Biased-Belief Equilibrium

Yuval Heller[*] Eyal Winter[†]

April 30, 2017

Abstract

We investigate how distorted, yet structured beliefs, emerge in strategic situations. Specifically, we study two-player games in which each player is endowed with a biased-belief function that represents the discrepancy between a player's beliefs about the opponent's strategy and the actual strategy. Our equilibrium condition requires that: (1) each player chooses a best response strategy to his distorted belief about the partner's strategy, and (2) the distortion functions form best responses to one another, in the sense that if one of the players is endowed with a different distortion function, then that player is outperformed in the game induced by this new distortion function. Our analysis characterizes equilibrium outcomes and identify the belief biases that support these equilibrium outcomes in different strategic environments.

JEL classification: C73, D03, D83.

1 Introduction

Standard models of equilibrium behavior attribute players with perfect rationality at two different levels: beliefs and actions. Players are assumed to form beliefs that are consistent with reality and to choose actions that maximize their utility given the beliefs that they hold. Much of the literature in behavioral and experimental economics that document violations of the rationality assumption at the level of beliefs is ascribing these violations to cognitive limitations. However, in interactive environments where one person's beliefs affect other persons' actions, beliefs' distortions are not arbitrary, and they may arise to serve some strategic purposes (see, e.g., the self-serving biases

[*]Department of Economics, Bar Ilan University, Israel. yuval.heller@biu.ac.il. URL: https://sites.google.com/site/yuval26/. The author is grateful to the European Research Council for its financial support (starting grant #677057).

[†]Center for the Study of Rationality and Department of Economics, Hebrew University of Jerusalem, Israel. mseyal@mscc.huji.ac.il . URL: http://www.ma.huji.ac.il/∼mseyal/. The author is grateful to the German Israeli Foundation for Scientific Research and to Google for their financial support.

analyzed in Babcock and Loewenstein, 1997, and the analysis of frustration and anger in Battigalli et al., 2015).

In this paper we investigate how distorted, yet structured beliefs, emerge in strategic situations. Our basic assumption here is that distorted beliefs often emerge because they offer a strategic advantage to those who are holding them even when these beliefs are wrong. More specifically, players often hold distorted beliefs as a form of a commitment device that affects the behavior of their counterparts. The precise cognitive process that is responsible for the formation of beliefs is complex, and it is out of the scope of this paper to draw it. We believe, however, that, in addition, to analytic assessment of evidence, preferences in the form of desires, fears and other emotions contribute to the process, and, to an extent, facilitate beliefs' biases. If the evidence is unambiguous and decisive, or if the consequence of beliefs' distortion is detrimental to the player's welfare, preferences may play less of a role and learning may work to calibrate beliefs to reality. But when beliefs are biased in ways that favor their holders by affecting the behavior of their counterparts, learning can actually reinforce biases rather than diminishing them.

Standard equilibrium notions in game theory draw a clear line between preferences and beliefs. The former are exogenous and fixed, the latter can be amended through Bayesian updating but are not allowed to be affected by preferences. However, phenomena such as wishful-thinking and overconfidence, where beliefs are tilted towards what their holder desires reality to be, suggest that in real life, beliefs and preferences can intermingle. Similarly, beliefs' rigidity and belief polarization (see, e.g., Lord et al., 1979; Ross and Anderson, 1982) refer to situations in which two people with conflicting prior beliefs both strengthen their beliefs in response to observing the same data. The parties' aversion to depart from their original beliefs can also be regarded as a form of preferences and beliefs intermingling.

It is easy to see how the belief biases described above can have strategic benefits in interactive situations. Wishful thinking and optimism can facilitate cooperation in interactions that require mutual trust. Overconfidence can deter competitors, and beliefs' rigidity may allow an agent to sustain a credible threat. An important objective of our analysis would be to identify the strategic environments that support biases such as wishful thinking and overconfidence as part of equilibrium behavior. It is worthwhile noting that not only individuals are susceptible to beliefs biases that are strategically motivated. Governments are prone to be affected by such biases as well. Bush administration's unsubstantiated confidence regarding Saddam Hussein's possession of "weapon of mass destruction" prior to the second Gulf war and the vast discrepancy between Israeli and US intelligence assessments regarding Iran's Nuclear intentions prior to the signing of the Iran nuclear deal can be easily interpreted as strategically motivated beliefs' distortion.

For biased beliefs to yield a strategic advantage to a player holding them, it is essential that counterparts to the interaction regard them as credible and believe that the player will act upon them.

A different body of empirical evidence consistent with strategic beliefs is offered by the Psychiatric literature on "Depressive Realism" (e.g., Dobson and Franche, 1989). This literature compares probabilistic assessments conveyed by psychiatrically healthy people with those suffering from clinical depression. Participants of both categories were requested to assess the likelihood of experiencing negative or positive events in both public and private setups. Comparing subjects' answers with the objective probabilities of these events revealed that in a public setup clinically depressed individuals were more realistic than their healthy counterparts for both types of events. The apparent belief bias among healthy individuals can be reasonably attributed to the strategic component of beliefs. Mood disorders negatively affect strategic reasoning (Inoue et al., 2004), which, to a certain extent, may diminish strategic belief distortion among clinically depressed individuals relative to their healthy counterparts.

The formation of biased beliefs and the process by which they are held credible by counterparts are two sides of the same coin. For the sake of tractability, we shall avoid specifying a concrete dynamic model that describes these processes. Instead, we shall adopt a static approach by imposing equilibrium conditions on players' beliefs and their interpretation by counterparts. This static approach is consistent with a large part of the literature on endogenous preferences (see, e.g., (Guth and Yaari, 1992; Dekel et al., 2007; Friedman and Singh, 2009; Herold and Kuzmics, 2009; Winter et al., forthcoming; Heller and Winter, 2016; Heller and Sturrock, 2017)). Nevertheless, we mention a few mechanisms that can facilitate these processes and turn biased beliefs into a credible commitment device:

1. Refraining from accessing information or using biased sources of information, e.g. subscribing to a newspaper with a specific political orientation or consulting with biased experts.

2. Following passionately a religion, a moral principle or an ideology that have belief implications regarding human behavior.

3. Possessing personality traits that have implications on beliefs (e.g. narcissism or naivety).

The mechanisms described above are not only likely to induce beliefs' biases, but are also prone to generate signals sent to the player's counterparts regarding these biases with a certain degree of verifiability. These mechanisms, the signals they induce and their interpretation are the main forces that facilitate biased belief equilibria.

Our notion of biased belief equilibrium uses a two-stage paradigm. In the first stage each player is endowed with a biased-belief function. This function represents the discrepancy between a player's beliefs about the strategy profile of others players and the actual profile. In the second stage each player chooses a best response strategy to his distorted belief about the partner's strategy (the chosen strategy profile is referred to as the equilibrium outcome). Finally, our equilibrium condition require that the distortion functions are not arbitrary, but form best responses to one another in the following sense. If one of the players is endowed with a different distortion function, then there exists an equilibrium of the induced biased game in which this player is outperformed. The stronger refinement of strong biased belief equilibrium requires a player endowed with a different biased-belief function to be outperformed in *all* equilibria of the induced biased game.

Our analysis of biased belief will go beyond characterizing equilibrium outcomes. An additional important objective is to identify the belief biases that support these equilibrium outcomes in different strategic environments. Central to our analysis will be belief distortion properties such as "wishful thinking" and "pessimism" that sustain biased belief equilibria in different strategic environments.

1.1 Summary of Main Results

We begin our analysis by studying the relations between biased-belief equilibrium outcomes and Nash equilibria. We show that any Nash equilibrium can be implemented as the outcome of a biased-belief equilibrium, though in some cases this requires at least one of the players to have a distorted belief about the opponent's strategy. This, in particular, implies that every game admits a biased-belief equilibrium. Next, we show that introducing biased beliefs do not change the set of equilibrium outcomes in games in which at least one of the players have a dominant action. In contrast, biased-belief equilibrium admits non-Nash behavior in most other games, including games in which both players always obtain identical payoffs.

Next we characterize the set of biased-belief equilibrium outcomes. We present two necessary conditions for a strategy profile to be a biased-belief equilibrium in any game: (1) no player uses a strictly dominated strategy, and (2) the payoff of each player is above the minmax payoff of the player while restricting the players to choose only undominated strategies (i.e., strategies that are not strictly dominated). We examine these condition by focusing on two classes of games: (1) games with two actions for each player, and (2) games in which the set of actions is an interval, and the payoff function is "well-behaved". We show that in those classes of games the above two conditions fully characterize the set of biased-belief equilibrium outcomes.

In Section 5 we focus on a class of games with strategic complementarity and spillovers (see, e.g., Bulow et al., 1985; Cooper and John, 1988), such as, input (or partnership) games, and price competition with differentiated goods. We show that in this class of games, there is a close relation between implementing outcomes that Pareto-dominate all Nash equilibria and wishful thinking (Babad and Katz, 1991; Budescu and Bruderman, 1995). We say that a biased belief exhibits wishful thinking if it distorts the perceived opponent's strategy in a way that yields the player a higher payoff relative to the payoff induced by the true strategy of the opponent. We show that any strategy profile in which both players use undominated strategies, and achieve a payoff higher than their best Nash equilibrium payoff, can be implemented as the outcome of biased-belief equilibria exhibiting wishful thinking, and, moreover, such a strategy profile can be implemented only by this kind of biased-belief equilibria.

Our final result shows an interesting class of biased-belief equilibria that exist in all games. We say that a strategy is undominated Stackelberg if it maximizes a player's payoff in a setup in which the player can commit to an undominated strategy, and his opponent reacts by best-replying to this strategy. We show that every game admits a biased-belief equilibrium in which one of the players is "rational" in the sense of having a constant belief about the opponent's strategy, and always playing his undominated Stackelberg strategy, while the opponent is "flexible" in the sense of having undistorted beliefs and best-replying to the player's true strategy.

We conclude by presenting additional examples of interesting biased-belief equilibria in three specific families of games: (1) Prisoner's Dilemma with an additional "withdrawal" action, (2) the Centipede, and (3) the traveler's dilemma.

The structure of this paper is as follows. Section 2 describes the model. In Section 3 we analyze the relations between biased-belief equilibria and Nash equilibria. In Section 4 we characterize the set of biased-belief equilibrium outcomes. Section 5 focuses on games with strategic complementarity and show the close relations between "good" biased-belief equilibrium outcomes and wishful thinking. In Section 6 we study the relation between biased-belief equilibrium and strategies played by a Stackelberg leader. Finally, we present a few additional examples of interesting biased-belief equilibrium in Section 7. We conclude with a discussion in Section 8.

2 Model

2.1 Underlying game

Let $i \in \{1, 2\}$ be an index used to refer to one of the players (he) in a two-player game, and let j be an index referring to the opponent (she). Let $G = (S, \pi)$ be a normal-form two-player game (henceforth, *game*), where $S = (S_1, S_2)$ and each S_i is a convex closed set of strategies. We denote by $\pi = (\pi_1, \pi_2)$ players' payoff functions, i.e. $\pi_i : S \to \mathbb{R}$ is function assigning each player a payoff for each strategy profile. We use s_i to refer to a typical strategy of player i. We assume each payoff function $\pi_i (s_i, s_j)$ to be twice differentiable in both parameters and weakly concave in the first parameter (s_i).

In most of the examples and applications presented in the paper, the set of strategies is, either:

1. a simplex over a finite set of pure actions, where each strategy corresponds to a mixed action (i.e., A_i is a finite set of pure actions, and $S_i = \Delta (A_i)$), and the vN-M payoff function is linear with respect to the mixing probability, or

2. an interval in $\mathbb{R}$ (e.g., each player chooses a real number representing quantity, price or effort).

Let BR (resp., BR^{-1}) denote the (inverse) best reply correspondence, i.e.,

$$BR\left(s_i\right) = \left\{s_j \in S_j | s_j = argmax_{s'_j \in S'_j} \left(\pi_j\left(s_i, s'_j\right)\right)\right\}$$

is the set of best replies against strategy $s_i \in S_i$, and

$$BR^{-1}\left(s_i\right) = \left\{s_j \in S_j | s_i = argmax_{s'_i \in S_i}\left(\pi_i\left(s'_i, s_j\right)\right)\right\}$$

is the set of strategies for which s_i is a best-reply against them.

2.2 Biased-Belief Function

We start here with the definition of biased belief functions that describe how players' beliefs are distorted. A *biased belief* $\psi_i : S_j \to S_j$ is a *continuous* function that assigns for each strategy of the opponent, a (possibly distorted) belief about the opponent's play. That is, if the opponent plays s_j , then player i believes that the opponent plays $\psi_i (s_j)$. We call s_j the opponent's real strategy, and we call $\psi_i (s_j)$ the opponent's perceived (or biased) strategy. Let I_d be the undistorted (identity) function, i.e., $I_d (s) = s$ for each strategy s. A biased belief ψ is *blind* if the perceived opponent's strategy is independent of the opponent's real strategy, i.e., if $\psi (s_j) = \psi \left(s'_j\right)$ *for each* $s_j, s'_j \in S_j$.

With a slight abuse of notation we use s_i to denote also the blind biased belief ψ_j that is always equal to s_i.

A biased game is a pair consisting of an underlying game and a profile of biased beliefs. Formally:

Definition 1. A *biased game* (G, ψ) is a pair where $G = (S, \pi)$ is a normal-form two-player game, and $\psi = (\psi_1, \psi_2)$ is a pair of biased beliefs.

A pair of strategies is a Nash equilibrium of a biased game, if each strategy is a best reply against the perceived strategy of the opponent. Formally,

Definition 2. A profile of strategies $s = (s_1^*, s_2^*)$ is a *Nash equilibrium of the biased game* (G, ψ) if each s_i^* is a best reply against the perceived strategy of the opponent, i.e., $s_i^* = argmax_{s_i \in \Delta(A_i)} \left(\pi_i \left(s_i, \psi_i \left(s_j^*\right)\right)\right)$.

Let $NE(G, \psi) \subseteq S_1 \times S_2$ denote the set of all Nash equilibria of the biased game (G, ψ).

A standard argument relying on Kakutani fixed-point theorem implies that any biased game (G, ψ) admits a Nash equilibrium (i.e., that $NE(G, \psi) \neq \emptyset$).

2.3 Biased-Belief Equilibrium

We are now ready to define our equilibrium concept. A biased-belief equilibrium is a pair consisting of a profile of biased beliefs and a profile of strategies, such that: (1) each strategy is a best reply to the perceived strategy of the opponent, and (2) each biased belief is a best reply to the partner's biased belief, in the sense that any agent who chooses a different biased-belief function is outperformed in at least one equilibrium in the new biased game (relative to the agent's payoff in the original equilibrium). The refinement of biased-belief equilibrium requires that such a deviator is outperformed in all equilibria of the induced biased game. Formally:

Definition 3. A *biased-belief equilibrium* (abbr., BBE) is a pair (ψ^*, s^*), where $\psi^* = (\psi_1^*, \psi_2^*)$ is a profile of biased beliefs and $s^* = (s_1^*, s_2^*)$ is a profile of strategies satisfying: (1) $\left(s_i^*, s_j^*\right) \in NE(G, \psi^*)$, and (2) for each player i and each biased belief ψ_i', there exists a strategy profile $\left(s_i', s_j'\right) \in NE\left(G, \left(\psi_i', \psi_j^*\right)\right)$, such that the following inequality holds: $\pi_i\left(s_i', s_j'\right) \leq \pi_i\left(s_i^*, s_j^*\right)$. A biased-belief equilibrium is *strong* if the inequality $\pi_i\left(s_i', s_j'\right) \leq \pi_i\left(s_i^*, s_j^*\right)$ holds for every strategy profile $\left(s_i', s_j'\right) \in NE\left(G, \left(\psi_i', \psi_j^*\right)\right)$.

It is immediate that any strong biased-belief equilibrium is a biased-belief equilibrium.

Strategy profile $s^* = (s_1^*, s_2^*)$ is a *(strong) biased-belief equilibrium outcome* if there exist a profile of biased beliefs $\psi^* = (\psi_1^*, \psi_2^*)$ such that (ψ^*, s^*) is a (strong) biased-belief equilibrium. In this case we say that the biased belief ψ^* supports (or implements) the outcome s^*.

3 Biased-Belief Equilibrium Outcomes and Nash Equilibria

In this section we present a few results that relate between Nash equilibria of the underlying game and biased-belief equilibria.

3.1 Nash Equilibria and Distorted Beliefs

We begin with a simple observation that shows that in any biased-belief equilibrium in which the outcome is not a Nash equilibrium, at least one of the players must distort the opponent's perceived strategy. The reason for this observation, is that if both players have undistorted beliefs, then it must be that each agent best replies to the partner's strategy, which implies that the outcome is a Nash equilibrium of the underlying game.

The following example demonstrates that some Nash equilibria cannot be supported as the outcomes of biased-belief equilibria with undistorted beliefs.

Example 1 (Cournot Equilibrium cannot be supported by undistorted beliefs, yet it can be supported by blind beliefs)**.** Consider the following symmetric Cournot game $G = (S, \pi)$: $S_i = [0, 1]$ and $\pi_i(s_i, s_j) = s_i \cdot (1 - s_i - s_j)$ for each player i. The interpretation of the game is as follows. Each s_i is interpreted as the quantity chosen by firm i, the price of both goods is determined by the linear inverse demand function $p = 1 - s_i - s_j$, and the marginal cost of each firm is normalized to be zero. The unique Nash equilibrium of the game is $s_i^* = s_j^* = \frac{1}{3}$, which yields both players a payoff of $\frac{1}{9}$. Assume to the contrary that this outcome can be supported as a biased-belief equilibrium by the undistorted beliefs $\psi_i^* = \psi_j^* = I_d$. Consider a deviation of player 1 to the blind belief $\psi_1' \equiv \frac{1}{4}$ (i.e., the strategy of the follower in a sequential Stackelberg game). The unique equilibrium of the biased game $\left(G, \left(\frac{1}{4}, I_d\right)\right)$is $s_1' = \frac{1}{2}$, $s_2' = \frac{1}{4}$, which yields the deviator a payoff of $\frac{1}{8} > \frac{1}{9}$.
The unique Nash equilibrium $s_i^* = s_j^* = \frac{1}{3}$ can be supported as the outcome of the strong biased-belief equilibrium $\left(\left(\frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{1}{3}\right)\right)$ with blind beliefs., in which each agent believes the opponent playing $\frac{1}{3}$ regardless of the opponent's actual play, and the agent plays the unique best reply to this belief, which is the strategy $\frac{1}{3}$.

3.2 Any Nash equilibrium is a BBE outcome

The following result generalizes the second part of Example 1, and shows that any Nash (Strict) equilibrium is an outcome of a (strong) biased-belief equilibrium in which both players have blind beliefs.

Proposition 1. *Let* (s_1^*, s_2^*) *be a Nash (strict) equilibrium of the game* $G = (S, \pi)$*. Let* $\psi_1^* \equiv s_2^*$ *and* $\psi_2^* \equiv s_1^*$*. Then* $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ *is a (strong) biased-belief equilibrium.*

Proof. The fact that (s_1^*, s_2^*) is a Nash equilibrium of the underlying game implies that (s_1^*, s_2^*) is an equilibrium of the biased game $(G, (\psi_1^*, \psi_2^*))$. The fact, that the beliefs are blind, implies that for any biased belief ψ_i' there is an equilibrium in the biased game $\left(G, \left(\psi_i', \psi_j^*\right)\right)$ in which player j plays s_j, and that player i gains at most $\pi_i\left(s_i^*, s_j^*\right)$, which implies that $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is a biased-belief equilibrium. Moreover, if (s_1^*, s_2^*) is a strict equilibrium, then in any equilibrium of any biased game $\left(G, \left(\psi_i', \psi_j^*\right)\right)$, player j plays s_j^*, and that player i gains at most $\pi_i\left(s_i^*, s_j^*\right)$, which implies that $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is a strong biased-belief equilibrium. □

An immediate corollary of Prop. 1 is that every game admits a biased-belief equilibrium.

Corollary 1. *Every game admits a biased-belief equilibrium.*

4 Characterization of BBE Outcomes

We begin by presenting two necessary conditions for a strategy profile to be a biased-belief outcome in all games. The following sections focus on specific classes of games, and fully characterize biased-belief equilibrium outcomes in these games.

4.1 Necessary Conditions for Being a BBE Outcome in all Games

Recall, that a strategy s_i of player i is strictly dominated if there exists another strategy s_i' of player i, such that $u_i(s_i, s_j) < u_i(s_i', s_j)$ for each strategy s_j of player j.

We say that a strategy is *undominated* if it is not strictly dominated. We say that a strategy profile is *undominated* if both strategies in the profile are undominated. We say that an undominated strategy profile (s_1^*, s_2^*) is an *undominated Pareto-optimal* profile if $\pi_i(s_1^*, s_2^*) \geq \pi_i(s_1', s_2')$ for each undominated strategy profile (s_1', s_2') and each player i.

Let $S_i^U \in S_i$ denote the *set of undominated strategies* of player i. Note, that S_i^U is not necessarily a convex set.

An undominated minmax payoff for player i is the maximal payoff player$i$ can guarantee to himself in the following process: (1) player j chooses an arbitrary *undominated* strategy, and (2) player i chooses a strategy (after observing player j's strategy). Formally:

Definition 4. Given game $G = (A, u)$, let M_i^U, the *undominated minmax* payoff of player i, be defined as follows:

$$M_i^U = \min_{s_j \in S_j^U} \left(\max_{s_i \in S_i} \pi_i \left(s_i, s_j\right) \right).$$

Observe that the undominated minmax is weakly larger than the standard maxmin, i.e., $M_i^U \geq \min_{s_j \in S_j} \left(\max_{s_i \in S_i} \pi_i \left(s_i, s_j\right)\right)$,with an equality if player j does not have any strictly dominated strategy (i.e., if $S_j^U = S_j$).[1]

Observe that $BR^{-1}\left(s_i\right) \neq \emptyset$ iff $s_i \in S_i^U$. The following simple result shows that any biased-belief equilibrium outcome is an undominated strategy profile that induces each player a payoff above the player's undominated minmax payoff. Moreover, within the family of interval games with monotone externalities, any undominated strategy profile with this property is the outcome of a strong biased-belief equilibrium.

Proposition 2. *If a strategy profile $s^* = \left(s_1^*, s_2^*\right)$ is a biased-belief equilibrium outcome, then (1) the profile s^* is undominated; and (2) $\pi_i\left(s^*\right) \geq M_i^U$.*

Proof. Assume that $s^* = \left(s_1^*, s_2^*\right)$ is a biased-belief equilibrium outcome. This implies that each s_i^* is a best reply to the player's distorted belief, which implies that each s_i^* is undominated. Assume to the contrary, that , $\pi_i\left(s^*\right) < M_i^U$. Then, by deviating to the undistorted function I_d, player i can guarantee a fitness of at least M_i^U in any distorted equilibrium. □

4.2 Zero-Sum Games, Dominant Strategies and Doubly Symmetric Games[2]

In this section we characterize biased-belief equilibrium outcomes in three classes of games: zero-sum games, games with a dominant action, and doubly symmetric games.

Zero-sum games. Recall that a game is *zero sum* if there exists $c \in R^+$ such that $\pi\left(s_i, s_j\right) = c$ for each strategy profile $\left(s_i, s_j\right) \in S$. It is immediate that the undominated minmax of a zero sum game coincides with the game's unique value. Thus, Proposition 2 implies that introducing biased beliefs to zero-sum games do not affect the equilibrium payoff.

Corollary 2. *The unique Nash equilibrium payoff of a zero-sum game is also the unique payoff in any biased-belief equilibrium.*

[1]Note, that the undominated minmax payoff might be strictly higher than the undominated *maxmin* payoff due to the non-convexity of S_U^i, i.e., player i might be able to guarantee only a lower payoff if player j would choose his undominated strategy after observing player i's chosen strategy.

[2]Y: Internal comment: I have tried the current location for these results, as your comment to the previous revision suggested they shouldn't be in Section 3, but later in the paper.

Games with a dominant strategy. Next we show that if at least one of the players has a dominant strategy, then any biased-belief equilibrium outcome must be a Nash equilibrium. Formally:

Proposition 3. *If a game admits a dominant strategy s_i^* for player i, then any biased-belief equilibrium outcome is a Nash equilibrium of the underlying game.*

Proof. Observe that s_i^* is the unique best-reply of player i to any perceived strategy of player j, and, as a result, player i plays the dominant action s_i^* in any biased-belief equilibrium. Assume to the contrary that there is a biased-belief equilibrium outcome in which player j does not best-reply against s_i^*. Consider a deviation of player j of choosing the undistorted belief I_d. Observe, that player i still plays his dominant action s_i^*, and that player j best-replies to s_i^* in any Nash equilibrium of the induced biased game, and as a result, player j achieves a strictly higher payoff, and we get a contradiction. $\square$

Doubly symmetric games. One might expect that biased beliefs do not play an essential role in games, in which the interests of both players perfectly coincide (as, is the case in related equilibrium notions in the literature, such as, the notion of rule-rational equilibrium in Heller and Winter, 2016). The following example shows that this is not the case, and that biased-beliefs equilibrium outcomes might differ from Nash equilibria even in *doubly symmetric* games (e.g., Weibull, 1997, Def. 1.11) in which $\pi(s_i, s_j) = \pi(s_j, s_i)$ for each strategy profile $(s_i, s_j) \in S$.

Example 2 (Doubly symmetric game with a non-Nash biased-belief equilibrium outcome)**.** Consider the doubly symmetric game presented in Table 1. In this example we show that the non-Nash strategy profile (a, a) is a biased-belief equilibrium outcome. We denote each strategy (=mixed action) $s_i \in S_i = \Delta(\{a, b, c, d\})$ as a vector $(\alpha, \beta, \gamma, \delta)$ (with $\alpha + \beta + \gamma + \delta = 1$), where α (resp., β, γ, δ) denotes the probability of playing action a, (resp., b, c, d), and we identify each action with the degenerate strategy assigning a mass of one to this action (i.e., $a \equiv (1, 0, 0, 0)$. Let ψ_i^* be the following (continuous) biased-belief function: $\psi_i^*(\alpha, \beta, \gamma, \delta) = (0, 0, \alpha, \beta + \gamma + \delta)$. We show that $((\psi_1^*, \psi_2^*), (a, a))$ is a biased-belief equilibrium. Observe first that a is a best-reply to the opponent's perceived strategy (c), i.e., $(a, a) \in NE(G, (\psi_1^*, \psi_2^*))$. Next consider a deviator who chooses a different belief bias ψ_i'. Observe that in order for the deviator to be able to archive a payoff higher than the original equilibrium payoff of 2, the deviator must play action b with a positive probability. This implies that the opponent's unique best-reply to the deviator's perceived strategy is action d, and, thus, the deviator's payoff is at most one.

	a	b	c	d
a	2	3	0	0
b	3	5	0	0
c	0	0	0	0
d	0	0	0	1

Table 1: Payoff matrix for both players in a doubly symmetric game

4.3 Games with Two Pure Actions

In this section we fully characterize biased-belief equilibrium outcomes in games with two pure actions. We say that game $G = (S, \pi)$ has two pure actions if the set of strategies is a simplex over two actions (i.e., $S_i = \triangle(\{a_i, b_i\})$ for each player i), and π is linear (i.e., a vN-M utility function).

The following result shows that, a strategy profile is a biased belief equilibrium outcome iff is is undominated and it yields each player a payoff weakly higher than the player's undominated minmax payoff. Formally,

Proposition 4. Let $G = (S_i = \triangle(\{a_i, b_i\}), \pi)$ be a game with two pure actions. Then the following two statements are equivalent:

1. *Strategy profile (s_1^*, s_2^*) is a biased-belief equilibrium outcome.*

2. *Strategy profile (s_1^*, s_2^*) is undominated, and $\pi_i(s_1^*, s_2^*) \geq M_i^U$ for each player i.*

Proof. Proposition 2 implies that "1.⇒2.". We now show that "2.⇒1.". Assume that (s_1^*, s_2^*) is undominated, and $\pi_i(s_1^*, s_2^*) \geq M_i^U$. For each player j, let s_j^p be an undominated strategy that guarantees that player i obtains, at most, his minmax payoff M_i^U, i.e., $s_j^p = \text{argmin}_{s_j \in S_j^U}(\max_{s_i \in S_i} \pi_i(s_i, s_j))$.

Assume first that one of the players has a dominant action. Say, without loss of generality, that action a_i is dominant for player i. This implies that $S_i^U = \{a_i\}$, and thus $M_j^U = \max_{s_j}(\pi_j(a_i, s_j))$. This implies that if strategy profile (s_1^*, s_2^*) is undominated, and satisfies $\pi_j(s_1^*, s_2^*) \geq M_j^U$ for each player i, then it must be that $s_i^* = a_i$, and (a_i, s_j^*) is a Nash equilibrium of the underlying game, which implies that $(s_i^* = a_i, s_j^*)$ is a biased-belief equilibrium outcome (due to Proposition 1).

We are left with the case in which the game does not admit any dominant actions. This implies that for each player j, there is a perceived strategy $\hat{s}_j \in S_j$ such that both actions are best replies against s_j, i.e., such that $S_i = BR^{-1}(\hat{s}_j)$. We conclude by showing that $((\hat{s}_2, \hat{s}_1), (s_1^*, s_2^*))$ is a biased-belief equilibrium (in which both players have blind beliefs). It is immediate that $(s_1^*, s_2^*) \in NE(\hat{s}_2, \hat{s}_1)$ (because any strategy is a best reply against each $\hat{s}_j$). Next, observe that for any deviation of player i to a different biased-belief ψ_i', there is a Nash equilibrium of the biased game $(G, (\psi_i', \hat{s}_i))$in which player j plays s_j^p, and, as a result, player i obtains a a payoff of at most M_i^U, which implies that the deviation is not profitable.

The following example demonstrates how to implement the best symmetric outcome in the Hawk-Dove game. □

Example 3 (Implementing cooperation as a BBE outcome in Hawk-Dove games[4]). Consider the Hawk-Dove game (AKA, "Chicken") described in Table 2. Let $\alpha \in [0,1]$ denote the mixed strategy assigning probability of α to action d_i. The best symmetric strategy profile of $(d_1, d_2) \equiv (1,1)$ can be supported as the outcome of the strong biased-belief equilibrium $((\psi_1^*, \psi_2^*), ((d_1, d_2)))$, where $\psi_i^*(\alpha) = \frac{2-\alpha}{2}$. On the equilibrium path each player i plays d_i and believes that his opponent is mixing equally between the two actions. If the opponent plays $s_j \neq d_j$, then player i believes that the opponent plays d_j with probability strictly more than 50%, and as a result player i plays the unique best reply to this belief, namely h_i, and the opponent gets a payoff of at most 1.

Table 2: Hawk-Dove Game

	d_2	h_2
d_1	$3,3$	$1,4$
h_1	$4,1$	$0,0$

4.4 Games with a Continuum Set of Actions

In this section we fully characterize biased-belief equilibrium outcomes in "well-behaved" games in which the set of strategies is an interval.

We say that a game $G = (S, \pi)$ is *well-behaved interval* if: (1) each S_i is a convex subset of $\mathbb{R}$ (i.e., an interval), (2) for each player i the payoff function $\pi_i(s_i, s_j)$ is strictly concave in the agent's strategy s_i, and weakly convex in the opponent's strategy s_j.

Well=behaved interval games are common in many economic environments. some examples include Cournot competition, Price competition with differentiated goods, public good games, and Tullock Contests.

The following result shows that in well-behaved interval games, any undominated strategy profile that induces each player a payoff strictly above the player's undominated minmax payoff can be implemented as an outcome of a strong biased-belief equilibrium. Formally,

Theorem 1. *Let* $G = (S, \pi)$ *be a* well-behaved interval game. *If profile* (s_1^*, s_2^*) *is undominated and* $\pi_i(s_1^*, s_2^*) > M_i^U$ *for each player* i*, then* (s_1^*, s_2^*) *is a strong biased-belief equilibrium outcome.*

[4] Y: Internal comment - I shortened and simplified the example in response to your previous comment that suggest that the (more general) example is not essential to the paper.

Proof. Assume that (s_1^*, s_2^*) is undominated and $\pi_i(s_1^*, s_2^*) > M_i^U$ for each player i. For each player j, let s_j^p be an undominated strategy that guarantees that player i obtains, at most, his minmax payoff M_i^U, i.e., $s_j^p = \text{argmin}_{s_j \in S_j^U} \left(\max_{s_i \in S_i} \pi_i(s_i, s_j)\right)$. The strict convexity of $\pi_i(s_i, s_j)$ with respect to s_i implies that the best-reply correspondence is a continuous one-to-one function. Thus, $BR^{-1}(s_i)$ is a singleton for each strategy s_i, and we identify $BR^{-1}(s_i)$ with the unique element in this singleton set.

For each $\epsilon > 0$ and each player i, let ψ_i^ϵ be defined as follows:

$$\psi_i^\epsilon\left(s_j'\right) = \begin{cases} \frac{\epsilon - \left|s_j' - s_j\right|}{\epsilon} \cdot BR_i^{-1}\left(s_i^*\right) + \frac{\left|s_j' - s_j\right|}{\epsilon} \cdot BR_i^{-1}\left(s_i^p\right) & \left|s_j' - s_j\right| \leq \epsilon \\ BR_i^{-1}\left(s_i^p\right) & \left|s_j' - s_j\right| > \epsilon. \end{cases}$$

We now show that for a sufficiently small $\epsilon > 0$, $((\psi_1^\epsilon, \psi_2^\epsilon), (s_1^*, s_2^*))$ is a strong biased-belief equilibrium. Observe first that the definition of $(\psi_1^\epsilon, \psi_2^\epsilon)$ immediately implies that $\{(s_1^*, s_2^*)\} = NE(G, (\psi_1^\epsilon, \psi_2^\epsilon))$. Next, consider a deviation of player i to a an arbitrary biased belief ψ_i'. Consider any equilibrium $\left(s_i', s_j'\right)$ of the biased game $\left(G, \left(\psi_i', \psi_j^\epsilon\right)\right)$. If $\left|s_j' - s_j\right| > \epsilon$, then the definition of $\psi_i^\epsilon\left(s_j'\right)$ implies that $s_i^p = s_i'$, and that player j achieves a payoff of at most $M_i^U < \pi_i(s_1^*, s_2^*)$. The convexity of the payoff function $\pi_i(s_1, s_2)$ with respect to the opponent's strategy s_j, and standard continuity arguments, imply that for a sufficiently small $\epsilon > 0$, player i's payoff is at most $\pi_i\left(s_i^*, s_j^*\right)$, which shows that $((\psi_1^\epsilon, \psi_2^\epsilon), (s_1, s_2))$is a strong biased-belief equilibrium.

An immediate corollary of Proposition 1 and Theorem 1 is the full characterization of biased-belief equilibrium outcomes in well-behaved interval games: A strategy profile is a BBE outcome, essentially, if and only if (1) it is undominated, and (2) it induces each player a payoff above the player's undominated minmax payoff. Formally, □

Corollary 3. *Let $G = (S, \pi)$ be a well-behaved one-denominational game.*

1. *If profile (s_1^*, s_2^*) is undominated and $\pi_i(s_1^*, s_2^*) > M_i^U$ for each player i, then (s_1^*, s_2^*) is a strong biased-belief equilibrium outcome.*

2. *If (s_1^*, s_2^*) is a biased-belief equilibrium outcome, then (s_1^*, s_2^*) is undominated and $\pi_i(s_1^*, s_2^*) \geq M_i^U$ for each player i.*

The following example demonstrates a biased-belief equilibrium that induces the undominated efficient outcome in Cournot competition.

Example 4 (Biased belief equilibrium that yields the efficient Outcome in Cournot game)**.** Consider the symmetric Cournot game with linear demand of Example 1: $G = (S, \pi)$: $S_i = [0, 1]$ and

$\pi_i(s_i, s_j) = s_i \cdot (1 - s_i - s_j)$ for each player i. Let ψ_i^* be defined for each player i as follows:

$$\psi_i^*(s_j) = \begin{cases} 0.5 & s_j \leq 0.25 \\ 1 - 2 \cdot s_j & 0.25 \leq s_j \leq 0.5 \\ 0 & 0.5 \leq s_j. \end{cases}$$

That is, on the equilibrium path the opponent plays 0.25, and the agent believes that the opponent plays 0.5, which implies that the agent's best-reply strategy is 0.25. If the opponent deviates and plays a lower strategy, it does not affect the agent's perceived strategy (which remains equal to 0.25). Finally, if the opponent deviates and plays a higher strategy than 0.25, then the agent's perceived strategy becomes lower, such that the agent's best-reply strategy becomes higher, and the opponent is outperformed.[5] This implies that $\left((\psi_1^*, \psi_2^*), \left(\frac{1}{4}, \frac{1}{4}\right)\right)$ is a biased-belief equilibrium, which induces the efficient symmetric outcome of $\left(\frac{1}{4}, \frac{1}{4}\right)$, in which both firms equally share the monopoly profit.

5 Wishful thinking and Strategic Complementarity

In this section we focus on the family of games with strategic complementarity and positive spillovers, and we show that in such games, there is a close relation between (1) achieving "socially desirable" outcomes that Pareto improve all Nash equilibria, and (2) monotone biased belief equilibria that rely on wishful thinking. This present a novel theoretical foundation for the tendency of people to exhibit wishful thinking in some situations (see, e.g., Babad and Katz, 1991; Budescu and Bruderman, 1995; Mayraz, 2013).

5.1 Games with Strategic Complementarity

We say that a game exhibits strategic complementarity and spillovers if: (1) the set of strategies of each player is an interval, (2) positive spillovers – each player strictly gains if the partner chooses a higher strategy (interpreted as a higher effort/contribution by the partner), (3) strategic complementarity (Supermodularity) – an increase in the opponent's strategy increases the marginal return to the agent's strategy, and (4) concavity – the payoff function is strictly concave in one's own strategy. Formally:

Definition 5. Game $G = (S, \pi)$ exhibits *strategic complementarity and spillovers* if: (1) each $S_i \subseteq \mathbb{R}$, and for each $s_i, s_j \in (0,1)$ (2) $\frac{\partial \pi_i(s_i,s_j)}{\partial s_j} > 0$, (3) $\frac{\partial^2 \pi_i(s_i,s_j)}{\partial s_i \partial s_j} > 0$, and (4) $\frac{\partial^2 \pi_i(s_i,s_j)}{\partial s_i^2} < 0$.

[5] Y: Internal comment: the recent couple of sentences are new in response to your comment for adding a couple of sentences with intuition about the function $\psi_i^*(s_j)$.

Games that exhibit strategic complementarity and positive spillovers are common in economics (see, e.g., Bulow et al., 1985; Cooper and John, 1988). The following two examples demonstrate two families of such games.

Example 5. *Input games* (aka, *partnership games*). Let $s_i \in \mathbb{R}^+$ be the effort (input) of player i in the production of a public good. The value of the public good, $f(s_1, s_2)$, which is enjoyed by both players, is a supermodular function that is increasing in the effort of each player. The payoff of each player is equal to the value of the public good minus a concave cost of the exerted effort (i.e., $\pi_i(s_i, s_j) = f(s_1, s_2) - g(s_i)$). A specific example for such an input game is presented in Example 7 below.

Example 6. *Price competition with differentiated goods. Let* $s_i \in \mathbb{R}^+$ denote the price of the good produced by firm i. The demand for good i is given by function $q_i(s_i, s_j)$, which is decreasing in s_i and increasing in s_j. The payoff of firm i is given by $\pi_i(s_i, s_j) = (s_i - c_i) \cdot q_i(s_i, s_j)$, where c_i is the marginal cost of production of firm i. Finally, we assume that marginal profit of a firm is increasing in the opponent's price (i.e., $\pi_i(s_i, s_j)$ is supermodular). For example, consider the payoff of symmetric linear city model (Hotelling) in which for each firm i, $S_i = \mathbb{R}^+$, $c_i = c$ and

$$q_i(s_i, s_j) = \begin{cases} 0 & \frac{s_i - s_j + t}{2 \cdot t} < 0 \\ \frac{s_i - s_j + t}{2 \cdot t} & 0 < \frac{s_i - s_j + t}{2 \cdot t} < 1 \\ 1 & \frac{s_i - s_j + t}{2 \cdot t} > 1, \end{cases}$$

where we interpret $t > 0$ as the consumer's travel cost per unit of distance (where a continuum of consumers are equally spaced on a unit interval, one of the firms is located at zero, and the other firm is located at one). One can show that the unique Nash equilibrium of this example is given by $s_i = s_j = c + t$.

It is well known that games with strategic complementarity admit pure Nash equilibria, and, that one of these equilibria $\bar{s}$ is *highest* in the sense that $\bar{s}_i \geq s'_i$ for each player i and each strategy s'_i that is played in a Nash equilibrium (see, e.g., Milgrom and Roberts, 1990). Under the assumption of positive spillovers, this equilibrium $\bar{s}_i$ Pareto-dominates all other Nash equilibria.

We say that a strategy profile (s_1, s_2) is *Nash improving* if it induces each player a payoff higher than the player's payoff in the highest Nash equilibrium (i.e., if $s_i > \bar{s}_i$ for each player i, where $\bar{s}$ is the highest Nash equilibrium).

5.2 Wishful Thinking and Monotonicity

In this section we define two properties of biased-belief equilibria: wishful thinking and monotonicity.

A biased belief equilibrium exhibits wishful thinking if the perceived opponent's strategy yields the agent a higher payoff relative to the real opponent's strategy for all strategy profiles. It exhibits wishful thinking in equilibrium if it satisfies this property with respect to the strategy the opponent plays on the equilibrium path. Formally:

Definition 6. Biased-belief equilibrium $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ *exhibits wishful thinking* (*in equilibrium*) if $\pi_i(s_i, \psi_i^*(s_j)) \geq \pi_i(s_i, s_j)$ for all s_i, s_j with a strict inequality for some s_i, s_j ($\pi_i(s_i, \psi_i^*(s_j)) \geq \pi_i(s_i, s_j^*)$ for all $s_i \in S_i$ with a strict inequality for some $s_i \in S_i$).

Next, we define monotone biased beliefs in interval games. A biased belief equilibrium is monotone if each bias function is increasing with respect to the opponent's strategy. It is monotone in equilibrium if it satisfies this monotonicity property with respect to opponent's strategies that improve the opponent's payoff relative to the equilibrium payoff . Formally:

Definition 7. Let $G = (S, \pi)$ be a game in which the set of strategies of each player is an interval (i.e., $S_i \subseteq \mathbb{R}$ for each player i). Biased belief equilibrium $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is *monotone* if $s_j \geq s_j' \Rightarrow \psi_i^*(s_j) \geq \psi_i^*(s_j')$ for each player i and each pair of strategies s_j and s_j' with a strict inequality for some $s_j > s_j'$ The biased-belief equilibrium $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is *monotone in equilibrium* if $s_j > s_j^* \Rightarrow \psi_i^*(s_j) > \psi_i^*(s_j^*)$ and $s_j < s_j^* \Rightarrow \psi_i^*(s_j) < \psi_i^*(s_j^*)$ for each strategy s_j that satisfies $\pi_j(s_i^*, s_j) > \pi_j(s_i^*, s_j^*)$).

To the extent that biased beliefs emerge through signals that players receive from their counterparts regarding their intentions, the monotonicity condition can be interpreted as requiring that these signals affect beliefs in the right direction but not necessarily in the right magnitude.

5.3 Results

The following result shows that any undominated Nash-improving strategy profile (s_1^*, s_2^*) can be supported by a monotone biased-belief equilibrium that exhibits wishful thinking. Moreover, any biased-belief equilibrium that yields Nash-improving strategy profile as its outcome must satisfy monotonicity and wishful thinking in equilibrium. The intuition is as follows.[6] In a supermodular game a player's incentives to cooperate increase with the level of cooperation of the opponent. Hence wishful thinking allows a player to credibly commit to a high level of cooperation which in turns

[6] Y (internal): new intuitive argument for the result, as you suggested in the previous version.

increases the level of cooperation of his opponent yielding a Pareto improvement over the Nash equilibrium. Formally,

Proposition 5. *Let $G = (S, \pi)$ be a game exhibiting* strategic complementarity and spillovers. *Let (s_1^*, s_2^*) be an undominated Nash-improving strategy profile. Then:*

1. *(s_1^*, s_2^*) is an outcome of a monotone strong biased-belief equilibrium exhibiting wishful thinking.*
2. *Any biased-belief equilibrium $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is monotone in equilibrium, and it exhibits wishful thinking in equilibrium.*

Proof. **Part 1**: The strict convexity of the payoff function π implies that the best-reply correspondence is a one-to-one function. The supermodularity of π implies that the function BR_i^{-1} is strictly increasing. The fact that (s_1^*, s_2^*) is Nash improving implies that $s_j^* < BR_i^{-1}\left(s_j^*\right)$ for each player j. For each $\epsilon > 0$ and each player i, let ψ_i^ϵ be defined as follows:

$$\psi_i^\epsilon\left(s_j'\right) = \begin{cases} s_j' & s_j' \leq s_j^* - \epsilon. \\ \left(1 - \frac{s_j^* - s_j'}{\epsilon}\right) \cdot BR^{-1}\left(s_i^*\right) + \frac{s_j^* - s_j'}{\epsilon} \cdot s_j' & s_j^* - \epsilon < s_j' \leq s_j^* \\ BR_i^{-1}\left(s_i^*\right) & s_j^* < s_j' \leq BR_i^{-1}\left(s_j^*\right) \\ s_j' & s_j' > BR_i^{-1}\left(s_i^*\right) \end{cases}$$

Observe that ψ_i^ϵ is monotone. The fact that the game has positive spillovers implies that ψ_i^ϵ exhibits wishful thinking. We now show that for a sufficiently small $\epsilon > 0$, $((\psi_1^\epsilon, \psi_2^\epsilon), (s_1^*, s_2^*))$ is a strong biased-belief equilibrium. Observe first that $(s_1^*, s_2^*) \in NE(G, (\psi_1^\epsilon, \psi_2^\epsilon))$. Consider a deviation of player j to ψ_j'. Consider any equilibrium of the biased game $\left(G, \left(\psi_i^\epsilon, \psi_j'\right)\right)$. If $\left|s_j' - s_j^*\right| > \epsilon$, then the definition of ψ_i^ϵ implies that player i best replies to the true strategy of player j (i.e., $\psi_i^\epsilon\left(s_j'\right) = s_j'$), and, thus, player j achieves at most the payoff of the highest Nash equilibrium, which is less than $\pi_j(s_1^*, s_2^*)$. The fact that the payoff function $\pi_j(s_1, s_2)$ is supermodular and has positive spillovers, and standard continuity arguments, imply that for a sufficiently small $\epsilon > 0$, player j's payoff is at most $\pi_j\left(s_i^*, s_j^*\right)$, which shows that $((\psi_1^\epsilon, \psi_2^\epsilon), (s_1^*, s_2^*))$ is a strong biased-belief equilibrium.

Part 2: Let $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ be a biased-belief equilibrium. The fact that $(s_1^*, s_2^*) \in NE(G, (\psi_1^*, \psi_2^*))$ implies that $\psi_i^*\left(s_j^*\right) = BR^{-1}\left(s_i^*\right) > s_j^*$ for each player i, which, due to the game being supermodular and having positive spillovers, implies the wishful thinking property in equilibrium. Next, let s_j' be a better reply of player j against s_i^* (relative to s_j^*), i.e., assume that $\pi_j\left(s_i^*, s_j'\right) > \pi_j\left(s_i^*, s_j^*\right)$. Due to the fact that (s_1^*, s_2^*) is Nash improving, this implies that $s_j' < s_j^*$. Assume to the contrary that $\psi_i^*\left(s_j'\right) \geq \psi_i^*\left(s_j^*\right)$. Due to the supermodularity of the game this inequality implies that

$BR\left(\psi_i^*\left(s_j'\right)\right) \geq BR\left(\psi_i^*\left(s_j^*\right)\right)$. Let $\psi_j' \equiv BR^{-1}\left(s_j'\right)$. Then the unique equilibrium of the biased game $\left(G, \left(\psi_i^*, \psi_j'\right)\right)$ is $\left(BR\left(\psi_i\left(s_j'\right)\right), s_j'\right)$, which due to positive spillovers

$$\pi_j\left(\left(BR\left(\psi_i\left(s_j'\right)\right), s_j'\right)\right) \geq \pi_j\left(\left(BR\left(\psi_i\left(s_j^*\right)\right), s_j'\right)\right) = \pi_j\left(s_i^*, s_j'\right) > \pi_j\left(s_i^*, s_j^*\right),$$

which contradicts $\left(\left(\psi_1^*, \psi_2^*\right), \left(s_1^*, s_2^*\right)\right)$ being a distribution equilibrium. □

The following example demonstrates a monotone biased-belief equilibrium exhibiting wishful thinking that induces the undominated efficient outcome in an input game.

Example 7 (Nash improving BBE in an input game)**.** Consider the following input game (which, is presented, and analysed in a different setup in Heller and Sturrock, 2017). Let $S_i = S_j = [0, M]$, and let the payoff function be $\pi_i(s_i, s_j, \rho) = s_i \cdot s_j - \frac{s_i^2}{2\rho}$, where the parameter $\frac{1}{\rho}$ is interpreted as the cost of effort. One can show that: (1) the best-reply function of each agent is playing an effort that is $\rho<1$ times smaller than the opponent (i.e., $BR(s_j) = \rho \cdot s_j$), (2) the unique Nash equilibrium is exerting no efforts $s_i = s_j = 0$, (3) the highest undominated strategy of each player i is $s_i = \rho \cdot M$, and (4) the undominated strategy profile $(\rho \cdot M, \rho \cdot M)$ is Nash improving and induce both players the best payoff among all undominated symmetric strategy profiles. Let ψ_i^*be the following biased-belief function

$$\psi_i^*\left(s_j\right) = \begin{cases} \frac{s_j}{\rho} & s_j < \rho \cdot M \\ 1 & s_j \geq \rho \cdot M. \end{cases}$$

Observe that ψ_i^* is monotone and exhibiting wishful thinking. We now show that $\left(\left(\psi_1^*, \psi_2^*\right), (\rho \cdot M, \rho \cdot M)\right)$ is a biased-belief equilibrium. Observe that $BR_i\left(\psi_i^*\left(s_j\right)\right) = BR_i\left(\frac{s_j}{\rho}\right) = s_j$ for any $s_j \leq \rho \cdot M$, and that $BR_i\left(\psi_i^*\left(s_j\right)\right) = BR(1) = \rho$ for any $s_j \geq \rho$. This implies that $(\rho \cdot M, \rho \cdot M) \in NE\left(G, \left(\psi_1^*, \psi_2^*\right)\right)$, and that for any player i, any biased-belief ψ_i', and any Nash equilibrium $\left(s_1', s_2'\right)$of the biased game $\left(G, \left(\psi_i', \psi_j\right)\right)$, $s_j' = min\left(s_i', \rho\right)$. This implies that $\pi_i\left(s_1', s_2'\right) \leq \pi_i(\rho, \rho)$, which shows that $\left(\left(\psi_1^*, \psi_2^*\right), (\rho \cdot M, \rho \cdot M)\right)$ is a biased-belief equilibrium. Observe that this biased-belief equilibrium induces only a small distortion in the belief of each player, assuming that ρ is sufficiently close to one:

$$\left|\psi_i^*\left(s_j\right) - s_j\right| < \left|\frac{s_j}{\rho} - s_j\right| < M \cdot \frac{1-\rho}{\rho}.$$

Remark 1. Proposition 5 shows the strong relation between Nash improving biased-belief equilibria and wishful thinking in games with strategies complementarity. If one studies the "opposite" family of games with strategic substitutability (i.e., games that satisfy conditions (1-2 and (4) in Definition

5, and the opposite inequality of condition (3), namely $\frac{\partial^2 \pi_i(s_i,s_j)}{\partial s_i \partial s_j} < 0$), then similar arguments to the one presented in Proposition 5 yield an analogous result about the strong relation between Nash improving biased-belief equilibria and pessimistic thinking in games with strategies substitutability.

6 BBE and Undominated Stackelberg Strategies

In this section we present an interesting class of biased-belief equilibria that exist in all games. In this class, one of the players is "rational" in the sense that he plays his undominated Stackelberg strategy (defined below) and has blind beliefs, while his opponent is "flexible" in the sense of having unbiased beliefs.

A strategy is undominated Stackelberg if it maximizes a player's payoff in a setup in which the player can commit to an undominated strategy, and his opponent reacts by choosing the best reply that maximizes player i's payoff. Formally:

Definition 8. The strategy s_i is an undominated Stackelberg strategy if it satisfies

$$s_i = \text{argmax}_{s_i \in S_i^U} \left(\max_{s_j \in BR(s_i)} \left(\pi_i \left(s_i, s_j \right) \right) \right).$$

Let $\pi_i^{\text{Stac}} = \max_{s_i \in S_i^U} \left(max_{s_j \in BR(s_i)} \left(\pi_i \left(s_i, s_j \right) \right) \right)$ be the undominated Stackelberg payoff. Observe that $\pi_i^{\text{Stac}} \geq \pi_i \left(s_1^*, s_2^* \right)$ for any Nash equilibrium $\left(s_1^*, s_2^* \right) \in NE\left(G \right)$.

Our next result shows every game admits a biased-belief equilibrium in which one of the players: (1) has a blind belief, (2) plays his undominated Stackelberg strategy, and (3) obtains his undominated Stackelberg payoff. The opponent has undistorted beliefs. Moreover, this biased-belief equilibrium is strong if the undominated Stackelberg strategy is a unique best-reply to some undominated strategy of the opponent. Formally:

Proposition 6. *Game $G = (S, \pi)$ admits a biased-belief equilibrium $\left(\left(\psi_i^*, Id \right), \left(s_i^*, s_j^* \right) \right)$ for each player i with the following properties: (1) ψ_i^* is blind, (2) s_i^* is an undominated Stackelberg strategy, and (3) $s_j^* = max_{s_j \in BR\left(s_i^*\right)} \left(\pi_i \left(s_i^*, s_j \right) \right)$. Moreover, this biased-belief equilibrium is strong if $\{ s_i^* \} = BR^{-1} \left(s_j^* \right)$.*

Proof. Let s_i^* be an undominated Stackelberg strategy of player i. Let $s_j^* = \text{argmax}_{s_j \in BR\left(s_i^*\right)} \left(\pi_i \left(s_i^*, s_j \right) \right)$. Let $s_j' \in BR^{-1} \left(s_i^* \right)$ ($\left\{ s_j' \right\} = BR^{-1} \left(s_i \right)$ with the additional assumption of the "moreover" part). We now show that $\left(\left(\psi_i^* \equiv s_j', Id \right), \left(s_i^*, s_j^* \right) \right)$ is a (strong) biased-belief equilibrium. It is immediate that $\left(s_i^*, s_j^* \right) \in NE \left(G, \left(\psi_i^* \equiv s_j', Id \right) \right)$. Next, observe that for any biased belief ψ_j' there is an equilibrium (in any equilibrium) of the biased game $\left(G, \left(\psi_i, \psi_j' \right) \right)$ in which player i plays s_i^*, and player j gain at most $\pi_j\left(s_i^*, s_j^*\right)$, which implies that the deviation to ψ_j' is not profitable to player j. If player i deviates to a biased belief ψ_i', then in any equilibrium of the biased game $(G, (\psi_i', \psi_j))$ player i plays some strategy s_i' and gains a payoff of at most $\max_{s_j' \in BR(s_i')}\left(\pi_i\left(s_i', s_j'\right)\right)$, and this implies that player i's payoff is at most π_i^{Stac} , and that he cannot gain by deviating. This shows that $((\psi_1^*, \psi_2^*), (s_1^*, s_2^*))$ is a (strong) biased-belief equilibrium. □

Example 8 (Biased-Belief equilibrium that yields the Stackelberg Outcome in Cournot game)**.** Consider the symmetric Cournot game with linear demand of Example 1: $G = (S, \pi)$: $S_i = \mathbb{R}^+$ and $\pi_i(s_i, s_j) = s_i \cdot (1 - s_i - s_j)$ for each player i. Then $\left((0, I_d), \left(\frac{1}{2}, \frac{1}{4}\right)\right)$ is a biased-belief equilibrium that induces the Stackelberg outcome $\left(\frac{1}{2}, \frac{1}{4}\right)$, and yields player 1 the Stackelberg leader's payoff of $\frac{1}{8}$ and yields player 2 the follower's payoff of $\frac{1}{16}$. This is because: (1) $\left(\frac{1}{2}, \frac{1}{4}\right) \in NE(0, I_d)$, (2) for any biased belief ψ_2', player 1 keeps playing $\frac{1}{2}$ and as a result player 2's payoff is at most $\frac{1}{16}$, and (3) for any biased belief ψ_1', player 2 would best-reply to player's 1 strategy, and thus player 1's payoff would be at most his Stackelberg payoff of $\frac{1}{8}$.

7 Additional Examples of Belief Biased Equilibria

In this section we present three examples of interesting biased-belief equilibria in specific games: (1) prisoner's dilemma with an additional "withdrawal" action, (2) the Centipede game, and (3) the traveler's dilemma.

7.1 Prisoner's Dilemma with an Additional "Withdrawal" Action

As we have argued earlier (Claim 3) biased belief equilibrium outcomes coincide with Nash equilibria in games that admit a dominant strategy. Hence defection is the unique biased belief equilibrium in the prisoner's dilemma game. However, in this section we show that adding weakly dominated strategies (interpreted as "withdrawal") to the prisoner's dilemma can sustain cooperation in the game as the outcome of a strong biased-belief equilibrium. This is done by means of biases under which a player believe that his opponent is planing to withdraw from the game whenever he intends to cooperate, which makes cooperation a rational move.

Table 3: Prisoner's Dilemma Game with a Withdrawal Action

	c	d	w
c	10,10	0,11	0,0
d	11,0	1,1	0,0
w	0,0	0,0	0,0

Consider the variant of the Prisoner's Dilemma game with a third "withdrawal" action as described in Table 3. In this symmetric game both players get a high payoff of 10 if they both play action c (interpreted as cooperation). If one player plays d (*defection*) and his opponent plays c, then the defector gets 11 and the cooperator gets 0. If both players defect, then each of them gets a payoff of 1. Finally, if either player plays action w (interpreted as *withdrawal*), then both players gets 0. Observe that defection is a weakly dominant action, and that the game admits two Nash equilibria: (w, w) and (c, c) inducing respective symmetric payoffs of zero and one.

We identify a mixed action with a vector $(\alpha_c, \alpha_d, \alpha_w)$, where $\alpha_c \geq 0$ (resp., $\alpha_d \geq 0$, $\alpha_w \geq 0$) denotes the probability of choosing action c (resp., d, w). For each player i, let ψ_i be the following biased-belief function:

$$\psi_i^* \left(\alpha_c, \alpha_d, \alpha_w\right) = \left(0, \alpha_d, \alpha_c + \alpha_w\right).$$

We now show $\left(\left(\psi_1^*, \psi_2^*\right), (c, c)\right)$ is a strong biased-belief equilibrium, in which both players obtain a high payoff of 10 (which is strictly better than the best Nash equilibrium payoff, and strictly better than the Stackelberg payoff of each player). Observe first that $c \in BR\left(\psi_i^*\left(c\right)\right) = BR\left(w\right)$, which implies that $(c, c) \in NE\left(G, \left(\psi_1^*, \psi_2^*\right)\right)$. Next, consider a deviation of player i to biased belief ψ_i'. Observe that player i can gain a payoff higher than 10, only if he plays action d with a positive probability, but this implies that the unique best-reply of player j to his biased belief about player i's strategy is defection, which implies that player i obtains a payoff of at most one.

7.2 Centipede Game

In this section we present a strong biased-belief equilibrium that implements the Pareto optimal undominated action profile in the centipede game (an asymmetric discrete game with strategic complementarity).

Consider the following normal-form version of the Centipede game (Rosenthal, 1981), in which each player has 101 actions $A_i = \{1, 2, ..., 100, 101\}$, and the payoff function if player 1 chooses action a_1 and player 2 chooses action a_2 is:

$$\pi_1\left(a_1, a_2\right) = \begin{cases} 2 \cdot \left(a_1 - 1\right) & a_1 \leq a_2 \\ 2 \cdot a_2 - 1 & a_1 > a_2 \end{cases} \qquad \pi_2\left(a_1, a_2\right) = \begin{cases} 2 \cdot \left(a_1 - 1\right) & a_1 \leq a_2 \\ 2 \cdot a_2 + 2 & a_1 > a_2. \end{cases}$$

The interpretation of the game is as follows. Each of the players has an "account" with an initial balance of \$0. At each stage, one of the players (in alternating order, starting with player one) has the right to stop the game. If a player stops the game, each player gets the current amount in his account. If a player chooses not to stop the game, then his account is debited by \$1 and the opponent's account is credited by \$3. The game lasts 200 stages, in which player one can stop in the odd stages and player two can stop in the even stages. Action $k < 101$ is interpreted as stopping in the k-th opportunity to stop. Action 101 is interpreted as not stopping at any point. We allow players to choose mixed strategies, and assume each player to be risk neutral. We identify a mixed strategy with the vector $(\alpha, \alpha_2, ..., \alpha_{101})$, where each $\alpha_k \geq 0$ is interpreted as the agent's probability of choosing action k (and $\sum_i \alpha_k = 1$).

It is well known, that player 1 chooses to stop in the first round in every Nash equilibrium, and both players get a payoff of zero. We say that action k is higher than action m if $k > m$, and we observe that the centipede game has positive spillovers and has strategic complementarity.

The highest undominated action of player one is $a_1 = 101$ (never stopping), which is a best-reply against an opponent who never stops. The highest undominated action of player two is $a_2 = 100$ (stopping in the last stage), which is a best-reply against an opponent who never stops. Observe that $(101, 100)$ is the undominated Pareto-optimal action profile, and that it yields the payoff profile $(199, 202)$

We define the biased beliefs ψ_1^* and ψ_2^* as follows:

$$\psi_1^* (\alpha_1, \alpha_2 ..., \alpha_{99}, \alpha_{100}, \alpha_{101}) = (\alpha_1, \alpha_2, ...\alpha_{99}, 0, \alpha_{100} + \alpha_{101}),$$

$$\psi_2^* (\alpha_1, \alpha_2 ..., \alpha_{99}, \alpha_{100}, \alpha_{101}) = \left(\frac{2}{3} \cdot \alpha_1, \frac{2}{3} \cdot \alpha_2, ..., \alpha_{99}, \frac{1}{3} + \frac{2}{3} \cdot \alpha_{100}, \frac{2}{3} \cdot \alpha_{101} \right).$$

The first player distorts player 2's strategy by perceiving an opponent stopping in the last stage as an opponent who never stops. This implies that never stopping (i.e., $a_1 = 101$) is a best reply of player 1 against player's 2 perceived strategy in equilibrium. The second player distorts player 1's strategy by adding a probability of $\frac{1}{3}$ to player 1 stopping in the last round (and normalizing all probabilities by multiplying them by $\frac{2}{3}$). This implies that in equilibrium player 2 is indifferent between stopping in the last round (i.e., $a_2 = 100$) and stopping in the penultimate round (i.e., $a_2 = 99$). This implies that (1) playing $a_2 = 100$ is a best reply against player 1's perceived strategy in equilibrium, and (2) if player 1 deviates and plays action 99 with a positive probability (which is the best reply against the undistorted equilibrium strategy of player 2, $a_2 = 100$), then $a_2 = 100$ is no longer a best reply against player 1's perceived strategy, and as a result player 2 stops earlier, and player 1 is outperformed.

Interestingly, in contrast to the input game discussed in Example 5, to sustain the efficient outcome in the Centipede game only player 1 can have distorted beliefs that represent wishful thinking. Indeed, to support the efficient outcome (101, 100) as a biased-belief equilibrium it is necessary that player 1 assigns sufficiently high probability that player 2 will act generously in his last decision node. Otherwise, player1's optimal response would be to stop at some earlier stage. But for player 1's optimism to be self-serving it is necessary that player 2 is endowed with pessimism regarding the behavior of player 1 in his last decision node. If player 2 is not pessimistic, then player 1 would be better off by possessing less-optimistic beliefs that allow him to to stop with a small positive probability in his last decision rule (without affecting player 2's equilibrium behavior), in which case the efficient outcome cannot be sustained. In contrast when player 2 is pessimistic about player 1's decision in his last round, this allows player 2 to incentivize player 1 to sustain player 1's optimistic belief and to continue with probability one in his last decision node.

In what follows we formally show that $((\psi_1^*, \psi_2^*), (101, 100))$ is a strong biased-belief equilibrium. Observe first, that $\psi_1^*(100) = 101$ and $\psi_2^*(101) = \left(0, ..., 0, \frac{1}{3}, \frac{2}{3}\right)$, which implies that $101 = BR_1(\psi_1^*(100))$, $100 \in BR_2(\psi_2^*(101))$, and $(101, 100) \in NE(G, (\psi_1^*, \psi_2^*))$. It is clear that player 2 cannot achieve a higher payoff by choosing a different biased belief, because his equilibrium payoff of 202 is the maximal feasible payoff. Let ψ_1' be an arbitrary biased belief of player 1. Observe that player 1 can obtain a payoff higher than 199 only if (1) player 2 chooses action 101 with a positive probability, and (2) player 1 chooses action 100 with a positive probability. However, the biased belief of player 2, ψ_2^*, implies that if player 1 chooses action 100 with a positive probability, then player 2 never chooses action 101 in any Nash equilibrium of the induced biased game because action 101 yields a strictly lower payoff against player 1's perceived strategy relative to the payoff of action 100.

7.3 The Traveler's Dilemma

In this section we present a strong biased-belief equilibrium exhibiting wishful thinking that implements the undominated Pareto-optimal action profile in the traveler's dilemma game (which is a discrete game with strategic complementarity).

Consider the following version of the traveler's dilemma game (Basu, 1994). Each player has 100 actions: $A_i = \{1, ..., 100\}$, and the payoff function of each player is:

$$\pi_i(a_i, a_j) = \begin{cases} a_i + 2 & a_i < a_j \\ a_i & a_i = a_j \\ a_j - 2 & a_i > a_j \end{cases}$$

The interpretation of the game is as follows. Two identical suitcases have been lost, each owned by one of the players. Each player has to evaluate the value of his own suitcase. Both players get a payoff equal to the minimal evaluation (as the suitcases are known to have identical values), and, in addition, if the evaluations differ, then the player who gave the lower (higher) evaluation gets a bonus (malus) of 2 to his payoff.

It is well known that the unique Nash equilibrium is $(1, 1)$, which yields a low payoff of one for each player. Observe, that the traveler's dilemma has positive spillovers, in the sense that it is always weakly better for a player if his opponent chooses a higher action. The traveler's dilemma has strategic complementarity in the sense that the best-reply of an agent is stop one stage before his opponent, and, thus, an agent has an incentive to choose a higher action if his opponent chooses a higher action.

Observe that action 99 is the "highest" undominated action of each player (as 99 is a best-reply against 100, and as action 100 is not a best-reply against any of the opponent's strategies). In what follows, we construct a strong biased-belief equilibrium exhibiting wishful thinking that yields the undominated Pareto-optimal strategy profile $(99, 99)$ with a payoff of 99 to each player.

We define the biased belief ψ_i^* as follows:

$$\psi_i^* \left(\alpha_1, \alpha_2 ..., \alpha_{99}, \alpha_{100}\right) = \left(\alpha_1, \alpha_2, ..., \frac{\alpha_{99}}{2}, \frac{\alpha_{99}}{2} + \alpha_{100}\right).$$

In what follows we show $\left(\left(\psi_1^*, \psi_2^*\right), (99, 99)\right)$ is a strong biased-belief equilibrium. Observe first, that $\psi_1^* (99) = \left(0, ..., 0, \frac{1}{2}, \frac{1}{2}\right)$, which implies that $100 \in BR\left(\psi_i^* (99)\right)$, and, thus, $(99, 99) \in NE\left(G, \left(\psi_1^*, \psi_2^*\right)\right)$. Let ψ_1' an arbitrary perception bias of player i. Observe that player i never plays action 100 in a any Nash equilibrium of any biased game, because action 100 is not a best reply against any strategy of player j. Next observe that player i can obtain a payoff higher than 99 only if (1) player j chooses action 99 with a positive probability, and (2) player 1 chooses action 98 with a probability strictly higher than his probability of playing action 100. However, the biased belief ψ_j^* of player j implies that if player i chooses action 98 with a probability strictly higher than his probability of playing 100, then player j never chooses action 99 in any Nash equilibrium of the induced biased game because action 99 yield player j a strictly lower payoff than action 98 against the perceived strategy of player i (because according to this perceived strategy, player i plays action 100 with a probability strictly less than player i's probability of playing either action 98 or action 99).

7.4 Auction

We next demonstrate the role of biased-belief equilibria in a typical game of competition. For this we consider a simple (first-price) auction with complete information. Our example here will demonstrate that collusive behavior can be sustainable as a biased belief equilibrium.

Consider the following discrete version of a symmetric two-player first-price sealed-bid auction. The two players compete over a single good that is worth $1 << V \in \mathbb{N}$ to each player.[7] Each player i submits a bid $a_i \in \{0, 1, 2, ..., V\}$. The player with the higher bid wins the auction and gets the object for the price he was biding. The opponents gets a payoff of zero. If both players submit the same bid, then the winner of the auction is chosen at random. Formally, the payoff function is:

$$\pi_1(a_1, a_2) = \begin{cases} 0 & a_1 < a_2 \\ \frac{1}{2} \cdot (V - a_1) & a_1 = a_2 \\ V - a_1 & a_1 > a_2. \end{cases}$$

Observe that the game admits 3 Nash equilibria: $(V-2, V-2)$, $(V-1, V-1)$ and (V, V), which induce a low expected payoff of at most 1 to each player. In what follows we show how to obtain the Pareto-optimal symmetric strategy profile $(0, 0)$, which yields a payoff of $\frac{V}{2}$ to each player, as the outcome of a strong belief-biased equilibrium.

We identify a mixed strategy with the vector $(\alpha_0, \alpha_1, ..., \alpha_V)$, where each $\alpha_k \geq 0$ is interpreted as the agent's probability of choosing action k (and $\sum_k \alpha_k = 1$). Let ψ_i^* be defined as follows: $\psi_i^*(\alpha_0, \alpha_1, ..., \alpha_{V-1}, \alpha_V) = \left(0, 0, ..., \sum_{i>0} \alpha_i, \alpha_0\right)$. That is, each player distorts the opponent's strategy, such that a bid of zero is perceived a bid of V, and any other bid is perceived as a bid of $V-1$.

The equilibrium we construct is based on the following intuition: Each player interprets the intention to bid zero as a deception coming from a bidder who will ultimately make the highest possible bid. Under such pessimistic beliefs avoiding making a competitive bid (i.e., bidding zero) is rational and it leads to collusion at a price of zero. The distorted interpretation is optimal because a more rational (or more optimistic) interpretation will lead one's opponent to be more competitive yielding an inferior outcome for both bidders.

We now formally show that $((\psi_1^*, \psi_2^*), (0, 0))$ is a strong belief-biased equilibrium, which yields each player an expected payoff of $\frac{V}{2}$. Observe that $0 \in BR(\psi_i^*(0)) = BR(V)$, which implies that

[7] The results that are presented in this example can be extended to a setup in which the two players have different evaluations for the good $V_i \neq V_j$.

$(0,0) \in NE(G,(\psi_1^*,\psi_2^*))$. Next consider an arbitrary deviation of player i to a biased belief ψ_i'. Let $\left(a_i',a_j'\right) \in NE\left(G,\left(\psi_i',\psi_j^*\right)\right)$ be a strategy profile played in the new biased game following the deviation of player i. If $a_i'=0$, then player i wins the auction with a probability of at most 0.5, and, thus, player i's payoff is at most $\frac{V}{2}$ and he does not gain from the deviation. If $a_i' \neq 0$, $\psi_j^*\left(a_i'\right)$ assigns strictly positive probability to $V-1$ and the remaining probability to V, which implies that player j's unique best reply to the perceived strategy of player i the action $V-1$, which implies that player i's payoff is at most 0.5, and therefore he has not gained from the deviation.

8 Discussion

Decision makers' preferences and beliefs may intermingle. In strategic environments distorted beliefs can take the form of a self-serving commitment device. Our paper introduces a formal model for the emergence of such beliefs and proposes an equilibrium concept that support them. Our analysis characterizes biased-belief equilibria in a variety of strategic environments. It also identifies strategic environments with equilibria that support belief distortions such as wishful thinking and pessimism.

Our analysis here deals with simultaneous games of complete information, but the idea of strategically distorted beliefs may play an important role also in sequential games and in Bayesian games. In these frameworks, belief distortion may violate Bayesian updating, and our concept here can potentially offer a theoretical foundation for some of the cognitive biases relating beliefs' updating. It can potentially also identify the strategic environments in which these biases are likely to occur. We view this as an important research agenda that we intend to undertake in the future.

A different research track that might shed more light on strategic belief distortion is experimental. Laboratory experiments often conduct belief elicitation with the support of incentives for truthful revelation. A strong evidence for strategic belief bias in experimental games can be obtained by showing that players assign different beliefs to the behavior of their own counter-part in the game and to a person playing the same role with someone else. In general our model would predict that beliefs about a third party's behavior are more aligned with reality than those involving one's counter-part in the game. Laboratory experiments can also test whether specific type of beliefs' distortions (such as wishful-thinking) arise in the strategic environments that are predicted by our model.

Finally, we point out that strategic beliefs may play an important role in the design of mechanisms and contracts. Belief distortions may destroy the desirable equilibrium outcomes that a standard mechanism aims to achieve. Mechanisms that either induce unbiased beliefs or adjust the rules of the game to account for possible belief biases are expected to perform better.

References

BABAD, E. AND Y. KATZ, "Wishful thinking – against all odds," *Journal of Applied Social Psychology* 21 (1991), 1921–1938.

BABCOCK, L. AND G. LOEWENSTEIN, "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *The Journal of Economic Perspectives* 11 (1997), 109–126.

BASU, K., "The traveler's dilemma: Paradoxes of rationality in game theory," *The American Economic Review* 84 (1994), 391–395.

BATTIGALLI, P., M. DUFWENBERG AND A. SMITH, "Frustration and anger in games," (2015).

BUDESCU, D. V. AND M. BRUDERMAN, "The relationship between the illusion of control and the desirability bias," *Journal of Behavioral Decision Making* 8 (1995), 109–125.

BULOW, J. I., J. D. GEANAKOPLOS AND P. D. KLEMPERER, "Multimarket oligopoly: Strategic substitutes and complements," *Journal of Political economy* 93 (1985), 488–511.

COOPER, R. AND A. JOHN, "Coordinating coordination failures in Keynesian models," *The Quarterly Journal of Economics* 103 (1988), 441–463.

DEKEL, E., J. C. ELY AND O. YILANKAYA, "Evolution of preferences," *Review of Economic Studies* 74 (2007), 685–704.

DOBSON, K. AND R.-L. FRANCHE, "A conceptual and empirical review of the depressive realism hypothesis.," *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 21 (1989), 419.

FRIEDMAN, D. AND N. SINGH, "Equilibrium vengeance," *Games and Economic Behavior* 66 (2009), 813–829.

GUTH, W. AND M. YAARI, "Explaining reciprocal behavior in simple strategic games: An evolutionary approach," in U. Witt, ed., *Explaining Process and Change: Approaches to Evolutionary Economics* (University of Michigan Press, Ann Arbor, 1992).

HELLER, Y. AND D. STURROCK, "Commitments and Partnerships," mimeo, 2017.

HELLER, Y. AND E. WINTER, "Rule rationality," *International Economic Review* 57 (2016), 997–1026.

Herold, F. and C. Kuzmics, “Evolutionary Stability of Discrimination under Observability,” *Games and Economic Behavior* 67 (2009), 542–551.

Inoue, Y., Y. Tonooka, K. Yamada and S. Kanba, “Deficiency of theory of mind in patients with remitted mood disorder,” *Journal of affective disorders* 82 (2004), 403–409.

Lord, C. G., L. Ross and M. R. Lepper, “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.,” *Journal of personality and social psychology* 37 (1979), 2098.

Mayraz, G., “Wishful Thinking,” Technical Report, The University of Melbourne, 2013.

Milgrom, P. and J. Roberts, “Rationalizability, learning, and equilibrium in games with strategic complementarities,” *Econometrica: Journal of the Econometric Society* (1990), 1255–1277.

Rosenthal, R. W., “Games of perfect information, predatory pricing and the chain-store paradox,” *Journal of Economic theory* 25 (1981), 92–100.

Ross, L. and C. Anderson, “Shortcomings in Attribution Processes: On the Origins and Maintenance of Erroneous Social Judgments,” in D. Kahiwmamm, P. Slovic and A. Tversky, eds., *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge University Press, 1982).

Weibull, J., *Evolutionary Game Theory* (MIT Press, Cambridge, MA, 1997).

Winter, E., I. Garcia-Jurado and L. Mendez-Naya, “Mental equilibrium and rational emotions,” *management science* (forthcoming).