Strong forward induction*

Bingyong Zheng†

Abstract

Forward induction, as defined by Govindan and Wilson (2009), places a local dominance condition on off-equilibrium path beliefs restricting relevant strategy profiles for an equilibrium outcome to be infinitely more likely than profiles that include irrelevant strategies. Meanwhile, it places no global dominance restrictions and thus leaves open the possibility that a dominated strategy is deemed more likely than strategies dominating it. This paper defines *strong forward induction* that improves upon forward induction by placing a global dominance condition on the lexicographical beliefs in addition to the local dominance condition. In some games strong forward induction can eliminate implausible outcomes that forward induction and an iterative procedure based on it fail to eliminate.

Keyword: forward induction, lexicographical belief system, equilibrium refinement.

JEL Classification Number: C72.

1 Introduction

Forward induction reflects a pattern of reasoning that players believe, even when confronted with unexpected events, that their opponents acted optimally in the past, as long

*I am deeply grateful to Hari Govindan for discussions and suggestions that have played a key role in shaping my ideas on this topic. I also thank Quan Wen, Adam Wong and Wenzhang Zhang for comments. Of course, remaining errors are my own.

†School of Economics, Shanghai University of Finance and Economics. Email: bingyongzheng@gmail.com

as the deviation can be rationalized. While game theorists often agree about what forward induction entails for particular games, to capture the intuition in a generally acceptable definition has proved to be elusive until the work of Govindan and Wilson (2009) (hereafter referred to as GW). The current work builds upon GW but corrects a deficiency with their definition.

As an example of forward induction, consider the Outside Option Game in Figure 1. The set of equilibria in which player 1 plays *Out* and player 2 plays R with probability greater than 1/3 (yielding the outcome (2, 2)) satisfies backward induction, and every equilibrium in the set is a sequential equilibrium of the game.[1] However, suppose that player 2 explains having been reached in terms of 1 playing some best response. Because B is strictly dominated by *Out* it is never a best response for player 1. Consequently player 2's belief consistent with the conviction that 1 played a best response must be that player 1 played T. Given this belief player 2's optimal response would then be L. But believing all of this, player 1 would do best to play T. Thus forward induction eliminates (2, 2) and leaves (3, 1) as the unique outcome of the game.

GW's definition of forward induction captures the notion of forward induction with a restriction on players' off-equilibrium beliefs. Take an equilibrium outcome of the game. A strategy of player i is said to be "relevant" for the outcome if it is optimal against a sequence of ϵ-perfect equilibria whose limit leads to the outcome; otherwise, it is "irrelevant." If players all share the belief that they are in some equilibrium yielding the outcome, a player, when being reached, would believe his opponents played relevant strategies, instead of irrelevant strategies. The concept of normal-form forward induction captures the intuition with a *relevance condition* restricting relevant strategy profiles to be infinitely more likely than profiles including irrelevant strategies. A set of equilibria satisfies normal-form forward induction if it contains a perfect equilibrium that is a limit of a sequence of ϵ-perfect equilibria satisfying the relevance condition.[2]

[1]Sequential equilibrium is a generalization of backward induction in games of perfect information to imperfect information extensive-form games with perfect recall.

[2]Blume et al. (1991) have showed that a normal-form perfect equilibrium can be taken as a Nash equilibrium with a lexicographical belief system. Therefore, restrictions placed on the lexicographical

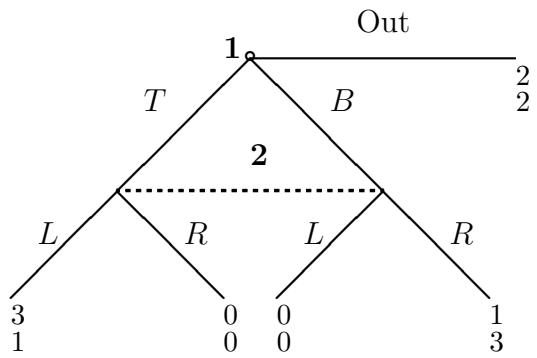


Figure 1: Outside Option Game

As an illustration, consider the Outside Option Game again. The set of normal-form perfect equilibria giving (2, 2) is $\{Out, (\alpha, 1-\alpha) | \alpha \in [0, 2/3]\}$. First, player 2's strategies L and R are equilibrium strategies and thus relevant. Player 1's strategies Out and T are also relevant, since both are optimal against the sequence $\{(1-\epsilon, 3\epsilon/4, \epsilon/4), (2/3, 1/3)\}$. But player 1's dominated strategy B is not a best response against any sequences and, thus, not relevant. When player 2 believes that player 1's relevant strategy T is infinitely more likely than the irrelevant strategy B, however, there exists no perfect equilibrium in which player 2 plays R and player 1 chooses Out, thereby ruling out (2, 2).

On the other hand, the restriction imposed by the concept of normal-form forward induction may be inadequate. In particular, in identifying the set of relevant strategies, it fails to put any restriction on players' lexicographical beliefs, and therefore leaves open the possibility that a player believes an opponent's dominated strategy is more likely than strategies dominating it.

Consider, for example, the Modified Outside Option Game in Figure 2. Compared to Figure 1, player 1 has an additional strategy Q that is strictly dominated by both T and B, and player 2 has an additional strategy M that becomes dominated when Q is deleted. Now, B is a relevant strategy for player 1 and thus (2, 2) becomes a forward induction outcome. To see this, note that B would be a best response for player 1 if he believed that

beliefs are equivalent to restrictions on the sequences of ϵ-perfect equilibria.

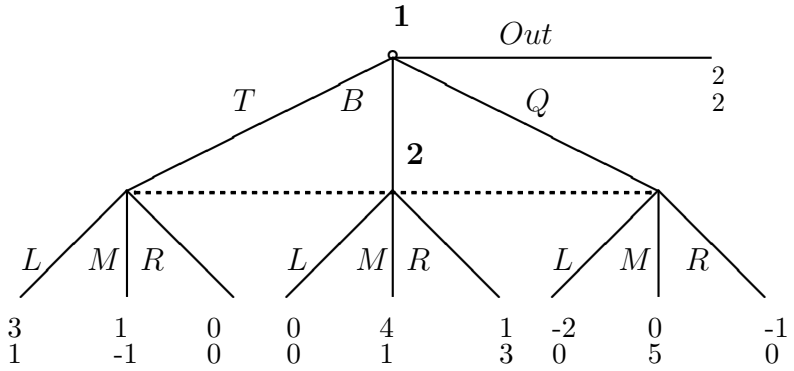


Figure 2: Modified outside option game

2 played M with probability greater than 1/3, which could be optimal for player 2 if he expected Q to be more likely than T in the event that 1 did not play Out.

Nevertheless, since Q is strictly dominated by T, if reached, player 2 would not play M unless he thought player 1 did not play optimally. Hence, B is relevant for player 1 only if his choice was allowed to be affected by 2's (wrong) belief about his own play. Indeed, neither B nor M would be relevant if we disallow the type of beliefs that deems player 1's dominated strategy Q to be more likely than the dominating strategy T. But beliefs of this sort would *go against the very essence of forward induction*, which, under the circumstances, would require player 2 maintain the assumption that player 1 played a best strategy among those consistent with the event that he is reached.

This example shows that the definition of normal-form forward induction lacks self-consistency in the following sense. On the one hand, player 2 is expected to update his belief so as to give infinitely more weight to player 1's relevant strategies than to his irrelevant strategy. And yet, in determining the set of relevant strategies, it allows players to hold almost any lexicographical beliefs.

Since the definition of forward induction suffers from this deficiency, the natural question that arises is: what are the conditions needed to solve the problem? The above example clearly shows that some restrictions on the lexicographical beliefs is also necessary in identifying the set of relevant strategies. At a minimum, we believe, the beliefs should consider a dominated strategy to be less likely than strategies dominating it, a condition

we shall call the *dominance condition*. An intuitive restriction like this will ensure the definition is conceptually consistent, and also enhance its refining power. For example, in Figure 2, insisting on player 2 holding fast to the conviction that player 1 plays optimally whenever possible would rule out (2, 2). Loosely speaking, this is how we plan to formulate the strengthened concept – *strong forward induction*.

The main contribution of this paper is to provide a definition of strong forward induction that incorporates both the relevance condition and dominance condition. To understand the connection between forward induction and strong forward induction, it helps to think the *relevance condition* as a "local dominance condition" and the *dominance condition* as a "global dominance condition."[3] By definition, for a given equilibrium outcome, an irrelevant strategy is *locally* dominated by and therefore a "worse" response than a relevant strategy. A dominated strategy, on the other hand, is a "worse" response than a dominating strategy for *any* equilibrium outcome. The local and global dominance conditions take into account the two dominating relationships respectively. Since it only imposes the relevance condition, forward induction satisfies the "local dominance condition," but may not satisfy the "global dominance condition." By contrast, strong forward induction includes both conditions and makes up the deficiency.

One may argue that lack of global dominance condition is not a defect of the definition of forward induction. The argument is that if players share common beliefs about the equilibrium outcome, then it is not clear why global dominance condition should matter. We do not agree. In some sense dismissing the global dominance condition is analogous to dismissing admissibility as a property of self-enforcing equilibria. After all, the concept of Nash equilibrium assumes that players share common beliefs about the equilibrium being played. Hence, if we were to reject the global dominance condition as unnecessary for the forward induction criterion, we should also accept equilibrium in (weakly) dominated strategies as self-enforcing equilibrium of the game.

The current work belongs to a large literature going back to Kohlberg and Mertens (1986), who show that backward induction solution concept like sequential equilibrium

[3] I am grateful to Hari Govindan for suggesting this point.

is flawed since unreasonable out-of-equilibrium beliefs might make implausible outcomes conform with backward induction. They use "forward induction" to denote a key feature of their stable set, but provide no formal definition to this term.[4] Other researches that discuss forward induction include van Damme (1989), Reny (1992) and Hillas and Kohlberg (2002), but none provides a formal definition to it. The "intuitive criterion" of Cho and Kreps (1987) (See also Banks and Sobel (1987) for a refinement based on it) captures the spirit of forward induction in signaling games. Battigalli and Siniscalchi (2002) derive the intuitive criterion from an epistemic model. Man (2012) develops forward induction equilibrium that is obtained from an iterative application of the relevance condition.

The rest of the paper is developed as follows. Section 2 introduces notations that will be used. In section 3, we define normal-form strong forward induction. Section 4 introduces a solution concept that incorporates forward induction and other aspects of rationality. Section 5 concludes. Appendix A includes the proof of the main result. Appendix B gives a definition of extensive-form strong forward induction equivalent to the one in the text.

2 Environment

As the formal description of an extensive-form game is fairly standard, we shall be very brief in introducing the extensive-form game and its corresponding normal-form. We consider only finite extensive-form games with perfect recall (Kuhn (1953)). Denote a finite game tree by Γ_E, which has a finite set of players $\mathbf{I} \equiv \{1, \ldots, N\}$. Let $\mathbf{A}$ and $\mathbf{X}$ be the finite set of actions and finite set of nodes respectively. Let $\mathbf{A}_i$ and $\mathbf{X}_i$ be the set of actions and set of decision nodes for player i. The set of terminal nodes is represented by $\mathbf{Z}$. An *outcome* of the game is a distribution over the terminal nodes of the game tree.

An information set h_i of player i is a partition of the player's set of nodes $\mathbf{X}_i$ that assigns each decision node x to the set of information set $\mathbf{H}_i$. We let $\mathbf{A}_i(h_i)$ denote the set of actions available at information set h_i. A pure strategy s_i of player i in Γ_E is a mapping

[4] According to Kohlberg and Mertens (1986, Prop.6), stable equilibria satisfy forward induction if a stable set contains a stable set of any game obtained by deletion of a strategy that is not optimal against any equilibrium in the set.

that prescribes an action at each of the player's information set, $s_i : \mathbf{H}_i \to \mathbf{A}_i$ provided $s_i(h_i) \in \mathbf{A}_i(h_i)$ for all h_i. A mixed strategy σ_i of player i is a probability distribution over the set of pure strategies. We use $\mathbf{S}_i$ ($\Delta(\mathbf{S}_i)$) to denote the set of pure strategies (mixed strategies) of player i.

We say a pure strategy $s_i \in \mathbf{S}_i$ is redundant if there exists a mixed strategy $\sigma_i \in \Delta(\mathbf{S}_i)$ such that for each strategy combination σ_{-i} of player i's opponents, σ_i and s_i lead to the same distribution over the terminal nodes. The *reduced normal-form* of the extensive-form game is the strategic form $G = [I, \{\tilde{\mathbf{S}}_i\}, \{u_i(\cdot)\}]$ in which for each $i \in \mathbf{I}$, set $\tilde{\mathbf{S}}_i$ contains no redundant strategies. We say that two games are equivalent if their reduced strategic forms are the same, except for relabeling of strategies. Denote $\tilde{s}_i$ as a pure strategy and $\tilde{\sigma}_i \in \Delta(\tilde{\mathbf{S}}_i)$ as a mixed strategy of player i in the reduced normal-form game G.

For the reduced normal-form game G, for all $i \in \mathbf{I}$, let $\tilde{\Sigma}_i^0$ be the set of probability distribution that gives positive weight to all elements in $\tilde{\mathbf{S}}_i$,

$$\tilde{\Sigma}_i^0 = \left\{ \tilde{\sigma}_i \middle| \sum_{\tilde{s}_i \in \tilde{\mathbf{S}}_i} \tilde{\sigma}_i(\tilde{s}_i) = 1, \tilde{\sigma}_i(\tilde{s}_i) > 0 \quad \forall \tilde{s}_i \in \tilde{\mathbf{S}}_i \right\}. \tag{1}$$

The set of totally mixed strategy profiles is represented by $\tilde{\Sigma}^0 \equiv \prod_{i \in I} \tilde{\Sigma}_i^0$.

3 Normal-form strong forward induction

While GW define two versions of forward induction, one in the context of games in extensive-form, another in the context of games in normal-form, we base our formulation mainly on reduced normal-form games, but provide an extensive version in the appendix.[5] The definition of normal-form forward induction of GW takes normal-form perfect equilibrium as a starting point and attempts to eliminate implausible outcomes by putting reasonable restrictions on out-of-equilibrium beliefs.[6] As Blume et al. (1991) show, a normal-form

[5]Normal-form forward induction is also stronger than extensive-form forward induction. Extensive-form forward induction fails admissibility as well as invariance. Nevertheless the two concepts are generically equivalent for two-player extensive-form game with perfect recall.

[6]According to the definition of Myerson (1978), an ϵ-perfect equilibrium is a totally mixed strategy profile $\tilde{\sigma} \in \tilde{\Sigma}^0$ such that for all i and for all $\tilde{s}_i' \in \tilde{S}_i$, $u_i(\tilde{s}_i, \tilde{\sigma}) > u_i(\tilde{s}_i', \tilde{\sigma})$ implies $\tilde{\sigma}_i(\tilde{s}_i') \leq \epsilon$. A *perfect*

perfect equilibrium is equivalent to a Nash equilibrium with lexicographical probability system satisfying the common prior and strong independence assumption. Restrictions placed on the sequence of ϵ-perfect equilibria are therefore equivalent to restrictions on players' lexicographical belief system of a perfect equilibrium.

3.1 Review of normal-form forward induction

In GW normal-form forward induction is taken as a property of a set of equilibria. Unfortunately, they do not explain how a set of equilibria is determined. Below we provide a brief explanation. Kohlberg and Mertens (1986, Prop. 1) show that any game has finitely many connected components of Nash equilibria. In principle (probably true of most interesting games) one can take any single connected component as a set of equilibria. However, this approach would run into difficulty in games of which different connected components induce the same outcome, for example, Figure 11 of Kohlberg and Mertens (1986).[7] Another issue is that some component may include Nash equilibria that put positive probability on (weakly) dominated strategies and, thus, fails to satisfy admissibility.

Hence, we let a set of equilibria $\mathbf{E}$ include all normal-form perfect equilibria in a closed connected component of Nash equilibria, except when more than one component induce the same outcome, in which case, $\mathbf{E}$ includes perfect equilibria in all components giving that outcome. With the set of equilibria $\mathbf{E}$ thus defined, a set $\mathbf{E}$ induces a single outcome (distribution over terminal nodes of the game tree) in generic games.[8] For non-generic games, however, a set of equilibria may induce a continuum of equilibrium outcomes.

One key assumption of GW is that players know the equilibrium outcome, but may be unsure of the equilibrium leading to the outcome. Thus, if reached unexpectedly, a player would expect his opponents to play optimally against some beliefs about what the others are playing, i.e., play relevant strategies for the set of equilibria being considered.

equilibrium is the limit of a sequence of ϵ-perfect equilibria as $\epsilon \to 0$.

[7] Kohlberg and Mertens attribute the example to Faruk Gul. In that game, two components of equilibria give the same outcome (2, 0, 0).

[8] By "generic game," we mean a game with the property that its game tree possesses finitely many equilibrium outcomes.

By definition, player i's pure strategy $\tilde{s}_i$ is *relevant* for a set of equilibria $\mathbf{E}$ if it is optimal against a sequence of ϵ-perfect equilibria whose limit $\tilde{\sigma}^*$ is contained in $\mathbf{E}$.

Let $\mathbf{R}_i(\mathbf{E})$ be i's set of relevant strategies, and $\mathbf{R}(\mathbf{E})$ be the set of relevant strategy profiles, i.e., $\mathbf{R}(\mathbf{E}) \equiv \prod_i \mathbf{R}_i(\mathbf{E})$. We say $\tilde{s}$ is a relevant strategy profile if $\tilde{s} \in \mathbf{R}(\mathbf{E})$ and a strategy profile $\tilde{s}'$ includes irrelevant strategies if $\tilde{s}' \in \tilde{\mathbf{S}}/\mathbf{R}(\mathbf{E})$.

DEFINITION 3.1. (GW Definition B.1) A set of perfect equilibria $\mathbf{E}$ satisfies *normal-form forward induction* if there is $\tilde{\sigma}^* \in \mathbf{E}$ whose lexicographical representation has all profiles of relevant strategies occurring before all profiles that include irrelevant strategies.

Differently put, $\mathbf{E}$ satisfies normal-form forward induction if it contains a perfect equilibrium $\tilde{\sigma}^*$ that is a limit of a sequence $\{\tilde{\sigma}^\epsilon\}$ such that, if $\tilde{s} \in \mathbf{R}(\mathbf{E})$ and $\tilde{s}' \in \tilde{\mathbf{S}}/\mathbf{R}(\mathbf{E})$,

$$\lim_{\epsilon \to 0} \frac{\tilde{\sigma}^\epsilon(\tilde{s}')}{\tilde{\sigma}^\epsilon(\tilde{s})} = 0. \tag{2}$$

Before proceeding, it is worthwhile to point out that the restriction on profile of strategies is stronger than one requiring every player's relevant strategies be infinitely more likely than irrelevant strategies.

To see this, note that restriction (2) on profile of strategies implies restriction on each player's relevant/irrelevant strategies: for all i, if $\tilde{s}_i \in \mathbf{R}_i(\mathbf{E})$ and $\tilde{s}'_i \in \tilde{\mathbf{S}}_i/\mathbf{R}_i(\mathbf{E})$,

$$\lim_{\epsilon \to 0} \frac{\tilde{\sigma}_i^\epsilon(\tilde{s}'_i)}{\tilde{\sigma}_i^\epsilon(\tilde{s}_i)} = 0. \tag{3}$$

Recall that for any mixed strategy profile $\tilde{\sigma}(\cdot)$ and any pure strategy profile $\tilde{s} = (\tilde{s}_1, \ldots, \tilde{s}_N)$, $\tilde{\sigma}(\tilde{s}) \equiv \prod_i^N \tilde{\sigma}_i(\tilde{s}_i)$. Thus, for any i and for some $\tilde{s}_{-i} \in \mathbf{R}_{-i}(\mathbf{E})$, $\tilde{s}_i \in \mathbf{R}_i(\mathbf{E})$ and $\tilde{s}'_i \notin \mathbf{R}_i(\mathbf{E})$ implies

$$\frac{\tilde{\sigma}_i^\epsilon(\tilde{s}'_i)}{\tilde{\sigma}_i^\epsilon(\tilde{s}_i)} = \frac{\tilde{\sigma}_i^\epsilon(\tilde{s}'_i)\prod_{j \neq i} \tilde{\sigma}_j^\epsilon(\tilde{s}_j)}{\tilde{\sigma}_i^\epsilon(\tilde{s}_i)\prod_{j \neq i} \tilde{\sigma}_j^\epsilon(\tilde{s}_j)} = \frac{\tilde{\sigma}^\epsilon(\tilde{s}'_i, \tilde{s}_{-i})}{\tilde{\sigma}^\epsilon(\tilde{s}_i, \tilde{s}_{-i})} \to 0.$$

On the other hand, the reverse is not true; condition (3) does not lead to condition (2).

As discussed before, placing the *relevance condition* eliminates the set of equilibria giving the outcome (2, 2) in the Outside Option Game in Figure 1. While the relevance condition is good enough for games like the Outside Option Game, however, it is inadequate to games

where the relevance of a player's strategy depends crucially on the assessment of the relative likelihood of an opponent's strategies, for example, Figure 2. In this game Out, T and L, R are all relevant strategies for the set of equilibria giving outcome (2, 2). Also, B and M are relevant as both are optimal against the sequence $\{(1-\epsilon-\epsilon^2, 2\epsilon^2, 5\epsilon/7-\epsilon^2, 2\epsilon/7), (\epsilon, (2-4\epsilon)/3, (1+\epsilon)/3)\}$. Hence there exist perfect equilibria satisfying the relevance condition and inducing (2, 2).[9]

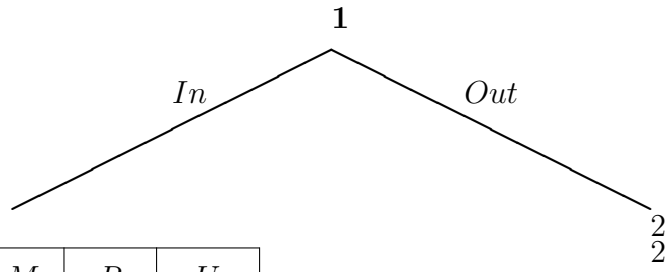
But to make B and M best responses and thus relevant, players need to share the belief that assigns more weight to a strictly dominated strategy than the dominating strategy. For example, to make M a best response, upon being reached unexpectedly, player 2 needs to hold the belief that player 1 is playing a strictly dominated strategy Q more likely than the dominating strategy T. But this is not what player 2 would really believe, if called upon to move and still thinking that player 1 played optimally. Holding the belief that Q is more likely than T can only result from the conviction that player 1 is not a rational player under the circumstances, which would go against the very essence of forward induction. In addition, player 1 would need to share this belief to make his choice B a relevant strategy. In this sense player 1 was allowed to be affected by the unreasonable belief about his own play.

Thus the concept of forward induction is not internally consistent. On one hand, it imposes restrictions on off-equilibrium path beliefs with the relevance condition, requiring strategies that are locally optimal (given the equilibrium outcome) is infinitely more likely than strategies that are not locally optimal. On the other hand, in identifying locally optimal (relevant) strategies, it places no restriction on the lexicographical beliefs and allows for beliefs that puts a dominated strategy as more likely than dominating strategies.

3.2 The definition of strong forward induction

Looking at the above example we see that without any global dominance restriction unreasonable belief can make a bad strategy relevant. Hence the definition of forward induction still leaves leeway for unreasonable beliefs. But what is the right restriction to

[9]For example, one such sequence of ϵ-perfect equilibria is $\{(1-\epsilon-\epsilon^2, 3\epsilon/4, \epsilon/4, \epsilon^2), (2/3-\epsilon, 1/3, \epsilon)\}$.



	L	M	R	U
T	3, 3	0, 0	0, 1	-1, -1
B	3, 0	0, 3	-1, 1	0, -1

Figure 3:

impose?

For games like Figure 2, it is clear how such a restriction can be applied. Since the problem is that a dominated strategy is deemed more likely than a dominating strategy, requiring the dominating strategy be more likely than the dominated strategy solves the problem. But in most games a strategy can also be dominated by mixed strategies. Thus the question becomes, whether similar condition can be applied when the dominating strategy is a mixed strategy, and what would be the right condition then.

At first, one might think an equivalent restriction would be asking every pure strategy in the support of the dominating mixed strategy be more likely than the dominated strategy. The problem, however, is that a dominating mixed strategy may have in its support a pure strategy that is also dominated. As an illustration, consider Figure 3. Note that player 2's mixed strategy $\tilde{\sigma}_2 = (0.5, 0.4, 0, 0.1)$, whose support includes U that is dominated by R, strictly dominates R. So we can not simply require every pure strategy in the support of a dominating mixed strategy be more likely than the dominated strategy.

On the other hand, since only L and M are necessary for any dominating strategy that dominates R whereas U is not, one might therefore conjecture that perhaps a less restrictive condition - - requiring strategies that are necessary for a dominating mixed strategy be more likely than the dominated strategy - - would correct the deficiency. It does not, as Figure 4 demonstrates. In this game there exist mixed strategies $\tilde{\sigma}_2$ with $supp(\tilde{\sigma}_2) = \{L, R\}$ that dominate M. Clearly both L and R are necessary for the dominating mixed strategies.

	L	M	R	U
A	4, 3	-1, 1	3, 1	1, 2
B	2, 5	-2, -1	-5, -2	0, -3
C	0, 0	5, 1	1, 2	-5, 5

Figure 4:

Meanwhile, there alo exist mixed strategies $\tilde{\sigma}_2'$ with $supp(\tilde{\sigma}_2') = \{M, U\}$ that dominate R. Thus, even if we were to ask only pure strategies necessary for a dominating mixed strategy be more likely than the dominated strategy, such a restriction would still lead to contradictory requirements on players' beliefs.

The two examples clearly show that not every dominating mixed strategy should be treated equally. To have a dominance condition that is free of contradictions we first construct $supp^m(\tilde{\sigma}_i, \tilde{s}_i')$ as a generalization of dominating strategy when mixed strategies are involved.

DEFINITION 3.2. For any strategy $\tilde{\sigma}_i$ dominating $\tilde{s}_i'$, a pure strategy $\tilde{s}_i \in supp^m(\tilde{\sigma}_i, \tilde{s}_i')$ if $\tilde{s}_i \in supp(\tilde{\sigma}_i)$ and one of the following is true:

(a) $\tilde{\sigma}_i$ puts probability 1 to $\tilde{s}_i$, i.e., $\tilde{\sigma}_i = \tilde{s}_i$; or

(b) there exists no $\tilde{\sigma}_i'$ such that $u_i(\tilde{\sigma}_i', \tilde{\sigma}_{-i}) > u_i(\tilde{\sigma}_i, \tilde{\sigma}_{-i})$ for all $\tilde{\sigma}_{-i} \in \tilde{\Sigma}_{-i}^0$.

The global dominance condition then requires every $\tilde{s}_i \in supp^m(\tilde{\sigma}_i, \tilde{s}_i')$ be infinitely more likely than the dominated strategy $\tilde{s}_i'$. Hence, when a pure strategy dominates another strategy, the condition simply ask the dominating strategy be infinitely more likely than the dominated strategy. When it comes to mixed strategies, however, the dominance condition only requires every pure strategy in the support of dominating strategies that are not dominated be more likely than the dominated strategy. As a result in some games there will be mixed strategies $\tilde{\sigma}_i$ that dominate $\tilde{s}_i'$ and $supp^m(\tilde{\sigma}_i, \tilde{s}_i') = \emptyset$.

To incorporate this new restriction we replace relevant strategy used in the definition of forward induction with *truly relevant strategy.*

DEFINITION 3.3. Given a set of perfect equilibria $\mathbf{E}$, player i's pure strategy $\tilde{s}_i$ is *truly relevant* if $\tilde{s}_i$ is optimal against a sequence of ϵ-perfect equilibria $\{\tilde{\sigma}^\epsilon\}$ that converges to

$\tilde{\sigma}^* \in \mathbf{E}$ and satisfies the condition that, for all i, if $u_i(\tilde{\sigma}_i, \tilde{\sigma}_{-i}) > u_i(\tilde{s}'_i, \tilde{\sigma}_{-i})$ for all $\tilde{\sigma}_{-i} \in \tilde{\Sigma}^0_{-i}$,

$$\lim_{\epsilon \to 0} \frac{\tilde{\sigma}^\epsilon_i(\tilde{s}'_i)}{\tilde{\sigma}^\epsilon_i(\tilde{s}_i)} = 0 \qquad \forall \quad \tilde{s}_i \in supp^m(\tilde{\sigma}_i, \tilde{s}_i). \tag{4}$$

To see its implementation, consider Figure 2. In this game, the set of perfect equilibria giving (2, 2) is $\{Out, (\alpha, 1-\alpha, 0)|3\alpha \leq 2\} \cup \{Out, (0, \gamma, 1-\gamma)|3\gamma \geq 2\} \cup \{Out, (0.5, 0, 0.5)\}$. Since player 1's strategy Q is strictly dominated by B and T, the dominance condition requires that $\lim_{\epsilon \to 0} \tilde{\sigma}^\epsilon_1(Q)/\tilde{\sigma}^\epsilon_1(B) = 0$ and $\lim_{\epsilon \to 0} \tilde{\sigma}^\epsilon_1(Q)/\tilde{\sigma}^\epsilon_1(T) = 0$. Thus imposing the dominance condition reduces the set of *truly relevant* strategies to $\{Out, T\}$ for player 1 and to $\{L, M\}$ for player 2.

Let $\overline{\mathbf{R}}^1_i(\mathbf{E})$ be i's set of truly relevant strategies, and $\overline{\mathbf{R}}^1(\mathbf{E})$ be the set of relevant strategy profiles, i.e., $\overline{\mathbf{R}}^1(\mathbf{E}) \equiv \prod_i \overline{\mathbf{R}}^1_i(\mathbf{E})$. We define the concept of strong forward induction as follows:

DEFINITION 3.4. A set of perfect equilibria $\mathbf{E}$ satisfies *strong forward induction* if there is $\tilde{\sigma}^* \in \mathbf{E}$ that is a limit of a sequence of ϵ-perfect equilibria $\{\tilde{\sigma}^\epsilon\}$ satisfying the *truly relevance condition*: for all ϵ, if $\tilde{s} \in \overline{\mathbf{R}}^1(\mathbf{E})$, $\tilde{s}' \in \tilde{\mathbf{S}}/\overline{\mathbf{R}}^1(\mathbf{E})$, then

$$\lim_{\epsilon \to 0} \frac{\tilde{\sigma}^\epsilon(\tilde{s}')}{\tilde{\sigma}^\epsilon(\tilde{s})} = 0. \tag{5}$$

In practice, to determine whether $\mathbf{E}$ satisfies strong forward induction, one can take a two step approach. The first step is to identify all normal-form perfect equilibria $\tilde{\sigma} \in \mathbf{E}$ satisfying the dominance condition. In case $\mathbf{E}$ contains no such perfect equilibria, it fails the dominance condition and, thus, does not satisfy strong forward induction. If $\mathbf{E}$ contains a perfect equilibrium satisfying the dominance condition, one then proceeds to the next step to identify the set of truly relevant strategies and apply the relevance condition.

Forward induction imposes a "local dominance condition" – the relevance condition – on players' lexicographical beliefs, but lacks any global dominance restriction, like the dominance condition we have here. The flaw not only makes forward induction conceptually inconsistent, but also diminishes its refining power, rendering it toothless in games like Figure 2 and Figure 3. By contrast, strong forward induction combines both the relevance condition and dominance condition together and makes up the deficiency.

At this point, we hope readers are convinced that imposing the extra dominance condition is not only appropriate but also necessary. After all, the same logic in imposing the relevance condition would recommend imposing the dominance condition as well. If one accepts the idea that a strategy that is locally dominated at the equilibrium is infinitely less likely than one that is not locally dominated, then similar consideration would apply to the relative probabilities of a dominated strategy and strategies dominating it. Since a dominated strategy is, by definition, a "worse" response than a dominating strategy, no reasonable assessment should put a dominated strategy before a dominating strategy. Based on this observation, we believe that any reasonable lexicographical beliefs should satisfy the dominance condition, and if a perfect equilibrium has lexicographical representation violating this condition, there is enough ground to reject it as a reasonable prediction of the game.

While a simple and intuitive restriction, the dominance condition nevertheless has a strong refining power. There is now only one equilibrium satisfying strong forward induction in Figure 2, (T, L). Note that the iterative procedure of Man (2012) can also eliminate the set of equilibria giving (2, 2). According to Man (2012), GW's normal form forward induction lacks self-consistency in the following sense. On the one hand, the players are expected to update their beliefs so as to give infinitely more weight to opponent's strategies that are optimal against some perfect equilibrium with the same outcome as the equilibrium outcome than to other strategies. However, after taking this into account the players ought to apply the same principle to the set of equilibria so derived. To correct the deficiency, she provides a refined definition based on an iterative technique. On the other hand, since Man's iterative procedure takes GW's definition of normal-form forward induction as a starting point, it suffers from the same deficiency as we pointed out.

For example, in the two-player game in Figure 3. Here the set of relevant strategies for the set of equilibria inducing (2, 2) is $\{Out, T, B, L, M\}$. To see this, note that Out, T, B L, M are all optimal against the sequence $\left\{\left(1-\epsilon, \frac{\epsilon}{2}, \frac{\epsilon}{2}\right), \left(\frac{2+\epsilon}{3}, \frac{1-7\epsilon}{3}, \epsilon, \epsilon\right)\right\}$. Thus the set of perfect equilibria inducing (2, 2) satisfies normal-form forward induction. Further, applying the iterative procedure of Man (2012) to the set of relevant strategies fails to prune any

strategies, thereby making every equilibrium in this set a forward induction equilibrium by her definition.

But since B is not a best response to any beliefs that puts R as more likely than U, imposing the dominance condition will eliminate B as a truly relevant strategy. When B is no longer relevant, however, there exists no perfect equilibrium inducing (2, 2). Therefore, strong forward induction alone eliminates the implausible outcome (2, 2) that the iterative procedure based on forward induction fails to rule out in this game.

4 Strong forward induction equilibrium

The definition of strong forward induction captures well the intuitive notion of what forward induction entails for the lexicographical beliefs. If one is unwilling or unable to place further restrictions upon the lexicographical beliefs, then outcomes resulted from equilibria satisfying strong forward induction will be the prediction of the game. On the other hand, one feels that at least on occasion certain restrictions upon players' beliefs may be appropriate.

Consider, for example, the game in Figure 5, a variant of Figure 2.[10] Compared to the game in Figure 2, both players now have one more strategy, V for player 1 and U for player 2. The new strategy V is strictly dominated and U becomes dominated after V is deleted. Because of the addition of U, however, Q is no longer a dominated strategy, thereby making B truly relevant for the set of equilibria giving (2, 2). Player 2's strategies M, R are too, truly relevant. Hence, the set of truly relevant strategies is $\{Out, T, B, L, M, R\}$ and the set of equilibria inducing (2, 2) satisfies strong forward induction.

On the other hand, the problem here should not be taken as a deficiency of the concept

[10] The normal-form is:

	L	M	R	U
Out	2, 2	2, 2	2, 2	2, 2
T	3, 1	0, 0	1, -1	0, -3
B	0, 0	1, 3	4, 1	1, -2
Q	-2, 0	-1, 0	0, 5	6, 3
V	-3, 0	-4, -2	-1, 2	-2, 4

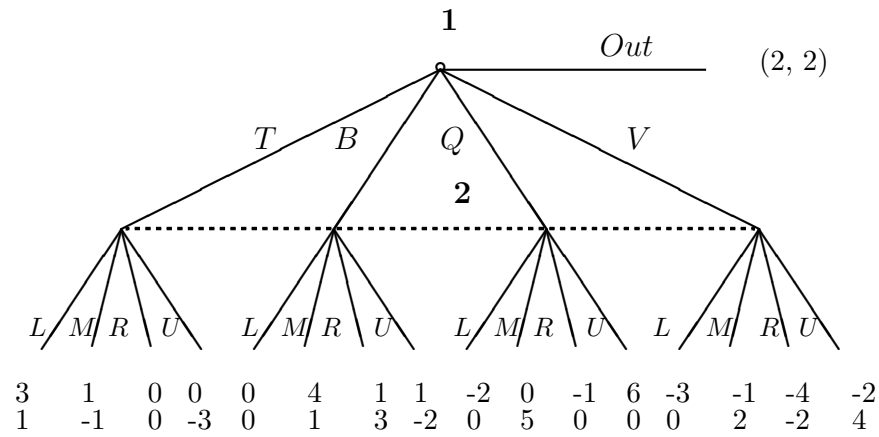


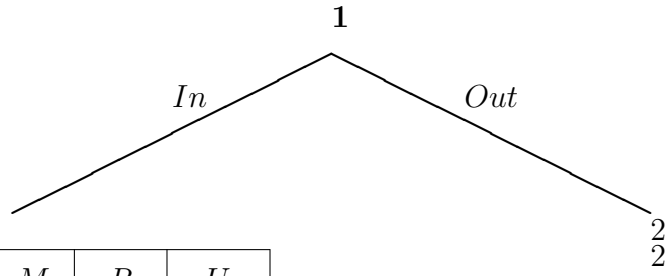
Figure 5:

of strong forward induction. Indeed, forward induction only requires inferences made by players about opponents' behaviors to be consistent with the assumption that one's opponent is playing optimally, and there is nothing in it says that after taking this inference upon rational behavior into account, they should further rationalize with the smaller choice set. Still, this example shows that conformity with strong forward induction, while being necessary for reasonable predictions, is not sufficient.

Then, a natural question arises. How can one adopt the forward induction reasoning while at the same time, not giving up other restrictions one feels appropriate for games under consideration? In what follows, we suggest a solution concept along this line.

The solution concept we propose, strong forward induction equilibrium, adopts an iterative approach to the set of equilibria satisfying strong forward induction. In the literature, similar iterative approaches to restricting off-equilibrium beliefs have been used by Pearce (1984), McLennan (1985), Reny (1992) and Man (2012).

At first, one may think that applying the relevance condition repeatedly to the set of truly relevant strategies would help eliminate unreasonable outcomes. Such a procedure does perform well in in Figure 5 and Figure 6, however, it may not perform well in some other games. As an example, consider modifying Figure 6 by replacing the outcome (1, 0) with a subgame in which two new players, 3 and 4, play a game of matching pennies. The payoff to player 1 depends on the outcome of the matching pennies game as follows. Player



	L	M	R	U
T	3, 3	0, 0	0, 1	-1, -1
B	3, 0	0, 3	-1, 1	0, -1
Q	1, -1	1, 0	-2, 2	-3, 3

Figure 6:

1 obtains a payoff of 6 if players 3 and 4 match, and obtains -4 if they do not match. Player 2's payoff is 0 regardless of the outcome of subgame. In this case, player 1's strategy Q is no longer dominated. Consequently player 2's strategies R and U would be truly relevant for (2, 2). But player 1's strategy Q remains irrelevant. Without any restriction on the relative probabilities of the two truly relevant strategies R and U no iterative procedure can rule out (2, 2).

Hence, in addition to applying the relevance condition repeatedly, one needs to apply the dominance condition in the smaller set of truly relevant strategies as well. In fact this would follow naturally from applying the strong forward induction criterion iteratively. If it is reasonable to presume that each player holds the belief that an opponent's dominated strategy is less likely than dominating strategies, then after taking this into account to obtain the set of truly relevant strategies, the same principle should be applied in the smaller set thus derived. Upon this observation, we construct an iterative procedure as follows.

First, given a set of equilibria $\mathbf{E}$, for all i, let $\overline{\mathbf{R}}_i^0(\mathbf{E}) \equiv \tilde{\mathbf{S}}_i$, and let $\overline{\mathscr{F}}^0(\mathbf{E}) \subseteq \mathbf{E}$ be a subset that contains all normal-form perfect equilibria $\tilde{\sigma}$ whose lexicographical representation puts

dominating strategies before dominated strategies. That is:

$$\overline{\mathscr{F}}^0(\mathbf{E}) \equiv \left\{ \begin{array}{l} \tilde{\sigma} \in \mathbf{E} \Big| \exists \ \{\tilde{\sigma}^\epsilon\} \text{ satisfying:} \\ \quad (i) \ \forall \ \epsilon, \ \tilde{\sigma}^\epsilon \in \tilde{\Sigma}^0 \quad \text{and } \tilde{\sigma} = \lim_{\epsilon \to 0} \tilde{\sigma}^\epsilon; \\ \quad (ii) \ \ \forall i, \ \text{if } u_i(\tilde{\sigma}_i, \tilde{\sigma}_{-i}) > u_i(\tilde{s}'_i, \tilde{\sigma}_{-i}) \text{ for all } \tilde{\sigma}_{-i} \in \tilde{\Sigma}^0_{-i}, \text{ then:} \\ \qquad \lim_{\epsilon \to 0} \dfrac{\tilde{\sigma}^\epsilon_i(\tilde{s}'_i)}{\tilde{\sigma}^\epsilon_i(\tilde{s}_i)} = 0 \qquad \forall \ \tilde{s}_i \in \ supp^m(\tilde{\sigma}_i, \tilde{s}'_i). \end{array} \right\} \tag{6}$$

Next, let ϵ be an arbitrarily small positive number, and let $\overline{\Sigma}^0_i(\epsilon) \equiv \tilde{\Sigma}^0_i$. For $n \geq 1$, define

$$\overline{\mathbf{R}}^n_i(\mathbf{E}) \equiv \left\{ \begin{array}{l} \tilde{s}_i \in \overline{\mathbf{R}}^{n-1}_i(\mathbf{E}) \Big| \tilde{s}_i \text{ is optimal against } \{\tilde{\sigma}^\epsilon\} \text{ that satisfies:} \\ \quad (i) \ \lim_{\epsilon \to 0} \tilde{\sigma}^\epsilon = \tilde{\sigma} \in \overline{\mathscr{F}}^{n-1}(\mathbf{E}); \\ \quad (ii) \ \forall \ j, \ \forall \ l \in \{0, \ldots, n-1\}, \ \forall \ \tilde{\sigma}_j \text{ with } supp(\tilde{\sigma}_j) \subset \overline{\mathbf{R}}^l_j(\mathbf{E}) \\ \qquad \text{and } \ u_j(\tilde{\sigma}_j, \tilde{\sigma}_{-j}) > u_j(\tilde{s}'_j, \tilde{\sigma}_{-j}) \text{ for all } \tilde{\sigma}_{-j} \in \overline{\Sigma}^l_{-j}(\epsilon) : \\ \qquad \lim_{\epsilon \to 0} \dfrac{\tilde{\sigma}^\epsilon_j(\tilde{s}'_j)}{\tilde{\sigma}^\epsilon_j(\tilde{s}_j)} = 0 \qquad \forall \tilde{s}_j \in supp^m(\tilde{\sigma}_j, \tilde{s}'_j) \subseteq supp(\tilde{\sigma}_j). \end{array} \right\} \tag{7}$$

$$\overline{\Sigma}^n(\epsilon) \equiv \left\{ \begin{array}{l} \tilde{\sigma} \in \overline{\Sigma}^0(\epsilon) \Big| \forall \ l \in \{1, \ldots, n\}, \quad \tilde{s} \in \overline{\mathbf{R}}^l(\mathbf{E}), \quad \tilde{s}' \in \tilde{\mathbf{S}}/\overline{\mathbf{R}}^l(\mathbf{E}) : \\ \qquad \tilde{\sigma}(\tilde{s}') < \epsilon \tilde{\sigma}(\tilde{s}). \end{array} \right\} \tag{8}$$

$$\overline{\mathscr{F}}^n(\mathbf{E}) \equiv \left\{ \begin{array}{l} \tilde{\sigma} \in \overline{\mathscr{F}}^{n-1}(\mathbf{E}) \Big| \exists \ \{\tilde{\sigma}^\epsilon\} \text{ such that:} \\ \quad (i) \ \lim_{\epsilon \to 0} \tilde{\sigma}^\epsilon = \tilde{\sigma}; \\ \quad (ii) \quad \forall \ \epsilon, \ \tilde{\sigma}^\epsilon \in \overline{\Sigma}^n(\epsilon). \end{array} \right\} \tag{9}$$

REMARK: When $n = 1$, $\overline{\mathbf{R}}^1_i(\mathbf{E})$ is the set of truly relevant strategies as defined in DEFINITION 3.3. And the set of equilibria $\mathbf{E}$ satisfies strong forward induction by DEFINITION 3.4 provided $\overline{\mathscr{F}}^1(\mathbf{E}) \neq \emptyset$.

DEFINITION 4.1. A strong forward induction equilibrium is a normal-form perfect equilibrium $\tilde{\sigma} \in \overline{\mathscr{F}}^n(\mathbf{E})$ for all $n \geq 1$.

Whereas strong forward induction alone does not eliminate the set of equilibria that induces (2, 2) in Figures 5 and Figure 6, the iterative procedure does. There is now a unique equilibrium in the two games: (3, 1) in Figure 5 and (3, 3) in Figure 6.

Proposition 1. *Every finite extensive-form game with perfect recall has at least one strong forward induction equilibrium.*

The proof can be found in the Appendix. Note that the existence result follows directly from Kohlberg and Mertens (1986), who show that every finite extensive-form game has a stable set contained in a single connected component of Nash equilibria. By definition, a stable set contains an equilibrium of any perturbed games in which every pure strategy is perturbed in the same amount towards the same completely mixed strategy. Thus, a stable set contained in a single component has at least one equilibrium that is the limit of a sequence of ϵ-perfect equilibria subject to our restrictions derived from the iterative procedure. After all, the iterative application of truly relevance and dominance conditions is just one way of perturbing the game.

Strong forward induction equilibrium is closely related to forward induction equilibrium of Man (2012). But forward induction equilibrium takes GW's definition of forward induction as a starting point and fails to take account of the global dominance restriction. Because of the two differences, forward induction equilibrium is weaker than strong forward induction equilibrium. A *strong forward induction equilibrium* of any game G is also a forward induction equilibrium of game G. This is obvious, as by construction, for all n, each n-th order truly relevant strategy satisfies the condition for n-th order relevant strategy, and an n-th order strong forward induction equilibrium meets the condition for n-th order forward induction equilibrium. Meanwhile, the definition of strong forward induction equilibrium imposes more restrictions than forward induction equilibrium; therefore, it is a stronger solution concept than forward induction equilibrium. This is true, for example, in Figures 3 and 6, where strong forward induction equilibrium eliminates more unreasonable outcomes than forward induction equilibrium.

5 Conclusion

In this paper, we point out a deficiency with the concept of forward induction defined by Govindan and Wilson (2009). We strengthen their definition by applying a dominance

condition and, as a result, give a definition of strong forward induction. By incorporating both the relevance condition and dominance condition, strong forward induction boasts stronger refining power compared with forward induction.

Meanwhile, strong forward induction criterion alone may not eliminate unreasonable predictions in some games. To remedy this shortcoming, we suggest a solution concept - - strong forward induction equilibrium - - that incorporates forward induction reasoning and other aspects of rationality. Strong forward induction equilibrium embodies an idea that, after applying the strong forward induction criterion to derive the set of equilibria satisfying strong forward induction, players should apply the same principle to the set of equilibria thus derived. Strong forward induction equilibrium is a stronger solution concept than forward induction equilibrium of Man (2012).

We do not claim that the solution concept is the right one to use, as it requires strong assumption on the sophistication of players. However, we believe that strong forward induction is a consistent concept that captures the idea of forward induction and an iterative procedure should take strong forward induction as a starting point.

Appendix A.

Proof of Proposition 1. We develop the proof in two steps. First, in Lemma 1 we show that every stable set contained in a connected component has a strong forward induction equilibrium. We then apply a result in Kohlberg and Mertens (1986) to complete the proof.

Lemma 1. *Every stable set contained in a single component of equilibria of game G has a strong forward induction equilibrium.*

Proof. The method of proof here employs in a fashion similar to Man (2012). Let $\mathbf{E}$ be a stable set contained in a single connected component of G. By definition, a closed set of Nash equilibria is stable in game G if it is minimal with respect to the following property. For any $\epsilon > 0$, there is $\delta_0 > 0$ such that for any $(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_N) \in \tilde{\Sigma}^0$ and for any $(\delta_1, \ldots, \delta_N)$ with $0 < \delta_i < \delta_0$, the perturbed game where every $\tilde{s}_i$ is replaced by $(1 - \delta_i)\tilde{s}_i + \delta_i\tilde{\sigma}_i$ has an equilibrium ϵ-close to the set.

What this implies is that for any perturbation G^ϵ of G, there exists NE of G^ϵ, which converges to an equilibrium in $\mathbf{E}$ as $\epsilon \to 0$. To apply this result, we first define sets of totally mixed strategies subject to the dominance and relevance conditions. Recall that for all i, $\overline{\mathbf{R}}_i^0(\mathbf{E}) = \tilde{\mathbf{S}}_i$ and $\overline{\Sigma}^0(\epsilon) = \tilde{\Sigma}^0(\epsilon)$ denotes the set of totally mixed strategies. For $n \geq 1$,

$$\overline{\Sigma}^n(\epsilon) \equiv \left\{ \begin{array}{l} \tilde{\sigma} \in \tilde{\Sigma}^0(\epsilon) \Big| \forall\, l \in \{1, \ldots, n\}, \quad \tilde{s} \in \overline{\mathbf{R}}^l(\mathbf{E}), \quad \tilde{s}' \in \tilde{\mathbf{S}}/\overline{\mathbf{R}}^l(\mathbf{E}) \\ \qquad \Longrightarrow \ \tilde{\sigma}(\tilde{s}') < \epsilon \tilde{\sigma}(\tilde{s}). \end{array} \right\}$$

Let $\underline{\Sigma}_i^0(\epsilon) = \overrightarrow{\Sigma}_i^0(\epsilon) = \tilde{\Sigma}_i^0(\epsilon)$. For $n \geq 1$, define

$$\underline{\Sigma}_i^n(\epsilon) \equiv \left\{ \begin{array}{l} \tilde{\sigma}_i \in \overline{\Sigma}_i^{n-1}(\epsilon) \Big| \ \forall\, l \in \{0, \ldots, n-1\}, \text{ if } \exists \tilde{\sigma}_j \text{ with } supp(\tilde{\sigma}_j) \subset \overline{\mathbf{R}}_j^l(\mathbf{E}) \\ \qquad \text{and} \quad u_j(\tilde{\sigma}_j, \tilde{\sigma}_{-j}) > u_j(\tilde{s}'_j, \tilde{\sigma}_{-j}) \text{ for all } \tilde{\sigma}_{-j} \in \overline{\Sigma}_{-j}^l(\epsilon) : \\ \qquad \lim_{\epsilon \to 0} \dfrac{\tilde{\sigma}_j^\epsilon(\tilde{s}'_j)}{\tilde{\sigma}_j^\epsilon(\tilde{s}_j)} = 0 \qquad \forall \tilde{s}_j \in supp^m(\tilde{\sigma}_j, \tilde{s}'_j). \end{array} \right\}$$

and

$$\overrightarrow{\Sigma}_i^n(\epsilon) \equiv \left\{ \begin{array}{l} \tilde{\sigma}_i \in \overline{\Sigma}_i^n(\epsilon) \Big| \forall\, l \in \{0, \ldots, n-1\}, \text{ if } \exists \tilde{\sigma}_j \text{ with } supp(\tilde{\sigma}_j) \subset \overline{\mathbf{R}}_j^l(\mathbf{E}) \\ \qquad \text{and} \quad u_j(\tilde{\sigma}_j, \tilde{\sigma}_{-j}) > u_j(\tilde{s}'_j, \tilde{\sigma}_{-j}) \text{ for all } \tilde{\sigma}_{-j} \in \overline{\Sigma}_{-j}^l(\epsilon) : \\ \qquad \lim_{\epsilon \to 0} \dfrac{\tilde{\sigma}_j^\epsilon(\tilde{s}'_j)}{\tilde{\sigma}_j^\epsilon(\tilde{s}_j)} = 0 \qquad \forall \tilde{s}_j \in supp^m(\tilde{\sigma}_j, \tilde{s}'_j). \end{array} \right\}$$

By construction, $\underline{\Sigma}_i^n(\epsilon)$ is the set of totally mixed strategies subject to the l-th order relevance condition and dominance condition for all $l \leq n-1$, and $\overrightarrow{\Sigma}_i^n(\epsilon)$ is the set of mixed strategies subject to the l-th order relevance condition for all $l \leq n$ and dominance condition for all $l \leq n-1$. For ϵ sufficiently small, $\overline{\Sigma}^n(\epsilon)$ is non-empty as long as $\overline{\mathbf{R}}_i^n(\mathbf{E})$ is non-empty for all i. Thus $\underline{\Sigma}^n(\epsilon)$ is non-empty if $\overline{\mathbf{R}}_i^{n-1}(\mathbf{E})$ is non-empty for all i, and $\overrightarrow{\Sigma}^n(\epsilon)$ is non-empty if $\mathbf{R}_i^n(\mathbf{E})$ is non-empty for all i. With proper scaling we have $\overrightarrow{\Sigma}^n(\epsilon''') \subseteq \underline{\Sigma}^n(\epsilon'') \subseteq \overrightarrow{\Sigma}^{n-1}(\epsilon') \subseteq \underline{\Sigma}^{n-1}(\epsilon)$ for all $n \geq 1$.

Next, let $\underline{\Omega}^n$ be the set of sequences of perturbations $\{\underline{G}^\epsilon\}$ of game G. The perturbations are constructed as follows: for every player i, replace each pure strategy $\tilde{s}_i \in \tilde{\mathbf{S}}_i$ by $(1-\epsilon)\tilde{s}_i + \epsilon\tilde{\tilde{\sigma}}_i^\epsilon$ where $\tilde{\tilde{\sigma}}_i^\epsilon \in \underline{\Sigma}_i^n(\epsilon)$. Also construct $\{\overrightarrow{G}^\epsilon\}$ and $\overrightarrow{\Omega}^n$ in a similar fashion: for every i, each $\tilde{s}_i$ is replaced by $(1-\epsilon)\tilde{s}_i + \epsilon\hat{\tilde{\sigma}}_i^\epsilon$ where $\hat{\tilde{\sigma}}_i^\epsilon \in \overrightarrow{\Sigma}_i^n(\epsilon)$. As long as $\overline{\mathbf{R}}_i^n(\mathbf{E})$ is non-empty for all i, $\underline{\Sigma}_i^n(\epsilon)$ and $\overrightarrow{\Sigma}_i^n(\epsilon)$ are non-empty and therefore, $\overrightarrow{\Omega}^n$ and $\underline{\Omega}^n$ are well-defined. By definition, $\overrightarrow{\Omega}^n \subseteq \underline{\Omega}^n \subseteq \overrightarrow{\Omega}^{n-1} \ldots \subseteq \overrightarrow{\Omega}^0 \subseteq \underline{\Omega}^0$.

For any game G', denote $NE(G')$ as the set of Nash equilibria. Let $\underline{\mathbf{E}}^0 = \overrightarrow{\mathbf{E}}^0 = \mathbf{E}$, and for $n \geq 1$, define

$$\underline{\mathbf{E}}^n \equiv \left\{ \underline{\sigma} \in \underline{\mathbf{E}}^{n-1} \;\middle|\; \begin{array}{l} \text{There exist } \{\underline{G}^\epsilon\} \in \underline{\Omega}^n \text{ and } \{\underline{\sigma}^\epsilon\} \text{ such that} \\ \text{for all } \epsilon,\ \underline{\sigma}^\epsilon \in NE(\underline{G}^\epsilon) \text{ and } \lim_{\epsilon\to 0} \underline{\sigma}^\epsilon = \underline{\sigma} \end{array} \right\}$$

and

$$\overrightarrow{\mathbf{E}}^n \equiv \left\{ \overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^{n-1} \;\middle|\; \begin{array}{l} \text{There exist } \{\overrightarrow{G}^\epsilon\} \in \overrightarrow{\Omega}^n \text{ and } \{\overrightarrow{\sigma}^\epsilon\} \text{ such that} \\ \text{for all } \epsilon,\ \overrightarrow{\sigma}^\epsilon \in NE(\overrightarrow{G}^\epsilon) \text{ and } \lim_{\epsilon\to 0} \overrightarrow{\sigma}^\epsilon = \overrightarrow{\sigma}. \end{array} \right\}$$

Since $\mathbf{E}$ is a stable set, by definition, there exists $\underline{\sigma} \in \underline{\mathbf{E}}^0$ ($\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^0$) for every sequence of perturbations in $\underline{\Omega}^0$ (in $\overrightarrow{\Omega}^0$).

We now show that first-order strong forward induction equilibrium exists. By definition of stable set, there exists $\underline{\sigma} \in \underline{\mathbf{E}}^1$ as a limit of a sequence of NE of the perturbed games $\{\underline{G}^\epsilon\} \in \underline{\Omega}^1$. Since every strategy of $\underline{\sigma} \in \underline{\mathbf{E}}^1$ satisfies the condition of first-order truly relevant strategy, $\overline{\mathbf{R}}_i^1(\mathbf{E})$ is nonempty for all i. Because $\overline{\mathbf{R}}^1(\mathbf{E})$ is nonempty, $\overrightarrow{\Omega}^1$ is well-defined and $\overrightarrow{\Omega}^1 \subseteq \overrightarrow{\Omega}^0$. The definition of stable set again implies that there exists $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^1$.

Furthermore, every $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^1$ is a first-order strong forward induction equilibrium. To see this, note that if $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^1$, then there exists a sequence $\{\overrightarrow{\sigma}^\epsilon\}$ as NE of a sequence of perturbed games such that $\lim_{\epsilon\to 0} \overrightarrow{\sigma}^\epsilon = \overrightarrow{\sigma}$. Clearly, for all i and for all ϵ, the NE $\overrightarrow{\sigma}^\epsilon$ of the perturbed game $\overrightarrow{G}^\epsilon$ would assign zero probability to any $\tilde{s}_i \notin \mathbf{R}_i^1(\mathbf{E})$, as by definition such $\tilde{s}_i$ can not be a best response. If we let the effective distribution resulted on $\tilde{\mathbf{S}}$ from the mixed strategy profile $\overrightarrow{\sigma}^\epsilon$ by $\widehat{\sigma}^\epsilon$, namely, for all i, $\widehat{\sigma}_i^\epsilon(\tilde{s}_i) = (1-\epsilon)\overrightarrow{\sigma}_i^\epsilon(\tilde{s}_i) + \epsilon\hat{\hat{\sigma}}_i^\epsilon(\tilde{s}_i)$, then the sequence $\{\widehat{\sigma}^\epsilon\}$ satisfies:

(i) If $\tilde{s} \in \overline{\mathbf{R}}^1(\mathbf{E})$ and $\tilde{s}' \in \tilde{\mathbf{S}}/\overline{\mathbf{R}}^1(\mathbf{E})$:

$$\frac{\widehat{\sigma}(\tilde{s}')}{\widehat{\sigma}(\tilde{s})} < \frac{\epsilon\hat{\hat{\sigma}}(\tilde{s}')}{\hat{\hat{\sigma}}(\tilde{s})} = \epsilon \to 0;$$

(ii) For all i, if $u_i(\tilde{\sigma}_i, \overrightarrow{\sigma}_{-i}) > u_i(\tilde{s}_i', \overrightarrow{\sigma}_{-i})$ for all $\overrightarrow{\sigma}_{-i} \in \overrightarrow{\Sigma}_{-i}^0$:

$$\frac{\widehat{\sigma}_i(\tilde{s}_i')}{\widehat{\sigma}_i(\tilde{s}_i)} < \frac{\epsilon\hat{\hat{\sigma}}_i(\tilde{s}_i)}{\hat{\hat{\sigma}}_i(\tilde{s}_i)} = \epsilon \to 0 \quad \forall\ \tilde{s}_i \in supp^m(\tilde{\sigma}_i, \tilde{s}_i').$$

Thus, by definition, an equilibrium $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^1$ is a first-order strong forward induction equilibrium, indicating $\overrightarrow{\mathbf{E}}^1 \subseteq \overline{\mathscr{F}}^1(\mathbf{E})$.

Now we are in a position to show that for all n, $\overrightarrow{\mathbf{E}}^n$ exists and in addition, $\overrightarrow{\mathbf{E}}^n \subseteq \overline{\mathscr{F}}^n(\mathbf{E})$. For $n = 1$, we have already shown that $\overrightarrow{\mathbf{E}}^1$ is non-empty and $\overrightarrow{\mathbf{E}}^1 \subseteq \overline{\mathscr{F}}^1(\mathbf{E})$. Suppose that for $l \leq n$, for every sequence of perturbation in $\overrightarrow{\Omega}^l$, there exists a $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^l$ and $\overrightarrow{\mathbf{E}}^l \subseteq \overline{\mathscr{F}}^l(\mathbf{E})$. We first show that $\overrightarrow{\mathbf{E}}^{n+1}$ is non-empty. By definition, any $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^l$ is the limit of a sequence of NE of the perturbed games. Since $\overrightarrow{\Omega}^{n+1} \subseteq \overrightarrow{\Omega}^n$, there must exist $\overrightarrow{\sigma} \in \overrightarrow{\mathbf{E}}^{n+1}$ if $\overline{\mathbf{R}}_i^{n+1}(\mathbf{E})$ is nonempty for all i. But as $\overrightarrow{\Omega}^n$ is well-defined by assumption, $\underline{\Omega}^{n+1}$ is also well-defined. Again by definition of stable set, there exists $\underline{\sigma} \in \underline{\mathbf{E}}^{n+1}$ for every sequence of perturbations in $\underline{\Omega}^{n+1}$. Moreover, since every strategy that is part of $\underline{\sigma} \in \underline{\mathbf{E}}^{n+1}$ is an $(n+1)$-th order truly relevant strategy, $\overline{\mathbf{R}}_i^{n+1}(\mathbf{E})$ is non-empty for all i. This in turn implies that $\overrightarrow{\Omega}^{n+1}$ is well-defined and the set $\overrightarrow{\mathbf{E}}^{n+1}$ is non-empty.

We have yet to show that $\overrightarrow{\mathbf{E}}^{n+1} \subset \overline{\mathscr{F}}^{n+1}(\mathbf{E})$. But similar argument that shows $\overrightarrow{\mathbf{E}}^1 \subseteq \overline{\mathscr{F}}^1(\mathbf{E})$ would also show $\overrightarrow{\mathbf{E}}^{n+1} \subseteq \overline{\mathscr{F}}^{n+1}(\mathbf{E})$. Hence, Lemma 1 holds by induction. □

Kohlberg and Mertens (1986) show that every finite extensive-form game with perfect recall has a stable set contained in a single connected component. Hence, we conclude that strong forward induction equilibrium always exists. □

Appendix B. Extensive-form forward induction

The extensive-form forward induction in GW is weaker than normal-form forward induction. Man (2012) shows that it is not an invariant concept. Moreover, as it places restriction on off equilibrium path belief in weakly sequential equilibrium, it fails admissibility as well.

We first provide a brief review of the definition of *extensive-form forward induction.* To start with, the extensive-form forward induction is defined as a property of an equilibrium outcome, and built upon *weakly sequential equilibrium* of Reny (1992).

DEFINITION 5.1. A consistent assessment (b, μ) constitutes a *weakly sequential equilibrium* in the extensive form game Γ_E if for all players i and for all player i's information

sets $h_i \in \mathbf{H}_i$ that her own behavioral strategy b_i does not exclude, b_i is a best response to b_{-i} given player i's belief μ,

$$b_i \in arg \max_{b_i' \in \mathbf{B}_i} E_{\mu,(b_i',b_{-i})}[u_i(Z)|h] \tag{A.1}$$

Unlike the sequential equilibrium of Kreps and Wilson (1982), in which players make optimal choice at any point in the game tree, weakly sequentiality requires each player to make optimal choice only at information sets not excluded by her own strategy but puts no restrictions at information sets excluded by her own previous choice. As a result, weakly sequential equilibrium allows players to have any beliefs about an opponent's continuation play at an information set excluded by her own strategy. As a refinement of weakly sequential equilibrium, forward induction imposes restriction on the beliefs players may hold at information sets excluded by an opponent's own strategy. As a first step, for an equilibrium outcome, the set of relevant strategies is determined as follows:

DEFINITION 5.2. (DEFINITION 3.3 in GW) A pure behavioral strategy b_i of player i is *relevant* for an outcome P if there exists a weakly sequential equilibrium (b^*, μ) giving outcome P such that b_i prescribes an optimal continuation play given i's equilibrium belief at every information set not excluded by b_i.

They then define *relevant information set* for an outcome (DEFINION 3.4 in GW) as those that can be reached with positive probability by every profile of relevant strategies for that outcome. The definition of extensive-form forward induction then places restrictions on player's off equilibrium path belief in the following fashion.

DEFINITION 5.3. (DEFINITION 3.5 in GW): An outcome satisfies *forward induction* if it results from a weakly sequential equilibrium in which at every information set that is relevant for that outcome the support of the belief of the player acting there is confined to profiles of Nature's strategies and other players' strategies that are relevant for that outcome.

Because of the restriction imposed by DEFFINITION 5.3, at any relevant information set h for outcome P, for any two pure strategy profiles s and s' reaching h with positive

probability, if s is relevant strategy profile for outcome P and s' is not, a player's expectation about opponents' behavior assigns positive weight to s and all its realization equivalent strategies, denoted by $[s]$, but assigns zero weight to s' and all its realization equivalent strategies, denoted by $[s']$. To understand the restrictions requires knowing all relevant information sets as well as relevant strategies for an outcome. For readers' convenience we therefore simplify the restriction as follows.

Let $\mathbf{R}_i(P)$ be the set of relevant strategies of player i for an outcome P and $\mathbf{R}(P)$ be the set of relevant strategy profiles. Denote by $prob([s_i]|b_i)$ the weight assigned to the set of pure strategies that are realization equivalent to s_i by the behavioral strategy b_i, and by $prob([s]|b) = \prod_{i\in\mathbf{I}} prob([s_i]|b_i)$. An outcome satisfies forward induction if it results from a weakly sequential equilibrium (b^*, μ) and there exists a sequence $\{(b^k, \mu^k)\}$ with $(b^*, \mu) = \lim_{k\to\infty}(b^k, \mu^k)$ such that if $s \in \mathbf{R}(P)$ and $s' \in \mathbf{S}/\mathbf{R}(P)$,

$$\lim_{k\to\infty} \frac{prob([s']|b^k)}{prob([s]|b^k)} = 0. \tag{A.2}$$

To define extensive-form forward induction that is equivalent to DEFINITION 3.4, we first strengthen weak sequentiality with a form of cautious play as Reny (1992, page 633) suggests. Instead of requiring each player's behavioral strategy b_i to be a best response to others' behavior strategies b_{-i}, this requires players' strategies to be a best response to a sequence of totally mixed behavior strategies b^k_{-i} whose limit is b_{-i}. We follow Man (2012) and call the strengthened version of weakly sequential equilibrium as *weakly quasi-perfect equilibrium.* Let the set of totally mixed behavioral strategies be Π^m, and denote the set of mixed consistent assessment by $\Psi^m \equiv \{(b, \mu)\}$ with $b \in \Pi^m$.

DEFINITION 5.4. An assessment (b, μ) constitutes a *weakly quasi-perfect equilibrium* in the extensive form game Γ_E if there exists a sequence of consistent assessment $(b^k, \mu^k) \in \Psi^m$ with $\lim_{k\to\infty}(b^k, \mu^k) = (b, \mu)$ such that, for all players i, and at all information sets $h_i \in \mathbf{H}_i$ not excluded by b_i,

$$b_i \in arg \max_{b'_i \in \mathbf{B}_i} E_{(\mu^k, b^k)}[u_i(Z)|h] \tag{A.3}$$

The requirement of cautious play suggests that the appropriate definition of *relevant strategy* become:

DEFINITION 5.5. A pure behavioral strategy b_i of a player i is *relevant* for outcome P if there exists a weakly quasi-perfect equilibrium (b^*, μ) with outcome P and a sequence of consistent assessment $(b^k, \mu^k) \in \Psi^m$ with $\lim_{k\to\infty}(b^k, \mu^k) = (b^*, \mu)$ such that at every information set h not excluded by b_i,

$$b_i \in arg \max_{b_i' \in B_i} E_{(\mu^k, b^k)}[u_i(Z)|h].$$

Next, we derive an extensive-form definition of forward induction equivalent to GW's normal-form version (DEFINITION 3.1):

DEFINITION 5.6. An outcome P satisfies *forward induction* if it results from a weakly quasi-perfect equilibrium (b^*, μ) and there exists a sequence $(b^k, \mu^k) \in \Psi^m$ with $(b^*, \mu) = \lim_{k\to\infty}(b^k, \mu^k)$ such that all $s \in \mathbf{R}(P)$ and all $s' \in \mathbf{S}/\mathbf{R}(P)$,

$$\lim_{k\to\infty} \frac{prob([s']|b^k)}{prob([s]|b^k)} = 0. \tag{A.4}$$

With the modification, extensive-form forward induction as defined in DEFINITION 5.6 satisfies invariance and admissibility. However, it still leaves open the relative probabilities of dominated strategies and dominating strategies. Below we provide an equivalent definition of *truly relevant strategy* in the extensive-form and then give the definition of extensive-form strong forward induction. First, a weakly quasi-perfect equilibrium (b, μ) *satisfies the dominance condition* if it is the limit of a sequence (b^k, μ^k) such that for all i, if $u_i(b_i, b_{-i}) > u_i(s_i', b_{-i})$ for all $b_{-i} \in \Pi_{-i}^m$, then

$$\lim_{k\to\infty} \frac{prob([s_i']|b_i^k)}{prob([s_i]|b_i^k)} = 0 \qquad \forall\ s_i \in supp^m(b_i, s_i'). \tag{A.5}$$

DEFINITION 5.7. A pure behavioral strategy b_i of player i is *truly relevant* for outcome P if: 1) there exists (b^*, μ) satisfying the dominance condition and inducing P; and 2) there is a sequence of consistent assessment $(b^k, \mu^k) \in \Psi^m$ with $\lim_{k\to\infty}(b^k, \mu^k) = (b^*, \mu)$ such that at every information set h not excluded by b_i,

$$b_i \in arg \max_{b_i' \in \mathbf{B}_i} E_{(\mu^k, b^k)}[u_i(Z)|h].$$

Finally, we define extensive-form forward induction that is free of inconsistent belief:

DEFINITION 5.8. An outcome P satisfies *strong forward induction* if it results from a weakly quasi-perfect equilibrium (b^*, μ) satisfying the relevance condition. That is, there exists a sequence of consistent assessment $(b^k, \mu^k) \in \Psi^m$ with $\lim_{k\to\infty}(b^k, \mu^k) = (b^*, \mu)$ such that for all $s \in \mathbf{R}(P)$ and all $s' \in \mathbf{S}/\mathbf{R}(P)$,

$$\lim_{k\to\infty} \frac{prob([s']|b^k)}{prob([s]|b^k)} = 0.$$

References

Banks, J., Sobel, J., 1987. Equilibrium selection in signaling games. Econometrica 55, 647–661.

Battigalli, P., Siniscalchi, M., 2002. Strong belief and forward induction reasoning. Journal of Economic Theory 106, 356–391.

Blume, L., Brandenberger, A., Dekel, E., 1991. Lexicographic probabilities and equilibria refinements. Econometrica 59, 81–98.

Cho, I.-K., Kreps, D. M., 1987. Signaling games and stable equilibria. Quarterly Journal of Economics 102, 179–222.

Govindan, S., Wilson, R., 2009. On forward induction. Econometrica 77, 1–28.

Hillas, J., Kohlberg, E., 2002. Conceptual foundations of strategic equilibria. In: Aumann, R., Hart, S. (Eds.), Handbook of Game Theory. Vol. 3. New York: Elsevier, pp. 1597–1663.

Kohlberg, E., Mertens, J.-F., 1986. On the strategic stability of equilibria. Econometrica 54, 1003–1037.

Kreps, D., Wilson, R., 1982. Sequential equilibria. Econometrica 50, 863–894.

Kuhn, H. W., 1953. Extensive games and the problem of information. In: Kuhn, H. W., Tucker, A. W. (Eds.), Contributions to the Theory of Games. Vol. 2. Princeton University Press, Princeton, NJ, pp. 193-216.

Man, P. T., 2012. Forward induction equilibrium. Games and Economic Behavior 75, 265–276.

McLennan, A., 1985. Justifiable beliefs in sequential equilibrium. Econometrica 53, 889–904.

Myerson, R., 1978. Refinements of the Nash equilibrium concept. International Journal of Game Theory 7, 73–80.

Pearce, D. G., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.

Reny, P. J., 1992. Backward induction, normal form perfection and explicable equilibria. Econometrica 60, 627–649.

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