To Learn or Not to Learn?

Incompetence and Learning in Matrix Games

Thomas Graham

Supervised by Jerzy Filar, Thomas Taimre

9th November, 2020

Abstract

Although the strategic insights offered by game theory are often helpful for planning and decision making, traditional game-theoretic frameworks generally ignore the participants' skill levels. Beck and Filar [2] introduce the notion of incompetence into matrix games to capture accidental deviations. How can incompetence be counteracted, either through short-term immediate strategies or long-term learning strategies? We address this question by investigating the equilibria of strategic interactions involving incompetence and learning. Consequently, this leads us to establish various properties of incompetence, including an explanation of game value plateaus and conditions on optimal learning parameters. Moreover, we describe a model of incremental learning and develop a backward induction procedure to exhaustively compute learning strategies. This model's insights are illustrated using a simple tennis game adapted from [1]. Hopefully, by better understanding the strategic considerations arising from players' incompetence, we can develop more realistic and robust strategies.

Contents

List of Figures

List of Tables

1 Introduction

Often, in a repeated strategic situation whose outcome is dependent on the skill of its participants, we might expect these participants to improve their skills by learning or training. Consider, for instance, a tennis player who must practice to increase their competitive performance or a military that must invest to increase their capabilities. These seemingly disparate examples share two common characteristics: short-term skill-dependent interactions with opponents (tennis matches or military engagements) and long-term opportunities to improve skills (practicing tennis shots or investing in military equipment). We are interested in studying situations with these characteristics from both a short-term perspective and a long-term perspective; that is, we want to provide insight into the strategies that can be employed to achieve desired outcomes.

A natural quantitative framework to explore this problem is provided by game theory, which investigates strategic interactions between rational agents whose objectives are not necessarily aligned. This mathematical notion of a game, inspired by its namesake, consists of multiple players who select actions with the intention of maximising their rewards. The modern game-theoretic paradigm was introduced in 1928 by von Neumann's "On the Theory of Games of Strategy" [25] and popularised in 1944 by von Neumann and Morgenstern's "Theory of Games and Economic Behaviour" [26]. Although their methodology was originally presented in the context of economic applications, it has become a useful tool for analysing problems in a variety of other domains, including political science, theoretical biology, and military planning [17]. Here, we will motivate our discussion of skill by reviewing the evolution and applications of game theory.

Unsurprisingly, several mathematical discussions of strategy—often in the context of parlour games—predate the methodology of modern game theory. Note that, similar to von Neumann and Morgenstern's [26] contributions, these discussions predominantly focus on two-player zero-sum games, or games in which total rewards are conserved. The earliest application of a minimax solution, a pair of strategies that minimise each player's maximum loss, is attributed to Waldegrave's analysis of a two-player variant of the card game *Le Her* in 1713 [8]. Precisely, each player is assigned a strategy that maximises their probability of winning regardless of their opponent's choices. Although Waldegrave expresses scepticism of the solution's need to randomise (or mix) over actions, these mixed strategies play a critical role in further developments [8]. Indeed, Borel [7] allows players to randomise over their actions in games whose "winnings depend [symmetrically] both on chance and the skill of the players." The existence of a minimax solution is proved in [7] and [5] under the assumption that players possess three and five actions, respectively. Borel [6] concludes by speculating that a minimax solution exists when players have seven available actions; however, he remains unconvinced as to whether the statement can be generalised to arbitrary finite collections of actions.

The scope of the problem is widened by von Neumann [25] to include two-player zero-sum games with finitely many actions and arbitrary, not necessarily symmetric, rewards. It is demonstrated, by way of a counterexample, that a minimax solution might not exist when players are restricted to using non-randomised (or pure) strategies. Then, in a similar vein to Borel [7], this motivates the need for unpredictable behaviour in the form of mixed strategies. The mathematical details behind this "shift" are justified by von Neumann and Morgenstern's [26] axiomatisation of player preferences and utility functions. Consequently, von Neumann [25] answers Borel's [6] earlier question in the affirmative; that is, he shows that a minimax solution exists in mixed strategies for any two-player zero-sum game. The concept of minimising a player's maximum loss, despite its successful application to zero-sum environments, begins to present problems in non-zero-sum (or general-sum) games. Although a minimax solution to a zero-sum game is stable in the sense that neither player has an incentive to deviate, the same cannot be said for the general-sum case [17]. Instead, Nash [18] introduces the now-ubiquitous Nash equilibrium, which guarantees that a player cannot benefit from unilaterally deviating from their prescribed strategy. The existence of a Nash equilibrium is proved for any general-sum game with finitely many actions [18]. Note that, since the set of minimax solutions and set of equilibria belonging to a zero-sum game are identical, we will henceforth prefer the terminology "equilibrium" over "minimax solution".

What does an equilibrium tell us when game theory is applied to real-world situations? Well, the answer depends on the lens through which equilibria are viewed; there are two possible interpretations:

- **Descriptive**, the view that game theory predicts (human or non-human) strategies, and
- **Normative**, the view that game theory recommends "best" strategies [28].

Here, by approaching the task of "solving" a game as individually recommending strategies to its players, we will adopt a normative interpretation of game theory. This viewpoint is particularly successful in zero-sum settings because equilibria offer both stability, dissuading unilateral deviations, and security, guaranteeing a minimum reward [17]. The security of equilibria is not necessary in general-sum games and, as a consequence, different equilibria might confer different rewards. A normative approach encounters problems when needing to compare these rewards and nominate a "best" equilibrium [17].[1]

Of course, whenever our intention is to suggest suitable strategies to a player, we should always account for their ability to implement these recommendations. Larkey, Kadane, Austin, and Zamir [16] observe that, in traditional game theoretic

[1] A response to the difficulty of recommending "best" strategies in general-sum games is to modify the concept of an equilibrium. For example, the subgame perfect equilibrium is an equilibrium refinement and the correlated equilibrium is an equilibrium generalisation (see [17, Chapter 7, Chapter 8]).

discussions, “[t]he cognitive or physical difficulties for players in devising and executing strategies for playing particular games are essentially assumed away.” Then, seeking to identify the difficulties a player might encounter, they propose a typology of skill consisting of:

- **Strategic Skill**, the ability to select which games should be played,
- **Planning Skill**, the ability to develop a desirable strategy within a game, and
- **Execution Skill**, the ability to execute desired actions throughout a game.

Larkey et al. [16] apply their typology to experimentally compare strategies in “Sum Poker” under different skill limitations; however, a precise description of skill is not provided. We are only interested in further exploring execution skill because strategic skill is not critical in individual games and planning skill is not critical in normative game theory.

A mathematical framework that incorporates execution skill—or lack thereof—is provided by Beck and Filar [2] for zero-sum settings and Beck, Ejov, and Filar [4] for general-sum settings. Essentially, a notion of incompetence is introduced that quantifies a player’s tendency to accidentally deviate from their intended actions. The key advantage of incompetence, despite being superficially similar to Selten’s [22] concept of a “slight mistake” (or, as it has since been called, a trembling hand), is that it is expressly designed to describe execution skill. Conversely, the motivation behind Selten’s [22] discussion of accidental deviations is to define an equilibrium refinement. So, while a trembling hand involves players making mistakes with negligible probability, incompetence allows players to make mistakes according to arbitrary probability distributions.

The application of incompetence to military planning is discussed by Beck in [1] and [3]. Moreover, within the context of evolutionary biology, incompetence has been applied by Kleshnina, Filar, Ejov, and McKerral in [14] to study the behaviour and adaptation of species that are prone to mistakes. Kleshnina, Streipert, Filar, and Chatterjee [15] also study the optimal stepwise learning strategies in evolutionary games under incompetence.

Presently, our main objective is to further explore the nature of incompetence in zero-sum games from a normative perspective. Chapter 2 begins by reviewing the relevant mathematical concepts in traditional games and incompetent games. Recall that, in the earlier examples of a tennis player and a military force, a distinction was established between short-term and long-term strategies. The short-term strategic horizon, which involves optimally playing games under incompetence, is discussed in Chapter 3. This mainly investigates various properties of incompetence by building upon the already-established properties from [1] and [2]. Additionally, the long-term strategic horizon, which involves improving execution skill and modifying incompetence, is discussed in Chapter 4. We describe a model of incremental learning, develop a backward induction technique to compute equilibria, and apply this process to a simple tennis game from [1]. The overarching goal is that, by understanding the short-term and long-term strategic considerations, we can create better strategies to mitigate and reduce incompetence.

2 Game Theory

What is a game? We certainly cannot promise a single mathematical model that captures the entire diversity of real-world strategic interactions. Instead, games can be broadly described as “situations involving several decision makers with different goals, in which the decision of each affects the outcome for all the decision makers” [17]. Adopting the terminology of traditional games, these decision makers are called *players* and their available choices are called *actions*. It is assumed that players select actions with the intention of maximising the outcome-dependent reward or *utility* they receive.[1]

After a game-theoretic model has been formulated, we seek to solve it by finding player strategies that satisfy a *solution concept*, which captures salient properties of rational behaviour. Accordingly, this chapter reviews several well-established models and solution concepts that are encountered throughout our exploration of incompetence and learning. Section 2.1 and Section 2.2 explain introductory game theory (from [17], [19], and [20]) and Section 2.3 explains incompetent matrix games (from [1] and [2]).

2.1 Game Representations and Strategies

Unsurprisingly, some real-world strategic interactions share similarities with popular parlour games—for example, chess, backgammon, or poker. These games involve players sequentially selecting actions in pursuit of a desired outcome but could also include randomness (dice in backgammon) or private information (face-down cards in poker). A situation with these characteristics may be described using an *extensive-form game* where, within a suitably constructed directed tree, each vertex is assigned to a player who must select an incident arc to traverse. This sequence of choices traces a path through the tree until a terminal vertex is reached and utility is awarded. Randomness is incorporated by adding vertices where a lottery determines the traversed arc and private information is incorporated by partitioning a player’s decision vertices into *information sets*. Precisely, the players are only aware of the current information set and, from among the vertices in this information set, are unable to discern the exact current vertex.

Recall the simple two-player game “Rock, Paper, Scissors” in which both players simultaneously reveal a rock, paper, or scissors hand sign. Then, a winner is

[1] Although a complete discussion of utility theory is beyond the scope of this brief introduction, the development of utility functions from player preferences is explained in [17, Chapter 2]. Here, utility can simply be viewed as an abstract quantity resembling a monetary incentive.

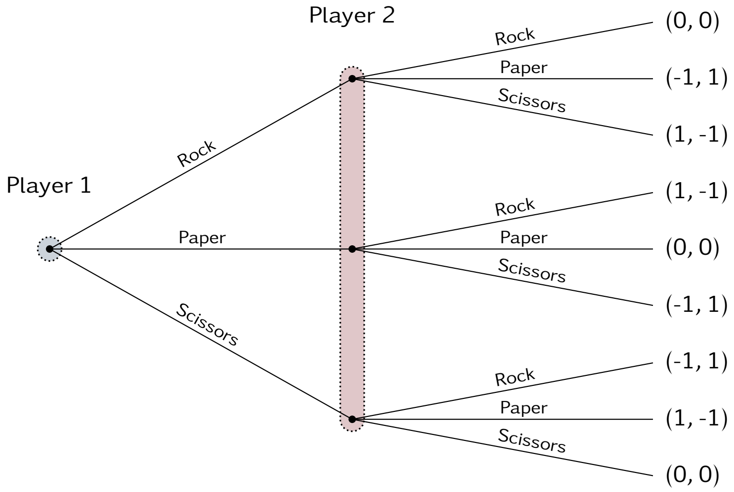


Figure 2.1. An extensive-form representation of "Rock, Paper, Scissors" [17].

determined by applying the dominance relationships: rock beats scissors, scissors beats paper, and paper beats rock. An extensive-form representation of this game is shown in Figure 2.1. The shaded regions represent information sets and a label (x, y) on a terminal vertex indicates that Player 1 receives utility x and Player 2 receives utility y. We imitate simultaneous action selection by combining the possible outcomes of Player 1's actions into a single information set such that Player 2 must act without knowing the choice that was made.

Already, from the example of "Rock, Paper, Scissors", it should be obvious that the complete definition of an extensive-form game quickly becomes cumbersome. A simplified model requires every player to implement a fixed strategy that will determine their behaviour throughout the game. Specifically, a *pure strategy* assigns unique actions to a player's information sets. Knowing that player behaviour is entirely determined by the implemented strategies, we may view the game as a simultaneous selection of strategies rather than a sequential selection of actions. This is captured using a *normal-form game* $G = (N, (S_k)_{k \in N}, (u_k)_{k \in N})$ with

- a set of players $N = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{Z}^+$,
- a set of actions S_k for each $k \in N$, and
- a utility function $u_k : S_k \to \mathbb{R}$ for each $k \in N$ where $S = S_1 \times S_2 \times \ldots, S_n$.

Here, every player $k \in N$ simultaneously chooses an action $s_k \in S_k$ to form an *action profile* (or *pure strategy profile*) $s = (s_1, s_2, \ldots, s_n) \in S$ containing the choices of all participants. The realisation of this action profile causes the player $k \in N$ to receive utility $u_k(s) = u_k(s_1, s_2, \ldots, s_n)$. Notice that the actions within the normal-form representation of an extensive-form game correspond to the available pure strategies.

A player wanting to behave unpredictably might employ a *mixed strategy*, which randomises among their pure strategies. The space of mixed strategies belonging to Player $k \in N$ is denoted by

$$\Delta_k = \left\{ \delta_k : S_k \to [0,1] : \sum_{s_k \in S_k} \delta_k(s_k) = 1 \right\}$$

and contains every probability distribution over the action set S_k.[2] If the mixed strategy $\delta_k \in \Delta_k$ is selected, then $\delta_k(s_k)$ is interpreted as the probability that Player $k \in N$ plays the action $s_k \in S_k$. Now, each player $k \in N$ can select a mixed strategy $\delta_k \in \Delta_k$ to form a *(mixed) strategy profile* $\delta \in \Delta$ where $\Delta = \Delta_1 \times \Delta_2 \times \ldots \times \Delta_n$. We say that a strategy profile $\delta \in \Delta$ is *completely mixed* whenever $\delta_k(s_k) > 0$ for all players $k \in N$ and actions $s_k \in S_k$.

How can a player value a strategy profile $\delta \in \Delta$ when the utility functions $u_1, u_2, \ldots, u_n$ are only defined on the set of action profiles S? Naturally, given any $k \in N$, we create an expected utility function $v_k : \Delta \to \mathbb{R}$ such that

$$v_k(\delta) = \sum_{s_1 \in S_1} \sum_{s_2 \in S_2} \cdots \sum_{s_n \in S_n} u_k(s_1, s_2, \ldots, s_n) \delta_1(s_1) \delta_2(s_2) \ldots \delta_n(s_n) \tag{2.1}$$

for all $\delta \in \Delta$. We say that $v_k(\delta)$ is the *value* of the strategy profile $\delta \in \Delta$ to Player $k \in N$.[3] Formally, the normal-form game $(N, (\Delta_k)_{k \in N}, (v_k)_{k \in N})$ is called the *mixed extension* of G but, because our players have access to their mixed strategies, we will use G itself to mean this mixed extension.

Intending to introduce incompetence to a narrower class of normal-form games in Section 2.3, we should clarify the notions of a finite game and a zero-sum game. We say that our normal-form game G is *finite* if, for every $k \in N$, the action set S_k is finite. We say that G is *zero-sum* when overall utility is conserved regardless of the game's outcome or, equivalently, whenever

$$\sum_{k \in N} u_k(s) = \sum_{k \in N} u_k(s_1, s_2, \ldots, s_n) = 0 \tag{2.2}$$

for all $s \in S$. It is straightforward to show from (2.1) and (2.2) that the mixed extension of a zero-sum game is also zero-sum.

The solution concept that we are generally interested in finding when working with a normal-form game is the *Nash equilibrium*. This is a collection of strategies for which no player would benefit from unilaterally deviating and adopting an alternative strategy. If $\delta \in \Delta$ is a strategy profile and $\delta_k \in \Delta_k$ is a strategy for Player $k \in N$, then (δ_k, δ_{-k}) denotes the strategy profile obtained by replacing the k^{th} entry of δ with δ_k. Using this notation to describe unilateral deviations, a Nash equilibrium is a profile $\delta^* = (\delta_1^*, \delta_2^*, \ldots, \delta_n^*) \in \Delta$ such that, for any player $k \in N$ and strategy $\delta_k \in \Delta(S_k)$, we have

$$v_k\left(\delta_k, \delta_{-k}^*\right) \leq v_k(\delta^*). \tag{2.3}$$

[2] This characterisation of mixed strategies, to avoid measure-theoretic complications, implicitly assumes that the action sets $S_1, S_2, \ldots, S_n$ are finite. An extension of this discussion to games with infinitely many actions can be found in [20, Chapter 4].

[3] The ability to consistently value mixed strategies via expected utility is a consequence of player preferences satisfying the von Neumann-Morgenstern axioms (see [17, Theorem 2.18]).

Nash [18] proved that an equilibrium always exists in a finite normal-form game. A useful tool when computing these equilibria is the *indifference principle*, which states that, under a mixed strategy equilibrium, any actions in S_k played with non-zero probability must yield equal expected utility to Player $k \in N$ (see, for instance, [17, Theorem 5.18]). If every equilibrium $\delta^* \in \Delta$ is completely mixed, then the game G is said to be *completely mixed*.

Figure 2.2 shows a normal-form representation of "Rock, Paper, Scissors" derived from the extensive-form game in Figure 2.1. The actions belonging to Player 1 are presented in rows, the actions belonging to Player 2 are presented in columns, and an entry (x, y) indicates that they receive utilities x and y, respectively. Aligning with the common usage of "Rock, Paper, Scissors" as a randomisation device, it is not a surprise that this game's single equilibrium—both players mixing uniformly over their actions—causes the players to win with equal probability.

Occasionally, the Nash equilibrium is not a suitable solution concept and a refinement is needed to further restrict the definition of rational behaviour. A useful refinement in extensive-form games is the *subgame perfect equilibrium*. This is a Nash equilibrium that remains resilient to unilateral deviations whenever it is restricted to an arbitrary *subgame*—a subtree that does not divide any information sets. The desirability of a subgame perfect equilibrium comes from its elimination of incredible threats or irrational strategies that attempt to dissuade opponents from particular actions. Schelling [21] gives an example of an incredible threat where

> "if I threaten to blow us both to bits unless you close the window, you know that I won't unless I have somehow managed to leave myself no choice in the matter."

We will always seek to eliminate these rational inconsistencies when solving games with sequential action selection.

		Player 2		
		Rock	Paper	Scissors
Player 1	Rock	0,0	−1,1	1,−1
	Paper	1,−1	0,0	−1,1
	Scissors	−1,1	1,−1	0,0

Figure 2.2. A normal-form representation of "Rock, Paper, Scissors" [17].

2.2 Matrix Games and Bimatrix Games

The tabular depiction of "Rock, Paper, Scissors" in Figure 2.2 correctly suggests that finite two-player normal-form games can be represented as matrices. Suppose G is a finite two-player normal-form game and, without loss of generality, take Player 1's action set to be $S_1 = A = \{a_1, a_2, \ldots, a_{m_1}\}$ and Player 2's action set to be $S_2 = B = \{b_1, b_2, \ldots, b_{m_2}\}$ for some $m_1, m_2 \in \mathbb{Z}^+$. The *utility matrices* $R_1 \in \mathbb{R}^{m_1 \times m_2}$ and $R_2 \in \mathbb{R}^{m_1 \times m_2}$ encode the utilities allocated to Player 1 and Player 2 for every possible combination of actions; that is,

$$R_1[i, j] = u_1(a_i, b_j) \quad \text{and} \quad R_2[i, j] = u_2(a_i, b_j)$$

		Player 2	
		Football	Concert
Player 1	Football	2,1	0,0
	Concert	0,0	1,2

Figure 2.3. A normal-form representation of "Battle of the Sexes".

for all $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$. We might think of G as a game wherein, after Player 1 chooses a row index $i = 1,2,\ldots,m_1$ and Player 2 chooses a column index $j = 1,2,\ldots,m_2$, they are awarded utilities $u_1(a_i,b_j)$ and $u_2(a_i,b_j)$, respectively. This interpretation allows us to write $u_1(i,j) = u_1(a_i,b_j)$ and $u_2(i,j) = u_2(a_i,b_j)$ for any $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$. A game admitting this representation is called an $m_1 \times m_2$ *bimatrix game*.

The mixed strategy spaces Δ_1 and Δ_2 contain probability distributions over the finite action sets A and B. It is convenient to represent these mixed strategies as stochastic row vectors from the sets

$$\mathbf{X} = \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_{m_1}) \in [0,1]^{m_1} : \sum_{i=1}^{m_1} x_i = 1 \right\}$$

and

$$\mathbf{Y} = \left\{ \mathbf{y} = (y_1, y_2, \ldots, y_{m_2}) \in [0,1]^{m_2} : \sum_{j=1}^{m_2} y_j = 1 \right\}.$$

If Player 1 chooses the mixed strategy $\mathbf{x} \in \mathbf{X}$ and Player 2 chooses the mixed strategy $\mathbf{y} \in \mathbf{Y}$, then the actions a_i and b_j are played with probability x_i and y_j, respectively. Adapting our expected utility functions $v_1 : \Delta \to \mathbb{R}$ and $v_2 : \Delta \to \mathbb{R}$ to the domain $\mathbf{X} \times \mathbf{Y}$, the value of these strategies to Player $k = 1,2$ is

$$v_k(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} x_i u_k(i,j) y_j = \mathbf{x} R_k \mathbf{y} \tag{2.4}$$

Next, by combining (2.3) and (2.4), observe that a strategy profile $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ is a Nash equilibrium in the bimatrix game G if and only if

$$\mathbf{x} R_1 (\mathbf{y}^*)^\intercal \leq \mathbf{x}^* R_1 (\mathbf{y}^*)^\intercal \quad \text{and} \quad \mathbf{x}^* R_2 \mathbf{y}^\intercal \leq \mathbf{x}^* R_2 (\mathbf{y}^*)^\intercal \tag{2.5}$$

given any deviations $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$.

A simple bimatrix game known as "Battle of the Sexes" is shown in Figure 2.3 and is represented by the utility matrices

$$R_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad R_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Here, without any communication, two players must individually decide between going to a football match or a concert. Although Player 1 prefers the football match and Player 2 prefers the concert, they must attend the same event to receive a non-zero utility reward. The three possible equilibrium solutions $(\mathbf{x}^*,\mathbf{y}^*)$, which each achieve different values of $\mathbf{v}(\mathbf{x}^*,\mathbf{y}^*) = (v_1(\mathbf{x}^*,\mathbf{y}^*), v_2(\mathbf{x}^*,\mathbf{y}^*))$, are:

- $\mathbf{x}^* = (1,0)$ and $\mathbf{y}^* = (1,0)$ with expected utility $\mathbf{v}(\mathbf{x}^*,\mathbf{y}^*) = (2,1)$,
- $\mathbf{x}^* = (0,1)$ and $\mathbf{y}^* = (0,1)$ with expected utility $\mathbf{v}(\mathbf{x}^*,\mathbf{y}^*) = (1,2)$, and
- $\mathbf{x}^* = (2/3,1/3)$ and $\mathbf{y}^* = (1/3,2/3)$ with expected utility $\mathbf{v}(\mathbf{x}^*,\mathbf{y}^*) = (2/3,2/3)$.

This demonstrates that a general bimatrix game might possess multiple Nash equilibria and that these do not necessarily award the same expected utilities.

Under the additional assumption that the bimatrix game G is zero-sum, we know that $R_1 = -R_2$ since $u_1(i,j)+u_2(i,j) = 0$ for all $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$. Hence, these utility allocations can be encoded in a single matrix $R \in \mathbb{R}^{m_1 \times m_2}$ where

$$R[i,j] = u_1(i,j) = -u_2(i,j)$$

for each $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$. A finite two-player zero-sum game G is called an $m_1 \times m_2$ *matrix game* and R is its *utility matrix*. We can describe a matrix game by giving a single utility function $u : A \times B \to \mathbb{R}$ where

$$u(a_i,b_j) = u(i,j) = u_1(i,j) = -u_2(i,j)$$

for all $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$. The expected utilities of the strategy profile $(\mathbf{x},\mathbf{y}) \in \mathbf{X} \times \mathbf{Y}$ are expressed by rewriting (2.4) as

$$v_1(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_i u_1(i,j) y_j = \mathbf{x}R\mathbf{y}^\intercal = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_i u_2(i,j) y_j = -v_2(\mathbf{x},\mathbf{y}). \tag{2.6}$$

This motivates our definition of a value function $v : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$ where $v(\mathbf{x},\mathbf{y}) = \mathbf{x}R\mathbf{y}^\intercal$ for all $(\mathbf{x},\mathbf{y}) \in \mathbf{X} \times \mathbf{Y}$. If Player 1 selects $\mathbf{x} \in \mathbf{X}$ and Player 2 selects $\mathbf{y} \in \mathbf{Y}$, then they expect to receive utilities $v(\mathbf{x},\mathbf{y})$ and $-v(\mathbf{x},\mathbf{y})$, respectively. We might view G as a game in which Player 1 chooses their strategy $\mathbf{x} \in \mathbf{X}$ to maximise $v(\mathbf{x},\mathbf{y}) = \mathbf{x}R\mathbf{y}^\intercal$ and Player 2 chooses their strategy $\mathbf{y} \in \mathbf{Y}$ to minimise $v(\mathbf{x},\mathbf{y}) = \mathbf{x}R\mathbf{y}^\intercal$. The equilibrium inequalities from (2.5) can be rearranged to show that the strategy profile $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ is an equilibrium if and only if

$$\mathbf{x}R(\mathbf{y}^*)^\intercal \le \mathbf{x}^*R(\mathbf{y}^*)^\intercal \le \mathbf{x}^*R\mathbf{y}^\intercal \tag{2.7}$$

for every $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$. The strategies $\mathbf{x}^*$ and $\mathbf{y}^*$ are called *optimal strategies* to emphasise their unique properties in zero-sum games. Namely, taking arbitrary equilibria $(\mathbf{x}^*,\mathbf{y}^*),(\mathbf{x}^\dagger,\mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$, we know that they have equal value $v(\mathbf{x}^*,\mathbf{y}^*) = v(\mathbf{x}^\dagger,\mathbf{y}^\dagger)$ and that $(\mathbf{x}^*,\mathbf{y}^\dagger)$ and $(\mathbf{x}^\dagger,\mathbf{y}^*)$ are also equilibria (see [20, Theorem 2.1.2]).[4] This common value shared among equilibria is called the *game value* of G and is denoted by $\mathrm{val}(G)$. The existence of optimal mixed strategies in matrix games is established by von Neumann's [25] minimax theorem, which proves the equality

$$\mathrm{val}(G) = \max_{\mathbf{x}\in\mathbf{X}} \min_{\mathbf{y}\in\mathbf{Y}} \mathbf{x}R\mathbf{y}^\intercal = \min_{\mathbf{y}\in\mathbf{Y}} \max_{\mathbf{x}\in\mathbf{X}} \mathbf{x}R\mathbf{y}^\intercal. \tag{2.8}$$

Accordingly, the equilibrium of a zero-sum game is also a *minimax solution* in which both players minimise their worst-case losses. We will reguarly need to find

[4] Generally, these properties cannot be extended to equilibrium solutions in bimatrix games. Recall that the equilibria in "Battle of the Sexes" did not yield the same expected utility and that their component strategies could not be interchanged to produce new equilibria.

optimal strategies and game values in matrix games, a task that is often achieved through linear programming. Specifically, if $\mathbf{x}^* \in \mathbf{X}$, $\mathbf{y}^* \in \mathbf{Y}$, and $\gamma \in \mathbb{R}$ solve the primal linear program

$$\begin{aligned} \text{maximise} \quad & \gamma \\ \text{subject to} \quad & \gamma - \sum_{i=1}^{m_1} u(i,j)x_i^* \leq 0, \quad j = 1,2,\ldots,m_2, \\ & \sum_{i=1}^{m_1} x_i^* = 1, \\ & x_i^* \geq 0, \quad i = 1,2,\ldots,m_1, \end{aligned} \tag{LP1}$$

and the dual linear program

$$\begin{aligned} \text{minimise} \quad & \gamma \\ \text{subject to} \quad & \gamma - \sum_{j=1}^{m_2} u(i,j)y_j^* \geq 0, \quad i = 1,2,\ldots,m_1, \\ & \sum_{j=1}^{m_2} y_j^* = 1, \\ & y_j^* \geq 0, \quad j = 1,2,\ldots,m_2, \end{aligned} \tag{LP2}$$

then $\mathbf{x}^*$ is an optimal strategy for Player 1, $\mathbf{y}^*$ is an optimal strategy for Player 2, and γ is the game value (see [20, Chapter 3]).

2.3 Incompetent Games

Beck and Filar [2] introduce incompetence to matrix games by allowing players to accidentally deviate from their proposed strategies.[5] They construct a pair of *incompetence matrices* $Q_1 \in \mathbb{R}^{m_1 \times m_1}$ and $Q_2 \in \mathbb{R}^{m_2 \times m_2}$ such that:

- $q_1(a_i, a_\alpha) = q_1(i,\alpha) = Q_1[i,\alpha]$ is the probability that Player 1 executes action a_α after selecting action a_i for all $i,\alpha = 1,2,\ldots,m_1$, and
- $q_2(b_j, b_\beta) = q_2(j,\beta) = Q_2[j,\beta]$ is the probability that Player 2 executes action b_β after selecting action b_j for all $j,\beta = 1,2,\ldots,m_2$.

If Player 1 chooses a_i and Player 2 chooses b_j for some $i = 1,2,\ldots,m_1$ and $j = 1,2,\ldots,m_2$, then the stochastic row vectors

$$\mathbf{q}_1(a_i) = \mathbf{q}_1(i) = (q_1(i,\alpha))_{\alpha=1}^{m_1} \quad \text{and} \quad \mathbf{q}_2(b_j) = \mathbf{q}_2(j) = (q_2(j,\beta))_{\beta=1}^{m_2}$$

are interpreted as probability distributions over the executable actions belonging to Player 1 and Player 2, respectively. What are the expected utilities of the action profile (a_i, b_j) under incompetence? Consider a function $u_{Q_1,Q_2} : A \times B \to \mathbb{R}$ where, by taking the probability-weighted sum over the possible action profiles, we have

$$u_{Q_1,Q_2}(a_i,b_j) = u_{Q_1,Q_2}(i,j) = \sum_{\alpha=1}^{m_1}\sum_{\beta=1}^{m_2} q_1(i,\alpha)u(\alpha,\beta)q_2(j,\beta) = \mathbf{q}_1(i)R\mathbf{q}_2(j)^{\mathsf{T}} \tag{2.9}$$

[5] We must, for the sake of brevity, only discuss incompetence in matrix games whose selectable and executable actions coincide. A broader definition of incompetence in matrix and bimatrix games can be found in [1], [4], and [2].

for all $i = 1, 2, \ldots, m_1$ and $j = 1, 2, \ldots, m_2$. Clearly, when accounting for the effects of incompetence, u_{Q_1,Q_2} becomes Player 1's utility function and $-u_{Q_1,Q_2}$ becomes Player 2's utility function. An *incompetent (matrix) game* G_{Q_1,Q_2} replaces the utility functions of the competent matrix game G with these incompetence-adjusted utility functions. A mixed strategy profile $(\mathbf{x}, \mathbf{y}) \in \mathbf{X} \times \mathbf{Y}$ in this incompetent game has value $v_{Q_1,Q_2}(\mathbf{x}, \mathbf{y})$ to Player 1 and value $-v_{Q_1,Q_2}(\mathbf{x}, \mathbf{y})$ to Player 2 where

$$v_{Q_1,Q_2}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} x_i u_{Q_1,Q_2}(i,j) y_j = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} x_i \mathbf{q}_1(i) R \mathbf{q}_2(j)^\intercal y_j = \mathbf{x} Q_1 R Q_2^\intercal \mathbf{y}^\intercal. \quad (2.10)$$

This shows that G_{Q_1,Q_2} is represented by the utility matrix $R_{Q_1,Q_2} = Q_1 R Q_2^\intercal$. Henceforth, whenever the choice of incompetence matrices is unambiguous, we substitute the subscript "Q" for "Q_1, Q_2", as in G_Q, u_Q, v_Q, and R_Q.

Primarily, we are interested in the behaviour of incompetent games under variations in their incompetence matrices. These variations are captured by Beck and Filar [2] using *learning trajectories*, which are functions that map from the interval $[0,1]$ to the set of $m \times m$ stochastic matrices for some $m \in \mathbb{Z}^+$. The learning trajectories $Q_1 : [0,1] \to \mathbb{R}^{m_1 \times m_1}$ and $Q_2 : [0,1] \to \mathbb{R}^{m_2 \times m_2}$ parameterise a family of incompetent games

$$\left\{ G_{\lambda,\mu} = G_{Q_1(\lambda), Q_2(\mu)} : \lambda, \mu \in [0,1] \right\}.$$

We shall refer to this family of games as a *parameterised incompetent (matrix) game* and write $G_{Q_1(\cdot),Q_2(\cdot)}$ or $G_{Q(\cdot)}$. Here, the expressions $Q_1(\cdot)$ and $Q_2(\cdot)$ serve as a reminder that these are learning trajectories, not incompetence matrices. Observe that, given any *learning parameters* $\lambda, \mu \in [0,1]$, the incompetent game $G_{\lambda,\mu}$ has $Q_1(\lambda)$ as Player 1's incompetence matrix, $Q_2(\mu)$ as Player 2's incompetence matrix, and $R_{\lambda,\mu} = Q_1(\lambda) R Q_2(\mu)^\intercal$ as its utility matrix.

Among the collection of $m \times m$ incompetence matrices for some $m \in \mathbb{Z}^+$, Beck and Filar [2] associate $1/m \cdot J_m$ with *uniform incompetence* and I_m with *complete competence* where J_m is the $m \times m$ all-one matrix and I_m is the $m \times m$ identity matrix. Intuitively, uniform incompetence causes a player to select actions uniformly at random and complete competence causes a player to select actions with absolute precision. They also restrict their discussion to linear learning trajectories $Q : [0,1] \to \mathbb{R}^{m \times m}$ satisfying

$$Q(\lambda) = Q(0)(1-\lambda) + Q(1)\lambda$$

for all $\lambda, \mu \in [0,1]$.[6]

Consider, for instance, a 2×2 matrix game G represented by the utility matrix $R \in \mathbb{R}^{2\times 2}$ where

$$R = \begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}.$$

We might introduce incompetence by assigning Player 1 the learning trajectory $Q_1 : [0,1] \to \mathbb{R}^{2\times 2}$ and Player 2 the learning trajectory $Q_2 : [0,1] \to \mathbb{R}^{2\times 2}$ where

$$Q_1(\lambda) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda \quad \text{and} \quad Q_2(\mu) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\mu$$

[6] An empirical exploration of several additional learning trajectories—namely, sigmoidal, exponential, power-law, and discontinuous learning trajectories—is provided in [1, Section 4.4].

for every $\lambda, \mu \in [0,1]$. Notice that, under these learning trajectories, both players transition linearly from uniform incompetence to complete competence. This produces a parameterised incompetent game $G_{Q(\cdot)}$ and, for all $\lambda, \mu \in [0,1]$, the incompetent game $G_{\lambda,\mu}$ is a matrix game represented by the utility matrix $R_{\lambda,\mu} = Q_1(\lambda)RQ_2(\mu)^{\mathsf{T}}$. If Player 1's learning parameter is $\lambda = 1$ and Player 2's learning parameter is $\mu = 0$, then the incompetent game $G_{1,0}$ has the utility matrix

$$R_{1,0} = Q_1(1)RQ_2(0)^{\mathsf{T}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 2 & 2 \end{pmatrix}.$$

Hence, $\mathbf{x}^* = (0,1)$ is an optimal strategy for Player 1, $\mathbf{y}^* = (q, 1-q)$ with $q \in [0,1]$ is an optimal strategy for Player 2, and the game value is $\operatorname{val}(G_{1,0}) = 2$. Alternatively, if Player 1's learning parameter is $\lambda = 0$ and Player 2's learning parameter is $\mu = 1$, then the utility matrix of $G_{0,1}$ is

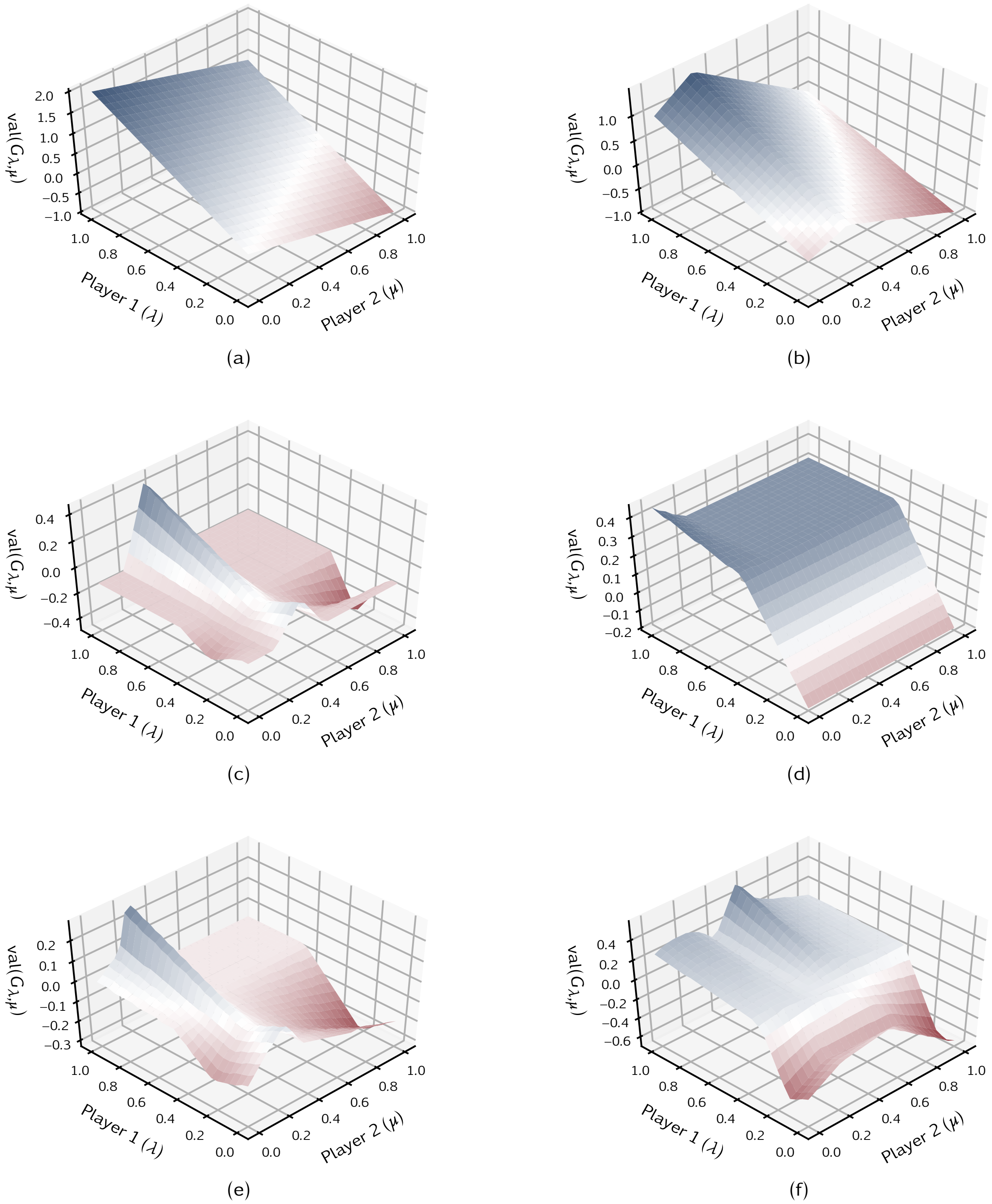
$$R_{0,1} = Q_1(0)RQ_2(1)^{\mathsf{T}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}.$$

This means that $\mathbf{x}^* = (p, 1-p)$ with $p \in [0,1]$ is an optimal strategy for Player 1, $\mathbf{y}^* = (0,1)$ is an optimal strategy for Player 2, and the game value is $\operatorname{val}(G_{0,1}) = 0$. The dependence of the game value $\operatorname{val}(G_{\lambda,\mu})$ on the learning parameters $\lambda, \mu \in [0,1]$ is shown in Figure 2.4(a).

Finally, to motivate our subsequent discussion of the variational properties of parameterised incompetent games, several additional examples have been compiled in Table 2.1 and Figure 2.4. A different $G_{Q(\cdot)}$ is produced for every combination of a utility matrix R, Player 1's learning trajectory $Q_1(\cdot)$, and Player 2's learning trajectory $Q_2(\cdot)$. We will further explore the features of these parameterised incompetent games in Chapter 3.

Table 2.1. The utility matrices (R) and learning trajectories (Q_1 and Q_2) that define a collection of parameterised incompetent games.

	R	$Q_1(\cdot)$	$Q_2(\cdot)$
Figure 2.4(a)	$\begin{pmatrix} -1 & -3 \\ 3 & 1 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda$	$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\mu$
Figure 2.4(b)	$\begin{pmatrix} 3 & 0 \\ -1 & -2 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda$	$\frac{1}{3}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\mu$
Figure 2.4(c)	$\begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}$	$\frac{1}{3}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda$	$\frac{1}{4}\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\mu$
Figure 2.4(d)	$\begin{pmatrix} 3 & -3 \\ -1 & 2 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}(1-\lambda) + \frac{1}{10}\begin{pmatrix} 9 & 1 \\ 1 & 9 \end{pmatrix}\lambda$	$\frac{1}{10}\begin{pmatrix} 2 & 8 \\ 5 & 5 \end{pmatrix}(1-\mu) + \frac{1}{5}\begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}\mu$
Figure 2.4(e)	$\begin{pmatrix} 0 & 3 & -2 \\ -2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$	$\frac{1}{7}\begin{pmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 1 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\lambda$	$\frac{1}{7}\begin{pmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 1 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\mu$
Figure 2.4(f)	$\begin{pmatrix} 2 & 3 & -1 \\ 0 & -3 & 0 \\ -3 & 0 & 2 \end{pmatrix}$	$\frac{1}{7}\begin{pmatrix} 1 & 3 & 3 \\ 2 & 2 & 3 \\ 1 & 4 & 2 \end{pmatrix}(1-\lambda) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\lambda$	$\frac{1}{7}\begin{pmatrix} 0 & 0 & 7 \\ 4 & 1 & 2 \\ 4 & 3 & 0 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\mu$



(a) (b) (c) (d) (e) (f)

Figure 2.4. The dependence of the game value val($G_{\lambda,\mu}$) on learning parameters $\lambda, \mu \in [0,1]$ for each parameterised incompetent game defined in Table 2.1. Generated using `incompetent_game_plot.py`.

3 A Strategic Perspective

A prominent feature in several of the game value graphs from Figure 2.4 is the presence of plateaus, or regions over which the game value is constant. Indeed, aligning with our observations, Beck [1] speculates that the game value might become constant as the players approach complete competence. Here, primarily motivated by these plateaus, we will display various properties of incompetent games by exploring incompetence from a strategic perspective.

The introduction of executed strategies and executable strategies in Section 3.1 establishes this "strategic perspective" of incompetence. This differs from previous approaches to incompetence in that it focuses on how a player's skill modifies their strategy space rather than their utility function. These concepts are used in Section 3.2 to explain the appearance of plateaus in the game value of parameterised incompetent games. Later, in Section 3.3, we use executable strategies to describe optimal learning parameter choices under certain conditions.

3.1 Executed Strategies

Consider a matrix game G and a corresponding incompetent game G_Q, with all definitions from Chapter 2 being reused. Recall that, in their introduction of incompetence to matrix games, Beck and Filar [2] distinguish between selectable and executable actions.[1] A player is required to select an action (called the selected action) that, after any accidental deviations occur, realises a potentially different action (called the executed action). We might say that, if Player 1 chooses a_i for some $i = 1, 2, \dots, m_1$ and Player 2 chooses b_j for some $j = 1, 2, \dots, m_2$, then $\mathbf{q}_1(i)$ and $\mathbf{q}_2(j)$ are the *executed strategies* realised by Player 1 and Player 2, respectively. This reflects the definition of $\mathbf{q}_1(i)$ and $\mathbf{q}_2(j)$ as the incompetence-induced probability distributions over each player's executable actions.

Similarly, we can extend the concept of an executed strategy to a player's entire mixed strategy space. Suppose $\mathbf{x}' \in \mathbf{X}$ is Player 1's selected strategy and $\mathbf{y}' \in \mathbf{Y}$ is Player 2's selected strategy. Define, for all $\alpha = 1, 2, \dots, m_1$ and $\beta = 1, 2, \dots, m_2$, the quantities

$$x_\alpha = \sum_{i=1}^{m_1} x'_i q_1(i, \alpha) = \mathbf{x}' Q_1 \mathbf{e}_\alpha^\mathsf{T} \quad \text{and} \quad y_\beta = \sum_{j=1}^{m_2} y'_j q_2(j, \beta) = \mathbf{y}' Q_2 \mathbf{e}_\beta^\mathsf{T} \tag{3.1}$$

[1] Unlike Beck and Filar [2], we require each player's set of selectable and executable strategies to coincide. Nevertheless, with some additional notation to distinguish between selectable and executable strategies, it is straightforward to adapt our conversation to their broader setting.

where $\mathbf{e}_\alpha \in \mathbb{R}^{1\times m_1}$ and $\mathbf{e}_\beta \in \mathbb{R}^{1\times m_2}$ are the standard basis vectors in the α^{th} and β^{th} directions. We interpret x_α as being the probability that Player 1 executes action a_α and y_β as being the probability that Player 2 executes action b_β. These probabilities are compiled to form the *executed strategies* $\mathbf{x} = (x_1, x_2, \dots, x_{m_1})$ and $\mathbf{y} = (y_1, y_2, \dots, y_{m_2})$, which can be expressed as

$$\mathbf{x} = \mathbf{x}'Q_1 \quad \text{and} \quad \mathbf{y} = \mathbf{y}'Q_2. \tag{3.2}$$

Hence, given that the strategy profile $(\mathbf{x}', \mathbf{y}') \in \mathbf{X} \times \mathbf{Y}$ has been selected, we know that the corresponding strategy profile $(\mathbf{x}'Q_1, \mathbf{y}'Q_2)$ is executed.

What strategies are the players able to execute? We will say that $\mathbf{x} \in \mathbf{X}$ is an *executable strategy* for Player 1 (under the incompetence matrix Q_1) whenever there exists $\mathbf{x}' \in \mathbf{X}$ satisfying $\mathbf{x} = \mathbf{x}'Q_1$. Analogously, we will say that $\mathbf{y} \in \mathbf{Y}$ is an *executable strategy* for Player 2 (under the incompetence matrix Q_2) whenever there exists $\mathbf{y}' \in \mathbf{Y}$ satisfying $\mathbf{y} = \mathbf{y}'Q_2$. The set of executable strategies belonging to Player 1 is

$$\mathbf{E}_1(Q_1) = \{\mathbf{x} \in \mathbf{X} : \mathbf{x} = \mathbf{x}'Q_1 \text{ for some } \mathbf{x}' \in \mathbf{X}\} \tag{3.3}$$

and the set of executable strategies belonging to Player 2 is

$$\mathbf{E}_2(Q_2) = \{\mathbf{y} \in \mathbf{Y} : \mathbf{y} = \mathbf{y}'Q_2 \text{ for some } \mathbf{y}' \in \mathbf{Y}\}. \tag{3.4}$$

If the choice of incompetence matrices is unambiguous, then it is often convenient to write $\mathbf{E}_1$ instead of $\mathbf{E}_1(Q_1)$ and $\mathbf{E}_2$ instead of $\mathbf{E}_2(Q_2)$. Moreover, provided a pair of learning trajectories $Q_1 : [0,1] \to \mathbb{R}^{m_1\times m_1}$ and $Q_2 : [0,1] \to \mathbb{R}^{m_2\times m_2}$, we will replace $\mathbf{E}_1(Q_1(\lambda))$ with $\mathbf{E}_1(\lambda)$ and $\mathbf{E}_2(Q_2(\mu))$ with $\mathbf{E}_2(\mu)$.

Next, to characterise the geometry of the spaces of executable strategies $\mathbf{E}_1$ and $\mathbf{E}_2$, observe that the strategies $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$ are executable if and only if

$$\mathbf{x} = \mathbf{x}'Q_1 = \sum_{i=1}^{m_1} x_i' \mathbf{q}_1(i) \quad \text{and} \quad \mathbf{y} = \mathbf{y}'Q_2 = \sum_{j=1}^{m_2} y_j' \mathbf{q}_2(j)$$

for some $\mathbf{x}' \in \mathbf{X}$ and $\mathbf{y}' \in \mathbf{Y}$. This means that $\mathbf{x}$ is a convex combination of the row vectors $\mathbf{q}_1(1), \mathbf{q}_1(2), \dots, \mathbf{q}_1(m_1)$ of Q_1 and $\mathbf{y}$ is a convex combination of the row vectors $\mathbf{q}_2(1), \mathbf{q}_2(2), \dots, \mathbf{q}_2(m_2)$ of Q_2. So, Player 1's set of executable strategies $\mathbf{E}_1$ is the convex hull of $\{\mathbf{q}_1(i) : i = 1, 2, \dots, m_1\}$ and Player 2's set of executable strategies $\mathbf{E}_2$ is the convex hull of $\{\mathbf{q}_2(j) : j = 1, 2, \dots, m_2\}$.

Suppose a game of "Rock, Paper, Scissors" (shown in Figure 2.1 and Figure 2.2) is being played between a pair of players with incompetence matrices

$$Q_1 = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \\ 1/3 & 2/3 & 0 \end{pmatrix}.$$

The executable strategy spaces $\mathbf{E}_1$ and $\mathbf{E}_2$ are the convex hulls of the rows within the matrices Q_1 and Q_2, respectively. These convex hulls are shown as shaded regions in Figure 3.1 and, in response to our previous question, we see that there are some strategies that neither player can execute.

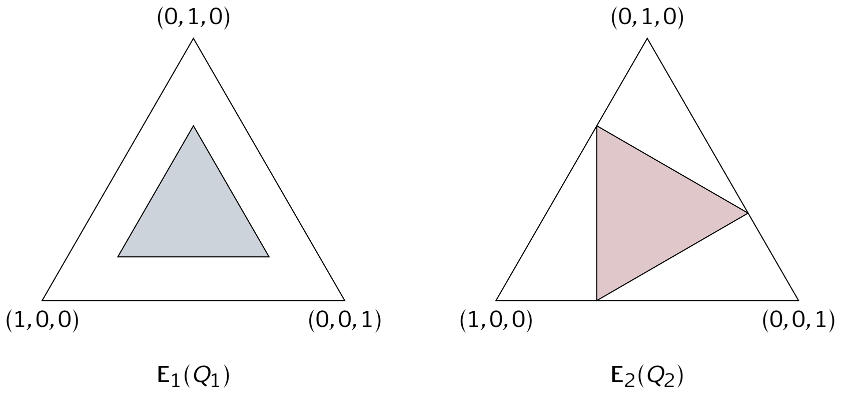


Figure 3.1. The executable strategy spaces belonging to Player 1 (blue) and Player 2 (red) under incompetence in "Rock, Paper, Scissors".

3.2 Game Value Plateaus

Now, we want to apply the concepts of executed and executable strategies to explain the presence of plateaus in Figure 2.4. Beck [1, Theorem 3.3] notes that, if a pair of completely mixed incompetent games are derived from the same matrix game, then their game values are equal. We will further explore this observation by showing that rectangular plateaus emerge when, despite their limited skills, players are able to execute their competent optimal strategies.

Again, reusing the notation introduced in Chapter 2, consider a matrix game G and an associated incompetent game G_Q. Our initial goal is to connect the equilibria of G and G_Q through the lens of executed strategies. This leads us to question when an equilibrium $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X}\times\mathbf{Y}$ of G_Q produces an equilibrium $(\mathbf{x}^*Q_1,\mathbf{y}^*Q_2)$ of G, and vice versa. Below, Proposition 3.1 answers the "backward direction" of this question by showing that, given an equilibrium $(\mathbf{x}^*Q_1,\mathbf{y}^*Q_2)$ of G for some $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X}\times\mathbf{Y}$, the strategy profile $(\mathbf{x}^*,\mathbf{y}^*)$ is always an equilibrium of G_Q.

Proposition 3.1. Consider a strategy profile $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X}\times\mathbf{Y}$ such that $(\mathbf{x}^*Q_1,\mathbf{y}^*Q_2)$ is an equilibrium of the competent game G. Then, $(\mathbf{x}^*,\mathbf{y}^*)$ is an equilibrium of the incompetent game G_Q.

Proof. We want to prove that neither player possesses a profitable unilateral deviation from the strategy profile $(\mathbf{x}^*,\mathbf{y}^*)$ in G_Q. First, to the contrary, assume that Player 1 has a strategy $\mathbf{x} \in \mathbf{X}$ such that $v_Q(\mathbf{x},\mathbf{y}^*) > v_Q(\mathbf{x}^*,\mathbf{y}^*)$ and $\mathbf{x}R_Q(\mathbf{y}^*)^\top > \mathbf{x}^*R_Q(\mathbf{y}^*)^\top$. Observe that

$$\begin{aligned} v(\mathbf{x}^*Q_1,\mathbf{y}^*Q_2) &= \mathbf{x}^*Q_1RQ_2^\top(\mathbf{y}^*)^\top = \mathbf{x}^*R_Q(\mathbf{y}^*)^\top \\ &< \mathbf{x}R_Q(\mathbf{y}^*)^\top = \mathbf{x}Q_1RQ_2^\top(\mathbf{y}^*)^\top = v(\mathbf{x}Q_1,\mathbf{y}^*Q_2). \end{aligned} \tag{i}$$

This contradicts the equilibrium inequalities for $(\mathbf{x}^*Q_1,\mathbf{y}^*Q_2)$ in G and, as a consequence, we must have $v_Q(\mathbf{x},\mathbf{y}^*) \leq v_Q(\mathbf{x}^*,\mathbf{y}^*)$ for all $\mathbf{x} \in \mathbf{X}$. Second, assuming that Player 2 possesses a strategy $\mathbf{y} \in \mathbf{Y}$ such that $v_Q(\mathbf{x}^*,\mathbf{y}) < v_Q(\mathbf{x}^*,\mathbf{y}^*)$ and $\mathbf{x}^*R_Q\mathbf{y}^\top <$

$\mathbf{x}^* R_Q (\mathbf{y}^*)^\top$, we obtain

$$\begin{aligned} v(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2) &= \mathbf{x}^* Q_1 R Q_2^\top (\mathbf{y}^*)^\top = \mathbf{x}^* R_Q (\mathbf{y}^*)^\top \\ &> \mathbf{x}^* R_Q \mathbf{y}^\top = \mathbf{x}^* Q_1 R Q_2^\top \mathbf{y}^\top = v(\mathbf{x}^* Q_1, \mathbf{y}, Q_2). \end{aligned} \tag{ii}$$

Again, this contradicts the equilibrium inequalities for $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ in G and allows us to conclude that $v_Q(\mathbf{x}^*, \mathbf{y}) \geq v_Q(\mathbf{x}^*, \mathbf{y}^*)$ for all $\mathbf{y} \in \mathbf{Y}$. Therefore, having established the necessary inequalities, we see that $(\mathbf{x}^*, \mathbf{y}^*)$ is an equilibrium of the incompetent game G_Q. It is straightforward to show that $\mathrm{val}(G) = \mathrm{val}(G_Q)$ by writing

$$\mathrm{val}(G) = v(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2) = \mathbf{x}^* Q_1 R Q_2^\top (\mathbf{y}^*)^\top = \mathbf{x}^* R_Q (\mathbf{y}^*)^\top = v_Q(\mathbf{x}^*, \mathbf{y}^*) = \mathrm{val}(G_Q), \tag{iii}$$

as required. □

A more interesting problem is encountered when answering the "forward" direction of our previous question; that is, when finding the conditions under which an equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$ of G_Q executes an equilibrium $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ of G. We may be unable to write every strategy profile $(\mathbf{x}, \mathbf{y}) \in \mathbf{X} \times \mathbf{Y}$ in the form $(\mathbf{x}' Q_1, \mathbf{y}' Q_2)$ for some $\mathbf{x}' \in \mathbf{X}$ and $\mathbf{y}' \in \mathbf{Y}$. This means that the constituent strategies of $(\mathbf{x}, \mathbf{y})$ cannot be executed, with either $\mathbf{x} \notin \mathbf{E}_1$ or $\mathbf{y} \notin \mathbf{E}_2$. Accordingly, whenever $\mathbf{E}_1 \subset \mathbf{X}$ or $\mathbf{E}_2 \subset \mathbf{Y}$ hold as strict inclusions, the equilibrium inequalities for $(\mathbf{x}^*, \mathbf{y}^*)$ in G_Q do not disprove the existence of a unilateral deviation from $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ in G.

Consider, for instance, a simple 2 × 2 matrix game G called "Matching Pennies", which has the utility matrix

$$R = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The players, who are each given their own penny, must simultaneously place these pennies to show either heads or tails. Player 1 is declared the winner if the face-up sides match and Player 2 is declared the winner if the face-up sides do not match. The game value of G is $\mathrm{val}(G) = 0$ and its unique equilibrium $(\mathbf{x}^\dagger, \mathbf{y}^\dagger)$ has $\mathbf{x}^\dagger = (1/2, 1/2)$ and $\mathbf{y}^\dagger = (1/2, 1/2)$. So, optimal play involves both players mixing uniformly at random between heads or tails. Suppose Player 1 and Player 2 are assigned the incompetence matrices

$$Q_1 = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 3/4 & 1/4 \\ 2/3 & 1/3 \end{pmatrix},$$

respectively. The resulting incompetent game G_Q is represented by the utility matrix

$$R_Q = Q_1 R Q_2^\top = \begin{pmatrix} -1/6 & -1/9 \\ -1/4 & -1/6 \end{pmatrix}$$

and has a unique equilibrium solution $(\mathbf{x}^*, \mathbf{y}^*)$ with $\mathbf{x}^* = (1, 0)$ and $\mathbf{y}^* = (1, 0)$. Clearly, under these incompetence matrices, optimal play requires that both players attempt to place their pennies showing heads. This implies that the strategies $\mathbf{x}^* Q_1 = (1/3, 2/3)$ and $\mathbf{y}^* Q_2 = (3/4, 1/4)$ do not form an equilibrium in G. Additionally, since

$$\mathbf{x}^\dagger Q_1^{-1} = (3, -2) \notin \mathbf{X} \quad \text{and} \quad \mathbf{y}^\dagger Q_2^{-1} = (-2, 3) \notin \mathbf{Y},$$

the unique optimal strategy $\mathbf{x}^\dagger$ in G belonging to Player 1 is not executable under Q_1 and the unique optimal strategy $\mathbf{y}^\dagger$ in G belonging to Player 2 is not executable under Q_2.

Shortly, we will provide a sufficient condition under which an equilibrium of G_Q executes an equilibrium of G. Specifically, Theorem 3.3 states that, given a completely mixed incompetent game G_Q, every equilibrium $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ of G_Q produces an equilibrium $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ of G.[2] Although this result is similar to both [1, Theorem 3.3] and [2, Lemma 4.2], it differs in that it describes not only the game value under optimal play but also the strategies capable of achieving optimality.

Lemma 3.2. If G_Q is completely mixed, then Q_1 and Q_2 are non-singular.

Proof. Here, to argue by contraposition, consider an arbitrary incompetent game G_Q. Additionally, we will assume that Player 1's incompetence matrix Q_1 is singular such that, for some distinct indices $I \subseteq \{1, 2, \ldots, m_1\}$, the row vectors $\{\mathbf{q}_1(i) : i \in I\}$ are linearly dependent. So, there exists non-zero coefficients $\{\theta_i \in \mathbb{R} : i \in I\}$ satisfying

$$\sum_{i \in I} \theta_i \mathbf{q}_1(i) = 0. \tag{i}$$

Let $I^+ = \{i \in I : \theta_i > 0\}$ denote the set of indices corresponding to positive coefficients and $I^- = \{i \in I : \theta_i < 0\}$ denote the set of indices corresponding to negative coefficients. Note that, since the rows of Q_1 are stochastic vectors, both positive and negative coefficients must be present, or $I^+ \neq \varnothing$ and $I^- \neq \varnothing$. We obtain

$$\sum_{i \in I^+} \theta_i \mathbf{q}_1(i) = -\sum_{i \in I^-} \theta_i \mathbf{q}_1(i). \tag{ii}$$

after rearranging the equality in (i). Next, to identify a relationship between the coefficients, observe that

$$\left(\sum_{i \in I^+} \theta_i \mathbf{q}_1(i) \right) \mathbf{1}_{m_1}^\top = -\left(\sum_{i \in I^-} \theta_i \mathbf{q}_1(i) \right) \mathbf{1}_{m_1}^\top \quad \text{implies} \quad \sum_{i \in I^+} \theta_i = -\sum_{i \in I^-} \theta_i \tag{iii}$$

where $\mathbf{1}_{m_1}^\top \in \mathbb{R}^{1 \times m_1}$ is the $1 \times m_1$ all-ones row vector. Of course, we can rescale these coefficients to ensure that both the left-hand side and right-hand side of (iii) sum to unity. This allows us to interpret the vectors $(\theta_i : i \in I^+)$ and $(-\theta_i : i \in I^-)$ as probability distributions over the sets of actions $\{a_i : i \in I^+\}$ and $\{a_i : i \in I^-\}$, respectively.

Let $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ be an equilibrium of the incompetent game G_Q. Moreover, define a constant $\alpha \in \mathbb{R}^+$ such that $x_i^* + \alpha \theta_i \geq 0$ for all $i \in I^-$ and $x_{i'}^* + \alpha \theta_{i'} = 0$ for some $i' \in I^-$. A suitable choice for this constant is

$$\alpha = \max_{i \in I^-} \left(-\frac{x_i^*}{\theta_i} \right). \tag{iv}$$

We can construct an alternative strategy $\mathbf{x}^\dagger \in \mathbf{X}$, which is not completely mixed, where

$$x_i^\dagger = \begin{cases} x_i^* + \alpha \theta_i, & i \in I, \\ x_i^*, & i \notin I, \end{cases} \tag{v}$$

[2] Implicitly, in the assumption that G_Q is completely mixed, we are insisting that both players possess the same number of actions, or $m_1 = m_2$ (see [13, Theorem 3]).

for all $i = 1, 2, \ldots, m_1$. Certainly, this is a stochastic vector because

$$\sum_{i=1}^{m_1} x_i^\dagger = \sum_{i \notin I} x_i^* + \sum_{i \in I} (x_i^* + \alpha\theta_i) = \sum_{i=1}^{m_1} x_i^* + \alpha \sum_{i \in I} \theta_i = \sum_{i=1}^{m_1} x_i^* = 1 \tag{vi}$$

and $x_i^\dagger \geq 0$ for every $i = 1, 2, \ldots, m_1$. Finally, to prove that $\mathbf{x}^\dagger$ is an optimal strategy for Player 1 in G_Q, it suffices to show that $\mathbf{x}^*$ and $\mathbf{x}^\dagger$ deliver equal expected utility regardless of Player 2's selected strategy. Fix an arbitrary strategy $\mathbf{y} \in \mathbf{Y}$ and write

$$\begin{aligned} v_Q(\mathbf{x}^\dagger, \mathbf{y}) &= \mathbf{x}^\dagger R_Q \mathbf{y}^\top = \mathbf{x}^\dagger Q_1 R Q_2^\top \mathbf{y}^\top \\ &= \sum_{i=1}^{m_1} x_i^\dagger \mathbf{q}_1(i) R Q_2^\top \mathbf{y}^\top = \left(\sum_{i \notin I} x_i^* \mathbf{q}_1(i) + \sum_{i \in I} (x_i^* + \alpha\theta_i) \mathbf{q}_1(i) \right) R Q_2^\top \mathbf{y}^\top \\ &= \left(\sum_{i=1}^{m_1} x_i^* \mathbf{q}_1(i) \right) R Q_2^\top \mathbf{y}^\top + \alpha \left(\sum_{i \in I} \theta_i \mathbf{q}_1(i) \right) R Q_2^\top \mathbf{y}^\top \\ &= \mathbf{x}^* Q_1 R Q_2^\top \mathbf{y}^\top = \mathbf{x}^* R_Q \mathbf{y}^\top = v_Q(\mathbf{x}^*, \mathbf{y}). \end{aligned} \tag{vii}$$

Hence, from an expected utility perspective, the strategies $\mathbf{x}^*$ and $\mathbf{x}^\dagger$ behave identically. This shows that $\mathbf{x}^\dagger$ is an optimal strategy for Player 1 in G_Q and, as a result, the incompetent game G_Q is not completely mixed.

We can conclude that, whenever G_Q is completely mixed, Player 1's incompetence matrix Q_1 is non-singular. Similarly, by an analagous argument, Player 2's incompetence matrix Q_2 will also be non-singular. □

Note that, because Q_1 and Q_2 are non-singular whenever G_Q is completely mixed, the set of Player 1's executable strategies $\mathbf{E}_1$ is m_1-dimensional and the set of Player 2's executable strategies $\mathbf{E}_2$ is m_2-dimensional. Furthermore, the row vectors $\{\mathbf{q}_1(i) : 1, 2, \ldots, m_1\}$ are extreme points of $\mathbf{E}_1$ and the row vectors $\{\mathbf{q}_2(j) : j = 1, 2, \ldots, m_2\}$ are extreme points of Q_2. Why? Certainly, if a row could be expressed as a convex combination of other rows, then these rows would be linearly dependent and the associated incompetence matrix would be singular.

Theorem 3.3. *Consider an equilibrium* $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ *of a completely mixed incompetent game* G_Q*. The strategy profile* $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ *is an equilibrium of the competent game* G *and* $\mathrm{val}(G_Q) = \mathrm{val}(G)$.

Proof. We will apply the indifference principle to the equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$ of G_Q. This states that Player 1's actions award equal expected utility γ_1 when Player 2 chooses $\mathbf{y}^*$ and Player 2's actions award equal expected utility γ_2 when Player 1 chooses $\mathbf{x}^*$. Mathematically, this is expressed in the equalities

$$R_Q(\mathbf{y}^*)^\top = Q_1 R Q_2^\top (\mathbf{y}^*)^\top = \gamma_1 \mathbf{1}_{m_1}^\top \quad \text{and} \quad \mathbf{x}^* R_Q = \mathbf{x}^* Q_1 R Q_2^\top = \gamma_2 \mathbf{1}_{m_2} \tag{i}$$

where $\gamma_1 = -\gamma_2 = \mathrm{val}(G)$ and $\mathbf{1}_{m_1} \in \mathbb{R}^{1 \times m_1}$ and $\mathbf{1}_{m_2} \in \mathbb{R}^{1 \times m_2}$ are all-ones row vectors. Given that Q_1 and Q_2 are stochastic matrices, setting $R Q_2^\top (\mathbf{y}^*)^\top = \gamma_1 \mathbf{1}_{m_1}^\top$ and $\mathbf{x}^* Q_1 R = \gamma_2 \mathbf{1}_{m_2}$ gives

$$Q_1 R Q_2^\top (\mathbf{y}^*)^\top = \gamma_1 Q_1 \mathbf{1}_{m_1}^\top = \gamma_1 \mathbf{1}_{m_1}^\top \quad \text{and} \quad \mathbf{x}^* Q_1 R Q_2^\top = \gamma_2 (Q_2 \mathbf{1}_{m_2}^\top)^\top = \gamma_2 \mathbf{1}_{m_2}. \tag{ii}$$

The uniqueness of these solutions is guaranteed by our finding in Lemma 3.2 that Q_1 and Q_2 are non-singular. Thus, we must have $RQ_2^\intercal(\mathbf{y}^*)^\intercal = \gamma_1 \mathbf{1}_{m_1}^\intercal$ and $\mathbf{x}^* Q_1 R = \gamma_2 \mathbf{1}_{m_2}$. We interpret these equalities as saying that, in the competent game G, Player 1 is indifferent between their actions when Player 2 selects $\mathbf{y}^* Q_2$ and Player 2 is indifferent between their actions when Player 1 selects $\mathbf{x}^* Q_1$. Evidently, by way of the indifference principle, this proves that $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ is an equilibrium of G and

$$\mathrm{val}(G) = v(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2) = \mathbf{x}^* Q_1 R Q_2^\intercal (\mathbf{y}^*)^\intercal = \mathbf{x}^* R_Q (\mathbf{y}^*)^\intercal = v_Q(\mathbf{x}^*, \mathbf{y}^*) = \mathrm{val}(G_Q), \tag{iii}$$

as desired. □

Briefly, shifting to the context of a parameterised incompetent game $G_{Q(\,\cdot\,)}$ with learning trajectories $Q_1 : [0,1] \to \mathbb{R}^{m_1 \times m_1}$ and $Q_2 : [0,1] \to \mathbb{R}^{m_2 \times m_2}$, define the set

$$\mathcal{C} = \Big\{ (\lambda, \mu) \in [0,1] \times [0,1] : G_{\lambda,\mu} \text{ is completely mixed} \Big\} \tag{3.5}$$

of learning parameter pairs. Theorem 3.3 tells us that the game value of $G_{Q(\,\cdot\,)}$ is constant on $\mathcal{C}$; that is, we have $\mathrm{val}(G_{\lambda,\mu}) = \mathrm{val}(G)$ for all $(\lambda, \mu) \in \mathcal{C}$. Our immediate task is to prove that, when $\mathcal{C}$ is non-empty, it constitutes the unions of rectangular plateaus present in Figure 2.4. Below, the rectangular structure is established in Proposition 3.5 and the plateauing behaviour is established in Proposition 3.6.

Lemma 3.4. If G_Q is completely mixed, then G is also completely mixed.

Proof. We know that the completely mixed game G_Q has a unique equilibrium $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ (see [13, Theorem 2]). Assume that, in pursuit of a contradiction, the corresponding competent game G possesses an equilibrium $(\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ that is not completely mixed.

First, having shown that the rows of Q_1 and Q_2 are linearly independent in Lemma 3.2, the executed strategies $\mathbf{x}^* Q_1 \in \mathbf{E}_1$ and $\mathbf{y}^* Q_2 \in \mathbf{E}_2$ are strictly positive convex combinations of the extreme points $\{\mathbf{q}_1(i) : i = 1, 2, \ldots, m_1\}$ and $\{\mathbf{q}_2(j) : j = 1, 2, \ldots, m_2\}$, respectively. Therefore, the interior of $\mathbf{E}_1$ includes $\mathbf{x}^* Q_1$ (or $\mathbf{x}^* Q_1 \in \operatorname{Int} \mathbf{E}_1$) and the interior of $\mathbf{E}_2$ includes $\mathbf{y}^* Q_2$ (or $\mathbf{y}^* Q_2 \in \operatorname{Int} \mathbf{E}_2$) by [24, Corollary 4.19].

Second, recall that $(\mathbf{x}^\dagger, \mathbf{y}^\dagger)$ is not a completely mixed strategy profile and, as a consequence, a constituent strategy must lie on the boundary of its associated strategy space: either $\mathbf{x}^\dagger \in \partial \mathbf{X}$ or $\mathbf{y}^\dagger \in \partial \mathbf{Y}$. Assume, without loss of generality, that $\mathbf{x}^\dagger$ is not completely mixed and $\mathbf{x}^\dagger \in \partial \mathbf{X}$. Given that the set of executable strategies $\mathbf{E}_1$ is a subset of the entire strategy space $\mathbf{X}$, we have $\operatorname{Int} \mathbf{E}_1 \cap \partial \mathbf{X} = \varnothing$ and $\mathbf{x}^\dagger \neq \mathbf{x}^* Q_1$. A strategy $\mathbf{x}^{\dagger\dagger} \in \mathbf{X}$ exists such that

$$\mathbf{x}^{\dagger\dagger} = \alpha \mathbf{x}^* Q_1 + (1 - \alpha) \mathbf{x}^\dagger \in \operatorname{Int} \mathbf{E}_1 \tag{i}$$

for some $\alpha \in (0,1)$ (see [24, Theorem 3.26]). Additionally, given the executability of $\mathbf{x}^{\dagger\dagger} \in \operatorname{Int} \mathbf{E}_1$, there exists $\mathbf{x}^{**} \in \mathbf{X}$ satisfying the equality $\mathbf{x}^{\dagger\dagger} = \mathbf{x}^{**} Q_1$. We know that $\mathbf{x}^{\dagger\dagger}$ is an optimal strategy in G (by closure under convex combinations, as in [10, Proposition G.5]) and that $\mathbf{x}^{**}$ is an optimal strategy in G_Q (by Proposition 3.1). This contradicts the uniqueness of the equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$ in G_Q because the distinct strategies $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are both supposedly optimal for Player 1.

Finally, after applying an analagaous argument when $\mathbf{y}^\dagger \in \partial \mathbf{Y}$, we find that G is always completely mixed whenever G_Q is completely mixed. □

Proposition 3.5. Suppose that, for some incompetence matrices $Q_1, T_1 \in \mathbb{R}^{m_1 \times m_1}$ and $Q_2, T_2 \in \mathbb{R}^{m_2 \times m_2}$, the incompetent games G_{Q_1,Q_2} and G_{T_1,T_2} are completely mixed. Then, G_{Q_1,T_2} and G_{T_1,Q_2} are also completely mixed.

Proof. Here, without loss of generality, assume that G has a non-zero game value.[3] Consider a pair of equilibria $(\mathbf{x}^*, \mathbf{y}^*), (\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ belonging to the incompetent games G_{Q_1,Q_2} and G_{T_1,T_2}, respectively. Theorem 3.3 tells us that the executed strategies $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ and $(\mathbf{x}^\dagger T_1, \mathbf{y}^\dagger T_2)$ are equilibria of G. Thus, $(\mathbf{x}^* Q_1, \mathbf{y}^\dagger T_2)$ is also an equilibrium of G and, after applying Proposition 3.1, we find that $(\mathbf{x}^*, \mathbf{y}^\dagger)$ is an equilibrium of G_{Q_1,T_2}. We want to prove the uniqueness of this equilibrium.

Note that the incompetence matrices Q_1 and T_2 are non-singular (by Lemma 3.2) and the utility matrix R is non-singular (by Lemma 3.4 and [13]). Fix an arbitrary equilibrium $(\mathbf{x}^\diamond, \mathbf{y}^\diamond) \in \mathbf{X} \times \mathbf{Y}$ of G_{Q_1,T_2}. Kaplansky [13, Theorem 1] states that, because both players possess completely mixed optimal strategies, we have

$$R_{Q_1,T_2} (\mathbf{y}^\diamond)^\intercal = Q_1 R T_2^\intercal (\mathbf{y}^\diamond)^\intercal = \gamma_1 \mathbf{1}_{m_1}^\intercal \quad \text{and} \quad \mathbf{x}^\diamond R_{Q_1,T_2} = \mathbf{x}^\diamond Q_1 R T_2^\intercal = \gamma_2 \mathbf{1}_{m_2} \tag{i}$$

where $\mathbf{1}_{m_1} \in \mathbb{R}^{1 \times m_1}$ and $\mathbf{1}_{m_2} \in \mathbb{R}^{1 \times m_2}$ are all-ones row vectors and $\gamma_1 = -\gamma_2 = \mathrm{val}(G)$. We can appeal to the invertability of Q_1, R, and Q_2 to show that the equalities in (i) are uniquely solved by

$$(\mathbf{y}^\diamond)^\intercal = \gamma_1 (Q_1 R T_2^\intercal)^{-1} \mathbf{1}_{m_1}^\intercal \quad \text{and} \quad \mathbf{x}^\diamond = \gamma_2 \mathbf{1}_{m_2} (Q_1 R T_2^\intercal)^{-1}. \tag{ii}$$

So, we must have $\mathbf{x}^\diamond = \mathbf{x}^*$ and $\mathbf{y}^\diamond = \mathbf{y}^\dagger$. This proves that the incompetent game G_{Q_1,T_2} has a unique, completely mixed equilibrium $(\mathbf{x}^*, \mathbf{y}^\dagger)$ and, as a result, it is a completely mixed game. □

Recall that the set $\mathcal{C}$ contains every pair of learning parameters $(\lambda, \mu) \in [0,1] \times [0,1]$ making $G_{\lambda,\mu}$ completely mixed. Proposition 3.5 tells us that the constituent parameters of these pairs are interchangeable; that is, given any $(\lambda, \mu), (\lambda', \mu') \in \mathcal{C}$, we know that $(\lambda, \mu') \in \mathcal{C}$ and $(\lambda', \mu) \in \mathcal{C}$. It is convenient to define the set

$$\mathcal{C}_1 = \{\lambda \in [0,1] : (\lambda, \mu) \in \mathcal{C} \text{ for some } \mu \in [0,1]\} \tag{3.6}$$

belonging to Player 1 and the set

$$\mathcal{C}_2 = \{\mu \in [0,1] : (\lambda, \mu) \in \mathcal{C} \text{ for some } \lambda \in [0,1]\} \tag{3.7}$$

belonging to Player 2. The interchangeability property immediately implies that $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$. Now, having explored the rectangular structure of $\mathcal{C}$, we investigate its plateauing behaviour in Proposition 3.6.

Proposition 3.6. If $Q_1 : [0,1] \to \mathbb{R}^{m_1 \times m_1}$ and $Q_2 : [0,1] \to \mathbb{R}^{m_2 \times m_2}$ are continuous, then the set $\mathcal{C}$ is open in $[0,1] \times [0,1]$.

Proof. Jansen [12, Theorem 3.15] proves that the set of $m_1 \times m_2$ utility matrix pairs associated with completely mixed bimatrix games is open in $\mathbb{R}^{m_1 \times m_2} \times \mathbb{R}^{m_1 \times m_2}$. This,

[3] If $\mathrm{val}(G) = 0$, then we can simply shift its utility matrix R by a non-zero constant, as in [17, Theorem 5.35]. This produces a strategically equivalent game G' with $\mathrm{val}(G') \neq 0$. The sets of equilibria in G and G' are identical.

in turn, implies that the set of $m_1 \times m_2$ utility matrices belonging to completely mixed matrix games is open in $\mathbb{R}^{m_1 \times m_2}$. Indeed, if a matrix $R \in \mathbb{R}^{m_1 \times m_2}$ exists such that the ball $\mathrm{B}_\epsilon(R)$ in $\mathbb{R}^{m_1 \times m_2}$ does not exclusively contain completely mixed utility matrices for all $\epsilon \in \mathbb{R}^+$, then the ball $\mathrm{B}_\epsilon(R, -R)$ in $\mathbb{R}^{m_1 \times m_2} \times \mathbb{R}^{m_1 \times m_2}$ would serve as a counterexample to [12, Theorem 3.15].

Fix a pair of learning parameters $(\lambda, \mu) \in \mathcal{C}$. Evidently, for some $\epsilon \in \mathbb{R}^+$, every matrix game represented by a utility matrix in $\mathrm{B}_\epsilon(R_{\lambda,\mu})$ is completely mixed. The continuity of $Q_1(\cdot)$ and $Q_2(\cdot)$ guarantees that the mapping from $[0,1] \times [0,1]$ to $\mathbb{R}^{m_1 \times m_2}$ given by $(\lambda', \mu') \mapsto R_{\lambda',\mu'}$ is continuous on its entire domain. Therefore, there exists $\delta \in \mathbb{R}^+$ such that, for all $(\lambda', \mu') \in \mathrm{B}_\delta(\lambda, \mu) \cap ([0,1] \times [0,1])$, we have

$$R_{\lambda',\mu'} = Q_1(\lambda') R Q_2(\mu')^\top \in \mathrm{B}_\epsilon(R_{\lambda,\mu}) \tag{i}$$

and $G_{\lambda',\mu'}$ is completely mixed. This allows us to conclude that $\mathcal{C}$ is open in the space $[0,1] \times [0,1]$ of learning parameter pairs, as required. □

Finally, through Proposition 3.5 and Proposition 3.6, we have found that, whenever $G_{\lambda,\mu}$ is completely mixed for some $\lambda, \mu \in [0,1]$, the game value of $G_{Q(\cdot)}$ forms a rectangular plateau around (λ, μ). This provides an explanation for the plateaus present in Figure 2.4; however, it is not a necessary condition for the appearance of plateaus.

3.3 Optimal Learning Parameters

The concept of an executable strategy can also be leveraged to describe a player's optimal learning parameter choices under certain conditions. Let $G_{Q(\cdot)}$ be a parameterised incompetent game in which Player 1 uses the learning trajectory $Q_1 : [0,1] \to \mathbb{R}^{m_1 \times m_1}$ and Player 2 uses the learning trajectory $Q_2 : [0,1] \to \mathbb{R}^{m_2 \times m_2}$. We will allow both players to freely choose their learning parameters before playing the associated incompetent game. This is modelled as a multi-stage game Γ where

Stage 1. (*Learning Parameter Selection*) Player 1 selects a learning parameter $\lambda \in [0,1]$ and Player 2 selects a learning parameter $\mu \in [0,1]$,

Stage 2. (*Action Selection*) Player 1 selects an action $a_i \in A$ and Player 2 selects an action $b_j \in B$ for some $i = 1, 2, \ldots, m_1$ and $j = 1, 2, \ldots, m_2$,

Utility. Player 1 expects to receive utility $u_{\lambda,\mu}(i,j)$ and Player 2 expects to receive utility $-u_{\lambda,\mu}(i,j)$.

The players are limited to their pure strategies in Stage 1 (where the action spaces are continuous) and their mixed strategies in Stage 2 (where the action spaces are discrete). We are exclusively interested in the subgame perfect equilibria of Γ, which require the players to select optimal strategies in $G_{\lambda,\mu}$ during Stage 2. Consequently, we are justified in treating Γ as a game wherein, after the players choose learning parameters $\lambda, \mu \in [0,1]$, they play $G_{\lambda,\mu}$ optimally to achieve expected utilities $\mathrm{val}(G_{\lambda,\mu})$ and $-\mathrm{val}(G_{\lambda,\mu})$ for Player 1 and Player 2, respectively.

The remaining problem is to identify optimal learning parameter choices in Stage 1. Thus, we seek learning parameters $\lambda^*, \mu^* \in [0,1]$ satisfying the usual zero-sum equilibrium inequalities

$$\mathrm{val}\left(G_{\lambda,\mu^*}\right) \leq \mathrm{val}\left(G_{\lambda^*,\mu^*}\right) \leq \mathrm{val}\left(G_{\lambda^*,\mu}\right) \tag{3.8}$$

for every $\lambda, \mu \in [0,1]$. We will investigate the learning parameters that allow the players to execute a competent optimal strategy. Let $\mathbf{O}_1$ denote Player 1's set of optimal strategies in G and $\mathbf{O}_2$ denote Player 2's set of optimal strategies in G. Then, define the sets of learning parameters

$$\mathcal{E}_1 = \{\lambda \in [0,1] : \mathbf{E}_1(\lambda) \cap \mathbf{O}_1 \neq \varnothing\} \tag{3.9}$$

and

$$\mathcal{E}_2 = \{\mu \in [0,1] : \mathbf{E}_2(\mu) \cap \mathbf{O}_2 \neq \varnothing\}. \tag{3.10}$$

Notice that the game value of $G_{Q(\cdot)}$ is constant on the region $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ by Proposition 3.1. Moreover, because Theorem 3.3 says that the equilibrium of a completely mixed incompetent game executes a competent optimal strategy, we have $\mathcal{C}_1 \subseteq \mathcal{E}_1$ and $\mathcal{C}_2 \subseteq \mathcal{E}_2$.

Next, Proposition 3.7 shows that every pair of learning parameters $(\lambda^*, \mu^*) \in \mathcal{E}$ is an equilibrium of the multi-stage game Γ. This means that a player should, whenever it is possible, choose a learning parameter that allows them to execute a competent optimal strategy.

Proposition 3.7. If $(\lambda^*, \mu^*) \in \mathcal{E}$, then (λ^*, μ^*) is an equilibrium in Γ.

Proof. We know that, for some $\mathbf{x}^* \in \mathbf{X}$ and $\mathbf{y}^* \in \mathbf{Y}$, the strategy profile $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ is an equilibrium of G. We will show that a contradiction arises whenever Player 1 or Player 2 possesses a profitable unilateral deviation from the strategy profile (λ^*, μ^*).

First, assume that Player 1 has an alternative learning parameter $\lambda \in [0,1]$ satisfying $\operatorname{val}(G_{\lambda,\mu^*}) > \operatorname{val}(G_{\lambda^*,\mu^*})$. Observe that, given an equilibrium $(\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ of $G_{\lambda^*,\mu}$, we have

$$\begin{aligned} v(\mathbf{x}^* Q_1(\lambda^*), \mathbf{y}^* Q_2(\mu^*)) &= \mathbf{x}^* Q_1(\lambda^*) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top = \mathbf{x}^* R_{\lambda^*,\mu^*} (\mathbf{y}^*)^\top \\ &= \operatorname{val}\left(G_{\lambda^*,\mu^*}\right) < \operatorname{val}\left(G_{\lambda,\mu^*}\right) = \mathbf{x}^\dagger R_{\lambda,\mu^*} (\mathbf{y}^\dagger)^\top \leq \mathbf{x}^\dagger R_{\lambda,\mu^*} (\mathbf{y}^*)^\top \\ &= \mathbf{x}^\dagger Q_1(\lambda) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top = v(\mathbf{x}^\dagger Q_1(\lambda), \mathbf{y}^* Q_2(\mu^*)). \end{aligned} \tag{i}$$

This implies that the strategy $\mathbf{x}^\dagger Q_1(\lambda)$ is a profitable deviation from the equilibrium $(\mathbf{x}^* Q_1, \mathbf{y}^* Q_2)$ in G. Hence, since this violates the equilibrium conditions, Player 1 cannot possess any profitable deviations from (λ^*, μ^*) in Γ.

Second, consider a learning parameter $\mu \in [0,1]$ for Player 2 with $\operatorname{val}(G_{\lambda^*,\mu}) < \operatorname{val}(G_{\lambda^*,\mu^*})$. If $(\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ denotes an equilibrium of $G_{\lambda^*,\mu}$, then

$$\begin{aligned} v(\mathbf{x}^* Q_1(\lambda^*), \mathbf{y}^* Q_2(\mu^*)) &= \mathbf{x}^* Q_1(\lambda^*) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top = \mathbf{x}^* R_{\lambda^*,\mu^*} (\mathbf{y}^*)^\top \\ &= \operatorname{val}\left(G_{\lambda^*,\mu^*}\right) > \operatorname{val}\left(G_{\lambda^*,\mu}\right) = \mathbf{x}^\dagger R_{\lambda^*,\mu} (\mathbf{y}^\dagger)^\top \geq \mathbf{x}^* R_{\lambda^*,\mu} (\mathbf{y}^\dagger)^\top \\ &= \mathbf{x}^* Q_1(\lambda^*) R Q_2(\mu)^\top (\mathbf{y}^\dagger)^\top = v(\mathbf{x}^* Q_1(\lambda^*), \mathbf{y}^\dagger Q_2(\mu)). \end{aligned} \tag{ii}$$

Again, this suggests that Player 2 would benefit from switching to the strategy $\mathbf{y}^\dagger Q_2(\mu)$ in G. This is a contradiction and, as a consequence, Player 2 also does not possess a profitable deviation from (λ^*, μ^*) in Γ.

Therefore, after showing that both players do not have any incentive to deviate from (λ^*, μ^*), we conclude that it is an equilibrium of Γ. □

Although Proposition 3.7 gives a sufficient condition for $(\lambda^*,\mu^*) \in [0,1]\times[0,1]$ to be an equilibrium, it does not exclude the possibility of other optimal learning parameters. Instead, to ensure that $\mathcal{E}$ contains every equilibrium of Γ, we can impose the additional restriction that the learning trajectories $Q_1(\cdot)$ and $Q_2(\cdot)$ must achieve complete competence on the domain $[0,1]$. This necessary condition is addressed in Theorem 3.8.

Theorem 3.8. Assume that the learning trajectories $Q_1(\cdot)$ and $Q_2(\cdot)$ achieve complete competence on the interval $[0,1]$, or $Q_1(\lambda^*) = I_{m_1}$ and $Q_2(\mu^*) = I_{m_2}$ for some $\lambda^\dagger, \mu^\dagger \in [0,1]$. If $(\lambda^\dagger,\mu^\dagger) \in [0,1]\times[0,1]$ is an equilibrium of Γ, then $(\lambda^\dagger,\mu^\dagger) \in \mathcal{E}$.

Proof. Notice that, because each player's entire strategy space is executable under the identity matrices $Q_1(\lambda^*)$ and $Q_2(\mu^*)$, we must have $\lambda^* \in \mathcal{E}_1$ and $\mu^* \in \mathcal{E}_2$. Then, Proposition 3.7 tells us that (λ^*,μ^*) is an equilibrium of Γ and, by the interchangeability of optimal strategies, $(\lambda^*,\mu^\dagger)$ and $(\lambda^\dagger,\mu^*)$ are also equilibria, with

$$\operatorname{val}(G) = \operatorname{val}\left(G_{\lambda^*,\mu^*}\right) = \operatorname{val}\left(G_{\lambda^*,\mu^\dagger}\right) = \operatorname{val}\left(G_{\lambda^\dagger,\mu^*}\right) = \operatorname{val}\left(G_{\lambda^\dagger,\mu^\dagger}\right) \tag{i}$$

(see [20, Theorem 2.1.2]).

Let the strategy profiles $(\mathbf{x}^*,\mathbf{y}^*) \in \mathbf{X}\times\mathbf{Y}$ and $(\mathbf{x}^\diamond,\mathbf{y}^\diamond) \in \mathbf{X}\times\mathbf{Y}$ be equilibria of G_{λ^*,μ^*} and $G_{\lambda^\dagger,\mu^*}$, respectively. Observe that

$$\begin{aligned} v(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}^*) &= \mathbf{x}^\diamond Q_1(\lambda^\dagger) R(\mathbf{y}^*)^\top \le \mathbf{x}^* R(\mathbf{y}^*)^\top = \mathbf{x}^* Q_1(\lambda^*) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top \\ &= \mathbf{x}^* R_{\lambda^*,\mu^*}(\mathbf{y}^*)^\top = v_{\lambda^*,\mu^*}(\mathbf{x}^*,\mathbf{y}^*) = \operatorname{val}\left(G_{\lambda^*,\mu^*}\right) = \operatorname{val}(G) \end{aligned} \tag{ii}$$

and

$$\begin{aligned} v(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}^*) &= \mathbf{x}^\diamond Q_1(\lambda^\dagger) R(\mathbf{y}^*)^\top = \mathbf{x}^\diamond Q_1(\lambda^\dagger) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top = \mathbf{x}^\diamond R_{\lambda^\dagger,\mu^*}(\mathbf{y}^*)^\top \\ &\ge \mathbf{x}^\diamond R_{\lambda^\dagger,\mu^*}(\mathbf{y}^\diamond)^\top = v_{\lambda^\dagger,\mu^*}(\mathbf{x}^\diamond,\mathbf{y}^\diamond) = \operatorname{val}\left(G_{\lambda^\dagger,\mu^*}\right) = \operatorname{val}(G). \end{aligned} \tag{iii}$$

The inequalities in (ii) and (iii) combine to prove that, in the competent game G, the expected utility of the strategy profile $(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}^*)$ is $\operatorname{val}(G)$. Hence, for any $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$, we have

$$\begin{aligned} v(\mathbf{x},\mathbf{y}^*) &= \mathbf{x} R(\mathbf{y}^*)^\top = \mathbf{x} Q_1(\lambda^*) R Q_2(\mu^*)^\top (\mathbf{y}^*)^\top = \mathbf{x} R_{\lambda^*,\mu^*}(\mathbf{y}^*)^\top \\ &\le \mathbf{x}^* R_{\lambda^*,\mu^*}(\mathbf{y}^*)^\top = v_{\lambda^*,\mu^*}(\mathbf{x}^*,\mathbf{y}^*) = \operatorname{val}\left(G_{\lambda^*,\mu^*}\right) = v(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}^*) \end{aligned} \tag{iv}$$

and

$$\begin{aligned} v(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}) &= \mathbf{x}^\diamond Q_1(\lambda^\dagger) R \mathbf{y}^\top = \mathbf{x}^\diamond Q_1(\lambda^\dagger) R Q_2(\mu^*)^\top \mathbf{y}^\top = \mathbf{x}^\diamond R_{\lambda^\dagger,\mu^*} \mathbf{y}^\top \\ &\ge \mathbf{x}^\diamond R_{\lambda^\dagger,\mu^*}(\mathbf{y}^\diamond)^\top = v_{\lambda^\dagger,\mu^*}(\mathbf{x}^\diamond,\mathbf{y}^\diamond) = \operatorname{val}\left(G_{\lambda^\dagger,\mu^*}\right) = v(\mathbf{x}^\diamond Q_1(\lambda^\dagger),\mathbf{y}^*) \end{aligned} \tag{v}$$

Clearly, since (iv) and (v) are the equilibrium inequalities, Player 1 possesses the optimal strategy $\mathbf{x}^\diamond Q_1(\lambda^\dagger)$ in G and $\lambda^\dagger \in \mathcal{E}_1$. A similar argument for the incompetent game $G_{\lambda^*,\mu^\dagger}$ constructs an optimal strategy for Player 2 such that $\mu^\dagger \in \mathcal{E}_2$. So, we obtain $(\lambda^\dagger,\mu^\dagger) \in \mathcal{E}_1 \times \mathcal{E}_2$, as desired. □

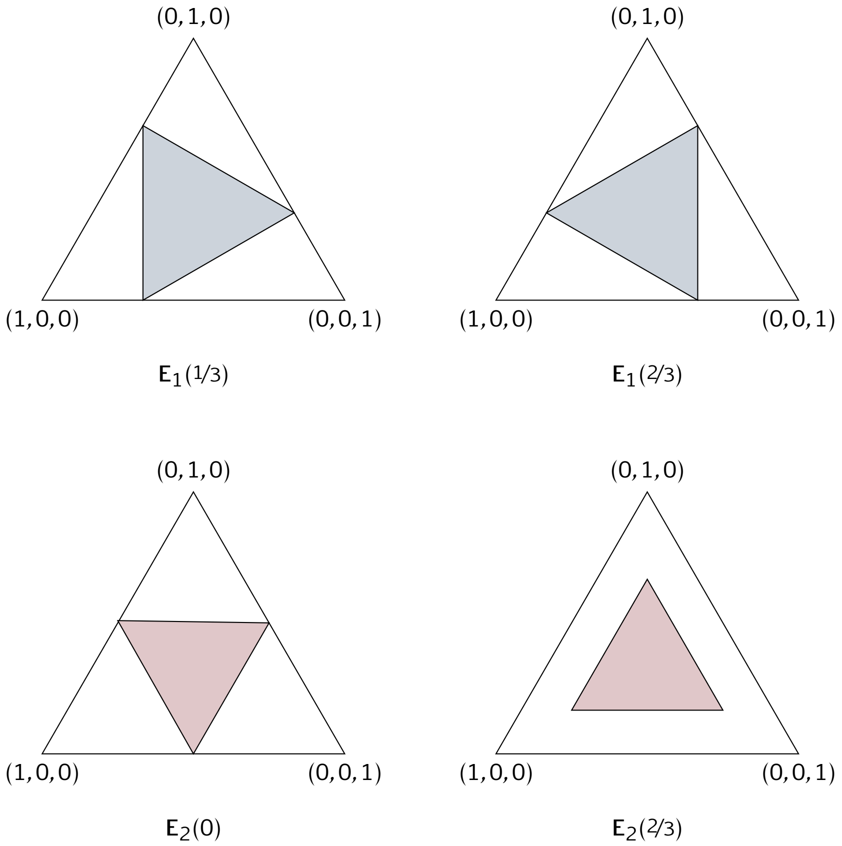


Figure 3.2. The executable strategy spaces belonging to Player 1 (blue) and Player 2 (red) at incomparable skill levels.

Lastly, to provide another sufficient condition for identifying optimal learning parameters, note that the mathematical notion of incompetence does not always admit straightforward comparisons of skill. An increase in a player's skill corresponds to an expansion of their executable strategy space without "forgetting" any previously executable strategy. Accordingly, we say that Player 1 and Player 2 achieve *greater skill* at $\lambda_2, \mu_2 \in [0,1]$ than at $\lambda_1, \mu_1 \in [0,1]$ whenever

$$\mathbf{E}_1(\lambda_1) \subseteq \mathbf{E}_1(\lambda_2) \quad \text{and} \quad \mathbf{E}_2(\mu_1) \subseteq \mathbf{E}_2(\mu_2), \tag{3.11}$$

respectively. The binary relation of having greater skill does not necessarily relate every pair of learning parameters. Consider, for example, the learning trajectories $Q_1 : [0,1] \to \mathbb{R}^{3\times 3}$ and $Q_2 : [0,1] \to \mathbb{R}^{3\times 3}$ where

$$Q_1(\lambda) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} (1-\lambda) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \lambda$$

for all $\lambda \in [0,1]$ and

$$Q_2(\mu) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}(1-\mu) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\mu$$

for all $\mu \in [0,1]$. Figure 3.2 shows the executable strategies available to Player 1 at $\lambda = 1/3, 2/3$ and the executable strategies available to Player 2 at $\mu = 0, 2/3$. Player 1's learning parameters are incomparable because $\mathbf{E}_1(1/3) \not\subseteq \mathbf{E}_1(2/3)$ and $\mathbf{E}_1(2/3) \not\subseteq \mathbf{E}_1(1/3)$. Similarly, Player 2's learning parameters are incomparable because $\mathbf{E}_2(0) \not\subseteq \mathbf{E}_2(2/3)$ and $\mathbf{E}_2(2/3) \not\subseteq \mathbf{E}_2(0)$. Although the ability to compare a player's skill at different parameters is not guaranteed, Lemma 3.9 relates the game value of $G_{Q(\cdot)}$ at those learning parameters satisfying (3.11).

Lemma 3.9. If $\mathbf{E}_1(\lambda_1) \subseteq \mathbf{E}_1(\lambda_2)$ and $\mathbf{E}_2(\mu_1) \subseteq \mathbf{E}_2(\mu_2)$ for some $\lambda_1, \lambda_2, \mu_1, \mu_2 \in [0,1]$, then

$$\operatorname{val}\left(G_{\lambda_1,\mu}\right) \le \operatorname{val}\left(G_{\lambda_2,\mu}\right) \quad \text{and} \quad \operatorname{val}\left(G_{\lambda,\mu_1}\right) \ge \operatorname{val}\left(G_{\lambda,\mu_2}\right) \tag{3.12}$$

for every $\lambda, \mu \in [0,1]$.

Proof. First, fix $\mu \in [0,1]$ and suppose $\mathbf{E}_1(\lambda_1) \subseteq \mathbf{E}_1(\lambda_2)$ for some $\lambda_1, \lambda_2 \in [0,1]$. Let $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ and $(\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ be equilibria of $G_{\lambda_1,\mu}$ and $G_{\lambda_2,\mu}$, respectively. Evidently, since $\mathbf{x}^* Q_1(\lambda_1) \in \mathbf{E}_1(\lambda_1) \subseteq \mathbf{E}_1(\lambda_2)$, there exists $\mathbf{x} \in \mathbf{X}$ such that $\mathbf{x} Q_1(\lambda_2) = \mathbf{x}^* Q_1(\lambda_1)$. Observe that

$$\begin{aligned} \operatorname{val}\left(G_{\lambda_1,\mu}\right) &= \mathbf{x}^* R_{\lambda_1,\mu} (\mathbf{y}^*)^\top \le \mathbf{x}^* R_{\lambda_1,\mu} (\mathbf{y}^\dagger)^\top = \mathbf{x}^* Q_1(\lambda_1) R Q_2(\mu)^\top (\mathbf{y}^\dagger)^\top \\ &= \mathbf{x} Q_1(\lambda_2) R Q_2(\mu)^\top (\mathbf{y}^\dagger)^\top = \mathbf{x} R_{\lambda_2,\mu} (\mathbf{y}^\dagger)^\top \le \mathbf{x}^\dagger R_{\lambda_2,\mu} (\mathbf{y}^\dagger)^\top = \operatorname{val}\left(G_{\lambda_2,\mu}\right), \end{aligned} \tag{i}$$

as required.

Second, fix $\lambda \in [0,1]$ and suppose $\mathbf{E}_2(\mu_1) \subseteq \mathbf{E}_2(\mu_2)$ for some $\mu_1, \mu_2 \in [0,1]$. Consider an equilibrium $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbf{X} \times \mathbf{Y}$ of G_{λ,μ_1} and an equilibrium $(\mathbf{x}^\dagger, \mathbf{y}^\dagger) \in \mathbf{X} \times \mathbf{Y}$ of G_{λ,μ_2}. Then, $\mathbf{y}^* Q_2(\mu_1) \in \mathbf{E}_2(\mu_1) \subseteq \mathbf{E}_2(\mu_2)$ and $\mathbf{y} Q_2(\mu_2) = \mathbf{y}^* Q_2(\mu_1)$ for some $\mathbf{y} \in \mathbf{Y}$. Again, we have

$$\begin{aligned} \operatorname{val}\left(G_{\lambda,\mu_1}\right) &= \mathbf{x}^* R_{\lambda,\mu_1} (\mathbf{y}^*)^\top \ge \mathbf{x}^\dagger R_{\lambda,\mu_1} (\mathbf{y}^*)^\top = \mathbf{x}^\dagger Q_1(\lambda) R Q_2(\mu_1)^\top (\mathbf{y}^*)^\top \\ &= \mathbf{x}^\dagger Q_1(\lambda) R Q_2(\mu_2)^\top \mathbf{y}^\top = \mathbf{x}^\dagger R_{\lambda,\mu_2} \mathbf{y}^\top \ge \mathbf{x}^\dagger R_{\lambda,\mu_2} (\mathbf{y}^\dagger)^\top = \operatorname{val}\left(G_{\lambda,\mu_2}\right), \end{aligned} \tag{ii}$$

which proves the desired conclusion. □

Next, we say that Player 1 achieves *maximum skill* at $\lambda^* \in [0,1]$ whenever $\mathbf{E}_1(\lambda) \subseteq \mathbf{E}_1(\lambda^*)$ for all $\lambda \in [0,1]$. Analogously, we say that Player 2 achieves *maximum skill* at $\mu^* \in [0,1]$ whenever $\mathbf{E}_2(\mu) \subseteq \mathbf{E}_2(\mu^*)$ for all $\lambda \in [0,1]$. This definition captures the idea that a maximally skilled player should be able to execute every learnable strategy. Proposition 3.10 proves that a pair of maximum skill learning parameters is necessarily an equilibrium of Γ.

Proposition 3.10. If Player 1 achieves maximum skill at $\lambda^* \in [0,1]$ and Player 2 achieves maximum skill at $\mu^* \in [0,1]$, then (λ^*, μ^*) is an equilibrium of Γ.

Proof. We know that, for all $\lambda, \mu \in [0,1]$, the spaces of executable strategy satisfy $\mathbf{E}_1(\lambda) \subseteq \mathbf{E}_1(\lambda^*)$ and $\mathbf{E}_2(\mu) \subseteq \mathbf{E}_2(\mu^*)$. Thus, according to Lemma 3.9, we obtain

$$\mathrm{val}\Big(G_{\lambda,\mu^*}\Big) \leq \mathrm{val}\Big(G_{\lambda^*,\mu^*}\Big) \leq \mathrm{val}\Big(G_{\lambda^*,\mu}\Big).$$

These are the necessary equilibrium inequalities from (3.8) and, as a result, (λ^*, μ^*) is an equilibrium of Γ. □

The existence of maximum skill learning parameters depends entirely on the selected learning trajectories; however, a special case occurs when Player 1 and Player 2 are capable of achieving complete competence. This means that $Q_1(\lambda^*) = I_{m_1}$ and $Q_2(\mu^*) = I_{m_2}$ for some $\lambda^*, \mu^* \in [0,1]$. Clearly, since

$$\mathbf{E}_1(\lambda^*) = \mathbf{E}_1(I_{m_1}) = \mathbf{X} \quad \text{and} \quad \mathbf{E}_2(\mu^*) = \mathbf{E}_2(I_{m_2}) = \mathbf{Y},$$

Player 1 achieves maximum skill at λ^* and Player 2 achieves maximum skill at μ^*. Therefore, Proposition 3.10 reduces to the "optimality of complete competence" discussed in [2, Section 4.4] and [1, Proposition 3.1].

4 Incremental Learning

It is often necessary to model the interaction of learning and playing through complex mechanisms. Consider, for instance, an athlete who must train between competitions or a poker player who must practice between tournaments. How can we describe these situations game theoretically? Our proposed solution is an incremental learning game in which, between repeated plays of an incompetent game, a player may increment their learning parameters to modify their incompetence.

We will begin in Section 4.1 by explaining stochastic games, the general framework used in the formulation of incremental learning. Then, Section 4.2 defines the incremental learning model and Section 4.3 and Section 4.4 propose a backward induction algorithm for computing equilibria. Lastly, this procedure is applied in Section 4.5 to analyse learning strategies in a simplified tennis game.

4.1 Stochastic Games

Firstly, before adding incremental learning to an incompetent game, we should explain the stochastic game model introduced by Shapley [23]. This is a mathematical framework for describing situations wherein multiple bimatrix games are played sequentially with player-influenced transitions. We will adopt the conventions used in Chapter 3 and Chapter 4 of [10] to accommodate some additional notation and, in particular, player association in a stochastic game will be indicated using a superscript.

A two-player *stochastic game* $\Gamma = (\mathcal{S}, \mathcal{A}, \mathcal{B}, r^1, r^2, p)$ evolves over a collection of *stages* $t = 0, 1, 2, \ldots$ and visits a *state* S_t at every stage. We view $\{S_t\}_{t=0}^{\infty}$ as a stochastic process that takes its values from a finite set $\mathcal{S}$ called the *state space*. If the game is in state $s \in \mathcal{S}$, then Player 1 and Player 2 must simultaneously choose *actions* from the finite sets $\mathcal{A}(s)$ and $\mathcal{B}(s)$, respectively. These choices are captured in the stochastic processes $\{A_t\}_{t=0}^{\infty}$ and $\{B_t\}_{t=0}^{\infty}$ where, at any stage $t = 0, 1, 2, \ldots$, the random variable A_t gives Player 1's action and the random variable B_t gives Player 2's action. Consequently, the event $\{S_t = s, A_t = a, B_t = b\}$ corresponds to Player 1 choosing action $a \in \mathcal{A}(s)$ and Player 2 choosing action $b \in \mathcal{B}(s)$ at stage $t = 0, 1, 2, \ldots$ and in state $s \in \mathcal{S}$.

After a pair of actions have been selected, the players are awarded utility and the game transitions to a potentially different state. Suppose that, at the stage $t = 0, 1, 2, \ldots$, the game is in state $S_t = s \in \mathcal{S}$, Player 1's action is $A_t = a \in \mathcal{A}(s)$, and Player 2's action is $B_t = b \in \mathcal{B}(s)$. The *immediate utility* allocations are denoted by $r^1(s, a, b)$ for Player 1 and $r^2(s, a, b)$ for Player 2. Then, the next state S_{t+1} is determined randomly with the *transition probability*

$$p(s'|s,a,b) = \mathbb{P}(S_{t+1} = s'|S_t = s, A_t = a, B_t = b) \tag{4.1}$$

being the probability that $s' \in \mathcal{S}$ is drawn. The quantity $p(s'|s,a,b)$ is well-defined because it is assumed that the transition dynamics are Markovian; that is, they are calculated using only the current state S_t, Player 1's action A_t, and Player 2's action B_t.

We are mostly interested in the space of *stationary strategies*, which are strategies that depend only on the current state.[1] Typically, the block row vectors $\mathbf{f} = (\mathbf{f}(s))_{s\in\mathcal{S}}$ and $\mathbf{g} = (\mathbf{g}(s))_{s\in\mathcal{S}}$ represent stationary strategies belonging to Player 1 and Player 2, respectively. If $s \in \mathcal{S}$ is the game's current state, then the stochastic row vectors $\mathbf{f}(s) = (f(s,a))_{a\in\mathcal{A}(s)}$ and $\mathbf{g}(s) = (g(s,b))_{b\in\mathcal{B}(s)}$ contain the probability $f(s,a)$ that Player 1 chooses action $a \in \mathcal{A}(s)$ and the probability $g(s,b)$ that Player 2 chooses action $b \in \mathcal{B}(s)$.[2] Aligning with our previous terminology for normal-form games, we call $\mathbf{f}$ and $\mathbf{g}$ *pure stationary strategies* whenever $f(s,a) \in \{0,1\}$ and $g(s,b) \in \{0,1\}$ for every $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$, and $b \in \mathcal{B}(s)$. The set of available stationary strategies is denoted by $\mathbf{F}$ for Player 1 and $\mathbf{G}$ for Player 2.

Fix a strategy profile $(\mathbf{f},\mathbf{g}) \in \mathbf{F}\times\mathbf{G}$. Seeing that the players' behaviours are entirely determined by these strategies, we extend the immediate utility of Player $k = 1,2$ to

$$r^k(s,\mathbf{f},\mathbf{g}) = \sum_{a\in\mathcal{A}(s)}\sum_{b\in\mathcal{B}(s)} f(s,a)r^k(s,a,b)g(s,b) \tag{4.2}$$

and the transition probabilities to

$$p(s'|s,a,b) = \sum_{a\in\mathcal{A}(s)}\sum_{b\in\mathcal{B}(s)} f(s,a)p(s'|s,a,b)g(s,b) \tag{4.3}$$

for all $s,s' \in \mathcal{S}$. Evidently, since these probabilities depend only on the present state, the process $\{S_t\}_{t=0}^{\infty}$ becomes a Markov chain whose one-step probability transition matrix is $P(\mathbf{f},\mathbf{g}) = (p(s'|s,\mathbf{f},\mathbf{g}))_{s,s'\in\mathcal{S}}$. This allows us to encode the streams of immediate utility rewards for Player 1 and Player 2 as stochastic processes $\{R_t^1\}_{t=0}^{\infty}$ and $\{R_t^2\}_{t=0}^{\infty}$. So, after starting at state $s \in \mathcal{S}$, the expected utility allocation for Player $k = 1,2$ at state $t = 0,1,2,\ldots$ is

$$\begin{aligned}\mathbb{E}_{s\mathbf{fg}}\left[R_t^k\right] &= \sum_{s'\in\mathcal{S}} \mathbb{E}_{s\mathbf{fg}}\left[R_t^k\middle|S_t = s'\right]\mathbb{P}_{s\mathbf{fg}}(S_t = s)\\ &= \sum_{s'\in\mathcal{S}} r^k(s',\mathbf{f},\mathbf{g})P(\mathbf{f},\mathbf{g})^t[s,s'] \end{aligned}\tag{4.4}$$

where $\mathbb{P}_{s\mathbf{fg}}$ is the probability measure induced by the dynamics of $(\mathbf{f},\mathbf{g})$ with initial state $S_0 = s$. How can this be used to value an arbitrary strategy profile $(\mathbf{f},\mathbf{g}) \in \mathbf{F}\times\mathbf{G}$?

[1] This definition of a strategy can be generalised to obtain Markov strategies, which may depend on the current state, and behaviour strategies, which may depend on the game's history. Fortunately, a suitable equilibrium solution always exists in the space of stationary strategies for the games we consider (see, for example, [10, Theorem 3.1.1, Theorem 4.6.4]).

[2] We are implicitly imposing a canonical ordering on the state space $\mathcal{S}$ and, for all $s \in \mathcal{S}$, the action sets $\mathcal{A}(s)$ and $\mathcal{B}(s)$. This allows $\mathcal{S}$, $\mathcal{A}(s)$, and $\mathcal{B}(s)$ to be used as index sets for various vectors and matrices.

Naively, we want to compute the expected total utility received throughout the game; however, the summations

$$\sum_{t=0}^{\infty} \mathbb{E}_{s\mathbf{fg}}\left[R_t^1\right] \quad \text{and} \quad \sum_{t=0}^{\infty} \mathbb{E}_{s\mathbf{fg}}\left[R_t^2\right]$$

might not converge.

Instead, consider a two-player *discounted stochastic game* Γ_β wherein future utility rewards are progressively diminished by a *discount factor* $\beta \in [0,1)$. The *discounted value* of $(\mathbf{f},\mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ to Player $k = 1,2$ is

$$v_\beta^k(s,\mathbf{f},\mathbf{g}) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{s\mathbf{fg}}\left[R_t^k\right] \tag{4.5}$$

after starting at the initial state $s \in \mathcal{S}$.[3] The convergence of the summation in (4.5) is guaranteed because the sequence of expected utilities is bounded between the minimum and maximum utility allotments. We call the vector $\mathbf{v}_\beta^k(\mathbf{f},\mathbf{g}) = (v_\beta^k(s,\mathbf{f},\mathbf{g}))_{s\in\mathcal{S}}$ the *discounted value vector* of $(\mathbf{f},\mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ to Player $k = 1,2$. Applying this valuation of player strategies, we say that a strategy profile $(\mathbf{f}^*,\mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ is a *Nash equilibrium* of Γ_β whenever the componentwise inequalities

$$\mathbf{v}_\beta^1(\mathbf{f},\mathbf{g}^*) \le \mathbf{v}_\beta^1(\mathbf{f}^*,\mathbf{g}^*) \quad \text{and} \quad \mathbf{v}_\beta^2(\mathbf{f}^*,\mathbf{g}) \le \mathbf{v}_\beta^2(\mathbf{f}^*,\mathbf{g}^*) \tag{4.6}$$

hold for every $\mathbf{f} \in \mathbf{F}$ and $\mathbf{g} \in \mathbf{G}$. Notice that, by requiring that the inequalities are satisfied regardless of the initial state, this condition eliminates incredible threats and mirrors the subgame perfect equilibrium in extensive-form games.

The existence of stationary equilibria in two-player zero-sum stochastic games—where $r^1(s,a,b)+r^2(s,a,b) = 0$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}(s)$, and $b \in \mathcal{B}(s)$—was established by Shapley [23]. Again, if $(\mathbf{f},\mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ is a Nash equilibrium of a zero-sum stochastic game, then $\mathbf{f}^*$ and $\mathbf{g}^*$ are called *optimal strategies*. Generally, we will not require that our stochastic games are zero-sum and, as a consequence, we must leverage Fink's [11] result showing that a stationary equilibrium always exists in a finite-player general-sum stochastic game.

4.2 Incremental Learning Games

We are now prepared to model the process of incremental learning in incompetent games. Consider a parameterised incompetent game $G_{Q_1(\cdot),Q_2(\cdot)}$ with learning trajectories $Q_1 : [0,1] \to \mathbb{R}^{m_1 \times m_1}$ and $Q_2 : [0,1] \to \mathbb{R}^{m_2 \times m_2}$. An incremental learning game allows Player 1 and Player 2 to increment their learning parameters through the ordered sets

$$\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_{n_1}\} \subset [0,1] \quad \text{and} \quad \mathrm{M} = \{\mu_1, \mu_2, \ldots, \mu_{n_2}\} \subset [0,1]$$

[3] A different solution to this problem is to construct a limiting average stochastic game Γ_α (see [10, Section 3.4, Chapter 5]). Here, the value of a strategy profile is the limit of its average rewards over finite time horizons. This alternative approach is ignored because, unlike in discounted stochastic games, the existence of stationary equilibria is not guaranteed in limiting average stochastic games (see [10, Example 3.4.1]).

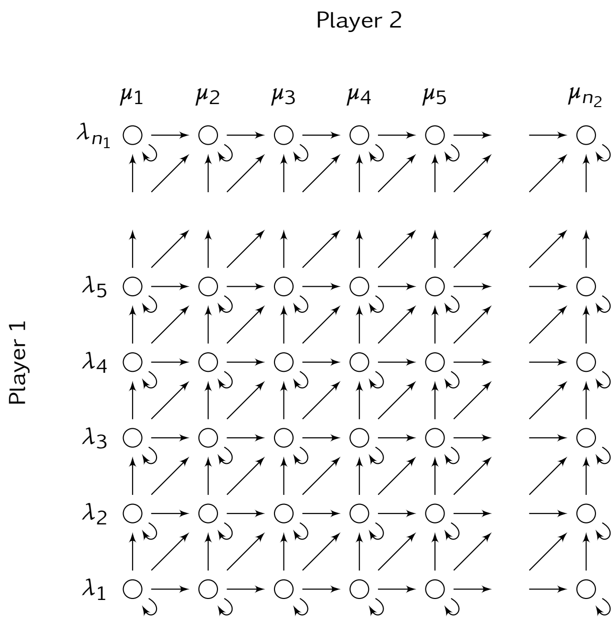


Figure 4.1. The transition structure of a general incremental learning game.

for some $n_1, n_2 \in \mathbb{Z}^+$. Recall that the learning parameters λ_i and μ_j correspond to the incompetence matrices $Q_1(\lambda_i)$ and $Q_2(\mu_j)$, as defined in Section 2.3. The elements of Λ and M are called *attainable learning parameters*.

Fix $i \in \{1, 2, \ldots, n_1\}$ and $j \in \{1, 2, \ldots, n_2\}$ such that λ_i is Player 1's current learning parameter and μ_j is Player 2's current learning parameter. We divide gameplay into two distinct phases: a *playing phase* and a *learning phase*. First, the playing phase involves playing the incompetent game G_{λ_i, μ_j} and receiving the utility allotments associated with its realised outcome. Second, unless a player's attainable learning parameters have been exhausted, the learning phase gives Player 1 and Player 2 the option to advance their learning parameters to λ_{i+1} and μ_{j+1}, respectively. The decision to increment a learning parameter might incur state-dependent *learning costs* $c^1(\lambda_i, \mu_j)$ and $c^2(\lambda_i, \mu_j)$. This process is repeated using the updated learning parameters and utility is gradually accumulated over time. The structure of an incremental learning game is illustrated in Figure 4.1, which represents learning parameter pairs as nodes and possible transitions as arcs. Note that a transition can only occur as a result of actions in the learning phase.

We are going to address incremental learning games with an infinite time horizon. The motivation, as it appears in [10, Section 2.2], behind our exclusion of finite time horizons is twofold:

- a stochastic game with a "short" time horizon can already be solved easily using dynammic programming, and
- a stochastic game with a "long" time horizon is computationally expensive to solve using the same method.

An alternative approach to finding equilibria in a "long" time horizon stochastic game is to approximate it over an infinite time horizon and leverage the various mathematical programming techniques capable of solving these models. This brings the added benefit of producing stationary strategies, which are often easier to implement than the time-dependent strategies produced when solving stochastic games with a finite time horizon [10].

Next, to develop a formal description of incremental learning, we define an *incremental learning game* as a stochastic game Γ constructed by the following process. The state space

$$\mathcal{S} = \Big\{(i,j) : i = 1,2,\dots,n_1 \text{ and } j = 1,2,\dots,n_2\Big\}$$

is chosen to index the attainable learning parameters; the state $(i,j) \in \mathcal{S}$ corresponds to the parameters $(\lambda_i, \mu_j) \in \Lambda \times \mathrm{M}$. Fix a state $s = (i,j) \in \mathcal{S}$. It is convenient to simplify our notation by writing i instead of λ_i and j instead μ_j whenever learning parameters are referenced. A player's action should consist of an action in the playing phase and an action in the learning phase. So, Player 1's action set is

$$\mathcal{A}(s) = \begin{cases} \{1,2,\dots,m_1\} \times \{0,1\}, & i \neq n_1, \\ \{1,2,\dots,m_1\} \times \{0\}, & i = n_1, \end{cases}$$

and Player 2's action set is

$$\mathcal{B}(s) = \begin{cases} \{1,2,\dots,m_2\} \times \{0,1\}, & j \neq n_2, \\ \{1,2,\dots,m_2\} \times \{0\}, & j = n_2. \end{cases}$$

If $a = (a_P, a_L) \in \mathcal{A}(s)$ and $b = (b_P, b_L) \in \mathcal{B}(s)$ are selected, then a_P, b_P are interpreted as playing phase actions and a_L, b_L are interpreted as learning phase actions. These actions award the utilities

$$r^1(s,a,b) = u_{i,j}(a_P, b_P) - a_L c^1(i,j) \quad \text{and} \quad r^2(s,a,b) = -u_{i,j}(a_P, b_P) - b_L c^2(i,j)$$

to Player 1 and Player 2, respectively.[4] Furthermore, they cause a guaranteed transition to the next state $(i + a_L, j + b_L)$ such that the transition probabilities are given by

$$p(s'|s,a,b) = \begin{cases} 1, & i' = i + a_L \text{ and } j' = j + b_L, \\ 0, & i' \neq i + a_L \text{ or } j' \neq j + b_L, \end{cases}$$

for every $s' = (i', j') \in \mathcal{S}$. The incremental learning game Γ is defined as the stochastic game $(\mathcal{S}, \mathcal{A}, \mathcal{B}, r^1, r^2, p)$. Again, to value a strategy profile, we use the *discounted incremental learning game* Γ_β with a predetermined discount factor $\beta \in [0,1)$.

Notice that, for all $s \in \mathcal{S}$ and $(a,b) \in \mathcal{A}(s) \times \mathcal{B}(s)$, the utilities $r^1(s,a,b)$ and $r^2(s,a,b)$ contain two terms: one that depends only on the playing phase actions and one that depends only on the learning phase actions. Thus, since these different types of actions do not interact, they are selected using independent probability distributions. Fix a state $s = (i,j) \in \mathcal{S}$ and a strategy profile $(\mathbf{f}, \mathbf{g}) \in \mathbf{F} \times \mathbf{G}$. We know

[4] Clearly, an incremental learning game is zero-sum if and only if there are no learning costs, or $c^1(i,j) = 0$ and $c^2(i,j) = 0$ for all $i = 1,2,\dots,n_1$ and $j = 1,2,\dots,n_2$.

that Player 1 chooses an action $a = (a_P, a_L)$ by independently drawing a_P from the distribution $\mathbf{f}_P(s)$ over $\{1, 2, \dots, m_1\}$ and a_L from the distribution $\mathbf{f}_L(s)$ over $\{0, 1\}$. Similarly, Player 2 chooses an action $b = (b_P, b_L)$ by independently drawing b_P from the distribution $\mathbf{g}_P(s)$ over $\{1, 2, \dots, m_2\}$ and b_L from the distribution $\mathbf{g}_L(s)$ over $\{0, 1\}$. This decomposition of player strategies is exploited in Proposition 4.1 to show that, given an equilibrium $(\mathbf{f}^*, \mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ of Γ_β, the strategy profile $(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s))$ is an equilibrium of $G_{i,j}$.

Proposition 4.1. Let the strategy profile $(\mathbf{f}^*, \mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ be an equilibrium of the discounted incremental learning game Γ_β. Then, at any state $s = (i, j) \in \mathcal{S}$, the strategy profile $(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s))$ is an equilibrium of the incompetent game $G_{i,j}$.

Proof. Consider a strategy profile $(\mathbf{f}, \mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ that is identical to $(\mathbf{f}^*, \mathbf{g}^*)$ except at $\mathbf{f}_P(s) \neq \mathbf{f}^*_P(s)$ and $\mathbf{g}_P(s) \neq \mathbf{g}^*_P(s)$. Observe that, after applying (4.2) and (4.5), we obtain

$$\begin{aligned} v^k_\beta(s, \mathbf{f}, \mathbf{g}) - v^k_\beta(s, \mathbf{f}^*, \mathbf{g}^*) &= \sum_{t=0}^{\infty} \beta^t \Big(\mathbb{E}_{s\mathbf{fg}}\big[R^k_t\big] - \mathbb{E}_{s\mathbf{f}^*\mathbf{g}^*}\big[R^k_t\big] \Big) \\ &= \sum_{t=0}^{\infty} \sum_{s' \in \mathcal{S}} \beta^t \Big(\mathbb{E}_{s\mathbf{fg}}\big[R^k_t \big| S_t = s'\big] \mathbb{P}_{s\mathbf{fg}}(S_t = s') - \mathbb{E}_{s\mathbf{f}^*\mathbf{g}^*}\big[R^k_t \big| S_t = s'\big] \mathbb{P}_{s\mathbf{f}^*\mathbf{g}^*}(S_t = s') \Big). \qquad \text{(i)} \end{aligned}$$

Now, since $\mathbf{f}_L(s') = \mathbf{f}^*_L(s')$ and $\mathbf{g}_L(s') = \mathbf{g}^*_L(s')$ for all $s' \in \mathcal{S}$, the strategy profiles $(\mathbf{f}, \mathbf{g})$ and $(\mathbf{f}^*, \mathbf{g}^*)$ induce the same transition dynamics such that $\mathbb{P}_{s\mathbf{fg}} = \mathbb{P}_{s\mathbf{f}^*\mathbf{g}^*}$. Moreover, the similarities between $(\mathbf{f}, \mathbf{g})$ and $(\mathbf{f}^*, \mathbf{g}^*)$ imply that their expected playing phase utility is equal on $\mathcal{S} \setminus \{s\}$ and their expected learning phase utility is equal on $\mathcal{S}$. This allows us to reduce (i) to

$$\begin{aligned} &v^k_\beta(s, \mathbf{f}, \mathbf{g}) - v^k_\beta(s, \mathbf{f}^*, \mathbf{g}^*) \\ &\quad = \sum_{t=0}^{\infty} \sum_{s' \in \mathcal{S}} \beta^t \mathbb{P}_{s\mathbf{f}^*\mathbf{g}^*}(S_t = s') \Big(\mathbb{E}_{s\mathbf{fg}}\big[R^k_t \big| S_t = s'\big] - \mathbb{E}_{s\mathbf{f}^*\mathbf{g}^*}\big[R^k_t \big| S_t = s'\big] \Big) \\ &\quad = (-1)^{k-1} \Big(v_{i,j}(\mathbf{f}_P(s), \mathbf{g}_P(s)) - v_{i,j}(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s)) \Big) \sum_{t=0}^{\infty} \beta^t \mathbb{P}_{s\mathbf{f}^*\mathbf{g}^*}(S_t = s) \qquad \text{(ii)} \end{aligned}$$

because the only remaining contribution is from the difference in expected playing phase utilities at the state s. Note that $\mathbb{P}_{s\mathbf{f}^*\mathbf{g}^*}(S_0 = s) = 1$ and, as a consequence, the summation in (ii) is strictly positive. So,

$$v^1_\beta(s, \mathbf{f}, \mathbf{g}) \leq v^1_\beta(s, \mathbf{f}^*, \mathbf{g}^*) \quad \text{implies} \quad v_{i,j}(\mathbf{f}_P(s), \mathbf{g}_P(s)) \leq v_{i,j}(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s)) \qquad \text{(iii)}$$

and

$$v^2_\beta(s, \mathbf{f}, \mathbf{g}) \leq v^2_\beta(s, \mathbf{f}^*, \mathbf{g}^*) \quad \text{implies} \quad v_{i,j}(\mathbf{f}_P(s), \mathbf{g}_P(s)) \geq v_{i,j}(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s)). \qquad \text{(iv)}$$

Set $\mathbf{g} = \mathbf{g}^*$ in (iii) and $\mathbf{f} = \mathbf{f}^*$ in (iv). Then, by the stochastic game equilibrium inequalities for $(\mathbf{f}^*, \mathbf{g}^*)$ in Γ_β, we conclude that $(\mathbf{f}^*_P(s), \mathbf{g}^*_P(s))$ satisfies the matrix game equilibrium inequalities in $G_{i,j}$. □

The message of Proposition 4.1 is that, whenever we are interested in finding equilibria of a discounted incremental learning game, we are able to assume that both players always select optimal strategies of $G_{i,j}$ in the playing phase at state $s = (i, j) \in \mathcal{S}$. Next, this realisation is used to simplify the description of incremental learning and "remove" the playing phase.

Fix an arbitrary state $s = (i, j) \in \mathcal{S}$. Noting that each player's behaviour during a playing phase is entirely determined by Proposition 4.1, they only need to select a learning phase action from the sets

$$\mathcal{A}(s) = \begin{cases} \{0,1\}, & i \neq n_1, \\ \{0\}, & i = n_1, \end{cases} \quad \text{and} \quad \mathcal{B}(s) = \begin{cases} \{0,1\}, & j \neq n_2, \\ \{0\}, & j = n_2, \end{cases}.$$

If Player 1 selects $a \in \mathcal{A}(s)$ and Player 2 selects $b \in \mathcal{B}(s)$, then they receive the immediate utilities

$$r^1(s,a,b) = \operatorname{val}\left(G_{i,j}\right) - ac^1(i,j) \quad \text{and} \quad r^2(s,a,b) = -\operatorname{val}\left(G_{i,j}\right) - bc^2(i,j),$$

respectively. Additionally, the game transitions to the state $(i+a, j+b)$ and the transition probabilities are given by

$$p(s'|s,a,b) = \begin{cases} 1, & i' = i + a \text{ and } j' = j + b, \\ 0, & i' \neq i + a \text{ or } j' \neq j + b, \end{cases}$$

for all $s' = (i', j') \in \mathcal{S}$. Henceforth, we will use this simplified formulation to describe an incremental learning game. Although Proposition 4.1 shows that both formulations are equivalent, the simplified version does not explicitly model the playing phase behaviour. The playing phase equilibrium strategies at the state $s = (i, j) \in \mathcal{S}$ can be found separately by computing the optimal strategies of $G_{i,j}$.

4.3 Backward Induction

After developing a mathematical model that captures the features of incremental learning in incompetent games, our immediate task is to identify a procedure to compute its equilibrium solutions. Typically, a single equilibrium of a general-sum stochastic game can be found using nonlinear programming (see, for example, [10, Section 3.8]); however, this approach complicates the process of finding multiple equilibria. Instead, we will propose a modified backward induction algorithm that, under certain conditions, is capable of exhaustively identifying every equilibrium in a discounted incremental learning game.

The process of *backward induction*, which develops a rational strategy by reasoning backward through time, is commonly used to solve game-theoretic problems. Consider a finite extensive-form game with perfect information; that is, a game wherein every information set is a singleton. Here, backward induction produces a subgame perfect equilibrium by proceeding upward through the game tree and assigning optimal actions contingent on future play [19]. Figure 4.2 shows a backward induction solution to a variant of "Battle of the Sexes" with perfect information. A selected action is indicated by a solid arc between nodes and an unselected action is indicated by a dashed arc between nodes. Notice that, at every decision point,

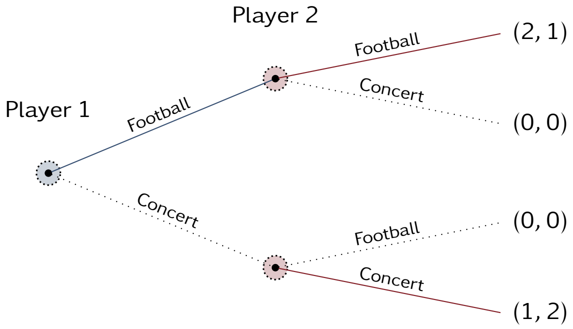


Figure 4.2. A backward induction solution to perfect-information "Battle of the Sexes".

the controlling player is always maximising their utility conditional on the already determined future behaviour.

A discounted incremental learning game, unlike the previous example of "Battle of the Sexes", cannot be solved using a standard backward induction procedure because it has infinitely many stages and no "last" stage to serve as a starting point. So, as an alternative to iterating through the game's stages, we will build a stationary equilibrium $(\mathbf{f}^*, \mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ by iterating through the game's states. Precisely, a formal description of this procedure requires an ordering $s_1, s_2, \ldots, s_n$ (with $n = n_1 n_2$) of the state space $\mathcal{S}$ and, for all $\ell = 1, 2, \ldots, n$, a method to find $(\mathbf{f}^*(s_\ell), \mathbf{g}^*(s_\ell))$ given an equilibrium of Γ_β restricted to $\{s_{\ell+1}, s_{\ell+2}, \ldots, s_n\}$.

First, to construct a suitable notion of "past" and "future" states, we want a sequence $s_1, s_2, \ldots, s_n$ of states such that

$$\ell' < \ell \quad \text{implies} \quad p(s_{\ell'} | s_\ell, a, b) = 0 \tag{4.7}$$

for every $\ell, \ell' = 1, 2, \ldots, n$ and $(a, b) \in \mathcal{A}(s_\ell) \times \mathcal{B}(s_\ell)$. The purpose of the condition in (4.7) is to ensure that a state s_ℓ can never transition to a preceding state $s_{\ell'}$. Equivalently, we want a topological ordering of the directed graph $D = (V, E)$ where $V = \mathcal{S}$ and

$$E = \Big\{(s, s') \in \mathcal{S} \times \mathcal{S} : s \neq s' \text{ and } p(s'|s, a, b) > 0 \text{ for some } (a, b) \in \mathcal{A}(s) \times \mathcal{B}(s)\Big\},$$

A *topological ordering* in D is a sequence of states such that s appears before s' for all $(s, s') \in E$. This sequence exists if and only if D is an acyclic directed graph [9]. Therefore, having constructed D to explicitly exclude self-loops, the structure of an incremental learning game (see Figure 4.1) suggests that a topological ordering is possible. Indeed, we can always use the lexicographic ordering $s_1, s_2, \ldots, s_n$ where, for any states $s_\ell = (i, j) \in \mathcal{S}$ and $s_{\ell'} = (i', j') \in \mathcal{S}$ with $\ell, \ell' = 1, 2, \ldots, n$, we have

$$\ell < \ell' \quad \text{if and only if} \quad (i < i') \text{ or } (i = i' \text{ and } j < j'). \tag{4.8}$$

It is straightforward to check that, when arranged in lexicographic order, the successors $(i+1, j)$, $(i, j+1)$, and $(i+1, j+1)$ must appear after (i, j).[5] So, assuming that

[5] Although a lexicographic ordering of the state space satisfies the required conditions, it is often possible to backward induct through the states in a different order. A general algorithm to compute topological orderings for an arbitrary acyclic directed graph is given by [9, Algorithm 6.11].

a suitable ordering has been selected, we will simplify our notation by relabelling the state s_ℓ as ℓ for every $\ell = 1, 2, \dots, n$.

Proposition 4.2. Let $(\mathbf{f}, \mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ be a strategy profile in the discounted incremental learning game Γ_β. Then, for any $\ell = 1, 2, \dots, n$ and $k = 1, 2$, we have

$$v_\beta^k(\ell, \mathbf{f}, \mathbf{g}) = \frac{r^k(\ell, \mathbf{f}, \mathbf{g}) + \beta \sum_{\ell'=\ell+1}^n v_\beta^k(\ell', \mathbf{f}, \mathbf{g}) p(\ell' | \ell, \mathbf{f}, \mathbf{g})}{1 - \beta p(\ell | \ell, \mathbf{f}, \mathbf{g})}. \tag{4.9}$$

Proof. Observe that, by conditioning on the outcome of the random variable S_1, the discounted value of $(\mathbf{f}, \mathbf{g})$ at ℓ can be written as

$$\begin{aligned} v_\beta^k(\ell, \mathbf{f}, \mathbf{g}) &= \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_t^k\right] \\ &= \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_0^k\right] + \sum_{t=0}^{\infty} \sum_{\ell'=1}^{n} \beta^t \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_t^k \middle| S_1 = \ell'\right] \mathbb{P}_{\ell \mathbf{f} \mathbf{g}}\left(S_1 = \ell'\right) \\ &= \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_0^k\right] + \sum_{\ell'=1}^{n} \mathbb{P}_{\ell \mathbf{f} \mathbf{g}}\left(S_1 = \ell'\right) \sum_{t=1}^{\infty} \beta^t \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_t^k \middle| S_1 = \ell'\right] \\ &= \mathbb{E}_{\ell \mathbf{f} \mathbf{g}}\left[R_0^k\right] + \beta \sum_{\ell'=1}^{n} \mathbb{P}_{\ell \mathbf{f} \mathbf{g}}\left(S_1 = \ell'\right) \sum_{t=1}^{\infty} \beta^t \mathbb{E}_{\ell' \mathbf{f} \mathbf{g}}\left[R_t^k\right] \\ &= r^k(\ell, \mathbf{f}, \mathbf{g}) + \beta \sum_{\ell'=1}^{n} v_\beta^k(\ell', \mathbf{f}, \mathbf{g}) p(\ell' | \ell, \mathbf{f}, \mathbf{g}). \end{aligned} \tag{i}$$

Of course, an immediate consequence of the ordering condition in (4.7) is that, for all $\ell' = 1, 2, \dots, \ell - 1$, we obtain

$$p(\ell' | \ell, \mathbf{f}, \mathbf{g}) = \sum_{a \in \mathcal{A}(\ell)} \sum_{b \in \mathcal{B}(\ell)} f(\ell, a) p(\ell' | \ell, a, b) g(\ell, b) = 0 \tag{ii}$$

and

$$v_\beta^k(\ell, \mathbf{f}, \mathbf{g}) = r^k(\ell, \mathbf{f}, \mathbf{g}) + \beta \sum_{\ell'=\ell}^{n} v_\beta^k(\ell', \mathbf{f}, \mathbf{g}) p(\ell' | \ell, \mathbf{f}, \mathbf{g}). \tag{iii}$$

Lastly, we can isolate the $v_\beta^k(\ell, \mathbf{f}, \mathbf{g})$ terms on the left-hand side of (iii) to obtain (4.9), as required. □

Fix an index $\ell = 1, 2, \dots, n$. We are allowed to restrict the discounted incremental learning game Γ_β to the smaller state space $\mathcal{S}_\ell = \{\ell, \ell + 1, \dots, n\}$. Why? The ordering condition in (4.7) guarantees that the excluded states cannot be accessed from within this restricted game. Thus, as in Proposition 4.2, a value can be assigned to incomplete strategy profiles that only specify a player's behaviour on $\mathcal{S}_\ell$. Precisely, *incomplete strategies* belonging to Player 1 and Player 2 are block row vectors

$$\mathbf{f}_\ell = \left(\mathbf{f}_\ell(\ell')\right)_{\ell'=\ell}^n \quad \text{and} \quad \mathbf{g}_\ell = \left(\mathbf{g}_\ell(\ell')\right)_{\ell'=\ell}^n$$

where, for all $\ell' = \ell, \ell + 1, \dots, n$, the blocks

$$\mathbf{f}_\ell(\ell') = \left(f_\ell(\ell', a)\right)_{a \in \mathcal{A}(\ell')} \quad \text{and} \quad \mathbf{g}_\ell(\ell') = \left(g_\ell(\ell', b)\right)_{b \in \mathcal{B}(\ell')}$$

are stochastic row vectors. The set of Player 1's incomplete strategies is denoted by $\mathbf{F}_\ell$ and the set of Player 2's incomplete strategies is denoted by $\mathbf{G}_\ell$. Clearly, using our previous observation, the discounted value $v_\beta^k(\ell',\mathbf{f}_\ell,\mathbf{g}_\ell)$ of $(\mathbf{f}_\ell,\mathbf{g}_\ell) \in \mathbf{F}_\ell \times \mathbf{G}_\ell$ to Player $k = 1,2$ is well-defined for every $\ell' = \ell, \ell+1, \dots, n$. An *incomplete equilibrium* $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*) \in \mathbf{F}_\ell \times \mathbf{G}_\ell$ satisfies

$$v_\beta^1\left(\ell',\mathbf{f}_\ell,\mathbf{g}_\ell^*\right) \le v_\beta^1\left(\ell',\mathbf{f}_\ell^*,\mathbf{g}_\ell^*\right) \quad \text{and} \quad v_\beta^2\left(\ell',\mathbf{f}_\ell^*,\mathbf{g}_\ell\right) \le v_\beta^2\left(\ell',\mathbf{f}_\ell^*,\mathbf{g}_\ell^*\right) \tag{4.10}$$

for all states $\ell' = \ell, \ell+1, \dots, n$, Player 1's alternatives $\mathbf{f}_\ell \in \mathbf{F}_\ell$, and Player 2's alternatives $\mathbf{g}_\ell \in \mathbf{G}_\ell$.

Now, since a backward induction algorithm builds an equilibrium by progressively adding to an incomplete equilibrium, some additional notation is needed to mathematically describe this process. Fix an index $\ell = 1,2,\dots,n-1$. A strategy $\mathbf{f}_\ell \in \mathbf{F}_\ell$ is said to *extend* $\mathbf{f}_{\ell+1} \in \mathbf{F}_{\ell+1}$ whenever $\mathbf{f}_\ell(\ell') = \mathbf{f}_{\ell+1}(\ell')$ for all $\ell' = \ell+1, \ell+2, \dots, n$. Similarly, a strategy $\mathbf{g}_\ell \in \mathbf{G}_\ell$ is said to *extend* $\mathbf{g}_{\ell+1} \in \mathbf{G}_{\ell+1}$ whenever $\mathbf{g}_\ell(\ell') = \mathbf{g}_{\ell+1}(\ell')$ for all $\ell' = \ell+1, \ell+2, \dots, n$. The sets of strategies that extend $\mathbf{f}_{\ell+1} \in \mathbf{F}_{\ell+1}$ and $\mathbf{g}_{\ell+1} \in \mathbf{G}_{\ell+1}$ are denoted by $\mathbf{F}_\ell(\mathbf{f}_{\ell+1})$ and $\mathbf{G}_\ell(\mathbf{g}_{\ell+1})$, respectively.

The modified backward induction procedure is outlined in Algorithm 4.3 using this terminology. Although the problem of finding a suitable extension in Step 2 is not resolved until Section 4.4, we assume that a method exists that is capable of computing these extensions. Below, Theorem 4.4 verifies that, depending on the choice of extensions, any stationary equilibrium of Γ_β can be returned by Algorithm 4.3.

Algorithm 4.3.

Input. An incremental learning game Γ_β with a state space $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$ satisfying the condition in (4.7).

Output. An equilibrium $(\mathbf{f}^*,\mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ of Γ_β.

Step 1. (*Initialisation*) Set $(\mathbf{f}_n^*,\mathbf{g}_n^*) \in \mathbf{F}_n \times \mathbf{G}_n$ such that $f_n^*(n,0) = 1$ and $g_n^*(n,0) = 1$.

Step 2. (*Extension*) Next, for each $\ell = n-1, n-2, \dots, 1$, find a strategy profile $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*) \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}) \times \mathbf{G}_\ell(\mathbf{g}_{\ell+1})$ satisfying the inequalities

$$v_\beta^1\left(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell^*\right) \le v_\beta^1\left(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell^*\right) \quad \text{and} \quad v_\beta^2\left(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell\right) \le v_\beta^2\left(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell^*\right) \tag{4.11}$$

for all $\mathbf{f}_\ell \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1})$ and $\mathbf{g}_\ell \in \mathbf{G}_\ell(\mathbf{g}_{\ell+1})$.

Step 3. (*Result*) Return $(\mathbf{f}_1^*,\mathbf{g}_1^*)$ as an equilibrium of Γ_β.

Theorem 4.4. Assume that we are able to find every solution to Step 2 in Algorithm 4.3. Then, a strategy profile $(\mathbf{f}^*,\mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ is an equilibrium of Γ_β if and only if it can be returned in Step 3.

Proof. We want to prove that, for any $\ell = 1,2,\dots,n$, the strategy profile $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*) \in \mathbf{F}_\ell \times \mathbf{G}_\ell$ is an equilibrium of Γ_β restricted to $\mathcal{S}_\ell$ if and only if it can be produced during the execution of Algorithm 4.3. First, in the base case ($\ell = n$), the only available actions are $\mathcal{A}(n) = \{0\}$ and $\mathcal{B}(n) = \{0\}$. Therefore, because the unique incomplete equilibrium $(\mathbf{f}_n^*,\mathbf{g}_n^*) \in \mathbf{F}_n \times \mathbf{G}_n$ has $f_n^*(n,0) = 1$ and $g_n^*(n,0) = 1$, the previous assertion holds trivially.

Second, taking an arbitrary index $\ell = n-1, n-2, \ldots, 1$, we will assume the validity of the inductive hypothesis for $\ell+1$; that is, $(\mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1}) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$ is an incomplete equilibrium if and only if it is produced during the execution of Algorithm 4.3. It remains to be shown that, under this assumption, the assertion also holds for the preceding state ℓ.

($\Longrightarrow$) Suppose $(\mathbf{f}^*_\ell, \mathbf{g}^*_\ell) \in \mathbf{F}_\ell \times \mathbf{G}_\ell$ is an incomplete equilibrium of the discounted incremental learning game. Define $(\mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1}) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$ such that

$$\mathbf{f}^*_{\ell+1}(\ell') = \mathbf{f}^*_\ell(\ell') \quad \text{and} \quad \mathbf{g}^*_{\ell+1}(\ell') = \mathbf{g}^*_\ell(\ell')$$

for all $\ell' = \ell+1, \ell+2, \ldots, n$. Note that $\mathbf{f}^*_\ell \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1})$ and $\mathbf{g}^*_\ell \in \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$. We know that $(\mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1})$ satisfies the incomplete equilibrium inequalities in (4.10) and, by the inductive hypothesis, it can be produced during the previous iteration of the extension procedure. Additionally, since $(\mathbf{f}^*_\ell, \mathbf{g}^*_\ell)$ satisfies the extension conditions in (4.11), it can be produced as a solution to Step 2, as required.

($\Longleftarrow$) Assume that, during the previous extension iteration, an incomplete strategy profile $(\mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1}) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$ is created. Obviously, by the inductive hypothesis, this strategy profile is an incomplete equilibrium of the discounted incremental learning game. Find a suitable extension $(\mathbf{f}^*_\ell, \mathbf{g}^*_\ell) \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1}) \times \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$. We already know that $(\mathbf{f}^*_\ell, \mathbf{g}^*_\ell)$ satisfies the equilibrium inequalities at $\ell+1, \ell+2, \ldots, n$, so it only remains to be shown that these conditions hold at the state ℓ. Observe that, when $\ell' = \ell+1, \ell+2, \ldots, n$, we have

$$v^1_\beta\big(\ell', \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big) \le v^1_\beta\big(\ell', \mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1}\big) = v^1_\beta\big(\ell', \mathbf{f}_\ell, \mathbf{g}^*_\ell\big) \tag{i}$$

for all $\mathbf{f}_\ell \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1})$ and $\mathbf{f}'_\ell \in \mathbf{F}_\ell$ with $\mathbf{f}_\ell(\ell) = \mathbf{f}'_\ell(\ell)$ and

$$v^2_\beta\big(\ell', \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big) \le v^2_\beta\big(\ell', \mathbf{f}^*_{\ell+1}, \mathbf{g}^*_{\ell+1}\big) = v^2_\beta\big(\ell', \mathbf{f}^*_\ell, \mathbf{g}_\ell\big) \tag{ii}$$

for all $\mathbf{g}_\ell \in \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$ and $\mathbf{g}'_\ell \in \mathbf{G}_\ell$ with $\mathbf{g}_\ell(\ell) = \mathbf{g}'_\ell(\ell)$. Then, as an immediate consequence of (i) and $\mathbf{f}_\ell(\ell) = \mathbf{f}'_\ell(\ell)$, it follows that

$$\begin{aligned} v^1_\beta\big(\ell, \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big) &= \frac{r^1\big(\ell, \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big) + \beta \sum_{\ell'=\ell+1}^n v^1_\beta\big(\ell', \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big) p\big(\ell' \big| \ell, \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big)}{1 - \beta p\big(\ell \big| \ell, \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big)} \\ &\le \frac{r^1\big(\ell, \mathbf{f}_\ell, \mathbf{g}^*_\ell\big) + \beta \sum_{\ell'=\ell+1}^n v^1_\beta\big(\ell', \mathbf{f}_\ell, \mathbf{g}^*_\ell\big) p\big(\ell' \big| \ell, \mathbf{f}_\ell, \mathbf{g}^*_\ell\big)}{1 - \beta p\big(\ell \big| \ell, \mathbf{f}_\ell, \mathbf{g}^*_\ell\big)} = v^1_\beta\big(\ell, \mathbf{f}_\ell, \mathbf{g}^*_\ell\big). \end{aligned} \tag{iii}$$

Analogously, as an immediate consequence of (ii) and $\mathbf{g}_\ell(\ell) = \mathbf{g}'_\ell(\ell)$, we obtain

$$\begin{aligned} v^2_\beta\big(\ell, \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big) &= \frac{r^2\big(\ell, \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big) + \beta \sum_{\ell'=\ell+1}^n v^2_\beta\big(\ell', \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big) p\big(\ell' \big| \ell, \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big)}{1 - \beta p\big(\ell \big| \ell, \mathbf{f}^*_\ell, \mathbf{g}'_\ell\big)} \\ &\le \frac{r^2\big(\ell, \mathbf{f}^*_\ell, \mathbf{g}_\ell\big) + \beta \sum_{\ell'=\ell+1}^n v^2_\beta\big(\ell', \mathbf{f}^*_\ell, \mathbf{g}_\ell\big) p\big(\ell' \big| \ell, \mathbf{f}^*_\ell, \mathbf{g}_\ell\big)}{1 - \beta p\big(\ell \big| \ell, \mathbf{f}^*_\ell, \mathbf{g}_\ell\big)} = v^2_\beta\big(\ell, \mathbf{f}^*_\ell, \mathbf{g}_\ell\big). \end{aligned} \tag{iv}$$

This means that, for any $\mathbf{f}'_\ell \in \mathbf{F}_\ell$ and $\mathbf{g}'_\ell \in \mathbf{F}_\ell$, the strategy profile $(\mathbf{f}^*_\ell, \mathbf{g}^*_\ell)$ satisfies the familiar equilibrium inequalities

$$v^1_\beta\big(\ell, \mathbf{f}'_\ell, \mathbf{g}^*_\ell\big) \le v^1_\beta\big(\ell, \mathbf{f}_\ell, \mathbf{g}^*_\ell\big) \le v^1_\beta\big(\ell, \mathbf{f}^*_\ell, \mathbf{g}^*_\ell\big) \tag{v}$$

and

$$v_\beta^2\big(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell'\big) \le v_\beta^2\big(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell\big) \le v_\beta^2\big(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell^*\big) \tag{vi}$$

where $\mathbf{f}_\ell \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*)$ and $\mathbf{g}_\ell \in \mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)$ are chosen such that $\mathbf{f}_\ell(\ell) = \mathbf{f}_\ell'(\ell)$ and $\mathbf{g}_\ell(\ell) = \mathbf{g}_\ell'(\ell)$. Hence, the extended strategy profile $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*)$ is an incomplete equilibrium of the incremental learning game.

We conclude by noting that $(\mathbf{f}_1^*,\mathbf{g}_1^*) \in \mathbf{F}_1 \times \mathbf{G}_1$, or equivalently $(\mathbf{f}^*,\mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$, is an equilibrium of Γ_β if and only if it can be returned at the termination of Algorithm 4.3. □

4.4 Finding Extensions

Lastly, before implementing Algorithm 4.3, we must identify a method for solving the inequalities in (4.11). Suppose that the modified backward induction algorithm has reached the current state $s_\ell = (i,j)$ with $\ell = 1,2,\ldots,n-1$ and, during previous iterations of the extension method, has produced an incomplete equilibrium $(\mathbf{f}_{\ell+1}^*,\mathbf{g}_{\ell+1}^*) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$. Player 1's set of available actions is either $\mathcal{A}(\ell) = \{0\}$ or $\mathcal{A}(\ell) = \{0,1\}$ and Player 2's set of available actions is either $\mathcal{B}(\ell) = \{0\}$ or $\mathcal{B}(\ell) = \{0,1\}$. So, the generic extensions $\mathbf{f}_\ell \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*)$ and $\mathbf{g}_\ell \in \mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)$ are entirely determined by the probabilities

$$f_\ell(\ell,0) = p_0 \in [0,1] \quad \text{and} \quad g_\ell(\ell,0) = q_0 \in [0,1]$$

of forgoing learning. Define $p_1 = 1 - p_0$ and $q_1 = 1 - q_0$ for the sake of notational compactness. Also, letting $s_{a,b} = (i+a, j+b)$ for all $a \in \mathcal{A}(\ell)$ and $b \in \mathcal{B}(\ell)$, define some additional auxiliary quantities

$$V_{a,b}^k = \begin{cases} r^k(\ell,a,b) + \beta v_\beta^k\big(s_{a,b},\mathbf{f}_{\ell+1},\mathbf{g}_{\ell+1}\big), & (a,b) \ne (0,0), \\ r^k(\ell,a,b), & (a,b) = (0,0), \end{cases} \tag{4.12}$$

for all $k = 1,2$ and $(a,b) \in \mathcal{A}(\ell) \times \mathcal{B}(\ell)$. Lemma 4.5 expresses the discounted value of $(\mathbf{f}_\ell,\mathbf{g}_\ell)$ in terms of the implemented strategies. Note that, since the expressions in (4.12) do not depend on the current strategies $\mathbf{f}_\ell(\ell)$ or $\mathbf{g}_\ell(\ell)$, they can be treated as constants throughout the process of extending the incomplete equilibrium.

Lemma 4.5. Fix a state $\ell = 1,2,\ldots,n-1$ and an incomplete equilibrium $(\mathbf{f}_{\ell+1}^*,\mathbf{g}_{\ell+1}^*) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$. Then,

$$v_\beta^k(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) = \frac{1}{1-\beta p_0 q_0} \sum_{a\in\mathcal{A}(\ell)} \sum_{b\in\mathcal{B}(\ell)} p_a V_{a,b}^k q_b \tag{4.13}$$

is the discounted value of $(\mathbf{f}_\ell,\mathbf{g}_\ell) \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*) \times \mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)$ to Player $k = 1,2$.

Proof. Recall from (4.2) that the expected immediate utility of $(\mathbf{f}_\ell,\mathbf{g}_\ell)$ to Player $k = 1,2$ is

$$r^k(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) = \sum_{a\in\mathcal{A}(\ell)} \sum_{b\in\mathcal{B}(\ell)} f_\ell(\ell,a) r^k(\ell,a,b) g_\ell(\ell,b) = \sum_{a\in\mathcal{A}(\ell)} \sum_{b\in\mathcal{B}(\ell)} p_a r^k(\ell,a,b) q_b \tag{i}$$

and from (4.3) that the transition probabilities are

$$p(\ell'|\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) = \begin{cases} f_\ell(\ell,a)g_\ell(\ell,b), & i' = i+a \text{ for some } a \in \mathcal{A}(\ell) \text{ and} \\ & j' = j+b \text{ for some } b \in \mathcal{B}(\ell), \\ 0, & \text{otherwise,} \end{cases}$$
$$= \begin{cases} p_a q_b, & i' = i+a \text{ for some } a \in \mathcal{A}(\ell) \text{ and} \\ & j' = j+b \text{ for some } b \in \mathcal{B}(\ell), \\ 0, & \text{otherwise,} \end{cases} \tag{ii}$$

for all $\ell' = 1,2,\ldots,n$ with $s_\ell = (i,j)$ and $s_{\ell'} = (i',j')$. Observe that, after considering the possible actions that can be selected to reach the future states, we have

$$\sum_{\ell'=\ell+1}^{n} v_\beta^k(\ell',\mathbf{f}_\ell,\mathbf{g}_\ell)p(\ell'|\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) = \sum_{\substack{a\in\mathcal{A}(\ell) \\ (a,b)\neq(0,0)}}\sum_{b\in\mathcal{B}(\ell)} p_a v_\beta^k(s_{a,b},\mathbf{f}_\ell,\mathbf{g}_\ell)q_b. \tag{iii}$$

Substitute (i), (ii), and (iii) into the expression for the discounted value of $(\mathbf{f}_\ell,\mathbf{g}_\ell)$ in (4.9) to obtain

$$\begin{aligned} v_\beta^k(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) &= \frac{r^k(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) + \beta\sum_{\ell'=\ell+1}^{n} v_\beta^k(\ell',\mathbf{f}_\ell,\mathbf{g}_\ell)p(\ell'|\ell,\mathbf{f}_\ell,\mathbf{g}_\ell)}{1-\beta p(\ell|\ell,\mathbf{f}_\ell,\mathbf{g}_\ell)} \\ &= \frac{1}{1-\beta p_0 q_0}\left(\sum_{a\in\mathcal{A}(\ell)}\sum_{b\in\mathcal{B}(\ell)} p_a r^k(\ell,a,b)q_b + \beta\sum_{\substack{a\in\mathcal{A}(\ell) \\ (a,b)\neq(0,0)}}\sum_{b\in\mathcal{B}(\ell)} p_a v_\beta^k(s_{a,b},\mathbf{f}_\ell,\mathbf{g}_\ell)q_b\right) \\ &= \frac{1}{1-\beta p_0 q_0}\sum_{a\in\mathcal{A}(\ell)}\sum_{b\in\mathcal{B}(\ell)} p_a V_{a,b}^k q_b. \end{aligned} \tag{iv}$$

Clearly, (iv) gives the desired valuation of the strategy profile $(\mathbf{f}_\ell,\mathbf{g}_\ell)$ to Player $k = 1,2$. □

Next, we will reformulate the inequalities in (4.11) to develop a coupled pair of equivalent maximisation problems. If $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*) \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*)\times\mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)$ is an incomplete equilibrium, then we will write

$$f_\ell^*(\ell,0) = p_0^* \in [0,1] \quad \text{and} \quad g_\ell^*(\ell,0) = q_0^* \in [0,1]$$

with $p_1^* = 1-p_0^*$ and $q_1^* = 1-q_0^*$. Evidently, from the valuation of player strategies in Lemma 4.5, the strategy profile $(\mathbf{f}_\ell^*,\mathbf{g}_\ell^*)$ satisfies (4.11) if and only if the probabilities p_0^* and q_0^* simultaneously solve the maximisation problems

$$p_0^* = \operatorname*{arg\,max}_{\mathbf{f}_\ell\in\mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*)} v_\beta^1\big(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell^*\big) = \operatorname*{arg\,max}_{p_0=1-p_1\in[0,1]} \frac{1}{1-\beta p_0 q_0^*}\sum_{a\in\mathcal{A}(\ell)}\sum_{b\in\mathcal{B}(\ell)} p_a V_{a,b}^1 q_b^* \tag{4.14}$$

and

$$q_0^* = \operatorname*{arg\,max}_{\mathbf{g}_\ell\in\mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)} v_\beta^2\big(\ell,\mathbf{f}_\ell^*,\mathbf{g}_\ell\big) = \operatorname*{arg\,max}_{q_0=1-q_1\in[0,1]} \frac{1}{1-\beta p_0^* q_0}\sum_{a\in\mathcal{A}(\ell)}\sum_{b\in\mathcal{B}(\ell)} p_a^* V_{a,b}^2 q_b. \tag{4.15}$$

A method for calculating solutions to (4.14) and (4.15) is provided in Algorithm 4.6 and verified in Proposition 4.7. Note that a *best response* is a strategy that maximises the player's utility conditional on their opponent's selected strategy. We claim that this method computes every solution when the discounted incremental learning game is *non-degenerate* at the state ℓ, that is, when every pure strategy has no completely mixed best responses. This assumption ensures that there are finitely many solutions, which allows us to exhaustively find every equilibrium.[6]

Algorithm 4.6.

Input. A state $\ell = 1,2,\ldots,n-1$ and an incomplete equilibrium $(\mathbf{f}^*_{\ell+1},\mathbf{g}^*_{\ell+1}) \in \mathbf{F}_{\ell+1} \times \mathbf{G}_{\ell+1}$.

Output. A set of strategy profiles $(\mathbf{f}^*_\ell,\mathbf{g}^*_\ell) \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1}) \times \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$ that solve (4.11).

Step 1. (*Pure Strategies*) Return every strategy profile $(\mathbf{f}^*_\ell,\mathbf{g}^*_\ell) \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1}) \times \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$ such that

$$p_0^* = \operatorname*{arg\,max}_{p_0=1-p_1\in\{0,1\}} \frac{1}{1-\beta p_0 q_0^*} \sum_{a\in\mathcal{A}(\ell)} \sum_{b\in\mathcal{B}(\ell)} p_a V_{a,b}^1 q_b^* \tag{4.16}$$

and

$$q_0^* = \operatorname*{arg\,max}_{q_0=1-q_1\in\{0,1\}} \frac{1}{1-\beta p_0^* q_0} \sum_{a\in\mathcal{A}(\ell)} \sum_{b\in\mathcal{B}(\ell)} p_a^* V_{a,b}^2 q_b. \tag{4.17}$$

Observe that, because $\mathbf{f}^*_\ell$ and $\mathbf{g}^*_\ell$ are pure strategies, these conditions can be checked by exhaustion.

Step 2. (*Mixed Strategies*) Return every strategy profile $(\mathbf{f}^*_\ell,\mathbf{g}^*_\ell) \in \mathbf{F}_\ell(\mathbf{f}^*_{\ell+1}) \times \mathbf{G}_\ell(\mathbf{g}^*_{\ell+1})$ such that

$$\sum_{a\in\mathcal{A}(\ell)} \left(p_a^* V_{a,0}^2 - p_a^*(1-\beta p_0^*) V_{a,1}^2\right) = 0 \quad \text{and} \quad \sum_{b\in\mathcal{B}(\ell)} \left(q_b^* V_{0,b}^1 - q_b^*(1-\beta q_0^*) V_{1,b}^1\right) = 0 \tag{4.18}$$

for some $p_0^* = 1-p_1^* \in (0,1)$ and $q_0^* = 1-q_1^* \in (0,1)$. These quadratic equations in p_0^* and q_0^* can be solved via the quadratic formula.

Proposition 4.7. If Γ_β is non-degenerate at the present state $\ell = 1,2,\ldots,n$, then there are finitely many solutions to the extension conditions in (4.11). Moreover, every solution is returned by Algorithm 4.6.

Proof. First, by the assumption that Γ_β is non-degenerate, a pure strategy can never have a best response in completely mixed strategies. It follows that a pure strategy solution to (4.11) can be found after only checking the necessary inequalities over the space of other pure strategies. Hence, every pure strategy solution—of which there are finitely many—is returned during Step 1 of Algorithm 4.6.

Second, given that the extension conditions in (4.11) have been reduced to a pair of maximisation problems in (4.14) and (4.15), it is useful to compute the partial derivatives of $v_\beta^k(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell)$ with respect to p_0 and q_0. So, by applying the quotient rule to the strategy valuation from Lemma 4.5, we obtain

$$\frac{\partial}{\partial p_0} v_\beta^1(\ell,\mathbf{f}_\ell,\mathbf{g}_\ell) = \frac{1}{(1-\beta p_0 q_0)^2} \sum_{b\in\mathcal{B}(\ell)} \left(q_b V_{0,b}^1 - q_b(1-\beta q_0) V_{1,b}^1\right) \tag{i}$$

[6] The assumption of non-degeneracy is used in several game-theoretic algorithms, including the support enumeration algorithm that inspires Algorithm 4.6 (see, for example, [27, Algorithm 3.4]).

and
$$\frac{\partial}{\partial q_0} v_\beta^2(\ell, \mathbf{f}_\ell, \mathbf{g}_\ell) = \frac{1}{(1-\beta p_0 q_0)^2} \sum_{a \in \mathcal{A}(\ell)} \Big(p_a V_{a,0}^2 - p_a (1-\beta p_0) V_{a,1}^2 \Big). \tag{ii}$$

The signs and zeros of these partial derivatives are entirely determined by the opponent's choice of strategy. Explicitly, q_0^* and p_0^* are roots of (i) and (ii) if and only if

$$\sum_{a \in \mathcal{A}(\ell)} \Big(p_a^* V_{a,0}^2 - p_a^* (1-\beta p_0^*) V_{a,1}^2 \Big) = 0 \quad \text{and} \quad \sum_{b \in \mathcal{B}(\ell)} \Big(q_b^* V_{0,b}^1 - q_b^* (1-\beta q_0^*) V_{1,b}^1 \Big) = 0. \tag{iii}$$

Clearly, under the conditions in (iii), the partial derivative of Player 1's valuation in (i) is zero regardless of p_0 and the partial derivative of Player 2's valuation in (ii) is zero regardless of q_0. This guarantees that, for all $\mathbf{f}_\ell \in \mathbf{F}_\ell(\mathbf{f}_{\ell+1}^*)$ and $\mathbf{g}_\ell, \mathbf{G}_\ell(\mathbf{g}_{\ell+1}^*)$, we have

$$v_\beta^1\Big(\ell, \mathbf{f}_\ell, \mathbf{g}_\ell^*\Big) = v_\beta^1\Big(\ell, \mathbf{f}_\ell^*, \mathbf{g}_\ell^*\Big) \quad \text{and} \quad v_\beta^2\Big(\ell, \mathbf{f}_\ell^*, \mathbf{g}_\ell\Big) = v_\beta^2\Big(\ell, \mathbf{f}_\ell^*, \mathbf{g}_\ell^*\Big). \tag{iv}$$

Thus, since the strategy profile $(\mathbf{f}_\ell^*, \mathbf{g}_\ell^*)$ satisfies the extension conditions whenever p_0^* and q_0^* solve (iii), every completely mixed solution to (4.11) is returned in Step 2 of Algorithm 4.6.

Lastly, to show that there are finitely many solutions in completely mixed strategies, notice that the existence of infinitely many solutions requires either p_0^* or q_0^* to solve (iii) for every $p_0^* \in [0,1]$ or $q_0^* \in [0,1]$. If this requirement holds for all $p_0^* \in [0,1]$, then Player 2 is always indifferent between their actions irrespective of Player 1's strategy. This contradicts our assumption of non-degeneracy because Player 2 has completely mixed best responses to Player 1's pure strategy choices: either $p_0^* = 0$ and $p_0^* = 1$. Therefore, there are only finitely many solutions to the quadratic equations in Step 2. □

Note that, as a consequence of its validity, we can use Algorithm 4.6 to find the solutions to (4.11) in Algorithm 4.3. The fact that we are capable of giving every suitable solution to these inequalities means that a backward induction procedure can exhaustively compute the equilibria of a non-degenerate incremental learning game. In particular, to find every equilibrium, we can simply branch Algorithm 4.3 whenever multiple solutions are returned from Algorithm 4.6. Unfortunately, because there is an uncountably infinite number of solutions to (4.11) in a degenerate game, we cannot possibly return every equilibrium. Instead, a compromise might involve only taking the pure strategy solutions whenever a continuum of solutions exists. The resulting backward induction procedure, with this compromise included, is implemented using Python in `incremental_solver.py`.

4.5 Incremental Learning in Tennis

Finally, to apply the recently described backward induction algorithm and illustrate an example of equilibrium behaviour in incremental learning games, we will look at a simple tennis game borrowed form [1, Section 4.2]. Consider a repeated sequence of tennis points in which Player 1 and Player 2 can select from among the actions

"Good Shot", "Safe Shot", and "Out". After Player 1 and Player 2 select a pair of actions, the probability of winning the point is determined by Figure 4.3.

Assume that, to convert these probabilities into a matrix game, a player is given -100 utility after losing a point and a player is given 100 utility after winning a point. Clearly, this situation can be expressed as a 3×3 matrix game G with the utility matrix

$$R = \begin{pmatrix} 0 & 40 & 100 \\ -40 & 0 & 100 \\ -100 & -100 & 0 \end{pmatrix}.$$

Unsurprisingly, the unique equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$ of the competent game G has $\mathbf{x}^* = (1,0,0)$ and $\mathbf{y}^* = (1,0,0)$; that is, both Player 1 and Player 2 should always select "Good Shot". Suppose that the incompetence of Player 1 is parameterised by $Q_1 : [0,1] \to \mathbb{R}^{3\times 3}$ where, for all $\lambda \in [0,1]$, we have

$$Q_1(\lambda) = \begin{pmatrix} 3/10 & 1/10 & 3/5 \\ 1/10 & 3/5 & 3/10 \\ 0 & 0 & 1 \end{pmatrix} (1-\lambda) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \lambda$$

and the incompetence of Player 2 is parametersied by $Q_2 : [0,1] \to \mathbb{R}^{3\times 3}$ where, for all $\mu \in [0,1]$, we have

$$Q_2(\mu) = \begin{pmatrix} 3/10 & 1/10 & 3/5 \\ 1/10 & 3/5 & 3/10 \\ 0 & 0 & 1 \end{pmatrix} (1-\mu) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mu.$$

Although an incompetent player is more likely to accidentally execute "Out" when they select "Good Shot" than when they select "Safe Shot", this is compensated for by having "Good Shot" being more likely to win than "Safe Shot". The game value of the resulting parameterised incompetent game is shown in Figure 4.4. Now, to create a discounted incremental learning game Γ_β from this tennis game, we allow Player 1 to attain the learning parameters

$$\Gamma = \left\{ \lambda_i = \frac{i-1}{5} : i = 1, 2, \ldots, 6 \right\}$$

and Player 2 to attain the learning parameters

$$\mathrm{M} = \left\{ \mu_j = \frac{j-1}{5} : j = 1, 2, \ldots, 6 \right\}.$$

		Player 2		
		Good Shot	Safe Shot	Out
Player 1	Good Shot	50%, 50%	70%, 30%	100%, 0%
	Safe Shot	30%, 70%	50%, 50%	100%, 0%
	Out	0%, 100%	0%, 100%	50%, 50%

Figure 4.3. The probabilities of Player 1 and Player 2 winning a tennis point for each combination of actions.

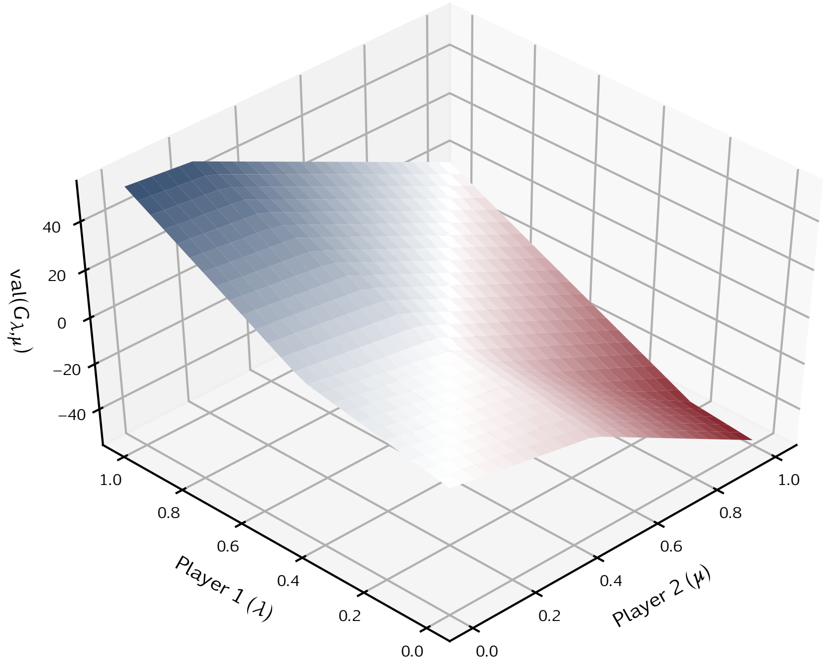
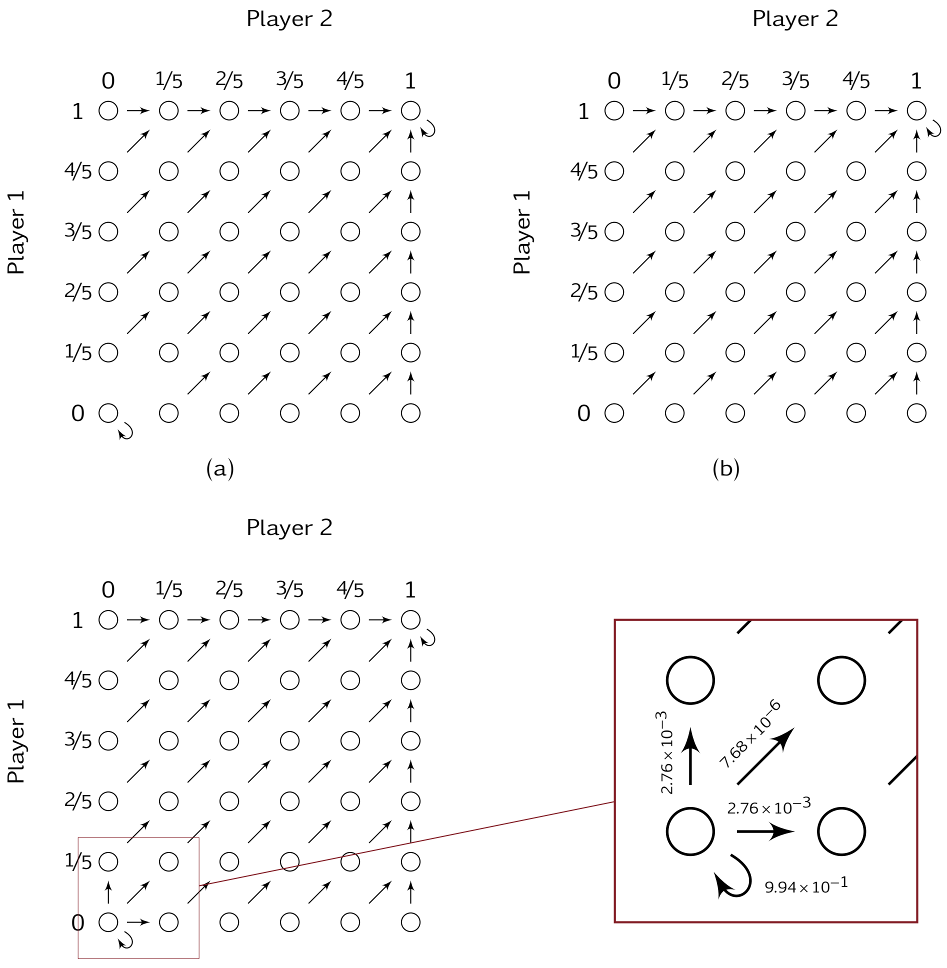


Figure 4.4. The dependence of the game value val($G_{\lambda,\mu}$) on learning parameters $\lambda, \mu \in [0,1]$ for a tennis point with parameterised incompetence. Generated using `incompetent_game_plot.py`.

Additionally, for any $i = 1, 2, \ldots, 6$ and $j = 1, 2, \ldots, 6$, the learning cost at state (i, j) is $c^1(i, j) = 10$ and $c^2(i, j) = 10$ for Player 1 and Player 2, respectively. The future rewards are progressively discounted by a discount factor of $\beta = {}^{19}\!/_{20}$.

Next, applying the aforementioned backward induction algorithm to the tennis game with incremental learning, we find that there are a total of three equilibria, which are shown in Figure 4.5. A node indicates a pair of learning parameters and an arc represents a transition realised by the equilibrium. So, a vertical arrow means that only Player 1 learns, a horizontal arrow means that only Player 2 learns, a diagonal arrow means that both players learn, and a loop means that neither player learns. An equilibrium in mixed strategies is shown in Figure 4.5(c) and the probabilities of each transition are included.

The only differences between the equilibria in Figure 4.5 are the strategies employed at the initial state $(0, 0)$ and, at the remaining states, the players always choose to learn whenever possible. Namely, at the initial state, neither player learns in Figure 4.5(a), both players learn in Figure 4.5(b), and learning is unlikely in Figure 4.5(c). How can these seemingly opposite equilibria occur within the same game? The equilibrium shown in Figure 4.5(a) arises because, when your opponent initially chooses not to learn, forgoing the benefits of learning is preferable to paying the immediate learning cost. Similarly, the equilibrium shown in Figure 2.4(b) arises because, when your opponent initially chooses to learn, paying



(a)

(b)

(c)

Figure 4.5. The learning strategies under various equilibria in a simple tennis game with incremental learning. Computed using `incremental_solver.py`.

the immediate learning cost is preferable to forgoing the benefits of learning.

Here, we encounter a problem that can often arise in general-sum games with multiple equilibria; the concept of a Nash equilibrium cannot be used alone to recommend a single strategy for each player. Recall that a similar observation was made about "Battle of the Sexes" (shown in Figure 2.3) whose three equilibria produced different expected utilities. Still, does the incremental learning game have an equilibrium that both players would prefer over the alternatives? This question is answered by comparing the discounted values of each equilibrium in Table 4.1. Notice that the discounted value awarded to both players is greatest when using the equilibrium in Figure 4.5(a) or, in other words, this equilibrium is *Pareto superior* to the remaining alternatives. Precisely, a strategy profile $(\mathbf{f}^*, \mathbf{g}^*) \in \mathbf{F} \times \mathbf{G}$ is Pareto superior to an alternative strategy profile $(\mathbf{f} \times \mathbf{g}) \in \mathbf{F} \times \mathbf{G}$ whenever, for every

Equilibrium	Player 1 Discounted Value	Player 2 Discounted Value
Figure 4.5(a)	0	0
Figure 4.5(b)	−45.24	−45.24
Figure 4.5(c)	−4.42	−4.42

Table 4.1. The discounted values of various equilibria in a simple tennis game with incremental learning.

$k = 1, 2$, we have

$$\mathbf{v}_\beta^k(\mathbf{f}^*, \mathbf{g}^*) \geq \mathbf{v}_\beta^k(\mathbf{f}, \mathbf{g}),$$

with a strict inequality holding for some $k = 1, 2$ [19]. Thus, since the players both prefer the equilibrium in Figure 2.4(a), we might recommend that Player 1 and Player 2 choose not to improve upon their initial levels of incompetence. Note that, as is the case with any general-sum equilibrium, this recommendation relies on a player's opponent following the same reasoning.

5 Conclusion

Adopting the perspective of normative game theory, we sought to address the strategic considerations arising from execution skill, or the possibility that a player accidentally deviates from their intended action. The mathematical notion of an incompetent matrix game, which was introduced by Beck and Filar [2], captured these accidental deviations as probability distributions over a player's available actions. We explored the properties of incompetent games in Chapter 3 and the "best" learning strategies in Chapter 4.

First, when exploring properties of incompetent games in Chapter 3, we focused on two interesting features: game value plateaus and optimal learning parameters. These problems were addressed by viewing incompetence as modifying each player's strategy space—to create the space of executable strategies—instead of modifying each player's expected utility. Accordingly, we were able to show that rectangular game value plateaus (as in Figure 2.4) arise whenever a pair of learning parameters create a completely mixed incompetent game. Additionally, under the assumption that complete competence is achievable, we proved that a learning parameter is optimal if and only if the corresponding space of executable strategies contains a competent optimal strategy. However, the established optimality conditions do not always apply and, as a consequence, further investigation might explore learning parameter choices in situations where competent optimal strategies are never executable. Furthermore, since the multi-stage game from Section 3.3 was free of cost, a possible extension could involve adding learning costs and finding properties of the resulting equilibria.

Second, in Chapter 4, we attempted to find the "best" learning strategies for playing incompetent games. An incremental learning model was proposed that required two players to repeatedly play an incompetent game with opportunities to increment their learning parameters between stages. We were able to leverage the resulting transition structure to apply a backward induction algorithm and exhaustively compute equilibria. Precisely, while traditional backward induction cannot be applied because incremental learning unfolds over infinitely many stages, we instead iterate through the game's finite collection of states. A simple tennis game was shown as an example of the potential strategic insights gained from understanding the equilibria in incremental learning games. Of course, given that we were primarily focused on finding methods to solve these games, we did not attempt to characterise structural patterns present in their equilibria. So, future research might explore the properties of the "best" learning strategies with the goal of developing easy-to-apply learning heuristics.

Overall, we have demonstrated the usefulness of the incompetence concept in modelling execution skill and expanded the scope of learning dynamics to include incremental learning. We provided insight into various properties of incompetence using executable strategies and insight into incremental learning strategies using backward induction. Hopefully, through continued incorporation of skill, the recommendations obtained from game-theoretic analysis will become more realistic and robust.

References

[1] J. Beck. *Incompetence, Training and Changing Capabilities in Game Theory*. PhD thesis, University of South Australia, Adelaide, Australia, 2013.

[2] J. Beck and J. A. Filar. Games, Incompetence, and Training. In S. Jørgensen, M. Quincampoix, and T. L. Vincent, editors, *Advances in Dynamic Game Theory*, Annals of the International Society of Dynamic Games, pages 93–110. Birkhäuser, Boston, MA, 2007.

[3] J. D. Beck. Game Theory Implementation of Capability Investment Problem. *Military Operations Research*, 16(1):41–55, 2011.

[4] J. D. Beck, V. Ejov, and J. A. Filar. Incompetence and Impact of Training in Bimatrix Games. *Automatica*, (10):2400–2408, 2012.

[5] E. Borel. On Games that Involve Chance and the Skill of the Players. Translated by L. J. Savage. *Econometrica*, 21(1):101–115, 1953.

[6] E. Borel. On Systems of Linear Forms of Skew Symmetric Determinant and the General Theory of Play. Translated by L. J. Savage. *Econometrica*, 21(1):116–117, 1953.

[7] E. Borel. The Theory of Play and Integral Equations with Skew Symmetric Kernels. Translated by L. J. Savage. *Econometrica*, 21(1):97–100, 1953.

[8] R. Dimand and M. A. Dimand. The Early History of the Theory of Strategic Games from Waldegrave to Borel. *History of Political Economy*:15–27, 1992.

[9] K. Erciyes. *Guide to Graph Algorithms*. Texts in Computer Science. Springer, 2018.

[10] J. A. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, New York, NY, 1997.

[11] A. M. Fink. Equilibrium in a Stochastic n-Person Game. *Journal of Science of the Hiroshima University*, 28(1):89–93, 1964.

[12] M. J. M. Jansen. Regularity and Stability of Equilibrium Points of Bimatrix Games. *Mathematics of Operations Research*, 6(4):530–550, 1981.

[13] I. Kaplansky. A Contribution to Von Neumann's Theory of Games. *Annals of Mathematics*, 46(3):474–479, 1945.

[14] M. Kleshnina, J. A. Filar, V. Ejov, and J. C. McKerral. Evolutionary Games under Incompetence. *Journal of Mathematical Biology*, 77(3):627–646, 2018.

[15] M. Kleshnina, S. S. Streipert, J. A. Filar, and K. Chatterjee. Prioritised Learning in Snowdrift-Type Games. *Mathematics*, 8(11), 2020.

[16] P. Larkey, J. B. Kadane, R. Austin, and S. Zamir. Skill in Games. *Management Science*, 43(5), 1997.

[17] M. Maschler, E. Solan, and S. Zamir. *Game Theory*. M. Borns, editor. Translated by Z. Hellman. Cambridge University Press, Cambridge, UK, 2013.

[18] J. F. Nash. Equilibrium Points in n-Person Games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950.

[19] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, MA, 1994.

[20] G. Owen. *Game Theory*. Emerald Group Publishing, Bingley, United Kingdom, 4th edition, 2013.

[21] T. C. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, 1980.

[22] R. Selten. Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory*, 4(1):25–55, 1975.

[23] L. S. Shapley. Stochastic Games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100, 1953.

[24] V. Soltan. *Lectures on Convex Sets*. World Scientific Publishing, Singapore, SG, 2020.

[25] J. von Neumann. On the Theory of Games of Strategy. In A. W. Tucker and R. D. Luce, editors. Translated by S. Bargmann, *Contributions to the Theory of Games*. Volume 4, number 40 in Annals of Mathematics Studies, pages 13–42. Princeton University Press, Princeton, NJ, 1959.

[26] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. With an introduction by W. H. Kuhn. With an afterword by A. Rubinstein. Princeton Classic Editions. Princeton University Press, Princeton, NJ, 60th anniversary edition, 2004.

[27] B. von Stengel. Equilibrium Computation for Two-Player Games in Strategic and Extensive Form. In N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*. Cambridge University Press, New York, NY, 2007.

[28] M. Wooldridge. Does Game Theory Work? *IEEE Intelligent Systems*, 27(6), 2012.