Università Commerciale "Luigi Bocconi"

PhD School

PhD program in: Economics and Finance

Cycle: 34

Disciplinary Field (code): SECS-P/01

Essays on Information Manipulation

Advisor: Massimo Morelli

PhD Thesis by

Daria Stepushina

ID number: 3037392

Year: 2024

Contents

1 Lie to Survive: When Propaganda is Efficient? 7

1.1 Introduction 7

1.2 Models of propaganda and limited rationality citizens 17

1.2.1 Soft propaganda and naïve citizens 18

1.2.2 Display of strength by the Dictator and partially rational citizens . 21

1.3 Models of propaganda and limited rationality élite 23

1.3.1 Soft propaganda and naïve oligarchs 24

1.3.2 Display of strength by the Dictator and partially rational élite . . . 25

1.4 Model of propaganda and sophisticated agents 26

1.4.1 Citizens and mass protests 28

1.4.2 Oligarchs and coups 30

1.5 Discussion and conclusions 30

1.6 Appendix 32

2 Trust but Verify: Quality Authentication in eCommerce 49

2.1 Introduction 49

2.2 Model 54

2.3 Buyer behaviour and seller's profits 60

2.4 Equilibrium Analysis 63

2.4.1 Benchmark: no evidence or authentication 63

2.4.2 Pure-strategy equilibria with perfect authentication technology . . . 63
2.4.3 Pure-strategy equilibria with imperfect authentication technology . 73
2.4.4 Mixed-strategy Equilibria 74
2.5 Discussion 75
2.6 Appendix 77

3 Controlling Protest Movements 93

3.1 Introduction 93
3.2 Model 97
3.2.1 Timing 97
3.2.2 Preferences 99
3.3 Equilibrium analysis 100
3.3.1 Homogeneous agents without any policy of the Dictator 100
3.3.2 Homogeneous agents with imperfect policy of the Dictator 102
3.4 Appendix 104
3.5 References 106

Acknowledgement

I thank Massimo Morelli, an amazing advisor and simply a person with a heart of gold, to whom I am infinitely grateful for his advice, support, and encouragement that guided me through the years and made this day possible.

I also thank Nenad Kos from the bottom of my heart for his patience, guidance, support, and advice, without whom this day would not be possible either.

I am thankful to have had the help and support of many other Bocconi faculty, especially Pierpaolo Battigalli and Marco Ottaviani, to whom I also want to express my gratitude and thank for their kindness and moral and professional support.

A special appreciation goes to my fellow PhD colleagues and friends for turning this lengthy endeavor into an experience of a lifetime.

Last but not least, I send my love and gratitude to my mom, my dad, and my brother, who have always had my back with their unconditional support and my significant other, Alberto, who makes it all worthwhile.

Introduction

The first chapter investigates how an autocratic (dictatorial) regime uses propaganda to stay in power by suppressing a mass protest by citizens or a coup d'état by the élite through the models of information manipulation. Agents (citizens or the élite members) participate in a collective action of an uprising against the status quo basing their decision on the belief about the strength of the Dictator which can be manipulated by the latter. I differentiate the agents depending on how they update beliefs: from fully rational to fully naïve. The manipulation of information is only partially successful in all cases. I find that the benefit of propaganda for the Dictator depends on how agents process propaganda messages. In fact, the efficiency of successful propaganda decreases in the strength of the Dictator if citizens are partially rational and increases in the strength of the Dictator if citizens are naïve. Moreover, the efficiency of propaganda does not matter in case of the coup by the élite members. Finally, when agents are fully rational and recognize the presence of propaganda, it becomes inefficient in increasing the probability of regime survival. These results indicate that understanding how agents respond to the manipulation of their beliefs is crucial also in understanding how propaganda works.

In the second chapter, which is a joint work with Francesco Massazza, we test the Grossmann-Stiglitz paradox in a setting with voluntary disclosure of partially verifiable evidence. We place our paper at the intersection of two strands of literature: the research on information certification and on the disclosure of private information. In particular. We propose an asymmetric information model in which the seller can provide partially

verifiable evidence about privately known quality and the buyer can request costly authentication of the quality. We characterize the equilibrium strategies of seller and buyer for both pooling and separating equilibria. We formalize necessary and sufficient conditions for each of the equilibria we find. We show the possibility of (partially) overcoming the Grossmann-Stiglitz paradox and show that voluntary disclosure of partially verifiable information can solve it. We study the role of the information cost in equilibrium formation. Finally, we analyze the effect of imperfect authentication technology and find that quality authentication can be acquired in equilibrium, unlike the perfect precision case.
In the third (unfinished) chapter, I try to analyze how dictatorial regimes can control media in order to suppress protest movements using a repeated public good game with information manipulation. At each round, citizens collectively participate in a mass protest aiming at overthrowing a dictator while the latter can defend her power through propaganda if she is not sure her regime is strong enough to survive the current attack. The propaganda can backfire due to only partial control over media. The preliminary results show that if protests unfold but repeatedly fail, agents learn that it is due to the strength of the regime and desist. The probability of regime survival increases over time but is not monotonic in the probability of the propaganda being successful.

Chapter 1

Lie to Survive: When Propaganda is Efficient?

1.1 Introduction

Propaganda in the most neutral sense of the Latin word ’to spread’ or ’to propagate’, that is, to promote ideas, has been known to the human race for centuries, if not millennia. Indeed, potentially one of the earliest instances of employing some techniques of propaganda dates back to the Old Babylonian Empire, as early as 1750 BC. The king of the city-state of Babylon, Hammurabi, in the epilogue to the famous Code of Hammurabi, justifies the necessity to follow the laws as they are the will of gods and gives a long list of promises of divine retribution for defacers during and after his reign. Jowett and O’Donnell (2012) argue that the creation of the Sphinx and the pyramids by the Egyptian pharaohs was also meant to persuade the population of their divine nature.

In a more modern sense of the word, however, which I understand as a tool of collective control by the government, extensive use of propaganda is clearly much more recent than 3700 years ago. Taking the definition of Jowett and O’Donnell (2012, p.1),

> “Propaganda is the deliberate, systematic attempt to shape perceptions, ma-

> nipulate cognition, and direct behavior to achieve a response that furthers the desired intent of the propagandist."

The rise of mass propaganda can, therefore, be attributed largely to the progressive elimination of illiteracy and the spread of mass media that occurred between the XIX and XX centuries. The newfound potential to easily convey ideas to large shares of the population made political propaganda an important tool of government control and an instrument to stay in power, extensively used in many countries but specifically in dictatorial and autocratic regimes. footnoteIn this paper, I am using interchangeably the notions of autocratic and dictatorial regimes, which is clearly a huge simplification. I allow myself to combine these two regimes into one notion in the paper on the basis of the absence of democratic transition of power in both to underline that the most popular ways the status quo (regime) is changed in these regimes is usually either by a protest or a coup and not through the general election. Both (but clearly not exclusively) Nazi Germany and the Soviet Union, the two probably most feared dictatorships in history, used propaganda extensively, firstly to come to power (Caprettini et al. (2021); Chogandaryan (2013)) and then to keep it by suppressing any possibility of popular revolt.

In this paper, I want to concentrate on the second goal of propaganda, that is, its role in containing a possible uprising against the government, which I assume is the only way to change the regime in the absence of general elections. I develop three very simple and tractable models of collective action with information manipulation used by a Dictator in order to inflate the beliefs about the strength of the former and, consequently, to suppress a mass protest by citizens or a coup d'état by the élite (which I also interchangeably call oligarchs). In particular, in the models, firstly, the Dictator can influence the beliefs about her strength, and then the agents (citizens or the élite members, depending on the model specification) participate in a collective action of a revolution to overthrow the current regime, basing their decision on the beliefs about the strength of the Dictator. I explore three different channels through which the propaganda can influence beliefs, depending on how sophisticated the agents are, and analyze the equilibrium probability to overthrow the regime. Specifically, I first study two different models of limited rationality citizens who do not treat the propaganda as rational (fully Bayesian) citizens (oligarchs) and conclude the analysis with a fully-fledged model of propaganda with rational citizens (oligarchs) who are Bayesian in their belief-updating process.

This paper contributes to three strands of literature. The first strand of literature this paper is mostly related to studies the specific mechanisms of propaganda and how they affect protests. Boussalis et al. (2023) examine two perspectives regarding political propaganda within autocratic governments. The authors argue that one strategic use of propaganda is to simply show the government's dominant control over its citizens, that is, to show the autocrat's strength: signaling. The second strategy is to convey a particular point of view by attracting or distracting the attention of the population. Through the text analysis of the North Korean propaganda texts, the study demonstrates that the dictatorial government can indeed interchange between the two propaganda strategies depending on the objective it wants to achieve. Carter and Carter (2022) study another strategic meaning of propaganda, arguing that propaganda can serve as a threat of repression in order to decrease the probability of mass protest. With the case study of a Chinese state-run newspaper, they find that propaganda as a tool of promising repression without the act of repression itself indeed decreases the probability of protest movements. In another paper that takes a more comprehensive view of propaganda, Carter and Carter (2021) study the effect of propaganda on mass protest through an original data-set of state-controlled newspapers from a variety of countries. Assuming that the propaganda's persuasion mechanism is threefold—to improve the beliefs about the autocrat's performance in office, to decrease the beliefs about other citizens' participation, and to threaten with repressions - they find that the propaganda does decrease the probability of civil unrest. Hollyer et al. (2015) study how controlling the disclosure of economic performance information as a tool of the strategic transparency decisions of autocratic regimes changes the propensity of the citizens to uprise against the government. The results show that a higher degree of transparency about poor economic performance increases the probability of mass protest. Mattingly and Yao (2020) analyze propaganda as a collection of various tools of emotional manipulation to achieve certain attitudes and points of view within the nation, contrary to looking at propaganda as a rational persuasion instrument. Through a number of experiments of exposing the subjects to nationalist propaganda in terms of news, movies, and social accounts, they record an increase in anti-foreign sentiment and behavior among the subjects, suggesting that propaganda can indeed manipulate public opinion to the desired sentiments. Shikikov (2023) introduces the notion of "affirmation propaganda" viewing it as a tool of consolidating the already existing views of the core government supporters by sending identity-consistent messages that, in turn, increase trust towards the state-controlled news outlets. Through three surveys in Russia, he shows that propaganda, by performing such belief affirmation strengthens trust in the government even further. Huang and Cruz (2022) study how propaganda can decrease the possibility of mass protests by influencing the beliefs of the citizens, not about the strength of the government but about the propensity of other citizens to protest. In particular, through a survey with Chinese internet users, they find that even if the subjects do not change their support for the government after being exposed to propaganda, they believe the other responders do, so the overall probability of protests decreases.

Rochlitz et al. (2023) study the effect of state propaganda in the age of social media and find its decreasing manipulation power over the population. Specifically, through a field experiment during the 2018 presidential elections in Russia, they show that if the citizens are exposed to alternative sources of information such as online news and social media, the effect of state propaganda decreases and can even sway the population away from supporting the autocrat. Although not formalized formally, it shows the potential backfiring effect of propaganda when, instead of strengthening the support of the autocrat, it weakens it. In this paper, I take this idea further and formalize a theoretical model of the possibility of backfiring. Frantz et al. (2020) study the same issue of using propaganda in increasingly digitalized societies. In particular, they show that digital repressions (in terms of censorship of information and bot-driven information campaigns online) in autocracies are more widespread than in democracies; moreover, they increase the probability of Dictator's survival.

The second strand of literature to which this paper contributes is the use of media control in general in non-democratic regimes. Gelbach and Sonin (2014) use the media in order to manipulate the citizens into taking a favorable action for the Dictator. Barberà and Jackson (2019) also use a public good game for modeling a revolution but allow agents to communicate among each other in an attempt to better understand the probability of a revolt being successful. Chen and Yang (2019) use a field experiment in China to evaluate the effect of access to the uncensored Internet on economic beliefs, political attitudes, and so on. Acemoglu et al. (2020) study the stability and persistance of political institutions with a dynamic game among different population groups that hold different political preferences. Guriev and Treisman (2020) use a signaling game played by a Dictator who tries to convince the public of her competence. Finally, Enikopov et al. (2020), through a case study in Russia, analyze the effect of social media on participation in protests, finding a positive causal relationship. Moreover, their results suggest that the causal effect goes through the lowering of coordination costs. Gehlbach et al. (2014) provide an overall survey for more papers studying nondemocratic politics and generalize the main key elements being asymmetries of information and commitment problems.

Finally, the third strand of literature covers the coordination games of regime change. It is common to model them as global games, first introduced by Carlsson and van Damme (1993) and extensively studied by Morris and Shin (1998, 2001): in addition to the prior on the strength of the regime, each agent also receives a private signal that is independent of others. However, one would expect that the active protest participants constitute a relatively homogeneous group in terms of the beliefs they hold about the Dictator's qualities. In addition, letting the agents have only the same prior gains in the tractability of the solution The closest papers to my setting in the global games literature are Edmond (2013) and Angeletos et al. (2006). The first one lets the policymaker influence the beliefs of the agents through additional signal(s) that increase the perceived strength of the regime. The main difference is that the policy of the autocrat does not have the possibility of backfiring; moreover, the model is static in the sense that the game does not move to the next round of protests if the current uprising fails. The second one gives the policymaker the opportunity to choose the cost of attacking. The cost of attacking in the current work is given exogenously. Angeletos and Werning (2006) study a game of regime change where, in addition to the private signals about the fundamentals, the agents also receive a noisy public signal available to all of them. Angeletos et al. (2007) introduce the dynamic structure into the standard static model, thus allowing the agents to learn over time about the regime depending on the previous outcomes of regime survival or failure. Boleslavsky et al. (2020) also introduce the possibility for the ruler to use the media for the manipulation of public opinion. Goldstein and Huang (2016) introduce Bayesian persuasion into the global game by enabling the policymaker to make the commitment to abandon the status quo before the agents coordinate on attacking for low values of regime strength. Huang (2015) studies how the presence of propaganda in itself is a signal about the regime's strength using the signaling model of Spence, where the amount of propaganda is the costly action by the sender (government). Analyzing the survey from the Chinese college students leads to the conclusion that exposure to propaganda indeed increases the expectation of the government's strength.

I begin by analyzing a model of fully naïve agents who are not rational in updating their beliefs about the strength of the Dictator and take the propaganda message at face value. In particular, the first approach takes propaganda as the "attempt to manipulate cognitions" and public opinion in the sense that Mattingly and Yao (2020) use propaganda. In this model specification, I study the manipulative communication aspect of propaganda, such as publishing fake news, withholding information about some events (for instance, postponing the announcement of Stalin's death) or deliberately lying about others (for instance, understating the death toll during the COVID pandemic), and its effect on the probability of an uprising against the regime. I will refer to this channel as "soft propaganda" and model it as an unobservable action (or series of actions) by the government that automatically shifts the beliefs of the citizens upwards in case propaganda is successful and downwards if it backfires. The need to depart from Bayesian updating, that is, the assumption that the citizens are rational in how they process the information upon being subject to propaganda, comes from the design of propaganda as a tool of collective mind manipulation. Specifically, among the methods of "propagandistic journalism" there are the following "tricks": First, the use of figurative and loaded language. The former, through the use of emotionally loaded words and figures of speech, helps to create a false image in the collective mind as it tends to get an emotive response from the audience instead of simply reporting the facts according to the literal language that is required in the "traditional" journalism. Secondly, propaganda extensively exploits cognitive biases (for instance, using illusory correlations, false analogies, dehumanization of the opposers to the official point of view, attention biases—the list is long) in order to manipulate the citizens and hide the lack of factual basis in the delivery of the news. footnoteMore on this topic (in Russian): M. Skulenko, "Journalism and Propaganda" (1987, Vyscha shkola); S. Kara-Murza, "The manipulation of the minds" (2000, Eksmo).

However, the novel feature of seeing propaganda as an emotional manipulation instrument that I introduce is the possibility that such "soft propaganda" can backfire and and shift downwards the beliefs about the downwards, contrary to the original goal. The rationale behind this comes from the fact that there still can be a share of savvy citizens who can "see it through" and identify that they are being manipulated or have evidence that the fake news is indeed fake. In this case, I assume they believe the opposite of what propaganda tries to convey. I find this backfiring effect extremely important to capture in the new digitized era. Given the increasing limits of audience reach of the standard propaganda offline tools (such as radio, TV, and newspapers) compared to the new online mass media channels (such as social networks including YouTube and online news outlets) and decreased ability to control (censor) information, the agents are more exposed to alternative media sources compared to the dictatorships of the past and, therefore, have more chances to critically analyze the propaganda and put its content into question. This observation of the limit of propaganda influence in the presence of access to alternative online media is shown by Rotchlitz et al. (2023), which I bring further into the current paper and formalize to study its effect on the probability of mass protests or coups. In particular, I allow for some stochasticity in the propaganda outcome: with some probability, the citizens indeed believe the propaganda messages and shift their beliefs about the strength towards high values, and with complementary probability, they shift their beliefs downward. I explain the model in detail in the next chapter.

Finally, the first approach can be seen as the propagandistic policy carried out by the "spin dictators", defined as such by Guriev and Treisman (2022). The autocratic regimes that try to keep the illusion of working democratic institutions rely massively on mass media in their attempt to create an alternative reality for the citizens in order to stay in power, but they do not resort to violent force.

In the second specification, I model the so-called "hard propaganda" which is some actions taken by the Dictator, oftentimes violent, in order to signal her strength that can be considered to have a propagandistic effect on the citizens. In this paper, I call this channel as "display-of-strength propaganda" and model propaganda as an observable public action that is usually associated with a powerful, all-mighty Dictator. Such actions can include, for instance, the prosecution of protesters and journalists, the elimination of opposition leaders, military parades, and other symbolic actions. The best illustration of such "display-of-strength" is the orchestrated (or not) airplane crash of the ex-leader of the "Wagner" paramilitary private company two months after his attempted rebellion.

The details of the plane crash are quite obscure, but the most probable version is that it was an order from Vladimir Putin to kill the traitor who dared to challenge his power. In this case, the message to the citizens and élite behind it is clear: Putin's regime is strong, and the agents in this case should recognize that this signal of the Dictator has a propagandist goal and update their beliefs towards higher values of strength.

However, in this model, I also allow the agents to doubt the credibility of the display of strength. Indeed, such a demonstration of power can be taken at face value, that is, the citizens or élite can believe the message and update their beliefs about the strength of the Dictator upwards (Putin in the example above is strong and can deal with any threat to his authority), but it can also have a backfiring effect and demonstrate the low strength of the regime because the citizens might think that only a weak Dictator can resort to such a display of strength while a truly strong leader would not need to since any challenge to her power would be unsuccessful anyway (Putin in the example above is weak since he is afraid of any threat to his authority). I model this possibility of backfiring by allowing the agents to recognize with some probability that such propaganda messages are sent by weaker regimes and update their beliefs about the strength in a Bayesian manner. In this sense, the citizens gain a higher degree of rationality compared to the first model of soft propaganda since they update the beliefs given to the backfiring effect, but the probability of backfiring is exogenous in the sense that they do not account for the fact that it should be connected to the real Dictator's strategy to send such messages.

Finally, I analyze the fully-fledged model of propaganda with rational agents who are fully aware of the presence of propaganda and attach the probabilities to observe a given propagandistic message in line with the actual probability of the Dictator sending that level of strength message. In this case, I do not differentiate between different propaganda methods, soft and hard, since the behavior of the citizens and their belief updating process are the same.

For each approach, I also consider two different types of agents, differentiated by the utility of a successful uprising. The first type of agents are common citizens who are only interested in overthrowing the Dictator and benefit from the change of the status quo independently of their personal participation in the protest. The second type is the élite or oligarchs, who only benefit from overthrowing the Dictator if they personally participate in the coup. The idea behind this distinction comes from the fact that usually only the élite members (apart from the Dictator) tend to benefit from the autocratic regime and, therefore, would only conspire against the Dictator if they would become the new élite afterwards.

The findings of the paper show that the probability of survival of the regime changes drastically with different model specifications. I find that, when agents are naïve, propaganda works generally well against the possible coups of the oligarchs since it increases the probability of survival for any regime type. The result for the agents is, however, paradoxical at first sight: the higher the type of regime, the higher should be the efficiency of propaganda to make it beneficial for the regime. In other words, for weak and middle regimes, even inefficient propaganda raises the chances of survival, while a strong regime gains from it only if it is highly efficient. This can look counter-intuitive since one could expect the opposite: a strong regime can survive even when propaganda is not very efficient. The results, however, show a very straightforward and intuitive implication that a strong dictatorship does not even need to resort to violent methods of repression to stay in power when citizens are naïve and subject to a lot of propaganda.

When the agents are partially rational and the propaganda methods are tougher, the results are more intuitive: with the citizens, the lower the effectiveness of propaganda, the higher should be the regime strength to make it profitable, so only strong regimes can benefit from inefficient propaganda, while for weak regimes it is pointless. The differentiation is strengthened even further with the oligarchs: the strong and middle-type regimes always use propaganda for any level of efficiency, while weak regimes never use it at all since it does not help in lowering the probability of being overthrown.

Finally, when agents are fully rational, the results are also straightforward. If the Dictator faces the threat of a mass protest, all regime types, irrespective of the type, send the same high-strength message, making propaganda inefficient on its own since it does not increase the initial probability of survival. What happens is that every type of regime is forced to pool on the signal of the highest strength since sending the true strength message will reveal lower strength and lower chances of survival, but rational citizens recognize the propaganda in all cases and disregard it completely. The main implication of such a result is that when citizens are sophisticated and can see through propaganda, to increase the chances of survival, the regime has to resort to more violent methods of suppressing the civil unrest—repressions. The same implication applies to the model of rational oligarchs, albeit the equilibrium is different: no regime type uses any propaganda since it is useless on its own against a potential coup d'état. Indeed, if the élite is rational and does not "buy" propaganda, the Dictator should threaten them with repression or provide them with even more financial resources to keep the oligarchs loyal to the current regime.

The rest of the paper is structured as follows. The next section develops the limited rationality models of mass protests and presents two different channels of propaganda. Section 3 develops the same limited rationality models of coups. Section 4 concludes, and all the proofs are in the appendix.

1.2 Models of propaganda and limited rationality citizens

There are two sets of players: three citizens indexed by i and a Dictator. Suppose there are also three states of the world: $\theta \in \{1, 2, 3\}$ which represent the strength of the Dictator, that is, how many citizens are required to attack in order to overthrow her. The strength is the Dictator's private information, while the citizens hold a shared prior putting the 1/3 probability on each state.

Each citizen has a binary decision a_i whether to participate in the riot to overthrow the dictatorial regime or not: $a_i = 1$ if he does and $a_i = 0$ if he does not, hence, it is an indicator function. Furthermore, attaching incurs a cost c while if the protest is successful, each agent yields a benefit of 1, irrespective of his participation in the protest. The cost values are private information of each agent and are distributed independently and identically according to a cdf $F(\dot{})$. For now, I assume that they are distributed uniformly for simplicity. Therefore, the payoff of each citizen can be written as follows:

$$u_i(a_i, a, \theta) = \mathbb{P}(\theta \leq a) - c_i a_i,$$

where a is the total size of attack: $a = a_1 + a_2 + a_3$.

1.2.1 Soft propaganda and naïve citizens

I begin the analysis of the effect of propaganda on mass protest with the case of soft propaganda and naïve citizens. With soft propaganda, the Dictator can manipulate the belief about her strength through soft propaganda *before* the protest happens, and by propaganda, I intend a simple message sent by the Dictator about her strength. In particular, she can persuade the citizens through pro-governmental media that the strength is at least medium; that is, she can convince the agents that $\mathbb{P}(\theta = 2) = \mathbb{P}(\theta = 3) = 1/2$ while the low state of the world is impossible. This persuasion policy is successful with probability α while with probability $1 - \alpha$ propaganda backfires: instead of trusting the Dictator that she is of an upper-medium strength, the citizens shift their beliefs downwards: they believe that only a weak Dictator would feel the need to employ the persuasion and start to believe that $\mathbb{P}(\theta = 1) = \mathbb{P}(\theta = 2) = 1/2$ and the highest state is impossible. Intuitively, the probability of success can be interpreted as the degree to which the audience is susceptible to the emotional manipulation of propaganda and is willing to trust the state media. The probability of back-firing, on the contrary, shows the degree of mistrust of the agents towards the Dictator when they take the message of the Dictator at the exactly opposite value. For now, the probability α is exogenously given; however, the next natural step would be to analyze the case when it can be increased by the Dictator through costly investment.

Finally, I imply that the updated belief about the strength of the regime (medium-high $\mathbb{P}(\theta = 2) = \mathbb{P}(\theta = 3) = 1/2$ or medium-low $\mathbb{P}(\theta = 1) = \mathbb{P}(\theta = 2) = 1/2$) is already the posterior belief of the citizens. In this sense, the model deals with naïve citizens insofar as the updating process is not a rational process in the citizens' minds but rather a black box. This assumption attempts to capture precisely the emotional component of propaganda intended to "zombify" or "brainwash" the citizens. The implication of this assumption is that there is no room for signaling: once the message is sent and the posterior belief is formed, the citizens do not try to understand which Dictator type sent the message but rather believe the message and use the medium-high posterior for calculating the probability of the success of the protest or do not believe it and use the medium-low posterior.

I search for a perfect Bayesian equilibrium under which 1) the propaganda decision maximizes the probability of the Dictator to survive, 2) the attacking strategy a^* maximizes the utility of each agent, and 3) is consistent with beliefs about other citizens' participation. More formally, let me denote the propaganda action of the Dictator (that is, the binary decision if to use propaganda or not) with $\Pi \in \{1, 0\}$ and beliefs of citizen i on the probability of any other agent j's attack with $b_i(p_j)$ for $i \neq j$, then the equilibrium of the game is defined as follows:

Definition 1. *A tuple* (Π^*, p^*, b^*) *is the equilibrium of the game if*

1. $\Pi^* \in \arg\max \mathbb{P}[survival|\alpha, p^*]$,

2. $p_i^* \in \arg\max \mathbb{E}[u(p_i, p_{-i}^*, c)|\alpha, \Pi^*]$, *where* $i = 1, 2, 3$, *and*

3. $p_i^* = b_j^*(p_i^*)$ *for all* $j \neq i$

I concentrate on the symmetric equilibrium, where the probability that an agent participates is the same for all: $p_i^* = p_j^* = p^*$ for all $i, j = 1, 2, 3$.

Interestingly, the threshold for propaganda efficiency (in the sense of increasing the probability of the regime to survive) increases with the strength of the regime. I summarize this result with a proposition:

Proposition 1. *The threshold for propaganda efficiency above which the Dictator opts for propaganda increases with the regime strength if the citizens are naïve.*

The result has a very intuitive rationale behind: a stronger regime can rely on its true strength in order to survive a mass protest, therefore, for propaganda to be useful, it has to be highly efficient. A weaker regime, however, will fall with much higher probability in case of no propaganda, making such a Dictator be willing to use it even if the probability of backfiring is high. The implication of the proposition is somehow discouraging if we think of how and when strong and old dictatorial regimes can fall apart. Indeed, if a Dictator (or a succession line of Dictators in the same regime) has been in power for long enough and managed to create a massive propaganda machine that has "zombified" the population enough (so it is efficient and sources of alternative information are limited) or if it shows the promise of repressions as taken in Carter and Carter (2022), the chances of the regime to survive much longer are relatively high even without much repressions. The example of post-Stalin USSR can be somehow a relevant example since it did not face any mass protests within the Soviet Union till the very last years of its existence and the actual repressions were relatively low (126 prisoners of conscience, for instance, in 1976 in the USSR according to Andrei Sakharov, a well-known Soviet dissident [1]) especially if compared to nowdays Russia (558 people in April 2023 [2] so more than four times as much) or the Stalinist era with millions of Soviet citizens in Gulag.

[1] Source: the Memorial Human Rights Centre, in Russian: https://www.memo.ru/ru-ru/

[2] Source: the Memorial Human Rights Centre, in Russian: https://www.memo.ru/ru-ru/

1.2.2 Display of strength by the Dictator and partially rational citizens

In this section, another policy of the Dictator will be analyzed where I allow the Dictator to take actual action in order to increase her apparent strength - “hard propaganda”. A violent public act, for instance, putting a leader of the opposition in prison or arresting the protesters, is different from propaganda in the sense that it indeed takes place and is not just persuasive in nature. Another display-of-strength example could be a continuous demonstration of military force. However, it can also convey a signal to the agents about the regime strength, which, in turn, can be ambiguous: sending Navalny to jail can be seen as a prove of Putin’s unquestionable power or as a desperate try to save a very vulnerable dictatorship. In either case, taking a repressive action sends a signal about the type of the regime and can be used instead of or in addition to manipulative propaganda, which indeed happens in many dictatorial states.

I adopt the approach of Gehlbach et al. (2016) and propose an extension to three-state world. Specifically, I assume that there is a constraint on the potential of this kind of propaganda: since it requires an actual observable action, I assume that it might not be feasible for some regime types. Specifically, I assume that the weakest regime cannot even display the level of strength corresponding to the strongest regime. The assumption seems reasonable: a weak Dictator with limited police forces might not be able to exhibit the same display of force available to a truly strong regime.

Each regime can take an observable action d_c which increases by 1 step the *displayed* level of strength. In other words, it makes θ_2-type appear as the strongest regime and θ_1-type appear as the medium-strength regime. The action is successful with probability δ, that is, with probability δ citizens are convinced by the display of power by the Dictator. With probability $1 - \delta$, instead, this action backfires and returns the true state of strength.

The updated belief distribution of the regime strength is the following:

- Given the highest displayed strength, it is either indeed the highest regime type with probability $\mathbb{P}(\theta = \theta_3 | d_u = \theta_3) = \frac{1}{1+\delta}$ or a successful medium regime type with probability $\mathbb{P}(\theta = \theta_2 | d_u = \theta_3) = \frac{\delta}{1+\delta}$.

- Given middle displayed strength, it is either indeed the medium regime type with probability $\mathbb{P}(\theta = \theta_2 | d_u = \theta_2) = 1 - \delta$ or a successful weakest regime type with probability $\mathbb{P}(\theta = \theta_1 | d_u = \theta_2) = \delta$

- Given the weakest displayed strength, it can only be the unsuccessful weakest regime: $\mathbb{P}(\theta = \theta_1 | d_u = \theta_1) = 1$.

Given the nature of propaganda in this model, I assume only partial rationality of the citizens. In particular, I assume that the citizens are able to realize the real updated probability of the types to display a certain level of strength but take it at face value in the sense that the signaling aspect is also missing in this model.

What is interesting in this case is that the benefit of using propaganda increases with the strength of the regime. Firstly, I find that the strongest type always wants to use propaganda. Secondly, for the middle type, there exists a threshold on the efficiency above which it is beneficial to use propaganda. Finally, the weakest type does not find it profitable to use propaganda, as the probability of surviving without it is higher.

Therefore, the results of proposition 1 are reversed when agents update rationally since the threshold for propaganda efficiency decreases with regime strength. This conclusion is summarized in the next proposition:

Proposition 2. *Suppose agents are partially rational, and the propaganda mechanism is as described above. The threshold for propaganda efficiency decreases with the strength of the regime. Moreover,*

- *The strong regime uses propaganda for any $\delta \in (0, 1)$.*

- *There exists* $\hat{\delta} \in (0, 1)$ *such that the middle-strength regime uses propaganda if and only if* $\delta \geq \hat{\delta}$*;*
- *The weak regime never uses propaganda.*

The implication that can be drawn from this model is as follows: a strong regime has to always prove her appearance and discourage the citizens from participating in a revolution; a middle-type regime can perform this display of strength only if it is credible enough; and a weak regime does not try to demonstrate artificial strength since it will not be successful since even partially rational citizens will see through it.

When the results are combined, the practical conclusion that can be drawn from the two models of soft and hard propaganda is as follows: a strong regime will always continue demonstrating its strength but use the propaganda only if it is highly efficient. A middle-strength regime will use both soft and hard propaganda even if is it less efficient but efficient enough that the citizens with limited rationality would believe it. Finally, a weak regime would tend to use any kind of soft propaganda even if the citizens mostly do not trust it but refrain from trying to display the artificial strength since the citizens do not believe in it even if partially rational.

In the next section, I study the implications of the same two models but with a different type of agents: oligarchs.

1.3 Models of propaganda and limited rationality élite

In the section above, the protest participants were assumed to be common citizens and value the regime change on its own. Now I consider a different utility function of the agents which captures the motives of the élite o, as I call them, oligarchs, who only benefit from the regime change if they are active participants in it. The rationale behind this assumption is that the élite are only interested in overthrowing the Dictator if they will be in the new government after a successful revolt, or a coup d'état.

Specifically, I assume that an agent gets a benefit from the change of the regime only if he himself participated in the protest. Therefore, the utility of a citizen i is

$$u_i(a_i) = \begin{cases} 1 - c & a_i = 1 \text{ and regime changes} \\ 0 & \text{otherwise} \end{cases}$$

The rationale behind this change comes from the fact that now only the coup d'état participants become the new élite in the next regime and benefit from the change.

As before, I first consider the case with the élite members face soft propaganda, which mechanism shifts the belief about the strength of the regime upwards in case it is successful and downwards in case not.

1.3.1 Soft propaganda and naïve oligarchs

Re-call, that if propaganda is successful, the agents believe that $\mathbb{P}(\theta = 2) = \mathbb{P}(\theta = 3) = 1/2$ while the low state of the world is impossible. This persuasion policy is successful with probability α while with probability $1 - \alpha$ propaganda backfires: instead of trusting the Dictator that she is of an upper-medium strength, the citizens shift their beliefs downwards: they believe that only a weak Dictator would feel the need to employ the persuasion and start to believe that $\mathbb{P}(\theta = 1) = \mathbb{P}(\theta = 2) = 1/2$ and the highest state is impossible.

The result is quite different in essence compared to the case of mass protests. Firstly, without propaganda, the regime cannot survive a coup. Secondly, in the case of failed propaganda, the regime always falls as well. However, whatever the strength of the regime, with effective propaganda, the probability of survival is strictly positive, leading to the conclusion that each regime type will always use it.

Proposition 3. *With naïve élite, any regime type always uses propaganda, for any $\alpha \in (0, 1)$.*

I find two interesting aspects needing more attention. Firstly, there is no distinction in how beneficial the propaganda is for different regime types. For each of them, propaganda strictly increases the probability to stay in power even when propaganda can backfire. Secondly, even for very ineffective propaganda with α being low, it is still worth using it compared to the case of mass protests when for the strongest type it had to be almost 1 to be beneficial.

1.3.2 Display of strength by the Dictator and partially rational élite

Now I assume that the élite members are partially rational in the sense described above and the display of strength works in the same way as with citizens. Re-call that the updated belief distribution of the regime strength is the following:

- Given the highest displayed strength, it is either indeed the highest regime type with probability $\mathbb{P}(\theta = \theta_3 | d_u = \theta_3) = \frac{1}{1+\delta}$ or a successful medium regime type with probability $\mathbb{P}(\theta = \theta_2 | d_u = \theta_3) = \frac{\delta}{1+\delta}$.
- Given middle displayed strength, it is either indeed the medium regime type with probability $\mathbb{P}(\theta = \theta_2 | d_u = \theta_2) = 1 - \delta$ or a successful weakest regime type with probability $\mathbb{P}(\theta = \theta_1 | d_u = \theta_2) = \delta$
- Given the weakest displayed strength, it can only be the unsuccessful weakest regime: $\mathbb{P}(\theta = \theta_1 | d_u = \theta_1) = 1$.

The results are close in essence to the case of partially rational citizens: the benefit of propaganda decreases in the regime strength, more sharply in this case. This conclusion is summarized in the next proposition:

Proposition 4. *Suppose élite members are partially rational, and the propaganda mechanism is as described above, then*

- *The strong and medium regimes use propaganda for any $\delta \in (0, 1)$.*

- *The weak regime never uses propaganda.*

The implication that can be drawn from this model is as follows: a strong or medium-strong regime has to always prove her appearance and discourage the oligarchs from plotting a coup, but it is enough to keep them away from a putsch while a weak regime does not try to demonstrate artificial strength since it will not be successful so the potential way to keep the élite united around the Dictator should be something else, for instance, buying loyalty or following the advice of Machiavelli and dividing the oligarchs and ruling.

1.4 Model of propaganda and sophisticated agents

In the previous chapter, the analysis was performed with the implicit assumption that, if the soft propaganda or display of strength is successful, it will be taken at face value by the agents. In other words, in the state of the world when the Dictator's action turns out to be successful, neither citizens nor the élite attach any probability to the fact that it could have been manipulation by the ruler and ignore the signaling part of the model. On the contrary, in the state of the world when the Dictator fails, both types of agents attach probability 1 to it. The assumption is quite reasonable if we think of propaganda as a manipulative tool that acts on the emotional part of human beings. In this case, indeed, we could think of successful propaganda as a manipulation that is not discovered by the agents and failed propaganda as a manipulation uncovered by the agents.

The assumptions are, however, very limitating. Firstly, it constrains the analysis to some specific cases with no space for signaling. Secondly, it gives little credit to the citizens or the élite of a country in recognizing propaganda, assuming them to be quite naïve. In many cases, it might not be the case, and the citizens, instead, can have the idea that the messages sent by the Dictator about her strength may not necessarily be truthful, but they can never know for sure, and neither can they disprove them. The formation of beliefs in this case should be different from the ones before.

Suppose the Dictator can send a propagandistic message about a certain strength of the regime, m_i, where $i = 1, 2, 3$ is the reported strength. In this case, I do not distinguish between soft propaganda or displaying strength but rather intend the message as any action or news that should support the declared level of strength. Moreover, suppose an agent, a citizen, or an oligarch, upon observing this message about the regime strength, thinks it is propaganda (that is, the real strength of the Dictator is strictly lower) with some probability μ and it is truthful, that is, matching the actual strength, with probability $1 - \mu$. The latter case can be again interpreted as a backfiring effect of propaganda. It means that if the message is about high strength, m_3, with probability μ agents think it is sent by either middle or weak regimes that try to pass as a strong regime and with $1 - \mu$ it is indeed the strong regime. If the message is about middle strength, m_2, agents think it is either the weak regime that try to pass as a middle regime with probability μ and with probability $1 - \mu$ it is indeed the middle regime. Finally, if the message is about low strength, m_1, then the agents always know it is the weak regime. Moreover, assume that if a strong Dictator sends a truthful high message, it is always believed with probability one.

Moreover, in equilibrium, this belief distribution about the sent message should correspond to the actual probabilities with which different regime types send different messages. For instance, the probability with which agents believe m_3 is sent by the middle regime should be equal to the equilibrium probability with which the middle type indeed sends this message. To formalize this requirement, firstly, define the beliefs about the type of Dictator sending different messages as follows: $\mu(m_3) = (\mu_3(m_3), \mu_2(m_3), \mu_1(m_3)) = (1, \mu_2(m_3), \mu_1(m_3))$ for the probability to send m_3 where I assume that the high type always sends the high message; $\mu(m_2) = (\mu_2(m_2), \mu_1(m_2))$ for the probabilities to send m_2, and $\mu(m_1) = (\mu_1(m_1))$ to send m_1. Secondly, let $\sigma_i(m_j)$ with $i, j = 1, 2, 3$ be the strategy of type-i regime to send strength-j propaganda message, and assuming $\sigma_3(m_3) = 1$, it should hold that

$$\begin{aligned}
&\text{Upon receiving } m_3 : \mu_3(m_3) = \sigma_3(m_3) = 1,\ \mu_2(m_3) = \sigma_2(m_3),\ \mu_1(m_3) = \sigma_1(m_3)\\
&\text{Upon receiving } m_2 : \mu_2(m_2) = \sigma_2(m_2),\ \mu_1(m_2) = \sigma_1(m_2)\\
&\text{Upon receiving } m_1 : \mu_1(m_1) = \sigma_1(m_1)
\end{aligned}$$

Moreover, to be valid probability distributions, the following conditions must hold:

$$\begin{aligned}
&\sigma_2(m_2) + \sigma_2(m_3) = \mu_2(m_3) + \mu_2(m_2) = 1\\
&\sigma_1(m_1) + \sigma_1(m_2) + \mu_1(m_3) = \mu_1(m_1) + \mu_1(m_2) + \mu_1(m_3) = 1
\end{aligned}$$

Given the prior structure and the strategies of the Dictator, the posterior beliefs are as follows:

$$\begin{aligned}
&\mathbb{P}(\theta_3|m_3) = \frac{1 + (1 - \mu_2(m_3))\sigma_2(m_3) + (1 - \mu_1(m_3))\sigma_1(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)}\\
&\mathbb{P}(\theta_2|m_3) = \frac{\mu_2(m_3)\sigma_2(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)}\\
&\mathbb{P}(\theta_1|m_3) = \frac{\mu_1(m_3)\sigma_1(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)}\\
&\mathbb{P}(\theta_2|m_2) = \frac{\sigma_2(m_2) + (1 - \mu_1(m_2))\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)}\\
&\mathbb{P}(\theta_1|m_2) = \frac{\mu_1(m_2)\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)}\\
&\mathbb{P}(\theta_1|m_1) = 1
\end{aligned}$$

1.4.1 Citizens and mass protests

I begin by analyzing the behavior of citizens. Re-call that the citizens are modelled as the agents such that their payoffs depend only on the survival of the regime while the cost depends on their participation: they benefit from the regime change independently of their attacking strategy but pay the cost only in case of participation.

I find that the unique existing equilibrium is in pure strategies where all types pool on the same message of the highest strength: $\sigma_1^*(m_3) = \sigma_2^*(m_3) = \sigma_3^*(m_3) = 1$.

Proposition 5. *There exists the unique equilibrium where* $\sigma_1^*(m_3) = \sigma_2^*(m_3) = \sigma_3^*(m_3) = 1$ *and citizens attack with probability 1/3.*

The intuition of the result is simple: if the Dictator faces the threat of a mass protest, all regime types independently of the type, are forced to send the same high-strength message, making propaganda inefficient on its own since it does not increase the initial probability of survival. What happens is that every regime type has to pool on the signal of the highest strength since otherwise the true strength message will reveal lower strength and lower chances of survival, but rational citizens recognize the propaganda in all cases and disregard it completely. The main implication of such a result is that when citizens are sophisticated and can see through propaganda, to increase the chances of survival, the regime has to resort to more violent methods of suppressing the civil unrest—repressions. The unique way to decrease the incentives for citizens to attack the regime is to increase the cost of protest participation. It can be done through different repressive acts, from tightening the repressive laws and increasing the punishment that follows participation in the protest to violently suppressing the protests with police forces or the army.

A second implication of the equilibrium is that with sophisticated citizens, propaganda becomes white noise and plays a necessary but useless role in trying to understand how strong the regime is. Indeed, if one wants to analyze the stability of the regime, any information from the state-run media (as well as potentially other official sources of information, including economic or social performance indicators) is quite unreliable since all regimes would try to embellish the state of the country as much as they can.

1.4.2 Oligarchs and coups

In this section, I study the equilibria of the sophisticated agents model who can recognize the presence of propaganda where the agents are the élite, that is, different from the mass protests where citizens gain utility just from the fact that the status quo is abandoned, they gain the benefits of the regime change only if they participate in the rebellion. Contrary to the mass protests case, there is no equilibrium at all where the regime can raise the chances of surviving a coup, in the sense that the propaganda is fruitless in trying to prevent the élite from overthrowing the regime if the members of the élite decide to attack the Dictator.

Proposition 6. *There is no equilibrium where the regime can survive the coup with a positive probability when the oligarchs are rational in recognizing propaganda.*

The same implication as with sophisticated rational citizens can be drawn from this model of rational oligarchs, albeit the equilibrium is different: no regime type uses any propaganda since it is useless on its own against a potential coup d'état. Indeed, if the élite is rational and does not "buy" propaganda, the Dictator should threaten them with repressions or buy loyalty.

1.5 Discussion and conclusions

With the models above, I attempted to analyze different propaganda instruments, differentiating them by the method of propaganda (soft vs. hard) and the degree of rationality of the agents, citizens in cases of mass protest and oligarchs in cases of coups. Given the complexity of the notion of propaganda, I find it important to treat propaganda differently and try to investigate one by one the effects it has on the population. Indeed, even if this word is deeply ingrained in the modern language, there is no unique definition of what propaganda is or single understanding of how exactly and through which tools it works; therefore, a careful consideration of cognitive assumptions and propaganda mechanisms is needed in order to analyze the issue. The drastic difference in the results of the models above proves this point even further: a change in cognitive assumptions can have drastically different implications.

If I try to combine all the results together and draw more general conclusions, I find that stronger regimes usually opt for hard propaganda when the agents' rationality is bounded, while weaker regimes prefer softer ways of persuasion since the show of artificial strength in case the agents are even a bit skeptical about the credibility of propaganda. It aligns well with the idea that autocrats generally demonstrate their strength in various ways if there is some secondary evidence that allows one to deduce they can be a strong regime, for instance, a large police body or an army. On the contrary, for weaker regimes, the display of strength can be an unreliable instrument since it is not credible; therefore, they can tend to persuade the citizens with softer propaganda.

I also find completely different equilibrium outcomes in the case of sophisticated agents, where propaganda on its own is not useful since both citizens and oligarchs can recognize it. A natural implication of the results is that with sophisticated citizens, a Dictator will have to use other ways to prevent the citizens from a revolution or the oligarchs from plotting a coup, with the immediate candidate of repression as an instrument of staying in power. It opens up a new line of research by studying a combination of propaganda and repression to suppress the mass protest. An interesting research idea for another paper could be the problem of the Dictator with limited resources that need to be split between propaganda and repression in the most efficient way.

Another interesting issue to study is the dynamics of propaganda since normally propagandist messages are sent every day, and they should have a cumulative effect on the degree of trust towards the Dictator and Dictator's produced news. I start analyzing the problem in the third chapter of the thesis.

1.6 Appendix

Proof of Proposition 1

Proof. Consider two cases:

1. The Dictator does not exhibit the policy, therefore, the beliefs of the agents about her strength remain equal to the prior. Then the probability of participating in the riot is pined down by the indifference condition:

$$\mathbb{E}[u_i(a_i = 1, c_i)] = \mathbb{E}[u_i(a_i = 0, c_i)]$$

Expected utility from participation is (where p_{NP} denotes the probability to attack without propaganda)

$$\begin{aligned}\mathbb{E}[u_i(a_i = 1, c)] &= \mathbb{P}(\theta = 1)(p^2 + 2p_{NP}(1 - p_{NP}) + (1 - p_{NP})^2) + \mathbb{P}(\theta = 2)(p_{NP}^2 + 2p_{NP}(1 - p_{NP}))\\ &+ \mathbb{P}(\theta = 3)(p_{NP}^2) - c_i\\ &= 1/3(p_{NP}^2 + 2p_{NP}(1 - p_{NP}) + (1 - p_{NP})^2) + 1/3(p_{NP}^2 + 2p_{NP}(1 - p_{NP})) + 1/3(p_{NP}^2) - c_i\\ &= p_{NP}^2 + 8/3p_{NP}(1 - p_{NP}) + 1/3(1 - p_{NP})^2 - c_i\end{aligned}$$

Expected utility from not participating is

$$\begin{aligned}\mathbb{E}[u_i(a_i = 0, c)] &= \mathbb{P}(\theta = 1)(p_{NP}^2 + 2p_{NP}(1 - p_{NP})) + \mathbb{P}(\theta = 2)(p_{NP}^2)\\ &= 2/3p_{NP}^2 + 2/3p_{NP}(1 - p_{NP}))\end{aligned}$$

It holds that

$$\mathbb{E}[u_i(a_i = 1, c_i)] > \mathbb{E}[u_i(a_i = 0, c_i)] \text{ iff } 1 > 3c_i$$

Hence, the equilibrium participation is, for $i = 1, 2, 3$,

$$a_i^* = \begin{cases} 1 & c_i < 1/3 \\ \Delta([0,1]) & c_i = 1/3 \\ 0 & c_i > 1/3 \end{cases} \tag{1.1}$$

Then, the survival probability is intermediate and depends on the strength of the regime:

- the weak regime survives with the following probability:

$$\begin{aligned} P_{survive_1} &= \mathbb{P}(\text{no agent attacks}) = \\ &= \mathbb{P}(\min c_i > 1/3) = (1 - 1/3)^3 = 8/27. \end{aligned}$$

- the medium regime survives with the following probability:

$$\begin{aligned} P_{survive_2} &= \mathbb{P}(\text{at most one agent attacks}) = \\ &= \mathbb{P}(\min c_i > 1/3) + \mathbb{P}(\text{“only one”} c_i < 1/3) = (1 - 1/3)^3 + 3 \times 1/3 \times (1 - 1/3)^2 = 20/27. \end{aligned}$$

- the strong regime survives with the following probability:

$$\begin{aligned} P_{survive_3} &= \mathbb{P}(\text{at most two agents attack}) = \\ &= 1 - \mathbb{P}(\max c_i < 1/3) = 1 - (1/3)^3 = 26/27 \end{aligned}$$

As it could be expected, the survival probability increases with the strength of the regime, however, each type of the government runs a positive risk of failing.

2. Consider now the case when the Dictator exhibits the policy, there are, again, two subcases to consider:

   (a) The policy is successful, then a posteriori, the belief of the agents is $\mathbb{P}_S(\theta = 2) = \mathbb{P}_S(\theta = 3) = 1/2$. Calling the new symmetric probability of participation by p_S, I have the following expected utility of participation:

$$\begin{aligned}\mathbb{E}[u_i(a_i = 1, c_i)] &= \mathbb{P}_S(\theta = 2)(p_S^2 + 2p_S(1 - p_S)) + \mathbb{P}_S(\theta = 3)(p_S^2) - c_i \\ &= 1/2(p_S^2 + 2p_S(1 - p_S)) + 1/2(p_S^2) - c_i \\ &= p_S^2 + p_S(1 - p_S) - c_i\end{aligned}$$

Expected utility from not participating is

$$\mathbb{E}[u_i(a_i = 0, c_i)] = \mathbb{P}_S(\theta = 2)(p_S^2) = 1/2(p_S^2)$$

The equilibrium participation probability solves

$$\begin{aligned}p_S^2 + p_S(1 - p_S) - c_i &= 1/2(p_S^2) \\ p_S &= 1 \pm \sqrt{1 - 2c_i}\end{aligned}$$

For p_S to be a valid probability, the only possible solution is $p_S = 1 - \sqrt{1 - 2c_i}$ that exists for $c_i \leq 1/2$. For a higher cost, the agents do not attack.

Moreover, it has to be consistent with the expectation each agent holds about the cost of participation of other agents. Specifically, there is another belief consistency condition to consider:

$$p_S = F(c) = c$$

that is, the participation probability p_S needs to be equal to the probability that the cost of any other agent $j \neq i$ falls below that threshold.

Combining the two equations, I get

$$\begin{aligned}1 - \sqrt{1 - 2c} &= c \\ 1 - c &= \sqrt{1 - 2c} \\ 1 - 2c + c^2 &= 1 - 2c \\ c &= 0\end{aligned}$$

Therefore, with effective propaganda, no agent attacks, and all regimes survive with probability one.

(b) Finally, the case where the propaganda is unsuccessful (failing): $\mathbb{P}_F(\theta = 2) = \mathbb{P}_F(\theta = 1) = 1/2$. Calling the new symmetric probability of participation by p_F, I have the following expected utility of participation:

$$\begin{aligned}\mathbb{E}[u_{iF}(a_i = 1, c_i)] &= \mathbb{P}_F(\theta = 2)(p_F^2 + 2p_F(1 - p_F)) + \mathbb{P}_F(\theta = 1)(p_F^2 + 2p_F(1 - p_F) + (1 - p_F)^2) - c_i \\ &= 1/2(p_F^2 + 2p_F(1 - p_F)) + 1/2(p_F^2 + 2p_F(1 - p_F) + (1 - p_F)^2) - c_i \\ &= p_F^2 + 2p_F(1 - p_F) + 1/2(1 - p_F)^2 - c_i\end{aligned}$$

Expected utility from not participating is

$$\begin{aligned}\mathbb{E}[u_{iF}(a_i = 0, c_i)] &= \mathbb{P}_F(\theta = 2)(p_F^2) + \mathbb{P}_F(\theta = 1)(p_F^2 + 2p_F(1 - p_F)) \\ &= 1/2(p_F^2) + 1/2(p_F^2 + 2p_F(1 - p_F)) \\ &= p_F^2 + p_F(1 - p_F)\end{aligned}$$

The equilibrium participation probability solves

$$\begin{aligned}p_F^2 + 2p_F(1 - p_F) + 1/2(1 - p_F)^2 - c_i &= p_F^2 + p_F(1 - p_F) \\ p_F &= \sqrt{1 - 2c_i}\end{aligned}$$

Combining with the consistency of beliefs condition, I get

$$\begin{aligned}\sqrt{1 - 2c} &= c \\ 1 - 2c &= c^2 \\ c &= .\sqrt{2} - 1 \\ p_F &= \sqrt{2} - 1\end{aligned}$$

Having solved separately two subcases, I can find the survival probabilities for each value of the regime strength and find the threshold of α above which using the propaganda increases the chances of survival of each regime:

- Consider the weak regime. Without propaganda, it survives only with probability 8/27, while with propaganda it survives with probability $\alpha + (1-\alpha)(\sqrt{2})^3$. Therefore, the threshold for the propaganda efficiency is given by

$$\begin{aligned}
8/27 &= \hat{\alpha}_1 + (1-\hat{\alpha}_1)(2-\sqrt{2})^3 \\
\hat{\alpha}_1 &= \frac{8/27-(2-\sqrt{2})^3}{1-(2-\sqrt{2})^3} \\
&\simeq 0.12
\end{aligned}$$

- Consider the medium regime now. The threshold for the propaganda efficiency is given by

$$\begin{aligned}
20/27 &= \hat{\alpha}_2 + (1-\hat{\alpha}_2)[(2-\sqrt{2})^3 + 3(\sqrt{2}-1)(2-\sqrt{2})^2] \\
\hat{\alpha}_2 &= \frac{20/27-(2-\sqrt{2})^3-3(\sqrt{2}-1)(2-\sqrt{2})^2}{1-(2-\sqrt{2})^3-3(\sqrt{2}-1)(2-\sqrt{2})^2} \\
\hat{\alpha}_2 &= \frac{20/27-(2-\sqrt{2})^2(8\sqrt{2}-10)}{1-(2-\sqrt{2})^2(8\sqrt{2}-10)} \\
&\simeq 0.53
\end{aligned}$$

- Finally, consider the strong regime now. The threshold for the propaganda efficiency is given by

$$\begin{aligned}
26/27 &= \hat{\alpha}_3 + (1-\hat{\alpha}_3)(\sqrt{2}-1)^3 \\
\hat{\alpha}_3 &= \frac{26/27-(\sqrt{2}-1)^3}{1-(\sqrt{2}-1)^3} \\
&\simeq 0.96
\end{aligned}$$

□

Proof of Proposition 2

Proof. Consider the conditions of a citizen to attack upon observing the display of strongest level of power assuming a symmetric equilibrium with probability of attacking q:

$$\begin{cases}
\frac{1}{1+\delta}q^2 + \frac{\delta}{1+\delta}\left[q^2 + 2q(1-q)\right] - c \geq \frac{\delta}{1+\delta}q^2 \\
q = c
\end{cases}$$

Combining the two together and simplifying, it should hold

$$c[c(1-3\delta)-(1-\delta)] \geq 0$$
$$c(1-2\delta) \geq 1-\delta$$

The condition cannot hold for $\delta \in (0,1)$ and, $c \in (0,1)$, therefore, no citizen attacks. Therefore, a strong regime always uses propaganda since it guarantees the survival. Moreover, conditional on being efficient, propaganda is beneficial for the middle type of the regime as well.

Similarly, upon observing the display of middle level of power, a citizen attacks if and only if:

$$\begin{cases} (1-\delta)[q^2+2q(1-q)]+\delta-c \geq (1-\delta)q^2+\delta[q^2+2q(1-q)] \\ q=c \end{cases}$$

The two conditions boil down to

$$c^2(3\delta-2)+c(1-4\delta)+\delta \geq 0$$

The roots are

$$\begin{aligned} c_1 &= \frac{4\delta-1-\sqrt{1-4\delta^2}}{6\delta-4} > 0 \\ c_2 &= \frac{4\delta-1+\sqrt{1-4\delta^2}}{6\delta-4} > 0 \text{ iff } \delta > 2/3 \end{aligned}$$

Consider three cases:

- If $\delta < 2/3$, I have an inverted parabola, and the attacking condition holds for $c \in [0, c_1]$;
- If $\delta = 2/3$, the condition is linear holding for $c \in [0, 1/2]$;

- If $\delta > 2/3$, c_2 is positive but above 1, therefore, the the attacking condition holds for $c \in [0, c_1]$.

Each agent attacks for $c \geq c_1$ whenever $\delta \neq 2/3$ and for $c \geq 1/2$ whenever $\delta = 2/3$. Moreover, it can be shown that c_1 is a decreasing function on the $[0, 1]$ interval taking the value of $1/2$ at 0 and $\frac{3-\sqrt{5}}{2}$ at 1. It means that for the middle type, the backfiring of propaganda is backfiring indeed: if the propaganda is unsuccessful, it increases the individual probability of propaganda compared to no-propaganda case (re-call that under no propaganda, the cost threshold on attacking is $1/3$).

Finally, upon observing the display of weak level of power, a citizen attacks if and only if:

$$\begin{cases} 1 - c \geq q^2 + 2q(1-q) \\ q = c \end{cases}$$

The two conditions boil down to

$$c^2 - 3c + 1 \geq 0$$

And a citizen attacks for $c \geq \frac{3-\sqrt{5}}{2}$. As before, this threshold is above $1/3$ meaning that the citizens' individual probability to attack is lower under no propaganda if the propaganda is successful or not if the regime is weak.

□

Proof of Proposition 3

Proof. Re-call that if the propaganda is successful, the agents believe that it is either medium or high strength with equal probabilities. As before, I consider first the case without any propaganda and look for a symmetric equilibrium denoting by v the probability of participation in the coup:

1. Let the probability that any agents attacked is v, it is pined down by the indifference condition:

$$\mathbb{E}[u_i(a_i = 1, c)] = \mathbb{E}[u_i(a_i = 0, c)]$$

Expected utility from participation is

$$\begin{aligned}\mathbb{E}[u_i(a_i = 1, c)] &= \mathbb{P}(\theta = 1) + \mathbb{P}(\theta = 2)(v^2 + 2v(1-v)) + \mathbb{P}(\theta = 3)(v^2) - c \\ &= \frac{1}{3}(1 + 2v^2 + 2v(1-v)) - c \\ &= \frac{1}{3}(1 + 2v) - c\end{aligned}$$

Expected utility from not participating in coup d'état is zero. Combined with the consistency of the beliefs condition, I have that with no propaganda, every oligarch should participate in the coup:

$$\begin{cases} \frac{1}{3}(1+2v) & \geq c \\ v & = c \end{cases}$$

$$1 \geq c$$

2. Consider now the case with successful propaganda. The expected utility from participation is

$$\begin{aligned}\mathbb{E}[u_{iS}(a_i = 1, c)] &= \mathbb{P}(\theta = 2)(v_S^2 + 2v_S(1-v_S)) + \mathbb{P}(\theta = 3)(v_S^2) - c \\ &= \frac{1}{2}(2v_S^2 + 2v_S(1-v_S)) - c \\ &= v_S - c\end{aligned}$$

Combined with the consistency of beliefs, $v_S = c$, the probability of participation is simply $v_S^* = c$. Then the expected probability of participation from the point of view of the regime is $\mathbf{E}(c) = 1/2$.

3. Finally, consider the unsuccessful propaganda.

The expected utility from participation is

$$\begin{aligned}\mathbb{E}[u_{iF}(a_i = 1, c)] &= \mathbb{P}(\theta = 2)(v_F^2 + 2v_F(1 - v_F)) + \mathbb{P}(\theta = 1) - c \\ &= \frac{1}{2}(v_F^2 + 2v_F(1 - v_F) + 1) - c \\ &= \frac{1}{2}(-v_F^2 + 2v_F + 1) - c\end{aligned}$$

The participation probability is then

$$v_F^* = 1 - \sqrt{2 - 2c}$$

Combined with the consistency of beliefs, $v_F = c$, I have that the élite attack for $c \leq 1$. Then the probability that a given oligarch participates in a coup is same as under no propaganda and is equal to 1.

This result leads to the conclusion that for any level of propaganda efficiency, any regime type will turn to it as it strictly increases the chance to survive when propaganda is successful. Indeed, of the propaganda fails, each élite member attack with probability 1 but if the propaganda is successful, the survival of any regime is strictly positive for any $\alpha > 0$. □

Proof of Proposition 4

Proof. Consider what is the problem of an oligarch denoting the individual probability of attacking as q_e:

$$\begin{cases} \frac{1}{1+\delta}q_e^2 + \frac{\delta}{1+\delta}\left[q_e^2 + 2q_e(1-q_e)\right] - c \geq 0 \\ q_e = c \end{cases}$$

Combining the two together and simplifying, it should hold

$$c[c(1-2\delta) - (1-\delta)] \geq 0$$
$$c(1-2\delta) \geq 1-\delta$$

Same as with citizens, no oligarch attacks. Therefore, the display of strength guarantees survival for strongest regime and for the medium type in case it is successful.

Similarly, upon observing the display of middle level of power, an élite member attacks if and only if:

$$\begin{cases} (1-\delta)[q_e^2 + 2q_e(1-q_e)] + \delta - c \geq 0 \\ q_e = c \end{cases}$$

The two conditions boil down to

$$c^2(1-2\delta) + c(\delta-1) + \delta \geq 0$$

The roots are

$$c_1 = \frac{2\delta - 2}{2\delta - 1}$$
$$c_2 = \frac{2\delta}{2\delta - 1}$$

Then if $\delta < 1/2$, the relevant root is c_1 as $c_1 > 0$ and $c_2 < 0$ in this case. However, $c_1 > 2$ for $\delta < 1/2$, therefore, oligarchs always attack. Similarly, if $\delta > 1/2$, the relevant root is c_2 as $c_2 > 0$ and $c_1 < 0$ in this case. However, $c_2 > 1$ for $\delta > 1/2$, and oligarchs always attack as well. Finally, if $\delta = 1/2$, the two conditions boil down to $-0.5c + 0.5 \geq 0$ which always holds for any $c \in (0, 1)$, and the oligarchs attack as well.

Finally, upon observing the display of weak level of power, an oligarch attacks always as independently of others' decision, attacking yields $1 - c > 0$.

To conclude, the strongest type always survives with the display of power. The middle type survives if the élite are convinced by the display of power, hence, for any level of $\delta \in (0, 1)$, she takes this action. Finally, the weakest regime is always overthrown with or without displaying of strength. □

Proof of Proposition 5

Proof. Denoting the (symmetric) probability of attacking as r than the relevant indifference conditions of an agent i read given message m_3:

$$\frac{1 + (1 - \mu_2(m_3))\sigma_2(m_3) + (1 - \mu_1(m_3))\sigma_1(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)} r^2 + \frac{\mu_2(m_3)\sigma_2(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)}(2r - r^2) + \frac{\mu_1(m_3)\sigma_1(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)} - c_i \geq \frac{\mu_2(m_3)\sigma_2(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)} r^2 + \frac{\mu_1(m_3)\sigma_1(m_3)}{1 + \sigma_2(m_3) + \sigma_1(m_3)}(2r - r^2)$$

The indifference condition given m_2 is

$$\frac{\sigma_2(m_2) + (1 - \mu_1(m_2))\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)}(2r - r^2) + \frac{\mu_1(m_2)\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)} - c_i \geq \frac{\sigma_2(m_2) + (1 - \mu_1(m_2))\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)} r^2 + \frac{\mu_1(m_2)\sigma_1(m_2)}{\sigma_2(m_2) + \sigma_1(m_2)}(2r - r^2)$$

Finally, the indifference condition after m_1 is very simple since the citizens know exactly the regime type:

$$1 - c_i \geq 2r - r^2$$

Re-call that the consistency of beliefs about an attack by other agents should be consistent with the beliefs about the cost, $r_i = c_i$, and there is the consistency of message-sending beliefs with the strategy of the Dictator types, then the indifference conditions can be simplified. Specifically, given messages m_3, m_2, and m_1, respectively, it reads:

$$[1+\sigma_2(m_3)+\sigma_1(m_3)-3\sigma_2(m_3)^2]c^2+[2\sigma_2(m_3)^2-2\sigma_1(m_3)^2-1-\sigma_2(m_3)-\sigma_1(m_3)]c+\sigma_1(m_3)^2\geq 0$$

$$[3\sigma_1(m_2)^2-2\sigma_2(m_2)-2\sigma_1(m_2)]c^2+[\sigma_1(m_2)+\sigma_2(m_2)-4\sigma_1(m_2)^2]c+\sigma_1(m_2)^2\geq 0$$

$$c^2-3c+1\geq 0$$

First, I prove that pooling on the high message is indeed an equilibrium. Assume that the middle regime sends the pure-strategy message m_3, so $\sigma_2(m_3)=\mu_2(m_3)=1, \sigma_2(m_2)=\mu_2(m_2)=0$. The indifference condition after observing m_3 becomes

$$[\sigma_1(m_3)-1]c^2+[-2\sigma_1(m_3)^2-\sigma_1(m_3)]c+\sigma_1(m_3)^2\geq 0$$

The LHS is a concave function since $\sigma_1(m_3)-1<0$, the maximizer $c_{max}=\frac{2\sigma_1(m_3)^2+\sigma_1(m_3)}{2\sigma_1(m_3)-2}<0$ and the value at zero is positive, therefore, the agents attack for any c between zero and the positive root. The roots are

$$\begin{aligned} c_{\pm} &= \frac{2\sigma_1(m_3)^2+\sigma_1(m_3)\pm\sqrt{D}}{2\sigma_1(m_3)-2} \\ &= \frac{2\sigma_1(m_3)^2+\sigma_1(m_3)\pm 2\sigma_1(m_3)\sqrt{1+(\sigma_1(m_3)-1)^2}}{2\sigma_1(m_3)-2} \end{aligned}$$

Then the positive root is

$$c_{+}=\frac{2\sigma_1(m_3)^2+\sigma_1(m_3)-2\sigma_1(m_3)\sqrt{1+(\sigma_1(m_3)-1)^2}}{2\sigma_1(m_3)-2}<0.2\ \forall\sigma_1(m_3)$$

And the low type survives with probability $(1-c_+)^3$.

Moreover, since high and middle types send the high message, the indifference conditions for the citizens will be the same whether the low type sends a middle or a low message since she would be recognized with probability 1. Therefore, upon m_2 or m_1 sent by the low type, the indifference condition for the citizens is

$$c^2-3c+1\geq 0$$

The LHS is a convex function with both positive roots, therefore the agents attack for c between zero and the smaller root $\frac{3-\sqrt{5}}{2}$, and the probability of survival for the low type is $\left(\frac{\sqrt{5}-1}{2}\right)^3$. However, for all $\sigma_1(m_3)$, the threshold c_+ is lower than $\frac{3-\sqrt{5}}{2}$, therefore, the low type will always send m_3 as a reply to $\sigma_2(m_3) = 1$.

Now I check if $\sigma_2(m_3) = 1$ is also the best reply to $\sigma_1(m_3) = 1$. Given $\sigma_1(m_3) = 1$, the indifference condition upon observing the high message becomes:

$$[\sigma_2(m_3) + 2 - 3\sigma_2(m_3)^2]c^2 + [2\sigma_2(m_3)^2 - \sigma_2(m_3) - 4]c + 1 \geq 0$$

It can be seen that the LHS is positive for all $\sigma_2(m_3) \times c \in [0,1] \times [0,1/3]$, therefore, the highest probability with which an agent attacks is below 1/3 when $\sigma_2(m_3)$ is positive.

If the middle type sends any other message, middle or low, she is recognized with probability one, and the indifference condition of the citizens becomes

$$r^2 + 2r(1-r) - c \geq r$$

Given $r = c$, it boils down to

$$r^2 + 2r(1-r) - c \geq r$$

It holds for $c \leq 1/2$, therefore, the probability one citizen attacks is 1/2. Since it is higher than the probability to attack when all types send the same high message, the middle type will also send m_3 as a reply to $\sigma_1(m_3) = 1$ since it yields a higher survival probability. Therefore, $\sigma_1^*(m_3) = \sigma_2^*(m_3) = 1$ is indeed an equilibrium.

Now I prove that it is a unique equilibrium in pure strategies, for which I assume $\sigma_2(m_1) = 1$ and check for the best replies of middle and low types.

The indifference condition upon observing m_3 becomes

$$[\sigma_1(m_3) + 1]c^2 - [2\sigma_1(m_3)^2 + \sigma_1(m_3) + 1]c + \sigma_1(m_3)^2 \geq 0$$

It is a convex function, the minimizer $c_{min} = \frac{2\sigma_1(m_3)^2+\sigma_1(m_3)+1}{2(\sigma_1(m_3)+1)}$ is positive and the discriminant $D = 4\sigma_1(m_3)^4 + 2\sigma_1(m_3)^2 + 2\sigma_1(m_3) + 1$ is positive, therefore, the range of the cost for which the agents attack is between zero and the lower root:

$$c_- = \frac{2\sigma_1(m_3)^2 + \sigma_1(m_3) + 1 - \sqrt{4\sigma_1(m_3)^4 + 2\sigma_1(m_3)^2 + 2\sigma_1(m_3) + 1}}{2(\sigma_1(m_3) + 1)}$$

It can be seen that the root is lower than 1/4 for any strategy of the low type, and also lower than the threshold $\frac{3-\sqrt{5}}{2}$ under which the citizens attack when the low type sends the low message with probability one, therefore, as a reply to $\sigma_2(m_1) = 1$, the low type prefer the high message compared to the low one. The only other relevant strategy of the low type is to send the middle message as the middle type does. The indifference condition then becomes

$$[3\sigma_1(m_2) - 2\sigma_1(m_2) - 2]c^2 - [1 + \sigma_1(m_2) - 4\sigma_1(m_2)^2]c + \sigma_1(m_2)^2 \geq 0$$

It can be seen that for any positive $\sigma_1(m_2)$, the threshold for attacking is at least 1/2, therefore, the low type would always prefer to send the high message as a reply to the middle message of the middle type, so there can be no other equilibrium in pure strategies.

Now I prove that no other equilibrium exists and the equilibrium above is unique. I already excluded all pure strategy profile candidates, so the task boils down to showing that the Dictator types do not mix either.

Re-call the indifference conditions upon receiving the high message m_3 and the middle message m_2:

$$[1 + \sigma_2(m_3) + \sigma_1(m_3) - 3\sigma_2(m_3)^2]c^2 + [2\sigma_2(m_3)^2 - 2\sigma_1(m_3)^2 - 1 - \sigma_2(m_3) - \sigma_1(m_3)]c + \sigma_1(m_3)^2 \geq 0$$
$$[3\sigma_1(m_2)^2 - 2\sigma_2(m_2) - 2\sigma_1(m_2)]c^2 + [\sigma_1(m_2) + \sigma_2(m_2) - 4\sigma_1(m_2)^2]c + \sigma_1(m_2)^2 \geq 0$$

Consider the left-hand side expressions: $\mathbb{E}[u_i(a_i = 1|m_3) - u_i(a_i = 0|m_3)]$ and $\mathbb{E}[u_i(a_i = 1|m_2) - u_i(a_i = 0|m_2)]$, which represent the net benefits of the citizens from attacking the regime upon receiving the high and middle messages. These expressions depend on the strategies of the middle and low types. Firstly, I can analyze the first partial derivative of the net benefit $\mathbb{E}[u_i(a_i = 1|m_3) - u_i(a_i = 0|m_3)]$ with respect to the strategy of the middle type to send m_3:

$$\frac{\partial \mathbb{E}[u_i(a_i = 1|m_3) - u_i(a_i = 0|m_3)]}{\partial \sigma_2(m_3)} = -c(1-c) - 2\sigma_2(m_3)c(2-3c)$$

It is negative for $c < 2/3$, that is, the higher is the probability of the middle type to send the high message when $c < 2/3$ the lower is the net benefit of the citizens and the lower probability for them to attack. Secondly, I analyze the first partial derivative of the net benefit $\mathbb{E}[u_i(a_i = 1|m_2) - u_i(a_i = 0|m_2)]$ in the strategy of the middle type to send m_2:

$$\frac{\partial \mathbb{E}[u_i(a_i = 1|m_2) - u_i(a_i = 0|m_2)]}{\partial \sigma_2(m_2)} = c(1-2c)$$

It is positive for $c < 1/2$, that is, the lower is the probability of the middle type to send the middle message (or the higher is the probability of the middle type to send the high message), the lower is the expected net benefit from attacking and the lower incentives for the citizens are to attack.

Combining two facts together, the dominant strategy of the middle type for $c < 1/2$ regardless of the low type's behavior is always to send the high message, $\sigma_2(m_3) = 1$, since it reduces the probability of the citizens to attack after both high and middle message. Then the best reply of the low type to this strategy of the middle type is to send the high message as well.

Moreover, for $c > 2/3$, the dominant strategy of the middle type regardless of the low type's behavior is to always send the middle message, $\sigma_2(m_2) = 1$, since it reduces the net benefit of the attack of the citizens after both m_2 and m_3 but, as shown before, it cannot be a part of any equilibrium.

Finally, for $c \in [1/2, 2/3]$, the analysis for the middle type is less straightforward since there is no dominant strategy: increasing $\sigma_2(m_2)$ or $\sigma_2(m_3)$ increases the net benefit of the citizens after one message but decreases for another so I analyze the first partial derivative of the net benefits $\mathbb{E}[u_i(a_i = 1|m_3) - u_i(a_i = 0|m_3)]$ with respect to the strategy of the low type instead:

$$\frac{\partial \mathbb{E}[u_i(a_i = 1|m_3) - u_i(a_i = 0|m_3)]}{\partial \sigma_1(m_3)} = 2\sigma_1(m_3)(1-2c) - c(1-c)$$

It is negative for the interval of interest so the higher is the probability to send the high message of the low type the lower is the net benefit of the citizens to attack after observing the high message. Moreover, the first partial derivative of the net benefit $\mathbb{E}[u_i(a_i = 1|m_2) - u_i(a_i = 0|m_2)]$ with respect to the strategy of the low type after the middle message is instead:

$$\frac{\partial \mathbb{E}[u_i(a_i = 1|m_2) - u_i(a_i = 0|m_2)]}{\partial \sigma_1(m_2)} = -2\sigma_1(m_2)(1-4c+3c^2) - c(1-2c)$$

For the range of the cost of interest, it is positive, therefore, it is negative with respect to the strategy of the low type after the high message. Then, the dominant strategy of the low is to send the high message since it reduces both the net benefit of the citizens after high and middle messages: $\sigma_1(m_3) = 1$, and the best reply of the middle type is to send the high message as a reply as well. Then, no other equilibrium apart from pooling on the high message exists. □

Proof of Proposition 6

Proof. I start by finding the threshold values of the cost for which the oligarchs attack the regime upon receiving the middle message. The indifference condition of an oligarchs reads:

$$\frac{\sigma_2(m_2)+(1-\mu_1(m_2))\sigma_1(m_2)}{\sigma_2(m_2)+\sigma_1(m_2)}(2r-r^2)+\frac{\mu_1(m_2)\sigma_1(m_2)}{\sigma_2(m_2)+\sigma_1(m_2)}-c\geq 0$$

$$-c^2(\sigma_2(m_2)+\sigma_1(m_2)-\sigma_1(m_2)^2)+c(\sigma_2(m_2)+\sigma_1(m_2)-2\sigma_1(m_2)^2)+\sigma_1(m_2)^2\geq 0$$

It is a convex function, the discriminant is simply equal to $(\sigma_2(m_2)+\sigma_1(m_2))^2$ and the roots are $\frac{-\sigma_1(m_2)^2}{-2(\sigma_2(m_2)+\sigma_1(m_2)-\sigma_1(m_2)^2)}$ which is a negative value and 1, so the oligarchs would attack for any value of the cost if they observe the middle message.

Then the middle type definitely sends the high message with probability one and the low type can mix between the high and low messages.

If the low type sends the low message with any strictly positive probability, she is recognized with probability one, and the indifference condition of the oligarchs is simply $1-c\geq 0$ which always holds, so the always attack. Therefore, the only possible remaining case is when both middle and low types send the same message m_3, therefore, the posterior of the citizens coincides with the prior. The probability to attack of any oligarch is one as proved before, therefore, it cannot be an equilibrium if there is any other profile that yields a strictly positive probability of survival.

Then there is no equilibrium where any regime type can have a positive probability of survival and, technically, all strategy profiles are in equilibrium since they yields the same zero payoff to the regime types. □

Chapter 2

Trust but Verify: Quality Authentication in eCommerce

2.1 Introduction

eCommerce platforms have become an extremely convenient alternative to the standard offline trade as a result of their multiple benefits for both the sellers and the buyers. For the sellers, the advantages come from cost minimization and the ability to reach a higher customer base since there is no need to maintain the sales or marketing department as well as the offline stores while the potential buyers are more numerous. Moreover, it gives the opportunity to re-sell used products which are becoming a steadily growing trend for both economic and environmental reasons. As for the customers, online retail provides a substantial search cost minimization and gives access to a larger number of retailers.

However, despite all the obvious advantages of eCommerce, they suffer from a substantial drawback which is the aggravating asymmetry of information between the seller and the buyer, especially when it comes to trading second-hand goods. More specifically, the quality of the goods on sale is more difficult to verify compared to a standard offline store as there is no direct way to control the quality of the good in person before the trade takes place and the product is delivered.

One well-studied way to deal with such an issue is to make the seller provide (possibly) verifiable evidence to prove that the actual quality indeed corresponds to the stated one. The verifiable disclosure literature is quite vast and can be divided into the class of games where the evidence is verifiable fully and partially, started by Grossman (1981), Milgrom (1981) and Dye (1985), respectively. The other way is to offer the buyer an opportunity to acquire information about the product, for instance, from an intermediary which can testify to the quality started by Matthews and Postlewaite (1985).

In this paper, we combine these two approaches to verify if both of them together are able to overcome the asymmetry of information when quality level is perfectly known only by the seller but not the buyer, and, as a result, the market failure when the first best prices are unattainable. Specifically, to make the model as credible as possible when notion of the eCommerce is analysed, we assume two main mechanisms to deal with information asymmetry: firstly, the quality disclosure provided by the seller is not fully verifiable ex-ante as it is often on the reselling platforms, secondly, the buyer can acquire additional verification of the quality at a certain cost which is fully reliable (again, as it is often allowed on the reselling platforms).

Related literature: Extensive research has been done on information acquisition by uninformed buyers [1]. Jackson (1991) tries to deal with the Grossman and Stiglitz paradox: when buyers are price takers and can acquire information about quality while prices are fully revealing, information acquisition does not occur in equilibrium. He relaxes the price-taking assumption to solve the paradox and get fully revealing prices and costly info acquisition. Bester and Ritzberger (2001) study a trade model with multiple buyers who have the option of perfect information acquisition upon observing the price of the good. They prove the Grossman–Stiglitz Paradox in the equilibrium with pure strategies under the Intuitive Criterion of Cho and Kreps but show that mixed strategies allow to

[1] Dranove and Jin (2010) provide a comprehensive review of quality disclosure and certification literature up to 2010.

approach the full information prices. Gertz (2016) analyses a similar model with a single buyer who can choose the amount (precision) of acquired information. He shows that information acquisition can lead to more efficient outcomes, but that there are limits to how much information can be acquired. Martinez-Gorricho (2020) studies the same problem where information acquisition is costless but always imperfect. Voorneveld and Weibull (2011) also study a case with incomplete information on both sides of the market where the buyer does not know the exact quality of the good but receives a private costless signal. Similarly, Figueroa and Guadalupi (2015) study a bilateral trade model where the quality of the good is not known to either but the seller holds a more informative prior about the quality while the buy can acquire a costly but imprecise signal about the quality. Hauswald and Marquez (2006) analyze a screening problem where competing decision-makers can invest in acquiring private information about the agents. Argenziano et al. (2016) study an environment where uninformed decision-makers can acquire private information from a biased expert showing that it is not always more efficient than direct information acquisition by the decision-maker. Persico (2000) studies the acquisition of information about the value of the object for sale in an auction by the bidders and sees in which auction format such acquisition occurs. Martin (2017) studies the pricing strategies when sellers face two types of irrational buyers who can be rationally inattentive and naive and finds that in both cases there are equilibria where information revelation fails. Similarly, Gabaix et al. (2006) study how the information acquisition process differs in limited rationality models compared to standardly assumed agents with full rationality and test the predictions in a lab experiment. Roesler and Szentes (2017) analyze a bilateral trade environment with take-it-or-leave-it offers where the buyer is uncertain about her valuation of the object but can be signaled about it. Stahl and Strauz (2017) add a third-party certification device available both for the seller and the buyer but not at the same time and find that seller-induced certification is more efficient in terms of total welfare when transparency of the market is beneficial for the welfare. Metthews and Persico (2005) study a bilateral trade of refundable goods, that is, a dissatisfied buyer can be reimbursed for a good if she wants, which is similar to our setting. Study if information can be acquired efficiently in a mechanism before agents decide to participate in it finding that it is generally achievable in private-value environments but not with common values. Bergemann and Välimäki (2002) see if information can be acquired efficiently in a mechanism before agents decide to participate in it finding that it is generally achievable in private-value environments but not with common values. Moscarini and Ottaviani (2001) characterize prices and profits in separating, pooling, and mixed-strategy equilibria in a Bertrand competition with the buyer's private information about her valuation when there is a common prior shared by the sellers and the buyer and a private signal only for the buyer. Liu (2011) studies a dynamic game where uninformed buyers can costly learn about sellers' product quality from their past trades. Guan and Chen (2017) analyze a two-sided information asymmetry in the producer-retailer problem: the producer knows the exam quality of the product while the retailer holds a prior about it; the buyer privately knows the preference of consumers for quality. The producer can acquire information (publicly) about the preferences at one cost and disclose information truthfully about the quality at a different cost. Finally, for a more detailed review see Capozza et al. (2021) who provide a survey on information acquisition literature across all the subfields of economics.

Finally, one of the two closest papers to our model is Hong et al. (2021) who analyze a similar problem to ours within a trade setting in a supply chain. Each agent can acquire information about the quality of the good from her upstream suppliers and disclose information to her downstream suppliers. The paper, however, does not fully characterize the equilibria nor give conditions for optimal information disclosure, which is the aim of our paper. Bester et al. (2019) study a competitive labor market application where both the workers can provide a costly (unlike our model) signal of their ability a la Spence while the firms can learn at a cost (audit) the exact ability of the workers. The auditing cost is always borne by the firms.

Indeed, there appears to be no work on combining the two approaches of information disclosure by one party and information acquisition by the second party with providing a full analysis. This paper will therefore contribute to the literature on both evidence provision and costly information verification, as a combination of the two. It is widely used by eCommerce platforms that are asking the sellers to provide the product description with photographs of the good and also allow the buyer to acquire additional authentication tests during the transaction.

However, in terms of the theoretical contribution to the literature on information acquisition from an uninformed buyer, the paper is meant to answer if the voluntary disclosure by an informed seller solves the Grossmann-Stiglitz paradox. Specifically, if buyers are price takers and can acquire information about the unknown quality while the prices are fully revealing, then information acquisition does not occur in the equilibrium.

To answer the above question, we develop an asymmetric information model that combines information acquisition and imperfectly verifiable quality disclosure. In our model of unilateral trade of a single indivisible good, the seller can provide partially verifiable evidence about privately known quality together with the price ex-ante while the buyer, upon observing a bundle "price - quality level" can request the costly authentication of the quality, that is, if the declared quality matches the asked price. We characterize the equilibrium strategies of seller and buyer for both pooling and separating equilibria in pure strategies and both necessary and sufficient conditions for each of the equilibria we find, which enables us to answer the question if the Grossmann-Stiglitz paradox can be dealt with in our model. What we find is that there exists the possibility of (partially) overcoming the Grossmann-Stiglitz paradox for a specific range of model parameters, and show that voluntary disclosure of partially verifiable information can solve it if the cost of authentication set by the platform belongs to that range. Additionally, to deal with the multiplicity (that is, infinity in our model) of equilibria, we adapt Bester and

Ritzberger (2001)'s extension of the intuitive criterion and apply it to our model to gain more robust predictions. Under the intuitive criterion, only two plausible equilibrium outcomes survive, which can also be efficient for some parameter combinations. Finally, we analyze the effect of imperfect authentication technology when the the buyer even after acquiring additional information on the quality, gets only a noise signal. The results show that in this case quality authentication can be acquired in equilibrium, unlike the perfect precision case.

The rest of the paper is structured as follows. In Section 2, we present the model and introduce the necessary notation. In Section 3, we describe the optimal behavior of the seller and the buyer. Section 4 provides the equilibrium analysis. Section 5 concludes.

2.2 Model

We consider an asymmetric information model with a buyer and a seller. The latter holds private information over the actual quality of a good. While setting the price, the seller can also provide some evidence about the quality of the product which generates a quality signal prior to the trade. This evidence technology is costless and monotone in quality so that, for a given level of evidence, a high-quality signal is more likely when the actual quality is high than when it is low. We will make this assumption explicit in the next paragraphs.

Upon observing the price and the evidence, the buyer has a few decisions to make. Firstly, she can order a costly quality authentication from an intermediary to check, at least to some extent, the value of the good. We consider two different types of authentication technology: perfect authentication, which always signals the actual quality, and imperfect authentication, which entails some degree of error. Then, having acquired the authentication or not, she decides if she wants to buy the good or not.

In the last part of the paper, we will consider an augmented game in which a third player, a profit-maximizer authenticator, can commit to an authentication structure, defining both the precision and the cost of the authentication, before the seller-buyer subgame takes place.

We now provide a formal outline of the setting and the assumptions of the model.

Preliminaries. We consider a good whose quality can be either low or high. Let $\theta \in \Theta = \{\theta_l, \theta_h\}$ denote the low and high quality, respectively, with $0 < \theta_l < \theta_h < \infty$, and $\Delta\theta := \theta_h - \theta_l$, representing the quality difference between the two types of good. The prior is $\mu_0 = \mathbb{P}(\theta = \theta_h)$. Both the seller and the buyer are risk-neutral. The seller does not value the good per se but is only interested in selling the good at the highest price p possible, while the buyer gets utility from the value of the product minus the price she pays, net of any other additional costs. The reservation utility in the case of non-trade is common to both players and equals zero. Clearly, efficiency would require that the trade always takes place.

The realization of the quality θ is private information of the seller, while the buyer holds a prior over its distribution, with $\mu_0 := \mathbb{P}(\theta = \theta_h)$. Moreover, the buyer uses Bayesian updating to form her posterior belief about the seller's type, whenever possible.

Seller's strategies and partially verifiable evidence. After observing the quality of the good, the seller needs to choose a price $p \in \mathbb{R}_+$ and an evidence level $e \in \{e_l, e_h\}$ where e_i represents evidence of the good being of quality θ_i. Allowing for mixed strategies, we define a strategy for the seller as a probability distribution over price-evidence vectors, and we denote by $\sigma_i(p, e) := \sigma(p, e|\theta_i)$ the probability assigned by the seller strategy to the vector (p, e) by the seller of type θ_i. Providing any type of evidence is costless regardless of the actual quality. However, the evidence can, in certain conditions, reveal the type of the seller, or equivalently, be verified.

We model this idea by assuming that the buyer not only observes the (p, e) vector prescribed by the seller strategy but also the realization of a binary (evidence) signal

$S^e \in \{S^e_l, S^e_h\}$ whose probability distribution depends jointly on the actual seller's quality θ and on the evidence provided by the seller e. In particular, we denote the conditional probability distributions of S^e by

$$\mathbb{P}(S^e_h|e_h, \theta_h) = \phi \qquad \mathbb{P}(S^e_h|e_h, \theta_l) = \chi$$
$$\mathbb{P}(S^e_h|e_l, \theta_h) = \psi \qquad \mathbb{P}(S^e_h|e_l, \theta_l) = \omega$$

and, accordingly, $\mathbb{P}(S^e_l|e, \theta) = 1 - \mathbb{P}(S^e_h|e, \theta)$ for every e and θ. Without loss of generality, we assume $\phi, \in \left[0, \frac{1}{2}\right]$,[2] so that when the high type seller provides high evidence, S^e_h is more likely to occur thanS^e_l. As stated above, we also assume the evidence technology to be monotone in the actual quality, i.e. $\phi \geq \chi$ and $\psi \geq \omega$. Given the same evidence, the high seller sends the high signal (weakly) more often than the low seller.

To simplify the analysis we assume that, when high evidence is provided, the high seller always conveys the high signal while the low type does not. Moreover, we assume the low evidence to be uninformative.

Assumption 1 (Evidence technology). *The conditional distribution of the evidence signal S^e is characterized by:*

(i) $\chi < \phi$ *or, equivalently,* $\chi = \alpha\phi$ *with* $\alpha \in (0, 1)$*,*

(ii) $\phi = 1$*,*

(ii) $\psi = \omega$*.*

We can interpret providing e_l as if the seller does not disclose evidence and thus the buyer cannot get any additional information about the good quality when she observes e_l. On the contrary, e_h generates a different probability of a high signal depending on the actual type. In particular, the buyer is able to detect a low seller providing high evidence with

[2] Otherwise, we could simply relabel the signal realizations.

probability $1 - \alpha = \mathbb{P}(S_l^e | e_h, \theta_l)$ since the high-quality seller never induces a low signal with high evidence. We will see that parameter α plays a key role in the analysis.

Beliefs. Given a strategy (σ_h, σ_l) of the seller, and an evidence signal realization, the buyer updates his beliefs about the good being of high quality following Bayes' rule whenever possible. Hence, the posterior beliefs upon observing a price p, given (σ_h, σ_l) are

$$\begin{aligned}
\mu(p, e_h, S_h^e) &= \frac{\mu_0 \sigma_h(p, e_h)\phi}{\mu_0 \sigma_h(p, e_h)\phi + (1-\mu_0)\sigma_l(p, e_h)\chi} \\
\mu(p, e_h, S_l^e) &= \frac{\mu_0 \sigma_h(p, e_h)(1-\phi)}{\mu_0 \sigma_h(p, e_h)(1-\phi) + (1-\mu_0)\sigma_l(p, e_h)(1-\chi)} \\
\mu(p, e_l, S_h^e) &= \frac{\mu_0 \sigma_h(e_l)\psi}{\mu_0 \sigma_h(p, e_l)\psi + (1-\mu_0)\sigma_l(p, e_l)\omega} \\
\mu(p, e_l, S_l^e) &= \frac{\mu_0 \sigma_h(p, e_l)(1-\psi)}{\mu_0 \sigma_h(p, e_l)(1-\psi) + (1-\mu_0)\sigma_l(p, e_l)(1-\omega)}
\end{aligned}$$

depending on the evidence provided and its signal realization whenever $\sigma_h(p, e) + \sigma_l(p, e) > 0$ for the corresponding (p, e). Given assumption 1, the posterior beliefs simplify to

$$\mu(p, e_h, S_h^e) = \frac{\mu_0 \sigma_h(p, e_h)}{\mu_0 \sigma_h(p, e_h) + (1-\mu_0)\sigma_l(p, e_h)\alpha} \tag{2.1}$$

$$\mu(p, e_h, S_l^e) = 0 \tag{2.2}$$

$$\mu(p, e_l, S_h^e) = \mu(p, e_l, S_l^e) = \frac{\mu_0 \sigma_h(e_l)}{\mu_0 \sigma_h(p, e_l) + (1-\mu_0)\sigma_l(p, e_l)} \tag{2.3}$$

whenever they are defined. Hence, low evidence e_l is uninformative on the type of the seller, while high evidence can be fully informative whenever the signal realization is low.

Quality authentication and buyer's strategies. In addition to the information possibly disclosed by the seller, the buyer has available a product authentication technology at a cost $c > 0$, which reveals additional information on the quality of the good. In particular, if the buyer purchases the authentication, she receives a binary authentication signal $S^a \in \{S^a_l, S^a_h\}$ whose distribution depends only on θ and is independent on S^e. The conditional distribution of S^a is defined by:

$$\mathbb{P}(S^a_h|\theta_h) = \eta \qquad\qquad \mathbb{P}(S^a_h|\theta_l) = \varepsilon$$

Similarly to the evidence-signal case, we assume, without loss of generality $\eta \in \left[0, \frac{1}{2}\right]$, and the authentication technology to be monotone in the good's quality, i.e. $\eta \geq \varepsilon$. Moreover, we limit our analysis to the case in which the authentication of the high-quality good always relieves the actual type, while we consider two different outcomes when the authentication is ordered on a low-quality good.

Assumption 2 (Authentication technology)**.** *The conditional distribution of the authentication signal S^a is characterized by $\zeta = 1$ and, alternatively,*

(i) $\varepsilon = 0$ in which case we define the authentication technology as perfect.

(ii) $\varepsilon \in (0, 1)$ in which case we define the authentication technology as imperfect.

Clearly, when $\varepsilon = 0$, acquiring the authentication fully reveals the type of seller to the buyer. When ε is positive, instead, there is a positive probability that a low-quality good pass the test, sending an authentication signal S^a_h. In this regard, we can consider ε as the authentication imprecision level, or error term.

Given this additional technology, the buyer has three alternatives upon observing (p, e, S^e): buying the good without authentication (b), not buying the good (n), acquiring the authentication technology and the buy the good if and only if $S^a = S^h$ (ba).[3] Therefore, given any realization (p, e, S^e) and the corresponding posterior $\mu := \mu(p, e, S^e)$, we define a strategy for the buyer as a probability distribution over the set of actions $\{b, n, ba\}$,

[3] As already pointed out in Stahl and Strausz (2017) for a perfect authentication technology environment, any other action, namely, acquiring the authentication and always buying the good, acquiring the authentication and never buying the good, acquiring the authentication and buying if and only if $S^a = S^l$ are dominated by the first three alternatives and hence are disregarded in the analysis.

and we denote by $\beta(x|p,\mu)$ the probability assigned by the buyer's strategy to action $x \in \{b, n, ba\}$.

Timing, payoffs, and solution concept. The game unfolds as follows:

1. Nature choose the quality of the good θ according to the prior distribution μ_0.
2. The seller observes θ and chooses a probability distribution over vectors (p, e).
3. The evidence signal S^e is realized according to its conditional probability distribution.
4. The buyer observes (p, e, S^e), (possibly) updates her belief according to (1), (2), and (3), and chooses a probability distribution over her available action set $\{b, n, ba\}$.
5. If $x = ba$, the authentication signal S^a is realized.
6. Finally, payoffs for the buyer, u, and the seller, π, are realized:
   - If $x = b$, the trade takes place and payoffs are: $u = \theta - p$ and $\pi = p$.
   - If $x = n$, there is no trade and payoffs are: $u = \pi = 0$.
   - If $x = ba$ and $S^a = S^a_h$, the buyer bears the cost of the authentication and the trade takes place and payoffs are: $u = \theta - p - c$ and $\pi = p$.
   - If $x = ba$ and $S^a = S^a_l$, the buyer bears the cost of the authentication but there is no trade and payoffs are $u = -c$ and $\pi = 0$.

The game can be solved by backward induction and our analysis considers (weak) Perfect Bayesian Equilibria (PBE) which specify the optimal strategies of the agents and the system of equilibrium beliefs consistent with the strategies and Bayes' rule.

First, we focus on pure-strategy equilibria. The procedure adopted to solve the model involves a series of steps. We specify candidate price-evidence strategies for low- and high-type sellers. Then, we determine the equilibrium posterior upon observing the equilibrium price-evidence vector of the buyer and her optimal strategy. This allows us to determine the optimal pricing posted by the two types of sellers in equilibrium. Finally, we check for possible deviations and, potentially, find necessary conditions determining the subset of the parameter space allowing for such equilibria.

Secondly, we consider the additional mixed-strategy equilibria that can arise in this environment.

2.3 Buyer behaviour and seller's profits

Let us first analyze what happens in the last stage of the game, when the buyer is called to play. Given any vector realization (p, e, S^e) and seller strategy (σ_h, σ_l), the buyer holds beliefs $\mu(p, e, S^e)$ on the good being of high quality according to (1), (2), and (3). Similarly to Gertz (2014), we define the expected quality of the good given some belief $\hat{\mu}$ as

$$\bar{\theta}_{\hat{\mu}} := \hat{\mu}\theta_h + (1 - \hat{\mu})\theta_l$$

and, accordingly, we denote the average quality given prior μ_0 as $\bar{\theta}_{\mu_0}$.

Assume that the buyer holds posterior $\hat{\mu}$ after observing (p, e, S^e). Then, the expected utility attainable by each action of the buyer is:

$$u(b|p, \hat{\mu}) = \bar{\theta}_{\hat{\mu}} - p \tag{2.4}$$

$$u(n) = 0 \tag{2.5}$$

$$u(ba|p, \hat{\mu}) = \hat{\mu}(\theta_h - p) + (1 - \hat{\mu})\varepsilon(\theta_l - p) - c \tag{2.6}$$

While (4) and (5) are straightforward, let us interpret (6). When the buyer acquires the authentication, the probability of receiving a high authentication signal S_h^a is one when $\theta = \theta_h$, and ε when $\theta = \theta_l$. Moreover, we can rewrite $u(ba|p, \hat{\mu})$ as $\bar{\theta}_{\hat{\mu}} - p - (1 - \hat{\mu})(1 - \varepsilon)(\theta_l - p) - c$. The following lemma characterizes the optimal strategy of the buyer as a function of p

and belief $\hat{\mu}$ held after observing a vector (p, e, S^e).

Lemma 1. *Suppose the strategy β is optimal for the buyer given price p and belief $\hat{\mu}$. Then,*

$$\beta(b|p,\hat{\mu}) \geq 0 \iff p \leq \min\left\{\bar{\theta}_{\hat{\mu}},\ \theta_l + \frac{c}{(1-\hat{\mu})(1-\varepsilon)}\right\} \tag{2.7}$$

$$\beta(ba|p,\hat{\mu}) \geq 0 \iff p \in \left[\theta_l + \frac{c}{(1-\hat{\mu})(1-\varepsilon)},\ \frac{\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c}{\hat{\mu} + (1-\hat{\mu})\varepsilon}\right] \tag{2.8}$$

$$\beta(n|p,\hat{\mu}) \geq 0 \iff p \geq \max\left\{\bar{\theta}_{\hat{\mu}},\ \frac{\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c}{\hat{\mu} + (1-\hat{\mu})\varepsilon}\right\} \tag{2.9}$$

While the proof of the lemma is straightforward, its implications are twofold. First, there exist three price thresholds that determine the optimal strategy of the buyer. Second, there exists an upper bound for the authentication costs above which acquiring the authentication is not optimal independently of the price.

Corollary 1. *Acquiring the authentication can be optimal for the buyer only if*

$$c \leq \hat{\mu}(1-\hat{\mu})(1-\varepsilon)(\theta_h - \theta_l) =: \bar{c}.$$

When the cost of the authentication is too high, in a sense specified by the corollary, acquiring the authentication is not optimal independently of the price, and the best of the other two alternatives, simply buying or not buying, yields a higher payoff to the buyer. Notice that the threshold $\bar{c}$ depends positively on the quality differential, which measures the opportunity cost of receiving θ_l, the posterior diffusion, which measures uncertainty on the true quality, and the precision of the signal $1-\varepsilon$, as the authentication technology becomes more valuable to the buyer. Moreover, whenever uncertainty is resolved, i.e. $\hat{\mu} \in \{0,1\}$, buying the authentication is clearly dominated since the buyer is already fully informed on the quality of the good.

Consider now a seller with strategy (σ_l, σ_h) and a buyer with belief system μ and strategy

β. Then, choosing a price-evidence vector (p, e), when the object quality is θ, yields to the following expected profits:

$$\pi(p, e|\theta) = \sum_{S^e \in \{S^e_l, S^e_h\}} \mathbb{P}(S^e|e, \theta) \left[\beta(b|p, \mu(p, e, S^e)) + \mathbb{P}(S^a_h|\theta_h)\beta(ba|p, \mu(p, e, S^e))\right] p \quad (2.10)$$

which depends on the conditional probabilities of the evidence signal, the probability assigned to b by the buyer strategy, the conditional probability of the high authentication signal times the probability assigned to ba by the buyer strategy, and the price p. Given our assumptions 1 and 2 and the expression for the posteriors we derived in (1), (2), and (3) we have

$$\begin{aligned}
\pi(p, e_h|\theta_h) &= \left[\beta(b|p, \mu(p, e_h, S^e_h)) + \beta(ba|p, \mu(p, e_h, S^e_h))\right] p \\
\pi(p, e_l|\theta_h) &= \left[\beta(b|p, \mu(p, e_l, S^e_h)) + \beta(ba|p, \mu(p, e_l, S^e_h))\right] p \\
\pi(p, e_h|\theta_l) &= \alpha \left[\beta(b|p, \mu(p, e_h, S^e_h)) + \varepsilon\beta(ba|p, \mu(p, e_h, S^e_h))\right] p + (1-\alpha)\beta(b|p, 0)p \\
\pi(p, e_l|\theta_l) &= \left[\beta(b|p, \mu(p, e_l, S^e_h)) + \varepsilon\beta(ba|p, \mu(p, e_l, S^e_h))\right] p
\end{aligned}$$

Then, by adopting a strategy (σ_h, σ_l) the high and the low type of the seller expect payoffs equal to

$$\begin{aligned}
\pi(\sigma_h|\theta_h) &= \sum_{(p,e)} \sigma_h(p, e)\pi(p, e|\theta_h) \\
\pi(\sigma_l|\theta_l) &= \sum_{(p,e)} \sigma_h(p, e)\pi(p, e|\theta_l)
\end{aligned}$$

Finally, we say that $\{\sigma^*_h, \sigma^*_l, \beta^*, \mu^*\}$ represent a PBE of the game if σ^*_h, and σ^*_l maximizes the expected payoffs of the high and the low-quality seller respectively, β^* maximizes the expected payoffs of the buyer, and the system of belief μ^* is consistent with σ^*_h, σ^*_l, β^* and Bayes' rule whenever possible.

2.4 Equilibrium Analysis

2.4.1 Benchmark: no evidence or authentication

Before we proceed with the fully-fledged analysis with evidence and authentication, let us describe the equilibrium in pure strategies where no such technology is available. Firstly, under complete information, the two types set the prices equal to their respective qualities: $p(\theta_h) = \theta_h$ and $p(\theta_l) = \theta_l$ while the buyer participates in the trade. The outcome is also efficient.

Under incomplete information, only pooling on the expected quality can be sustained in the equilibrium:

Proposition 1. *Under no evidence and no authentication, in the unique equilibrium, it holds* $p(\theta_h) = p(\theta_l) = p^* = \mu_0\theta_h + (1-\mu_0)\theta_l$ *with a posterior belief* $\mu := \mathbb{P}(\theta = \theta_h|p) = \mu_0 \ \forall \ p \geq p^*$.

Indeed, separation cannot be obtained since the low type will always mimic the high type and set equal prices resulting in the buyer's negative utility. Pooling on a lower price leads to lower profits for both types while pooling on a higher price leads to zero profits since the buyer will not participate in the trade.

We now turn to the equilibrium analysis in the original model. We first analyze equilibria in pure strategy, with both perfect and imperfect authentication technology. Then we considered the same cases when extending the possibility of mixing between strategies.

2.4.2 Pure-strategy equilibria with perfect authentication technology

In the first part of this section, we study the model where the authentication technology acquired by the buyer is perfect, that is, it provides the degenerate distribution with mass

1 on the true type. We begin by characterizing the pure-strategy equilibria and providing the necessary conditions for their existence. We consider first separating and then pooling PBEs. The main result is that authentication is never bought in pure-strategy equilibria, and the parameter spaces allowing for pooling-on-the-high type and separating equilibria are not overlapping.

For ease of analysis of the buyer's out-of-equilibrium beliefs, we will consider them to be evidence-dependent and price-independent. Specifically, we assume that, for any price-evidence vector (p, e) different from the equilibrium vector(s), the posterior of the buyer has the following structure

$$\mu(p, e) = \begin{cases} \delta & \text{if } e = e_h \\ \\ \gamma & \text{if } e = e_l \end{cases}$$

where δ and γ are values to be specified in each equilibrium characterization[4]. This means that the buyer evaluates the likelihood of facing a high-quality good off-path given the evidence provided rather than the price posted. This is consistent with the fact that only the evidence disclosed by the seller can possibly reveal information about the good quality, while the price cannot. The assumption also simplifies the analysis, since the evidence-space is binary, and makes the problem tractable without the need for additional hypotheses and how the buyer forms his out-of-equilibrium expectations when observing a price different from the equilibrium price. We will re-evaluate this assumption later on.

[4] Note that analyzing out-of-equilibrium beliefs is not strictly necessary in a signaling setting such as the one we are considering here, so this assumption is rather innocuous. In fact, it is always possible to sustain a large number of equilibria given out-of-equilibrium beliefs that are *enough* pessimistic. We will drop this assumption in the relevant section on mixed-strategy equilibria when we will use an adaptation of the refinement criterion developed in Bester and Ritzberger (2001).

Separating equilibrium

In a separating equilibrium, different types of sellers provide different price-evidence combinations so that the quality of the good is revealed to the buyer. With an abuse of notation, let us define with $(p^h, e^h) = \sigma(\theta_h)$ and $(p^l, e^l) = \sigma(\theta_l)$ the equilibrium strategies of low and high type sellers in the pure-strategy context. Then, in a separating equilibrium, it must be that $(p^h, e^h) \neq (p^l, e^l)$ so that $\mu(p^h, e^h) = 1$ and $\mu(p^l, e^l) = 0$.
A natural candidate for a separating equilibrium is the one in which the high type of seller provides evidence of high quality, while the low one provides low-quality evidence. We will now characterize such separating equilibria.

Proposition 2. *The following strategies of the agents and beliefs of the buyer represent a continuum of separating PBEs:*

- $\sigma(\theta_h) = (\theta_h, e_h)$, $\sigma(\theta_l) = (\theta_l, \theta_l)$;
- $\mu((\theta_h, e_h)) = 1$, $\mu((\theta_l, \theta_l)) = 0$;
- $\delta = \mu((p, e_h)) \in [0,1]$ *with* $p \neq \theta_h$, *and* $\gamma = \mu((p, e_l)) = 0$ *if* $p \neq \theta_l$;
- $\beta(\theta_h, e_h) = b$, $\beta(\theta_l, \theta_l) = b$;
- $$\beta(p, e_l) = \begin{cases} b & \textit{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ nb & \textit{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases} \quad \textit{if } \; p \neq \theta_l;$$
- $$\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ nb & \textit{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases} \quad \textit{if } \; \delta \geq 1 - \tfrac{c}{\Delta\theta} \;\; \textit{and} \;\; p \neq \theta_h;$$
- $$\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \theta_l + \frac{\delta}{1-\delta}c \\ ba & \textit{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ nb & \textit{if } p > \theta_h - c \end{cases} \quad \textit{if } \; \delta < 1 - \tfrac{c}{\Delta\theta} \;\; \textit{and} \;\; p \neq \theta_h.$$

Moreover, a necessary condition for such equilibria is:

$$\Delta\theta \leq \frac{1-\alpha}{\alpha}\theta_l \iff \alpha(\theta_h - \theta_l) \leq (1-\alpha)\theta_l \tag{2.11}$$

Firstly and mainly, we are able to state that the first-best prices in general are achievable but not for the whole parameter space. The condition under which separation is obtained comes from the incentive compatibility of the low type and relies on the fact that the expected loss from unsuccessful high-evidence mimicking is greater than the potential gain from price increase to $p = \theta_h$:

$$\alpha(\theta_h - \theta_l) \leq (1-\alpha)\theta_l$$

The left-hand side of the inequality is the expected net gain of the low type if he successfully provides high-quality evidence while the right-hand side is the expected net loss. Clearly, the higher is the quality variance the higher is the incentive for the low type to risk mimicking the high type. Moreover, recall that α is the probability of trade when the declared quality-evidence exceeds the actual quality. This probability can be interpreted as the ability of the low type to pass as the high type and fool the customer, alternatively, $1-\alpha$ can be interpreted as the probability of returns if the customer is able to prove to the platform the mismatch between the evidence and the real quality. The higher is the probability of such returns the lower is the incentive for the low-quality seller to lie and mimic the high type.

Finally, note that the cost of authentication does not enter in the condition since, in the equilibrium, once the seller chooses the type-specific price, the buyer knows with certainty the quality he faces and saves on the cost of authentication. This cost, instead, enters in off-the-equilibrium price deviations of the low type: if he sets the price below θ_h but still above the real quality θ_l, the strategy of the buyer must be such that if she expects the low type with relatively high probability her threat of using authentication prevents the low type from doing so.

The next result shows that the set of separating equilibria characterized above is the unique set of such equilibria.

Proposition 3. *All separating equilibria are characterized by Proposition 1.*

Above we stated that the necessary condition for the existence of separating equilibria is that the distance between θ_h and θ_l is small, namely $\Delta\theta \leq \frac{1-\alpha}{\alpha}\theta_l$. The next session will argue that this condition is also sufficient: if an equilibrium exists, in the parameter region defined by condition (1), then it must be a separating equilibrium.

Pooling Equilibria

We now turn our attention to equilibria in which different types of sellers provide the same price-evidence vector. Unlike the separating case, the buyer might have the incentive to buy the authentication when purchasing the good in order to avoid disguised low-quality goods. However, the perfectly precise nature of the authentication technology prevents this to happen in equilibrium. We show that this result holds for pooling on high-quality evidence equilibria first, and then we do the same for the pooling on low-quality evidence counterpart.

Pooling on high-quality evidence. First, we consider pooling equilibria in which both types present evidence of high quality. Necessary conditions for such equilibria require both an upper and a lower bound for $\Delta\theta$. On the one hand, the quality difference must be large enough to encourage low-type sellers to mimic the high type despite the evidence technology could reveal the provided untruthful evidence. On the other hand, $\Delta\theta$ needs to be small compared to the authentication cost c, otherwise, the high type could have an incentive to deviate and set the maximum possible price with authentication, $\theta_h - c$. This happens because the high type does not fear facing the authentication process. We show that these conditions are not compatible with the necessary conditions for the existence of separating equilibria. Hence, the parameter regions of separating and pooling on θ_h equilibria are not overlapping. The next proposition characterizes this first set of pooling equilibria.

Proposition 4. *The following strategies of the agents and beliefs of the buyer represent a continuum of pooling PBEs on θ_h:*

- $\sigma(\theta_h) = \sigma(\theta_l) = (\bar{\theta}_{\tilde{\mu}}, \theta_h)$, *where* $\bar{\theta}_{\tilde{\mu}} := \tilde{\mu}\theta_h + (1-\tilde{\mu})\theta_l$, *and* $\tilde{\mu} := \frac{\mu_0}{\mu_0 + (1-\mu_0)\alpha}$;
- $\mu(\bar{\theta}_{\tilde{\mu}}, \theta_h) = \tilde{\mu}$;
- $\delta = \mu((p, e_h)) \in [0, \tilde{\mu}]$ *with* $p \neq \bar{\theta}_{\tilde{\mu}}$, *and* $\gamma = \mu((p, e_l)) \in [0, \alpha\tilde{\mu} - \frac{(1-\alpha)\theta_l}{\Delta\theta}]$;
- $\beta(\bar{\theta}_{\tilde{\mu}}, \theta_h) = b$;
- $\beta(p, e_l) = \begin{cases} b & \textit{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ nb & \textit{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases},$
- $\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ nb & \textit{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases}, \textit{ if } \delta \geq 1 - \frac{c}{\Delta\theta} \textit{ and } p \neq \bar{\theta}_{\tilde{\mu}};$
- $\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \theta_l + \frac{\delta}{1-\delta}c \\ ba & \textit{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ nb & \textit{if } p > \theta_h - c \end{cases}, \textit{ if } \delta < 1 - \frac{c}{\Delta\theta} \textit{ and } p \neq \bar{\theta}_{\tilde{\mu}}.$

Moreover, necessary conditions for such equilibria are:

$$\Delta\theta \leq \frac{c}{1-\tilde{\mu}}, \tag{2.12}$$

$$\Delta\theta \geq \frac{1-\alpha}{\alpha\tilde{\mu}}\theta_l, \tag{2.13}$$

which require

$$c \geq \frac{(1-\alpha)(1-\tilde{\mu})}{\alpha\tilde{\mu}}\theta_l =: \underline{c}. \tag{2.14}$$

The pooling equilibrium described above sets the price equal to the expected quality once adjusted for the probability of the low type to successfully pass as the high type. Indeed, if both types pool on the high evidence, the posterior belief of the quality should be equal to the expected value but taking into account that the (posterior) probability of the low type is the prior objective probability multiplied by α. The authentication is not acquired in this case either, instead it serves as a threat to guarantee the low type's incentive compatibility. This equilibrium can only be sustained if the expected payment of checking the quality for the customer exceeds the expected gain from it passing from the average quality to the high one (both happen with probability the good of high type):

$$\tilde{\mu}c \geq \tilde{\mu}(1-\tilde{\mu})(\theta_h - \theta_l) = \tilde{\mu}(\theta_h - \tilde{\theta}),$$

which is exactly the condition (2.12). .

Finally, condition (2.13) pins down the lower bound on the quality difference such that the low type wants to risk providing the high evidence. If re-written, it requires that the expected gain from pooling on the high type (left-hand side) exceeds the gain from submitting the honest low evidence:

$$\alpha\bar{\theta}_{\tilde{\mu}} \geq \theta_l$$

Pooling on the high evidence is not, however, the unique pooling strategy that can arise in the equilibrium.

Pooling on low-quality evidence. We characterize the pooling equilibria in which high and low-type sellers provide evidence of low quality. This set of equilibria requires more (less) stringent conditions on the out-of-equilibrium beliefs of the buyer when observing high (low) quality evidence when observing an unexpected price-evidence vector compared to the pooling on high evidence case. Moreover, this type of equilibria does not require a lower bound on $\Delta\theta$, while the upper bound is stricter with respect to the other pooling set of equilibria.

Proposition 5. *The following strategies of the agents and beliefs of the buyer represent a continuum of pooling PBEs on θ_l:*

- $\sigma(\theta_h) = \sigma(\theta_l) = (\bar{\theta}_{\mu_0}, \theta_l)$, *where* $\bar{\theta}_{\mu_0} := \mu_0\theta_h + (1-\mu_0)\theta_l$;
- $\mu(\bar{\theta}_{\mu_0}, \theta_l) = \mu_0$;
- $\delta = \mu((p, e_h)) \in [0, \mu_0]$, *and* $\gamma = \mu((p, e_l)) \in [0, \mu_0]$ *with* $p \neq \bar{\theta}_{\mu_0}$;
- $\beta(\bar{\theta}_{\mu_0}, \theta_h) = b$;
- $\beta(p, e_l) = \begin{cases} b & \textit{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ nb & \textit{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases} \quad \textit{and} \quad p \neq \bar{\theta}_{\mu_0},$
- $\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ nb & \textit{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases}, \ \textit{if } \delta \geq 1 - \frac{c}{\Delta\theta};$
- $\beta(p, e_h) = \begin{cases} b & \textit{if } p \leq \theta_l + \frac{\delta}{1-\delta}c \\ ba & \textit{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ nb & \textit{if } p > \theta_h - c \end{cases}, \ \textit{if } \delta < 1 - \frac{c}{\Delta\theta}.$

Moreover, a necessary condition for such equilibria is:

$$\Delta\theta \leq \frac{c}{1-\mu_0}. \tag{2.15}$$

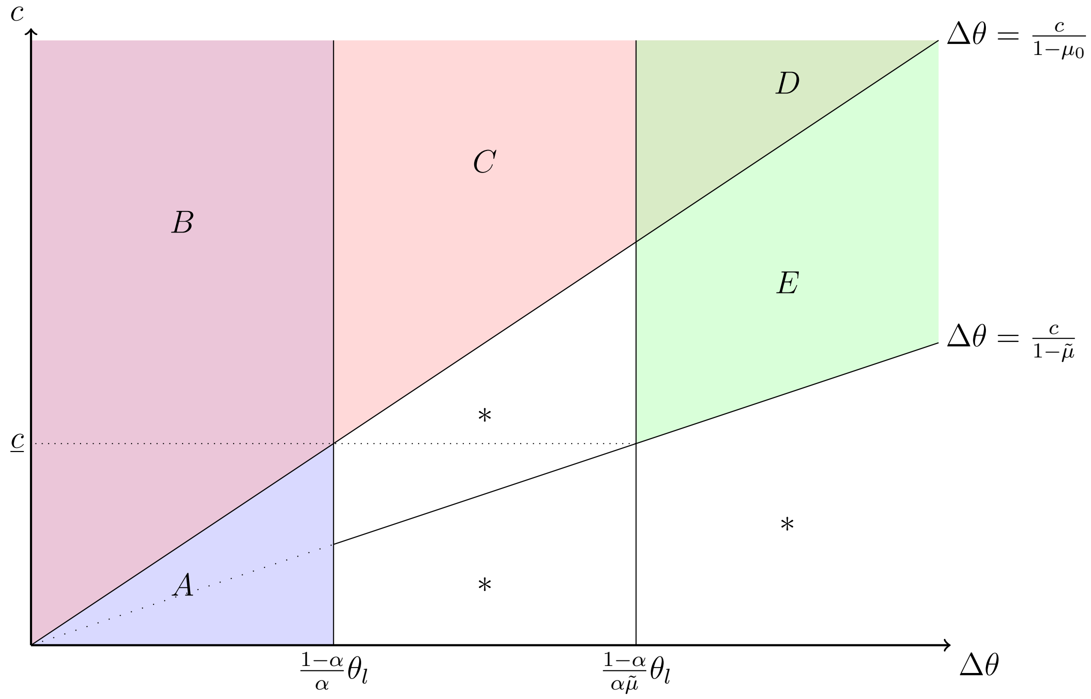
Notice that condition (2.15) is very similar to condition (2.12) but in this case it guarantees that the high type does not want to deviate to showing the evidence of his true quality inducing the buyer to confirm his high type with authentication. Indeed, for this equilibrium to be sustained, the expected quality set as the price should be higher than the price $\theta_h - c$ when the high type submits the high evidence and prepays the cost of authentication for the buyer:

$$\begin{aligned} \mu_0\theta_h + (1-\mu_0)\theta_l \geq \theta_h - c& \\ c \geq (1-\mu_0)\Delta\theta& \end{aligned}$$

Notice that, by combining the results of all pure-strategy equilibria, we have two important implications. Firstly, authentication is never acquired in the equilibrium since is it always sub-game dominated for the buyer. The intuition is simple: if in the equilibrium the sellers set choose a combination of price-quality expecting the buyer to buy authentication (that is, only the high type chooses $e = p = \theta_h$), the buyer attributing probability 1 to the high type upon observing $e = p = \theta_h$ does not buy the authentication to save the cost. Instead, it serves the role of disciplining the low type in the off-equilibrium paths from choosing a higher price.

The result is not quite intuitive but provides an explanation of why the online re-sale platforms with moderately expensive objects with a relatively little quality variation like clothes satisfying condition (2.11) for the separating equilibrium offer the possibility of buying authentication but allow for the trade to happen without it. Instead, when it comes to the markets of more valuable goods like cars or real estate where the variation in quality can be high so condition (2.11) is not satisfied, this tool does not work and either authentication is not optional or the seller has to provide perfect evidence.

Secondly, the necessary conditions for the existence of the separating (2.11) and high-evidence pooling (2.13) equilibria are mutually exclusive, therefore, there are no overlaps in the parameter regions of separating and pooling equilibria on high evidence. We can draw the relation of c and $\Delta\theta$ to illustrate what kinds of equilibria exist under any combination of the parameters:


c
$\Delta\theta = \frac{c}{1-\mu_0}$
D
C
B
E
$\Delta\theta = \frac{c}{1-\tilde{\mu}}$
$\underline{c}$
$*$
$*$
A
$*$
$\frac{1-\alpha}{\alpha}\theta_l$
$\frac{1-\alpha}{\alpha\tilde{\mu}}\theta_l$
$\Delta\theta$

In the graphs above, the existence of the equilibria is as follows

- Separation is obtained in the regions A and B;
- Pooling on the high evidence is obtained in the regions E and D;
- Pooling on the low evidence is obtained in the regions B, C, and D;
- Regions with an asterisk have no equilibrium in pure strategies.

The important implication of the results is that generally it is possible to achieve separation if quality dispersion of the products for sale is not high: indeed, for the regions A and B which represent exactly this condition, it is easy for the platform to set a proper cost of authentication that yields the first-best prices. It means that for a large category of products, such as non-luxury clothes or footwear, for instance, just the presence of the potential quality control is enough to gain efficiency and prevent market shutdown. They buyers can be more or less sure of the declared quality and avoid bearing extra costs of quality authentication, while the sellers earn exactly their products' worth.

The situation is becoming more complicated when the potential quality variance is high: in this case, separation is not achievable through the mechanisms we propose, and alternative model specification should be considered.

2.4.3 Pure-strategy equilibria with imperfect authentication technology

The main result when considering a setting in which the authentication technology is not perfect, i.e. $\varepsilon > 0$, is that the authentication can be acquired in equilibrium with positive probability. Clearly, purchasing the authentication remains a dominated action in separating equilibria even in this context, since the equilibrium beliefs are degenerate. However, there are two possible outcomes of a polling equilibrium on high evidence depending on the parameter space. The first one is the imperfect-technology counterpart of the separating equilibrium characterized in Proposition 4 in which the optimal strategy of the buyer is b. The second outcome allows for authentication acquisition and is defined by the following proposition.

Proposition 6. *The following strategies of the agents and beliefs of the buyer represent a continuum of pooling PBEs on θ_h when the authentication technology is not perfect, i.e. $\varepsilon > 0$:*

- $\sigma(\theta_h) = \sigma(\theta_l) = (\bar{\theta}_{\hat{\mu}} - c, \theta_h)$*, where* $\bar{\theta}_{\hat{\mu}} := \hat{\mu}\theta_h + (1 - \hat{\mu})\theta_l$*, and* $\hat{\mu} := \frac{\mu_0}{\mu_0 + (1-\mu_0)\alpha\varepsilon}$*;*
- $\mu(\bar{\theta}_{\hat{\mu}} - c, \theta_h) = \hat{\mu}$*;*
- $\beta(\bar{\theta}_{\hat{\mu}} - c, \theta_h) = ba.$

Moreover, a necessary condition for such equilibria is:

$$c \leq \frac{\tilde{\mu}(1-\tilde{\mu})(1-\varepsilon)}{\tilde{\mu}(1-\tilde{\mu})\varepsilon}\Delta\theta \tag{2.16}$$

A few important implications follow the proposition. Firstly and mainly, the authentication is bought in the equilibrium for the first time but its cost is borne by the agents. The intuition comes from the fact that it does not guarantee the buyer to perfectly learn the quality of the good thus allowing also thus low type to pass the authentication, which stops being sub-game dominated for the buyer.

Notice also that the asked price is lower in this equilibrium compared to the perfect-authentication case for two reasons: one is already mentioned since the seller offers a discount c to the expected quality to allow the authentication to happen. Moreover, the expected quality is lower as well as in some cases the low type passes the authentication as well due to its noisy technology.

2.4.4 Mixed-strategy Equilibria

We extend the equilibrium analysis by allowing the players to mix among optimal strategies. In order to refine the set of equilibria we intend to use an adaptation of the Che and Kreps (1987) intuitive criterion found in Bester and Ritzberger (2001) and subsequently used in Stahl and Strauz (2017). Obviously, we have to adapt this refined criterion to our setting.

Assumption 3 (Belief refinement). *We say that a PBE $\{\sigma_h^*, \sigma_l^*, \beta^*, \mu^*\}$ survived the refinement criterion if, for any belief $\mu \in [0,1]$ and any out-of-equilibrium price-evidence vector (p,e),*

$$\pi(p,\mu|\theta_h) > \pi(\sigma_h^*|\beta^*,\theta_h) \quad \text{and} \quad \pi(p,\mu|\theta_l) < \pi(\sigma_l^*|\beta^*,\theta_l) \implies \mu^*(p,e) \geq \mu.$$

Conjecture 1. *After applying the belief refinement specified in assumption 3, there are only two possible equilibrium outcomes: the outcome of the separating equilibrium described in Proposition 2, and a semi-pooling equilibrium in which $\sigma_h(\hat{p}) = 1$ and $\sigma_l(\hat{p}) + \sigma_l(\hat{\theta}_l) = 1$ with $\sigma_l(\hat{p}) \in (0, 1)$.*

2.5 Discussion

From the analysis of pure-strategy equilibrium, a few important implications can be noticed. Firstly, authentication is never acquired in equilibrium with perfect authentication but it is still can be considered a useful tool since it serves the role of disciplining the low type in the off-equilibrium paths from choosing a higher price. Once allowing for an imperfect authentication technology, however, authentication is finally arising in a pooling equilibrium since it stops being sub-game dominated for the buyer with probability one. Secondly, a fully revealing outcome is possible (regardless of c) when $\Delta\theta$ is relatively small which is a common case for the re-sale markets of moderately expensive goods with little variation in prices. Indeed, if we look at two different platforms of second-hand clothes, Vinted which in 2021 had an average order value below 100 dollars [5] and Vestiaire Collective whose customers in the same 2021 were spending on average around 300 euro per order [6], we notice an interesting fact: the former offers voluntary authentication which the buyer is free not to buy while the second obliges the buyer to acquire it. This fact indeed can be explained through our analysis of the condition for separation to be sustainable.

Finally, as it often happens, our model generates a multiplicity of equilibria, which does not provide much explanatory power. However, once we apply the Intuitive Criterion of Cho and Kreps, only the separating equilibrium remains in pure strategies, which is an

[5] Source https://secondmeasure.com/datapoints/fashion-resale-platforms-outperform-retailers-during-pandemic/

[6] https://www.scmp.com/business/banking-finance/article/3123657/second-hand-fashion-app-vestiaire-collective-turns

optimistic results since separation yields efficiency.

However, in order to complete the analysis, a number of extentions can be considered. Firstly, one could study the augmented version of the game in which an additional player, the authenticator, selects the subgame to be played by the seller and the buyer, by choosing the authentication parameters (ε, c) that maximize her payoff in equilibrium. In particular, the interesting part in such analysis would be the role of the parameter α in shaping the authenticator's optimal decision.

Another interesting case to study is the one where α is endogenously chosen by the low-type seller. Since this parameter defines the range in which separation is possible, such endogenous choice can potentially reduce the possibility to achieve fully-revealing prices and raise the question what is the optimal mechanism to overcome the issue.

Finally, the case in which providing untruthful evidence by the high type might reveal her type with positive probability is also worth considering. The idea behind his assumption is as follows: if α is low, the high type might have an incentive to provide low-quality evidence in order to reveal her actual type, then the equilibrium structure will be different, and fully-revealing prices can be gained through pooling instead of separating, which is paradoxical yet not impossible.

2.6 Appendix

Proof of Lemma 1

Proof. Consider expressions (4), (5) and (6). First, $u(b|p,\hat{\mu}) \geq u(n)$ if and only if $p \leq \bar{\theta}_{\hat{\mu}}$. Moreover, $u(b|p,\hat{\mu}) \geq u(ba|p,\hat{\mu})$ if and only if $\bar{\theta}_{\hat{\mu}} - p \geq \bar{\theta}_{\hat{\mu}} - p - (1-\hat{\mu})(1-\varepsilon)(\theta_l - p) - c$ or $p \leq \theta_l + c(1-\hat{\mu})^{-1}(1-\varepsilon)^{-1}$. Hence, b is optimal whenever both conditions on are satisfied:

$$p \leq \min\left\{\bar{\theta}_{\hat{\mu}},\ \theta_l + \frac{c}{(1-\hat{\mu})(1-\varepsilon)}\right\}.$$

Second, $u(ba|p,\hat{\mu}) \geq u(n)$ if and only if $[\hat{\mu} + (1-\hat{\mu})\varepsilon]p \leq \hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c$ or $p \leq [\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c][\hat{\mu} + (1-\hat{\mu})\varepsilon]^{-1}$. Moreover, $u(ba|p,\hat{\mu}) \geq u(b|p,\hat{\mu})$ if and only if $p \geq \theta_l + c(1-\hat{\mu})^{-1}(1-\varepsilon)^{-1}$. Hence, ba is optimal whenever both conditions are satisfied:

$$p \in \left[\theta_l + \frac{c}{(1-\hat{\mu})(1-\varepsilon)},\ \frac{\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c}{\hat{\mu} + (1-\hat{\mu})\varepsilon}\right]$$

Finally, $u(n) \geq u(b|p,\hat{\mu})$ if and only if $p \geq \bar{\theta}_{\hat{\mu}}$, while$u(n) \geq u(ba|p,\hat{\mu})$ if and only if $p \geq [\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c][\hat{\mu} + (1-\hat{\mu})\varepsilon]^{-1}$ Hence, ba is optimal whenever both conditions are satisfied:

$$p \geq \max\left\{\bar{\theta}_{\hat{\mu}},\ \frac{\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c}{\hat{\mu} + (1-\hat{\mu})\varepsilon}\right\}.$$

□

Proof of Corollary 1

Proof. Given Lemma 1, the set of values of p for which ba is optimal is non-empty if and only if:

$$\begin{aligned}
\theta_l + \frac{c}{(1-\hat{\mu})(1-\varepsilon)} &\leq \frac{\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c}{\hat{\mu} + (1-\hat{\mu})\varepsilon} \\
(1-\hat{\mu})(1-\varepsilon)[\hat{\mu} + (1-\hat{\mu})\varepsilon]\theta_l + [\hat{\mu} + (1-\hat{\mu})\varepsilon]c &\leq (1-\hat{\mu})(1-\varepsilon)[\hat{\mu}\theta_h + (1-\hat{\mu})\varepsilon\theta_l - c] \\
\hat{\mu}(1-\hat{\mu})(1-\varepsilon)\theta_l + [\hat{\mu} + (1-\hat{\mu})\varepsilon]c &\leq (1-\hat{\mu})(1-\varepsilon)[\hat{\mu}\theta_h - c] \\
c &\leq \hat{\mu}(1-\hat{\mu})(1-\varepsilon)(\theta_h - \theta_l)
\end{aligned}$$

which proves the claim. □

Proof of Proposition 2

Proof. Consider a separating PBE in which the high-quality seller chooses (p_h, θ_h), and the low-quality seller chooses (p_l, θ_l) in equilibrium. Therefore, the posteriors of the buyer upon observing these two vectors are $\mu((p_h, \theta_h)) = 1$ and $\mu((p_l, \theta_l)) = 0$ respectively. First, consider (p_h, θ_h). The utility of the buyer given each of the actions in her choice set $\{b, ba, nb\}$ is equal, respectively, to

$$\begin{aligned}
u^B(b|(p_h, \theta_h)) &= \theta_h - p_h, \\
u^B(ba|(p_h, \theta_h)) &= \theta_h - p_h - c, \\
u^B(nb|(p_h, \theta_h)) &= 0.
\end{aligned}$$

Hence, the best response of the buyer to (p_h, θ_h) is

$$\beta(p_h, \theta_h) = \begin{cases} b & \text{if } p_h \leq \theta_h \\ nb & \text{if } p_h > \theta_h \end{cases}$$

Similarly, we have that

$$\beta(p_l, \theta_l) = \begin{cases} b & \text{if } p_l \leq \theta_l \\ nb & \text{if } p_l > \theta_l \end{cases}$$

so that the optimal strategy for the seller requires $p_h = \theta_h$ and $p_l = \theta_l$, which implies $u_h^S = \theta_h$ and $u_l^S = \theta_l$.

Let us now consider the off-equilibrium strategies. Given our assumption on the structure of off-path beliefs, we have

$$\mu((p, e)) = \begin{cases} \delta & \text{if } e = \theta_h \\ \gamma & \text{if } e = \theta_l \end{cases}$$

Consider (p, e_l), $p \neq \theta_l$, with $\mu((p, e_l)) = \gamma$. Then, the expected utility of the buyer, given each of her possible actions is equal to

$$\begin{aligned} u^B(b|(p, e_l)) &= \gamma\theta_h + (1-\gamma)\theta_l - p \\ u^B(ba|(p, e_l)) &= \gamma\theta_h + (1-\gamma)\theta_l - p - c \\ u^B(nb|(p, e_l)) &= 0 \end{aligned}$$

so that the best response to (p, e_l) is

$$\beta(p, e_l) = \begin{cases} b & \text{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ nb & \text{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases}$$

Finally, consider (p, e_h), $p \neq \theta_h$, with $\mu((p, e_h)) = \delta$. Then, the expected utility of the buyer, given each of her possible actions is equal to

$$\begin{aligned} u^B(b|(p,e_h)) &= \delta\theta_h + (1-\delta)\theta_l - p \\ u^B(ba|(p,e_h)) &= \delta(\theta_h - p - c) \\ u^B(nb|(p,e_h)) &= 0 \end{aligned}$$

The difference now is that the authentication technology can potentially kick in and reveal low-type sellers providing high-quality evidence. This happens, given the beliefs of the buyer, with probability $1-\delta$, while the trade takes place with probability δ. The best response to (p, e_h) is then

$$\beta(p,e_h) = \begin{cases} b & \text{if } p \leq \min\{\delta\theta_h + (1-\delta)\theta_l, \theta_l + \frac{\delta}{1-\delta}c\} \\ ba & \text{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ nb & \text{if } p > \max\{\delta\theta_h + (1-\delta)\theta_l, \theta_h - c\} \end{cases}$$

which corresponds to

$$\beta(p,e_h) = \begin{cases} b & \text{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ nb & \text{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases} \quad \text{if } \delta \geq 1 - \frac{c}{\Delta\theta}$$

and

$$\beta(p,e_h) = \begin{cases} b & \text{if } p \leq \theta_l + \frac{\delta}{1-\delta}c \\ ba & \text{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ nb & \text{if } p > \theta_h - c \end{cases} \quad \text{if } \delta < 1 - \frac{c}{\Delta\theta}$$

Now, we should consider all the possible deviations given the two cases depending on the value of δ. First, notice that the high type never has the incentive to deviate from her equilibrium strategy since she attains the maximum possible payoff, i.e. θ_h. Let us focus on the low-quality seller. Consider the case in which $\delta \geq 1 - \frac{c}{\Delta\theta}$. Then, the low type compares her equilibrium payoff, θ_l, to her best possible deviations. If she deviates to (θ_h, e_h) mimicking the high-quality seller, she could get $\alpha\theta_h$, since, in such case, $e_l > \theta_l$ and the evidence will not reveal her true type with probability α. If she deviates to (p, e_l), $p \neq \theta_l$ she could get up to $\gamma\theta_h + (1-\gamma)\theta_l$. Finally deviating to (p, e_h), $p \neq \theta_h$, would yield up to $\alpha[\delta\theta_h + (1-\delta)\theta_l]$. To prevent the low type from deviating then, it must be that

$$\begin{aligned}
&\theta_l \geq \alpha\theta_h, \\
&\theta_l \geq \gamma\theta_h + (1-\gamma)\theta_l, \\
&\theta_l \geq \alpha[\delta\theta_h + (1-\delta)\theta_l],
\end{aligned}$$

which imply

$$\gamma = 0, \tag{2.17}$$

$$\theta_h \leq \frac{\theta_l}{\alpha} \quad \Leftrightarrow \quad \Delta\theta \leq \frac{1-\alpha}{\alpha}\theta_l. \tag{2.18}$$

If $\delta < 1 - \frac{c}{\Delta\theta}$, the conditions for preventing deviations of low-type sellers are the same except for the last one. In this second case, the buyer finds it optimal to buy the good with authentication for some intermediate values of p. A low-quality seller cannot charge a price higher than $\theta_l + \frac{\delta}{1-\delta}c$ The best alternative for a low-quality seller when deviating to (p, e_h), $p \neq \theta_h$, yields an expected payoff of $\alpha(\theta_l + \frac{\delta}{1-\delta}c)$. In fact, posting a price higher than $\theta_l + \frac{\delta}{1-\delta}c$ when providing evidence of high quality induces the buyer to either buy the good with authentication or not buy the good at all. In both cases, the payoff of the seller is equal to zero.

Therefore, the no-deviation conditions for the low-type seller when $\delta < 1 - \frac{c}{\Delta\theta}$ are

$$\begin{aligned}
&\theta_l \geq \alpha\theta_h \\
&\theta_l \geq \gamma\theta_h + (1-\gamma)\theta_l \\
&\theta_l \geq \alpha\left(\theta_l + \frac{\delta}{1-\delta}c\right)
\end{aligned}$$

Even in this second case, the three conditions translate into conditions (2) and (3) since

$$\alpha\left(\theta_l + \frac{\delta}{1-\delta}c\right) < \alpha\theta_h$$

as we are assuming $\delta < 1 - \frac{c}{\Delta\theta}$. Finally, notice that no restrictions on δ are required. □

Proof of Proposition 3

Proof. A separating equilibrium requires $\sigma(\theta_h) = (p_h, e_h)$ and $\sigma(\theta_l) = (p_l, e_l)$ to be different. Hence, there are four possible separating alternatives when considering the type of evidence provided by the two types of sellers in equilibrium:

1. (p_h, θ_h), (p_l, θ_l),

2. (p_h, θ_l), (p_l, θ_h),

3. (p_h, θ_h), (p_l, θ_h), with $p_h \neq p_l$,

4. (p_h, θ_l), (p_l, θ_l), with $p_h \neq p_l$,

where the first vector corresponds to $\sigma(\theta_h)$ and the second one to $\sigma(\theta_l)$. Let us consider these four cases separately.

1. The first case corresponds to the equilibria characterized by Proposition 1.

2. Suppose the equilibrium price-evidence combination of the high and low seller are (p_h, θ_l), (p_l, θ_h), respectively. Then the best response of the buyer is defined by

$$\beta(p_h, \theta_l) = \begin{cases} b & \text{if } p_h \leq \theta_h \\ nb & \text{if } p_h > \theta_h \end{cases}$$

and

$$\beta(p_l, \theta_h) = \begin{cases} b & \text{if } p_l \leq \theta_l \\ nb & \text{if } p_l > \theta_l \end{cases}$$

Clearly, in equilibrium it must hold that $p_h = \theta_h$, $p_l = \theta_l$, so that equilibrium payoffs are $u_h^S = \theta_h$ and $u_l^S = \alpha\theta_l$. However, if the low type deviated to (p_h, θ_l), she would not be afraid of providing false evidence and get θ_h which is always larger than $u_l^S = \alpha\theta_l$. Therefore, there cannot be an equilibrium in which both high and low-quality sellers provide untruthful evidence.

3. Suppose the equilibrium price-evidence combination of high and low sellers are (p_h, θ_h), (p_l, θ_h), respectively, with $p_h \neq p_l$. Then, the best response of the buyer is

$$\beta(p_h, \theta_h) = \begin{cases} b & \text{if } p_h \leq \theta_h \\ nb & \text{if } p_h > \theta_h \end{cases}$$

and

$$\beta(p_l, \theta_h) = \begin{cases} b & \text{if } p_l \leq \theta_l \\ nb & \text{if } p_l > \theta_l \end{cases}$$

As before, this implies that in equilibrium $p_h = \theta_h$, $p_l = \theta_l$, which yield to $u_h^S = \theta_h$ and $u_l^S = \alpha\theta_l$. If the low type deviated to (p_h, θ_h), she would get $\alpha\theta_h$ which is always larger than $u_l^S = \alpha\theta_l$. Hence, separating equilibria in which only the high type provides truthful evidence are not sustainable.

4. Suppose the price-evidence combinations of the high and low seller are (p_h, θ_l), (p_l, θ_l), respectively, with $p_h \neq p_l$. Then the best response of the buyer is defined by

$$\beta(p_h, \theta_l) = \begin{cases} b & \text{if } p_h \leq \theta_h \\ nb & \text{if } p_h > \theta_h \end{cases}$$

and

$$\beta(p_l, \theta_l) = \begin{cases} b & \text{if } p_l \leq \theta_l \\ nb & \text{if } p_l > \theta_l \end{cases}$$

In equilibrium, it must be that $p_h = \theta_h$, $p_l = \theta_l$, $u_h^S = \theta_h$ and $u_l^S = \theta_l$. However, if the low type deviated to (p_h, θ_l), she would not be afraid of providing false evidence and she would get θ_h which is always larger than $u_l^S = \theta_l$. Therefore, separating equilibria in which only the low type provides truthful evidence are not sustainable.

Since we exhausted the set of possible separating equilibria in pure strategies, we proved the proposition. □

Proof of Proposition 4

Proof. Consider a pooling PBE in which both high and low-quality sellers choose to provide high-quality evidence in equilibrium, so that $(p_h, e_h) = (p_l, e_l) = (p_{e=\theta_h}, \theta_h)$ where $p_{e=\theta_h}$ is the equilibrium price. In this case, the low-type seller provides untruthful quality evidence, $e_l > \theta_l$, which makes her susceptible to being detected by the buyer. This happens with probability $1 - \alpha$, while her true type is not revealed with probability α. Therefore, upon observing the equilibrium vector $(p_{e=\theta_h}, \theta_h)$, the buyer updates her belief following Bayes rule as follows:

$$\mu((p_{e=\theta_h}, \theta_h)) = \frac{\mu_o}{\mu_o + (1-\mu_0)\alpha} =: \tilde{\mu} \tag{2.19}$$

Clearly, the updated belief when observing the price-evidence vector in this case, $\tilde{\mu}$, is larger than the prior μ_0 since low-quality sellers only *pass* the evidence test with probability α while high-quality sellers do it with probability one. This implies that the expected quality of the good faced by the buyer in a pooling equilibrium on θ_h is equal to

$$\tilde{\theta} := \tilde{\mu}\theta_h + (1-\tilde{\mu})\theta_l. \tag{2.20}$$

Given $(p_{e=\theta_h}, \theta_h)$ and the actions in her choice set, the utility of the buyer equals, respectively,

$$\begin{aligned} u^B(b|(p_{e=\theta_h}, \theta_h)) &= \tilde{\theta} - p_{e=\theta_h}, \\ u^B(ba|(p_{e=\theta_h}, \theta_h)) &= \tilde{\mu}(\theta_h - p_{e=\theta_h} - c), \\ u^B(nb|(p_{e=\theta_h}, \theta_h)) &= 0. \end{aligned}$$

Then, her best response is

$$\beta(p_{e=\theta_h}, \theta_h) = \begin{cases} b & \text{if } p_{e=\theta_h} \leq \tilde{\theta} \\ nb & \text{if } p_{e=\theta_h} > \tilde{\theta} \end{cases} \quad \text{if } \tilde{\mu} \geq 1 - \frac{c}{\Delta\theta}$$

and

$$\beta(p_{e=\theta_h}, \theta_h) = \begin{cases} b & \text{if } p_{e=\theta_h} \leq \theta_l + \frac{\tilde{\mu}}{1-\tilde{\mu}}c \\ ba & \text{if } \theta_l + \frac{\tilde{\mu}}{1-\tilde{\mu}}c < p_{e=\theta_h} \leq \theta_h - c \\ nb & \text{if } p_{e=\theta_h} > \theta_h - c \end{cases} \quad \text{if } \tilde{\mu} < 1 - \frac{c}{\Delta\theta}$$

In the first case, the optimal strategy for both sellers requires $p_{e=\theta_h} = \tilde{\theta}$ which implies

$u_h^S = \tilde{\theta}$ and $u_l^S = \alpha\tilde{\theta}$. However, when $\tilde{\mu} < 1 - \frac{c}{\Delta\theta}$ the optimal price for the high-quality and the low-quality seller differs so that a pooling equilibrium cannot be supported. Therefore, a necessary condition for pooling equilibria on high evidence is $\tilde{\mu} \geq 1 - \frac{c}{\Delta\theta}$, or equivalently,

$$\Delta\theta \leq \frac{c}{1-\tilde{\mu}}. \tag{2.21}$$

Consider now the off-equilibrium strategies, and recall we assumed

$$\mu((p,e)) = \begin{cases} \delta & \text{if } e = \theta_h \\ \gamma & \text{if } e = \theta_l. \end{cases}$$

In a pooling equilibrium on $(\tilde{\theta}, \theta_h)$, there are two possible deviations - conditional on a generic price p - for both types: (p, e_l), and (p, e_h) with $p \neq \tilde{\theta}$. Symmetrically to the separating case, the best response of the buyer to (p, e_l) given her out-of-equilibrium beliefs is

$$\beta(p, e_l) = \begin{cases} b & \text{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ nb & \text{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases}$$

The optimal price when deviating to (p, e_l) is then common to both types of seller and equals $\gamma\theta_h + (1-\gamma)\theta_l$. To prevent this type of deviation, the following two conditions must hold:

$$\alpha\tilde{\theta} \geq \gamma\theta_h + (1-\gamma)\theta_l,$$
$$\tilde{\theta} \geq \gamma\theta_h + (1-\gamma)\theta_l,$$

which boil down to a condition for γ:

$$\gamma \leq \alpha\tilde{\mu} - (1-\alpha)\frac{\theta_l}{\Delta\theta}. \tag{2.22}$$

Since $\gamma := \mu((p, e_l))$ represents an off-path posterior, a necessary condition for (2.22) is $\alpha\tilde{\mu} \geq (1-\alpha)\frac{\theta_l}{\Delta\theta}$ or, equivalently,

$$\Delta\theta \geq \frac{1-\alpha}{\alpha\tilde{\mu}}\theta_l. \tag{2.23}$$

On the other hand, buyer's best response to (p, e_h), when $p \neq \tilde{\theta}$, is

$$\beta(p, e_h) = \begin{cases} b & \text{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ \\ nb & \text{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases} \quad \text{if } \delta \geq 1 - \frac{c}{\Delta\theta}$$

and

$$\beta(p, e_h) = \begin{cases} b & \text{if } p \leq \theta_l + \frac{\delta}{1-\delta}c \\ \\ ba & \text{if } \theta_l + \frac{\delta}{1-\delta}c < p \leq \theta_h - c \\ \\ nb & \text{if } p > \theta_h - c \end{cases} \quad \text{if } \delta < 1 - \frac{c}{\Delta\theta}$$

If $\delta \geq 1 - \frac{c}{\Delta\theta}$, the optimal price when deviating to (p, e_h), with $p \neq \tilde{\theta}$, is common between the two types and equals $\delta\theta_h + (1-\delta)\theta_l$. Then, deviating to (p, e_h) would not be profitable for low and high types, respectively, if

$$\alpha\tilde{\theta} \geq \alpha[\delta\theta_h + (1-\delta)\theta_l],$$
$$\tilde{\theta} \geq \delta\theta_h + (1-\delta)\theta_l,$$

which correspond to the following condition on δ:

$$\delta \leq \tilde{\mu}. \tag{2.24}$$

Instead, optimal pricing is different between the two groups if $\delta < 1 - \frac{c}{\Delta\theta}$ since the buyer may find it optimal to purchase the authentication depending on the posted price when δ is relatively small. In this case, the optimal price is $\theta_l + \frac{\delta}{1-\delta}c$ for the low type, while it equals $\theta_h - c$ for the high type. Then, the no-deviation conditions for low and high types become, respectively,

$$\alpha\tilde{\theta} \geq \alpha\left(\theta_l + \frac{\delta}{1-\delta}c\right),$$
$$\tilde{\theta} \geq \theta_h - c.$$

When $\delta < 1 - \frac{c}{\Delta\theta}$, $\theta_l + \frac{\delta}{1-\delta}c$ is smaller than $\theta_h - c$, so the first condition is implied by the second, which corresponds to condition (2.21). Finally, for (2.21) and (2.23) to simultaneously hold, it must be that $\frac{c}{1-\tilde{\mu}} \geq \frac{1-\alpha}{\alpha\tilde{\mu}}\theta_l$, which corresponds to a lower bound $\underline{c}$ to the cost of the authentication necessary to sustain a pooling equilibrium on high evidence:

$$c \geq \frac{(1-\alpha)(1-\tilde{\mu})}{\alpha\tilde{\mu}}\theta_l =: \underline{c}. \tag{2.25}$$

□

Proof of Proposition 5

Proof. Consider a pooling PBE in which both high and low-quality sellers choose to provide low-quality evidence, $(p_h, e_h) = (p_l, e_l) = (p_{e=\theta_l}, \theta_l)$ where $p_{e=\theta_l}$ is the equilibrium price. In this case, the low-type seller provides truthful quality evidence, $e_l = \theta_l$, and her type is never revealed to the buyer by the evidence technology. Therefore, upon observing the equilibrium vector (p, e_l), the buyer cannot update her belief which remains equal to her prior:

$$\mu((p_{e=\theta_l}, \theta_h)) = \mu_o \tag{2.26}$$

Hence, the expected quality of the good faced by the buyer in a pooling equilibrium on θ_l is equal to the unconditional expected quality, i.e.

$$\bar{\theta} := \mu_0\theta_h + (1-\mu_0)\theta_l. \tag{2.27}$$

Following the same reasoning adopted above, the best response of the buyer to $(p_{e=\theta_l}, \theta_h)$ is

$$\beta(p_{e=\theta_l}, \theta_h) = \begin{cases} b & \text{if } p_{e=\theta_l} \leq \bar{\theta} \\ \\ nb & \text{if } p_{e=\theta_l} > \bar{\theta} \end{cases}$$

and the optimal strategy for the seller requires $p_{e=\theta_l} = \bar{\theta}$, so that $u_h^S = u_l^S = \bar{\theta}$.

There are two possible deviations - conditional on a generic price p - for both types: (p, e_l) with $p \neq \bar{\theta}$, and (p, e_h). Symmetrically to the previous cases, the best response of the buyer to (p, e_l) with $p \neq \bar{\theta}$, given her out-of-equilibrium beliefs, is

$$\beta(p, e_l) = \begin{cases} b & \text{if } p \leq \gamma\theta_h + (1-\gamma)\theta_l \\ \\ nb & \text{if } p > \gamma\theta_h + (1-\gamma)\theta_l \end{cases}$$

The optimal price when deviating to (p, e_l) is then common to both types of seller and equals $\gamma\theta_h + (1-\gamma)\theta_l$. The no-deviation condition, then, is the same for both seller types, and reads:

$$\bar{\theta} \geq \gamma\theta_h + (1-\gamma)\theta_l$$

This corresponds to:

$$\gamma \leq \mu_0. \tag{2.28}$$

We know that buyer's best response to (p, e_h) is

$$\beta(p, e_h) = \begin{cases} b & \text{if } p \leq \delta\theta_h + (1-\delta)\theta_l \\ \\ nb & \text{if } p > \delta\theta_h + (1-\delta)\theta_l \end{cases} \quad \text{if } \delta \geq 1 - \frac{c}{\Delta\theta}$$

and

$$\beta(p,e_h)=\begin{cases} b & \text{if } p\leq \theta_l+\frac{\delta}{1-\delta}c \\ ba & \text{if } \theta_l+\frac{\delta}{1-\delta}c<p\leq\theta_h-c \\ nb & \text{if } p>\theta_h-c \end{cases} \quad \text{if } \delta<1-\frac{c}{\Delta\theta}$$

Again, when $\delta \geq 1-\frac{c}{\Delta\theta}$, the optimal price when deviating to (p,e_h), with $p\neq\tilde{\theta}$, is common between the two types and equals $\delta\theta_h+(1-\delta)\theta_l$. Then, deviating to (p,e_h) would not be profitable for low and high types, respectively, if

$$\begin{aligned}\bar{\theta}_{\mu_0} &\geq \alpha[\delta\theta_h+(1-\delta)\theta_l], \\ \bar{\theta} &\geq \delta\theta_h+(1-\delta)\theta_l,\end{aligned}$$

where the second condition implies the first one and corresponds to:

$$\delta\leq\mu_0. \tag{2.29}$$

Then, we know that optimal pricing is different between the two groups if $\delta<1-\frac{c}{\Delta\theta}$. In this case, the no-deviation conditions for low and high type become, respectively,

$$\begin{aligned}\bar{\theta}_{\mu_0} &\geq \alpha\left(\theta_l+\frac{\delta}{1-\delta}c\right), \\ \bar{\theta} &\geq \theta_h-c.\end{aligned}$$

Clearly, the condition on the high type implies the condition on the low type and can be expressed as

$$\Delta\theta\leq\frac{c}{1-\mu_0}. \tag{2.30}$$

□

Proof of Proposition 6

Proof. Consider the optimal buyer's behavior. Suppose the strategies of the two seller types described in the equilibrium are indeed optimal, that is, both pool on the high evidence and set the price equal to $\bar{\theta}_{\hat{\mu}} - c$. Then, firstly, the posterior beliefs are equal exactly to $\hat{mu} = \frac{\mu_0}{\mu_0 + (1-\mu_0)\alpha\varepsilon}$.

Moreover, by Lemma 1, the buyer should buy the good (from condition (2.8)) on the one hand, on the other, acquire the authentication technology, so $\beta(\bar{\theta}_{\hat{\mu}} - c, \theta_h) = ba$.

Now we check, if given the best reply of the buyer, the strategies of the sellers to pool on high evidence with the asked price are indeed in equilibrium.

Firstly, conditional on both sending the high evidence, the price is indeed optimal: a downward deviation yields strictly lower profits, an upward deviation results in buyer refusing to buy the good.

Secondly, consider the best possible deviation the low type can do, which is, to send the true signal of low quality and ask the low quality price θ_l:

$$(\bar{\theta}_{\hat{\mu}} - c)\varepsilon - c(1-\varepsilon) \geq \theta_l \hat{\mu}\theta_h\varepsilon + \theta_l\varepsilon - \hat{\mu}\theta_l\varepsilon - \theta_l \qquad\qquad \geq c$$

Then substituting $\hat{\mu}$ by the original belief $\tilde{\mu}$ and after some algebra, we get exactly the necessary condition. □

Chapter 3

Controlling Protest Movements

3.1 Introduction

Defined as the “dissemination of information - facts, arguments, rumours, half-truths, or lies - to influence public opinion” by the Encyclopædia Britannica, propaganda has always been widely used by autocratic regimes to ensure their power. Indeed, by manipulating public opinion through mass media, the autocratic political elite can alienate the interests of the public in accordance with their own. In that case, the actions taken by the citizens will also serve to the benefit of the autocrat.

However, compared to the dictatorships of the past, modern non-democracies have only *imperfect* control over how much they can influence the public opinion. The reason for this is twofold:

- On the one hand, the sources of information tent to be rather decentralized even in the countries which are considered autocratic (apart from extreme cases). Firstly, nowadays citizens can have access to foreign media, which hinders the manipulation of the public opinion into the direction that is most favorable for autocrat. Moreover, it limits the ability of the autocrat to cover the incidences that can considerable damage her reputation as a capable ruler. Indeed, if in the Soviet Union it was possible to conceal the Chernobyl disaster for 36 hours, in the present-day Russia it would not be feasible. Indeed, the information about the explosion would have been spread in minutes. Secondly, there can be local opposition press. It enables citizens who might not speak foreign languages (hence, read foreign media) to learn the news from a point of view that is different from the official state propaganda and form their own opinion about the competence of the dictatorial government. Finally, social media which are more difficult to control that the traditional media outlets also promote the heterogeneity in the public opinion.

- On the other hand, the propaganda in the consolidated autocratic regimes can be rather illogical and contrary to common sense even without being challenged by alternative points of view. For instance, bringing up again the example of Russia, when the leader of the opposition, Alexey Navalny, was poisoned in the summer of 2020, the first official diagnosis published by RT explained his coma was a "metabolic disorder due to low sugar level".

Nevertheless, even if media control is partial, can it still can be able to the attempts to overthrow the regime. Indeed, despite not having full trust from the citizens, many dictatorships successfully stay in power for years (like Belarus that has Aleksandr Lukashenko as the head of state for 26 years) being able to suppress even numerous mass protests). To understand why, I model the ongoing protests as a repeated public good game with information manipulation. In particular, there is a finite set of potentially heterogeneous agents (citizens). Agents want to overthrow the status quo, the dictatorial regime by repeatedly participating in protest movements. At each point of time, each agent decides whether to participate in the attack against the Dictator. If the critical mass of agents attacks, the regime falls and revolution occurs. The threshold for the riots to be successful is given by the strength of the dictatorial government. It can be interpreted as the "hard power", or brute force, of the dictator: the amount of army and police at disposal to suppress the riots and their willingness to do so when they are called to contain the rebellion, that is, the degree of their compliance with orders. The agents do not know the exact strength of the regime, but only its distribution. At the beginning of the game, they are given with a prior distribution of the strength, but as time (and protests) unfolds, they update the beliefs.

Another player on stage is the Dictator, who is interested in saving the status quo, knows her strength and can take a hidden action at each round of protests in order to try to stay in power. On the contrary to military strength, this policy measure represents "soft power": through propaganda and media censorship, it is aimed at manipulating the beliefs of the citizens about the strength of the regime. However, to capture the imperfection of media control, this policy can backfire: instead of shifting upwards the perception of the regime strength by the agents, with a positive probability, it shifts it downwards.

Finally, after the realised size of the current attack and the realised policy of the dictator, the riots either result to be successful, and the game ends at that point with the regime being overthrown, or they fail, and the game moves to the next period.

The paper is yet incomplete. Specifically, only the easiest case is fully worked out when the Dictator does not use any policy and the citizens are homogeneous, and the dynamics of the citizens' participation behavior and of the regime survival are shown. In the case of Dictator's interference with propaganda prior to the protest, the dynamics of the first two periods are shown with the optimal attacking strategy of the citizens and the probability of regime survival. In order to conclude the paper, a number of steps are needed: firstly, to conclude the analysis of the homogeneous citizens case, secondly, to analyze the the main and fully-fledged model where agents are heterogeneous in how they benefit from the regime change.

This paper will contribute (upon conclusion of the analysis) to two strands of literature. The first one covers coordination games of regime change. It is common to model them as global games first introduced by Carlsson and van Damme (1993) and extensively studied by Morris and Shin (1998, 2001): in addition to the prior on the strength of the regime, each agent also receives a private signal which is independent of others. However, one would expect that the active protest participants constitute a relative homogeneous group in terms of the beliefs they hold about the dictators qualities. In addition, letting the agents have only the same prior gains in the tractability of the solution. The closest paper to my setting in the global games literature is Edmond (2013) and Angeletos et al. (2006). The first one lets the policy maker to influence the beliefs of the agents through additional signal(s) that increase the perceived regime strength. The main difference is that the policy of the autocrat does not have the backfiring possibility, moreover, the model is static in the sense that the game does not move to the new round of protests were the current uprising to fail. The second one gives the policy maker the opportunity to choose the cost of attacking. The cost of attacking in the current work is given exogenously.

Angeletos and Werning (2006) study a game of regime change where, in addition to the private signals about the fundamentals, the agents also receive a public noisy signal available to all of them. Angeletos et al. (2007) introduce the dynamic structure into the standard static model thus allowing the agents to learn over time about the regime depending on the previous outcomes of regime survival or failure. Finally, Boleslavsky et al (2020) also introduce the possibility of the ruler to use the media for the manipulation of public opinion. Finally, Goldstein and Huang (2016) introduce Bayesian persuasion into the global games by enabling the policy maker to make the commitment to abandon the status quo before the agents coordinate on attacking for the low values of regime strength.

The second strand of literature studies directly the economics of non-democracies in general and the use of media control in particular. Gehlbach et al (2014) provide a survey of the papers studying the nondemocratic politics and generalize the main the key elements being asymmetries of information and commitment problems. Gelbach and Sonin (2014) use media in order to manipulate the citizens into taking the favorable for the dictator action. Barberà and Jackson (2019) also use a public good game for modelling a revolution but allow agents to communicate among each other in the attempt of learning better the probability of a revolt being successful. Chen and Yang (2019) use a field experiment in China to evaluate the effect of access to uncensored Internet on economic beliefs, political attitudes and so on. Acemoglu et al (2020) study the stability and persistence of political institutions with a dynamic game among different population groups that hold different political preferences. Guriev and Treisman (2020) use a signalling game played by a dictator who tries to convince the public of her competence. Finally, Enikopov et al (2020) through a case study in Russia, analyze the effect of social media on the participation in protests finding a positive causal relation. Moreover, their results suggest that causal effect goes through the lowering of coordination costs.

3.2 Model

3.2.1 Timing

There is a finite set of agents (citizens) $I = \{1, \ldots, i, \ldots, N\}$ who want to overthrow the regime, and a Dictator, D, who wants to stay in power. At $t = 0$, the Dictator learns her strength θ, while the agents only learn the prior distribution given by $F_0(\theta) := U([0, N])$ (that is, if every agent attacks, the regime falls with probability one). This parameter represents the "hard power" of the Dictator, that is, the amount of police and even army prepared to suppress the revolts if needed. At each $t \geq 1$, each agent $i \in I$ takes a binary decision a_{it} if to participate in the current protest:

$$a_{it} = \mathbb{1}_{\{i \text{ participates in revolt at } t\}} = \begin{cases} 1 & \text{if } i \text{ participates in the protest at time } t \\ 0 & \text{if not} \end{cases}$$

The size of the current attack a_t is then given by the sum of current participants:

$$a_t := \sum_{i \in I} a_{it} \in \{0, 1, ..., N\},$$

where it takes the value of 0 in case no one participates in the attack, N in case all agents participate, and an intermediate integer between 0 and N if some fraction of the agents attack.

At the same time with the current attack a_t, the Dictator can take the policy measure $r_t \in [0, N - \theta]$ at some small fixed cost $\kappa > 0$ to increase the perceived strength of the regime. The upper bound depends on the realisation of θ: the Dictator cannot oversell her strength more than the upper bound known to the citizens. The policy represents the "soft power": media propaganda and censorship aimed at making the Dictator look strong.

The policy can backfire on the Dictator with an exogenous probability $p \in (0, 1)$: if the information "fed to the agents" from the pro-government media is too absurd or diverges too much from what they can learn from the alternative sources (social media, foreign media, their own experience from the protests), it undermines the status of the Dictator and maker her look weaker. Formally, for any chosen $r_t \in [0, N - \theta]$, the realisation of the policy r'_t is as follows:

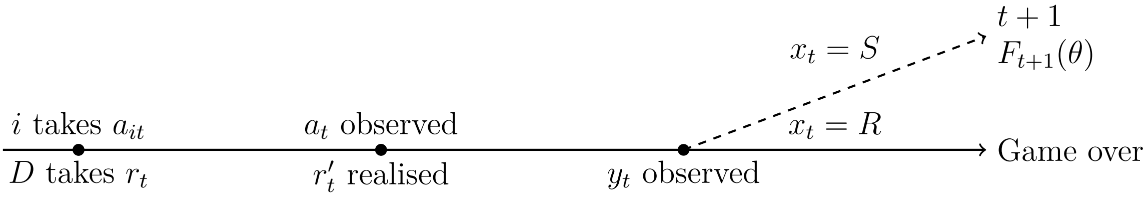
$$r'_t = \begin{cases} r_t & \text{with probability } p \\ -r_t & \text{with probability } 1 - p \end{cases}$$

The meaning behind the probability of success p is twofold: it captures the overall control over information sources, on the one hand, but also the level of trust of the citizens towards their government, on the other hand.

The citizens do not observe the chosen policy r_t nor the realisation of the policy r'_t, but only what I define as "the perceived strength of the regime" which combines the actual strength θ and the realised policy r'_t:

$$y_t := \theta + r'_t$$

Finally, let $x_t \in \{R, S\}$ be the set of possible outcomes at time t: R occurs if the riots are successful at the end of period t which happens if and only if $a_t > y_t$, while S represents the outcome where the status quo survives, that is, whenever $a_t \leq y_t$. If $x_t = R$, the game ends at period t and the revolution is successful. If $x_t = S$, the game moves to period $t+1$, with agents updating their beliefs about θ according to Bayes' rule, and then the same steps are repeated. To summarize, the timing at each round t of protests can be represented as follows:



3.2.2 Preferences

The utility of each agent $i \in I$ quasi-linear and depends on the outcome of the protests, her action, her type, and the cost of participating in the protests equal and constant for all agents: $c \in (0, 1)$. As mentioned before, agents are potentially heterogeneous in their marginal benefit from overthrowing the regime. At the beginning of the game, that is, at $t = 0$, each agents learns her private type $\tau_i \in \mathcal{T}_i$. The type spaces are same for all agents an normalized to between 0 and 1: $\mathcal{T}_i = \mathcal{T}_j = [0, 1]$ for all $i, j \in I$. Agents do not know the types of others know the distributions: each i knows that each other agent $\tau_j \sim G_j(\cdot)$ (with the probability mass function $g_j(\cdot)$, and the Dictator only knows the distributions of agents. Finally, types are distributed independently but not identically.

I assume the agents myopic: they only consider the current costs and benefits of protesting at each round t without taking into consideration past or future possible gains and losses. The utility at time t for an agent i is given as follows:

$$u_{it}(a_{it}) = x_t \tau_i - c a_{it}$$

Alternatively, at each $t \geq 1$, the payoff matrix is:

	R	S
$a_{it} = 1$	$\tau_i - c$	$-c$
$a_{it} = 0$	τ_i	0

Finally, the utility of the Dictator is zero in case she stays in power and infinitely negative if the revolution occurs: it is a simplifying assumption to assure that she would take the policy in order stay in power if there is a chance to do so, but if she is sure that regime will survive without any policy, she will choose $r_t = 0$ not to bear the cost κ.

Finally, the solution concept is perfect Bayesian equilibrium at each round of protests; and the updated beliefs about the strength of the regime at the end of period t becomes the "prior" belief at the beginning of $t + 1$.

3.3 Equilibrium analysis

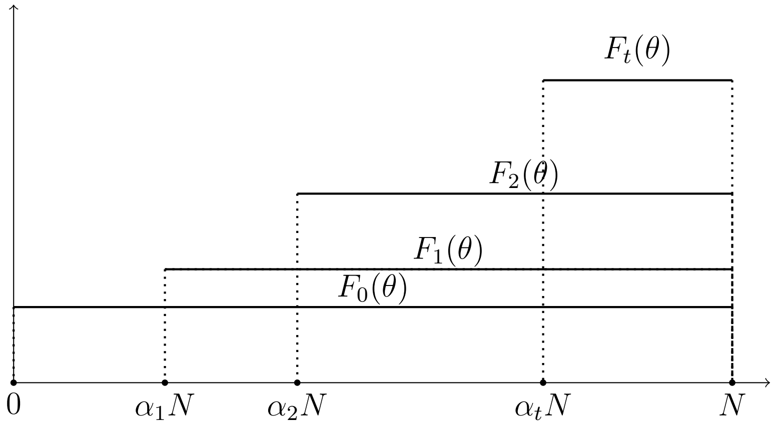
3.3.1 Homogeneous agents without any policy of the Dictator

The first benchmark to consider is the case where the Dictator cannot use even the imperfect policy while all the agents want the regime to fall: $\tau_i = 1$ for all $i \in I$. Formally, each agent's type has a degenerate distribution with probability mass 1 on having type equal to 1: $g_i(\tau_i) = \mathbb{1}_{\{1\}}$ for all $i \in I$, hence, all agents are IID.

In this case, at each $t \geq 1$, equating the expected utility of attacking with the expected utility of refraining from participation at protests pins down the probability with which each agent attacks:

$$\mathbb{E}[u_{it}(1)] = \mathbb{E}[u_{it}(0)] \tag{3.1}$$

If I denote such a probability α_t, the total expected size of the attack at t is $\alpha_t N$, and the weak regimes such that $\theta < \alpha_t N$ fall while the strong ones $\theta \geq \alpha_t N$ survive. Furthermore, upon the survival of the regime, the agents correctly update their beliefs about the strength of the regime upwards, that is, they realise that the regime is uniformly distributed on $[\alpha_t N, N]$, and use the new distribution to calculate the probability of being pivotal. Such a probability increases over time: the higher is the expected strength, the higher is the participation incentive as agents understand that their "contribution" becomes crucial for the regime to fail. Finally, as the participation rate increases over time and tends to 1, the regime will eventually fall. The beliefs dynamics are presented below:



This intuition is captured by the following lemma:

Lemma 1. *In a symmetric equilibrium, at each $t \geq 1$, all agents attack with probability α_t^* solving (with $\alpha_0 = 0$)*

$$\sum_{k=\alpha_{t-1}^* N}^{N-1} \binom{N-1}{k} (\alpha_t^*)^k (1-\alpha_t^*)^{N-1-k} \mathbb{P}_{t-1}(k \leq \theta \leq k+1) = c \tag{3.2}$$

The probability that the regime survives the attack conditional om previous survival is *given by*

$$\mathrm{P}_t(S) = \frac{1-\alpha_t^*}{1-\alpha_{t-1}^*} \tag{3.3}$$

where $(\alpha_t^*)_{t=0}^{\infty}$ *is an increasing sequence.*

The updated (conditional on $x_t = S$*) beliefs about* θ *is given by a uniform* $U([\alpha_t^* N, N])$

3.3.2 Homogeneous agents with imperfect policy of the Dictator

In this subsection, I consider the case where all the agents still want the regime to fall but the Dictator can use the imperfect policy.

Suppose the Dictator always chooses the highers policy to save the regime, that is, $r^* = N - \theta$. The realizations of the manipulated strength of the world (the state *theta* plus the realized policy) will be

$$y_1 = \theta + r_1' = \begin{cases} N & \text{with probability } p \\ 2\theta - N & \text{with probability } 1-p \end{cases}$$

If the policy is successful, the regime survives any size of the attack. If the policy backfires, there can be two cases:

- For $\theta \in [0, N/2]$, $y_1 = \theta + r_1' = 2\theta - N \leq 0$. Without loss, let me assume that it is normalized to be 0, that is, it fails after any positive size of the attack.
- For $\theta \in [N/2, N]$, $y_1 = \theta + r_1' = 2\theta - N \geq 0$, moreover, $y_1 = 2\theta - N$ is distributed uniformly over $[0, N]$ with the density $\frac{1}{2N}$. To see that, firstly note that

$$N/2 \leq \theta \leq N$$
$$N \leq 2\theta \leq 2N$$
$$0 \leq 2\theta - N \leq N$$

Moreover, it holds

$$\begin{aligned}
F_{y_1}(x) &= P(y_1 \leq x \,\big|\, \theta > N/2) \\
&= P(2\theta - N \leq x \,\big|\, \theta > N/2) \\
&= P\left(N/2 < \theta < \frac{x+N}{2}\middle|\theta > N/2\right) \\
&= \frac{P\left(N/2 < \theta < \frac{x+N}{2}\right)}{P\left(\theta > N/2\right)} \\
&= \frac{x}{2N}
\end{aligned}$$

Therefore, the distribution of the realized state of the world is given as follows:

$$F(y_1) = \frac{1}{2} + \frac{y_1}{2N} \quad \text{for } y_1 \in [0, N]$$

As for the agents, the equilibrium probability $\hat{\alpha}_1$ is found again from the indifference condition with a only addition that the left-hand side is multiplied by $1 - p$, that is, the probability that the policy of the Dictator fails:

$$\frac{1}{2}\sum_{k=0}^{N-1}\binom{N-1}{k}(\hat{\alpha}_1)^k(\hat{\alpha}_1)^{N-1-k}\left(\frac{1-p}{2N}\right) + \frac{1}{2}(1-\hat{\alpha}_1)^{N-1} = c$$

The probability that the regime survives is given by

$$\begin{aligned}
P_0(S) &= p\int_0^{\hat{\alpha}_1 N} f(y_1)\mathrm{d}y_1 + \int_{\hat{\alpha}_1}^{N} f(y_1)\mathrm{d}y_1 \\
&= \frac{p\hat{\alpha}_1 + 1 - \hat{\alpha}_1}{2}
\end{aligned}$$

Interestingly, the introduction of policy does not necessarily yields a higher probability of regime survival as it depends on the relation of $\alpha_1^*, \hat{\alpha}_1$ and p.

3.4 Appendix

Proof of Lemma 1

Proof. Let $t = 1$. Consider the behavior of an agent i, and let a_{-1i} denote the size of the attack by other agents:

$$a_{-1i} = \sum_{j \in Ii} a_{j1}. \tag{3.4}$$

Moreover, let α_1 be the probability of attacking by each agent. The expected utility from attacking is given as follows:

$$\begin{aligned}
\mathbb{E}[u_{it}(1)] &= \mathbb{P}_0(\theta \leq 1 + a_{-1i}) - c \\
&= \sum_{k=0}^{N-1} \binom{N-1}{k} \alpha_1^k (1-\alpha_1)^{N-1-k} \mathbb{P}_0(\theta \leq k+1) - c \\
&= \sum_{k=0}^{N-1} \binom{N-1}{k} \alpha_1^k (1-\alpha_1)^{N-1-k} \frac{k+1}{N} - c
\end{aligned}$$

The expected utility from refraining is given as follows:

$$\begin{aligned}
\mathbb{E}[u_{it}(1)] &= \mathbb{P}_0(\theta \leq 1 + a_{-1i}) - c \\
&= \sum_{k=0}^{N-1} \binom{N-1}{k} \alpha_1^k (1-\alpha_1)^{N-1-k} \mathbb{P}_0(\theta \leq k) \\
&= \sum_{k=0}^{N-1} \binom{N-1}{k} \alpha_1^k (1-\alpha_1)^{N-1-k} \frac{k}{N}
\end{aligned}$$

Therefore, the equilibrium probability of attacking solves the following equation:

$$\sum_{k=0}^{N-1} \binom{N-1}{k} (\alpha_1^*)^k (1-\alpha_1^*)^{N-1-k} \frac{1}{N} = c$$

Furthermore, the expected attack is $\alpha_1^* N$, and the outcome of the first round of protests is:

$$x_1 = \begin{cases} R & \text{if } \theta \leq \alpha_1^* N \\ S & \text{otherwise} \end{cases}$$

The updated distribution is

$$\begin{aligned} \mathbb{P}_1(\theta) &= \mathbb{P}_0(\theta | x_1 = S) \\ &= \frac{\theta - \alpha_1^* N}{N - \alpha_1^* N} \quad \text{for } \theta \in [\alpha_1^* N, N] \end{aligned}$$

Let $t = 2$ now. To find the probability with which agents attack, I need to solve a similar equation as before:

$$\sum_{k=\alpha_2^* N}^{N-1} \binom{N-1}{k} (\alpha_2^*)^k (1-\alpha_2^*)^{N-1-k} \frac{1}{N(1-\alpha_1^*)} = c$$

The lowest number of riot participants, however, must be above the new lower bound of the regime strength for the attack to be successful, and we have that $\alpha_2^* > \alpha_1^*$.
The outcome of the first round of protests is:

$$x_2 = \begin{cases} R & \text{if } \theta \leq \alpha_2^* N \\ S & \text{otherwise} \end{cases}$$

The updated distribution is

$$\begin{aligned} \mathbb{P}_1(\theta) &= \mathbb{P}_1(\theta | x_2 = S) \\ &= \frac{\theta - \alpha_2^* N}{N - \alpha_2^* N} \quad \text{for } \theta \in [\alpha_2^* N, N] \end{aligned}$$

Continuing the argument for a general t, the results of the lemma obtain. □

3.5 References

Acemoglu, Daron and Egorov, Georgy and Sonin, Konstantin, Institutional Change and Institutional Persistence (September 14, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-127.

Angeletos, George-Marios and Hellwig, Christian and Pavan, Alessandro, Dynamic Global Games of Regime Change: Learning, Multiplicity and Timing of Attacks (May 2007). ECONOMETRICA, Vol. 75, No. 3, pp. 711-756, May 2007.

Angeletos, George-Marios and Hellwig, Christian and Pavan, Alessandro, Signaling in a Global Game: Coordination and Policy Traps (June 1, 2006). Journal of Political Economy, Vol. 114, pp. 452-484, June 2006.

Angeletos, George-Marios, and Iván Werning. 2006. "Crises and Prices: Information Aggregation, Multiplicity, and Volatility." American Economic Review, 96 (5): 1720-1736.

Argenziano, Rossella, Severinov, Sergei, and Squintani, Francesco. 2016. "Strategic Information Acquisition and Transmission." American Economic Journal: Microeconomics, 8 (3): 119-55.

Barberà Sàndez, Salvador and Jackson, Matthew O., A Model of Protests, Revolution, and Information (October 2019).

Bergemann, Dirk and Valimaki, Juuso. 2000. "Information Acquisition and Efficient Mechanism Design." Yale University, Department of Economics Working Paper No. 1248.

Bester, Helmut and Ritzberger, Klaus. 2001. "Strategic Pricing, Signalling, and Costly Information Acquisition." International Journal of Industrial Organization 19 (9): 1347-1361.

Bester, Helmut, Lang, Matthias, and, Li, Jianpei. 2019. "Signaling versus Auditing." CESifo Working Paper Series No. 7183.

Caesmann, Marcel & Caprettini, Bruno & Voth, Hans-Joachim & Yanagizawa-Drott, David, 2021. "Going Viral: Propaganda, Persuasion and Polarization in 1932 Hamburg," CEPR Discussion Papers 16356, C.E.P.R. Discussion Papers.

Carlsson, H., & Van Damme, E. (1993). Global Games and Equilibrium Selection. Econometrica, 61(5), 989-1018.

Carter, E. B., & Carter, B. L. (2021). Propaganda and Protest in Autocracies. Journal of Conflict Resolution, 65(5), 919–949.

Carter, E., & Carter, B. (2022). When Autocrats Threaten Citizens with Violence: Evidence from China. British Journal of Political Science, 52(2), 671-696.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" American Economic Review, 109 (6): 2294-2332.

Constantine Boussalis, Alexander Dukalskis & Johannes Gerschewski (2023). Why It Matters What Autocrats Say: Assessing Competing Theories of Propaganda, Problems of Post-Communism, 70:3, 241-252.

Dranove, David, and Zhe Jin, Ginger. 2010. "Quality Disclosure and Certification: Theory and Practice." Journal of Economic Literature 48 (4): 935-963.

Edmond, C. (2013). Information Manipulation, Coordination, and Regime Change. The Review of Economic Studies, 80(4 (285)), 1422-1458.

Enikolopov, R., Makarin, A. and Petrova, M. (2020), Social Media and Protest Participation: Evidence From Russia. Econometrica, 88: 1479-1514.

Erica Frantz, Andrea Kendall-Taylor, Joseph Wright, Digital Repression in Autocracies (2020). Working paper.

Figueroa, Nicolás, and Guadalupi, Carla. 2022. "Price Signaling with Information Acquisition." Working paper.

Francesco Capozza, Haaland, Ingar, Roth, Christopher, Wohlfart, Johannes. 2021. "Studying Information Acquisition in the Field: A Practical Guide and Review," ECONtribute Discussion Papers Series 124, University of Bonn and University of Cologne, Germany.

Gabaix, Xavier, Laibson, David, Moloche, Guillermo, and, Weinberg, Stephen. 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." American Economic Review, 96 (4): 1043-1068.

Gehlbach, Scott and Sonin, Konstantin and Svolik, Milan, Formal Models of Nondemocratic Politics (August 1, 2015). Annual Review of Political Science, Forthcoming.

Gehlbach, Scott and Sonin, Konstantin, Government Control of the Media (April 20, 2014). Journal of Public Economics, vol. 118, pp. 163-171, October 2014.

Gertz, Christopher, 2014. Quality Uncertainty with Imperfect Information Acquisition, Center for Mathematical Economics Working Papers 487, Center for Mathematical Economics, Bielefeld University

Goldstein, Itay, and Chong Huang. 2016. "Bayesian Persuasion in Coordination Games." American Economic Review, 106 (5): 592-96.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. The American Economic Review, 70(3), 393–408

Guan, Xiaolan, and, Chen, Yehning. 2017. "The Interplay between Information Acquisition and Quality Disclosure." Production and Operations Management 26: 389-408.

Guriev, S. M. and Daniel, Treisman. 2022. Spin Dictators: The Changing Face of Tyranny in the 21st Century. Princeton; Oxford, Princeton University Press.

Guriev, S., & Treisman, D. (2019). Informational autocrats. Journal of Economic Perspectives, 33(4), 100-127.

Hollyer, J., Rosendorf, B., & Vreeland, J. (2015). Transparency, Protest, and Autocratic Instability. American Political Science Review, 109(4), 764-784.

Huang, H., Cruz, N. Propaganda, Presumed Influence, and Collective Protest. Polit Behav 44, 1789–1812 (2022).

Huang, Haifeng. “Propaganda as Signaling.” Comparative Politics 47, no. 4 (2015): 419–37.

Jackson, Matthew O. 1991. "Equilibrium, Price Formation, and the Value of Private Information." Review of Financial Studies 4 (1): 1-16.

Jowett, G. S., & O'Donnell, V. (2012). Propaganda and Persuasion (5th ed.). London: Sage.

Kara-Murza Sergei, Manipulyatsiya soznaniem [Consciousness manipulation]. Moscow: Eksmo, 2000.

Marquez, Robert S. and Hauswald, Robert B.H., Competition and Strategic Information Acquisition in Credit Markets (March 2002).

Martin, Daniel. 2017. "Strategic Pricing with Rational Inattention to Quality." Games and Economic Behavior 104: 131-145.

Martinez-Gorricho S. Signalling, Information and Consumer Fraud. Games. 2020; 11(3):29.

Matthews, Steven A. and Persico, Nicola G., Information Acquisition and the Excess Refund Puzzle (March 28, 2005).

Mattingly, Daniel and Yao, Elaine, How Propaganda Manipulates Emotion to Fuel Nationalism: Experimental Evidence from China (January 6, 2020).

Morris, S., & Shin, H. (1998). Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks. The American Economic Review, 88(3), 587-597.

Morris, Stephen Edward and Shin, Hyun Song, Global Games: Theory and Applications (August 2001). Cowles Foundation Discussion Paper No. 1275R.

Moscarini, Giuseppe, Ottaviani, Marco. (2001). Price Competition for an Informed Buyer. Journal of Economic Theory. 101. 457-493.

Persico, N. (2000). Information Acquisition in Auctions. Econometrica, 68(1), 135–148.

Qingmin Liu, Information Acquisition and Reputation Dynamics, The Review of Economic Studies, Volume 78, Issue 4, October 2011, Pages 1400–1425.

Raphael Boleslavsky, Mehdi Shadmehr, Konstantin Sonin, Media Freedom in the Shadow of a Coup, Journal of the European Economic Association, 2020.

Rochlitz, Michael and Schoors, Koen J. L. and Zakharov, Nikita, Can Authoritarian

Propaganda Compete with the Opposition on Social Media? Experimental Evidence from Russia (April 29, 2023).

Roesler, Anne-Katrin, and Balázs Szentes. 2017. "Buyer-Optimal Learning and Monopoly Pricing." American Economic Review, 107 (7): 2072-80.

Shirikov, Anton, Rethinking Propaganda: How State Media Build Trust Through Belief Affirmation. Working paper.

Skulenko, M. Y. (1987). Journalism and Propaganda. Kyev: Yzdateljstvo pry KGhU YO «Vyshha shkola».

Stahl, Konrad and Strausz, Roland, (2017). Certification and Market Transparency, Review of Economic Studies, 84, issue 4, p. 1842-1868

Tore Ellingsen. Price signals quality: The case of perfectly inelastic demand, International Journal of Industrial Organization, Volume 16, Issue 1, 1997, Pages 43-61, ISSN 0167-7187.

Voorneveld M, Weibull JW. A Scent of Lemon—Seller Meets Buyer with a Noisy Quality Observation. Games. 2011; 2(1):163-186.

Xianpei Hong, Xinlu Cao, Yeming Gong, Wanying Chen. Quality information acquisition and disclosure with green manufacturing in a closed-loop supply chain, International Journal of Production Economics, Volume 232, 2021, 107997, ISSN 0925-5273.