Extensive Form Abstract Economies and Generalized Perfect Recall*

Nicholas Butler†

September 13, 2015

Abstract

I introduce the notion of an extensive form abstract economy (EFAE) to model dynamic strategic settings where players’ feasible strategies depend on those chosen by others. Extensive form abstract economies are motivated by models of limited strategic complexity, and nest standard extensive form games, games played by finite automata, and games with rational inattention. I provide sufficient conditions for existence of an equilibrium in behavioral strategies in extensive form abstract economies, the most salient of which is a generalized convexity condition that is equivalent to perfect recall in extensive form games. I demonstrate that in an extensive form abstract economy players may engage in incredible self restraint, a novel type of incredible threat. In response, I propose a refinement related to perfect equilibrium, and show that such an equilibrium exists under slightly stronger conditions.

---

*Economics Department, Princeton University. Email: nob@princeton.edu

†Economics Department, Princeton University. Email: nob@princeton.edu
1 Introduction

In the theory of noncooperative games, the simplest model of a strategic situation is a normal form game. The dynamic extension of these models are extensive form games, where the timing of players’ decisions and information are modeled explicitly. Another variant, where the strategies available to players depend on those chosen by others, is the notion of an abstract economy introduced by Debreu (1952). In this paper, I develop a theory of extensive form abstract economies (EFAE), a class of models which combine these features.

In extensive form games, the actions and local strategies available at a given moment are determined by the structure of the game tree and information partition. Implicit in this framework lie the following assumptions. First, the available actions at a given node depend exclusively on the realized history of actions. Second, restrictions on local strategies across nodes are limited to those that impose equality. Third, the nodes for which these equality restrictions apply are specified exogenously by the information partition.

These assumptions are relaxed in extensive form abstract economies. In lieu of an information partition, each player is endowed with a correspondence associating a set of feasible strategies to each strategy profile of others. Feasibility correspondences allow general endogenous restrictions to be imposed on players’ strategies, rather than the exogenous equality constraints imposed by information sets. Depending on the particular setting, players’ feasibility correspondences may be used to model a variety of salient phenomena including physical constraints, bounded rationality, complexity constraints, nonstandard consistency conditions, and endogenous information acquisition. The relationship between extensive form abstract economies and abstract economies is analogous to the relationship between extensive form games and normal form games. A minor difference is that a normal form game is an abstract economy with trivial feasibility correspondences, whereas an extensive form game is an extensive form abstract economy with feasibility correspondences induced by players’ information partitions.

Extensive form abstract economies are a general framework encompassing numerous classes of models with nonstandard features. In particular, they are the natural framework for modeling dynamic settings where the complexity and historical dependence of players’ strategies are restricted. Models where players choose the distribution of their own signals are ubiquitous, and can be easily mapped into this framework. Games played by finite automata (Rubinstein [1986]; Abreu and Rubinstein [1988]) can be naturally represented as extensive form abstract economies, with feasibility correspondences derived from the struc-

---

1I utilize the same abbreviation (EFAE) to refer to the singular and plural of extensive form abstract economies throughout.
tural assumptions imposed on the automata. Multi-self models of decision making with imperfect recall (Gilboa [1997]; Aumann, Hart, and Perry [1997]) are extensive form abstract economies, with feasibility correspondences derived from the consistency restrictions imposed on the strategies of the various selves. Extensive form equilibrium refinements, including perfect equilibrium and subgame perfect equilibrium, can be obtained as equilibrium of appropriately defined extensive form abstract economies. Dynamic general equilibrium models can also be formulated as extensive form abstract economies by introducing an auctioneer, where players’ feasibility correspondences combine their information sets and budget sets.

A primary motivation for this paper is the distinct suitability of extensive form abstract economies for modeling strategic situations with endogenous information: where players acquire information subject to constraints determined in equilibrium. In these settings, unlike in extensive form games, the strategies available to players may depend on the strategies chosen by others and not the realized history alone.\(^2\) In particular, this paper is motivated by the growing literature on games with rational inattention (Sims 1998), where the mutual information between players’ strategies and the history is limited. Martin (2013) considers a game where consumers are rationally inattentive towards the quality of a product, and producers take this into account. Yang (2012) considers optimal security design and coordination games with rational inattention. In these papers, players are rationally inattentive towards the actions of nature, but not towards the actions of other players. I demonstrate that in an extensive form abstract economy, feasibility correspondences can be constructed in which players are rationally inattentive towards both.

The focus of this paper is on determining sufficient conditions for existence of an equilibrium in behavioral strategies in extensive form abstract economies. The analogous condition in finite extensive form games is known as perfect recall, under which Kuhn (1953) showed that mixed and behavioral strategies are equivalent. As a result, an equilibrium in behavioral strategies exists in any game with perfect recall, since an equilibrium in mixed strategies always exists in finite games. However, there are apparently weaker conditions than perfect recall which are sufficient for existence of an equilibrium in behavioral strategies. For instance, an equilibrium in behavioral strategies exists if players do not exhibit absentmindedness (Piccione and Rubinstein [1997]), and players’ behavioral strategies are equivalent to a convex set of mixed strategies.\(^3\) The first major result of this paper, Theorem 1, demonstrates that there is no gap between these conditions. I show that in extensive form

\(^2\)Modeling these situations as standard extensive form games gives rise to conceptual difficulties regarding existence and interpretation of equilibria.

\(^3\)Notice this condition is weaker than perfect recall since under perfect recall mixed and behavioral strategies are equivalent, and the set of mixed strategies is convex.
games, perfect recall is equivalent to players’ sets of behavioral strategies being *behavioral convex*, a novel generalized convexity condition. In addition, I prove that behavioral convexity is the weakest restriction on players’ behavioral strategies under which the hypotheses of standard fixed point theorems are necessarily satisfied.

Since it is a geometric property unlike perfect recall, behavioral convexity can be applied to the feasibility correspondences of extensive form abstract economies. This motivates the following definition: a feasibility correspondence satisfies *generalized perfect recall* if it is behavioral convex valued. The second major result of this paper, Theorem 2, provides sufficient conditions for existence of an equilibrium in behavioral strategies in an extensive form abstract economy, the most salient being generalized perfect recall. The remaining hypotheses include a nonstandard notion of continuity necessary to accommodate rational inattention, and some standard technical conditions. An extensive form abstract economy satisfying these remaining hypotheses is called *canonical*.

Generalized perfect recall is a global restriction on players’ behavioral strategies. As such, verifying that it is satisfied by an arbitrary feasibility correspondence is nontrivial. In response, I introduce the notion of *partitional convexity*, and provide a simple sufficient condition for generalized perfect recall which utilize this notion. Finally, I show that various types of useful feasibility correspondences satisfy this condition, including those describing rational inattention.

To demonstrate that generalized perfect recall is tight, I consider a sequence of canonical extensive form abstract economies that converge to a game with perfect recall, whose members have no equilibrium in behavioral strategies. This analysis is related to Wichardt (2008), which provided an explicit example of a finite extensive form game with no equilibrium in behavioral strategies; however, the example presented in this paper differs in two ways. First, in the game considered in Wichardt (2008) players have imperfect recall regarding their own actions, whereas players can recall their own actions in my example. Second, the non-existence result of this paper is stronger, since there is a natural notion of convergence of feasibility correspondences.

I conclude by demonstrating that in an extensive form abstract economy players may engage in a novel type of incredible threat, which I refer to as *incredible self restraint*. The intuition behind this notion is the following. Holding fixed the strategy profile of others, players choose their behavioral strategies from feasibility sets that do not necessarily have a Cartesian product structure. As a result, a player’s choice at a node may limit his choices at another. Players can exploit this by selecting strategies that constrain their choices off the equilibrium path, allowing them to make seemingly incredible threats. I demonstrate that natural extensions of standard refinements, such as subgame perfect equilibrium, do not
rule out this type of threat. In response, I propose an extension of extensive form perfect equilibrium introduced by Selten (1975), which I refer to as *generalized perfect equilibrium*, and show that it precludes incredible self restraint. The final major result of this paper, Theorem 3, provides sufficient conditions for existence of a generalized perfect equilibrium. These include the hypotheses of Theorem 2, and a minor condition which ensures that players’ feasibility correspondences allow for trembling.

2 Extensive Form Abstract Economies

An *extensive form abstract economy* is a septuple \( G = \langle \mathcal{I}, A, T, \rho, b_0, u, Q \rangle \), with constituents as given as follows. The set of players is \( \mathcal{I} = \{0, ..., I\} \), where 0 denotes nature. The set of all actions is specified by \( A \), and the set of all sequences with members in \( A \) is given by \( A^* \). The tree \( T \subset A^* \) is a finite set of sequences with members in \( A \). I assume that the empty sequence \( \phi \in T \) and that if \( (a_1, ..., a_k) \in T \) then \( (a_1...a_{k-1}) \in T \). Elements of the tree are called *nodes*, and identify the sequence of actions taken up to that point. The set of available actions at a particular node \( w \) is given by \( A(w) = \{a \in A : (w, a) \in T\} \). The nodes with no available actions are called *terminal nodes* and are designated by \( Z \).

The player function \( \rho : T \setminus Z \rightarrow \mathcal{I} \) maps nonterminal nodes to the player who acts at that node. The set of nodes where player \( i \) acts is denoted by \( T_i = \{w \in T : \rho(w) = i\} \). The distribution of actions of nature is given by \( b_0 \). The utility index \( u : Z \rightarrow \mathbb{R}^I \) maps terminal nodes to a utility value for each player.

A *pure strategy* of player \( i \) assigns an available action to every node where player \( i \) acts. The set of pure strategies of player \( i \) is given by,

\[
S_i = \prod_{w \in T_i} A(w)
\]

A *mixed strategy* of player \( i \) is a probability distribution over the pure strategies of player \( i \). The set of mixed strategies of player \( i \) is given by \( \Sigma_i = \Delta(S_i) \). A *behavioral strategy* of player \( i \) assigns a local strategy \( b_i(\cdot|w) \in \Delta(A(w)) \) to every node \( w \in T_i \). The set of behavioral strategies is given by,

\[
B_i = \prod_{w \in T_i} \Delta(A(w))
\]

4If any constituent of \( G \) satisfies some condition I will say that \( G \) satisfies said condition as well.

5The set of probability distributions over a sample space \( X \) is denoted by \( \Delta(X) \).

6I utilize the standard notation that \( B = \prod_{i \in I} B_i \) and \( B_{-i} = \prod_{j \neq i} B_j \), and similarly for players’ mixed and pure strategies.
The final component of an extensive form abstract economy is a collection of feasibility correspondences \( Q = \{Q_1, \ldots, Q_I\} \). In this paper, I focus on the case where the feasibility correspondences are defined over players’ behavioral strategies, since they are the most natural strategic notion in dynamic settings. Analogous results where players’ feasibility correspondences are defined over their mixed strategies can be found in the appendix.

A feasibility correspondence \( Q_i : B_{-i} \to 2^{B_i} \) is a mapping from the behavioral strategies of players other than \( i \) to the sets of behavioral strategies of player \( i \). The interpretation of these correspondences is that if other players choose \( b_{-i} \in B_{-i} \), then player \( i \) is compelled to choose some behavioral strategy \( b_i \in Q_i(b_{-i}) \). For a given strategy profile \( b \in B \), the behavioral strategy of player \( i \) is called feasible if \( b_i \in Q_i(b_{-i}) \), and the profile \( b \) is called jointly feasible if \( b_i \) is feasible for all \( i \in I \).

Since players are endowed with their own feasibility correspondence, they need only consider the feasibility of their own strategy in general. Situations where players account for the joint feasibility of the strategy profile resulting from a deviation can also be formulated as an EFAE, and is equivalent to imposing a single constraint in the space of behavioral strategy profiles. Extensive form abstract economies of this form are the dynamic analogue of games with coupled constraints introduced by Rosen (1965). Coupled constraints are the natural notion for describing physical restrictions; however, when modeling endogenous information acquisition the more general formulation is more natural. In these settings, there is nothing which explicitly prevents a deviation from resulting in a jointly infeasible strategy profile. Rather, a deviation forces players to revise their conjectures regarding the strategy profile of others, which might subsequently affect their sets of feasible strategies.

### 2.1 Extensive Form Games

By carefully choosing players’ feasibility correspondences, extensive form games can be formulated as extensive form abstract economies. These correspondences must be formulated differently depending on if mixed or behavioral strategies are the preferred notion in a particular setting. Formally, an extensive form game is a septruple \( G = \langle I, A, T, \rho, b_0, u, H \rangle \), whose first six constituents are identical to those of an extensive form abstract economy.

The final component of an extensive form game is a collection of information partitions \( H = \{H_1, \ldots, H_I\} \). Each information partition \( H_i \) is a partition of \( T_i \). Elements of \( H_i \) are called the information sets of player \( i \). The information set containing \( w \in T \) is designated by \( h(w) \). The sets of actions available at nodes in the same information set are assumed to be identical. The interpretation of an information set \( h \in H_i \) is that when a node \( w \in h \) is realized, player \( i \) cannot distinguish \( w \) from any other \( w' \in h \).
Since strategies are functions of nodes rather than information sets in extensive form abstract economies, it is necessary to restrict players’ strategies to be measurable with respect to their information partitions when formulating an extensive form game as an EFAE. Accordingly, a pure strategy is called standard if it implements the same action at nodes in the same information set. Given an information partition $H_i$, the set of standard pure strategies of player $i$ is denoted by $S^*_i(H_i)$. A mixed strategy is called standard if its support is a subset of $S^*_i(H_i)$. The set of standard mixed strategies is given by $\Sigma^*_i(H_i)$. Similarly, a behavioral strategy is called standard if it implements identical local strategies at nodes in the same information set. The set of standard behavioral strategies is given by $B^*_i(H_i)$. By construction, standard strategies are precisely the types of strategies that are feasible in extensive form games.

Definition 1. An feasibility correspondence is behavioral standard if $Q_i(\cdot) = B^*_i(H_i)$, and mixed standard if $Q_i(\cdot) = \Sigma^*_i(H_i)$.

Any extensive form game can be formulated as an extensive form abstract economy using either behavioral standard or mixed standard feasibility correspondences. The choice between these two conventions is left to the modeler, and depends on the natural notion of a strategy in the setting in question.

3 Behavioral Convexity and Perfect Recall

In this section I examine the relationship between convexity and perfect recall in extensive form games. I begin by providing an overview of well known results regarding perfect recall and equivalence between mixed and behavioral strategies. I then present some examples illustrating the relationship between perfect recall and notions of convexity. Finally, in Theorem 1, I demonstrate that this relationship holds generally.

3.1 Perfect Recall in Extensive Form Games

In extensive form games, behavioral strategies are the more natural notion than mixed strategies, since nodes represent points of decision rather than points of execution. However, the usual fixed point arguments cannot be used in the space of behavioral strategies, since players’ utility functions are not quasiconcave. Consequently, the natural method for proving existence of an equilibrium in behavioral strategies involves determining conditions under which mixed and behavioral strategies are equivalent. Since an equilibrium in mixed
Definition 2. Strategies \( x_i \) and \( x'_i \) are equivalent, written as \( x_i \sim x'_i \), if \( P(z|x_i, x_{-i}) = P(z|x'_i, x_{-i}) \) for every \( z \in Z \) and strategy profile of others \( x_{-i} \in X_{-i} \).

Definition 3. Sets of strategies \( X_i \) and \( Y_i \) are equivalent if for every \( x_i \in X_i \) there exists \( y_i \in Y_i \) such that \( x_i \sim y_i \) and conversely.

Definition 4. A player has perfect recall if for every \( h \in H_i \)

(R1) If \( w' \in h(w) \) then \( w \) is not a predecessor of \( w' \) and conversely.

(R2) If \( w' \in h(w) \) and \( \tilde{w} \) is a predecessor of \( w \), then there exists a \( \tilde{w}' \in h(w') \) such that \( \tilde{w}' \) is a predecessor of \( w' \) and \( a(\tilde{w}, w) = a(\tilde{w}', w') \)

Informally, players have perfect recall if they remember their own actions and their past knowledge. A player whose information partition violates either (R1) or (R2) is said to have
imperfect recall. If in particular $H_i$ violates (R1), then the player $i$ said to be absentminded. (R2) states that for nodes in the same information set, the acting player must visit the same information sets and take the same actions at those information sets on the path to both. It is straightforward to verify that in finite extensive forms (R2) implies (R1). An extensive form game $G$ has perfect recall if every player has perfect recall. Bonanno (2004) demonstrated that (R2) can be further decomposed into conditions describing perfect recall of past actions and perfect recall of past knowledge separately, an important distinction in the following sections.

3.2 Imperfect Recall and Non-Convexity

In this section, I seek to illustrate the relationship between convexity and perfect recall in extensive form games. Three simple extensive form decision problems are considered, each exhibiting a different type of imperfect recall: absentmindedness, imperfect recall of past actions, and imperfect recall of past knowledge. I show that the player’s feasible mixed strategies are not convex in all three. The feasibility correspondences in each example are assumed to be behavioral standard, and payoffs at terminal nodes are omitted since they are not necessary to illustrate the relevant phenomena.

Consider the extensive form depicted in Figure 1. Intuitively, the decision maker has imperfect recall of past actions, since upon reaching the second information set he can no longer distinguish between his initial actions.

![Figure 1: Imperfect Recall of Past Actions](image)

Proposition 1. The set of feasible mixed strategies of Figure 1 is not convex.

Proof. The pure strategies of this decision problem are $S_1 = \{(L, L), (L, R), (R, L), (R, R)\}$, where the action at the unreached node is omitted. The information set imposes that
Consider the following mixed strategies:

\[ \sigma_1(L, L) = \frac{1}{2}, \quad \sigma_1(L, R) = \frac{1}{4}, \quad \sigma_1(R, L) = \frac{1}{6}, \quad \sigma_1(R, R) = \frac{1}{12} \]

\[ \sigma'_1(L, L) = \frac{1}{4}, \quad \sigma'_1(L, R) = \frac{1}{12}, \quad \sigma'_1(R, L) = \frac{1}{2}, \quad \sigma'_1(R, R) = \frac{1}{6} \]

Let \( b_1 \sim \sigma_1 \) and \( b'_1 \sim \sigma'_1 \). By (1) the following must hold

\[ b_1(L|L) = \frac{\sigma_1(L, L)}{\sigma_1(L, L) + \sigma_1(L, R)} = \frac{2}{3} = \frac{\sigma_1(R, L)}{\sigma_1(R, L) + \sigma_1(R, R)} = b_1(L|R) \]

\[ b'_1(L|L) = \frac{\sigma'_1(L, L)}{\sigma'_1(L, L) + \sigma'_1(L, R)} = \frac{3}{4} = \frac{\sigma'_1(R, L)}{\sigma'_1(R, L) + \sigma'_1(R, R)} = b'_1(L|R) \]

So both mixed strategies are feasible. However, a convex combination of these mixed strategies is not necessarily feasible. Let \( \sigma''_1 = \frac{1}{2} \sigma_1 + \frac{1}{2} \sigma'_1 \) and \( b''_1 \sim \sigma''_1 \). By (1)

\[ b''_1(L|L) = \frac{\sigma''(L, L)}{\sigma''(L, L) + \sigma''(L, R)} = \frac{9}{13} < \frac{8}{11} = \frac{\sigma''(R, L)}{\sigma''(R, L) + \sigma''(R, R)} = b''_1(L|R) \]

Non-convexity arises in Figure 2 due to the conjunction of two factors. First, the probability of reaching nodes in the second information set differs under \( \sigma_1 \) and \( \sigma'_1 \). Second, the local strategies induced by \( \sigma_1 \) and \( \sigma'_1 \) at the second information set are different. Intuitively, since the probability of reaching \( R \) is higher under \( \sigma'_1 \) than under \( \sigma_1 \), the former has more influence over the local strategy at \( R \) than the latter.\(^8\)

The extensive form depicted in Figure 2 is identical to the well known “Paradox of the Absentminded Driver”. This decision problem has two standard pure strategies, namely playing \( R \) at both nodes or playing \( D \) at both nodes. Notice that under any mixed strategy that places positive weight on both pure strategies, the player always chooses \( R \) at the second node but chooses \( D \) with nonzero probability at the first.

**Proposition 2.** The set of feasible mixed strategies of Figure 2 is not convex.

In the example depicted in Figure 3, the player has imperfect recall of his past knowledge. Intuitively, since the nodes \( l \) and \( r \) are in distinct information sets, the player knows the action taken by nature. However, after choosing \( In \) he forgets the action made by nature upon arriving at his subsequent information set. Non-convexity arises in this example for the

\(^8\)Of course the opposite holds at the other node in the second information set \( L \).
Figure 2: Absentmindedness

Figure 3: Imperfect Recall of Past Knowledge

same reasons as in the first. Specifically, if two feasible strategies choose \( In \) with different probabilities at \( l \) and \( r \), and choose different local strategies at the subsequent information set, a nontrivial convex combination of the two is not feasible.

**Proposition 3.** The set of feasible mixed strategies of Figure 3 is not convex.

### 3.3 Behavioral Convexity

The examples of the previous section are suggestive of a relationship between convexity and perfect recall in extensive form games. In this section, I demonstrate that this relationship holds generally. The primary result of this section, Theorem 1, states that in essentially any extensive form game a player has perfect recall if and only if his set of behavioral strategies is behavioral convex. Moreover, I prove that behavioral convexity of players’ sets of behavioral strategies is equivalent to convexity of their sets of feasible mixed strategies.

**Definition 5.** A tree \( T \) is irreducible if it contains no nodes having only one available action.

An extensive form game or extensive form abstract economy is called irreducible if its constituent tree is irreducible. Intuitively, a node with only one available action can be
pruned from the tree without any loss of generality since the action taken at that node is predetermined.

The probability of reaching \( w \) given that the other players choose the action on the path to \( w \) with probability one wherever possible is,

\[
P_i^*(w|b_i) = \prod_{w' \in P_i(w)} b_i(a(w', w)|w')
\]  

(2)

The node dependent mixing coefficient of player \( i \) is \( \Lambda_i : B_i \times B_i \times [0, 1] \times T_i \rightarrow [0, 1] \), defined by

\[
\Lambda_i(b_i, b'_i, \lambda, w) = \frac{\lambda P_i^*(w|b_i)}{\lambda P_i^*(w|b_i) + (1 - \lambda) P_i^*(w|b'_i)}
\]  

(3)

The behavioral mixing function of player \( i \) is \( B_i : B_i \times B_i \times [0, 1] \rightarrow B_i \), defined componentwise using (3) as follows,

\[
B_i(b_i, b'_i, \lambda)(a|w) = \Lambda_i(b_i, b'_i, \lambda, w)b_i(a|w) + (1 - \Lambda_i(b_i, b'_i, \lambda, w))b'_i(a|w)
\]  

(4)

A behavioral convex combination of two behavioral strategies \( b_i, b'_i \in B_i \) is a behavioral strategy \( b''_i = B_i(b_i, b'_i, \lambda) \) for some \( \lambda \in [0, 1] \). By (4) each component of a behavioral convex combination \( B_i(b_i, b'_i, \lambda)(a|w) \) is a convex combination of \( b_i(a|w) \) and \( b'_i(a|w) \). The node dependent mixing coefficient \( \Lambda_i(b_i, b'_i, \lambda, w) \) adjusts for the relative probability of reaching \( w \) under \( b_i \) versus \( b'_i \). In general, \( B_i(b_i, b'_i, \lambda) \) is not a convex combination of \( b_i \) and \( b'_i \), since \( \Lambda_i \) depends on \( w \).

Definition 6. A set \( V_i \subset B_i \) is behavioral convex if \( v_i, v'_i \in V_i \) implies that any behavioral convex combination of \( v_i \) and \( v'_i \) is also an element of \( V_i \).

The relationship between the relevant notions of convexity for behavioral and mixed strategies is given by the following proposition, which relies on the characterization of equivalent mixed and behavioral strategies stated in (1).

Proposition 4. The set of feasible mixed strategies is convex if and only if the set of behavioral strategies is behavioral convex.

By Proposition 4, the sets of behavioral strategies in the previous examples are not behavioral convex. The following theorem demonstrates that the relationship between perfect recall and convexity observed in these examples holds generally.

Theorem 1. If \( G \) is irreducible then \( G \) satisfies perfect recall if and only if the set of behavioral strategies is behavioral convex.
The proof of sufficiency is straightforward. Perfect recall implies that the set of feasible mixed strategies is equivalent to the set of mixed strategies, which is convex by construction. To complete the proof, I show that the set of feasible mixed strategies is convex and apply Proposition 4. The proof of necessity mirrors the examples of the previous section. For each type of imperfect recall: absentmindedness, imperfect memory of past actions, and imperfect memory of past knowledge, I construct feasible behavioral strategies that have infeasible behavioral convex combinations. This demonstrates that imperfect recall implies that the set of behavioral strategies is not behavioral convex, the contrapositive of which is the desired result.

The primary significance of Theorem 1 is that it establishes a previously unknown geometric characterization of perfect recall, which can be extended to the more general framework of extensive form abstract economies. By revealing the underlying geometry implicit in the assumption of perfect recall, Theorem 1 provides a more complete and intuitive understanding of why perfect recall is sufficient for existence of an equilibrium in behavioral strategies.

The results of this section demonstrate that as far as existence of equilibrium in behavioral strategies is concerned perfect recall is tight. Proposition 4 and Theorem 1 imply that there is no weaker condition than perfect recall which guarantees that players’ feasible mixed strategy best response correspondences are convex valued, a necessary condition for applying standard fixed point arguments.

4 Equilibrium in Extensive Form Abstract Economies

In this section I formally introduce the notion of an equilibrium in behavioral strategies for extensive form abstract economies. I then provide sufficient conditions for such an equilibrium to exist. Analogous results for mixed strategies can be found in the appendix.

The probability of reaching a node \( w \) under the behavioral strategy profile \( b \) is given by,

\[
P(w|b) = \prod_{i \in I} \prod_{w' \in P_i(w)} b_i(a(w', w)|w')
\]

and the expected utility of player \( i \) under \( b \) is given by,

\[
U_i(b) = \sum_{z \in Z} P(z|b)u_i(z).
\]
Definition 7. An equilibrium in behavior strategies is a $b \in B$ such that for every $i \in \mathcal{I}$

$$b_i \in \arg \max_{b'} U_i(b_i, b_{-i})$$

Equilibria in behavioral strategies in extensive form abstract economies are the natural analogue of a Nash equilibria in behavioral strategies in extensive form games. In equilibrium, players take the strategy profiles of the others as given and choose feasible behavioral strategies to maximize their expected utility.

4.1 Existence of Equilibrium

In this section I determine sufficient conditions for existence of an equilibrium in behavioral strategies in extensive form abstract economies. While some of the conditions are standard, I utilize a novel notion of continuity, which is necessary to accommodate certain types of useful feasibility correspondences.

Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be endowed with the usual metric. Let $C : Y \to 2^X$ be a correspondence, and $f : X \times Y \to \mathbb{R}$.

Definition 8. A correspondence $C$ is $f$-lower hemicontinuous if for any $y_n \to y$ and $x \in C(y)$, there exists a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ and $x_{n_k} \in C(y_{n_k})$ such that $f(x_{n_k}, y_{n_k}) \to f(x, y)$.

This definition is identical to the sequential characterization of lower hemicontinuity, except it requires that $f(x_{n_k}, y_{n_k}) \to f(x, y)$, rather than $x_{n_k} \to x$. Notice that if $C$ is lower hemicontinuous and $f$ is jointly continuous then $C$ is $f$-lower hemicontinuous. Intuitively, $f$-lower hemicontinuity of $C$ means that the values of $f$ that are achievable in the graph of $C$ do not change discontinuously in the limit. This notion of continuity is relevant because some interesting feasibility correspondences, including those describing rational inattention, are $U_i$-lower hemicontinuous but not lower hemicontinuous.

Definition 9. A compact valued correspondence $C$ is upper hemicontinuous if for any $y^n \to y$ with $x^n \in C(y^n)$, there exists a convergent subsequence of $\{x^n\}$ with limit in $C(y)$.

A correspondence is called $f$-continuous if it is both upper hemicontinuous and $f$-lower hemicontinuous. The following lemma establishes that if $f$ is the objective function, then the maximum theorem holds under $f$-continuity of the feasibility correspondence.

Lemma 1. Let $f : X \times Y \to \mathbb{R}$, be jointly continuous and $C : Y \to 2^X$ be a compact valued correspondence. Let $f^*(y) = \max_{x \in C(y)} f(x, y)$ and $C^*(y) = \arg \max_{x \in C(y)} f(x, y)$.

If $C$ is $f$-continuous then $f^*$ is continuous, and $C^*$ is non-empty, compact valued, and upper hemicontinuous.
The proof of Lemma 1 is similar to standard proofs of the maximum theorem. Some of the hypotheses of Lemma 1 can be weakened, but the form above is sufficiently general for the purpose of proving Theorem 2. In particular Lemma 1 holds if $f$-continuity is weakened to apply only to the constrained maximizers of $f$.

**Definition 10.** A feasibility correspondence $Q_i$ is *canonical* if it satisfies the following,

(Q1) $Q_i$ is compact and nonempty valued

(Q2) $Q_i$ is $U_i$-continuous

(Q3) If $b_j \sim b'_j$ for all $j \neq i$ then $Q_i(b_{-i}) = Q_i(b'_{-i})$

An extensive form abstract economy is called *canonical* if the feasibility correspondence of each player is canonical. The interpretation of a canonical extensive form abstract economy is as follows. (Q1) is a standard technical condition, and ensures that players’ best response correspondences are well defined. Since $U_i$ denotes the expected utility of player $i$ it is continuous, so (Q2) is weaker than standard continuity. By Lemma 1, (Q1) and (Q2) ensure that players’ best response correspondences are upper hemicontinuous. While these conditions are primarily technical, they do have natural interpretations. (Q1) means that regardless of what the other players do, a player can always choose to do something. (Q2) has the interpretation that players’ feasible behavioral strategies and achievable utilities do not change discontinuously when other players change their strategies. (Q3) imposes equivalence classes on $Q$ which correspond to the equivalence classes of $B$ induced by the equivalence relation. This restriction is quite weak because if $b_i \sim b'_i$ then they can only differ at nodes that are reached with zero probability for any behavioral strategy of the other players.\(^9\) (Q3) can be interpreted as requiring that whatever a player claims he will do at a node that could never be realized, based on his strategy alone, does not affect the feasible strategies of other players. In the absence of (Q3), players can costlessly manipulate the constraints of other players.\(^10\) In practice, (Q3) ensures that any equilibrium in mixed strategies of the induced normal form abstract economy is equivalent to an equilibrium in behavioral strategies of the extensive form.

**Definition 11.** A player has *generalized perfect recall* if his feasibility correspondence is behavioral convex valued.

\(^9\)From the perspective of player $i$ it is impossible to reach histories where this indeterminacy manifests.

\(^10\)These manipulations are not only costless holding the strategy profile of the other players fixed, but also costless for any strategy profile of the other players.
An extensive form abstract economy satisfies \textit{generalized perfect recall} if each player has generalized perfect recall. If the feasibility correspondence of a player is not behavioral convex valued then that player is said to have \textit{generalized imperfect recall}. These definitions are inspired by Theorem 1, which implies that an extensive form game has perfect recall if and only if its behavioral standard representation has generalized perfect recall. Moreover, generalized perfect recall plays the same role as perfect recall in extensive form games regarding existence of an equilibrium in behavioral strategies. Specifically, each ensures that players’ feasible mixed strategy best response correspondences are convex valued.

\textbf{Theorem 2.} If $G$ is finite, canonical, and satisfies generalized perfect recall then it has an equilibrium in behavioral strategies.

The proof of Theorem 2 is relatively straightforward. I begin by associating with $G$ a normal form abstract economy by transforming players’ feasibility correspondences into the space of mixed strategies. Players’ best response correspondences are well defined since their feasibility correspondences are canonical, and are convex valued since they have generalized perfect recall. Care must be taken to ensure that players’ transformed feasibility correspondences satisfy the hypotheses of Lemma 1, which implies that players’ transformed best response correspondence is upper hemicontinuous. Applying a standard fixed point argument to players’ best response correspondences in the space of mixed strategies, and confirming that the fixed point is feasible, yields an equilibrium of the normal form abstract economy. The primary difficulty arises in ensuring that an equilibrium in mixed strategies of the associated normal form abstract economy yields an equilibrium in behavioral strategies in the extensive form, since the mapping between equivalent mixed and behavioral strategies is not bijective. (Q3) is sufficient to show that this is indeed the case.

A second approach to proving Theorem 2 is to first apply Lemma 1 to players’ behavioral strategy best response, then transform it into a mixed strategy best response. Care must be taken to ensure that this transformation preserves upper hemicontinuity. The primary difficulty in this approach involves showing that the resulting mixed strategy best response is convex valued. The critical lemma for this approach is the following.

\textbf{Lemma 2.} $U_i(\mathcal{B}(b_i, b'_i, \lambda), b_{-i}) = \lambda U_i(b_i, b_{-i}) + (1 - \lambda) U_i(b'_i, b_{-i})$

Intuitively, Lemma 2 means that utility is “linear” in behavioral convex combinations of behavioral strategies. This ensures that players’ behavioral strategy best response is behavioral convex valued, which implies their mixed strategy best response is convex valued. Both methods are presented in detail in the appendix.
5 Tightness of Generalized Perfect Recall

Since Theorem 2 provides sufficient conditions for existence of an equilibrium, a natural question to ask is whether these conditions are tight. In this section, I seek to answer this question as it pertains to generalized perfect recall. I fix an extensive form and consider a sequence of canonical EFAE with generalized imperfect recall which converge to an EFAE with generalized perfect recall. I show that none of the EFAE along the sequence have an equilibrium in behavioral strategies, whereas the limiting EFAE does. I also define a natural notion of distance, which measures the degree to which an EFAE departs from generalized perfect recall, and show that this distance converges to zero in my example.

Consider the extensive form abstract economy depicted in Figure 4. The dashed lines indicate equality constraints, whereas the dashed enclosure indicates a looser constraint specified below. I assume that the feasibility correspondence of player 2 imposes the following restrictions,

\[ b_2(l|Ll) = b_2(l|Rl) \quad b_2(l|Lr) = b_2(l|Rr) \quad |b_2(l|L) - b_2(l|R)| \leq \epsilon, \]

In addition, I assume that the feasibility correspondence of player 1 requires that,

\[ b_1(L|Ll) = b_1(L|Rl) = b_1(L|Lr) = b_1(L|Rr) \]
\[ b_1(L|Rl) = b_1(L|Rr) = b_1(L|Rrl) = b_1(L|Rrr) \]
For a given \( \epsilon \geq 0 \), I refer to the EFAE depicted in Figure 4 as \( G^\epsilon \). Intuitively, these EFAE describe type of sequential matching pennies. Player 2 wins if his second action matches the first action of player 1, and his first action matches the second action of player 1. The feasibility correspondences of both players are canonical; however, for \( \epsilon > 0 \) player 2 has generalized imperfect recall. In effect he forgets his past knowledge, since he is able to partially condition his first action on the first action of player 1, but is unable to do so at subsequent nodes. Since player 2 seeks to match his second action to the first action of player 1, he can use his first action to signal to himself at his future information set; however, this may come at the cost of poorly predicting the second action of player 1.

The intuition for why an equilibrium does not exist for \( \epsilon > 0 \) is as follows. Fixing a strategy of player 1, player 2 best responds by utilizing his self signalling capability. Player 2 effectively encodes his partial knowledge using his first action by choosing different local strategies at \( L \) and \( R \). Since player 2 remembers his first action when he chooses his second action, he can use this information at subsequent nodes to partially infer the first action of player 1. However, holding the strategy of player 2 fixed, player 1 may alter his action at his second information set so that the signalling scheme of player 1 predicts the second action of player 2 poorly. In fact, an equilibrium in behavioral strategies does not exist unless this unusual signalling capability is completely suppressed by setting \( \epsilon = 0 \).

**Proposition 5.** If \( \epsilon > 0 \) then \( G^\epsilon \) does not have an equilibrium in behavioral strategies.

The proof of Proposition 5 proceeds in two steps. Showing that for any \( \epsilon > 0 \) and any behavioral strategy of player 1, player 2 can secure a payoff of strictly more than 1/4 is relatively straightforward. The more challenging step, which relies on a recursive argument, is showing that player 1 can secure a payoff of at least 3/4 for any strategy of player 2. These steps imply that there exists no strategy profile where both players are best responding, since the total payoff is 1.

Proposition 5 is a striking result, and is considerably stronger than existing examples of extensive form games with no equilibrium in behavioral strategies. Intuitively, while there is no natural notion of convergence for information partitions, there are natural ways to measure the distance between a given feasibility correspondence and those feasibility correspondences satisfying generalized perfect recall. Consider for example the following distance function.

\[
D(Q_i, \tilde{Q}_i) = \inf_{\tilde{Q}_i} \sup_{b_{-i}} d^H(Q_i(b_{-i}), \tilde{Q}_i(b_{-i}))
\]

Where \( Q_i \) is the set of feasibility correspondences of player \( i \) which satisfy generalized
perfect recall, $d$ is a metric on $B_i$ and $d^H$ is the Hausdorff distance given by,

$$d^H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$$

Intuitively, the distance given in (5) measures the amount of imperfect recall exhibited by $Q_i$ by finding a feasibility correspondence $\tilde{Q}_i$ with generalized recall that minimizes the maximum Hausdorff distance between $Q_i$ and $\tilde{Q}_i$ as a function of $b_{-i}$. It is straightforward to verify that $D(Q_2, Q_i) \rightarrow 0$, since $G^0$ is a behavioral standard extensive form game with perfect recall. This implies there are canonical extensive form abstract economies whose feasibility correspondences exhibit arbitrarily small departures from generalized perfect recall which have no equilibrium in behavioral strategies.

The results of this section demonstrate that care must be exercised when utilizing feasibility correspondences that do not satisfy generalized perfect recall, since arbitrarily small departures may lead to non-existence of equilibrium. In this sense, generalized perfect recall can be considered tight.

6 Partitional Convexity and Endogenous Information

Since feasibility correspondences are global restrictions on players' behavioral strategies, checking that a given extensive form abstract economy satisfies generalized perfect recall is nontrivial. On the other hand, determining whether an information partition satisfies perfect recall is relatively simple. In this section, I formulate a simple and intuitive sufficient condition for generalized perfect recall. Like perfect recall, it is defined via a partition of players' decision nodes, which makes constructing feasibility correspondences satisfying generalized perfect recall much simpler.

I begin by introducing a notion of partitional convexity, and show that under certain conditions if a players' feasibility correspondence is partitional convex valued, then that player has generalized perfect recall. Like perfect recall in extensive form games, partitional convexity imposes structure on the types of partitions restrict players' feasible strategies. However, unlike an information partition it does not require that players choose the same local strategy at nodes in the same partition element. Rather, it requires the sets of feasible local strategies within each partition element to be convex. I conclude by utilizing partitional convexity to demonstrate that a variety of natural feasibility correspondences, including rational inattention, satisfy the hypotheses of Theorem 2.
6.1 Partitional Convexity

Let \( Z \subset \mathbb{R}^m \), \( I = \{1, \ldots, m\} \) and \( P = \{P_1, \ldots, P_K\} \) be a partition of \( I \). If \( i, j \in I \) are members of the same partition element denote this by \( i \sim j \).

**Definition 12.** A \( P \)-partitional convex combination of elements \( z, z' \in Z \) is a \( z'' \in Z \) such that \( z'' = \lambda_i z_i + (1 - \lambda_i) z'_i \) for some \( \lambda \in [0, 1]^m \) satisfying if \( i \sim j \) then \( \lambda_i = \lambda_j \).

A \( P \)-partitional convex combination is formed by taking different convex combinations of the components of \( z \) and \( z' \) at every equivalence class induced by \( P \). A set \( Z \) is said to be \( P \)-partitionally convex if \( z, z' \in Z \) implies every \( P \)-partitional convex combination of \( z \) and \( z' \) is also an element of \( Z \). Partitional convexity imposes significant structure on the form of \( Z \), as the following lemma demonstrates.

**Lemma 3.** A set \( Z \) is \( P \)-partitionally convex if and only if \( Z = \prod_{k=1}^K Z^k \) where each \( Z^k \subset \mathbb{R}^{|P_k|} \) is convex.

Intuitively, a \( P \)-partitionally convex set must have a product structure that respects the partition \( P \). The proof of necessity is immediate, and sufficiency is obtained through an inductive argument on \( K \). Notice that any behavioral standard feasibility correspondences is \( H_i \)-partitionally convex valued by construction.

6.2 Partitional Convexity and Perfect Recall

A partitional decomposition of a feasibility correspondence \( Q_i \) is a pair \((\tilde{Q}_i, H_i)\) satisfying \( Q_i = \tilde{Q}_i \cap B_i^*(H_i) \). Partitional decompositions are generally not unique, and every feasibility correspondence has at least one, since \( H_i \) can be taken to be the set singletons. An extensive form abstract economy with nontrivial partitional decomposition can be thought of as a behavioral standard extensive form game with residual constraints specified by \( \tilde{Q}_i \).

In settings with endogenous information acquisition, \( H_i \) describes the nodes that are a priori indistinguishable, whereas \( \tilde{Q}_i \) describes how players can partially differentiate certain nodes upon acquiring information.

Perfect recall is the relevant restriction on players’ information partitions ensuring existence of an equilibrium in extensive form games. An analogous restriction for extensive form abstract economies is generated by the following partition.

\[
\mathcal{R}_i = \{W \subset T_i \mid w, w' \in W \iff R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i)\} \tag{6}
\]

Each element of \( \mathcal{R}_i \) is a maximal set of decision nodes which can be reached under the same standard pure strategies of player \( i \). Intuitively, player \( i \) cannot distinguish between
nodes contained in an element of $\mathcal{R}_i$ based on his past actions or past knowledge. However, a player may be able to partially distinguish between nodes in $W_i$ based on his current knowledge. This suggests a relationship between $\mathcal{R}_i$ and perfect recall, which is given in the following proposition.

**Proposition 6.** If $G$ is irreducible, an information partition $H_i$ satisfies perfect recall if and only if $H_i$ is finer than $\mathcal{R}_i$.

Proposition 6 is analogous to Theorem 1, and its proof is similar. Intuitively, $\mathcal{R}_i$ characterizes the coarsest information partitions of player $i$ which satisfy perfect recall. A residual constraint is $\mathcal{R}_i$-partitional convex valued if the local strategy chosen at a node affects the permissible local strategies at another node only if the information sets and actions taken along the path to both are identical. Moreover, any restriction the local behavioral strategies must be convex. The relationship between $\mathcal{R}_i$-partitional convexity and generalized perfect recall is as follows.

**Proposition 7.** If $H_i$ satisfies perfect recall and $\tilde{\mathcal{Q}}_i$ is $\mathcal{R}_i$-partitionally convex valued then $\mathcal{Q}_i = \tilde{\mathcal{Q}}_i \cap B_i^*(H_i)$ satisfies generalized perfect recall.

### 6.3 Endogenous Information Constraints

Strategic situations where players devote a fixed amount of resources or attention to gathering information are naturally modeled as extensive form abstract economies. Players that fit these criteria include government agencies with fixed appropriations, divisions of corporations with a fixed budget, and individuals with a fixed attention capacity. In this section I consider two natural types of feasibility correspondences, and demonstrate that they satisfy the hypotheses of Theorem 2. For both types, players are not generally free to choose an arbitrary behavioral strategy. Rather, the degree to which players’ local strategies can depend on the realized node is limited, and depends on the strategies of others. As a result, both types of constraints can be viewed as describing endogenous information acquisition or processing.

The *constraint partition* of player $i$ is a partition of $T_i$ given by $W_i = \{W_i^1, ..., W_i^K\}$. Elements of a constraint partition are referred to as *constraint sets*. For a given feasibility correspondence and partitional decomposition, a *partition trio* is a triple $(\mathcal{R}_i, W_i, H_i)$. A partition $X$ is said to be *finer* than a partition $Y$, if for any $x \in X$ there is a $y \in Y$ such that $x \subset y$. If the opposite holds $X$ is said to be *coarser* than $Y$.

**Definition 13.** A partition trio is *standard* if $W_i$ is both finer than $\mathcal{R}_i$ and coarser than $H_i$. 

21
In a standard partition trio $\mathcal{H}_i$ is finer than $\mathcal{R}_i$, so by Proposition 6 $\mathcal{H}_i$ satisfies perfect recall. The set of local behavioral strategies of player $i$ at constraint set $W^k_i$ is given by,

$$B^k_i = \prod_{w \in W^k_i} \Delta(A_i(w))$$

A local feasibility correspondence is a $Q^k_i : B_{\equiv i} \rightarrow B^k_i$. For the remainder of this section I consider feasibility correspondence of the following form.

$$Q_i(\cdot) = \hat{Q}_i(\cdot) \cap B^*_i(H_i) = \prod_{k=1}^{K_i} Q^k_i(\cdot) \cap B^*_i(H_i)$$

A feasibility correspondences of this form has the following interpretation. Upon arriving at a node in a constraint set that is not a singleton, the player cannot immediately determine which node has realized. Moreover, his behavioral strategy at nodes outside the constraint set have no effect on the local behavioral strategies he can choose at the constraint set and conversely. If a local constraint imposes equality of the local strategies across the nodes in the constraint set, then the constraint set can be interpreted exactly like an information set. However, imposing weaker local constraints allows players’ behavioral strategies to depend on the realized history to a limited extent. This can be interpreted as describing either information acquisition or information processing. Under the former interpretation, the realized history is hidden but players have some capacity to investigate. Under the latter interpretation, information regarding the realized history is readily available; however, players are unable to map this information into actions with arbitrary precision due to limited cognitive capabilities, bounded rationality, or limitations in the channel of transmission. In either case, the optimal way to gather or process information arises endogenously.

The model of player $i$ at constraint set $W^k_i$, denoted by $p^k_i$, is the distribution over nodes in $W^k_i$ induced by the conjectured strategy profile conditioning on everything player $i$ knows upon arriving at $W^k_i$. The mutual information between a constraint set $W^k_i$ and the local behavioral strategy of player $i$ is given by the following.\(^{11}\)

$$I(W^k_i, A(W^k_i)) = \sum_{w \in W^k_i} \sum_{a \in A(w)} b_i(a|w)p^k_i(w)\log\left(\frac{b_i(a|w)}{\sum_{w \in W^k_i} b_i(a|w)p^k_i(w)}\right)$$

A constraint set $W^k_i$ is on the path of play of $b$ if there exists a node $w \in W^k_i$ such that $P(w|b) > 0$. A constraint set on the potential path of play of $b_{\equiv i}$ if there exists a $b_i \in B_i$ such that it is on the path of play of $(b_i, b_{\equiv i})$. The constraint sets of player $i$ that are on the

\(^{11}\)See the appendix for a derivations of $I$ and $p^k_i$. 

22
potential path of $b_{-i}$ are denoted by $\Pi_i(b_{-i})$.

**Definition 14.** A rational inattention constraint is a weakly positive vector of capacities $c_i \in \mathbb{R}^{K_i}$ and a feasibility correspondence $Q_i$ satisfying,

$$Q^k_i(b_{-i}) = \begin{cases} \{b^k_i \in B^k_i \mid I(W^k_i, A(W^k_i)) \leq c_{ik}\}, & \text{if } W^k_i \in \Pi_i(b_{-i}) \text{.} \\ B^k_i, & \text{otherwise.} \end{cases}$$

For each constraint set, the capacity specifies the maximum mutual information between players’ local behavioral strategy and the history of actions up to that point. The assumption that players are unconstrained off the path of play ensures that $Q_i$ is upper hemicontinuous, but it also has an intuitive interpretation. Off the potential path of play, players’ models are not pinned down by Baye’s rule, hence any model is permissible. It is straightforward to verify that if $p_k^i = 1$ for some node in a constraint set, then the mutual information between any local behavioral strategy and that constraint set is zero. In that case every local behavioral strategy is feasible since capacities are assumed to be weakly positive. Hence, no behavioral strategy can be excluded off the potential path of play without making additional assumptions on players’ off path models.

**Proposition 8.** If $(R_i, W_i, H_i)$ is standard, and $G$ is irreducible, then rational inattention constraints satisfy the hypotheses of Theorem 2.

The proof of Proposition 8 relies on Proposition 7 and the fact that mutual information is a convex function of the conditional distribution. Notice that rational inattention constraints are not lower hemicontinuous, since they are unconstrained off the path of play. However, they are $U_i$-lower hemicontinuous since their feasibility correspondences only change discontinuously at nodes that are off the path of play in the limit.

Rational inattention constraints have the following interpretation inspired by information theory. When a player arrives at a constraint set, he has a model in mind governing the nodes in the constraint set that is consistent with everything he knows, as well as the equilibrium strategy profile. The player then observes the realized history through a channel, which is designed optimally subject to his model and exogenously specified capacity, and acts accordingly. Under the information acquisition interpretation, the capacity represents the limited resources the player has at his disposal to acquire the hidden information. Under the information processing interpretation, the capacity represents the limited computational and reasoning faculties of the player. In either case, in an equilibrium each players’ model coincides with the conditional distribution over constraint sets induced by the equilibrium strategy profile, and players design their channels optimally given their models and capacities.
While rational inattention constraints are intuitively appealing due to their information theoretic foundations, their functional form may not be applicable in all situations involving endogenous information acquisition. A variant on the notion of rational inattention, with a similar interpretation, are feasibility correspondences of the following form.

**Definition 15.** A *continuous diameter* constraint is a $Q_i$ satisfying

(CD1) $Q_i(b_{-i}) = \{b_i^k \in B_i^k : |b_i^k(w) - b_i^k(w')| \leq d_i^k(b_{-i})\}$

(CD2) $d_i^k : B_{-i} \to \mathbb{R}$ is continuous and weakly positive

(CD3) If $b_{-i} \sim b_{-i}'$ then $d_i^k(b_{-i}) = d_i^k(b_i)$

**Proposition 9.** If $(R_i, W_i, H_i)$ is standard, and $G$ is irreducible, then continuous diameter constraints satisfy the hypotheses of Theorem 2.

The proof of Proposition 9 is similar to that of Proposition 8. However, unlike rational inattention constraints, continuous diameter constraints are lower hemicontinuous. Continuous diameter constraints have the following intuitive interpretation. At a constraint set $W_i^k$, the local strategies of player $i$ are required to fit inside a closed ball of diameter $d_i^k$. This can be interpreted as a two step procedure where players first specify a status quo for each constraint set, given by the center of the closed ball, and subsequently choose their local strategies to fit inside the balls. If the diameter is equal to zero then the constraint set behaves like an information set. At the other extreme, if the diameter is greater than one then any local behavioral strategy is feasible. A particularly tractable type of continuous diameter constraints, utilized in previous sections, are *constant diameter constraints* satisfying $d_i^k(\cdot) = c_i^k$ for some constants $c_i^k \in \mathbb{R}^+$.

Both rational inattention and continuous diameter constraints possess the feature that every behavioral strategy composed of local strategies that are equal at nodes in the same constraint set are feasible. Intuitively, a player who utilizes a strategy of this form is not gathering any new information beyond that of his information partition. In addition, these types of constraints have the natural property that a behavioral strategy constructed by randomizing independently over feasible local behavioral strategies is also feasible. For these reasons, both of these types of constraints are natural for describing strategic situations with endogenous information acquisition.

### 7 Generalized Perfection and Incredible Self Restraint

In this section I introduce the notion of a *generalized perfect equilibrium* (GPE) of an extensive form abstract economy, an extension of the notion of extensive form perfect equi-
librium in an extensive form game. Theorem 3 provides sufficient conditions for existence of a generalized perfect equilibrium that are nearly identical to those of Theorem 2. In addition to the usual concerns regarding incredible actions off the equilibrium path, the notion of a generalized perfect equilibrium is motivated by the possibility of novel incredible threats that are particular to extensive form abstract economies. Specifically, a player may use his local strategy at one node to limit his feasible actions at another node off the equilibrium path. I refer to this type of threat as incredible self restraint. Moreover, while naive extensions of many standard notions of perfection do not rule out incredible self restraint, generalized perfection does.

Consider the extensive form abstract economy depicted in Figure 5 below. The feasibility correspondence of player 2 imposes the constraint that $|b_2(L|M) - b_2(L|R)| \leq 1/2$. The interpretation is that player 2 is able to distinguish the node $L$ from the other two, but can only partially distinguish the nodes $M$ and $R$ from each other.

![Figure 5: Incredible Self Restraint](image_url)

**Proposition 10.** The following strategy profile is an equilibrium of Figure 5

$$b_1(L|\epsilon) = 1 \quad b_2(R|L) = 1 \quad b_2(R|M) = 1/2 \quad b_2(R|R) = 1$$

This equilibrium is peculiar in the following sense. Player 1 chooses not to play $M$ because player 2 follows by playing $R$ half the time, to the detriment of both players. At first glance, the local strategy of player 2 at $M$ does not appear to be credible; however, due to his feasibility correspondence he cannot choose an arbitrary local strategy $M$. Fixing his local strategy at $R$, player 2 is in fact best responding at $M$ subject to the induced constraint $b_2(L|M) \leq 1/2$. Since player 2 is indifferent between his actions at $R$, and since he is best responding at $L$, his strategy is credible at these nodes as well. In this sense, the equilibrium described in Proposition 10 is “subgame perfect”, since player 2 is best responding at each of his decision nodes subject to the constraints generated by fixing his behavioral strategy out-
side the node. This example suggests that naive extensions of standard subgame perfection are too permissive, as the following thought experiment confirms.

Suppose player 2 was informed that play had arrived at either $M$ or $R$, with positive probability on each. Fixing his local strategy at $L$, he is afforded the opportunity to deviate at his other nodes simultaneously, provided the resulting behavioral strategy remains feasible. In order to be best responding player 2 must choose $b_2(L|M) = 1$ and $b_2(L|R) \geq 1/2$, so the equilibrium of Proposition 10 is not credible in this sense. Intuitively, the strategy of player 2 exhibits incredible self restraint since his local strategy at $R$ restrains his local strategy at $M$, in an effort to deter player 1 from playing $M$ in the first place. Moreover, if player 2 was ever faced with the situation described above he would deviate.

Standard notions of perfection are insufficient in an extensive form abstract economy, since they only address continuation play beginning at individual nodes (subgame perfection) or at equality constraints (sequential equilibrium). While sequential equilibrium is usually considered sufficient in extensive form games, in an EFAE players face a more general feasibility correspondence which affords them additional opportunities to make incredible threats.

A generalized perfect equilibrium addresses this issue by requiring players’ equilibrium strategies to be the limit of equilibrium strategy profiles of $\epsilon$-modified games. This is the natural extension of perfect equilibrium (Selten 1975) to the setting of extensive form abstract economies, and has a similar interpretation. Since every node is reached with positive probability in an $\epsilon$-modified game, in an equilibrium players’ must be best responding conditional on arriving at any set of nodes. This remains true in the limit, which rules out standard incredible threats as well as incredible self restraint.

For a feasibility correspondence $Q_i$ denote its $\epsilon$-modification by $Q_i(b_{-i}, \epsilon) = Q_i(b_{-i}) \cap \Gamma_i(\epsilon)$, where $\Gamma_i(\epsilon)$ are the behavioral strategies of player $i$ in which $b_i(a|w) \geq \epsilon$ for any node $w \in T_i$ and $a \in A(w)$. Denote by $G^\epsilon$ the extensive form abstract economy obtained by replacing the feasibility correspondences of $G$ with its $\epsilon$-modification.

**Definition 16.** A strategy profile $b \in B$ is a generalized perfect equilibrium of $G$ if $b$ is an equilibrium of $G$, and there exists $\epsilon_n \to 0$ and $b_n \to b$ such that $b_n$ is an equilibrium of $G^{\epsilon_n}$.

**Definition 17.** A feasibility correspondence admits trembling if there exists a $\delta > 0$ such that for any $\epsilon \leq \delta$, its $\epsilon$-modification is nonempty valued and $U_i$-lower hemicontinuous.\(^{(12)}\)

An extensive form abstract economy admits trembling if every player’s feasibility correspondence admits trembling. While this is a relatively weak restriction it is not without consequence. First, it requires that if the probability of trembling is taken to be small enough,\(^{(12)}\)This means $U_i$-lower hemicontinuous in both $b_{-i}$ and $\epsilon$, where $U_i(b, \epsilon) = U_i(b, \epsilon')$, for any $\epsilon, \epsilon'$

26
then the $\epsilon$-modified feasibility correspondence is well defined. Second, since the intersection of $U_i$-lower hemicontinuous correspondences is not necessarily $U_i$ lower hemicontinuous, it imposes some structure on how $Q_i(\cdot, \cdot)$ varies.

Theorem 3. If $G$ is finite, canonical, admits trembling, and satisfies generalized perfect recall then it has generalized perfect equilibrium.

The proof of this theorem resembles that of Selten (1975), except I operate in the space of behavioral strategies rather than in the space of mixed strategies of the agent normal form. The primary difficulty lies in showing that the $\epsilon$-modified abstract economies of $G$ possesses an equilibrium. Care must be taken to ensure that the $\epsilon$-modified feasibility correspondences satisfy the hypotheses of Theorem 2, $U_i$-lower hemicontinuity in particular. It is straightforward to show that the equilibrium of Proposition 10 is not totally perfect. For any full support strategy of player 1, a best response for player 2 in $G^\epsilon$ must satisfy $b_2(R|M) = \epsilon$, hence any totally perfect equilibrium must satisfy $b_2(R|M) = 0$.

8 Conclusion

This paper lays the groundwork for the theory of extensive form abstract economies, a family of models describing dynamic strategic situations with nonstandard constraints. Extensive form abstract economies generalize the notion of an extensive form game, and a variety of existing nonstandard game theoretic models can be formulated as EFAE. In this paper, I showed that in extensive form games perfect recall is equivalent to a notion of behavioral convexity. I then introduced the notion of an extensive form abstract economy and provided sufficient conditions for existence of an equilibrium in behavioral strategies. The critical sufficient condition is a generalized convexity condition referred to as generalized perfect recall, which is equivalent to perfect recall in extensive form games. I showed that perfect recall is tight in extensive form games, and that generalized perfect recall is tight in extensive form abstract economies. Using the notion of partitional convexity, I provided a simple sufficient condition for generalized perfect recall. Specific feasibility correspondences, including rational inattention and continuous diameter constraints were shown to satisfy this condition. Finally, motivated by the phenomenon of incredible self restraint, a novel type of incredible threat, I introduced the notion of a generalized perfect equilibrium and provided sufficient conditions for such an equilibrium to exist.

There are many avenues for further inquiry in both the theory and applications of extensive form abstract economies. Perhaps the most obvious theoretical direction to pursue

---

13It is straightforward to show that the correspondences considered in the previous section admit trembling.
is extending the results of this paper to infinite extensive form abstract economies. Another possible direction is to formulate a theory of extensive form abstract economies with generalized imperfect recall. Such a theory might be useful for games in which players almost satisfy generalized perfect recall. Additionally, extensive form abstract economies may be used to construct equilibrium refinements for standard extensive form games by perturbing players’ feasibility correspondences. There might also exist a satisfactory extension of sequential equilibrium to extensive form abstract economies. In any case, while this paper lays a foundation for the theory of extensive from abstract economies, there are many more questions to be asked and answered regarding their theory and applications.

9 Appendix

9.1 Definitions

The set of terminal nodes is given by $Z = \{w \in T \mid A(w) = \emptyset\}$. The nodes where player $i$ acts are denoted by $T_i = \{w \in T \mid \rho(w) = i\}$. Recall that $P_i(w)$ is the set of predecessors of $w$ where player $i$ acts. For $w \in P_i(w')$ let $a(w, w')$ be the action satisfying either $(w, a(w, w')) \in P_i(w')$ or $(w, a(w, w')) = w'$. Intuitively, $a(w, w')$ is the action taken at $w$ on the path to $w'$.

For a node $w = (a_1^w a_2^w \ldots a_k^w)$, define inductively for $n \leq k$, $w_0 = \phi$, and $w_n = (w_{n-1}, a_n^w)$. Let $R(w)$ be the set of pure strategy profiles $s$ such that $s_i \in R_i(w)$ for every $i \in I$. The probability of reaching a node $w$ under a mixed strategy profile $\sigma$ is given by the following:

$$P(w|\sigma) = \sum_{s \in R(w)} \sigma(s) \prod_{w' \in P_i(w)} b_i(a(w', w)|w')$$

The expected utility of player $i$ under $\sigma$ is given by

$$U_i(\sigma) = \sum_{z \in L_T} P(z|\sigma) u_i(z)$$

The probability of reaching $w$ if other players choose the actions on the path to $w$ is

$$P_i(w|\sigma_i) = \sum_{s_i \in R_i(w)} \sigma_i(s_i)$$

An immediate consequence is that for any $\sigma \in \Sigma$ and $b \in B$ that if $\sigma \sim b$ then $U_i(\sigma) =$
The set of standard pure strategies is denoted by

\[ S_i^*(H_i) = \{ s_i \in S_i \mid w' \in h(w) \Rightarrow s_i(w) = s_i(w') \} \]

The set of standard mixed strategies is denoted by

\[ \Sigma_i^*(H_i) = \{ \sigma_i \in \Sigma_i \mid s_i \notin S_i^*(H_i) \Rightarrow \sigma_i(s_i) = 0 \} \]

The set of standard behavioral strategies is denoted by

\[ B_i^*(H_i) = \{ b_i \in B_i \mid w' \in h(w) \Rightarrow b_i(\cdot|w) = b_i(\cdot|w') \} \]

For a behavioral standard game with information partition \( H \), denote the set of feasible mixed strategies of player \( i \) by \( \Sigma_i^* \). Similarly denote his feasible behavioral strategies by \( B_i^* \).

### 9.2 Imperfect recall and non-convexity

**Proposition 2.** The set of feasible mixed strategies of Figure 2 is not convex.

**Proof.** Omitting the actions taken at unreached nodes, the pure strategies are given by

\[ S_1 = \{(D), (R, D), (R, R)\} \]

Consider the mixed strategies \( \sigma_1(D) = 1 \) and \( \sigma'_1(R, R) = 1 \). Notice that \( \sigma_1 \sim b_1 \) where \( b_1(D|\phi) = b_1(D|R) = 1 \), and \( \sigma'_1 \sim b'_1 \) where \( b_1(R|\phi) = b_1(R|R) = 1 \), so they are feasible. However \( \sigma''_1 = \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma'_1 \sim b''_1 \) where \( b''_1(R|\epsilon) = \frac{1}{2} \) and \( b''_1(R|R) = 1 \), so \( \sigma''_1 \) is not feasible. \( \square \)

**Proposition 3.** The set of feasible mixed strategies of Figure 3 is not convex.

**Proof.** Assume that \( 0 < b_0(l) < 1 \). Omitting actions taken at unreached nodes the pure strategies are given by

\[ S_1 = \{(Out), (In, L), (In, R)\}^2 \]

Consider the following mixed strategies.

\[ \sigma_1(\text{In}, L|l) = \sigma_1(\text{Out}|r) = 1 \]
\[ \sigma'_1(\text{Out}|l) = \sigma'_1(\text{In}, R|r) = 1 \]

29
So \( \sigma_1 \sim b_1 \) where \( b_1(L|l, In) = b_1(L|r, In) = 1 \), and \( \sigma'_1 \sim b'_1 \) where \( b_1(R|l, In) = b_1(R|r, In) = 1 \); however, \( \sigma''_1 = \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma'_1 \sim b''_1 \) where \( b''_1(L|l, In) = b''_1(R|r, In) = 1 \). So \( \sigma''_1 \) is not feasible.

\[ \square \]

9.3 Behavioral Convexity

**Proposition 4.** The set of feasible mixed strategies are convex if and only if the set of behavioral strategies is behavioral convex.

**Proof.** I first establish the following claim.

**Claim.** If \( b_i, b'_i \in B_i \) and \( \sigma_i, \sigma'_i \in \Sigma_i \) such that \( b_i \sim \sigma_i \) and \( b'_i \sim \sigma'_i \), then \( B_i(b_i, b'_i, \lambda) \sim \lambda \sigma_i + (1 - \lambda) \sigma'_i \).

Assume the hypotheses of the claim above, this implies that

\[
P^i(w|\sigma_i) > 0 \implies b_i(a_i|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)}
\]

And similarly for \( \sigma'_i \). Let \( \lambda \in [0, 1] \), let \( \sigma''_i = \lambda \sigma_i + (1 - \lambda) \sigma'_i \), and let \( b''_i \sim \sigma''_i \).

If \( P^i(w|\sigma''_i) > 0 \) then it follows that

\[
b''_i(a_i|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma''_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma''_i(s_i)} = \frac{\sum_{s_i \in R_i(w)} \lambda \sigma_i(s_i) + (1 - \lambda) \sigma'_i(s_i)}{\sum_{s_i \in R_i(w)} \lambda \sigma_i(s_i) + (1 - \lambda) \sigma'_i(s_i)}
\]

\[
\frac{\lambda P^i(w|\sigma_i)}{\lambda P^i(w|\sigma_i) + (1 - \lambda) P^i(w|\sigma'_i)} b_i(a_i|w) + \frac{(1 - \lambda) P^i(w|\sigma'_i)}{\lambda P^i(w|\sigma_i) + (1 - \lambda) P^i(w|\sigma'_i)} b'_i(a_i|w) =
\]

\[
\frac{\lambda P^i(w|b_i)}{\lambda P^i(w|b_i) + (1 - \lambda) P^i(w|b'_i)} b_i(a_i|w) + \frac{(1 - \lambda) P^i(w|b'_i)}{\lambda P^i(w|b_i) + (1 - \lambda) P^i(w|b'_i)} b'_i(a_i|w) = B(b, b', \lambda)(a_i|w)
\]

So \( \lambda \sigma_i + (1 - \lambda) \sigma''_i \sim B(b_i, b'_i, \lambda) \) for every \( \lambda \in [0, 1] \), which proves the claim.

Let \( b_i, b'_i \in B_i^*(H_i) \) and \( \sigma_i, \sigma'_i \in \Sigma_i^* \) such that \( b_i \sim \sigma_i \) and \( b'_i \sim \sigma'_i \).

Suppose \( \Sigma_i^* \) is convex. By the claim for any \( \lambda \in [0, 1] \) the convex combination \( \lambda \sigma_i + (1 - \lambda) \sigma'_i \sim B(b_i, b'_i, \lambda) \). Since \( \Sigma_i^* \) is convex \( \lambda \sigma_i + (1 - \lambda) \sigma'_i \in \Sigma_i^* \) which implies \( B(b_i, b'_i, \lambda) \in B_i^*(H_i) \)
by construction, so $B_i^*(H_i)$ is behavioral convex.

Now suppose $B_i^*(H_i)$ is behavioral convex. By the claim above for any $\lambda \in [0,1]$ the convex combination $\lambda \sigma_i + (1-\lambda)\sigma'_i \sim B(b_i, b'_i, \lambda)$. Since $B_i^*(H_i)$ is behavioral convex $B(b_i, b'_i, \lambda) \in B_i^*(H_i)$ which implies that $\lambda \sigma_i + (1-\lambda)\sigma'_i \in \Sigma_i^*$, so $\Sigma_i^*$ is convex. 

Lemma 2: $U_i(B(b_i, b'_i, \lambda), b_{-i}) = \lambda U_i(b_i, b_{-i}) + (1-\lambda)U_i(b'_i, b_{-i})$

Proof. By the claim in Proposition 4 if $b_i, b'_i \in B_i$ and $\sigma_i, \sigma'_i \in \Sigma_i$ such that $b_i \sim \sigma_i$ and $b'_i \sim \sigma'_i$, then $B_i(b_i, b'_i, \lambda) \sim \lambda \sigma_i + (1-\lambda)\sigma'_i$. Clearly if $b \sim \sigma$ then $U_i(b) = U_i(\sigma)$.

Let $b \sim \sigma$ and $b' = (b'_i, b_{-i}) \sim \sigma' = (\sigma'_i, \sigma_{-i})$, and $\sigma'' = \lambda \sigma + (1-\lambda)\sigma'$. Since $U_i$ is linear in mixed strategies, and since $B_i(b_i, b'_i, \lambda) \sim \sigma''$, it follows that.

$$U_i(B_i(b_i, b'_i, \lambda), b_{-i}) = U_i(\sigma'', \sigma_{-i}) = \lambda U_i(\sigma_i, \sigma_{-i}) + (1-\lambda)U_i(\sigma'_i, \sigma_{-i}) = \lambda U_i(b_i, b_{-i}) + (1-\lambda)U_i(b'_i, b_{-i})$$

Theorem 1. If $G$ is irreducible then $G$ satisfies perfect recall if and only if the set of behavioral strategies is behavioral convex.

Proof. Suppose by way of contradiction that player $i$ does not have perfect recall.

Step 1: Absentmindedness

Suppose that player $i$ is absentminded. So there exists $w, w' \in T_i$ such that $w \in P_i(w')$ and $w' \in h(w)$. By irreducibility there exists $s_i \in S_i^*(H_i)$ such that $s_i \in R_i(w)$ and $s_i(w) \neq a(w, w')$. Let $s'_i \in R_i(w')$, $b_i \sim s_i$ and $b'_i \sim s'_i$. By construction $b_i, b'_i \in B_i^*(H_i)$, and $b_i(a(w, w')|w) = 0$ and $b'_i(a(w, w')|w) = 1$. Let $b''_i = B(b_i, b'_i, \frac{1}{2})$. Notice that $b''_i(a(w, w')|w) = \frac{1}{2}$ and that $b''_i(a(w, w')|w') = 1$ so $b''_i \notin B_i^*(H_i)$.

Step 2: Imperfect Recall of Past Information

Without loss of generality, assume that player $i$ is not absentminded. Let $w' \in h(w)$ and suppose there is a $\hat{w} \in P_i(w)$ such that there is not a $\hat{w}' \in P_i(w')$ such that $\hat{w}' \in h(\hat{w})$. Construct the behavioral strategies $b_i$ and $b'_i$ as follows.

For $x \in P_i(w)$ such that $\exists x' \in P_i(w')$ satisfying $x \in h(x')$, let $b'(a|x) = b(a|x) = 1/|A(x)| \ \forall a \in A(x)$. For $x \in P_i(w)$ such that $\nexists x' \in P_i(w')$ satisfying $x \in h(x')$, let $b(a(x, w)|x) = 1$ and $b'(a(x, w)|x) = 1$ if $x \neq \hat{w}$ and $b'(a(\hat{w}, w)|\hat{w}) = 0$. For $x' \in P_i(w')$ such that $\nexists x \in P_i(w)$ satisfying $x \in h(x')$, let $b'(a(x', w')|x') = b(a(x', w')|x') = 1$. Notice that this completely specifies $b_i$ and $b'_i$ at the predecessors of $w$ and $w'$.
By construction \( b_i, b'_i \in B_i^*(H_i) \). Since player \( i \) is not absentminded there is a bijection between the predecessors of \( w \) and \( w' \) in the same information set, hence 
\[
P^i(w|b) = P^i(w'|b) = P^i(w'|b') > 0,
\]
and by construction \( P^i(w|b') = 0 \). Since \( G \) is irreducible there are unique \( a_1, a_2 \in A(w) = A(w') \). Let \( b(a_1|w) = b(a_1|w') = 1 \), and let \( b'(a_2|w) = b'(a_2|w') = 1 \). Consider \( b''_i = B(b, b', \frac{1}{2}) \). Since \( P^i(w|b') = 0 \) and \( P^i(w'|b) > 0 \) it follows that \( b''(a_1|w) = b(a_1|w) = 1 \). Moreover, since \( P^i(w'|b) = P^i(w'|b') > 0 \) it follows that \( b''(a_1|w') = \frac{1}{2}b(a_1|w') + \frac{1}{2}b'(a_1|w) = \frac{1}{2} \). Since \( w' \in h(w) \) this implies that \( b''_i \notin B_i^*(H_i) \).

**Step 3: Imperfect Recall of Past Actions**

Now without loss of generality assume that player \( i \) is not absentminded, and does not have imperfect recall of past information. Since player \( i \) has imperfect recall, it follows that there is a \( w' \in h(w) \) such that there is a \( \hat{w} \in P_i(w) \) and \( \hat{w}' \in P_i(\hat{w}) \), such that \( \hat{w} \in h(\hat{w}') \), and \( a(\hat{w}, w) \neq a(\hat{w}', w') \).

By irreducibility there are unique actions \( a_1, a_2 \in A(w) \). Since \( a(\hat{w}, w) \neq a(\hat{w}', w') \) there exists \( s_i \in R_i(w) \) such that \( s_i(w) = s_i(w') = a_1 \), and \( s_i \notin R_i(w') \). Similarly there exists \( s'_i \in R_i(w') \) such that \( s'_i(w') = s_i(w) = a_2 \), and \( s'_i \notin R_i(w) \). Let \( b_i \sim s_i \) and \( b'_i \sim s'_i \). By construction \( P^i(w|b_i) = P^i(w|b_i') = 1 \). Consider \( b''_i = B_i(b_i, b'_i, \frac{1}{2}) \). Notice that 
\[
\text{So if player } i \text{ has imperfect recall, then } B_i^*(H_i) \text{ is not behavioral convex. By the contraposition, it follows that if } B_i^*(H_i) \text{ is behavioral convex then } G \text{ satisfies perfect recall. I now consider the converse.}
\]

Suppose \( G \) satisfies perfect recall. By Kuhn’s Theorem \( \Sigma^*_i(H_i) \) is equivalent to \( B_i^*(H_i) \). By construction \( B_i^*(H_i) \) is equivalent to \( \Sigma^*_i \) hence \( \Sigma^*_i(H_i) \) is equivalent to \( \Sigma^*_i \). By construction \( \Sigma^*_i(H_i) \) is convex. Suppose \( \sigma_i, \sigma'_i \in \Sigma^*_i(H_i) \) and \( \varsigma_i, \varsigma'_i \in \Sigma^*_i \) such that \( \sigma_i \sim \varsigma_i \) and \( \sigma'_i \sim \varsigma'_i \). By construction if \( P^i(w|\sigma_i) > 0 \) or \( P^i(w|\varsigma_i) > 0 \) then \( P^i(w|\sigma_i) = P^i(w|\varsigma_i) \) and 
\[
\frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)} = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \varsigma_i(s_i)}{\sum_{s_i \in R_i(w)} \varsigma_i(s_i)}
\]

And similarly for \( \sigma_i \) and \( \varsigma_i \). Consider \( \sigma''_i = \lambda \sigma_i + (1 - \lambda) \varsigma_i \) and \( \varsigma''_i = \lambda \varsigma_i + (1 - \lambda) \varsigma_i \). Notice that if \( P^i(w|\sigma''_i) > 0 \) or \( P^i(w|\varsigma''_i) > 0 \) then \( P^i(w|\sigma''_i) = P^i(w|\varsigma''_i) \) and 
\[
\frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} (\lambda \sigma_i(s_i) + (1 - \lambda) \sigma'_i(s_i))}{\sum_{s_i \in R_i(w)} (\lambda \sigma_i(s_i) + (1 - \lambda) \sigma'_i(s_i))} = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} (\lambda \varsigma_i(s_i) + (1 - \lambda) \varsigma'_i(s_i))}{\sum_{s_i \in R_i(w)} (\lambda \varsigma_i(s_i) + (1 - \lambda) \varsigma'_i(s_i))}
\]

This implies that \( \sigma''_i \sim \varsigma''_i \). Notice that \( \sigma''_i \in \Sigma^*_i(H_i) \) since \( \Sigma^*_i(H_i) \) is convex. Since \( \Sigma^*_i(H_i) \)
is equivalent to $\Sigma^*_i$ it follows that $\zeta_i'' \in \Sigma^*_i$, hence $\Sigma^*_i$ is convex. By proposition 4, this implies that $B_i^*(H_i)$ is behavioral convex. 

\[ \]

9.4 Properties of Correspondences

Definition 18. A correspondence $C : Y \rightarrow 2^X$ is compact/closed valued if $C(y)$ is compact/closed for every $y \in Y$.

Definition 19. A correspondence $C$ is closed if $x_n \rightarrow x$, $y_n \rightarrow y$, and $x_n \in C(y_n)$ then $y \in C(x)$.

Fact 1. If $C$ is closed valued, and $Y$ is compact then $C$ is upper hemicontinuous if and only if $C$ is closed.

Proof. See Border (1985) Proposition 11.9 (a)(b)

The image of $K \subset Y$ under $C$ is given by

$$C(K) = \bigcup_{k \in K} C(k)$$

Fact 2. If $C$ is upper hemicontinuous and compact valued and $K \subset X$ is compact then $C(K)$ is compact.

Proof. See Border (1985) Proposition 11.16

Definition 20. A correspondence $C$ is lower hemicontinuous if for any $y_n \rightarrow y$ and $x \in C(y)$, there exists $\{y_{nk}\}$ a subsequence of $\{y_n\}$, and $x_{nk} \in C(y_{nk})$ such that $x_{nk} \rightarrow x$.

The composition of two correspondences $C_1 : A \rightarrow 2^B$ and $C_2 : B \rightarrow 2^C$ written as $C_1 \circ C_2 : A \rightarrow 2^C$ is given by

$$(C_2 \circ C_1)(a) = \bigcup_{b \in C_1(a)} C_2(b)$$

Fact 3. Let $C_1 : A \rightarrow 2^B$, and $C_2 : B \rightarrow 2^C$. If $C_1$ and $C_2$ are upper hemicontinuous (resp lower hemicontinuous) then $C_2 \circ C_1$ is upper hemicontinuous (resp lower hemicontinuous).

Proof. See Border (1985) Proposition 11.23 (a)(b)

Fact 4. Let $C_1 : Y \rightarrow 2^X$, and $C_2 : Y \rightarrow 2^X$ be closed valued. If $C_1$ and $C_2$ are upper hemicontinuous at $x$ and $C_1(x) \cap C_2(x)$ is nonempty then $C_1(x) \cap C_2(x)$ is upper hemicontinuous at $x$. 

33
For a collection of correspondences \( C_i : Y \to 2^{X_i} \) indexed by \( i = 1, ..., k \) their product correspondence \( C : Y \to 2^X \), where \( X = \prod_{i=1}^k X_i \) is given by

\[
\prod_{i=1}^k C_i : y \to \prod_{i=1}^k C_i(y)
\]

**Fact 5.** If each \( C_i \) is upper hemicontinuous and compact valued then their product correspondence is upper hemicontinuous and compact valued. Moreover, if each \( C_i \) is lower hemicontinuous then their product correspondence is lower hemicontinuous.

**Proof.** See Border (1985) Proposition 11.25

### 9.5 Equivalence Correspondences

For \( b_i \in B_i \), let \( \Psi_i(b_i) = \{ \sigma_i \in \Sigma_i : \sigma_i \sim b_i \} \). By equation (1) this can be expressed as

\[
\Psi_i(b_i) = \{ \sigma_i \in \Sigma_i : \sum_{s_i \in R(w)} \sigma_i(s_i) > 0 \implies b_i(a_i|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)} \}
\]

For \( \sigma_i \in \Sigma_i \), let \( \Phi_i(\sigma_i) = \{ b_i \in B_i : b_i \sim \sigma_i \} \). By equation (1) this can be expressed as

\[
\Phi_i(\sigma_i) = \{ b_i \in B_i : \sum_{s_i \in R(w)} \sigma_i(s_i) > 0 \implies b_i(a_i|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)} \}
\]

Let \( \Psi = \prod_{i \in I} \Psi_i \) and \( \Phi = \prod_{i \in I} \Phi_i \).

**Remark 1.** \( \Psi_i \) and \( \Psi \) are continuous, non-empty valued, and compact valued.

**Proof. Upper Hemicontinuity:** Let \( b_i^n \to b_i, \sigma_i^n \to \sigma_i \), with \( \sigma_i^n \in \Psi_i(b_i^n) \). By construction this implies that

\[
P^i(w|\sigma_i^n) > 0 \implies b_i^n(a_i|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i^n(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i^n(s_i)}
\]

So if \( P^i(w|\sigma_i) > 0 \) there exists an integer \( N \) such that if \( n > N \) then \( P^i(w|\sigma_i^n) > 0 \). So there is an open neighborhood around \( \sigma_i \) such that \( b_i^n(a_i|w) \) is a continuous function of \( \sigma_i^n \). Hence

\[
b_i(a_i|w) = \lim_{n \to \infty} b_i^n(a_i|w) = \lim_{n \to \infty} \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i^n(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i^n(s_i)} = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)}
\]

34
Let $\tilde{\sigma}$ by construction. Therefore, $$\tilde{\sigma} = \tilde{b}_i.$$ I proceed by induction on the number of nodes of $T_i$. 

**Claim.** If $\sigma_i \sim b_i$ and $|b_i - \tilde{b}_i| < \epsilon$, then there exists a $\tilde{\sigma}_i \sim \tilde{b}_i$ and $\delta(\epsilon)$ such that $|\sigma_i - \tilde{\sigma}_i| < \delta(\epsilon)$. Moreover $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Let $|b_i - \tilde{b}_i| < \epsilon$, and let $w_1, \ldots, w_N$ be an ordering of the elements of $T_i$ such that if $i > j$ then $w_i$ is not a predecessor of $w_j$. Let $T_i' = \{w_1, \ldots, w_n\}$ be the restriction of $T_i$ to the first $n$ nodes. I proceed by induction on the number of nodes of $T_i$.

For the base case assume $n = 1$. Set $\tilde{\sigma}_i(a_i) = \tilde{b}_i(a_i|w_i)$. By construction $\tilde{\sigma}_i \sim \tilde{b}_i$, and notice that $|\sigma_i - \tilde{\sigma}_i| = |b_i - \tilde{b}_i| < \epsilon$. So let $\delta(\epsilon) = \delta_1(\epsilon) = \epsilon$ proving the base case.

Assume the result holds for $n < N$, and consider mixed strategies defined over $T_i^{N-1}$. By the inductive hypothesis there exists a $\tilde{\sigma}_i$ and $\delta_{N-1}(\epsilon)$, such that $|\sigma_i - \tilde{\sigma}_i| < \delta_{N-1}(\epsilon)$ and $\delta_{N-1}(\epsilon) \to 0$, as $\epsilon \to 0$. Let $(s_i, a)$ be the pure strategy defined over $T_i^N$ where player $i$ follows $s_i$ at the first $N - 1$ nodes and chooses $a$ at $w_N$, so that by construction $\sigma_i(s_i) = \sum_{a_i \in A(w_N)} \sigma_i(s_i, a_i)$. Construct a candidate mixed strategy $\sigma'_i \sim b'_i$ as follows.

$$\sigma'_i(s_i, a_i) = \begin{cases} \sigma_i(s_i, a_i) \frac{\tilde{\sigma}_i(s_i)}{\sigma_i(s_i)} & \text{if } \sigma_i(s_i) > 0 \\ \sigma_i(s_i) \tilde{b}_i(a_i|w_N) & \text{if } \sigma_i(s_i) = 0 \end{cases}$$

In both cases $\sigma'_i(s_i) = \sum_{a_i \in A(w_N)} \sigma'_i(s_i, a_i) = \tilde{\sigma}_i(s_i)$ so $b'_i(a_i|w_n) = \tilde{b}_i(a_i|w_n)$ for $n < N$.

**Case 1:** Suppose that $P^i(w_N|\sigma_i) = 0$ and $P^i(w_N|\tilde{\sigma}_i) = 0$.

Let $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i)$. Since $P^i(w_N|\tilde{\sigma}_i) = 0$ it follows that $\tilde{\sigma}_i \sim \tilde{b}_i$. If $s_i \in R_i(w_N)$ then $\tilde{\sigma}_i(s_i, a_i) = \sigma_i(s_i, a_i) = 0$, so $|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = 0$.

If $s_i \notin R_i(w_N)$ and $\sigma_i(s_i) > 0$, then $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i) = \sigma_i(s_i, a_i) \frac{\tilde{\sigma}_i(s_i)}{\sigma_i(s_i)}$ by construction. Therefore, $$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = \sigma_i(s_i, a_i)|1 - \frac{\tilde{\sigma}_i(s_i)}{\sigma_i(s_i)}| \leq \max_{s_i \in S_i, \delta \in [-\delta_{N-1}(\epsilon), \delta_{N-1}(\epsilon)]} \left| 1 - \frac{\sigma_i(s_i) + \theta}{\sigma_i(s_i)} \right| = \delta_{N-1}(\epsilon).$$

Notice that as $\epsilon \to 0$, $\delta_{N-1}(\epsilon) \to 0$.

If $s_i \notin R_i(w_N)$ and $\sigma_i(s_i) = 0$, then $\sigma_i(s_i, a_i) = 0$ and $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i) = \tilde{\sigma}_i(s_i) \tilde{b}_i(a_i|w_N)$ by construction. Therefore, $$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = \tilde{\sigma}_i(s_i, a_i) \leq \tilde{\sigma}_i(s_i) = |\tilde{\sigma}_i(s_i) - \sigma_i(s_i)| < \delta_{N-1}(\epsilon).$$
Case 2: Suppose that $P^i(w_N|\sigma_i) = 0$ and $P^i(w_N|\tilde{\sigma}_i) > 0$
Let $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i)$. If $s_i \in R_i(w_N)$ then $\sigma_i(s_i) = 0$, so $\tilde{\sigma}_i(s_i, a_i) = \tilde{\sigma}_i(s_i)\tilde{b}_i(a_i|w_N)$. By construction

$$\frac{\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i, a_i)}{\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i)} = \frac{\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i)\tilde{b}_i(a_i|w_N)}{\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i)} = \tilde{b}_i(a_i|w_N)$$

So $\tilde{\sigma}_i \sim \tilde{b}_i$. Since $\sigma(s_i) = 0$ it follows that

$$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = |\tilde{\sigma}_i(s_i) - \sigma_i(s_i)| < \delta_{N-1}(\epsilon)$$

If $s_i \notin R_i(w_N)$ then the same argument from Case 1 applies.

Case 3: Suppose that $P^i(w_N|\sigma_i) > 0$ and $P^i(w_N|\tilde{\sigma}_i) = 0$
Let $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i)$. Since $P^i(w_N|\tilde{\sigma}_i) = 0$ it follows that $\tilde{\sigma}_i \sim \tilde{b}_i$. If $s_i \in R_i(w_N)$ then $\tilde{\sigma}_i(s_i) = \sum_{a_i \in A(w_N)} \tilde{\sigma}_i(s_i, a_i) = 0$, so $\sigma_i(s_i) = |\tilde{\sigma}_i(s_i) - \sigma_i(s_i)| < \delta_{N-1}(\epsilon)$. Since $\sigma_i(s_i) = \sum_{a_i \in A(w_N)} \sigma_i(s_i, a_i)$, it follows that

$$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = |\sigma_i(s_i, a_i)| < \delta_{N-1}(\epsilon)$$

For $s_i \notin R_i(w_N)$ the same argument from Case 1 applies.

Case 4: Suppose that $P^i(w_N|\sigma_i) > 0$ and $P^i(w_N|\tilde{\sigma}_i) > 0$
For $s_i \notin R_i(w_N)$ let $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i)$, and apply the argument in Case 1. For $s_i \in R_i(w_N)$, look for a solution of the form $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i) + \gamma(s_i, a_i)$, satisfying the following properties.

(C1) $\sum_{a_i \in A(w_N)} \gamma(s_i, a_i) = 0$
(C2) $\sum_{s_i \in R_i(w_N)} \gamma(s_i, a_i) = (\tilde{b}_i(a_i|w_N) - \tilde{b}'_i(a_i|w_N)) \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)$
(C3) $\gamma(s_i, a_i) > 0 \implies \gamma(s'_i, a_i) > 0$
(C4) $\gamma(s_i, a_i) < 0 \implies \gamma(s'_i, a_i) < 0$
If such a $\gamma$ exists, (C1) implies that $\tilde{\sigma}_i(s_i) = \sigma'_i(s_i)$, and (C2) implies that

$$\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i, a_i) = \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) + \gamma(s_i, a_i)$$

$$= \frac{\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i)}{\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)} + \frac{(\tilde{b}(a_i|w_N) - b'(a_i|w_N)) \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)}{\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)}$$

$$= b'_i(a_i|w_N) + \tilde{b}_i(a_i|w_N) - b'_i(a_i|w_N) = \tilde{b}_i(a_i|w_N)$$

So $\tilde{\sigma}_i \sim \tilde{b}_i$. Let $A^- = \{a_i \in A(w_N) : (\tilde{b}_i(a_i|w_N) - \tilde{b}_i(a_i|w_N)) < 0\}$, and let $A^+ = \{a_i \in A(w_N) : (\tilde{b}_i(a_i|w_N) - \tilde{b}_i(a_i|w_N)) \geq 0\}$. For $a_i \in A^-$ the following holds.

$$\sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i, a_i) = \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) + \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)(\tilde{b}_i(a_i|w_N) - \tilde{b}'_i(a_i|w_N))$$

$$\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) - \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)\tilde{b}_i(a_i|w_N) = \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) - \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) = 0$$

Therefore for $a_i \in A^-$ there exists a $\gamma(s_i, a_i)$ satisfying (C1), (C2), and (C4) such that $\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i) + \gamma(s_i, a_i) \geq 0$. Fix such a solution for $a_i \in A^-$. Notice that if (C3) and (C2) hold it follows that.

$$\tilde{\sigma}_i(s_i, a_i) = \sigma'_i(s_i, a_i) + \gamma(s_i, a_i) \leq \sigma'_i(s_i, a_i) + \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)(\tilde{b}_i(a_i|w_N) - \tilde{b}'_i(a_i|w_N))$$

$$\sigma'_i(s_i, a_i) + \sum_{s_i \in R_i(w_N)} \tilde{\sigma}_i(s_i, a_i) - \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) \leq \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i, a_i) \leq 1$$

Define the following coefficients for each $a_i \in A^+$

$$\phi_{a_i} = \frac{\tilde{b}_i(a_i|w_N) - \tilde{b}'_i(a_i|w_N)}{\sum_{a_i \in A^+} (\tilde{b}_i(a_i|w_N) - \tilde{b}'_i(a_i|w_N))}$$

For $a_i \in A^+$ let $\gamma(s_i, a_i) = \phi_{a_i}(-\sum_{a_i \in A^-} \gamma(s_i, a_i)) \geq 0$, so (C3) is satisfied. By construction (C1) is satisfied since.

$$\sum_{a_i \in A^+} \gamma(s_i, a_i) = -\sum_{a_i \in A^-} \gamma(s_i, a_i) \Longrightarrow \sum_{a_i \in A(w_N)} \gamma(s_i, a_i) = 0$$
For $a_i \in A^+$ observe that (C2) is satisfied since

$$\sum_{s_i \in R_i(w_N)} \gamma(s_i, a_i) = \phi_{a_i}(\sum_{s_i \in R_i(w_N)} - \gamma(s_i, a_i)) = \phi_{a_i}(\sum_{s_i \in R_i(w_N)} - (b_i(a_i|w_N) - b_i(a_i|w_N))) = \phi_{a_i}(\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)) = \frac{\sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)}{\sum_{s_i \in R_i(w_N)} \sigma_i(s_i)} \leq \frac{\sum_{s_i \in R_i(w_N), \sigma(s_i) > 0} \sigma_i(s_i)}{\sum_{s_i \in R_i(w_N)} \sigma_i(s_i)} \leq \frac{\max_{\theta_i \in [-\delta_{N-1}(\epsilon), \delta_{N-1}(\epsilon)]} \sum_{s_i \in R_i(w_N), \sigma(s_i) > 0} \sigma_i(s_i)}{\sum_{s_i \in R_i(w_N)} \sigma_i(s_i)} \leq \max_{\theta_i \in [-\delta_{N-1}(\epsilon), \delta_{N-1}(\epsilon)]} \sum_{s_i \in R_i(w_N), \sigma(s_i) > 0} \sigma_i(s_i) = \delta^m_{N-1}(\epsilon)$$

Notice that $\delta^m_{N-1}(\epsilon) \to 0$ as $\epsilon \to 0$. So $|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)|$ can be bounded as follows

$$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| = |\sigma'_i(s_i, a_i) + \gamma(s_i, a_i) - \sigma_i(s_i, a_i)| \leq |\sigma'_i(s_i, a_i) - \sigma_i(s_i, a_i)| + |\gamma(s_i, a_i)| \leq |\sigma'_i(s_i, a_i) - \sigma_i(s_i, a_i)| + |\tilde{b}_i(a_i|w_N) - b'_i(a_i|w_N)| \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i) \leq |\sigma'_i(s_i, a_i) - \sigma_i(s_i, a_i)| + (\delta^m_{N-1}(\epsilon) + \epsilon) \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i)$$

By the previous cases if $\sigma_i(s_i) > 0$ then $|\sigma'_i(s_i, a_i) - \sigma_i(s_i, a_i)| \leq \delta^m_{N-1}(\epsilon)$, and if $\sigma_i(s_i) = 0$ then $|\sigma'_i(s_i, a_i) - \sigma_i(s_i, a_i)| \leq \delta_{N-1}(\epsilon)$. This implies that

$$|\tilde{\sigma}_i(s_i, a_i) - \sigma_i(s_i, a_i)| \leq \max\{\delta_{N-1}(\epsilon), \delta^m_{N-1}(\epsilon)\} + (\delta^m_{N-1}(\epsilon) + \epsilon) \sum_{s_i \in R_i(w_N)} \sigma'_i(s_i) = \delta^m_{N-1}(\epsilon)$$

Notice that $\delta^m_{N-1}(\epsilon) \to 0$ as $\epsilon \to 0$. Let $\delta_N(\epsilon) = \max\{\delta_{N-1}(\epsilon), \delta'_{N-1}(\epsilon), \delta^m_{N-1}(\epsilon), \delta^m_{N-1}(\epsilon)\}$. By construction $\delta_N(\epsilon) \to 0$ as $\epsilon \to 0$, which confirms the inductive hypothesis, proving the
claim.
Let \( \sigma_i \in \Psi_i(b_i) \), and \( b_i^n \rightarrow b_i \). By the claim if \( |b_i^n - b_i| \leq \epsilon \), then there exists a \( \sigma^n_i \in \Psi_i(b_i^n) \) such that \( |\sigma_i^n - \sigma_i| \leq \delta(\epsilon) \). Moreover \( \delta(\epsilon) \rightarrow 0 \) as \( \epsilon \rightarrow 0 \), which implies that \( \sigma_i^n \rightarrow \sigma_i \), so \( \Psi_i \) is indeed lower hemicontinuous.

Proof. Compact Valued: Let \( \sigma_i^n \in \Psi_i(b_i) \) for all \( n \) and let \( \sigma_i^n \rightarrow \sigma_i \). Since \( \sigma_i^n \rightarrow \sigma_i \), there exists an integer \( N \) such that for \( n \geq N \) then

\[
\sum_{s_i \in R_i(w)} \sigma_i^n(s_i) > 0 \iff \sum_{s_i \in R_i(w)} \sigma_i(s_i) > 0
\]

If \( \sum_{s_i \in R_i(w)} \sigma_i(s_i) > 0 \) then there exists an open neighborhood around \( \sigma_i \) such that \( b_i^n(a|w) \) is a continuous function of \( \sigma_i^n \). Since \( \sigma_i^n \in \Psi_i(b_i) \) this implies that

\[
b_i(a_i|w) = \lim_{n \rightarrow \infty} \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i^n(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i^n(s_i)} = \lim_{n \rightarrow \infty} b_i^n(a|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)}
\]

So \( \sigma_i \in \Psi_i(b_i) \). So \( \Psi_i \) is closed valued and hence compact valued since \( B_i \) is compact.

Proof. Extending to \( \Psi \): By definition \( \Psi = \prod_{i=1}^{N} \Psi_i \). Since \( \Psi_i \) is nonempty valued for all \( i \), \( \Psi \) is nonempty valued. Since \( \Psi_i \) is compact valued for all \( i \), \( \Psi \) is compact valued in the product topology. Moreover, since \( \Psi_i \) is continuous for all \( i \), \( \Psi \) is continuous in the product topology by Fact 5.

Remark 2. \( \Phi_i \) and \( \Phi \) are upper hemicontinuous, nonempty and compact valued.

Proof. Upper Hemicontinuity: Let \( \sigma_i^n \rightarrow \sigma_i \), \( b_i^n \in \Phi_i(\sigma_i^n) \), and \( b_i^n \rightarrow b_i \). If \( \sum_{s_i \in R_i(w)} \sigma_i(s_i) > 0 \) then for \( n \) sufficiently large \( \sum_{s_i \in R_i(w)} \sigma_i^n(s_i) > 0 \). So there exists an open neighborhood around \( \sigma_i \) such that \( b_i^n(a|w) \) is a continuous function of \( \sigma_i^n \). This implies that

\[
b_i(a_i|w) = \lim_{n \rightarrow \infty} b_i^n(a_i|w) = \lim_{n \rightarrow \infty} \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i^n(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i^n(s_i)} = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)}
\]

So \( b_i \in \Phi_i(\sigma_i) \), which implies that \( \Phi_i \) is upper hemicontinuous by Fact

Proof. Compact Valued: Let \( b_i^n \in \Phi_i(\sigma_i) \) and \( b_i^n \rightarrow b \). If \( \sum_{s_i \in R_i(w)} \sigma_i(s_i) > 0 \) then

\[
b_i(a|w) = \lim_{n \rightarrow \infty} b_i^n(a|w) = \frac{\sum_{s_i \in R_i(w), s_i(w) = a_i} \sigma_i(s_i)}{\sum_{s_i \in R_i(w)} \sigma_i(s_i)}
\]

So \( b_i \in \Phi_i(\sigma_i) \), which implies \( \Phi_i \) is closed valued, so \( \Phi_i \) is compact valued since \( B_i \) is compact.
Proof. Properties of $\Phi$: $\Phi = \prod_{i=1}^{N} \Phi_i$. Since $\Phi_i$ is nonempty for all $i$ then $\Phi$ is nonempty. Since $\Phi_i$ is compact valued for all $i$ then $\Phi$ is compact valued in the product topology. Moreover, since $\Phi_i$ is upper hemicontinuous for all $i$ $\Phi$ is continuous in the product topology by Fact 5. \square

9.6 Modified Maximum Theorem

The original maximum theorem of Berge (1963) requires the feasibility correspondence to be lower hemicontinuous. The following lemma weakens this condition slightly, in order to prove the existence of an equilibrium for EFAE with feasibility correspondences with that are not lower hemicontinuous. Intuitively, Lemma 1 says that for the maximum theorem to hold the constraint correspondence need not be lower hemicontinuous, but any violations of lower hemicontinuity must be irrelevant to players’ payoffs.

Lemma 1. Let $f : X \times Y \to \mathbb{R}$, be jointly continuous and $C : Y \to 2^X$ be a compact valued correspondence. Let $f^*(y) = \max_{x \in C(y)} f(x, y)$ and $C^*(y) = \arg \max_{x \in C(y)} f(x, y)$. If $C$ is $f$-continuous then $f^*$ is continuous, and $C^*$ is non-empty, compact valued, and upper hemicontinuous.

Proof. Since $C$ is compact valued and $f$ is continuous, $f^*$ is well defined, and $C^*(y)$ is non-empty for any $y \in Y$. Let $y_n \to y$, and $x_n \in C^*(y_n)$. Since $C$ is upper hemicontinuous there exists $\{x_{n_k}\}$ a subsequence of $\{x_n\}$ such that $x_{n_k} \to x \in C(y)$.

Suppose by way of contradiction that $x \notin C^*(y)$. So there exists $\hat{x} \in C(y)$ such that $f(\hat{x}, y) > f(x, y)$. Since $C$ is $f$-lower hemicontinuous there exists $\hat{x}_{n_j} \in C(y_{n_j})$ where $\{x_{n_j}\}$ is a subsequence of $\{x_{n_k}\}$ such that $f(\hat{x}_{n_j}, y_{n_j}) \to f(\hat{x}, y)$. Since $f$ is continuous, for $n_j$ sufficiently large $f(\hat{x}_{n_j}, y_{n_j}) > f(x_{n_j}, y_{n_j})$, a contradiction. So $x \in C^*(y)$, so $C^*$ is upper hemicontinuous. Since $f$ is continuous it follows that,

$$\lim_{k \to \infty} f^*(y_{n_k}) = \lim_{k \to \infty} f(x_{n_k}, y_{n_k}) = f(x, y) = f^*(y)$$

So $f^*$ is continuous and $C^*$. $C^*$ closed valued by the continuity of $f$. Moreover, since $C^*(\cdot) \subset C(\cdot)$ and $C$ is compact valued it follows that $C^*$ is compact valued. \square

9.7 Partitional Convexity

Recall that $Z \subset \mathbb{R}^m$ and $P$ is a partition of $I = \{1, \ldots, m\}$ with elements $P_1, \ldots, P_K$, and $Z^k = \prod_{i \in P_k} Z_i$. Lemma 3 demonstrates that $P$-partitionally convex sets must take a specific cartesian product form. Intuitively, if $P$ is a partition $T_i$ then a $P$-partitionally convex
feasibility correspondence can be thought of as \( K \) separate feasibility correspondences, one governing each element of \( P \).

**Lemma 3.** A set \( Z \) is \( P \)-partitionally convex if and only if \( Z = \prod_{k=1}^{K} Z^k \) where each \( Z^k \subset \mathbb{R}^{|P_k|} \) is convex.

**Proof.** If \( Z = \prod_{k=1}^{K} Z^k \) each \( Z^k \) is convex, then the fact that \( Z \) is \( P \)-partitionally convex is immediate.

To prove the converse proceed by induction on \( K \). If \( K = 1 \), notice that any \( P \)-partitional convex combination is also convex combinations. Since \( Z = Z^1 \) is convex, it follows that \( Z \) is \( P \)-partitionally convex, which proves the base case.

Suppose the Lemma holds for \( K \leq N \). Let \( Z \) be \( P \)-partitionally convex with \( K = N+1 \). By construction, any \( z \in Z \) can be written as \( z = (z_1, \ldots, z_{N+1}) \) where \( z_k \in Z^k \). Define \( Z_{N+1}(a) = \{ z \in Z : z_{N+1} = a \} \), and let

\[
A = \{ a \in Z^{N+1} : \exists z \in Z \text{ s.t. } z = (z_1, \ldots, z_N, a) \}
\]

Since \( Z \) is \( P \)-partitionally convex so is \( Z_{N+1}(a) \). Let \( P' = P \setminus P_{N+1} \), and notice that \( Z_{N+1}(a) \) can be written as,

\[
Z_{N+1}(a) = Z'(a) \times \{ a \}
\]

Where \( Z'(a) \) is \( P' \)-partitionally convex. By the inductive hypothesis \( Z_{N+1}(a) = \prod_{k=1}^{N} Z^k(a) \times \{ a \} \). Let \( (z_1, \ldots, z_N, a) \in Z_{N+1}(a) \) and \( (z'_1, \ldots, z'_N, a') \in Z_{N+1}(a') \). Since \( Z \) is \( P \)-partitionally convex \( (z'_1, \ldots, z'_N, a) \in S \) hence \( (z'_1, \ldots, z'_N, a) \in S_{N+1}(a) \). So for any \( a, a' \in A \)

\[
\prod_{k=1}^{N} Z^k(a) = \prod_{k=1}^{N} Z^k(a') = \prod_{k=1}^{N} Z^k
\]

Defining \( Z^{N+1} = A \), notice that

\[
Z = \bigcup_{a \in A} Z_{N+1}(a) = \bigcup_{a \in A} (\prod_{k=1}^{N} Z^k \times \{ a \}) = \prod_{k=1}^{N} Z^k \times A = \prod_{k=1}^{N+1} Z^k
\]

Moreover, if \( (z_1, \ldots, z_N, a), (z_1, \ldots, z_N, a') \in Z \) then \( (z_1, \ldots, z_N, \lambda a + (1 - \lambda) a') \in Z \) since \( Z \) is \( P \)-partitional convex. Therefore \( Z^K = A \) is convex, confirming the inductive hypothesis \( \square \)
9.8 Normal Form Abstract Economies

A normal form abstract economy is a tuple $G = \langle I, S, u, C \rangle$. The set of players is $I = \{1, 2, ..., m\}$. The set of pure strategies is $S = \prod_{i=1}^{m} S_i$. The set of mixed strategies is $\Sigma = \Delta(S)$. The utility index is given by $u : S \rightarrow \mathbb{R}^n$. The expected utility of player $i$ is given by $U_i(\sigma) = \sum_{s \in S} \sigma(s) u_i(s)$. The collection of feasibility correspondences is $C = \{C_1, ..., C_N\}$, where $C_i : \Sigma_{-i} \rightarrow 2^{\Sigma_i}$.

**Definition 21.** An equilibrium in mixed strategies is a $\sigma \in \Sigma$ satisfying for every $i \in I$

$$\sigma_i \in \arg \max_{\tilde{\sigma}_i \in C_i(\sigma_{-i})} U_i(\tilde{\sigma}_i, \sigma_{-i})$$

(N1) $C_i$ is non-empty and compact valued

(N2) $C_i$ is $U_i$-continuous

(N3) $C_i$ is convex valued

**Lemma 3:** If $G$ is finite and satisfies (N1) – (N3) then $G$ has a constrained equilibrium in mixed strategies.

**Proof.** $\Sigma$ is convex and non-empty by construction, and compact since $G$ is finite. By (N1) $C_i$ is compact valued. Since $U_i$ is continuous,

$$BR_i(\sigma_{-i}) = \arg \max_{\sigma_i \in C_i(\sigma_{-i})} U_i(\sigma_i, \sigma_{-i})$$

is nonempty. So $BR = \prod_{i=1}^{m} BR_i$ is nonempty valued. Since $U_i$ is linear and $C_i$ is convex valued $BR_i$ is convex valued, which implies that $BR = \prod_{i=1}^{m} BR_i$ is convex valued.

By construction $U_i$ is jointly continuous. By (N1) and (N2), $C_i$ is upper hemicontinuous, $U_i$-lower hemicontinuous, non-empty and compact valued. By Lemma 1, $BR_i$ is non-empty, compact valued, and upper hemicontinuous. So $BR = \prod_{i=1}^{m} BR_i$ is upper hemicontinuous by Fact 5. So $BR$ satisfies all the hypotheses of Kakutani’s Fixed Point Theorem, hence there exists a $\sigma^* \in \Sigma$ such that $\sigma^* \in BR(\sigma^*)$. By definition $\sigma^*$ is an equilibrium in mixed strategies. □

9.9 Equilibrium in Extensive Form Abstract Economies

An extensive form abstract economy with mixed strategy feasibility correspondences is a tuple $G = \langle I, A, T, \rho, \pi_0, H, u, C \rangle$, where $C = \{C_1, ..., C_N\}$, and $C_i : \Sigma_{-i} \rightarrow 2^{\Sigma_i}$. An
equilibrium in mixed strategies is defined in the natural way, where $U_i$ is given in section 9.1.

(C1) $C_i$ is nonempty and compact valued

(C2) $C_i$ is $U_i$-continuous

(C3) $C_i$ is convex valued

(C4) $C_i = \Psi_i \circ Q_i \circ \Phi_i$ for some canonical $Q_i : B_{-i} \to 2^{B_i}$

**Lemma 4:** If $G$ is finite and satisfies (C1) – (C3) then $G$ has an equilibrium in mixed strategies.

**Proof.** The induced normal form of $G$ is given by the tuple $\tilde{G} = \langle N, S, \tilde{u}, C \rangle$, where $\tilde{u}(s) = u(R^{-1}(s))$. By Lemma 3, $\tilde{G}$ has an equilibrium in mixed strategies $\sigma^*$. For $z \in Z$, if $s = R(z)$ then by construction $P(z|\sigma) = \sigma(s)$, so

$$U_i(\sigma) = \sum_{z \in Z} u_i(z)P(z|\sigma) = \sum_{s \in S} \tilde{u}_i(s)\sigma(s) = \tilde{U}_i(\sigma)$$

Since the feasibility correspondences and utility functions of $G$ and $\tilde{G}$ are identical, for any $\sigma_{-i} \in \Sigma_{-i}$, $\sigma_i \in BR_i(\sigma_{-i}) \iff \sigma_i \in \tilde{BR}_i(\sigma_{-i})$. So $\sigma^*$ is a constrained equilibrium in mixed strategies of $G$. □

**Lemma 5:** If $G$ satisfies (C1)-(C4) then $G' = \langle I, A, T, \rho, \pi_0, H, u, Q \rangle$ has an equilibrium in behavioral strategies, and any equilibrium in mixed strategies of $G$ is equivalent to an equilibrium in behavioral strategies of $G'$.

**Proof.** By Lemma 4, $G$ has an equilibrium in mixed strategies $\sigma^*$. By (C4) there exists $b^* \in B$ such that $\sigma^* \in \Psi(b^*)$. If $b^*_i \in Q_i(\Phi_i(\sigma^*_{-i}))$ then by construction there exists $b_{-i} \in \Phi_{-i}(\sigma^*_{-i})$ such that $b^*_i \in Q_i(b_{-i})$. Since $b_{-i} \sim \sigma^*_{-i} \sim b^*_{-i}$, and $Q_i$ is canonical it follows that $Q_i(b_{-i}) = Q_i(b^*_{-i})$ which implies that $b^*_i \in Q_i(b^*_{-i})$. So $b^*$ is feasible in $G'$.

Suppose by way of contradiction that $b^*$ is not an equilibrium in behavioral strategies. Then there exists a $b'_i \in Q_i(b^*_{-i})$ such that $U_i(b'_i, b^*_{-i}) > U_i(b^*_i, b^*_{-i})$. Let $\sigma'_{-i} \sim b'_i$. Since $b'_i \in Q_i(b^*_{-i})$ by construction $\sigma'_{-i} \in \Psi_i(Q_i(b^*_{-i}))$, and since $b^*_{-i} \in \Phi_{-i}(\sigma^*_{-i})$ it follows that $\sigma'_{-i} \in \Psi_i(Q_i(\Phi_{-i}(\sigma^*_{-i}))) = C_i(\sigma^*_{-i})$ by (C4). Since $U_i(\sigma'_i, \sigma^*_{-i}) = U_i(b'_i, b^*_{-i}) > U_i(b^*_i, b^*_{-i}) = U_i(\sigma^*_i, \sigma^*_{-i})$, $\sigma^*$ is not a constrained equilibrium in mixed strategies, a contradiction. So $b^*$ is an equilibrium in mixed strategies. □

Let $K_i : \Sigma_{-i} \to 2^{B_i}$, and define $U_i : B_i \times \Sigma_{-i} \to \mathbb{R}$ in the obvious way.
(K1) \( K_i \) is nonempty and compact valued

(K2) \( K_i \) is \( U_i \)-continuous

(K3) \( K_i \) is behavioral convex valued

**Lemma 6:** If \( K_i \) satisfies (K1)-(K3) then \( C_i = \Psi_i \circ K_i \) satisfies (C1)-(C3).

**Proof.** Upper Hemicontinuity By (K1) and (K2) \( K_i \) is upper hemicontinuous, non-empty, and closed valued. By Remark 1, \( \Psi_i \) is upper hemicontinuous, nonempty and closed valued. So \( C_i \) is upper hemicontinuous by Fact 2.

\( U_i \)-Lower Hemicontinuity: Let \( \sigma^n_{i-1} \to \sigma_{i-1} \) and suppose that \( \sigma_i \in C_i(\sigma_{i-1}) = \Psi_i(K_i(\sigma_{i-1})) \). By construction there exists a \( b_i \in K_i(\sigma_{i-1}) \) such that \( \sigma_i \in \Psi_i(b_i) \). By (K2) there exists a subsequence \( b^n_i \in K_i(\sigma^n_{i-1}) \) such that \( U_i(b^n_i, \sigma^n_{i-1}) \to U_i(b_i, \sigma_{i-1}) \). Let \( \sigma^n_i \in \Psi_i(b^n_i) \), by construction \( \sigma^n_i \in C_i(\sigma^n_{i-1}) \), so \( U_i(\sigma^n_i, \sigma^n_{i-1}) = U_i(b^n_i, \sigma^n_{i-1}) \to U_i(b_i, \sigma_{i-1}) = U_i(\sigma_i, \sigma_{i-1}) \).

Convex Valued: Let \( \sigma_i, \sigma'_i \in C_i(\sigma_{i-1}) \). By construction there exists \( b_i, b'_i \in K_i(\sigma_{i-1}) \) such that \( \sigma_i \in \Psi_i(b_i) \) and \( \sigma'_i \in \Psi_i(b'_i) \). Let \( \sigma^n_i = \lambda \sigma_i + (1-\lambda)\sigma'_i \) where \( \lambda \in [0,1] \). By the claim in Proposition 4, \( \sigma^n_i \in \Psi_i(B_i(b_i, b'_i, \lambda)) \). By (K3) \( B_i(b_i, b'_i, \lambda) \in K_i(\sigma_{i-1}) \) so \( \sigma^n_i \in C_i(\sigma_{i-1}) \), so \( C_i \) is convex valued.

Compact and Nonempty Valued: \( C_i \) is nonempty valued by construction. Since \( K_i \) is compact valued, and since \( \Psi_i \) is upper hemicontinuous and compact valued \( C_i \) compact valued by Fact 2.

**Theorem 2:** If \( G \) is finite, canonical, and satisfies generalized perfect recall then it has an equilibrium in behavioral strategies.

**Proof.** Transform Then Optimize: Let \( K_i = Q_i \circ \Phi_i \). Since \( Q_i \) is canonical it is nonempty valued so by the definition of \( \Phi_i \), \( K_i \) is nonempty valued. Since \( Q_i \) is canonical it is upper hemicontinuous, compact valued, and nonempty valued. By Remark 2, \( \Phi_{i-1} \) is compact valued, so \( K_i = Q_i \circ \Phi_{i-1} \) is compact valued by Fact 2. Moreover, by Remark 2, \( \Phi_{i-1} \) is upper hemicontinuous, compact valued, and nonempty. So \( K_i = Q_i \circ \Phi_{i-1} \) is upper hemicontinuous by Fact 3.

Let \( \sigma^n_{i-1} \to \sigma_{i-1} \) and \( b_i \in K_i(\sigma_{i-1}) \). By construction there exists a \( b_{i-1} \in \Phi_{i-1}(\sigma_{i-1}) \) such that \( b_i \in Q_i(b_{i-1}) \). Let \( b^n_{i-1} \in \Phi_{i-1}(\sigma^n_{i-1}) \), and since \( B_{i-1} \) is compact there is a convergent subsequence \( b^{n_k}_{i-1} \to \tilde{b}_{i-1} \). \( \Phi_{i-1} \) is upper hemicontinuous by Remark 2, and hence closed so \( \tilde{b}_{i-1} \in \Phi_{i-1}(\sigma_{i-1}) \).

By the definition of \( \Phi_{i-1} \), \( b_{i-1} \sim \tilde{b}_{i-1} \), so since \( Q_i \) is canonical \( b_i \in Q_i(\tilde{b}_{i-1}) \). Since \( Q_i \) is \( U_i \)-lower hemicontinuous, there exists a further subsequence \( b^n_{i-1} \) such that \( b_i^{n} \in Q_i(b^n_{i-1}) \) and \( U_i(b_i^{n}, b^n_{i-1}) \to U_i(b_i, \tilde{b}_{i-1}) = U_i(b_i, b_{i-1}) \). Since \( b_{i-1} \sim \tilde{b}_{i-1} \) and \( b^n_{i-1} \sim \sigma_{i-1} \) it follows that

\[
U_i(b_i^{n}, \sigma^n_{i-1}) = U_i(b_i^{n}, b^n_{i-1}) \to U_i(b_i, \tilde{b}_{i-1}) = U_i(b_i, b_{i-1}) = U_i(b_i, \sigma_{i-1})
\]
Since \( b_i^\sigma \in Q_i(\Phi_{-i}(\sigma^\sigma)) = K_i(\sigma^\sigma) \) it follows that \( K_i \) is \( U_i \)-lower hemicontinuous.

By construction \( K_i(\sigma_{-i}) = Q_i(\Phi_{-i}(\sigma_{-i})) = \bigcup_{b_{-i} \in \Phi_{-i}(\sigma_{-i})} Q_i(b_{-i}) = Q_i(b_{-i}) \) for any \( b_{-i} \sim \sigma_{-i} \) since \( Q \) is canonical. So \( K_i \) is behavioral convex valued since \( Q_i \) satisfies generalized perfect recall. So \( K_i \) satisfies (K1)-(K3).

Let \( C_i = \Psi_i \circ K_i = \Psi_i \circ Q_i \circ \Phi_{-i} \), so (C4) is satisfied. By Lemma \( 6 \) \( C_i \) satisfies (C1)–(C3). So by Lemma \( 5 \) \( G \) has an equilibrium in behavioral strategies.

\[ \text{Proof. Optimize then Transform: Since } Q_i \text{ is canonical it satisfies the hypotheses of Lemma } 1. \text{ It follows that} \]
\[ BR_i(b_{-i}) = \arg \max_{b_i \in Q_i(b_{-i})} U_i(b_i, b_{-i}) \]

is upper hemicontinuous non-empty and compact valued. Moreover, by Lemma \( 2 \) \( BR_i \) is behavioral convex valued.

Let \( \beta_i = \Psi_i \circ BR_i \circ \Phi_{-i} \). \( \beta_i \) is upper hemicontinuous by Remark 1, Remark 2, and Fact 3. Let \( \lambda \in [0, 1], \sigma'_i, \sigma''_i \in \beta_i(\sigma_{-i}) \), and \( \sigma_i = \lambda \sigma'_i + (1 - \lambda) \sigma''_i \). Since \( Q_i \) is canonical, if \( \sigma'_i \sim \sigma_{-i} \) then \( \sigma'_i, \sigma''_i \in \beta_i(\sigma_{-i}) \). Let \( b_{-i} \sim \sigma_{-i} \), \( \sigma'_i \sim b'_i \in Q_i(\Phi_{-i}(\sigma_{-i})) = Q_i(b_{-i}) \), and \( \sigma''_i \sim b''_i \in Q_i(\Phi_{-i}(\sigma_{-i})) = Q_i(b_{-i}) \). Since \( Q_i \) is behavioral convex valued \( B_i(b'_i, b''_i, \lambda) \in Q_i(b_{-i}) \).

By the claim of proposition 4, \( B_i(b'_i, b''_i, \lambda) \sim \lambda \sigma'_i + (1 - \lambda) \sigma''_i \), which implies that
\[ \lambda \sigma'_i + (1 - \lambda) \sigma''_i \in \Psi_i(BR_i(b_{-i})) = \Psi_i(BR_i(\Phi_i(\sigma_{-i}))) = \beta_i(\sigma_{-i}) \]

So \( \beta_i \) is convex valued. \( \beta_i \) is non-empty valued since it is a composition of non-empty valued correspondences. \( \beta_i \) is compact valued by Fact 2. Define \( \beta(\sigma) = \prod_{i=1}^N \beta_i(\sigma_{-i}) \). By Fact 5 \( \beta \) satisfies the hypotheses of the Kakutani Fixed point theorem, so there exists a \( \sigma^* \in \beta(\sigma^*) \).

Suppose \( \sigma^* \) is not an equilibrium in mixed strategies for \( C_i = \Psi_i \circ Q_i \circ \Phi_{-i} \), then there exists and \( \hat{\sigma}_i \in C_i(\sigma^*_i) \) such that \( U_i(\hat{\sigma}_i, \sigma^*_i) > U_i(\sigma^*_i) \). By construction there exists \( \tilde{b}_{-i}, \tilde{b}_i \) such that \( \hat{b}_{-i} \in \Phi_{-i}(\sigma^*_i), \hat{b}_i \in Q_i(\tilde{b}_{-i}), \) and \( \hat{\sigma}_i \in \Psi_i(\hat{\sigma}_i) \). Since \( Q \) is canonical \( Q_i(\tilde{b}_{-i}) = Q_i(b^*_i) \). Let \( b^*_i \sim \sigma^*_i \) and \( b^*_{-i} \sim \sigma^*_{-i} \), so \( U_i(b^*_i, b^*_{-i}) > U_i(\sigma^*_i, \sigma^*_{-i}) \).

This implies that \( b^*_i \notin BR_i(\Phi_{-i}(\sigma^*_{-i})) \) for any \( b^*_{-i} \sim \sigma^*_{-i} \), hence \( \sigma^* \notin \beta_i(\sigma^*_i) \) a contradiction. Therefore \( \sigma^* \) is an equilibrium in mixed strategies, and by Lemma \( 5 \) there exists a \( b^* \sim \sigma^* \) such that \( b^* \) is an equilibrium in behavioral strategies.

\[ \square \]

9.10 Tightness of Generalized Perfect Recall

**Proposition 5:** If \( \epsilon > 0 \) then \( G^* \) does not have an equilibrium in behavioral strategies.

**Proof.** Let \( B = \{ z \in Z : u_2(z) = 1 \} \). Denote the probability of \( B \) under behavioral strategy
b by $P(B|b)$, and notice that $U_2(b) = P(B|b)$, and $U_1(b) = 1 - P(B|b)$. By construction $P(B|b)$ is given by

$$P(B|b) = b(L|\phi)b(l|L)b(l|LL)b(L|LLL) + b(L|\phi)b(r|L)b(l|Lr)b(R|Llr) +$$

$$b(R|\phi)b(r|R)b(r|Rr)b(R|Rrr) + b(R|\phi)b(l|L)b(r|Rr)b(L|Lrr)$$

Let $b(L|\phi) = p_1 \Rightarrow b(R|\phi) = 1 - p_1$, $b(l|LL) = p_2 \Rightarrow b(r|Rl) = 1 - p_2$, $b(l|Lr) = p_3 \Rightarrow b(r|Rr) = 1 - p_3$, $b(L|LLL) = p_4 \Rightarrow b(R|Lrl) = 1 - p_4$, $b(R|Rrr) = p_5 \Rightarrow b(L|Rlr) = 1 - p_5$, $b(l|L) = p_6 \Rightarrow b(r|L) = 1 - p_6$, $b(r|R) = p_7 \Rightarrow b(l|R) = 1 - p_7$. So $P(B|b)$ can be written as,

$$P(B|b) = p_1p_6p_2p_4 + p_1(1 - p_6)p_3(1 - p_4) + (1 - p_1)p_7(1 - p_3)p_5 +$$

$$(1 - p_1)(1 - p_7)(1 - p_2)(1 - p_5)$$

Fixing a strategy $b_2$ for player 2, the problem of player 1 is given by

$$\min_{b_1 \in B_1} P(B|b_1, b_2) = \min_{p_1, p_4, p_5 \in [0,1]} p_1p_6p_2p_4 + p_1(1 - p_6)p_3(1 - p_4) + (1 - p_1)p_7(1 - p_3)p_5$$

$$+(1 - p_1)(1 - p_7)(1 - p_2)(1 - p_5) = \min\{p_6p_2, (1 - p_6)p_3, p_7(1 - p_3), (1 - p_7)(1 - p_2)\}$$

Denote the solution of this problem by $P(B|b_2)$. Suppose by way of contradiction that $P(B|b_2) \geq .25$, so the following must hold.

$$p_6p_2 \geq .25 \quad (1 - p_6)p_3 \geq .25 \quad p_7(1 - p_3) \geq .25 \quad (1 - p_7)(1 - p_2) \geq .25$$

Let $P = \{p_2, p_3, p_6, p_7, 1 - p_2, 1 - p_3, 1 - p_6, 1 - p_7\}$, and notice that for any $p \in P$ that $1 - p \in P$. Since every $p \in P$ appears in one of the expressions above, and since $p \leq 1$ for any $p \in P$ it follows that $(1 - p) \geq .25$ for any $p \in P$ which implies that $p \leq .75$ for any $p \in P$.

Similarly, suppose that $p \leq k$ for every $p \in P$. Then $(1 - p)k \geq .25$ for every $p \in P$ so $p \leq 1 - \frac{1}{4k}$. Let $\varphi(k) = 1 - \frac{1}{4k}$, and define $k_n$ recursively as follows, $k_0 = 1, k_{n+1} = \varphi(k_n)$ for $n > 0$.

Notice that $\Delta(k) = k - \varphi(k) = k - 1 + \frac{1}{4k} = \frac{1}{k}(k^2 - k + \frac{1}{4}) = \frac{1}{k}(k - \frac{1}{2})^2 > 0$ for $k > 0$. If $k > .5$ then $\varphi(k) = 1 - \frac{1}{4k} \geq .5$. Since $k_0 = 1$ this implies that $k_n \geq .5$ for any $n > 0$ so $\lim_{n \to \infty} k_n = \bar{k}$ is well defined.

Suppose by way of contradiction that $\bar{k} > .5$. Since $\bar{k} - \varphi(\bar{k}) = \frac{1}{k}(\bar{k} - \frac{1}{2})^2 > 0$ it must be that $k_n > \bar{k}$ for all $n$. Notice that $\Delta(k)$ is increasing in $k$ for $.5 < k \leq 1$ since $\frac{d\Delta}{dk} = 1 - \frac{1}{4k^2} > 0$. So if $\bar{k} < k_n \leq \bar{k} + \Delta(\bar{k})$ then $k_{n+1} = k_n - \Delta(k_n) \leq \bar{k} + \Delta(\bar{k}) - \Delta(k_n) < \bar{k}$. Since the sequence is decreasing it must satisfy $k_n > \bar{k} + \Delta(\bar{k})$ for all $n$, hence $k_n \to \bar{k}$ a contradiction. So
\( k = .5 \). By the same argument as above if \( p \leq .5 \) for every \( p \in P \) then \( p \geq .5 \) for every \( p \in P \), and hence \( p = .5 \).

So the unique \( b_2 \in Q_2 \) satisfying \( P(B|b_2) \geq .25 \) is given by

\[
\begin{align*}
 b_2'(l|L) &= .5 & b_2'(r|R) &= .5 \\
 b_2'(l|Ll) &= .5 & b_2'(r|Rr) &= .5 
\end{align*}
\]

Notice that \( P(B|b_1, b_2') = .25 \) for any \( b_1 \in Q_1 \). Suppose by way of contradiction that \( b^* \) is an equilibrium in behavioral strategies in which \( b_2' \neq b^*_2 \). Since player 1 is best responding, \( P(B|b^*) = P(B|b^*_1, b_2') \), so player 2 is not best responding, a contradiction. So in any equilibrium player 2 must choose \( b_2' \).

The final step is to show that for any \( \epsilon > 0 \) and any \( b_1 \in Q_1 \) there exists a \( b_2 \in Q_2 \) such that \( P(B|b_1, b_2) > .25 \). This implies that \( G^* \) has no equilibrium in behavioral strategies, since player 2 cannot be best responding when choosing \( b_2' \).

Let \( P(LL) = b(L|\phi)b(L|LL) \), \( P(LR) = b(L|\phi)b(R|LL) \), \( P(RL) = b(R|\phi)b(L|RL) \), and \( P(RR) = b(R|\phi)b(R|RL) \). If \( P(LL) > .25 \) then if player 2 chooses \( b_2(l|L) = b_2(l|LL) = 1 \), then \( P(B|b_1, b_2) > .25 \). A similar argument shows that if \( P(LR) > .25 \) or \( P(RL) > .25 \) or \( P(RR) > .25 \) then there exists a \( b_2 \in Q_2 \) such that \( P(B|b_1, b_2) > .25 \). Suppose that \( P(LL) = P(RL) = P(LR) = P(RR) = .25 \), which implies that \( b_1(L|\epsilon) = b_1(L|LL) = b_1(L|LL) = .5 \). Then the strategy given by \( b_2(l|L) = 1 \) \( b_2(r|R) = \epsilon \) \( b_2(l|LL) = 1 \) \( b_2(r|Rr) = 1 \) is feasible and \( P(B|b_1, b_2) = .25 + \frac{\epsilon}{4} > .25 \) for any \( \epsilon > 0 \).

\[ \Box \]

### 9.11 Properties of \( \mathcal{R}_i \)-convexity

**Proposition 6:** If \( G \) is irreducible, an information partition \( H_i \) satisfies perfect recall if and only if \( H_i \) is finer than \( \mathcal{R}_i \).

**Proof.** By the claims below, if \( G \) is irreducible then \( H_i \) satisfies perfect recall if an only if \( w \in h(w') \) implies \( R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i) \).

Suppose \( H_i \) is finer than \( \mathcal{R}_i \) then for any \( h \in H_i \) there exists \( W_i \in \mathcal{R}_i \) such that \( h \subseteq W_i \). The definition of \( \mathcal{R}_i \), implies that \( R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i) \) for any \( w, w' \in h \). Now suppose that \( w \in h(w') \) implies \( R_i(w) \cap S^*_i(H_i) = R_i(w) \cap S^*_i(H_i) \). Let \( h \in H_i \) and notice that there exists a \( W_i \in \mathcal{R}_i \) such that \( h \subseteq W_i \) by construction, so \( H_i \) is finer than \( \mathcal{R}_i \). \[ \Box \]

**Claim:** If \( H_i \) satisfies perfect recall then \( w \in h(w') \) implies \( R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i) \).

**Proof.** Let \( w' \in h(w) \) and \( s_i \in R_i(w) \cap S^*_i(H_i) \). Since \( s_i \in R_i(w) \) if \( \tilde{w} \in P_i(w) \) then \( s_i(\tilde{w}) = a(\tilde{w}, w) \). Let \( \tilde{w}' \in P_i(w') \). Since \( H_i \) satisfies perfect recall there exists a \( \tilde{w} \in P_i(w) \) such that \( \tilde{w}' = \tilde{w}(\tilde{w}) \) and \( a(\tilde{w}, w) = a(\tilde{w}', w') \). Since \( s_i \in S^*_i(H_i) \) then \( s_i(\tilde{w}') = s_i(\tilde{w}) = a(\tilde{w}, w) = a(\tilde{w}', w') \). Since \( \tilde{w}' \) was arbitrary \( s_i \in R_i(w') \) and therefore \( s_i \in R_i(w') \cap S^*_i(H_i) \). \[ \Box \]
Claim: If $G$ is irreducible and $w \in h(w')$ implies $R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i)$ then $H_i$ satisfies perfect recall.

Proof. Suppose by way of contradiction that $H_i$ does not satisfy perfect recall, $R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i)$, and $G$ is irreducible.

Case 1: Suppose that $w \in h(\hat{w})$ and $\hat{w}$ is a predecessor of $w$, and without loss of generality suppose there exists no $w' \in h(w)$ such that $w'$ is a predecessor of $\hat{w}$. Let $s_i \in R_i(\hat{w}) \cap S^*_i(H_i)$ and since $G$ is irreducible there exists $a \in A(\hat{w})$ such that $a \neq a(\hat{w}, w)$. Let $s_i(\hat{w}) = a$, and notice that $s_i \notin R_i(w) \cap S^*_i(H_i)$ and therefore $R_i(w) \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$.

Case 2: Let $w'' \in h(w')$ and $w$ be predecessor of $w'$ where player $i$ acts. Suppose there is no $\hat{w} \in h(w)$ such that $\hat{w}$ is a predecessor of $w''$ and $a(w, w') = a(\hat{w}, w'')$. So either there is not a $\hat{w} \in h(w)$ such that $\hat{w}$ is a predecessor of $w''$, or if $\hat{w} \in h(w)$ such that $\hat{w}$ is a predecessor of $w''$ then $a(w, w') \neq a(\hat{w}, w'')$.

In the first case $h(w) \in h(P_i(w'))$ and $h(w) \notin h(P_i(w''))$ so $R_i(w'') \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$ by Lemma 6, a contradiction.

In the second case, let $\hat{w} \in h(w)$ be a predecessor of $w''$ such that $a(w, w') \neq a(\hat{w}, w'')$. So $w \in P_i(\hat{w})$ and $\hat{w} \in P_i(w'')$ such that $h(w) = h(\hat{w})$ but $a(w, w') \neq a(\hat{w}, w'')$. So $R_i(w) \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$ by Lemma 6, a contradiction.

Lemma 6: If $G$ is irreducible, then $R_i(w) \cap S^*_i(H_i) = R_i(w') \cap S^*_i(H_i)$ if and only if $h(P_i(w)) = h(P_i(w'))$ and $\hat{w} \in P_i(w) \Rightarrow \hat{w} \in P_i(w')$ such that $h(\hat{w}) = h(\hat{w}')$ implies $a(\hat{w}, w) = a(\hat{w}', w)$.

Proof. Suppose $h(P_i(w)) \neq h(P_i(w'))$. Since $G$ is irreducible there exists a $s_i(\hat{w}) = a(\hat{w}, w)$ for $\hat{w} \in h(\hat{w}) \in h(P_i(w))$ and $s_i(\hat{w}) = a(\hat{w}', w')$ for $\hat{w}' \in h(\hat{w}') \in h(P_i(w')) \setminus h(P_i(w)) \neq \emptyset$. By construction $s_i \in R_i(w) \cap S^*_i(H_i)$ but $s_i \notin R_i(w') \cap S^*_i(H_i)$ and therefore $R_i(w) \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$.

Suppose $a_i(\hat{w}, w) \neq a(\hat{w}', w')$ for some $\hat{w} \in P_i(w) \hat{w}' \in P_i(w')$ such that $h(\hat{w}) = h(\hat{w}')$. Let $s_i(\hat{w}) = a(\hat{w}, w)$ for any $\hat{w} \in P_i(w)$, then by definition $s_i \in R_i(w)$ but since $h(\hat{w}) = h(\hat{w}')$, if $s_i \in S^*_i(H_i)$ then $s_i(\hat{w}') = a(\hat{w}', w) \neq a(\hat{w}', w)$ so $s_i \notin R_i(w') \cap S^*_i(H_i)$ and therefore $R_i(w) \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$.

Let $s_i \in R_i(w) \cap S^*_i(H_i)$, so by definition if $\hat{w} \in P_i(w)$ then $s_i(w) = a(\hat{w}, w)$. Suppose that $h(P_i(w)) = h(P_i(w'))$ and that if $\hat{w} \in P_i(w)$ and $\hat{w}' \in P_i(w')$ such that $h(\hat{w}) = h(\hat{w}')$ then $a(\hat{w}, w) = a(\hat{w}', w')$. So for any $\hat{w}' \in P_i(w')$ there exists a $\hat{w} \in P_i(w)$ such that $h(\hat{w}) = h(\hat{w}')$.

So $s_i \in S^*_i(H_i)$ then $s_i(\hat{w}') = s_i(\hat{w}) = a(\hat{w}, w) = a(\hat{w}', w)$ and therefore $s_i \in R_i(w') \cap S^*_i(H_i)$, so $R_i(w) \cap S^*_i(H_i) \neq R_i(w') \cap S^*_i(H_i)$.
Proposition 7: If $H_i$ satisfies perfect recall, and $\tilde{Q}_i$ is $\mathcal{R}_i$-partitionally convex valued then $Q_i = \tilde{Q}_i \cap B_i^\ast(H_i)$ satisfies generalized perfect recall.

Proof. Let $b_i, b'_i \in Q_i(b_{-i}) = \tilde{Q}_i(b_{-i}) \cap B_i^\ast(H_i)$ and suppose that $H_i$ satisfies perfect recall and $\tilde{Q}_i$ is $\mathcal{R}_i$-partitionally convex valued.

Suppose that $R_i(w) \cap S_i^\ast(H_i) = R_i(w') \cap S_i^\ast(H_i)$. Suppose that $\sigma_i \in \Sigma_i^\ast(H_i)$.

$$P_i^i(w|\sigma_i) = \sum_{s_i \in R_i(w)} \sigma_i(s_i) = \sum_{s_i \in R_i(w')} \sigma_i(s_i) = P_i^i(w'|\sigma_i)$$

Since $H_i$ satisfies perfect recall, for any $b_i, b'_i \in Q_i$ there exists $\sigma_i' \sim b'_i$ and $\sigma_i \sim b_i$, so $P_i^i(w'|b_i) = P_i^i(w|b_i)$ for any $b_i \in Q_i(b_{-i})$. This implies that $\Lambda_i(b_i, b'_i, \lambda, w) = \Lambda_i(b_i, b'_i, \lambda, w')$ for any $b_i, b'_i \in Q_i(b_{-i})$.

By the same argument as in Theorem 1, $B_i^\ast(H_i)$ is behavioral convex. Since $\tilde{Q}_i$ is $\mathcal{R}_i$-convex valued, if $b_i, b'_i \in \tilde{Q}_i(b_{-i})$ then $b''_i \in \tilde{Q}_i(b_{-i})$ if $b''_i$ is a $\mathcal{R}_i$-partitionally convex combination of $b_i$ and $b'_i$. That is $b''_i(a|w) = \lambda(w)b_i(a|w) + (1 - \lambda(w))b'_i(a|w)$, and $\lambda(w) = \lambda(w')$ if $R_i(w) \cap S_i^\ast(H_i) = R_i(w') \cap S_i^\ast(H_i)$. The previous paragraph implies that $B_i(b_i, b'_i, \lambda)$ is a $\mathcal{R}_i$-partitionally convex combination of $b_i$ and $b'_i$ for any $\lambda \in [0, 1]$, and hence $\tilde{Q}_i$ is behavioral convex valued. Clearly the intersection of behavioral convex sets is behavioral convex, hence $Q_i(b_{-i}) = \tilde{Q}_i(b_{-i}) \cap B_i^\ast(H_i)$ is behavioral convex, and therefore $Q_i$ satisfies generalized perfect recall since $b_{-i}$ was arbitrary. \hfill \Box

9.12 Rational Inattention and Continuous Diameter Constraints

Lemma 7: If $G$ is irreducible, $Q_i^k$ is convex valued, and $(\mathcal{R}_i, W_i, H_i)$ is standard then $Q_i$ satisfies the hypotheses of Proposition 7.

Proof. Since $G$ is irreducible and $H_i$ is finer than $\mathcal{R}_i$, $H_i$ satisfies perfect recall by Proposition 6. By construction $Q_i$ is given by

$$Q_i(\cdot) = \tilde{Q}_i(\cdot) \cap B_i^\ast(H_i) = \bigcap_{k=1}^{K_i} Q_i^k(\cdot) \cap B_i^\ast(H_i)$$

Let $b_i, b'_i \in Q_i(b_{-i})$ and $b''_i$ be a $\mathcal{R}_i$-partitionally convex combination of the two, with node dependent mixing coefficient given by $\lambda(w)$. Since $W_i$ is finer than $\mathcal{R}_i$, if $w, w' \in W_i^k \in W_i$ then $R_i(w) \cap S_i^\ast(H_i) = R_i(w') \cap S_i^\ast(H_i)$, and therefore $\lambda(w) = \lambda(w')$. Since $b''_ik = \lambda(w)b^k_i + (1 - \lambda(w))b''_i$, and $Q_i^k$ is convex valued it follows that $b''_ik \in Q_i^k(b_{-i})$. By construction $b''_i \in B_i^\ast(H_i)$, and therefore $b''_i \in Q_i(b_{-i})$ which implies that $Q_i$ is $\mathcal{R}_i$-partitionally convex valued. \hfill \Box

Lemma 8: If $b_i \sim b'_i$ then $b_i(\cdot|w) = b'_i(\cdot|w)$ for $w$ on the potential path of play of $b_i$. 49
Proof. Let \( b_i \sim b'_i \), so if \( P_i(w|b) > 0 \) then \( b_i(a_i|w) = b'_i(a_i|w) \). So it is sufficient to show that if \( w \) is on the potential path of play of \( b_i \). By construction \( s_i \in R_i(w) \implies s_{i-1} \) such that \( s = (s_i, s_{-i}) \in R(w) \). Pick a \( s_i \in R_i(w) \) such that \( \sigma_i(s_i) > 0 \), and let \( b_{-i} \in \Phi_{-i}(s_{-i}) \). Then \( P(w|b) \geq \sigma_i(s_i) > 0 \), so \( w \) is on the path of play of \( b_i \), and hence on the potential path of play of \( b_i \).

Let \( \tilde{w} \) be a predecessor of \( w \), suppose \( \tilde{w} \) and \( w \) are on the path of play of \( b_i \), and let \( \tilde{w} \in h \in H \). The probability under \( b_i \) of reaching \( w \) given play previously reached \( \tilde{w} \) and the realized actions of player \( i \) are consistent with \( w \) is

\[
P_i(w|\tilde{w}, b) = \prod_{j \in \Omega \setminus \{i\}} \prod_{w' \in P_j(w)} b_j(a(w', w)|w')
\]

The probability under \( b_i \) of reaching \( w \) given that play has previously reached \( h \), and player \( i \)'s realized actions are consistent with \( w \) is

\[
P_i(w|h, b) = \frac{P_i(w|\tilde{w})}{\sum_{w' \in h(\tilde{w})} P_i(w'|h)} P_i(w|\tilde{w}, b)
\]

The probability under \( b_i \) of reaching \( w \) given that play previously reached \( h \), player \( i \)'s realized actions are consistent with \( w \), and play has reached \( W^k_i \) is denoted by

\[
P_i(w|h, b, W^k_i) = \frac{P_i(w|h, b)}{\sum_{w' \in W^k_i} P_i(w'|h, b)}
\]

Let \( (\Omega, \Sigma, p) \) be a probability space, and let \( A, B \) be random variables with joint distribution \( p(a, b) \), marginal distributions \( p(a) \) and \( p(b) \), and conditional distribution \( p(a|b) \). The mutual information \( I(A; B) \) of \( A \) and \( B \) is given by,

\[
I(A; B) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log(\frac{p(a, b)}{p(a)p(b)}) = \sum_{a \in A} \sum_{b \in B} p(a|b)p(b) \log(\frac{p(a|b)}{p(a)}) = \sum_{a \in A} \sum_{b \in B} p(a|b)p(b) \log(\frac{p(a|b)}{\sum_{b \in B} p(a|b)p(b)})
\]

Where it is assumed that \( 0 \log(0) = 0 \). The following facts are well known facts about the mutual information function \( I(A; B) \).

1. \( I(A; B) \) is a convex function of \( p(a|b) \)
2. \( I(A; B) \) is a concave function of \( p(b) \)
3. \( I(A; B) \) is a continuous in \( p(b) \) and \( p(a|b) \)
(4) \( I(A; B) \geq 0 \)

(5) If \( A \) and \( B \) are independent then \( I(A; B) = 0 \)

By Lemma 6, if \( W_i \) is finer than \( R_i \) then \( w, w' \in W_i \) implies \( h(P_i(w)) = h(P_i(w')) \), and therefore the last information where player \( i \) acted before reaching the constraint set \( W_i^k \), denoted by \( h'(W_i^k) \), is well defined. Let \( p_i^k(w) = P^i(w|h'(W_i^k), b, W_i^k) \). These are the relevant probabilities for constructing rational inattention constraints since they on everything player \( i \) knows upon reaching a node in \( W_i^k \). The mutual information between a constraint set \( W_i^k \) and the local behavioral strategy of player \( i \) is given by,

\[
I(W_i^k; A(W_i^k)) = \sum_{w \in W_i^k} \sum_{a_i \in A(w)} b_i(a_i | w) p_i^k(w) \log \left( \frac{b_i(a_i | w)}{\sum_{\hat{a} \in W_i^k} b_i(a_i | \hat{w}) p_i^k(\hat{w})} \right)
\]

**Proposition 8:** If \((R_i, W_i, H_i)\) is standard, and \( G \) is irreducible, then rational inattention constraints satisfy the hypotheses of Theorem 2.

**Proof.** Suppose \( b_{i-1} \sim b'_{i-1} \), then by Lemma 8 \( b_{i-1}(-|w) = b'_{i-1}(-|w) \) for \( w \) on the potential path of play of \( b_{i-1} \). Clearly if \( w \) is on the potential path of play of \( b_{i-1} \), and \( \tilde{w} \) is a predecessor of \( w \) then \( \tilde{w} \) is on the potential path of play of \( b_{i-1} \). So for \( w \) on the potential path of play of \( b_{i-1} \), and for any \( b_i \in B_i \), \( P^i(w|h, (b_i, b_{i-1}), W_i^k) = P^i(w|h, (b_i, b'_{i-1}), W_i^k) \). So \( Q_i^k(b_{i-1}) = Q_i^k(b'_{i-1}) \) for all \( W_i^k \) on the potential path of play of \( b_{i-1} \). Off the potential path of play \( Q_i^k(b_{i-1}) = B_i^k = Q_i^k(b'_{i-1}) \) so \( Q_i(b_{i-1}) = Q_i(b'_{i-1}) \).

\( Q_i \) is nonempty valued since \( c_{ik} \geq 0 \) and \( I(W_i^k; A(W_i^k)) \geq 0 \). \( Q_2 \) is compact valued since \( I \) is continuous in \( b_i \), and \( B_i \) is compact.

Let \( b^n_{i-1} \rightarrow b_{i-1} \), and \( b^{nk}_i \rightarrow b^k_i \) with \( b^{nk}_i \in Q_i(b^{n}_{i-1}) \). Let \( I^k(b^k_i, b_{i-1}) = I(W_i^k; A(W_i^k)) \). Since \( p_i^k(w) \) is continuous in \( b_{i-1} \), and \( I(W_i^k, A_i) \) is continuous in \( b_i^k \) and \( p_i^k(w) \), \( I^k \) is continuous in both arguments. So if \( W_i^k \) is on the potential path of play of \( b^n_{i-1} \) then \( I^k(b^n_{i-1}, b^{nk}_i) \leq c_{ik} \). If \( W_i^k \) is not on the potential path of play of \( b_{i-1} \), then \( b^k_i \in Q_i^k(b_{i-1}) \) by construction.

If \( W_i^k \) is on the potential path of play of \( b^n_{i-1} \), then there exists an integer \( N \) such that if \( n > N \) then \( W_i^k \) is on the potential path of play of \( b^n_{i-1} \). So if \( n > N \) then \( I^k(b^n_{i-1}, b^{nk}_i) \leq c_{ik} \Rightarrow I^k(b^n_{i-1}, b^k_i) \leq c_{ik} \) since \( I^k \) is continuous, so \( b^k_i \in Q_i^k(b_{i-1}) \), so \( Q_i^k \) is closed and hence upper hemicontinuous by Fact 1. So \( Q_i \) is upper hemicontinuous by Fact 4 and Fact 5.

Let \( b^n_{i-1} \rightarrow b_{i-1} \), and \( b_i \in Q_i(b_{i-1}) \), so \( b^k_i \in Q_i^k(b_{i-1}) \). Since \( b^n_{i-1} \rightarrow b_{i-1} \) there exists a \( N \) such that if \( n \geq N \) and \( w \) is on the potential path of play of \( b_{i-1} \) then \( w \) on the potential path of play of \( b^n_{i-1} \).

Let \( \tilde{b}_i(a_i | w) = \tilde{b}_i(a_i | w') \) for all \( w, w' \in W_i^k \). So \( \tilde{b}_i \) is constant over every constraint set, and therefore \( b_i^k \) is independent of \( W_i^k \) so \( I^k(\tilde{b}_i^k, b_{i-1}) = 0 \) for any \( W_i^k \in \Pi_i(b_{i-1}) \) and

51
\( b_{-i} \in B_{-i} \). Define

\[
\lambda^n = \min_{W^k \in \Pi_i(b_{-i})} \left\{ \frac{c_{ik}}{I^k(b^k_i, b^n_{-i})}, 1 \right\}
\]

And for \( W^k_i \in \Pi_i(b_{-i}) \) let \( b^{nk}_i = \lambda^n b^k_i + (1 - \lambda^n)\bar{b}^k_i \). Since \( I^k \) is concave in \( b^k_i \) it follows that

\[
I^k(b^{nk}_i, b^n_{-i}) = I^k(\lambda^n b^k_i + (1 - \lambda^n)\bar{b}^k_i, b^n_{-i}) \leq \\
\lambda^n I^k(b^k_i, b^n_{-i}) + (1 - \lambda^n)I^k(\bar{b}^k_i, b^n_{-i}) \leq c_{ik}
\]

\( I^k \) is a continuous function of \( b^n_{-i} \) and \( b^n_{-i} \to b_{-i} \), so \( \lambda^n \to 1 \). For \( W^k_i \notin \Pi_i(b_{-i}) \) let \( b^{nk}_i = \bar{b}^k_i \), so \( I^k(b^{nk}_i, b^n_{-i}) = 0 \). Therefore \( b^n_i \in Q_i(b^n_{-i}) \) since the capacities are weakly positive.

Let \( \tilde{b}^k_i = b^k_i \) for \( W^k_i \in \Pi_i(b_{-i}) \), and \( \tilde{b}^k_i = \bar{b}^k_i \) for \( W^k_i \notin \Pi_i(b_{-i}) \). Since \( U_i \) is continuous and \( \lambda^n \to 1 \), it follows that \( U_i(\tilde{b}^k_i, b^n_{-i}) \to U_i(\bar{b}^k_i, b_{-i}) \). Moreover since \( \tilde{b}^k_i = b^k_i \) for \( W^k_i \in \Pi_i(b_{-i}) \), it follows that \( U_i(\tilde{b}^k_i, b_{-i}) = U_i(\bar{b}^k_i, b_{-i}) \), and therefore \( Q_i \) is \( U_i \)-lower hemi-continuous.

Let \( b^k_i, b^{k^*}_{-i} \in Q^k_i(b_{-i}) \) so \( I^k(b^k_i, b_{-i}) \leq c_{ik} \) and \( I^k(b^{k^*}_{i}, b_{-i}) \leq c_{ik} \). Let \( b^{k^n}_{i} = \lambda b^k_i + (1 - \lambda) b^{k^*}_{-i} \). Since \( I^k \) is convex in \( b_i \) we have \( I^k(b^{k^n}_{i}, b_{-i}) = I^k(\lambda b^k_i + (1 - \lambda) b^{k^*}_{i}, b_{-i}) \leq \lambda I^k(b^k_i, b_{-i}) + (1 - \lambda) I^k(b^{k^*}_{i}, b_{-i}) \leq c_{ik} \). So \( b^{k^n}_i \in Q^k_i(b_{-i}) \) so \( Q^k_i \) is convex valued.

Since the partition trio is standard and \( G \) is irreducible Lemma 7 implies that \( Q_i \) satisfies the hypotheses of Proposition 7, and therefore \( Q_i \) satisfies generalized perfect recall. The arguments above show that \( Q_i \) is canonical, so \( Q_i \) satisfies the hypotheses of Theorem 2. \( \square \)

**Proposition 9:** If \( (R_i, W_i, H_i) \) is standard, and \( G \) is irreducible then continuous diameter constraints satisfy the hypotheses of Theorem 2.

**Proof.** \( (Q3) \) is satisfied by \( (CD3) \). \( Q_i \) is nonempty by \( (CD1) \) and \( (CD2) \), and \( Q_i \) is compact valued by \( (CD1) \).

Let \( b^n_{-i} \to b_{-i} \), and suppose \( b^n_i \in Q_i(b^n_{-i}) \) and \( b^n_i \to b_i \). So if \( w, w' \in W^k_i \) then \( |b^n_i(a_i|w) - b^n_i(a_i|w')| \leq d^k_i(b^n_{-i}) \). Since each \( d^k_i \) is continuous if \( w, w' \in W^k_i \) then \( |b_i(a_i|w) - b_i(a_i|w')| \leq d^k_i(b_{-i}) \) which implies that \( b_i \in Q_i(b_{-i}) \), so \( Q_i \) is upper hemi-continuous.

Let \( b_i \in Q_i(b_{-i}) \) and \( b^n_{-i} \to b_{-i} \), and let

\[
\tilde{b}^i_n = \arg \min_{\tilde{b}_i \in Q_i(b^n_{-i})} |b_i - \tilde{b}_i|
\]

which is well defined since the supremum metric is continuous and \( Q_i \) is compact. Since \( B_i \) is compact there exists a convergent subsequence \( b^n_i \to \tilde{b}_i \). Suppose by way of contradiction that \( \tilde{b}_i \neq b_i \), so there exists a \( c > 0 \) such that \( |b_i - \tilde{b}_i| \geq 2c > 0 \).

Without loss of generality take \( c < 1 \). Since \( b^n_i \to \tilde{b}_i \) there exists a \( L \) such that for \( l > L \),
\[ |b_i - b_i^n| > c. \] Since \( b_i^n \to b_{-i} \) there exists a \( N \) such that if \( n > N \) then \( |d_t^k(b_i^n) - d_t^k(b_{-i})| < d_t^k(b_{-i})c \) for all \( W_i^k \in W_i \). Let \( F = \max\{n_L, M_i\} \), and let \( f > F \).

So \( |b_i - b_i'| > c \) and \( |d_t^k(b_i') - d_t^k(b_{-i})| < d_t^k(b_{-i})c \). If \( d_t^k(b_i') > d_t^k(b_{-i}) \), then \( b_i \in Q_i(b_i') \) so

\[ b_i' \notin \arg \min_{b_i \in Q_i(b_i')} |b_i - b_i'| \]
a contradiction. If \( d_t^k(b_i') < d_t^k(b_{-i}) \) and \( b_i \in Q_i(b_i') \) the same argument is valid.

If \( d_t^k(b_i') < d_t^k(b_{-i}) \) and \( b_i \notin Q_i(b_i') \) there exists \( w, w' \in W_i^k \), and \( a_i \in A(w) \) such that

\[ |b_i(a_i|w) - b_i(a_i|w')| > d_t^k(b_i') > d_t^k(b_{-i})(1 - c) \]

Let \( b_i(a_i|w) = \tilde{b}_i(a_i|w') \) for any \( w, w' \in W_i^k \) and for any \( W_i^k \in W_i \). Let \( b_i' = (1-c)b_i + c\tilde{b}_i \), then for \( w, w' \in W_i^k \),

\[ |b_i'(a_i|w) - b_i'(a_i|w')| = (1-c)|b_i(a_i|w) - b_i(a_i|w')| < d_t^k(b_{-i})(1 - c) < d_t^k(b_i') \]

and therefore \( b_i' \in Q_i(b_i') \). By construction \( |b_i - b_i'| = |(1-c)b_i - c\tilde{b}_i| < c \), so \( |b_i - b_i'| < |b_i - b_i'| \) a contradiction. So \( b_i^n \to b_i \) and therefore \( Q_i \) is lower hemicontinuous. Moreover, since \( U_i \) is continuous \( Q_i \) is \( U_i \)-lower hemicontinuous, so \( Q_i \) is canonical.

Let \( b_i^k, b_i^k' \in Q_i^k(b_{-i}) \), so \( |b_i^k(a_i|w) - b_i^k(a_i|w')| \leq d_t^k(b_{-i}) \) and \( |b_i^k'(a_i|w) - b_i^k'(a_i|w')| \leq d_t^k(b_{-i}) \). Let \( b_i^{k''} = \lambda b_i^k + (1 - \lambda)b_i^{k''} \), so

\[ |b_i^{k''}(a_i|w) - b_i^{k''}(a_i|w')| \leq |\lambda b_i^k(a_i|w) + (1-\lambda)b_i^{k'}(a_i|w) - \lambda b_i^k(a_i|w') - (1-\lambda)b_i^{k'}(a_i|w')| \leq \lambda|b_i^k(a_i|w) - b_i^k(a_i|w')| + (1-\lambda)|b_i^{k'}(a_i|w) - b_i^{k'}(a_i|w')| \leq d_t^k(b_{-i}) \]

which implies that \( b_i^{k''} \in Q_i^k(b_{-i}) \), so \( Q_i^k \) is convex valued.

The partition trio is standard and \( G \) is irreducible so by Lemma 7 \( Q_i \) satisfies the hypotheses of Proposition 7, which implies that \( Q_i \) satisfies generalized perfect recall. Moreover, since \( Q_i \) is canonical it hypotheses of Theorem 2. \( \blacksquare \)

### 9.13 Perfection in Extensive Form Abstract Economies

**Theorem 3:** If \( G \) is finite, canonical, admits trembling, and satisfies generalized perfect recall then it has generalized perfect equilibrium.

**Proof.** Since \( G \) admits trembling there exists \( \delta > 0 \) such that for any \( \epsilon \leq \delta \), \( Q_i(\cdot, \epsilon) \) is non-empty valued. Since \( Q_i \) is canonical it is closed valued. \( \Gamma_i(\epsilon) \) is closed by construction, so
\[ Q_i(b_{-i}, \epsilon) = Q_i(b_{-i}) \cap \Gamma_i(\epsilon) \text{ is closed, so } Q_i(\cdot, \cdot) \text{ is compact valued.} \]

\[ \Gamma_i(\epsilon) \text{ is behavioral convex by construction. } Q_i \text{ is behavioral convex valued since it satisfies generalized perfect recall, so } Q_i(b_{-i}, \epsilon) = Q_i(b_{-i}) \cap \Gamma_i(\epsilon) \text{ is behavioral convex, and therefore } Q_i(\cdot, \cdot) \text{ is behavioral convex valued.} \]

\[ \Gamma_i(\epsilon) \text{ is upper hemicontinuous, non-empty valued for } \epsilon \leq \delta \text{ and compact valued by construction. } Q_i \text{ is upper hemicontinuous, non-empty valued, and compact valued since it is canonical. Since } Q_i \text{ admits trembling, } Q_i(b_{-i}, \epsilon) \text{ is nonempty for } \leq \delta, \text{ so by Fact 4 } Q_i(\cdot, \cdot) \text{ is upper hemicontinuous for } \epsilon \leq \delta. \]

Since \( Q_i \) admits trembling \( Q_i(\cdot, \cdot) \) is \( U_i \)-lower hemicontinuous.

If \( b_{-i} \sim b'_{-i} \) then \( Q_i(b_{-i}) = Q_i(b'_{-i}) \) by (Q3). So if \( b_{-i} \sim b'_{-i} \) then

\[ Q_i(b_{-i}, \epsilon) = Q_i(b_{-i}) \cap \Gamma_i(\epsilon) = Q_i(b_{-i}) \cap \Gamma_i(\epsilon) = Q_i(b'_{-i}, \epsilon) \]

so \( Q_i(\cdot, \cdot) \) satisfies (Q3).

So for any \( \epsilon \leq \delta \), \( G^\epsilon \) satisfies the hypotheses of Theorem 2, and therefore \( G^\epsilon \) has an equilibrium in behavioral strategies. Let \( \epsilon_n \to 0 \) and let \( b_n^* \) be an equilibrium of \( G^{\epsilon_n} \). Since \( B \) is compact there exists a convergent subsequence \( b_{n_k}^* \to b^* \). By the upper hemicontinuity of \( Q_i(\cdot, \cdot) \), \( b_k^* \in Q_i(b_{n_k}^*, 0) = Q_i(b_{n_k}^*, 0) \).

Suppose that \( b_i^* \notin BR(b_{n_k}^*) \). Then there exists a \( \tilde{b}_i \in Q_i(b_{n_k}^*) \) such that \( U_i(\tilde{b}_i, b_{n_k}^*) > U_i(b_i^*, b_{n_k}^*) \). By the \( U_i \)-lower hemicontinuity of \( Q_i(\cdot, \cdot) \), there exists a \( \tilde{b}_i^{n_k} \in Q_i(b_{n_k}^*, b_{n_k}^*) \) such that \( U_i(\tilde{b}_i^{n_k}, b_{n_k}^* - 1) \to U_i(\tilde{b}_i, b_{-i}^*) \). Let \( U_i(\tilde{b}_i, b_{n_k}^*) - U_i(b_i^*, b_{n_k}^*) = 2\gamma > 0 \).

Since \( U_i \) is continuous, there exists \( N_1 \) such that for \( n_k > N_1 \), \( |U_i(\tilde{b}_i^{n_k}, b_{n_k}^* - 1) - U_i(\tilde{b}_i, b_{-i}^*)| < \gamma \). Since \( U_i \) is continuous there exists a \( N_2 \) such that for \( n_k > N_2 \), \( |U(b_{n_k}^*, b_{n_k}^* - 1) - U(b_{n_k}^*, b_{-i}^*)| < \gamma \). So if \( n_k > \max\{N_1, N_2\} \) then \( U_i(\tilde{b}_i^{n_k}, b_{n_k}^* - 1) > U_i(b_{n_k}^*, b_{n_k}^*) \), so \( b_{n_k}^* \) is not an equilibrium, a contradiction. So \( b_i^* \in BR_i(b_{n_k}^*) \) for all \( i \), so \( b^* \) is an equilibrium of \( G^0 = G \), so by definition \( b^* \) is a generalized perfect equilibrium of \( G \).

\[ \square \]

References


