

**WISHFUL THINKING IN STRATEGIC ENVIRONMENTS**

MUHAMET YILDIZ

ABSTRACT. Towards developing a theory of systematic biases about strategies, I analyze strategic implications of a particular bias: wishful thinking about the strategies. I identify a player as a wishful thinker if she hopes to enjoy the highest payoff that is consistent with her information about the others' strategies. I develop a straightforward elimination process that characterizes the strategy profiles that are consistent with wishful thinking, mutual knowledge of wishful thinking, and so on. Every pure-strategy Nash equilibrium is consistent with common knowledge of wishful thinking. For generic two-person games, I further show that the pure Nash equilibrium strategies are the only strategies that are consistent with common knowledge of wishful thinking. My analysis also illustrates how one can characterize the strategic implications of general decision rules using the tools of game theory.

*Keywords:* optimism, strategic uncertainty, wishful thinking

*JEL Classifications:* C72, D80

---

*Date:* July, 2004.

MIT Economics Department, 50 Memorial Drive, Cambridge, MA 02142, USA.

E-mail: myildiz@mit.edu

I thank Daron Acemoglu, Glenn Ellison, Haluk Ergin, Casey Rothschild, Marzena Rostek, Jonathan Weinstein, Robert Wilson, two anonymous referees, and an editor for detailed, thoughtful comments.

## 1. INTRODUCTION

Traditionally, game theory assumes rationality and focuses on Nash equilibrium. In real life and in experiments, however, economic agents have systematically biased beliefs<sup>1</sup> and use decision rules that are inconsistent with rationality. This has inspired many non-rational theories of individual decision making. Clearly, in order for these theories to be useful, we need to know how they can be incorporated into strategic environments with multiple players. For the most part, the papers that consider such decision rules in strategic environments resort to equilibrium analysis. But equilibrium analysis assumes that players correctly guess the other players' strategies.<sup>2</sup> This assumption is problematic in the presence of irrationality: if there are enough repetitions for players to learn each others' strategies, then they would also learn to play a best response. Moreover, if a player uses a non-rational rule of thumb as a decision rule, then she may also use some rule of thumb in predicting other players' strategies. Such rules of thumb may introduce systematic biases, and the resulting beliefs may not correspond to an equilibrium.

Similarly, the papers that investigate the role of optimistic beliefs in economic applications assume that players hold optimistic beliefs about some underlying parameters, and then apply equilibrium analysis.<sup>3</sup> The equilibrium assumption is again problematic. Firstly, it is hard to interpret such equilibria as outcomes of a learning process (Dekel, Fudenberg, and Levine (2004)). More importantly, most observers agree that it is often harder to predict the outcome of strategic interactions than to predict aspects of physical reality. This suggests that players

---

<sup>1</sup>For example, there is a large empirical literature on self-serving biases (e.g., Larwood and Wittaker (1977), Weinstein (1980), and Babcock and Loewenstein (1997)). The common-prior assumption rules out such biases (Aumann (1976,1987) and Feinberg (2000)).

<sup>2</sup>Even when there is uncertainty about strategies in the form of mixed strategies, in order to justify equilibrium analysis, one needs to make similar assumptions, such as mutual knowledge of players' conjectures—which implies that players guess other players' conjectures correctly—or a common prior about strategies (Aumann and Brandenburger (1995)).

<sup>3</sup>There are applications on financial markets (Harrison and Kreps (1978), Morris (1996)), bargaining (Posner (1972), Landes (1971), Yildiz (2003,2004), Ali (2003) Watanabe (2004)), collective action (Wilson (1968), Banerjee and Somanathan (2001)), lending (Manove and Padilla (1999)), and theory of the firm (van den Steen (2005)).

have more substantial self-serving biases about the strategies of others than they do regarding physical reality. Hence, it would be methodologically preferable to model players having biased beliefs about strategic uncertainty whenever they have biased beliefs about the underlying parameters of the real world.

In this paper, I take an epistemic approach to investigate the strategic implications of non-rational decision rules and systematic biases, allowing the players to hold systematic biases about the other players' strategies. I focus on the example of wishful thinking, i.e., the extreme form of optimism, and characterize the strategy profiles that are consistent with this decision rule, mutual knowledge of this decision rule and so on. Similar analyses have been done for rationality and some variations of it (see e.g., Bernheim (1984), Pearce (1984), Tan and Werlang (1988), and Epstein (1997), who considers maximizing behavior without expected utility). These studies allowed heterogeneous beliefs about strategies, but they have not considered *systematic* biases, the focus of the present work.

In my analysis, I use standard models of strategic uncertainty, as in Aumann (1976, 1987). Any such model consists of several components. First, there are states, each of which is the complete description of a strategic situation, including what strategy each player plays, what each player believes about these strategies, and so on. Players do not necessarily know the state. At each state, each player has an information cell consisting of the states that she cannot rule out at that state. This cell represents the set of correct assumptions that she takes as given, which is sometimes referred to as her information. She also has a (conditional) probability distribution on this cell, which is taken to represent her "subjective" beliefs (about other players' strategies, etc.).

I define wishful thinking as follows. Consider an information cell of a player. This player takes the set of strategy profiles played by the other players on this cell as given. She also knows that she can choose any strategy from the set of her own strategies. The product of these two sets is the set of all possible outcomes according to the information cell. I identify a player as *a wishful thinker* at a state if her expected payoff (according to her own probability distribution) at that state

coincides with the highest possible expected payoff one can ever expect within the set of these possible outcomes.

My analysis involves determining strategy profiles that are played in models in which the players are wishful thinkers. To do that, for each strategy profile, one checks whether that strategy profile is *consistent* with wishful thinking, i.e., whether there exists a model with a state at which the strategy profile is played and the players are wishful thinkers. We might also be interested in knowing which strategy profiles are consistent with wishful thinking and some awareness on the part of the players that their opponents are wishful thinkers. Indeed, when a decision rule or a particular bias such as wishful thinking is common to an environment, the players' beliefs may reflect the fact that the other players use that decision rule or have that bias. For example, Babcock and Loewenstein (1997) report cases in which the seasoned negotiators are not only optimistic but also aware that the other parties are optimistic. One can analyze such cases by determining the strategy profiles that are consistent with mutual knowledge of wishful thinking. As wishful thinking becomes more widely appreciated by the players, one must further narrow the set of strategy profiles that can reasonably emerge by requiring them to be consistent with higher order mutual knowledge of wishful thinking. Again this will correspond to considering models where there are more states at which players are wishful thinkers. Researchers often assume a decision rule or a bias throughout the model (e.g. in order to disentangle the implications of self-serving biases from those of asymmetric information). The strategy profiles played in such models are consistent with common knowledge of wishful thinking. These are the strategy profiles that remain possible as we assume more and more layers of knowledge of wishful thinking.

It is clearly too cumbersome to determine such strategy profiles through a model of strategic uncertainty, which consists of a state space, partitions and conditional probability distributions. More problematically, the resulting set depends on the model, as models contain some "common knowledge" assumptions. To overcome this, I develop a straightforward iterated elimination procedure directly on the strategy profiles. The first round of this elimination procedure yields the set of strategy profiles that are consistent with wishful thinking. The second round

of elimination yields the set of strategy profiles that are consistent with mutual knowledge of wishful thinking, and so on. The strategy profiles that survive the elimination process correspond to common knowledge of wishful thinking.

Consider the simple case of the battle of the sexes game:

	$l$	$r$
$t$	<b>2,1</b>	<b>0,0</b>
$b$	0,0	<b>1,2</b>

The strategy profile  $(b, l)$  is not consistent with wishful thinking: it is impossible for Player 1 to simultaneously be a wishful thinker, play  $b$ , and believe that it is possible that Player 2 plays  $l$ . For if it is possible that Player 2 plays  $l$ , then Player 1, being a wishful thinker, imagines that she could get a payoff of 2. But the highest payoff that she can get from playing  $b$  is 1. Therefore, it is impossible for a wishful thinking Player 1 to play  $b$  and to believe that it is possible that Player 2 plays  $l$ . In accordance with this inconsistency, the strategy profile  $(b, l)$  is eliminated at the first round of my elimination process. All the remaining strategy profiles are consistent with wishful thinking, and consistent with common knowledge of wishful thinking. Hence, no other strategy profile is eliminated.

To see that strategy profiles  $(t, l)$ ,  $(t, r)$ , and  $(b, r)$  are consistent with common knowledge of wishful thinking, consider the three state model  $\Omega = \{(t, l), (t, r), (b, r)\}$  (where each state is fully specified by the strategy profile played at that state). Each player knows only her own strategy, e.g., the partition of Player 1 consists of  $\{(t, l), (t, r)\}$  and  $\{(b, r)\}$ . As a wishful thinker, at  $\{(t, l), (t, r)\}$ , Player 1 assigns probability 1 to state  $(t, l)$ , where she gets the highest possible payoff. At the cell  $\{(b, r)\}$ , she plays a best reply to  $r$ , the only possible strategy of Player 2, and hence she is again a wishful thinker. Similarly, Player 2 is also a wishful thinker at all states. Since both players are wishful thinkers at all states, it is common knowledge that they are wishful thinkers. Notice that non-equilibrium profile  $(t, r)$  is consistent with common knowledge of wishful thinking. Indeed, in this model, at state  $(t, r)$  players exhibit a clear form of wishful thinking. Each player is certain that her own favored equilibrium— $(t, l)$  for Player 1 and  $(b, r)$

for Player 2—is being played, but they both are wrong due to the mismatch of expectations.

I show that a strategy profile  $(s_i, s_{-i})$  is consistent with wishful thinking of player  $i$  if and only if there is  $s'_{-i}$  against which  $s_i$  is a best reply and such that the payoff of  $i$  at  $(s_i, s'_{-i})$  is at least as high as her payoff when she plays a best reply to  $s_{-i}$ . For example, in the battle of the sexes game, finding that  $(t, l)$  is better than  $(b, r)$  for Player 1, one can conclude that  $(b, l)$  is not consistent with wishful thinking of Player 1 and that  $(t, r)$  is consistent with wishful thinking of Player 1. In the elimination procedure I will develop, this means that  $(b, l)$  will be eliminated and  $(t, r)$  will be kept. One will then iteratively apply this test to strategy profiles until all the remaining profiles pass the test. Note that we eliminate *strategy profiles*, rather than strategies. This is because, as in the battle of the sexes,  $(b, l)$  may not be consistent with wishful thinking, while both  $(t, l)$  and  $(b, r)$  are.

It turns out that there is a strong relationship between common knowledge of wishful thinking and Nash equilibria in pure strategies. Firstly, every pure-strategy Nash equilibrium is consistent with common knowledge of wishful thinking, simply because it does not leave any room for any strategic uncertainty. Hence, they survive the elimination process. More interestingly, for generic, two-person games, I show that if a strategy is not played in a pure-strategy Nash equilibrium, then it must be eliminated, eventually. That is, each strategy that is consistent with common knowledge of wishful thinking must be played in some pure-strategy Nash equilibrium. In other words, if players are wishful thinkers throughout a model, then each player must be playing a Nash equilibrium strategy. This yields a characterization: a strategy is consistent with common knowledge of wishful thinking if and only if it is played in a pure-strategy Nash equilibrium. This characterization reflects the broader fact that if one assumes a decision rule or a bias throughout a model, then there will be relatively small strategic uncertainty, and the scope of biases will be relatively small.

A number of authors have developed general game theoretical models that incorporate deviations from expected utility maximization (Dekel, Safra, and Segal

(1991), Dow and Werlang (1994), Feinberg (2004)) and psychological motivations (Geanakoplos, Pearce, Stacchetti (1989), Rabin (1993)). Another literature developed solution concepts that are based on bounded rationality, such as quantal-response equilibrium (McKelvey and Palfrey (1995)), limited-foresight models (Jehiel (1995)), and games with procedurally-rational players (Osborne and Rubinstein (1998)). In the latter games, the players use a heuristic to guess the other players' strategies, just as random variables. In a more related work, Eyster and Rabin (2005) propose an "equilibrium" notion in which players underestimate the correlation between the other players' actions and their private information while they correctly estimate the distribution of actions. Jehiel (2005) proposes another "equilibrium" notion in which players play a best reply to the average behavior in coarse analogy groups. Rostek (2004) proposes some set-valued solution concepts and discusses various decision rules on these sets. One of these decision rules reflects wishful thinking, though with a somewhat different formulation. Speigler (2004) analyzes a model in which consumers do not take the firms' strategies into account in forming their expectations.

I formulate the problem in the next section and investigate the strategic implications of wishful thinking in Section 3. In Section 4, I investigate the relation between Nash Equilibrium and common knowledge of rationality. In Section 5, I analyze the case that only some of the players are wishful thinkers. Section 6 concludes. All proofs can be found in the appendix.

## 2. FORMULATION

Consider a game  $(N, S, u)$  where  $N = \{1, 2, \dots, n\}$  is the set of players,  $S = S_1 \times \dots \times S_n$  is the finite set of strategy profiles,<sup>4</sup> and  $u_i : S \rightarrow \mathbb{R}$  is the utility function of player  $i$  for each  $i \in N$ . I consider the standard models of strategic uncertainty, as formulated by Aumann (1976).<sup>5</sup> A *model* is any quadruple  $(\Omega, I, p, \sigma)$  where

---

<sup>4</sup>I use the notational convention of  $x = (x_1, \dots, x_n) \in X_1 \times \dots \times X_n$ ,  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in X_{-i} = \prod_{j \neq i} X_j$ , and  $x = (x_i, x_{-i})$ . For any function  $f$  and any set  $A$ , I write  $f(A) = \{f(a) \mid a \in A\}$ .

<sup>5</sup>See Dekel and Gul (1997) for a survey of alternative ways to model strategic uncertainty.

- $\Omega$  is a state space, where each state  $\omega \in \Omega$  contains all information about players' strategies, including what each player knows and believes, at  $\omega$ .
- $I = (I_i)_{i \in N}$  is a collection of partitions  $I_i$  of  $\Omega$ .  $I_i$  is called the *information partition* of player  $i$ , and  $I_i(\omega)$  will denote the cell of  $I_i$  that contains  $\omega$ . The states in  $I_i(\omega)$  are indistinguishable from  $\omega$  for player  $i$ .
- $p = (p_{i,\omega})_{i \in N, I_i(\omega) \in I_i}$ , where  $p_{i,\omega}$  is a probability distribution on information cell  $I_i(\omega)$ , representing the (subjective) beliefs of player  $i$  at  $\omega$ , and  $p_{i,\omega'} = p_{i,\omega}$  whenever  $\omega' \in I_i(\omega)$ . The expectation operator with respect to  $p_{i,\omega}$  is denoted by  $E_{i,\omega}$ .
- Finally,  $\sigma : \Omega \rightarrow S$ , where  $\sigma(\omega) = (\sigma_1(\omega), \dots, \sigma_n(\omega))$  is the strategy profile played at  $\omega$  and  $\sigma_i$  is constant on each cell  $I_i(\omega)$ .

Here  $(\Omega, I, p)$  is equivalent to usual models of incomplete information, where players observe signals (see Aumann (1999)). The mapping  $\sigma$  gives strategic meaning to the abstract states in the model. The condition that  $\sigma_i$  is constant on each information cell of  $i$  assures that player  $i$  knows her own strategy. The players' behavior at a state is governed by their subjective beliefs at that state, and these beliefs are represented by  $p$ . Using  $p_{i,\omega}$ , one can compute the entire hierarchy of beliefs of player  $i$  at  $\omega$ , namely, her beliefs about the other players' strategies, her beliefs about other players' beliefs about the strategy profile that is being played, and so on.

A model also contains information partitions. I use information partitions to define wishful thinking, the decision rule this paper focuses on. I also use them to characterize the strategic behavior implied by this decision rule in conjunction with various assumptions, for example the assumption that a player correctly assumes that the other player is a wishful thinker. An information partition may be interpreted in two ways. First, it may represent all the "objective" information player  $i$  has according to the model. In this interpretation, at  $\omega$ , player  $i$  *knows* that one of the states in  $I_i(\omega)$  occurs but cannot rule out any of these states. Alternatively, an information partition represents a set of assumptions a player may take as given. In this interpretation,  $I_i(\omega)$  corresponds to states at which a certain set of assumptions holds. Player  $i$  takes these assumptions as given



at  $\omega$ , i.e., she does not assign positive probability outside  $I_i(\omega)$ . Notice that, since  $\omega \in I_i(\omega)$ , these assumptions indeed hold at  $\omega$ . But no such restriction is imposed on players' actual beliefs, and these beliefs may incorporate other (and possibly incorrect) assumptions. In particular, a wishful thinker will be often certain of some state, making many other strong assumptions. In most states these extra assumptions will be false and the player will be "delusional," in the sense of assigning zero probability to the true state.

**Decision rules and Rationality.** A *decision rule* is any function that maps a player's information cells to her strategies (Aumann (1995)). Rationality simply means expected utility maximization. Formally, a player  $i$  is said to be *rational at  $\omega$*  if and only if

$$(2.1) \quad E_{i,\omega} [u_i(\sigma)] = \max_{s_i \in S_i} E_{i,\omega} [u_i(s_i, \sigma_{-i})].$$

That is,  $\sigma_i(\omega)$  maximizes the expected value of  $u_i$  given the beliefs of player  $i$  about the other players' strategies (induced by  $p_{i,\omega}$  and  $\sigma_{-i}$ ). The set of states at which  $i$  is rational is denoted by  $R_i$ . Rationality is a restriction on decision rules: the outcome of the decision rule at an information cell  $I_i(\omega)$  is a best response to a belief that puts probability one on  $\sigma_{-i}(I_i(\omega))$ . Since there are many such beliefs, there are many decision rules consistent with rationality.

**Wishful Thinking.** I will define a wishful thinker as a player who expects to enjoy the highest payoff that is possible given her information about the other players' strategies and given that she can choose any of her own strategies. Formally, a player  $i$  is said to be *a wishful thinker at  $\omega$*  if and only if

$$(2.2) \quad E_{i,\omega} [u_i(\sigma)] = \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s).$$

Player  $i$  is a wishful thinker if her expected payoff (with respect to  $p_{i,\omega}$ ) happens to be the highest payoff that she could get if one considers all the strategy profiles  $s_{-i} \in \sigma_{-i}(I_i(\omega))$  of other players that are possible according to  $I_i(\omega)$  and all the strategies of player  $i$ . I write  $W_i$  for the set of states at which  $i$  is a wishful thinker and  $W = \bigcap_{i \in N} W_i$  for the set of states at which everybody is a wishful thinker.

Clearly, the maximization in (2.2) can be written as a maximization over strategies  $s_i$  and beliefs  $\mu$  on  $\sigma_{-i}(I_i(\omega))$ . This leads to other interpretations of wishful

thinking. First, the strategy and the beliefs of a wishful thinker are *as if* she chooses her strategies and beliefs in order to make herself feel as happy as possible. Second, wishful thinking is maximal optimism under rationality. The belief of a wishful thinker maximizes her expected payoff among all allowed beliefs, given that she will play a best response to these beliefs. Such extreme beliefs themselves can be considered as "irrational" in the popular sense, as they may involve self deception and misperception of the world. Third, wishful thinking can also be thought of as maximizing one's own strategy while thinking that the other players will react to one's own strategy by playing strategies that are favorable to her. Believing that one's own strategy will have an impact on the other players' strategies is clearly inconsistent with reality, and thus wishful thinking can be interpreted as a form of irrationality in that sense. Nevertheless, wishful thinking *formally implies* rationality in the sense of expected utility maximization:

**Fact 1.** *If a player  $i$  is a wishful thinker at  $\omega$ , then player  $i$  is rational at  $\omega$ , i.e.,  $W_i \subseteq R_i$ .*

*Proof.* All the proofs are presented in the appendix. □

The next section explores the strategic implications of using a decision rule that is consistent with wishful thinking. Fact 1 establishes that these decision rules are also consistent with rationality. For generic utility functions, there exists a unique decision rule consistent with wishful thinking, which picks the unique strategy  $s_i^*$  with  $u_i(s^*) = \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s)$  at each information cell  $I_i(\omega)$ . Hence, I will sometimes refer to wishful thinking as a decision rule.

**Notation 1.** Given any  $F \subseteq \Omega$ ,

$$K_i(F) = \{\omega | I_i(\omega) \subseteq F\}$$

denotes the set of states at which player  $i$  knows (or correctly assumes) that  $F$  occurs. The mutual knowledge at any order  $m \geq 0$  is represented by operator  $K^m$  where  $K^0(F) = F$  and  $K^m(F) = \bigcap_{i \in N} K_i(K^{m-1}(F))$ . The set of states at which  $F$  is common knowledge is denoted by  $CK(F)$ , which is defined as  $CK(F) = \bigcap_{m=0}^{\infty} K^m(F)$ .

**Remark 1.** At a state  $\omega$ , a player  $i$  knows that a player  $j$  is a wishful thinker (i.e.,  $\omega \in K_i(W_j)$ ) if and only if player  $j$  is a wishful thinker at each state  $\omega' \in I_i(\omega)$ . In that sense, player  $i$ 's knowing that player  $j$  is a wishful thinker is related to commonness of wishful thinking of player  $j$  in the model. Such commonness may reflect an environment where  $j$  is known to have a tendency to hold excessively optimistic beliefs. For example, this may be due to earlier interactions, or an environment where wishful thinking is common. Also, a player always knows whether she is to be identified as a wishful thinker in this paper: if  $\omega \in W_i$ , then  $I_i(\omega) \subseteq W_i$ . Therefore,

$$(2.3) \quad K_i(W_i) = W_i.$$

Wishful thinking is defined with respect to the information set of the player. In certain situations, e.g., when players face a new decision problem, there is significant uncertainty about players' beliefs, their susceptibility towards certain biases, and about what they would do. In that case, biases play significant role in forming their beliefs. In models that describe such situations, the information sets will be relatively large. A wishful thinker will form a strong belief that is consistent with her information by often being certain that a particular state is true. In that case, there will be large gap between what the player "knows" and what she assumes on faith. Since the latter determines her behavior, its impact will be large and her beliefs can be significantly different from what the analyst would find reasonable. In certain situations, uncertainty will be small, and even a wishful thinker may be aware of many facts of the case. For example, this may be the case when the decision problem is similar to many earlier decisions. In the models that describe those situations, the information sets will be small, and the gap between what the player knows and what she takes on faith will be small. In that case, biases will play relatively small role. For example, if it is clear that a certain equilibrium is played (e.g., after a long repetition of a game), then the relevant model may contain a single state, and even a wishful thinker's beliefs and action corresponds to an equilibrium. In this paper, I will allow all possible models and find the outcomes of these models when players are wishful thinkers.

### 3. STRATEGIC IMPLICATIONS OF WISHFUL THINKING

In this section, I explore the strategic implications of wishful thinking. I develop an iterative elimination process that characterizes the strategy profiles that are consistent with wishful thinking, the strategy profiles that are consistent with mutual knowledge of wishful thinking, and so on.

Though I develop the elimination procedure in the context of wishful thinking, I have a broader class of problems in mind. Consider a researcher who wants predict the outcome of strategic play when players obey a general decision rule—for example when they are wishful thinkers. To do this one can write a model  $M = (\Omega, I, p, \sigma)$  as described in the previous section and identify the states at which the players use such a decision rule, checking which strategy profiles are played at these states. In the case of wishful thinking, the set of such states will be  $W$ , and the set of strategy profiles will be  $\sigma(W)$ . The latter set consists of the strategy profiles implied by wishful thinking and the model  $M$ , and it depends on  $M$ . In order to find out the strategic implications of wishful thinking alone, one considers the union of all the sets  $\sigma(W)$  over all models. The resulting set will be denoted by  $X^0$  in this paper. A strategy profile  $s$  is in  $X^0$  if and only if there is a model  $M$  and a state  $\omega$  such that players are wishful thinkers at  $\omega$  and they play  $s$  at  $\omega$ . Such strategy profiles are said to be *consistent with wishful thinking*.

The researcher may be interested in applications where a decision rule is more commonly used so that the players take into account that the other players also use the decision rule. In that case, the researcher considers again a model  $M$ , but now identifies only the states such that the decision rule is applied and the players know (or correctly assume) that the other players apply the decision rule. In the case of wishful thinking, this set is

$$W \cap K^1(W).$$

The researcher is interested in the set of strategy profiles that are played at these states, namely  $\sigma(W \cap K^1(W))$ , which is a subset of  $\sigma(W)$ . This set characterizes the strategic implications of mutual knowledge of wishful thinking and the model.

Again, in order to characterize the strategy profiles that are consistent with mutual knowledge of wishful thinking, one takes the union of all sets of the form  $\sigma(W \cap K^1(W))$  over all models. That set will be denoted by  $X^1$  in this paper. As the researcher considers the situations where the decision rule is more and more common, strategic implications of higher-order knowledge of the decision rule become more relevant.

In application, a researcher will often assume throughout the model that a particular decision rule is applied. To find out what predictions the researcher can get from such models, one characterizes the set of strategy profiles played in such models. In the case of wishful thinking, this corresponds to finding the union of all the sets of the form  $\sigma(CK(W))$  over all models  $M$ , which will be denoted by  $X^\infty$ . Such strategy profiles are said to be *consistent with common knowledge of wishful thinking*. I will next develop a simple elimination procedure to compute  $X^0, X^1, X^2, \dots, X^\infty$  without going through the cumbersome task of considering arbitrary models. I need one more bit of notation.

**Notation 2.** For each  $s_{-i}$ ,

$$BR_i(s_{-i}) = \arg \max_{s_i \in S_i} u_i(s_i, s_{-i})$$

denotes the set of best responses to  $s_{-i}$ , and the graph of  $BR_i$  is denoted by

$$B_i = \{(\hat{s}_i, s_{-i}) \mid \hat{s}_i \in BR_i(s_{-i}), s_{-i} \in S_{-i}\}.$$

The next lemma describes a defining property of the strategy profiles consistent with wishful thinking of a player. I will use this property to define the elimination procedure.

**Lemma 1.** *For any  $X \subseteq S$  and any  $i \in N$ , let  $\hat{s}$  be a strategy profile played at a state  $\omega$  at which player  $i$  is a wishful thinker and knows that a strategy profile in  $X$  is played. Then, there exists  $(\hat{s}_i, s_{-i}) \in B_i \cap X$  such that*

$$(3.1) \quad u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i}).$$

The reasoning behind this lemma is simple. Since player  $i$  is a wishful thinker at  $\omega$ , by definition, she plays  $\hat{s}_i$  as a best reply to some  $s_{-i}$  played by others at

some  $\omega'$  in her information set. By definition of  $B_i$ ,  $(\hat{s}_i, s_{-i}) \in B_i$ . Since  $\omega'$  is in her information set and she knows that the outcome is in  $X$ , the outcome  $(\hat{s}_i, s_{-i})$  at  $\omega'$  must also be in  $X$ . Since she is a wishful thinker, her expected payoff,  $u_i(\hat{s}_i, s_{-i})$ , cannot be lower than the payoff  $\max_{s_i} u_i(s_i, \hat{s}_{-i})$  that she would get when she played a best response to the belief that the state is  $\omega$ , a belief that is consistent with her information.

Using Lemma 1, one can rule out certain strategy profiles as possible strategic outcomes when some of the players are wishful thinkers. For example, for the battle of the sexes game in the Introduction, Lemma 1 implies that if there were a state at which Player 1 is a wishful thinker and strategy profile  $(b, l)$  is played, then  $u_1(t, l) \leq u_1(b, r)$ . Since the latter is not true, one then concludes that  $(b, l)$  is not consistent with wishful thinking of Player 1. Iterative application of this idea leads to a powerful elimination procedure, which characterizes the strategy profiles that are consistent with wishful thinking, mutual knowledge of wishful thinking, and so on.

### Elimination Procedure.

- (1) Initialization: Set  $X^{-1} = S$ .
- (2) Elimination: For any  $m \geq 0$ , for each strategy profile  $\hat{s}$ , eliminate  $\hat{s}$  if there exists  $i$  for which there does not exist any  $(\hat{s}_i, s_{-i}) \in B_i \cap X^{m-1}$  with  $u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i})$ . Call the remaining strategy profile  $X^m$ .
- (3) Iterate step (2).

I will explain the logic of the elimination procedure and its relation to strategic implications to wishful thinking in greater detail in the rest of this section. First, let me indicate some of its properties.

- One eliminates *strategy profiles*, rather than strategies.
- If  $\hat{s} \in B_i$ , then  $\hat{s}$  is not eliminated for player  $i$  because in that case  $\hat{s} \in B_i \cap X^{m-1}$  and  $u_i(\hat{s}) = \max_{s_i} u_i(s_i, \hat{s}_{-i})$ .
- If  $\hat{s}_i$  is not a best reply to any strategy  $s_{-i}$  where  $(\hat{s}_i, s_{-i})$  is still available, then strategy  $\hat{s}_i$  is eliminated. That is, all strategy profiles in which player  $i$

plays  $s_i$  are eliminated. Hence, the elimination procedure contains iterated elimination of strategies that are never a best reply to a pure strategy. Therefore,  $X^\infty$  refines rationalizability.

More formally, I define a mapping  $\phi : 2^S \rightarrow 2^S$  by setting

$$(3.2) \quad \phi(X) = X \cap \left\{ \hat{s} \mid \forall i : \max_{s_i} u_i(s_i, \hat{s}_{-i}) \leq \max_{(\hat{s}_i, s_{-i}) \in B_i \cap X} (u_i(\hat{s}_i, s_{-i})) \right\}$$

at each  $X$ , where I use the convention that the maximum over the empty set yields  $-\infty$ . The sequence  $(X^{-1}, X^0, \dots)$  is recursively defined by  $X^{-1} = S$  and

$$X^m = \phi(X^{m-1}) \quad (m \geq 0).$$

The limit of the sequence is

$$X^\infty = \bigcap_{m=0}^{\infty} X^m.$$

Since there are only finitely many strategy profiles, the elimination process stops at some iteration  $m$ , and we have  $X^\infty = X^m$  for some  $m$ .

**Remark 2.** The elimination process is monotonic, i.e.,  $X \subseteq Y \Rightarrow \phi(X) \subseteq \phi(Y)$ . Hence, the limit set  $X^\infty$  does not change if one fails to eliminate certain strategies at some step or applies different orders.

The next result states that  $X^m$  is precisely the strategies that are consistent with  $m$ th-order mutual knowledge of wishful thinking. Therefore, using the elimination procedure above, a researcher can investigate the strategic implications of wishful thinking directly from the strategy profiles—without dealing with abstract, complicated models of strategic uncertainty.

**Proposition 1.** *For any model  $(\Omega, I, p, \sigma)$ , and any  $m \geq 0$ ,*

$$\sigma(K^m(W)) \subseteq X^m;$$

*in particular,*

$$\sigma(CK(W)) \subseteq X^\infty.$$

*Moreover, there exist models  $(\Omega, I, p, \sigma)$  in which the above inclusions are equalities.*

The first statement is given by inductive applications of Lemma 1. The proof of the second part involves constructing a submodel for each strategy profile in  $X^m$ , where the strategy profile is played at a state in which wishful thinking is  $m$ th-order mutual knowledge. One constructs such a model using information sets with only one or two states. At such information cells, Lemma 1 characterizes the wishful thinking behavior. Integrating these models into one model, one obtains the desired model.

In the following example, I will illustrate how the elimination procedure is applied and why each round of elimination corresponds to a new layer of knowledge of wishful thinking, as stated in the previous result.

**Example 1.** Consider the following two-person game, where Player 1 chooses between the rows, and Player 2 chooses between the columns:

	$\alpha$	$\beta$	$\gamma$	$\delta$
$a$	$3^*, 0$	$-1, 0$	$0, 0$	$0, 2^*$
$b$	$0, 0$	$2^*, 1^*$	$0, 0$	$0, 0$
$c$	$0, 0$	$0, 0$	$1^*, 2^*$	$1^*, 0$
$d$	$2, 3^*$	$1, 0$	$0, 0$	$0, 0$

In this table, the asterisk after a player's payoff indicates that the player is playing a best reply to the other player's strategy at that profile. Notice that no strategy is weakly dominated, and hence all strategies are rationalizable. Let us apply the above elimination procedure. Take  $m = 0$ . In this round, we eliminate the strategy profiles that are not played at a state at which both players are wishful thinkers. Let us start with Player 1. Consider the strategy profile  $(b, \alpha)$ . For Player 1,  $b$  is a best reply only to  $\beta$ , and her best reply to  $\alpha$  is  $a$ . Since  $u_1(b, \beta) < u_1(a, \alpha)$ , according to the elimination procedure, we eliminate  $(b, \alpha)$ . (By Lemma 1, if  $(b, \alpha)$  is played at a state at which Player 1 is a wishful thinker, then  $\max_{(b, s_2) \in B_1} u_1(b, s_2) \geq \max_{s_1} u_1(s_1, \alpha)$ .) Similarly,  $(c, \alpha)$  and  $(c, \beta)$  are eliminated. Finally, the strategy  $d$  is eliminated because  $d$  is not a best reply to any  $s_2$ . (Recall that a wishful thinker plays a best reply to a pure strategy.) No other profile is eliminated for Player 1 at this round. For example,  $(a, \beta)$  is not eliminated. For Player 1,  $a$  is a best reply to  $\alpha$ , and the best reply to  $\beta$  is  $b$ .



Since  $u_1(a, \alpha) \geq u_1(b, \beta)$ , one does not eliminate  $(a, \beta)$  for Player 1.<sup>6</sup> However,  $(a, \beta)$  is eliminated for Player 2. For Player 2,  $\beta$  is a best reply to only  $b$ , and her best reply to  $a$  is  $\delta$ . Since  $u_2(a, \delta) > u_2(b, \beta)$ , we eliminate  $(a, \beta)$ . No other profile is eliminated.  $X^0$  consists of the strategy profiles in bold:

	$\alpha$	$\beta$	$\gamma$	$\delta$
$a$	<b><math>3^*, 0</math></b>	$-1, 0$	<b><math>0, 0</math></b>	<b><math>0, 2^*</math></b>
$b$	$0, 0$	<b><math>2^*, 1^*</math></b>	<b><math>0, 0</math></b>	<b><math>0, 0</math></b>
$c$	$0, 0$	$0, 0$	<b><math>1^*, 2^*</math></b>	<b><math>1^*, 0</math></b>
$d$	$2, 3^*$	$1, 0$	$0, 0$	$0, 0$

By Proposition 1,  $X^0$  is precisely the set of strategy profiles that are played at states at which players are wishful thinkers. The following model, denoted by  $M^0$ , contains such states. (For a visual description, see Figure 1, where the information cells of Player 1 are rectangular.) Take  $\Omega = X^0 \cup \{(d, \alpha)\}$ ; set  $\sigma(\omega) = \omega$  at each  $\omega$ , and let each player know precisely her own strategy, i.e.,

$$I_1 = \{(a, \alpha), (a, \gamma), (a, \delta)\}, \{(b, \beta), (b, \gamma), (b, \delta)\}, \{(c, \gamma), (c, \delta)\}, \{(d, \alpha)\}$$

$$I_2 = \{(a, \alpha), (d, \alpha)\}, \{(b, \beta)\}, \{(a, \gamma), (b, \gamma), (c, \gamma)\}, \{(a, \delta), (b, \delta), (c, \delta)\}.$$

Player 1 assigns probability 1 to the first state in each information cell. Player 2 assigns probability 1 to  $(d, \alpha)$  at  $\{(a, \alpha), (d, \alpha)\}$ , to  $(c, \gamma)$  at the cell with  $\gamma$ , and to  $(a, \delta)$  at the cell with  $\delta$ . In her information cell with  $a$ , Player 1 expects the highest payoff in the game, hence she is a wishful thinker. In the cell with  $b$ , she expects the payoff of 2, the highest payoff she could expect given that Player 2 does not play  $\alpha$ , as stipulated by the information cell. Hence, she is a wishful thinker. At  $\{(c, \gamma), (c, \delta)\}$ , she is again a wishful thinker as her expected payoff of 1 is the highest payoff given that Player 2 plays  $\gamma$  or  $\delta$ . Similarly, one can check that Player 2 is a wishful thinker at each state. Notice that Player 1 is *not* a wishful thinker at  $(d, \alpha)$ , as she could get 3 by playing a best reply to  $\alpha$ . Hence, at  $(a, \alpha)$ , Player 2 does not know that Player 1 is a wishful thinker, although both of them are wishful thinkers.

<sup>6</sup>Indeed, if Player 1 has a information cell  $\{\omega, \omega'\}$  with  $\sigma(\omega) = (a, \beta)$  and  $\sigma(\omega') = (a, \alpha)$  and if she puts probability 1 on  $\omega'$ , then she is a wishful thinker.

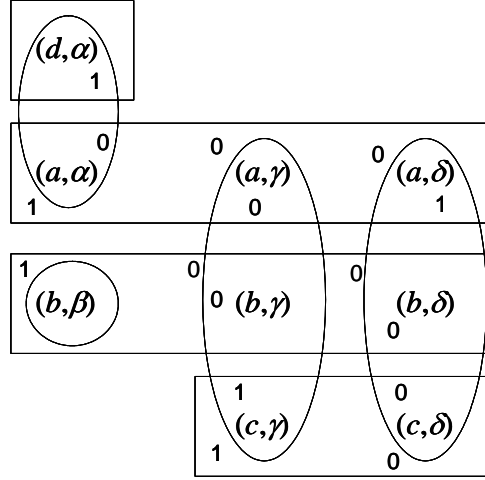


FIGURE 1. Model  $M^0$  of Example 1; players are wishful thinkers at all states except for  $(d, \alpha)$ , where Player 1 is not a wishful thinker.

Now consider  $m = 1$ . In this round we eliminate the strategy profiles that are not played at a state at which players are wishful thinkers and know that their fellow players are wishful thinkers. Since each player  $i$  knows that she is also a wishful thinker, at each such state  $\omega$ ,  $\sigma(I_i(\omega)) \subseteq X^0$ . Then, we ignore the strategy profiles eliminated in the previous round and apply Lemma 1. For example, the strategy profile  $(a, \alpha)$  is eliminated for Player 2. Mechanically, this is because  $\alpha$  is a best reply only against  $d$ , but  $(d, \alpha)$  has been eliminated. The deeper reason for this elimination is that there cannot be a state  $\omega$  at which Player 2 is a wishful thinker, knows that Player 1 is also a wishful thinker, and plays  $\alpha$ . For such  $\omega$ , at each  $\omega' \in I_2(\omega)$ , we would have  $\sigma(\omega') \in X^0$ —as she knows that players are wishful thinkers—and  $\sigma_2(\omega') = \alpha$ , as she knows her own strategy. But the only strategy profile  $(s_1, s_2) \in X^0$  with  $s_2 = \alpha$  is  $(a, \alpha)$ . Hence,  $\sigma(I_2(\omega)) = \{(a, \alpha)\}$ . Then, the expected payoff of Player 2 at  $\omega$  would be  $u_2(a, \alpha) = 0$ . But the higher payoff of  $u_2(a, \delta) = 2$  is consistent with  $\sigma_1(I_2(\omega))$ . Hence, Player 2 would not be a wishful thinker at  $\omega$ —a contradiction.

It turns out that there are no more eliminations for  $m = 1$ , yielding  $X^1 = X^0 \setminus \{(a, \alpha)\}$ . Indeed, in model  $M^0$ , both players are wishful thinkers and know that they are wishful thinkers at each state  $\omega \in X^1$ . For  $m = 2$ , strategy  $a$

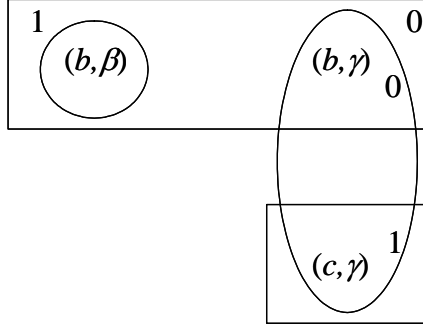


FIGURE 2. A model with common knowledge of wishful thinking in Example 1.

is eliminated as  $a$  is not a best reply for Player 1 in any remaining profile in the first row. This is the only elimination. Hence,  $X^2$  consists of the boldface strategy profiles in the second and third rows above. For  $m = 3$ , strategy  $\delta$  is eliminated. The elimination process stops here. Therefore,  $X^3 = X^4 = \dots = X^\infty = \{(b, \beta), (b, \gamma), (c, \gamma)\}$ .

In the following model, which is described in Figure 2, wishful thinking is common knowledge, and each strategy profile in  $X^\infty$  is played at some state. Take  $\Omega = X^\infty$  and  $\sigma(\omega) = \omega$  at each  $\omega$ . Each player knows her own strategy:  $I_1 = \{(b, \beta), (b, \gamma)\}, \{(c, \gamma)\}$  and  $I_2 = \{(b, \beta)\}, \{(b, \gamma), (c, \gamma)\}$ . Take also  $p_{1,(b,\beta)}((b, \beta)) = 1$  and  $p_{2,(c,\gamma)}((c, \gamma)) = 1$ . Clearly, each player is a wishful thinker at each state, and hence  $W = K^1(W) = K^2(W) = \dots = CK(W) = \Omega = X^\infty$ . Therefore,  $\sigma(CK(W)) = X^\infty$ .

When a strategy profile  $(s_1, s_2)$  is not consistent with wishful thinking while  $(s_1, s'_2)$  is, it is not clear that we can ignore strategy profile  $(s_1, s_2)$  when we compute the strategy profiles that are consistent with wishful thinking and the fact that Player 2 knows this. After all, strategy  $s_1$  is consistent with the assumption that Player 1 is a wishful thinker. How can then Player 2 rule out the possibility that Player 1 plays  $s_1$ ? The reason is as follows. We are trying to find out whether there is a state  $\omega$  of a model where Player 2 knows that players are wishful thinkers and plays  $s_2$ . At  $\omega$ , knowing that players are wishful thinkers,

she must know that a strategy profile in  $X^0$  is played. That is, at each  $\omega' \in I_i(\omega)$ , the outcome  $(\sigma_1(\omega'), s_2)$  is in  $X^0$ . Since  $(s_1, s_2) \notin X^0$ ,  $\sigma_1(\omega') \neq s_1$ . Therefore, at  $\omega$ , Player 2 knows that Player 1 does not play  $s_1$ , and she can ignore  $s_1$ . In Example 1, using this reasoning, I concluded that  $(a, \alpha)$  is not consistent with wishful thinking and the assumption that Player 2 knows this.

A wishful thinker is often delusional as she often assigns probability one to a state, assigning zero probability at all other states. Since the assumption that players do not assign zero probability to the true state plays significant role in many problems in the epistemic literature, one may wonder if this certainty of players plays a significant role in the present analysis. The answer turns out to be No. Given any state  $\omega$ , say that a player  $i$  is *strongly non-delusional at  $\omega$*  if  $p_{i,\omega}(\omega') > 0$  for each  $\omega' \in I_i(\omega)$ . For some  $\varepsilon \geq 0$ , say that  $i$  is an  *$\varepsilon$ -wishful thinker at  $\omega$*  if  $E_{i,\omega}[u_i(\sigma)] \geq \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s) - \varepsilon$ . Let  $Y^m$  be the strategy profiles that are consistent with mutual knowledge of strong non-delusionality and  $\varepsilon$ -wishful thinking at order  $m$ . Clearly, in the models constructed in the proof of Proposition 1, one can put very small probabilities to the states with zero probability, making sure that wishful thinkers remain  $\varepsilon$ -wishful thinkers. Hence,  $Y^m$  contains  $X^m$ . When  $\varepsilon$  is large,  $Y^m$  may contain more strategy profiles. On the other hand, when the game is generic in the sense that payoffs at different strategy profiles are always different and  $\varepsilon$  is sufficiently small, the conclusion of Lemma 1 remains valid for  $\varepsilon$ -wishful thinkers (as shown by Lemma 3 in the Appendix). In that case,  $Y^m$  is also contained in  $X^m$ , and therefore  $X^m = Y^m$ .

#### 4. NASH EQUILIBRIUM AND THE COMMON KNOWLEDGE OF WISFUL THINKING

In this section, I will focus on the strategy profiles that are consistent with common knowledge of wishful thinking. These strategy profiles are important for two reasons. Firstly, these are the strategy profiles that remain possible as we allow more and more levels of knowledge of wishful thinking. More importantly, it is highly desirable from a methodological point of view to disentangle the implications of heterogeneous priors or self-serving biases from those of asymmetric

information. Researchers commonly accomplish this by assuming that there is no asymmetric information.<sup>7</sup> (This also allows the researcher to focus on belief differences without having to deal with asymmetric information.) In keeping with this methodology, it is desirable to examine the implications of wishful thinking when there is no asymmetric information about it—i.e., when wishful thinking is common knowledge.

I will establish a close relationship between the pure strategy Nash equilibria and common knowledge of wishful thinking. Firstly, since wishful thinking is stronger than rationality, common knowledge of wishful thinking will lead to a refinement of rationalizability, i.e., all non-rationalizable strategies will eventually be eliminated. Secondly, any Nash equilibrium in pure strategies will survive the iterated elimination process. More surprisingly, I will show that for generic two person games, these are the only strategies that survive. This will yield an unlikely epistemic characterization for pure-Nash-equilibrium strategies.

Let

$$NE = \bigcap_{i \in N} B_i$$

be the set of all pure-strategy Nash equilibria; recall that  $B_i$  is the set of profiles in which player  $i$  plays a best reply to others' strategies. Let also

$$NE_i = \{s_i | \exists s_{-i} : (s_i, s_{-i}) \in NE\}$$

be the set of all Nash-equilibrium strategies of a player  $i$ . Similarly, let

$$X_i^\infty = \{s_i | \exists s_{-i} : (s_i, s_{-i}) \in X^\infty\}$$

be the set of all strategies of a player  $i$  consistent with common knowledge of wishful thinking.

Recall that, if  $s \in B_i$ , we do not eliminate  $s$  for player  $i$ . But for any  $s \in NE$ ,  $s \in B_i$  for each player  $i$ , and hence  $s$  is not eliminated for any player. Therefore,  $s$  survives the elimination process. This yields the following result.

---

<sup>7</sup>For example, all of the papers mentioned in Footnote 3 make this assumption. See Squintani (2006) for an exception.

**Proposition 2.** *Every pure strategy Nash equilibrium is consistent with common knowledge of wishful thinking:*

$$NE \subseteq X^\infty.$$

In a pure-strategy Nash equilibrium there is no strategic uncertainty, and hence players do not have any freedom of entertaining different beliefs. Hence, all players, independent of their level of optimism, hold the same correct beliefs. Formally, any pure-strategy Nash equilibrium is consistent with a model with a single state at which each player plays according to the equilibrium. In such a model, it is common knowledge that each player is a wishful thinker. The next proposition states a more surprising and substantive fact about two-player games. The assumptions of this proposition generically hold.<sup>8</sup>

**Proposition 3.** *For any two-person game, assume (i) for each  $s_{-i}$ , there exists a unique best reply  $s_i \in BR_i(s_{-i})$ , and (ii) Player 1 is not indifferent between any two distinct strategy profiles  $s, s' \in B_1$ . Then, only pure Nash equilibrium strategies are consistent with common knowledge of wishful thinking:*

$$(4.1) \quad X_i^\infty = NE_i \quad (\forall i \in N).$$

Moreover,

$$(4.2) \quad X^\infty = \phi(NE_1 \times NE_2) \\ = \{(\hat{s}_1, \hat{s}_2) | \exists (s_1, s_2), (s_1, \hat{s}_2) \in NE : u_1(\hat{s}_1, s_2) \geq u_1(s_1, \hat{s}_2), \\ u_2(s_1, \hat{s}_2) \geq u_2(\hat{s}_1, s_2)\}.$$

The first part of the proposition characterizes common knowledge of wishful thinking in terms of strategies. It states that in a generic two-person game, the only strategies that are consistent with common knowledge of wishful thinking are the pure Nash equilibrium strategies. The second part gives a practical characterization for the strategy profiles that are consistent with common knowledge

---

<sup>8</sup>By definition, wishful thinking is an ordinal notion. The strategic implications of wishful thinking are invariant to monotonic transformations of payoff functions. The only non-generic situations one must ever rule out regarding wishful thinking are indifferences between certain pure strategy profiles.

of wishful thinking. It states that this set can be computed by simply applying the elimination procedure only once to the set  $NE_1 \times NE_2$ .

The proof can be summarized as follows. First, under assumption (i), one shows that the restriction of the best response function of each player  $i$  to  $X_j^\infty$  is well-defined, one-to-one, and onto  $X_i^\infty$ . Together with assumption (ii), this allows one to rearrange the strategies so that  $B_1 \cap X^\infty$  is equal to the diagonal of  $X_1^\infty \times X_2^\infty$  and the payoff of Player 1 is decreasing along this diagonal—e.g.  $(b, \beta)$  and  $(c, \gamma)$  in Example 1. This implies that all the strategy profiles that are under the diagonal and in  $X_1^\infty \times X_2^\infty$  must have been eliminated—e.g.  $(c, \beta)$  in Example 1. Using arguments similar to the one used to eliminate strategy profile  $(a, \alpha)$ , one then shows that if Player 2 does not give a best reply in a strategy profile on the diagonal, then the strategy of Player 2 must have been eliminated. Therefore, the diagonal of  $X_1^\infty \times X_2^\infty$  is equal to  $NE$ . This proves (4.1). Since the elimination process is monotonic, (4.1) implies that the result would not change if one started elimination from the set  $NE_1 \times NE_2$ . The latter elimination stops at the first step, yielding (4.2).

This provides an unusual epistemic characterization for pure-Nash-equilibrium strategies in terms of common knowledge of wishful thinking. More importantly, it suggests that there is little room left for optimism or pessimism when the wishful thinking is common knowledge. In fact, (4.2) establishes that the strategic uncertainty is reduced to uncertainty about the equilibrium played. Finally, this characterization provides partial support for the theoretical literature that uses equilibrium analysis to study the behavior of optimistic players. It shows that, given a generic, two-player model, if a researcher allows wishful thinking but sticks to the methodology in this literature, then she can simply focus on equilibrium strategies.

I must emphasize that common knowledge of wishful thinking is characterized by equilibrium strategies—*not by equilibria*. As stated in (4.2), two players may play equilibrium strategies that correspond to two different equilibria at some state in which wishful thinking is common knowledge. In such a state there is still substantial strategic uncertainty remaining, and players exhibit a clear form of

wishful thinking. For example, at state  $(b, \gamma)$  in Example 1, each player incorrectly believes that they will play *her* favorite equilibrium. This is a general fact. By (4.2), if the outcome  $\hat{s} = (\hat{s}_1, \hat{s}_2) \in X^\infty$  is not an equilibrium already, then there are two equilibria  $(\hat{s}_1, s_2)$  and  $(s_1, \hat{s}_2) \in NE$  that the players consider as possible and rank in diagonally opposing orders:  $u_1(\hat{s}_1, s_2) > u_1(s_1, \hat{s}_2)$  and  $u_2(s_1, \hat{s}_2) \geq u_2(\hat{s}_1, s_2)$ . Each player plays according to her own favorite equilibrium, believing that her own favorite equilibrium is to be played.

If the equilibria are strictly Pareto-ranked, then the players cannot have such opposing rankings. In that case, the outcome is necessarily an equilibrium.

**Corollary 1.** *Under the assumptions of Proposition 3, if the equilibria are Pareto-ranked with strict inequalities, then*

$$X^\infty = NE.$$

Proposition 3 has established already that for generic two-person games, when wishful thinking is common knowledge, strategic uncertainty is reduced to uncertainty about which equilibrium strategies are played. As in Example 1, one can indeed construct a model in which wishful thinking is common knowledge, all strategy profiles that are consistent with common knowledge of wishful thinking are played, and at each state each player assigns probability 1 to an equilibrium. That is, the players are in agreement that an equilibrium is played, but they may disagree about which equilibrium is played. This is stated by the following corollary, which was suggested by Haluk Ergin.

**Corollary 2.** *Under the assumptions of Proposition 3, there exists a model in which  $\sigma(CK(W)) = X^\infty$  and at each  $\omega \in CK(W)$ , each player  $i$  is certain that an equilibrium is played, i.e.,  $p_{i,\omega}(\sigma^{-1}(NE) \cap I_i(\omega)) = 1$ .*

**Remark 3.** Common knowledge of wishful thinking differs from usual epistemic foundations for equilibrium, such as the sufficient conditions of Aumann and Brandenburger (1995). For example, at state  $(b, \gamma)$  in Example 1, rationality is common



knowledge, but the players do not know each other's conjectures, and these conjectures do not form a Nash equilibrium. Notice that each player is certain about the other player's conjecture, but she is wrong.

**Counterexamples.** The next three examples illustrate that the assumptions in Proposition 3 are not superfluous. The first example illustrates that, with three or more players, there may be non-equilibrium strategies that are consistent with common knowledge of wishful thinking.

**Example 2.** Consider the following three-player game where Player 3 chooses between the matrices  $\lambda$  and  $\rho$ .

	$\lambda$			$\rho$	
	$l$	$r$		$l$	$r$
$t$	$2^*, -\varepsilon_2, -\varepsilon_3$	$-\varepsilon_1, 0^*, -\varepsilon_3$	$t$	$-\varepsilon_1, \varepsilon_2^*, \varepsilon_3^*$	$-\varepsilon_1, -\varepsilon_2, \varepsilon_3^*$
$b$	$-\varepsilon_1, -\varepsilon_2, -\varepsilon_3$	$1^*, 1^*, 1^*$	$b$	$\varepsilon_1^*, \varepsilon_2^*, \mathfrak{Z}^*$	$\varepsilon_1^*, -\varepsilon_2, -\varepsilon_3$

Above, in order to satisfy the genericity conditions in Proposition 3, I take  $\varepsilon_i : S \rightarrow (0, 1)$  as an arbitrary function of strategy profiles for each  $i \in N$ ; the arguments are suppressed for simplicity. Check that  $NE = \{(b, r, \lambda), (b, l, \rho)\}$ . I will show that  $X^\infty = S \setminus \{(b, l, \lambda)\}$ , so that the non-equilibrium strategy  $t$  is consistent with common knowledge of wishful thinking. In the first round of elimination, strategy profile  $(b, l, \lambda)$  is eliminated. This is because the highest payoff Player 1 gets when she plays  $b$  is 1, and this payoff is smaller than the payoff she gets by playing a best response to  $(l, \lambda)$  (i.e. a payoff of 2). No other strategy profile is eliminated in any round. To see this, ignore the eliminated strategy profile  $(b, l, \lambda)$ . Consider  $(t, l, \lambda)$ . Since  $(t, l, \lambda) \in B_1$ , it is not eliminated for Player 1. When Player 2 plays  $l$ , she gets a positive payoff at  $(b, l, \rho) \in B_2$ , and her payoff is zero when she plays a best response to  $(t, \lambda)$ . Hence,  $(t, l, \lambda)$  is not eliminated for Player 2, either. We do not eliminate  $(t, l, \lambda)$  for Player 3 because Player 3 gets payoff of 1 at  $(b, r, \lambda) \in B_3$ , while she would get only  $\varepsilon_3(t, l, \rho) < 1$  if she played a best response to  $(t, l)$ . One can check similarly that no other strategy profile is eliminated. Therefore,  $X^\infty = S \setminus \{(b, l, \lambda)\}$ .

The fact that  $(t, l, \lambda)$  is not eliminated in this example while  $(a, \alpha)$  is eliminated in Example 1 illustrates a key difference between the case of two players and the case of many players. In the case of two players, when we fix the strategy of Player 2 as in the top-left corner, namely at  $(a, \alpha)$ , we can only consider the changes in strategies of Player 1. When  $(a, \alpha)$  is not in  $B_2$  already,  $\alpha$  can be a best reply to a strategy  $s_1$  only if  $(s_1, \alpha)$  is under the graph of  $B_1$ , and such strategy profiles have been eliminated in the first round. On the other hand, with more players, when we fix the strategy of Player 2, we can consider joint changes in strategies of multiple players and thereby consider other strategy profiles that are not under the graph of  $B_1$ , and therefore need not have been eliminated in the first round. Some of these strategy profiles, such as  $(b, l, \rho)$ , may be in  $B_2$  and may yield high payoff for Player 2. Hence, the proof of Proposition 3 is not valid for more than two players, and the proposition is not necessarily true in these games.

The next two examples show that the genericity assumptions in Proposition 3 are not superfluous. Recall that these assumptions are (i) for each  $s_{-i}$ , there exists a unique best reply  $s_i \in BR_i(s_{-i})$  and (ii) Player 1 is not indifferent between any two distinct strategy profiles  $s, s' \in B_1$ . First, I consider assumption (i).

**Example 3.** Consider the game

	$l$	$r$
$t$	$2^*, 2^*$	$0, 2^*$
$b$	$0, 1^*$	$1^*, 0$

where assumption (i) fails while (ii) is satisfied. In this game, the only pure-strategy Nash equilibrium is  $(t, l)$ , while  $X^\infty = S \setminus \{(b, l)\}$ , containing non-equilibrium strategies  $b$  and  $r$ . Hence, assumption (i) is not superfluous. The strategy profile  $(b, l)$  is eliminated because  $u_1(t, l) > u_1(b, r)$  (as in Example 1). Ignoring  $(b, l)$ , let us check that no other strategy profile is eliminated. The Nash Equilibrium  $(t, l)$  is not eliminated. The strategy profile  $(t, r)$  is not eliminated for Player 1 because  $(t, l) \in B_1$  gives Player 1 the highest possible payoff (more than that of  $(b, r)$ ); it is not eliminated for Player 2 because  $(t, r) \in B_2$ . Since  $(b, r) \in B_1$ ,  $(b, r)$  is not eliminated for Player 1. Since  $(t, r) \in B_2$  yields the highest payoff for Player 2,  $(b, r)$  is not eliminated for Player 2, either.

The next example illustrates that assumption (ii) is not superfluous.

**Example 4.** Consider the matching-pennies game

	$l$	$r$
$t$	$1^*, -1$	$-1, 1^*$
$b$	$-1, 1^*$	$1^*, -1$

where assumption (ii) fails, while assumption (i) holds. In this game,  $NE = \emptyset$ , but no strategy is eliminated, and  $X^\infty = S$ . To see this, consider  $(b, l)$ . The payoff of Player 1 at  $(b, r) \in B_1$  is not less than what she would get if she played a best response to  $l$ . Hence,  $(b, l)$  is not eliminated for Player 1. Since  $(b, l) \in B_2$ , it is not eliminated for Player 2, either.

**Non-emptiness of  $X^\infty$ .** By perturbing the payoff of Player 1 at  $(t, l)$  in the matching-pennies game, one can satisfy assumption (i). Since there is no pure-strategy Nash equilibrium, by Proposition 3,  $X^\infty$  must be empty in the new game. This is illustrated in the following example.

**Example 5.** Consider the game

	$l$	$r$
$t$	$2^*, -1$	$-1, 1^*$
$b$	$-1, 1^*$	$1^*, -1$

where there is no pure-strategy Nash equilibrium. In the first round  $(b, l)$  is eliminated for Player 1, as the best reply to  $l$  now yields a higher payoff of 2 than the highest payoff she can get when she plays  $b$ . No other strategy is eliminated in this round. In the second round,  $(t, l)$  is eliminated for Player 2 as there is no remaining strategy profile  $(s_1, l)$  with  $(s_1, l) \in B_2$ . In the third round,  $(t, r)$  is eliminated for Player 1 because, since  $(t, l)$  is eliminated in the previous round, there is no remaining strategy profile  $(t, s_2)$  with  $(t, s_2) \in B_1$ . In the fourth round, we eliminate  $(b, r)$  for Player 2 because now there is no strategy profile  $(s_1, r)$  with  $(s_1, r) \in B_2$ . At this point, all of the strategy profiles have been eliminated, and therefore,  $X^3 = X^\infty = \emptyset$ .

The underlying reason for the last elimination is as follows. If Player 2 knows at a state  $\omega$  that it is second-order mutual knowledge that players are wishful thinkers, then  $\sigma(I_2(\omega)) = \{(b, r)\}$ , for  $X^2 = \{(b, r)\}$ . In that case, the expected payoff of Player 2 at  $\omega$  is  $-1$ , while she could obtain the higher payoff of 1 by playing a best response to  $b$ , a contradiction. This example illustrates that there are games for which one cannot write a logically consistent model in which wishful thinking is mutually known at high orders. (Similar examples have been found for some intuitive rationality notions.) More generally, Propositions 2 and 3 immediately yield a characterization of the games for which  $X^\infty$  is non-empty:

**Corollary 3.** *If  $NE \neq \emptyset$ , then  $X^\infty \neq \emptyset$ . Under the assumptions of Proposition 3, the converse is also true, and therefore,*

$$X^\infty \neq \emptyset \iff NE \neq \emptyset.$$

**Mixed Strategies.** This paper considers only pure strategies. If players have access to randomization devices on strategies, then one can introduce corresponding mixed strategies as pure strategies. In the new game the payoffs are computed as the expected payoffs when some player uses one of these previously mixed strategies. If players have access to devices that lead to equilibrium mixing, then  $X^\infty$  will be non-empty in the resulting game, as the next example illustrates.

**Example 6.** In Example 5, there is a Nash equilibrium  $(m_1, m_2)$  in mixed strategies, where  $m_1$  assigns equal probabilities to  $t$  and  $b$  while  $m_2$  assigns probabilities  $2/5$  and  $3/5$  to  $l$  and  $r$ , respectively. Now, consider the game in which these equilibrium strategies introduced as pure strategies:

	$l$	$r$	$m_2$
$t$	$2^*, -1$	$-1, 1^*$	$1/5^*, 1/5$
$b$	$-1, 1^*$	$1^*, -1$	$1/5^*, -1/5$
$m_1$	$1/2, 0^*$	$0, 0^*$	$1/5^*, 0^*$

In the new game, the previous mixed-strategy equilibrium is a Nash equilibrium in pure strategies, and  $NE = \{(m_1, m_2)\}$ . By Proposition 2,  $(m_1, m_2) \in X^\infty$ , so that  $X^\infty$  is no longer empty. Indeed, one can easily check that  $X^\infty = X^4 = \{(m_1, m_2)\}$ .

In the first round, the strategy profiles  $(b, l)$ ,  $(m_1, l)$ , and  $(m_1, r)$  are eliminated for Player 1 as in Example 1. In this round, we also eliminate  $(t, m_2)$  and  $(b, m_2)$  for Player 2 because the payoff from  $(m_1, m_2)$ —the only strategy profile with  $(s_1, m_2) \in B_2$ —is lower than the payoff of 1 that she would get if she played a best response to the strategy of Player 1. Given these eliminations, strategy profiles  $(t, l)$ ,  $(t, r)$ , and  $(b, r)$  are eliminated in the second, third, and fourth rounds, respectively, as in Example 5. The only remaining strategy profile is  $(m_1, m_2)$ , which is never eliminated.

Analyzing strategic outcomes in the presence of systematic biases like wishful thinking necessitates a re-examination of modeling mixed strategies. Traditionally, mixed strategies are considered to capture deliberate randomizations by players. The more current understanding is that mixed strategies are the other players' common conjecture about the player's strategy. (Deliberate randomizations can be introduced as pure strategies.) Consequently, in epistemic literature, it is usual to restrict attention to pure strategies. But, under the traditional paradigm of rationality, this is only a philosophical issue. The results are not fundamentally altered if we allow explicit mixing to be part of a player's strategy set. For example, when we allow mixed strategies, the strategies that are consistent with common knowledge of rationality are the ones with support on the set of pure strategies that survive the usual iterated elimination of strictly dominated strategies. On the other hand, the issue of explicit mixing is a real issue with significant consequences when we consider wishful thinkers (or players with other biases or general decision rules). As illustrated in the above example, the outcome is now highly sensitive to whether mixed strategies are allowed as explicit pure strategies. When a researcher analyzes the general decision rules or biases, such as wishful thinking, he might need to understand which randomization devices the players can use. Introduction of mixed strategies has such a great impact because when we allow explicit mixing, the opponent is restricted to play the mixed strategy rather than the pure strategies in its support. This restricts the beliefs a wishful thinker may entertain.

A more basic issue regarding biases is that, unlike in the case of equilibrium analysis, we also ought to reconsider the beliefs players would form on these randomization devices, as the beliefs of players with biases need not coincide with "objective" probabilities that the researcher assigns. In particular, if players are truly wishful thinkers, then they would also assign probability 1 on the most favorable outcome of the randomization, and introduction of mixed strategies would become an unnecessary complication.

**Robustness to higher-order knowledge assumptions.** It is important to understand how the characterization in Proposition 3 changes when we relax the assumption that wishful thinking is common knowledge, as it is difficult to check whether such assumptions are satisfied. As in the examples so far, the elimination process often stops quickly after few rounds. In that case, Nash equilibrium strategies will be played whenever wishful thinking is mutually known at few orders. I will now present a general class of games in which the elimination process stops just at the *third* round, and hence second-order mutual knowledge of wishful thinking will be sufficient to conclude that players will play some Nash equilibrium strategies.

**Definition 1.** A two-player game  $(N, S, u)$  with linearly-ordered strategy sets is said to be a *monotone game with strategic substitutes* iff

- (*monotonicity*) for all  $(s_1, s_2), (s'_1, s'_2) \in B_i$  with  $s_i > s'_i$ ,  $u_i(s_1, s_2) > u_i(s'_1, s'_2)$ , and
- (*strategic substitutes*) for each  $s_{-i}$ , there exists a unique best reply  $BR_i(s_{-i})$ , and  $BR_i(s_{-i})$  is non-increasing in  $s_{-i}$ .

That is, if a player  $j$  increases her strategy, the best response of player  $i$  weakly decreases, and the overall change weakly decreases the payoff of  $i$ . Cournot competition and the battle of the sexes are in this large class of games. These games possess pure Nash equilibria and satisfy the conditions of Proposition 3. Hence,  $X^\infty$  is non-empty and involves only equilibrium strategies. Indeed, Nash equilibrium strategies are played whenever everyone knows that everyone knows that everyone is a wishful thinker, as the next result establishes.

**Proposition 4.** *For any monotone game with strategic substitutes, there are only three rounds of elimination:*

$$(4.3) \quad X^\infty = X^2 = \{(\hat{s}_1, \hat{s}_2) | \exists (\hat{s}_1, s_2), (s_1, \hat{s}_2) \in NE : \hat{s}_1 \geq s_1, \hat{s}_2 \geq s_2\}.$$

*In particular,  $X_i^2 = NE_i$  for each  $i$ .*

The proof shows that all non-equilibrium strategies are eliminated in the first three rounds, i.e.,  $X^2 \subseteq NE_1 \times NE_2$ . The elimination in the first round therefore ensures that  $X^2$  is as in (4.3). Since  $X^2$  is identical to the set in (4.2), by Proposition 3, there is no further elimination. The following example illustrates this.

**Example 7.** Consider a Cournot duopoly in which each player  $i$  produces  $q_i$  units of a good, where the best-response functions  $BR_1$  and  $BR_2$  are as plotted in Figure 3, with  $q_1$  and  $q_2$  are measured on the horizontal and vertical axes, respectively. There are three Nash equilibria:  $(q_1^1, q_2^1)$ ,  $(q_1^2, q_2^2)$ , and  $(q_1^3, q_2^3)$ . In the first round, strategies  $q_1$  with  $q_1 > \bar{q}_1$  and  $q_2$  with  $q_2 > \bar{q}_2$  are eliminated because they are never a best reply. Moreover, every strategy profile that is under the graph of some best-response function is eliminated. To see this, consider  $(\hat{q}_1, \hat{q}_2)$  in Figure 3. If  $(\hat{q}_1, \hat{q}_2)$  is played at a state at which Player 2 is a wishful thinker, then her expected payoff at that state is  $u_2(q_1', \hat{q}_2)$ . If she plays a best reply to  $\hat{q}_1$ , then her payoff will be  $u_2(\hat{q}_1, BR_2(\hat{q}_1))$ , which is higher than  $u_2(q_1', \hat{q}_2)$  by monotonicity. Therefore,  $(\hat{q}_1, \hat{q}_2)$  is eliminated. Now consider any  $\tilde{q}_1$  with  $q_1^1 < \tilde{q}_1 < q_1^2$ . At  $q_1 = \tilde{q}_1$ , the graph of  $BR_1$  lies under the graph of  $BR_2$ , and hence the profile  $(\tilde{q}_1, q_2)$  with  $\tilde{q}_1 = BR_1(q_2)$  is eliminated in the first round. Therefore, strategy  $\tilde{q}_1$  is eliminated in the second round. Similarly, in the second round, we eliminate all strategies  $q_1$  with  $q_1^3 < q_1 < \bar{q}_1$  and  $q_2$  with  $q_2^3 < q_2 < q_2^2$  or  $q_2^1 < q_2 < \bar{q}_2$ . Now, in the third round, every strategy  $q_1$  with  $q_1^2 < q_1 < q_1^3$  is eliminated because  $q_1$  can be a best reply to only a strategy  $q_2$  with  $q_2^3 < q_2 < q_2^2$ , which was eliminated in the last round. Similarly each  $q_2$  with  $q_2^1 < q_2 < q_2^2$  is eliminated in this round. Notice that the only remaining strategies are the Nash equilibrium strategies, and  $X^\infty = X^2 = \{(q_1^1, q_2^1), (q_1^2, q_2^2), (q_1^2, q_2^1), (q_1^3, q_2^3), (q_1^3, q_2^2), (q_1^3, q_2^1)\}$ . When  $(q_1^3, q_2^1)$  is

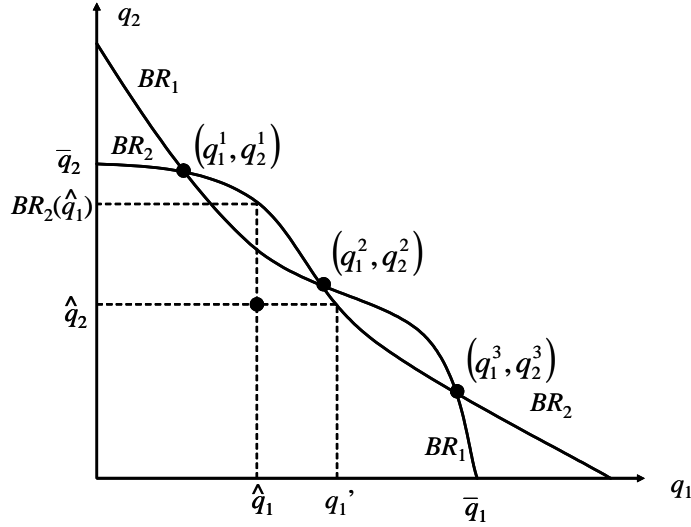


FIGURE 3. Wishful thinking in a general Cournot duopoly.

played, the total supply is more than what any equilibrium predicts, leading to a price that is lower than any equilibrium price.

Notice that in the case of linear Cournot duopoly, there will be only two rounds of elimination. Moreover, by merely knowing that players are wishful thinkers, we can conclude that the total supply weakly exceeds the equilibrium supply, leading to a price weakly lower than the equilibrium price.

Proposition 4 shows that, in monotone games with strategic substitutes, knowledge assumptions beyond the second order do not play any role. This stems from the fact that, in these games, a wishful thinker tends to gain when her own wishful thinking is known, and she tends to assume it whenever she can. Towards illustrating this, notice that, by monotonicity, given any two states  $\omega$  and  $\omega'$  with  $\sigma(\omega), \sigma(\omega') \in B_i$ , we have  $u_i(\sigma(\omega)) > u_i(\sigma(\omega'))$  whenever  $\sigma_i(\omega) > \sigma_i(\omega')$ . Hence, roughly speaking, say that player  $i$  is *more optimistic at  $\omega$  than  $\omega'$*  iff  $\sigma_i(\omega) > \sigma_i(\omega')$ . Assume also that  $u_i(BR_i(s_{-i}), s_{-i}) < u_i(BR_i(s'_{-i}), s'_{-i})$  whenever  $s_{-i} > s'_{-i}$ , an assumption that is implied by monotonicity and strategic substitutes when  $BR_i(s_{-i}) \neq BR_i(s'_{-i})$ . Now consider a state  $\omega$  at which a



player  $i$  is a wishful thinker and  $\sigma_j$  is one-to-one on  $I_i(\omega)$ , where  $i \neq j$ . Being a wishful thinker, player  $i$  then assigns probability 1 to the state  $\omega^*$  with  $\sigma_j(\omega^*) = \min \sigma_j(I_i(\omega))$ . That is, player  $i$  believes that the other player  $j$  is of the most pessimistic type within  $I_i(\omega)$ . Moreover, since  $BR_j$  is non-increasing, if player  $j$  assigns probability 1 to states  $\hat{\omega}^*$  and  $\hat{\omega}'$  at states  $\omega^*$  and  $\omega' \in I_i(\omega)$ , respectively, then  $\sigma_i(\hat{\omega}^*) > \sigma_i(\hat{\omega}')$ . In that sense, a wishful thinker  $i$  believes that the other player is of the type that finds  $i$  most optimistic.

Now imagine that player  $i$  happens to know that the other player  $j$  is also a wishful thinker. Firstly, since  $j$  plays a best reply to a pure strategy,  $\omega^*$  is indeed the state in  $I_i(\omega)$  at which player  $j$  is of the type that is most pessimistic and finds  $i$  most optimistic. More interestingly, there cannot be any state  $\omega' \in I_i(\omega) \setminus \{\omega^*\}$  such that  $I_j(\omega') \subseteq I_j(\omega^*)$ .<sup>9</sup> In other words, player  $i$  believes that the other player  $j$  has a minimal amount of strategic uncertainty. The intuition is simple. Since optimism of  $j$  hurts player  $i$ , player  $i$  wishes that  $j$  faces relatively small uncertainty because a wishful thinker  $j$  becomes more optimistic as her uncertainty increases. In particular, if there is a state  $\omega' \in I_i(\omega)$  at which player  $j$  knows the state (i.e.,  $I_j(\omega') = \{\omega'\}$ ), then player  $i$  assigns probability 1 to that state. In that case, at  $\omega'$ , there will be common certainty of the state, and an equilibrium will be played.

On the other hand, in a monotone game with *strategic complementarity*, optimism of a player is helpful to the other player. In that case, a wishful thinker  $i$  tends to believe that the other player  $j$  is optimistic, but that optimism now will predict that player  $i$  also holds optimistic beliefs. Once again, a wishful thinker tends to believe that she is perceived as very optimistic.

## 5. AN EXTENSION

In some applications, it might be appropriate to assume that only a subset of players have systematic biases. For example, one might want to consider the

---

<sup>9</sup>For in that case, by definition of wishful thinking, we have  $E_{i,\omega'}[u_j(\sigma)] \leq E_{i,\omega^*}[u_j(\sigma)]$ , showing by monotonicity that  $\sigma_j(\omega') \leq \sigma_j(\omega^*)$ , and contradicting that  $\sigma_j(\omega^*)$  is the unique minimum.

case that consumers have optimistic views about firms, while firms' have unbiased expectations about consumers. Assuming that each player plays a best reply to his beliefs, one can easily extend the analysis above to cover such cases.

Let  $\hat{N}$  be the set of wishful thinkers, and assume that the remaining players are rational. Formally, we consider the states  $\omega \in W_{\hat{N}}$  where

$$W_{\hat{N}} \equiv \left( \bigcap_{i \in \hat{N}} W_i \right) \cap \left( \bigcap_{i \notin \hat{N}} R_i \right).$$

Now, Lemma 1 summarizes the restrictions on strategies of wishful thinkers. The remaining players play a best reply to some belief that is consistent with their knowledge. That is, their strategies must not be strictly dominated given the strategies that may be "available" to the other players. This is the only restriction on rational players' strategies.

### Elimination Procedure.

- (1) Initialization: Set  $X_{\hat{N}}^{-1} = S$ .
- (2) Elimination ( $m \geq 0$ ): For all  $i \in \hat{N}$ , eliminate all strategy profiles  $\hat{s}$  with  $\max_{s_i} u_i(s_i, \hat{s}_{-i}) > \max_{(\hat{s}_i, s_{-i}) \in B_i \cap X_{\hat{N}}^{m-1}} u_i(\hat{s}_i, s_{-i})$ . For all  $i \notin \hat{N}$ , eliminate each strategy  $\hat{s}_i$  that is strictly dominated given that the set of available strategies for that other players is  $\{s_{-i} \mid (\hat{s}_i, s_{-i}) \in X_{\hat{N}}^{m-1}\}$ . Call the remaining strategy profile  $X_{\hat{N}}^m$ .
- (3) Iterate step (2).

Notice that the set of "available" strategies for other players depends on the strategy for which we check rationality of the player. This is because we eliminate strategy profiles, rather than strategies, when we consider wishful thinking, so that  $X_{\hat{N}}^{m-1}$  is not rectangular. The next result states that  $X_{\hat{N}}^m$  is indeed the set of strategy profiles that is consistent with  $m$ th-order mutual knowledge of the situation that the players in  $\hat{N}$  are wishful thinkers and the remaining players play a best reply.

**Proposition 5.** *For any model  $(\Omega, I, p, \sigma)$ , any  $\hat{N}$ , and any  $m \geq 0$ ,*

$$\sigma(K^m(W_{\hat{N}})) \subseteq X_{\hat{N}}^m \text{ and } \sigma(CK(W_{\hat{N}})) \subseteq X_{\hat{N}}^\infty.$$

Moreover, there exist models  $(\Omega, I, p, \sigma)$  in which the above inclusions are equalities.

The first statement is implied by Lemma 1 and the definition of  $R_i$ . One can construct models with equalities as in the proof of Proposition 1. The next example illustrates how the modified elimination procedure is applied.

**Example 1** (*continued*). In the original game

	$\alpha$	$\beta$	$\gamma$	$\delta$
$a$	<b><math>3^*, 0</math></b>	<b><math>-1, 0</math></b>	<b><math>0, 0</math></b>	<b><math>0, 2^*</math></b>
$b$	$0, 0$	<b><math>2^*, 1^*</math></b>	<b><math>0, 0</math></b>	<b><math>0, 0</math></b>
$c$	$0, 0$	$0, 0$	<b><math>1^*, 2^*</math></b>	<b><math>1^*, 0</math></b>
$d$	$2, 3^*$	$1, 0$	$0, 0$	$0, 0$

assume that Player 1 is a wishful thinker and Player 2 is rational; i.e.,  $\hat{N} = \{1\}$ . Take  $m = 0$ . For Player 1, as before, the strategy  $d$  and strategy profiles  $(b, \alpha)$ ,  $(c, \alpha)$ , and  $(c, \beta)$  are eliminated. No strategy is eliminated for Player 2. The remaining strategy profiles,  $X_{\{1\}}^0$ , are typed in bold in the above table. For  $m = 1$ , strategy  $\alpha$  is eliminated because  $\alpha$  is not a best reply to  $a$ , the only “remaining strategy” of Player 1 for  $\alpha$ . ( $\{s_1 \mid (s_1, \alpha) \in X_{\{1\}}^0\} = \{(a, \alpha)\}$ .) No other strategy of Player 2 is eliminated because each of them is clearly a best reply to a remaining (pure) strategy. For  $m = 2$ , strategy  $a$  is eliminated as it is not a best reply to any pure strategy, and hence the inequality in the elimination procedure is vacuously satisfied. For  $m = 3$ , strategy  $\delta$  is eliminated, as it becomes strictly dominated by a mixture of  $\beta$  and  $\gamma$  when strategies  $a$  and  $d$  are ignored. The elimination process stops here. Therefore,  $X_{\{1\}}^3 = \dots = X_{\{1\}}^\infty = \{(b, \beta), (b, \gamma), (c, \gamma)\} = X^\infty$ , as in the original case.

**Remark 4.** Notice that the outcome is not affected by whether Player 2 is a wishful thinker or not. The power of elimination process comes from elimination of strategy profiles for a player at the first stage. In fact, given a generic, two-player game in which a strategy of Player 1 can be a best reply to at most one strategy of Player 2, if Player 1 is a wishful thinker, we can re-label the strategies so that the strategy profiles under the graph of best reply correspondence for

Player 1 are eliminated for  $m = 0$ . When we know that Player 2 plays a best reply to a pure strategy, all of the arguments in the proof of Proposition 3 would go through. Therefore, once again, for generic, two-player games, we can characterize the Nash equilibrium strategies as the strategies that are consistent with common knowledge of the case that one of the players is a wishful thinker and the other plays a best reply to a pure strategy. Since only Player 1 is a wishful thinker, I do not assume here that a wishful thinker knows that some other player is a wishful thinker. Hence, the results in this paper do not rely on such assumptions, which some may find problematic.

## 6. CONCLUSION

Self-serving biases are reportedly common. If this is indeed the case, one would expect these biases to manifest themselves in the beliefs about the other players strategies. Even in the literature on optimism, however, this sort of systematic biases about the other players' strategies are typically assumed away. What is worse, there is no game theoretical framework that incorporates such systematic deviations from the common-prior assumption—although the use of heterogenous priors about strategies is becoming mainstream in game theory. This paper takes a first step towards a theory of such deviations, focusing on the extreme form of optimism, namely wishful thinking. In particular, it develops a framework for analyzing one type of deviation, namely wishful thinking, the extreme form of optimism. I use the canonical model for strategic uncertainty to identify whether a player is a wishful thinker and develop a straightforward elimination process directly on strategy profiles to characterize the set of strategy profiles that are consistent with wishful thinking, mutual knowledge of wishful thinking, and so on. I further show that in generic two-person games, pure Nash-equilibrium strategies are the only strategies that are consistent with common knowledge of wishful thinking. In such games, wishful thinking can be common knowledge only in cases in which strategic uncertainty is reduced to uncertainty about the equilibrium that is played, and if a researcher assumes away asymmetric information about whether a player is a wishful thinker in order to disentangle the effects of wishful thinking

from that of informational differences, then he may as well focus on equilibrium strategies.

My analysis of the example of wishful thinking illustrates how one can use an epistemic approach to find out the strategic implications of common knowledge of a general decision rule—while allowing systematic biases about the strategies. It also illustrates that, in such an analysis, we will need to face new issues, such as elimination of strategy profiles and sensitivity to modeling of strategies. The elementary nature of my analysis of this particular bias suggests that it will be possible to develop a tractable framework for incorporating more general types of systematic biases about strategies into game theory.

#### APPENDIX A. TECHNICAL APPENDIX—PROOFS

*Proof of Fact 1.* For any  $i$  and  $\omega$  as in (2.2), we have (2.1):

$$E_{i,\omega} [u_i(\sigma)] = \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s) = \max_{s_i \in S_i, \mu \in \Delta(I_i(\omega))} E_\mu [u_i(s_i, \sigma_{-i})] = \max_{s_i \in S_i} E_{i,\omega} [u_i(s_i, \sigma_{-i})],$$

where  $\Delta(I_i(\omega))$  is the set of all probability distributions on  $I_i(\omega)$ , and  $E_\mu$  is the expectation with respect to  $\mu$ . [By (2.2), the second maximum is obtained at  $(\sigma_i(\omega), p_{i,\omega})$ , and hence the maximum does not change if we fix  $\mu = p_{i,\omega}$  in this maximization, yielding the last equality.]  $\square$

**A.1. Strategic Implications of Wishful Thinking.** I will now prove Lemma 1 and Proposition 1. Lemma 1 immediately follows from the following result.

**Lemma 2.** *For any  $F \subseteq \Omega$ ,  $i \in N$ , and any  $\hat{s} \in \sigma(K_i(F) \cap W_i)$ , there exists  $(\hat{s}_i, s_{-i}) \in B_i \cap \sigma(F)$  such that  $u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i})$ .*

*Proof.* Let  $\hat{s} = \sigma(\omega)$  for some  $\omega \in K_i(F) \cap W_i$ . Since  $\sigma_i(I_i(\omega)) = \{\hat{s}_i\}$  and the expectation of a random variable cannot be strictly higher than the variable everywhere, there exists  $(\hat{s}_i, s_{-i}) \in \sigma(I_i(\omega))$  such that

$$(A.1) \quad u_i(\hat{s}_i, s_{-i}) \geq E_{i,\omega} [u_i(\sigma)] = \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s),$$

where the equality is due to the fact that  $\omega \in W_i$ . Now, since  $s_{-i} \in \sigma_{-i}(I_i(\omega))$ , (A.1) implies that  $u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, s_{-i})$ , showing that  $\hat{s}_i \in BR_i(s_{-i})$ ,

and hence  $(\hat{s}_i, s_{-i}) \in B_i$ . Moreover, since  $\omega \in K_i(F)$ , we have  $I_i(\omega) \subseteq F$ , and hence  $(\hat{s}_i, s_{-i}) \in \sigma(I_i(\omega)) \subseteq \sigma(F)$ . Thus,  $(\hat{s}_i, s_{-i}) \in B_i \cap \sigma(F)$ . But, since  $\hat{s}_{-i} = \sigma_{-i}(\omega) \in \sigma_{-i}(I_i(\omega))$ , (A.1) implies that

$$u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i}).$$

□

The next result shows that for generic games, if we require players to put positive probability at all states and consider  $\varepsilon$ -wishful thinking for sufficiently small  $\varepsilon$ , the conclusion of Lemma 1 remains valid.

**Lemma 3.** *For some  $i$ , assume that  $u_i(s) \neq u_i(s')$  for all  $s, s' \in S$ . Let  $\varepsilon < \min_{s, s' \in S} |u_i(s) - u_i(s')|$ . Then, for any  $F \subseteq \Omega$  and any  $\hat{s} \in \sigma(\omega)$  where  $\omega \in K_i(F)$  and player  $i$  is an  $\varepsilon$ -wishful thinker at  $\omega$ , there exists  $(\hat{s}_i, s_{-i}) \in B_i \cap \sigma(F)$  such that  $u_i(\hat{s}_i, s_{-i}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i})$ .*

*Proof.* Let  $\tilde{s} = \max_{s \in S_i \times \sigma_{-i}(I_i(\omega))} u_i(s) \in B_i$ . Since  $i$  is an  $\varepsilon$ -wishful thinker at  $\omega$ , we have  $E_{i, \omega}[u_i(\sigma)] \geq u_i(\tilde{s}) - \varepsilon$ . Hence, there exists  $\omega' \in I_i(\omega)$  such that  $u_i(\sigma(\omega')) \geq u_i(\tilde{s}) - \varepsilon$ . By definition of  $\tilde{s}$  and  $\varepsilon$ , it must be that  $\sigma(\omega') = \tilde{s}$ . Hence,  $\tilde{s} = (\hat{s}_i, s_{-i})$  for some  $s_{-i}$  (for  $\sigma_i(\omega') = \sigma_i(\omega) = \hat{s}_i$ ) and  $\tilde{s} \in \sigma(F)$  (for  $\omega \in K_i(F)$  and hence  $\omega' \in F$ ). Moreover, by definition,  $u_i(\tilde{s}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i})$ . □

*Proof of Proposition 1.* For  $m = 0$ , the statement,  $\sigma(W) \subseteq X^0$ , is immediately implied by Lemma 1. For any  $m$ , assume that  $\sigma(K^{m-1}(W)) \subseteq X^{m-1}$ . Take any  $\hat{s} = \sigma(\omega)$  for some  $\omega \in K^m(W)$ . Firstly, since  $K^m(W) \subseteq K^{m-1}(W)$ ,  $\hat{s} \in X^{m-1}$ . Moreover, for any  $i$ ,  $\omega \in K_i(K^{m-1}(W)) \cap W_i$ , so that, at  $\omega$ , player  $i$  is a wishful thinker and knows that a strategy profile in  $\sigma(K^{m-1}(W)) \subseteq X^{m-1}$  is played (i.e.,  $\sigma(I_i(\omega)) \subseteq \sigma(K^{m-1}(W)) \subseteq X^{m-1}$ ). Hence, by Lemma 1, there exists  $(\hat{s}_i, s_{-i}) \in B_i \cap X^{m-1}$  such that

$$\max_{s_i} u_i(s_i, \hat{s}_{-i}) \leq \max_{(\hat{s}_i, s'_{-i}) \in B_i \cap X^{m-1}} u_i(\hat{s}_i, s'_{-i}).$$

Thus,  $\hat{s} \in \phi(X^{m-1}) = X^m$ . Therefore,  $\sigma(K^m(W)) \subseteq X^m$ .

For any given  $m$ , I will now construct a model  $(\Omega, I, p, \sigma)$  in which  $\sigma(K^m(W)) = X^m$ . Specifically, for each  $\hat{s} \in X^m$ , I will construct a model  $(\Omega^{\hat{s}}, I^{\hat{s}}, p, \sigma)$  in which

$\hat{s} \in \sigma(K^m(W))$ . Then, the model with  $\Omega = \cup_{\hat{s} \in X^m} \Omega^{\hat{s}}$  and  $I = \cup_{\hat{s} \in X^m} I^{\hat{s}}$  satisfies the desired property. Fix any  $\hat{s} \in X^m$ . Take some  $\omega_0$  as the first member of  $\Omega^{\hat{s}}$  and set  $\sigma(\omega_0) = \hat{s}$ . Set  $F^m = \{\omega_0\}$ . (I will define a sequence  $F^0, \dots, F^m$  such that  $F^k \subseteq K^k(W)$  for each  $k$ .) For any player  $i$ , if  $\hat{s} \in B_i$ , then set  $I_i^{\hat{s}}(\omega_0) = \{\omega_0\}$ . Clearly,  $\omega_0 \in W_i$  for such a player  $i$ . If  $\hat{s} \notin B_i$ , since  $\hat{s} \in X^m$ , there exists  $s^{i,m-1} \equiv (\hat{s}_i, s_{-i}^{i,m-1}) \in B_i \cap X^{m-1}$  such that  $u_i(s^{i,m-1}) \geq \max_{s_i} u_i(s_i, \hat{s}_{-i})$ . For each such  $i$ , consider a new state  $\omega_{i,m-1}$  and set  $\sigma(\omega_{i,m-1}) = s^{i,m-1}$ ,  $I_i^{\hat{s}}(\omega_0) = \{\omega_0, \omega_{i,m-1}\}$ , and  $p_{i,\omega_0}(\omega_{i,m-1}) = 1$ . By construction,  $\omega_0 \in W_i$  for each such  $i$ . Write  $F^{m-1}$  for the set of states that are defined so far. Recall that, for each  $\omega_{i,m-1}$ ,  $I_i^{\hat{s}}(\omega_{i,m-1}) = I_i^{\hat{s}}(\omega_0)$  and  $\sigma_i(\omega_{i,m-1}) = \hat{s}_i$  have been defined already. Now for each  $j \neq i$ , conduct the last operation again: if  $s^{i,m-1} \in B_j$ , then set  $I_j^{\hat{s}}(\omega_{i,m-1}) = \{\omega_{i,m-1}\}$ , yielding  $\omega_{i,m-1} \in W_j$ . If  $s^{i,m-1} \notin B_j$  and  $m-1 \geq 0$ , then there exists  $s^{j,m-2} = (s_j^{i,m-1}, s_{-j}^{j,m-2}) \in B_j \cap X^{m-2}$  such that  $u_j(s^{j,m-2}) \geq \max_{s_j} u_j(s_j, s_{-j}^{j,m-2})$ . For each such  $j$ , consider a new member  $\omega_{j,m-2}$  and set  $\sigma(\omega_{j,m-2}) = s^{j,m-2}$ ,  $I_j^{\hat{s}}(\omega_{i,m-1}) = \{\omega_{i,m-1}, \omega_{j,m-2}\}$ , and  $p_{j,\omega_{i,m-1}}(\omega_{j,m-2}) = 1$ . Once again  $\omega_{i,m-1} \in W_j$ . Conduct this for each  $\omega_{i,m-1}$ , and let  $F^{m-2}$  be the set of states that are defined so far. Clearly, one can define such a sequence of sets  $F^m, F^{m-1}, \dots, F^0, F^{-1}$  following the above procedure. Set  $\Omega^{\hat{s}} = F^{-1}$ . For the states  $\omega_{i,-1} \in F^{-1} \setminus F^0$ , for which  $\sigma(\omega_{i,-1}) = s^{i,-1} \in X^{-1} = S$ ,  $I_j^{\hat{s}}$  remains to be defined for  $j \neq i$ ; set  $I_j^{\hat{s}}(\omega_{i,-1}) = \{\omega_{i,-1}\}$ . Clearly, such  $j$  need not be a wishful thinker or rational at  $\omega_{i,-1}$ . But by construction each player is a wishful thinker at each state in  $F^0$ . Therefore,  $W \supseteq F^0$ . Hence, for each  $i$ ,  $K_i(W) \supseteq K_i(F^0) \supseteq F^1$ , so that  $K^1(W) \supseteq F^1$ . Similarly,  $K^k(W) \supseteq F^k$  for each  $k \leq m$ . In particular,  $\omega_0 \in F^m \subseteq K^m(W)$ , showing that  $\hat{s} = \sigma(\omega_0) \in \sigma(K^m(W))$ . Finally, for the case of  $X^\infty$ , such a sequence of increasing sets could be defined indefinitely without ever going out of  $X^\infty$ , and each player will be a wishful thinker at each state, so that wishful thinking is common knowledge, and at the initial state the fixed profile  $\hat{s} \in X^\infty$  is played.  $\square$

**A.2. Wishful Thinking and Nash Equilibrium.** Here, I will explore the relationship between Nash equilibrium and common knowledge of wishful thinking

and prove Propositions 2 and 3. The next lemma states some straightforward but very useful facts about the elimination process and its relationship to equilibrium.

**Lemma 4.** *The following are true.*

- (1)  $\phi$  is monotonic (i.e.,  $X \subseteq Y \Rightarrow \phi(X) \subseteq \phi(Y)$ );
- (2)  $NE$  is a fixed point of  $\phi$  (i.e.,  $\phi(NE) = NE$ );
- (3)  $X^\infty$  is a fixed point of  $\phi$  (i.e.,  $\phi(X^\infty) = X^\infty$ ).

Lemma 4 immediately implies Proposition 2.

*Proof of Proposition 2.* By Lemmas 4.2, 4.1, and the definition of  $X^\infty$ ,

$$NE = \phi^\infty(NE) \subseteq \phi^\infty(S) = X^\infty.$$

□

Recall that

$$X_i^\infty = \{s_i \mid \exists s_{-i} : (s_i, s_{-i}) \in X^\infty\}$$

is the set of strategies for player  $i$  that are consistent with common knowledge of wishful thinking. Since  $\phi(X^\infty) = X^\infty$ , each such strategy must be a best reply to a surviving strategy:

**Lemma 5.** *For each  $i$  and  $s_i \in X_i^\infty$ , there exists  $s_{-i}$  such that  $(s_i, s_{-i}) \in B_i \cap X^\infty$ .*

The next lemma establishes some useful facts for two-player games with unique best replies. It states that  $X^\infty$  is closed under best reply and that the restriction of the best-response function to  $X_i^\infty$  is a bijection. Most notably, part 3 states that, when applied to  $NE_1 \times NE_2$ , the elimination process stops at the first iteration.

**Lemma 6.** *For any two-player game assume that, for each  $i$ ,  $BR_i$  is singleton-valued. Then, the following are true.*

- (1)  $|X_1^\infty| = |X_2^\infty|$ ;
- (2) For each  $i$ , there exists a one-to-one and onto mapping  $\rho_i : X_j^\infty \rightarrow X_i^\infty$  such that  $BR_i(s_j) = \{\rho_i(s_j)\}$  for each  $s_j$ ;
- (3)  $\phi^\infty(NE_1 \times NE_2) = \phi(NE_1 \times NE_2)$ .



*Proof.* By Lemma 5, for each  $s_i \in X_i^\infty$ , there exists a  $\rho_i^{-1}(s_i) \in X_j^\infty$  such that  $s_i \in BR_i(\rho_i^{-1}(s_i))$ . Since  $BR_i$  is singleton-valued,  $\rho_i^{-1}$  is one-to-one. Hence,  $|X_i^\infty| \leq |X_j^\infty|$ . Since  $i$  is arbitrary, this yields (1). But (1) in turn implies that the one-to-one function  $\rho_i^{-1}$  is also onto. Thus,  $\rho_i : X_j^\infty \rightarrow X_i$  is well-defined. Since  $\rho_i^{-1}$  is a bijection, so is  $\rho_i$ , yielding (2). To show (3), check that, when  $BR_i$  is singleton-valued, for each  $\hat{s}_i \in NE_i$ , there exists a unique  $(\hat{s}_i, s_{-i}) \in B_i \cap (NE_1 \times NE_2)$ . In that case,  $\phi(\phi(NE_1 \times NE_2)) = \phi(NE_1 \times NE_2)$ , and hence  $\phi^k(NE_1 \times NE_2) = \phi(NE_1 \times NE_2)$  for each  $k$ .  $\square$

*Proof of Proposition 3.* To prove (4.1), first note that if  $X^\infty = \emptyset$ , then by Proposition 2,  $NE = \emptyset$ , and thus  $X^\infty = NE$ . Assume that  $X^\infty \neq \emptyset$ . Using Lemma 6 and assumption (ii), one can rename the strategies as  $X_1^\infty = \{s_1^1, \dots, s_1^k\}$  and  $X_2^\infty = \{s_2^1, \dots, s_2^k\}$  for some  $k \geq 1$  so that  $B_1 \cap X^\infty$  is the diagonal of  $X_1^\infty \times X_2^\infty$ , i.e.,

$$(A.2) \quad B_1 \cap X^\infty = \{(s_1^l, s_2^l) \mid 1 \leq l \leq k\} \subseteq X_1^\infty \times X_2^\infty,$$

and

$$(A.3) \quad u_1(s_1^l, s_2^l) \text{ is strictly decreasing in } l.$$

Now, for any  $l > m$ , (A.2) and (A.3) imply that

$$\max_{s_1} u_1(s_1, s_2^m) = u_1(s_1^m, s_2^m) > u_1(s_1^l, s_2^l) = \max_{(s_1^l, s_2) \in X^\infty \cap B_1} u_1(s_1^l, s_2),$$

which in turn implies that  $(s_1^l, s_2^m) \notin \phi(X^\infty) = X^\infty$ . Therefore,

$$(A.4) \quad X^\infty \subseteq \{(s_1^l, s_2^m) \mid 1 \leq l \leq m \leq k\}.$$

Now, I will use mathematical induction (on  $l$ ) to show that  $NE = B_1 \cap X^\infty$ . Together with (A.2), this implies (4.1). For  $l = 1$ , by Lemma 5, there exists  $s_1$  such that  $(s_1, s_2^1) \in B_2 \cap X^\infty$ . But (A.4) states that  $s_1^1$  is the only strategy that can satisfy this. Hence,  $(s_1^1, s_2^1) \in B_2$ . Together with (A.2), this shows that  $(s_1^1, s_2^1) \in NE$ . Assume that  $(s_1^1, s_2^1), \dots, (s_1^{l-1}, s_2^{l-1}) \in NE$  for some  $l > 1$ . Since  $BR_2$  is singleton-valued (by assumption (i)),  $s_2^l \notin BR_2(s_1)$  for any  $s_1 \in \{s_1^1, \dots, s_1^{l-1}\}$ . Hence,  $s_2^l \in BR_2(s_1)$  for some  $s_1 \in \{s_1^l, \dots, s_1^k\}$ ; recall from Lemma 6.2 that  $\rho_2$  is onto. But since  $\{(s_1^{l+1}, s_2^l), \dots, (s_1^k, s_2^l)\} \cap X^\infty = \emptyset$  (by (A.4)), it must be that

$s_2^l \in BR_2(s_1^l)$ , showing that  $(s_1^l, s_2^l) \in B_1 \cap B_2 = NE$ . Therefore,  $B_1 \cap X^\infty \subseteq NE$ . (And  $NE \subseteq B_1 \cap X^\infty$  by Proposition 2.)

To prove (4.2), write

$$X^\infty = \phi(X^\infty) \subseteq \phi(NE_1 \times NE_2) = \phi^\infty(NE_1 \times NE_2) \subseteq \phi^\infty(S) = X^\infty,$$

where the first equality is by Lemma 4.3, the next inclusion is by Lemma 4.1 and the first part of the proposition (i.e.,  $X^\infty \subseteq NE_1 \times NE_2$ ), the next equality is by Lemma 6.3, the next inclusion is again by Lemma 4.1, and the last equality is by definition.  $\square$

*Proof of Corollary 2.* Let  $\Omega = X^\infty$ , and let  $\sigma$  be the identity mapping. Set  $I_i(\omega) = \{s \in X^\infty \mid s_i = \sigma_i(\omega)\}$  for each  $i$  and  $\omega$ . By (4.1), for each  $i$  and  $\omega$ , there exists a (unique) Nash equilibrium  $\hat{s}$  with  $\hat{s} \in I_i(\omega)$ . Set  $p_{i,\omega}(\hat{s}) = 1$ . Now, by (4.2), for any  $s \in I_i(\omega)$ , there exists  $(\tilde{s}_i, s_{-i}) \in NE$  such that  $u_i(\hat{s}) \geq u_i(\tilde{s}_i, s_{-i})$ . Since  $u_i(\tilde{s}_i, s_{-i}) = \max_{s'_i} u_i(s'_i, s_{-i})$  (by definition of  $NE$ ) for each  $s_{-i}$ , this shows that  $u_i(\hat{s}) = \max_{s_i, s_{-i} \in \sigma_{-i}(I_i(\omega))} u_i(s)$ , showing that  $\omega \in W_i$ .  $\square$

*Proof of Proposition 4.* Each strategy profile  $(s_1, s_2)$  with  $BR_i(s_j) > s_i$  is eliminated in the first round for player  $i$  because in that case, by monotonicity,  $u_i(BR_i(s_j), s_j) > u_i(s_i, s'_j)$  for every  $(s_i, s'_j) \in B_i$ . Hence,

$$(A.5) \quad s_i \geq BR_i(s_j) \quad (\forall (s_1, s_2) \in X^0, \forall i, j \in N).$$

I will show that  $X^2 \subseteq NE_1 \times NE_2$ ; i.e., for each  $(\hat{s}_1, \hat{s}_2) \in X^1$ , there exist  $(\hat{s}_1, s_2), (s_1, \hat{s}_2) \in NE$ . Since  $(\hat{s}_1, \hat{s}_2) \in X^0$ , by (A.5),  $\hat{s}_1 \geq BR_1(\hat{s}_2) = s_1$  and  $\hat{s}_2 \geq BR_2(\hat{s}_1) = s_2$ , yielding the expression for  $X^2$  in (4.3). Then, by Proposition 3,  $\phi(X^2) = X^2$ , showing that  $X^\infty = X^2$ .

Take any  $s_1 \in S_1 \setminus NE_1$ . I will show that  $s_1$  is eliminated in the first three rounds, so that  $s_1 \notin X_1^2$ . Since this will also be true for non-equilibrium strategies of Player 2, this will show that  $X^2 \subseteq NE_1 \times NE_2$ . Now, take any  $(s_1, s_2) \in B_1 \cap X^0$ . By (A.5),  $s_2 \geq BR_2(s_1)$ . Since  $s_1 = BR_1(s_2)$  and  $(s_1, s_2)$  is not a Nash equilibrium,  $s_2 \neq BR_2(s_1)$ , so that  $s_2 > BR_2(s_1)$ . Then, for each  $(s'_1, s_2) \in B_2$ , we have  $s'_1 < s_1 = BR_1(s_2)$ , as  $B_2$  is non-increasing (by strategic substitutes) and  $s_2 = BR_2(s'_1)$ . By (A.5), such  $(s'_1, s_2)$  is not in  $X^0$ , showing that there is no  $s'_1$

with  $(s'_1, s_2) \in B_2 \cap X^0$ . Therefore, strategy  $s_2$  must be eliminated in the second round (for  $m = 1$ ). In particular,  $(s_1, s_2) \notin X^1$ . Here,  $(s_1, s_2) \in B_1 \cap X^0$  was arbitrary. Thus, there is no  $s_2$  with  $(s_1, s_2) \in B_1 \cap X^1$ . Therefore,  $s_1$  must be eliminated in the third round (for  $m = 2$ ), or before.  $\square$

## REFERENCES

- [1] Ali, S. Nageeb M. (2003): "Multilateral Bargaining with Subjective Biases: Waiting To Settle," *Journal of Economic Theory*, forthcoming.
- [2] Aumann, Robert (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics* 1, 67-96.
- [3] Aumann, Robert (1976): "Agreeing to Disagree," *Annals of Statistics* 4, 1236-1239.
- [4] Aumann, Robert (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1-18.
- [5] Aumann, Robert (1995): "Interactive Epistemology," *Hebrew University Discussion Paper* 67.
- [6] Aumann, Robert (1999): "Interactive Epistemology I: Knowledge," *International Journal of Game Theory* 28, 263-300.
- [7] Aumann, R. and A. Brandenburger (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161-1180.
- [8] Babcock, Linda and George Loewenstein (1997): "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives* 11, 109-126.
- [9] Bar-Hillel, Maya and David Budescu (1995): "The Elusive Wishful Thinking Effect," *Thinking and Reasoning*, 1, 71-103.
- [10] Banerjee, Abhijit and Rohini Somanathan (2001): "A Simple Model of Voice," *Quarterly Journal of Economics*, 116, 189-227.
- [11] Bernheim, Douglas (1984): "Rationalizable Strategic Behavior," *Econometrica* 52, 1007-1028.
- [12] Brandenburger, Adam and Eddie Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391-1402.
- [13] Dekel, Eddie, Drew Fudenberg, and David Levine (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46, 282-303.
- [14] Dekel, Eddie and Faruk Gul (1997): "Rationality and Knowledge in Game Theory," in Kreps, D. and K. Wallis (eds): *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, UK.

- [15] Dekel, Eddie, Zvi Safra, and Uzi Segal (1991): "Existence and Dynamic Consistency of Nash Equilibrium with Non-expected Utility Preferences," *Journal of Economic Theory* 55, 229-246.
- [16] Dow, James and Sergio Werlang (1994): "Nash Equilibrium under Knightian Uncertainty: Braking Down Backward Induction," *Journal of Economic Theory* 64, 305-324.
- [17] Epstein, Larry (1997): "Preference, Rationalizability, and Equilibrium," *Journal of Economic Theory* 73, 1-29.
- [18] Eyster, Eric and Matthew Rabin (2005): "Cursed Equilibrium," *Econometrica* 73, 1623-1672.
- [19] Farber, Henry S. and Max H. Bazerman (1989): "Divergent expectations as a cause of disagreement in bargaining: Evidence from a comparison of arbitration schemes," *Quarterly Journal of Economics* 104, 99-120.
- [20] Feinberg, Yossi (2000): "Characterizing Common Priors in the Form of Posteriors," *Journal of Economic Theory* 91, 127-179.
- [21] Feinberg, Yossi (2004): "Subjective Reasoning—Games with Unawareness," mimeo.
- [22] Geanakoplos, John, David Pearce, and Ennio Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60-79.
- [23] Harrison, Michael and David Kreps (1978): "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations" *Quarterly Journal of Economics* 92, 323-36.
- [24] Jehiel, Philippe (1995): "Limited Horizon Forecast in Limited Alternate Games," *Journal of Economic Theory* 67, 497-519.
- [25] Jehiel, Philippe (2005): "Analogy-based Expectation Equilibrium," *Journal of Economic Theory* 123, 81-104.
- [26] Kaplan, Todd and Bradley Ruffle (2004): "The Self-serving Bias and Beliefs about Rationality," *Economic Inquiry* 42, 237-246.
- [27] Landes, William M. (1971): "An economic analysis of the courts," *Journal of Law and Economics* 14, 61-107.
- [28] Larwood, Laurie and William Whittaker (1977): "Managerial Myopia: Self-serving Biases in Organizational Planning," *Journal of Applied Psychology*, 194-198.
- [29] Manove, Michael and Jorge Padilla (1999): "Banking (Conservatively) with Optimists," *Rand Journal of Economics*, 30, 324-350.
- [30] Mc Kelvey, Richard and Thomas Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior* 10, 6-38.
- [31] Morris, Stephen (1996): "Speculative Investor Behavior and Learning" *Quarterly Journal of Economics* 111, 1111-33.
- [32] Osborne, Martin and Ariel Rubinstein (1998): "Games with Procedurally Rational Players," *American Economic Review* 88, 834-847.

- [33] Pearce, David (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52, 1029-1050.
- [34] Posner, Richard (1972): "The Behavior of Administrative Agencies," *Journal of Legal Studies* 1, 305.
- [35] Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 5, 1281-1302.
- [36] Squintani, Francesco (2006): "Mistaken self-perception and Equilibrium," *Economic Theory* 27, 615-641.
- [37] Rostek, Marzena (2004): " $p$ -dominant sets", mimeo.
- [38] Tan, T. and S. Werlang (1988): "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory* 45, 370-391.
- [39] Van den Steen, Eric (2005): "Organizational Beliefs and Managerial Vision," *Journal of Law, Economics, and Organization* 21, 256-283.
- [40] Watanabe, Yasutora (2004): "Learning and Bargaining in Dispute Resolution: Theory and Evidence from Medical Malpractice Litigation," mimeo.
- [41] Weinstein, Neil (1980): "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology*, 860-20.
- [42] Wilson, Robert (1968): "The Theory of Syndicates," *Econometrica*, 36, 119-132.
- [43] Yildiz, Muhamet (2003): "Bargaining without a Common Prior — An Agreement Theorem," *Econometrica* 71, 793-811.
- [44] Yildiz, Muhamet (2004): "Waiting to Persuade," *Quarterly Journal of Economics* 119, 223-248.

MIT

URL: <http://econ-www.mit.edu/faculty/myildiz/index.htm>