Essays in Microeconomic Theory

Luis Frones

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2021

To Deborah.

Abstract

This thesis uses game theory and microeconomics to investigate empirical puzzles. In Chapter 1 we provide a formal model to explain why new legislation that has widespread quickly in the last decades failed to achieve its objectives in accordance to recent evidence. In Chapter 2 we uncover a new class of equilibria in the canonical social learning setting with endogenous timing of decisions. We argue how our results can offer a social learning explanation for two applications: delays in the adoption of policy measures during the Covid-19 pandemic, and the timing of investments in the venture capital industry. In Chapter 3 we provide a formal model to explain recent evidence that correlates high levels of residential segregation based on income with low intergenerational mobility.

Apologies are considered as a cheap and strong mechanism to restore broken relationships. Between 1999 and 2011, the number of US states with *apology laws*, legislation that excludes the admissibility of apologies in court, increased from 2 to 38, along with all the Australian jurisdictions, the United Kingdom, most of the Canadian provinces, and Hong Kong. Legislators' hope is that by passing these laws apologies will be encouraged, with the consequence that civil disputes will settle more often and lawsuits will be prevented. However, recent evidence from US shows that these laws have had the opposite effect: apology laws have increased the number of lawsuits. In Chapter 1 we provide an explanation for why apology laws fail that is consistent with the best available evidence. We show that apology laws may reduce settlements by encouraging insincere apologies which in turn induce plaintiffs not to accept apologies. We contribute to show on which type of relationships apology laws fail: apology laws preclude the settlement of cases that are socially valuable to be settled. Moreover, for the cases where these laws increase litigation we show that apology laws induce more miscarriages of justice and deter inter-party communication.

In Chapter 2 we ask: Does waiting to observe others' action delay profitable choices? If so, for how long? We characterize long delays in a social learning environment. In contrast with previous work, we show the existence of equilibria in which agents end up adopting a profitable and risky policy with substantial delay. These results point to social learning as a plausible explanation for delays evidenced in the adoption of policy measures during the Covid-19 pandemic. Next, we allow agents to choose the quality of their information before deciding. We show how in this setting long delays may also exists, and how our equilibrium sheds light on the investment timing patterns evidenced in the venture capital industry.

Recent evidence shows a negative association between social mobility and residential segregation based on income. In Chapter 3 we provide a theory that explains this link based on beliefs in a just world. Our argument is that segregated communities exhibit more polarized and pessimistic views that hard work pays off than integrated ones because families in those communities learn differently about the value of effort. This polarization and pessimism in segregated communities make in turn mobility lower, as those families with low beliefs in effort have higher income inertia. We model agents as trying to learn the relative importance of effort and predetermined factors in the generation of income. They learn from two sources, by socialization in neighbourhoods and from their dynastic income mobility experience. In a dynamic model, we characterize conditions on initial beliefs under which the society exhibits in the long run income segregation with low rates of social mobility, or income integration with high social mobility rates. We provide evidence for U.S. that support our theoretical results. Using survey-data with beliefs in a just world we show that more segregated communities are correlated with more polarized and pessimistic views about the value for effort.

Acknowledgements

I am grateful to Daniel Seidmann, Matthias Dahm, Silvia Sonderegger and Mikhail Safronov for their invaluable support, patience and guidance during my PhD study. I want to thank Facundo Albornoz and Omar Licandro for their support and advice at different stages of this process.
I am thankful for financial support provided by the School of Economics at the University of Nottingham, ANII at Uruguay, and for research assistantships from Matthias Dahm and Facundo Albornoz.
Finally, I am grateful to my family for constant support, specially to Deborah, Ada and Isabella for joining me in this adventure.

Contents

Chapter 1

Why do apology laws fail?

1.1 Introduction

Should apologies be admissible into evidence as proof of fault in civil cases? In other words, in a civil case where Smith has sued Jones, and Jones had apologized for what happened, should we allow Smith to use Jones's apologies as evidence of Jones's fault at trial? This question has been debated in the legal arena for more than 20 years. The answer seems negative in the light of what happened the lasts decades in common law jurisdictions. Legislation that limits the admissibility of apologies in court for proving civil liability, broadly referred as *apology laws*, has become the norm. Between 1999 and 2011, the number of US states with apology laws increased from 2 to 38, along with all the Australian jurisdictions, the United Kingdom, most of the Canadian provinces, and Hong Kong (Vines (2015); Myers (2016); Kleefeld (2017)). But why they have became so popular? As the administration of civil justice is increasingly compromised by high litigation cost and delays, legislators hope that by passing these laws civil disputes will be resolved more amicably and less expensively. In other words, legislators expected to prevent lawsuits and speed up settlements with apology laws.[1] Research in psychology and social sciences shows that an apology is a strong and cheap device to restore social or economic relationships that have been disturbed. Drawing on these results, legislators passed these laws with the following intended mechanism in mind. By "protecting" apologies from admissibility, defendants will apologize more often, and then, given this proliferation of apologies, plaintiff will take fewer cases to trial.

Despite legislators' intentions, recent empirical evidence shows that apology laws have the opposite effect: they decrease out-of-court settlements and increase cases at trial (Ho and Liu (2011a); McMichael et al. (2019); McMichael (2021)). In this paper we offer an explanation for why and when apology laws fail. We use a signaling model to analyse the decision of defendants to apologize and plaintiffs to accept it in the shadow of the courtroom, when apologies are or are not admissible in court. This allows us to examine the effect of making apologies inadmissible on settlement/trial outcomes. We show that apology laws have the *intended effect* of legislators -more settlements/less trial- but also an *unintended effect* -fewer settlements/more trial. In the latter case, we show that apology laws reduce settlement by encouraging insincere apologies that in turn motivate plaintiffs to reject apologies. Importantly, which effect is prevalent depends on the type of relationship con-

[1] "One new strategy to promote the early, effective and affordable resolution of disputes that is being considered by the Ministry of Attorney General is apology legislation." (Discussion paper on apology legislation, British Columbia Ministry of Attorney General (2006), p.1). These laws seek to "reduce lawsuits and encourage settlements" (California Assembly Commission on Judiciary, TENN. R. EVID. 409.1 cmt.).

sidered and the context of the conflict. We show that the unintended effect of apology laws arises in relationships where parties benefit more if the relationship is restored -long term relationships-, and in context where the defendant's reputation concerns are not too high. In contrast, the intended effect of apology laws arises for comparatively less valuable -short term- relationships and higher reputational concerns.

Previous work has showed that an apology may serve to restore broken and possibly fruitful relationships. However, as the sincerity of an apology and willingness to forgive are generally parties' private information, apologies might not work and could make things worse.[2] Our argument is based on a two-sided incomplete information environment between litigants that captures this apologies' dilemma.[3] On the one side of the conflict, the plaintiff has incomplete information about the sincerity of an apology and the liability of the defendant. Drawing from previous research on apologies, we allow for sincere and insincere apologies. We consider *sincere apologies* as a communication device designed to restore broken relationships. In our framework, sincere and insincere defendants are differentiate by their respective gains from restoring the relationship. The latter ones issue *insincere apologies* tendered just to avoid trial.[4] On the other side of the conflict, the defendant is not sure if an apology would be accepted. Whereas a subset of plaintiff will be ready to accept just sincere apologies, i.e. the ones tendered by sincere defendants, other plaintiffs do not. Then, as we allow for plaintiffs that do not care about re-establishing the relationship, apologizing is risky: some plaintiffs may use apologies at trial if they are admissible as they only care about indemnity payment. In this context, we uncover two channels through which apology laws induce insincere defendants to apologize more often: i) by cheapening apologies, as now apologies cannot be used as evidence in courts; and ii) by increasing the expected average damage awards for insincere-guilty defendants at trial, as now the judge cannot observe a rejected sincere apology at trial, more guilty defendants are pooled with innocent ones and expected awards at trial increase.

Importantly, we show that the unintended effect of apology laws arises in context where parties could benefit more if the relationship is restored. Before an apology law, apologies were costly because they could be used in court by plaintiffs if rejected. This potential cost of apologies coupled with innocent defendants that refuse to apologize enable plaintiffs to infer the sincerity of an apology. On the one hand, insincere defendants preferred to remain silent, as they would be pooled at with innocent defendants and pay less. On the other hand, sincere defendants with high enough valuation of the relationship take the risk of apologizing. In this sense, by encouraging insincere apologies, we show that apology laws reduce the settlement of cases that are socially valuable to be settled. On the other hand, we show that apology laws may induce more settlement for cases where future gains derived from a restored relationship are not so high. The potential risk of plaintiffs us-

[2]There is a large body of evidence showing that a harmdoer who sends an apology is much more likely to be forgiven than a non-apologizer (Weiner et al., 1991; Ohbuchi and Sato, 1994; Ohtsubo and Watanabe, 2009; Abeler et al., 2010; Fischbacher and Utikalc, 2013; Ho and Liu, 2011a). "Victims seem to trust that an apology is more than the attempt to get around trouble or punishment."(Fischbacher and Utikalc (2013), p.593). However, lab experiments showed that apologies can sometimes backfire if they are seen as insincere (Fischbacher and Utikalc, 2013; Gilbert et al., 2018)

[3]The model captures the vicious cycle that Benjamin Ho explains at freakonomics: "...doctors are afraid to apologize because they're scared of getting sued. But the patients, the only reason they sue is perhaps because they never got an apology. To combat this, a lot of states started passing what are called 'I'm sorry' laws" (`https://freakonomics.com/podcast/apologies/`, accessed on 25/09/2020).

[4]Cohen (2002), p.872: "The main arguments against these laws are that such laws could: (1) induce insincere, manipulative apologies from unremorseful injurers...". Fischbacher and Utikalc (2013), p.593: "We find that offenders primarily apologize if they fear punishment for the offense. Evidently it is not remorse that makes a harmdoer apologize but the hope to prevent punishment."

ing apologies in court is what prevented settlement in these cases absent an apology law: it was too costly for sincere defendants. Therefore, the exemption of apologies in court may work. This highlights the original motivation of apology laws legislators', and we provide conditions for this to hold.

Our results offer an explanation for the existing evidence of the unintended effect of apology laws. When a lawsuit is litigated fully, there is a public record of the events. In contrast, apologies often lead to "non-lawsuits" and private settlements about which data is more difficult to obtain. McMichael et al. (2019) explore the unintended effect of apology laws in the context of medical malpractice cases through a dataset that includes information on claims that resulted in lawsuits as well as those that did not −the ones that settled privately.[5] For the particular speciality they have data, some physicians focus primarily on seeing patients in an office setting, referred as non-surgeons, while others both see patients in an office and perform surgery, surgeons.[6] They found that apology laws do not affect settlement/trial outcomes for surgeons, but reduce settlement −and increase lawsuits− against non-surgeons. We argue that this is consistent with an explanation based on the physicians' valuation of the relationship. In their database non-surgeons treat patients for several year, whereas surgeons interact in discrete events.[7] In this sense, our theory predicts that the unintended effect of apology laws arises for cases where physicians have high gains from restoring the relationship, provided apologies are not too costly. This explains the result for non-surgeons. On the other hand, when this valuation is lower, as for surgeons, our results can also show that these laws have no effect on litigation outcomes. We discuss the explanation of McMichael et al. (2019) for their results in Section 1.4.1 and argue that the evidence is more consistent with the one provided in this work. Finally, our predictions also explain the results of McMichael (2021) that uses novel information about medical malpractice insurance premiums charged to physicians across three separate specialties: general surgery, internal medicine, and obstetrics/gynecology.[8] McMichael (2021) shows that apology laws have caused an increase in malpractice premiums for all specialties, but the effect is higher for internists. As the focus of this specialty is on dealing with long term doctor-patient relationships, our results based on differences in physician's valuation of the relationship offer a plausible explanation.[9]

These findings have implications for legislators. Apology laws' intended effects are dependant on the type of relationship that has been broken. In particular, the doctor-patient relationship may enter in the type where apology laws reduce possible settlement. This is relevant as 28 of the 38

[5]McMichael et al. (2019), p.341: "This dataset includes substantially more information than is publicly available and, thus, presents a unique opportunity to understand the effect of apology laws on the entire litigation landscape in ways that are not possible using publicly available data." Ho and Liu (2011a) finds mixed evidence of the effect of apology laws in the medical malpractice context using pubic available data. This data only includes cases that ended with a positive payment to the claimant.

[6]For confidentiality reasons McMichael et al. (2019) are not able to identify either the insurer or the specialty.

[7]McMichael et al. (2019), p.369: "Surgeons generally interact with and treat patients in a discrete event, i.e., the surgery they are performing plus any pre-operative and post-operative care. ... On the other hand, non-surgeons generally treat their patients over the course of years or may interact with patients a number of times when attempting to resolve an injury or illness."

[8]McMichael (2021), p.4 footnote 22: "...the data analyzed here offer two important advantages: (1) malpractice premiums represent an amalgamation of the factors that influence malpractice liability risk compiled by insurance companies, whose profitability depends on accurately capturing this risk, and (2) these data are not subject to the substantial problem of missing information that affects publicly available malpractice claims data."

[9]See the American College of Physicians, about internal medicine, detailing the care provided by internists: "Their training uniquely qualifies them to practice primary care and follow patients over the duration of their adult lives and establish long and rewarding personal relationships with their patients.", "...a discipline focused ... in caring for patients in the context of thoughtful, meaningful doctor-patient relationships." (https://www.acponline.org/about-acp/about-internal-medicine, accessed on 25/09/ 2020)

US states' apology laws were approved for healthcare providers. We also uncover other negative outcomes of apology laws. First, our model allows us to explore whether apology laws cause greater miscarriages of justice, an implication that has not been addressed in the public discussion of apology laws. We show that for cases where the unintended effect on settlement arises, apology laws result in more miscarriages of justice. Whereas before apology laws some guilty defendants did not pay, after their introduction more guilty defendants are exempted or innocent ones end up paying damages. Second, we show that if there is a small cost to apologize for defendants when they know that an apology will be rejected, apology laws also reduce communication. This is relevant as, apart from the pretended effect on litigation, legislators were also motivated by the idea that apology laws would encourage more apologies because its intrinsic social value.[10] Therefore, our results predict that for cases where apology law reduce settlements, as for the healthcare context, innocent doctor would pay more often and less apologies would be tendered. Finally, our model highlights two other negative consequences of apology laws that were raised in the literature. First, when apology laws have the intended effect, public confidence in the courts could be adversely affected.[11] Second, if frivolous claims are prevalent, apology law may encourage the acceptance of insincere apologies. This last result may be relevant for the discussion of apology laws in the context of the tort reform debate.[12]

The chapter is organized as follows. In the next subsection we review the literature. In Section 1.2 we provide an overview of the main elements of apology laws. In Section 1.3 we present the model. In Section 1.4 we analyse the effects of apology laws. In Section 1.5 and 1.6 we respectively discuss our modelling choices and conclude. Proof of Proposition 2 is in the appendix.

1.1.1 Related literature

Our work fits in the law and economics literature studying settlement/trial outcomes under asymmetric information.[13] Most models focus on bilateral bargaining following early models by Bebchuk (1984); Reinganum and Wilde (1986); Spier (1992). Our focus is on settlements/trial outcomes induced by apologies and not on bargaining arguments. More closely papers are Daughety and Reinganum (1995) and Seidmann (2005) that analyse outcomes of institutional rules that preclude particular evidence from being used at trial. Daughety and Reinganum (1995) studies the rule that exempt settlements offer as evidence in trials on settlement outcomes, Seidmann (2005) the effect of the right to silence on convictions rate and accuracy. As in Daughety and Reinganum (1995) we analyse signaling games with two audiences, a plaintiff and a judge, and compare the outcomes of making the signal private or public. In this sense, our model contributes to the literature on strategic information transmission with multiple audiences (Farrell and Gibbons (1989); Goltsman and Pavlov (2011)). Finally, there is a literature regarding apology laws from legal scholars, either discussing the pros and cons of these laws (Taft (2000); Cohen (2002); O'Hara (2004); Heimreich

[10]"Factors in favour of apology legislation include: a. To avoid litigation and encourage the early and cost-effective resolution of disputes; b. To encourage natural, open and direct dialogue between people after injuries; and c. To encourage people to engage in the moral and humane act of apologizing after they have injured another and to take responsibility for their actions"(British Columbia Ministry of Attorney General (2006), p.4).

[11]"Negative factors include: a. Public confidence in the courts could be adversely affected if a person who has admitted liability in an apology is found not liable;..." (British Columbia Ministry of Attorney General (2006), p.4).

[12]Some scholars consider apology laws as another tort reform design to reduce litigation and favour the health-care industry. Cohen (2002), p.856: "Sophisticated defendants are going take advantage of naive injured parties through these laws. They'll issue apologies knowing that there's no real risk involved, but naive injured parties will think these apologies are meaningful - that they do involve risk. Injured parties will think the injurers are putting their necks on the line when in fact they aren't." See the discussion in McMichael (2018).

[13]Spier (2007) and Daughety and Reinganum (2012) provide excellent overviews of the literature.

(2012); Vines (2015)) or testing their implications (Robbennolt (2003, 2006); Ho and Liu (2011a,b); McMichael et al. (2019); McMichael (2021)).

More broadly, our work is related to an existing apology literature in psychology, sociology, medicine, and law. In economics (Ho, 2012) introduces a costly signaling model to study apologies in a principal-agent framework where apologies are exogenously costly. Apart from the fact that our model has two audiences, we add a judge that makes apologies endogenously costly, there is a difference in terms of the interpretation of apologies, admissions or excuses, that we believe may be relevant for readers interested in modelling apologies. In his model, an effective apology shifts the principal's attribution of the cause of the bad outcome from the agent's disposition to the external situation. In this sense, after an accepted apology the principal do not find the agent 'guilty' and apologies can be interpreted as excuses. However, an apology usually implies admissions of fault where the agents acknowledge her responsibility. Our framework allows for the distinction between both interpretation of apologies, admissions and excuses. In Section 1.5 we discuss this modelling difference in detail. Ho (2012); Fischbacher and Utikalc (2013) study in experiments when apologies may work. Finally, there is a recent literature that focuses on how firms should apologize (see Halperin et al. (2019) for a large-scale field experiment with Uber's customers and Abeler et al. (2010) for a field experiment with ebay).

1.2 Apology laws: how they work and for which cases

Apology legislation includes statutory provisions that remove legal disincentives to offering an apology in the context of civil disputes.[14] The regulatory framework differs to a large extent between various jurisdictions. The points of difference concern the type of apology that is protected, the interference with other areas of law (evidence law as well as substantive law) and the scope of coverage. We explain each of these features below. Despite the disparity between apology laws legislation, there has been a trend towards a broader legal protection and a wider scope of application. The first apology act enacted in Massachusetts in 1986 provided solely for inadmissibility of benevolent statements as evidence of fault or liability in the context of an accident. While US states mainly followed this path (with some alterations), the introduction of apologies acts in the Australian states and territories (as of 2002) and in the Canadian provinces (as of 2006) expanded the scope remarkably . A large part of those acts apply to civil liability of any kind, cover apologies which may include admissions of fault and sometimes even encompass specific provisions on insurance and limitation. The Hong Kong Apology Ordinance of 2017 is described as "the most ambitious apology law yet".[15]

How they work. Excluding apologies from admissibility into evidence is one of the most widespread forms of evidentiary adjustments. Almost all jurisdictions limit the admissibility of apologies for proving fault or liability.[16] Inadmissibility means that apologies cannot be part of the decision-making process. A court will not receive nor consider them in deciding on liability. When there is a jury trial, the apologies shall not be available to jurors. The laws in most jurisdictions stipulate that apologies are not admissible "as evidence of fault or liability in connection with the matter" or something similar. From a common law perspective, an admission is an oral or written statement

[14] For this section we rely on Kleefeld (2017)'s recent survey of apology laws from an international law point of view.

[15] However, there is not an unique trend, the aspirations of the Scottish Apologies Act of 2016 and the amendment of the Irish Civil Liability and Courts Act of 2015 are much more moderate.

[16] The exceptions are England and Wales.

or conduct made directly by or on behalf of a party which goes against the interests of that party. Admissions may be formal or informal. Apologies made outside the court room are associated with informal admissions. Unlike formal admissions, they are tendered as substantive evidence by the opponent, subject to contradiction or explanation and may be denied, explained away or contradicted by the maker (e.g. by establishing that it was made for some secondary reason and not true). Legislation in almost all jurisdictions stipulates in more or less similar terms that an apology does not constitute "an express or implied admission of fault or liability". Finally, some jurisdictions have made the policy choice to broaden the act's ambit beyond the amendment of evidence laws. The apology acts in the Canadian provinces, Hong Kong and Ireland expressly bars using an apology to void or otherwise affect insurance coverage. Insurers would regularly tell their clients that an apology will render insurance coverage void.[17]

Definition of apology: full vs partial apology laws. The 'law and apology' literature often makes a distinction between full and partial apologies. Accordingly, the basic distinction mostly comes down to the question whether the legal provisions merely protect expressions of sympathy and benevolence, defined in the literature as "partial apology laws", or also include fault-admitting apologies, "full apology laws".[18] The former category provides for protection of expressions of a general sense of sympathy or benevolence. It encompasses statements such as "I am sorry this happened to you". The latter refers to apologies which are not confined to expressions of sympathy or regret, but also incorporate admissions of fault or wrongdoing. From a legal perspective, such an apology is much more burdensome than a mere communication of sympathy. Australian and Canadian legislation underline this point by adding to the definition "whether or not the apology admits or implies an admission of fault in connection with the matter". In six US states, the definition of apology extends to admissions of fault as well.[19]

Scope of coverage. There is a diversity in scope of application of apology legislation as well. The Massachusetts act in 1986 provided solely for inadmissibility in the context of accidents. While some US states have followed this path, most of them have even narrowed down the scope to medical malpractice cases (28 out of 38). Other common law systems refer to civil liability of any kind; this path is taken by Australian and Canadian lawmakers.[20]

[17]This would sometimes even be reflected in insurance clauses hindering insured persons from making apologies to those whom they injured.

[18]However, in accordance with apology theorists, an acknowledgement of fault without any expression of sympathy or regret should also be considered a partial apology. According to these scholars, apologies consist of different building blocks: an affirmation or acknowledgement of fault; an expression of regret, remorse or sorrow; a willingness to repair and a promise to adapt future behaviour. In this context, reference is also made to the four "Rs" (responsibility, remorse, reparation and resolution). Whereas partial apologies would consist of some, but not all of these components, full apologies would enclose all or at least the majority of them. Although all jurisdictions have defined the term apology in their legislation, the willingness to repair (the action component) and the commitment to change future behaviour (the articulation of forbearance) are almost always lacking. See Kleefeld (2007) and the references there for a discussion.

[19]This happens in a more subtle way than under Australian or Canadian law. Definitions of apologies in the US normally consist of a list of emotions which are expressed or affirmed (i.e. apology, sympathy, commiseration, condolence, compassion, regret or a general sense of benevolence). Those six US states integrate terms such as fault, mistake and error or responsibility and liability into that list.

[20]A limited number of US states also refers to "any claim growing out of the event" or "a cause of action in tort". The UK Compensation Act applies to "negligence or breach of statutory duty". The Scottish Apologies Act and Hong Kong Ordinance even have a broader scope, covering "all civil proceedings" or "applicable proceedings". These apology acts can potentially impact other areas of law (such as family law, contracts, commercial litigation and administrative law).

1.3 The Model

We consider a simple model of litigation in which litigants may settle by tendering and accepting apologies. This simple model allows us to analyse the effects of apology laws on settlement/trial outcomes. The objective is to characterize equilibria of two games: Γ^a, in which apologies are admissible in court, and Γ^{-a}, in which they are inadmissible. There are three players in both games: a defendant (D), a plaintiff (P) and a judge (J), that we refer to as he, she and it respectively. Our model is motivated by the following the history. The history starts after P has suffered damages and wants to initiate a formal claim against D. After she notifies D about the decision, D decides whether or not to apologize to P, an action which J does not directly observe. If D does not apologize then the case goes to trial; otherwise P decides whether to accept or reject the apology. If she accepts it, then the litigants settle; if she rejects it, the case proceeds to trial. The next figure shows the time line of both games:[21]

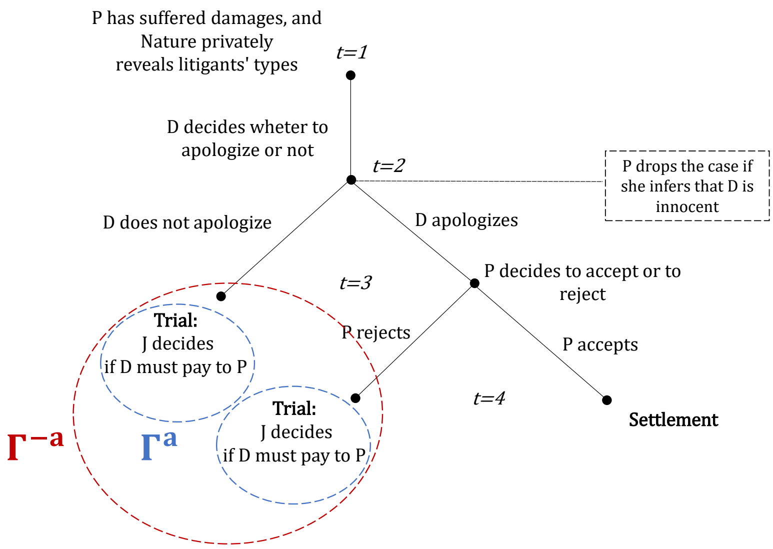

Figure 1.1: Timeline

We start our formal presentation by describing the elements of both games in the following order: the time line, strategies for players, how the award is decided at trial and the difference between games and, finally, the payoffs for each player.

Timing. At $t = 1$ Nature selects the litigants' type. At $t = 2$ D decides whether to apologize or not. The subgame played at $t = 3$ depends on D's decision. On the one hand, if D does not apologize at $t = 2$, then the case goes to trial at $t = 3$ where J decides, the players payoffs are realized and the game ends. On the other hand, if D has apologized at $t = 2$, P decides at $t = 3$. If P decides to reject an apology, the case goes to trial at $t = 4$ where J decides and the game ends. If P decides to accept an apology the game ends without trial, an outcome that we call settlement.

[21] In the standard models of settlement bargaining with asymmetric information, any settlement bargaining outcome between litigants is determined by the judge's (equilibrium) decision at trial. Then, the trial outcome of our model can be interpreted as a trial/settlement bargaining outcome. What is relevant for our discussion of apology laws is that any settlement-bargaining/trial outcome (unmodeled) reached by litigants implies a higher cost to the civil justice system than the settlement-induced-by-apologies outcome (that we model), which is what legislators tried to promote.

Notice that there are three final nodes in the game: trial if D has not apologized, trial if D has apologized and P has rejected it, and settlement.

Litigants' type. The model intends to capture the following apologies' dilemma. On the one hand, some plaintiffs, but not all, would accept sincere apologies and settle because there are mutual gains from restoring the relationship. As not all plaintiffs accept apologies, apologizing is risky. On the other hand, not every apology tendered may be sincere. We assume that the sincerity of an apology and willingness-to-forgive are litigants' private information, and therefore the two litigants are each privately informed.[22] Specifically, D has private information about her liability, i.e. whether he is innocent or not, and also whether he values the relationship with P. D has 3 possible types, denoted by d: d_i (the innocent type), d_s (the sincere guilty type) and d_n (the insincere guilty type). The prior probability of type d_x is denoted by p_x for $x \in \{i, s, n\}$. On the other hand, P has private information about whether she would accept a sincere apology, defined as an apology tendered by d_s. P has two possible types: P_F (the forgiving type) and P_L (the litigious type), with the prior probability of type P_L denoted by q_L.

Strategies. D must choose whether to apologize or not at $t = 2$. P chooses at $t = 3$ only if D has apologized at $t = 2$. She must choose whether to accept an apology or to reject it. (As it will be clear later, the decision of P to present the apologies at trial or not when admissible (in Γ^a) is irrelevant for the analysis). We assume that P drops the case at $t = 3$ if she infers that D is innocent. Because P will never drop the case in equilibrium, we simplify the exposition of the model by not including this decision stage in the timeline and by not defining the payoffs of the players when the case is dropped. Finally, J must decide whenever the game reaches trial. It must decide whether to award damages of $m \in \{0, M\}$ to P, where $m = M$ (resp. $m = 0$) implies that D was found liable (resp. not liable) for the damages in court.

Trial and difference between games. At trial, J observes whether an apology has been made in Γ^a, but not in Γ^{-a}, if the case goes to trial. We also assume that there is an exogenous probability, $\lambda \in (0, 1)$, that the trial itself reveals the true liability of D in the case, whereas with probability $(1 - \lambda)$, J must infer the liability by the (equilibrium) strategies of litigants. When we refer in the text to the award m decided by J, we mean J's decision when the trial does not reveal the true liability of D in the case.

Payoffs. Now we describe the payoffs for each player in the three relevant outcomes of the game. *(Judge)* Whenever any game reaches a trial, J wants to award her best guess about the damages given all admissible evidence. We assume that it gains zero when a correct award is made, whereas the losses incurred by court mistakes are symmetric: it loses the same amount either when a guilty D is exempted from indemnity payments or when an innocent one pays it. Because of this symmetry defining the amount of the losses is irrelevant as J will only consider whether D is more probable guilty or innocent to decide.[23] J's loss if the case does not go to trial (the settlement outcome) is irrelevant in equilibrium so we do not define this payoff. *(Defendant)* We treat apologies

[22] Proposition 0 in the appendix shows that two-sided incomplete information is needed for apology laws to have an effect on settlement/trial outcomes.

[23] We believe the symmetry in J's payoff is not an unreasonable assumption. This implies a standard of proof, i.e. the critical probability of guilt at which the jury is indifferent between acquitting and convicting at trial, of 0.5. Note that the standard of proof in a civil case is proof on the balance of probabilities, which means that the party bearing the burden of proof, the innocent, must prove that her case is more probable than not. When we discuss the miscarriages of justice of apology laws in section 1.4.1 we introduce other asymmetric standards of proof.

as a confession, that is, an apology communicates that a fault has been made. In this sense, we assume that an innocent type never apologizes, he will not provide a false confession (we discuss this assumption in Section 1.5). The following payoffs apply to guilty defendants. We allow for future payoffs derived from a restored relationship, denoted by $v_d \in \{v_s, v_n\}$, and this payoff is what differentiates the sincere and insincere types. Type d_s receives a positive payoff, $v_s > 0$, if he apologizes and the relationship is restored; d_n does not care about the relationship, $v_n = 0$, and the only reason for her to apologize is to avoid trial. Finally, we allow for apologies to be directly costly $-$besides other potential legal consequences of an apology in court. This cost depends on whether apologies are accepted by P, with a cost of $t \geq 0$ for D; or not accepted, in which case D incurs the cost of c. When an apology is accepted, and parties settle, we interpret t as a monetary transfer to compensate P and restore the relationship. On the other hand, when an apology is rejected, we interpret c as a "reputation" cost, a psychological cost of admitting fault, or the risk of future litigation with other potentially damaged parties involved in the incident.[24] In a trial outcome, D loses the award $m \in \{0, M\}$. Given our discussion, as d_i never apologizes he ends at trial where he pays $-m$. On the other hand, the payoffs of each guilty D in the three possible outcomes of the game are as follows:

$$\text{(guilty) D's payoffs:} \begin{cases} v_d - t & \text{(Settlement)} \\ -c - m & \text{(Trial if he apologizes)} \\ -m & \text{(Trial if he does not apologize)} \end{cases}$$

(Plaintiff) We suppose that P does not litigate if she believes that D is innocent, and her payoff when she drops the claim in this case is irrelevant for the analysis. We introduce the possibility of frivolous claims in Section 1.4.3. Now we describe each P type's payoff at settlement. When they settle, there are two potential payoffs. First, both types receive the compensation included in the apology, t. Second, there are potential gains from restoring the relationship with a sincere D who has apologized, denoted by $v_P \in \{v_F, v_L\}$. The forgiving type, P_F, receives a positive payoff $v_F > 0$ from accepting an apology from d_s, but not from d_n; P_L does not gain a positive payoff from restoring the relationship with any guilty type, $v_L = 0$. Therefore, P_F receives $v_F + t$ if she settles with d_s and just t when she settles with d_n, and P_L receives t in any settlement agreement irrespective of the defendant type. At trial, both types receive the award m. Then, the payoffs of P are given in the following expression:

$$\text{P's payoffs:} \begin{cases} v_P + t & \text{(Settlement)} \\ m & \text{(Trial -any trial outcome)} \end{cases}$$

Unspecified payoffs. We have not defined two payoffs: P's payoffs when she receives the award at trial from d_i, and J's payoff when litigants settle. These payoffs will not be relevant for the analysis of the equilibria, but they are for equilibrium selection issues. We come back to this when we discuss our selection criteria. The payoffs of the players if P drops the case are irrelevant for the analysis, so we omit them.

[24] Evidence shows that transgressors who apologize in situations in which competence is relevant suffer a negative impact on their perceived competence, and speakers are aware of this (Weiner et al., 1991; Kim et al., 2006). Physicians are reluctant to give apologies because it damages their reputation (Gallagher et al. (2003)). See Chaudhry and Loewenstein (2019) for a theory, and evidence that explain why apologizing is not costless. We consider that a trial process risks D to a higher level of public exposure in comparison to a settlement agreement, which is why we choose not to include c when an apology is accepted (in a settlement outcome). Our results hold if different guilty defendant types incur different apology costs.

Parameters restrictions. As mentioned in the Introduction, an apology law is a mechanism designed to encourage settlement through apologies. We take two premises about plaintiffs and apologies that legislators considered when they designed these laws and then we analyse the effects of apology laws under these assumptions. The premises are two. First, apologies may work: there is a subset of plaintiffs, P_F, that is ready to accept an apology and settle. Second, apologizing is risky. The plaintiff's willingness-to-forgive is private information and some plaintiffs, i.e. P_L, may use an apology as evidence at trial when it is allowed, as they only care about indemnity payments. We could have motivated our P_L as a litigant who receives gains by taking the case to court. For example, a patient may want to ensure that the physician is sanctioned in court and does not practice medicine any more, or to damage his reputation as a public trial implies a higher exposure for defendants in comparison with private settlement arrangements.[25] However, to ease the exposition, we rely on compensation as P_L's motivation. These two premises require setting relevant values for the compensation included in an apology, t, relative to P's expected award at trial. We adopt the following assumption:[26]

Assumption (A): $max\{0; M - v_F\} < t < \lambda M$

The lower bound on t refers to the apologies-may-work premise: given $M < v_F + t$, P_F will accept an apology tendered by d_s irrespective of the trial outcome. The payoff from settlement with d_s, $v_F + t$, exceeds M, the maximum amount of expected indemnity payments at trial.[27] The upper bound on t refers to the risk of apologizing: given $t < \lambda M$, P_L prefers to reject the compensation included in the apology, t, and to take the case to trial where she can expect an award of at least λM from a guilty type. This upper bound has two implications. First, as $t < M$, d_n will prefer to apologize and pay t to P_F, if apologies are accepted, rather than paying the positive indemnity payments $m = M$. Second, for the same reason, P_F will prefer to not accept an apology from this type.[28]

Solution concept. Our solution concept is pure strategy perfect Bayesian equilibrium.

1.4 Analysis

We start the analysis by introducing the following Lemmas that characterize the two types of equilibria that exist in both games, equilibria with and without settlement. Lemma 1 describes the unique pair of litigant types that may settle in equilibrium. Lemma 2 characterizes the pooling equilibria that exists when there is no settlement outcome in equilibrium. After presenting these Lemmas, we discuss how we compare the outcomes of the two games. In order to simplify the exposition we present the analysis for the case where $c > 0$, but all the results hold for $c = 0$.[29]

[25] Legal systems have always allowed for the possibility that a plaintiff may sue driven by a desire to harm the defendant's finances and reputation, and have even adopted countermeasures to deal with some possibilities (see the discussion of such measures in Guha (2016)).

[26] Note that we consider t as exogenous and not the result of a bargaining outcome between litigants. We analyse the games for different values of t. (A) implicitly assumes $\lambda > 1 - v_F/M$ if $M > v_F$: the trial process has a prior positive probability to uncover the truth.

[27] Notice that we do not analyse the possibility that apology laws encourage settlement because courts become less efficient as less evidence is available at trial (this would require allowing for $M > v_F + t$).

[28] Our main results for why and when apology laws induce less settlement will hold whenever $\lambda M < t < M$. The assumption that $t < \lambda M$ simplifies the exposition.

[29] When $c = 0$ we need to consider equilibria in which defendants apologize despite knowing that apologies are rejected.

Lemma 1: *(Litigant types in settlement) In any game, any settlement must be between d_s and P_F.*

PROOF. d_i never settles as we have assumed that he never apologizes and that the case goes to trial when no apologies have been tendered. P_L can also not settle, as we have assumed that $t < \lambda M$ in (A). She prefers an expected award of at least λM at trial when she receives an apology from a guilty type rather than the settlement's payments. d_n cannot be the unique defendant type that settles through apologizing in equilibrium. If this were the case, P_F would infer that an apology comes from an insincere type and therefore would not accept it by (A). Finally, we need to prove that there cannot be an equilibrium where both guilty types, d_s and d_n, apologize and the apologies are accepted by P_F. The key assumption in the argument is that there are no frivolous claims. Assume that we have an equilibrium where both guilty types settle. If that were the case, both plaintiffs would drop the case after receiving no apologies, as only d_i would not apologize. As the case without apologies would be dropped, d_n would prefer to deviate to not apologizing and pay nothing instead of paying t to compensate P_F -or the cost of a rejected apology and subsequent trial when he encounters a P_L. ■

Notice that the argument about the impossibility of having settlement with both guilty types relies on the assumption that both types of plaintiffs drop the case whenever the only d that not apologizes is d_i. The result can also be proved by considering cases where P_F will not accept a pool of sincere and insincere apologies. To avoid the introduction of a new condition for this we rely on the above argument. However, we sometimes refer to the impossibility of having an equilibrium where P_F accepts both apologies as *apologies are non credible.*[30]

Lemma 2: *(Pooling equilibria without settlement) For all the parameters of the model, Γ^a and Γ^{-a} have an equilibrium without settlement which prescribes all of D's types not to apologize and J to award damages of $m = 0$ (resp. $m = M$) if $p_i > p_n + p_s$ (resp. $p_i < p_n + p_s$). These are the unique equilibria without settlement.*

PROOF. Because the equilibrium prescribes that no apology is tendered, P_F does not make a decision. This equilibrium holds for P_F's off-the-path beliefs specifying that an apology must come from d_n. Notice that given P_F's off-the-path beliefs, she would reject an apology if tendered. Therefore, as rejected apologies are costly, $c > 0$, and they are always rejected by any P, guilty defendants prefer to not apologize. As every defendant ends up in court, J's award will only depend on the prior evidence, and therefore on whether p_i is higher or lower than $p_n + p_s$.

(Uniqueness) Notice that an equilibrium without settlement must feature all D's types at trial. This can happen because no type apologizes or because some type apologizes but apologies are rejected. The equilibria in the above paragraph cover the case where no type apologizes. For the other case, we have argued that if D knows that P will not accept an apology, he prefers to not apologize because $c > 0$. Hence, there cannot be an equilibrium with rejected apologies. ■

We say that apology laws have an effect when the settlement/trial equilibrium outcomes differ between the games. In the appendix we provide existence conditions for the equilibria with settle-

[30] Notice that P_F will reject apologies from both guilty types whenever $v_F \frac{p_s}{p_s+p_n} < T - t$, where T is the expected award at trial against a guilty type. The benefit from keeping the relationship with d_s is lower than the difference between the award and the settlement compensation with both types. When this condition holds the result in Lemma 1 applies. The analysis of section 1.4.3 uses this condition.

ment in both games. Notice that the equilibria without settlement of Lemma 2 always exist. By contrast, the equilibria with settlement do not cover the whole space of parameters in any game, so there are parameters of the model under which just an equilibrium without settlement exists. On the other hand, because of the existence of equilibria without settlement of Lemma 2, when an equilibrium with settlement exists there are multiple equilibria. Therefore, to discuss the effects of apology laws (differences between games) we need to specify our equilibrium selection criterion, that is, which equilibrium is played in each game. Our selection criterion is Pareto dominance. Specifically, we select the equilibrium that results in higher payoffs for every type of every player (D,P, and J). This criterion implies that we need to define two payoffs that were unspecified in Section 1.3. First, the payoffs for P when she receives the indemnity payments at trial from an innocent defendant and, second, the payoff of J when there is settlement. For the former, we assume that every P-type loses an amount if the award m at trial comes from d_i. For the latter, that J receives the same payoff that P receives if P decides to settle.[31] Assuming these payoffs, we can show that if P loses enough when she receives the indemnity payments at trial from innocent types, an equilibrium with settlement always results in higher payoffs for every player-type than an equilibrium without settlement (when both equilibria exists simultaneously and for both games). Therefore, given our discussion, whenever there is an equilibrium with settlement in any game, this is the equilibrium that we assume is played.

Now that we have specified how we select an equilibrium for each game, we can compare the games and analyse the effects of apology laws. By changing J's available evidence, the introduction of apology laws will change D's incentives to apologize. In Figure 1.2, panel a, we provide an overview of the apology laws effects that we prove formally in the next subsections. The Figure illustrates for each game (before and after apology laws) and for each guilty D, the expected payoffs of apologizing minus the expected payoff of not doing so (denoted by Π in the Figure). The unintended effect arises because before apology laws just d_s apologizes (his net payoffs from apologizing are positive). However, after apology laws d_n also apologizes, as the Figure shows; but by Lemma 1 there cannot be an equilibrium that features settlement with this type. That is, an unintended effect of apology laws arises when players were playing an equilibrium with a settlement outcome before apology laws that is broken after the introduction of apology laws. After the introduction of an apology law, the pooling equilibrium without settlement of Lemma 2 is played. On the other hand, the intended effect arises when the opposite happens: from a pooling equilibrium without settlement to an equilibrium with settlement.

To understand the incentives of D let us consider the trade-off involved in the decision of whether to apologize or not when apologies are admissible−in Γ^a. Notice that in equilibrium D will apologize if apologies are accepted by P_F. Otherwise, he will not apologize as apologies are costly when rejected ($c > 0$). In equilibrium the guilty-defendant's apology decision depends on the following inequality:

$$(1 - q_L)(v_d - t) - q_L(c + M) \gtrless -\lambda M - (1 - \lambda)m \tag{1}$$

The left−hand−side of (1) represents the payoff from apologizing. Whenever type d believes that P_F will accept an apology, with probability $(1 - q_L)$ he must compensate P_F by paying t and obtains v_d, the gains from restoring the relationship. However, he also faces the risk of encountering P_L with probability q_L. In this latter case P_L will then take the apology to trial where, as

[31] Notice that, by Lemma 1, any equilibrium with settlement features P_F accepting an apology from d_s and receiving the payoffs of $v_F + t$. In this sense, P is compensated by D in a settlement outcome, and this compensation is higher than any award at trial as $v_F + t > M$ by (A).

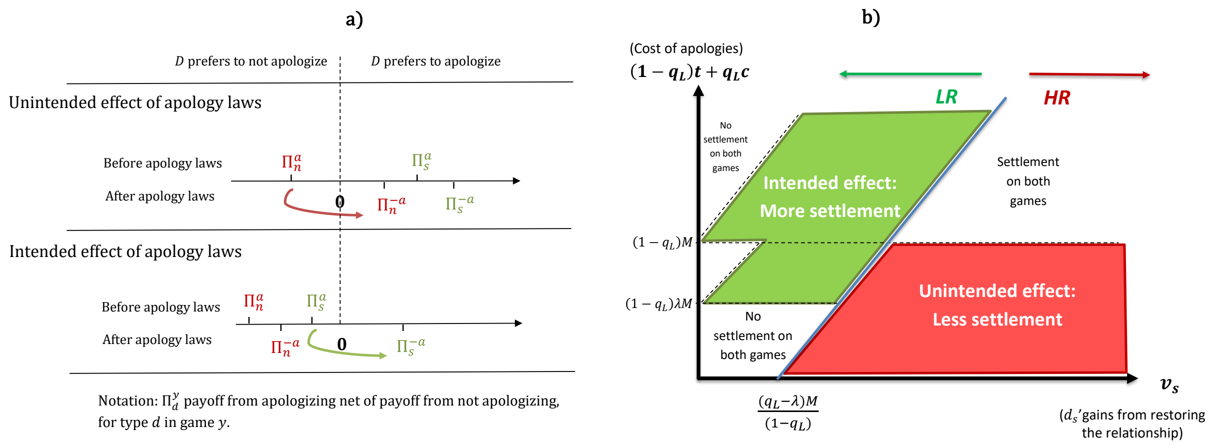

Figure 1.2: The effects of apology laws. a) How incentives to apologize change. b) Types of relationships affected (the graph is for a fix q_L, λ, M and $q_L > \lambda$)

apologies are admissible and innocent types never apologize, he must pay $m = M$. D also incurs the cost of a rejected apology, c. On the other hand, the right−hand−side represents the expected payoff from not apologizing and facing trial with an expected cost of $\lambda M + (1-\lambda)m$: the trial uncovers the liability with probability λ, so he expects to pay λM; and with probability $(1-\lambda)$ J awards $m \in \{0, M\}$ depending on its inference about which types end up in court without apologies.

Now we label an important condition that will be relevant for the identification of the effects of apology laws on settlement and on the types of relationships affected.

Definition: We will say that d_s has *high gains from restoring the relationship* (HR) with P_F (resp. *low gains* (LR)) when the following condition is satisfied (resp. not satisfied)

$$(1-q_L)(v_s-t) - q_L(c+M) > -\lambda M \quad \text{or equivalently,}$$
$$v_s > \frac{(q_L-\lambda)M + q_Lc}{1-q_L} + t$$

That is, whenever d_s has HR he prefers to apologize and face the risk of paying M at trial when he encounters P_L, instead of not doing it and paying λM. Notice that when d_s has HR he prefers to apologize even though the award at trial when no apology is presented is the lowest that he can expect to pay, λM. On the other hand, whenever d_s has LR he prefers to remain silent, apologies are too costly. In Figure 1.2, panel b, we show that the effects of apology laws differ on both sides of this condition. The Figure illustrates, for fixed values of M, λ, q_L, the effects of apology laws as a function of v_s (d_s's value of restoring the relationship with P_F), and the expected cost of apologies, $(1-q_L)t + q_Lc$ (if apologies are accepted by P_F). Condition HR is a straight line (with slope $(1-q_L)$) that divides the plane. Then, as can be seen in the Figure, when d_s has HR just the unintended effect of apology laws may arise; whereas when d_s has LR is the intended effect that may happen.

We provide the complete characterization of the equilibria with settlement of each game in the appendix. In the main text we focus on the subset of parameters where the two games have different equilibrium outcomes. In the next subsections we explain the effects of apology laws in detail.

1.4.1 The unintended effect of apology laws

We start by constructing an equilibrium with settlement when apologies are admitted, and then show how the introduction of apology laws breaks the previous settlement.

Lemma 3: *(Settlement before apology laws)* Γ^a *has an equilibrium which prescribes type* d_s *alone to apologize, type* P_F *to accept an apology, and* J *to award damages of* $m = 0$ *at trial when no apology is presented and* $m = M$ *when it is presented, if and only if*

i) d_s *has* HR,
ii) $(1 - q_L)t + q_L c > (\lambda - q_L)M$, *and*
iii) $p_i > p_n$

PROOF. (Sufficiency) *(Plaintiff)* By construction, P and J would infer D's type from an apology; so type P_F cannot profitably deviate to not accepting an apology, from which settlement follows. P_L always rejects an apology by assumption (A).
(Judge) By construction, the case would only go to trial without apologies if D's type is either d_i or d_n. Hence, given $p_i > p_n$, J would award $m = 0$ at trial when no apologies are presented. On the other hand, when it observes an apology at trial it awards $m = M$, as just the guilty type d_s apologizes.
(Defendant) Consequently, given J's best reply, d_s and d_n expect to pay λM by not apologizing. To see why notice that the case must then proceed to trial without apologies where they pay $m = 0$ with probability $1 - \lambda$, i.e. whenever the trial does not uncover the true liability. Then, given that P_L would always turn down an apology, we have

$$d_s\text{'s apologies payoff}: \ (1 - q_L)(v_s - t) - q_L(c + M) > -\lambda M$$
$$d_n\text{'s apologies payoff}: \ (1 - q_L)(-t) - q_L(c + M) < -\lambda M$$

The first inequality holds by condition (i), d_s has HR; hence d_s cannot profitably deviate to remain silent. The second inequality follows by (ii), and d_n cannot profitable deviate to apologizing. Finally, d_i never apologizes by assumption.

(Necessity) Suppose d_s has LR (condition (i) does not hold). Then d_s can profitably deviate to apologizing as the cost of apologizing, $-(1 - q_L)(v_s - t) + q_L(M + c)$, exceeds λM, the expected award he must pay at trial without apologies. This proves necessity of condition (i). By an analogous argument, if (ii) does not hold then d_n would prefer to deviate to apologizing. Finally, assume $p_n > p_i$. By our prescribed equilibrium the types that end up in court without apologies presented are d_i and d_n. Therefore, now J would award $m = M$ at trial when no apology is presented. This change in the expected trial outcome would modify D's incentives to apologize (and therefore conditions (i) and (ii)). Equilibrium a2 in the appendix considers this case. ■

Notice two key conditions that allow settlement in this equilibrium by separating the guilty defendant's incentives to apologize. First, d_s must have HR: he prefers to take the risk of apologizing even though apologies are admissible and presented by litigious plaintiffs. Second, there must be high enough prior evidence of innocence in the case ($p_i > p_n$), that makes the expected trial award without apologies low enough for d_n, as he is pooled with d_i.

Consider now that apology laws are introduced in an environment where conditions $i - iii$ of the above equilibrium hold. Proposition 1 below shows that apology laws reduce settlements, the *unintended effect* of legislators, while the effects of these laws on miscarriages of justice are analysed in Section 1.4.1 We will show that apology laws induce insincere apologies in these environments for two different reasons. First, because apology laws reduce the legal consequences of apologies as they cannot be presented as evidence; we call this *cheaper apologies*. Second, because of *J's higher awards*: as J can now not observe sincere apologies when they are tendered and rejected by P_L at trial, it awards positive indemnity payments more often than before because of an adverse inference about liability. Both effects cause the previous settlement outcome to fail by Lemma 1.

Proposition 1: *(The unintended effect) The introduction of apology laws reduces settlement between litigants whenever the equilibrium of Lemma 3 with settlement before apology laws exists (d_s has HR, $(1-q_L)t + q_L c > (\lambda - q_L)M$, and $p_i > p_n$) and*

- *a) (Cheaper apologies)* $(1-q_L)t + q_L c < (1-q_L)\lambda M$ *and* $p_i > p_n + q_L p_s$*, or*
- *b) (J's higher awards)* $(1-q_L)t + q_L c < (1-q_L)M$ *and* $p_i < p_n + q_L p_s$

PROOF. *(Cheaper apologies.)* Suppose, per contra, that Γ^{-a} has the same outcome as that constructed for Γ^a in Lemma 3. Notice that given these strategies and $p_i > p_n + q_L p_s$, J would award $m = 0$ at trial. To see why note that d_i and d_n end up in court because they do not apologize, and d_s because he apologizes and P_L rejects. Then, as J cannot distinguish between cases that reach trial with and without apologies in Γ^{-a}, $p_i > p_n + q_L p_s$ implies that her best reply is $m = 0$ when only prior evidence can be used by J to decide. Consider next d_n's incentives to apologize. Given the best reply of J, d_n will pay λM at trial whether he apologizes or not, as there is just one information set for J in this game. Therefore, d_n faces now a lower risk from apologizing than before the introduction of apology laws: the new cost from apologizing is $(1-q_L)t + q_L(c + \lambda M)$, which is less than $(1-q_L)t + q_L(c + M)$ (before apology laws). This lower cost arises as apologies cannot be used by P_L as evidence after the introduction of apology laws, and J awards $m = 0$ in the unique trial outcome given $p_i > p_n + q_L p_s$. Then, if the new cost of apologizing is low enough, $(1-q_L)t + q_L(c + \lambda M) < \lambda M$, he will prefer to apologize as this cost is lower than the cost of remaining silent, λM. But we have already proved in Lemma 1 that no equilibrium prescribes P_F to accept an apology offered by both guilty types.

(J's higher awards). Suppose, per contra, that P and D choose the same strategies as in Lemma 3. Notice that given these strategies and $p_i < p_n + q_L p_s$, J would award $m = M$ at trial. To see why note that d_i and d_n end up in court because they do not apologize, and d_s because he apologizes and P_L rejects it. Then, as J cannot distinguish between cases that reach trial with and without apologies in Γ^{-a}, $p_i < p_n + q_L p_s$ implies that J chooses $m = M$ when only prior evidence can be used by J to decide. Consider next d_n's incentives to apologize. If he does not apologize and ends up in court, he must pay M; whereas if he apologizes he pays M to P_L alone, as P_F accepts apologies in our prescribed equilibrium. Then if the cost of apologizing is low enough, in particular, if $(1-q_L)t + q_L(c + M) < M$, d_n will prefer to apologize as this cost is lower than the cost of remaining silent and paying M at trial. But we have already proved in Lemma 1 that there cannot be settlement with both guilty types. ■

Relationships affected by apology laws

Apology laws affect on the type of relationships than can be restored. Figure 1.2, panel b, shows the effects of apology laws as a function of d_s's valuation in restoring the relationship with P. Notice that by Proposition 1, apology laws decrease settlement for cases where d_s has high gains from restoring the relationship with P_F, HR. Now let us define the *value of a restored relationship* by $v_F + v_s$. The higher this amount, the higher the social value of rules that promote the recovery of relationships. Recall from (A) that $t > max\{0; M - v_F\}$. Then, the higher the value P_F obtains by recovering the relationship with d_s, v_F, the lower may be the compensation needed to restore a relationship, t. In this sense, notice from Figure 1.2, panel b, that apology laws start working when the ex-ante cost of apologies is high enough, and do not work whenever this cost is low (for example whenever $(1 - q_L)t + q_L c < -(1 - q_L)\lambda M$ apology laws have either null or negative effects on settlement). Therefore, a higher v_F allows for lower values of t that in turn increases the set of parameters at which the unintended effect holds. The following corollary follows from this observation and Proposition 1:

Corollary 1: *(Relationships affected) Apology laws can be detrimental to restoring socially valuable relationships.*

Explaining the evidence

We focus on the evidence provided in McMichael et al. (2019) which is the best available evidence as argued in the Introduction (see footnote 5). The data comes from a national malpractice insurer that contains information on 90% of all US physicians practising in a single specialty from 2004 through 2014. For this speciality, they have data, some physicians focus primarily on seeing patients in an office setting, referred to as non-surgeons, while others both see patients in an office and perform surgery, surgeons. They found that apology laws do not affect settlement/trial outcomes for surgeons, but reduce settlement and increase lawsuits against non-surgeons. As non-surgeons treat patients repeatedly, whereas surgeons interact in discrete events, we argue that this is consistent with an explanation based on the physicians' valuation of the relationship. Recall that a sincere defendant is defined as one who receives a positive payoff if the relationship is restored. The difference in the valuation between both types of physician can be translated to the model in two different ways. First, the ex-ante proportion of sincere and insincere guilty physicians in the case, p_s and p_n (the extensive margin of valuation). Second, the value of restoring the relationship, v_s, for sincere physicians (the intensive margin of valuation).[32] We explain both possibilities below:

1. Prior "evidence" of sincerity. The difference in a physician's valuation of the relationship can be translated into how J perceives the prior sincerity of defendants for each physician. Suppose that the ex ante probability of facing a sincere defendant conditional on facing a guilty defendant, $p_s/(p_s + p_n)$, is lower for surgeons than for non-surgeons. That is, p_s is lower and p_n higher for surgeons than for non-surgeons from J's perspective. Then, condition (iii) of Lemma 3, $p_i > p_n$, is more relaxed for non-surgeons than for surgeons for the same prior evidence about liability p_i vs $p_s + p_n$. Hence, if the conditions for the unintended effect holds for non-surgeons, but there were no cases for surgeons before apology laws that satisfy condition (iii) in Lemma 3, a high enough $p_n/(p_s + p_n)$, the empirical results will show that apology laws only have a detrimental effect on settlement's outcomes for non-surgeons and no effect for surgeons (and the others conditions of the

[32] In our explanation we assume that c, λ and q_L are the same for both types of physician as we do not have arguments to set differences in these parameters between both physicians. As "Surgeons pay higher malpractice premiums than non-surgeons since they face higher malpractice liability risk." (McMichael (2021), p.364) we could consider that c, the reputation or physiological cost of an apology, is lower for surgeons as litigation is more common for them. Introducing this difference in c between physicians would go in line with the argument of this section, as fixing the parameters of the unintended effect for non-surgeons, lower c and valuation for surgeons imply no effects for surgeons.

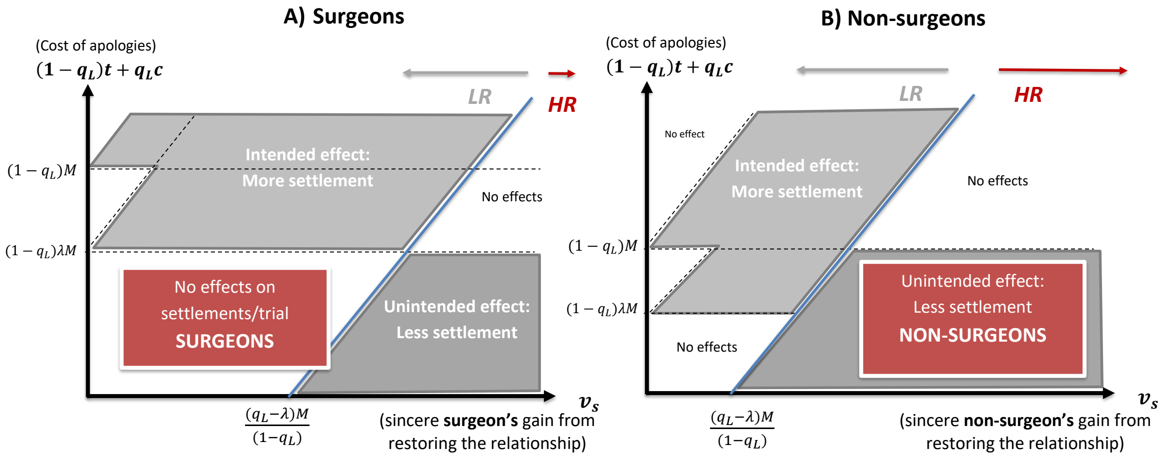

Figure 1.3: Explaining the evidence. Same parameters for panel A and B except that M is higher in A.

unintended effect are satisfied).

2. Value of restoring the relationship. This is relevant for d_s's decision to apologize. Assume that for the same level of sincerity for each type of physician, i.e. same p_s and p_n, v_s is lower for surgeons than for non-surgeons. This makes the unintended effect more easily to arise for non-surgeons, all else equal, as HR (condition (i)) of Lemma 3 is more relaxed the higher v_s is. We illustrate this argument in Figure 1.3, where in line with McMichael et al. (2019)'s results, we also consider that indemnity payments, M, are on average higher for surgeons than for non-surgeons irrespective of the introduction of an apology law.[33]

Comments on McMichael et al. (2019).

1. McMichael et al. (2019) explain their results by the hypothesis that combines: i) more apologies from physicians after apology laws, and ii) symmetric (surgeons) vs asymmetric (non-surgeons) information environments for each type of malpractice case. Their argument is that apology laws induced physicians to "rush to apologize following an error without completely understanding the risks and complexities of apologizing in the wake of an error."[34] The difference in outcomes between surgeons and non-surgeons is explained by considering apologies as a signal of malpractice for plaintiffs. They argue that asymmetric information about liability, i.e. whether the plaintiff knows the defendant is guilty or not, is more likely to be present in malpractice claims involving non-surgeons than surgeons.[35] Then, if physicians apologize more often after apology laws, previous malpractice cases that were not filed because they were not noticed by patients, mostly with non-surgeons ones, would end up in court more often.[36] Proposition 0 in the appendix shows that our model is also

[33]See table C.3 in the appendix of McMichael (2021), p.403.
[34]McMichael et al. (2019), p.390.
[35]McMichael et al. (2019), p.369: "Surgeons generally interact with and treat patients in a discrete event, i.e., the surgery they are performing plus any pre-operative and post-operative care. Because of this discrete interaction, patients who suffer an injury will likely have little trouble tracing that injury to an error that occurred during surgery. On the other hand, non-surgeons generally treat their patients over the course of years or may interact with patients a number of times when attempting to resolve an injury or illness. Thus, observing the malpractice of non-surgeon physicians may be more difficult."
[36]This hypothesis is not completely consistent with their empirical results. Once a patient decides to pursue a

consistent with this explanation. When liability is common knowledge, i.e. symmetric information, apology laws have no effect on litigation outcomes. However, our results offer an alternative explanation based on the physicians' valuation of the relationship when there is asymmetric information on both sorts of physician's cases, and provide conditions under which the unintended effect arises.

We argue that our explanation is more consistent with their evidence. McMichael et al. (2019) argue "If apology laws increase malpractice risk and, on balance, are not in their best interests, why would physicians continue to apologize? While future research should investigate this question in detail, the most likely answer is that physicians have simply been conditioned to apologize with little training on how to do so effectively...Moreover, many physicians are not involved in multiple malpractice actions, so they have little reason to know—particularly given the positive treatment of apologies from a variety of sources—that apologizing can increase their malpractice liability risk."[37] Note that if their hypothesis holds, it is plausible to expect a short term impact of apology laws on litigation through this mechanism. Once physicians learn that apologies increase malpractice risk, no more apologies would be tendered. Then, if apologies as a signal of malpractice is the main force for the unintended effect, one should expect that the effect will diminish over time. In contrast, the mechanism uncovered by our theory, as an equilibrium result, predicts a long term effect of apology laws. Once players learn how to play, the unintended effect should be more stable. In this sense, McMichael et al. (2019) find that the unintended effect for non-surgeons did decrease over time: "we find no additional evidence from our state-level analysis that suggests the net effect of apology laws on medical malpractice liability risk is zero (or possibly negative) in the long run (many years after an apology law is passed). Examining the probability that a physician will be subject to a claim directly, we find evidence that apology laws simply increase the probability of lawsuits for non-surgeons in general and no evidence that this effect dissipates over time."[38] Finally, as argued in the Introduction, the physician-valuation explanation of the apology's laws effect is also consistent with the results between different medical specialties of McMichael (2021).

2. Policy implication. The hypothesis of more physicians' apologies after apology laws causes McMichael et al. (2019) to moderate the advice of repealing these laws.[39] "To the extent that apology laws promote transparency in the physician-patient relationship through the revelation of otherwise hidden malpractice, they may benefit society. Indeed, this transparency may elucidate errors that would have been repeated but for the apology that was offered. Because this increased transparency comes at the cost of increased malpractice liability risk, state lawmakers must weigh transparency against liability in deciding whether to repeal apology laws."[40] Our results predict that apology laws do not necessarily increase transparency. If there is a small cost to apologizing

claim, she notifies the physician of her claim and the physician notifies his malpractice insurer. They show that apology laws have no effect on the number of claims made by the patient, before the formal process of litigation starts, for both sorts of physicians. What apology laws change, for non-surgeons, is the composition of cases that are dropped and fully-litigated once a claim has started. But then, if a claim was already made, plaintiff should be already aware of a potential malpractice case and the 'apologies as signal of malpractice' explanation cannot explain the change in the composition of cases for the same number of claims. Our theory can account for this. Before apology laws, some cases were dropped after the claim started because of sincere apologies; by making apologies non credible apology laws induce fewer cases to be dropped.

[37] McMichael et al. (2019), p.389.

[38] McMichael et al. (2019), p.386.

[39] McMichael et al. (2019), p.390: "The most natural course of action may be to repeal these laws, given their specific inability to achieve their stated purpose....If apology laws—even unintentionally—promote apologies that improve people's lives, they may generate a net social benefit with respect to patients' well-being, despite their failure to achieve their primary financial goal."

[40] McMichael et al. (2019), p.391.

for physicians when they know that an apology will be rejected ($c > 0$), apology laws also reduce the number of apologies in our model.

Miscarriages of justice

In this part we compare the miscarriages of justice in each game when the unintended effect holds by considering J's payoff. For the analysis we extend the payoffs of J in line with previous literature (see Feddersen and Pesendorfer (1998)). We assume that there is $\delta \in (0, 1)$ such that the jury earns a pay-off of: 0 whenever the award is correct, $-\delta$ whenever an innocent defendant must pay, and $-(1-\delta)$ whenever it exempts a liable defendant from paying. δ is the standard of proof: the critical probability of guilt at which the judge is indifferent between awarding positive indemnity payments, $m = M$, or not, $m = 0$. (The analysis above has implicitly assumed that $\delta = 0.5$.)

Let us first consider the miscarriages of justice in the equilibrium with settlement of Lemma 3. This is the equilibrium played before apology laws for cases where apology laws induce less settlement. This equilibrium prescribes d_s to settle with P_F and pay $m = M$ to P_L as the latter presents the apologies in court. On the other hand, d_i and d_n end up in court without apologies where they do not pay indemnity payments, $m = 0$. So the miscarriage of justice is d_n's exemption from paying damages with probability $(1 - \lambda)$. On the other hand, after the introduction of an apology law, players in these cases start playing one of the pooling equilibrium described in Lemma 2. Which equilibrium is played depends on whether $\delta \gtrless p_n + p_s$ (when $\delta = 0.5$, $p_i \gtrless p_n + p_s$. See Lemma 2).[41]

1. If $\delta > p_n + p_s$, then $m = 0$ at trial after apology laws. In these cases, d_n and d_s are exempted from paying damages with probability $(1 - \lambda)$. So after apology laws the miscarriage of justice increases by $(1 - \lambda)p_s$ $[(1 - \lambda)(p_s + p_n) - (1 - \lambda)p_n]$.
2. If $\delta < p_n + p_s$, then $m = M$ at trial after apology laws. Here d_i must pay damages with probability $(1 - \lambda)$, but also both guilty types. To compare the miscarriages of justice between games we need to consider whether J prefers to exempt d_n to pay before apology laws or to award indemnity payments against a d_i after apology laws, that is, $-(1-\delta)p_n \gtrless -\delta p_i$. In order to have higher payoffs for J under apology laws we need $\delta < p_n/(p_i + p_n)$. However, notice that we are considering cases where $\delta p_i > (1 - \delta)p_n$ (condition (iii) of Lemma 3), that implies $\delta > p_n/(p_i + p_n)$. Therefore, J's payoffs are higher before apology laws for any δ when the unintended effect arises.

Corollary 2: *(Miscarriages of justice) When the unintended effect holds (Proposition 1) apology laws cause more miscarriages of justice.*

McMichael et al. (2019) shows that apology laws increase indemnity payments for non-surgeons. Our results predict that this is because innocent physicians that before apology laws did not pay, are now found liable at trial after apology laws because of J's inability (in equilibrium) to observe sincere apologies (case $\delta < p_n + p_s$). See Proposition 1, the case of J's higher awards, for a detailed explanation.

[41] The decision of J in a pooling equilibrium is to award no indemnity payments, $m = 0$, with an expected payoff of $-(1 - \delta)(p_s + p_n)$ (as guilty defendants are exempted); or $m = M$, with an expected payoff of $-\delta p_i$ (as innocent defendants must pay). Therefore it compares $-\delta p_i \lesseqgtr -(1 - \delta)(p_s + p_n)$ or $\delta \lesseqgtr p_s + p_n$.

1.4.2 The intended effect of apology laws

In this section we show that apology laws may increase settlement, that is, the intended effect of legislators. Then, we show that for some cases they achieve settlement but at the cost of affecting the reliability of courts.

The reasons why apology laws may increase settlement are analogous to the ones provided in Proposition 1, but now the effects work through d_s's motives to apologize rather than on d_n's ones. First, *cheaper apologies*: apology laws may increase settlement by reducing the legal consequences of apologies for sincere defendants who did not to apologize when apologies were admissible. This is what motivated legislators to introduce apology laws in the first place. Second, *J's higher awards*: apology laws may increase settlement by increasing the cost of not apologizing. J's higher awards under apology laws encourage apologies from d_s more often than before these laws. Importantly, for the intended effect to arise, these mechanisms should encourage apologies from d_s but not from d_n; so there must be a high enough non-legal cost of apologies for d_n. This can be seen in Figure 1.2, panel b, where the intended effect holds for a high enough cost of apologizing. The following Proposition, which we prove in the appendix, formalizes these observations:

Proposition 2: *(The intended effect) Apology laws induce more settlement by reducing the legal cost of apologies whenever d_s has LR and*

a) (Cheaper apologies) $v_s > t - \lambda M + \frac{q_L c}{1-q_L} > 0$, $p_i > p_n + q_L p_s$; *or*
b) (J's higher awards) $v_s > t - M + \frac{q_L c}{1-q_L} > 0$, $p_n < p_i < p_n + q_L p_s$; *or*
c) (No insincere D) $p_n = 0$ *and* $v_s > t - \lambda M + \frac{q_L c}{1-q_L}$

In the case of *cheaper apologies*, in the equilibrium with settlement played after apology laws, P_L has evidence that proves the liability of d_s, i.e. d_s's apologies. However, she cannot use it at trial, and is therefore not awarded any damages. This highlights one of the arguments raised against apology laws: Public confidence in the courts could be adversely affected if a person who has admitted liability in an apology is found not liable (Cohen (2002)).[42]

Corollary 3: *Apology laws may induce settlement at the cost of affecting public confidence in courts.*

Consider the role play by c, the cost of a rejected apology. Chaudhry and Loewenstein (2019) proposes that apologies are effective because they involve giving up something valuable: being perceived as competent. Transgressors who apologize in situations in which competence is relevant suffer a negative impact on their perceived competence (Kim et al. (2006); Weiner et al. (1991)), and speakers are aware of this. In this sense, c is related to competence and its value will be dependant on the type of relationship considered. For example, physicians are reluctant to give apologies in part because it damages their reputation (Gallagher et al. (2003)). Notice that whenever $c \sim 0$, as we assume $t < \lambda M$ in (A), apologies cannot work, as $(1-q_L)t < (1-q_L)\lambda M$ (see Figure 1.2, panel b). This suggests that apology laws can only work on relationships where competence is damaged by apologizing or, more generally, when the non-legal cost of an apology is high enough).

[42]"Consider what will happen when your rule actually gets implemented - when an injurer has admitted his fault in an apology, but at trial invokes the 'apology exception' to have the apology excluded....And can you imagine how this will damage public respect for our courts? Suppose that, with the apology excluded, the plaintiff cannot prevail at trial. Can you see what will happen? The offender admitted his fault in the apology, but he'll go scott-free. What could more greatly tarnish the image of the court?" (Cohen (2002), p.859.)

1.4.3 Apology laws in the tort reform debate

Apology laws have been considered by previous scholars as a another tort reform designed to reduce litigation.[43] There is a long-standing debate about tort reform in US regarding whether there is too much or too little medical malpractice litigation. Those in favour of tort reform tend to be physicians, hospitals, and others connected with the health care industry who "..argue that too many medical malpractice claims are filed in the United States and that a large percentage of these claims are frivolous." (McMichael (2018), p.17). On the other side of the debate, the most salient advocates tend to be plaintiffs' attorneys who argue that too few medical malpractice claims are filed in the US. Because of the relevance of this debate we extend the model of Section 1.3 by allowing for frivolous claims, that is, we allow for the possibility that plaintiffs take innocent defendants to court. Our purpose is twofold: to analyse how frivolous claims impact the effects of apology laws, and to identify the law's beneficiaries.

Usually attorneys in civil claims receive a contingent fee: the attorney receives a percentage of any settlement or judgement but receives nothing if the case is lost. Moreover, attorneys sometimes have control over the settlement decision (see Spier (2007) for a discussion). We model frivolous claims by assuming that with probability $x \in (0,1)$ P has a contract which stipulates that the attorney decides whether to settle or take the case to trial because she earns a share of the award. We assume that the attorney will take innocent types to trial if she expects a positive payment. The probability x is independent of the P's type, that is, the probability of facing a plaintiff with this type of contract is $xq_F + xq_L = x$. Finally, whether P has this type of contract is P's private information, that is, neither D nor J can observe the contract, but the ex ante probability of x is common knowledge. We say that frivolous claims are possible whenever $x > 0$.

In Lemma 1 we showed that if there are no frivolous claims, $x = 0$, we cannot have an equilibrium where both guilty types settle in any game. Now we show that when frivolous claims are possible, apology laws may facilitate settlement through insincere apologies. That is, the next proposition shows first that there cannot be an equilibrium with this feature in Γ^a, and then that an equilibrium where this holds exists in Γ^{-a}. Therefore, after the introduction of these laws, settlements where plaintiffs accept insincere apologies are possible.

Proposition 3: *(Settlement with insincere defendants) When frivolous claims are possible, apology laws encourage settlement through the acceptance of insincere apologies whenever:*

i) $x < \frac{q_L(p_s+p_n)}{p_i-(1-q_L)(p_s+p_n)}$
ii) $x < \frac{(1-q_L)M-((1-q_L)t+q_Lc)}{(1-q_L)M-(1-q_L)(t-c)}$
iii) $M < v_F\frac{p_s}{p_s+p_n} + t$

PROOF. We show first that we cannot have an equilibrium with this feature in Γ^a. Towards a contradiction, assume that we have an equilibrium in which both guilty types settle. Note that if both guilty types apologize then $m = 0$ is J's best reply when no apologies are presented at trial.

[43]Tracing the development of apology laws, Arbel and Kaplan (2016), p.1200-01, explain that "tort reformers have . . . co-opt[ed] the rhetoric and discourse on apologies and the law—independently developed by ethicists, dispute resolution specialists, and legal theorists," and have thereby "found a path into the hearts of legislators and the public.". See also the discussion in McMichael (2018); McMichael et al. (2019)

Thus, frivolous claims against innocent types will not be filed as attorneys do not expect gains. Therefore, d_n will prefer to deviate and not apologize; by mimicking innocent defendants he pays nothing, just as in the benchmark model.

We now construct an equilibrium of Γ^{-a} which prescribes settlement with both guilty types: J's inability to distinguish between cases with and without apologies in Γ^{-a} permits frivolous claims in equilibrium. Now, by mimicking d_i, guilty types could end at trial. Consider a putative equilibrium where both guilty types apologize and settle, and attorneys take innocent defendants to court. We prove that under the specified conditions this equilibrium can exist.

(Judge) Let us consider J's best reply in her unique information set. The probability of having innocent defendants at trial is xp_i, that is, the probability that an innocent defendant faces a plaintiff with a frivolous claim contract. On the other hand, the probability of guilty defendants at trial is $(1-(1-x)q_F)(p_s+p_n) = (q_L + xq_F)(p_s+p_n)$: both guilty types apologize and these apologies are accepted just by P_F when she has no contract signed, that is, with probability $(1-x)$. Notice that P_L will always reject an apology, and a fraction x of forgivers and litigants have a contract that allows attorneys to decide whether to settle or not. Then, if positive payments are expected at trial attorneys will take these cases to trial. Finally, the award at trial will depend on whether the probability of frivolous claims is higher or lower than cases with rejected apologies, in particular $m = M$ if $xp_i < (q_L + xq_F)(p_s + p_n)$ or $x < q_L(p_s+p_n)/(p_i - q_F(p_s+p_n))$. The upper bound on x in condition (i) implies that J's best reply is $m = M$.

(Attorney) By construction, the attorney would infer D's liability from the defendant's apology decision. Condition (i) implies that by taking innocent types to trial she receives a percentage of $\lambda M > 0$, so filing a frivolous claims is her best reply. Notice that against guilty defendants the attorney receives a share of the expected payoff M.

(Plaintiff) As both guilty types apologize in our putative equilibrium we need to consider the best reply of P_F without a contract when she receives an apology. The payoff from accepting an apology is $v_F \frac{p_s}{p_s+p_n} + t$, where she receives compensation t from both guilty defendants. However, she only receives the value of restoring the relationship, v_F, when she accepts an apology from a sincere defendant, which happens with probability $p_s/(p_s+p_n)$. The payoff of rejecting an apology is the award at trial of $m = M$. Condition (iii) implies that accepting an apology is P_F's best response. P_L always rejects an apology by (A).

(Defendants) d_i never apologizes by assumption. d_n compares the cost of apologizing with the return from not doing so, which depends on the trial award. The former cost is $(1-q_L)(1-x)t + (q_L + xq_F)(c+M)$: an apology is only accepted by a P_F without a contract, $(1-q_L)(1-x)$, and with the remaining probability he faces trial with an expected cost of M and the cost of a rejected apology, c. By not apologizing he pays the positive indemnity payments at trial, M. The upper bound on x in condition (ii), that lies in (0,1) if $c > 0$, implies that apologizing is the best reply. Finally, if type d_n prefers to apologize in equilibrium it must be that type d_s also does so, as d_s receives v_s by restoring the relationship with P_F, which makes her expected payoff from apologizing higher for this type.■

Proposition 3 shows that the prevalence of frivolous claim in the health care context, as argued by tort reform advocates, could potentially promote settlement through insincere apologies that benefit the health-care industry once apology laws are introduced. This resembles an argument made by apology laws detractors: apology laws will encourage the manipulation of naive injured parties (not to pursue money damages, not to speak to a lawyer and not to become educated about their legal rights) by sophisticated actors (insurers, hospitals, big companies) (see Cohen (2002); Arbel and Kaplan (2016)).[44]

On the other hand, Proposition 3 shows that apology laws should increase settlement if the conditions (i-iii) are met for most medical malpractice cases. However, as this conflicts with the available evidence, this suggests that some of the conditions are not empirically relevant. Let us consider each condition. If $c \sim 0$, condition (ii) holds for any values of x. The restrictive conditions are (i) and (iii). Condition (iii) implies that some plaintiffs accept a mixture of sincere and insincere apologies. Apologies should be credible enough in order for apology laws to induce more settlement by this mechanism. Condition (i) sets an upper bound on x, that is relevant from J's perspective. When this condition is met, few innocent types end in court as a result of frivolous claims. Then, J awards positive indemnity payments at trial which in turn motivate attorneys to file frivolous claims against innocent defendants. Notice that there are no lower bound conditions for frivolous claims. This implies that apology laws could paradoxically favour pro-tort reform parties, who argue that too many frivolous claims are filed, when frivolous claims are infrequent.

1.5 Discussion

Modelling apologies: excuses vs admissions. Ho (2012) provides a costly signaling model of apologies in a principal-agent framework. He introduces two types of agents, high and low in terms of future gains for the principal, what he call the *disposition* of the agent. In addition, he introduces another component for the agent's type referred as the agent's *situation*. The agent knows both dimensions of her type, the principal knows neither. He connects these types to psychologists' Attribution Theory, which predicts that in settings where an outcome is observed, e.g. Amy is late for a meeting, and there could be two possible causes: (1) a dispositional cause, Amy is lazy and inconsiderate; or (2) a situational cause, Amy was held up by unexpected traffic; individuals attribute too much blame to the disposition of the actor and not enough to the situation. In his model, "An apology is the agent's attempt to shift the principal's attribution of a bad outcome from the agent's disposition to the agent's situation." He shows that in any separating equilibrium it is the high-disposition type that apologizes, "in our model, an effective apology shifts the principal's attribution of the cause of the bad outcome from the agent's disposition to the external situation." We call this *apologies as excuses*, as an accepted apology implies that the wrongdoer is innocent from the damaged party's point of view.[45]

Now notice that we model separately the liability of the defendant and her valuation of the

[44] Proposition 3 shows that some plaintiffs are accepting apologies from insincere defendants with positive probability in equilibrium under apology laws. If we consider that the compensation included in the apology in these cases, t, is similar to the one included when just sincere defendants apologize−before apology laws, then this result relates to another argument against apology laws: "Apologies encouraged by such legislation might create an emotional vulnerability in some plaintiffs who may accept settlements that are inappropriately low." (British Columbia Ministry of Attorney General (2006), p.4).

[45] Ho (2012): "Some argue that an excuse is not the same as an apology. In fact, the word apology derives from the Greek word apologos or story, and came into usage to describe an account used to excuse a transgression (Tavuchis 1991). In more common contemporary usage, the Merriam-Webster Dictionary (10th ed., s.v. "apology") defines apology firstly as '1a: a formal justification, b: excuse' and only secondly as '2: an admission of error...accompanied by an expression of regret.' ". See footnote 18 and Kleefeld (2007) for a discussion of the elements of an apology according to apology theorists.

relationship, i.e. disposition in terms of Ho (2012). This modelling choice allows for two interpretations of apologies that depend on which type sends the message. On the one hand, apologies as excuses, if just the innocent type apologizes in equilibrium, or apologies as admissions of fault, if only guilty types do. It would be interesting to allow for this distinction in Ho (2012)'s model and compare the results. For example, Ho (2012) shows that there cannot be a separating equilibrium if apologies are cheap talk in his model. On the contrary, in our litigation framework there can be a separating equilibrium when apologies are cheap talk. The equilibrium stated in Lemma 3 holds for $t = 0$ and $c = 0$ (this can be seen in Figure 1.2, panel b, when the cost of apologies is zero). In this equilibrium apologies are used as admissions of fault, as sincere guilty type alone apologizes. We conjecture that this type of separating equilibrium may exist when apologies are tendered to one audience as in Ho (2012), and apologies are costless.

Apologies from innocent types. One of the conditions that allows for settlement before apology laws is that innocent types prefer not to apologize. This makes insincere defendants better off by not apologizing as J awards no payments when it observes a case without apologies. By assuming that an innocent type does not apologize and that an apology necessarily involves a compensation, we are implicitly assuming that an apology is an admission of fault.[46] In some contexts innocent types may also prefer to take responsibility for what happened if they can thereby avoid a negative award at trial. Our results are robust to a small proportion of innocents being prepared to falsely apologize (as this will not change guilty defendants' incentives to apologize), and we argue below that the results still hold if all innocents are prepared to do so. There are two reasons why the strategies in Lemma 3 are robust to allowing all innocents to apologize. First, d_i always faces a lower expected trial payments whenever $\lambda > 0$, so her incentives to admit fault would generally be lower if trial uncovers the truth with some probability. Second, if we assume that d_i has $v_i > 0$, i.e. he also obtains a benefit from restoring the relationship with P_F, it seems plausible to assume that v_i is small. Notice that our history starts after the relationship is broken, P suspects liability and will take the case to trial if no apologies are tendered.

By introducing gains from restoring the relationship with the plaintiff for d_i, $v_i > 0$, the strategies of Lemma 3 form an equilibrium if the following condition is added:

$$(1 - q_L)(v_i - t) - q_L(c + M) < 0 \tag{2}$$

The left-hand side is the expected payoff from apologizing, that must be lower than the gains from remaining silent and paying nothing. Note that (2) holds whenever v_i is low enough, $v_i < \frac{q_L(c+M)}{(1-q_L)} + t$.

Finally, we argue how the results of Proposition 1 can be made robust to potential apologies from innocent types. Assume that both the premise of Lemma 3 and (2) hold. Then, there exists an equilibrium with settlement before apology laws (in Γ^a) where d_i and d_n end up in court and type d_s settles. Consider now that apology laws are introduced and litigants play the strategies of Lemma 3. By changing the evidence available at trial, apology laws may induce d_i to apologize by

[46] In the context of medical malpractice, the assumption that innocent physicians are reluctant to provide false confession has some support. When a random sample of 3.985 physicians were asked if saying "I'm sorry" would have helped to avoid litigation, the answer was 'No' in 81% of responses. "Among the verbal comments to this question, most physicians reported that they didn't say they were sorry because it wasn't their fault..." This results remains for a similar survey in years 2017 and 2019. (https://www.medscape.com/features/slideshow/public/malpractice-report-2015#page=25, accessed on 25/09/2020). Golann (2011) shows that 58% of claims are dropped by the plaintiff. As plaintiffs acquire more information in the course of a lawsuit, they often conclude that a claim is weaker than they had first thought. This may deter false admissions in the medical malpractice context.

increasing the expected award at trial. Notice that in the unique trial outcome of Γ^{-a}, d_i must pay at trial 0 (if J awards $m = 0$) or $(1-\lambda)M$ (if J awards $m = M$). If d_i prefers to remain silent after the introduction of apology laws, the same arguments as in Proposition 1 can be made: apology laws will induce apologies from d_n which preclude settlement. We state sufficient conditions for d_i to prefer to not apologize after the introduction of apology laws in the parameters where there is settlement in Γ^a. For the case where $m = 0$, conditions $(1 - q_L)(v_i - t) - q_L c < 0$ and $\lambda > q_L$ are required. For the case where $m = M$, conditions $(1-q_L)(v_i-t)-q_L(c+(1-\lambda)M) < (1-\lambda)M$ and $\lambda > 1/(2-q_L)$. If these conditions hold then apology laws induce less settlement by the arguments made in Proposition 1.[47]

1.6 Conclusion

We have provided a novel explanation for why apology laws fail based on the valuation of restoring the affected relationship. We showed that apology laws may induce insincere apologies by two mechanisms which in turn reduce settlement in the context of highly-valuable affected relationships. This explanation is consistent with the available evidence for US medical malpractice. We also illustrate two other negative consequences of apology laws when the conditions under which these law decrease settlement are met. First, apology laws result in more miscarriages of justice. Second, as no apologies are tendered after the introduction of an apology law, there is less communication between parties. These results call for a better design of these laws, at least in the medical malpractice context where there is evidence that these laws do not promote settlement. On the other hand, our results also provide conditions under which apology laws may work. We believe these results may be useful for policymakers in terms of understanding how the apologies dilemma works in the context of litigation.

Outside the context of litigation, we believe that our modelling strategy which allows the interpretation of apologies as excuses and admissions, deserves attention for readers interested in apologies. Finally, our results could also inform readers interested in dispute resolution mechanisms. We discuss two potential applications. First, making apologies or some private information endogenously costly could serve to restore valuable broken economic relationships. For example, the model can be cast in terms of a history of customer dissatisfaction with the service provided, and that the firm wants to encourage direct agreements or reconciliation (settlement). In such contexts, the preservation of a long term relationship may benefit the firm. The admissibility of apologies in courts can then be translated into the verifiability of apologies by the firm or a third party. We conjecture that the design of a communication system between the customer and the employee where apologies can be verified by the firm if tendered, as in 'this communication may be recorded' could benefit a firm which seeks long-term relationships.[48] Our model may also provide useful guidance is in the context of transitional justice: in particular, the design of optimal truth and reconciliation

[47] However, if these conditions are not met then Γ^{-a} might have no equilibrium with settlement. For example, assume that apology laws induce apologies from d_i in the space of parameters of Lemma 3. Then, every type apologizes (recall that when litigants play the strategies of Lemma 3, apology laws induce apologies from d_n). Therefore, in order to construct an equilibrium with settlement, P must accept these apologies. Moreover, we need to define reasonable off-the-path beliefs for the plaintiff when no apologies are tendered, and refinements to select this equilibrium over the pooling one.

[48] In 2018 Uber introduced UberChat where customers can communicate directly with drivers (https://eng.uber.com/one-click-chat/, accessed on September 2020). Our results indicate that Uber should encourage drivers to apologize when they fail with frequent customers. A credible justice system may not be necessary. If the firm (or J) always prefer to award positive payments to the buyer to keep the customer, as long as insincere guilty sellers and innocent ones are not punished, separating strategies for sincere/non-sincere guilty sellers may exist (see Lemma 3).

commissions (TRCs). We can interpret the role of P in our model as a political motivated TRC that may grant amnesty (settlement) in exchange for truth (apologies). Our model may serve to explore whether allowing or not the TRC to present confessions in criminal courts (admissibility of apologies) could affect the truth and justice outcomes of the reconciliation process. In South Africa, information and evidence obtained by the TRC, by testimony or by subpoena, was not admissible in domestic courts.

1.7 Appendix

Characterization of equilibria in each game

Equilibrium a1 in Γ^a was presented as Lemma 3 in the text.

Lemma A: *Settlement in the admissible case (Γ^a)*

a1) Γ^a *has an equilibrium which prescribes type* d_s *alone to apologize, type* P_F *to accept an apology, and* J *to award damages of* 0 *at trial if no apology is presented and* M *if it is presented, if and only if*

i) $(1 - q_L)(v_s - t) - q_L c > (q_L - \lambda)M$,
ii) $(q_L - \lambda)M > -(1 - q_L)t - q_L c$, and
iii) $p_i > p_n$

a2) Γ^a *has an equilibrium which prescribes type* d_s *alone to apologize, type* P_F *to accept an apology, and* J *to award damages of* M *at trial irrespective of the presentation of an apology, if and only if*

i) $(1 - q_L)(v_s - t) - q_L c > -(1 - q_L)M$,
ii) $-(1 - q_L)M > -(1 - q_L)t - q_L c$, and
iii) $p_i < p_n$

These are the unique equilibria of Γ^a *with settlement*

PROOF: (Uniqueness) By Lemma 1, we just need to consider equilibria in which d_s apologizes and P_F accepts the apology. As d_i never apologizes by assumption, any case in which apologies are presented implies $m = M$. Then guilty D's apology decision will depend on J's award, m, when no apology is presented and trial does not reveal the true liability. So we have to consider two potential equilibria: $m = 0$ (equilibrium a1) and $m = M$ (equilibrium a2). Equilibrium a1 was proved in Lemma 3.

Proof of equilibrium a2): (Sufficiency) By construction, P would infer D's type from an apology; so type P_F cannot profitably deviate from accepting.
Moreover, by construction, the case would only go to trial without an apology if D's type is either d_i or d_n. Hence, as $p_n > p_i$, J would award M at trial when no apology is presented. On the other hand, it would award $m = M$ when observes an apology, as only d_s apologizes.
Consequently, type d_n loses M by not apologizing and $(1 - q_L)t + q_L(M + c)$ by apologizing (as P_L never accepts an apology). Hence, condition (ii) implies that he cannot profitably deviate. Type d_s cannot profitably deviate either: he loses $-(1 - q_L)(v_s - t) + q_L(M + c)$ by apologizing, lower than M, the payoff from not doing it, given condition (i).
(Necessity) Suppose that condition (i) does not hold. Then, type d_s can deviate to not apologizing as the cost of apologizing exceeds $-M$, the expected award he must pay at trial without apologies. This proves necessity of (i). By an analogous argument, if (ii) does not hold, then d_n would prefer to deviate to apologizing. Finally, assume the prescribed equilibrium with $p_n < p_i$ (condition (iii) does not hold). According to the putative equilibrium, the types that end at trial without apologies presented are d_i and d_n, hence, J would award 0 at trial. Equilibrium a1 already considered this case. ■

Lemma B: *Settlement in the inadmissible case (Γ^{-a})*

a1) Γ^{-a} *has an equilibrium which prescribes type* d_s *alone to apologize, type* P_F *to accept an apology, and* J *to award damages of* 0 *at trial, if and only if*

i) $(1-q_L)(v_s-t)-q_Lc > -(1-q_L)\lambda M$,
ii) $-(1-q_L)\lambda M > -(1-q_L)t-q_Lc$, and
iii) $p_i > p_n + q_L p_s$

a2) Γ^{-a} *has an equilibrium which prescribes type* d_s *alone to apologize, type* P_F *to accept an apology, and* J *to award damages of* M *at trial, if and only if*

i) $(1-q_L)(v_s-t)-q_Lc > -(1-q_L)M$,
ii) $-(1-q_L)M > -(1-q_L)t-q_Lc$, and
iii) $p_i < p_n + q_L p_s$

These are the unique equilibria of Γ^{-a} *with settlement*

PROOF: (Uniqueness) By Lemma 1, we just need to consider equilibria in which d_s apologizes and P_F accepts the apology. Then guilty D's apology decision will depend on J's award when trial does not reveal the true liability. So we have to consider two potential equilibria: $m = 0$ (equilibrium a1) and $m = M$ (equilibrium a2).

Let us start with a1. By construction, P would infer D's type from an apology; so type P_F cannot profitably deviate from accepting.
The cases that go to trial are between P_L and every type of D; and between P_F and types d_i and d_n (who do not apologize). Hence, J would award $m = 0$ at trial because $p_i > p_n + q_L p_s$ (which proves necessity of $p_i > p_n + q_L p_s$).
Consequently, types d_n and d_s would respectively lose $(1-q_L)t + q_L(\lambda M + c)$ and $(1-q_L)(t-v_s) + q_L(\lambda M + c)$ if they apologize. As $m = 0$ when trial does not reveal the true liability, type d_s prefers to apologize given $(1-q_L)(v_s-t)-q_Lc > -(1-q_L)\lambda M$ (this proves necessity of condition i). In the same sense, type d_n prefers to not apologize as $-(1-q_L)\lambda M > -(1-q_L)t-q_Lc$.

Consider now a2). By construction, P would infer D's type from an apology; so type P_F cannot profitably deviate from accepting. The cases that go to trial are between P_L and every type of D; and between P_F and types d_i and d_n (who do not apologize). Hence, J would award $m = M$ at trial because $p_i < p_n + q_L p_s$ (this proves necessity of $p_i < p_n + q_L p_s$).
Consequently, as F forgives, types d_n and d_s would respectively lose $(1-q_L)t + q_L(M + c)$ and $(1-q_L)(t-v_s) + q_L(M + c)$ if they apologize. Type d_n prefers to not apologize as by apologizing he loses a higher amount than M, the cost of not doing so (this proves necessity of condition ii). Type d_s prefers to apologize as $(1-q_L)(v_s-t)-q_L(c+M) > -M$ (this proves necessity of condition i). ■

Proof of Proposition 2

Based on the above characterization, let us start with part a) of proposition 2, cheaper apologies. Consider the equilibrium a1 of Γ^{-a}. Condition (iii), $p_i > p_n + q_L p_s$, precludes the existence of equilibrium a2 in Γ^a (as $p_i > p_n$). On the other hand, if d_s has LR equilibrium a1 of Γ^a cannot exist. Then whenever equilibrium a1 of Γ^{-a} exists and d_s has LR we have an equilibrium with settlement under apology laws, and just the pooling equilibria of Lemma 2 when apologies are admissible.

For part b), J's higher awards, consider the equilibrium a2 of Γ^{-a}. It has the same necessary conditions (i) and (ii) as equilibrium a2 in Γ^{a}. However, this equilibrium holds for more cases regarding condition (iii); it exists whenever $p_n < p_i < p_n + q_L p_s$ whereas a2 in Γ^{a} does not. Then whenever equilibrium a2 of Γ^{-a} exists, $p_i > p_n$, and d_s has LR, we have an equilibrium with settlement under apology laws, and just the pooling equilibria of Lemma 2 when apologies are admissible.

Finally, part c). If $d_n = 0$, an apology will always be accepted by P_F as every guilty defendant is sincere. We just need to consider the incentives of d_s to apologize that depends on the cost of apologies, $(1-q_L)(v_s-t)+q_L c$. Then, as apologies are cheaper under Γ^{-a}, d_s will apologize for more cases in the inadmissible case. Specifically, whenever $-(1-q_L)M > (1-q_L)(v_s-t)-q_L c > -(1-q_L)\lambda M$ he prefers to apologize in Γ^{-a} and not in Γ^{a}.$\blacksquare$

Cases where apology laws have no effects

Here we characterize environments where both games induce the same settlement/trial outcomes:

Proposition 0: *Apology laws have no effects on settlement/trial outcomes on the following cases:*

a) $p_i = 0$, or liability is common knowledge by litigants, or the trial uncovers the true liability of D with probability one ($\lambda = 1$).
b) there is just one type of plaintiff in the case,
c) $p_s = 0$.

PROOF. Notice first that what is relevant for settlement in any game is D's ex-ante comparison between a potential settlement by apologizing and trial without apologies. Then, a difference in outcomes between games must arise whenever the award at trial in any of the two information set of J in game Γ^{a}, trial with and without apologies, is different from the unique information set of J at Γ^{-a}. Otherwise, if the trial award is the same in both games, the incentives of the defendants are the same on both games.

(a) If $p_i = 0$, then as J knows that there are just guilty types in the case, it will award $m = M$ whenever a case reaches trial in any game. Then apologies' admission is not relevant and both games have the same outcomes. Consider now cases where P can observe whether D is innocent or liable. Then, given that there are no frivolous claims, that is, if no P takes an innocent D to trial, then admissibility of apologies does not have any effect. This is analogous to the case where $p_i = 0$, as J will always award $m = M$ at trial in both games. Finally, whenever $\lambda = 1$, i.e. trial always discovers the true liability, admissibility of evidence does not matter by the same argument.

(b) Assume first that $q_L = 1$, that is, there are no plaintiffs ready to forgive sincere apologies. As apologies are never accepted, and so never tendered, then both games induce the same outcome of no settlement. Second, assume $q_L = 0$: all plaintiffs would be ready to forgive a sincere apology. Lemma 1 applies in this environment, and so the only possible equilibrium with settlement is when just d_s apologizes. Then, assume we have an equilibrium with this feature in Γ^{a}. Notice that as there are only forgivers, no apology is rejected and there are no cases with apologies at trial. Then, as J just observes cases without apologies in Γ^{a}, it has the same information to decide as in Γ^{-a}.

Therefore, there cannot be a difference between games in terms of settlement, either settlement exists in both games for the same parameters or it does not exist.

(c) If $p_s = 0$, given (A), P_F will not accept an apology and thus no apology will be tendered in the first place.■

Chapter 2

Waiting for others: Long delays in social learning

"In analyzing properties of social learning, the literature has often focused on whether learning converges to the truth or not. This focus is legitimate for theorists, but it is seriously misleading. What is the difference between a slow convergence to the truth and a fast convergence to an error? From a welfare point of view and for many people, it is not clear."

— Christophe P. Chamley*, Rational Herds -Economcis Models of Social Learning*

2.1 Introduction

In the current COVID-19 pandemic countries need to decide whether and when to adopt urgent and far-reaching policy decisions in a context where available information is limited and there is high uncertainty about the effects of these policies. Policymakers face the trade-off of waiting to see what others are doing or act quickly based on their private information. Learning from the behaviour of others, or social learning, has been the subject of extensive research in the social sciences in the last decades. Social learning models have served as an explanation for economic phenomena such as imitation, bubbles and crashes in financial markets, entry to new markets and technology adoption. As with the diversity of applications of these ideas, social learning models differ widely in the observational structure of the environment. The most studied setting is the one where agents must decide in a specified period and observe the action of all predecessors. However, in several of the examples cited above the timing of the decision is crucial. A politician that delays social distancing measures may have caused unnecessary deaths; similarly the success of entrepreneurs is linked to selecting the right time to start new ventures.[1] In this paper we study a social learning environment where agents can delay their decision in order to observe others' actions. The central question that we ask is, when private information points to an action, do agents delay their decision in order to observe others' actions?; and if so then for how long?

This question is not new. For example, Chamley and Gale (1994),Chamley (2004) and Wang (2017) have addressed it. The prediction of previous work is that agents do not delay profitable

[1] Recent studies support the effectiveness of social distancing measures, suggesting that measures adopted by US states during the spring of 2020 reduced mobility by as much as 23%, potentially saving many lives (Malik et al. (2020); White and Hébert-Dufresne (2020); Siedner et al. (2020)). For how market timing skills are relevant for successful entrepreneurs see Gompers et al. (2010).

choices in social learning environments. For example, in settings with two agents, all the relevant decisions are taken in the first two periods. In the case that someone takes the risky choice, this must be done in the first period, after which the other agent may follow. Hence, we should observe an early resolution of decisions in social learning contexts. Our main contribution is to show that in social learning environments there exist equilibria with possibly substantial delays in the adoption of the risky choice. In contexts where agents do not receive information in the same period or some agents can decide before others, agents may delay for many, possibly infinite periods; and so we may observe the first adoption decision in periods far way from the start. Let us illustrate our point using the pandemic example. In this setting some countries were affected earlier than others. Earlier affected countries faced the decision to implement some risky and uncertain policy measure based solely on their private information, or they could wait in the hope of gathering more information about the policy from others' decisions. Consider the case where the risky policy is in fact the better choice to take and everyone has private information pointing to this. Existing social learning models that allow delays predict either that the policy is adopted without delay or is never adopted. Our results highlight that this is because previous models assume that every country can decide and receive information in the same period. When this symmetry is not imposed, we show that there are equilibria where countries end up adopting the risky policy after (possibly) long delays. This inefficiency of delaying a profitable choice, due to the social learning environment, may have substantial detrimental welfare consequences and echoes what Chamley refers to in the epigraph.

We consider a standard model of social learning in which two players learn about the state of nature, the payoff of a risky and irreversible policy choice, through private information and the observation of other's actions. In contrast to previous models, we introduce an asymmetry in the timing of when they can adopt the risky choice or receive their private information. A first mover, F, receives her private information in period one and can decide whether or not to adopt the policy from then on. A second mover, S, receives his private information in the second period and faces the same decision as F from period two onwards. They can wait in order to observe the other player's decision to learn more about the risky policy. In this setting, we say that an equilibrium exhibits *long delay*, or simply *delay*, if it is possible to observe the first adoption of the policy after period two. The arbitrary selection of period two in this definition is just to contrast with previous results. We characterize all equilibria with long delay. We show that in any equilibrium with delay, the second mover adopts the policy first, and then F may follow him depending on her private information. These equilibria are in mixed strategies.[2] If F has favourable information about the policy, she adopts it with positive probability in the first period, otherwise she waits for S. We say that in the first period she is *dynamically* indifferent between adopting or not because her trade-off is intertemporal: whether to adopt the policy immediately or wait for the information that S is going to provide. On the other hand, if S observes that F did not invest in period one, and also has favourable information about the policy, he becomes *statically* indifferent whether to adopt the policy from then on. S is not indifferent between adopting or waiting for F, as F will not reveal more information. His beliefs after seeing no adoption in period one makes the expected payoff of adopting the policy equal to the payoff of not adopting it and, as F just reveals information in period one, S becomes *statically* indifferent forever. This implies that S can use any sequence of probabilities from period two as long as he provides enough incentives for F to make her willing to delay a profitable choice. The different equilibria with delay differ precisely in the sequence of

[2] Apart from the equilibria with delay the model exhibits two pure strategy equilibria. In one of them, F reveals her information in the first period and then S has all the information available for his decision in period two. The other equilibrium is the reverse: F waits in the first period as S reveals his information in the second period, leaving the first mover with all the information available at period three.

probabilities used by S. For example, in one such equilibrium, S adopts the policy with positive probability just in some period, say period ten; in another, S adopts with the same probability in every period forever. Therefore, if F did not invest in period one, an outside observer may see the first adoption in period ten in the former equilibrium, whereas in the latter it can be observed in any period.

Our results on delay allow us to offer a social learning explanation for two applications: the timing of policy measure adoptions during the current pandemic, and the investments patterns in the venture capital industry. For the former, Adolph et al. (forthcoming) provide evidence on the differences in the timing of US-states' social distancing policy adoptions (such as school closures and stay-at-home orders among others).[3] They find that after controlling for differences in social, economic and political costs of implementing social distancing measures, states may be more likely to adopt social distancing policies when neighbouring states also act.[4] A state with no neighbors announcing a policy on a given day was 26% less likely to announce compared with a state with 50% or more of its neighbors adopting the policy. Taking context into account, states with no neighbouring states taking action delayed each policy by an average of 1 day. They argue this is relevant as one day delay of a fully effective response to the epidemic could result in an increase in future caseloads between 8% to 26% higher, all else equal.[5] For the case of Europe, many observers believe that Italy responded too late in the initially affected region.[6] As France and Germany reported their first cases 6 and 3 days before Italy respectively, our results provide at least a partial explanation for Italy's delays.[7] Apart from the Italian social and political idiosyncrasies that may have delayed the implementation of the first lock-down, the hope of waiting to observe other' decisions could also be part of the history.[8]

Our second application refers to the venture capital (VC) industry. There are two competing theories in the literature that explain the high volatility of investments documented in the VC industry. The "overreaction view", which argues that volatility is a symptom of overreaction by venture capitalists and entrepreneurs to perceived investment opportunities. A contrasting view is the "fundamentals view", where volatility of the industry stems from the inherent volatility of fundamentals. In trying to test which theory is more relevant, Gompers et al. (2008) shows that more industry-experienced VC firms are more likely to increase their funding in their industry when corresponding public market valuations are high, i.e. in "hot markets" (we explain this evidence in detail in Section 2.2).[9] We interpret the industry-experienced firm as F, the first mover (or the in-

[3]The other policy measures are: restaurant restrictions, non-essential business closures, and restrictions on gatherings. As the authors argue, in the absent of pre-existing immunity to the virus, available vaccine, or effective treatment, social distancing policies, which aim to reduce close contact among individuals, have emerged as the primary tool to mitigate the epidemic.

[4]Covariates include the number of confirmed cases of COVID-19 in the state, logged gross state product per capita, whether the state has a Republican governor, and population density.

[5]Estimates of the uncontrolled doubling-time of COVID-19 cases vary and are complicated by the slow rollout of effective testing in the US, but many studies find doubling times in the range of three days absent any intervention to nine days if social distancing measures are in place (see the references in (Adolph et al., forthcoming).

[6]See for example `https://www.nytimes.com/interactive/2020/04/05/world/europe/italy-coronavirus-lockdown-reopen.html` and `https://www.aa.com.tr/en/europe/covid-19-what-went-wrong-in-italy-and-spain/1797461`.

[7]The data with the first positive cases dates for each European country was taken from the ECDC, the European Center of Disease Prevention and Control (`www.ecdc.europa.eu`).

[8]For a recent work on political herding in the context of the pandemic see Herrera and Ordoñez (2020).

[9]Gompers et al. (2008) consider an investment to be the first time a VC firm invests in a particular company. Venture organizations typically raise new funds every three to five years in their sample.

cumbent) in our model, and S, the second mover, as the less-experienced.[10] Moreover, we associate the entry decision of more industry-experienced firms as delay: they wait until the market is hot.[11] As our F firm ends up investing late in any equilibrium with delay, our results can explain this investment timing pattern. We argue that industry-experienced firms are waiting for information that will be provided by less-experienced firms once the market heats up. Moreover, we show that the better informed is F the more likely one is to observe the timing of investments described, which is consistent with the interpretation of F as the industry-experienced and better informed firm. We note that previous social learning models cannot explain this evidence. While discrete time models do not feature delay, continuous time models that may have delay always feature the more informed player investing first. Next, Gompers et al. (2008) conjecture that their evidence is most consistent with investment patterns being driven by shifts in industry fundamentals rather than by "overreaction" or social learning because the investment performance of more experienced firms does not seem to suffer as a result. We argue that our equilibria can reconcile both explanations. First, we show that in any equilibrium with delay, whenever an investment with delay by F is observed, this decision is well informed and relies on more (or better) information than the early investment decision of S (recall that F invests after S does). This implies that Gompers et al. (2008)'s evidence of good outcomes for experienced firms should not be considered sufficient evidence to dismiss social learning as at least a partial explanation for the observed pattern (an argument that also has been made by Khanna and Mathews (2011)). Second, as explained above, in any equilibrium with delay S is statically indifferent about investing or not from period two onwards. Then, a small shift in industry fundamentals can easily break this indifference towards investing, from which the first mover's entry follows. This is consistent with the shifts in fundamentals explanation of volatility.

Finally, we consider two extensions of the benchmark model. First we analyse the same model but with a "monitoring" cost. That is, players must incur a fixed cost each period if they want to observe the other's action. In this setting, an equilibrium with delay analogous to the one described above exists under some conditions; and we show a novel phenomenon that arises in an equilibrium, which we call *paying for waiting*. We show that after some histories F ends up paying forever to observe S's action in the hope of seeing an adoption. However, if S has received bad news about the policy, she will never adopt it. We provide an example in which with an ex ante probability of 20%, F is trapped in this situation of paying forever. In our second extension, we show that delay may also occur when players decide how much information to acquire. Most existing models of social learning assume that agents are either endowed with a given quality of free information or can acquire a fixed quality after paying a fixed cost. The assumption of exogenously endowed information seems particularly inappropriate for market entry and technology development decisions.[12] In this environment we show the existence of an equilibrium with delay analogous to the one described above and thus consistent with the available evidence for VC firms' entry decisions. In this respect,

[10]As Gompers et al. (2008) argue, their findings on the importance of industry-specific rather than overall experience suggests that a critical part of VC investing is the network of industry contacts to identify good investment opportunities as well. Notice that F receives her signal before the other player (and as we argue later, may be better informed).

[11]Their results are obtained when public market signals are lagged one year, that is, most industry-experienced firms increase their activity one period after the public market signals becomes more favourable.

[12]Firms undertaking such investments, along with entities such as venture capitalists who provide financing, generally undertake costly studies of both market size and technical feasibility prior to making capital investments. Furthermore, being able to acquire only a fixed quality of information for a given cost rules out situations where a leader and follower may optimally prefer to acquire different qualities. This becomes important when agents have opportunities to free ride on each others' effort by delaying their entry decision. See the discussion in Khanna and Mathews (2011).

we compare our results with Khanna and Mathews (2011). They show in the same model but with exogenous timing of decisions that a herding and a non-herding equilibrium both exists, and that the herding equilibrium features better investments decisions and quality of aggregate information. We show that the herding equilibrium of Khanna and Mathews (2011) exists in our endogenous timing environment and that our equilibrium with delays results in even better decisions for late entry firms (who "herd" on early entries) in comparison to their herding equilibrium. We see this result as supporting the idea that good performance of more experienced firms is not enough to dismiss social learning as an explanation for the high volatility in the VC industry.

The paper is organized as follows. In Section 2.2 we review the literature and present the evidence for the VC industry. In Section 2.3 we introduce the game. In Section 2.4 we analyse the game with exogenous quality of information and relate our results to the applications described above. Finally, in Section 2.5 and 2.6 we present the extensions and conclude respectively. Some proofs are relegated to the appendix.

2.2 Literature review

Our work contributes to the social learning literature. This literature studies how agents learn from each other through the observation of others' actions. Economic models with these features can be categorized into two types, exogenous and endogenous timing models. In the former ones, players make their choices according to an exogenous order, they must act in their assigned slot. The exogenous timing models, the literature that follows Bikhchandani et al. (1992) and Banerjee (1992), introduced the concepts of herding behaviour and informational cascades. On the other hand, in the endogenous timing models, players choose whether and when to invest. Our works fits into this category. This literature started with Chamley and Gale (1994). The question addressed within this framework is how information is revealed through time, and the main results highlight the inefficiencies due to social learning. These can take two forms: i) *herding*: as in the exogenous timing models, at some point players stop providing valuable information for others as they follow early movers and disregard their own information, so nothing more can be learned after herding starts and; ii) *delays*, players may delay profitable risky choices just to observe what others will do. We focus on the latter inefficiency.

The most closely related models to this work are Chamley and Gale (1994) and Wang (2017). Both studies analyse the same model as the present one: the only difference being in the timing of decisions or the periods in which players receive their private information. In these papers each player receives one signal at the beginning of the game and can decide from then on.[13] Chamley and Gale (1994) analyse symmetric equilibria, Wang (2017) asymmetric ones. While the symmetric timing of movements and information may not be an unreasonable assumption in certain settings, in many contexts players receive their information or decide asynchronously. For example, financial

[13]Our results will hold if S can also decide in the first period but will receive the information in period two. This timing was studied by (Chari and Kehoe, 2004) where they argue that information arrives "slowly" in the economy: one signal in each period. Each period the signal is randomly distributed to one and only one agent among the set of investors who have not already received a signal and is privately observed by that agent. They mention, "We can imagine that agents are randomly drawn without replacement from the pool of agents and assigned a number designating the period in which each will receive a signal. Neither the names of the agents who will receive the signals nor the periods in which these agents will receive signals are observed, but the process for assigning names and periods is common knowledge." Another timing where our results hold is the alternating moves: F takes an action in odd periods, and S in even periods. There is a vast literature on alternating move games in industrial organization and coordination games (see Dutta (2012) and the papers cited there).

investment decisions by someone living in New York and someone in Tokyo are asynchronous for the simple reason that they are made in financial markets that operate in different time zones. Entry investment decisions which involve incumbents and entrants are also asynchronous. The current pandemic is another obvious setting in which our timing is reasonable.

Several other researchers have also used different endogenous timing models to investigate various herding and related issues. Chari and Kehoe (2004) use endogenous timing to generate herding in a continuous actions investment model. In Rosenberg et al. (2007); Murto and Välimäki (2011), players play a one-armed bandit and continuously receive private signals. Chamley (2004) addresses the non-existence of symmetric pure strategy equilibrium in Chamley and Gale (1994) by allowing players to have different beliefs drawn from a continuous distribution, and establishes the existence of multiple symmetric pure strategy equilibria. However, the result on delay is the same as in his original work. Levin and Peck (2008) also establishes the existence of symmetric pure strategy equilibrium by adding a second signal (an idiosyncratic cost of investment) to Chamley and Gale's model, but with the same results on delay as previous work. Zhang (1997) presents another endogenous timing model where players with different precision of signals choose their optimal timing to act in continuous time. His paper shows that the player with the highest precision invests first after a initial delay. Similar results are also found in Aoyagi (1998) and Aghamolla and Hashimoto (2020) in models with continuous time. The reason that a delay exists in these models is different from our results. It relies on the heterogeneity of private signals' precision and the symmetric strategies equilibrium analysed. Agents delay their action because they presume correctly that another agent with better information will decide before them. Moreover, agents with the same information invest in the same period and, if there exists an agent with high enough precision, delays disappear. In our model, using asymmetric strategies, delays occur because of the indifference of late movers after seeing the no investment decision of first movers. Moreover, in contrast with these studies, our equilibrium with delay can feature better informed players delaying more than less informed ones. Finally, to the best of our knowledge there is no work that studies the endogenous timing model with a cost of observing other's action.[14]

All the work mentioned above assumes exogenous information quality. However, making information quality depend on agents' effort choices has previously been shown to significantly change results from models where agents are endowed with a given quality of information.[15] In this sense, this article is also closely related to the literature on market entry and/or competition with private information. Early articles that consider the importance of private information for entry and investment decisions include Cukierman (1980) and Jovanovic (1981), but in their analyses there is no ability to learn from others' actions. Later articles have considered how the ability to learn from others' decisions may affect leader versus follower advantages (Gal-Or (1987); Hoppe (2000); Décamps and Mariotti (2004); Thijssen et al. (2006)), and the timing and rate of entry or technology adoption (Rob (1991); Hirokawa and Sasaki (2001); Choi (1997)). However, these articles do not consider endogenous information acquisition. Ridley (2008) considers a market entry model with endogenous information acquisition where firms may herd on early movers' location choices. The closest work to our endogenous signal acquisition setting is Khanna and Mathews (2011). We discuss this work in more detail in Section 2.5. None of the papers mentioned above allow for delays in entry decisions. Two recent studies Kirpalani and Madsen (2021); Aghamolla and Hashimoto (2020)

[14] For an analysis of the exogenous timing setting with this feature see (Song, 2016) and the papers cited there.

[15] For example Burguet and Vives (2000) show that whereas more precise public information is good when an agent has endowed private information, it may hurt when the agent has to expend increasing costs to acquire better-quality signals.

allow for endogenous investment timing and signal acquisition in continuous time. The former analyse the unique pure symmetric equilibrium, and focuses on information aggregation inefficiencies in large populations. The latter allows for information acquisition effort in continuous time; but their equilibria do not feature delay once players receive their signals.

2.2.1 Venture capital industry – evidence on investment patterns

Social learning models have been used by previous work to explain VC industry features (see for example Khanna and Mathews (2011); Kirpalani and Madsen (2021)).[16] Because we are also going to relate our results with the investment patterns in the VC industry, we describe the relevant evidence in this section. VC has been a central source of finance for commercializing radical innovations in the U.S. economy over the past several decades (Lerner and Kortum (2000); Samila and Sorenson (2011)), and thus has been receiving an increasing attention from scholars. One of the salient features of the VC industry is the high volatility of the investments made by VC firms in companies. These swings result in periods in which too many competing companies are funded, followed by ones in which not enough companies have access to capital. There are two competing theories that seek to explain this volatility. The "overreaction view", which argues that the volatility of the VC industry is a symptom of overreaction by venture capitalists and entrepreneurs to perceived investment opportunities (see, for instance, Gupta (2000)). A contrasting view is the "fundamentals view", where volatility of the industry stems from the inherent volatility of fundamentals. According to this view, fluctuation in VC investment activity is simply a response to changes in investment opportunities. For instance, there may be shocks to the investment opportunities of existing entrepreneurial firms or entry by new entrepreneurs, both of which increase the demand for capital.

Gompers et al. (2008) "takes a step towards distinguishing between the 'overreaction view' and the 'fundamentals view' by examining the responses of different classes of venture investors."[17] To answer this question they focus on the investment patterns of VC firms with high and low industry-experience. Their evidence can be divided into two components: (i) the timing pattern of each type of firm entry decision and, (ii) their respective performance. For the former they ask: Are the most experienced investors more likely to increase their investments when the market heats ups? "Our empirical results indicate that investment by the most experienced venture capital firms—notably, those with the most industry experience—are most responsive to public market signals of investment opportunities." (i.e "hot market"). For the latter they ask: How well do they do with these investments relative to less experienced venture capitalists? "Although the success rate for deals associated with a hot market is lower than that for deals associated with a cold market, the difference is small. Experienced venture capital firms perform slightly better in hot markets, while less experienced venture capital firms do somewhat worse." Then, their test of the competing theories about volatility is the following: "If we find that the most experienced investors are more likely to increase their investment levels when the market heats up, this would suggest that shifts in

[16] Two features of the VC industry make social learning models appropriate to this end. First, payoff are non-rival. Evidence from empirical analysis of venture capital contracts indicates that venture-backed startups typically accept capital from more than one firm during fundraising periods. Second, social learning is important. Lerner (1994), in his study of syndicated investing (i.e. multi-investor) by venture capitalists, argues that social learning is so important that venture capitalists actively seek investing partners for the purpose of validating their investment decisions: "Another venture capitalist's willingness to invest in a potentially promising firm may be an important factor in the lead venture capitalist's decision to invest" (Lerner (1994), p.16). This assertion is reinforced by the survey evidence of Gompers et al. (2020), who report that 77% of venture capitalists surveyed cite "complementary expertise" of other investors as an important factor in deciding to join a syndicated round with multiple investors.

[17] All the quotations in this section are taken from Gompers et al. (2008).

fundamentals are an important component of venture capital investing. This interpretation would find further support if there is also little degradation in their performance. If we observe instead that the least experienced venture capitalists are most likely to increase their investment activity during hot markets, this would lend more credibility to the view that overreaction is a more important cause of volatility in the venture capital industry." Finally, based on their evidence they conclude: "These findings suggest that an important component of volatility in venture capital investment activity is driven by volatility of fundamentals" We will relate our results with this evidence in Section 2.4.

2.3 The model

The game is played by two players, F and S, over an infinite number of periods, indexed by $t \in \{0, ..., \}$. (We will refer to F and S as she and he respectively.) In period 0, Nature chooses the state, $\theta \in \{G, B\}$, and two signals, s_F and s_S: We write μ_0 for the prior probability that $\theta = G$, and suppose that the two signals are independently and symmetrically distributed, conditional on the state. We write

$$q_i = Pr(s_i = g \mid \theta = G) = Pr(s_i = g \mid \theta = B) \quad \text{for} \quad i \in \{F, S\}$$

In each period $t \geq 1$, the "active" players in period t simultaneously choose whether to invest (I) or wait (W) after privately observing their signal and the public history at t, denoted h^t, which consists of the choices made by each active player in every period before t. We denote player i's action as $a_i \in A = \{I, W\}$: $i = F, S$. Player F (the first mover) is active in every period $t \geq 1$ until she invests (and leaves the game); player S (the second mover) is active in every period $t \geq 2$ until she invests (and leaves the game). Player i's private history at t, denoted h_i^t, consists of the public history at t and s_i: $i = F, S$. (By convention, h_S^1 is empty.)

The return to investing in any period $t \geq 1$ depends on the state of nature. If a player chooses $a_i = I$ she receives $R - c$ when the state is G(ood), and $-c$ when the state is B(ad). We assume that $c < R \leq 2$, and $c \leq 1$; and normalize the return to either waiting or not choosing an action to 0. (Recall that S cannot choose an action in period 1.) Hence, the yield of the investment is positive in the good state and negative in the bad state. Players are impatient and their payoffs are discounted by a common discount factor $\delta \in (0, 1)$. Player i's strategy is a sequence $\sigma_i = \{\sigma_{it}(h_i^t)\}_{t=1}^{\infty}$, where $\sigma_{it}(h_i^t)$ is a mapping from the private history of the game up to t, h_i^t, to $\triangle(\{I, W\})$, the set of probability distributions over A.

Since players have private signals, this is a game of incomplete information. Each player i's beliefs that $\theta = G$ after private history h_i^t, denoted by μ_i^t, must be specified as part of the equilibrium, where $\mu_i(h_i^t) \in [0, 1]$. A perfect Bayesian equilibrium (PBE) consists of a strategy σ_i and probability assessments $\{\mu_i(.)\}$ for every private history h_i^t such that (i) each player's strategy is a best response at every information set and (ii) the beliefs are consistent with Bayes' rule at every information set that is reached with positive probability.

2.4 Analysis

2.4.1 Preliminaries

We denote player i's expected payoff of investing by $\pi_i(I \mid \mu_i)$. Notice first that without the presence of the other player, a player invests if and only if her beliefs about $\theta = G$ after the signal, denoted by μ_i, are high enough. To see this, notice that the expected payoff of investing after receiving $\hat{s}_i$ is

$$\pi_i(I \mid \mu_i) = Pr(\theta = G|\hat{s}_i)(R-c) + Pr(\theta = B|\hat{s}_i)(-c) = \mu_i(R-c) + (1-\mu_i)(-c) = \mu_i R - c$$

Hence, a player will invest whenever $\mu_i > c/R$ ($\mu_i R - c > 0$), as she can always get zero by waiting.

Next, let us denote by μ_i^+ (resp. μ_i^-) the players' beliefs about G after receiving $s_i = g$ (resp. $s_i = b$). Using Bayes' rule

$$\mu_i^+ = \frac{Pr(\theta = G)Pr(s_i = g|\theta = G)}{Pr(\theta = G)Pr(s_i = g|\theta = G) + Pr(\theta = B)Pr(s_i = b|\theta = B)} = \frac{\mu_0 q_i}{\mu_0 q_i + (1-\mu_0)(1-q_i)}$$

Because of the applications discussed in the Introduction, we will be interested in cases where players have different information, that is, when $q_F \neq q_S$ (which implies $\mu_F^+ \neq \mu_S^+$ by the above expression). Let us denote by μ^{+-} and μ^{++} the beliefs that a player holds after she infers a good signal and a bad signal or two good signals, respectively, in the game. (Notice that these beliefs must be the same for both players, so we omit the subscript in the notation.) The relevant assumptions for a strategic analysis of waiting are:

Assumptions: For $i = F, S$:
(A1) $\pi_i(I \mid \mu_i^+) = \mu_i^+(R-c) + (1-\mu_i^+)(-c) > 0$
(A2) $\pi_i(I \mid \mu_0) = \mu_0(R-c) + (1-\mu_0)(-c) < 0$
(A3) $\pi_i(I \mid \mu^{+-}) = \mu^{+-}(R-c) + (1-\mu^{+-})(-c) < 0$

With these assumptions, without learning any other information, it is profitable for a player to invest if her private signal is good, (A1), and not to invest if she has no private signal, (A2). Note that (A2) implies that $\pi_i(I \mid \mu_i^-) = \mu_i^- R - c < 0$. If a player will not invest when she has no signal, she will definitely not invest when she has a bad signal (where μ_i^- denotes this belief). By assuming (A3), when players infer one good signal and one bad signal in the game, the best reply is to not invest. (A3) implies that both players care about the other player's signal when they receive a good signal. If (A3) does not hold, then one player must have a much more informative signal than the other, and so when the well informed player receives a good signal she will invest right away and the game ends.[18]

Given the above discussion, if a player receives a bad signal the best reply is to not invest even if she could infer a good signal from the other player's action (by (A3)). Therefore, the interesting case to analyse delay is when some player receives $s_i = g$. Consider that player i receives $s_i = g$. In this case, the player's decision problem will consist of investing in the first period without further information, with an expected payoff of $\pi_i(I \mid \mu_i^+)$, or waiting for j's decision to gather more

[18]Notice that when (A1) does not hold, no one will invest ever (the best reply is to wait), and thus there are no strategic interactions in the game. Note that (A1) requires a high enough q :

$$\mu_i^+ > c/R \Leftrightarrow q_i > \frac{c/R(1-\mu_0)}{c/R(1-\mu_0) + \mu_0(1-c/R)}$$

The case where (A2) do not hold is also simple. If $\pi_i(I \mid \mu_0) = \mu_0 R - c > 0$ and $s_F = g$, F will invest in $t = 1$ as there are no gains by waiting for the other's action. To see why, notice that the signals have similar precision by (A3) and if S receives $s_S = b$ and reveals it, F's beliefs will be $\mu^{+-} \sim \mu_0 > c/R$, so it is still better to invest. If $s_F = b$, and $\pi_F(I \mid \mu_F^-) < 0$, F will not invest in $t = 1$, S will infer $s_F = b$ at $t = 2$ and decides with all the information. Then, F will infer both signals at $t = 3$.

information. In the latter case, if she can infer that $s_j = g$ in a given period, the expected payoffs of investing after observing two good signals is $\pi_i(I \mid \mu^{++}) = \mu^{++}R - c > 0$, where

$$\mu^{++} = \frac{q_j\mu_i^+}{q_j\mu_i^+ + (1-q_j)(1-\mu_i^+)} = \frac{q_F q_S \mu_0}{q_F q_S \mu_0 + (1-q_F)(1-q_S)(1-\mu_0)}$$

On the other hand, if she can infer $s_j = b$, i will prefer to not invest by (A3). Then, if we assume that j will reveal information at say $t = 3$, the expected payoff of waiting in $t = 1$ until $t = 3$ to observe the other players' action is:

$$\begin{aligned}\delta^2[\pi_i(I \mid \mu^{++})Pr(s_j = g \mid s_i = g) + 0Pr(s_j = b \mid s_i = g)] = \\ \delta^2(\mu^{++}R - c)[\mu_i^+ q_j + (1-\mu_i^+)(1-q_j)],\end{aligned} \tag{1}$$

that is, the payoff of investing after having observed two signals, $\pi_i(I \mid \mu^{++})$, times the probability that j has a good signal conditional on having a good signal, $Pr(s_j = g \mid s_i = g)$. Using the independence of the private information, $Pr(s_j = g \mid s_i = g) = \mu_i^+ q_j + (1-\mu_i^+)(1-q_j)$. In this chapter we will analyse this decision in detail.

2.4.2 Pure strategy equilibria

We start by characterizing the two pure strategy equilibria of this game.

Proposition 1. *There are two pure strategy equilibria that differ in the index of the player for the following strategies:*

Player i reveals her signal in her first move: she invests if $s_i = g$, and waits if $s_i = b$. Player $j \neq$ waits until she infers i's signal from her decision. Then, she decides in accordance with both signals.

The equilibrium in which S plays as j exists for every δ. The one in which player F plays as j exists for $\delta > \delta^*$ where.

$$\delta^* = \sqrt{\frac{\pi_F(I \mid \mu_F^+)}{\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)}}$$

PROOF: As mentioned in the previous section, if player i receives signal b the dominant strategy is to not invest. Thus, we focus on the case where $s_i = g$.

Equilibrium 1: F plays as i and S as j. Consider first the decision of F given S's strategy and $s_F = g$. If she deviates to not invest, S will not invest no matter his signal as S infers $s_F = b$ from F's action. Then, F cannot infer S's signal by waiting. Given this, as she will take her decision conditioned only on her information, she will be better off by not delaying a profitable investment decision (because of the discount factor). Next, consider the best reply of S when he moves at $t = 2$. Given the strategy of F, player S can infer F's signal from her action, so he cannot profitably deviate. Notice that as S is not waiting because he starts playing at $t = 2$, these strategies are independent of δ.

Equilibrium 2: F plays as j and S as i. Consider first the decision of S given F's strategy and $s_S = g$. Similarly to the previous equilibrium, if S deviates to not invest at $t = 2$, F will not invest no matter her signal as she infers $s_S = b$ from S's action. Then, S cannot infer F's signal by waiting. Given this, as he will take his decision conditioned only on his information, he will be better off by not delaying a profitable investment decision (because of the discount factor). Next, consider the best reply of F. Given the strategy of S, player F will be able to infer S's signal if she waits until $t = 3$. Thus, we need to show that the expected payoff of waiting (see (1)) is higher than the expected payoff of investing at $t = 1$. We need to show that:

$$\delta^2 \pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g) > \pi_F(I \mid \mu_F^+)$$

Next, setting $\delta = 1$ and using the definition of μ^{++} we can show that

$$\begin{gathered}\pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g) - \pi_F(I \mid \mu_F^+) = \\ (\mu^{++}R - c)[\mu_F^+ q_S + (1-\mu_F^+)(1-q_S)] - (\mu_F^+ R - c) = \\ \left(\frac{q_F(1-q_S)\mu_0 + q_S(1-q_F)(1-\mu_0)}{q_F\mu_0 + (1-q_F)(1-\mu_0)}\right)[\mu^{+-}(c-R) + (1-\mu^{+-})c] > 0\end{gathered}$$

as $\mu^{+-}(c-R)+(1-\mu^{+-}) > 0$ by (A3). Then, if $\delta > \delta^*$, defined in the statement, F's best response is to wait until $t = 3$. Notice that if waiting is the best reply at $t = 1$ it also must be at $t = 2$: F has the same beliefs as in $t = 1$, and the information that is going to be available at $t = 3$ is less discounted at $t = 2$.

Off-the-equilibrium path beliefs: The only case to consider is when a player deviates to invest when the equilibrium prescribes the opposite. Notice that if a player deviates to invest, it exits the game and so the beliefs of the other player after the deviation cannot affect its decision.

Uniqueness: We show that the equilibria stated above are the only pure strategy ones that exist. Consider first an equilibrium where both players invest in the same period. This can only happen at $t > 1$. But then F will prefer to deviate to invest in the first period as she cannot infer S's signal if they invest simultaneously. Next, consider an equilibrium where after each receiving signal g neither player invests till the other player invests. This cannot be an equilibrium for the same reason as above: a player can deviate to invest in its first move as there are no gains to waiting and there is a positive expected payoff of investing. Thus, in any equilibrium, there must be a first investment if players receive signal g. Then, as the first investment must be made conditioned only on one good signal, because of the discount factor the first player to invest must do so as soon as possible. Hence, in any equilibrium, one player reveals its signals on its first move. This leaves us with the two equilibria stated. □

2.4.3 Characterizing long delays

We start by defining precisely what we mean by long delay.

Definition 1. *We say that an equilibrium exhibits long delay, or simply delay, if there exists an equilibrium path such that the first investment in the game is made in period $t > 2$.*

The arbitrary selection of period two in the above definition is to contrast with previous results in the literature in which if there is no investment in the first two periods then there is later investment. In this section we show that long delay is possible in equilibrium. The next Theorem characterizes equilibria with long delay. Part (A) provides a class of equilibria with long delay, and part (B) proves that these equilibria are the only ones with this feature. Let us provide some intuition for the Theorem before the formal proof.

There is a key property of the game showed (cf. Lemma 1 below) that, one the one hand, serves to prove the existence of an equilibrium and, on the other, rules out many potential mixed strategy equilibria. Suppose that we are in a history where i has a profitable expected payoff of investing. Then, consider an equilibrium which prescribes that player i waits at t in order to observe the action of j at $t+1$. We can then ask whether an equilibrium which prescribes player i to invest with positive probability after seeing that player j did not invest at t can exists? Trivially, if i has a non positive expected payoff of investing before t, after seeing that j has not invested at $t+1$ her beliefs would be even lower than before and thus she is not going to invest. However, when i has a positive expected payoff of investment at t, one may think that after seeing no investment she can still have high enough beliefs that makes her willing to invest at $t+1$. We show in Lemma 1 that this is not possible in equilibrium. The intuition for why it cannot happen is the following. In order to be worthwhile for player i to wait for more information when it is already profitable to invest, the information that j will reveal through her action in the next period must have the prospect of changing i's decision after observing it. Otherwise, if every action that j will take implies the same decision for i after she observes it, i.e. to invest in any case, i can anticipate this and can deviate to take the investment decision earlier instead of waiting. This property implies that if a player waits to see the other player's actions in equilibrium, after no investment, the player who waits will not reveal more information, i.e. will not invest with positive probability, absent more information. She will wait forever absent more news.

Next, in part B of the Theorem we argue that in any equilibrium with long delay, F must be the first player that reveals information, i.e. the first that invests with positive probability (but less than one), and must do it in the first period. Finally, part A of the Theorem asserts the existence of equilibria with long delay that have the features discussed. Notice that because F reveals information in the first period, for long delay to exist, F must reveal bad news (i.e. she must wait) in the first period, otherwise she leaves the game. So, how long can delay be in equilibrium? The only possibility is that after F reveals bad news, S keeps revealing information. But how can player S be indifferent between investing and waiting after receiving a good signal, if F did not invests in the first period and will not reveal information in the future by Lemma 1? The only possibility is that S becomes "statically" indifferent between investing or waiting from period two onwards. That is, given the beliefs about the good state after observing no investment by F, the expected payoff of investing in a given period must be exactly zero, i.e. the payoff of waiting. Notice that this is different from F's indifference in the first period, as F is indifferent between a positive payoff of investing in period one and waiting to obtain more information. We call these *static* and *dynamic* indifference conditions respectively. As it will become clear in the proof below, the *static* indifference of player S after this history allows for different combinations of probabilities used in equilibrium by S over the remaining periods, as long as these probabilities satisfy the indifference condition of player F.

Theorem 1. *A) (Equilibria with long delay) Take any $T \geq 3$. If δ is high enough then there exists an equilibrium in which the first investment is made with positive probability until T, and that is described by the following strategies:*

Both players choose to wait forever in case they receive $s_i = b$. If F receives $s_F = g$, with probability z_F she invests in $t = 1$, and with probability $1 - z_F$ she waits. In period $t \geq 2$ if she has not invested in period 1 and observes that S has invested in period $t-1$, she invests; otherwise she decides to wait. If S receives $s_S = g$, for any $t \geq 2$ where he observes that F has invested, he *invests; otherwise he invests (resp. waits) with probability $z_S(t) \geq 0$ (resp. $(1 - z_S(t)) \geq 0$) for every $t \in [2, T]$ and $z_S(t) = 0$ for $t > T$. The probabilities used in equilibrium satisfy:*

$$z_F = \frac{\pi_S(I \mid \mu_S^+)}{\pi_S(I \mid \mu^{++})Pr(s_F = g \mid s_S = g)} \quad \textit{and} \tag{2}$$

$$\delta^2 z_S(2) + \sum_{n=3}^{T} \left[\delta^n z_S(n) \prod_{i=2}^{n-1} (1 - z_S(i)) \right] = \frac{\pi_F(I \mid \mu_F^+)}{\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)} \tag{3}$$

B) (Uniqueness) These are the only equilibria with long delay.

PROOF: Let us first prove the following Lemma that will be useful for proving part A and part B.

Lemma 1: *Suppose we are in a history at $t - 1$ or t where player i has received a good signal and the expected payoff of investing is positive. Also, assume that an equilibrium prescribes that player j invests with probability $z_j > 0$ at time t in the case she has received a good signal and no one has invested before. Then, at $t + 1$, after seeing no investment at t, i will not invest in any period after t unless j does so.*

PROOF: See the appendix.$\diamondsuit$

(Part (A)). As mentioned before, if a player receives $s_i = b$ the dominant strategy is to not invest in every period (by (A2) and (A3)). Let us consider the best reply of the each player in each period when they receive $s_i = g$. Below we show that the strategies in the statement form an equilibrium with long delay. Notice that this equilibrium is in mixed strategies. We first show the best reply of F and S for every t. Then, we characterise each player's probabilities used in equilibrium, and finally we discuss the off-the-path beliefs.

F's indifference condition, period 1:

Consider first the expected payoff of F in $t = 1$ given the strategy of player S. When she receives $s_F = g$ she mixes, thus she must be indifferent between investing in $t = 1$, with an expected payoffs of $\pi_F(I \mid \mu_F^+)$, and the expected payoff of waiting for more information. In equilibrium the following expression must be satisfied:

$$\begin{aligned}
\pi_F(I \mid \mu_F^+) &= \{\delta^2 z_S(2) + \delta^3[1 - z_S(2)]z_S(3) + \delta^4[1 - z_S(2)][1 - z_S(3)]z_S(4) + ... \\
&\quad ... + \delta^T[1 - z_S(2)]...[1 - z_S(T-1)]z_S(T)\}\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g) \\
&= \{\delta^2 z_S(2) + \sum_{n=3}^{T} \left[(\delta^n z_S(n)) \prod_{i=2}^{n-1} (1 - z_S(i)) \right] \}\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g),
\end{aligned} \tag{4}$$

where the expected payoff of waiting at $t = 1$ (right hand side of (4)) is the same as in (1) with the difference that now, as S uses investments probabilities ($z_S(t)$) if she received a good signal, F needs to consider the possibility of observing S's investment in any period where S uses these positive probabilities of investment. For example, the probability of receiving the two-signal-payoff, $\pi_F(I \mid \mu^{++})$, in $t = 4$ is $z_S(3)(1 - z_S(2))Pr(s_S = g \mid s_F = g)$: the probability that S does not invest in period 2, $(1 - z_S(2))$, times the probability the S invests in period 3, $z_S(3)$, times the probability that $s_S = g$ given $s_F = g$. We have already shown in the proof of Proposition 1 that,

$$\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g) > \pi_F(I \mid \mu_F^+)$$

Hence, the equilibrium prescribes that S uses probabilities $z(t) > 0$ (for $t \in [2, T]$) that satisfy:

$$\delta^2 z_S(2) + \sum_{n=3}^{T} \left[(\delta^n z_S(n)) \prod_{i=2}^{n-1} (1 - z_S(i)) \right] = \frac{\pi_F(I \mid \mu^+)}{\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)} < 1$$

Notice that (4) implicitly assumes that when S does not invest in a given t, F will prefer to wait. Below we show that this is the best response of F these histories.

F best reply, periods $t > 1$:

In period 2, as she is not already able to see the action of S, her beliefs must remain equal to those in period 1. This implies that at $t = 2$ she prefers to wait: if she was previously indifferent between investing or waiting for information, now she must strictly prefer to wait as the information will arrive sooner. Then, consider the best reply of F in period $t \geq 3$ when $s_F = g$. If she observes that S has invested in $t - 1$ the best reply is to invest in period t: she can infer $s_S = g$ from S's strategy, and thus investing yields an expected payoff of $\mu^{++}R - c > 0$ (the expected payoff of two good signals). If she observes no investment in t we proved in Lemma 1 that she will strictly prefer to wait if her indifference condition (4) holds. If F is indifferent at $t - 1$ and S may invest with positive probability in t, after seeing no investment from S, F prefers to wait in $t+1$. Notice that this is the case when F is player i, S is j and $t = 2$ in the statement of Lemma 1.

S's indifference condition, period 2:

Now consider the best reply of S when he receives $s_S = g$ in $t = 2$. If he observes that F has invested in $t = 1$ he will invest in period 2, with an expected payoff of $\pi_S(I \mid \mu^{++})$. Otherwise, he mixes and thus must be indifferent between investing in $t = 2$ and waiting for more information. Notice that after seeing no investment in period 1 his beliefs must be updated downwards from μ_S^+ to $\mu_S^2(z_F)$, where the latter denotes S's belief in period 2 and depends on z_F. This implies that the expected payoff of investing after bad news $\pi_S(I \mid \mu_S^2(z_F)) = \mu_S^2(z_F)R - c$ must be lower than $\pi_S(I \mid \mu_S^+)$. Hence, in equilibrium the following equation needs to be satisfied in $t = 2$:

$$\pi_S(I \mid \mu_S^2(z_F)) = \left\{ \delta^2 z_S(2) + \sum_{n=3}^{T} \left[(\delta^n z_S(n)) \prod_{i=2}^{n-1} (1 - z_S(i)) \right] \right\} \pi_S(I \mid \mu_S^2(z_F)), \quad (5)$$

where the right hand side of (5) is the expected payoff of waiting given the strategies prescribed by the equilibrium: F will not invest in period $t \geq 2$ if S has not invested, and S will invest with the sum of probabilities above described. Note that as F will not invest after period 1 unless S does so, S's beliefs remain unchanged from period 2 and so the expected payoff of investing after period 2 is the same as in period 2. Then, as the sum of probabilities used by S (curly brackets) must be lower than 1 by the arguments made for F's indifference condition, the only way to satisfy the above condition is when $\pi_S(I \mid \mu_S^2(z_F)) = 0$. That is

$$\pi_S(I \mid \mu_S^2(z_F)) = \mu_S^2(z_F)R - c = 0$$

Hence, the indifference condition of S can only be satisfied if $\mu_S^2(z_F) = c/R$. Moreover, as F does not provide more information after period 1 (i.e. F waits with probability one), $\mu_S^t(z_F) = \mu_S^2(z_F)$ for $t \geq 3$. Below we provide a strategy for F, z_F, that makes $\mu_S(z_F) \equiv \mu_S^t(z_F) = c/R$.

Probabilities used in equilibrium:

Consider the expression for $\mu_S(z_F)$, i.e. S's belief that $\theta = G$ given that he has received $s_S = g$ and F has not invested in period 1, $a_F^1 = W$:

$$\mu_S(z_F) = Pr[\theta = G|(s_S = g, a_F^1 = W)] = \frac{Pr[(s_S = g, a_F^1 = W)|\theta = G]Pr(\theta = G)}{Pr[(s_S = g, a_F^1 = W)|\theta = G]Pr(\theta = G) + Pr[(s_S = g, a_F^1 = W)|\theta = B]Pr(\theta = B)} \tag{6}$$

Notice that:

i) $Pr[(s_S = g, a_F^1 = W)|\theta] = Pr(s_S = g|\theta)Pr(a_F^1 = W|\theta)$
ii) $Pr(s_S = g|\theta = G) = q_S$
iii) $Pr(a_F^1 = W|\theta = G) = (1 - q_F) + q_F(1 - z_F) = 1 - q_F z_F$,

where (i) uses the fact that the signal received, s_S, and F's action are independent conditional on the state. And (iii) uses the fact that the probability of observing $a_F^1 = W$ given $\theta = G$ is the probability that $s_F = b$, $1 - q_F$, plus the probability that $s_F = g$ and F has not invested, $q_F(1 - z_F)$ (By a similar argument $Pr(a_F^1 = W|\theta = B) = 1 - (1 - q_F)z_F$). Substituting these expressions in (6) and equating it to c/R we obtain:

$$\mu_S(z_F) = \frac{q_S\mu_0[1 - q_F z_F]}{q_S\mu_0[1 - q_F z_F] + (1 - q_S)(1 - \mu_0)[1 - (1 - q_F)z_F]} = c/R,$$

and solving for z_F,

$$z_F = \frac{(c - R)q_S\mu_0 + c(1 - q_S)(1 - \mu_0)}{(c - R)q_F q_S\mu_0 + c(1 - q_F)(1 - q_S)(1 - \mu_0)} \tag{7}$$

Finally, rearranging terms we obtain (2). Notice that the right hand side of (2) is similar to the right hand side of (3), the only difference being that (2) considers the expected payoffs for S instead of F. In the proof of Proposition 1 we show that the right hand side of (3) lies in (0,1). By repeating the same arguments of the proof we can show that (2) lies in (0.1).

Finally we need to show that for high enough δ, we can always find a sequence of $\{z_S(t)\}_{t=2}^T$ that satisfies (4). Notice that if $\delta = 1$ (4) becomes

$$z_S(2) + (1 - z_S(2))z_S(3) + (1 - z_S(2))(1 - z_S(3))z_S(4) + ... = \frac{\pi_F(I \mid \mu_F^+)}{\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)} < 1,$$

where again the last inequality was proved in the proof of Proposition 1. Then, there are several degrees of freedom for choosing the values of $\{z_S(t)\}_{t=2}^T$. For example, suppose that S invests with positive probability just in period T ($z_S(t) = 0$ for $t < T$ and $z_S(T) > 0$). Then $z_S(T)$ must be equal to

$$z_S(T) = \frac{\pi_F(I \mid \mu^+)}{\delta^T \pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)}$$

that lies in (0,1) if

$$\delta > \left(\frac{\pi_F(I \mid \mu_F^+)}{\pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g)} \right)^{1/T}$$

In Corollary 1 below we provide an example where $T = \infty$ and $z_S(t) > 0$ for every $t > 1$.

Off the path beliefs: The only case to consider is when a player deviates to invest when the equilibrium prescribes the opposite. Notice that if a player deviates to invest, it exits the game and so the beliefs of the other player after the deviation cannot affect its decision.

(Part B). We first prove the following Lemma:

Lemma 2. *In any equilibrium that features long delay, F must be the first player to invest with positive probability, and she does so in the first period.*

PROOF: Notice first that because of discounting, the player who reveals information earlier in equilibrium must do so in her first move. Suppose this is not the case, that is, the first period of information revelation, i.e. an investment with positive probability, occurs in some period $t \geq 3$. Then, if this player is indifferent in period t between investing or waiting, before t, as beliefs are the same as in t because the other player has not revealed information, she must prefer investing. Finally, we argue that F must be the first player who invests with positive probability in an equilibrium with long delay. Suppose per contra that S is the player who reveals information earlier. As argued, this must happen in equilibrium in period 2 (his first move). However, by Lemma 1 we know that if F has received good news and S will reveal information (i.e. $z_S > 0$ if $s_S = g$), after bad news from S, F will not invest. Then notice that as F will not reveal information unless S does if he waits, S does not gain by waiting and thus he will invest in period $t = 2$ with probability one when he receives a good signal. Then, an equilibrium with long delay is not possible if S is the first player to invest with positive probability. $\diamondsuit$

From Lemma 1 and 2, we know that in any equilibrium with long delay, F must invest with positive probability in $t = 1$ (Lemma 2) and then if she does not invest, F is going to wait forever if S does not invest (Lemma 1). Part A characterizes the equilibria with these feature, which proves uniqueness. $\square$

Theorem 1 part A gives the strategies that characterize equilibria with long delay without specifying the probabilities used in equilibrium for player S. In order to contrast with previous work, let us construct an equilibrium with infinite periods of information revelation, i.e. $z_S(t) > 0$ for all t and $T = \infty$. This implies the possibility of observing the first investment far away from the beginning of the game. We consider the case where the information revealed is the same in each period, that is, $z_S(t) = z_S$ for all $t \geq 2$, as close solutions can be obtained. Notice then that,

$$\delta^2 z_S(2) + \sum_{n=3}^{T} \left[(\delta^n z_S(n)) \prod_{i=2}^{n-1} (1 - z_S(i)) \right] = \delta^2 z_S + \delta^3 z_S(1 - z_S) + \delta^4 z_S(1 - z_S)^2 + \ldots = \frac{\delta^2 z_S}{(1 - \delta(1 - z_S))}$$

Replacing this in (4) and solving for z_S we have the following result that, because it highly contrasts with previous results in the literature, we summarize it in the following Corollary:

Corollary 1. *If $\delta > \delta^*$ then there exists an equilibrium with long delay with the possibility of infinite periods of information revelation. In equilibrium, if F does not invest in period 1, S invests with the following probability in every period after receiving a good signal:*

$$z_S = \frac{(1-\delta)\pi_F(I \mid \mu_F^+)}{\delta^2 \pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g) - \delta \pi_F(I \mid \mu_F^+)} \quad \text{and}$$

$$\delta^* = \sqrt{\frac{\pi_F(I \mid \mu_F^+)}{\pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g)}}$$

z_S lies in (0,1) if $\delta > \delta^*$. Notice that in this equilibrium, if F did not invest in period 1 she keeps updating her beliefs for a possibly infinite number of periods (for example in histories where S has received a bad signal).

Finally, notice that the possibility of observing a history where both players invest after period 2 depends on z_F. If z_F is high then there is a high probability that F invests in period 1, and so investments after long delay are unlikely to be observed. Then, by taking the derivative of z_F with respect to q_F (see expression (7)), we have that

$$\frac{\partial z_F}{\partial q_F} < 0$$

To see why, notice that z_F is the value that makes S pessimistic enough after having received a good signal (see (7)). For fixed z_F, the higher q_F the more pessimistic S becomes after seeing no investment, as it is more likely that F has received a bad signal after no investment ($Pr(a_F^1 = W|\theta = G) = 1 - q_F z_F$). Then, a higher q_F allows for a lower z_F. Moreover, the higher is q_F the lower the right hand side of (3), which implies that the investment probabilities $z_S(t)$ that S uses in equilibrium could be lower (this result follows by taking the derivative of the right hand side of (3) respect of q_F). Finally, notice that in any equilibrium with long delay is S is the first investor. From these observations we have the following Corollary that will be useful when we apply our results.

Corollary 2. *An equilibrium with long delay where both players end up investing is more likely the more precise the signal of the first mover, q_F. Moreover, all of these equilibria prescribe S to invest first and F second.*

Differences with previous work

Discrete time models. Chamley and Gale (1994) and Wang (2017) analyse the same model but both players receive their signals in period one and can decide from then on. Wang (2017) analyse pure strategies. As in pure strategies the information is revealed through investment, the game trivially ends in at most two periods (see the strategies of Proposition 1). Chamley and Gale (1994) analyses symmetric mixed strategy equilibrium and shows that the game must end after at most two periods. Using an analogous argument as for Lemma 1, they show that if a player waits in one period to observe the other's action when investing is profitable, she will wait in the following periods if she observes bad news. Hence, why in the equilibria of Theorem 1 S invests with positive probability in period two after having observed that F did not invest? The proof of Lemma 1 does not apply for player S as she is not waiting. In period 1 S has not received his signal yet, and thus the premise $\mu_i^{t-1}R - c > 0$ (i.e. that investment is profitable at $t-1$) in the statement of Lemma 1 does not hold for S.

Notice that in the pure strategy equilibria of Proposition 1, if both players receive good signals both end up adopting the policy. In contrast, in all except one of the mixed strategy equilibria with long delay there is a positive probability that players wrongly believe that the state is bad despite both receiving good signals. This is the same inefficiency studied by (Chamley and Gale, 1994) in their symmetric mixed strategy equilibrium. The exception is the equilibrium of Corollary 1, in which S reveals information in every period, and therefore we must see an eventual adoption with probability one.

Continuous time models. Zhang (1997); Aoyagi (1998); Aghamolla and Hashimoto (2020) analyse endogenous timing social learning models in continuous time. In these papers the player with the highest precision invests first after a possible initial delay. The reason that a delay may occur in these models relies on the heterogeneity of private signals' precision and the symmetric strategies equilibria analysed. Players delay their action because they presume correctly that another player with better information will decide before them. Our results are different. Players use asymmetric strategies and there may be delay because of the indifference of late movers after seeing the no investment decision of first movers. Moreover, previous models have better informed players investing first (and if some player has high enough precision, delays disappear). In contrast with these studies, the equilibria with delay of Theorem 1 can feature better informed players delaying more than less informed ones, that is, q_F can be higher than q_S (and recall that S invests first in all the equilibria).

Applications

Pandemic: delays in the adoption of policy measures. We discussed in the Introduction that delays in the adoption of policy measures has been documented for U.S. states as well as for Europe countries. For the case of the U.S. Adolph et al. (forthcoming) find that after controlling for differences in social, economic and political costs of implementing social distancing measures, states may be more likely to adopt social distancing policies when neighbouring states also act. A state with no neighbors announcing a policy delayed policy measures. This points to social learning as a plausible explanation for the delay. However, previous model of social learning cannot explain delays, states must adopt the policy without delays or never adopt it. The equilibria of Theorem 1 are more consistent with this pattern of adoption, especially as states were not all affected at the same time and so there were timing asymmetries between them. In the case of Europe, although there are no similar studies as for U.S., many observers believe that Italy responded too late in the initially affected region. Corollary 2 offers a plausible explanation. As France and Germany (first movers) reported their first cases before Italy (second mover), and these countries were potentially more informed than Italy (high q_F), Italy could have been delaying the adoption of policy measures to observe others' decision.

VC industry's investment patterns. Now we discuss our results in light of the evidence of Gompers et al. (2008) described in Section 2.2.1. Let us consider first the timing pattern of each type of firm entry decision. They show that more industry-experienced venture capital firms delay their investment until the market is hot, whereas the less-experienced ones do not. We argue that our results for the equilibrium with delay can explain this timing pattern. We interpret player F in the model as the industry-experienced venture capitalist (the incumbent), as this player can decide or receives her information before S.[19] Then notice that Corollary 2 shows that in any equilib-

[19](Gompers et al., 2008, p. 2): "Of independent interest is our finding on the importance of industry-specific rather

rium with delay F is the one who invests late. Our results provide an explanation for this delay: industry-experienced firms are waiting for information that may be provided by less-experienced firms. Gompers et al. (2008) argue that a natural question to ask is why less experienced venture capitalists do not scale up their investment in a sector as much as more industry-experienced firms when investment opportunities appear to be more attractive. They propose the hypothesis that this pattern reflects a crowding out effect: the less experienced venture capitalists may wish to invest in the sector as well, but cannot get a "seat at the table" in the transactions being completed.[20] We provide an alternative explanation. Less experienced firms are first to invest in this industry, so their regressions do not show a significant correlation between increasing investments and previous year favourable public market signal. A simple empirical test of our explanation consists of regressing contemporaneous measures of hot markets with the investments of less-experienced firms, where we expect to see a positive correlation.

Second, Gompers et al. (2008) show that more experienced VC firms perform slightly better in hot markets, while less experienced venture capital firms do somewhat worse. They argue that this result is most consistent with investment patterns being driven by shifts in industry fundamentals rather than by "overreaction" or social learning, as they relate social learning with poor performances and inexperienced firms. (See Section 2.2.1 for the description of the two theories of volatility and Gompers et al. (2008)'s arguments) We argue that our equilibrium can reconcile both explanations of volatility. First, as mentioned above, in any equilibrium with delay is S (the less experienced firm) is the first to invest. Moreover, Corollary 2 implies that an equilibrium with delay in which both firms invest is more likely the more informed F is (the higher q_F). Then, in an equilibrium with delay, whenever an investment by an experienced firm (F) is observed, the investment decision is based on two signals with precision q_F and q_S. On the other hand, the less experienced firm (S) based its investment decision on just one signal q_S. Also notice that if F has received a bad signal she will not follow the investment of S, and so when late investment decisions from F are observed (the ones that Gompers et al. (2008) shows are empirically relevant), these must be well informed decisions. These results are consistent with the "overreaction" explanation of volatility in which the more experienced firm performed better but invested later by waiting for others. Finally, we argue that our equilibrium with delay is also consistent with the shift in fundamental explanation. Notice that in an equilibrium with delay, S is indifferent from period two onwards after seeing no entry in the first period. On the other hand, F's beliefs during long delay are pessimistic, as when S waits, F updates her beliefs downwards. Then if we consider a small shift in the fundamentals, for example a small increase in R, S's indifference can be easily break towards investment, but not F's decision. Therefore, we argue that this history of an equilibrium with delay is consistent with both explanations. If F does not enter in the first period (with probability z_F), then a small shift in fundamentals triggers the investment decision of S if she has acquired good news. Then, F will follow this decision through herding.

than overall experience. This result points to the importance of industry-specific human capital and suggests that a critical part of venture capital investing is the network of industry contacts to identify good investment opportunities as well as the know-how to manage and add value to these investments. These contacts and know-how come only from long-standing experience doing deals in an industry."

[20](Gompers et al., 2008, p. 15): "While ultimately the supply of transactions in a given sector may adjust to accommodate demand, in the short run there may be intense competition for transactions."

2.5 Extensions

2.5.1 Costly observation

In this section we introduce to the model of Section 2.3 a cost of monitoring the other player's action. We assume that players must pay an observational cost, denoted by $c_o > 0$, to see the other player's past decisions. (Notice that in the model of Section 2.3 observation of the other player's actions was free, $c_o = 0$.) This cost can be incurred more than once, that is, players decide each period whether they want to pay c_o to observe previous actions. We assume for simplicity that just player F must pay a monitoring cost to observe player S's action. We will show that when players play the strategies of Corollary 1 an interesting equilibrium outcome arises that we call *paying for waiting.*

Suppose that player F receives a good signal, and player S a bad one. Based on the equilibrium stated in Corollary 1, we will show in the next proposition that under some conditions if player F does not invest in period 1 then she will pay to observe S's action for ever. She pays in the hope of observing an investment that reveals good news. However, as S has received bad news, F will pay forever just for waiting. Hence, she incurs an infinite loss on this history.[21] As this extension has not been analysed by models with symmetric timing between players, i.e. where both players can invest from period 1, we also show that this outcome only arises by allowing for timing asymmetries.

Proposition 2. *In the game where there is a cost of observing S' action, $c_o > 0$, if $z_S > \frac{c_o}{\pi_F(I|\mu^{++})(1-q)}$ and $\delta < 0.5$, there exists an equilibrium which prescribes the strategies of Corollary 1 regarding investments, z_F is given by (2) and*

$$z_S = \frac{(1-\delta)\pi_F(I \mid \mu_F^+) + \delta^2 c_o Pr(s_S = b \mid s_F = g)}{\delta^2 \pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g) - \delta\pi_F(I \mid \mu_F^+) - \delta^3 c_o Pr(s_S = b \mid s_F = g)/(1-\delta)},$$

It also prescribes F to pay c_o in every period with positive probability; and, if both players can invest from $t = 1$ then c_o is paid by F at most once.

PROOF: See the appendix. □.

Notice that the expression for z_S in the statement of Proposition 2 is the same as the one in Corollary 1 for the case where $c_o = 0$. Let us now provide some intuition for the stated conditions. Notice that F faces a new trade-off in this extension. As S invests with the same probability (z_S) in every period in case $s_S = g$, then F must decide whether to pay c_o to observe S's action in the previous periods or wait to let the probability of seeing S's investment accumulate over time. For example, if she waits one more period instead of paying c_o right now, the probability of investment conditional on a good signal increases from z_S to $z_S+(1-z_S)z_s$. Then, in the appendix we show that if $\delta < 0.5$ F will prefer to pay right now rather than wait to pay. Finally, we need to consider for how many periods F will pay c_o when she has received a good signal and no one has invested yet. Notice that in these histories, after seeing no investment from S, her belief about the good state $\mu_F(z_S)$ decreases each period. Then, if S has received a bad signal, F's belief about the good state tends to zero (as S reveals information in every period on the equilibrium path). We show in the appendix that if $z_S > \frac{c_o}{\pi_F(I|\mu^{++})(1-q)}$ then F will pay forever in spite of having low beliefs about the good state.

[21] Notice that a similar outcome arises when S receives good news and she fails to invest given the mixed strategies used in equilibrium. However, in this case S will invest at some point and F will follow it.

Now we provide an example where F pays forever c_o with a high ex-ante probability. Consider the following parameters of the game: $R = 1$, $c = 0.5$, $\mu_0 = 0.3$, $q = 0.7$, $\delta = 0.4$ and $c_o = 0.005$. Then, from (2) and the expression for z_S we have that $z_F = 0.06, z_S = 0.3$. These values for δ and z_S imply that paying now to observe S's action is better than waiting one more period ($\delta < 0.5$), and that paying to observe is always profitable, ($z_S > \frac{c_o}{\pi_F(I|\mu^{++})(1-q)}$). Hence, F will pay c_o forever. In this example, the ex ante probability of a history where F receives a good signal and S a bad one is $0.2 = q(1-q)$. Then, the ex ante probability of a history where F pays forever is $q(1-q)(1-z_F) = 0.19$, $q(1-q)$ times the probability that F does not invest in $t = 1$, $(1-z_F)$.

2.5.2 Endogenous signal's precision acquisition

In this section we analyse the same model as Section 2.3 except that now players can collect private information by incurring a variable cost. The more they spend on information collection the better their ex ante information quality. We have two objectives. First, to show that the strategies of Corollary 1 form an equilibrium in this environment. Second, to relate our results to the evidence for investment timing patterns in the VC industry. For the latter purpose we compare our equilibrium with the work of Khanna and Mathews (2011) as they also relate their results to the evidence mentioned. To ease the comparison with Khanna and Mathews (2011) we adopt their modelling assumptions as them regarding information acquisition. We assume that each player can buy at most one signal and decides how much effort to expend to improve the precision, q_i, of its signal. In particular, a level of precision equal to $q_i \in [0.5, 1]$, where $Pr[s_i = g|\theta = G] = Pr[s_i = b|\theta = B] = q_i$, requires player i to pay $C(q_i)$ in effort cost. The cost function is common across the two players. We assume a quadratic cost function: $C(q) = \frac{1}{2\beta}(q_i - 0.5)^2$.[22] Notice that $C(.)$ is twice continuously differentiable and convex, $C(1/2) = C'(1/2) = 0$, and $C''(.) = 1/\beta$ (we assume $\beta \in (0, 0.5)$ to ensure interior solutions). Finally, without loss of generality, we assume as in Khanna and Mathews (2011): $u_0 = 0.5$, $c = 1$. (As in Section 2.3, $c < R \leq 2$).

Khanna and Mathews (2011) study the same model as above with two main differences. First, there is an exogenous timing of decisions. F must decide the precision of her signal and whether to enter a new market in the first period, whereas S takes the same decisions in the second period. In our model players may acquire the signal in any period that they are active (and also can invest in any period as in Section 2.3). Second, Khanna and Mathews (2011) introduce a follow-up investment stage. At $t = 3$, both players must decide simultaneously whether to make an additional investment that will affect their final payoff. If $\theta = G$, the return to either firm assuming it has paid both to enter and for the additional investment is $R \in [1, 2]$. A firm that chooses to enter but does not undertake the additional investment receives a scaled-down payoff of $(1-\alpha)R$, where $\alpha \in [0, 1]$. α simply measures the relative size of the two investments. If $\theta = B$, the firms receive nothing, irrespective of their investment decisions. Hence, our model is one in which $\alpha = 0$ and players can decide to delay their entry decision.

Notice first that in this extension of the benchmark model, players make two decisions in each period: whether to acquire a signal and whether to invest or not. The following Lemma simplifies the analysis:

Lemma 3. *In any equilibrium, if player i acquires a signal in some period then she must invest that period if the signal is good; if the signal is bad, absent another investment, i will wait forever.*

[22] The results of Khanna and Mathews (2011) are for generic convex functions and use the quadratic function for illustration.

PROOF: See the appendix. □

Notice that by Lemma 3 we only need to consider for how long players will delay their decision of acquiring a signal as, after acquiring it, players either invest if the signal is good or wait forever (absent other investment) if it is bad. We show first that the same equilibria of Khanna and Mathews (2011) (when $\alpha = 0$) exists in our endogenous timing framework. These equilibria are without delay, as in their model there is no option to delay decisions. Finally, we show the existence of an equilibrium with long delay and compare the properties of both equilibria that co-exist in our model.

Free-riding equilibrium without delay

Following Khanna and Mathews (2011) we define a *free-riding equilibrium* without delay as one in which one firm copies the other firm's decision without acquiring information.[23] More specifically, a free-riding equilibrium is defined by the following strategies. Firm i acquires its signal in its first move, invests if the signal is good, and otherwise waits. Firm j waits until it infers i's signal from its decision, and then copies i's decision without acquiring information: if i has not invested, j does not invest; if i has invested, j invests.

To ease exposition, let us focus on the equilibrium where F plays as i and S as j in the above definition of a free-riding equilibrium, but the other way round is also an equilibrium for a high enough δ (similar to Proposition 1). The existence of the free-riding equilibrium hinges the following conditions: (i) F's prefers to acquire information, (ii) and to do it without delay at $t = 1$, (iii) S will optimally gather no information when F does not invest, and (iv) S will optimally gather no information when F invests. The proof of existence that we relegate to the appendix follows closely the one in Khanna and Mathews (2011). Below we discuss each condition.

Condition (i) is necessary to make the analysis relevant, otherwise no firm would acquire information. Hence, let us consider the conditions on parameters that satisfy condition (i). Consider first the signal's precision that F acquires at $t = 1$ if players play the above strategies. We denote by q_F^* the chosen precision. It must solve the following problem:

$$\max_{q_F \in [\frac{1}{2},1]} \quad \frac{1}{2} q_F (R-1) + \frac{1}{2}(1-q_F)(-1) - C(q_F) \tag{8}$$

If the state is good, which F believes to be true with probability $\mu_0 = 1/2$, F receives a good signal with probability q_F and makes the investment, leading to the full payoff of $R-1$. If the state is bad, she receives a good signal with probability $1 - q_F$, after which she invest and obtains -1. This is reflected in the first two term of (8). Notice that after a bad signal F waits given the initial parameters considered ($\mu_0 R - c < 0$). Using the quadratic cost function and taking the derivative of (8) with respect to q_F, we find F's optimal precision value,

$$q_F^* = (\beta R + 1)/2 \quad , \tag{9}$$

where the second-order condition is satisfied given $C''(.) = 1/\beta > 0$. Thus, F will prefer to acquire information whenever expression (8) evaluated at q_F^* is positive. That is, when

[23] Khanna and Mathews (2011) refer to it as a *herding equilibrium*. We choose our label because the word herding has been used differently by other authors.

$$\frac{1}{2}q_F^*(R-1)+\frac{1}{2}(1-q_F^*)(-1)-\frac{1}{2\beta}(q_F^*-0.5)^2>0 \tag{10}$$

This expression depends on two parameters of the model, R and β. Notice that a high enough R is necessary to make the project profitable (if $R=1$ both remaining terms are nonpositive). Also we need a high enough β to make the information acquisition not too costly for a given R (if $\beta=0$, the cost is infinite, making $q_F^*=0.5$ and the left-hand nonpositive). We show in the appendix that the above inequality holds if $R>1.46$ and $\beta>(4-2R)/R^2$.

Let us now consider condition (ii). As S will not reveal any information unless F invests, F will not gain by delaying her decision of acquiring a signal in period $t=1$. Then, by Lemma 3, she will invest (resp. wait) if the signal is good (resp. bad). Next, consider condition (iii), the decision of S in period two after observing no investment from F. As F invests only if $s_F=g$, S can infer that $s_F=b$ whenever she observes no investment. Thus, μ_S, his posterior about the probability of the good state at this point (before deciding whether to acquire a signal), by Bayes' rule is

$$\mu_S=\frac{\frac{1}{2}Pr[s_F=b|\theta=B]}{\frac{1}{2}Pr[s_F=b|\theta=G]+\frac{1}{2}Pr[s_F=b|\theta=B]}=1-q_F^* \tag{11}$$

where S correctly infers the precision acquired by F at $t=1$, q_F^*. Then, if he prefers to acquire a signal, he will acquire a signal with precision q_S^* that solves the following problem:

$$\max_{q_S\in[\frac{1}{2},1]} \quad (1-q_F^*)q_S(R-1)+q_F^*(1-q_S)(-1)-C(q_S) \tag{12}$$

Expression (12) is similar to (8) except that now S believes the good state to be true with probability $1-q_F^*$ using (11). Taking the derivative with respect to q_S gives the first order condition, from which

$$q_S^*=\beta((1-q_F^*)(R-2)+1)+0.5$$

and the second order condition is satisfied given $C''(.)>0$. Notice that $q_S^*>0.5$ (as $q_F^*<1$). Then, S will prefer to not acquire a signal if the expected benefit of acquiring it is lower than zero (the payoff of not acquiring a signal). That is, if the following condition is satisfied (we show that this condition holds in the appendix):

$$(1-q_F^*)q_S^*(R-1)+q_F^*(1-q_S^*)(-1)-C(q_S^*)<0 \tag{13}$$

Let us consider now condition (iv). We need to show that S's expected payoff of investing without acquiring a signal is higher than the expected payoff of acquiring it. Using Bayes' Rule as in (11), after seeing F's investment S's posterior belief is $\mu_S=q_F^*$. Then, the expected payoff of investing without acquiring a signal is $q_F^*(R-1)-(1-q_F^*)$. On the other hand, the expected payoff of acquiring a signal, using a similar argument as for (12) is:

$$q_F^*q_S^*(R-1)+(1-q_F^*)(1-q_S^*)(-1)-C(q_S^*) \quad \text{where } q_S^*=\beta((q_F^*(R-2)+1)+0.5$$

In the appendix we show that the former payoff is higher than the latter.

Finally, we show that the free-riding equilibrium is the unique equilibrium in pure strategies when it exists.

Proposition 3. *The free-riding equilibrium exists for $R > 1.46$ and $\beta > (4 - 2R)/R^2$. When it exists, it is the only equilibrium in pure strategies.*

PROOF: See the appendix.□

Equilibrium with long delays

Now we provide an equilibrium with long delay for the endogenous signal acquisition game. We show the existence of an equilibrium in which S invests with positive probability in every period (analogous to the one shown in Corollary 1).[24] Consider the following strategies:

F's strategy. At $t = 1$ F mixes. With probability z_F she buys a signal with precision q_F^* (as defined in (9)). If she receives a good signal she invests, and otherwise waits. With probability $(1 - z_F)$ she waits to observe S's decision. If S does not invest in $t \geq 2$, then she keeps waiting. If S invests at some point after $t = 2$, then F invests without acquiring information.[25]

S's strategy. At $t = 2$, if S observes that F has invested then he invests without acquiring information. Otherwise, he mixes in every period. With probability z_S, the same in every period, he buys a signal with precision q_S^* (to be determined below) and invests (resp. waits) if the signal is good (resp. bad). With probability $(1 - z_S)$ he does not acquire a signal and waits.

We will show that these strategies form an equilibrium. We use the following notation: $\pi_i(I \mid s(q_k) = g)$ denotes player i's payoff of investing after having acquired/inferred $s(q_k) = g$ for $k \in \{i, j\}$, where $s(q_k)$ denotes that a signal with precision q_k has been either acquired by i ($k = i$), or inferred from the other player ($k = j$). Notice that given the strategies considered, when i can infer the signal of the other player she saves the cost of acquiring a signal.

Proposition 4. *If $\delta > \delta^*$, $R > 1.46$ and $\beta > (4 - 2R)/R^2$ there exists an equilibrium with long delay in which S, after seeing no investment from F in period one, acquires a signal with precision q_S^* with probability z_S in every period and invests if this signal is good. F acquires a signal with precision q_F^* with probability z_F in period one, and invests if this signal is good. When players do not acquire information, or receive bad signals after acquiring it, they wait. The probabilities and the precisions used in equilibrium are:*

$$z_F = \frac{-2(4 - 2R + 2(-1+\beta)\beta R^2 - \beta^2 R^3 - \sqrt{\beta^2 R^2 (R^2 + 8\beta(2 - 3R + R^2))})}{-4 + 2R + \beta(3 - 4\beta + 4\beta^2)R^2 + 2(1 - 2\beta)\beta^2 R^3 + \beta^3 R^4} \tag{14}$$

$$z_S = \frac{(1-\delta)\pi_F(I \mid s(q_F^*) = g)}{\frac{\delta^2}{2}\pi_F(I \mid s(q_S^*) = g) - \delta\pi_F(I \mid s(q_F^*) = g)} \tag{15}$$

$$q_F^* = \beta R/2 + 0.5, \quad q_S^* = \beta(\mu_S(R-2) + 1) + 0.5 \quad \text{and}$$

$$\delta^* = \sqrt{\frac{\pi_F(I \mid s(q_F^*) = g)}{\frac{1}{2}\pi_F(I \mid s(q_S^*) = g)}}$$

[24] We conjecture that we can prove a more general result as in Theorem 1, where S uses a sequence of probabilities that do not need to be equal in every period and do not need to be positive in every period. To this end we would need to extend Lemma 1 for this game.

[25] Notice that we do not specify F's choices on the history where she has acquired a bad signal in period one and then observes an investment from S. As this does not change players' incentives it does not matter for the analysis. For completeness see the end of the proof of Proposition 4 where we discuss this best reply.

PROOF: See the appendix.□.

EXAMPLE: Let us consider the following parameters of the model: $R = 1,5$, $\beta = 0.45$, $\delta = 0.8$. Then, using the above expressions: $z_F = 0.015$ and $z_S = 0.019$. The acquired signals' precisions are $q_F^* = 0.8375$ and $q_S^* = 0.8381$. Therefore, conditional on the state being good, the ex ante probability that neither F nor S enter in the first $t = 10$ and $t = 100$ periods are 0.85 and 0.25 respectively.

Notice that Proposition 4 is the analogous equilibrium of Corollary 1. Moreover, if we introduce asymmetries in the cost of acquiring information, $\beta_F \neq \beta_S$, then it is possible to show that as β_F increases z_F decreases (notice in (9) that q_F^* is increasing in β_F). Therefore, similar results as for Corollary 2 are obtained, that is, the better informed is F (or the better its capacity to acquire information) the more likely is a cluster of investments with delay in which the last investment is made by F. Notice then that our explanation for the timing of investment in the VC industry of Section 2.4.3 is robust to this extension.

2.5.3 Comparison with Khanna and Mathews (2011)

Khanna and Mathews (2011) show that when $\alpha = 0$ and the free-riding equilibrium exists, it is unique. This result carries over to our model as shown in Proposition 3. However, when $\alpha > 0$ apart from the free-riding equilibrium of Proposition 3 there is also a *non-free-riding equilibrium* (both in pure strategies). The non-free-riding equilibrium is defined as one in which S also acquires a signal, and then S enters only if F has entered and his own signal is good. Notice then that in a free-riding equilibrium F's second-stage decision depends only on its own signal (as S will not acquire information), whereas in a non-free-riding equilibrium F can use the information conveyed by S's entry decision for its own second-stage decision.[26] We conjecture that this equilibrium also exists in our endogenous timing framework if we introduce a follow-up investment decision, $\alpha > 0$, but we leave this extension for future work.[27]

Khanna and Mathews (2011) show that the free-riding equilibrium results in an improvement in aggregate information quality and decision quality compare to the non-free-riding equilibrium. They argue that their work adds a new perspective to Gompers et al. (2008)'s evidence as it can be caused by social learning (or overreaction) without implying worse investment performance on average. In other words, the evidence of good outcomes for experienced firms should not be considered sufficient evidence to dismiss social learning as at least a partial explanation for the observed pattern. Their results for the free-riding equilibrium regarding information quality and decision improvements are independent of follow-up investment, and thus they carry over to our free-riding equilibrium. In the free-riding equilibrium F acquires information with precision q_F^*. On the other hand, in our equilibrium with delay F collects information with the same precision as in the previous one, q_F^*, with probability z_F; and with probability $1 - z_F$ she decides to invest with an available information of higher quality, $q_S^* > q_F^*$. Therefore, in our equilibrium with long delay industry-experienced firms

[26] In particular, in a non-free-riding equilibrium, if it is optimal given q_F and q_S for S to stay out of the market after receiving a bad signal despite F's good signal, F will optimally make the second investment only if S enters.

[27] To provide some intuition for why, notice that in the non-free-riding equilibrium there also must be a first entry, and because this is in pure strategies, it must be done in the first period by F or in the second period by S. Consider the equilibrium where F acquires a signal first and enters only if this signal is good. Khanna and Mathews (2011) shows that when F waits after acquiring a signal in a non-free-riding equilibrium, S will not acquire more information and thus she will not invest either. Then, F will not gain by deviating to delay her decision of investment when she acquires a good signal. Therefore, the possibility of delay does not play any role in this equilibrium.

invest with higher quality of information in comparison to the free-riding equilibrium without delay of Khanna and Mathews (2011), from which better performances can be expected. We see this results as providing support for the social learning explanation.

2.6 Conclusion

We uncover a new class of equilibria with long delay in the canonical social learning setting with endogenous timing of decision. Previous models predict that the game must end quickly. In particular, in a game with two players, the relevant decisions are observed in the first two periods. We show that if one player receives the information or can decide before others, there are equilibria where the first adoption is made in periods far away from the start of the game. This long delay can have detrimental welfare consequences if the risky choice is the better one. We argue how our results can offer a social learning explanation for two applications: delays in the adoption of policy measures during the Covid-19 pandemic, and the timing of investments in the VC industry. We show that the equilibria with delay are robust to two extensions. First, we introduce a monitoring cost of observing others' actions. In this case we uncover a new phenomenon: a player can be trapped into paying forever to see the other player's decision in the hope of observing good news that will never come. Second, we show how our results are also robust to the case where players decide whether to acquire a signal and its precision. We believe our results on delay highlight a potential inefficiency of the social learning environment that can be relevant for any outsider who tries to encourage a risky adoption/entry, such as a scientific community promoting social distancing measures in the case of the pandemic, entrepreneurs trying to attract funds for their ventures.

Finally, we believe that our equilibria with delay can be easily extended to more players. To provide some intuition for why, assume the same game of Section 2.3 with the difference that there are more players receiving their signals in period 1 and 2. Assume that players F and S play the strategies defined in Theorem 1. Then, assume that the rest of the players wait unless they observe an investment from one player. Given these strategies, notice that players F and S will have the same incentives as before to reveal information, as the other players are just waiting for their actions. On the other hand, the other players, given the information that is going to be reveal by F and S, may prefer to wait by similar arguments as we state for F and S. Therefore, we conjecture that equilibria with long delay may exist.[28]

[28] There is another timing possibility for a game with more players that may generate more delay. Consider a model where after a player decides, some remaining players can observe this information earlier than others. For example, if we assume that players are connected in a network and the observational lag depends on the distance between them. Therefore, in this kind of setting after one of the players decides there will be an asymmetry in the timing between players analogous to the model of Section 2.3. Notice then that we can set the strategies of two remaining players with asymmetric timing as in Theorem 1, generating an additional delay after an initial delay.

2.7 Appendix

Proof of Lemma 1

The proof is the same if we consider whether player i is deciding to invest or wait at $t-1$ or at t, we prove the statement for the case where i is deciding at $t-1$. Suppose we are in a history where player i has received a positive signal, no one has invested before, and the expected payoff of investing is positive at $t-1$: $\pi_i^{t-1}(I \mid \mu^{t-1}) = \mu^{t-1}R - c > 0$. Also assume the equilibrium prescribes that player j invests with probability $z_j > 0$ at time t in the case she has received a positive signal and no one has invested before. Let $\mu_i^{t+1}(a_j^t)$ denote player i's posterior beliefs about $\theta = G$, if action a_j^t has been taken by j at t. Thus, denote by $\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t)]$ the payoff of investing for player i after having observed a_j^t. The undiscounted payoff of waiting at $t-1$ can be defined as follows:

$$\pi_i^{t-1}(W \mid \mu_i^{t-1}, z_j) \equiv \sum_{a_j^t \in \{I,W\}} Pr(a_j^t|\mu_i^{t-1}) Max\{\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t)]; \delta^2 \pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t), \sigma_j^{t+1}]\} \quad (D)$$

where $Pr(a_j^t|\mu_i^{t-1})$ denotes i's probability assessment that j will choose a_j^t at a history with beliefs μ_i^{t-1}, and $\pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t), \sigma_j^{t+1}]$ is i's undiscounted expected payoff of waiting at $t+1$ on history $\mu_i^{t+1}(a_j^t)$, given some strategy of j for the next periods, σ_j^{t+1}. In what follows we omit the waiting payoff's dependence on on the future strategy of j, z_j and σ_j^{t+1}.

With this notation in hand, notice first that by the law of iterated expectations,

$$\pi_i^{t-1}(I \mid \mu_i^{t-1}) = \sum_{a_j^t \in \{I,W\}} Pr(a_j^t|\mu_i^{t-1}) \pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t)] \quad (16)$$

That is, the payoff of investing at $t-1$ is equal to the undiscounted expected payoff of investing at $t+1$ given the beliefs holds at $t-1$. Next, let us define the option value of waiting to see j's action before making an investment decision as:

$$\delta^2 \pi_i^{t-1}(W \mid \mu_i^{t-1}) - \pi_i^{t-1}(I \mid \mu_i^{t-1}) \qquad (0\text{VW})$$

Notice that i is going to wait with positive probability at $t-1$ whenever OVW≥ 0. We show below that OVW≥ 0 if i's decision depends in a non trivial way on the value of a_j^t. Now we restate the statement of Lemma 1 in the following claim:

Claim: In order to OVW≥ 0 at $t-1$, the payoff of investing at $t+1$ must be lower than the expected payoff of waiting at $t+1$ after seeing that player j has not invested at t, that is

$$\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t = W)] < \delta^2 \pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t = W)]$$

Proof of Claim: Notice first that because of discounting, in order to be indifferent at $t-1$, OVW=0 (or to prefer to wait, OVW>0), it must be that

$$\pi_i^{t-1}(I \mid \mu_i^{t-1}) < \pi_i^{t-1}(W \mid \mu_i^{t-1})$$

We will show that for this to hold, the inequality of the statement is necessary. The proof follows by using the definition of $\pi_i^{t-1}(W \mid \mu_i^{t-1})$ in (D) and (16). Notice first that,

$$\pi_i^{t-1}(I \mid \mu^{t-1}) \leq \pi_i^{t-1}(W \mid \mu^{t-1})$$

as, by (16), $\pi_i^{t-1}(I \mid \mu_i^{t-1})$ is equal to one of the arguments in the max function of $\pi_i^{t-1}(W \mid \mu_i^{t-1})$'s definition (see D). Furthermore, it also follows from (16) that the above inequality is strict if and only if

$$\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t = x)] < \delta^2 \pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t = x)]$$

for some value of a_j^t, say x. To see why notice that otherwise, if $\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t = x)] > \delta^2 \pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t = x)]$ for every $x \in (I, W)$, then using (D)

$$\pi_i^{t-1}(W \mid \mu_i^{t-1}) = \sum_{a_j^t \in \{W,I\}} Pr(a_j^t|\mu_i^{t-1})\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t)] = \pi_i^{t-1}(I \mid \mu_i^{t-1})$$

where the last equality uses (16). But this implies that OVW<0 because of discounting, that is, it is strictly better to invest at $t-1$ rather than wait. Therefore, for OVW≥ 0, it is needed some action of j at t after which waiting further is preferred than investing at $t+1$. Finally, we show that this must be when $a_j^t = W$. Notice that $a_j^t = I$ can only be observed whenever $s_j = g$. Thus, if i observes an investment:

$$\pi_i^{t+1}[I \mid \mu_i^{t+1}(a_j^t = I)] = \mu^{++}R - c > \delta^2 \pi_i^{t+1}[W \mid \mu_i^{t+1}(a_j^t)] = 0$$

the last inequality follows from the fact that there is no more information to wait for once j has invested. Therefore, we have proved that for OVW$\geq$0 it is necessary that the inequality in the statement of the Claim holds. $\diamondsuit$

The above Claim implies that when there is a non negative option value of waiting at $t-1$, OVW$\geq$0, that includes the case where i is indifferent between investing or waiting at $t-1$ (OVW$=0$), i will invest with positive probability at $t+1$ only if j has invested on the previous period. Or similarly, i will not invest in any period after $t+1$ if j does not invest at some point, which proves Lemma 1. $\square$

Proof of Proposition 2

(Aymmetric timing game) We show an equilibrium where F and S play the strategies defined for Corollary 1 and in which F pays c_o forever. Notice first that as S does not need to pay to observe F decision ($c_o = 0$ for S) and F is playing the same strategy as in Theorem 1, then the same arguments that we used in Theorem 1 hold for S's best reply. Thus, let us focus on F's best reply. Assume we are in a history where F has received good news and she has not invested in $t = 1$. We first characterise the best reply of F after observing S's action. Then, we provide conditions for i) when c_o is paid, ii) for how many period F will pay it, and finally, iii) we provide the value of z_S that makes F indifferent in period 1.

F's best reply after observing a_S. We first prove the following Lemma that follows the same arguments as Lemma 1:

Lemma 1.B: *Suppose we are in a history in period t where F has received a good signal and the expected payoff of investing is positive. Also, assume that an equilibrium prescribes that S invests with probability $z_S > 0$ at time t when he has received a good signal and no one has invested before. Then, if F pays c_o to observe S's action, she will invest only if S has invested.*

PROOF: The argument is the same as in Lemma 1 where we just relabel some definitions. Notice that F will compare the expected payoff of investing without paying to observe S's action, $\pi_i^t(I \mid \mu_i^t)$, with the payoff of paying to observe. Define the expected payoff of paying to observe S's action as

$$\pi_i^t(P \mid \mu_i^t, z_j) = \pi_i^{t-1}(W \mid \mu_i^{t-1}, z_j) - c_o,$$

where $\pi_i^{t-1}(W \mid \mu_i^{t-1}, z_j)$ was defined in the proof of Lemma 1 (equation (D)). (Replace the time superscript of (D) by t whenever it says $t+1$, as now F's decision is made at t.) Notice that (16) holds. Then, set $\delta = 1$ in the proof of Lemma 1. Notice that in this case OVW is defined as

$$\pi_i^t(P \mid \mu_i^t) - \pi_i^t(I \mid \mu_i^t)$$

With these modifications, Lemma 1.B follows by repeating the same arguments inf the proof of Lemma 1. $\diamondsuit$

Then, notice that the expected payoff of observing S' actions is $z_S[\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)]$. She will observe an investment with probability $z_S Pr(s_S = g \mid s_F = g)$, after which she invests and receives $\pi_F(I \mid \mu^{++})$. By Lemma 1.B if she does not observe an investment, she will wait with a payoff of zero.

i) When c_o is paid. Notice that F faces a new trade-off. As S invests with the same probability, z_S, in every period when $s_S = g$, F must decide whether to pay c_o to observe S's action in the previous periods or wait more time to let the probability of seeing S's investment accumulate over time. For example, if she waits one more period instead of paying c_o right now, the probability of investment conditional on a good signal increases from z_S to $z_S + (1 - z_S)z_s$. If the following condition is satisfied then F prefers, in every t, to pay c_o in period t instead of waiting one more period to pay:

$$\begin{aligned} &z_S[\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)] - c_o \\ &\qquad > \delta[(z_S + (1 - z_S)z_S)(\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)] - c_o \end{aligned} \tag{17}$$

To see why, notice that the expression on the left of (17) is the expected payoff of paying right now to observe S's past actions. The expression on the right is the payoff of waiting one more period to pay c_o and have more information revealed. Whenever waiting one more period to observe the action is better (i.e. when (17) does not hold), then we cannot have an equilibrium where S reveals information in every period. Notice that if that were the case, in each period t F will prefer to wait one more period to let the probability of information revelation accumulate. Hence, in period 1 F will forecast this situation and will not be indifferent between investing and waiting forever, as she must discount the future observation infinitely (that is, (4) cannot be satisfied as the right hand side would be zero). Then, notice that condition (17) can be simplified to $z_S(1 - 2\delta + \delta z_S) > 0$. A sufficient condition for this to hold is $\delta < 0.5$. Finally, we show that if $\delta < 0.5$, F will also prefer to pay c_o at t instead of paying it in more distant periods. Compare F's expected payoff in period t of paying c_o at $t + n - 1$ or at $t + n$. As in (17), we can show that the former is higher than the latter, that is,

$$\begin{aligned} &\delta^{t+n-1}(z_S + ... + (1 - z_S)^{t+n-1}z_s)[\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)] - c_o \\ &\qquad > \delta^{t+n}(z_S + ... + (1 - z_S)^{t+n}z_s)(\pi_F(I \mid \mu^{++})Pr(s_S = g \mid s_F = g)] - c_o \end{aligned}$$

which can be simplified to

$$(1 + (1 - z_S) + ...(1 - z_S)^{t+n-1})(1 - \delta) > (1 - z_S)^{t+n}\delta$$

if $\delta < 0.5$. To see why, notice that if $\delta < 0.5$, $(1-\delta) > \delta$ and $(1-z_S)^{t+n}$ are lower than $(1-z_S)^{t+n-1}$ for $n > 0$.

ii) How many periods c_o *is paid.* Assume that F and S play the strategies of Corollary 1 and $\delta < 0.5$. We need to consider for how many periods player F will pay c_o in the case that she has received a good signal and no one has invested yet. Notice that in these histories, after seeing no investment from S, her beliefs about the good state, $\mu_F(z_S)$, decrease in every period. Then, as S reveals information in every period on the equilibrium path, F's belief that the state is good tends to zero. This implies that there may be a period after which F will prefer to not pay the monitoring cost: her beliefs about the good state may be too low to pay c_o one more period. Formally, F will pay c_o until period $\hat{T}$, where $\hat{T}$ is the period that satisfies the following inequalities:

$$z_S[\pi_F(I \mid \mu^{++})(\mu_F^{\hat{T}}(z_S)q + (1-\mu_F^{\hat{T}}(z_S))(1-q))] - c_o > 0$$
$$z_S[\pi_F(I \mid \mu^{++})(\mu_F^{\hat{T+1}}(z_S)q + (1-\mu_F^{\hat{T+1}}(z_S))(1-q))] - c_o < 0$$

The first inequality shows that the expected payoff of paying c_o is profitable at $\hat{T}$, but it is not worthwhile after having observed no investment at $\hat{T}$ (second inequality). After $\hat{T}$, F prefers to not pay the cost. Now we argue that if $z_S(\pi_F(I \mid \mu^{++})(1-q)) - c_o > 0$, F will pay forever, that is, $\hat{T} = \infty$. As mentioned, whenever F observes no investment, her belief $\mu_F^t(z_S)$ decreases. Hence, if $s_S = b$ $\mu_F^t(z_S)$ tends to zero, and the above inequalities converge to $z_S(\pi_F(I \mid \mu^{++})(1-q)) - c_o$. Thus, when this expression is positive F will prefer to pay c_o in every t.

Probabilities used by S. S needs to take into account that F will forecast that she may pay c_o forever. This implies that we need to modify F's indifference condition in the proof of Theorem 1 (expression (4)). Given the prescribed equilibrium strategies, F must be indifferent between investing in period 1 and waiting (and paying c_o) to observe S's action in every period. Formally,

$$\begin{aligned}\pi_F(I \mid \mu_F^{+}) = \delta^2 z_S &\left[\sum_{i=0}^{\infty}(\delta(1-z_S))^i \pi_F(I \mid \mu^{++}) Pr(s_S = g \mid s_F = g)\right] \\ &- \delta^2 c_o \left[\sum_{i=0}^{\infty}(\delta(1-z_S))^i Pr(s_S = g \mid s_F = g) + \delta^i Pr(s_S = b \mid s_F = g)\right] \end{aligned} \quad (18)$$

The first term in the right-hand-side of (18) is the same term in (4). The second term is the expected cost of observing S's action. The first part considers the probability that S has received good news, so she must pay c_o with discounted probability of

$$\delta^2 Pr(s_S = g \mid s_F = g)\sum_{i=0}^{\infty}(\delta(1-z_S))^i = \frac{\delta^2 Pr(s_S = g \mid s_F = g)}{1-\delta(1-z_S)}$$

The last term of (18) represents the possibility that S has received bad news and thus will never invest; so this cost is not discounted by the probabilities of investment:

$$\delta^2 Pr(s_S = b \mid s_F = g)\sum_{i=0}^{\infty}\delta^i = \frac{\delta^2 Pr(s_S = b \mid s_F = g)}{1-\delta}$$

Solving for z_S, we obtain the expression for z_S in the statement.

(Symmetric timing game) As mentioned above, by (A2) and (A3), if a player receives a bad signal she does not care about the other signal. Then, she will not pay c_o. Thus, we focus on the case where both players receive good signals. Let us first show that c_o is paid by F at most once when both players can invest from the first period. Consider first the case where both players use pure strategies. (Pure strategies of this game are analogous as the ones defined for Proposition 1). Notice that by the same argument that we used in Proposition 1 (see the uniqueness part), in any pure strategy equilibrium one player must reveal her signal in the first period by investing (resp. waiting) whenever she receives a good (resp. bad) signal. Then, if F is this player, she will invest without paying c_o. On the other hand, if the equilibrium prescribes that S plays this strategy, F may pay c_o just in period 1, as she will infer the signal of S from S's action in this period. Finally, consider the case where both players use mixed strategies (so they must be indifferent between investing and waiting). By Lemma 1 (see the proof of Theorem 1), whenever player i receives a good signal, if she waits one period to observe j's action and observes that j waits, i will wait. Then, notice that if both use mixed strategies, F will pay c_o at most once. To see why, notice that when F waits, pays c_o and observes an investment from S, she will invest (as she infers two good signals). In the other case, whenever she observes no investment she will wait by Lemma 1. Moreover, notice that Lemma 1 also applies to S: if S does not invest (with some probability) to observe F's decision, and observes that F waits, S will wait forever absent more information. Then, on this history, as S will not invest unless F does, and vice versa, F will not pay c_o any more. Therefore, we have proved that in the symmetric game c_o is paid by F at most once. □

Proof of Lemma 3

By (A2), if a player acquires a bad signal she will never invest ever unless some (highly precise) good signal can be inferred from the other player. Let us consider the case where i has acquired a good signal. If j has already decided before t, i will acquire a signal only if she expects to follow it: to invest (resp. to wait) if the signal is good (resp. bad). Otherwise, if any realization of the signal will imply the same decision after observing it, she will take this decision without paying the acquisition cost (as the signal is irrelevant). Therefore, in this case, if she acquires a good signal she will invest. Consider next the case where j has not already decided before t and i has acquired a good signal. If the equilibrium prescribes that j will not invest unless i invests, then there are no gains by delaying the investment decision if the signal acquired is good. Finally, consider the case where the equilibrium prescribes that j, with some probability, will acquire a signal and invest if the signal is good at some future period. Suppose per contra that i has acquired a good signal but decides to wait and see j's decision. This implies that the payoff of waiting to see is higher than investing right away. But then, she will be better off by waiting to acquire information after seeing j's decision instead of doing it at t. Notice that beliefs will change after seeing j's action, and so the information that she has already acquired in t will not longer be optimal once she observe j's decision. Therefore, as players can acquire information once, i will prefer to wait to gather information in this case. We have proved that i will invest right away if she acquires a good signal. □

Proof of Proposition 3

Condition (i). By Replacing $q_F^* = (\beta R + 1)/2$ in (10) and simplifying terms, the condition that needs to be satisfied becomes $\frac{\beta R^2}{8} + \frac{R}{4} - 0.5 > 0$. Finding the roots for R, we obtain that the positive one is

$$R^{r+} = \frac{-1+\sqrt{1+4\beta}}{\beta}$$

Therefore, for $R > R^{r+}$ condition (i) holds. Expressing the above condition for β gives $\beta > (4-2R)/R^2$, that lies in (0,0.5) whenever $R > 1.46$.

Condition (iii). We need to show that (13) holds. If $q_F^* = 1/2$, (13) is equivalent to condition (i) for F. Thus, from our analysis of condition (i) we know that gathering information and following his signal would be optimal for S when (i) holds. Furthermore, holding q_F^* constant, the derivative of (13) with respect to R is $q_S^*(1-q_F^*)$ using the envelope theorem. This implies that for any given q_F^* there will be a cutoff level of R above which the maximized value of (13) is positive, and it is therefore optimal for S to gather information, and below which it is not. Taking the derivative of the maximized (13) with respect to q_F^* using the envelope theorem yields $-q_F^*(R-1)-(1-q_F^*) < 0$, which means the cutoff value of R is increasing in q_F^*. Thus, it suffices to show that this cutoff reaches $R=2$ at some $q_F < q_F^*$. Setting $R=2$, S's optimal precision if it gathers information becomes $q_S^* = \beta + 0.5 = q_F^*$. Finally, notice that (13) is negative at $R=2$ and $q_S = q_F$ (the first two terms cancel out), providing the result.

Condition (iv). We need to show that:

$$q_F^*(R-1)+(1-q_F^*)(-1) > q_F^* q_S^*(R-1)+(1-q_F^*)(1-q_S^*)(-1)-(0.5\beta)(q_S^*-0.5)^2$$

Replacing q_F^* (given in (9)) and $q_S^* = \beta((q_F^*(R-2)+1)+0.5$ on it and rearranging terms, the above inequality becomes

$$-\frac{\beta^3 R^4}{8}+\left(-\frac{1}{4}+\frac{\beta}{2}\right)\beta^2 R^3+\left(\frac{1}{8}+\frac{\beta}{2}-\frac{\beta^2}{2}\right)\beta R^2+\frac{R}{4}-\frac{1}{2}>0$$

This expression is increasing for $R \in (0,2)$ and $\beta \in (0,0.5)$. Therefore, we only need to show that this expression is positive for low values of R and β. Setting $R = 1.46$ and β equal to the lower bound of $(4-2R)/R^2$ gives the result.

(*Uniqueness*). As we argued in the proof of Proposition 1 (uniqueness part), one player must be the first who reveals information in equilibrium and must do so at the first opportunity. Let us focus on the case where F is the first player who reveals information, that is, she acquires a signal and invests if the signal is good. Notice that conditions (iii) and (iv) rule out the possibility that S would prefer to acquire more information after inferring the signal of F when this signal is chosen with precision q_F^* (given in (9)). Then, the other equilibrium candidate is one in which F acquires a lower precision level $\hat{q}_F < q_F^*$, that makes S prefer to acquire information after seeing F's decision. However, F cannot infer S's signal given that she is the first player to reveal information. Then, F will acquire a signal with precision q_F^*. Therefore, we cannot have an equilibrium where F is the first player who reveals information and acquires a precision level different from q_F^*. This proves the uniqueness of the equilibrium.□

Proof of Proposition 4

Notice first that by Lemma 3, whenever players acquire a good (resp. bad) signal in a given period, they will invest in that period (resp. wait forever absent more news). Then, let us focus on when they decide to acquire a signal.

F's indifference condition: Consider the best reply of F at $t = 1$ given the strategy of S. As she mixes, she must be indifferent between buying a signal in $t = 1$ with precision q_F^* or waiting to observe S's decision and follow it without acquiring a signal. Notice that the former payoff is the same as expression (10). Next, we need to compute F's expected payoff of waiting until period 3 to observe S's decision, and then deciding according to the prescribed equilibrium. This expected payoff is:

$$\delta^2 \left(\frac{z_S}{1 - \delta(1 - z_S)} \right) \pi_F(I \mid s(q_S^*) = g) Pr(s(q_S^*) = g \mid \mu_0)$$

That is, the expected payoff of investing after inferring a good signal with precision q_S^* from S's investment, $\pi_F(I \mid s(q_S^*) = g)$, times the probability that S has received a good signal and invest (given F's beliefs at $t = 1$), $\delta^2(z_S/(1 - \delta(1 - z_S))Pr(s(q_S^*) = g \mid \mu_0)$. Hence, F is indifferent in period 1 if the following condition is satisfied:

$$\pi_F(I \mid s(q_F^*) = g) = \delta^2 \left(\frac{z_S}{1 - \delta(1 - z_S)} \right) \pi_F(I \mid s(q_S^*) = g) Pr(s(q_S^*) = g \mid \mu_0) \tag{19}$$

Then, notice that

$$\pi_F(I \mid s(q_S^*) = g) = q_S^*(R - 1) + (1 - q_S^*)(-1)$$

Let us explain why. By Bayes' Rule (similar to (11)) F believes that the state is good with probability q_S^* after seeing an investment (that reveals a good signal given S's strategy). Notice that for computing this payoff, F correctly infers that S will prefer to acquire a signal with precision q_S^* when she waits in period 1, where q_S^* is the signal's precision that solves:

$$\max_{q_S \in [\frac{1}{2}, 1]} \quad \mu_S^2 q_S(R - 1) + (1 - \mu_S^2)(1 - q_S)(-1) - C(q_S)$$

and therefore, $q_S^* = \beta(\mu_S(R - 2) + 1) + 0.5$. Notice that q_S^* depends on S's posterior belief after observing no investment from F in period 1, denoted by μ_S. (This is similar to the argument for Theorem 1, as F mixes S cannot infer a negative signal completely when F waits. We omit to write $\mu_S(z_F)$ as in Theorem 1 to simplify the notation.) Finally, in (19) this expected payoff is multiplied by the probability that S receives a good signal when he acquires one with precision q_S^* given F's beliefs in $t = 1$: $Pr(s(q_S^*) = g \mid \mu_0) = (\mu_0 q_S^* + (1 - \mu_0)(1 - q_S^*)) = (\frac{1}{2} q_S^* + \frac{1}{2}(1 - q_S^*)) = \frac{1}{2}$.

Now we are able to compare the payoff of investing in period 1, (10), with waiting until t=3. We need to show that

$$\frac{1}{2} q_F^*(R - 1) + \frac{1}{2}(1 - q_F^*)(-1) - C(q_F^*) < \delta^2 \left(\frac{z_S}{1 - \delta(1 - z_S)} \right) (q_S^*(R - 1) + (1 - q_S^*)(-1)) \frac{1}{2}$$

Assume $\delta^2 \left(\frac{z_S}{1-\delta(1-z_S)} \right) = 1$, we will show that this inequality is satisfied if $q_S^* > q_F^*$. Recall that $q_S^* = \beta(\mu_S(R - 2) + 1) + 0.5$ and $q_F^* = \beta R/2 + 0.5$. Then, if $\mu_S < 1/2$, $q_S^* > q_F^*$. Note that $\mu_S < 1/2$ must hold as after seeing no investment from F, and given that F invests with positive probability in period one, S adjusts her beliefs downwards from $\mu_0 = 1/2$ to μ_S. Therefore, for high enough δ, $\delta^2 \left(\frac{z_S}{1-\delta(1-z_S)} \right) \sim 1$, and we can find a $z_S \in (0, 1)$ that equalizes both payoffs, which implies that F's indifference condition, (19), can be satisfied.

S's indifference condition: Next, consider S's best response in period two. Using the same proof as for condition (iv) in the free-riding equilibrium without delay, we know that if S observes that F has invested, he will invest in period 2 without acquiring a signal. On the other hand, whenever F not invests in the first period, S mixes. Therefore, he must be indifferent between acquiring a signal with precision q_S^* and deciding as in the prescribed equilibrium, or waiting forever with a payoff of zero, given that F will not reveal any information after period one. That is,

$$\pi_S(I \mid s(q_S^*) = g) = 0 \tag{20}$$

As mentioned above, notice that after seeing no investment in period 1 his belief about the good state must be updated downwards from $\mu_0 = 1/2$ to μ_S^2. Notice that this belief depends on z_F, the probabilities used by F in equilibrium. Thus, we need to find a z_F that satisfies the above indifference condition (equation (20)).

Probabilities used in equilibrium: We need to solve for z_F in (20). Consider S's expected payoff of acquiring a signal and deciding accordingly in period 2 after seeing no investment:

$$\pi_S(I \mid s(q_S^*) = g) = \mu_S^2 q_S^*(R-1) + (1-\mu_S^2)(1-q_S^*)(-1) - C(q_S^*) \tag{21}$$

where $q_S^* = \beta(\mu_S^2(R-c) + (1-\mu_S^2)c) + 0.5$. Then, as we show in the proof of Theorem 1 (using now $\mu_0 = 0.5$), by Bayes' Rule,

$$\mu_S^2 = \frac{\mu_0[(1-z_F) + z_F(1-q_F^*)]}{\mu_0[(1-z_F) + z_F(1-q_F^*)] + (1-\mu_0)[(1-z_F) + z_F q_F^*)]} = \frac{1 - z_F q_F^*}{2 - z_F}$$

Replacing μ_S^2, q_S^* and $C(.)$ in (21) and equalizing it to zero−the expected payoff of waiting (see (20)), we obtain the following quadratic equation in z_F:

$$a z_F^2 + b z_F + c = 0, \text{where}$$

$$\begin{aligned}
a &= -4 + 2R + (3 - 4\beta + 4\beta^2)\beta R^2 + (2 - 4\beta)\beta^2 R^3 + \beta^3 R^4 \\
b &= 16 - 8R - (8 - 8\beta)\beta R^2 - 4\beta^2 R^3 \\
c &= -16 + 8R + 4\beta R^2
\end{aligned}$$

Solving for the roots of the quadratic function, the one that lies in $(0,1)$ is expression (14) provided $R > 1.46$ and $\beta > \frac{4-2R}{R^2}$.

Finally, we need to solve for z_S. Solving for it in F's indifference condition (19) we obtain expression (15). z_S lies in $(0,1)$ provided $\delta > \delta^*$ where δ^* was defined in the statement of the Proposition.

First mover, periods $t > 1$: We will show that after S's investment F prefers to invest without gathering information, and after no investment F prefers to wait. For the former, note that we have proved for the free-riding equilibrium that after seeing an investment with precision q_F^*, the other player will prefer to invest without gathering information (condition (iv)). This implies that if a player prefers to invest without acquiring information when she infers a good signal with precision q_F^* from the other player's action, she must prefer the same when inferring a good signal with a higher precision. Recall from the previous paragraph that $q_S^* > q_F^*$. Therefore, after S's investment,

F will follow it without acquiring information.

Next, consider the best reply of F when she observes no investment from S. In period two, as she is not already able to see S's action in this period, she does not update her belief. This implies that at $t = 2$ she prefers to wait: if before she was indifferent between investing or waiting for information, now she must strictly prefer to wait as the information will arrive sooner. Next, consider the best reply of F in period $t \geq 3$, when she observes no investment in $t-1$. We need to prove that she still prefers to wait instead of acquiring a signal.[29] We show this by taking the derivative on both sides of her indifference condition (19) with respect to F's beliefs, and then we show that the payoff of acquiring a signal whenever F observes no investment from S decreases faster than the payoff of waiting.

Denote by x F's belief that the state is good at t after having observed no investment by S at $t-1$. Then, the expected payoff of acquiring a signal instead of waiting more is:

$$xq_F(x)(R-1) + (1-x)(1-q_F(x))(-1) - 0.5\beta^{-1}(q_F(x) - 0.5)^2$$

where $q_F(x) = \beta(x(R-2)+1) + 0.5$. Taking the derivative with respect of x, we have that when F's becomes more pessimistic, this payoff is reduced by

$$0.5R + \beta(R-2)(1 + (-2+R)x) > 0$$

Next, let us consider how the payoff of waiting changes when F's beliefs decrease. The payoff of waiting until S invests is:

$$\delta^2 \left(\frac{z_S}{1 - \delta(1 - z_S)} \right) \pi_F(I \mid s(q_S^*) = g)(xq_S^* + (1-x)(1-q_S^*))$$

where $\pi_F(I \mid s(q_S^*) = g) = q_S^*(R-1) + (1-q_S^*)(-1)$ does not depend on x. Taking the derivative with respect of x we have

$$\delta^2 \left(\frac{z_S}{1 - \delta(1 - z_S)} \right) \pi_F(I \mid s(q_S^*) = g)(2q_S^* - 1)$$

Then, as $\delta^2 \left(\frac{z_S}{1-\delta(1-z_S)} \right) (2q_S^* - 1) < 1$,if we show that

$$0.5R + \beta(R-2)(1 + (-2+R)x) > \pi_F(I \mid s(q_S^*) = g) = q_S^* R - 1$$

then we have the result. Next, note that the left-hand side is decreasing in x, and the right-hand side is increasing in q_S^*. Moreover, as q_S^* increases x moves downward, as the posterior beliefs of F after observing no investment depends on the precision q_S^* (this is analogous to how μ_S decreases as q_F^* increases, shown above). Therefore, setting $x = 0$ and $q_S^* = 1$, the inequality simplifies to $0.5R + \beta(R-2) > R - 1$, that holds for $\beta < 1/2$. Hence, we have proved that the left-hand side of F's indifference condition (19) decreases faster than the right-hand side after seeing no investment. That is, waiting is better in every t after seeing no investment if F is indifferent in period 1.

Finally, for completeness we must provide the best reply of F on the history that she has acquired a bad signal in period 1 but observes an investment from S at some t. Recall that $q_S^* > q_F^*$: the

[29] As we do not have an analogue of Lemma 1 for this game, we cannot use the same argument as in Theorem 1.

precision that S acquires is higher than the precision gathered by F. F's belief after seeing an investment is:

$$\mu_F = \frac{(1-q_F^*)q_S^*}{(1-q_F^*)q_S^* + q_F^*(1-q_S^*)}$$

Then, the expected payoff of investing is $\mu_F(R-1)-(1-\mu_F)$. If this expression is positive F will prefer to invest; otherwise she will wait. Notice that μ_F is increasing in q_S^* and decreasing in q_F^*, both of which are increasing in R. However, the effect of q_S^* on μ_F dominates the effect of q_F^*. Hence, this expected payoff is increasing in R. We do not provide exact conditions, but we know from numerical examples that there is a cutoff value for R above which it is profitable to invest and below it is not.

Off the path beliefs: Same as in the proof of Proposition 1. □

Chapter 3

Segregation and Beliefs in a Just World

3.1 Introduction

Recent work has shown that upward mobility in U.S. communities is highly negative correlated with residential segregation (Wilson (2012); Sampson et al. (2002); Chetty et al. (2014); Chetty and Hendren (2018a,b)). This becomes relevant because of most industrialized countries have seen a remarkable increase in income and wealth inequality over the past 35 to 40 years (see e.g. Piketty (2018)), and this rising inequality has frequently been accompanied by an increase in socio-economic segregation.[1] Often, middle-income neighbourhoods have made way for both rich and poor communities, and segregation and the erosion of the middle class have gone hand in hand. However, despite these facts that threaten the "American dream", the mechanism that underlies the link between mobility and segregation is not well understood which represents an obstacle to look for potential policy interventions aimed at increasing the social mobility rates.

On the other hand, previous work has revealed striking differences between the views held by people concerning the causes of wealth and poverty, the extent to which individuals are responsible for their own fate, and the long-run rewards to personal effort. People differ substantially on practical assessments concerning the key to personal success, the poor emphasizing structural factors; the rich, personal qualities such as effort and ambition (see Rytina et al. (1970); Kluegel and Smith (1986); Miller (1992); Alesina et al. (2018)). In the context of educational attainment, recent work has documented that beliefs about returns are decisive determinants of individual schooling decisions (Jensen, 2010; Attanasio and Kaufmann, 2014; Kaufmann, 2014; Belfield et al., 2020). In this paper we extend the model of Levy and Razin (2017) to offer an explanation for the relationship between segregation and low mobility based on these different assessments about the value of effort to escalate the income ladder, what Benabou and Tirole (2006) call *beliefs in a just world*. Our argument is that this different assessments about the value of effort between the poor and rich, or more and less educated individuals, is more pronounced in segregated communities than in integrated ones. We argue that segregated communities exhibit more polarized and pessimistic views that hard work pays off than integrated ones because families in those communities learn differently about the value of effort. Then, this polarization and pessimism in segregated communities makes

[1] Watson (2009) and Reardon and Bischoff (2011) demonstrate that both income inequality and income segregation have risen sharply in the US between 1970 and 2000, especially in metropolitan areas. An income segregated community refers to an uneven spatial distribution of individuals based on income. For example, if there are two groups, e.g. rich and poor, and two neighbourhoods in the community, a perfectly segregated community has each group living in a different neighbourhood, whereas a perfectly integrated one the same proportion of each group in each neighbourhood.

mobility lower, as those families with low beliefs in effort end up with a higher income inertia, i.e. children and parents have a similar income. We provide evidence for U.S. suggesting that more segregated communities have more polarized and pessimistic beliefs in a just world than integrated ones.[2]

We consider that families may develop conflicting views about whether economic success is based on luck (predetermined factors) or individual effort, not because they are maximizing different objective functions, but rather because through lifetime experience they happen to learn and to believe different things. The model works as follows. First, beliefs in a just world are shaped by peers in neighbourhoods. Neighbourhoods usually involve intense socialization, where beliefs are formed and moulded by peers. Second, young adults decide their level effort before entering the labour market. This early-stage effort is taken under uncertainty when it is difficult to predict how the labour market will value these decisions and to change it later once the decision is made. Third, actual mobility experiences in the labour market is informative about the value of effort, but the information depends on the level of effort previously decided. We assume that choosing low effort is uninformative about the value of effort, whereas choosing high-effort is risky as effort may not be the key to success and could be better to choose low effort.[3] Finally, after labor market experience, families decide where to live based on how neighbourhood's beliefs will influence their offspring. Taking both learning determinants into account, neighbourhood socialization and labour market experience, we analyse how the learning process about the underlying opportunities of social mobility is affected by the level of segregation in the community, that we allow to be determined endogenously by families.

Following Levy and Razin (2017), our analysis relies on two important "behavioral" assumptions. First, we assume that individuals during childhood have selection bias. In particular, we assume that individuals exchange information only with those in the same neighbourhood and neglect to take into account that the selection into the neighbourhood depends on parents' beliefs. Selection bias is not sufficient to create segregation and polarized beliefs as this depends on how people select into neighbourhoods. The second assumption below implies homophily, which together with selection bias will create the "echo chamber" effect. Specifically, we assume that parents decide on a neighbourhood for their child using "imperfect empathy", as in Bisin and Verdier (2001). Parents base their decisions on their expectation about how the neighbourhood will affect their child's future beliefs and behavior. Parents understand that the environment will influence their offspring's beliefs but think that the optimal course of behavior is the one that follows their own beliefs (hence, empathy is imperfect). This creates homophily, that is, parents would prefer that their children segregate with like-minded others so that their child's belief does not end up too far from their own. This endogenous homophily, along with selection bias, may imply that dynasties do segregate themselves in the city and beliefs become polarized.

In this framework, we analyse the following questions: Are income segregation and low social mobility sustainable in the long run? That is, if some agents overestimate the value of effort for economic success, do they indeed segregate so that beliefs in social mobility perspective become

[2]When we refer to belief polarization we mean, as in Levy and Razin (2017), that if two groups start with some average beliefs, as time passes these average group beliefs become more different between groups and more extreme.

[3]As Lemieux (2006) points out, relative wages among the highly educated have become much more dispersed since the mid 1970s. He suggests this is the result of a strong heterogeneity in the returns to higher education. Much of the increase in labor market inequality is the result of this heterogeneity and reflects higher wages at the very top of the distribution. Altonji et al. (2012) show that the wage differences between degree graduates in various fields of study are as large as the differences between those with a degree and those without a degree.

polarized and self-fulfilling in the long term? If this is the case, it implies low-mobility? Specifically, we assume that families in one neighbourhood start with strong beliefs in the reward of effort, and with higher beliefs than the families from the other neighbourhood. We uncover two types of initial conditions for the low-beliefs neighbourhood that when satisfied, in all equilibria, there is income segregation in the log run and polarized beliefs about the value of effort in each neighbourhood. Dynasties in each neighbourhood believe that the value of effort is different, choose different effort levels, receive different income and choose to live in different neighbourhoods. This also implies low social mobility: each cohort income is more correlated with their parents' income than with the income of parents from others neighbourhoods. The first case is one in which the families from the low-beliefs neighbourhood start with beliefs that points to exert low effort. Then, as choosing low effort is uninformative about the value of high-effort in the society, they are trapped in this situation and never take the risk of choosing high-effort. The second case is one in which families from the low-beliefs neighbourhood have also relatively strong beliefs in effort and start choosing high-effort. However, as we explain below, families from the high-beliefs neighbourhood have incentives to segregate, and so the socialization process between families in the low-beliefs neighbourhood makes their beliefs pessimistic in the long run. We argue below that this case is a plausible explanation for the segregation mechanism in U.S.

Next, we provide evidence for U.S. that supports our theoretical results. Using U.S. survey-data with beliefs in a just world and the level of segregation in the community in which respondents live, we show first that more segregated communities are correlated with more polarized and pessimistic views about the value of effort. Second, we provide suggestive evidence about the segregation mechanism highlighted in the previous paragraph. Our theoretical results predict that when initial beliefs in a just world are high in the population, as has been argued for U.S. history, segregation in the long run arises when high-income individuals start from relatively more pessimistic views in comparison with the case where integration is the long run equilibrium.[4] The theoretical argument is that when beliefs of high-income individuals are not so high, they have more incentives to segregate as they infer that their child is going to be highly influence by the neighbourhood. In contrast, if beliefs are optimistic enough, high-income parents do not care much about the neighbourhood's influence and more integration is possible. In line with this, we show in the data that U.S. segregated communities exhibit more pessimistic views for individuals with high-income in comparison with integrated communities. Moreover, we argue that this segregation mechanism offers a plausible explanation for the results of Rothstein (2019) that explores what drives the massive heterogeneity in social mobility rates across U.S.'s geography. He shows that what mostly explains the difference in social mobility rates between U.S.'s communities, 40%, is the relative likelihood that children from high- and low-income families have working spouses. Low mobility communities have more high-income households with working spouses.[5] In this sense, our model captures the idea that the difference

[4] Benabou and Tirole (2006), p.726: "...as suggested by many analysts of American history, the most likely role of land abundance was to shape initial views on opportunity and social mobility.", and footnote 29, p.726: "Long and Ferrie [2005] indeed confirm that between 1850 and 1880 the United States was indeed much more mobile than Great Britain (with occupational and geographical mobility going hand in hand) but that mobility declined significantly after 1920, bringing to an end the reality of "American exceptionalism" with respect to social mobility—though not its popular perceptions." The American "exceptionalism", manifested by the widely held belief in that effort and hard work pay off, is well documented in international surveys comparisons as well. Data from the World Values Survey (Alesina et al. (2001)) shows that only 29 percent of Americans believe that the poor are trapped in poverty and only 30 percent that luck, rather than effort or education, determines income. The figures for Europeans are nearly double: 60 percent and 54 percent, respectively.

[5] He finds that this reflects differences in the likelihood of having a working spouse, not differences in assortative matching or inheritances.

between integrated or segregated communities is that in the latter ones high-income dynasties are more worried about the effects of mixing with low-income dynasties because the potential effects on their offspring (as they start with relatively low initial beliefs in effort). Then notice that this mechanism can be amplified if these high-income parents of segregated communities have higher cost to transmit their beliefs to their child, perhaps because they must work for longer hours and spend more time outside home.[6] This offers an explanation based on segregation for the results of Rothstein (2019): household with working spouses and less time to transmit beliefs in effort to their child are more prone to segregate, from which low mobility in the long run follows.

Our theory of the link between segregation and mobility is also consistent with other three pieces of evidence. Chetty et al. (2014) show that social mobility is strongly correlated with income segregation. However, a priori it is unclear whether the correlations between segregation and mobility is driven by the causal effects of place or selection effects. For instance, is growing up in a less segregated area beneficial for a given child or do families who choose to live in less segregated areas simply have better unobservable characteristics? Chetty and Hendren (2018b) find that 80% of the association between segregation and upward mobility across commuting zones in observational data is driven by the causal effect of place; and 20% is due to sorting. In line with this, our model includes both factors that explain the link between segregation and mobility, the initial beliefs on neighbourhoods and peers' influence (causal effect) and parents' neighbourhood choice (selection). Second, Chetty and Hendren (2018b) examine how much more one has to pay for housing to live in an area that generates better outcomes for one's children. Within communities, counties that produce better outcomes for children have slightly higher rents, especially in highly segregated cities. However, rents explain less than 5% of the variance in counties' causal effects for low-income families and it shows that some areas are "opportunity bargains"- counties within a labor market that offer good outcomes for children without higher rents. They suggest that the opportunity bargains may partly exist because families do not know which neighbourhoods have the highest value added. Our theory points to an alternative explanation, "opportunity bargains" are not taken by families because they believe are not worthwhile. Finally, an emerging body of evidence shows that countries with more inequality at one point in time also experience less earnings mobility across the generations, a relationship that has been called "The Great Gatsby Curve" (see Corak (2013)). Our theory predicts that segregated communities with less social mobility also features more income inequality.

The work is organized as follows. In the next Section we review the literature. In Section 3.3 we present the model. In Section 3.4 we analyse the model and show conditions for segregation and integration. In Section 3.5 we present the data, empirical strategy and results. Finally we conclude in Section 3.6.

3.2 Related work

Our work contributes to the recent literature that seeks to explain what drives the differences between social mobility rates across U.S. regions (Chetty et al. (2014); Chetty and Hendren (2018a,b); Rothstein (2019)). We contribute to this literature by explaining the link between segregation and mobility observed in the data.

[6] In our simple model we have assumed that parents transmit their beliefs at no cost, and also that they do not interfere with their child's socialization. Introducing such a cost would not change the intuition of the segregation mechanism that we highlight.

Second, our work is related to the literature on neighborhood effects and endogenous socioeconomic segregation explaining how local interactions drive spatial segregation and persistent income inequality (see for instance, Loury (1976); Benabou (1993, 1996b,a); Borjas (1998); Durlauf (1996)). In these analyses, the dynamics of income inequality rely on human-capital accumulation, and individual human capital is determined by both that of their parents and local spillovers. More related to our work are Durlauf and Seshadri (2018); Fogli and Guerrieri (2019) that explain the link between inequality and mobility through segregation. The idea is that inequality allows rich families to afford better neighbourhoods. Then, segregation implies better resources for children of rich neighbourhood, such as better schools, and finally this results in lower social mobility. The key ingredient of the model is a local spillover: investment in education yields higher returns in neighbourhoods with higher average level of human capital.[7] As these works mention, such a spillover can capture a variety of mechanisms, such as differences in the quality of public schools or peer effects, and it is not the scope of these models to determine which is more important or to explain how they work. We contribute to this literature by focusing on the microeconomics of one channel, beliefs in effort. Moreover, Rothstein (2019) shows that human capital plays a relatively small role in the geographic variation of social mobility rates, and much of this variation appears to reflect differences due to local labour market institutions or differences in access to good jobs, and cultural tendencies of having working spouses.[8] In line with this, we abstract from human capital accumulation issues and sorting mechanisms driven by income, and put the focus on the information channel of the labour market and how this is reinforced by neighbourhoods.

Finally, our work is related with the literature on beliefs in a just world. This works mostly focuses on how these beliefs may explain preferences for redistribution, and the different policies between U.S. and Europe (Piketty (1995); Alesina et al. (2001); Alesina and La Ferrara (2005); Alesina and Angeletos (2005); Benabou and Tirole (2006); Alesina et al. (2018)). We borrow from this literature the idea of the underlying uncertainty about the true value of effort, or social mobility parameters, that dynasties may learn over time. The closest works are Piketty (1995) and Benabou and Tirole (2006). The former focuses on learning from own experience, not allowing peer effect (nor endogenous segregation of dynasties), and how this affects preferences for redistribution. The latter, on why the need to believe in a "just world" varies considerably across countries and the implications of this phenomenon for international differences in political ideology, redistribution.

Our model is an extension of Levy and Razin (2017). In their model there are two kinds of schools: state and private. All dynasties know the productivity of private school graduates but are uncertain about the productivity of a state school graduate. Then, all individuals go through the following phases. First, peers influence in school, where they communicate their beliefs about public schools. Next, in the labour market school graduates become employees or employers and are randomly matched. Employers observe the schooling history of employees and make employment decisions given their beliefs. Labour market experience is potentially informative about the pro-

[7] The relevant assumption is that the local spillover is complementary to the children's innate ability and to their level of education. This generates sorting in equilibrium: richer parents with more talented children choose to pay higher rents to live in the neighbourhood with higher average human capital. It follows that in equilibrium one neighbourhood becomes endogenously the "good" one and hence the one where houses are more expensive. This means that in this kind of model, the residential choice is a form of human capital investment.

[8] Rothstein (2019),p.122: "The evidence points to other factors as potentially more important, including cultural tendencies toward early marriage and local labor market factors that influence the labor force participation rate and the ability of children from high-income families to match into high-earnings jobs conditional on their education and skills."

ductivity of state school graduates. Finally, individuals have one offspring each and choose a school for their offspring, private or state, to maximize their perceived offspring's payoff in the labour market. In this sense, the same learning forces of their work are present in our model, learning from experience and from peers (and the behavioural assumptions mentioned in the Introduction). We depart from their analysis in three directions. First, we introduce a different underlying uncertainty, the rewards of effort, and incorporate into the model elements relevant to our case accordingly (e.g. effort decision, income conditional on effort).[9] The focus of their model is on discrimination. In their labour market stage, discrimination may arise based on beliefs. Employers decide whether to hire an employee, based on the school he graduated from and their own beliefs about the schools' effect on productivities. Second, we allow decisions from agents from a given cohort (the effort choice) to influence how much they learn from later experience. They only allow parents' decision to influence learning.[10] These differences change the conditions for segregation and result in a different mechanism for why dynasties segregate in both models.[11]

3.3 Theoretical framework

3.3.1 The model

We consider a non-overlapping generations model with a continuum of dynasties, each indexed by $i \in [0, 1]$. Each dynasty consists of one individual at any period $t \in \{1, ...\}$ who is replaced by one offspring at the end of the period. The city is comprised of two neighbourhoods indexed by $J \in \{P, R\}$. Houses are identical across the city, and the supply of houses is inelastic. We consider that the land rent is paid to absentee landlords. It is assumed that neighbourhood R can accommodate a fixed share ρ of dynasties, and the dynasties with the highest willingness to pay for living at R are the ones that live there (we will explain this assumption in detail later).[12] The focus of our analysis will be on how (endogenous) segregation in different neighbourhoods can foster polarised beliefs about the value of effort, and how this is in turn reinforces residential segregation. The value of effort, $\theta \in \{L, H\}$, is determined in the labour market. It is profitable to take costly effort before entering the labour market if $\theta = H$, and to take low effort otherwise. Crucially, dynasty i does not know the value of effort when she chooses her effort level and believes that $\theta = H$ with probability $q_i^t \in (0, 1]$, where some $F^t(.)$ describes the distribution of q_i^t in the population. The beliefs of different dynasties will change within periods and across time as we describe below, and we will analyse conditions for beliefs and behaviour to converge. At any period t, all individuals go through the following phases:

Stage A: Peer influence in childhood. Individuals start with the belief q_i^t inherited from their parents. They communicate with peers in the neighbourhood and update their beliefs.
Stage B: Effort choice. Individuals decide their effort level based on their beliefs.
Stage C: Labour market. Individuals become employees and receive a wage that depends on the level of effort chosen and the state θ. Labour market experience is informative about the value of effort.

[9] In their model the wage is drawn randomly for every individual, in our model it depends on the effort decided and on the underlying uncertainty.

[10] While their dynasties learn from an exogenous matching procedure with the other group, what can be learnt from experience in our case is endogenously decided through effort decisions.

[11] For example, integration (i.e. non-segregation) arises in their model if wrong-beliefs parents realize that their offspring, when become employers, can incur in losses by underestimating the productivity of the other group.

[12] Our city is the same as in Bezin and Moizeau (2017), with the difference that we do not introduce exogenous land rents and dynasties move based on endogenous willingness to pay for living in a neighbourhood.

Stage D: Parental neighbourhood choice. Following labour market experience, individuals have one offspring each, to whom they transmit their beliefs. They then choose a neighbourhood for raising their offspring to maximize their offspring's labour market returns.

We now describe the specific stages of the model.

A. Peer influence in neighbourhood. We follow Levy and Razin (2017) for the modelling of peer influence. The key assumption is that there is peer influence in neighbourhoods and that residents have "selection bias" when updating their beliefs based on what they learn from peers. As in their model, we suppose that some individuals in the same neighbourhood transmit their beliefs q_i^t truthfully to each other. Specifically, let $f_J^t(.)$ denote the distribution of the inherited beliefs of residents in neighbourhood J at period t. We assume that individual i in neighbourhood J is randomly matched and exchanges beliefs simultaneously with $n > 0$ others in the same neighbourhood.[13] Note that individuals have a selection bias: they only learn from those that they interact with while not taking into account that those who chose to live in the other neighbourhood may have different beliefs. This will play an important role if the neighbourhood choice decision leads to segregation of beliefs.[14] Apart from selection bias, the belief updating process follows from Bayesian updating assuming conditional independence in the *initial* beliefs.[15] We denote by $q_{i,g}^t$ the beliefs that the child from dynasty i holds after neighbourhood interaction (when she has grown up), and $f_{J,g}^t$ the distribution of beliefs of individuals from neighbourhood J after childhood.

B. Effort choice. Once individuals have finished their childhood and have updated their beliefs to $q_{i,g}^t$, they must decide on their level of effort $e \in \{\underline{e}, \overline{e}\}$ where $\overline{e} > \underline{e}$. We normalize these values such that $\underline{e} = 0$.[16]

C. The labour market. Following the effort decision, individuals enter the labour market and receive a wage $w(\theta, e) \in \{\underline{w}, \overline{w}_L, \overline{w}_H\}$ based on their effort choice, e, and on the unknown value of effort, θ. We assume that $\underline{w} < \overline{w}_L < \overline{w}_H$. An individual who chooses a low effort level ($\underline{e}$) receives a low wage, denoted by $\underline{w}$, with certainty, that is, independent of θ. On the other hand, whenever individuals incur high effort ($\overline{e}$) they may receive two possible wages, $\overline{w}_L$ or $\overline{w}_H$, that depend on θ. If $\theta = k \neq l$ then an individual who chooses high effort $\overline{e}$ receives wage $\overline{w}_k$ with probability $\tau \in (0.5, 1)$ and wage $\overline{w}_l$ otherwise. High-effort wages are therefore imperfect signals about the value of effort θ, with symmetric precision denoted by $\tau \in (0.5, 1)$, i.e. $P(\overline{w}_\theta = \overline{w}_k | \theta = k) = \tau$ for $k \in \{L, H\}$.[17]

D. Parental neighbourhood choice. We denote by $q_{i,e}^t$ the beliefs after labour market experience.

[13]We assume that each matched individual simultaneously communicates their initial (inherited) beliefs, and after the communication phase they update their beliefs.

[14]Peer influence occurs among individuals in the same neighbourhood. As in the network literature, which assumes that individuals communicate with those they are connected to, in this model information exchange arises only among those that interact with one another. An important difference, however, is that in this model "the network", i.e. the identity of the individuals in the different neighbourhoods, will be endogenously determined by the parents, given their beliefs.

[15]Note that given the continuum of dynasties, two dynasties will meet with probability 0. Thus, conditional on initial beliefs being derived from independent sources, all future exchanged beliefs would still be conditionally independent. Hence, there is no correlation neglect in the model.

[16]One natural interpretation of the level of effort is whether to pursue higher education, $\overline{e}$, or not, $\underline{e}$. However, the effort level could also include, for example, networking efforts in early stages of life in order to have better labour market connections, or an unpaid learning period in a company.

[17]See footnote 3 for a motivation of these assumptions.

Following employment, individuals become parents and transmit their beliefs to their children directly. That is we assume $q_i^{t+1} = q_{i,e}^t$. At this stage parents decide which neighbourhood to raise them in. A parent only cares about her direct offspring. Let $\triangle_i^t$ be the willingness of a parent from dynasty i to pay for the neighbourhood R. In other words, $\triangle_i^t$ is the perceived future payoff of the child in the labour market in period $t+1$, conditional on forming beliefs in neighbourhood R, relative to that payoff conditional on forming beliefs in P. The parent will calculate this expected labour market payoff from each neighbourhood given her forecast of how the offspring's beliefs will change in each neighbourhood. Importantly, the parent would compute $\triangle_i^t$ given what she believes to be the "true" expected value of effort, $q_{i,e}^t$. Parents will therefore prefer to raise their children in the neighbourhood that will induce the best beliefs from their own point of view. Following Levy and Razin (2017), this assumption of "imperfect empathy" implies that they would prefer their offspring's beliefs to stay as close as possible to their own. Note that even though parents realize that their offspring will be influenced by others' beliefs, they do not update their beliefs any more. Imperfect empathy implies "old age rigidity" of beliefs. This phenomenon is recently documented in the literature (see Ortoleva and Snowberg (2015)). In our model, young individuals are influenced by the environment they grow up in, but when they grow old and after they have accumulated job market experience, parents believe that their assessment is the correct one. They are aware however that the environment they choose to raise their children in will affect their children's beliefs and hence their future behaviour.[18]

Who lives in R. We assume that the dynasties in neighbourhood R at time t are the individuals with the highest $\triangle_i^t$. If a share of individuals greater than ρ have the same $\triangle_i^t > 0$, the individuals who end up at R are chosen randomly.

Payoffs. Let us define the following utility function at period t: $U^t(w,e) = w(\theta, e) - e$. Then, for any t, the young adults choose their effort level (see Stage B) in order to maximize the following expected payoffs based on the beliefs that they hold at that life stage , $q_{i,g}^t$:

$$E\{U^t(w,e) + \delta E[U^{t+1}(w,e)] | q_{i,g}^t\} \tag{1}$$

where $\delta \in [0,1]$ is a discount factor. Notice that when a parent maximizes (1) by choosing her effort level, she considers her own utility ($U^t(w,e)$) and the utility of her future child ($U^{t+1}(w,e)$) at $t+1$. Next, once the young adult becomes a parent, in Stage D she must decide where to live, R or P, in order to maximize the expected utility of the child at $t+1$, i.e. $E[U^{t+1}(w,e)|q_{i,e}^t]$, based on the beliefs that the parent holds at that life stage, $q_{i,e}^t$.

Equilibrium definition. An equilibrium in the infinite game given some initial state in period 1 (an allocation of a measure ρ of dynasties to the neighbourhood R) is a dynamic process of neighbourhood peer influence, effort choice, labour market experience, and parental neighbourhood choice, in which, at any period t, the following is satisfied:

(i) *Optimal effort choice:* At any period $t = 1, 2, ...,$, individuals maximize their expected payoffs given in (1), using the belief $q_{i,g}^t$.

[18] Notice that a rational parent would infer something about θ if she has information about the distribution of beliefs in each neighbourhood when deciding where to live. To simplify the analysis of the model we follow Levy and Razin (2017) and assume that parents do not update their beliefs any more after the labour market experience. The relaxation of this assumption is left for future work.

(ii) *Optimal neighbourhood choice:* Parents at period $t = 1, 2, ...$ compute $\triangle_i^t$ using imperfect empathy and their own beliefs $q_{i,e}^t$, given correct expectations of how their offspring's beliefs will change in each neighbourhood, and the correct expectation of her offspring's effort choice in period $t+1$.[19] The measure ρ of dynasties with the highest $\triangle_i^t$ move in to neighbourhood R.

(iii) *Belief updating:* First, following Sobel (2014); Levy and Razin (2017) the evolution of beliefs of an individual i in neighbourhood J at time t, from q_i^t, before she interacts with others, to $q_{i,g}^t$ after communicating with n peers, is given by the following process:[20]

$$q_{i,g}^t \equiv C(\mathbf{q}^t, q_i^t) = \frac{q_i^t \prod\limits_{k=1,\dots,n} q_k^t}{q_i^t \prod\limits_{k=1,\dots,n} q_k^t + (1-q_i^t) \prod\limits_{k=1,\dots,n} (1-q_k^t)} \tag{2}$$

where $\mathbf{q}^t$ is a vector of beliefs $(q_1^t, ...q_n^t)$ of length n, each belief drawn from $f_J^t(.)$. Second, as wages are informative about the value of effort, following the employment phase, all those who received a wage $\overline{w}_H$ update their beliefs towards $\theta = H$, i.e. $q_{i,g}^t$ increases, as $\overline{w}_H$ is more likely under $\theta = H$. Similarly, those who received a wage $\overline{w}_L$ update their beliefs towards $\theta = L$, i.e. $q_{i,g}^t$ decreases. By Bayes' rule $q_{i,e}^t$. in each case, is given by:

$$q_{i,e}^t = \frac{q_{i,g}^t \tau}{q_{i,g}^t \tau + (1-q_{i,g}^t)(1-\tau)} \quad \text{if } w(\theta, e) = \overline{w}_H \tag{3}$$

$$q_{i,e}^t = \frac{q_{i,g}^t (1-\tau)}{q_{i,g}^t (1-\tau) + (1-q_{i,g}^t)\tau} \quad \text{if } w(\theta, e) = \overline{w}_L \tag{4}$$

In equilibrium, for each dynasty i and period $t = 1, 2, ...,$ q_i^t is updated to $q_{i,g}^t$ according to (2) given the distribution of beliefs f_R^t and f_P^t in period t. Finally, if high-effort is chosen, $q_{i,g}^t$ is updated following employment to $q_{i,e}^t$ according to (3) and (4).

3.3.2 Analysis

We start by stating that an equilibrium exists.

Proposition 0: *An equilibrium exists.*

PROOF. See the appendix.□

Next, we restrict the relative values for wages and effort in order to consider interesting cases. It is assumed that the optimal effort decision depends on whether high-effort is rewarded in the market ($\theta = H$), or luck mostly drives economic success ($\theta = L$). We assume that it is better to

[19] That is, parents know the present distribution of beliefs f_J^t, the equilibrium neighbourhood choices of other parents (and hence the correct distribution of beliefs in neighbourhood J in period $t+1$, f_J^{t+1}), and have the correct forecast of the distribution of beliefs following childhood from neighbourhood J.

[20] For the proof that Bayesian updating of the communication stage results in (2) see Proposition 5 of Sobel (2014) and Proposition 1 of Levy and Razin (2016). To provide some background for it, both works assumed that each individual receives a private signal $s \in S$ (S finite) about the state of the world. Let $P(\theta,\mathbf{s})$ be any joint probability that the state is θ and the profile of signals is $\mathbf{s} = (s_1, ..., s_N)$. Then, they shows that for any P if individuals believe that they all started with a common uniform prior for θ (on (0,1)), and that their posterior beliefs q_i^t were formed by receiving private conditionally independent signals, then these posterior beliefs are sufficient statistics for rational Bayesian updating. Finally, they show that the posteriors after communication are given by (2).

choose $\overline{e}$ whenever $\theta = H$, and $\underline{e} = 0$ whenever $\theta = L$. This requires the following assumptions:

Assumption:

$$\overline{w}_H - \overline{e} > \underline{w}, \quad \overline{w}_L - \overline{e} < \underline{w} \tag{A1}$$

$$(\overline{w}_H - \overline{e}) - \underline{w} = \underline{w} - (\overline{w}_L - \overline{e}) \equiv \pi > 0 \tag{A2}$$

These assumptions imply that choosing high-effort is risky as if the true state is $\theta = L$, it would be better to choose low-effort (as $\tau(\overline{w}_L - \overline{e}) + (1-\tau)(\overline{w}_H - \overline{e}) < \underline{w}$). The assumed symmetry between the payoff of choosing the "right" effort level knowing $\theta = k$ in (A2) simplifies the exposition but our results hold with asymmetry as long it is better to choose low-effort when the true state is $\theta = L$. Notice that if individuals choose high effort under $\theta = L$, some dynasties may obtain $\overline{w}_H$ (the income that is more likely to receive if $\theta = H$) with probability $1-\tau$. In this sense, high effort is ex-post rewarded, but driven by luck.[21]

Finally, we assume the following initial state of beliefs and size of neighbourhoods. Suppose that all dynasties in neighbourhood R (resp. P) start period 1 with belief q_R (resp. q_P), where $q_R > q_P$. Where will society converge in the long-run? Below we analyse each of the stages of the model, the child's effort decision, the parents' neighbourhood choice, and how beliefs evolve. After this, we provide conditions on the initial beliefs q_R and q_P that generate segregated or integrated communities based on income.

The effort and neighbourhood decisions

Consider first the effort decision. Notice that the effort decision at t will be taken based on beliefs that the young adult holds at t, $q^t_{i,g}$. Let us refer to the young adult as parent. Notice that if a parent chooses high effort she will receive an informative signal about θ, whereas if she chooses low effort she does not. So choosing $\overline{e}$ has an *information value*. This implies that the parent's optimal decision when $\delta = 0$ might not be the same if $\delta > 0$ (see (1)). To see why notice that the parent may want to choose a suboptimal level of effort for her utility, $U^t(.)$, to receive a signal and then make a more informative decision for her child when deciding about the neighbourhood later. Lemma 1 part A below shows that the information value of high effort is irrelevant for parents' effort decision and provides a threshold for beliefs that induces different effort choices on each side of it. Second, consider the parental neighbourhood choice. Denote by $P(\overline{e}|f^{t+1}_J)$ the probability that their children will choose $\overline{e}$ given that the distribution of beliefs at J next period is f^{t+1}_J for $J \in \{P, R\}$. Lemma 1 part B provides an expression for $\triangle^t_i$ and states that when parents believe that high-effort is the best choice for their offspring they will prefer to raise their child in the neighbourhood that makes this decision more likely.

Lemma 1: *A) (Optimal effort choice) For any $\delta \in [0, 1]$, parents will choose $\overline{e}$ whenever $q^t_{i,g} > 0.5$ and $\underline{e}$ otherwise.*

B) (Optimal neighbourhood choice) The i-parent's willingness to pay for living at neighbourhood R is given by the following expression:

$$\triangle^t_i = \{P(\overline{e}|f^{t+1}_R) - P(\overline{e}|f^{t+1}_P)\}\{(2\tau - 1)\pi(2q^t_{i,e} - 1)\}$$

[21] Given that $\overline{e} > 0$, (A1) implies $\overline{w}_H > \underline{w}$. On the other hand, $\overline{e} > 0$ does not impose restrictions on wages whenever $\theta = L$. Readers may keep in mind the natural assumption that $\overline{w}_L > \underline{w}$; but the analysis in the paper holds for $\overline{w}_L < \underline{w}$.

Moreover, whenever $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\} > 0$, *then* $\triangle_i^t > 0$ *if and only if* $q_{i,e}^t > 0.5$.

PROOF: (Part A). We first characterise the optimal parents' decision when $\delta = 0$ and then argue that she will never take a suboptimal decision to gather more information for future decisions. Whenever $\delta = 0$ the parent maximizes their expected payoffs given their beliefs:

$$\max_{e \in \{0, \overline{e}\}} E[w(\theta, e) - e|q_{g,i}^t]$$

In other words, they compare the expected payoff of choosing $\overline{e}$, $E[\overline{w}_\theta - \overline{e}|q_{i,g}^t]$, with the payoff of choosing $\underline{e} = 0$, $\underline{w}$. That is,

$$E[\overline{w}_\theta - \overline{e}|q_{i,g}^t] = P(\theta = H)[P(\overline{w}_H|\theta = H)(\overline{w}_H - \overline{e}) + P(\overline{w}_L|\theta = H)(\overline{w}_L - \overline{e})] + \\ P(\theta = L)[P(\overline{w}_L|\theta = L)(\overline{w}_L - \overline{e}) + P(\overline{w}_H|\theta = L)(\overline{w}_H - \overline{e})] \lesseqgtr \underline{w}$$

Using $P(\theta = H) = q_{i,g}^t$, $P(\overline{w}_k|\theta = k) = \tau$, $P(\overline{w}_l|\theta = k) = 1 - \tau$ for $l \neq k$, and substituting above, we have:

$$q_{i,g}^t[\tau(\overline{w}_H - \overline{e}) + (1 - \tau)(\overline{w}_L - \overline{e})] + (1 - q_{i,g}^t)[\tau(\overline{w}_L - \overline{e}) + (1 - \tau)(\overline{w}_H - \overline{e})] \lesseqgtr \underline{w}$$

Using (A2) in the above inequality and solving for $q_{i,g}^t$, we obtain the condition that relates beliefs to optimal effort:

$$q_{i,g}^t \lesseqgtr 0.5$$

Whenever the beliefs after childhood exceed a half, young adults prefer to choose high-effort, $\overline{e}$, and $\underline{e}$ when their beliefs are lower. Now, assume that $\delta > 0$. Notice that if $q_{i,g}^t > 0.5$, then there is no discrepancy between her optimal effort choice when she maximizes just $E[U^t(.)]$ ($\delta = 0$) and the information value of high-effort. Then, in this case, choosing $\overline{e}$ is the optimal decision also when $\delta > 0$ and $q_{i,g}^t > 0.5$. Next, consider the case where $q_{i,g}^t < 0.5$. Here the parent believes that the optimal effort is $\underline{e}$ absent any information value of $\overline{e}$ ($\delta = 0$). If the parent forecast that she will not be able to influence her offspring by choosing the neighbourhood (because of peers influence in both neighbourhoods at $t + 1$), then there is no information value motive and $\underline{e}$ is also the optimal decision when $\delta > 0$. Finally, assume that she forecast that she will be able to influence the child's effort decision by choosing the neighbourhood in which to raise the child. That is, she will be able to decide for her child by choosing the neighbourhood. Then, if the parent chooses $\underline{e}$ she receives a payoff of $\underline{w} + \delta\underline{w}$ (notice that she forecasts that the child will choose $\underline{e}$ at $t + 1$). On the other hand, if she chooses $\overline{e}$ she receives

$$E[\overline{w}_\theta - \overline{e}|q_{i,g}^t] + \delta E\{E[w(\theta, e^*) - e^*|q_{i,e}^t]|q_{i,g}^t\}$$

where $E[w(\theta, e^*) - e^*|q_{i,e}^t]$ is the expected payoff of choosing the optimal effort level, e^*, after receiving a signal in the labour market. Notice that by the law of iterated expectations, $E\{E[w(\theta, e^*) - e^*|q_{i,e}^t]|q_{i,g}^t\} = E[w(\theta, e^*) - e^*|q_{i,g}^t]$. That is, the expected optimal decision after labour market experience from the perspective of a parent before she receives the signal (when she holds beliefs $q_{i,g}^t$) must be the same as the expected optimal effort level before entering the labour market. Hence, as we are in the case where $q_{i,g}^t < 0.5$, she chooses $\underline{e} = 0$ and so $E\{E[w(\theta, e^*) - e^*|q_{i,e}^t]|q_{i,g}^t\} = \underline{w}$. Finally,

$$\underline{w} + \delta\underline{w} > E[\overline{w}_\theta - \overline{e}|q_{i,g}^t] + \delta\underline{w}$$

as $q_{i,g}^t < 0.5$ implies that $\underline{w} > E[\overline{w}_\theta - \overline{e}|q_{i,g}^t]$ by our discussion of the optimal decision when $\delta = 0$. Taking stock, we have proved that condition $q_{i,g}^t \lesseqgtr 0.5$ is the relevant threshold for beliefs to induce a different level of effort for any $\delta \in [0, 1]$.

(Part B) Consider now the parents' neighbourhood decision. Once parents from dynasty i have received their labour market experience, the decision of where to raise their children depends on how the expected beliefs of their child after childhood, $q_{i,g}^{t+1}$, in each neighbourhood will affect her effort decision given the parental belief $q_{i,e}^{t}$. Then, the expected payoff of raising the child at J is:

$$q_{i,e}\{P(\overline{e}|f_J^{t+1})[\tau(\overline{w}_H - \overline{e}) + (1-\tau)(\overline{w}_L - \overline{e})] + (1 - P(\overline{e}|f_J^{t+1}))\underline{w}\} +$$
$$(1 - q_{i,e})\{P(\overline{e}|f_J^{t+1})[\tau(\overline{w}_L - \overline{e}) + (1-\tau)(\overline{w}_H - \overline{e})] + (1 - P(\overline{e}|f_J^{t+1}))\underline{w}\}$$

Computing the same payoff for the other neighbourhood and taking the difference, the i-parent's willingness to pay for living at neighbourhood R, $\triangle_i^t$, is given by the expression in the statement of the Lemma. Then, whenever the probability that their offspring chooses high-effort in R is higher than the same probability in P, that is, when $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\} > 0$, parent i will prefer to live in R, $\triangle_i^t > 0$, if $\{(2\tau-1)\pi(2q_{i,e}^t - 1)\} > 0$. Finally, this last term is positive whenever $q_{i,e}^t > 0.5$, as $\tau > 0.5$ and $\pi > 0$, and negative otherwise. $\square$

The evolution of beliefs

In this part, we show how beliefs evolve after socialization and labour market experience, assuming that there is no mobility between neighbourhoods. Figure 3.1, panel a, illustrates how beliefs evolve over time.

Lemma 2: *(The evolution of beliefs) Whenever there is no mobility between neighbourhoods, for (initial) q_J:*

i) Beliefs decrease over time if $q_J < 0.5$.
ii) Beliefs increase over time if $q_J \in (L(\overline{e}), 1)$ where

$$L(\overline{e}) = \frac{\tau^{\frac{1}{n}}}{(1-\tau)^{\frac{1}{n}} + \tau^{\frac{1}{n}}},$$

and increase (resp. decrease) on average if $q_J \in [0.5, L(\overline{e})]$ and $\theta = H$ (resp. $\theta = L$)

PROOF: As the effort decision influences how much workers learn from labour market experience, we need to consider both cases separately. Consider first the case where dynasties choose $\underline{e}$. In this case they do not receive a signal about the state in the labour market. Beliefs evolve just by the socialization process in the neighbourhood. If all dynasties in the neighbourhood are choosing $\underline{e}$, by Lemma 1 (part A) we know that $q < 0.5$, and so beliefs decrease over time. To see why they decrease, notice that beliefs evolve using equation (2); so if all beliefs are lower than 0.5, after socialization beliefs at $t + 1$ should be even lower.

Consider now how beliefs evolve when dynasties choose high effort, $\overline{e}$. To ease exposition, we assume that the true state is $\theta = L$, and so if beliefs increase (decrease) over time we say that dynasties are *not learning* (resp. *learning*). We discuss below the case for $\theta = H$, which is similar. We are going to show that there is a fixed point of initial beliefs such that after socialization and labour market experience beliefs remains at the same level. Then, we will show that if beliefs start from a level different from this fixed point, they increases or decrease over time. Let us denote the initial beliefs $q = q_{fp} > 0.5$. When a young individual interacts with n others, her beliefs after childhood become by (2):

$$q_g(q_{fp}) = \frac{(q_{fp})^{n+1}}{(q_{fp})^{n+1} + (1-q_{fp})^{n+1}} > q_{fp} \tag{5}$$

where the inequality follows from $q_{fp} > 0.5$. Then assume that after choosing $\overline{e}$, individuals receive the most informative signal in the labour market, $\overline{w}_L$ (recall that we have assumed $\theta = L$). Her beliefs after the labour experience become q_e by (4). Hence, there are two forces that point in different directions: beliefs increase after socialization, as we start from $q = q_{fp} > 0.5$, but decrease after labour market experience (see (4)). In order to obtain a fixed point, denoted as $L(\overline{e})$, substitute $q_g(q_{fp})$ from (5) into (4), and equate this to the initial beliefs q_{fp}. That is,

$$q_e(q_{fp}) = \frac{q_g(q_{fp})(1-\tau)}{q_g(q_{fp})(1-\tau) + (1-q_g(q_{fp}))\tau} = q_{fp}$$

Finally, solving for q_{fp} we obtain the expression for the fixed point $L(\overline{e})(= q_{fp})$ in the statement of the Lemma.

Notice that the mapping from q to $\frac{q_g(q)(1-\tau)}{q_g(q)(1-\tau)+(1-q_g(q))\tau}$ has a unique fixed point in $(0,1)$. For initial beliefs $q > L(\overline{e})$, $\frac{q_g(q)(1-\tau)}{q_g(q)(1-\tau)+(1-q_g(q))\tau} > q$, and for $q < L(\overline{e})$, $\frac{q_g(q)(1-\tau)}{q_g(q)(1-\tau)+(1-q_g(q))\tau} < q$. Then, beliefs decrease to the left of $L(\overline{e})$ and increase otherwise.

$L(\overline{e})$ was obtained by considering dynasties that have received the most informative wage given that $\theta = L$, $\overline{w}_L$. Now we consider the dynasties that receive the "incorrect" signal in the labour market, $\overline{w}_H$. Notice that if beliefs increase for the former dynasties, the case where $q > L(\overline{e})$ and hence they are not-learning, it must also increase for individuals who have received $\overline{w}_H$. However, the opposite is not true. If we start from an initial belief $q < L(\overline{e})$, the dynasties considered for threshold $L(\overline{e})$ will keep their beliefs decreasing (learning) if they interact; but the dynasties that have received $\overline{w}_H$ could still have higher beliefs than the initial ones after the labour experience. Repeating the same arguments as for $L(\overline{e})$, we can find a another threshold such that if the initial q is lower than this threshold even dynasties that have received the wrong signal in the labour market keep their beliefs decreasing (learning) over time. Socialization is strong enough and outweighs the wrong labour market experience. Following the same arguments that we applied to $L(\overline{e})$, we can show that this learning threshold is lower than 0.5.[22] However, by Lemma 1 (part A), whenever $q < 0.5$ dynasties choose low effort, and so this threshold is not relevant.[23] Finally, we need to consider the case where $q \in [0.5, L(\overline{e})]$. As $\tau \in (0.5, 1)$, and there is independent interaction between dynasties in childhood, there are more dynasties learning, i.e decreasing their beliefs. To see why, notice that a mass of τ receives $\overline{w}_L$ and a mass of $(1-\tau)$ receives $\overline{w}_H$. Then, on the one hand, there is greater probability of receiving the correct signal, $\tau > 0.5$, and on the other, it is also more likely that the individual interacts with other dynasties who have received the informative signal.

[22] To do this, consider a group of dynasties that have received $\overline{w}_H$. Using q_g from (5) and then (3) (instead of (4)) we obtain:

$$l(\overline{e}) = \frac{(1-\tau)^{\frac{1}{n}}}{(1-\tau)^{\frac{1}{n}} + \tau^{\frac{1}{n}}}$$

Notice that $l(\overline{e}) < 0.5 < L(\overline{e})$.

[23] This is because we have assumed symmetry of payoffs between states in (A2). When payoffs are asymmetric, for example a higher $\overline{w}_H$, q could be close to zero and dynasties still choose $\overline{e}$. In this case, we would need to consider this threshold for the analysis.

Therefore, we have an "average" learning region in this case.

The analysis for the case where $\theta = H$ is similar. The expressions for $L(\overline{e})$ is the same. If beliefs start to the right of $L(\overline{e})$ then they increase. The difference is in the region $q \in [0.5, L(\overline{e})]$. In this case, as more dynasties receive $\overline{w}_H$ by $\tau > 0.5$, on average dynasties keep their beliefs increasing instead of decreasing over time. $\square$

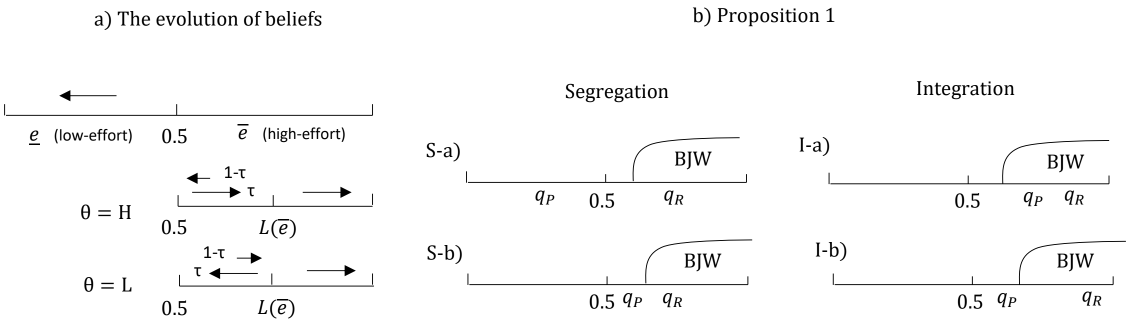

Figure 3.1

Segregated and integrated communities

Notice that if dynasties in one neighbourhood end up choosing a different level of effort respect to the dynasties from the other neighbourhood, then they will receive a different income level. We will say that there is *segregation* (resp. *integration*) if in the long run dynasties from each neighbourhood receive a different wage (resp. same wage on average). Now we present conditions for initial beliefs at R and P that generate segregated and integrated communities in the long run, and show that segregated communities exhibit polarized beliefs, higher income inequality, and lower social mobility.

We restrict the analysis to high-enough beliefs for dynasties at R. This is because our empirical analysis is for the U.S., which exhibits a widely held belief in the American Dream as documented by previous work, cited in footnote 4. Let us define first the Strong Beliefs in a Just World set, denoted by BJW(θ). Whenever initial belief q lies in BJW(θ), *all* dynasties in the neighbourhood choose high-effort and beliefs increase over time on average towards a just world, i.e. towards $\theta = H$ (effort is worthwhile). The set is defined formally as:

$$\text{BJW}(\theta) = \{q \mid q \in [0.5, 1] \text{ if } \theta = H; \text{ or } q \in [L(\overline{e}), 1] \text{ if } \theta = L\}$$

Recall that dynasties choose high-effort if $q > 0.5$ (by Lemma 1). Then, by Lemma 2, whenever $\theta = H$ and dynasties choose $\overline{e}$ ($q > 0.5$), the majority of dynasties have their beliefs increasing over time if $q \in [0.5, 1]$. Therefore, this is the relevant set for strong beliefs in a just world when $\theta = H$. On the other hand, when $\theta = L$, beliefs needs to be higher as the learning threshold is $L(\overline{e}) > 0.5$. Hence, by Lemma 2, the relevant interval is $q \in [L(\overline{e}), 1]$ and the majority of dynasties will have increasing beliefs over time. Notice that whenever initial beliefs $q \notin \text{BJW}(\theta)$, dynasties may choose high or low effort, but beliefs in a just world decreases over time on average if there is no mobility.

The next proposition provides conditions for initial beliefs at P and R, q_P and q_R respectively, that determine the two types of long-run equilibria of the model: segregation or integration.

Proposition 1: *A) Whenever $n = 0$, there is segregation if and only if $q_P < 0.5 < q_R$ and $\theta = H$. Otherwise, there is integration.*

B) For a given θ and a high enough n and ρ, if $q_P < q_R$ and $q_R \in BJW(\theta)$, then there are two types of equilibria in the long run:

i) Segregation and polarization. In all equilibria there is no mobility between neighbourhoods, the dynasties living in neighbourhood P end with beliefs $q^\infty = 0$ and choose low effort, whereas dynasties at R end with beliefs $q^\infty = 1$ and choose high effort if:

S-a) $q_P \notin BJW(\theta)$ and $q_P < 0.5$, or
S-b) $q_P \notin BJW(\theta)$, $q_P > 0.5$, and q_R low enough.

ii) Integration and no-polarization. In all equilibria there is mobility between neighbourhoods, all the dynasties end with beliefs $q^\infty = 1$ and choose high effort if:

I-a) $q_P \in BJW(\theta)$, or
I-b) $q_P \notin BJW(\theta)$, $q_P > 0.5$, and q_R high enough.

PROOF: See the appendix for the proof of part A) and cases *S-b* and *I-b* of part B).

Let us consider first case *S-a*. As $q_P < 0.5$, by Lemma 1 (part A), all dynasties in P start choosing $\underline{e}$. Moreover, also because of $q_P < 0.5$ and Lemma 2, beliefs decrease over time if there is no mixing with dynasties in R. Then, notice that as $q_R \in \mathrm{BJW}(\theta)$, dynasties in R will choose $\overline{e}$ and beliefs increase over time. Hence, $\triangle_i^t > 0$ for all parents in R, they believe in high effort; and $\triangle_i^t < 0$ for all parents in P, they do not believe that effort pays and do not want to move to R (where the neighbourhood may influences their child towards exerting high-effort). This implies that beliefs will converge to opposite values in each neighbourhood in the long run, and dynasties in each neighbourhood receive a different wage. Therefore, we have a segregated community with polarized beliefs. Now consider case *I-a*. As $q_P, q_R \in \mathrm{BJW}(\theta)$, all dynasties in P and R choose $\overline{e}$ and beliefs increase over time. This implies that beliefs will converge to one in both neighbourhoods, the community will be integrated with both neighbourhood receiving a similar wage and having strong beliefs in a just world. $\square$

Now we provide the intuition for the proof of cases *S-b* and *I-b*. In these cases, as $q_P > 0.5$, dynasties in P start choosing high-effort. Moreover, as $q_P \notin \mathrm{BJW}(\theta)$, their beliefs are decreasing over time on average. We will show that for high enough q_R only integrated communities exist in equilibrium, some parents from P will move to R and vice versa until there is integration in the long run (case *I-b*); whereas for low enough q_R only segregated communities exist, where there is no neighbourhood mobility (case *S-b*). Denote by $\triangle_{i,J}^t$ the willingness to pay for living in R for a dynasty i living in J. Notice that as dynasties in P start choosing $\overline{e}$, then $\triangle_{i,P}^t \geq 0$ at $t = 1$. $\triangle_{i,P} = 0$ whenever they infer that the child will choose $\overline{e}$ in both neighbourhoods, and $\triangle_{i,P} > 0$ when high-effort will be more likely chosen in R. However, as beliefs are decreasing in P, for no segregation P's dynasties must mix with R's ones before they end with beliefs that make $\triangle_{i,P}^t < 0$

and trap these dynasties. The difficulty for integration is that if $q_R < 1$, dynasties in R also have $\triangle_{i,R}^t \geq 0$, and thus we need to consider which parents have a higher willingness to pay to live in R, i.e. $\triangle_{i,R}^t - \triangle_{j,P}^t$. Notice that whenever $\triangle_{i,R}^t - \triangle_{j,P}^t < 0$ for some j and i, dynasties will mix as some parents from P will move to R. Let us consider the expression $\triangle_i^t$ given in Lemma 1 (part B) in more detail:

$$\triangle_i^t = \underbrace{\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\}}_{\text{Decreasing in q}}\underbrace{\{(2\tau - 1)\pi(2q_{i,e}^t - 1)\}}_{\text{Increasing in q}}$$

The terms in curly brackets move in opposite directions as beliefs increase. First, the term $(2\tau - 1)\pi(2q_{i,e}^t - 1)$ is related to the expected payoff of choosing $\overline{e}$ rather than $\underline{e}$. This expression is increasing in beliefs, $q_{i,e}^t$. Dynasties with higher beliefs that $\theta = H$ have higher expected payoffs from choosing $\overline{e}$, and hence will prefer their offspring to choose $\overline{e}$ more strongly than a parent with lower beliefs. As we have assumed that $q_R > q_P$, this term is always higher for a parent living in R than for a parent living in P. Second, the expression $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\}$ is related with the expected neighbourhood's "contagion" effect on the child. Recall that $P(\overline{e}|f_J^{t+1})$ is the expected probability that the offspring chooses $\overline{e}$ in neighbourhood J, given the correct expectations of beliefs in neighbourhood J in the next period, f_J^{t+1}. This term is decreasing in $q_{i,e}^t$ whenever $q > 0.5$. To see why, notice that the "contagion" in each neighbourhood is given by the function $C(\mathbf{q}^t, q_{i,e}^t)$ in (2). This function is increasing in $q_{i,e}^t$ (the prior) and converge to one as $q_{i,e}^t$ increases. Moreover, it is concave in $q_{i,e}^t$ for $\mathbf{q}^t > 0.5$, where $\mathbf{q}$ is the mean (expected) of others' beliefs in that neighbourhood. Then, by concavity, when $q_{i,e}^t$ is close to 1, others' belief will not have much influence on the child in any neighbourhood, i.e. $C(\mathbf{q}^t, q_{i,e}^t)$ will not change much by changing the neighbourhood. The concavity of the contagion function implies that $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\}$ will be close to zero for high enough beliefs. On the other hand, when $q_{i,e}^t$ is not extreme, others' beliefs will have a greater impact. Therefore, for different $\mathbf{q}^t$ in each neighbourhood, parents with low enough beliefs will have a higher $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\}$ than a parent with high enough beliefs. Therefore, this term is decreasing in $q_{i,e}^t$. See Figure 3.2.

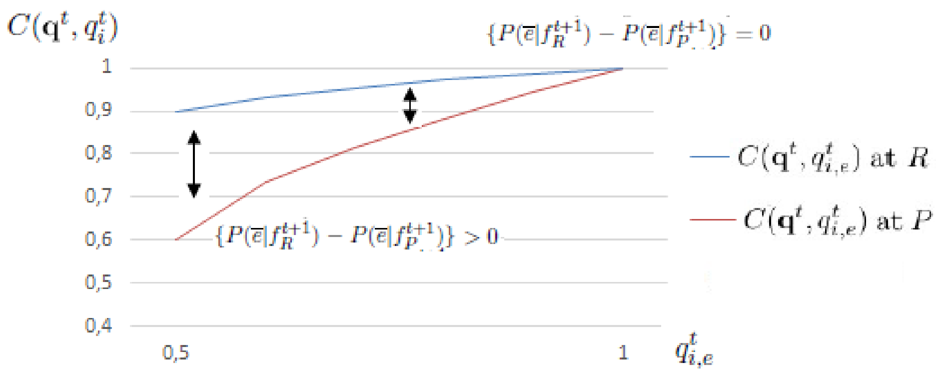

Figure 3.2

Given the discussion above, for high enough beliefs of parents in R, $\triangle_i^t$ would be close to zero as $\{P(\overline{e}|f_R^{t+1}) - P(\overline{e}|f_P^{t+1})\}$ is close to zero by concavity of $C(\mathbf{q}^t, q_{i,e}^t)$. (Notice that $q_R \in \mathrm{BJW}(\theta)$ implies $q_R > 0.5$ from which concavity follows.) This would allow that dynasties in P with $\triangle_{i,P} > 0$ move to R, and that R's dynasties with high enough beliefs move to P and "contaminate" both neighbourhoods. This is the argument for case *I-b* on integrated communities that we provide in the appendix. For the case *S-b*, the argument is the opposite. If R's parents have low enough beliefs, they will have higher $\triangle_i^t$ than parents in P because $(2\tau-1)\pi(2q_{i,e}^t-1)$, which is increasing in q, dominates. Then, as R-dynasties do not want to mix with P's ones, the latter will remain in P and end up with low beliefs in effort.

Social mobility, income inequality and segregation. Notice that there are two sources of inequality in the model. First, the one that arises from luck/bad-luck in the labour market for high-effort workers. That is, when individuals choose $\overline{e}$ agents can receive two possible incomes, $\overline{w}_H$ and $\overline{w}_L$. Second, inequalities that arise from the effort decision (voluntary choice). In this sense, integrated communities only experience intertemporal inequalities based on temporary luck, but all dynasties exert the same level of effort. Low-income dynasties in a cohort, $\overline{w}_L$, can become high-income ones, $\overline{w}_H$, in the next period. Therefore, social mobility is high in integrated communities. On the other hand, segregated communities experience both sources of inequality between dynasties. The correlation between parents' and children's income is higher than in the previous case. In P, low-effort children have the same income that their parents, whereas dynasties living in R have the same mobility as in the integrated communities that we have discussed. From Proposition 1 and this observation, we have the following corollary:

Corollary 1: *Segregated communities exhibit lower rates of social mobility and high income inequality, whereas non segregated ones exhibit higher rates of social mobility and low income inequality.*

Parent-child belief transmission. Notice that in case *S-b*, parents in R would not want their children to mix with children in P, whereas in case *I-b* they do not care. In this sense the model captures the idea that the difference between integrated and segregated communities is that in the latter ones R's dynasties start with relatively low initial beliefs in effort and hence are worried about the effects of mixing with P's dynasties because of the potential effects on their offspring. We have assumed that parents transmit their beliefs at no cost, and also that they do not interfere with their child's socialization. An alternative way to interpret this result is that segregated communities arise when R's parents with a relatively high cost to transmit their beliefs to their child, perhaps because they must work for longer hours and spend more time outside home. This is a plausible explanation for the results of Rothstein (2019), who finds that the differences between low and high social mobility communities is mostly explained by whether both parents work in the community. The higher the share of high-income households where both parents work, the lower the mobility. Our results based on beliefs offer an explanation for this: if households with a higher cost of transmitting beliefs in effort do not want to mix with low beliefs ones, then segregation is more likely and mobility lower in the long run.

Income restrictions. Proposition 1 is a lower bound for segregation. The assumption that whenever some dynasty has $\triangle_i^t > 0$, and this is higher than the same payoff for other dynasties, she can move to R is restrictive. Consider the opposite extreme case: dynasties cannot move to the other neighbourhood despite of having $\triangle_i^t > 0$ higher than some dynasties in R. It can be seen from the proof of Proposition 1 that this is not relevant for cases where there is segregation (cases *S-a*, *S-b*) as there is no mobility between dynasties; and also for case *I-a*, where mobility is not relevant for integration. Income restrictions affects case *I-b*. Here P-dynasties would like to raise their children in R but cannot do it because of income restrictions. This last case implies that if there is no public policy towards helping dynasties in P to move on early stages, segregation may arise because of income restrictions.

Integration and optimism/pessimism. Proposition 1 provides conditions under which dynasties from integrated communities end up in the long run with optimistic beliefs about effort, $q^{\infty} = 1$. However, there may be cases where integrated communities end up with low beliefs in effort. The relevant condition for the link between integration and optimism is a high enough ρ. To see why, consider a low enough ρ, that is, a small original proportion of high-beliefs dynasties in the community (recall that ρ refers to the size of neighbourhood R or, equivalently, the share of original dynasties with high beliefs in effort given that $q_R > q_P$). In this case, if dynasties move from R to P (and vice versa), then original R's dynasties cannot fully "contaminate" original P's dynasties (as in case I-b) before the latter ones end up with very low beliefs about effort. Because ρ is low there are not enough R-dynasties to influence P's ones. Therefore, low-beliefs dynasties (those who start at P) will influence high-beliefs dynasties by socialization through time, and thus the community may end up integrated and with pessimistic views about the value of effort.

3.4 Empirical analysis

In this section we present our data, empirical strategy and results.

3.4.1 Data

We use the survey-data from Alesina et al. (2018) that includes beliefs in a just world for U.S., and also community characteristics, in particular, the income segregation of the community. We first describe our measures for beliefs in a just world. We use the answers to the four questions described in Table 3.1.[24]

Table 3.1: Beliefs in a just world measures

Variable Name	Description
(AD) American Dream	Discrete variable from 1 to 5 (Strongly disagrees (1)- to Strongly agree (5)) if respondent disagrees or agrees with the statement "In U.S. everybody has a chance to make it and be economically successful".
(EP) Effort Reason Poor	Dummy equal to one if respondent believes that "Lack of effort on his or her own part" is a more important determinant of why a person is poor than "Circumstances beyond his or her control".
(ER) Effort Reason Rich	Dummy equal to one if respondent believes that "Because she or he worked harder than others" is a more important determinant of why a person is rich than "Because she or he had more advantages than others".
(F) Economic System Fair	Dummy equal to one if respondent believes that the economic system in U.S. is basically fair, since all have an equal opportunity to succeed.

Second, we use community and individual variables. Community variables in the database are computed at the Commuting Zone (CZ) spatial unit. CZs are aggregations of counties based on commuting patterns in the 1990 Census. Since CZs are designed to span the area in which people live and work, they provide a natural counterpart for the community in our theoretical model.[25]

[24] We choose to present the result for each measure because although the correlation between them is high, it is far from perfect. We report the correlation matrix in the appendix.

[25] CZs are similar to metropolitan statistical areas (MSA), but unlike MSAs, they cover the entire U.S., including rural areas. In our data 274 of the 380 CZs are MSAs.

There are 741 CZs in the U.S.; on average, each CZ contains 4 counties and has a population of 380,000. This is our measure for the community. In our final dataset where we have complete data for control variables there are 380 CZ, that have an average population of 684,748. There is a mean of 20 respondents in the CZs. Table 3.2 presents descriptive statistics of the variables used for the empirical analysis.

Table 3.2: Summary statistics

Variable	N	Mean	S.D	Min	Max
Dependent var					
AM	3311	3.29	1.14	1.00	5.00
EP	3311	0.46	0.50	0.00	1.00
ER	3311	0.41	0.49	0.00	1.00
F	3311	0.52	0.50	0.00	1.00
AM-Pol	172	3.29	1.14	1.00	5.00
EP-Pol	172	0.46	0.50	0.00	1.00
ER-Pol	172	0.41	0.49	0.00	1.00
F-Pol	172	0.52	0.50	0.00	1.00
Community (CZ) var					
(Income) Segregation	3311	0.09	0.03	0.01	0.14
Social Capital Index	3311	-0.41	0.90	-2.69	2.76
Gini	3311	0.48	0.08	0.26	0.85
Racial Segregation	3311	0.24	0.11	0.01	0.47
College Grad Rate	3311	-0.02	0.11	-0.35	0.47
Manufacturing Share	3311	0.14	0.06	0.02	0.42
Median Household Income (USD 2016)	3311	61,506	13,283	26,645	103,043
Individual var					
Male	3311	0.49	0.50	0.00	1.00
Age	3311	42.79	14.57	18.00	69.00
Immigrant	3311	0.15	0.36	0.00	1.00
Number of children	3311	1.22	1.31	0.00	5.00
Income	3311	6.62	2.91	1.00	12.00
Moved-up	3311	0.42	0.49	0.00	1.00
Live in R	3311	0.47	0.50	0.00	1.00

3.4.2 Empirical strategy

Our data allow us to contrast three hypothesis implied by Proposition 1. We state each hypothesis and describe our empirical strategy for each below.

Segregation and polarization. Notice first that Proposition 1 shows that segregated communities exhibit polarized beliefs between neighbourhoods P and R, whereas integrated communities do not. Then,

H1) *Segregated communities exhibit more polarized beliefs in a just world than integrated ones.*

In order to test H1 we construct measures of belief polarization for each CZ in the database where we have more than four respondents in the following way. Notice that a household with high income is more likely to live in a high-income neighbourhood in a segregated community than in a low-income neighbourhood. Therefore, we assume that respondent i lives in R if her household income is higher than the median of the CZ's income, and lives in P if it is lower. Then, for each measure of beliefs Y$\in$ {AD,EP,ER,F} we construct beliefs polarization at each CZ, denoted by Y-Pol, by taking the mean of each measure Y in each neighbourhood (i.e. for individuals above and below the CZ's income median) and taking the difference. Formally, beliefs polarization in each CZ is computed as:

$$\text{Y-Pol}_{\text{cz}} = \left| \frac{\sum_i^{N_R} Y_i}{N_R} - \frac{\sum_i^{N_P} Y_i}{N_P} \right|$$

where $|\ .\ |$ denotes absolute value, and N_J the number of respondents living at $J \in \{P, R\}$ for the corresponding CZ. Table 3.2 provides summary statistics for each Y-Pol measure. Then we estimate the following equation for each Y-Pol:

$$\text{Y-Pol}_{\text{cz}} = \alpha + \beta \text{Segregation}_{\text{cz}} + \delta \mathbf{Z}_{\text{cz}} + \epsilon_{\text{cz}} \tag{6}$$

where $\mathbf{Z}_{\text{cz}}$ is a vector of CZ's control variables. We use the same CZ's controls as in Chetty et al. (2014); Chetty and Hendren (2018b); Alesina et al. (2018), that are shown in Table 3.2. The coefficient of interest is β: if β is positive, this provides supports for H1, as more segregated communities are correlated with a higher level of belief polarization.

Segregation and pessimism. Proposition 1 shows that segregated communities exhibit on average lower beliefs in a just world than integrated communities, as dynasties in P end up with low beliefs in segregated communities but high beliefs in integrated ones. Then,

H2) *Segregated communities exhibit more pessimistic beliefs in a just world than integrated ones.*

To test this hypothesis we estimate the following equation at the individual level for each Y:

$$\text{Y}_i = \alpha + \beta \text{Segregation}_{\text{cz}} + \gamma \mathbf{X}_i + \delta \mathbf{Z}_{\text{cz}} + \epsilon_i \tag{7}$$

where $\mathbf{Z}_{\text{cz}}$ are the same CZ control variables that we use for (12), and $\mathbf{X}_i$ is a vector of individual characteristics described at Table 3.2. The coefficient of interest is β: if β is negative, this provides supports for H2, as more segregated communities are correlated with lower beliefs in a just world on average.

Finally, notice that comparing cases *S-b* with *I-b* in Proposition 1, segregated communities exhibit lower beliefs for dynasties living in R than integrated ones. This is because we need low enough initial beliefs at R when segregation arises, and high enough q_R when integration does. Then,

H3) *For dynasties living in R, segregated communities exhibit more pessimistic beliefs in a just world than integrated ones.*

To test this hypothesis we add to equation (12) an interaction variable between segregation and households living in R, where the variable (Live in R) is a dummy variable that takes value one if the household income of the respondent is higher than the median of the CZ's income. That is, we estimate the following equation at the individual level for each Y:

$$\mathrm{Y}_i = \alpha + \beta_1 \mathrm{Segregation}_{\mathrm{cz}} + \beta_2 (\mathrm{Segregation}_{\mathrm{cz}})(\text{Live in } R)_i + \beta_3 (\text{Live in } R)_i + \gamma \mathbf{X}_i + \delta \mathbf{Z}_{\mathrm{cz}} + \epsilon_i \tag{8}$$

where $\mathbf{Z}_{\mathrm{cz}}$ and $\mathbf{X}_i$ are the same control variables that we use for (12). The coefficient of interest is β_2: if β_2 is negative, this provides supports for H3, as more segregated communities are correlated with lower beliefs in a just world for households living in R.

3.4.3 Results

We report the estimation of (11), segregation and polarization, in Table 3.3. The OLS estimation results in a positive and significant correlation between segregation and polarization when we use AD and ER as beliefs in a just world (columns 1 and 3). Positive but insignificant for F (column 4), and negative and significant when we use EP (column 2). These results are robust when we add the poor share, the total population and the population density for each CZ as additional control variables, and when we consider CZ with 8 beliefs' observation (the median in the database is 9 respondents per CZ). Notice that the belief measure with more variation is AD, 1 to 5, whereas the others are dummy variables, so we consider that AD is a better measure to test polarization. Although our data has limitations to test the link between segregation and polarization, we see these results as providing some support for it.[26]

Finally, we report the results for the link between segregation and pessimism in Table 3.4. We report the OLS estimation but the same results are obtained if we use logit and probit models. The first four columns show the OLS estimation of equation (12). The correlation between segregation and beliefs is negative for all measures of beliefs and significant for three of them (columns 1,3 and 4). This suggests that segregated communities are more pessimist on average than integrated ones regarding beliefs in a just world. This provides support for H2. The last four columns report the OLS estimation of equation (13). The coefficient of Segregation*(Live in R) is negative and significant for the four measures of beliefs. This provides supports for H3. Dynasties that live in high-income neighbourhoods are more pessimistic in segregated communities than in integrated ones. Notice two other interesting results of these estimations. First, the coefficient of household income is positive and significant in the eight columns. This is in line with the theoretical model assumption that a household that receives a high income believes more in high effort. Second, for those that lives in R, the coefficient is also positive and significant. This suggests that socialization also plays a role, as after controlling for income, the households that lives in high-income neighbourhoods have more optimistic beliefs.

3.5 Conclusion

We have provided a theoretical framework to understand the observed link between segregation and social mobility. Based on a theory of learning about the value of effort, or beliefs in a just world, in the labour market and neighbourhoods we show that dynasties in segregated communities learn differently than dynasties in integrated communities. In particular, we show that segregated

[26] Notice that we do not know where the individual actually lives. We proxy this by assigning each observation to a low-income or high-income neighbourhood based on whether the household income is above or below the CZ's median income.

Table 3.3: Segregation and Belief Polarization

	(1) AD-Pol	(2) EP–Pol	(3) ER–Pol	(4) F-Pol
Income Segregation	6.352**	-1.934*	1.976*	0.374
	(2.452)	(1.034)	(1.182)	(1.181)
Social Capital Index	-0.071	0.010	-0.027	0.043
	(0.071)	(0.029)	(0.033)	(0.030)
Racial Segregation	-1.286**	-0.190	-0.200	-0.575**
	(0.599)	(0.245)	(0.272)	(0.248)
College Grad Rate	0.552	0.394**	0.254	0.350*
	(0.462)	(0.164)	(0.223)	(0.199)
Manufacturing Share	1.440	-0.569**	0.185	-0.506
	(0.883)	(0.278)	(0.388)	(0.361)
Median HH Income	-0.000***	-0.000	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gini	-1.570	-0.074	-1.051**	0.374
	(0.996)	(0.410)	(0.459)	(0.460)
Observations	172	172	172	172
R^2	0.139	0.188	0.091	0.103

Standard errors in parentheses

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

communities exhibit more polarized and hence pessimistic beliefs on average than integrated communities. Using survey data from the U.S. we find empirical support for these results. Moreover, we provide an explanation for the mechanism that prevents integration in communities: parents with relatively low beliefs in effort in high-income neighbourhoods do not want to mix with low-belief parents in different neighbourhoods as they are concerned about the influence of these low-belief families on their child. In line with this explanation, we find evidence that households from high-income neighbourhoods have more pessimistic beliefs in segregated communities than in integrated ones. Our results also provide an explanation for the recent evidence that points to working parents as the relevant channel to explain different rates of social mobility across U.S. regions. We argue that communities with more high-income parents with a higher cost to transmit their beliefs to their children due to longer working hours outside home amplify the segregation mechanism described, which in turn affects social mobility rates.

Table 3.4: Segregation and Pessimism

	(1) AD	(2) EP	(3) ER	(4) F	(5) AD	(6) EP	(7) ER	(8) F
Segregation	-2.075**	-0.738*	-0.740*	-0.115	-0.217	0.057	-0.193	0.658
	(0.939)	(0.411)	(0.416)	(0.383)	(1.107)	(0.490)	(0.519)	(0.464)
Income	0.049***	0.026***	0.026***	0.026***	0.028**	0.016***	0.014**	0.018***
	(0.009)	(0.004)	(0.004)	(0.003)	(0.013)	(0.005)	(0.006)	(0.005)
Seg*(Live in R)					-3.083**	-1.244**	-0.638	-1.349**
					(1.296)	(0.580)	(0.607)	(0.529)
Live in R					0.431***	0.190***	0.146**	0.175***
					(0.137)	(0.059)	(0.061)	(0.055)
Observations	3311	3311	3311	3311	3311	3311	3311	3311
R^2	0.043	0.049	0.041	0.058	0.047	0.053	0.044	0.061
Indiv. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CZ controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CZ obs	380	380	380	380	380	380	380	380

Standard errors in parentheses, clustered at the commuting zone.
* p<0.1, ** p<0.05, *** p<0.01

3.6 Appendix

Proof of Proposition 0: The proof follows the same argument as in Levy and Razin (2017), where we just adapt the beliefs updating process for a given t to our model extension. Notice first that we can always compute $\triangle_i^t$ for each dynasty using the beliefs updating process described in our equilibrium definition. That is, for each t, each dynasty update their (initial stage) beliefs using (2), chooses her optimal effort level by maximizing (1), and then, if high-effort is decided, they update their beliefs after the labour market experience using (3) and (4). Second, we will show that for any t it is possible to find a measure ρ of dynasties with the highest $\triangle_i^t$ living at R. At any period t, construct the self correspondence $\Omega : [0,1] \to 2^{[0,1]}$ in the following way. For any set of beliefs S of measure ρ that is composed of intervals in $[0,1]$, the correspondence assigns a measure ρ of dynasties that have the highest $\triangle_i^t$, given that the set S lives in R. To complete the description of the correspondence, assign any individual who is indifferent between living in R or P to R. For any S, $\Omega(S)$ is never empty. Thus, this correspondence is closed. Moreover, the set $[0,1]$ is compact. By Ok (2004), the set $[0,1]$ has the fixed set property so that every closed self correspondence has a fixed set. Therefore, we can always construct for $[0,1]$ a sequence of fixed set for $t = 1, 2, ...$ in which $\triangle_i^t$ is computed as described and where a measure ρ of dynasties with the highest $\triangle_i^t$ lives in R. Hence, an equilibrium exists. $\square$

Proof of Proposition 1: Consider part (A). (Sufficiency) As $q_P < 0.5$, by Lemma 1 P-dynasties choose $\underline{e}$. Then, because $n = 0$ and they do not receive a signal, their beliefs remain constant over time. On the other hand, as $q_R > 0.5$ and by Lemma 1, R-dynasties choose $\overline{e}$ and so they receive an informative signal. Whenever $\theta = H$, belief increase over time. Therefore, there is segregation in the long run.

(Necessity) Suppose $q_R < 0.5$. Then, by Lemma 1 and $n = 0$, dynasties in each neighbourhood choose low effort and so they have their beliefs constant over time. Hence, there is no segregation if $q_R < 0.5$. Next, suppose $q_P > 0.5$. Then, in both neighbourhoods dynasties receive informative signals and so they end with the same decision regarding effort. Either they learn the truth whenever $\theta = H$ as beliefs increase over time, or, whenever $\theta = L$, their beliefs becomes lower and eventually enter the low effort region. Hence, there is no segregation if $q_P > 0.5$. Finally, suppose $q_P < 0.5 < q_R$ and $\theta = L$. Then, dynasties in R receive informative signals, and so their beliefs decrease over time and eventually end up choosing low effort. Hence, there is no segregation if $q_P < 0.5 < q_R$ and $\theta = L$.

(Part B) We denote by $\triangle^t_{i,J}$ the willingness to pay for living in R for a dynasty i living in J. Cases *S-a* and *I-a* were proved in the text. Consider cases *S-b* and *I-b*. Notice first that if $\theta = H$ then we cannot have $q_P > 0.5$ and $q_P \notin$ BJW(θ): for if $\theta = H$ then BJW(θ)=[0.5,1]. This is an implication of the assumed symmetry between the "correct" net payoff in each state at (A2). By allowing for asymmetry both cases exist for $\theta = H$ and the argument is analogous to the one that we state below.

Consider case *I-b*. Notice that when $q_R = 1$, the willingness to pay for living in R for these dynasties is zero, $\triangle^t_{i,R} = 0$, as these dynasties are not influenced by other's beliefs. Therefore, we can always pick some high enough initial beliefs q_R such that without mobility of dynasties $\{P(\bar{e}|f^{t+1}_R) - P(\bar{e}|f^{t+1}_P)\}$ at t is close to zero, from which the willingness to pay for living at R, $\triangle^t_i$, is close to zero at $t = 1$. On the other hand, as $q_P \notin$ BJW(θ), initial beliefs between neighbourhoods are far away from each other. Then, $\triangle^t_{i,R} - \triangle^t_{j,P} < 0$ at $t = 1$ and parents from P will move to R allowing for "contagion". At this point we need to construct distributions of beliefs in each neighbourhood at $t = 2$, $f^{t=2}_J$, that are consistent with equilibrium decisions. We do it in the following way. We first move dynasties at R that replace P's parents with lower beliefs, the ones that receive a signal towards $\theta = L$ in the labor market, as these P dynasties are the ones that have higher willingness to escape from P. Notice that as we move high-belief dynasties from R to P, $\{P(\bar{e}|f^{t=2}_R) - P(\bar{e}|f^{t=2}_P)\}$ decreases, and so $\triangle^t_i$ also does. Then, we move dynasties until we have the dynasties with the highest willingness to pay for R living in R at $t = 2$. As we are considering initial beliefs for R-dynasties which are close to one, we have a positive measure of high-belief parents that have moved to P (recall that we have assumed a high enough ρ). Notice from the belief updating function $C(\mathbf{q}^t, q^t_{i,e})$ (given in (2)) that confident individuals are persuasive. (For example, if an individual has extreme beliefs at one or zero, then he fully convinces all others.) Therefore, this positive measure of high-belief parents will increase beliefs in each neighbourhood by socialization at $t = 2$ and ensures that the expected beliefs in each neighbourhood reaches the BJW(θ) set eventually, from which beliefs increase over time for all dynasties after that. This proves integration.

Consider now case *S-b*. As $\theta = L$, BJW(θ)=[$L(\bar{e})$,1]. Notice that as n increases $L(\bar{e})$ decreases towards 0.5. Then, pick low enough $L(\bar{e})$ such as $L(\bar{e}) = 0.5 + \epsilon$ and a low initial q_R such that $q_R = L(\bar{e}) + \epsilon$. Given these choices and $q_P \notin$ BJW(θ), q_P must lie in (0.5,0.5+ϵ). Then, as $q_P \notin$ BJW(θ), beliefs in P are decreasing over time and because they are close to 0.5, at $t = 2$ all children in P will prefer $\underline{e}$ after socialization. This implies that $P(\bar{e}|f^{t=2}_P) = 0$ for parents living in P if there is no movement between neighbourhoods. However, as we pick q_R close to q_P, parents in R will also have $P(\bar{e}|f^{t=2}_P) = 0$. Moreover, without movement between neighbourhoods, $P(\bar{e}|f^{t=2}_R) = 1$ for every parent in R and P as by $q_R \in$ BJW(θ) every dynasty is choosing high-effort at R and beliefs increase over time. Then, we have

$$\triangle^{t=2}_i = \{P(\bar{e}|f^{t=2}_R) - P(\bar{e}|f^{t=2}_P)\}\{(2\tau - 1)\pi(2q^{t=2}_{i,e} - 1)\} = (2\tau - 1)\pi(2q^{t=2}_{i,e} - 1)$$

for every parent. Notice that this is increasing in $q_{i,e}$ and so $\triangle^t_{i,R} - \triangle^t_{j,P} > 0$ at $t = 2$. Therefore, there cannot be integration as parents from R will refuse to move to P at $t = 1$. Finally, at $t = 2$ the majority of dynasties in P will prefer to choose low effort and to live in P as they do not believe in high-effort as R's parents do. This proves segregation.

Finally, we now prove that there are no other type of equilibrium. As we are considering a high enough n, $L(\overline{e})$ is low and so, as $0.5 < q_P < L(\overline{e})$ (recall that $q_P \notin \mathrm{BJW}(\theta)$), q_P must also be low. Then, suppose q_R is close to one. Case *I-b* considers this case. Here $\triangle^t_{i,R} = 0$ because $P(\overline{e}|f^t_P) = 1$ for R-dynasties, and so $\{P(\overline{e}|f^t_R) - P(\overline{e}|f^t_P)\} = 0$. This implies $\triangle^t_{i,R} - \triangle^t_{j,P} < 0$, which allows integration. Now, let us consider how $\triangle^t_{i,R} - \triangle^t_{j,P}$ moves as q_R decreases for a fix q_P. We will show that there must be segregation eventually, which proves that these are the two possible equilibria. Notice first that as q_R decreases $\triangle^t_{i,R}$ will increase eventually because, at some point, R-dynasties will have $P(\overline{e}|f^t_P) < 1$ which implies that $1 > \{P(\overline{e}|f^t_R) - P(\overline{e}|f^t_P)\} > 0$. That is, R-dynasties' children may choose low effort if their parents move to P. (Notice that $P(\overline{e}|f^t_R) = 1$ for R-dynasties because $q_R \in \mathrm{BJW}(\theta)$.) Moreover, as q_R decreases P-dynasties will also have a lower $\triangle^t_{i,P}$, as $P(\overline{e}|f^t_R)$ decreases for them. Then, there must exists a $\hat{q}_R$ such that $\triangle^t_{i,R} - \triangle^t_{j,P} = 0$. Below that level, $\triangle^t_{i,R} - \triangle^t_{j,P} > 0$: parents from R have a higher willingness to pay for living in R. Hence, dynasties from P cannot move to R, their beliefs decrease through time and enter the low-effort region eventually. (Case *S-b* considers the extreme case where $P(\overline{e}|f^t_P) = 0$ for R-dynasties from the start, but this is not necessary.) $\square$

Description of variables

CZ variables are included in the database of Alesina et al. (2018) and come from https://www.opportunityinsights.org. The database of Alesina et al. (2018) is available at https://www.aeaweb.org/articles?id=10.1257/aer.20162015.

(Income) Segregation: Rank-Order index estimated at the census-tract level using equation (13) in Reardon (2011); the δ vector is given in Appendix A4 of Reardon's paper. H(p_k) is computed for each of the income brackets given in the 2000 census. See Appendix D for further details. Original source: 2000 Census.
Social Capital: Standardized index combining measures of voter turnout rates, the fraction of people who return their census forms, and measures of participation in community organizations. Original source: Rupasingha and Goetz (2008).
Gini: Gini coefficient computed using parents of children in the core sample, with income topcoded at $100 million in 2012 dollars. Original source: Tax Records.
Racial Segregation: Multi-group Theil Index calculated at the census-tract level over four groups: White alone, Black alone, Hispanic, and Other. Original source: 2000 Census.
College Graduation Rate: Residual from a regression of graduation rate (the share of undergraduate students that complete their degree in 150% of normal time) on household income per capita in 2000. Original source: IPEDS 2009.
Manufacturing Share: Share of employed persons 16 and older working in manufacturing. Original source: 2000 Census.
Move-up: Dummy equal to one if the level of status of the respondent's job is higher than his father's one. Source: Alesina et al. (2018).
Income: Total household income, before taxes, year 2015. Discrete variable with values between 1 and 12, with 1 if income belongs to $0 - $9,999; 2 if $10,000 - $14,999; ...and 12 if $200,000 or higher. Source: Alesina et al. (2018).
Live at R: Dummy equal to one if the income of the respondent is higher than the median of the CZ's income. Source: Own construction.

Correlation matrix of beliefs in a just world measures:

Table 3.5

	AD	EP	ER	F
AD	1			
EP	0.42	1		
ER	0.44	0.44	1	
F	0.54	0.43	0.46	1

Bibliography

Abeler, J., J. Calaki, K. Andree and C. Basek, "The power of apology," *Economics Letters* 107 (2010), 233–235.

Adolph, C., K. Amano, B. Bang-Jensen, N. Fullman and J. Wilkerson, "Pandemic politics: Timing state-level social distancing responses to COVID-19," *Journal of Health Politics, Policy and Law* (forthcoming).

Aghamolla, C. and T. Hashimoto, "Information arrival, delay, and clustering in financial markets with dynamic freeriding," *Journal of Financial Economics* 138 (2020), 27–52.

Alesina, A. and G.-M. Angeletos, "Fairness and redistribution," *American Economic Review* 95 (2005), 960–980.

Alesina, A., E. Glaeser and B. Sacerdote, "Why doesn't the United States have a european-style welfare state?," *Brookings Papers on Economic Activity* (2001).

Alesina, A. and E. La Ferrara, "Preferences for redistribution in the land of opportunities," *Journal of Public Economics* 89 (2005), 897–931.

Alesina, A., S. Stantcheva and E. Teso, "Intergenerational mobility and preferences for redistribution," *American Economic Review* 108 (2018), 521–54.

Altonji, J., E. Blom and C. Meghir, "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics* 4 (2012), 185–223.

Aoyagi, M., "Equilibrium delay in a simple dynamic model of investment," *Economic Theory* 12 (1998), 123–146.

Arbel, Y. and Y. Kaplan, "Tort reform through the back door: A critique of law and apologies," *Southern California Law Review* 90 (2016), 1199.

Attanasio, O. and K. Kaufmann, "Education choices and returns to schooling: Intrahousehold decision making, gender and subjective expectations," *Journal of Development Economics* 109 (2014), 203–216.

Banerjee, A., "A simple model of herd behavior," *The Quarterly Journal of Economics* 107 (1992), 797–817.

Bebchuk, L. A., "Litigation and settlement under imperfect information," *The RAND Journal of Economics* (1984), 404–415.

Belfield, C., T. Boneva, C. Rauh and J. Shaw, "What drives enrolment gaps in further education? the role of beliefs in sequential schooling decisions," *Economica* 87 (2020), 490–529.

BENABOU, R., "Workings of a city: location, education, and production," *The Quarterly Journal of Economics* 108 (1993), 619–652.

———, "Equity and efficiency in human capital investment: the local connection," *The Review of Economic Studies* 63 (1996a), 237–264.

———, "Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance," *American Economic Review* (1996b), 584–609.

BENABOU, R. AND J. TIROLE, "Belief in a just world and redistributive politics," *The Quarterly Journal of Economics* 121 (2006), 699–746.

BEZIN, E. AND F. MOIZEAU, "Cultural dynamics, social mobility and urban segregation," *Journal of Urban Economics* 99 (2017), 173–187.

BIKHCHANDANI, S., D. HIRSHLEIFER AND I. WELCH, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy* 100 (1992), 992–1026.

BISIN, A. AND T. VERDIER, "The economics of cultural transmission and the dynamics of preferences," *Journal of Economic Theory* 97 (2001), 298–319.

BORJAS, G., "To ghetto or not to ghetto: Ethnicity and residential segregation," *Journal of Urban Economics* 44 (1998), 228–253.

BURGUET, R. AND X. VIVES, "Social learning and costly information acquisition," *Economic Theory* 15 (2000), 185–205.

CHAMLEY, C., "Delays and equilibria with large and small information in social learning," *European Economic Review* 48 (2004), 477–501.

CHAMLEY, C. AND D. GALE, "Information revelation and strategic delay in a model of investment," *Econometrica* 62 (1994), 1065–1085.

CHARI, V. AND P. KEHOE, "Financial crises as herds: overturning the critiques," *Journal of Economic Theory* 119 (2004), 128–150.

CHAUDHRY, S. AND G. LOEWENSTEIN, "Thanking, apologizing, bragging, and blaming: Responsibility exchange theory and the currency of communication," *Psychological Review* 126 (2019), 313–344.

CHETTY, R. AND N. HENDREN, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics* 133 (2018a), 1107–1162.

———, "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics* 133 (2018b), 1163–1228.

CHETTY, R., N. HENDREN, P. KLINE AND E. SAEZ, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics* 129 (2014), 1553–1623.

CHOI, J., "Herd behavior, the "penguin effect", and the suppression of informational diffusion: an analysis of informational externalities and payoff interdependency," *The RAND Journal of Economics* 28 (1997), 407–425.

Cohen, J., "Legislating apology: The pros and cons," *University of Cincinnati Law Review* 70 (2002), 819–872.

Corak, M., "Income inequality, equality of opportunity, and intergenerational mobility," *Journal of Economic Perspectives* 27 (2013), 79–102.

Cukierman, A., "The effects of uncertainty on investment under risk neutrality with endogenous information," *Journal of Political Economy* 88 (1980), 462–475.

Daughety, A. and J. Reinganum, "Keeping society in the dark: On the admissibility of pretrial negotiations as evidence in court," *The RAND Journal of Economics* 26 (1995), 203–221.

———, "Settlement," *The Encyclopedia of Law and Economics (Second Ed.) – Ch 15, Vol. 8 (Procedural Law and Economics, Ed. by Chris William Sanchirico), published by Edward Elgar Publishing Ltd.* (2012), 386–471.

Décamps, J.-P. and T. Mariotti, "Investment timing and learning externalities," *Journal of Economic Theory* 118 (2004), 80–102.

Durlauf, S., "A theory of persistent income inequality," *Journal of Economic Growth* 1 (1996), 75–93.

Durlauf, S. and A. Seshadri, "Understanding the Great Gatsby curve," *NBER Macroeconomics Annual* 32 (2018), 333–393.

Dutta, P., "Coordination need not be a problem," *Games and Economic Behaviour* 76 (2012), 519–534.

Farrell, J. and R. Gibbons, "Cheap talk with two audiences," *American Economic Review* 79 (1989), 1214–1223.

Feddersen, T. and W. Pesendorfer, "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting," *American Political Science Review* (1998), 23–35.

Fischbacher, U. and V. Utikalc, "On the acceptance of apologies," *Games and Economic Behavior* 82 (2013), 592–608.

Fogli, A. and V. Guerrieri, "The end of the American Dream? inequality and segregation in us cities," Technical Report, National Bureau of Economic Research, 2019, available at https://www.nber.org/papers/w26143.

Gal-Or, E., "First mover disadvantages with private information," *The Review of Economic Studies* 54 (1987), 279–292.

Gallagher, T., A. Waterman, A. Ebers, V. Fraser and W. Levinson, "Physicians are reluctant to give apologies because it damages their reputation," *The Journal of the American Medical Association* 289 (2003), 1001–1007.

Gilbert, B., A. James and J. Shogren, "Corporate apology for environmental damage," *Journal of Risk and Uncertainty* 56 (2018), 51–81.

Golann, D., "Dropped medical malpractice claims: Their surprising frequency, apparent causes, and potential remedies," *Health Affairs* 30 (2011), 1343–1350.

GOLTSMAN, M. AND G. PAVLOV, “How to talk to multiple audiences,” *Games and Economic Behavior* 72 (2011), 100–122.

GOMPERS, P., W. GORNALL, S. KAPLAN AND I. STREBULAEV, “How do venture capitalists make decisions?,” *Journal of Financial Economics* 135 (2020), 169–190.

GOMPERS, P., A. KOVNER, J. LERNER AND D. SCHARFSTEIN, “Venture capital investment cycles: The impact of public markets,” *Journal of Financial Economics* 87 (2008), 1–23.

———, “Performance persistence in entrepreneurship,” *Journal of Financial Economics* 96 (2010), 18–32.

GUHA, B., “Malicious litigation,” *International Review of Law and Economics* 47 (2016), 24–32.

GUPTA, U., *Done deals: Venture capitalists tell their stories* (Harvard Business School Press Cambridge, MA, 2000).

HALPERIN, B., B. HO, J. LIST AND I. MUIR, “Towards an understanding of the economics of apologies: evidence from a large-scale natural field experiment,” available at https://www.nber.org/papers/w25676,, 2019.

HEIMREICH, J., “Does ‘sorry’ incriminate - Evidence, harm and the protection of apology,” *Cornell Journal of Law and Public Policy* 21 (2012), 567–609.

HERRERA, H. AND G. ORDOÑEZ, “The political economy of pandemics,” available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3613256,, 2020.

HIROKAWA, M. AND D. SASAKI, “Endogenously asynchronous entries into an uncertain industry,” *Journal of Economics & Management Strategy* 10 (2001), 435–461.

HO, B., “Apologies as signals: With evidence from a trust game,” *Management Science* 58 (2012), 141–158.

HO, B. AND E. LIU, “Does sorry work? The impact of apology laws on medical malpractice,” *Journal of Risk and Uncertainty* 43 (2011a), 141–167.

———, “What’s an apology worth? decomposing the effect of apologies on medical malpractice payments using state apology laws,” *Journal of Empirical Legal Studies* 8 (2011b), 179–199.

HOPPE, H., “Second-mover advantages in the strategic adoption of new technology under uncertainty,” *International Journal of Industrial Organization* 18 (2000), 315–338.

JENSEN, R., “The (perceived) returns to education and the demand for schooling,” *The Quarterly Journal of Economics* 125 (2010), 515–548.

JOVANOVIC, B., “Entry with private information,” *The Bell Journal of Economics* 12 (1981), 649–660.

KAUFMANN, K., “Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns,” *Quantitative Economics* 5 (2014), 583–630.

KHANNA, N. AND R. D. MATHEWS, “Can herding improve investment decisions?,” *The RAND Journal of Economics* 42 (2011), 150–174.

Kim, P., K. Dirks, C. Cooper and D. Ferrin, "When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation," *Organizational Behavior and Human Decision Processes* 99 (2006), 49–65.

Kirpalani, R. and E. Madsen, "Strategic investment evaluation," available at https://www.rkirpalani.com/, 2021.

Kleefeld, J., "Thinking like a human: British Columbia's Apology Act," *UBC Law Review* 40 (2007), 769.

———, "Promoting and protecting apologetic discourse through law: A global survey and critique of apology legislation and case law," *Oñati Socio-legal Series* 7 (2017), 455–496.

Kluegel, J. and E. Smith, *Beliefs about inequality: Americans' views of what is and what ought to be* (Transaction Publishers, 1986).

Lemieux, T., "Postsecondary education and increasing wage inequality," *American Economic Review* 96 (2006), 195–199.

Lerner, J., "The syndication of venture capital investments," *Financial Management* (1994), 16–27.

Lerner, J. and S. Kortum, "Assessing the contribution of venture capital," *The RAND Journal of Economics* 31 (2000), 674–692.

Levin, D. and J. Peck, "Investment dynamics with common and private values," *Journal of Economic Theory* 143 (2008), 114–139.

Levy, G. and R. Razin, "Correlation capacity, ambiguity and group shifts," available at https://ethz.ch/content/dam/ethz/special-interest/mtec/cer-eth/cer-eth-dam/documents/research/seminar/2016/Ronny%20Razin.pdf, 2016.

———, "The coevolution of segregation, polarized beliefs, and discrimination: The case of private versus state education," *American Economic Journal: Microeconomics* 9 (2017), 141–70.

Loury, G., "A dynamic theory of racial income differences," Technical Report, Discussion paper, 1976, available at https://www.econstor.eu/bitstream/10419/220585/1/cmsems-dp0225.pdf.

Malik, A., C. Couzens and S. Omer, "COVID-19 related social distancing measures and reduction in city mobility," *MedRxiv* (2020).

McMichael, B., "The failure of "sorry": An empirical evaluation of apology laws, health care, and medical malpractice," *Lewis & Clark Law Review* 22 (2018), 1199–1281.

———, "Insuring apologies," *Florida State University Law Review* 48 (2021).

McMichael, B., L. V. Horn and K. Viscusi, ""Sorry" is never enough: How state apology laws fail to reduce medical malpractice liability risk," *Stanford Law Review* 71 (2019), 341–409.

Miller, D., "Distributive justice: What the people think," *Ethics* 102 (1992), 555–593.

Murto, P. and J. Välimäki, "Learning and information aggregation in an exit game," *The Review of Economic Studies* 78 (2011), 1426–1461.

MYERS, C., "Apology, sympathy, and empathy: The legal ramifications of admitting fault in U.S. public relations practice," *Public Relations Review* 42 (2016), 176–183.

O'HARA, E., "Apology and thick trust: What spouse abusers and negligent doctors might have in common," *Chicago-Kent Law Review* 79 (2004), 1055.

OHBUCHI, K.-I. AND K. SATO, "Children's reactions to mitigating accounts: Apologies, excuses, and intentionally of harm," *The Journal of Social Psychology* 134 (1994), 5–17.

OHTSUBO, Y. AND E. WATANABE, "Do sincere apologies need to be costly? Test of a costly signaling model of apology," *Evolution and Human Behavior* 30 (2009), 114–123.

OK, E., "Fixed set theory for closed correspondences with applications to self-similarity and games," *Nonlinear Analysis: Theory, Methods & Applications* 56 (2004), 309–330.

ORTOLEVA, P. AND E. SNOWBERG, "Overconfidence in political behavior," *American Economic Review* 105 (2015), 504–35.

PIKETTY, T., "Social mobility and redistributive politics," *The Quarterly Journal of Economics* 110 (1995), 551–584.

———, *Capital in the twenty-first century* (Harvard University Press, 2018).

REARDON, S., "Measures of income segregation," *Unpublished Working Paper. Stanford Center for Education Policy Analysis* (2011), available at https://inequality.stanford.edu/sites/default/files/reardon_measures-income-seg.pdf.

REARDON, S. AND K. BISCHOFF, "Income inequality and income segregation," *American Journal of Sociology* 116 (2011), 1092–1153.

REINGANUM, J. AND L. WILDE, "Settlement, litigation, and the allocation of litigation costs," *The RAND Journal of Economics* (1986), 557–566.

RIDLEY, D., "Herding versus Hotelling: Market entry with costly information," *Journal of Economics & Management Strategy* 17 (2008), 607–631.

ROB, R., "Learning and capacity expansion under demand uncertainty," *The Review of Economic Studies* 58 (1991), 655–675.

ROBBENNOLT, J., "Apologies and legal settlement: An empirical examination," *Michigan Law Review* 102 (2003), 560–516.

———, "Apologies and settlement levers," *Journal of Empirical Legal Studies* 3 (2006), 333–373.

ROSENBERG, D., E. SOLAN AND N. VIEILLE, "Social learning in one-arm bandit problems," *Econometrica* 75 (2007), 1591–1611.

ROTHSTEIN, J., "Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income," *Journal of Labor Economics* 37 (2019), S85–S123.

RUPASINGHA, A. AND S. GOETZ, "US county-level social capital data, 1990-2005," *The northeast regional center for rural development, Penn State University, University Park, PA* (2008).

RYTINA, J., W. FORM AND J. PEASE, "Income and stratification ideology: Beliefs about the American opportunity structure," *American Journal of Sociology* 75 (1970), 703–716.

Samila, S. and O. Sorenson, "Venture capital, entrepreneurship, and economic growth," *The Review of Economics and Statistics* 93 (2011), 338–349.

Sampson, R., J. Morenoff and T. Gannon-Rowley, "Assessing "neighborhood effects": Social processes and new directions in research," *Annual Review of Sociology* 28 (2002), 443–478.

Seidmann, D., "The effects of a right to silence," *The Review of Economic Studies* 72 (2005), 593–614.

Siedner, M., G. Harling, Z. Reynolds, R. Gilbert, A. Venkataramani and A. Tsai, "Social distancing to slow the US COVID-19 epidemic: An interrupted time-series analysis," *MedRxiv* (2020).

Sobel, J., "On the relationship between individual and group decisions," *Theoretical Economics* 9 (2014), 163–185.

Song, Y., "Social learning with endogenous observation," *Journal of Economic Theory* 166 (2016), 324–333.

Spier, K., "The dynamics of pretrial negotiation," *The Review of Economic Studies* 59 (1992), 93–108.

———, "Litigation," in North-Holland, ed., *Handbook of law and economics, volume 1* (Amsterdam, 2007), 259–342.

Taft, L., "Apology subverted: The commodification of apology," *Yale Law Journal* 109 (2000), 1135–1160.

Thijssen, J., K. Huisman and P. Kort, "The effects of information on strategic investment and welfare," *Economic Theory* 28 (2006), 399–424.

Vines, P., "Apologising for personal injury in law: Failing to take account of lessons from psychology in blameworthiness and propensity to sue," *Psychiatry, Psychology and Law* 22 (2015), 624–634.

Wang, T., "Information revelation through bunching," *Games and Economic Behaviour* 102 (2017), 568–582.

Watson, T., "Inequality and the measurement of residential segregation by income in American neighborhoods," *Review of Income and Wealth* 55 (2009), 820–844.

Weiner, B., S. Graham, O. Peter and M. Zmuidinas, "Public confession and forgiveness," *Journal of Personality* 59 (1991), 281–312.

White, E. and L. Hébert-Dufresne, "State-level variation of initial COVID-19 dynamics in the United States: The role of local government interventions," *MedRxiv* (2020).

Wilson, W., *The truly disadvantaged: The inner city, the underclass, and public policy* (University of Chicago Press, 2012).

Zhang, J., "Strategic delay and the onset of investment cascades," *The RAND Journal of Economics* 28 (1997), 188–205.