning should increase deliberation times. However, our model predicts that at the same time higher incentives decrease the time required for each single step, which can explain this finding. Indeed, in that treatment we find that when incentives are increased the deliberation time per step decreases. Further, we show that a decrease in incentives only leads to longer deliberation times and can be explained by longer deliberation times per step, as predicted by our model, if it is systematic and sufficiently large. Further, we find a systematic effect of incentives on the depth of reasoning, that is in line with previous findings in the literature (Alaoui and Penta, 2016a).

The paper is structured as follows. Section 3.2 introduces our model and relates it to existing models in the literature. Section 3.3 describes the experimental design. Section 3.4 presents the results of our experiment. In Section 3.5 we presents additional analysis. Section 3.6 discusses and summarizes our findings.

3.2 The Model

We explicitly model decision making as a process of iterative reasoning as put forward in the literature on iterative thinking (Nagel, 1995; Stahl, 1993; Stahl and Wilson, 1995; Ho et al., 1998). Our model yields testable predictions linking deliberation times to choices and incentives in a specific class of strategic games.

Consider a symmetric, two-player game $\Gamma = (\pi, S)$ with finite strategy space S and payoff function $\pi : S \times S \longrightarrow \mathbb{R}$. Assume that for any $s \in S$ there is a unique best-reply, denoted by $BR(s)$, that maximizes $\pi(\cdot, s)$. The *best-reply structure* of Γ for a starting point $s_0 \in S$ is a sequence $(s_k^*)_{k\in\mathbb{N}}$ such that $s_0^* = s_0$ and $s_k^* = BR(s_{k-1}^*)$. Fix a best-reply structure $(s_k^*)_{k\in\mathbb{N}}$ with starting point s_0. We model a process of iterative thinking as a sequence of binary "choices," where in each step a player evaluates the current strategy s_{k-1}^*, reached after $k-1$ steps of thinking, against strategy s_k^* by comparing $\pi(s_{k-1}^*, s_{k-1}^*)$ against $\pi(s_k^*, s_{k-1}^*)$. In other words, the player considers the case where his opponent has also conducted $k-1$ steps of thinking, hence uses strategy s_{k-1}^*, and then evaluates the potential gain from conducting an additional step of thinking, that is $\pi(s_k^*, s_{k-1}^*) - \pi(s_{k-1}^*, s_{k-1}^*)$. Note that this last evaluation does not necessarily involve conscious calculations, but should rather be understood as a proxy that determines whether to engage in additional deliberation. For example, one way to think about is that this evaluation happens automatically and that the controlled process of iterative thinking only takes over when the payoff is large enough. In addition, we assume that each step of thinking comes with a cognitive cost. The *cognitive cost* associated with the kth step of thinking is given by $c_i(k)$ with $c_i : \mathbb{N} \longrightarrow \mathbb{R}_+$ weakly increasing. Thus the maximal number of steps of thinking player i is willing (or able) to conduct is given by $T_i = \min\{k \in \mathbb{N} \mid \pi(s_{k+1}^*, s_k^*) - \pi(s_k^*, s_k^*) < c_i(k)\}$.

Denote by $u_k = \pi(s_k^*, s_{k-1}^*) - \pi(s_{k-1}^*, s_{k-1}^*)$ the potential gain of the kth step of thinking. We link this simple model of iterative thinking to deliberation times via two basic assumptions. First, we assume that the deliberation time for conducting k steps of thinking is the sum of the deliberation times required for each step. Second, we assume that the deliberation time for a given step of thinking is larger the smaller the potential gain/loss for that step. This models a well-established fact in psychology that "decisions" closer to indifference are slower (Dashiell, 1937; Mosteller and Nogee, 1951; Moyer and Landauer, 1967). The deliberation time of player i for choosing strategy s^*_k is given by

$$DT_i(s^*_k) = \sum_{i=0}^{k} f_i(|u_i|) \text{ with } f_i : \mathbb{R}_+ \longrightarrow \mathbb{R}_{++} \text{ strictly decreasing and positive.}$$

We say that a strategy s requires *more cognitive effort* compared to s', if it is the result of more steps of reasoning, that is $s = s^*_k$ and $s' = s^*_{k'}$ with $k > k'$. In that case, our model implies that $DT_i(s) > DT_i(s')$ if s requires more cognitive effort than s'.

Prediction 1 (Cognitive Effort)**.** Deliberation time is increasing in cognitive effort.

This prediction is straightforward and is also a prediction of Alaoui and Penta (2016b). The following prediction, however, is particular to our model where we in addition assume that that deliberation time per step is a decreasing function of the utility differences. This allows us to derive predictions on how changing the incentives, that is, the payoff structure of the game, affects deliberation times. Consider two symmetric, two-player games $\Gamma = (\pi, S)$ and $\Gamma' = (\pi', S)$ with the same strategy space S and the same best reply structure $(s^*_k)_{k\in\mathbb{N}}$ with starting point s_0. We say that Γ' has *higher payoff differences* than Γ for step k if $u'_k > u_k$ where $u_k = \pi(s^*_k, s^*_{k-1}) - \pi(s^*_{k-1}, s^*_{k-1})$ and $u'_k = \pi'(s^*_k, s^*_{k-1}) - \pi'(s^*_{k-1}, s^*_{k-1})$. Suppose Γ' has higher payoff differences than Γ for any step $l \leq k$ for some k, then $DT'_i(s^*_k) = \sum_{i=0}^{k} f_i(|u'_i|) < \sum_{i=0}^{k} f_i(|u_i|) = DT_i(s^*_k)$ because $f(|u'_l|) < f(|u_l|)$ for all $l \leq k$.

Prediction 2 (Incentives)**.** The deliberation time for a choice corresponding to k steps of thinking is shorter (longer) if the payoff differences for all steps up to k are increased (decreased).

For a fixed number of steps of thinking our model predicts shorter deliberation times for higher payoff differences, because a player requires less time for each step. Note that this does not necessarily imply that one should observe longer total deliberation times for higher payoff differences. This is because higher payoff differences might increase the gain from conducting another step of thinking as well, hence subjects potentially conduct more steps of thinking, which in turn increases deliberation time.

It is conceivable that individual differences in cognitive ability affect deliberation times. In our model cognitive ability could affect deliberation times in two ways: First, higher cognitive ability could translate into uniformly lower cognitive costs of reasoning c_i. In that case, players with higher cognitive ability are likely to conduct more steps of reasoning, because $T_i \geq T_i'$ if $c_i(k) \geq c_i'(k)$ for all $k \leq T_i$, which would increase overall deliberation time. Second, higher cognitive ability could also translate into shorter deliberation times per step, which would decrease deliberation times. However, that does not mean that deliberation time for players with high cognitive ability will generally be shorter (independently of the number of steps). This is because higher cognitive ability might also result in more steps of reasoning requiring additional deliberation time, so that the overall effect on deliberation times is indeterminate.

3.2.1 Related Models

In this subsection, we discuss related models that account for response or deliberation times. The model closest to ours is Alaoui and Penta (2016b) which extends the model of "Endogeneous Depth of Reasoning" in Alaoui and Penta (2016a) to account for deliberation times. Alaoui and Penta (2016a) provide a model of iterative thinking where the depth of reasoning is endogenously determined and results from a cost-benefit analysis. Alaoui and Penta (2016b) discuss how this model can be used to make comparative statics predictions for deliberation times. Total response time for a given number of steps of reasoning is the sum of the times required to attain the necessary unit of understanding for each step. Hence their model also predicts (for sufficiently similar games) that response time is increasing in the depth of reasoning. Like we do, they assume that the depth of reasoning is determined by the "value of reasoning" and the "cost of reasoning." The former is linked to the payoff structure of the game, whereas the latter depends on the complexity of the game. They assume that the value of reasoning has a maximum gain representation (Alaoui and Penta, 2015), that is, it equals the highest possible payoff improvement that an agent could obtain by using the "next step strategy" instead of the current one. The key difference to our model is that we assume that deliberation time of a given step is decreasing in the utility differences. In their model the value of reasoning is also related to differences in payoffs between alternatives, however, a higher value of reasoning only affects total deliberation times because it increases the probability of conducting another step, but not the time required for a given step, which is a key assumption in our model.

Chabris et al. (2009) study the allocation of time across decision problems. Their model is similar in spirit to ours in that it is motivated by the closeness-to-indifference effect. They also model response time as a decreasing function of differences in expected utility. However, in contrast to our model they focus on binary intertemporal choices and do not consider iterative reasoning. They report empirical evidence that choosing among two options whose expected utilities are close requires longer decision times than when expected utilities are far apart, thus indicating an inverse relationship between average response time and utility differences. They argue that their results support the view that decision-making is a cognitive costly activity that allocates time according to cost-benefit principles, which is also in line with the interpretation in Alaoui and Penta (2016b).

Echenique and Saito (2017) give an axiomatic characterization for when data on choices and deliberation times is consistent with a monotonic relationship between response time and differences in utility. Their model is related to ours in that we also assume that there is a monotone relationship between deliberation time and the utility difference between staying with the current strategy or conducting an additional step of reasoning. However, their model focuses on binary, discrete choices, while the focus of our model is iterative thinking of the level-k type.

3.3 Experimental Design

In the previous section we have introduced a simple model linking depth of reasoning to deliberation times. In this section we report results from an experiment which allows us to test the predictions of this model. Our main objective is to study whether depth of reasoning can be linked to a simple measure of cognitive effort, namely deliberation times. The motivation for this is to use an independent measure as a potential validation of iterative thinking. We use two games commonly employed to study cognitive sophistication, the beauty contest game (Nagel, 1995), the workhorse model of iterative thinking, and the 11-20 money request game, a more recent alternative that was explicitly designed to study level-k behavior (Arad and Rubinstein, 2012). We ask whether a higher level of reasoning is reflected in higher cognitive effort, or in other words, is there a direct link between higher levels of reasoning and deliberation times? We use different variations of the 11-20 game, including a variant introduced in Goeree et al. (2016), that vary the incentives for iterative thinking, but do not affect the underlying best-reply structure of the game. This allows us to study how behavior and deliberation times react to systematic changes in the payoff structure.

3.3.1 The Games

The Beauty Contest Game

The standard work horse for the study of cognitive sophistication is the guessing game, or p-beauty contest game (Nagel, 1995). We use a standard, one-shot p-beauty game with $p = 2/3$. In the (discrete) beauty contest game a population of players has to simultaneously guess an integer number between 0 and 100. The winner is the person whose guess is closest to p times the average of all chosen numbers. The winner receives a fixed price, which is split equally among all winners in case of a tie. The beauty contest game is a game with usually more than two players. Since our model is about bilateral

A_9	A_8	A_7	A_6	A_5	A_4	A_3	A_2	A_1	20
9	8	7	6	5	4	3	2	1	0

Figure 3.1: Generalized 11-20 game.

interactions it cannot be directly applied to this situation. However, in the beauty contest game a player's payoff depends only on the average number chosen by all other players and iterative thinking in this game is typically based on beliefs about an representative agent. Thus we can apply our model by viewing this game as a two-player game against a representative agent, who selects that average number.

In this game it is usually assumed that non-strategic (level-0) players pick a number at random from the uniform distribution over $\{0, \ldots, 100\}$, which yields an expected average of 50. Hence, we assume that the starting point for iterative thinking is given by $s_0^* = 50$. If all players choose s_0^*, then the average of all numbers chosen is 50 and hence the best-reply to s_0^* is to choose $s_1^* = 33$, that is the integer closest to $(2/3) \cdot 50$. Iterating, this defines the best-reply structure of the beauty contest game (s_k^*) where s_k^* is the integer closest to $(2/3)^k \cdot 50$. This game has two Nash equilibria at 0 and 1.

The 11-20 Game

The second part of our experiment focuses on variants of the 11-20 money request game, that was introduced by Arad and Rubinstein (2012) as a two-player game that is specifically well-suited to study iterative reasoning. Alaoui and Penta (2016a) used the 11-20 game to test their model of endogenous depth of reasoning. Goeree et al. (2016) introduced a graphical version of the 11-20 game that allows to vary the payoff structure without affecting the underlying best-reply structure of the game. We now describe a generalized version of this graphical 11-20 game. In what follows, we will refer to this game (and variants thereof) simply as "11-20 game."

Consider ten boxes horizontally aligned and numbered from 9 (far left) to 0 (far right) as depicted in Figure 3.1. Each box $b \in \{1, \ldots, 9\}$ contains

BASE	11	12	13	14	15	16	17	18	19	20
EXTR	19	18	17	16	15	14	13	12	11	20
FLAT	17	17	17	17	17	17	17	17	17	20

Figure 3.2: Payoff structure for the different variants with low cost.

an amount $A_b < 20$ and the rightmost box contains the highest amount of $A_0 = 20$. There are two players, $i = 1, 2$ and each has to choose a box $b_i \in \{0, \ldots, 9\}$. Each player receives the amount A_{b_i} in the box he chose plus a bonus of R if he chose the box that is exactly one to the left of his opponent's box. That is, payoffs are given by

$$\Pi_i(b_i \mid b_{-i}) = \begin{cases} A_{b_i} & \text{if } b_i \neq b_{-i} + 1 \\ A_{b_i} + R & \text{if } b_i = b_{-i} + 1 \end{cases}.$$

A feature of this game is that choosing box 0 is the salient and obvious candidate for a non-strategic level-0 choice, because it awards the highest "sure payoff" of 20 that can be obtained without any strategic considerations. Thus, the rightmost box 0 is a natural anchor serving as a starting point for iterative thinking. If the bonus R is large enough, that is, $R > 20 - \min\{A_b | b = 1, \ldots 9\}$, then the best-reply structure for the salient starting point $s_0^* = 0$ is $(s_k^*)_k$ with $s_k^* = k$ for $k = 1, \ldots, 9$.[2] In other words, for a sufficiently large bonus the best-reply is always to choose the box that is exactly one to the left of your opponent (if there is such a box). Note that for $s_0^* = 0$ the best-reply structure is independent of the specific payoff structure given by $A_9, \ldots, A_1$ if $R > 20 - \min\{A_b | b = 1, \ldots 9\}$ and $A_0 = 20$, that is, if the bonus is large enough and the rightmost box contains the salient amount of 20.

[2] Note that the best-reply to an opponent choosing box 9 is to choose box 0, hence for $k > 9$ the best-reply structure cycles repeatedly from 0 to 9. For simplicity, we abstract from this issue and focus only on steps 1-9. Alaoui and Penta (2016a) use a slightly different payoff structure with an additional bonus in case of a tie that breaks this best-reply cycle.

We use the three versions of the 11-20 game shown in Figure 3.2.[3] The sure payoffs given by the amounts $A_0, \dots, A_9$ differ across versions, however, they are always chosen such that they feature the best-reply structure just described above. In the baseline version (BASE) the amounts are increasing from the left box to the rightmost box, containing the highest amount of 20. BASE corresponds to the original version of Arad and Rubinstein (2012) and to the baseline version of Goeree et al. (2016). In BASE there is a natural trade-off between the sure payoffs $A_1, \dots, A_9$ and the bonus, that can be obtained by outsmarting ones opponent, with each incremental step of reasoning being equally costly in terms of sure payoff. The extreme version (EXTR) was previously used in Goeree et al. (2016). In this version, the amounts in the boxes are rearranged so that they are decreasing from left to right with the second highest amount in the leftmost box. While this rearrangement does not alter the underlying best-reply structure, it crucially affects the cost in terms of sure payoff associated with different levels of reasoning. Choosing box 1 is now disproportionately expensive, and further increments come, in terms of sure payoff, at no cost but instead at a gain. Moreover, this asymmetry potentially opens the door for alternative heuristics, such as choosing the highest amount that still grants the possibility for a bonus, which in this case would imply choosing the leftmost amount. The third version we use was introduced to remove the trade-off between higher steps of reasoning and sure payoff. This flat cost version (FLAT) is a modification of BASE, where the first iteration results in a flat cost, but after that all additional steps are identical and come at no further cost in terms of sure payoff. FLAT could be viewed as a modification of Arad and Rubinstein's (2012) costless iterations version. In FLAT all boxes except the rightmost box contain the same amount, which is by some fixed amount lower than the salient amount of 20. Thus, choosing any box except the rightmost gives the same sure payoff and, hence, after the first step any additional step is "costless."

[3]For each of the three versions BASE, EXTR, and FLAT there is a unique mixed strategy Nash equilibrium. For the low cost and low bonus versions those are given by $(0, 0, 0, 0, 0.25, 0.25, 0.2, 0.15, 0.1, 0.05)$, $(0, 0, 0, 0, 0, 0, 0.15, 0.40, 0.45)$, and $(0, 0, 0, 0.10, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15)$, respectively.

BASE	2	4	6	8	10	12	14	16	18	20
EXTR	18	16	14	12	10	8	6	4	2	20
FLAT	14	14	14	14	14	14	14	14	14	20

Figure 3.3: Payoff structure for the different variants with high cost.

We varied these three versions of the 11-20 game along two additional dimensions: First, for each treatment we have a "high cost" version, where for BASE and EXTR the amounts $A_1, \ldots, A_9$ range from 2 to 20 in increments of 2 instead of from 11 to 20 in increments of 1, and for FLAT all amounts other than 20 were set to 14 in the high cost version instead of 17 (see Figure 3.3). Depending on the treatment, the trade-off between bonus and sure payoff for an additional step of reasoning is decreased or increased under high cost. Second, we vary the incentive to reason in that we change the size of the bonus for choosing the box exactly one to the left of the other subject. In the high bonus condition, subjects obtain $R = 40$ additional points for the "correct" box, while in the low bonus condition they only get $R = 20$ additional points.

3.3.2 Design and Procedures

The experiment consisted of three parts during which subjects could earn points with 10 points being worth €0.25. First, each subject played a series of different versions of the money request game. Each treatment BASE, EXTR, and FLAT was played four times, once for each bonus-cost combination. Second, subjects participated in a single p-beauty contest with $p = (2/3)$. In the third part we collected several individual correlates intended to control for cognitive ability, social value orientation, aversion to strategic uncertainty, swiftness, and demographics. There was no feedback during the course of the experiment, that is, subjects did not learn the choices of their opponents nor did they get any information regarding their earnings until the very end of the experiment. All decisions were made independently and at a subject's individual pace. In particular, subject's never had to wait for the decisions of another subject except for the very end of the experiment, at that point they had to wait until everybody had completed the experiment so that outcomes and payoffs could be realized.

We now describe each part of the experiment in detail. For the 11-20 games, we randomly assigned the subjects within a session to one of four randomized sequences of the games to control for order effects.[4] Subjects were informed that for every game they would be randomly matched with a new opponent to determine their payoff for that round, hence preserving the one-shot character of the interaction. Each of the three variants BASE, EXTR and FLAT was played exactly four times, once for each possible combination of cost (low/high) and bonus (low/high).

In the second part, subjects played a single p-beauty game with $p = 2/3$ among all 32 participants in the session. The winner, that is, the subject whose guess was closest to 2/3 times the average of all choices, received 500 points. In case of a tie, that is, when there were multiple winners, the amount was split equally among all winners.

In the final part of the experiment, participants answered a series of questions. First, subjects completed a combined version of the CRT with nine items consisting of the classical three items from Frederick (2005), three additional items taken from Toplak et al. (2014), and two further items introduced by Primi et al. (2015).[5] Subjects received 5 points for each correct answer. Next, we elicited aversion to strategic uncertainty using the method by Heinemann et al. (2009) with random groups of four. The task involves measuring certainty equivalents, similarly to Holt and Laury (2002a)'s multiple price list method, for a situation where payoffs depend on the decision of another subject, that is, strategic uncertainty. In ten situations subjects have to choose between different amounts of sure payoffs (5 to 50 points) and a binary coordination game, in which they can earn 50 points if at least

[4]The exact sequences are provided in Table 3.9 in Appendix 3.A. Besides our main treatments the sequences contained a further treatment with four additional games discussed in Section 3.5.2.

[5]Other studies (Cappelen et al., 2013; Gill and Prowse, 2016) have also used the Raven test as a proxy for cognitive ability. Brañas-Garza et al. (2012) used the Raven test and the CRT by Frederick (2005) in a series of six one-shot p-beauty games and find that CRT predicts lower choices (i.e. higher level), while performance in the Raven test does not.

two other members of their group have also chosen the coordination game and zero points otherwise. Subjects were randomly allocated into groups of four, and for each group one of the decision situations was randomly selected for payment. Finally, we collect a measure to control for differences in mechanical swiftness (Cappelen et al., 2013). To that end we recorded the time needed to complete four simple demographic questions on gender, age, field of study, and native language. This part was integrated into a larger questionnaire, which also comprised questions regarding subjects' understanding of the tasks, their perception of its complexity, number of university semesters, left- or right-handedness, average amount of money needed per month, and previous attendance of a lecture in game theory.

To determine a subject's earning in the experiment the payoffs from each part were added up and converted to € at rate of € 0.25 for each 10 points. In addition subjects received a show-up fee of € 4 for an average total renumeration of € 15.67. A session lasted on average 60 minutes including instructions and payment.

A total of 128 subjects (79 female) participated in 4 experimental sessions with 32 subjects each. Participants were recruited from the student population of the University of Cologne using ORSEE (Greiner, 2015), excluding students of psychology, economics, and economics related fields, as well as experienced subjects who already participated in more than 10 experiments. The experiment was conducted at the Cologne Laboratory for Economic Research (CLER) and was programmed in z-Tree (Fischbacher, 2007).

3.4 Results

In Subsection 3.4.1 we analyze behavior and deliberation times in the beauty contest game. Subsection 3.4.2 presents the results for the 11-20 games. Results regarding the effect of incentives in the 11-20 game are presented in Subsection 3.4.3.

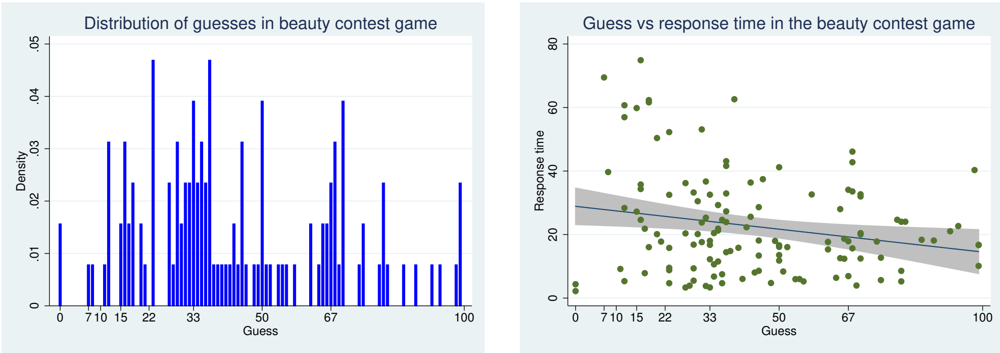


Figure 3.4: Choices and deliberation times in the beauty contest game.
Notes: Left panel shows histogram of guesses (0-100) in the beauty contest game ($N = 128$). Right panel shows a scatter plot of guesses (0-100) vs associated deliberation time (in s) in the beauty contest game and plots the result of a linear regression of choice on deliberation times with 95% confidence interval.

3.4.1 Results for the Beauty Contest Game

We first explore the relation of depth of reasoning, as revealed by choices, cognitive ability and deliberation times in the beauty contest game.

The left panel of Figure 3.4 depicts the distribution of choices in the beauty contest game. Of the 128 subjects only 2 chose the Nash Equilibrium strategy of 0, 23 chose a number close to 33 (level-1), 9 chose a number close to 22 (level-2), 15 chose a number close to 15 (level-3), and 3 chose a number corresponding to higher levels. The target numbers in our four sessions were 27, 28, 29 and 32 and the respective winning numbers were 28, 27, 30 and 32. Hence, the best-performing strategy (among the level-k strategies) would have been the level-1 choice of 33. We classify all choices that are at a distance of at most 2 from the level-k strategy as level-k.[6] Overall behavior is in line with previous results in the literature, that commonly observe mostly one to three steps of reasoning and a significant amount of unclassified (random) choices usually thought of as level-0. The right panel of Figure 3.4 shows a scatter plot of subjects' guesses and the corresponding time taken for

[6]Choices between 31 and 35 were classified as level-1, choices between 20 and 24 as level-2, choices between 13 and 17 as level-3, choices between 8 and 12 as level-4, choices between, 5 and 7 as level-5, and choices of 0 as Nash Equilibrium. There were no choices between 1 and 5. Our results are unchanged for narrower classifications, e.g. where only the level-k strategy ± 1 are classified as level-k.

Table 3.1: Linear regressions on log DT for the beauty contest game.

log DT	1	2	3
Level		0.1264**	0.1352**
		(0.0593)	(0.0617)
CRTExtended	0.0017		−0.0173
	(0.0316)		(0.0323)
Swiftness	−0.2836	−0.2780	−0.3225
	(0.4463)	(0.4306)	(0.4399)
Female	−0.2640*	−0.2237	−0.2420*
	(0.1418)	(0.1353)	(0.1400)
Constant	3.1202***	3.0158***	3.1016***
	(0.2353)	(0.1668)	(0.2319)
Adjusted R^2	0.0114	0.0468	0.0412
F-Test	1.4784	3.0456**	2.3422*
Observations	126	126	126

Notes: Standard errors in parentheses. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Two subjects with very fast choices of 0 were excluded from the analysis. Our results are robust when those choices are included and classified as level-0. Classification of levels: Level 1 (31-35), Level 2 (20-24), Level 3 (12-17), Level 4 (7-11), Level 0 (rest). There were no choices in the range 1-6. CRTExtended (number of correct answers; 0-7); Swiftness (time needed to answer 3 demographic questions); Female (dummy).

that choice. The slope of the regression line suggests a negative correlation between deliberation times and "higher-level" choices. Hence, indicating that choices corresponding to more steps of reasoning require longer deliberation times.

Our model predicts that deliberation time is increasing in cognitive effort (Prediction 1), that is, we expect longer deliberation times for choices associated with more steps of reasoning. We now test this prediction using a series of three linear regressions with log transformed deliberation times (log DT) as dependent variable and controls for gender and individual differences in mechanical swiftness (Cappelen et al., 2013). The results of those regressions are presented in Table 3.1. In model 1, we see that cognitive ability, measured by the extended CRT, has no effect on deliberation times. Recall that in our model the overall effect of cognitive ability on deliberation time is indeterminate because there are two possible effects that might be countervailing. Model 2 shows a significant positive effect of higher-level choices on deliberation time. Thus, we find, in line with Prediction 1, that deliberation time is increasing in cognitive effort. This result remains robust when we additionally control for cognitive ability (model 3).[7]

Performance in the CRT was previously found to be correlated with level in the beauty contest (Brañas-Garza et al., 2012). Conducting an additional linear regression with level as dependent variable on CRT, we find a significant and positive coefficient for CRT ($\beta = 0.1462$, $p = 0.009$). This result indicates that, in line with previous results in the literature, subjects with higher cognitive ability conduct more steps of reasoning.

3.4.2 Results for the 11-20 Games

In this section we analyze the relation between deliberation times, choices and cognitive ability in the three versions of the 11-20 game.

We first consider the observed behavior across the different variants of the 11-20 game.[8] Choices in BASE closely resemble the behavioral patterns found in Arad and Rubinstein (2012) and Goeree et al. (2016). Most subjects selected one of the 3 rightmost boxes corresponding to levels 0 to 3. In EXTR, behavior is comparable to that observed in Goeree et al. (2016) and vastly different from that observed in BASE. There is a large fraction of subjects (between 38% and 62%) choosing the rightmost box containing the salient amount of 20, but box 1 and 2 to its left are chosen very rarely compared to BASE. Instead between 25% and 33% of subjects chose one of the leftmost boxes 8 and 9, which were essentially not chosen at all in BASE. These choices correspond to eight or nine steps of reasoning. In contrast, behavior in FLAT is again very similar to that in BASE with most choices corresponding to not more than three steps of reasoning. Compared to BASE there is a larger

[7]Throughout the paper we use the results from the 7-item CRT by Toplak et al. (2014) as measure of cognitive ability. Subject's also answered the two additional items proposed by Primi et al. (2015). Our results do not change if we use their version or a combination of both instead.

[8]Figure 3.6 in Appendix 3.B shows the distribution of choices across all instances of the 11-20 game.

Table 3.2: Random effects log DT regressions on level (full sample).

log DT	1	2	3
Level	0.0385***	0.0272***	0.0282***
	(0.0057)	(0.0060)	(0.0060)
EXTR		0.1631***	0.1614***
		(0.0351)	(0.0352)
FLAT		−0.0309	−0.0308
		(0.0337)	(0.0338)
CRTExtended			0.0572***
			(0.0165)
Swiftness	0.2607	0.2626	0.4155*
	(0.2406)	(0.2407)	(0.2335)
Female	0.0943	0.0932	0.1654**
	(0.0744)	(0.0744)	(0.0738)
Period	−0.0884***	−0.0889***	−0.0889***
	(0.0030)	(0.0030)	(0.0030)
Constant	2.5943***	2.5788***	2.2628***
	(0.0928)	(0.0944)	(0.1284)
R^2 (overall)	0.3056	0.3132	0.3417
Wald-Test	1299.4016***	1371.6146***	1385.4388***
Observations	1536	1536	1536

Notes: Standard errors in parentheses. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. A choice of box $k \in \{0, \ldots, 9\}$ is classified as level k; EXTR and FLAT are treatment dummies; CRTExtended (number of correct answers; 0-7); Swiftness (time needed to answer 3 demographic questions); Female (dummy); Period controls for position in the sequence of games.

fraction of level-0 choices in FLAT, which is most likely due to the first step being more costly in terms of sure payoff. When playing against the empirical distribution of choices the best-performing strategies for BASE, EXTR and FLAT would correspond to level 2, level 1 and level 1, respectively.[9]

Table 3.2 shows a series of four GLS random-effects regressions with log DT as the dependent variable including as observations all 12 choices in BASE, EXTR, and FLAT. In all models we control for mechanical swiftness, gender and the position within the sequence of games (Period). We find a significant and positive relation between deliberation times and depth of rea-

[9]Controlling for empirical payoffs does not affect our results.

soning. As predicted, choices associated with more steps of thinking require more deliberation (model 1). This relation is unaffected when we include the treatment dummies EXTR and FLAT in model 2. Further, we observe a significant positive coefficient on EXTR indicating that choices in EXTR generally required more deliberation time. In fact, the average deliberation time in EXTR was 12.6 seconds, whereas the average deliberation time in BASE and FLAT was only 9.9 seconds. Pairwise Wilcoxon signed rank tests confirm that the average deliberation time in EXTR is significantly higher compared to both BASE ($N = 128$, $z = 4.6778$, $p < 0.001$) and FLAT ($N = 128$, $z = 4.3746$, $p < 0.001$). In model 3 we include performance in the CRT, as measured by the number of correct answers, as an independent variable to control for cognitive ability. The coefficient of CRT is significant and positive, that is, subjects scoring higher on the CRT take more time to make their decisions. More importantly, controlling for cognitive ability does not alter the effects of level and EXTR on deliberation times.

In a next step, we repeat the same exercise separately for each of the three treatments BASE, EXTR and FLAT. To that end, we run separate regressions considering only the four choices taken for each of the variants. Table 3.3 presents the results of these regressions. The results confirm our previous findings, there is a positive significant relation between deliberation times and higher-level choices in all three variants of the game. There is no effect of cognitive ability on deliberation times in BASE, whereas we find a positive and significant effect in EXTR and FLAT.

Considering the different variants separately has the advantage that we now can check for potential explanations for the previously observed relation between deliberation times and higher-level choices. In all variants a choice of the rightmost box is appealing for various reasons. First, it maximizes the sure payoff as the requested sure amount by selecting this box is 20 and thus maximal. Second, it minimizes strategic uncertainty as it yields a payoff of 20 independent of the choice of the other player, which is also what makes it a salient level-0 strategy. To better understand how the differences in the payoff structure between the treatments affect the relation of deliberation times, choices and cognitive ability we ran additional regressions where we

Table 3.3: Random effects log DT regressions on level.

log DT	BASE 1	EXTR 2	FLAT 3
Level	0.0449**	0.0277***	0.0532***
	(0.0182)	(0.0079)	(0.0152)
CRTExtended	0.0280	0.0848***	0.0628***
	(0.0186)	(0.0207)	(0.0186)
Swiftness	0.5335**	0.2254	0.4919*
	(0.2632)	(0.2925)	(0.2617)
Female	0.1313	0.1926**	0.1883**
	(0.0834)	(0.0925)	(0.0830)
Period	−0.0878***	−0.0843***	−0.0934***
	(0.0049)	(0.0053)	(0.0050)
Constant	2.3286***	2.3157***	2.1742***
	(0.1511)	(0.1649)	(0.1497)
R^2 (overall)	0.3298	0.3148	0.3621
Wald-Test	352.7159***	300.5368***	378.8169***
Observations	512	512	512

Notes: Standard errors in parentheses. $^{*}\ p < 0.1,^{**}\ p < 0.05,^{***}\ p < 0.01$. Models are restricted to subsamples including only the four decisions in BASE, EXTR or FLAT, respectively. A choice of box $k \in \{0, \ldots, 9\}$ is classified as level k; CRTExtended (number of correct answers; 0-7); Swiftness (time needed to answer 3 demographic questions); Female (dummy); Period controls for position in the sequence of games.

add further variables controlling for specific features of the payoff structure. First, to check whether choices of the rightmost box are particularly fast, we include a dummy indicating those choices, denoted *Rightmost20*, in the regressions. Further, in EXTR a choice of the leftmost box could also be salient because it combines a high sure payoff with a chance of getting the bonus. We thus include another dummy, denoted *LeftmostBox*, into the regression for EXTR. Table 3.4 shows the results of these regressions.

In BASE we still observe a clear positive relation between deliberation time and level after controlling for choices of the rightmost box, which in turn are not significantly faster. In EXTR choices of the rightmost box are significantly faster and this explains most of the effect of level on deliberation times, in particular, level becomes insignificant when adding the dummy

Table 3.4: Random effects log DT regressions with controls for the payoff structure.

	BASE	EXTR	FLAT
log DT	1	2	3
Level	0.0658^{***}	-0.0164	0.0202
	(0.0210)	(0.0183)	(0.0182)
Rightmost20	0.1634^{*}	-0.3438^{***}	-0.2356^{***}
	(0.0835)	(0.0991)	(0.0728)
LeftmostBox		0.1010	
		(0.1201)	
CRTExtended	0.0330^{*}	0.0860^{***}	0.0570^{***}
	(0.0189)	(0.0207)	(0.0183)
Swiftness	0.5424^{**}	0.1841	0.4671^{*}
	(0.2646)	(0.2893)	(0.2571)
Female	0.1359	0.1912^{**}	0.2034^{**}
	(0.0839)	(0.0911)	(0.0817)
Period	-0.0862^{***}	-0.0845^{***}	-0.0921^{***}
	(0.0049)	(0.0053)	(0.0050)
Constant	2.2255^{***}	2.6161^{***}	2.3138^{***}
	(0.1605)	(0.1881)	(0.1534)
R^2 (overall)	0.3294	0.3330	0.3785
Wald-Test	360.2338^{***}	317.7823^{***}	394.5066^{***}
Observations	512	512	512

Notes: Standard errors in parentheses. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Models are restricted to subsamples including only the four decisions in BASE, EXTR or FLAT, respectively. A choice of box $k \in \{0, \ldots, 9\}$ is classified as level k; Rightmost20 (dummy for choosing rightmost box); LeftmostBox (dummy for choosing leftmost box); CRTExtended (number of correct answers; 0-7); Swiftness (time needed to answer 3 demographic questions); Female (dummy); Period controls for position in the sequence of games.

Rightmost20. Also in FLAT we observe very fast level-0 choices, and no further relation between deliberation times and level. In BASE iterative thinking is associated with an increasing cost and in this variant we observe the strongest link between deliberation times and level. This effect also goes beyond the observation that choices of the rightmost box containing the salient amount of 20 are faster (in fact they are even slower in BASE). In contrast, for EXTR and FLAT most of the effect is explained by fast level-0 choices. In EXTR we find no systematic relation between deliberation times and higher level choices. Also there is no effect on deliberation times for the leftmost box. This result suggests that iterative thinking might be less dominant in EXTR than for example in BASE although they feature the same best-reply structure. In FLAT the comparably large cost associated with the first step of reasoning results in a large fraction of subjects choosing the rightmost box and these choices are faster. However, beyond that we do not find evidence for a similar relation between higher-level choices and deliberation times in FLAT.

In summary, we find generally longer deliberation times for higher-level choices, which is in line with Prediction 1. But, this result has to be qualified. Although deliberation time increases with cognitive effort in all three treatments, we find this relation to be strongest in BASE. In EXTR, we find that level-0 choices are significantly faster, however, we find no evidence that there is a relation between cognitive effort and deliberation times for higher-level choices. In particular, choices of the leftmost box, while frequent, are not accompanied by longer deliberation times. It is possible that the particular payoff structure of EXTR renders the best-reply structure less salient, which might explain why we find only limited support for iterative thinking in EXTR. Also in FLAT, most of the effect of level on deliberation times is explained by fast level-0 choices. However, in FLAT choices corresponding to more than two steps of reasoning are very rare, which might explain the absence of a relationship for higher-level choices. This does not apply to EXTR where a large fraction of choices corresponds to nine steps of reasoning.

3.4.3 Effect of Incentives in the 11-20 Game

According to Alaoui and Penta (2016a) a higher bonus increases the value of reasoning and hence leads to a higher depth of reasoning. Following this argument, one would expect that the observed level is weakly higher for a high bonus compared to a low bonus for all treatments. On the other hand, choices should correspond to weakly lower levels for high cost for a similar reason.

Table 3.5 shows the results of three random-effects Tobit regressions, one

Table 3.5: Random effects Tobit regressions of level with controls for bonus and cost.

Level	BASE	EXTR	FLAT
HighBonus	0.2811**	0.7425	0.2829
	(0.1295)	(0.8704)	(0.1943)
HighCost	−0.2485*	−3.0640***	−0.8270***
	(0.1292)	(0.8871)	(0.1947)
CRTExtended	−0.0474	−0.4359	−0.0968
	(0.0642)	(0.3992)	(0.0892)
StratUnc	−0.0186	−0.0330	−0.0864
	(0.0522)	(0.3257)	(0.0722)
Gametheory	0.1071	4.2876	0.5795
	(0.4282)	(2.6932)	(0.5932)
Female	−0.3388	0.6374	−0.7885**
	(0.2867)	(1.8018)	(0.3966)
Period	−0.0185	−0.2085**	−0.0334
	(0.0137)	(0.0954)	(0.0209)
Constant	2.0234***	4.2326	2.7137***
	(0.4745)	(2.9572)	(0.6579)
Log likelihood	−892.1128	−803.4438	−897.4692
Wald-Test	13.2723*	21.3988***	30.3227***
Observations	512	512	512

Notes: Standard errors in parentheses. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Models are restricted to subsamples including only the four decisions in BASE, EXTR, or FLAT, respectively. A choice of box $k \in \{0, \ldots, 9\}$ is classified as level k; HighBonus (dummy for high bonus); HighCost (dummy for high cost); CRTExtended (number of correct answers; 0-7); StratUnc (number of B choices in the strategic uncertainty task; 0-10); Gametheory (previous knowledge of game theory; dummy); Female (dummy); Period controls for position in the sequence of games.

for each treatment, where we control for the size of the bonus and the size of the increment (cost). In treatment BASE there is a significant and positive effect, with more high-level choices for a high bonus. For EXTR and FLAT we find no effect of high bonus on level. For high cost, we consistently find that observed levels are lower in all three treatments. These results are consistent with the effect of changing incentives on the depth of reasoning previously observed in Alaoui and Penta (2016a).

Table 3.6: Random effects log DT regressions with interaction of level and bonus.

	BASE		EXTR		FLAT	
log DT	1	2	3	4	5	6
Level	0.0664***	0.0936***	−0.0046	−0.0141	0.0224	0.0128
	(0.0209)	(0.0264)	(0.0126)	(0.0145)	(0.0181)	(0.0212)
HighBonus	−0.1153**	−0.0278	0.0272	−0.0294	0.1318***	0.0975
	(0.0461)	(0.0679)	(0.0483)	(0.0642)	(0.0468)	(0.0613)
Level × HighBonus		−0.0536*		0.0171		0.0224
		(0.0304)		(0.0128)		(0.0257)
CRTExtended	0.0326*	0.0329*	0.0887***	0.0891***	0.0574***	0.0575***
	(0.0189)	(0.0185)	(0.0205)	(0.0203)	(0.0183)	(0.0184)
Rightmost20	0.1407*	0.1532*	−0.3117***	−0.3218***	−0.2107***	−0.2120***
	(0.0836)	(0.0839)	(0.0957)	(0.0960)	(0.0729)	(0.0729)
Swiftness	0.5413**	0.5395**	0.2083	0.2079	0.4695*	0.4690*
	(0.2646)	(0.2583)	(0.2892)	(0.2863)	(0.2572)	(0.2580)
Female	0.1363	0.1451*	0.1938**	0.1927**	0.2012**	0.2026**
	(0.0839)	(0.0820)	(0.0914)	(0.0905)	(0.0817)	(0.0819)
Period	−0.0876***	−0.0874***	−0.0838***	−0.0839***	−0.0908***	−0.0907***
	(0.0049)	(0.0049)	(0.0053)	(0.0053)	(0.0050)	(0.0050)
Constant	2.3001***	2.2481***	2.5451***	2.5815***	2.2234***	2.2369***
	(0.1631)	(0.1634)	(0.1837)	(0.1846)	(0.1565)	(0.1577)
R^2 (overall)	0.3352	0.3418	0.3310	0.3343	0.3850	0.3857
Wald-Test	371.3798***	370.6184***	317.6695***	318.7950***	409.2548***	410.3835***
Observations	512	512	512	512	512	512

Notes: Standard errors in parentheses. $^{*}\ p < 0.1,^{**}\ p < 0.05,^{***}\ p < 0.01$. Models are restricted to subsamples including only the four decisions in BASE, EXTR or FLAT, respectively. A choice of box $k \in \{0, \dots, 9\}$ is classified as level k; HighBonus (dummy for high bonus); CRTExtended (number of correct answers; 0-7); Rightmost20 (dummy for choosing rightmost box); Swiftness (time needed to answer 3 demographic questions); Female (dummy); Period controls for position in the sequence of games.

We now analyze the effect of a change in the incentives resulting from an increased bonus on deliberation times, separately for each treatment. Increasing the bonus has a twofold effect on deliberation times: First, it increases the potential gain from an additional step of reasoning by 20 and thus increases the payoff differences for the first nine steps. Hence, according to Prediction 2 deliberation times per step should be shorter for a high bonus. On the other hand, assuming that the cognitive cost is unaffected by a change in the bonus, subjects potentially conduct more steps of reasoning, which in turn increases overall deliberation time. As a consequence the aggregate effect on deliberation times is indeterminate. Controlling for the size of the bonus and the interaction of level with bonus allows us to dissect these two explanations.

Table 3.6 shows the results of a series of random-effects GLS regressions of log DT on level, separately for each variant, where we additionally control for the size of the bonus (models 1, 3 and 5) and the interaction of level with bonus (model 2, 4 and 6). We have already seen in Subsection 3.4.2 that level-0 choices are significantly faster and, being non-strategic, they are likely not to be affected by changes in the bonus. For that reason, we control for non-strategic choices by including a dummy for the rightmost box. In BASE, we find shorter deliberation times when the bonus is high (model 1), however, this effect becomes non-significant when we include the interaction of level with high bonus (model 2). The coefficient for the latter is significant with negative sign, that is, when the bonus is high subjects require less additional deliberation time per step. We find no evidence that bonus has any systematic effect on deliberation times in EXTR (model 3 and 4). In FLAT, subjects overall deliberate longer in the high bonus condition (model 5), however, this effect becomes non-significant when we additionally control for the interaction of level with bonus (model 6). The latter is not significant. Summarizing, we find that increasing the bonus decreases deliberation times in BASE, increases deliberation times in FLAT, and has no effect on deliberation times in EXTR. The decrease in BASE is a result of shorter deliberation times per step, as predicted by our model, which can also explain why overall deliberation time decreases although observed levels are higher.

Next, we study the effect of an increase in the cost, in terms of sure payoff, of an additional step of reasoning, again separately for each treatment. The predicted effect of an increase in cost differs across treatments as it depends on the particular payoff structure. In BASE high cost has a twofold effect: First, the potential gain for conducting an additional step decreases by 1 for high cost for the first nine steps. Hence according to Prediction 2 we expect shorter deliberation times per step for high cost. However, the decrease in payoff differences is very small compared to the one resulting from a change in the bonus, and hence this effect is likely to be small. On the other hand, because high cost implies smaller payoff differences, subjects potentially conduct less steps of reasoning (again assuming that cognitive cost is unaffected), which would decrease overall deliberation time. The

Table 3.7: Random effects panel log DT regressions with interaction of level and cost.

	BASE		EXTR		FLAT	
log DT	1	2	3	4	5	6
Level	0.0653***	0.0505**	−0.0067	−0.0115	0.0233	0.0066
	(0.0211)	(0.0242)	(0.0124)	(0.0137)	(0.0181)	(0.0204)
HighCost	−0.0179	−0.0781	0.1550***	0.1203*	0.1343***	0.0639
	(0.0462)	(0.0669)	(0.0481)	(0.0640)	(0.0474)	(0.0626)
Level × HighCost		0.0374		0.0105		0.0457*
		(0.0301)		(0.0128)		(0.0265)
CRTExtended	0.0330*	0.0338*	0.0898***	0.0895***	0.0571***	0.0585***
	(0.0189)	(0.0189)	(0.0205)	(0.0205)	(0.0183)	(0.0182)
Rightmost20	0.1640**	0.1730**	−0.3628***	−0.3587***	−0.2608***	−0.2404***
	(0.0836)	(0.0839)	(0.0947)	(0.0949)	(0.0729)	(0.0737)
Swiftness	0.5425**	0.5480**	0.2062	0.2024	0.4655*	0.4846*
	(0.2646)	(0.2647)	(0.2888)	(0.2889)	(0.2571)	(0.2547)
Female	0.1357	0.1389*	0.1939**	0.1946**	0.2079**	0.2085***
	(0.0839)	(0.0839)	(0.0913)	(0.0913)	(0.0817)	(0.0808)
Period	−0.0860***	−0.0855***	−0.0857***	−0.0856***	−0.0933***	−0.0927***
	(0.0049)	(0.0049)	(0.0052)	(0.0052)	(0.0050)	(0.0050)
Constant	2.2338***	2.2469***	2.5255***	2.5429***	2.2581***	2.2659***
	(0.1619)	(0.1622)	(0.1792)	(0.1805)	(0.1544)	(0.1532)
R^2 (overall)	0.3296	0.3294	0.3413	0.3410	0.3856	0.3903
Wald-Test	321.5910***	327.7389***	367.0965***	369.0819***	356.5129***	365.6349
Observations	512	512	512	512	512	512

Notes: Standard errors in parentheses. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Models are restricted to subsamples including only the four decisions in BASE, EXTR or FLAT, respectively. A choice of box $k \in \{0, \ldots, 9\}$ is classified as level k; HighCost (dummy for high bonus); CRTExtended (number of correct answers; 0-7); Rightmost20 (dummy for choosing rightmost box); Swiftness (time needed to answer 3 demographic questions); Female (dummy); Period controls for position in the sequence of games.

payoff structure in EXTR, does not allow for a clear-cut prediction for the effect of high cost on deliberation times. The reason is that for high cost, the potential gain for the first step decreases sharply, but the potential gain for all further steps increases slightly. This would lead to longer deliberation times for the first step, and shorter deliberation times for all subsequent steps. It is unclear, which of these countervailing effects should dominate. In FLAT, only the potential gain from the first step is lower for high cost, while the remaining steps are unaffected. Hence, we expect longer deliberation times for the first step. Again, the decrease in potential gain for the first step might lead to subjects conducting less steps of reasoning, which in turn might increase overall deliberation time.

Table 3.7 reports the results of a series of random-effects GLS regressions of log DT on level, separately for each variant, where we additionally include a dummy for high cost (models 1, 3 and 5) and the interaction of level with high cost (model 2, 4 and 6). Again we also control for fast level-0 choices. We find no effect of high cost on deliberation time in BASE (model 1 and 2). This is not surprising, since, as explained above, the resulting change in potential gains per step is very small. In EXTR, deliberation times are longer for high cost with no systematic effect of the interaction with level. For FLAT, model 5 indicates longer deliberation times for high cost. This effect becomes non-significant when we additionally control for the interaction of level with high cost (model 6). The coefficient for the latter is significant and positive, as predicted by our model. Summarizing we find overall longer deliberation times in EXTR and FLAT for high cost, but not in BASE. The increase in FLAT is a result of longer deliberation times per step, as predicted by our model. That is, our model can explain why deliberation times in FLAT are increasing for high cost although observed choices correspond to less steps of reasoning.

3.5 Additional Observations

3.5.1 Other Level-0 Specifications in the 11-20 Game

Arad and Rubinstein (2012) argue that choosing 20 in the 11-20 game is a natural anchor for an iterative reasoning process. A further appeal of the original 11-20 game, which essentially corresponds to BASE, is that it is fairly robust to the level-0 specification. That is, choosing 19 in the original 11-20 game, or box 1 in BASE, is the level-1 strategy for a wide range of level-0 specifications. This robustness of course depends on the particular payoff structure of the game and hence might be different across the various versions used in our experiment. In this subsection we explore the robustness of BASE, EXTR and FLAT to the level-0 specification.

Let $\sigma_0 = (p_i)_{i \in \{0,\ldots,9\}}$ denote a level-0 specification that assigns probability p_i to box i. Recall that box 0 always contains the salient amount of 20. We

Table 3.8: Lower bounds on p_0

	BASE		EXTR		FLAT	
bonus	low	high	low	high	low	high
low cost	5%	2.5%	45%	22.5%	15%	7.5%
high cost	10%	5%	90%	45%	30%	15%

want to study the range of specifications σ_0 such that choosing box 1 is still the unique level-1 strategy, that is, $BR(\sigma_0) = \{1\}$. Since the sure payoff of box 1 is strictly smaller than 20, any such specification has to satisfy $A_1 + p_0 R > 20$. If the probability of obtaining the bonus when choosing box 1 times the bonus can not compensate for the loss in sure payoff by moving away from box 0, giving 20 for sure, then choosing box 0 is weakly preferred to choosing box 1. This already provides a lower bound $\underline{p_0}$ on the probability assigned to the rightmost box. Table 3.8 gives an overview over the values of $\underline{p_0}$ across BASE, EXTR and FLAT for each combination of bonus and cost.

Thus, we have identified $p_0 > \underline{p_0}$ as a necessary condition. Note, however, that in general this condition is not sufficient, because some box $j \neq 0$ might be a best reply to σ_0 if the probability p_{j+1} assigned to box $j+1$ is high enough. It turns out that as long as box 1 contains the second highest sure amount, that is, $A_1 \geq A_j$ for all $j \neq 0, 1$, and $p_0 > \underline{p_0}$, it is sufficient that no box $j \neq 0$ is assigned a probability larger than p_0. This condition is satisfied for BASE as long as box 0 is most probable under σ_0 (note that p_0 being most probable already implies $p_0 > 10\%$, hence $p_0 > \underline{p_0}$). For FLAT this condition is satisfied if box 0 is most likely under σ_0 and p_0 is larger than the corresponding lower bound $\underline{p_0}$ (similar to BASE, this latter condition is void for high bonus and low cost). EXTR does not satisfy this condition because box 1 contains the lowest sure amount, hence the probability assigned to the rightmost box has to exceed the probability of any box j by more than $(A_j - A_1)/R$. This condition together with $p_0 > \underline{p_0}$ is sufficient to make box 1 the unique best-response in EXTR.

We identified sufficient conditions on the level-0 specification that ensure that choosing box 1 is the unique level-1 strategy. The requirements for

BASE are fairly weak, in particular they are satisfied when σ_0 is assumed to be uniform randomization, which is often done in games without a salient strategy.[10] Thus BASE can be considered robust to a wide range of level-0 specifications. The conditions for FLAT are slightly stronger, because the lower bounds $\underline{p_0}$ are tighter. In particular, when σ_0 is uniform randomization the level-1 strategy remains to choose box 1 only for high bonus and low cost, while it prescribes to stay with box 0 for the other conditions. The same cannot be said with regard to EXTR. Here both, the lower bounds and the additional condition are very demanding. In particular, choosing the leftmost box that grants the second highest sure payoff is the level-1 strategy for a relatively wide range of specifications that include uniform randomization. We conclude that BASE and FLAT can be considered robust to the level-0 specification, but not EXTR. In EXTR, there is an alternative level-1 strategy for a relatively wide range of alternative level-0 specifications.

In Section 3.4.2 we have assumed that the starting point in the 11-20 game for our model of iterative thinking is to choose the rightmost box containing the salient amount of 20. As just illustrated, the best-reply structure in BASE and FLAT is robust for a wide range of alternative level-0 specifications. Thus, even if, different from our assumption, the starting point does not assign probability one to choosing the rightmost box, the best-reply structure and hence our results are unaffected as long as p_0 is not too small. However, this is not true for EXTR. Here, the best-reply structure is less robust to changes in the level-0 specification, and there is a clear alternative best-reply structure where the leftmost box is the level-1 strategy.

To check for robustness we consider an alternative best-reply structure in the following way: We assume that the level-0 specification is mixed in such a way that the best-reply is to choose the leftmost box, containing the second highest sure payoff, which we then classify as the level-1 strategy. The best-reply to that is to choose the rightmost box containing the salient amount of 20, now classified as level 2. From there the best-reply structure follows the known pattern from right to left. We then repeat the complete analysis

[10] For high cost and low bonus, choosing box 1 is a best-reply, but not a unique one, because it ties with choosing the rightmost box.

conducted for EXTR in Section 3.4.2 for this alternative classification. Comparing the so obtained results to those presented earlier for EXTR, we find no qualitative difference.[11] Hence, we conclude that our results, presented in the previous section, cannot be explained by differences in the robustness to the level-0 specification between treatments.

3.5.2 A Social Preference Variant - SOCP

Our experiment included an additional treatment to check for an alternative explanation of the frequent "high-level" choices of the two leftmost boxes in EXTR, as previously observed by Goeree et al. (2016). By choosing the leftmost box in EXTR a subject can get the second highest sure amount, while at the same time granting his opponent the chance to receive the bonus. If a subject is motivated by some form of other-regarding preferences, then choosing the leftmost box is very attractive because it grants somebody else the chance to get a bonus that is relative large in comparison to the subject's own sacrifice in terms of sure payoff. We thus added an additional treatment, denoted SOCP, that is a variation of FLAT where the *two* rightmost boxes contain both the salient amount of 20. Figure 3.5 shows both the low and high cost version of SOCP. Here choosing the rightmost box guarantees the highest safe amount of 20, while also, at least theoretically, granting the other player the chance to obtain the bonus by selecting the second, inner box that also contains 20. On the other hand, a purely self-interested individual should never choose the rightmost box, since it is weakly dominated by the inner 20 for all possible beliefs. However, it is conceivable that an altruist would select the rightmost alternative to grant his opponent the chance to get the bonus by selecting the inner box with the second 20.[12]

As a proxy for prosociality we measured the social value orientation (SVO) of each subject using a computerized version (Crosetto et al., 2012) of the scale developed by Murphy et al. (2011). We used a scaled version of their

[11] We do not present these alternative regressions here. They are available from the authors upon request.

[12] Note that this is also the most efficient outcome in the sense that it maximizes joint profits.

SOCP	17	17	17	17	17	17	17	17	20	20
SOCP	14	14	14	14	14	14	14	14	20	20

Figure 3.5: Payoff structure for SOCP in the low (top panel) and high (bottom panel) cost version.

six primary items in which subjects were asked to choose among different allocations of points between themselves and a randomly selected partner. For the SVO task one of the six items was randomly selected and paid out using a ring matching procedure, that is, each subject received two payments of up to 25 points, one as a sender and one as a receiver. A higher SVO score indicates that a subject is more prosocial.

In SOCP 36 out 128 subjects chose the rightmost box at least once. To explore whether behavior in SOCP is driven by other-regarding preferences, we first test whether subjects choosing the rightmost box have a higher SVO score. Conducting separate Wilcoxon rank sum tests for each of the four instances of SOCP we find no difference in SVO score for subjects choosing the rightmost box compared to those choosing another box. Next, we consider the relative frequency of choosing the rightmost box across all four instances of SOCP per subject. We run a fractional logit regression of this relative frequency with the SVO score as a independent variable. The coefficient of social value orientation is positive but not significant. Summarizing, we find no evidence that the prosocial motive of granting the opponent the chance to obtain a bonus is a driver of behavior in the 11-20 game.

3.6 Discussion

In this work, we have introduced a simple model linking cognitive sophistication, incentives, and deliberation times, incorporating stylized facts from the psychophysiological literature on response times. We model the total deliberation time of an observed choice as the sum of the deliberation times resulting from a sequence of binary hypothetical decisions that model steps of reasoning. As an immediate consequence we obtain the prediction that exerting higher cognitive effort, that is, conducting more steps of reasoning implies longer deliberation times. They key assumption then builds on the closeness-to-indifference effect, that is, decisions take longer the smaller the utility difference of the options. We assume that deliberation time for a given step is a decreasing function of the potential gain (or loss) of that step. This model provides empirically testable predictions regarding the relation of deliberation times, cognitive sophistication, as revealed by choices, and incentives.

We then test the predictions of our model using experimental data. We find longer deliberation times for choices associated with more steps of reasoning, confirming our prediction that deliberation time is increasing in cognitive effort. This link is strongest, when the payoff structure of the underlying game is such that iterative thinking is salient. However, for games without a salient iterative structure, there is no clear relation between deliberation times and cognitive effort. That is, features besides the best-reply structure matter as well, but deliberation times can serve as a tool to identify such situations. Further, we find effects of changes in the incentives that systematically vary the utility difference of a step of reasoning which are consistent with the predictions of our model. Hence, our results suggest that our closeness-to-indifference account can serve as a helpful tool to better understand processes of iterative thinking.

Overall, the answer to the question whether deliberation times support level-k models is "yes, but." If the underlying processes are clearly identified, we observe a clear link between deliberation times and steps of reasoning supporting level-k thinking. Additionally, however, deliberation times also allow us to detect when other elements enter the picture, and hence are also helpful for further theory development.

Appendix 3.A: Sequence of Games

To control for order effects we counter-balanced the order of the different 11-20 games using the following four randomized sequences. We denote the low cost, low bonus version of BASE, EXTR, FLAT, and SOCP by B, E, F, and SP, respectively. Similarly for $X \in \{B, E, F, S\}$ we use the notation $+X$ to indicate high cost, and $X+$ to indicate high bonus, e.g. $+B+$ denotes BASE with high cost and high bonus.

Table 3.9: Pseudo-randomized sequences of the 11-20 games used in the experiment.

Sequence 1	B	F	E	S	E+	B+	S+	F+	+F	+S	+B	+E	+S+	+E+	+F+	+B+
Sequence 2	+E	+B	+S	+F	+B+	+F+	+E+	+S+	S	E	F	B	F+	S+	B+	E+
Sequence 3	+F+	+S+	+B+	+E+	B+	F+	E+	S+	+S	+E	+F	+B	E	B	S	F
Sequence 4	S+	E+	F+	B+	F	S	B	E	+E+	+B+	+S+	+F+	+B	+F	+E	+S

Appendix 3.B: Behavior in the 11-20 Games

Figure 3.6 shows the distribution of choices in our experiment across all 16 instances of the 11-20 game including the four instances of the additional treatment SOCP described in Section 3.5.

Figure 3.6: Distribution of choices in the 11-20 games.

Chapter 4
Institutional History, Leniency and Collusive Tax Evasion

4.1 Introduction

Corruption and tax evasion are among the most pervasive forms of illicit behavior inducing negative externalities on both the economic and societal level (Banerjee, 2016a; Slemrod, 2007). Understanding their drivers and implementing suitable institutional measures to curb their severity has been at the center of the past decade's theoretical, empirical and experimental research.

In this paper, we focus on the effectiveness of providing legal immunity to the bribe-giver for blowing the whistle as a measure to deter collusive bribery. In our experiment corruption is embedded in a tax evasion framework, in which underreporting taxes is only possible through collusive cooperation among tax payers and public officials. We study the exchange of bribes as one explicit collaboration-inducing mechanism. This has previously been found to be effective in sustaining illicit cooperation. This literature also highlights the importance of studying the collaborative roots of deviant behavior due to their inherent negative economic and societal externalities (Weisel and Shalvi, 2015).

Our results shed light on the effectiveness of leniency programs as a means to distort collusive relationships between public officials and tax payers and to reduce tax fraud. We consider a mechanism that offers tax payers a "a safe way out" by blowing the whistle on the corrupt public official and cooperating with the auditors. This mechanism resembles a leniency program for tax evasion in which audited tax fraudsters can turn state's evidence. In many countries, the introduction of some form of leniency mechanism represents an integral institutional feature aimed at suppressing criminal behavior, for example, in the context of collusion among firms (Buccirossi and

Spagnolo, 2006; Abbink and Hennig-Schmidt, 2006; Bigoni et al., 2012). We are interested in examining the effects of a leniency program for tax payers on collusive bribery and tax evasion. We contribute to the corruption and tax evasion literature by shedding light on how collusive tax evasion is affected by the specifics of the strategic interaction between a tax payer and an intermediary, a dimension not present in a setting of individual tax evasion.

While most of the economic research in that literature has focused on deterrence of income tax evasion or its related variants, other forms of tax evasion, such as trade/import or custom taxes, where taxes are in some way collected through the direct intermediation of a third party (for example custom duties), have received little attention (Banuri and Eckel, 2012). This is particularly true for the case of "corruption within tax evasion." Existing experimental studies have for example focused on the role of the fear of being caught or public disclosure in deterring tax evasion (Orviska and Hudson, 2003; Bø et al., 2015).

In a related setting, Abbink and Wu (2017) study whether rewarding self-reports is effective in reducing collusive bribery. They find this mechanism to be effective in some circumstances, especially in a context of repeated interaction. However, they study different mechanisms with a focus on rewards for reporting, whereas we focus on the shift of the risk of being caught between two colluding parties. Further, in our experiment bribe-givers face two decisions, whether and how much to bribe, and also how much taxes to declare, which in turn determines the consequences of bribery. Christöfl et al. (2017) study the possibility to cooperate with the authorities (principal witness) in combination with a leniency policy that offers reduced fines for cooperation in a setup where two bidders compete for a contract. They find a lower number of bribes when a leniency policy is present, while at the same time offering a bribe becomes more profitable for a corrupt bidder. Closely related to our work, Heinemann and Kocher (2013) study the effects of regime changes on tax compliance, however, they focus on changes in the tax rate and consider neither corruption nor reforms that incentivize whistleblowing. By and large, the economics of whistleblowing are understudied and have only recently attracted attention (see Spagnolo, 2004; Apesteguia et al., 2007; Spagnolo, 2006; Heyes and Kapur, 2009; Breuer, 2013; Schmolke and Utikal, 2016). In particular, Butler et al. (2017) study the effectiveness of financial rewards and public scrutiny as triggers to motivate employees to blow the whistle against their managers. Their findings indicate that both financial rewards and public visibility increase the likelihood of whistleblowing (see also Bartuli et al., 2016). The recent surge in cases of whistleblowing and the lack of international institutional uniformity to achieve sufficient protection for whistleblowers renders the importance to further study the economics of whistleblowing (Dyck et al., 2010).

We use a controlled laboratory experiment modeling an income reporting scenario that requires the interaction between two parties, a tax payer and a tax officer, thus opening the door for collusive corruption. Our experimental design employs a collusive bribery game (Abbink et al., 2002) nested in a tax evasion scenario, in which corrupt tax officers face little to no consequences for accepting bribes and for providing assistance to the tax payer in order to evade taxes. This mimics a situation where tax authorities do not have the means to sufficiently control the tax officers, for example due to the institutional environment rendering enforcement of adequate consequences impossible. Excessive costs of monitoring are among the reasons why authorities might be unable to detect dishonest officers.

In the basic bribery game without leniency each tax payer receives a fixed income, taxed at a fixed rate, that has to be reported to the authorities represented by a tax officer. A distinct feature of our design is that underreporting requires the cooperation of the tax officer. Thus, the tax payer can offer the tax officer a bribe as reward for his assistance in evading taxes. Tax reports are subject to audits with a known probability.[1] Detection of tax evasion during an audit results in a penalty for the tax payer, but not for the tax officer. This game is then extended by adding an additional stage in the spirit of a leniency mechanism. The resulting bribery game with leniency

[1]Our focus is on the effectiveness of a leniency mechanism as a policy intervention, thus we decided to keep a fixed audit probability instead of implementing an endogenously determined audit probability, e.g. by modeling the tax authority as an additional player. For a theoretical analysis of endogenous audit probabilities see Landsberger and Meilijson (1982) or Raymond (1999). For an experimental treatment see Alm et al. (1993).

follows the same rules except for the situation when an audit occurs. In that case a tax payer can report the corrupt tax officer and avoid the pending penalty. Instead, now the reported tax officer incurs a fine. This whistle-blowing mechanism offers a save way out for the tax payer and shifts the risk of a being detected and fined to the tax officer. Moreover, it renders the tax officer formally responsible for engaging in collusive bribery as she now faces the threat of a fine as well.

The goals of this study are twofold: First, we seek to analyze collusive bribery and its drivers under a regime with and without leniency for the tax payer in a setting where corruption is feasible due to the interaction between tax payers and tax officers. Second, we investigate the effectiveness of the introduction of such a mechanism and the consequences of its removal on collusion, the frequency of bribe offers and their size, the tax officers' willingness to accept bribes as well as overall tax compliance.

Our main results can be summarized as follows. We find that in the presence of a leniency mechanism successful collusion between tax payers and tax officers is less frequent and this is mainly driven by a lower willingness of tax officers to accept bribes. Further, we find no support that leniency for tax payers encourages them to offer bribes, that is, there is no significant increase in the frequency of bribes being offered. Thus, our results suggest that leniency is effective in deterring tax officers from engaging in bribery and that this translates into more taxes being collected. Our results regarding the role of institutional changes also highlight the importance of institutional history for the evaluation of policy measures. We show that the introduction of the opportunity to blow the whistle decreases collusion, deters tax officers from accepting bribes, as reflected in a lower acceptance rate of bribe offers and increases the tax yield collected, while at the same time it does not encourage bribe offers. In contrast, the removal of the institutional mechanism does not cause similar effects in the opposite direction, which suggests a positive spillover effect of leniency that persists even after the mechanism has been removed (see also d'Adda et al., 2017).

The paper is organized as follows: Section 4.2 describes the experimental design. Section 4.3 presents the analysis of our empirical results. In Section

4.4 we discuss our results and conclude.

4.2 Experimental Design

Both of our institutional setups mimic a scenario where collusive bribery is nested in a tax evasion framework. Taxes are collected through an intermediary, the tax officer. Hence, to successfully evade taxes the tax payer requires the cooperation of the tax officer, for example by "looking the other way." We now give a detailed description of the two institutional frames used in our experiment.

4.2.1 The Bribery Game with and without Leniency

The upper part of Figure 4.1 illustrates the bribery game (BG).[2] A tax payer (TP) receives an income of 80 Experimental Currency Units (ECU) and has to submit a declaration of his income to the tax authorities. The tax officer (TO), acting as an intermediary, is in charge of processing the tax report. Declared income D is subject to a tax rate of 50%.[3] The TP can decide whether he wants to truthfully declare his full income of 80 or whether he wants to evade taxes, that is, potentially declare a lower income $D \leq 80$. In order to evade taxes, the TP has to convince the TO to collude with him. To that end, the TP can offer a bribe b to the TO that can range from 0 to 30 ECU. The situation we have in mind is one, where the TP can vastly increase the chance of his false tax declaration not being detected by colluding with the TO, who is in charge of processing the report. For simplicity, we assume that it is impossible for the TP to evade taxes without the TO's support. That is, declaring less than the full income is only possible if the TO accepts the TP's bribe offer and hereby agrees to collude with the TP, e.g. by manipulating the report. If a bribe is offered, the TO observes the

[2]The TO only observes the bribe b but not the declared income D as indicated by the dashed line. The stage below the dotted line is only available in the bribery game with leniency. For the sake of a simpler exposition the tax officer's fixed wage of 50 is not depicted.

[3]Subjects were informed that this tax rate is in line, according to a recent study of Confcommercio, with the mean tax burden in Italy.

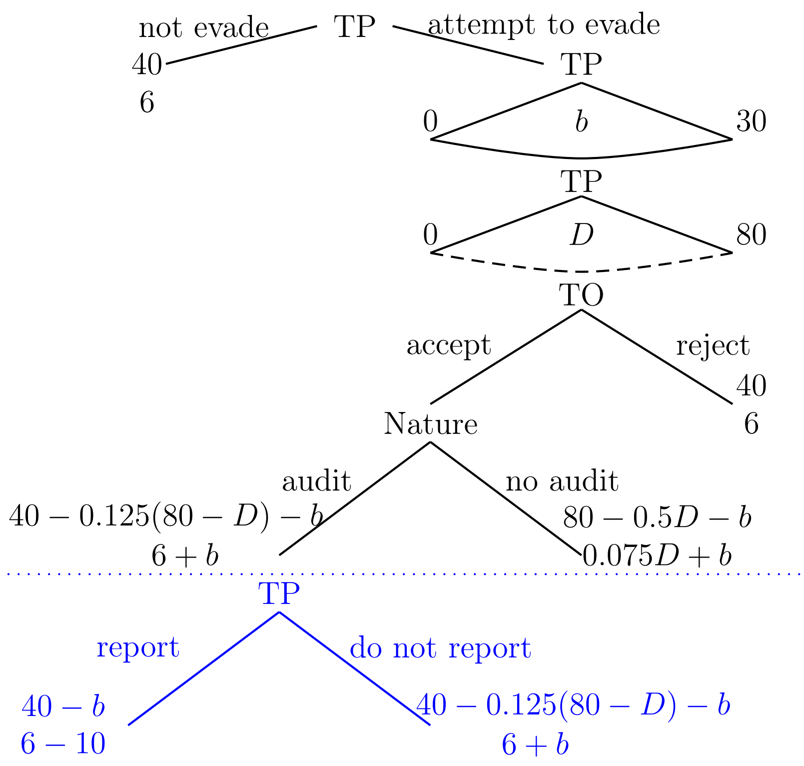


Figure 4.1: Representation of the bribery game (with leniency).

amount that is offered and can accept or reject it. It is important to note that the TO cannot observe the amount of taxes declared prior to his decision, hence cannot condition her decision on the amount of taxes evaded.[4] Not informing the TO about the exact amount the TP intends to evade allows us to establish a minimal level of uncertainty regarding the TO's payoffs, which are fully determined (in the absence of whistleblowing) by the bribe and the amount of taxes declared as described in more detail below. If the TO rejects a bribe, then she refuses to collude with the TP, which forces the TP to truthfully declare his full income of 80. Upon acceptance the TP is able file the original report declaring D.[5] Tax reports are audited by the tax authorities with an exogenous probability of 20%. In case of an audit

[4]For example, imagine a situation where the TO does not know the actual income of the TP, which is only known to the official tax authority conducting the audits.

[5]Note that this differs from Abbink and Wu (2017) in that the tax officer is not able to pocket the bribe without delivering the corrupt favor of colluding with the tax payer.

incorrect reports are detected and the TP has to pay both the evaded amount of taxes $0.5(80 - D)$ and an additional fine proportional to the amount of evaded taxes.[6] The fine is set to 25% evaded taxes, hence the maximum fine is 10 ECU. The fine rate of 25% was chosen such that together with the upper bound (of 30) of bribe payments the TP can never incur a net loss. Thus, the TP's payoff is his income minus taxes on the declared income D and potentially the bribe and/or fine paid. The TO's payoff consists of three components: a fixed wage of 50, a commission of 15% of the taxes collected, and the amount of bribes accepted.[7]

The bribery game with leniency (BGL) is very similar to the bribery game just described, but with one important difference. In the BGL we add an additional stage to the BG intended to mimic a leniency program for blowing the whistle.[8] Decisions in BGL are identical to those in BG, however, following detection of an incorrect tax report during an audit the TP now has the opportunity to "blow the whistle" by reporting the TO. If the TP chooses to report, he has to correct the (false) tax report and declare taxes truthfully, however, he does not incur an additional monetary punishment, that is, the fine is waived. A TO that has been reported, on the other hand, incurs a fine for colluding with the TP to evade taxes. This fine equals the bribe received from the TP plus an additional penalty of 10 ECU.

In Appendix 4.A we analyze the one-shot bribery game with and without a leniency mechanism assuming standard preferences based on maximization of own payoffs. Under that assumption, attempting collusion, that is, bribing the tax officer and evading taxes, is always optimal for a tax payer in the bribery game with and without leniency. In the bribery game with leniency the tax payer always reports the tax officer in equilibrium, resulting in a

[6]Proportional fines are an institutional feature often observed in developed countries (Mittone, 2006).

[7]The introduction of a commission for the TO mimics something existing in reality. In Italy for example the tax authority delegates inspections and audits to a private organization (Equitalia) and pays Equitalia with a percentage of the money collected.

[8]In the BG punishment can be viewed as asymmetric as only tax payers are running the risk of being fined, however, in BGL leniency shifts, at least partially, this risk towards the tax officer, hence creating a situation that might be perceived as more symmetric. See also Engel et al. (2013) for a discussion of symmetric vs asymmetric punishment regimes.

higher bribe acceptance threshold on the side of the tax officer. As a result optimal bribe offers are higher when leniency is in place. Recall that the tax officer is only able to observe the bribe but not the amount of taxes declared, hence her exact acceptance threshold also depends on her belief regarding the amount of taxes declared by the tax payer. The threshold is increasing in her belief regarding the amount of taxes evaded, since this negatively affects her payoff. As a consequence there are many equilibria involving different levels of tax compliance by the tax payer and beliefs by the tax officer. For example there is one equilibrium in which the tax payer declares zero taxes and the tax officer correctly anticipates this behavior, hence expects declared taxes to be zero.

4.2.2 Treatments

One can to think of the introduction of a leniency mechanism as a stylized situation where tax authorities decide to invest in establishing control mechanisms that allow for better monitoring of public officials. Hence, allowing them to enforce legal consequences not only on tax payers but also on corrupt tax officers, for example via improved monitoring. We mimic transitions of that type by employing not only static treatments, where exactly one regime is present for the whole duration of the experiment, but also dynamic treatments involving a regime change from one to the other. This allows us to study both the effectiveness of either setup in isolation and how subjects react to a change in either direction. For example, we are interested in whether the transition from a scenario without the opportunity to blow the whistle to a situation in which this is feasible can break collusive behavior established during a earlier periods. If that is the case, then this would provide strong evidence that such a measure can serve as a tool to reduce collusive corruption and tax evasion in a world where the absence of such a mechanism is the status quo.

In our experiment subjects repeatedly played the bribery game and/or its extended version (with leniency) over the course of a total of 20 rounds. We ran four different treatments. In treatment *NoLEN*, participants play the bribery game without leniency for 20 rounds. In treatment *LEN* subjects play the bribery game with leniency instead, also for 20 rounds. These two treatments allow a between-subject comparison of the role that leniency plays with respect to collusive bribery and tax compliance. In addition, these treatments represent a benchmark for treatments *NoL-L* and *L-NoL* in which institutional shocks occur. These treatments were designed to study the effects of institutional transitions, e.g. potential spillover effects from one regime to another, since in those treatments the rules of the game change unannounced midway through the experiment after round 10. In particular, in treatment *NoL-L* subjects start with the basic bribery game and are then transitioned into an environment in which reporting the tax officer becomes feasible. Treatment *L-NoL* captures the same dynamics but in reverse order, that is, first the option to report is available and is then abolished after round 10. These two treatments involve a regime change that allows us to analyze the effectiveness of both the introduction and the removal of leniency relative to a "status quo," that is, the regime present during the first block of 10 rounds. Table 4.1 summarizes the four treatments.

Table 4.1: Overview over the treatments and number of subjects assigned to each treatment.

Treatment	Round 1-10	Round 11-20	Tax Payers	Tax Officers
NoLEN	BG	BG	30	10
LEN	BGL	BGL	36	12
NoL-L	BG	BGL	66	22
L-NoL	BGL	BG	42	14

4.2.3 Experimental Procedures

Subjects were randomly assigned either the role of a tax payer or the role of a tax officer. Participants were randomly matched in groups of four consisting of one tax officer and three tax payers, that is, each tax officer assigned three tax payers to interact with simultaneously. There was no direct interaction between different tax payers in the same group. Groups remained fixed throughout the experiment, which consisted of 20 rounds. Subjects were informed that the number of rounds was predetermined, but were not informed about the exact number of rounds.[9] In each period subjects played, depending on the treatment, the bribery game with or without leniency. For treatment *NoLEN* and treatment *LEN* no institutional change occurred. In treatments *NoL-L* and *L-NoL* the participants were informed about a change in the institutional setting after the 10th round via an announcement on screen that provided a detailed description of the new institutional environment. We emphasized that there would be no additional change of the institution until the end of the experiment. Subjects were informed in the instructions that the existing institution may be subject to change but no information regarding the nature of the change was provided.[10] Thus, we use both within- and between-subject variations of the institutional setting to study the effect of leniency on corruption and tax compliance. Since initial tax declarations were not observable by the tax officer, we also elicited the tax officer's beliefs about the amount of taxes evaded by each of the tax payers offering a bribe. Beliefs were elicited in each round after the tax officer's decision to accept or reject a bribe offer, but before any feedback regarding the outcome of this round was provided. At the end of each round tax officers were informed about whether they were reported, how much they were fined (if at all) and how much they earned from the tax yield collected. Tax payers received information regarding whether their bribe was accepted, whether they were audited and how much (if at all) they were fined.

To make tax evasion more salient in the laboratory setting, we introduced a third party that incurs a monetary damage as a result of tax evasion. All participants were informed that the total tax yield collected would be used to finance future research of doctoral students at the University of Trento.[11] That is, tax evasion in the experimental laboratory translates into an actual

[9] We choose not to announce the number of rounds to avoid potential end-game effects.

[10] Subjects in treatments *NoLEN* and *NoL-L* were provided with identical information at the start of the experiment. The same holds for subjects in treatments *LEN* and *L-NoL*. In particular, participants assigned to treatment *NoLEN* and *LEN* were informed about the possibility of a change although, ultimately, they would not experience one.

[11] This is a common procedure in tax evasion experiments in order to link tax evasion to a negative externality, for example see Fortin et al. (2007) or Coricelli et al. (2010).

social welfare loss outside the lab (Eckel and Grossman, 1996; Lambsdorff and Frank, 2010).

The experiment was conducted at the Cognitive and Experimental Economics Laboratory at the University of Trento. A total of of 268 undergraduate students (46% females) participated in the experiment, each in exactly one treatment. Table 4.1 shows the distribution of subjects over the four experimental treatments. Sessions consisted of 20 rounds followed by an incentivized risk-elicitation task (Holt and Laury, 2002b) and a demographic questionnaire. The final payoff of each subject was determined as the sum of all earnings over the 20 rounds plus their earnings from the risk-elicitation task, which were then converted to Euro at a rate of 100 ECU = €0.7. All participants were paid their final payoff plus an additional show-up fee of €3 in cash at the end of the experiment. On average, a session lasted about 60 minutes and subjects earned €12 excluding the show-up fee of €3.

4.3 Results

An important feature of our experiment is that tax evasion is nested within a corruption framework that requires collusive behavior for tax evasion to be successful. We believe that this additional layer of interaction is important to help us to better understand unethical behavior in situations in which cooperation is necessary. This interaction possibly increases the impact of behavioral factors such as psychological costs and uncertainty on tax compliance and the willingness to engage in collusive bribery.

We structure our analysis in the following way: first, we will discuss the effectiveness of leniency in affecting collusive agreements between public officials and tax payers. In a next step, we will break down the behavior of tax payers and public officials individually. We employ a very cautious approach in our data analysis. Following our design, we regard the behavior of one group (consisting of one public official and three tax payers) averaged over all rounds, or over all rounds in the first and second part, respectively, as one independent observation. This allows us to conduct between- as well as within-subject comparisons.

4.3.1 Collusive Behavior

One of our main objectives is to study the effectiveness of a leniency mechanism as a means to hinder collusive corruption. The presence of the possibility to blow the whistle effectively reduces the risk the tax payer faces when evading taxes, while shifting responsibility to the tax officer, thus potentially reducing the tax payer's psychological costs. Intuitively, in the bribery game with leniency the possibility to report the tax officer offers the tax payer a "safe way out". Leniency effectively allows the tax payer to avoid an additional fine when evading taxes, and if the fine is what is keeping a tax payer from offering a bribe and evading taxes this should encourage the tax payer to engage in collusive bribery. On the other hand, leniency also affects the chances that an attempt to collude is successful, since this requires the cooperation of the tax officer, who now faces the additional threat of being reported. Thus, leniency is likely to decrease the tax officers willingness to engage in collusive corruption. It is unclear which of these opposing effects will dominate.

In line with our primary interest to study the effectiveness of leniency on collusive arrangements, our experimental design allows us to approach this question from two perspectives:

1. Is collusion generally different in an environment where leniency exists?
2. How does an institutional change from an environment with (without) leniency to an environment without (with) leniency affect collusive behavior?

To address these two questions, we compare the rate of collusion between *NoLEN* and *LEN*, between *NoLEN* and *NoL-L*, and between *LEN* and *L-NoL*, respectively. We define collusion as the successful exchange of bribes in return for the avoidance of taxes.

We first analyze the effect of leniency in absence of institutional history by comparing treatments *NoLEN* and *LEN*. To that end, we calculated for each group the collusion rate as the proportion of successful illicit agreements

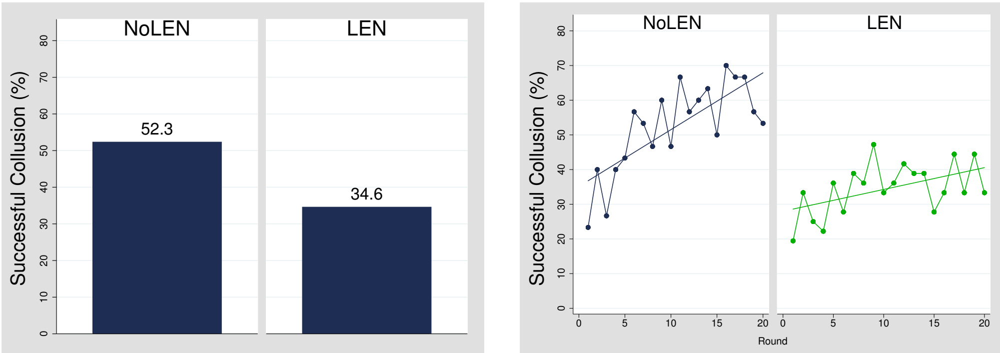


Figure 4.2: Average collusion in *NoLEN* and *LEN*.

relative to all rounds in which paying a bribe and evading taxes was possible. Figure 4.2 shows the average collusion rate for each treatment as well as the evolution of the average collusion rate, calculated for each round. In the *NoLEN* treatment, the average collusion rate per group was 52.3%. In contrast, in the presence of a leniency mechanism the incidence of collusion was only 34.6% in the *LEN* treatment. This difference is significant according to a Mann-Whitney-Wilcoxon test ($N = 22$, $z = 2.1448$, $p = 0.0320$) indicating that collusion is less frequent in *LEN* compared to the *NoLEN*. The right panel of Figure 4.2 suggests that collusion is increasing over the course of the experiment in both treatments. To check whether this is indeed the case we calculated for each group the average collusion rate for the first and second half of the experiment, separately. In the first part of *NoLEN* the average collusion rate is 43.7% and rises to 61% in the second half. This difference is significant according to a Wilcoxon Signed Rank test ($N = 10$, $z = -2.6711$, $p = 0.0076$). In *LEN* the collusion rate is 31.9% and 37.2% in first and second half of the experiment, respectively, with the difference not being significant (WSR, $N = 12$, $z = -1.2183$, $p = 0.2231$). With respect to the first question, we find less collusion when a leniency mechanism is in place compared to an environment without such a mechanism. Moreover, we find that collusion is increasing significantly over time in the *NoLEN* treatment, while there is no significant increase in *LEN* when leniency is in place.

Next, we turn to our second question regarding the effect of an institutional change on collusion. Treatment *NoL-L* allows us to study the effect of the introduction a leniency mechanism into a setting in which corrupt behavior has already been able to thrive in the absence of leniency. There is some evidence that fear of being reported has a deterrent effect, which might decrease the tax officer's acceptance rate (Engel et al., 2013; Abbink et al., 2014). On the other hand, we are able to study whether a period in which leniency was implemented affects behavior even after it was removed, for example because a successful relationship is harder to build after developing mistrust in earlier periods. In treatment *L-NoL* subjects start under a regime with leniency followed by its removal. Following the same logic one would expect low acceptance rates in the first part when facing the bribery game with leniency, but an increased acceptance rate as a result of the removal of the mechanism in the second part of treatment *L-NoL*.

The left panel of Figure 4.3 shows the average rate of collusion for each part of treatment *NoL-L* and *L-NoL*. The right panel illustrates how collusion evolves over the course of the experiment in each of the treatments. We again observe that collusion is increasing over time, moreover, the graph suggests that the introduction of a reporting option in *NoL-L* causes a drop in collusion. Since collusion is increasing over time, we cannot simply compare the means before and after an institutional change has occurred. Thus, we evaluate the effect of the introduction or removal of a leniency mechanism by comparing the change in collusion rates resulting from the introduction or removal of leniency to the corresponding change in the absence of an regime change. Thus, we calculated the change of the collusion rate between the first and second half of the experiment for each group in all treatments. We then compare the changes between *NoLEN* and *NoL-L*, and between *LEN* and *L-NoL*, respectively. This difference in differences analysis is necessary to account for the increase in collusion over time.

In treatment *NoL-L* the average collusion rate before and after a leniency mechanism was introduced are 43.2% and 41.7%, respectively. Hence, the introduction of a leniency mechanism in *NoL-L* results in a decrease of collusion by 1.5 percentage points. Recall that in *NoLEN* there was an increase in collusion by 16.3% from the fist part to the second part. Comparing the change from part one to part two between *NoLEN* and *NoL-L* reveals that the

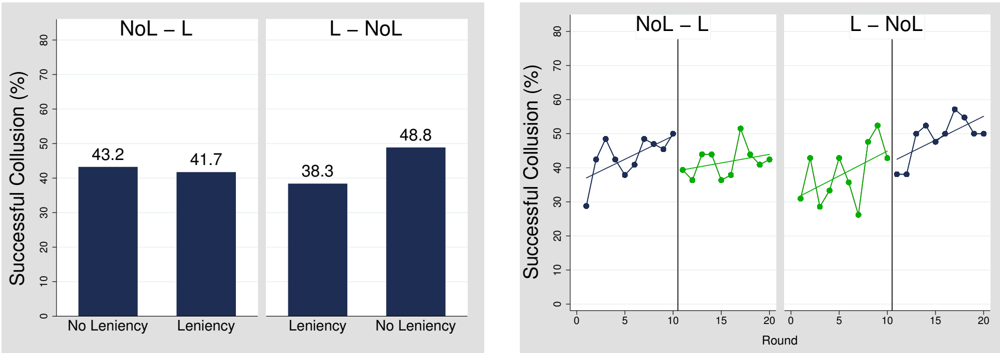


Figure 4.3: Average collusion in *NoL-L* and *L-NoL*.

introduction of leniency has a significant negative effect on collusion (MWW, $N = 32$, $z = 2.2249$, $p = 0.0261$).

Similarly, we now consider the effect of the removal leniency. In the first part of treatment *L-NoL* the collusion rate is 38.3% when leniency is present and following its removal rises to 48.8% in the second part. Thus we observe an increase in collusion by 10.5 percentage points in *L-NoL* compared to an increase of 5.3 percentage points in *LEN* from part one to part two. There is no significant difference between the increase of collusion in *LEN* and *L-NoL* (MWW, $N = 26$, $z = -0.6978$, $p = 0.4853$). Thus we find no evidence that the removal of the possibility to blow the whistle does lead to an additional increase in collusion that goes beyond the gradual increase over time observed in *LEN* in the absence of a regime change. In particular, there is no upwards "jump" in the frequency of collusive cooperation following the removal of the leniency mechanism.

Summarizing, our results suggest that the presence of a leniency mechanism indeed deters collusion. Interestingly, we also see some evidence for an increase in successful collusive cooperation over time in the absence of leniency, while under leniency we see no such effect. This is in line with the idea that leniency makes it more difficult to reach a collusive agreement, that is honored by both parties. Regarding the effects of a regime change we find that the introduction of a leniency mechanism in treatment *NoL-L* has a deterrent effect on collusion. This result suggests that implementing such a measure is likely to hinder collusive bribery. On the other hand, the removal of whistleblowing in *L-NoL* does not foster collusion. That is, collusion rates show no significant "jump" upwards after the mechanism is removed. This points towards a potential positive spillover effect from the first part, where a leniency mechanism was in place, that persists even after its removal. A potential explanation for this spillover effect is that leniency saws mistrust between the tax officer and the tax payer, hence reduces the tax officer's willingness to cooperate also in later periods although reporting is not feasible anymore.

4.3.2 What Are the Drivers of Collusion?

Collusion requires the cooperation of both, the tax payer and the tax officer. In order to pin down the drivers of the effects on collusion found in the previous section we now analyze the behavior of tax payers and tax officers separately. To that end, we first consider the rate of collusion attempts initiated by the tax payer, that is, the incidence of bribe offers relative to all relevant situations where offering a bribe was feasible. Since collusion requires the cooperation of the tax officer, in a second step we investigate the bribe acceptance rate, that is, the fraction of bribes that were accepted by the tax officer relative to the number of bribe offers received. Clearly, collusion is the result of a combination of both, the frequency of bribe offers and the fraction of bribe offers that are accepted. Moreover, the size of the bribes is likely to affect the acceptance rate, since it is natural that tax officers accept large bribes more often than small bribes. Hence, we also consider the treatment effects on the size of the bribes offered by the tax payer and how they affect the acceptance rate.

In the absence of leniency the tax payers decision to collude with the tax officer and evade taxes comes at the risk of being detected and fined. The presence of a reporting opportunity effectively reduces this risk, while shifting responsibility to the tax officer. This not only renders tax evasion more profitable, but also potentially reduces the tax payer's psychological cost associated with paying a bribe in order to evade taxes. Intuitively, leniency offers the tax payer a "safe way out" when getting caught, hence, they are likely to offer bribes more frequently. On the other hand, accepting a bribe is more risky for a tax officer when a leniency mechanism is in place, since she now faces the threat of being reported and fined. Thus, we expect tax officers to reject more bribes when blowing the whistle is possible.

As in the previous section, we first seek to understand the drivers of collusion in treatments *NoLEN* and *LEN*, where there was no regime change. In a second step we analyze the role of the institutional history, that is, the effect of the introduction of a leniency mechanism in *NoL-L* and the effect of the removal of such a mechanism in *NoL-L*.

Incidence of bribe offers and acceptance rate in NoLEN and LEN

We first consider the behavior of the tax payer in that we analyze the incidence of bribe offers. The left panel of Figure 4.4 shows the dynamics of the average frequency of bribe offers per round over the course of the experiment. Surprisingly, we see that bribe offers are not more frequent but rather less frequent in *LEN* compared to *NoLEN*. In fact, the average incidence of bribe offers per group over all rounds was 67.7% in *NoLEN* and 55.1% in *LEN* and thus even lower in the presence of leniency. However, this difference fails to reach significance according to a Mann-Whitney-Wilcoxon test ($N = 22$, $z = 1.4870$, $p = 0.137$). Further, the graphs suggest that the frequency of bribe offers is increasing over time in *NoLEN*, whereas it appears to be slightly decreasing in treatment *LEN*. According to Spearman rank order correlations there is a positive trend in *NoLEN* ($\rho = 0.3803$, $p < 0.001$), whereas the bribe offers exhibit a negative trend in *LEN* ($\rho = -0.2496$, $p < 0.001$). For both treatments we again calculated the average frequency of bribe offers for the first and second half of rounds, separately. In *NoLEN* the average incidence of bribe offers is 68.3% in rounds 1-10 and 73.7% in rounds 11-20 with the difference not being statistically different. In *LEN* the frequency of bribe offers is with 56.6% in the second half of the experiment slightly lower than in the first ten rounds, where it is 58.1%. Again this difference is not statistically significant.

In treatment *LEN* the tax payer not only faces less risk than in treatment

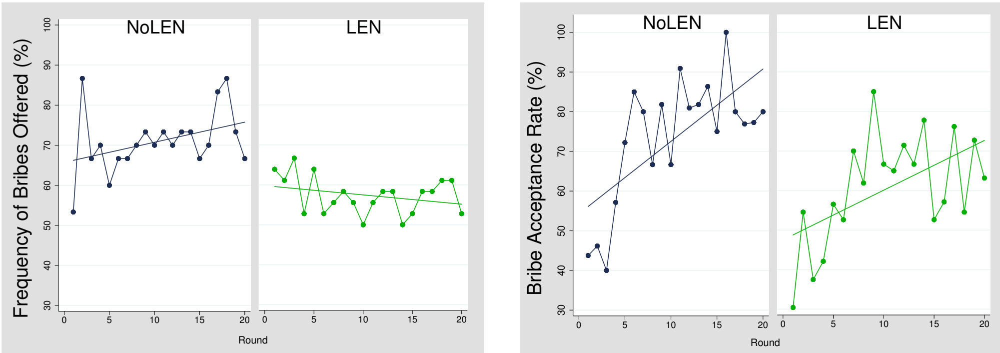


Figure 4.4: Frequency of bribes offered and bribe acceptance rate in *NoLEN* and *LEN*.

NoLEN but this risk is also effectively shifted to the tax officer as leniency exposes her to the possibility of being reported and fined. In Appendix 4.A we show that this raises the optimal bribe acceptance threshold in equilibrium. A failure of the tax payers to acknowledge this increased risk for the tax officers is likely to result in more rejections of bribes. Next, we consider the behavior of the tax officer, more precisely we look at the average fraction of bribes accepted by tax officers. The right panel of Figure 4.4 shows the evolution of the bribe acceptance rate over the 20 rounds for the two treatments without a regime change. The graphs indicate a higher acceptance rate in *NoLEN* compared to *LEN* and clearly show that the acceptance rate is increasing in both treatments over time. The average acceptance rate in *LEN* is 58.8% and thus lower compared to the average acceptance rate of 73.2% in *NoLEN*, however, this difference fails to reach significance (MWW, $N = 22$, $z = 1.51657$, $p = 0.1294$). Spearman rank order correlations confirm our observation of a significant positive trend in both treatments that is of about the same magnitude in *NoLEN* ($\rho = 0.5245$, $p < 0.001$) and *LEN* ($\rho = 0.5073$, $p < 0.001$). Comparing the average acceptance rate for the first ten rounds with the average acceptance rate in the second part, we find a significant increase in *NoLEN* from 64.4% to 81.6% (MWW, $N = 10$, $z = -2.7557$, $p = 0.0059$) as well as an significant increase in *LEN* from 52.9% to 65.0% (MWW, $N = 12$, $z = -2.353393$). Thus, our results suggest that the acceptance rate is increasing over time in both treatments without a regime change.

In combination these results suggest that the decrease collusion in *LEN* compared to *NoLEN* is likely the result of both, the absence of an increase in the number of bribes offered by the tax payer (which are even slightly less frequent, but not statistically significant) and a reduced acceptance rate by the tax officer in *LEN*, which, however, fails to reach significance. The increase in collusion over time seems to be mainly driven by an increase in the acceptance rate of the tax officer, especially for *LEN* where the number of bribes offered is even slightly decreasing. In *NoLEN* there is a positive trend also for the frequency of bribe offers, which might explain why in *NoLEN* collusion seems to be increasing more rapidly than in *LEN*.

Incidence of bribe offers and acceptance rate in NoL-L and L-NoL

We now study the effects of the introduction and the removal of a leniency mechanism that allows for whistleblowing on the frequency of bribe offers by the tax payer and on the bribe acceptance rate by the tax officer. The left panel of Figure 4.5 shows the dynamics of the frequency of bribe offers in treatments *NoL-L* and *L-NoL*. We observe that there is an upward jump in the frequency of bribe offers after the introduction of leniency in *NoL-L*, from that point on we see a steep decrease until the end of the experiment. Overall bribe offers seem to be more frequent in the presence of leniency for both treatments, but more so for *L-NoL*. In the latter there is a positive trend before and after the removal of the reporting option, but the frequency of bribe offers drops sharply.

In treatment *NoL-L* the average frequency of bribe offers is 65.0% in the first part and rises by 5.5 percentage points to 70.5% in the second part when a leniency mechanism is introduced. This increase is identical to the increase observed in *NoLEN* where no regime change occurred. A Mann-Whitney-Wilcoxon test comparing the increase in *NoL-L* to the increase in *NoLEN* from part one to part two confirms this observation ($N = 32$, $z = 0.0001$, $p = 1.000$). We conclude that the introduction of a leniency mechanism has no significant effect on the average incidence of bribe offers. However, Figure 4.5

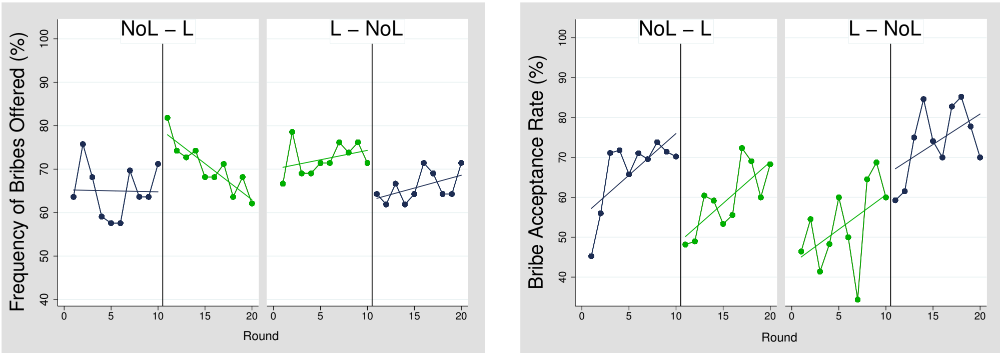


Figure 4.5: Frequency of bribes offered and bribe acceptance rate in *NoL-L* and *L-NoL*.

suggests that the introduction of whistleblowing affects the dynamics of bribe offers over rounds. Spearman rank order correlations reveal that during the first part of *NoL-L* bribe offers show no clear trend ($\rho = 0.0369$, $p = 0.5859$), whereas there is a significant negative trend following the introduction of a reporting mechanism ($\rho = -0.8924$, $p < 0.001$).

We now turn to the effect of the removal of leniency on bribe offers. In treatment *L-NoL* on average tax payers offered bribes in about 72.5% of all cases when whistleblowing was possible. This rate is 6.5 percentage points higher than in the second part, where this number falls to 66.0% following the removal of the mechanism. This change is very close to the decrease by 1.5 percentage points observed in *LEN* and indeed the difference in the effects from the first part to the second part between *LEN* and *L-NoL* is not statistically significant (MWW, $N = 26$, $z = 0.7232$, $p = 0.4696$). We also observe from Figure 4.5 that there is a positive similar positive trend, both before ($\rho = 0.4075$, $p < 0.001$) and after the removal of the leniency mechanism ($\rho = 0.4909$, $p < 0.001$).

Next, we consider how the behavior of tax officers, as revealed by the average acceptance rate of bribe offers, is affected by the introduction and the removal of a leniency mechanism. The right panel of Figure 4.5 shows the evolution of the acceptance rate of tax officers over the course of the experiment for treatments *NoL-L* and *L-NoL*. The graphs suggest that tax officers accept less bribes after the introduction of leniency, which is likely due to the potential risk of being reported and fined. Our findings in treatment *NoL-L* indicate that the average acceptance rate of bribes decreases from 64.4% to 59.7% following the transition to an institutional environment with leniency. Recall that in *NoLEN*, where no such measure was introduced, we have seen that the acceptance rate increases by 17.2 percentage points from the first part to the second part. The difference in the change between parts across *NoLEN* and *NoL-L* is highly significant (MWW, $N = 32$, $z = 3.0910$, $p = 0.0020$). This result suggests that the introduction of a reporting option is an efficient deterrent for the tax officer as reflected by a stark negative effect in bribe acceptance rates. Moreover, the dynamic pattern in *NoL-L* confirms our earlier observation that acceptance rates are increasing over time independently of the presence of a leniency mechanism.

The removal of the reporting mechanism in *L-NoL* appears to have a different effect as revealed by the dynamics in the right panel of Figure 4.5. Acceptance rates are increasing over time in a similar fashion as we have observed in treatment *LEN* where the mechanism was not removed. Most importantly, in treatment *L-NoL* the dynamics does not indicate any behavioral change in acceptance rates from the first to the second part, but only a steady increase over time. The average acceptance rate increases from 52.2% in part one where whistleblowing was possible to 74.2% in the second part without such a mechanism. This increase is not statistically different from the increase observed in *LEN* (MWW, $N = 26$, $z = -0.4115$, $p = 0.6807$). Thus, we find no evidence that the removal of leniency significantly increases acceptance rates. Further, also in treatment *L-NoL* the dynamic pattern over the course of the experiment confirms that acceptance rates are increasing as subjects gain more experience.

We find no evidence that the introduction of leniency for the tax payer has a strong effect on bribe offers, in particular that leniency encourages tax payers to offer bribes more frequently is not supported by our data. At most, there is weak evidence for a temporary increase in bribe offers following the introduction of leniency, but this is coupled with a sharp and steady decrease over later periods. On the other hand, our data suggests that the introduction of a whistleblowing mechanism, that renders the tax officer formally responsible, is able to discourage tax officers from accepting bribes. Moreover, we find no evidence that the removal of such a mechanism triggers an effect in the opposite direction, that is, acceptance rates show no significant jump upwards when the threat of whistleblowing is removed, which indicates a positive spillover effect of whistleblowing. We find consistent evidence for a general increase of acceptance rates over time that is independent of the presence of a leniency mechanism. Thus, our results identify a deterrent effect of leniency on tax officers as the driver behind the effects on collusion rates reported in Subsection 4.3.1. That is, tax officers reject more bribe offers after leniency is introduced and do not accept more bribes when it is removed. Moreover, this effect outweighs any potential encouragement for tax payers to offer more bribes under leniency, for which our data offers only limited support.

Effects of Bribe size, Reporting and Beliefs on the Bribe Acceptance Rate

Let us now consider the amount of bribes paid. Recall that in the bribery game with leniency, the optimal bribe acceptance threshold is higher, hence in order to sustain collusion the tax payer has to compensate the tax officer for the additional risk with higher bribe payments. As shown in Appendix 4.A in equilibrium bribe payments are by about 3.7 ECU higher in the bribery game with leniency compared to when it is absent. In line with these theoretical predictions we observe that average bribe payments are 14.4 ECU in treatment *NoLEN* compared to 16.6 ECU in treatment *LEN*. Although this difference is smaller than predicted we find that the difference is statistically significant (MWW, $N = 22$, $z = -1.7808$, $p = 0.0749$). Similarly, there is a significant upwards shift in the size of bribes paid following the introduction of a whistleblowing mechanism in treatment *NoL-L* from 13.2 ECU to 15.4 ECU (WSR, $N = 22$, $z = -2.3538$, $p = 0.0186$). Analogously, bribe payments are 16.6 ECU during the first part of treatment *L-NoL*, whereas they decrease to 15.2 ECU following the removal of whistleblowing, however, this difference is not statistically significant. Evidently, taxpayers acknowledge the higher risk that public officials have to bear in the presence of a

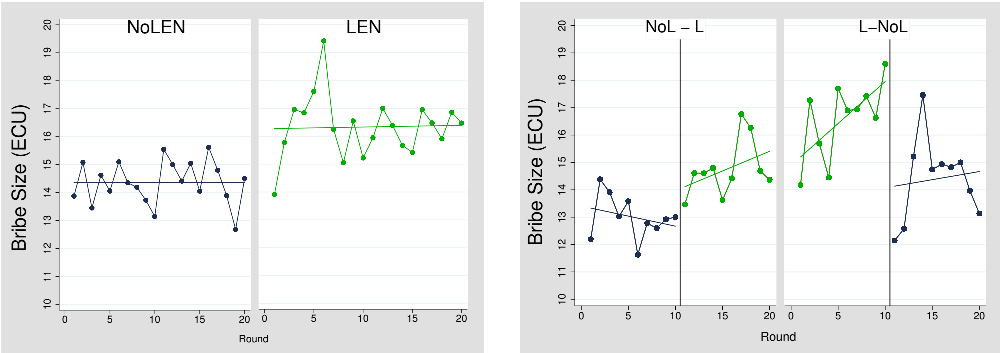


Figure 4.6: Bribe size over rounds across treatments.

leniency mechanism and as a consequence compensate them, at least partially, with higher bribes. It is important to note that for a tax officer all bribe payments above 6 ECU, respectively 8.16 ECU, are profitable in the presence, respectively absence, of leniency. Bribe offers below the respective threshold occurred only in about 10% of the cases in both treatments *NoLEN* and *LEN*, hence those were relatively rare. Non-profitable bribe offers were slightly more common but equally likely in treatments *NoL-L* and *L-NoL* occurring in 15.5% and 16.3% of all cases, respectively. Hence, differences in the frequency of non-profitable bribe offers cannot explain the effects of leniency on collusion and bribe acceptance rates.

Figure 4.6 illustrates the evolution of bribe payments over the course of the experiment across treatments. In both *NoLEN* and *LEN* the size of bribe payments remains fairly constant over time, apart from an initial adjustment period during the first five rounds of treatment *LEN*. In treatment *NoL-L* bribe payments show some positive trend following the introduction of the reporting option. There is a similar trend in treatment *L-NoL*, but also only in the presence of leniency. In our setting, a tax payer's decision on whether to evade taxes goes hand in glove with the decision to pay a bribe and make the tax officer look the other way. Ceteris paribus, higher bribe payments should naturally lead to higher collusion rates. To test whether this is indeed the case we ran a logistic panel regression with random effects and standard errors clustered at the group level for each treatment separately. The dependent variable is whether a bribe was accepted or not, that is, any instance of a bribe offer is one observation. We include the size of the bribe offer as an independent variable and for treatments *NoL-L* and *L-NoL* we also include a dummy for the presence of a leniency mechanism and the interaction with bribe size. Table 4.3 in 4.4 reports the results of these four regressions. The regression results show that larger bribes are more likely to be accepted by the tax officer across all treatments and independent of the possibility to blow the whistle, which confirms our intuition. The effect of bribe size appears to be smaller in the presence of a leniency mechanism, but is still positive and significant. This somewhat suggests that when reporting is possible tax officers react to a lesser extend to the size of the bribe offer, possibly because some tax officers are sufficiently deterred by the threat of being reported that the size of the bribe becomes less relevant.

We briefly discuss also the use of the whistleblowing mechanism among tax payers. Tax payers made use of the possibility to report almost to the full extent with an overall average propensity to report the tax officer of about 91.4%. Reporting was most frequently used in treatment *L-NoL* (98.6%), but not significantly different from the frequency observed in *NoL-L* (87.4%) and *LEN* (90.0%). We thus do not find any evidence for reciprocity among tax payers and tax officers, which may partially be attributed to the fact that in our setting tax payers who chose to report were granted partial anonymity. Tax officers were only informed that and by how many tax payers they were reported, but not exactly by whom. Hence, depending on the particular situation tax officers were not able to determine whether a particular tax payer did blow the whistle or not. This limits the scope for retaliation, for example via withholding future cooperation, and hence may explain the high rate of reporting decisions. In contrast it has been argued that betrayal, such as reporting, is associated with a moral or psychological cost (see also Coricelli et al., 2010), which is not supported by our data.

4.3.3 Effects on Tax Evasion

Experimental evidence suggests that subjects' tax compliance usually is well above the theoretically optimal level, for example due to moral costs of en-

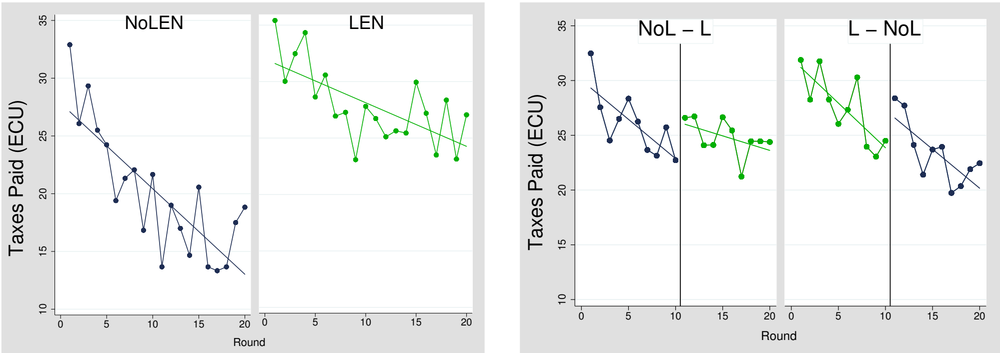


Figure 4.7: Amount of taxes finally paid over rounds and across treatments.

gaging in illicit behavior.[12] In our experiment tax payers had to make two decisions: First, whether and how much to bribe the tax officer, and second, how much taxes they wanted to declare. Since in our setup tax evasion is nested within a framework of collusive bribery, the amount of taxes actual paid is the result of a tax payer's decision about the amount of taxes declared as well as tax officer's decision to accept or reject the report (and thus the bribe offered). We hence have to distinguish between attempted tax evasion, as revealed by the amount of taxes declared, and actual tax evasion, as revealed by the amount of taxes finally reported. Recall that following a rejection by the tax officer, the tax payer is forced to truthfully report taxes. While attempted tax evasion is also of some interest and can surely cause moral damage to society as a whole, it is the actual amount of taxes evaded what directly causes a negative externality on society. In this subsection we therefore focus on actual tax evasion, that is, the amount of taxes finally reported. The results for attempted tax evasion are very similar to those of actual tax evasion and hence we omit them for brevity.

The average amount of taxes paid in *NoLEN* is 20.1 ECU and is thus smaller than the average of 27.9 ECU observed in treatment *LEN*. This difference is statistically significant (MWW, $N = 22$, $z = -2.4397$, $p = 0.0147$) showing that the lower rate of collusion in *LEN* observed previously also

[12]It was shown in Banerjee (2016b) that a loaded frame that creates the right sense of entitlement significantly decreases corruption, suggesting that moral costs are indeed at work.

translates into a higher tax yield collected. As illustrated by Figure 4.7 the amount of taxes paid also shows a negative trend across all treatments independent of the presence of a leniency mechanism. In fact, in *NoLEN* the amount of taxes paid decreases from 23.9 ECU to 16.2 ECU from part one to part two, whereas in *LEN* this decrease is smaller with average tax payments of 29.7 ECU in part one and 26.2 ECU in part two. In treatment *NoL-L* taxes paid show almost no decrease following the introduction of whistleblowing being 26.1 ECU before and 24.8 ECU on average after the mechanism was introduced, respectively. Comparing the changes in taxes paid between *NoLEN* and *NoL-L*, we find a significantly smaller decrease in treatment *NoL-L* (MWW, $N = 32$, $z = -1.8093$, $p = 0.0704$). Thus the introduction of a whistleblowing mechanism has a significant positive effect on the tax yield collected. In contrast, in treatment *L-NoL* paid taxes decrease from 27.5 ECU on average in the first part of the experiment to 23.4 ECU on average in the second half of the experiment where leniency was removed. This decrease is similar in size to the one observed in treatment *LEN* and a Mann-Whitney-Wilcoxon test confirms that the removal of the reporting option has no significant negative effect on the amount of taxes paid ($N = 26$, $z = 0.4115$, $p = 0.6807$).

4.4 Discussion and Conclusion

Our results shed light on the effects of a leniency mechanism on collusive bribery in a tax evasion framework utilizing a controlled laboratory setting. We nest collusive corruption in a tax evasion framework, in which tax payers require the cooperation of a tax officer to evade taxes, thus opening the door for collusive bribery. The leniency mechanism we consider offers leniency to tax payers for reporting corrupt tax officers. In our setup leniency not only shifts the risk and negative consequences (fines) of collusive bribery from the tax payer to the tax officer, who otherwise faces little to no consequences, but also renders her formally responsible. Compared to most studies in the tax evasion literature we add a dimension of strategic interaction that allows us to capture a richer strategic environment, which is applicable to other domains, such as custom duties, that are understudied so far. Further, we investigate the dynamics of institutional changes and their effects on both corruption and tax evasion by considering not only environments with and without leniency, but also the introduction and the removal of such a policy. By doing so we have identified a positive spillover effect of the presence of a whistleblowing mechanism present from the first half of the experiment to the second half of the experiment where it is no longer in place.

Comparing settings with and without leniency, in the absence of an institutional change, we found leniency to be effective in combating collusive bribery. When leniency for a tax payer is in place successful collusion between tax payer and tax officer is less frequent. Further, it effectively deters the tax officer from accepting bribes, while at the same time we find no evidence that leniency encourages the tax payer to offer bribes. We identify a lower willingness of the tax officer to accept bribes as the main driver behind the observed effects on collusion. We also find a positive effect of leniency on tax compliance with more taxes being collected when such a mechanism is in place. In addition, our results highlight the role of institutional changes and its importance to the evaluation of policy measures. We show that the introduction of the opportunity to blow the whistle is effective in breaking up already established collusive pattern by sowing distrust between the colluding parties, which prevents collusive bribery and tax evasion to thrive further. In contrast, the removal of the institutional mechanism does not cause similar effects in the opposite direction, which points towards a positive spillover effect of the particular institutional mechanism we consider. That is, the positive effects of offering leniency to whistleblowers persists even after the mechanism has been removed. This in in line, with some recent evidence emphasizing the importance of spillover effects (e.g., see d'Adda et al., 2017; Engl et al., 2017).

We provide empirical evidence emphasizing that a political measure should not be judged in isolation by disregarding the reference point provided by the pre-reform system, since this might lead to an incomplete or even flawed assessment of its effectiveness. It is therefore crucial to consider the history of political or legal systems when deciding upon means to combat corruption and tax evasion. The classical economic model of tax evasion does not consider the fact that individuals are "born into" a certain legal system, but exactly this status quo might determine whether a potential reform is effective or not. Taking this evidence into account will be crucial for understanding why sometimes reforms are highly effective in a certain country or cultural environment, while they are ineffective in others. This might be related to the echo effect found in Mittone (2006), that is, a change in the audit sequence affects behavior because subjects "learn" to be risk-averse or risk-seeking through experiencing early or late first audits. This indicates that past experience can create some sort of reference behavior that cannot easily be "unlearned," and hence might enhance or hinder the effectiveness of a subsequent reform. Following that line of argument reforms can turn out to be a one-way street, once implemented their effects cannot simply be undone by reestablishing the pre-reform regime. Hence rolling out reforms is a process that ought to be taken with great caution by policy makers.

Appendix 4.A: Theoretical Analysis of the Bribery Game

Consider the bribery game with and without leniency described in Subsection 4.2.1 above as one-shot interaction between a TP and a TO, both assumed to be rational in the sense of being risk-neutral expected payoff-maximizers. Assuming the rational model of crime (Allingham and Sandmo, 1972) we now derive theoretical predictions regarding tax compliance and bribe exchange. Our analysis shows that predicted tax compliance of the TP is the same for both institutional frames. On the other hand, the optimal bribe payment is higher in the BGL where reporting is possible. Moreover, bribe exchange (collusion) is optimal under both regimes. Denote the amount of taxes declared by D and the bribe offered by b.

In the BG a rational TO will accept any bribe b that is (weakly) above the expected foregone commission of 15% from the taxes declared, that is 7.5% of the declared income D. Since the TO does not observe the income declared by the TP we assume that she holds a belief $\mu : \{0, \ldots, 80\} \longrightarrow [0, 1]$ over D. The expected amount of declared income given this belief μ is then

$D(\mu) = \sum \mu(D)D$. Hence, the TO will accept a bribe if she beliefs that the bribe is larger than her foregone commission, that is if and only if

$$b \geq 4.8 - 0.06D(\mu).$$

The bribe acceptance threshold, which we denote by $b_{BG}(\mu)$, depends only on the expected amount of declared income $D(\mu)$. For example, if the TO expects the TP to declare zero taxes, that is $D(\mu) = 0$, then only bribes of at least 4.8 are accepted. Note that the threshold is strictly increasing in $D(\mu)$. On the other hand, if the TP offers a bribe b and the TO accepts (which is the case for $b \geq b_{BG}(\mu)$), the TP's expected payoff for reporting an amount of D is

$$\Pi_{\text{TP}}(D, b \mid accept) = 70 - b - 0.375D.$$

Note that Π_{TP} is decreasing in D and b, hence a rational TP will optimally declare an income of $D = 0$ and pay the smallest bribe that is accepted by the TO, which is $b = 4.8 - 0.06D(\mu)$.

In the BGL leniency introduces the possibility for a TP to report a corrupted TO following an audit. In the one-shot scenario it is optimal for the TP to report the TO when being audited, in which case the TP now has an expected payoff of

$$\Pi_{\text{TP}}(D, b, report \mid accept) = 72 - b - 0.4D.$$

This payoff is still decreasing in D and b, and thus the TP prefers to declare zero taxes and pay the smallest bribe that is accepted by the TO. However, the bribe threshold in the BGL is not the same as in the BG. To see this, suppose the TO anticipates that the TP will always report her when audited, then a rational TO will accept a bribe if and only if

$$b \geq 8.5 - 0.075D(\mu).$$

We denote this threshold by $b_{BGL}(\mu)$. Intuitively, now the TP has to compensate the TO not only for his forfeited (expected) salary, but also for the risk of being reported and its consequences.

The game described above is a game of imperfect information (the TO does not observe D) and as such it has many equilibria. We use Perfect Bayesian Equilibrium (PBE) as our solution concept of choice. Given a point belief μ with $\mu(0) = 1$ and $\mu(D) = 0$ for $D \neq 0$ there is a PBE of BG where the TP declares exactly $D = 0$ and offers a bribe $b = b_{BG}(\mu)$, which the TO accepts. Similarly, given the same belief there is a PBE for BGL where the TP declares $D = 0$, offers a bribe $b = b_{BGL}(\mu)$, which the TO accepts, and always reports the TO when audited. In both, BG and BGL, collusion is an equilibrium of the one-shot game. However, since $b_{BGL}(\mu) > b_{BG}(\mu)$ for any μ, the bribe acceptance threshold in BGL is higher compared to BG. It is important to note that for both games the bribe acceptance threshold is decreasing in the mean of the TO's belief μ.

Appendix 4.B: Summary Statistics

Table 4.2 provides an overview of the behavior in all four treatments. We report the frequency of successful bribe exchange (collusion), the frequency of bribe offers, the amount of bribes paid, the proportion of bribes accepted by the tax officer, tax compliance, both attempted and effective, and the propensity of tax payers to report tax officers when given the chance.

Table 4.2: Summary statistics across treatments.

Treatment	*NoLEN*	*LEN*	*NoL-L*		*L-NoL*	
Rounds	1-20	1-20	1-10	11-20	1-10	11-20
Collusion (in %)	52.3	34.6	43.2	41.7	38.3	48.8
BribeOffered (in %)	71.0	57.3	65.0	70.5	72.4	66.0
BribeSize (in ECU)	14.4	16.6	13.1	15.4	16.6	15.2
AccRate (in %)	73.1	58.8	64.5	59.7	52.2	74.2
TaxDeclared (in ECU)	13.9	20.6	20.2	14.9	18.1	18.3
TaxPaid (in ECU)	20.1	27.9	26.8	24.8	27.5	23.4
Reporting (in %)	-	85.5	-	91.5	98.7	-

Note: Collusion denotes the incidence of successful bribe exchange (bribe offered and accepted); BribeOffered denotes the incidence of a bribe being offered relative to all situation where this was possible; BribeSize is the average size of the offered bribes (0-30 ECU); AccRate denotes the fraction of bribe offers that were accepted by tax officers; TaxesDeclared denotes the amount of taxes initially reported (0-40 ECU); TaxesPaid denotes that taxes actually paid according to the final, accepted report (0-40 ECU); Reporting denotes the fraction of reporting decision by tax payers when audited.

Appendix 4.C: Additional Analysis

Table 4.3: Logistic panel regression with random effects of acceptance on bribe size.

Accepted	*NoLEN*	*LEN*	*NoL-L*	*L-NoL*
BribeSize	0.3987***	0.1613***	0.2434***	0.2403***
	(0.1219)	(0.0445)	(0.0501)	(0.0578)
Leniency			−0.1954	−2.5350*
			(0.5717)	(1.3143)
Leniency × BribeSize			−0.0607**	0.0229
			(0.0296)	(0.0750)
Constant	−4.1264**	−2.1187***	−2.1590***	−1.9292*
	(1.8780)	(0.7271)	(0.7008)	(1.0607)
Linear combination test			0.1828***	0.2632**
BribeSize + Leniency × BribeSize			(0.03703)	(0.1142)
Observations	426	413	894	581

Note: Standard errors clustered at the group level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 5
Timing, Uncertainty and Institutional Deterrence

5.1 Introduction

Governments all over the world use substantial resources to keep society safe and punish people for criminal acts. Annually, the US spends approximately $75 billion on incarceration (not including costs for courts, trials, etc.). Thus, it is hardly surprising that extensive research has been done to understand the determinants of deviant behavior and shed light on alternative deterrence mechanisms. Existing economic literature not only stresses the relevance of institutional environments in shaping prosperity and growth (La Porta et al., 1999; Acemoglu et al., 2005), but also their importance in effectively deterring criminal and immoral behavior in ways that include staff rotations in public administration, crown witness regulations, and changes in punishment regimes (Shleifer and Vishny, 1993; Abbink, 2004; Abbink et al., 2014; Engel et al., 2016). Due to the inherent methodological challenges of studying deviant behavior, where reliable observational data is unavailable, economists have turned to controlled experiments to address these pressing questions (Abbink, 2006). We follow this methodological approach in our paper.

There is vast literature on criminal deterrence that focuses on the relevance of the certainty and severity of punishment in deterring deviant behavior (see e.g. Becker, 1968; Baker et al., 2004; DeAngelo and Charness, 2012; for a recent review of economic research see Chalfin and McCrary, 2017 and for a cross-disciplinary discussion of experimental work see Engel, 2016). However, the swiftness of punishment (often referred to as celerity), frequently mentioned alongside certainty and severity (Bailey, 1980; Howe and Brandau, 1988; Yu, 1994; Nagin and Pogarsky, 2001, 2004), has been under researched by those in the economic field. Understanding the mechanisms underlying deterrence of deviant behavior yields important policy implications. Given the high costs involved in increasing punishment's certainty (e.g. costs for an executive body) or punishment's severity (e.g. incarceration costs), we argue that the timing of conviction and punishment, that is, their delay with respect to the transgression in question, can potentially serve as a powerful tool for deterrence that is often available at a relatively low cost.

The classic theoretical approach towards the deterrence of criminal activity (e.g. Becker, 1968) is based on the assumption that potential offenders mainly weigh the potential gains against the potential adverse consequences of an offense. In the standard framework of discounted expected utility, delayed punishment should reduce deterrence due to a discounting effect, whereas the timing of resolution of uncertainty should have no effect on behavior. Starting with the seminal paper of Loewenstein (1987), several theories propose that anticipation of future events is an important determinant of inter-temporal utility (see e.g., Wu, 1999; Lovallo and Kahneman, 2000; Caplin and Leahy, 2001; Dillenberger, 2010; Strzalecki, 2013; Golman and Loewenstein, 2015). These models are based on the idea that a non-negligible proportion of the overall consequences from future consumption (be it negative or positive) is already consumed in the form of so-called anticipatory utility before actual consumption takes place. While there is growing theoretical literature supporting anticipatory utility theory and its implications, there is little empirical work being done and even less experimental investigation. [1]

The goal of the present paper is to experimentally test the implications of anticipatory utility in the context of institutional deterrence mechanisms. In particular, we are interested in how the timing of sanctions (be it conviction or sentencing) and the timing of the resolution of uncertainty surrounding these sanctioning mechanisms affects deterrence. We systematically vary the celerity of a sanction within a new, stylized, experimental paradigm along the following two dimensions: first, we vary the delay between offense and

[1] Two recent exceptions are Falk and Zimmermann (2016), who experimentally tested the implications of anticipatory utility in the context of information preferences and Kogler et al. (2016), who showed that delayed resolution of a tax audit results in higher tax compliance.

detection/conviction; second, we vary the delay between offense and sanctioning. Our main objective is to better understand the role of celerity, in our opinion, an important dimension of most deterrence mechanisms, that has received surprisingly little attention in previous literature. We argue that celerity could potentially serve as a useful tool for policy makers to design more efficient and/or less expensive institutional deterrence mechanisms. However, delayed punishment is not necessarily less deterrent (due to discounting) if utility from anticipation is taken into account. Additionally, we study the role of the timing of resolution of uncertainty. We vary the point in time when the information about whether or not a transgression was detected is revealed to subjects. We show that in theory, depending on the impact of anticipatory utility, delayed resolution of uncertainty may increase deterrence.

Our experimental analysis is based on a simple guessing game where subjects may cheat in some periods to increase payoffs. After these periods there is an investigation such that cheaters will be detected and fined with a given probability. In the single treatments, we vary both the timing of the potential fine, as well as, the timing of the resolution of uncertainty, i.e. when the participants learn the results of the investigation. We analyze behavior alongside two dimensions: total cheating behavior and recidivism (conditional cheating). Our results show that delayed resolution has no systematic impact on cheating. With respect to the relation between the delay of punishment and deterrence, we observe an inverted U-shape relationship where deterrence is lowest for a short delay of punishment and significantly lower for either no delay or a long delay when combined with a late resolution of uncertainty.

This result is at odds with discounted expected utility and theories of anticipatory utility, but can be explained by the recent model of Baucells and Bellezza (2016). They extended anticipatory utility by a reference point, a utility of recall and a magnitude effect in discounting. We conclude that in order to increase the deterrence of sanctioning mechanisms, punishment should either be swift or sufficiently delayed and paired with the psychological dread of uncertainty.

The paper is organized as follows. The next section provides a brief review of the theoretical and empirical background on the relation between celerity and deterrence. Section 3 details our experimental procedures and discusses the hypotheses we aim to test. Results are presented in Section 4. The final section discusses our results and derives some conclusions.

5.2 Theoretical and Empirical Background

The benefits of criminal behavior are usually immediate. Any proceeding detection, conviction, and implementation of legal consequences are generally delayed and stochastic. This poses an inter-temporal decision problem under uncertainty. Classically, celerity meant only the temporal delay of a potential sanction following a transgression. We will adopt a wider definition of celerity, using it as a catch-all phrase for the timing of the various facets of a deterrence mechanism. There are several prominent economic theories of inter-temporal decision making. Here we want to focus on two. First, theories of temporal discounting suggest that future costs or benefits receive a lower weight than immediate ones; this weight decreases as one moves further into the future (Frederick et al., 2002). The implications are simple. If a potential offender discounts delayed legal consequences, then deterrence decreases the longer the delay. As a consequence, higher celerity (less delay) would increase the efficiency of legal sanctions, which is the classical hypothesis in criminological literature (Nagin and Pogarsky, 2004; Paternoster, 2010).

Second, theories of anticipatory utility that incorporate anticipatory feelings such as excitement, fear or dread into classical expected utility theory suggest that one might want to bring forward an unpleasant event to shorten the period of dread (or delay a positive event to enjoy the excitement for a longer period of time). The idea is that future events influence current utility. More precisely, negative (positive) future events cause negative (positive) utility today the further away the event is (at least up to a certain point). Caplin and Leahy (2001) extend Loewenstein's model by allowing for uncertainty and point toward the importance of anticipatory feelings prior to the resolution of uncertainty. However, anticipatory emotions, such as anxiety, are often predicated on an uncertain future. Thus, they are mainly relevant prior to the resolution of uncertainty. This suggests that the point in time at which uncertainty is resolved is particularly important. For example, Kreps and Porteus (1978) and Kocher et al. (2014) show that preferences over temporal lotteries also depend on the point in time when the uncertainty is resolved. That is, agents can show a preference for earlier or delayed resolution of uncertainty. Further evidence comes from consumer literature. Anticipatory emotions, compared with outcome-based emotions, are central in prospective consumption situations. Furthermore, the uncertainty associated with anticipatory emotions negatively affects intentions (Bee and Madrigal, 2013). Psychological learning theories (Skinner, 1963; Tversky and Kahneman, 1986; Ehrlich, 1996; Hackenberg, 2009) second the argument that the time between a transgression and the punishment and the uncertainty that is associated with the punishment are driving forces for effective behavioral changes. If this is indeed the case, then the classical interpretation of celerity as the time between committing an offense and the actual punishment (e.g. fine or imprisonment) should be complemented by the time the uncertainty is resolved, thus, the time of sentencing.

The implications of the timing of a sanction on deterrence derived from anticipatory utility theory could oppose those suggested by temporal discounting. Clearly this is an important point that has to be taken into consideration for the design of legal institutions. A systematic study of the role of celerity for deterrence poses a serious empirical challenge, because changing the celerity of an enforcement mechanism would most likely impact existing institutional structures on multiple levels. For that reason, isolating the impact of such an intervention is hardly possible in the field. In addition, it is unclear whether an actual or would-be offender is aware of this change or not, making identification almost impossible. Thus, a systematic study of celerity calls for a highly controlled environment that allows for the isolation of the direct effect of institutional changes varying celerity on behavior. Fortunately, the experimental laboratory provides such a controlled environment.

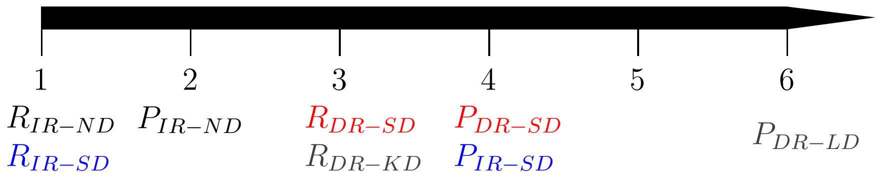
5.3 Design and Hypotheses

5.3.1 Experimental Design

We use a simple guessing game that is played repeatedly by our subjects. In certain rounds subjects are presented with the option to "cheat." Cheating guarantees them the maximum possible payoff for that round. Our goal was to design a simple game where the option to cheat was not integral; we wanted the game to be easy-to-understand, but meaningful regardless of whether or not the option to cheat was presented. Specifically, we wanted to make sure that cheating was not considered part of the game, but a clear violation of said games rules. In our guessing game a card is randomly drawn from a deck of 32 cards and subjects have to guess which card was drawn. A subject received 10 Experimental Currency Units (ECU) for a correct guess and 4 ECU for an incorrect guess. In some rounds participants are given the option to cheat. By cheating, participants are allowed to uncover the randomly drawn card before making one's guess, ensuring a correct answer and the maximum payoff of 10 ECU less a possible fine if detected.[2] Participants were informed that each instance of cheating would be followed by an "investigation" that would detect cheating with a fixed probability of 25%. Hence, cheating exposes them to the risk of being caught. If caught the consequences are two-fold. First, the subject has to pay a fine of 10 ECU. Second, the subject is suspended from the game for one round, is not allowed to make any decision and cannot earn any ECU. Furthermore, suspended participants are forced to wait 60 seconds before they are allowed to continue in the next period. We deliberately chose suspension as part of the sanctioning mechanism to increase salience with regard to the timing of sanctions. While one might argue that a delayed fine in a laboratory context where all "actual" payments are realized at the very end of the experiment decreases the result's robustness, such concerns do not apply to the suspension as it is clearly linked to the particular round

[2]When subjects decide to cheat, we automatically implement the "right guess" for them. Subjects are informed about this procedure in the instructions. We implement this forced guess to avoid "second thoughts" where a subject cheats, views the drawn card, but chooses a different card.

Figure 5.1: Timeline of the experiment for each treatment.



Notes: P and R indicate the timing of resolution of uncertainty and timing of punishment for each of the treatments *IR-ND*, *IR-SD*, *DR-SD*, and *DR-LD*, respectively.

a subject is suspended.

In order to make the moral dimension of cheating more salient in our laboratory context we introduce a third party, represented by a charity, that incurs a monetary damage as a result of cheating. Specifically, for each experimental session there is a charity pool of 250 ECU (worth $25) from which 50 ECU is deducted each time a particular subject decides to cheat.[3] At the end of the experiment one subject is randomly selected whose decisions determine the charity pool, the remainder of which will be donated to "Doctors without Borders."

In our experiment, we vary the timing along the following two dimensions: the timing of punishment and the timing of the resolution of uncertainty. Punishment is either immediate, delayed by two rounds or delayed by four rounds. In addition, the resolution of uncertainty regarding whether cheating is detected (and hence whether there are sanctions) is either immediate or delayed by two periods. Figure 5.1 illustrates the timeline for each treatment.

All treatments consist of 28 rounds: four training rounds followed by four blocks of six rounds each. In the first four rounds participants play the guessing game without cheating to familiarize themselves with the game and the interface. In the first round of each block subjects can cheat. In the remaining rounds of a block (rounds 2-6) they play the guessing game without the option to cheat. Using blocks of 6 rounds allows us to vary both the timing of the resolution of uncertainty, as well as, the timing of

[3] For each subject there are exactly four cheating opportunities, all in the first round of a "block" of six rounds. That is, in rounds 5, 11, 17 and 23 subjects are given the opportunity to cheat. Subjects are informed that "occasionally" they will be presented with the option to cheat, but not about the exact timing and frequency of this option.

Table 5.1: Overview of timing of resolution of uncertainty and punishment in the different treatments.

Treatment	Timing of resolution of uncertainty	Timing of punishment
IR-ND	immediate	no delay
IR-SD	immediate	short delay (2 rounds)
DR-SD	delayed (2 rounds)	short delay (2 rounds)
DR-LD	delayed (2 rounds)	long delay (4 rounds)

punishment without an overlap with subsequent cheating decisions.

Table 5.1 summarizes the four treatments. In treatment *IR-ND*, we have immediate resolution of uncertainty and no delay of punishment. Subjects receive immediate feedback within the same round about whether cheating was detected and there is no delay in punishment. That is, the fine (if due) is deducted and a potential suspension is implemented immediately for the next period.[4] In treatment *IR-SD*, resolution of uncertainty is again immediate, but now there is a short delay in punishment of two periods; when cheating in period t the uncertainty will be resolved immediately, but the potential fine and suspension are executed only in period $t+3$ (as opposed to $t+1$ in *IR-ND*). We will also refer to *IR-SD* as immediate resolution of uncertainty and short delay of punishment. In treatment *DR-SD*, the investigation into cheating does not conclude immediately, but lasts for two additional periods. Only after that is the participant informed about whether his cheating was detected or not. As in *IR-SD*, there is a short delay of punishment. We hence refer to *DR-SD* as delayed resolution of uncertainty and short delay of punishment. Finally, in treatment *DR-LD* resolution is again delayed, but now punishment is delayed for four periods rather than two. That is, cheating in period t results in resolution of uncertainty in period $t+2$, followed by the actual punishment (if due) in period $t+5$.

[4]Clearly punishment cannot precede the resolution of uncertainty which determines whether a subject was detected and hence will have to face a punishment.

5.3.2 Experimental Procedures

We conducted 32 experimental sessions at the Decision Science Lab at Harvard University. Participants were recruited vie e-mail invitation from the laboratory's database which contains students, as well as, non-students. A total of 296 subjects (out of which 46.6 % were males) participated in the experiment split between treatments as follows: 66 subjects in *IR-ND*, 85 subjects in *IR-SD*, 69 subjects in *DR-SD* and 76 subjects in *DR-LD*. The experiment was programmed and run using z-Tree (Fischbacher, 2007).[5] Within each session participants were randomly assigned to a computer booth in which they would participate in the experiment anonymously. The consent forms and instructions for the corresponding treatment were distributed. Upon agreeing to the informed consent page the participants were given sufficient time to read the instructions carefully. Before the start of the experiment subjects had to answer a series of comprehension questions in order to check their understanding of the game and its payoff structure. Subjects then played 28 periods after which they were informed of their total earnings via a detailed summary screen. One subject was randomly drawn to determine the charity pool and all participants were informed about the final amount left in the pool to be donated to "Doctors without Borders."[6] At the end of the experiment subjects completed a questionnaire containing questions on personal characteristics (demographics, education, income, age), risk-attitudes (SOEP), consideration of future consequences (Strathman et al., 1994) and self-control (Tangney et al., 2004).

Sessions lasted approximately 45 minutes excluding the time for payment. A participant's payoff was determined by the sum of his earnings over all 28 rounds. The total payoff in ECU was then converted to dollars at a rate of

[5]It is worth noting that we observed an influx of disproportionately older participants due to a bug in the recruitment software in our first sessions. This was quickly resolved. Participants of 41 years and older represent around 11% of our data set. Unless noted otherwise, our results are robust with respect to this subgroup.

[6]Prior to the experiment subjects received a short description of the work of "Doctors without Borders." Although we cannot know for sure that all participants endorse their work, we wanted to enforce a minimal level of common knowledge to increase salience. A receipt of the amount actually donated was made available to all participants via email.

10 ECU = \$1. The average payment was \$14.29 which includes a show-up fee of \$2.50.

5.3.3 Hypotheses

In this section, we further detail our main hypothesis on how deterrence could be affected by the delay of punishment and the timing of resolution of uncertainty. In the standard discounted expected utility (DEU) model, optimal decisions do not depend on the timing of resolution of uncertainty. In our model a delay of punishment should decrease deterrence. The utility of not cheating (NC) is identical in all treatments and is given by

$$\text{DEU}(NC) = \frac{31}{32}4 + \frac{1}{32}10 \tag{5.1}$$

where we assume for convenience a linear utility function.[7] In what follows we denote DEU(NC) by $\bar{u}_{NC}$. We restrict attention to a single block consisting of six periods, where cheating was possible in the first round of that block. Further, we only consider the utility generated from the decision about cheating in the first period of such a block in all our analyses. The remaining utility components within a block are identical across treatments. In *IR-ND*, detected cheaters are fined (10 ECU plus one round suspension) directly in the next period. For a discount factor $\delta < 1$, the utility of cheating (C) amounts to

$$\text{DEU}(C, \text{IR-ND}) = 10 - \frac{1}{4}\delta(10 + \bar{u}_{NC}) \tag{5.2}$$

as cheating is not possible in the next period. Compared to *IR-ND*, punishment is delayed by two further periods in *IR-SD*. The same is true for *DR-SD*. As the timing of resolution of uncertainty is immaterial under DEU, we get

$$\text{DEU}(C, \text{IR-SD}) = \text{DEU}(C, \text{DR-SD}) = 10 - \frac{1}{4}\delta^3(10 + \bar{u}_{NC}). \tag{5.3}$$

[7] While risk aversion modeled by a concave utility function certainly influences the decision between cheating and not cheating, it does not imply differences between treatments.

Finally, we have

$$\text{DEU}(C, \text{DR-LD}) = 10 - \frac{1}{4}\delta^5(10 + \text{DEU}(NC)). \tag{5.4}$$

as punishment is delayed by a total of four periods in *DR-LD*. Since $\text{DEU}(C, \text{IR-ND}) < \text{DEU}(C, \text{IR-SD}) = \text{DEU}(C, \text{DR-SD}) < \text{DEU}(C, \text{DR-LD})$ where the utility of not cheating is independent of the treatments, we get the following hypothesis:

Hypothesis 1. Increasing the delay of punishment decreases deterrence, leading to more violations in *IR-SD* compared to *IR-ND* and in *DR-LD* compared to *DR-SD*.

Hypothesis 2. The timing of resolution of uncertainty does not affect behavior, implying that violations in treatments *IR-SD* and *DR-SD* are identical.

Hypothesis 3. Since the timing of resolution of uncertainty does not change deterrence and increasing the delay of punishment decreases deterrence, we will have more violations in *DR-SD* than in *IR-ND* and more violations in *DR-LD* than in *IR-SD*.

Following Loewenstein (1987) negative future outcomes can cause immediate disutility through negative anticipatory emotions such as fear, dread or anxiety. DEU fails to take this into consideration. Suppose you were cheating in treatment *IR-ND*. Then you dread in the first period that you will be fined in the next one, i.e. you dread a loss of $10 + \bar{u}_{NC}$. For a discount rate γ which measures the degree to which current utility is influenced by anticipated emotions from consumption in the next period, the utility of cheating is given by

$$\text{UAE}(C, \text{IR-ND}) = 10 - \frac{1}{4}(\delta + \gamma)(10 + \bar{u}_{NC}) \tag{5.5}$$

where UAE denotes utility with anticipated emotions. We now consider *IR-SD* where there is a short delay of punishment by two periods. Note that the utility from anticipation is discounted with discount factor δ. While the discounting effect in (5.3) increases utility compared to *IR-ND*, anticipation leads to decreasing utility as dread is now experienced in more than one period. More specifically, we get

$$\text{UAE}(C, \text{IR-SD}) = 10 - \frac{1}{4}\delta^3(10 + \bar{u}_{NC}) - \frac{1}{4}(\gamma^3 + \delta\gamma^2 + \delta^2\gamma)(10 + \bar{u}_{NC}) \quad (5.6)$$

Comparing (5.5) and (5.6), it may well be that the utility of cheating is lower in *IR-SD* than in *IR-ND* if γ is sufficiently high. Since the utility of not cheating is identical across treatments, we get the following hypothesis as alternative to Hypothesis 1:

Hypothesis 1*. If the effect of anticipation is sufficiently high, delaying punishment increases deterrence leading to less violations in *IR-SD* compared to *IR-ND* and in *DR-LD* compared to *DR-SD*.

Anticipated emotions in the model of Loewenstein (1987) refers to future consumption under certainty. In treatments *DR-SD* and *DR-LD* resolution of uncertainty is delayed which may alter anticipatory emotions. While in *IR-SD* a detected cheater may feel dread in periods 1-3 due to anticipating the punishment in period 4, in *DR-SD* a cheater may experience the anxiety of being detected in the later investigation. Following Caplin and Leahy (2001) the anxiety experienced one period before resolution should depend on the probability of being detected and the size of the fine. As all these parameters are identical in treatments *DR-SD* and *DR-LD* we simply use the terms A to denote the anxiety of a cheater one period before resolution. We now introduce a third discount rate α, such that anxiety experienced t periods before resolution is given by $\alpha^t A$. This yields the following utility of cheating in *DR-SD*:

$$\text{UAE}(C, \text{DR-SD}) = 10 - \frac{1}{4}\delta^3(10 + \bar{u}_{NC}) - (\alpha + \delta\alpha^2)A - \frac{1}{4}\delta^2\gamma(10 + \bar{u}_{NC}) \quad (5.7)$$

Typically, it is observed that people prefer early resolution of uncertainty for negative outcomes. In our model this is the case if

$$(\alpha^2 + \delta\alpha)A > \frac{1}{4}(\gamma^3 + \delta\gamma^2)(10 + \bar{u}_{NC}) \quad (5.8)$$

and leads to the following hypothesis:

Hypothesis 2*. Delayed resolution of uncertainty increases deterrence leading to less violations in *DR-SD* compared to *IR-SD*.

Obviously, if the resolution of uncertainty should be delayed in order to increase deterrence, punishment has to be delayed as it cannot precede the resolution of uncertainty. The combined effect of delayed resolution and delayed punishment can be grasped by comparing *DR-SD* to *IR-ND*. If both delaying punishment according to Hypothesis 1* and delaying resolution according to Hypothesis 2* increases deterrence, our model implies the following:

Hypothesis 3*. If delaying punishment increases deterrence due to dread and delayed resolution also increases deterrence due to anxiety, then the combined effect of delaying punishment and resolution results in less cheating and, therefore, less violations in *DR-SD* compared to *IR-ND*.

Let us finally consider the utility of cheating in *DR-LD*. Here we get

$$\mathrm{UAE}(C, \text{DR-LD}) = 10 - \frac{1}{4}\delta^5(10+\bar{u}) - \frac{1}{4}(\gamma^3 + \delta\gamma^2 + \delta^2\gamma^3 + \delta^3\gamma^2 + \delta^4\gamma)(10+\bar{u}) \tag{5.9}$$

The cheater experiences anxiety prior to the resolution of uncertainty as in *DR-SD*, but there is also an extended period where he may experience dread due to delayed punishment. The second component is similar to the dread experienced in *IR-SD*, additionally discounted as the experience starts two periods later. Assuming (5.8), a comparison of (5.9) and (5.6), reveals that the utility of cheating in *DR-LD* will be smaller than that of cheating in *IR-SD* under the conditions of Hypothesis 1*. This results in the following hypothesis:

Hypothesis 4*. If (5.8) holds and the effect of anticipation is sufficiently high (γ is large enough), then delayed resolution combined with delaying punishment results in less cheating leading to less violations in *DR-LD* compared to *IR-SD* and less violations in *DR-LD* compared to *IR-ND*.

5.4 Results

Here, we present our results using parametric and non-parametric comparisons,[8] as well, various regression techniques to analyze differences in cheating behavior, as motivated by our hypotheses. Please note that not only the number of cheating opportunities (4) were the same in all treatments, but also their timing (always in the first round of each block). Hence, any difference in behavior can only result from our systematic variation in the timing of punishment and the timing of resolution of uncertainty.

First, we look at the mean differences in total cheating across all treatments as outlined in the theory part of our paper. Total cheating is defined as the total number of individual cheating incidences across all rounds. We calculate the percentage as the ratio of actual individual cheating decisions to the maximum possible number of cheating opportunities (4). We present the test results in Table 5.2 and a graphical illustration in Figure 5.2. Results illustrate that the amount of cheating is 15 percentage higher in *IR-SD* when compared to cheating in *IR-ND* (BSM, $p = 0.03$). Cheating is 13 percentage lower in *DR-LD* compared to *DR-SD* (BSM, $p = 0.07$). Furthermore, cheating is 12 percentage lower in *IR-ND* than in *DR-SD* (BSM, $p = 0.09$) and roughly 15 percentage lower in *DR-LD* than in *IR-S* (BSM, $p = 0.02$). We test the theoretical predictions derived from our two theoretical frameworks (DEU and UAE) in Table 5.4 and Table 5.5, respectively. This can be found in Appendix 5.A. Overall, our hypotheses are partially supported by both theoretical approaches. We discuss the implications in greater detail in the next section.

In order to check for robustness, we ran a series of regressions to analyze the behavioral motivations that result in cheating and the total amount of cheating that took place. Treating decisions across rounds in the fashion of

[8] We follow Moffatt (2015) and employ the bootstrap two-sample t-test method (hereafter BSM) with 9999 replications to analyze mean differences of average return behavior. This has the advantage that we can retain the rich cardinal information in the data without making any assumptions about the distribution. Unless noted otherwise, the use of non-parametric Mann-Whitney-Wilcoxon (hereafter MWW) tests yields results that are in line with our bootstrap approach.

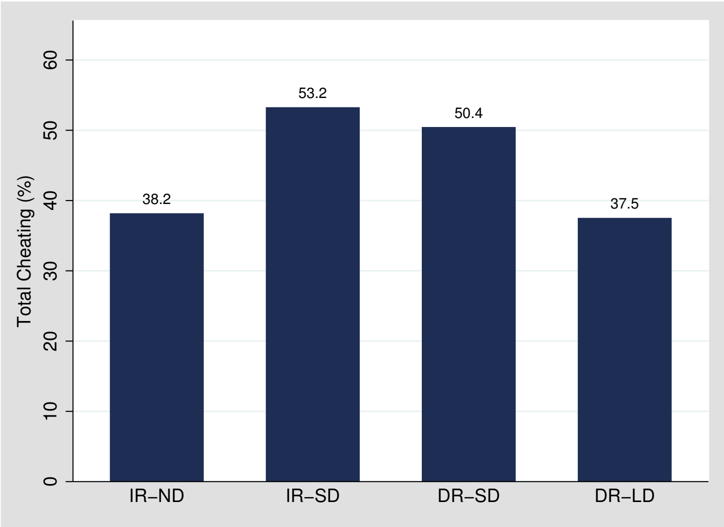


Figure 5.2: Average Total Cheating.

panel data, our dependent variable in Table 5.2 is a count variable adding up the total amount of cheating decisions across blocks.We present two types of regressions. The first analyzes behavior by timing (models 1 and 2), while the second analyzes behavior across treatment specifications (model 3 and 4). This allows us to dissect the impact of the timing of punishment from the timing of resolution of uncertainty, as well as, the effect of their interaction on total cheating behavior. To this end, we use *IR-SD* with a short delay of punishment and no delayed resolution of uncertainty as our reference category. The extended form regressions (column 2) include a battery of relevant covariates (gender, age, number of correct card guesses, experience with punishment from past cheating, round indicator, risk tendencies, awareness of future consequences, self-control, and a dummy indicating a participant's previous participation in economic experiments).

Our analysis in Table 5.2 suggests that, relative to a short delay of punishment, both swifter and more delayed punishment renders individual cheating decisions significantly less likely. The introduction of delayed uncertainty resolution itself does not significantly affect cheating behavior. A direct comparison of our treatments mirrors this finding, indicating that higher deterrence can be achieved by either implementing swift punishment (*IR-ND*) or

Table 5.2: Total Cheating using GLS Random Effects Regressions

	Analysis by timing				Analysis by treatment			
TotalCheating	(1)		(2)		(3)		(4)	
No Delay	-0.3066^{**}	(0.1560)	-0.3277^{**}	(0.1566)				
Long Delay	-0.2849^{*}	(0.1577)	-0.2813^{*}	(0.1582)				
Uncertainty	-0.0212	(0.1615)	-0.0437	(0.1643)				
IR-ND					-0.3066^{**}	(0.1560)	-0.3277^{**}	(0.1566)
DR-SD					-0.0212	(0.1615)	-0.0437	(0.1643)
DR-LD					-0.3062^{**}	(0.1527)	-0.3250^{**}	(0.1523)
Male	0.4020^{***}	(0.1147)	0.4361^{***}	(0.1252)	0.4020^{***}	(0.1147)	0.4361^{***}	(0.1252)
Age	-0.4847^{***}	(0.1671)	-0.4800^{***}	(0.1816)	-0.4847^{***}	(0.1671)	-0.4800^{***}	(0.1816)
GuessCorrect	-0.0668	(0.0681)	-0.0510	(0.0693)	-0.0668	(0.0681)	-0.0510	(0.0693)
Punishment	0.3404^{***}	(0.1157)	0.2881^{***}	(0.0689)	0.3404^{***}	(0.1157)	0.2881^{***}	(0.0689)
Round			0.4263^{***}	(0.0242)			0.4263^{***}	(0.0242)
Risk			-0.0625	(0.0624)			-0.0625	(0.0624)
FutCons			-0.0249	(0.0610)			-0.0249	(0.0610)
SelfControl			-0.0692	(0.0595)			-0.0692	(0.0595)
ExpParticipation			0.1472	(0.1204)			0.1472	(0.1204)
Constant	1.1450^{***}	(0.1373)	-0.0318	(0.1545)	1.1450^{***}	(0.1373)	-0.0318	(0.1545)
Observations	1184		1184		1184		1184	

Note: Standard errors in parentheses and clustered at the individual level. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Reference categories are Short Delay and IR-SD, respectively. Age is 1 for participants older than 40 years. Higher values for Risk, Future Consequences, and Self-Control depict higher willingness to take risks, to be forward-looking and to exhibit higher self-control, respectively. These values are standardized. Total Punishment relates to the overall frequency of inflicted punishment on the individual if caught cheating.

through the combination of delayed uncertainty resolution and significantly delayed punishment (*DR-LD*). Post estimation tests yield no difference between the coefficients of *IR-ND* and *DR-LD* ($p = 0.88$), suggesting that the effectiveness of deterrence is comparable in both cases. It is worth noting that we observe substantial gender heterogeneity indicating that males cheat significantly more than females. The results also suggest that deviant behavior increases with punishment inflicted for caught cheating. This finding indicates that individuals try to make up for incurred losses by increasing the frequency of cheating and taking larger risks, thus being more risk-seeking in losses. Additionally, a participant's age is inversely and significantly correlated with cheating, while our other covariates cannot explain deviant behavior in our sample. Noteworthy, the amount of correct guesses in non-cheating rounds, which are the driving force behind wealth accumulation in our setting, has no significant predictive power for cheating. This

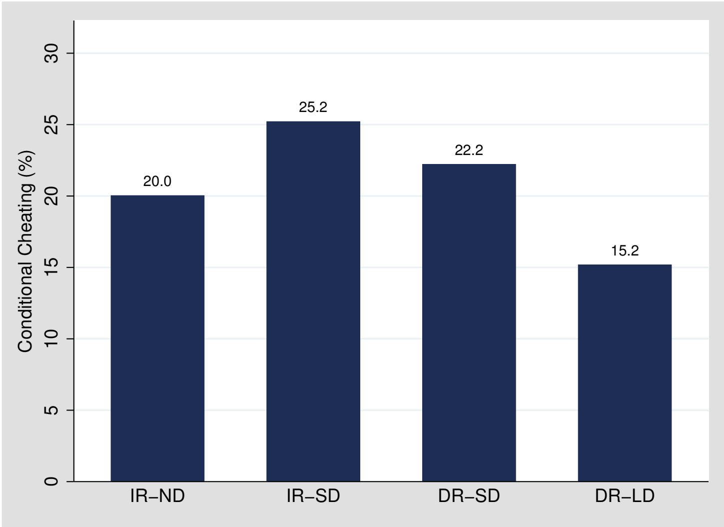


Figure 5.3: Conditional Cheating.

indicates that potential wealth effects cannot explain cheating behavior. All this suggests that swifter punishment or delayed punishment in combination with delayed resolution of uncertainty significantly increases the deterrence of deviant behavior. The delay of uncertainty alone remains non-effective.

We conclude that both very efficient (no delays of punishment) and very inefficient (long delays of punishment in combination with long uncertainty about the status of discovery) punishment institutions are equally effective in deterring deviant behavior.

It is worth noting that one could also plausibly assume the presence of learning effects. A large body of existing literature suggests that the learning effects that emerge through experience are shaped by the timing of rewards and punishments. Due to this, they affect subsequent behavior (cf. Camp et al., 1967; Parke and Deur, 1972). This is of particular importance in the punishment context, because such learning effects would directly speak to the occurrence of recidivism among former felons. Following this logic, the experience of uncertainty and punishment following transgressive behavior could lead to differences in subsequent transgressions. We call this *Conditional Cheating*. Conditional Cheating is defined as the number of individual cheating decisions that proceed the first cheating decision (which can oc-

Table 5.3: Conditional Cheating using OLS

	Analysis by timing				Analysis by treatment			
ConditionalCheating	(1)		(2)		(3)		(4)	
No Delay	−0.1666**	(0.0742)	−0.1757**	(0.0743)				
Long Delay	−0.2129**	(0.0844)	−0.1932**	(0.0851)				
Uncertainty	−0.0782	(0.0709)	−0.0871	(0.0717)				
IR-ND					−0.1666**	(0.0742)	−0.1757**	(0.0743)
DR-SD					−0.0782	(0.0709)	−0.0871	(0.0717)
DR-LD					−0.2911***	(0.0785)	−0.2803***	(0.0789)
Male	0.0881	(0.0570)	0.1178*	(0.0625)	0.0881	(0.0570)	0.1178*	(0.0625)
Age	−0.1298	(0.1033)	−0.1362	(0.1083)	−0.1298	(0.1033)	−0.1362	(0.1083)
GuessCorrect	−0.0831**	(0.0348)	−0.0830**	(0.0360)	−0.0831**	(0.0348)	−0.0830**	(0.0360)
Punishment	0.0759	(0.0738)	0.0684	(0.0736)	0.0759	(0.0738)	0.0684	(0.0736)
Risk			−0.0185	(0.0306)			−0.0185	(0.0306)
FutCons			0.0343	(0.0313)			0.0343	(0.0313)
SelfControl			−0.0151	(0.0320)			−0.0151	(0.0320)
ExpParticipation			0.0691	(0.0627)			0.0691	(0.0627)
Constant	0.7549***	(0.0587)	0.6975***	(0.0762)	0.7549***	(0.0587)	0.6975***	(0.0762)
Observations	189		189		189		189	

Note: Odds ratio reported. Standard errors in parentheses and clustered on the individual level. * $p < 0.1$,** $p < 0.05$,*** $p < 0.01$. Reference categories are Short Delay and IR-SD, respectively. Age is 1 for participants older than 40 years. Higher values for Risk, Future Consequences, and Self-Control depict higher willingness to take risks, to being forward-looking and to exhibit higher self-control, respectively. These values are standardized. Total Punishment relates to the overall frequency of inflicted punishment on the individual if caught cheating.

cur at the beginning of any of the first three blocks). The idea behind this measure is to understand whether experiencing the drain of uncertainty of punishment following their first cheating decision will affect the individual's subsequent propensity to cheat. Our results do not indicate that any such learning effect exists. In fact, cheating behavior following the experience of uncertainty and punishment is congruent to our previous findings on general cheating behavior. We present a graphical illustration in Figure 5.3.

In order to shed light on this mechanism, we employ a series of OLS regressions. Through these regressions we look to analyze the total amount of cheating that took place following the individual's first cheating decision and any resulting punishment that he or she incurred. In our attempt to proxy recidivism, our dependent variable measures the amount of cheating that occured after one's first cheating decision. The less frequent or the later participants recidivise, the lower the value of our dependent variable.

Our results for conditional cheating are consistent with our previous findings, suggesting that the recidivism of individuals is lowest when punishment is either immediate or late when paired with uncertainty. The delay of uncertainty alone is non-effective. In particular, relative to immediate resolution and immediate punishment, a short-term delay of punishment (*IR-SD*) leads to a significant increase in deviant behavior, while the additional introduction of uncertainty (*DR-SD*) alone does not affect cheating rates relative to *IR-ND*. We again find an inverted U-shape relationship; when combining the long delay of punishment with uncertainty of resolution, cheating rates return to levels similar to those found when immediate punishment is paired with no uncertainty resolution (*IR-ND*). In support of this, post estimation tests show that the drop in cheating rates in *DR-LD* is significant compared to cheating in *IR-SD* ($p < 0.01$) and *DR-SD* ($p = 0.02$). In contrast to total cheating behavior, we do not observe robust gender heterogeneity or a traceable impact of age, self-control or experienced punishment. The latter finding indicates that it is not the experience of punishment that affects recidivism rates, but the combined initial experience of uncertainty and timing of punishment.

In summary, we can conclude that the same institutional settings that are capable of reducing recidivism are also the ones deterring deviant behavior in the first place. Our results demonstrate that swift or sufficiently delayed punishment, where the latter is accompanied by an extensive dread of uncertainty regarding one's detection, reduces future criminal behavior.

5.5 Discussion and Conclusion

We investigate along two dimensions how timing can impact the effectiveness of sanctions. We use a controlled laboratory experiment designed to study the effect of delayed punishment and delayed resolution of uncertainty on deterrence. Our experimental findings show that the timing of resolution of uncertainty has no effect on deterrence. For the delay of punishment, we observe the following inverted U-shape relationship: deterrence is highest for no delay or a large delay of punishment and lowest for a short delay of punishment.

The observed inverted U-shape is at odds with both discounted expected utility theory and anticipatory utility theory. According to the first theory, deterrence should decrease monotonically with the delay of punishment. According to the second, there should also be a monotonous relation between deterrence and delay which would be the inverse of that in the previous case if the effect of anticipation is sufficiently high. Recently, Baucells and Bellezza (2016) proposed a new theory of inter-temporal decision making. They extend the existing models of anticipatory utility by a reference point which adjusts. It does so during the anticipation phase by altering a utility of recall in the periods succeeding the consumption and changing the magnitude effect in discounting. In this theory it is possible that the utility maximizing timing of an unpleasant event is somewhere in the middle of the time horizon, i.e. fines in earlier or later periods hurt more and should, therefore, lead to higher deterrence. While our experiment was not designed to test the theory of Baucells and Bellezza (2016) it is the only theory which is compatible with the findings of our experiment.

It is important to note that the effects of the treatments on the total cheating behavior can be obtained by two different, possibly simultaneously operating processes. First, the variations in the experimental treatments could have affected anticipatory reasoning in the participants about how a possible punishment would impact them. If the impact is anticipated to be severe, this could lead to no or delayed cheating. Second, learning processes may have affected cheaters (who at least once underwent the respective treatments) differently by experiencing the (non)waiting for a resolution of uncertainty and the potential execution of an immediate or delayed punishment. This may have influenced their likelihood to cheat again in the future. Inspecting the results for conditional cheating (i.e. future cheating upon having cheated before) shows that they closely mirror the results of the total cheating behavior. Even if some experience for the treatments to become effective would be needed, basic learning theories (e.g. Azrin, 1956; Banks and Vogel-Sprott, 1965) are at odds with the inverted U-shaped relation between deterrence and delay of punishment which is also observed for conditional cheating. Arguably, the highly effective deterrence of deviant behavior in *DR-LD* could be interpreted in one of the following two ways: one, only an extensive delay of punishment, and not the existence of uncertainty resolution, is responsible for the decrease in cheating; two, it is the combination of both the extensive delay in punishment and the existence of uncertainty that imposes additional dread and, thus, the interaction of both is driving the strength of deterrence. Our regression analysis and theoretical foundation suggests that it is most likely the former. We consider this as a promising venue for future research.

Our findings yield important insights for optimally designing sanctioning schemes in legal systems. Existing deterrence literature has almost exclusively focused on the role of severity and certainty of legal consequences in deterring proscribed actions. Our study shows that celerity, the timing of sanctions through sentencing, may also be a crucial component of an effective legal system. Our results imply that punishment should either follow the criminal act quickly or be sufficiently delayed if deterrence is to be maximized. As immediate punishment may be relatively costly, an optimally delayed punishment could be the most efficient solution.

Our study provides a first step into analyzing the effects of deterrence in a sanctioning system. In order to make conclusions for an optimal policy in the real world, future research needs to tackle several limitations of our study. In particular, it seems necessary to study celerity when the delay of punishment extends to the real payout of subjects. Also, the optimal delay may be very sensitive to the type of punishment, e.g. the optimal delay may be rather different for monetary fines than for imprisonment. Despite these limitations we think that our study highlights the role of celerity in designing optimal sanctioning systems and points to fruitful avenues for future research.

Appendix 5.A: Overview Predictions

Table 5.4: Predictions for Total Cheating under DEU

Hypothesis	Predictions	Confirmed?	Sign. Level
H1	$IR-SD > IR-ND$	Yes	**
	$DR-LD > DR-SD$	No	-
H2	$IR-SD = DR-SD$	Yes	Not rejected
H3	$DR-SD > IR-ND$	Yes	**
	$DR-LD > IR-SD$	No	-

Note: Significance levels are the result of one-sided t-tests examining the direction of mean differences based on the theoretical predictions.

Table 5.5: Predictions for Total Cheating under UAE

Hypothesis	Predictions	Confirmed?	Sign. Level
H1*	$IR-SD < IR-ND$	No	-
	$DR-LD < DR-SD$	Yes	**
H2*	$DR-SD < IR-SD$	No	-
H3*	$DR-SD < IR-ND$	No	-
H4*	$DR-LD < IR-SD$	Yes	***
	$DR-LD < IR-ND$	No	-

Note: Significance levels are the result of one-sided t-tests examining the direction of mean differences based on the theoretical predictions.

References

Abbink, K. (2004): "Staff Rotation as an Anti-Corruption Policy: an Experimental Study," *European Journal of Political Economy*, 20, 887–906.

——— (2006): "14 Laboratory experiments on corruption," in *The International Handbook on the Economics of Corruption*, ed. by S. Rose-Ackermann, Edward Elgar, 418–438.

Abbink, K., U. Dasgupta, L. Gangadharan, and T. Jain (2014): "Letting the Briber Go Free: an Experiment on Mitigating Harassment Bribes," *Journal of Public Economics*, 111, 17–28.

Abbink, K. and H. Hennig-Schmidt (2006): "Neutral Versus Loaded Instructions in a Bribery Experiment," *Experimental Economics*, 9, 103–121.

Abbink, K., B. Irlenbusch, and E. Renner (2002): "An Experimental Bribery Game," *Journal of Law, Economics, and Organization*, 18, 428–454.

Abbink, K. and K. Wu (2017): "Reward Self-Reporting to Deter Corruption: An Experiment on Mitigating Collusive Bribery," *Journal of Economic Behavior & Organization*, 133, 256–272.

Acemoglu, D., S. Johnson, and J. Robinson (2005): "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *The American economic review*, 95, 546–579.

Agastya, M. (1999): "Perturbed Adaptive Dynamics in Coalition Form Games," *Journal of Economic Theory*, 89, 207–233.

Alaoui, L. and A. Penta (2015): "Cost-Benefit Analysis in Reasoning," Mimeo: University of Wisconsin.

——— (2016a): "Endogenous Depth of Reasoning," *The Review of Economic Studies*, 83, 1297–1333.

——— (2016b): "Endogenous Depth of Reasoning and Response Time, with an application to the Attention-Allocation Task," Mimeo: Universitat Pompeu Fabra.

Allingham, M. G. and A. Sandmo (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323–338.

Alm, J., M. B. Cronshaw, and M. McKee (1993): "Tax Compliance with Endogenous Audit Selection Rules," *Kyklos*, 46, 27–45.

Alós-Ferrer, C. (2004): "Cournot vs. Walras in Oligopoly Models with Memory," *International Journal of Industrial Organization*, 22, 193–217.

Alós-Ferrer, C. and A. B. Ania (2005): "The Evolutionary Stability of Perfectly Competitive Behavior," *Economic Theory*, 26, 179–197.

Alós-Ferrer, C. and G. Kirchsteiger (2010): "General Equilibrium and the Emergence of (Non) Market Clearing Trading Institutions," *Economic Theory*, 44, 339–360.

——— (2015): "Learning and Market Clearing: Theory and Experiments," *Economic Theory*, 60, 203–241.

Alós-Ferrer, C., G. Kirchsteiger, and M. Walzl (2010): "On the Evolution of Market Institutions: The Platform Design Paradox," *Economic Journal*, 120, 215–243.

Alós-Ferrer, C. and N. Netzer (2015): "Robust Stochastic Stability," *Economic Theory*, 58, 31–57.

Alós-Ferrer, C. and F. Shi (2012): "Imitation with Asymmetric Memory," *Economic Theory*, 49, 193–215.

Alós-Ferrer, C. and S. Weidenholzer (2014): "Imitation and the Role of Information in Overcoming Coordination Failures," *Games and Economic Behavior*, 87, 397–411.

Amir, R. (1996): "Cournot Oligopoly and the Theory of Supermodular Games," *Games and Economic Behavior*, 15, 132–148.

Amir, R. and V. E. Lambson (2000): "On the Effects of Entry in Cournot Markets," *Review of Economic Studies*, 67, 235–254.

Anderson, S. P. and M. Engers (1992): "Stackelberg versus Cournot Oligopoly Equilibrium," *International Journal of Industrial Organization*, 10, 127–135.

Apesteguia, J., M. Dufwenberg, and R. Selten (2007): "Blowing the Whistle," *Economic Theory*, 31, 143–166.

Arad, A. and A. Rubinstein (2012): "The 11–20 Money Request Game: A Level-k Reasoning Study," *The American Economic Review*, 102, 3561–3573.

Arifovic, J. (1994): "Genetic Algorithm Learning and the Cobweb Model," *Journal of Economic Dynamics and Control*, 18, 3–28.

Arifovic, J. and M. K. Maschek (2006): "Revisiting Individual Evolutionary Learning in the Cobweb Model – an Illustration of the Virtual Spite-Effect," *Computational Economics*, 28, 333–354.

Azrin, N. H. (1956): "Some Effects of Two Intermittent Schedules of Immediate and Non-Immediate Punishment," *The Journal of Psychology*, 42, 3–21.

Bailey, W. C. (1980): "Deterrence and the Celerity of the Death Penalty: A Neglected Question in Deterrence Research," *Social Forces*, 58, 1308–1333.

Baker, T., A. Harel, and T. Kugler (2004): "The Virtues of Uncertainty in Law: An Experimental Approach," *Iowa Law Review*, 89, 443.

Banerjee, R. (2016a): "Corruption, Norm Violation and Decay in Social Capital," *Journal of Public Economics*, 137, 14–27.

——— (2016b): "On the Interpretation of Bribery in a Laboratory Corruption Game: Moral Frames and Social Norms," *Experimental Economics*, 19, 240–167.

Banks, R. K. and M. Vogel-Sprott (1965): "Effect of Delayed Punishment on an Immediately Rewarded Response in Humans," *Journal of Experimental Psychology*, 70, 357.

Banuri, S. and C. Eckel (2012): "Experiments in Culture and Corruption: A Review," *New Advances in Experimental Research on Corruption*, 15, 51.

Bartuli, J., B. Djawadi, and R. Fahr (2016): "Business Ethics in Organizations: An Experimental Examination of Whistleblowing and Personality," IZA Discussion Paper No. 10190.

Baucells, M. and S. Bellezza (2016): "Temporal Profiles of Instant Utility during Anticipation, Event, and Recall," *Management Science*, 63, 729–748.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169–217.

Bee, C. C. and R. Madrigal (2013): "Consumer Uncertainty: The Influence of Anticipatory Emotions on Ambivalence, Attitudes, and Intentions," *Journal of Consumer Behaviour*, 12, 370–381.

Bergin, J. and D. Bernhardt (2004): "Comparative Learning Dynamics," *International Economic Review*, 45, 431–465.

——— (2009): "Cooperation through Imitation," *Games and Economic Behavior*, 67, 376–388.

Bigoni, M., S.-O. Fridolfsson, C. Le Coq, and G. Spagnolo (2012): "Fines, Leniency, and Rewards in Antitrust," *The RAND Journal of Economics*, 43, 368–390.

Blume, L. (1993): "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5, 387–424.

Bø, E. E., J. Slemrod, and T. O. Thoresen (2015): "Taxes on the Internet: Deterrence Effects of Public Disclosure," *American Economic Journal: Economic Policy*, 7, 36–62.

Brañas-Garza, P., T. García-Muñoz, and R. H. González (2012): "Cognitive Effort in the Beauty Contest Game," *Journal of Economic Behavior and Organization*, 83, 254–260.

Breuer, L. (2013): "Tax Compliance and Whistleblowing: The Role of Incentives," *The Bonn Journal of Economics*, 2, 7–44.

Buccirossi, P. and G. Spagnolo (2006): "Leniency Policies and Illegal Transactions," *Journal of Public Economics*, 90, 1281–1297.

Butler, J., D. Serra, and G. Spagnalo (2017): "Motivating Whistleblowers," Mimeo: Available at SSRN: https://ssrn.com/abstract=2970551.

Camp, D. S., G. A. Raymond, and R. M. Church (1967): "Temporal Relationship Between Response and Punishment," *Journal of Experimental Psychology*, 74, 114.

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

Cappelen, A. W., K. O. Moene, E. Ø. Sørensen, and B. Tungodden (2013): "Needs versus Entitlements—An International Fairness Experiment," *Journal of the European Economic Association*, 11, 574–598.

Chabris, C. F., C. L. Morris, D. Taubinsky, D. Laibson, and J. P. Schuldt (2009): "The Allocation of Time in Decision-Making," *Journal of the European Economic Association*, 7, 628–637.

Chalfin, A. and J. McCrary (2017): "Criminal Deterrence: A Review of the Literature," *Journal of Economic Literature*, 55, 5–48.

Christöfl, A., U. Leopold-Wildburger, and A. Rasmussen (2017): "An Experimental Study on Bribes, Detection Probability and Principal Witness Policy," *Journal of Business Economics*, forthcoming.

Coricelli, G., M. Joffily, C. Montmarquette, and M. C. Villeval (2010): "Cheating, Emotions, and Rationality: An Experiment on Tax Evasion," *Experimental Economics*, 13, 226–247.

Coricelli, G. and R. Nagel (2009): "Neural Correlates of Depth of Strategic Reasoning in Medial Prefrontal Cortex," *Proceedings of the National Academy of Sciences*, 106, 9163–9168.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69, 1193–1235.

Crawford, V. and M. Costa-Gomes (2006): "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *American Economic Review*, 96, 1737–1768.

Crosetto, P., O. Weisel, and F. Winter (2012): "A Flexible z-Tree Implementation of the Social Value Orientation Slider Measure (Murphy et al. 2011)," Jena Economic Research Paper.

d'Adda, G., V. Capraro, and M. Tavoni (2017): "Push, Don't Nudge: Behavioral Spillovers and Policy Instruments," *Economics Letters*, 154, 92–95.

Dashiell, J. F. (1937): "Affective Value-Distances as a Determinant of Aesthetic Judgment-Times," *American Journal of Psychology*, 50, 57–67.

DeAngelo, G. and G. Charness (2012): "Deterrence, Expected Cost, Uncertainty and Voting: Experimental Evidence," *Journal of Risk and Uncertainty*, 44, 73–100.

Dillenberger, D. (2010): "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica*, 78, 1973–2004.

DYCK, A., A. MORSE, AND L. ZINGALES (2010): "Who Blows the Whistle on Corporate Fraud?" *The Journal of Finance*, 65, 2213–2253.

ECHENIQUE, F. AND K. SAITO (2017): "Response Time and Utility," *Journal of Economic Behavior & Organization*, 139, 49–59.

ECKEL, C. AND P. GROSSMAN (1996): "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 181–191.

EHRLICH, I. (1996): "Crime, Punishment, and the Market for Offenses," *The Journal of Economic Perspectives*, 10, 43–67.

ELLISON, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047–1071.

——— (2000): "Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution," *Review of Economic Studies*, 67, 17–45.

——— (2006): "Bounded Rationality in Industrial Organization," in *Advances in Economics and Econometrics: Theory and Applications*, ed. by R. Blundell, W. K. Newy, and T. Persson, Cambridge: Cambridge University Press, 142–180.

ENGEL, C. (2016): "Experimental Criminal Law. A Survey of Contributions from Law, Economics and Criminology," *Mimeo.*

ENGEL, C., S. J. GOERG, AND G. YU (2013): "Symmetric vs. Asymmetric Punishment Regimes for Bribery," Mimeo.

——— (2016): "Symmetric vs. Asymmetric Punishment Regimes for Collusive Bribery," *American Law and Economics Review*, 18, 506–556.

ENGL, F., A. RIEDL, AND R. A. WEBER (2017): "Spillover Effects of Institutions on Cooperative Behavior, Preferences and Beliefs," IZA Discussion Paper No. 10781.

EVANS, R. (1993): "Observability, Imitation and Cooperation in the Repeated Prisoners' Dilemma," Working Paper, University of Cambridge.

Ewerhart, C. (2014): "Cournot Games with Biconcave Demand," *Games and Economic Behavior*, 85, 37–47.

Falk, A. and F. Zimmermann (2016): "Beliefs and Utility: Experimental Evidence on Preferences for Information," .

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

Fortin, B., G. Lacroix, and M.-C. Villeval (2007): "Tax Evasion and Social Interactions," *Journal of Public Economics*, 91, 2089–2112.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *The Journal of Economic Perspectives*, 19, 25–42.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002): "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40, 351–401.

Fudenberg, D. and D. Levine (1998): *The Theory of Learning in Games*, Cambridge, Massachusetts: The MIT Press.

Gabaix, X. and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121, 505–540.

Gill, D. and V. Prowse (2016): "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis," *Journal of Political Economy*, 124, 1619–1676.

Gill, D. and V. L. Prowse (2017): "Using Response Times to Measure Strategic Complexity and the Value of Thinking in Games," Mimeo: https://ssrn.com/abstract=2902411.

Goeree, J. K., P. Louis, and J. Zhang (2016): "Noisy Introspection in the '11–20' Game," *Economic Journal*, forthcoming.

Golman, R. and G. Loewenstein (2015): "The Demand For, and Avoidance of, Information," Mimeo.

Greiner, B. (2015): "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

Hackenberg, T. D. (2009): "Token Reinforcement: A Review and Analysis," *Journal of the Experimental Analysis of Behavior*, 91, 257–286.

Hamilton, W. (1970): "Selfish and Spiteful Behavior in an Evolutionary Model," *Nature*, 228, 1218–1220.

Heinemann, F. and M. G. Kocher (2013): "Tax Compliance under Tax Regime Changes," *International Tax and Public Finance*, 20, 225–246.

Heinemann, F., R. Nagel, and P. Ockenfels (2009): "Measuring Strategic Uncertainty in Coordination Games," *The Review of Economic Studies*, 76, 181–221.

Heyes, A. and S. Kapur (2009): "An Economic Model of Whistle-Blower Policy," *Journal of Law, Economics, and Organization*, 25, 157–182.

Ho, T.-H., C. Camerer, and K. Weigelt (1998): "Iterated Dominance and Iterated Best Response in Experimental 'p-Beauty Contests'," *The American Economic Review*, 88, 947–969.

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

——— (2002b): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Howe, E. S. and C. J. Brandau (1988): "Additive Effects of Certainty, Severity, and Celerity of Punishment on Judgments of Crime Deterrence Scale Value," *Journal of Applied Social Psychology*, 18, 796–812.

Huck, S., H.-T. Normann, and J. Oechssler (2004): "Through Trial and Error to Collusion," *International Economic Review*, 45, 205–224.

Kandori, M., G. J. Mailath, and R. Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61, 29–56.

Karlin, S. and H. M. Taylor (1975): *A First Course in Stochastic Processes, 2nd Ed.*, San Diego: Academic Press.

Klaus, B. and J. Newton (2016): "Stochastic Stability in Assignment Problems," *Journal of Mathematical Economics*, 62, 62–74.

Kocher, M. G., M. Krawczyk, and F. van Winden (2014): "'Let Me Dream on!' Anticipatory Emotions and Preference for Timing in Lotteries," *Journal of Economic Behavior and Organization*, 98, 29–40.

Kogler, C., L. Mittone, and E. Kirchler (2016): "Delayed Feedback on Tax Audits Affects Compliance and Fairness Perceptions," *Journal of Economic Behavior & Organization*, 124, 81–87.

Krajbich, I., B. Bartling, T. Hare, and E. Fehr (2015): "Rethinking Fast and Slow Based on a Critique of Reaction-time Reverse Inference," *Nature Communications*, 6, 7455.

Krajbich, I., B. Oud, and E. Fehr (2014): "Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference," *The American Economic Review*, 104, 501–506.

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

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1999): "The Quality of Government," *The Journal of Law, Economics, and Organization*, 15, 222–279.

Lambsdorff, J. G. and B. Frank (2010): "Bribing versus Gift-Giving – an Experiment," *Journal of Economic Psychology*, 31, 347–357.

Landsberger, M. and I. Meilijson (1982): "Incentive Generating state Dependent Penalty System: The Case of Income Tax Evasion," *Journal of Public Economics*, 19, 333–352.

Lindner, F. and M. Sutter (2013): "Level-k Reasoning and Time Pressure in the 11–20 Money Request Game," *Economics Letters*, 120, 542–545.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 97, 666–684.

Lovallo, D. and D. Kahneman (2000): "Living with Uncertainty: Attractiveness and Resolution Timing," *Journal of Behavioral Decision Making*, 13, 179.

Mandel, A. and H. Gintis (2014): "Stochastic Stability in the Scarf Economy," *Mathematical Social Sciences*, 67, 44–49.

Mittone, L. (2006): "Dynamic Behaviour in Tax Evasion: An Experimental Approach," *The Journal of Socio-Economics*, 35, 813–835.

Moffatt, P. G. (2015): *Experimetrics: Econometrics for Experimental Economics*, Palgrave Macmillan.

Mosteller, F. and P. Nogee (1951): "An Experimental Measurement of Utility," *Journal of Political Economy*, 59, 371–404.

Moyer, R. S. and T. K. Landauer (1967): "Time Required for Judgements of Numerical Inequality," *Nature*, 215, 1519–1520.

Murphy, R. O., K. A. Ackermann, and M. J. J. Handgraaf (2011): "Measuring Social Value Orientation," *Judgment and Decision Making*, 6, 771–781.

Nagel, R. (1995): "Unraveling in Guessing games: An Experimental Study," *The American Economic Review*, 85, 1313–1326.

Nagin, D. S. and G. Pogarsky (2001): "Integrating Celerity, Impulsivity, and Extralegal Sanction Threats into a Model of General Deterrence: Theory and Evidence," *Criminology*, 39, 865–892.

——— (2004): "Time and Punishment: Delayed Consequences and Criminal Behavior," *Journal of Quantitative Criminology*, 20, 295–317.

Nax, H. H. and B. S. R. Pradelski (2015): "Evolutionary Dynamics and Equitable Core Selection in Assignment Games," *International Journal of Game Theory*, 44, 903–932.

Newton, J. (2012): "Recontracting and Stochastic Stability in Cooperative Games," *Journal of Economic Theory*, 147, 364–381.

Orviska, M. and J. Hudson (2003): "Tax Evasion, Civic Duty and the Law Abiding Citizen," *European Journal of Political Economy*, 19, 83–102.

Papke, L. E. and J. M. Wooldridge (2008): "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics*, 145, 121–133.

Parke, R. D. and J. L. Deur (1972): "Schedule of Punishment and Inhibition of Aggression in Children," *Developmental Psychology*, 7, 266.

Paternoster, R. (2010): "How Much Do We Really Know About Criminal Deterrence?" *The Journal of Criminal Law and Criminology*, 765–824.

Polonio, L., S. Di Guida, and G. Coricelli (2015): "Strategic Sophistication and Attention in Games: An Eye-Tracking Study," *Games and Economic Behavior*, 94, 80–96.

Primi, C., K. Morsanyi, F. Chiesi, M. A. Donati, and J. Hamilton (2015): "The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT)," *Journal of Behavioral Decision Making*.

Raymond, M. (1999): "Enforcement Leverage when Penalties are Restricted: a Reconsideration under Asymmetric Information," *Journal of Public Economics*, 73, 289–295.

Riechmann, T. (2006): "Cournot or Walras? Long-Run Results in Oligopoly Games," *Journal of Institutional and Theoretical Economics*, 162, 702–720.

Samuelson, L. (1997): *Evolutionary Games and Equilibrium Selection*, Cambridge, Massachusetts: The MIT Press.

Sandholm, W. H. (2010): *Population Games and Evolutionary Dynamics*, Cambridge, Massachusetts: The MIT Press.

Schaffer, M. (1989): "Are Profit-Maximisers the Best Survivors?" *Journal of Economic Behavior and Organization*, 12, 29–45.

Schmolke, K. U. and V. Utikal (2016): "Whistleblowing: Incentives and Situational Determinants," Mimeo: FAU.

Shi, F. (2015): "Long-run Technology Choice with Endogenous Local Capacity," *Economic Theory*, 59, 377–399.

Shleifer, A. and R. W. Vishny (1993): "Corruption," *The Quarterly Journal of Economics*, 108, 599–617.

Skinner, B. F. (1963): "Operant Behavior," *American Psychologist*, 18, 503.

Slemrod, J. (2007): "Cheating Ourselves: The Economics of Tax Evasion," *The Journal of Economic Perspectives*, 21, 25–48.

Spagnolo, G. (2004): "Divide et Impera: Optimal Leniency Programs," Mimeo: CEPR.

——— (2006): "Leniency and Whistleblowers in Antitrust," in *Handbook of Antitrust Economics*, ed. by P. Buccirossi, Cambridge, Massachusetts: MIT Press.

Spiegler, R. (2006): "Competition Over Agents with Boundedly Rational Expectations," *Theoretical Economics*, 1, 207–231.

Spiliopoulos, L. and A. Ortmann (2017): "The BCD of Response Time Analysis in Experimental Economics," *Experimental Economics*, forthcoming.

Stahl, D. O. (1993): "Evolution of Smart n Players," *Games and Economic Behavior*, 5, 604–617.

Stahl, D. O. and P. W. Wilson (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

Strathman, A., F. Gleicher, D. S. Boninger, and C. S. Edwards (1994): "The Consideration of Future Consequences: Weighing Immediate and Distant Outcomes of Behavior," *Journal of Personality and Social Psychology*, 66, 742.

Strzalecki, T. (2013): "Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion," *Econometrica*, 81, 1039–1074.

Tangney, J. P., R. F. Baumeister, and A. L. Boone (2004): "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success," *Journal of personality*, 72, 271–324.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014): "Assessing Miserly Information Processing: An Expansion of the Cognitive Reflection Test," *Thinking & Reasoning*, 20, 147–168.

Tversky, A. and D. Kahneman (1986): "Rational Choice and the Framing of Decisions," *Journal of business*, S251–S278.

Vallée, T. and M. Yildizoğlu (2009): "Convergence in the Finite Cournot Oligopoly with Social and Individual Learning," *Journal of Economic Behavior & Organization*, 72, 670–690.

——— (2013): "Can They Beat the Cournot Equilibrium? Learning with Memory and Convergence to Equilibria in a Cournot Oligopoly," *Computational Economics*, 41, 493–516.

VEGA-REDONDO, F. (1997): "The Evolution of Walrasian Behavior," *Econometrica*, 65, 375–384.

VRIEND, N. J. (2000): "An Illustration of the Essential Difference Between Individual and Social Learning, and Its Consequences for Computational Analyses," *Journal of Economic Dynamics and Control*, 24, 1–19.

WEISEL, O. AND S. SHALVI (2015): "The Collaborative Roots of Corruption," *Proceedings of the National Academy of Sciences of the United States of America*, 112, 10651.

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

YOUNG, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61, 57–84.

——— (1998): *Individual Strategy and Social Structure*, Princeton, New Jersey: Princeton University Press.

YU, J. (1994): "Punishment Celerity and Severity: Testing a Specific Deterrence Model on Drunk Driving Recidivism," *Journal of Criminal Justice*, 22, 355–366.