Resisting Persuasion

Elias Tsakas [*1], Nikolas Tsakas [†2], and Dimitrios Xefteris [‡2]

[1]Department of Economics, Maastricht University

[2]Department of Economics, University of Cyprus

February 20, 2020

Abstract

Agents that are subject to persuasion attempts often employ strategies that allow them to effectively resist. In the context of Bayesian Persuasion (Kamenica and Gentzkow, 2011), we argue that if appropriate action-contingent payoff adjustments are available to the subject of persuasion (e.g., through public commitment), then payoff improvements are achieved. Remarkably, payoff-improving resistance strategies *need not involve adding benefits to any action.* We characterize the optimal resistance strategy when only costly payoff adjustments are allowed and we show that it always induces a substantial increase in the agent's welfare and it often induces the design of a perfectly informative signal.

Keywords: Bayesian persuasion; resistance; uncertainty; public commitment.

JEL classification: D72, D82, D83, K40, M38

1. Introduction

Persuasive communication describes the process in which an agent (the Sender, male) intends to alter the behavior of another agent (the Receiver, female) in his favor. Attempted persuasion is commonly observed in communication related to economic and political decisions, such as product advertisement (Bertrand et al., 2010), or information provided by politicians or media to potential voters (DellaVigna and Kaplan, 2007). For this reason it has attracted a lot of academic interest both in economics (DellaVigna and Gentzkow, 2010; Glazer and Rubinstein, 2006; Kamenica and

[*] Department of Economics (AE1), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; E-mail: e.tsakas@maastrichtuniversity.nl

[†] Department of Economics, University of Cyprus, P.O. Box 20537, 1678, Nicosia, Cyprus; E-mail: tsakas.nikolaos@ucy.ac.cy

[‡] Department of Economics, University of Cyprus, P.O. Box 20537, 1678, Nicosia, Cyprus; E-mail: xefteris.dimitrios@ucy.ac.cy

Gentzkow, 2011) and psychology (Petty and Cacioppo, 1986; Perloff, 2017). A common theme in the literature is that the attempt of the Sender to change the behavior of the Receiver for his own benefit might not have a positive impact on the welfare of the Receiver. In such cases, the Receiver's natural reaction is to resist the intended persuasion. Resistance to persuasion may take several forms that depend on the context of communication and the incentives of the involved parties, and as a result it has been the subject of extensive research in several disciplines, including economics, psychology, communication and marketing (see, for instance, Knowles and Linn, 2004; Fransen et al., 2015; Jacks and Cameron, 2003).[1]

In the context of Bayesian persuasion (Kamenica and Gentzkow, 2011), communication occurs through informative signals designed by the Sender and intended to alter the beliefs of the Receiver regarding the state of nature, with both agents being Bayes-rational. More specifically, a Sender intends to persuade a Receiver to choose his most preferred action.[2] The optimal persuasion strategy, from the point of view of the Sender, provides a substantial increase in the Sender's payoff, without reducing the welfare of the Receiver. Indeed, a Sender cannot convince a Receiver to make "more mistakes" – with respect to the Receiver's preferences – compared to when persuasion does not take place, but he can influence the types of mistakes that the Receiver makes, thereby increasing his expected utility. In this framework a Receiver understands that a persuasion attempt allows a Sender to reap all the value of the informative message, and it is natural to expect that she will attempt to resist by claiming a piece of the value of the informative message for herself, provided appropriate resistance strategies are available.

In this paper, we introduce resistance in the Bayesian persuasion model, in a simple setup with a binary state space and a binary action space. Within this setup, we exhaustively study a specific class of resistance strategies, namely action-contingent payoff adjustments, which we regard as theoretically challenging and empirically relevant. The optimal signal designed by the Sender crucially depends on the preferences of the Receiver. More specifically, it depends on the degree of the prior bias to the action that the Receiver exhibits. In a binary setting (that is, if we have two alternative actions; say g and b), when the Receiver is moderately biased in favor of an action and the Sender wants to persuade her to choose the other one, the Sender can succeed in convincing her even by employing a message that is not very informative. However, when the Receiver feels very strongly in favor of her preferred action, the Sender must provide a very informative message to successfully persuade her. Additional information is always beneficial to the Receiver, since it reduces the chances of mistakes, and thus, appearing biased against the action preferred by the Sender is appealing to the Receiver. For an initially unbiased Receiver, though, to become credibly biased against the alternative preferred by the Sender (say, for example, against action g), it would involve either additional benefits when

[1] It may arise even absent explicit payoff incentives, as it is also related to the natural tendency of people to avoid being influenced. See Ringold (2002), as well as the large literature on psychological reactance (Brehm, 1966).

[2] The main difference between models of Bayesian persuasion (Kamenica and Gentzkow, 2011) and the earlier literature on cheap-talk (Crawford and Sobel, 1982) is that the former assumes that the Sender commits to his signal before observing his own type.

she chooses b or additional costs when she chooses g (or both). But are such payoff adjustments possible in applied settings of interest? The most straightforward example of a process that fits these descriptions is *public commitment.*

Public commitment is considered to be an efficient tool to resist persuasion in several settings. For instance, in international relations, Leventoglu and Tarar (2005) and Tarar and Leventoglu (2009) find public commitment to provide bargaining leverage in international negotiations, because it creates a cost for the agent when taking a certain action that she has ex–ante committed not to. This feature is also related to the theory of "audience costs", which refer to the costs suffered domestically by a leader who first escalates a situation to the status of an international crisis and then backs down (Fearon, 1994, 1997; Tomz, 2007). Resistance to persuasion via public commitment is also studied in marketing, as it is a common feature of consumer behavior. For instance, Gopinath and Nyer (2009) discuss several psychological explanations of this behavior. Despite not mentioning explicitly the feature of costs (benefits) induced by revoking on (sticking to) a public commitment, they relate it to social influence and preference towards consistency, which bears the idea of attention paid by individuals on the reactions of others regarding ones own decisions.[3] More generally, public commitment has broader implications in diverse areas, like in the promotion of socially beneficial behaviors (Lokhorst et al., 2009), the efficacy of selling techniques (Cialdini et al., 1978) and the formation of opinions (Jellison and Mills, 1969).

So, in many cases a Receiver has the power to strategically adjust her bias before the Sender attempts to persuade her. The natural next question is then: Can such a strategy be welfare improving for the Receiver? In this paper, we show that deterministic resistance strategies ("decrease my utility by κ if I take the action preferred by the Sender and increase my utility by β if I take the opposite action") always improve the Receiver's welfare and the relative informativeness of the message, compared to the case when no resistance takes place, *for any arbitrarily small* benefit attached to making the choice least desired by the Sender. This is, arguably, a very strong result since it does not require that the benefit, β, depends on the cost, κ, in order for the resistance strategy to improve the Receiver's welfare. In fact, it might very well be the case that a resistance strategy involves adding very large costs, when choosing according to the Sender's will, and only tiny benefits when choosing against it, and that it still provides a larger expected welfare to the Receiver! Indeed, making the Receiver better off simply by increasing the benefits of choosing some action is somewhat uninteresting, but the fact that even the smallest increase in the payoff when choosing against the Sender's will can offset any substantially large cost of choosing according to his will, makes the finding relevant to real world persuasion settings.

Moreover, it is sometimes true that a resistance strategy brings along payoff adjustments that are subject to uncertainty. For instance, there has been a large amount of research in political science

[3] Resistance to persuasion via public commitment is also associated with susceptibility to normative influence (Batra et al., 2001), preference consistency (Wells and Iyengar, 2005), source credibility (Tormala and Petty, 2004) and attitude certainty (Tormala and Petty, 2002).

on the effect of exogenously imposed domestic constraints on international negotiations (Putnam, 1988). Domestic constraints on the ratification of proposals –such as the approval of parliament, a referendum, or even veto power to other political entities– and potential legal constraints create uncertainty during negotiations, as the negotiating parties cannot be sure about the future behavior of other entities (Iida, 1993; Mo, 1994, 1995). It is sometimes possible for an authority to strategically call for a ratification process (e.g. propose a referendum) or request for external legal advice. If this occurs prior to negotiations, it can prove beneficial, as the final agreement should take into account the potential reactions from the other involved agents.

For this reason, we consider a more general spectrum of resistance strategies that includes not only deterministic action-contingent adjustments but also probabilistic ones. We find that the introduction of uncertainty has non-trivial implications on the persuasion process: A Receiver can substantially increase her expected welfare by never adding benefits to any action and by only undertaking occasional costs! We study cost-only resistance strategies in detail ("my utility from taking the action least preferred by the Sender is not adjusted and my utility from choosing the action preferred by the Sender is reduced by κ, where κ is randomly drawn from a distribution F with support in $\mathbb{R}_+$") and we try to characterize the optimal F, allowing for the support of F to be (a) a unique point, (b) binary, and (c) any subset of $\mathbb{R}_+$.

For the first case, we show that the expected welfare of the Receiver is constant independent of where the unique point of the support of F is. That is, introducing a deterministic cost when choosing the alternative most preferred by the Sender, leaves the Receiver's expected utility invariant, while the informativeness of the message is strictly increasing in the size of this cost.[4] For each of the other two cases, we show that the optimal cost-only resistance strategy strictly improves the Receiver's welfare and in some cases it induces the design of a perfectly informative signal. That is, uncertainty regarding cost-bearing can be beneficial for the Receiver, and this does not happen only when the Receiver can design sophisticated strategies (that is, when the support of F is allowed to be any subset of $\mathbb{R}_+$) but also in the simplest case with non-degenerate uncertainly (that is, when the support of F is binary). The globally optimal cost-only resistance strategy is a rather intriguing one: It assigns a strictly positive probability of taking a cost equal to zero and distributes the rest of the probability continuously to costs from zero to some positive threshold that depends on the players' preferences. When the Sender has state-independent preferences, the globally optimal strategy induces the design of perfectly informative signal irrespective of the exact preferences of the Receiver. This might also be achieved with a binary distribution, as long as the Receiver cares predominantly about choosing the "correct" action in the state in which the Sender would like to persuade her to do the opposite.

Our work is related to research on Bayesian persuasion (Kamenica and Gentzkow, 2011), which has developed in several different directions. Alonso and Câmara (2017) study Bayesian persuasion with multiple receivers and briefly discuss commitment as a potential welfare enhancing strategy

[4] This is the driving force behind our result that even the tiniest additional payoff when choosing the Sender's least preferred alternative can induce an increase in the Receiver's expected utility.

for the receivers, which in that context takes a different form than in our environment and, more importantly, is not incentive compatible for the receivers. Kolotilin et al. (2017) study persuasion with a privately informed receiver, which can reduce the persuading ability of the sender. This is in contrast to our setup in which the two agents possess the same amount of information at any point during the process. Nevertheless, similar to our results, under certain conditions, the sender may choose to design a signal that reveals the true state of nature. Increased signal informativeness may sometimes also be a result of competition between senders (Gentzkow and Kamenica, 2017a,b) or noise in the communication between the Sender and the Receiver (Tsakas and Tsakas, 2017). Furthermore, Alonso and Câmara (2016) have considered an extended model with heterogeneous priors, whereas Perez-Richet (2014) and Hedlund (2017) consider a similar game with a privately informed sender. Finally, Hagmann and Loewenstein (2017) study biases in information processing that affect persuasion, based on whether the provided information supports the receiver's prior beliefs. For a recent review of this literature, we refer to Kamenica (2019).

Our results also relate to those of the wider money-burning literature and entail that persuasion approaches should control for potential resistance strategies that might be available to the subjects of the persuasion attempts. Indeed, it is known that in many instances destroying own utility is an effective means of convincing other players to behave according to one's interests (Ben-Porath and Dekel, 1992; van Damme, 1989). For instance, as far as communication frameworks are concerned, money burning is proved to expand the set of equilibrium outcomes (Austen–Smith and Banks, 2000; Kartik, 2007), bringing along possibilities for payoff enhancements.

Moreover, in the context of optimal delegation contract design (Amador and Bagwell, 2016) it has recently been shown that a principal can enhance her utility by inducing action-contingent money-burning to the agent (Ambrus and Egorov, 2017). That is, a contract designer may be better off by just punishing the agent for taking specific actions – and not claiming any benefit from the loss in utility experienced by the agent – compared to simply imposing a transfer to her benefit. In a way, our work combines these intuitions and, to our knowledge, this paper is the first to consider action-contingent burning of own-utility. Additionally, our approach is linked to more traditional communication settings, related to price discrimination, in which a receiver can make the sender indifferent among all signals that are optimal for some cost realization (see for instance Bergemann et al., 2015; Roesler and Szentes, 2017). This is rather reassuring as it guarantees that the nature of the optimal resistance strategy in the context of Bayesian persuasion is well accepted in alternative frameworks of information transmission.

Finally, since we allow resistance strategies to take the form of probability distributions with an arbitrary unidimensional support, our attempt to characterize the optimal cost-only resistance strategy relates to setups whose objectives are technically the same. Beyond games with only mixed equilibria, which naturally fall into this category, there are a number of setups which directly consider that a strategy is a function defined over a continuous set. Famously, Myerson (1993) characterizes the optimal distribution of transfers in a probabilistic redistribution game for an office motivated candidate and finds it to be uniform, while the non-linear income taxation literature tries to identify optimal univariate taxation schemes (e.g. Lehmann et al., 2014).

In what follows, we first present an example that helps establish the ideas behind our model (Section 2). Next, we describe the model (Section 3) and the formal results (Section 4) and then we conclude (Section 5).

2. Motivating Example

Consider a regulator (the Receiver, female) that is about to enter a series of meetings with lobbyists of a given industry (the Sender, male) regarding whether this industry maintains a reduced-tax regime or not. The lobbyists obviously prefer the reduced-tax regime to be maintained, whereas the regulator prefers to make the right choice for the local economy, which depends on the state of nature. Namely, the global economic environment in the industry may provide opportunities for relocation to some emerging market or not. If such opportunities exist then an abolishment of the reduced–tax regime will lead major companies to move their headquarters to this emerging market. If it does not exist, then even with higher taxes, these same companies will choose to keep their headquarters in the country.

A common strategy employed by the firms and their lobbyists is to hire an independent consulting agency to provide a report that constitutes a noisy signal of the actual global economic environment. Although the results of the analysis must be reported truthfully, the lobbyists could design strategically the type of analysis they ask for, as this could affect the chances of persuading the regulator to maintain the favorable regime. On her side, the regulator would like to maintain the reduced–tax regime only if revoking it would drive a significant share of the firms to move their headquarters out of the country. If the regulator and the lobbyists share a common prior belief regarding the state of the economic environment, then the lobbyists can sometimes persuade the regulator to maintain the current regime.

For instance, let the regulator enjoy one unit of utility when she makes the right decision and no utility otherwise and let the lobbyists enjoy one unit of utility if the reduced–tax regime is maintained and no utility otherwise. In this scenario, the regulator and the lobbyists share a common prior belief that the environment in the emerging market is favorable towards relocation of the firms with a relatively small probability, say, $p_0 = 0.3$. In this case, absent of persuasive attempts from the lobbyists, the regulator would choose to revoke with certainty, which guarantees a higher expected utility for herself. However, as shown in Kamenica and Gentzkow (2011), the lobbyists can request a strategically designed analysis that would lead the regulator to maintain the reduced–tax regime with probability 0.6. They achieve this not by inducing the regulator to more frequently make a mistaken call (both with and without persuasion the regulator decides correctly with a probability equal to 0.7), but by changing the distribution of correct and mistaken calls. That is, when persuasion takes place, the reduced-tax regime is always maintained, if revoking will indeed lead the companies relocate and is also maintained with positive probability if this is not the case.

Public Commitment: The regulator is aware of the potential attempts of the lobbyists to persuade her, as well as that these attempts will probably influence her behavior in their favor. Thus, the natural dilemma that she faces is whether she has any way to resist the persuasion attempts and actually increase the probability that she makes the correct decision. A viable solution to this dilemma could be *public commitment.* For instance, assume that the regulator is an elected politician and before meeting with the lobbyists she can make a public commitment that she will revoke the reduced-tax regime. When making such a commitment, the politician knows that not following it will induce a political cost, $\kappa > 0$, since her credibility will be reduced. On the contrary, keeping her promise may provide a boost to her credibility, that can be translated to a benefit $\beta > 0$. Therefore, such a commitment makes an ex-ante unbiased regulator, to become biased in favor of revoking the reduced-tax regime.

This strategy can have an important effect on the persuasion attempts of lobbyists because they now know that the analysis they should provide to the regulator should be more informative than before, in order to succeed in persuading her with positive probability. In fact, the stronger the commitment of the politician (higher κ and/or higher β), the more informative the provided analysis should be.[5] If stronger public commitment leads to a greater decline in credibility when the politician breaks her promise and also to a higher boost in credibility when the politician keeps her promise (β is increasing in κ), then the lobbyists have to ask for a fully informative analysis. This makes the regulator always choose correctly and, importantly, enjoy a larger expected utility compared to when she makes no public commitment! That is, making such a public commitment before the talks makes a regulator effectively resist persuasion, and the result is an increase in both social welfare (in the sense that it increases the probability of taking the correct decision) and her own private utility.

Introducing Uncertainty: The regulator can improve her expected welfare even more by also *introducing uncertainty* regarding her ex–post action. For instance, instead of making a public commitment, she informs the lobbyists that she has asked for legal advice regarding whether she has the right to maintain the favorable tax regime or not, and that this advice will arrive before she makes any decisions. If the legal advice suggests that the regulator can maintain the favorable tax regime, then any choice –maintaining it or not– induces no extra cost or benefit. If the legal advice though suggests that the regulator cannot maintain the favorable tax regime, then things are substantially more complicated: If the regulator decides not to maintain the tax regime, there is no additional cost/benefit, but if she decides to ignore the legal advice and maintain it she will incur a significant cost. Notice that this action–contingent cost (which is never undertaken when the regulator decides not to maintain the tax regime) is not to be realized with certainty, but it might still influence the relative informativeness of the lobbyists persuasion attempt. Indeed, as we

[5] This is provided, of course, that the politician does not become excessively committed and cannot be persuaded even with perfect information available.

will show, if the regulator has such resistance strategies at her disposal she can make the lobbyists provide more accurate information and, perhaps more importantly, she can enjoy a larger expected utility. The most interesting feature of these probabilistic mechanisms is that the welfare of the regulator increases even without explicit positive gains, as was the case with public commitment. That is, there is no need to enjoy benefits when deciding against the lobbyists's interests to enjoy a strict increase in her welfare along with a more informative message.

3. The Model

The benchmark persuasion game (without resistance): Let $\Omega = \{G, B\}$ be a binary state space and $A = \{g, b\}$ be a binary action space. There are two agents, a male Sender and a female Receiver with utility functions $v : A \times \Omega \to \mathbb{R}$ and $u : A \times \Omega \to \mathbb{R}$, respectively. Both agents are Bayesian expected utility maximizers and share a common prior $\mu_0 \in \Delta(\Omega)$ assigning probability $p_0 := \mu_0(G) \in (0, 1)$ to the state G.

Before the Receiver chooses an action, the Sender chooses a signal/experiment $\pi : \Omega \to \Delta(S)$, which is represented by a pair of distributions $\pi(\cdot|G)$ and $\pi(\cdot|B)$ over a finite set of signal realizations, S. The choice of the signal is observed by both players, as is the actual realization. Hence, information is symmetric throughout the game. Formally, given the signal π, upon observing a realization $s \in S$, both agents update their beliefs to a posterior $\mu_s \in \Delta(\Omega)$ that attaches probability

$$p_s := \frac{p_0 \pi(s|G)}{p_0 \pi(s|G) + (1 - p_0)\pi(s|B)}$$

to the state being G. Then, using her updated belief μ_s, the Receiver chooses an action $a \in A$ that maximizes her expected utility,

$$u_s(a) := \sum_{\omega \in \Omega} \mu_s(\omega) u(a, \omega).$$

Whenever the Receiver is indifferent between the two actions she chooses the action most preferred by the Sender. If the Sender is also indifferent between the two, then he chooses arbitrarily. Let us denote the Receiver's action for an arbitrary posterior μ by $\hat{a}(\mu)$, and denote the Sender's respective utility at a state $\omega \in \Omega$ by $v(\hat{a}(\mu), \omega)$. Hence, the Sender's problem reduces to choosing a signal π that maximizes his (ex ante) expected utility

$$V(\pi) := \sum_{\omega \in \Omega} \sum_{s \in S} \mu_0(\omega) \pi(s|\omega) v(\hat{a}(\mu_s), \omega). \tag{1}$$

Note that the Sender's optimal signal strategy always exists and is characterized by means of the standard concavification technique (Kamenica and Gentzkow, 2011).

Utility specifications: We naturally assume that the Receiver's preferences are state-dependent.

Otherwise, the analysis is trivial, as the Receiver always chooses her preferred action irrespective of her beliefs or the signal sent by the Sender. In particular, let us assume that the Receiver wants "to match the true state", i.e., let $u(g, G) > u(b, G)$ and $u(b, B) > u(g, B)$.

We assume that the Sender strictly prefers action g over action b irrespective of the true state, i.e., $v(g, G) > v(b, G)$ and $v(g, B) > v(b, B)$. Therefore, the meaning of persuasion in this context is that the Sender wants to persuade the Receiver to choose action g more often than she would otherwise do given her prior. For our main results we consider that *the Sender has state-independent preferences*, i.e. $v(g, G) = v(g, B)$ and $v(b, G) = v(b, B)$. Later on, we relax this assumption and discuss the impact of the Sender's preferences on the results.

It is helpful to define the quantities $\Delta u_G := u(g, G) - u(b, G)$, $\Delta u_B := u(b, B) - u(g, B)$, $\Delta v_G := v(g, G) - v(b, G)$ and $\Delta v_B := v(g, B) - v(b, B)$, which signify the excess utility that each agent gets at each state of nature if her/his most preferred action for that state is chosen. Note that, by construction, these four quantities are always strictly positive and when the Sender has state-independent preferences it holds that $\Delta v_G = \Delta v_B$.

Finally, throughout the paper, the common prior p_0 is assumed to be sufficiently low to ensure that the Sender has an incentive to attempt persuading the Receiver. The formal condition that describes this is that $p_0 < \frac{\Delta u_B}{\Delta u_G + \Delta u_B}$ (see Remark 1 below).

Resistance strategies: The Receiver is aware of the Sender's upcoming persuasion attempt. Thus, prior to the design of the signal, she may set up a *resistance strategy* against persuasion, which is based on *commitment*. More specifically, we define a *commitment mechanism* as a vector $c = (c(g), c(b)) \in \mathbb{R}^2$ of utils to be gained or lost by the Receiver for each of the two actions.[6] The Receiver's overall utility is assumed to be additively separable, i.e., for each $c \in \mathbb{R}^2$, her utility is given by

$$u^c(a, \omega) := c(a) + u(a, \omega).$$

Throughout the paper, we focus on commitment mechanisms in the (convex and compact) set

$$\mathcal{M} := \{c \in \mathbb{R}^2 : (c(g), c(b)) = (-\kappa, \beta), \text{ for } \kappa \geq 0 \text{ and } \beta \geq 0 \text{ and } \kappa + \beta \leq \Delta u_G\}.$$

That is, the Receiver commits to bear a cost if she chooses the Sender's preferred action and will (perhaps) have a benefit if she chooses the alternative action. The condition $\kappa + \beta \leq \Delta u_G$ guarantees that persuasion is possible, so that the problem is non-trivial (see Equation (4) below). Moreover, we define the no-commitment mechanism $c_0 := (0, 0) \in \mathcal{M}$. Notice that the Sender's utility function is not affected by the commitment mechanism, i.e.,

$$v^c(a, \omega) := v(a, \omega)$$

[6] Formally, there is an underlying set of outcomes, together with an unbounded vNM utility function. Then, the Receiver commits to receive a vNM lottery conditional on each of her own actions.

for all $(a, \omega) \in A \times \Omega$ and all $c \in \mathcal{M}$. In what follows, we mainly focus on a special type of commitment mechanisms, viz., those that yield no benefit to the Receiver when b is chosen. Formally, we define

$$\mathcal{C} := \{c \in \mathcal{M} : \beta = 0\}.$$

These mechanisms bear striking similarities to the extensive literature on "burning money" (e.g. Amador and Bagwell, 2016; Ambrus and Egorov, 2017; Austen–Smith and Banks, 2000; Kartik, 2007). We refer to those as sets of *cost–only commitment mechanisms*.

Then, we define a *resistance strategy* as a distribution $r \in \Delta(\mathcal{M})$ over the space of commitment mechanisms. Resistance strategies that put probability one to a single commitment mechanism are called *deterministic* and the rest are called *stochastic*. Formally, the set of deterministic resistance strategies is denoted by

$$\mathcal{DR} := \{r \in \Delta(\mathcal{M}) : r(c) = 1 \text{ for some } c \in \mathcal{M}\},$$

i.e., it is the set of Dirac measures over $\mathcal{M}$. The set of stochastic resistance strategies is obviously denoted by $\mathcal{SR} := \Delta(\mathcal{M}) \setminus \mathcal{DR}$. The set $\mathcal{DCR}$ of *deterministic cost-only resistance strategies* is naturally defined by

$$\mathcal{DCR} := \{r \in \mathcal{DR} : r(\mathcal{C}) = 1\},$$

i.e., these are (degenerate) resistance strategies that put probability one to a commitment mechanism that yields some cost $\kappa \geq 0$ for the Receiver if g is chosen, and no benefit if b is chosen. We use r_0 to denote the deterministic strategy that puts probability one to the no-commitment mechanism $c_0 = (0, 0)$.

The set of *binary stochastic cost–only resistance strategies* contains all strategies that assign probability $l \in [0, 1]$ to a mechanism $c = (-\kappa, 0) \in \mathcal{C}$ and the remaining probability $1 - l$ to the degenerate mechanism $c_0 = (0, 0) \in \mathcal{C}$, i.e.,

$$\mathcal{BSCR} = \{r \in \Delta(\mathcal{C}) : r(\{c_0, c\}) = 1, \text{ for some } c \neq c_0\}$$

Finally, the set of *general stochastic cost–only resistance strategies* contains all strategies that are distributed over cost–only commitment mechanisms, with positive mass on (at most) finitely many mechanisms $c \in \mathcal{C}$, i.e.,

$$\mathcal{GSCR} = \{r \in \Delta(\mathcal{C}) : \text{there is a finite } C \in \mathcal{C} \text{ such that } r(c) = 0 \text{ for all } c \notin C\}.$$

For each $r \in \mathcal{GSCR}$ let $F_r : [0, \Delta u_G] \to [0, 1]$ denote the cumulative distribution function, i.e., $F_r(\kappa) = r(\{c \in \mathcal{C} : -c(g) \leq \kappa\})$. Obviously each $r \in \mathcal{GSCR}$ is identified by F_r, and therefore the set $\mathcal{F}$ of all such CDF's represents $\mathcal{GSCR}$.

Notice that, by construction, $\mathcal{DCR} \subseteq \mathcal{BSCR} \subseteq \mathcal{GSCR}$. The essential difference between these sets of available resistance strategies is that they allow the Receiver to introduce more uncertainty, which will turn out to be beneficial for her. Moreover, all sets of resistance strategies are essentially based on commitment by the Receiver against the preferred action of the Sender.

Persuasion game with resistance: The timing of our game is as follows (see Figure 1). First, the Receiver chooses a resistance strategy r from a choice set $\mathcal{R} \subseteq \Delta(\mathcal{M})$ that always contains the degenerate no-commitment strategy r_0. The resistance strategy becomes commonly known. Then, the Sender chooses a signal. They both observe the signal realization, $s \in S$, and update their beliefs. Similar to the benchmark case, both the signal and the realization are common knowledge. Subsequently, a commitment mechanism, $c \in \text{supp}(r)$, is drawn and observed by both agents. Finally, the Receiver chooses an action that maximizes her (ex-post) expected utility,

$$u_s^c(a) := c(a) + u_s(a).$$

The optimal action of the Receiver depends both on her posterior belief μ_s and on the realized commitment mechanism c and is denoted by $\hat{a}(\mu_s, c)$. The Sender's expected utility at a state $\omega \in \Omega$ is now denoted by $v(\hat{a}(\mu_s, c), \omega)$. Hence, the Sender's expected utility from choosing a signal π when the Receiver has chosen a resistance strategy r becomes

$$V_r(\pi) = \sum_{\omega \in \Omega} \sum_{s \in S} \mu_0(\omega)\pi(s|\omega) \int_{\mathcal{M}} v(\hat{a}(\mu_s, c), \omega) dr(c)$$

The optimal signal for the Sender (given a resistance strategy r) is denoted by $\hat{\pi}_r$. Finally, the ex-ante expected utility of the Receiver from choosing a resistance strategy $r \in \mathcal{R}$ becomes:

$$U(r) = \sum_{\omega \in \Omega} \sum_{s \in S} \mu_0(\omega) \int_{\mathcal{M}} \hat{\pi}_r(s|\omega) u_s^c(\hat{a}(\mu_s, c), \omega) dr(c)$$

Graphically, the timing of the persuasion game with a set of resistance strategies $\mathcal{R}$ looks as follows. Events above the line are observed by both players, whereas events below the line are not observed by anyone. The draw of the signal realization is conditionally independent to the draw of the state of nature, whereas the draw of the commitment mechanism is independent to the other two. The order of steps that correspond to $t = 3$ and $t = 4$ can be reversed, provided the realization of the commitment mechanism takes place after the choice of the signal by the Sender and before

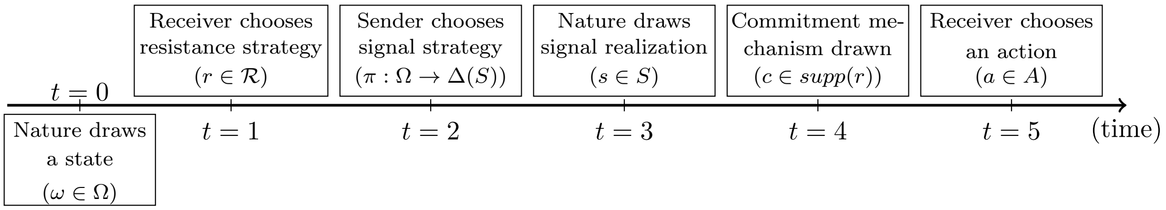
Receiver chooses resistance strategy ($r \in \mathcal{R}$)
Sender chooses signal strategy ($\pi : \Omega \to \Delta(S)$)
Nature draws signal realization ($s \in S$)
Commitment mechanism drawn ($c \in supp(r)$)
Receiver chooses an action ($a \in A$)
$t = 0$
$t = 1$
$t = 2$
$t = 3$
$t = 4$
$t = 5$
(time)
Nature draws a state ($\omega \in \Omega$)

Figure 1: Persuasion game with resistance.

the choice of the action by the Receiver. Moreover, step 4 is trivial for deterministic commitment strategies, in which case it is omitted.

4. Results

4.1. Preliminary Findings

Before proceeding to the main results of our study regarding cost-only resistance strategies, it is helpful to provide some preliminary general results that hold true in all the cases we consider. These results do not depend on whether the Sender has state-independent preferences or not, yet they bear similarities with previous results on Bayesian persuasion with a binary choice (see Alonso and Câmara, 2016; Kolotilin, 2015).

First, it can be shown that the Sender can design an optimal signal that puts positive probability to two signal realizations.[7] Hence, we can restrict the set of signal realizations to some $S = \{s_G, s_B\}$. [8] Thus, a signal π is represented by a pair of probabilities $q := \pi(s_G|G)$ and $z := \pi(s_G|B)$ and will sometimes be mentioned as *signal* (q, z).

Second, for a signal (q, z), the Receiver may form two posteriors regarding the probability that the state is G, one for each signal realization,

$$p_{s_G} = \frac{p_0 q}{p_0 q + (1-p_0)z} \qquad \text{or} \qquad p_{s_B} = \frac{p_0(1-q)}{p_0(1-q) + (1-p_0)(1-z)}. \tag{2}$$

For a realization $s \in \{s_G, s_B\}$ and the respective posterior $p_s \in \{p_{s_G}, p_{s_B}\}$, the expected utility of the Receiver from choosing action g or action b, respectively, is

$$u_s^c(g) = p_s \cdot u(g, G) + (1 - p_s) \cdot u(g, B) - \kappa, \tag{3a}$$
$$u_s^c(b) = p_s \cdot u(b, G) + (1 - p_s) \cdot u(b, B) + \beta. \tag{3b}$$

Therefore, the Receiver chooses action g if and only if $u_s^c(g) \geq u_s^c(b)$, or equivalently whenever

$$p_s \geq \widetilde{p} := \frac{\kappa + \beta + \Delta u_B}{\Delta u_G + \Delta u_B}. \tag{4}$$

[7] The argument is similar to the one of Kamenica and Gentzkow (2011). In particular, the only complications of restricting to only two realizations appear when a stochastic resistance strategy has been used by the Receiver. For instance, consider a binary stochastic cost-only strategy. In this case, as we will see in our analysis later in this section, the Receiver is never persuaded for low posterior beliefs, she is always persuaded for high beliefs, and is sometimes persuaded for intermediate beliefs (depending on the mechanism that will be drawn). So, from the Sender's point of view, we can consider the Receiver's response at these intermediate beliefs as a third artificial action of the Receiver, and finally apply the standard concavification technique in a game with two states and three actions. Thus, using the usual Caratheodory-based argument, two signal realizations suffice for the Sender to achieve his maximum expected utility.

[8] Given that there are only two states, the Sender can maximize his expected using a signal with merely two realizations. The argument is similar to the one of Kamenica and Gentzkow (2011). The only potential complication may appear in cases where the Receiver applies a stochastic resistance strategy. In particular, consider a

Recall that when the Receiver chooses her action, the commitment mechanism has already been drawn. Hence she knows the values of κ and β.

Remark 1. The assumption $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ guarantees that the Sender cannot persuade the Receiver in both signal realizations irrespective of the commitment mechanism. Indeed, $p_{s_G} \geq \widetilde{p}$ (resp., $p_{s_B} \geq \widetilde{p}$) implies $p_{s_B} < \widetilde{p}$ (resp., $p_{s_G} < \widetilde{p}$). Hence, the Sender focuses on persuading the Receiver to take action g in one realization, say realization s_G. $\triangleleft$

Thus, the Sender chooses a signal that maximizes his expected utility, subject to the constraint $p_{s_G} \geq \widetilde{p}$. For such a signal, the Receiver chooses action g upon observing signal realization s_G and action b upon observing s_B.

The next lemma characterizes the optimal signal and the ex-ante expected utilities of the two agents for an arbitrary deterministic resistance strategy, $r \in \mathcal{DR}$.

Lemma 1. *Assume $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ and let the Receiver choose a deterministic resistance strategy, $r \in \mathcal{DR}$, that assigns probability one to a mechanism $c = (-\kappa, \beta) \in \mathcal{M}$. Then, the optimal signal for the Sender is $\hat{\pi}_r(s_G|G) = 1$ and $\hat{\pi}_r(s_G|B) = p_0(1-\widetilde{p})/\left[(1-p_0)\widetilde{p}\right]$, and if the Receiver responds optimally to $\hat{\pi}_r$, the ex-ante expected utilities of the Receiver and the Sender are*

$$\begin{aligned} U(r) &= \beta + p_0 u(b,G) + (1-p_0)u(b,B), \\ V_r(\hat{\pi}_r) &= p_0 v(g,G) + (1-p_0)v(b,B) + p_0\left(\frac{1}{\widetilde{p}} - 1\right)\Delta v_B, \end{aligned}$$

respectively.

Some immediate observations can be made from the previous result. First, the Sender still designs the signal in a way that makes the Receiver indifferent between choosing each of the actions when the signal realization is s_G, as she did in the case without resistance. Obviously, the informativeness of the signal should be higher in order to compensate for the cost induced to the Receiver when choosing g. Moreover, the expected utility of the Receiver is independent of the cost κ since she is compensated for this cost because the signal is more informative. However, the gain from choosing action b is capitalized on by the Receiver, as it enters her expected utility function positively. This leads immediately to our next result.

Proposition 1. *Let $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$. Then, for deterministic strategies $r, r' \in \mathcal{DR}$ that put probability one to mechanisms $(-\kappa, \beta), (-\kappa', \beta') \in \mathcal{M}$ respectively, $U(r) > U(r')$ if and only if $\beta > \beta'$.*

A direct corollary of the above result is that $U(r) > U(r_0)$ if and only if $\beta > 0$. That is, as long as a resistance strategy allows for persuasion to take place and assigns a non–degenerate benefit to the Receiver when she chooses the action least preferred by the Sender, then this resistance strategy strictly improves the welfare of the Receiver. The reading of this result becomes even stronger when one notices that this is true independently of the size of the costs undertaken by the Receiver when she decides against the Sender's preference. Indeed, this makes the analysis empirically relevant, since if only large benefits (that is, large values of β) were necessary for a successful resistance, then it would be arguably hard to claim that resistance strategies are available in many real-life instances. "Destroying" own-utility is far easier than generating additional own-utility, in any possible context.

An interesting set of such resistance strategies arises when κ and β are (strictly) positively correlated. In particular, fix an arbitrary strictly increasing continuous function $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(0) = 0$, and consider the following set of deterministic resistance strategies:

$$\mathcal{DR}_h := \{r \in \mathcal{DR} : r(-\kappa, h(\kappa)) = 1 \text{ for some } \kappa \geq 0\}$$

It is apparent that for each set of strategies associated with such a function the following result holds as a direct consequence of Lemma 1.

Proposition 2. *Assume that $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ and let $\mathcal{R} = \mathcal{DR}_h$ for some strictly increasing continuous function $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(0) = 0$. Then the (unique) optimal resistance strategy $\hat{r}$ assigns probability 1 to the mechanism $(-\hat{\kappa}, \hat{\beta})$ that satisfies $\hat{\kappa} + \hat{\beta} = \Delta u_G$.[9] Moreover, the optimal resistance strategy makes the Sender provide a fully informative signal, i.e. $\hat{\pi}_{\hat{r}}(s_G|G) = \hat{\pi}_{\hat{r}}(s_B|B) = 1$.*

The essence of this result is that the Receiver is willing to commit to taking the highest admissible cost κ when choosing the Sender's preferred option, as this cost is associated with the highest potential benefit β when choosing against it, the value of which is shown in Lemma 1 to be the only one that matters. This suggests that a Receiver might be willing to commit very strongly against a given action, as long as this commitment will lead to more accurate information from the Sender and to a higher benefit when she keeps her promise. We will see this to be a recurrent theme in subsequent results. For a graphical representation of a set $\mathcal{DR}_h$ and the respective optimal resistance strategy see Figure 2.

In light of the general result of Proposition 1, the investigation for the optimal resistance strategy in $\Delta(\mathcal{M})$ becomes almost trivial. In fact, the following holds true.[10]

Remark 2. Assume that $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ and let $\mathcal{R} = \Delta(\mathcal{M})$. Then the (unique) optimal resistance strategy $\hat{r}$ is deterministic and assigns probability 1 to the mechanism $(0, \Delta u_G)$. Moreover, $\hat{\pi}_{\hat{r}}(s_G|G) = \hat{\pi}_{\hat{r}}(s_B|B) = 1$. ◁

Essentially, this observation is based on the fact that the threshold probability in Equation 4, which drives the choice of the Sender, depends on the sum of cost and benefit $\kappa + \beta$. This means that the Receiver can achieve any threshold probability by choosing only among mechanisms without cost. Given that, it is apparent that it is optimal to focus on the one that yields the highest benefit

[9] Given that h is continuous and strictly increasing, with $h(0) = 0$, there always exists a unique $\hat{k} \in (0, \Delta u_G)$ such that $\hat{k} + h(\hat{k}) = \Delta u_G$.

[10] The result is presented without proof, which is available upon request

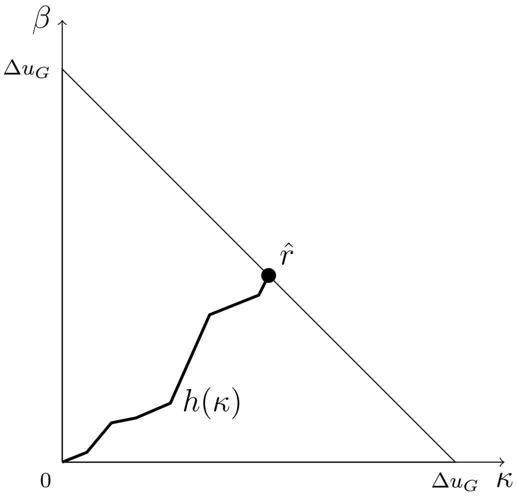

Figure 2: An example of a set of strategies $\mathcal{DR}_h$.

β, as this also leads the Sender to design a fully informative signal. For this reason, for the rest of the paper we focus on trying to detect the optimal resistance strategies in the most empirically relevant subset –that is, among *cost–only resistance strategies.*

4.2. Optimal Cost-Only Resistance Strategies

As argued above, and as is commonly accepted in the literature (see, for instance, Amador and Bagwell, 2016; Ambrus and Egorov, 2017; Austen–Smith and Banks, 2000; Kartik, 2007), the most interesting way of inducing adjustments in incentives is by undertaking own costs. Indeed, reducing one's own payoff is always feasible compared to increasing it. Of course, here we consider action-contingent adjustments which are arguably a more complex version of burning money. Even so, it is true that public commitment and other similar strategies can induce action-contingent costs more easily than action-contingent benefits. For this reason, we exhaustively analyze optimality among cost–only resistance strategies both when any of them is feasible and when choice is limited to some of the most interesting subsets.

We proceed directly to the statement of the main result of our analysis.

Theorem 1 (Optimal cost-only resistance)**.** *Let $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ and $\Delta v_G = \Delta v_B$. Then, the following hold:*

(i) DETERMINISTIC COST–ONLY RESISTANCE: *The Receiver has no incentive to resist persuasion. Formally, if $\mathcal{R} = \mathcal{DCR}$ then $\arg\max_{r\in\mathcal{R}} U(r) = \mathcal{R}$.*

(ii) BINARY STOCHASTIC COST–ONLY RESISTANCE: *There is a unique optimal resistance strategy, which assigns a strictly positive probability to a mechanism that bears a strictly positive cost and if $\Delta u_G \leq \Delta u_B$ it leads to a fully informative signal. Formally, if $\mathcal{R} = \mathcal{BSCR}$ then $\arg\max_{r\in\mathcal{R}} U(r) = \{\hat{r}\}$. Then, if $\Delta u_G > \Delta u_B$, $\hat{r}$ assigns probability $\hat{l} = 1/2$ to the commitment*

mechanism $(-\Delta u_B, 0)$*, whereas if* $\Delta u_G \leq \Delta u_B$*,* $\hat{r}$ *assigns probability* $\hat{l} = \Delta u_G/(\Delta u_G + \Delta u_B)$ *to the commitment mechanism* $(-\Delta u_G, 0)$ *and* $\hat{\pi}_{\hat{r}}(s_G|G) = \hat{\pi}_{\hat{r}}(s_B|B) = 1$*.*

(iii) General stochastic cost–only resistance: *There is a unique optimal resistance strategy, which leads to a fully informative signal. Formally, if* $\mathcal{R} = \mathcal{GSCR}$ *then* $\arg\max_{r\in\mathcal{R}} U(r) = \{\hat{r}\}$*, where* $\hat{r}$ *is identified by the CDF* $\hat{F}_{\hat{r}}(k) = (\Delta u_B + \kappa)/(\Delta u_G + \Delta u_B)$ *for each* $\kappa \in [0, \Delta u_G]$*, and* $\hat{\pi}_{\hat{r}}(s_G|G) = \hat{\pi}_{\hat{r}}(s_B|B) = 1$*.*

Moreover, the Receiver benefits strictly from uncertainty, i.e., $\max_{r\in\mathcal{GSCR}} U(r) > \max_{r\in\mathcal{BSCR}} U(r) > \max_{r\in\mathcal{DCR}} U(r)$*.*

Below we treat each of the cases of our main theorem separately.

4.2.1. Deterministic Cost-Only Resistance

In this case, it is straightforward by Lemma 1 and Proposition 1 that undertaking costs improves the informativeness of the message but has no effect on the Receiver's expected utility. Actually, this result holds irrespective of whether the Sender has state-independent preferences or not.

Indeed, the direct loss in expected utility induced by an increase in the cost is counterbalanced by an indirect positive effect –in particular, an increase in the informativeness of the message.

4.2.2. Binary Stochastic Cost-Only Resistance

The inability of the Receiver to increase her expected utility by committing to any deterministic cost–only resistance strategy puts under question the effectiveness of commitment as a successful mean to resist persuasion. However, this is rushing to a false conclusion.

Continuing to focus on cost–only resistance strategies, we observe that resistance can be beneficial if it makes the Sender uncertain of the exact cost that the Receiver will eventually bear (in case he chooses g). This can happen through stochastic resistance strategies. The important feature of stochastic resistance strategies is that the commitment mechanism – and hence, the Receiver's cost – is realized after the Sender has designed the signal, but before the Receiver chooses her action. As we will see, this is enough to increase the expected utility of the Receiver, compared to the no-commitment case, despite the absence of any explicit benefit.

Note that in this equilibrium, the expected utilities of the Receiver and the Sender are

$$U(\hat{r}) = p_0 u(b, G) + (1 - p_0) u(b, B) + \begin{cases} p_0 \frac{\Delta u_G + \Delta u_B}{4} & \text{if } \Delta u_G > \Delta u_B \\ p_0 \frac{\Delta u_G \Delta u_B}{\Delta u_G + \Delta u_B} & \text{if } \Delta u_G \leq \Delta u_B \end{cases}$$

$$V_{\hat{r}}(\hat{\pi}_{\hat{r}}) = p_0 v(g, G) + (1 - p_0) v(b, B) + \begin{cases} p_0 \Delta v_B \frac{\Delta u_G - \Delta u_B}{2\Delta u_B} & \text{if } \Delta u_G > \Delta u_B \\ 0 & \text{if } \Delta u_G \leq \Delta u_B \end{cases}$$

respectively.

Theorem 1(ii) shows not only that the Receiver can improve her expected utility by using stochastic resistance, but also that, interestingly, even a binary costly strategy is sufficient to make the Sender provide a perfectly informative signal, when this is desirable by the Receiver. This is the case when the Receiver cares more about not "making a mistake" in state B –i.e., the state in which her preferences are essentially misaligned to those of the Sender– than in state G.

The proof of the result is constructive and provides a clear intuition as to why this occurs. The final decision of the Receiver depends on her posterior and the realized commitment mechanism. This generates a dilemma to the Sender, as he has to choose between designing a more informative signal that would be sufficient to persuade the Receiver (persuasion is always conditional on the realization of s_G) for both realizations of the commitment mechanism, or a less informative signal that would persuade the Receiver only when mechanism $(0, 0)$ is realized. Naturally, the Sender prefers to design a more informative signal only when the probability l of mechanism $(-\kappa, 0)$ being realized is sufficiently high. The Receiver is able to anticipate this behavior and choose her resistance strategy accordingly. In fact, it turns out that for any cost κ, choosing a sufficiently high probability to incentivize the design of the more informative signal is always preferred. Intuitively, the optimal probability is the one that is just as high as needed in order to achieve that, as any choice higher than that would induce the same signal by the Sender, while being costly more often.

Therefore, the problem is in essence one of finding the cost κ of the optimal mechanism for the Receiver. This cost is associated with an optimal probability, which in turn determines the optimal signal for the Sender. It then determines the action of the Receiver conditional on the signal and the commitment mechanism realization. This cost κ has two negative effects and one positive. On the one hand, it directly makes it more costly to choose action g when mechanism $(-\kappa, 0)$ is realized. In fact, when κ increases the Receiver must also increase the probability with which mechanism $(-\kappa, 0)$ is realized, so as to ensure that the Sender will choose a sufficiently informative signal to persuade her under both mechanisms. On the other hand, it indirectly increases the Receiver's expected utility when mechanism $(0, 0)$ is realized, by inducing the design of a more informative signal. It turns out that when the Receiver cares predominantly about choosing the correct action in state B, the latter effect always overwhelms the other two, which means that the Receiver prefers to include a more costly mechanism in her strategy in order to incentivize the design of a more informative signal.

4.2.3. General Stochastic Cost-Only Resistance

We have shown that the Receiver can increase her expected utility and induce the design of a fully informative signal via a binary stochastic costly strategy. But, could the Receiver do even better by constructing some more general strategy? The answer is yes. The improvement in the Receiver's expected utility was mainly due to the introduction of uncertainty to the Sender, with respect to the strength of the Receiver's commitment.

Theorem 1(iii) characterizes the optimal resistance strategy when the Receiver is able to choose any general stochastic cost–only resistance strategy. The optimal distribution is continuous. It places a positive mass only at c_0 and the remaining probability is distributed continuously (and in fact uniformly) among the rest of the admissible costs. The expected utility of the Receiver is equal to

$$U(\hat{r}) = p_0 u(b, G) + (1 - p_0) u(b, B) + p_0 \Delta u_G \left(1 + \frac{\Delta u_B}{\Delta u_G + \Delta u_B}\right),$$

Furthermore,

$$V_{\hat{r}}(\hat{\pi}_{\hat{r}}) = p_0 v(g, G) + (1 - p_0) v(b, B)$$

is the Sender's expected utility. Note that the Receiver's expected utility is always strictly higher than the utility under the optimal binary stochastic cost–only resistance strategy. The reason for this is twofold: First, the optimal resistance strategy makes it optimal for the Sender to provide a signal that is sufficiently informative to persuade the Receiver even for the most costly mechanism in the distribution, although this will almost certainly never be realized. Second, in some cases it even allows the Receiver to induce the design of a more informative signal by the Sender, compared to the binary stocahstic case. Actually, the optimal general cost-only resistance strategy always leads to a fully informative signal.

The proof is again constructive and bears similarities to the proof of Theorem 1(ii). Namely, for each mechanism realization, the Receiver's choice is characterized by a cut–off posterior, above which she chooses action g upon observing s_G. Therefore, by choosing a signal, the Sender implicitly chooses the maximum cost at which the Receiver may be persuaded. Thus, the dilemma remains the same. A more accurate signal increases the potential costs for which persuasion is possible but reduces the expected utility gained by a persuaded Receiver. Now, we can characterize the optimal distribution for the Receiver among all distributions that lead to the choice of a signal that induces the same maximum cost for which persuasion is possible, call it $\widetilde{\kappa}$. Essentially, that is a distribution with support $[0, \widetilde{\kappa}]$ (thus persuasion is possible for any mechanism realization) and makes the Sender indifferent between inducing any maximum cost within this range. Overall, this analysis reduces the problem to the Receiver choosing the maximum cost at which he wishes to be potentially persuaded. Like before, this turns out to be the maximum cost $\kappa = \Delta u_G$. Therefore, the optimal distribution is the one that allows for persuasion for any mechanism realization and induces the Sender to design a fully informative signal by making him indifferent between this and any other potentially optimal signal that would induce a different maximum cost for which persuasion would be possible. This distribution is unique.

4.3. Sender with State-Dependent Preferences

The effect of resistance is still prevalent when the Sender has state-dependent preferences. All results of Section 4.1 carry on to this case, meaning that deterministic mechanisms are still beneficial as long as they incur direct benefits and not only costs. With stochastic resistance mechanisms, the Receiver can still increase her expected utility even when restricted to cost-only mechanisms, although it is not guaranteed –but still often possible– that her optimal strategy would induce the design of a fully informative signal by the Sender.

Proposition 3. *Assume that* $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ *and let* $\mathcal{R} = \mathcal{BSCR}$*. Then, there is a unique optimal resistance strategy, which assigns a strictly positive probability to a mechanism that bears a strictly positive cost and if* $\frac{\Delta u_G}{\Delta u_B} \leq \frac{\Delta v_G}{\Delta v_B}$ *it leads to a fully informative signal.*

Proposition 3 generalizes Theorem 1(ii) for general Sender's preferences. Not surprisingly the result depends on the interplay between the relative gains of the two agents from each action. In general, a fully informative signal is again induced when the Receiver really cares about making the correct decision in state B (high Δu_B), yet these gains need to be quite substantial when the Sender has also much to gain by luring her towards his preferred action in that state (high Δv_B).

Proposition 4. *Assume that* $p_0 < \Delta u_B/(\Delta u_G + \Delta u_B)$ *and let* $\mathcal{R} = \mathcal{GSCR}$*. Then, there is a unique optimal resistance strategy, according to which the cost* k *of the resistance mechanism is drawn from a continuous distribution with full support in* $k \in [0, \widetilde{\kappa}]$*, for some* $\widetilde{\kappa} \in (0, \Delta u_G]$*.*

Proposition 4 generalizes Theorem 1(iii) for general Sender's preferences. As expected, the optimal strategy depends also on the preferences of the Sender, yet it is still optimal for the Receiver to resist. In fact, we show that the optimal strategy of the Receiver will still contain costs drawn from a continuous distribution, which may however not include such high costs that could induce the design of a fully informative signal.

5. Conclusion

In this paper, we have shown that a Receiver, in the context of Bayesian Persuasion, is able to resist persuasion by the Sender by using strategies based on public commitment and uncertainty. This form of resistance, albeit plausible and empirically relevant, may not be the only successful strategy. Thus, it would be interesting to consider other types of strategies that can be employed by the Receiver and do not share the same characteristics as the action–contigent payoff adjustments analyzed here.

Overall, the results suggest that the Receiver wants to force the Sender to provide accurate information, in a way that also allows her to capitalize on the benefits from the increased accuracy of the information, and at the expense of the Sender who sees a decrease in his welfare.

A. Proofs

In all proofs, we denote $u_{a,\omega} := u(a, \omega)$ and $v_{a,\omega} := v(a, \omega)$, for all $a \in A$ and $\omega \in \Omega$.

Proof of Lemma 1: The condition $p_{s_G} \geq \widetilde{p}$ is equivalent to $z \leq \frac{p_0(1-\widetilde{p})}{(1-p_0)\widetilde{p}}q$. Moreover, the expected utility of Sender from selecting a signal (q, z) is as follows:

$$\begin{aligned} V_r(q,z) &= p_0\left[qv_{g,G} + (1-q)v_{b,G}\right] + (1-p_0)\left[zv_{g,B} + (1-z)v_{b,B}\right] = \\ &= p_0 v_{b,G} + (1-p_0)v_{b,B} + qp_0\Delta v_G + z(1-p_0)\Delta v_B \end{aligned} \tag{A.1}$$

$V_r(q, z)$ increases in both q and z, as long as the abovementioned condition holds, which implies that the optimal signal should satisfy $\hat{q}_r = 1$ and $\hat{z}_r = \frac{p_0(1-\widetilde{p})}{(1-p_0)\widetilde{p}}$, for which the Receiver chooses action g when observing s_G and action b otherwise. Substituting this into V_r, we directly obtain:

$$V_r(\hat{q}_r, \hat{z}_r) = p_0 v_{g,G} + (1-p_0)v_{b,B} + p_0\left(\frac{1}{\widetilde{p}} - 1\right)\Delta v_B$$

Analogously, the Receiver's ex–ante expected utility, anticipating that the Sender will choose optimally, is as follows:

$$\begin{aligned} U(r) &= p_0\left[\hat{q}_r(u_{g,G} - \kappa) + (1-\hat{q}_r)(u_{b,G} + \beta)\right] + (1-p_0)\left[\hat{z}_r(u_{g,B} - \kappa) + (1-\hat{z}_r)(u_{b,B} + \beta)\right] \\ &= p_0(u_{g,G} - \kappa) + (1-p_0)(u_{b,B} + \beta) - (1-p_0)\hat{z}_r(\Delta u_B + \kappa + \beta) \\ &= p_0(u_{g,G} - \kappa) + (1-p_0)(u_{b,B} + \beta) - \frac{p_0(1-\widetilde{p})}{\widetilde{p}}(\Delta u_B + \kappa + \beta) \\ &= p_0(u_{g,G} - \kappa) + (1-p_0)(u_{b,B} + \beta) + p_0(\Delta u_B + \kappa + \beta) - \frac{p_0}{\widetilde{p}}(\Delta u_B + \kappa + \beta) \\ &= p_0(u_{g,G} - \kappa) + (1-p_0)(u_{b,B} + \beta) + p_0(\Delta u_B + \kappa + \beta) - p_0(\Delta u_G + \Delta u_B) \\ &= \beta + p_0 u_{b,G} + (1-p_0)u_{b,B} \end{aligned}$$

□

Proof of Theorem 1(ii). We begin the proof allowing for general preferences of the Sender and subsequently restrict to the case of state-independent preferences.

Consider a strategy $r \in \mathcal{BSCR}$ that assigns probability l to mechanism $(-\kappa, 0)$ and probability $1 - l$ to $(0, 0)$. By Equation (4), the Receiver will choose action g upon observing commitment mechanism $(0, 0)$ if her posterior satisfies $p \geq \frac{\Delta u_B}{\Delta u_G + \Delta u_B} = \widetilde{p}(0)$ and upon observing $(-\kappa, 0)$ if her posterior satisfies $p \geq \frac{\Delta u_B + \kappa}{\Delta u_G + \Delta u_B} = \widetilde{p}(\kappa)$, where $\widetilde{p}(\kappa) > \widetilde{p}(0)$ for all $\kappa > 0$.

The Sender designs a signal (q, z) which yields two possible posteriors p_{s_G} and p_{s_B} as in Equation 2 and, because of the low prior (see Remark 1), he may persuade the Receiver to choose action g for at most one signal realization, say s_G. Given that $\widetilde{p}(\kappa) > \widetilde{p}(0)$, the Receiver, upon observing s_G, is persuaded for both commitment mechanism realizations if $p_{s_G} \geq \widetilde{p}(\kappa)$ and only for realization $(0, 0)$ if $\widetilde{p}(\kappa) > p_{s_G} \geq \widetilde{p}(0)$. Equivalently, the Receiver is persuaded for both mechanism realizations if $z \leq \frac{p_0}{1-p_0}\frac{\Delta u_G - \kappa}{\Delta u_B + \kappa}q = \widetilde{z}(\kappa)$ and is persuaded only for realization $(0, 0)$ if $\widetilde{z}(\kappa) < z \leq \frac{p_0}{1-p_0}\frac{\Delta u_G}{\Delta u_B}q = \widetilde{z}(0)$.

Therefore, the Sender's expected utility from a signal (q, z), given resistance strategy r, is equal to:

$$V_r(q,z) = \begin{cases} p_0 v_{b,G} + (1-p_0)v_{b,B} + qp_0\Delta v_G + z(1-p_0)\Delta v_B & \text{if } z \leq \widetilde{z}(\kappa) \\ l[p_0 v_{b,G} + (1-p_0)v_{b,B}] + (1-l)[p_0 v_{b,G} + (1-p_0)v_{b,B} + qp_0\Delta v_G + z(1-p_0)\Delta v_B] & \text{if } \widetilde{z}(\kappa) < z \leq \widetilde{z}(0) \\ p_0 v_{b,G} + (1-p_0)v_{b,B} & \text{if } z > \widetilde{z}(0) \end{cases}$$

Observe that, similarly to the deterministic case, irrespective of the value of z, it is always optimal to choose $\hat{q}_r = 1$. Moreover, persuasion is always beneficial for the Sender, given that $qp_0\Delta v_G + z(1-p_0)\Delta v_B > 0$, therefore it is never optimal to choose $z > \widetilde{z}(0)$. Furthermore, it is never optimal either to choose $z < \widetilde{z}(\kappa)$, as it always yields lower expected utility compared to $z = \widetilde{z}(\kappa)$. Analogously, it is never optimal to choose $\widetilde{z}(\kappa) < z < \widetilde{z}(0)$ because it always yields lower expected utility than $z = \widetilde{z}(0)$. Overall, this leaves two potential optimal choices for the Sender, either $z = \widetilde{z}(0)$ or $z = \widetilde{z}(\kappa)$. After some calculations we get that:

$$\begin{aligned} V_r[1, \widetilde{z}(0)] &= p_0 v_{g,G} + (1-p_0)v_{b,B} + p_0\left[(1-l)\frac{\Delta u_G}{\Delta u_B}\Delta v_B - l\Delta v_G\right] \\ V_r[1, \widetilde{z}(\kappa)] &= p_0 v_{g,G} + (1-p_0)v_{b,B} + p_0\frac{\Delta u_G - \kappa}{\Delta u_B + \kappa}\Delta v_B, \end{aligned}$$

thus directly implying

$$V_r[1, \widetilde{z}(0)] > V_r[1, \widetilde{z}(\kappa)] \;\Leftrightarrow\; l\left[\frac{\Delta u_G}{\Delta u_B}\Delta v_B + \Delta v_G\right] < \left[\frac{\Delta u_G}{\Delta u_B} - \frac{\Delta u_G - \kappa}{\Delta u_B + \kappa}\right]\Delta v_B \tag{A.2}$$

Hence, the Sender chooses $(1, \widetilde{z}(0))$ as long as l is sufficiently small and chooses $(1, \widetilde{z}(\kappa))$ otherwise.

Given that the optimal choice of the Sender depends only on l and κ, the Receiver can anticipate the signal that the Sender will choose for each commitment strategy. Recall, that the Sender, when indifferent, is assumed to choose the most preferred signal for the Receiver.

The Receiver's expected value when choosing a strategy that assigns probability l to a mechanism $(-\kappa, 0)$ such that the Sender will then choose the signal $(1, \widetilde{z}(0))$, denoted by $U_0(l, \kappa)$, is as follows:

$$\begin{aligned} U_0(l,\kappa) &= l\left[p_0 u_{b,G} + (1-p_0)u_{b,B}\right] + (1-l)\left\{p_0 u_{g,G} + (1-p_0)\widetilde{z}(0)u_{g,B} + (1-p_0)\left[1-\widetilde{z}(0)\right]u_{b,B}\right\} \\ &= l\left[p_0 u_{b,G} + (1-p_0)u_{b,B}\right] + (1-l)\left[p_0 u_{g,G} + p_0\frac{\Delta u_G}{\Delta u_B}u_{g,B} + (1-p_0)\left(1 - \frac{p_0}{1-p_0}\frac{\Delta u_G}{\Delta u_B}\right)u_{b,B}\right] \\ &= l\left[p_0 u_{b,G} + (1-p_0)u_{b,B}\right] + (1-l)\left[p_0 u_{g,G} + (1-p_0)u_{b,B} - p_0\Delta u_G\right] \\ &= p_0 u_{b,G} + (1-p_0)u_{b,B} \end{aligned} \tag{A.3}$$

This result is not surprising given that $\widetilde{z}(0)$ is designed so as to make the Receiver exactly indifferent between being persuaded by signal realization s_G to choose action g and choosing always action b.

On the other hand, when the Receiver chooses a strategy that assigns probability l to mechanism $(-\kappa, 0)$ such that the Sender will then choose the signal $(1, \widetilde{z}(\kappa))$, then the expected utility she gets, denoted by $U_k(l,\kappa)$, is as follows:

$$\begin{aligned}
U_k(l,\kappa) &= p_0 u_{g,G} + (1-p_0)\widetilde{z}(\kappa)u_{g,B} + (1-p_0)\left[1-\widetilde{z}(\kappa)\right]u_{b,B} - l\kappa\left[p_0 + (1-p_0)\widetilde{z}(\kappa)\right] \\
&= p_0 u_{g,G} + (1-p_0)u_{b,B} - p_0\frac{\Delta u_G - \kappa}{\Delta u_B + \kappa}\Delta u_B - l\kappa p_0 - l\kappa p_0\frac{\Delta u_G - \kappa}{\Delta u_B + \kappa} \\
&= p_0 u_{b,G} + (1-p_0)u_{b,B} + p_0(1-l)(\Delta u_G + \Delta u_B)\frac{\kappa}{\Delta u_B + \kappa} \qquad \text{(A.4)}
\end{aligned}$$

Equations (A.3) and (A.4) suggest that the Receiver prefers for all κ to choose l sufficiently high to ensure that the Sender chooses signal $(1,\widetilde{z}(\kappa))$. Moreover, again for each κ, among all l that ensure such a choice, the Receiver chooses the smallest one, which is the one that satisfies $V_r[1,\widetilde{z}(0)] = V_r[1,\widetilde{z}(\kappa)]$, denoted as $\widetilde{l}(\kappa)$.[11] After some calculations, this takes the following form:

$$\widetilde{l}(\kappa) = \frac{(\Delta u_G + \Delta u_B)\Delta v_B}{\Delta u_G \Delta v_B + \Delta v_G \Delta u_B}\cdot\frac{\kappa}{\Delta u_B + \kappa} \qquad \text{(A.5)}$$

When the Sender has state-independent preferences, i.e. $\Delta v_G = \Delta v_B$, we get that $\widetilde{l}(\kappa) = \frac{\kappa}{\Delta u_B + \kappa}$. If we plug $\widetilde{l}(\kappa)$ in U_κ and differentiating with respect to κ, we get that $U_\kappa(\widetilde{l}(\kappa),\kappa)$ attains a unique maximum for $\kappa \in [0,\Delta u_G]$ whose value depends on the relative size of Δu_B and Δu_G. Namely,

$$\begin{aligned}
&(\Delta u_G > \Delta u_B)\colon\ k^* = \Delta u_B\ ,\ \widetilde{l}(k^*) = 1/2\ ,\ \widetilde{z}(k^*) = \frac{p_0}{2(1-p_0)}\left(\frac{\Delta u_G - \Delta u_B}{\Delta u_B}\right) \\
&(\Delta u_G \leq \Delta u_B)\colon\ k^* = \Delta u_G\ ,\ \widetilde{l}(k^*) = \frac{\Delta u_G}{\Delta u_G + \Delta u_B}\ ,\ \widetilde{z}(k^*) = 0
\end{aligned}$$

□

Proof of Theorem 1(iii): We begin the proof allowing for general preferences of the Sender and subsequently restrict to the case of state-independent preferences.

By Equation 4, for any $\kappa \in [0,\Delta u_G]$ that is drawn the Receiver chooses action g if her posterior satisfies $p \geq \widetilde{p}(\kappa) = \frac{\kappa + \Delta u_B}{\Delta u_G + \Delta u_B}$. According to this and given that the prior is low, the Sender chooses a signal (q,z) that persuades the Receiver in one of the two signal realizations, say s_G. Following the same reasoning as in the proof of Theorem 1(ii), potential optimal signals are those that satisfy $\widetilde{q} = 1$ and $\widetilde{z}(\kappa) = \frac{p_0}{1-p_0}\frac{\Delta u_G - \kappa}{\Delta u_B + \kappa}$ for some $\kappa \in [0,\Delta u_G]$, which corresponds to the maximum commitment cost for which the Receiver is persuaded by signal realization s_G. Hence, the problem of the Sender is equivalent to choosing a threshold value $\widetilde{\kappa}$ above which persuasion does not take place. The threshold value that maximizes his expected utility is denoted by $\widetilde{\kappa}^*$. For some $\widetilde{\kappa} \in [0,\Delta u_G]$ and given the distribution F associated to the strategy r of the Receiver, the expected utility of the Sender can be

[11] Here is where the assumption that the Sender, when indifferent, chooses the Receiver's preferred choice plays an important role, because it guarantees the existence of an optimal resistance strategy.

rewritten as a function of $\widetilde{\kappa}$ as follows:

$$\begin{aligned} V_r(\widetilde{\kappa}) &= [1-F(\widetilde{\kappa})]\cdot[p_0 v_{b,G}+(1-p_0)v_{b,B}]+F(\widetilde{\kappa})\big(p_0 v_{g,G}+(1-p_0)\widetilde{z}(\widetilde{\kappa})v_{g,B}+(1-p_0)[1-\widetilde{z}(\widetilde{\kappa})]v_{b,B}\big) \\ &= p_0 v_{b,G}+(1-p_0)v_{b,B}+p_0F(\widetilde{\kappa})\left(\Delta v_G+\frac{\Delta u_G-\widetilde{\kappa}}{\Delta u_B+\widetilde{\kappa}}\Delta v_B\right) \end{aligned} \tag{A.6}$$

Assuming that the Sender, when indifferent, chooses the Receiver's preferred signal, then any choice of distribution F by the Receiver induces some $\widetilde{\kappa}$ to be chosen by the Sender. Thus, let $\mathcal{F}_{\widetilde{\kappa}}$ be the set of all available distributions that induce $\widetilde{\kappa}$.[12] The expected utility of the Receiver when choosing a strategy $r\in\mathcal{GSCR}$ associated to a distribution $F\in\mathcal{F}_{\widetilde{\kappa}}$ is as follows:

$$\begin{aligned} U(r) =&[1-F(\widetilde{\kappa})]\cdot[p_0u_{b,G}+(1-p_0)u_{b,B}] \\ &+F(\widetilde{\kappa})\left\{p_0u_{g,G}+(1-p_0)\widetilde{z}(\widetilde{\kappa})u_{g,B}+(1-p_0)[1-\widetilde{z}(\widetilde{\kappa})]u_{b,B}\right\} \\ &-[p_0+(1-p_0)\widetilde{z}(\widetilde{\kappa})]\int_{[0,\widetilde{\kappa}]}\kappa dF(\kappa) \end{aligned} \tag{A.7}$$

where the (Lebesque) integral essentially refers to the average cost for the Receiver in the region where persuasion is possible.

Our next step is to find the optimal distribution within each set $\mathcal{F}_{\widetilde{\kappa}}$, recalling that also the Receiver, when indifferent, chooses the most preferred resistance strategy to the Sender. Recall also that a signal chosen by the Sender that induces $\widetilde{\kappa}$, will be structured such that the Receiver will be indifferent between "always choosing action b" and "choosing action g when observing s_G". Therefore, the Receiver is indifferent between two distributions that distribute probability identically up to $\widetilde{\kappa}$ and one of them assigns positive mass to values $\kappa\in(\widetilde{\kappa},\Delta u_G]$ while the other one puts the same mass exactly on $\widetilde{\kappa}$. Yet, the latter distribution is preferred by the Sender, because it increases the probability with which the Receiver will get persuaded, without affecting his optimal choice.[13]

Hence, all potentially optimal distributions that induce $\widetilde{\kappa}$ satisfy $F(\widetilde{\kappa})=1$. Moreover, by definition each of these distributions should satisfy the following condition for all $\kappa\in[0,\widetilde{\kappa})$:

$$V_r(\widetilde{\kappa})\geq V_r(\kappa) \quad\Leftrightarrow\quad \Delta v_G+\frac{\Delta u_G-\widetilde{\kappa}}{\Delta u_B+\widetilde{\kappa}}\Delta v_B\geq F(\kappa)\left(\Delta v_G+\frac{\Delta u_G-\kappa}{\Delta u_B+\kappa}\Delta v_B\right) \tag{A.8}$$

This is because, we have considered the distributions for which it is optimal for the Sender to induce $\widetilde{\kappa}$ as the maximum cost for which the Receiver can be persuaded. In fact, the equivalence relation

[12] This set is always non–empty because it always contains the trivial distribution in which the Receiver puts probability one to the mechanism $(-\widetilde{\kappa},0)$. If a distribution F induces several $\widetilde{\kappa}$, then is included in all relevant sets $\mathcal{F}_{\widetilde{\kappa}}$.

[13] On the one hand, the Sender would not choose a signal that would induce $\widetilde{\kappa}'>\widetilde{\kappa}$, because that would require the signal to be more informative, without increasing the probability of persuasion (as the Receiver is always persuaded by s_G). On the other hand, the Sender would not choose a signal that would induce some $\widetilde{\kappa}'<\widetilde{\kappa}$, because if $\widetilde{\kappa}'$ is optimal now, then it should have also been optimal for the initial distribution, which cannot happen since $\widetilde{\kappa}$ was by definition the induced value that maximizes the expected utility of the Sender for the chosen distribution. Therefore, the new distribution does not alter the subsequent signal choice of the Sender.

guarantees that this inequality characterizes the set of all distributions in $\mathcal{F}_{\widetilde{\kappa}}$.

Among all the distributions satisfying expression A.8, the Receiver prefers the one with the minimum expected value (as this enters negatively in her expected utility, in expression A.7). It is straightforward to see that there is a unique such distribution, which is the one that satisfies expression A.8 with equality for all $\kappa \in [0, \widetilde{\kappa}]$ and is denoted by $\widetilde{F}_{\widetilde{\kappa}}$, i.e.

$$\widetilde{F}_{\widetilde{\kappa}}(\kappa) = \begin{cases} \left(\Delta v_G + \frac{\Delta u_G - \widetilde{\kappa}}{\Delta u_B + \widetilde{\kappa}} \Delta v_B\right) \cdot \frac{1}{\Delta v_G + \frac{\Delta u_G - \kappa}{\Delta u_B + \kappa} \Delta v_B} & \text{if } \kappa \in [0, \widetilde{\kappa}) \\ 1 & \text{if } \kappa \in [\widetilde{\kappa}, 1] \end{cases} \tag{A.9}$$

This distribution would make the Sender indifferent between signals that induce any $\kappa \in [0, \widetilde{\kappa}]$. Yet, given that, when indifferent he chooses the most preferred to the Receiver, his choice will be $\widetilde{\kappa}$.

It is important to notice that this distribution is differentiable in $[0, \Delta u_G]$ and puts positive mass only at $\kappa = 0$. Given these observations, we know that the distribution also has an associated well–defined continuous probability distribution function $\widetilde{f}_{\widetilde{\kappa}}$ for every $\kappa \in (0, \widetilde{\kappa})$.

Therefore, we have shown that the Receiver can induce any $\widetilde{\kappa} \in [0, \Delta u_G]$ and we have found the optimal distribution for achieving so. Hence, the problem is summarized in finding the value of $\widetilde{\kappa}$ that would maximize the expected utility of the Receiver, if it is induced. Given our previous findings, we can rewrite the expected utility of the Receiver as a function of $\widetilde{\kappa}$ as follows:

$$\begin{aligned} U(\widetilde{\kappa}) &= p_0 u_{g,G} + (1-p_0)\widetilde{z}(\widetilde{\kappa}) u_{g,B} + (1-p_0)[1 - \widetilde{z}(\widetilde{\kappa})] u_{b,B} - [p_0 + (1-p_0)\widetilde{z}(\widetilde{\kappa})] \int_0^{\widetilde{\kappa}} \kappa \widetilde{f}_{\widetilde{\kappa}}(\kappa) d\kappa \\ &= p_0 u_{g,G} + (1-p_0) u_{b,B} - p_0 \frac{\Delta u_G - \widetilde{\kappa}}{\Delta u_B + \widetilde{\kappa}} \Delta u_B - p_0 \frac{\Delta u_G + \Delta u_B}{\Delta u_B + \widetilde{\kappa}} \int_0^{\widetilde{\kappa}} \kappa \widetilde{f}_{\widetilde{\kappa}}(\kappa) d\kappa \end{aligned} \tag{A.10}$$

When the Sender has state-independent preferences, i.e. $\Delta v_G = \Delta v_B$, then $\widetilde{F}_{\widetilde{\kappa}}(\kappa) = \frac{\Delta u_B + \kappa}{\Delta u_B + \widetilde{\kappa}}$ for $\kappa \in [0, \widetilde{\kappa}]$, which means that the optimal distribution puts a positive mass $\frac{\Delta u_B}{\Delta u_B + \widetilde{\kappa}}$ at 0 and is uniform in $(0, \widetilde{\kappa})$. Therefore, the expected utility of the Receiver is the following:

$$U(\widetilde{\kappa}) = p_0 u_{g,G} + (1-p_0) u_{b,B} - p_0 \left[\frac{\Delta u_G - \widetilde{\kappa}}{\Delta u_B + \widetilde{\kappa}} \Delta u_B + \frac{\Delta u_G + \Delta u_B}{2(\Delta u_B + \widetilde{\kappa})^2} \widetilde{\kappa}^2 \right] \Rightarrow U'(\widetilde{\kappa}) \quad = p_0 \Delta u_B^2 \frac{\Delta u_B + \Delta u_G}{(\Delta u_B + \widetilde{\kappa})^3} > 0$$

The expected utility is strictly increasing in $[0, \Delta u_G]$. Hence, the Receiver wants to induce $\widetilde{\kappa} = \Delta u_G$, which means that she chooses distribution $\widetilde{F}_{\Delta u_G}$. Therefore, the optimal strategy $\hat{r}$ among all $r \in \mathcal{GSCR}$ is characterized by the cumulative distribution function

$$\hat{F}_{\hat{r}}(\kappa) = \frac{\Delta u_B + \kappa}{\Delta u_B + \Delta u_G}, \quad \text{for } \kappa \in [0, \Delta u_G]$$

For this resistance strategy the Sender designs a fully informative signal, i.e. $\hat{q}_{\hat{r}} = 1$ and $\hat{z}_{\hat{r}} = 0$. □

Proof of Proposition 3: The proof is identical to that of Theorem 1(ii) up to the expression (A.5).

Starting from this point, after plugging $\widetilde{l}(\kappa)$ in U_κ and differentiating with respect to κ, we get that $U_\kappa(\widetilde{l}(\kappa), \kappa)$ attains a unique maximum for $\kappa \in [0, \Delta u_G]$ whose value depends on the relative size of the ratios $\Delta v_G/\Delta v_B$ and $\Delta u_G/\Delta u_B$. Namely,

$$\left(\frac{\Delta u_G}{\Delta u_B} > \frac{\Delta v_G}{\Delta v_B}\right):\ k^* = \frac{\Delta u_B(\Delta u_G \Delta v_B + \Delta u_B \Delta v_G)}{\Delta v_B(2\Delta u_B + \Delta u_G) - \Delta v_G \Delta u_B}\ ,\ \widetilde{l}(k^*) = 1/2\ ,\ \widetilde{z}(k^*) = \frac{p_0}{2(1-p_0)}\left(\frac{\Delta u_G}{\Delta u_B} - \frac{\Delta v_G}{\Delta v_B}\right)$$

$$\left(\frac{\Delta u_G}{\Delta u_B} \leq \frac{\Delta v_G}{\Delta v_B}\right):\ k^* = \Delta u_G\ ,\ \widetilde{l}(k^*) = \frac{\Delta u_G \Delta v_B}{\Delta u_G \Delta v_B + \Delta u_B \Delta v_G}\ ,\ \widetilde{z}(k^*) = 0$$

□

Proof of Proposition 4: The proof is identical to the one of Theorem 1(iii) up to expression A.10. Starting from this expression, we show that even if the Sender's preferences are state dependent, it is still always optimal for the Receiver to resist to persuasion, i.e. to induce $\widetilde{\kappa} > 0$, although inducing a fully informative signal might not be optimal for the Receiver, i.e. the Receiver might prefer to induce $\widetilde{\kappa} < \Delta u_G$. First, note that $U(\widetilde{\kappa})$ is a differentiable function of $\widetilde{\kappa} \in [0, \Delta u_G]$. Thus, it admits a global maximum in this interval. In principle, there might be several $\widetilde{\kappa}$ at which the global maximum is achieved. Given that when indifferent the Receiver chooses the most preferred choice for the Sender, who prefers a $\widetilde{\kappa}$ as small as possible, and that by differentiability of $U(\widetilde{\kappa})$ we can find the minimum of $\text{argmax}_{\widetilde{\kappa}\in[0,\Delta u_G]} U(\widetilde{\kappa})$ exists, there is a unique optimal $\widetilde{\kappa}^*$ for the Receiver. Moreover, if we calculate the derivative of the expected utility and evaluate its limit at zero, we get that: $\lim\limits_{\widetilde{\kappa}\to 0} U'(\widetilde{\kappa}) = p_0 \frac{\Delta u_B + \Delta u_G}{\Delta u_B} > 0$. Thus, $\widetilde{\kappa}^*$>0, which means that resisting persuasion always yields higher profits to the Receiver. □

References

Alonso, R. & Câmara, O. (2016). Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory* 165, 672–706.

——— (2017). Persuading voters. *American Economic Review* 106, 35-90–3605.

Amador, M. & Bagwell, K. (2016). Money Burning in the Theory of Delegation. *mimeo.*

Ambrus, A. & Egorov, G. (2017). Delegation and nonmonetary incentives. *Journal of Economic Theory* 171, 101–135.

Austen–Smith, D. & Banks, J.S. (2000). Cheap Talk and Burned Money. *Journal of Economic Theory* 91(1), 1–16.

Batra, R., Homer, P.M. & Kahle, L.R. (2001). Values, Susceptibility to Normative Influence, and Attribute Importance Weights: A Nomological Analysis. *Journal of Consumer Psychology* 11(2), 115–128.

Ben-Porath, E. & Dekel, E. (1992). Signaling future actions and potential for sacrifice. *Journal of Economic Theory* 57, 36–51.

Bergemann, D., Brooks, B. & Morris, S. (2015). The Limits of Price Discrimination. *American Economic Review* 105 (3), 921–57.

Bertrand, M., Karlan, D.S., Mullainathan, S., Shafir, E. & Zinman, J. (2010). What's advertising content worth? Evidence fromn a consumer credit marketing field experiment. *Quarterly Journal of Economics* 125, 263–305.

Brehm, J.W. (1966). A theory of psychological reactance. New York: Academic Press.

Cialdini, R.B., Cacioppo, J.T., Bassett, R. & Miller, J.A. (1978). Low-ball procedure for producing compliance: Commitment then cost. *Journal of Personality and Social Psychology*, 36(5), 463–476.

Crawford, V. & Sobel, J. (1982). Strategic information transmission. *Econometrica* 50(6), 1431–1451.

DellaVigna, S. & Gentzkow, M. (2010). Persuasion: Empirical Evidence. *Annual Review of Economics* 2, 643–669.

DellaVigna, S. & Kaplan, E. (2007). The Fox News effect: media bias and voting. *Quarterly Journal of Economics* 122(3), 1187–1234.

Fearon, J. (1994) Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88(3), 577–592.

Fearon, J. (1997) Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *The Journal of Conflict Resolution* 41(1), 68–90.

Fransen, M.L., Smit, E.G. & Verlegh, P.W.J. (2015). Strategies and motives for resistance to persuasion: an integrative framework. *Frontiers in Psychology* 6:1201.

Gentzkow, M. & Kamenica, E. (2017a). Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior* 111, 411–429.

Gentzkow, M. & Kamenica, E. (2017b). Competition in persuasion. *Review of Economic Studies* 84, 300-322.

Glazer, J. & Rubinstein, A. (2006). A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics* 1, 395–410.

Gopinath, M. & Nyer, N.U. (2009). The effect of public commitment on resistance to persuasion: The influence of attitude certainty, issue importance, susceptibility to normative influence, preference for consistency and source proximity. *International Journal of Research in Marketing* 26, 60–68.

Hagmann, D. & Loewenstein, G. (2017). Persuasion with Motivated Beliefs. working paper.

Hedlund, J. (2017). Bayesian persuasion by a privately informed sender. *Journal of Economic Theory* 167, 229–268.

Iida, K. (1993) When and How Do Domestic Constraints Matter? Two–Level Games with Uncertainty. *The Journal of Conflict Resolution* 37(3), 403–426.

Jacks, J.Z. & Cameron, K.A. (2003) Strategies for resisting persuasion. *Basic and Applied Social Psychology*, 25(2), 145–161.

Jellison, M.J. & Mills, J. (1969) Effect of public commitment upon opinions. *Journal of Experimental Social Psychology* 5, 340–346.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics* (forthcoming).

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review* 101, 2590–2615.

Kartik, N. (2007). A note on cheap talk and burned money. *Journal of Economic Theory* 136(1), 749–758.

Knowles, E.S. & Linn, J.A. (eds.) (2004). Resistance and persuasion. Psychology Press. New Jersey.

Kolotilin, A. (2015). Experimental design to persuade. *Games and Economic Behavior* 90, 215-226.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A. & Li, M. (2017). Persuasion of a privately informed receiver. *Econometrica* 85, 1949-1964.

Lehmann, E., Simula, L., & Trannoy, A. (2014). Tax me if you can! Optimal nonlinear income tax between competing governments. *The Quarterly Journal of Economics* 129(4), 1995-2030.

Leventoglu, B. & Tarar, A. (2005) Prenegotiation Public Commitment in Domestic and International Bargaining. *American Political Science Review* 99(3), 419–433.

Lokhorst, A.M., Van Dijk, E. & Staats, H. (2009) Public commitment making as a structural solution in social dilemmas. *Journal of Environmental Psychology* 29, 400–406.

Mo, J. (1994) The Logic of Two–Level Games with Endogenous Domestic Coalitions. *The Journal of Conflict Resolution* 38(3), 402–422.

Mo, J. (1995) Domestic Institutions and International Bargaining: The Role of Agent Veto in Two-LevelGames. *American Political Science Review* 89(4), 914–924.

Myerson, R. B. (1993). Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review* 87(4), 856-869.

Perez-Richet, E. (2014). Interim Bayesian persuasion: First steps. *American Economic Review, Papers and Proceedings* 104, 469–474.

Perloff, R.M. (2017). The dynamics of persuasion: communication and attitudes in the 21st Century, Sixth Edition. Routledge, New York.

Petty, R.E. & Cacioppo, J.T. (1986). Communication and persuasion: central and peripheral routes to attitude change. Springer, New York.

Putnam, R.D. (1988). Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization* 42(3), 427–460.

Ringold, D.J. (2002). Boomerang Effects in Response to Public Health Interventions: Some Unintended Consequences in the Alcoholic Beverage Market. *Journal of Consumer Policy* 25(1), 27–63.

Roesler, A. K. & Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review* 107(7), 2072-80.

Stigler, G.J. (1961). The Economics of information. *Journal of Political Economy* 69, 213–225.

Tarar, A. & Leventoglu, B. (2009) Public Commitment in crisis bargaining. *International Studies Quarterly* 53, 817–839.

Tomz, M. (2007) Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization* 61, 821–840.

Tormala, Z.L. & Petty, R.E. (2002) What Doesn't Kill Me Makes Me Stronger: The Effects of Resisting Persuasion on Attitude Certainty. *Journal of Personality and Social Psychology* 83(6), 1298–1313.

Tormala, Z.L. & Petty, R.E. (2004) Source Credibility and Attitude Certainty: A Metacognitive Analysis of Resistance to Persuasion. *Journal of Consumer Psychology* 14(4), 427–442.

Tsakas, E. & Tsakas, N. (2017). Noisy persuasion. *Working paper.*

van Damme, E. (1989). Stable equilibria and forward induction. *Journal of Economic Theory* 48, 476–496.

Wells, R.E. & Iyengar, S.S. (2005) Positive illusions of preference consistency: When remaining eluded by one's preferences yields greater subjective well-being and decision outcomes. *Organizational Behavior and Human Decision Processes* 98, 66–87.