Persuading a Wishful Thinker*

Victor Augias† Daniel M. A. Barreto‡

February 15, 2022

Abstract

We analyze a model of persuasion in which Receiver forms wishful non-Bayesian beliefs. The effectiveness of persuasion depends on Receiver's material stakes: it is more effective when intended to encourage risky behavior that potentially lead to a high payoff and less effective when intended to encourage more cautious behavior. We illustrate this insight with applications showing why informational interventions are often ineffective in inducing greater investment in preventive health treatments, how financial advisors might take advantage of their clients overoptimistic beliefs and why strategic information disclosure to voters with different partisan preferences can lead to belief polarization in an electorate.

JEL classification codes: D82; D83; D91.

Keywords: non-Bayesian persuasion; motivated thinking; overoptimism; optimal beliefs.

*This paper formerly circulated under the title "Wishful Thinking: Persuasion and Polarization." We thank Jeanne Hagenbach and Eduardo Perez-Richet for their support. We also thank S. Nageeb Ali, Roland Bénabou, Michele Fioretti, Alexis Ghersengorin, Simon Gleyze, Emeric Henry, Deniz Kattwinkel, Frédéric Koessler, Laurent Mathevet, Meg Meyer, Daniel Monte, Nikhil Vellodi, Adrien Vigier and Yves Le Yaouanq for their valuable feedbacks and comments as well as seminar audiences at Sciences Po, Paris School of Economics, São Paulo School of Economics (FGV) and at the Econometric Society European Meeting 2021. All remaining errors are ours.

†Sciences Po and CNRS, e-mail: victor.augias@sciencespo.fr. Victor Augias thanks the European Research Council (grant 850996 – MOREV) for financial support.

‡Sciences Po, e-mail: daniel.barreto@sciencespo.fr.

1 Introduction

It is generally assumed in models of strategic communication that receivers update beliefs in a perfectly rational manner, as would a Bayesian statistician. Yet, a substantial literature in psychology and behavioral economics shows that the process by which individuals interpret information and form beliefs is not guided solely by a desire for accuracy but often depends on their motivations and material incentives. This phenomenon is generally referred to as *motivated inference* (Kunda, 1987, 1990), and a common manifestation of it is *wishful thinking*: the tendency of individuals to let their *preferences about outcomes* influence the way they process information, leading to beliefs that are systematically biased towards outcomes they wish to be true.[1] In this paper we investigate how wishful thinking affects the effectiveness of persuasion, i.e., the probability or frequency with which a sender is able to induce a receiver to take her preferred action.

Following Caplin and Leahy (2019), we propose a model in which the receiver's belief updating rule is non-bayesian: after observing an informative signal, Receiver forms beliefs by trading off their anticipatory value against the psychological cost of distorting beliefs away from Bayesian ones. As a result, Receiver's beliefs are stakes-dependent, i.e., they depend on his preferences, and overweight the state associated with the highest payoff, giving rise to overoptimism.

Distortions in beliefs lead to distortions in Receiver's behavior: some actions end up being favored, meaning that they are taken more often (i.e., after the reception of a strictly greater set of possible signals) relative to a Bayesian decision-maker. When he only has two available actions, wishful thinking leads Receiver to favor the action associated with the highest payoff and the highest payoff variability. If one of the two actions induces the highest possible payoff and the other induces the highest payoff variability, then which of the two is favored depends on the magnitude of Receiver's belief distortion cost. As such, the effectiveness of information provision as a tool to incentivize agents might vary with individuals' material stakes: *persuasion is more effective when it is aimed at encouraging behavior that is risky but can potentially*

[1] There exists abundant experimental evidence of wishful thinking. See in particular Bénabou and Tirole (2016), page 150 and Benjamin (2019) Section 9, as well as, e.g., Weinstein (1980), Mijović-Prelec and Prelec (2010), Mayraz (2011), Heger and Papageorge (2018), Coutts (2019), Engelmann et al. (2019) or Jiao (2020).

yield very high returns and less effective when it is aimed at encouraging more cautious behavior. We illustrate this insight in applications in which wishful beliefs can play an important role.

Application 1: Information Provision and Preventive Health Care. In this application a public health agency designs an information policy about the risk of infection of an illness in order to promote a preventive treatment that can be adopted by individuals at some cost. Since not adopting the treatment is the action that can potentially yield the highest payoff (in case the illness is not severe) and also the action with the highest payoff variability, it is favored by wishful receivers. As such, information campaigns aimed at promoting preventive behavior are less effective. We also show how the effectiveness of information campaigns are impacted by the severity of the disease and the effectiveness of the treatment.

This application sheds light on the stylized fact that individuals are consistently investing too little in preventive health care treatments, even if offered at low prices (especially in developing countries, see Dupas, 2011; Chandra et al., 2019; Kremer et al., 2019, Section 3.1) and that informational interventions are often ineffective in inducing more investment in preventive health care devices (see, in particular, Dupas, 2011, Section 4, and Kremer et al., 2019, Section 3.3). Recent literature conjectures that individuals might not be responsive to such information campaigns because they prefer to hold optimistic prospects about their health risks (see Schwardmann, 2019 and Kremer et al., 2019, Section 3.3).[2] Our model formalizes this argument.

Application 2: Persuading a Wishful Investor. In this application, we consider the interaction between a financial broker and her potential client. The broker designs reports about the (continuously distributed) return of some risky financial product to persuade the client to buy the asset. We show that a financial broker interested in selling a risky product is always more effective when persuading a wishful investor.

This application formalizes why some professional financial advisors might sometimes not act in the best interest of their clients by making investment recommen-

[2]There exists compelling experimental evidence that such self-deception exists in the medical testing context (Lerman et al., 1998; Oster et al., 2013; Ganguly and Tasoff, 2017).

dations that take advantage of their biases and mistaken beliefs (see, for instance, Mullainathan et al., 2012 or Beshears et al., 2018, Section 9) as well as why some consulting firms seem to specialize in advice misconduct and cater to biased consumers (Egan et al., 2019). It also helps explaining why the online betting industry puts so much effort into persuasion. Indeed, Babad and Katz (1991) document that individuals generally display wishful thinking when they take part in lotteries: they prefer to think they will win and are therefore more receptive to information encouraging risky bets.

Application 3: Public Persuasion and Political Polarization. Belief polarization along partisan lines is a pervasive and much debated feature of contemporary societies. Although such polarization can be partly caused by differential access to information, evidence suggests that it is exacerbated by the fact that individuals tend to make motivated inferences about the *same* piece of information (Babad, 1995; Thaler, 2020).

In this application we explore the relationship between optimal information disclosure to wishful citizens and belief polarization. Following Alonso and Câmara (2016), we model a majority voting setting in which an electorate, differentiated in terms of partisan preferences, uses information disclosed by a politician to vote on a proposal. Wishful thinking leads voters with different preferences to adopt different beliefs after being exposed to a public signal: those voting against or for the proposal distort their beliefs in opposite directions, giving rise to polarization. Sender's optimal public experiment consists in persuading the median voter, which maximizes the number of voters distorting beliefs in opposite directions. We show that if partisan preferences are symmetrically distributed around the median, then Sender's optimal information policy generates maximal belief polarization in the electorate as a byproduct. This adds nuance to the argument that motivated thinking is one of the drivers of polarization: not only can motivated thinking lead to polarization, but the strategic disclosure of information to a motivated electorate can also accentuate this tendency[3].

[3]This application is related to the paper by Le Yaouanq (2021) who constructs a model of large elections with motivated voters. As in our model, the formation of motivated beliefs by citizens leads voters with different preferences to hold different beliefs after observing the same information. We find, as he does, that greater heterogeneity in partisan preferences increases belief polarization but has

Related literature. The persuasion and information design literature[4] has initially focused on the problem of influencing rational Bayesian decision-makers as in the seminal contributions of Kamenica and Gentzkow (2011) and Bergemann and Morris (2016). By introducing non-Bayesian updating in the form of motivated beliefs formation, we contribute to the literature studying persuasion of receivers subject to mistakes in probabilistic inferences.[5][6] Levy et al. (2018) analyze a Bayesian persuasion problem where a sender can send multiple signals to a receiver subject to correlation neglect. Benjamin et al. (2019) provide an example of persuasion game where Receiver exhibits base-rate neglect when updating beliefs. In de Clippel and Zhang (2020) the receiver holds subjective beliefs which belong to a broader class of distorted Bayesian posteriors. In contrast, in our model, Receiver's belief formation process optimally trades-off the benefits and costs associated with maintaining non-Bayesian beliefs as in the work of Caplin and Leahy (2019).

On the one hand, we assume that Receiver's value from maintaining inaccurate beliefs comes from the anticipation of the payoff he will achieve in equilibrium. Intuitively, it represents the idea that individuals might derive utility from the anticipation of future outcomes, be them good or bad. This hypothesis has been widely used in the literature to study how anticipatory emotions affect physical choices (see, e.g., Loewenstein, 1987; Caplin and Leahy, 2001) as well as choices of beliefs (Bénabou and Tirole, 2002; Brunnermeier and Parker, 2005; Bracha and Brown, 2012; Caplin and Leahy, 2019). Receiver's choice of beliefs is thus a way of satisfying his psychological need to be optimistic about the best-case outcomes or, on the contrary, to avoid the dread and anxiety associated with the worst-case outcomes. This hypothesis is

no effect on the policy implemented in equilibrium. This is, however, the consequence of a different modelling assumption. Namely, that information is endogenously designed to persuade the median voter, whose vote is not distorted relative to a Bayesian voter.

[4] See Bergemann and Morris (2019) and Kamenica (2019) for reviews of this literature.

[5] See Benjamin (2019) for a review of the literature. In particular, wishful thinking belongs to preference-biased inferences reviewed in Benjamin (2019), Section 9.

[6] It is interesting to note that an active literature also explores how errors in strategic reasoning (Eyster, 2019) affect equilibrium outcomes in strategic communication games. Although in our model Receiver understands all the strategic issues, we believe, nevertheless, that it is important to mention that players' misunderstanding of their strategic environment might also lead them to make errors in statistical inference even if they update beliefs via Bayes' rule, as in Mullainathan et al. (2008), Ettinger and Jehiel (2010), Hagenbach and Koessler (2020) and Eliaz et al. (2021a,b) who consider communication games where players make inferential errors because of a coarse understanding of their environment.

supported experimentally by Engelmann et al. (2019), who find significant evidence that wishful thinking is caused by the desire to reduce anxiety associated with anticipating bad events. It is important to note that while anticipatory utility may be a strong motive for manipulating one's beliefs, it is not the only possible one. This differentiates wishful thinking from the more general concept of motivated reasoning, which is usually defined as the degree to which individuals' cognition is affected by their motivations.[7] Different motivations from anticipated payoffs have been explored in the literature such as cognitive dissonance avoidance (Akerlof and Dickens, 1982; Golman et al., 2016), preference to believe in a "Just World" (Bénabou and Tirole, 2006), maintaining high motivation when individuals are aware of being subject to a form of time-inconsistency (Bénabou and Tirole, 2002, 2004) or satisfying the need to belong to a particular identity (Bénabou and Tirole, 2011).

On the other hand, we assume distorting beliefs away from the Bayesian benchmark is subject to some psychological cost. This assumption reflects the idea that, under a motivated cognition process (Kunda, 1987, 1990), individuals may use sophisticated mental strategies such as manipulating their own memory (Bénabou, 2015; Bénabou and Tirole, 2016)[8], avoiding freely available information (Golman et al., 2017) or creating elaborate narratives supporting their bad choices or inaccurate claims to justify their preferred beliefs.[9] Our assumptions on the cost function captures, in "reduced form", the fact that implementing such mental strategies comes at a cost when desired beliefs deviate from from the Bayesian rational ones. In contrast, Brunnermeier and Parker (2005) model the cost of erroneous beliefs as the instrumental loss associated with the inaccurate choices induced by such beliefs. It is worth noting that Coutts (2019) provides experimental evidence in favor of the psychological rather than instrumental costs associated with belief distortion.

[7] See Krizan and Windschitl (2009) for a more detailed discussion on the differences between wishful thinking and motivated reasoning.

[8] For experimental evidence on memory manipulation see, e.g., Saucet and Villeval (2019), Carlson et al. (2020) and Chew et al. (2020).

[9] One can relate this possible microfoundation of the belief distortion cost to the literature on lying costs (Abeler et al., 2014, 2019) since, when Receiver is distorting away his subjective belief from the rational Bayesian beliefs, he is essentially lying to himself. We thank Emeric Henry for suggesting us this interpretation of the cost function.

2 Model

States and prior belief. A state of the world θ is drawn by Nature from a state space Θ according to a prior distribution $\mu_0 \in \text{int}(\Delta(\Theta))$.[10] Receiver (he) and Sender (she) do not observe the state ex-ante but its prior distribution is common knowledge.

Actions and payoffs. Receiver chooses an action a from a compact space A with at least two actions. His material payoff is given by $u(a,\theta)$.[11] Receiver's choice affects Sender's payoff, which is given by $v(a)$. Before Receiver takes his action, Sender can commit to any signal structure (σ, S) given by an endogenously chosen set of signal realizations S and a stochastic mapping $\sigma\colon \Theta \to \Delta(S)$ associating any realized state θ to a conditional distribution $\sigma(\theta)$ over S.

Receiver's behavior. For any belief $\eta \in \Delta(\Theta)$, Receiver's optimal action correspondance is given by

$$A(\eta) = \arg\max_{a\in A} \int_\Theta u(a,\theta)\,\eta(\mathrm{d}\theta).$$

Without loss of generality, we assume that no action is dominated, i.e., for any action $a \in A$ there always exists some belief η such that $a \in A(\eta)$. When the set $A(\eta)$ has more than one element we break the tie in favor of Sender. That is, for any belief η, the action played by Receiver in equilibrium is given by a selection $a(\eta) \in A(\eta)$ which maximizes Sender's expected payoff.[12]

Receiver's beliefs. After observing any signal realization $s \in S$, a Bayesian decision-maker's belief is given by

$$\mu\left(\tilde{\Theta}|s\right) = \frac{\int_{\tilde{\Theta}} \sigma(s|\theta)\,\mu_0(\mathrm{d}\theta)}{\int_\Theta \sigma(s|\theta)\,\mu_0(\mathrm{d}\theta)},$$

[10] In what follows, for any nonempty Polish space X, we denote $\Delta(X)$ the set of Borel probability measures over the measure space $(X,\mathscr{B}(X))$. We always endow $\Delta(X)$ with the weak*-topology. If the support of a measure $\mu \in \Delta(X)$ is finite we adopt the shorthand notation $\mu(\{x\}) = \mu(x)$ for any $x \in \text{supp}(\mu)$.

[11] We assume the map $u(a,\cdot)\colon \Theta \to \mathbb{R}$ to be Borel measurable, continuous and bounded for any $a \in A$.

[12] There might be more than one such selection if there exists some $\eta \in \Delta(\Theta)$ at which Sender is indifferent between some actions in $A(\eta)$. In that case, we pick arbitrarily one of those.

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

In contrast, we assume that, when forming beliefs, Receiver trades-off the psychological benefit against the psychological cost of holding possibly non-Bayesian beliefs. The psychological benefit of Receiver under a certain belief η is given by his *anticipated material payoff*

$$U(\eta) = \int_{\Theta} u(a(\eta), \theta)\, \eta(\mathrm{d}\theta).$$

However, holding belief η when the Bayesian belief generated by some signal is μ comes at a psychological cost $C(\eta, \mu)$ for Receiver. We assume that this cost is given by the Kullback-Leibler divergence between η and μ, formally defined by

$$C(\eta, \mu) = \int_{\Theta} \frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta) \ln\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) \mu(\mathrm{d}\theta),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\mathrm{d}\eta/\mathrm{d}\mu$ is the Radon-Nikodym derivative of η with respect to μ, defined whenever η is absolutely continuous with respect to μ. This assumption is made for tractability but does not qualitatively affect our main results.[13] Accordingly, we define Receiver's *psychological payoff* as

$$\Psi(\eta, \mu) = U(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\rho \in \mathbb{R}_+^*$ parametrizes the extent of Receiver's *wishfulness*. Receiver's belief η must maximize his psychological payoff given any Bayesian belief μ. Therefore, it must belong to the optimal beliefs correspondence

$$B(\mu) = \underset{\eta \in \Delta(\Theta)}{\arg\max}\, \Psi(\eta, \mu),$$

for any $\mu \in \Delta(\Theta)$, and Receiver's psychological payoff when he holds a belief $\eta \in B(\mu)$ is

$$\Psi(\mu) = \max_{\eta \in \Delta(\Theta)} \Psi(\eta, \mu),$$

[13] We show that our results on Receiver's equilibrium beliefs and behavior continue to hold when the psychological cost functions belongs to a more general class of statistical divergences in Appendix A.

for any Bayesian posterior $\mu \in \Delta(\Theta)$.[14] We assume that when Receiver is psychologically indifferent between several beliefs in $B(\mu)$ he picks the one that maximizes Sender's expected utility. Therefore, Receiver's *equilibrium belief* is given by a selection $\eta(\mu) \in B(\mu)$ which maximizes Sender's expected payoff.[15] This tie breaking rule ensures that the Receiver's equilibrium belief is uniquely defined and simplifies the characterization of the optimal information policy.

Persuasion problem. We can equivalently think of Sender committing ex-ante to a signal structure (σ, S) or to an *information policy* $\tau \in \mathcal{T}(\mu_0)$, where

$$\mathcal{T}(\mu_0) = \left\{ \tau \in \Delta(\Delta(\Theta)) : \int_{\Delta(\Theta)} \mu\left(\tilde{\Theta}\right) \tau(\mathrm{d}\mu) = \mu_0\left(\tilde{\Theta}\right) \text{ for any Borel set } \tilde{\Theta} \subseteq \Theta \right\},$$

is the set of Bayes-plausible distributions over posterior beliefs given the prior μ_0.

We assume *Sender knows Receiver is a wishful thinker*. Accordingly, she correctly anticipates the belief Receiver holds in equilibrium. Since Receiver's equilibrium belief characterizes how he would distort his belief away from any realized Bayesian posterior, Sender can choose the best information policy by backward induction, knowing: (i) which belief $\eta(\mu)$ Receiver holds in equilibrium after a posterior $\mu \in \text{supp}(\tau)$ is realized and (ii) which action $a(\eta(\mu))$ Receiver chooses in equilibrium given the distorted belief $\eta(\mu)$. Sender's indirect payoff function is therefore given by

$$v(\mu) = v\left(a(\eta(\mu))\right)$$

for any $\mu \in \Delta(\Theta)$ and, hence, Sender's value from persuading a wishful Receiver under

[14] As already noted by Bracha and Brown (2012) as well as Caplin and Leahy (2019), this optimization problem has a similar mathematical structure to the multiplier preferences developed in Hansen and Sargent (2008) and axiomatized in Strzalecki (2011). Precisely, the agent in Strzalecki (2011) solves

$$\max_{a \in A} \min_{\eta \in \Delta(\Theta)} \int_{\Theta} u(a,\theta)\, \eta(\mathrm{d}\theta) + \frac{1}{\rho} C(\eta, \mu), \tag{1}$$

for any given $\mu \in \Delta(\Theta)$. In that model, the parameter ρ measures the degree of confidence of the decision-maker in the belief μ or, in other words, the importance he attaches to belief misspecification. Conclusions on the belief distortion in that setting are naturally reversed with respect to our model: a receiver forming beliefs according to Equation (1) would form overcautious beliefs. Studying how a rational Sender would persuade a Receiver concerned by robustness seems an interesting path for future research.

[15] Again, if Sender is indifferent between some beliefs we pick arbitrarily one of those.

the prior μ_0 is

$$V(\mu_0) = \max_{\tau \in \mathcal{T}(\mu_0)} \int_{\Delta(\Theta)} v(\mu)\, \tau(\mathrm{d}\mu). \tag{2}$$

3 Receiver's wishful beliefs and behavior

In this section, we first extend Caplin and Leahy (2019) results by characterizing Receiver's equilibrium beliefs and behavior without imposing any restrictions on the action or state space.

To begin with, let Receiver's anticipated material payoff under action a and belief η be defined by

$$U_a(\eta) = \int_{\Theta} u(a,\theta)\, \eta(\mathrm{d}\theta).$$

Moreover, let

$$\eta_a(\mu) = \operatorname*{arg\,max}_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's belief motivated by action a under posterior μ and

$$\Psi_a(\mu) = \max_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's maximal psychological payoff motivated by action a under posterior μ. We identify Receiver's equilibrium belief $\eta(\mu)$ by: (i) finding the belief motivated by action a under μ, resulting in psychological payoff $\Psi_a(\mu)$, for any a and μ; (ii) finding which action it is optimal to motivate by maximizing $\Psi_a(\mu)$ with respect to a. Proposition 1 characterizes $\eta_a(\mu)$ and $\Psi_a(\mu)$ in closed-form.

Proposition 1. *Receiver's maximal psychological payoff motivated by action a under the Bayesian posterior μ is given by*

$$\Psi_a(\mu) = \frac{1}{\rho} \ln \left(\int_{\Theta} \exp\left(\rho\, u(a,\theta)\right) \mu(\mathrm{d}\theta) \right), \tag{3}$$

and is attained uniquely at the belief

$$\eta_a(\mu)\left(\tilde{\Theta}\right) = \frac{\int_{\tilde{\Theta}} \exp\left(\rho\, u(a,\theta)\right) \mu(\mathrm{d}\theta)}{\int_{\Theta} \exp\left(\rho\, u(a,\theta)\right) \mu(\mathrm{d}\theta)}. \tag{4}$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

Proof. See Appendix A. □

Remark now that if the action a uniquely maximizes Receiver's psychological payoff under Bayesian posterior μ we have $\eta(\mu) = \eta_a(\mu)$. If, on the other hand, $\Psi_a(\mu) = \Psi_{a'}(\mu)$ at μ for some $a' \neq a$, meaning that Receiver is psychologically indifferent between two beliefs, then Sender breaks the tie. As a consequence, if $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) > \Psi_{a'}(\mu), \tag{5}$$

for all $a' \neq a$, meaning that Receiver psychologically prefers action a to any other action a', then Receiver's equilibrium belief is given by

$$\eta(\mu)\left(\tilde{\Theta}\right) = \eta_a(\mu)\left(\tilde{\Theta}\right),$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$. If $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) = \Psi_{a'}(\mu),$$

for some $a' \neq a$, meaning that Receiver is psychologically indifferent between some actions a' and a, then Sender picks her preferred belief given by

$$\eta(\mu)\left(\tilde{\Theta}\right) = \eta_{a^*}(\mu)\left(\tilde{\Theta}\right),$$

where $a^* \in \arg\max_{\tilde{a} \in \{a,a'\}} v(\tilde{a})$.

First, we can see from Equation (4) that Receiver only distorts beliefs that induce actions with state-dependant payoffs, i.e., Receiver's beliefs are *stakes-dependent*. Formally, for any $a \in A$, we have $\eta_a(\mu) \neq \mu$ if, and only if, there exists $\theta \neq \theta'$ such that $u(a,\theta) \neq u(a,\theta')$. Second, Receiver forms beliefs that overweight the states associated with the highest payoff, giving rise to *overoptimism*. Formally, we always have $\eta_a(\mu)(\Theta_a) \geq \mu(\Theta_a)$ for any $a \in A$ where $\Theta_a = \arg\max_{\theta \in \Theta} u(a,\theta)$. Moreover, Receiver's belief about payoff maximizing states $\eta_a(\mu)(\Theta_a)$ grows monotonically and eventually converges to 1 as Receiver's wishfulness ρ grows from 0 to $+\infty$.[16]

[16] This property comes from the fact that wishful beliefs take the form of a soft-max function. For the sake of completeness we provide a proof of this result in Appendix B.

As Proposition 1 shows, wishful thinking leads Receiver to hold overoptimistic beliefs. The next result shows that wishful thinking distorts Receiver's behavior accordingly.

Corollary 1. *Under his equilibrium belief, Receiver's optimal action correspondence is given by*

$$A(\eta(\mu)) = \arg\max_{a \in A} \int_{\Theta} \exp\left(\rho\, u(a,\theta)\right) \mu(\mathrm{d}\theta),$$

for any $\mu \in \Delta(\Theta)$ so Receiver's equilibrium action $a(\eta(\mu))$ corresponds to Sender's preferred selection in $A(\eta(\mu))$.

Remark that this result comes as a direct consequence of Proposition 1 as, by definition, any action a is optimal under the belief motivated by action a. As already observed by Caplin and Leahy (2019), the previous result states, in essence, that a Receiver forming wishful beliefs behaves as a Bayesian agent whose preferences are distorted by the function $z \mapsto \exp(\rho z)$ for any $z \in \mathbb{R}$. Importantly, from Sender's point of view, a wishful Receiver's behavior is indistinguishable from that of a Bayesian rational agent with payoff function $\exp(\rho\, u(a,\theta))$. Accordingly, since the function $z \mapsto \exp(\rho z)$ is strictly convex as soon as $\rho > 0$, an agent forming wishful beliefs is less risk averse than his Bayesian self.

Corollary 1 also shows that wishful thinking materializes in the form of "motivated errors" in the sense of Exley and Kessler (2019): by choosing psychologically desirable beliefs, Receiver commits systematic errors in his decision-making, i.e., acts as if he had cognitive limitations or behavioral biases relatively to a Bayesian decision-maker.

4 Sender's value from persuasion

In this section, we assume that the action space of Receiver is binary, so $A = \{0,1\}$, and that Sender wants to induce $a = 1$, so $v(a) = a$. We characterize conditions on Receiver's preferences under which he would take action 1 under a greater set of beliefs than a Bayesian Receiver. This allows us to compare Sender's value from persuading a wishful rather than a Bayesian Receiver as a function of the model's primitives, that is: Receiver's preferences and wishfulness. The restriction to a binary set of actions is

with loss of generality, but this assumption turns out to be necessary for tractability.

We start by defining the two following sets of beliefs:

$$\Delta_a^B = \left\{\mu \in \Delta(\Theta) : a \in A(\mu)\right\},$$

and

$$\Delta_a^W = \left\{\mu \in \Delta(\Theta) : a \in A(\eta(\mu))\right\},$$

for any $a \in A$. The set Δ_a^B (resp. Δ_a^W) is the subset of posterior beliefs supporting an action a as optimal for a Bayesian (resp. wishful) Receiver. We say that an action is *favored* by a wishful receiver if that action is supported as optimal on a strictly larger set of posterior beliefs by a wishful Receiver compared to a Bayesian.

Definition 1 (Favored action)**.** *An action $a \in A$ is favored by a wishful Receiver if $\Delta_a^B \subset \Delta_a^W$.*

Assume for now on that $\Theta = \{\underline{\theta}, \overline{\theta}\}$. We first characterize when a wishful Receiver favors action $a = 1$ when the state space is binary and show afterwards that our results extend to any finite state space. Let us denote $u(a, \underline{\theta}) = \underline{u}_a$ and $u(a, \overline{\theta}) = \overline{u}_a$ for any $(a, \theta) \in A \times \Theta$. Assume that Receiver wants to "match the state," such that $\overline{u}_1, \underline{u}_0 > \overline{u}_0, \underline{u}_1$. Define the *payoff variability under action* 0 by $u_0 = \underline{u}_0 - \overline{u}_1$, the *payoff variability under action* 0 by $u_1 = \overline{u}_1 - \underline{u}_1$ and the indicator of the *highest achievable payoff* by $u_{\max} = \underline{u}_0 - \overline{u}_1$. With a small abuse of notation, denote $\eta = \eta(\overline{\theta})$ and $\mu = \mu(\overline{\theta})$.

By Corollary 1, comparing how a wishful Receiver behaves compared to a Bayesian one is equivalent to comparing the behavior of two Bayesian receivers with respective payoff functions $\exp(\rho u(a, \theta))$ and $u(a, \theta)$. Thus, denote μ^B (resp. $\mu^W(\rho)$) the belief at which a Receiver with preferences $u(a, \theta)$ (resp. $\exp(\rho u(a, \theta))$) is indifferent between the two actions. Those beliefs are respectively equal to

$$\mu^B = \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0}$$

and

$$\mu^W(\rho) = \frac{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1)}{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1) + \exp(\rho \overline{u}_1) - \exp(\rho \overline{u}_0)}.$$

With only two states, a wishful Receiver favors action $a = 1$ if and only if $\mu^W < \mu^B$, since whenever that condition is satisfied a wishful Receiver takes action $a = 1$ under a larger set of beliefs than a Bayesian. Next proposition characterizes when this is the case.

Lemma 1. *Action $a = 1$ is favored by a wishful Receiver if, and only if:*

(i) $u_{\max} \leq 0$ *and* $u_0 < u_1$*, or;*

(ii) $u_{\max} < 0$, $u_0 > u_1$ *and* $\rho > \overline{\rho}$*, or;*

(iii) $u_{\max} > 0$, $u_0 < u_1$ *and* $\rho < \overline{\rho}$*.*

where $\overline{\rho}$ is a strictly positive threshold such that

$$\mu^W(\overline{\rho}) = \mu^B.$$

Proof. See Appendix C. □

Two key aspects of Receiver's material payoff thus determine which action he favors: *the highest achievable payoff* as well as *the payoff variability* for both actions. It is easy to grasp the importance of the highest payoff. Since the wishful thinker always distorts his beliefs in the direction of the most favorable outcome, in the limit, when there is no cost of distorting the Bayesian belief, Receiver would fully delude himself and always play the action that potentially yields such a payoff. The payoff variability u_a, on the other hand, is precisely Receiver's marginal psychological benefit from distorting his belief under action a. Hence, the higher the payoff variability associated with action a, the more the uncertainty about θ is relevant when such action is played and the bigger the marginal gain in anticipatory payoff the wishful thinker would get from distorting beliefs.

Lemma 1 states that if an action a has both the highest payoff $\underline{u}_0$ or $\overline{u}_1$ and the greatest payoff variability u_a among all actions $a \in A$, it is always favored. If an action has either the highest payoff or the greatest payoff variability, then the wishfulness parameter ρ defines whether or not it is favored: for high wishfulness the action with the highest payoff is favored, whereas for low wishfulness it is the action with the greatest payoff variability that is favored. The intuition is the following: for sufficiently high

values of Receiver's wishfulness, Receiver can afford stronger overoptimism about the most desired outcome, thus favoring the action that potentially yields this outcome despite such action not being associated with the highest marginal psychological benefit. In contrast, for sufficiently low values of ρ, Receiver cannot afford too much overoptimism about the most desired outcome. Hence, he prefers to distort beliefs at the margin that yields the highest marginal psychological benefit, such that the action associated with the highest payoff variability is favored.

The next proposition extends Lemma 1 to an arbitrary finite number of states.

Proposition 2. *Assume Θ is a finite set with more than two elements. Receiver favors action $a = 1$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in Lemma 1.*

Proof. See Appendix D. □

Proposition 2 can easily be visualized graphically in an example with three states. Assume $\Theta = \{0,1,2\}$ and denote $\mu^B_{\theta,\theta'}$ (resp. $\mu^W_{\theta,\theta'}$) the belief making a Bayesian (resp. wishful) Receiver indifferent between actions $a = 0$ and $a = 1$ when $\mu(\theta), \mu(\theta') > 0$ but $\mu(\theta'') = 0$ for any $\theta, \theta', \theta'' \in \Theta$. In Figure 1 we illustrate how Δ^W_1 compares to Δ^B_1 when Receiver's payoff function is given by:

$u(a,\theta)$	$\theta = 0$	$\theta = 1$	$\theta = 2$
$a = 0$	2	3	-1
$a = 1$	1	0	4

Notice that for the two pairs of states $(0,2)$ and $(1,2)$, the associated payoffs satisfy property (i) in Lemma 1. That is, action $a = 1$ is associated with the highest payoff $u(1,2) = 4$ as well as the highest payoff variability $u(1,2) - u(0,2) = 5$, under both pair of states. As a consequence, Lemma 1 applies whenever focusing on those two pairs of states letting the other one being assigned probability zero. Then, we have $\mu^W_{0,2} > \mu^B_{0,2}$ and $\mu^W_{1,2} > \mu^B_{1,2}$. Remark now, that $\Delta^B_1 = \text{co}(\{\mu^B_{0,2}, \mu^B_{1,2}, \delta_2\})$ and $\Delta^W_1 = \text{co}(\{\mu^W_{0,2}, \mu^W_{1,2}, \delta_2\})$, where δ_θ denotes the Dirac distribution on state $\theta \in \Theta$. Consequently, $\Delta^B_1 \subset \Delta^W_1$ so action $a = 1$ is favored by Receiver. If one of the conditions highlighted in Lemma 1 were not satisfied for at least one of the pairs of states $(0,2)$ or $(1,2)$ then one of the

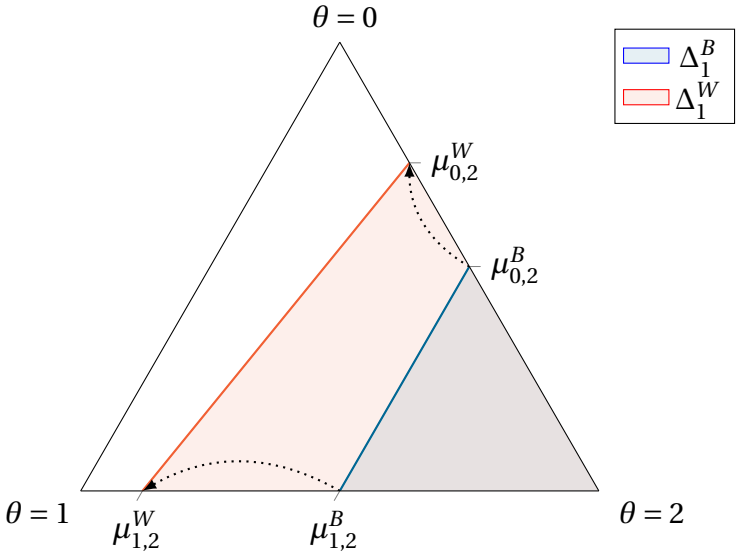


Figure 1: Comparison of supporting sets of beliefs. In blue, the set of Bayesian posteriors supporting action $a = 1$ for a Bayesian Receiver. In red, the set of Bayesian posteriors supporting action $a = 1$ for a wishful Receiver.

thresholds $\mu_{\theta,\theta'}^W$ would be less or equal than $\mu_{\theta,\theta'}^B$ in which case Δ_1^W would not be a superset of Δ_1^B anymore.

Let us now turn our attention to the following questions: when is Sender better-off facing a wishful Receiver compared to a Bayesian and how does the (Blackwell) informativeness of Sender's optimal policy compare when persuading a wishful or a Bayesian Receiver? Remember that Sender chooses an information policy $\tau \in \Delta(\Delta(\Theta))$ maximizing

$$\int_{\Delta(\Theta)} v(\mu)\,\tau(\mathrm{d}\mu),$$

where

$$v(\mu) = \begin{cases} 1 & \text{if} \quad \mu \in \Delta_1^W \\ 0 & \text{otherwise} \end{cases},$$

subject to the Bayes plausibility constraint

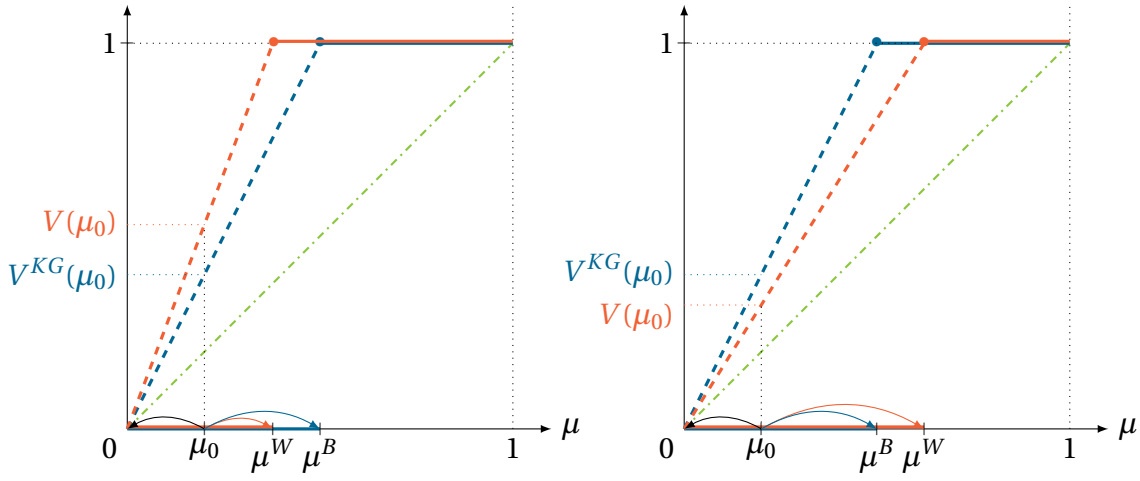
$$\int_{\Delta(\Theta)} \mu\,\tau(\mathrm{d}\mu) = \mu_0.$$

In the binary state case, it means that the threshold belief μ^W corresponds to the

smallest Bayesian posterior Sender needs to induce to persuade a wishful Receiver to take action $a = 1$. Therefore, Lemma 1 and Proposition 2 have immediate consequences for Sender.

Corollary 2. *Let Θ be an arbitrary finite space with at least two elements. Then, Sender always achieves a weakly higher payoff when interacting with a wishful Receiver compared to a Bayesian for any prior $\mu_0 \in]0,1[$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in Lemma 1. Moreover, when the state space is binary, Sender's optimal information policy is always weakly less (Blackwell) informative than in the Bayesian case.*

To illustrate Corollary 2 we represent in Figure 2 the concavifications of Sender's indirect utility when Receiver is wishful or Bayesian in two different cases. The case



(a) At least one property in Lemma 1 is satisfied.

(b) No property in Lemma 1 is satisfied.

Figure 2: Expected payoffs under optimal information policies. Red curves: expected payoffs under wishful thinking. Blue curves: expected payoffs when Receiver is Bayesian. Dashed-dotted green lines: expected payoffs under a fully revealing experiment.

corresponding to Lemma 1 is represented in Figure 2a. Sender is always better-off persuading a wishful compared to a Bayesian receiver as $V(\mu_0) \geq V^{KG}(\mu_0)$ for any $\mu_0 \in]0,1[$. On the other hand, if Receiver's preferences or wishfulness do not satisfy

any of the properties in Lemma 1, then Sender is weakly worse-off under any prior. This case is represented on Figure 2b.

When Sender wants to induce an action that is (resp. is not) favored by a wishful Receiver, persuasion is always "easier" (resp. "harder") for Sender in the following sense: Sender needs a strictly less (resp. strictly more) Blackwell informative policy than KG to persuade Receiver to take his preferred action. Equivalently, if experiments were costly to produce, as in Gentzkow and Kamenica (2014), then Sender would always need to consume less (resp. more) resources to persuade a wishful Receiver to take his preferred action than a Bayesian. The hypothesis of a binary state space facilitates the comparisons between the Bayesian-optimal and the wishful-optimal information policies as it ensures that the Bayesian-optimal and the wishful-optimal information policies are Blackwell comparable. Although the informativeness comparisons in Corollary 2 do not necessarily extend when the state space contains more than two elements, Sender's welfare comparisons, in contrast, still hold under any arbitrary finite state space. We compare in Figure 3 Sender's optimal information policies when Receiver is Bayesian and wishful, with the same payoff function as in Figure 1. When the state space is finite, a policy $\tau \in \mathcal{T}(\mu_0)$ such that all elements in

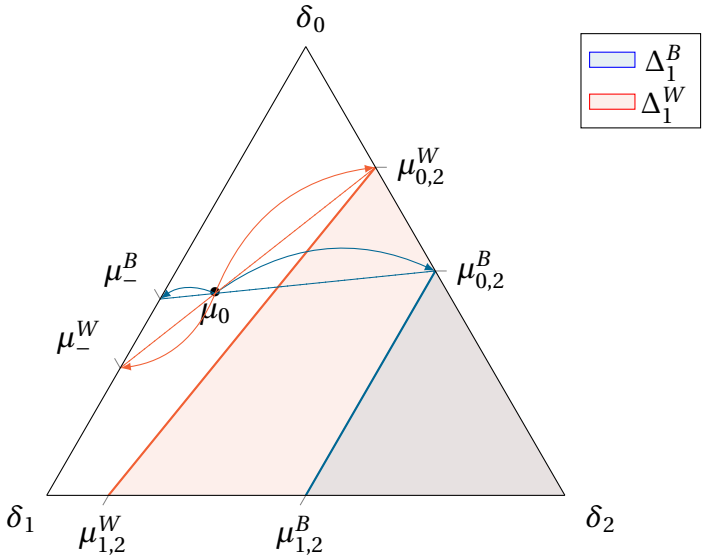


Figure 3: The Bayesian-optimal policy τ^B (in blue) vs. the wishful-optimal policy τ^W (in red) with respective supports $\{\mu_-^B, \mu_{0,2}^B\}$ and $\{\mu_-^W, \mu_{0,2}^W\}$.

$\text{supp}(\tau)$ are affinely independent is (weakly) more Blackwell-informative than a pol-

icy $\tau' \in \mathcal{T}(\mu_0)$ if, and only if, and $\text{supp}(\tau') \subset \text{co}(\text{supp}(\tau))$ (see Lipnowski et al., 2020, Lemma 2). The support of the Bayesian-optimal policy τ^B (resp. wishful-optimal policy τ^W) is $\{\mu_-^B, \mu_{0,2}^B\}$ (resp. $\{\mu_-^W, \mu_{0,2}^W\}$). Hence, $\text{co}(\text{supp}(\tau^W)) = \{\mu \in \Delta(\Theta) : \exists t \in [0,1], \mu = t\mu_-^W + (1-t)\mu_{0,2}^W\}$. It is visible on Figure 3 that $\{\mu_-^B, \mu_{0,2}^B\} \not\subset \text{co}(\text{supp}(\tau^W))$. Hence, τ^B and τ^W are not Blackwell comparable. However, since Sender is interested in inducing action $a = 1$ and Receiver's favors that action, Sender's expected payoff is higher for any prior when Receiver is wishful.

5 Applications

In this section, we expose in three applications that Corollary 2 might have important economic consequences.

5.1 Information provision and preventive health care

A public health agency (Sender) informs an individual (Receiver) about the prevalence of a certain disease. Receiver forms beliefs about the infection risk, which can be either high or low: $0 < \underline{\theta} < \overline{\theta} < 1$. The probability of contracting that illness also depends on whether the individual adopts a preventive treatment or not, where $a = 1$ designates adoption. Investment in the treatment entails a cost $c > 0$ to Receiver.[17] Moreover, let us assume that the effectiveness of the treatment, i.e., the probability that the treatment works, is $\alpha \in [0,1]$ so that the probability of falling ill, conditional on adoption, is $(1-\alpha)\theta$. The payoff from staying healthy is normalized to 0 whereas the payoff from being infected equals $-\varsigma < 0$ where ς is the severity of the disease. Receiver's payoff function is

$$u(a,\theta) = (1-a)(-\varsigma\theta) + a(-(1-\alpha)\theta\varsigma - c)$$

for any $(a,\theta) \in A \times \Theta$. We assume that $\varsigma\alpha\underline{\theta} < c < \varsigma\alpha\overline{\theta}$ so Receiver faces a trade-off: he would prefer not to invest if he was sure the probability of infection was low and, conversely, would prefer to invest in the treatment if he was sure the risk of infection

[17] One might interpret that cost to be the price of the treatment or the either material or psychological cost from undertaking medical procedures.

is high. Also remark that Receiver always expects to experience a negative payoff, as $u(a,\theta) < 0$ for any $(a,\theta) \in A \times \Theta$.

The public health agency wants to maximize the probability of individuals adopting the preventive treatment.[18] The agency informs individuals about the prevalence of the disease by designing and committing to a Bayes-plausible information policy τ. A Bayesian Receiver would be indifferent between adopting or not the treatment at belief

$$\mu^B = \frac{c - \alpha\underline{\theta}\varsigma}{\alpha(\overline{\theta} - \underline{\theta})\varsigma}.$$

In contrast, by Proposition 1 and Corollary 1, the equilibrium beliefs and behavior of a wishful Receiver are given by

$$\eta(\mu) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho\varsigma(\overline{\theta} - \underline{\theta}))} & \text{if} \quad \mu < \mu^W \\ \\ \dfrac{\mu\exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta}))}{\mu\exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta})) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W \end{cases},$$

and

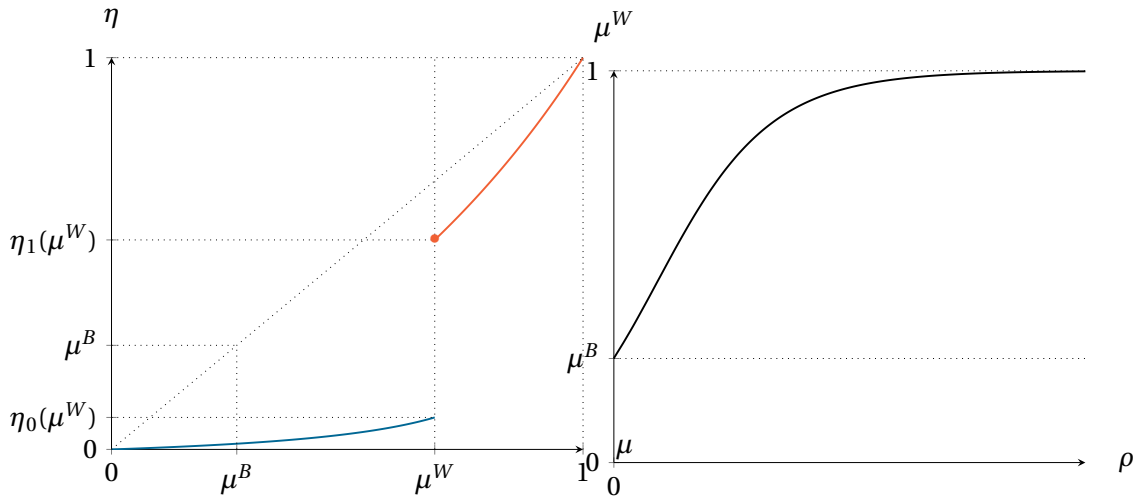
$$a(\eta(\mu)) = \mathbb{1}\left\{\mu \geq \mu^W\right\}$$

for any posterior belief $\mu \in [0,1]$, where

$$\mu^W = \frac{\exp(-\rho\underline{\theta}\varsigma) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c))}{\exp(-\rho\varsigma\underline{\theta}) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c)) + \exp(\rho(-(1-\alpha)\overline{\theta}\varsigma - c)) - \exp(-\rho\overline{\theta}\varsigma)}.$$

We illustrate the belief distortion of Receiver in Figure 4a. Receiver is always overoptimistic about his probability of staying healthy, as $\eta(\mu) \leq \mu$ for any $\mu \in [0,1]$. Remark that non-adoption is associated with the highest possible payoff $-\varsigma\underline{\theta}$ as well as the highest payoff variability $\varsigma(\overline{\theta} - \underline{\theta})$. Accordingly, by Lemma 1, Receiver always favors non adoption as illustrates Figure 4b. As a result of Corollary 2, Sender always

[18] Maximizing the probability of adoption is a sensible objective since most infections cause negative externalities due to their transmission through social interactions. Therefore, a benevolent planner who wants to reduce the likelihood of transmission of an infection would do well to maximize the rate of adoption of the preventive treatment (for example, maximize condom distribution to control AIDS transmission, maximize injection of vaccines to control viral infections, or maximize mask use to control the spread of airborne diseases).



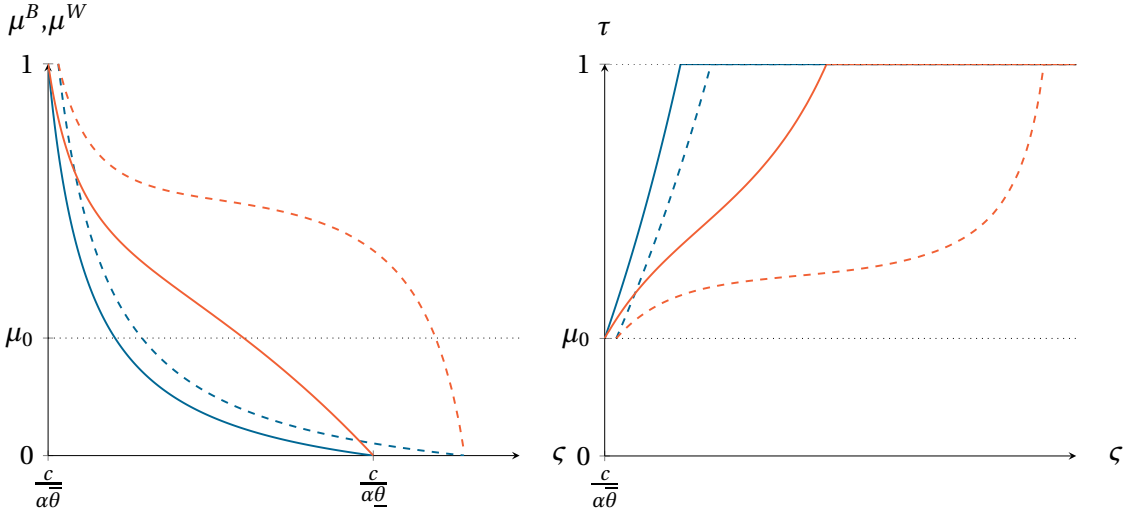
(a) equilibrium belief $\eta(\mu)$ as a function of μ. (b) Behavioral threshold μ^W as a function of ρ.

Figure 4: The belief correspondence for $\varsigma = 2$, $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Receiver is always overoptimistic concerning his health risk for any induced posterior, except at $\mu = 0$ or $\mu = 1$. Moreover, the belief threshold μ^W as a function of ρ is strictly increasing and admits μ^B as a lower bound.

needs to induce higher beliefs for Receiver to adopt the treatment than she would need if she faced a Bayesian agent, all the more so when Receiver's wishfulness ρ becomes larger. Therefore in this example, overoptimism of Receiver always goes against Sender's interest.

It is interesting to see how Sender's probability of inducing the adoption of the treatment evolves with respect to the severity of the disease ς, as well as the effectiveness of the treatment α.[19] We represent on Figure 5b the probability that Sender induces adoption of the treatment under the optimal information policy as a function of ς. Notice that the probability of inducing adoption is less sensitive to the severity of the disease, i.e., becomes "flatter," when facing a wishful Receiver compared to the Bayesian when the treatment becomes less effective. The intuition is the following: when the treatment is fully effective, i.e., $\alpha = 1$, Receiver's payoff in case he invests in the treatment becomes state independent. Therefore, he does not have any incentive to distort beliefs when taking action $a = 1$. As a result, μ^W decreases and Receiver

[19] This probability is pinned down by the Bayes-plausibility constraint and equal to $\tau^{KG} = \mu_0/\mu^B$ in the Bayesian case and $\tau = \mu_0/\mu^W$ in the wishful case.



(a) Behavioral thresholds μ^B (in blue) and μ^W (in red) as functions of severity ς.

(b) Probability τ of inducing treatment adoption as a function of severity ς.

Figure 5: Red (resp. blue) curves correspond to wishful (resp. Bayesian) Receiver. We set parameters to $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Full lines correspond to the case where $\alpha = 1$ whereas dashed curves correspond to $\alpha = 0.8$.

holds perfectly Bayesian beliefs when $\mu \geq \mu^W$. However, whenever there is uncertainty about the treatment efficacy, i.e., $\alpha < 1$, uncertainty about infection risk matters and gives room to belief distortion even when taking the treatment. Decreasing α increases the anticipated anxiety of Receiver leading to more optimistically biased beliefs, a higher μ^W and, in turn, complicates persuasion for Sender for any severity s. Remark on Figure 5b that τ decreases sharply with α for a fixed s. In fact, one could show that as α decreases, τ becomes closer and closer to μ_0 for any ς, meaning that the agency cannot achieve a substantially higher payoff than under full disclosure.[20]

In the next subsection we extend out framework to the case of a continuous state space and linear preferences. We show that results in the finite state space case extend to this setting. We also highlight why we might expect persuasion to be more

[20] One additional implication of this result is the following. Assume the true treatment efficacy is α but Receiver perceives the efficacy to be $\hat{\alpha} < \alpha$ (e.g. because Receiver adheres to anti-vaccines movements or generally mistrusts the pharmaceutical industry). In that case, the doubts expressed by Receiver about the treatment efficacy makes him even more anxious which, in turn, makes belief distortion stronger and, thus, downplays the effectiveness of the agency's information policy whatever is the severity of the disease.

effective in the context of risky investment decisions.

5.2 Persuading a wishful investor

A financial broker (Sender) designs reports about the return of some risky financial product to inform a potential client (Receiver). The return of the product is $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, where $\underline{\theta} < 0 < \overline{\theta}$. Returns are distributed according to the prior distribution μ_0. Let F be the cumulative distribution function associated with μ_0 and let us assume that μ_0 admits a continuous and strictly positive density function f over $[\underline{\theta}, \overline{\theta}]$. Receiver has some saved up money he is willing to invest and chooses action $a \in A = \{0,1\}$, where $a = 0$ represents the choice of non-investing in which case Receiver's payoff is 0 and $a = 1$ represents investing, in which case Receiver's payoff is the realized return θ. The broker is remunerated on the basis of a flat fee $v > 0$ that is independent of the true product's profitability. Hence, Receiver's payoff is $u(a,\theta) = a\theta$ while Sender's payoff is $v(a,\theta) = va$ for any $(a,\theta) \in A \times \Theta$.

Receiver forms motivated beliefs about the return of the financial product. By Proposition 1 his equilibrium beliefs are given by

$$\eta(\mu)\left(\tilde{\Theta}\right) = \begin{cases} \mu\left(\tilde{\Theta}\right) & \text{if } \int_{\Theta} \exp(\rho\theta)\,\mu(\mathrm{d}\theta) < 1 \\ \dfrac{\int_{\tilde{\Theta}} \exp(\rho\theta)\,\mu(\mathrm{d}\theta)}{\int_{\Theta} \exp(\rho\theta)\,\mu(\mathrm{d}\theta)} & \text{if } \int_{\Theta} \exp(\rho\theta)\,\mu(\mathrm{d}\theta) \geq 1 \end{cases},$$

for any $\mu \in \Delta(\Theta)$ and any Borel set $\tilde{\Theta} \subseteq \Theta$, and, by Corollary 1, his equilibrium behavior is given by

$$a(\eta(\mu)) = \mathbb{1}\left\{\int_{\Theta} \exp(\rho\theta)\,\mu(\mathrm{d}\theta) \geq 1\right\}.$$

Therefore, Sender's indirect utility is equal to

$$v(\mu) = v\mathbb{1}\left\{\int_{\Theta} \exp(\rho\theta)\,\mu(\mathrm{d}\theta) \geq 1\right\}.$$

for any $\mu \in \Delta(\Theta)$. To make the problem interesting, we assume that neither a Bayesian nor a wishful Receiver would take action $a = 0$ under the prior. That is, $\hat{m} = \int_{\underline{\theta}}^{\overline{\theta}} \theta\,\mu_0(\mathrm{d}\theta) <$

0 and $\hat{x} = \int_{\underline{\theta}}^{\overline{\theta}} \exp(\rho\theta)\mu_0(\mathrm{d}\theta) < 1$.[21]

Under these assumptions, remark that a signal structure σ that induces a distribution τ over posterior beliefs μ matters for Receiver and Sender only through the *distribution of exponential moments* $x = \int_\Theta \exp(\rho\theta)\,\mu(\mathrm{d}\theta)$ *it induces*. Let X be the space of such moments, that is, $X = \mathrm{co}(\exp(\rho\Theta))$, where $\exp(\rho\Theta)$ is the graph of the function $\theta \mapsto \exp(\rho\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. That is, $X = [\underline{x}, \overline{x}]$ where $\underline{x} = \exp(\rho\underline{\theta})$ and $\overline{x} = \exp(\rho\overline{\theta})$. Let G be the prior cumulative distribution function over the random variable $\exp(\rho\theta)$ induced by F, that is

$$G(x) = F\left(\frac{\ln(x)}{\rho}\right),$$

for any $x \in [\underline{x}, \overline{x}]$. By standard arguments (Gentzkow and Kamenica, 2016), the problem of finding an optimal signal structure σ reduces to finding a cumulative distribution function H that maximizes

$$\int_{\underline{x}}^{\overline{x}} v(x)\,\mathrm{d}H(x)$$

subject to

$$\int_{\underline{x}}^{z} H(x)\,\mathrm{d}x \leq \int_{\underline{x}}^{z} G(x)\,\mathrm{d}x$$

for every $z \in [\underline{x}, \overline{x}]$. The solution to such a problem is well-known and can be found either using techniques from optimization under stochastic dominance constraints (Gentzkow and Kamenica, 2016; Ivanov, 2020; Kleiner et al., 2021) or linear programming (Kolotilin, 2018; Dworczak and Martini, 2019; Dizdar and Kováč, 2020). In our context, the optimal signal is a binary partition of the state space. That is, the broker reveals whether the return is above or below some threshold state.

Proposition 3. *There exists a unique* $\theta^W \in [\underline{\theta}, \overline{\theta}]$ *verifying*

$$\frac{1}{1 - F(\theta^W)} \int_{\theta^W}^{\overline{\theta}} \exp(\rho\theta) f(\theta)\,\mathrm{d}\theta = 1$$

and such that Sender pools all states $\theta \in [\theta^W, \overline{\theta}]$ *under the same signal* $s = 1$*, i.e.,* $\sigma(1|\theta) = 1$ *for all* $\theta \in [\theta^W, \overline{\theta}]$*, and similarly pools all states* $\theta \in [\underline{\theta}, \theta^W]$ *under the same*

[21] It is in fact always true that $\hat{m} < 0$ when $\hat{x} < 1$. Hence, assuming $\hat{m} < 0$ additionally to $\hat{x} < 1$ is without loss.

signal $s = 0$. *Hence, the probability of inducing action* $a = 1$ *for Sender is equal to*

$$\int_{\theta^W}^{\overline{\theta}} \sigma(1|\theta) f(\theta)\, \mathrm{d}\theta = 1 - F(\theta^W).$$

Proof. See Ivanov (2020), Section 3. □

It is optimal for Sender to partition the state space at the threshold state making Receiver indifferent between investing or not at the prior. Such an information policy can intuitively be seen as the investment recommendation rule which maximizes the probability that Receiver invests given the prior distribution of returns F.

Using the exact same arguments as above, one can deduce that the probability of inducing action $a = 1$ when Receiver is Bayesian is given by $1 - F(\theta^B)$ where θ^B is the unique threshold verifying the equation

$$\frac{1}{1 - F(\theta^B)} \int_{\theta^B}^{\overline{\theta}} \theta f(\theta)\, \mathrm{d}\theta = 0.$$

Therefore, Sender is more effective at persuading a wishful Receiver if and only if $\theta^W < \theta^B$.

Proposition 4. *It is always true that* $\theta^W < \theta^B$. *Hence, Sender is always more effective at persuading a wishful rather than a Bayesian investor.*

Proof. See Appendix F. □

The above result relates to Proposition 2: buying the risky product is favored by the wishful investor since it is the action that yields both the highest possible payoff and the highest payoff variability. This example thus illustrates how the results in the finite state space case naturally extend to an infinite state space setting with linear preferences. It further helps explaining the pervasiveness of persuasion efforts in financial and betting markets, illustrating why some financial consulting firms seem to specialize in advice misconduct and cater to biased consumers.

5.3 Public persuasion and political polarization

A Sender (e.g., a politician, a lobbyist) persuades an odd-numbered finite group of voters $N = \{1, \ldots, n\}$ (e.g., a committee or parliamentary members) to adopt a pro-

posal $x \in X = \{0,1\}$, where $x = 0$ corresponds to the status-quo. The state space is binary, $\Theta = \{0,1\}$, and the audience uses only the information disclosed by Sender to vote on the proposal. Let $a^i \in A = \{0,1\}$ be the ballot cast by voter i, where $a^i = 0$ designates voting for the status-quo. The proposal is accepted if it is supported by a simple majority of voters. We assume Sender is only interested in the proposal being accepted, so her utility is $v(x) = x$. In contrast, any voter $i \in N$ has payoff function

$$u^i(x,\theta) = x\theta\beta^i + (1-x)(1-\theta)(1-\beta^i)$$

for any $(x,\theta) \in X \times \Theta$ where $\beta^i \in [0,1]$ parametrizes the partisan preference of voter i. That is, all voters agree that the proposal should be implemented only when $\theta = 1$, but they vary in how much they value the implementation of the proposal. We assume β^i is symmetrically distributed around $1/2$ in the population. Denote $\beta^m = 1/2$ the median voter's preference.

All voters form wishful beliefs and ρ is assumed homogeneous among the electorate. As a result, the direction as well as the magnitude of voters' belief distortion depends only on their partisan preferences β.[22] By Proposition 1, voter i's belief under posterior $\mu \in [0,1]$ is given by

$$\eta(\mu,\beta^i) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho(1-\beta^i))} & \text{if} \quad \mu < \mu^W(\beta^i) \\ \\ \dfrac{\mu\exp(\rho\beta^i)}{\mu\exp(\rho\beta^i) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W(\beta^i) \end{cases}.$$

where

$$\mu^W(\beta^i) = \frac{\exp(\rho(1-\beta^i)) - 1}{\exp(\rho(1-\beta^i)) + \exp(\rho\beta^i) - 2}.$$

Remark that, similarly as in Alonso and Câmara (2016), since the policy space is binary and voters do not hold private information there is no room for strategic voting in our model. Hence, citizen i's voting strategy under belief $\eta(\mu,\beta^i)$ is given by

$$a(\eta(\mu,\beta^i)) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^i)\right\}.$$

[22] It has been shown in psychology (Babad et al., 1992; Babad, 1995, 1997) as well as in behavioral economics (Thaler, 2020) that voters political beliefs are often motivated by their partisan orientation.

Due to the heterogeneity in β, there is always some level of belief polarization among wishful voters for any $\mu \in]0,1[$. Let us measure such polarization by the sum of the absolute difference between each pair of beliefs in the audience

$$\pi(\mu) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |\eta(\mu, \beta^i) - \eta(\mu, \beta^j)| \tag{6}$$

for any $\mu \in [0,1]$.

Proposition 5. *Under Sender's optimal information policy, the signal that leads to the implementation of the proposal also generates the maximum polarization among voters.*

Proof. See Appendix E. □

To build an intuition of why this is the case, let's first note that, in our model, belief polarization and action polarization are closely related. Agents voting for the implementation of the proposal distort their beliefs upwards, whereas agents voting for the status quo distort their beliefs downwards. We can thus see that maximum belief polarization should be attained for some belief for which action polarization is maximized, that is, for some belief at which $(n+1)/2$ agents are voting one way and the remaining $(n-1)/2$ are voting another way. This is the case for any $\mu \in [\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$.

Due to sincere voting, the result of the election always coincides with the vote of the median voter under posterior belief μ. Accordingly, Sender's indirect utility is

$$v(\mu) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^m)\right\},$$

for any $\mu \in [0,1]$. The optimal information policy for Sender is thus supported on $\{0, \mu^W(\beta^m)\}$ whenever $\mu_0 \in]0,1/2[$, and on $\{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^m), 1[$. The posterior $\mu^W(\beta^m)$, which leads to the implementation of the proposal, belongs to the interval $[\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$ and, as such, is in the neighbourhood of the belief that maximizes polarization for any distribution of preferences. When such distribution is symmetric around the median voter, polarization is maximized exactly at the middle point in that interval, which is $\mu^W(\beta^m)$.

We illustrate Proposition 5 below in Section 5.3 in a setup with 3 voters. Fol-

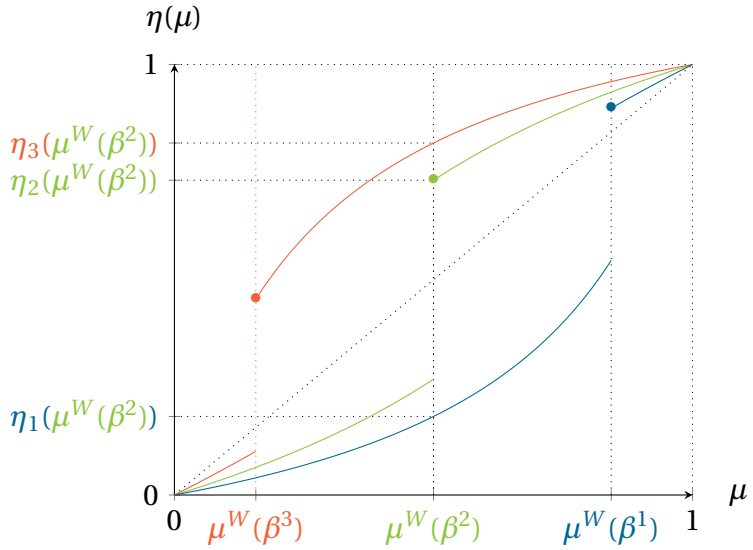


Figure 6: Beliefs distortions in the electorate for $\rho = 2$, $\beta_1 = 1/4$, $\beta_2 = 1/2$ and $\beta_3 = 3/4$. Polarization equals $\pi(\mu) = 2(\eta(\mu, \beta^1) - \eta(\mu, \beta^3))$ which is maximized at $\mu^W(\beta^2) = 1/2$.

lowing Corollary 1, wishful thinking induces voters to switch from disapproval to approval at different Bayesian posteriors $\mu^W(\beta^i)$. The optimal information policy τ for Sender is the one that maximizes the probability of the median voter voting for the approval. That is, $\text{supp}(\tau) = \{0, \mu^W(\beta^m)\}$ and $\mu^W(\beta^m) = 1/2$ is induced with probability $\tau = \mu^W(\beta^m)/\mu_0$ whenever $\mu_0 \in]0, \mu^W(\beta^2)[$ and $\text{supp}(\tau) = \{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^2), 1[$.

Let us now turn to polarization. First, it is quite easy to see in Section 5.3 that

$$\pi(\mu) = 2\left(\eta(\mu, \beta^1) - \eta(\mu, \beta^3)\right)$$

for any $\mu \in [0, 1]$, as the distances to the median belief add up to $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$. Thus, it suffices to check where $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$ is maximized. Quite naturally, polarization is maximized when the posterior belief induced by Sender is in between $\mu^W(\beta^3)$ and $\mu^W(\beta^1)$. In particular, it is exactly maximized at the posterior belief $\mu^W(\beta^2) = 1/2$ which is exactly the posterior belief Sender induces to obtain the approval of the proposal under her optimal policy.

Proposition 5 establishes that the intuition developed in this example is generally valid when the partisan preferences of voters are symmetrically distributed around

the median. In other words, attempts by a rational sender to maximize the probability of approval induces, as an externality, maximal belief polarization among wishful voters. This result differs from the literature studying the possible heterogeneity of beliefs due to deliberate attempts at persuasion which tends to focus on polarization arising from differential access to information.[23] Our model gives an alternative mechanism to the rise of polarization, based on motivated beliefs: a sender can induce polarization involuntarily when her message is subject to motivated interpretations, and such polarization might be especially large whenever sender's strategy involves targeting an agent with a median preference.

6 Conclusion

In this paper we study optimal persuasion in the presence of a wishful Receiver. By modeling wishful thinking as a process that optimally trades-off gains in anticipatory utility with the cost of distorting beliefs, we characterize the correspondence between wishful and Bayesian beliefs, highlighting the particularities that such belief formation process entails.

In particular, we show that wishful thinking impacts behavior, causing some actions to be favored in the sense that they are taken at a greater set of beliefs. This has important implications for the strategic design of information, as it adds some nuance on the way preferences and information determine behavior. Concretely, we show that, in the presence of wishful thinking, persuasion is more effective when it is aimed at inducing actions that are risky but can potentially yield a very large payoff and less effective when it is aimed at inducing more cautious actions. We use this model to illustrate why information disclosure seems less effective than expected at inducing preventive health behavior and more effective than expected at inducing dubious financial investments. Wishful thinking opens a channel for preferences to interfere in belief formation, raising the question of what kind of belief polarization could we observe in a population in which agents have access to the same information but vary in their preferences. We show in an application that an information designer interested in the approval of a proposal would, by optimally targeting the

[23] See Arieli and Babichenko (2019) for general considerations on the private persuasion of multiple receivers and see Chan et al. (2019) for an application to voting.

median voter in her choice of signal structure, induce, as an externality, maximum polarization among the electorate whenever the proposal is approved.

Some studies already investigate the effects of wishful thinking on the outcomes of strategic interactions (see, Yildiz, 2007; Banerjee et al., 2020; Heller and Winter, 2020). Further investigation on ways in which individual preferences might impact information processing and how these may impact social phenomena such as belief polarization in non-strategic and strategic settings seem to be promising paths for future research.

References

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113:96–104. 6

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, 87(4):1115–1153. 6

Akerlof, G. A. and Dickens, W. T. (1982). The Economic Consequences of Cognitive Dissonance. *American Economic Review*, 72(3):307–319. 6

Alonso, R. and Câmara, O. (2016). Persuading Voters. *American Economic Review*, 106(11):3590–3605. 4, 26

Arieli, I. and Babichenko, Y. (2019). Private Bayesian persuasion. *Journal of Economic Theory*, 182:185–217. 29

Babad, E. (1995). Can Accurate Knowledge Reduce Wishful Thinking in Voters' Predictions of Election Outcomes? *The Journal of Psychology*, 129(3):285–300. 4, 26

Babad, E. (1997). Wishful thinking among voters: motivational and cognitive Iinfluences. *International Journal of Public Opinion Research*, 9(2):105–125. 26

Babad, E., Hills, M., and O'Driscoll, M. (1992). Factors Influencing Wishful Thinking and Predictions of Election Outcomes. *Basic and Applied Social Psychology*, 13(4):461–476. 26

Babad, E. and Katz, Y. (1991). Wishful Thinking—Against All Odds. *Journal of Applied Social Psychology*, 21(23):1921–1938. 4

Banerjee, S., Davis, J., and Gondhi, N. (2020). Motivated Beliefs in Coordination Games. *SSRN Electronic Journal*. 30

Bénabou, R. (2015). The Economics of Motivated Beliefs. *Revue d'économie politique*, 125(5):665–685. 6

Bénabou, R. and Tirole, J. (2002). Self-Confidence and Personal Motivation. *Quarterly Journal of Economics*, 117(3):871–915. 5, 6

Bénabou, R. and Tirole, J. (2004). Willpower and Personal Rules. *Journal of Political Economy*, 112(4):848–886. 6

Bénabou, R. and Tirole, J. (2006). Belief in a Just World and Redistributive Politics. *Quarterly Journal of Economics*, 121(2):699–746. 6

Bénabou, R. and Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets *. *Quarterly Journal of Economics*, 126(2):805–855. 6

Bénabou, R. and Tirole, J. (2016). Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, 30(3):141–164. 2, 6

Benjamin, D., Bodoh-Creed, A., and Rabin, M. (2019). Base-Rate Neglect: Foundations and Implications. 5

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Applications and Foundations 2*, volume 2, chapter 2, pages 69–186. Elsevier B.V. 2, 5

Bergemann, D. and Morris, S. (2016). Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium. *American Economic Review*, 106(5):586–591. 5

Bergemann, D. and Morris, S. (2019). Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1):44–95. 5

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2018). Behavioral Household Finance. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 1*, chapter 3, pages 177–276. Elsevier B.V. 4

Bracha, A. and Brown, D. J. (2012). Affective decision making: A theory of optimism bias. *Games and Economic Behavior*, 75(1):67–80. 5, 9

Broniatowski, M. and Keziou, A. (2006). Minimization of ϕ-divergences on sets of signed measures. *Studia Scientiarum Mathematicarum Hungarica*, 43(4):403–442. 39

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal Expectations. *American Economic Review*, 95(4):1092–1118. 5, 6

Caplin, A. and Leahy, J. (2001). Psychological Expected Utility Theory and Anticipatory Feelings. *Quarterly Journal of Economics*, 116(1):55–79. 5

Caplin, A. and Leahy, J. (2019). Wishful Thinking. *NBER Working Paper Series*. 2, 5, 9, 10, 12

Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., and Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, 11(1):2100. 6

Chan, J., Gupta, S., Li, F., and Wang, Y. (2019). Pivotal persuasion. *Journal of Economic Theory*, 180:178–202. 29

Chandra, A., Handel, B., and Schwartzstein, J. (2019). Behavioral economics and health-care markets. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 2*, chapter 6, pages 459–502. Elsevier B.V. 3

Chew, S. H., Huang, W., and Zhao, X. (2020). Motivated False Memory. *Journal of Political Economy*, 128(10):3913–3939. 6

Coutts, A. (2019). Testing models of belief bias: An experiment. *Games and Economic Behavior*, 113:549–565. 2, 6

de Clippel, G. and Zhang, X. (2020). Non-Bayesian Persuasion. *Working Paper*. 5

Dizdar, D. and Kováč, E. (2020). A simple proof of strong duality in the linear persuasion problem. *Games and Economic Behavior*, 122:407–412. 24

Dupas, P. (2011). Health Behavior in Developing Countries. *Annual Review of Economics*, 3(1):425–449. 3

Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley. 38

Dworczak, P. and Martini, G. (2019). The Simple Economics of Optimal Persuasion. *Journal of Political Economy*, 127(5):1993–2048. 24

Egan, M., Matvos, G., and Seru, A. (2019). The Market for Financial Adviser Misconduct. *Journal of Political Economy*, 127(1):233–295. 4

Eliaz, K., Spiegler, R., and Thysen, H. C. (2021a). Persuasion with endogenous misspecified beliefs. *European Economic Review*, 134:103712. 5

Eliaz, K., Spiegler, R., and Thysen, H. C. (2021b). Strategic interpretations. *Journal of Economic Theory*, 192:105192. 5

Engelmann, J., Lebreton, M., Schwardmann, P., van der Weele, J. J., and Chang, L.-A. (2019). Anticipatory Anxiety and Wishful Thinking. *SSRN Electronic Journal*. 2, 6

Ettinger, D. and Jehiel, P. (2010). A Theory of Deception. *American Economic Journal: Microeconomics*, 2(1):1–20. 5

Exley, C. and Kessler, J. (2019). Motivated Errors. *NBER Working Paper Series*. 12

Eyster, E. (2019). Errors in strategic reasoning. In *Handbook of Behavioral Economics: Applications and Foundations 2*, volume 2, chapter 3, pages 187–259. Elsevier B.V. 5

Ganguly, A. and Tasoff, J. (2017). Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future. *Management Science*, 63(12):4037–4060. 3

Gentzkow, M. and Kamenica, E. (2014). Costly Persuasion. *American Economic Review: Papers & Proceedings*, 104(5):457–462. 18

Gentzkow, M. and Kamenica, E. (2016). A Rothschild-Stiglitz Approach to Bayesian Persuasion. *American Economic Review: Papers & Proceedings*, 106(5):597–601. 24

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, 55(1):96–135. 6

Golman, R., Loewenstein, G., Moene, K. O., and Zarri, L. (2016). The Preference for Belief Consonance. *Journal of Economic Perspectives*, 30(3):165–188. 6

Hagenbach, J. and Koessler, F. (2020). Cheap talk with coarse understanding. *Games and Economic Behavior*, 124:105–121. 5

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton University Press. 9

Heger, S. A. and Papageorge, N. W. (2018). We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology*, 67(July):177–190. 2

Heller, Y. and Winter, E. (2020). Biased-Belief Equilibrium. *American Economic Journal: Microeconomics*, 12(2):1–40. 30

Ivanov, M. (2020). Optimal monotone signals in Bayesian persuasion mechanisms. *Economic Theory*. 24, 25

Jiao, P. (2020). Payoff-Based Belief Distortion. *The Economic Journal*, 130(629):1416–1444. 2

Kamenica, E. (2019). Bayesian Persuasion and Information Design. *Annual Review of Economics*, 11:249–272. 5

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, 101(6):2590–2615. 5

Kleiner, A., Moldovanu, B., and Strack, P. (2021). Extreme Points and Majorization: Economic Applications. *Econometrica*, 89(4):1557–1593. 24

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635. 24

Kremer, M., Rao, G., and Schilbach, F. (2019). Behavioral development economics. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics: Applications and Foundations 2*, chapter 5, pages 345–458. Elsevier B.V. 3

Krizan, Z. and Windschitl, P. D. (2009). Wishful Thinking about the Future: Does Desire Impact Optimism? *Social and Personality Psychology Compass*, 3(3):227–243. 6

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53(4):636–647. 2, 6

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498. 2, 6

Le Yaouanq, Y. (2021). Motivated cognition in a model of voting. *Working Paper*. 4

Lerman, C., Hughes, C., Lemon, S. J., Main, D., Snyder, C., Durham, C., Narod, S., and Lynch, H. T. (1998). What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. *Journal of Clinical Oncology*, 16(5):1650–1654. 3

Levy, G., Moreno de Barreda, I., and Razin, R. (2018). Persuasion with Correlation Neglect. *Working Paper*. 5

Lipnowski, E., Mathevet, L., and Wei, D. (2020). Attention Management. *American Economic Review: Insights*, 2(1):17–32. 19

Loewenstein, G. (1987). Anticipation and the Valuation of Delayed Consumption. *The Economic Journal*, 97(387):666–684. 5

Mayraz, G. (2011). Wishful Thinking. *SSRN Electronic Journal*. 2

Mijović-Prelec, D. and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):227–240. 2

Mullainathan, S., Noeth, M., and Schoar, A. (2012). The Market for Financial Advice: An Audit Study. Technical report, National Bureau of Economic Research, Cambridge, MA. 4

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse Thinking and Persuasion *. *Quarterly Journal of Economics*, 123(2):577–619. 5

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease. *American Economic Review*, 103(2):804–830. 3

Panik, M. J. (1993). *Fundamentals of Convex Analysis*. Springer Netherlands, Dordrecht. 45

Saucet, C. and Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, 117:250–275. 6

Schwardmann, P. (2019). Motivated health risk denial and preventative health care investments. *Journal of Health Economics*, 65:78–92. 3

Strzalecki, T. (2011). Axiomatic Foundations of Multiplier Preferences. *Econometrica*, 79(1):47–73. 9

Thaler, M. (2020). The 'Fake News' Effect: Experimentally Identifying Motivated Reasoning Using Trust in News. *SSRN Electronic Journal*. 4, 26

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5):806–820. 2

Yildiz, M. (2007). Wishful Thinking in Strategic Environments. *Review of Economic Studies*, 74(1):319–344. 30

Appendix

A Proof for Proposition 1

Let Θ be any Polish space and let $\Delta(\Theta)$ be the set of probability measures on Θ endowed with its Borel σ-algebra, let also $\mathscr{C}_b(\Theta)$ be the set of bounded continuous and Borel-measurable real-valued functions on Θ.

For any $\eta, \mu \in \Delta(\Theta)$, by application of the Donsker-Varadhan variational formula (see Dupuis and Ellis, 1997, Lemma 1.4.3) we have

$$C(\eta,\mu) = \sup_{u(a,\cdot)\in\mathscr{C}_b(\Theta)} \int_\Theta \rho u(a,\theta)\,\eta(\mathrm{d}\theta) - \ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)\right). \tag{7}$$

Taking the Legendre-Fenchel's dual to the variational equality (7) (see Dupuis and Ellis, 1997, Proposition 1.4.2) we get

$$\ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)\right) = \sup_{\eta\in\Delta(\Theta)} \int_\Theta \rho u(a,\theta)\,\eta(\mathrm{d}\theta) - C(\eta,\mu). \tag{8}$$

Hence, we have

$$\Psi_a(\mu) = \frac{1}{\rho}\ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)\right),$$

for any $a \in A$, any $\mu \in \Delta(\Theta)$ and any $\rho \in \mathbb{R}_+^*$. Moreover, the supremum in Equation (8) is attained uniquely by the probability measure $\eta_a(\mu) \in \Delta(\Theta)$ defined by

$$\eta_a(\mu)\left(\tilde{\Theta}\right) = \frac{\int_{\tilde{\Theta}} \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)}{\int_\Theta \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)},$$

for any Borel set $\tilde{\Theta}$ (see, again, Dupuis and Ellis, 1997, Proposition 1.4.2).

In fact, we can extend the result beyond the case of the Kullback-Leibler divergence. Define the φ-divergence between η and μ as

$$\mathrm{D}_\varphi(\eta||\mu) = \int_\Theta \varphi\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right)\mu(\mathrm{d}\theta),$$

where $\varphi\colon \mathbb{R} \to \mathbb{R}_+$ is a proper, closed, convex and essentially smooth function such that $\varphi(1) = 0$ and such that its domain is an interval with endpoints $a < 1 < b$ (which

may be finite or infinite). Let us also define the Legendre-Fenchel conjugate of φ, denoted φ^*, by

$$\varphi^*(y) = \max_{x \in \mathbb{R}} xy - \varphi(x)$$

for any $y \in \mathbb{R}$. Then, the following proposition holds.

Proposition 6. *Receiver's belief motivated by action a under posterior μ uniquely satisfies*

$$\varphi'\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) = \rho u(a,\theta),$$

for any $\theta \in \Theta$, any $a \in A$ and any $\mu \in \Delta(\Theta)$, while Receiver's optimal psychological payoff equals

$$\Psi_a(\mu) = \frac{1}{\rho}\int_\Theta \varphi^*\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta),$$

for any $a \in A$ and any $\mu \in \Delta(\Theta)$.

Proof. This proposition is a direct application of Theorem 4.4 in Broniatowski and Keziou (2006). □

B Overoptimism about preferred outcomes

Fix an $a \in A$ and let Θ_a be the (measurable) set of states such that $\Theta_a = \arg\max_{\theta\in\Theta} u(a,\theta)$. Define $\delta(a,\theta) = u(a,\theta) - u(a,\theta^*)$ for all θ and some $\theta^* \in \Theta_a$. Remark that $\eta_a(\mu)(\Theta_a)$ can be expressed as follows:

$$\begin{aligned}\eta_a(\mu)(\Theta_a) &= \frac{\displaystyle\int_{\Theta_a} \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)}{\displaystyle\int_\Theta \exp\left(\rho u(a,\theta)\right)\mu(\mathrm{d}\theta)} \\ &= \frac{\mu(\Theta_a)}{\mu(\Theta_a) + \displaystyle\int_{\Theta\setminus\Theta_a} \exp(\rho\delta(a,\theta))\,\mu(\mathrm{d}\theta)}.\end{aligned}$$

Let's define the function

$$h(\rho) = \frac{\mu(\Theta_a)}{\mu(\Theta_a) + \displaystyle\int_{\Theta\setminus\Theta_a} \exp\left(\rho\delta(a,\theta)\right)\mu(\mathrm{d}\theta)}$$

for any $\rho \in \mathbb{R}_+^*$.

First, remark that $h(0) = \mu(\Theta_a)$. Moreover, by Leibniz integral rule, we have

$$h'(\rho) = \frac{-\mu(\Theta_a)}{\displaystyle\int_{\Theta \setminus \Theta_a} \delta(a,\theta) \exp\big(\rho\delta(a,\theta)\big)\, \mu(\mathrm{d}\theta)} \geq 0$$

for any $\rho \in \mathbb{R}_+^*$, since $\delta(a,\theta) \leq 0$. Finally, we also have that $\lim_{\rho \to +\infty} h(\rho) = 1$. Hence the probability of payoff maximizing states is bounded below by the Bayesian posterior $\mu(\Theta_a)$, is always increasing and is converging to 1 from below. Hence, a wishful Receiver always puts more probability mass on Θ_a than a Bayesian and eventually believes that the state belongs to Θ_a with probability 1 when ρ becomes large.

C Proof for Lemma 1

Let us study the properties of the belief threshold μ^W as a function of ρ and payoffs. First of all, let us define the function

$$\mu^W(\rho) = \frac{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1)}{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1) + \exp(\rho \overline{u}_1) - \exp(\rho \overline{u}_0)}.$$

for any $\rho \in \mathbb{R}_+^*$. To avoid notational burden, we omit the superscript W in the proof. We can find the limit of $\mu(\rho)$ at 0 by applying l'Hôpital's rule

$$\begin{aligned}
\lim_{\rho \to 0} \mu(\rho) &= \lim_{\rho \to 0} \frac{\underline{u}_0 \exp(\rho \underline{u}_0) - \underline{u}_1 \exp(\rho \underline{u}_1)}{\underline{u}_0 \exp(\rho \underline{u}_0) - \underline{u}_1 \exp(\rho \underline{u}_1) + \overline{u}_1 \exp(\rho \overline{u}_1) - \overline{u}_0 \exp(\rho \overline{u}_0)} \\
&= \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0} \\
&= \mu^B.
\end{aligned}$$

So, we are back to the case of a Bayesian Receiver whenever the cost of distortion becomes infinitely high. After multiplying by $\exp(-\rho \underline{u}_0)$ at the numerator and the denominator of $\mu(\rho)$ we get

$$\mu(\rho) = \frac{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0))}{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0)) + \exp(\rho(\overline{u}_1 - \underline{u}_0)) - \exp(\rho(\overline{u}_0 - \underline{u}_0))}.$$

So the limit of μ^W at infinity only depends on the sign of $\overline{u}_1 - \underline{u}_0$ as, by assumption, $\underline{u}_1 - \underline{u}_0 < 0$ and $\overline{u}_0 - \underline{u}_0 < 0$. Hence, $\lim_{\rho\to+\infty}\mu(\rho) = 1$ when $\overline{u}_1 - \underline{u}_0 < 0$ and $\lim_{\rho\to+\infty}\mu(\rho) = 0$ when $\overline{u}_1 - \underline{u}_0 > 0$. Finally, in the case where $\underline{u}_0 = \overline{u}_1$ we have

$$\begin{aligned}\lim_{\rho\to+\infty}\mu(\rho) &= \lim_{\rho\to+\infty}\frac{1-\exp(\rho(\underline{u}_1-\underline{u}_0))}{2-\exp(\rho(\underline{u}_1-\underline{u}_0))-\exp(\rho(\overline{u}_0-\underline{u}_0))}\\ &= \frac{1}{2}.\end{aligned}$$

Let us now check the variations of the function. After differentiating with respect to ρ and rearranging terms, one can remark that the derivative of $\mu(\rho)$ must verify the following logistic differential equation with varying coefficient

$$\mu'(\rho) = \alpha(\rho)\mu(\rho)(1-\mu(\rho)),$$

where

$$\alpha(\rho) = \frac{\underline{u}_0\exp(\rho\underline{u}_0) - \underline{u}_1\exp(\rho\underline{u}_1)}{\exp(\rho\underline{u}_0)-\exp(\rho\underline{u}_1)} - \frac{\overline{u}_1\exp(\rho\overline{u}_1) - \overline{u}_0\exp(\rho\overline{u}_0)}{\exp(\rho\overline{u}_1)-\exp(\rho\overline{u}_0)},$$

for all $\rho \in \mathbb{R}_+^*$, together with the initial condition $\mu(0) = \mu^B$. Hence, α completely dictates the variations of $\mu(\rho)$. Let us study the properties of the function α defined on $\mathbb{R}_+^*$. First, still applying again l'Hôpital's rule, its limits are given by

$$\begin{aligned}\lim_{\rho\to 0}\alpha(\rho) &= \frac{\underline{u}_0 - \overline{u}_0 - (\overline{u}_1 - \underline{u}_1)}{2}\\ &= \frac{1}{2}(u_0 - u_1)\end{aligned}$$

and

$$\begin{aligned}\lim_{\rho\to+\infty}\alpha(\rho) &= \underline{u}_0 - \overline{u}_1\\ &= u_{\max}.\end{aligned}$$

Second, after rearranging terms, its derivative is given by

$$\alpha'(\rho) = \frac{(\underline{u}_0-\underline{u}_1)^2}{\cosh(\rho(\underline{u}_0-\underline{u}_1))-1} - \frac{(\overline{u}_1-\overline{u}_0)^2}{\cosh(\rho(\overline{u}_1-\overline{u}_0))-1},$$

for any $\rho \in \mathbb{R}_+^*$, where cosh is the hyperbolic cosine function defined by

$$\cosh(x) = \frac{e^x + e^{-x}}{2},$$

for any $x \in \mathbb{R}$. Remark that the function defined by

$$f(x) = \frac{x^2}{\cosh(\rho x) - 1} \tag{9}$$

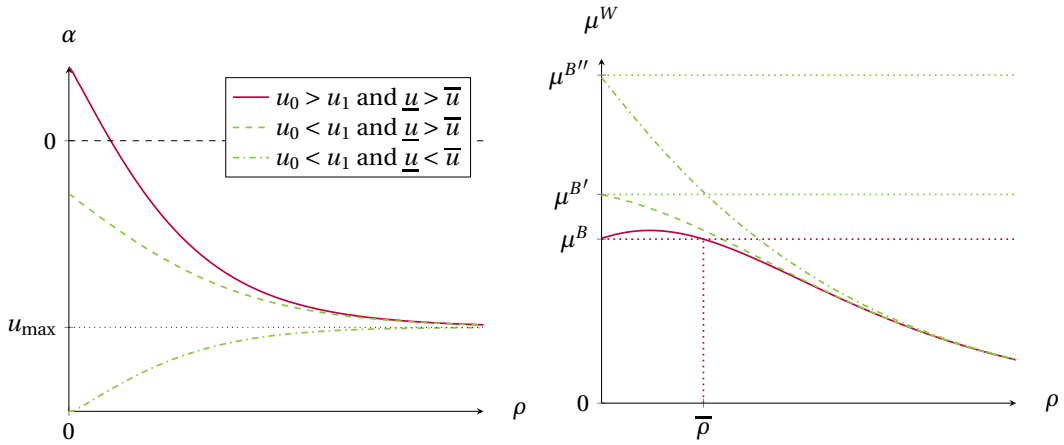
is strictly decreasing on $\mathbb{R}_+^*$. So, we have $\alpha'(\rho) < 0$ and therefore μ^W strictly decreasing for all $\rho \in \mathbb{R}_+^*$ if and only if $\underline{u}_0 - \underline{u}_1 > \overline{u}_1 - \overline{u}_0$. Accordingly, α is always a strictly monotonic function if and only if $\underline{u}_0 \neq \overline{u}_1$ and $\overline{u}_0 \neq \underline{u}_1$. Hence, excluding the extreme case where $\underline{u}_0 = \overline{u}_1$ and $\overline{u}_0 = \underline{u}_1$ so $\alpha'(\rho) = 0$ and $\mu(\rho) = \mu^B$ for all $\rho \in \mathbb{R}_+^*$, three interesting cases arise, all depicted on Figure 7 for different payoff matrices:

(i) If $u_{\max} < 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only if $u_0 < u_1$, in which case μ^W is strictly decreasing from μ^B to 0. In case $u_0 > u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly increasing and strictly decreasing toward 0.

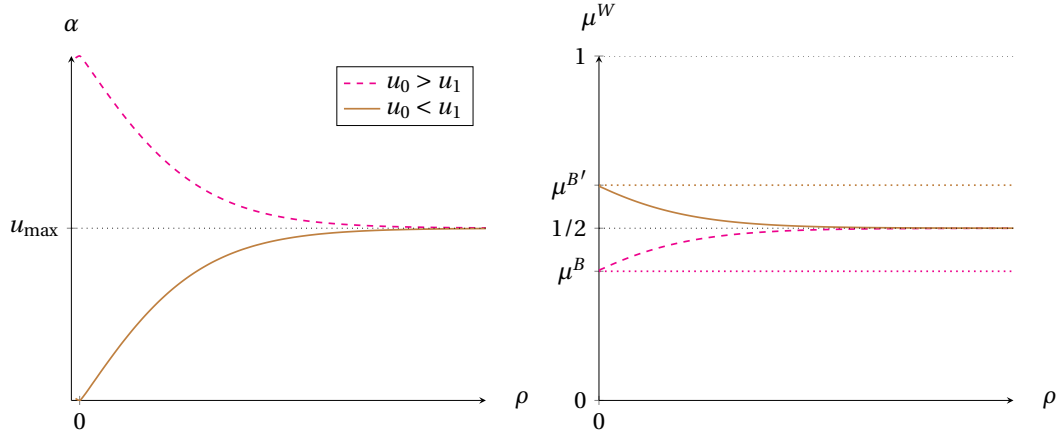
(ii) If $u_{\max} = 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$. In this case μ^W is strictly increasing from μ^B to 1/2 if and only if $u_0 > u_1$.

(iii) If $u_{\max} > 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only if $u_0 > u_1$, in which case μ^W is strictly increasing from μ^B to 1. In case $u_0 < u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly decreasing and strictly increasing toward 1.

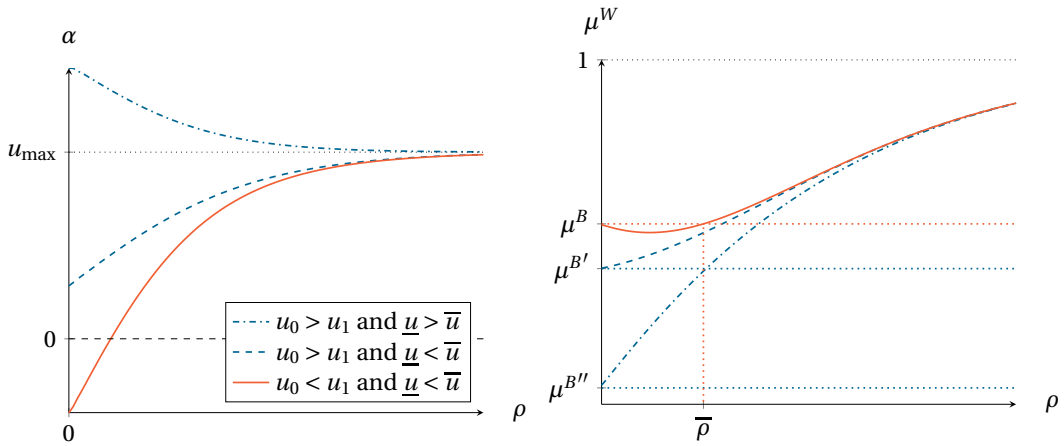
Accordingly, in case μ^W is non-monotonic in ρ, there always exists some $\overline{\rho} > 0$ such that $\mu^W(\overline{\rho}) = \mu^B$. This concludes the proof.



(a) Functions α and μ^W when $u_{\max} < 0$.



(b) Functions α and μ^W when $u_{\max} = 0$.



(c) Functions α and μ^W when $u_{\max} > 0$.

Figure 7: Functions α and μ^W for different payoff matrices $(u_a^\theta)_{a,\theta \in A \times \Theta}$. Action $a = 1$ is favored by a wishful Receiver whenever $\mu^W < \mu^B$.

D Proof for Proposition 2

Assume $|\Theta| = n$ where $2 \leq n \leq \infty$. We want to show that $\Delta_1^B \subset \Delta_1^W$ if, and only if, the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in Lemma 1 for every pair of states $\theta, \theta' \in \Theta$.

Extreme point representation for Δ_1^B and Δ_1^W. First, remark that Δ_a^B and Δ_a^W are both convex polytopes in $\mathbb{R}^{|\Theta|}$ defined by

$$\Delta_a^B = \Delta(\Theta) \cap \left\{ \mu \in \mathbb{R}^{|\Theta|} : \forall a' \in A, \sum_{\theta\in\Theta} u(a,\theta)\mu(\theta) \geq \sum_{\theta\in\Theta} u(a',\theta)\mu(\theta) \right\},$$

and

$$\Delta_a^W = \Delta(\Theta) \cap \left\{ \mu \in \mathbb{R}^{|\Theta|} : \forall a' \in A, \sum_{\theta\in\Theta} \exp\left(\rho u(a,\theta)\right)\mu(\theta) \geq \sum_{\theta\in\Theta} \exp\left(\rho u(a',\theta)\right)\mu(\theta) \right\}.$$

The sets Δ_a^B and Δ_a^B are thus compact and convex sets in $\mathbb{R}^{|\Theta|}$ with finitely many extreme points. Let us now characterize the sets of extreme points of Δ_1^B and Δ_1^W. For any $\mu \in \mathbb{R}^{|\Theta|}$, define the systems of equations

$$\mathbf{A}^B \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

and

$$\mathbf{A}^W \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

where

$$\mathbf{A}^B = \begin{pmatrix} u^B(\theta_1) & \dots & u^B(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

and

$$\mathbf{A}^B = \begin{pmatrix} u^W(\theta_1) & \dots & u^W(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

are $2\times n$ matrices, where $u^B(\theta) = u(1,\theta) - u(0,\theta)$ and $u^W(\theta) = \exp(\rho u(1,\theta)) - \exp(\rho u(0,\theta))$ for any $\theta \in \Theta$, and

$$\mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In what follows, we always assume that $(u^B(\theta))_{\theta\in\Theta}$ and $(u^W(\theta))_{\theta\in\Theta}$ are such that $\text{rank}(\mathbf{A}^B) = \text{rank}(\mathbf{A}^W) = 2$.[24] Let us recall some mathematical preliminaries.

Definition 2 (Basic feasible solution)**.** *Let $\theta,\theta' \in \Theta$ be any pair of states. A vector μ^* is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu \geq 0$, for θ,θ' if $\mathbf{A}^B \cdot \mu^* = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu^*(\theta), \mu^*(\theta') > 0$ and $\mu^*(\theta'') = 0$ for any $\theta'' \neq \theta,\theta'$.*

Lemma 2 (Extreme point representation for convex polyhedra)**.** *A vector $\mu \in \mathbb{R}^{|\Theta|}$ is an extreme point of the convex polyhedron Δ_1^B (resp. Δ_1^B) if, and only if μ is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$, $\mu \geq 0$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$, $\mu \geq 0$).*

Proof. See Panik (1993) Theorem 8.4.1. □

Therefore, to find extreme points of Δ_1^B, we just have to solve the system of equations

$$\begin{cases} \mu(\theta)\, u^B(\theta) + \mu(\theta')\, b(\theta') = 0 \\ \mu(\theta) + \mu(\theta') = 1 \\ \mu(\theta), \mu(\theta') \geq 0 \end{cases} \tag{10}$$

for any pair of states θ,θ'. When either $\mu(\theta) = 0$ or $\mu(\theta') = 0$, the solution to (10) is given by the Dirac measure δ_θ only if $u^B(\theta) \geq 0$. Denote $\mathscr{E}_1^B$ the set of such beliefs. The set $\mathscr{E}_1^B$ then corresponds to the set of degenerate beliefs under which a Bayesian Receiver would take action $a = 1$. Now, if $\mu(\theta), \mu(\theta') > 0$ then the solution to (10) is given by

$$\mu^B_{\theta,\theta'} = \frac{u(0,\theta') - u(1,\theta')}{u(0,\theta') - u(1,\theta') + u(0,\theta) - u(1,\theta)}.$$

Such a belief is exactly the belief on the edge of the simplex between δ_θ and $\delta_{\theta'}$ at which a Bayesian decision-maker is indifferent between action $a = 0$ and $a = 1$. Denote $\mathscr{I}^B$ the set of such beliefs. Hence, we have

$$\text{ext}(\Delta_1^B) = \mathscr{E}_1^B \cup \mathscr{I}^B.$$

[24] This amounts to assuming that payoff are not constant across states.

Following the same procedure, the set of extreme points of Δ_1^W is given by $\mathscr{E}_1^W \cup \mathscr{I}^W$, where $\mathscr{E}_1^W$ is the set of degenerate beliefs at which $u^W(\theta) \geq 0$ and $\mathscr{I}^W$ is the set of beliefs

$$\mu^W_{\theta,\theta'}(\rho) = \frac{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta'))}{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta')) + \exp(\rho u(0,\theta)) - \exp(\rho u(1,\theta))},$$

for any $\theta, \theta' \in \Theta$. Now, applying Krein-Milman theorem, we can state that

$$\Delta_1^B = \text{co}\left(\mathscr{E}_1^B \cup \mathscr{I}^B\right)$$

and

$$\Delta_1^W = \text{co}\left(\mathscr{E}_1^W \cup \mathscr{I}^W\right)$$

Sufficiency. Assume the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in Lemma 1 for every pair of states $\theta, \theta' \in \Theta$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$. This implies $\mathscr{I}_1^B \subset \Delta_1^W$, since action $a = 1$ is favored by a wishful Receiver on each edge of the simplex. Moreover, it is trivially satisfied that $\mathscr{E}_1^B = \mathscr{E}_1^W$. Hence, since any point in Δ_1^B can be written as a convex combination of points in $\mathscr{E}_1^B \cup \mathscr{I}^B \subset \Delta^W$, it follows that $\Delta_1^B \subset \Delta_1^W$.

Necessity. Assume now that $\Delta_1^B \subset \Delta_1^W$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$ which implies that $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in Lemma 1 for every pair of states $\theta, \theta' \in \Theta$.

E Proof for Proposition 5

First, note that we can always index the voters in an ascending order of β, such that $\eta(\mu, \beta^i) \geq \eta_j(\mu)$ for all $\mu \in \Delta(\Theta)$ whenever $i < j$, such that

$$\pi(\mu) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta(\mu, \beta^i) - \eta(\mu, \beta^j)$$

does indeed represent the absolute difference between each pair of beliefs. Now, remark that the sum can be rearranged in the following way:

$$\begin{aligned}\pi(\mu) &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \eta(\mu,\beta^i) - \eta(\mu,\beta^j) \\ &= (n-1)\eta^1(\mu) + (n-2)\eta^2(\mu) - \eta^2(\mu) + \\ &\quad \cdots + \frac{n-1}{2}\eta(\mu,\beta^m) - \frac{n-1}{2}\eta(\mu,\beta^m) + \cdots + \\ &\quad \eta(\mu,\beta^{n-1}) - (n-2)\eta(\mu,\beta^{n-1}) - (n-1)\eta^n(\mu) \\ &= \sum_{i=1}^{m}(n+1-2i)(\eta(\mu,\beta^i) - \eta(\mu,\beta^{n+1-i})),\end{aligned}$$

for any $\mu \in [0,1]$, where $m = (n+1)/2$. That is, we can express it in terms of the differences in beliefs among voters who are equidistant from the median. To see that this is true, we need to first realize that each belief appears $n-1$ times in Equation (6) (since each belief is paired once with each of the other $n-1$ beliefs). The beliefs of voters below the median appear more often as positive than negative (the belief of the first voter is positive in all of its pairings, the belief of the second voter is positive in all of its pairing except for the pairing with the first voter, etc.), whereas the beliefs of voters above the median are more often negative than positive. If we rearrange the terms of the sum in order to pair symmetric voters, the term $(\eta(\mu,\beta^1) - \eta_n(\mu))$ appears $n-1$ times, whereas the term $(\eta_2(\mu) - \eta(\mu,\beta^{n-1}))$ appears $n-3$ times, since out of the $n-1$ times $\eta_2(\mu)$ appears on Equation (6), $n-2$ of them are positive and 1 is negative (the converse is true for $\eta(\mu,\beta^{n-1})$). One can continue the same reasoning for all the pairs of symmetric voters, and get to the formulation of $\pi(\mu)$ presented above. Note, also, that the belief of the median voter is summed and subtracted at the same rate, such that it does not matter in our measure of polarization.

Consider the distance between beliefs of any pair of symmetric voters $\eta(\mu,\beta^i) - \eta(\mu,\beta^{n+1-i})$ for $i \in \{1,\ldots,m\}$. Given our symmetry assumption these two agents are symmetric, such that $\beta^i = 1-\beta^{n+1-i}$. It is not difficult to show that any of those pairwise distances is maximized when agent i is distorting its belief upwards and agent $n+1-i$ is distorting its belief downwards. That is, when $\mu \in [\mu^W(\beta^i), \mu^W(\beta^{n+1-i})]$.

First, the distance between symmetric beliefs in such an interval can be rewritten

as

$$\eta(\mu,\beta^i)-\eta(\mu,\beta^{n+1-i})=\frac{\mu\exp(\rho\beta^i)}{\mu\exp(\rho\beta^i)+(1-\mu)}-\frac{\mu}{\mu+(1-\mu)\exp(\rho\beta^i)}.$$

for any $i\in\{1,\dots,m\}$ and $\mu\in[\mu^W(\beta^i),\mu^W(\beta^{n+1-i})]$.

Second, by taking the first order condition in this interval and rearranging it we get

$$\frac{\mu+(1-\mu)\exp(\rho\beta^i)}{\mu\exp(\rho\beta^i)+(1-\mu)}=1,$$

such that the difference between symmetric beliefs is maximized uniquely at

$$\mu=\mu^W(\beta^m)=\frac{1}{2},$$

for any $i\in\{1,\dots,m\}$, $\beta^i\in\,]0,1[$ and any $\rho\in\mathbb{R}_+^*$. Since

$$\mu^W(\beta^m)=\underset{\mu\in[0,1]}{\arg\max}\,\eta(\mu,\beta^i)-\eta(\mu,\beta^{n+1-i})$$

for any $i\in\{1,\dots,m\}$, we get

$$\mu^W(\beta^m)=\underset{\mu\in[0,1]}{\arg\max}\,\pi(\mu),$$

which concludes the proof.

F Proof for Proposition 4

First, we define the function

$$\psi(z)=\frac{1}{1-F(z)}\int_z^{\overline{\theta}}\exp(\rho\theta)f(\theta)\,\mathrm{d}\theta,$$

for any $z\in[\underline{\theta},\overline{\theta}[$ and adopt the convention that $\psi(\overline{\theta})=\exp(\rho\overline{\theta})$. It is not difficult to show that ψ is a continuous and strictly increasing function from $\psi(\underline{\theta})=\hat{x}<1$ to $\psi(\overline{\theta})=\exp(\rho\overline{\theta})$. Define similarly the function

$$\varphi(z)=\frac{1}{1-F(z)}\int_z^{\overline{\theta}}\theta f(\theta)\,\mathrm{d}\theta,$$

for any $z \in [\underline{\theta}, \overline{\theta}[$ and $\varphi(\overline{\theta}) = \overline{\theta}$. Again, it is not difficult to show that φ is a continuous and strictly increasing function from $\varphi(\underline{\theta}) = \hat{m} < 0$ to $\varphi(\overline{\theta}) = \overline{\theta}$.

Since ψ is strictly increasing, it thus suffices to show that $\psi(\theta^B) > 1 = \psi(\theta^W)$ to prove that $\theta^W < \theta^B$. Applying Jensen's inequality, it follows that

$$\psi(z) > \exp(\rho\varphi(z)),$$

for any $z \in]\underline{\theta}, \overline{\theta}[$, where the strict inequality comes from the strict convexity of $z \mapsto \exp(\rho z)$ and the non degeneracy of F. In particular, Jensen's inequality holds with equality at $\underline{\theta}$ and $\overline{\theta}$, but, by the intermediate value theorem, it must be that θ^B (as well as θ^W) lie in the open interval $]\underline{\theta}, 0[$. Thus, we have

$$\psi(\theta^B) > 1,$$

since $\varphi(\theta^B) = 0$ and $\theta^B \neq \underline{\theta}, \overline{\theta}$.