Εθνικο Μετσοβιο Πολυτεχνειο
Τμημα Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων

ΤΟΜΕΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΛΟΓΙΚΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ

Θεωρία Παιγνίων και Πολυπλοκότητα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του Γεωργίου Πήλιουρα

Επιβλέπων: Ευστάθιος Ζάχος
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2004

Εθνικο Μετσοβιο Πολυτεχνειο
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΛΟΓΙΚΗΣ ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ

Θεωρία Παιγνίων και Πολυπλοκότητα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του Γεωργίου Πήλιουρα

Επιβλέπων: Ευστάθιος Ζάχος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11\10\2004

....	....	....
Ευστάθιος Ζάχος Καθηγητής	Φώτω Αφράτη Καθηγήτρια	Τίμος Σελλής Καθηγητής

Αθήνα, (10,2004).

....
(Γεώργιος Α. Πήλιουρας)
(Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών)

Περίληψη

Η παρούσα διπλωματική, με αφετηρία την αστοχία της κλασσικής θεωρίας παιγνίων να ερμηνεύσει ορισμένα σενάρια αλληλεπίδρασης μεταξύ "λογικών" οντοτήτων, επιχειρεί να παρουσιάσει με τρόπο προσιτό την πορεία της έρευνας που επακολούθησε την συνειδητοποίηση των περιορισμένων ικανοτήτων μοντελοποίησης της κλασσικής θεωρίας παιγνίων.

Κατά την διάρκεια της προσπάθειάς μας αυτής εφαρμόζουμε διάφορες έννοιες που επισυνάπτονται της θεωρίας πολυπλοκότητας στο πλαίσιο της θεωρίας παιγνίων και εξετάζουμε το βαθμό κατά τον οποίο αυτές επιτυγχάνουν την υπέρβαση των δεδομένων σκοπέλων και την ικανοποιητική ερμηνεία των διαφόρων παραδόξων και κυρίως αυτό του Διλήμματος των Φυλακισμένων.

Συγκεκριμένα ξεκινούμε με μία μικρή εισαγωγή στην θεωρία παιγνίων και συγκεκριμένα στην έννοια της Nash ισορροπίας (Κεφ. 2). Στην συνέχεια παραθέτουμε διάφορα παραδείγματα που αποδεικνύουν την ισχύ της ως σκεπτικό λύσης, ενώ κλείνουμε το κεφάλαιο παρουσιάζοντας το Δίλημμα των Φυλακισμένων.

Στο επόμενο κεφάλαιο (Κεφ. 3) αναπτύσσουμε την έννοια των επαναλαμβανόμενων παιγνίων, αναδεικνύοντας την ικανότητα τους να μοντελοποιούν πολύπλοκες συμπεριφορές, αλλά και ταυτόχρονα καθορίζοντας τους περιορισμούς τους που συνεπάγεται η απουσία της έννοιας της πολυπλοκότητας από το πλαίσιό τους.

Στο τέταρτο κεφάλαιο προχωρούμε στη σύνθεση των ιδεών της θεωρίας παιγνίων και της θεωρίας πολυπλοκότητας, βελτιώνοντας έτσι τα αποτελέσματα της θεωρίας των επαναλαμβανόμενων παιγνίων.

Στο πέμπτο κεφάλαιο παραθέτουμε διάφορες προσεγγίσεις που έχουν χρησιμοποιηθεί για την επίλυση γνωστών παραδόξων, αλλά και αναπτύσσουμε την πλήρη απόδειξη του πεπερασμένου Διλήμματος των Φυλακισμένων.

Τέλος, κλείνουμε την εργασία με μία εισαγωγή σε ζητήματα υπολογισιμότητας και εξετάζουμε κατά πόσο είναι εφικτή μία σύνθεση τους με τις έννοιες της θεωρίας παιγνίων.

Λέξεις κλειδιά: θεωρία παιγνίων, πολυπλοκότητα, επαναλαμβανόμενα παίγνια, αυτόματα, μηχανές Turing.

Abstract

The present thesis, in response to the failure of classical game theory to interpret certain scenarios of interaction amongst rational players (such as the Prisoner's Dilemma), undertakes the task of presenting the line of research that followed the disturbing awareness of the limited modelling scope of the classical game theory.

The effort of resolving such paradoxes led to the successful combination of several ideas from two apparently diverse fields, that of game theory and complexity theory, that led to a more sophisticate examination of the paradoxes in question.

Namely, we begin by introducing several basic game theoretic concepts (Chapter 2) and specifically the notion of Nash equilibria. We resume by proving the generality of the Nash equilibria concept via the use of examples and finally we present the Prisoner's Dilemma paradox.

In the following chapter (Chapter 3) we cover the area of repeated games, presenting its power in representing real life scenarios as well as several problems that arise due to the lack of the notion of complexity in it.

In Chapter 4 we proceed with the strengthening of certain game theoretic results by applying ideas from complexity theory.

Next in Chapter 5, we present several methods that have been applied in a strenuous effort to resolve finite horizon paradoxes and we provide the complete proof to the finite Prisoner's Dilemma.

Lastly in Chapter 6, we attempt a brief introduction to certain computability issues and explore the possibility of utilizing them into game theory.

Keywords: game theory, complexity, bounded rationality, repeated games, automata, Turing machines.

στους γονείς μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω την οικογένειά μου για την αμέριστη συμπαράσταση στην πορεία μου. Τους καθηγητές μου, Στάθη Ζάχο για την γνήσια αγάπη που μου ενέπνευσε για τη θεωρητική πληροφορική και Τίμο Σελλή για την υποστήριξή του καθόλη την διάρκεια των σπουδών μου καθώς και την καθηγήτριά μου Φώτω Αφράτη για την άψογη συνεργασία και το ενδιαφέρον της. Όλα τα μέλη του Εργαστηρίου Λογικής και Επιστήμης Υπολογισμών για την πολύτιμη αρωγή τους.

Ακόμη θέλω να ευχαριστήσω όλους τους συγγενείς και φίλους που με στήριξαν και με στηρίζουν όλα αυτά τα χρόνια. Τέλος θέλω να ευχαριστήσω τον Άκη Παπαπαύλου για την βοηθειά του κατά τα τελευταία στάδια συγγραφής της παρούσας διπλωματικής.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Εισαγωγή

Η θεωρία παιγνίων, ως επιστημονικός κλάδος, επικεντρώνεται στην ερμηνεία και την ανάλυση των φαινομένων που αναδύονται κατά την την αλληλεπίδραση οντοτήτων. Οι οντότητες αυτές διεκπεραιώνουν στρατηγικές, λαμβάνοντας υπόψη την γνώση ή τις εκτιμήσεις που έχουν για την συμπεριφορά των υπόλοιπων οντοτήτων, αποσκοπώντας πάντα να μεγιστοποιήσουν τα οφέλη τους.

Δεδομένου λοιπόν του γεγονότος ότι τα υπό μελέτη μοντέλα που εξετάζονται στην θεωρία παιγνίων αποτελούν απλουστεύσεις σεναρίων που ανακύπτουν στην πραγματική καθημερινή μας ζωή κατά την συναναστροφή μας με τους γύρω μας, προφανώς θα επιθυμούσαμε οι στρατηγικές που αναδεικνύονται από τα μοντέλα μας να επιβεβαιώνονται από τις συμπεριφορές αναδύονται στον πραγματικό κόσμο.

Δυστυχώς συχνά όμως η θεωρία παιγνίων προβλέπει συμπεριφορές που έρχονται σε σύγκρουση με την εμπειρία μας. Ο κύριος σκόπελος στον οποίο προσκρούει η προσπάθειά μας για ορθή και επαρκή μοντελοποίηση της αλληλεπίδρασης οντοτήτων κρύβεται μέσα στον ίδιο τον ορισμό της θεωρίας παιγνίων. Πραγματικά τόσο η επεξεργασία των εκτιμήσεων για τις συμπεριφορές των υπολοίπων οντοτήτων-παικτών, όσο και η επιλογή και η εκτέλεση κάποιου στρατηγικού σχεδίου εμπεριέχουν μέσα τους την έννοια της πολυπλοκότητας. Έτσι λοιπόν θα επικεντρώσουμε το ενδιαφέρον μας, στο πάντρεμα των ιδεών της κλασσικής θεωρίας παιγνίων με έννοιες της πολυπλοκότητας, σε μια προσπάθεια να ενισχύσουμε την ικανότητα της θεωρίας παιγνίων να μοντελοποιεί πραγματικά σενάρια αλληλεπίδρασης αλλά και να επιλύσουμε γνωστά παράδοξα.

Κεφάλαιο 2

Παίγνια και Ισορροπία

Θα αφιερώσουμε το κεφάλαιο αυτό στην ανάπτυξη ορισμένων θεμελιωδών ιδεών της θεωρίας παιγνίων. Έτσι λοιπόν θα ξεκινήσουμε με τον ορισμό των κυριοτέρων εννοιών του γνωστικού αυτού κλάδου. Στο δεύτερο τμήμα του κεφαλαίου θα παρουσιάσουμε την έννοια της ισορροπίας Nash καθώς και ορισμένα χαρακτηριστικά παραδείγματα παιγνίων, ώστε να διευκολύνουμε την αφομοίωση των ορισμών αλλά και να εκθέσουμε τους διάφορους τρόπους με τους οποίους αναδύεται η έννοια της ισορροπίας Nash στα παίγνια. Στην τελευταία ενότητα παραθέτουμε το παράδοξο του "Διλήμματος των Φυλακισμένων" το οποίο πέρα ότι αναδεικνύει καταφανέστατα τις αδυναμίες μοντελοποίησης που αντιμετώπιζε η "κλασσική" θεωρία παιγνίων, ταυτόχρονα συνιστά και ένα σημείο αναφοράς αφού λειτούργησε ιστορικά ως αφετηρία για μια καρποφόρα σύμπραξη ανάμεσα στην θεωρία παιγνίων και την θεωρία πολυπλοκότητας.

2.1 Στρατηγικά παίγνια

Ένα στρατηγικό παίγνιο συνιστά ουσιαστικά ένα μοντέλο αλληλεπίδρασης ανάμεσα σε διακριτές οντότητες-παίκτες. Έτσι το μοντέλο αποτελείται από ένα πεπερασμένο σύνολο N παικτών όπου σε κάθε παίκτη i αντιστοιχεί ένα σύνολο A_i από ενέργειες (κινήσεις) και μια σχέση προτίμησης (διάταξης) πάνω στο σύνολο των διανυσμάτων των ενεργειών. Έτσι αναφερόμαστε σε ένα διάνυσμα ενεργειών $a = (a_j)_{j \in N}$ ως έκβαση ή αποτέλεσμα και συμβολίζουμε το σύνολο $\times j \in NA_j$ των δυνατών αποτελεσμάτων με το A. Το γεγονός ότι κάθε παίκτης ορίζει την σχέση προτίμησής του πάνω στο A και όχι στο A_i, αποτελεί την ειδοποιό διαφορά ανάμεσα σε ένα στρατηγικό παίγνιο και σε ένα πρόβλημα απόφασης. Πραγματικά λοιπόν εδώ προϋποθέτουμε ότι κάθε παίκτης λαμβάνει υπόψη του και τις επιλογές των υπολοίπων παικτών.

Ορισμός 2.1 (Στρατηγικό Παίγνιο). *Ένα στρατηγικό παίγνιο αποτελείται* *από*

- *ένα (πεπερασμένο) σύνολο N* **παικτών**.
- *για κάθε παίκτη $i \in N$ ένα μη κενό σύνολο A_i, δηλαδή το σύνολο των δυνατών* **κινήσεων** *για τον i-οστό παίκτη.*
- *για κάθε παίκτη $i \in N$ μία σχέση προτίμησης $\succ_i$ πάνω στο $A = \times_{j \in N} A_j$, δηλαδή η* **σχέση προτίμησης** *του i-οστού παίκτη.*

Συχνά αντί της σχέσης προτίμησης $\succ_i$ για τον i-οστό παίκτη χρησιμοποιούμε μία **συνάρτηση αποτίμησης** $u_i : A \rightarrow \Re$, η οποία αντιπροσωπεύει την αξία που αντιστοιχεί ο συγκεκριμένος παίκτης σε κάθε δυνατή έκβαση του παιγνίου. Έτσι έχουμε ότι $a \succ b$ εφόσον $u_i(a) \geq u_i(b)$. Στην περίπτωση που χρησιμοποιούμε την συνάρτηση αποτίμησης προκειμένου να περιγράψουμε τις προτιμήσεις ενός παίκτη, τότε θα συμβολίζουμε το παίγνιο $\langle N, (A_\imath), (u_\imath) \rangle$ αντί για $\langle N, (A_\imath), (\succ_\imath) \rangle$.

Τέλος πριν ξεκινήσουμε την προσπάθειά μας να βρούμε την λύση ενός παιγνίου, δηλαδή να προβλέψουμε την συμπεριφορά των εμπλεκομένων παικτών θα δώσουμε τον τρόπο απεικόνισης ενός παιγνίου. Χαρακτηριστικό είναι το παρακάτω παράδειγμα ενός παιχνιδιού δύο παικτών. Έτσι λοιπόν τα παιχνίδια δύο παικτών συνήθως αναπαριστάνονται με την βοήθεια ενός πίνακα του οποίου οι διαστάσεις αντιστοιχούν στον αριθμό των δυνατών ενεργειών κάθε παίκτη. Στον παρακάτω 2×3 πίνακα έχουμε ότι οι δύο γραμμές του πίνακα αντιστοιχούν στις δύο διαφορετικές ενέργειες (ή στρατηγικές) που μπορεί να εκτελέσει ο πρώτος παίκτης, ενώ ανάλογα οι 3 στήλες αντιστοιχούν στις τρεις διαφορετικές ενέργειες (ή στρατηγικές) που μπορεί να εκτελέσει ο δεύτερος παίκτης .

$$G = \begin{array}{c} \\ A^1 \\ A^2 \end{array} \begin{array}{c} \begin{array}{ccc} B^1 & B^2 & B^3 \end{array} \\ \begin{bmatrix} (1,6) & (2,-2) & (2,-4) \\ (7,2) & (-5,2) & (0,-2) \end{bmatrix} \end{array}$$

Από τον παραπάνω πίνακα, βλέπουμε τι κερδίζουν οι δύο παίκτες για κάθε συνδυασμό στρατηγικών τους. Αν ο πρώτος παίκτης παίξει την i-οστή στρατηγική του (δηλαδή την A^i) και ο δεύτερος παίκτης τη j-οστή στρατηγική του (δηλαδή την B^j), το όφελος των δύο παικτών δίνεται από το διάνυσμα $G[i][j]$. Για παράδειγμα αν ο πρώτος παίκτης επιλέξει την στρατηγική A^1 και ο δεύτερος τη στρατηγική B^2, το όφελος του πρώτου παίκτη θα είναι 2 και του δεύτερου -2, δηλαδή ο πρώτος παίκτης θα κερδίσει 2 και ο δεύτερος θα χάσει 2.

Έχοντας λοιπόν επιτύχει να αναπαραστήσουμε το παίγνιο με τρόπο σαφή και απλό και δεδομένου τώρα ότι οι εμπλεκόμενοι παίκτες έχοντας πλήρη γνώση του και με πάντα απώτερο σκοπό την μεγιστοποίηση του προσλαμβανόμενου κέρδους, γεννάται εύλογα το ερώτημα "Τί συμφέρει να παίξουν οι παίκτες;"

2.2 Ισορροπία Nash

2.2.1 Εισαγωγή: Σκεπτικά Λύσης

Προς απάντηση του παραπάνω ερωτήματος, δηλαδή το πώς να βρίσκουμε τη λύση ενός παιχνιδιού (solution concept) έχουν κατά καιρούς προταθεί αρκετές μέθοδοι. Πραγματικά ένας από αυτούς είναι η κυριαρχία (dominance). Η κυριαρχία όπως φανερώνει και το όνομά της εντοπίζεται σε περιπτώσεις παιγνίων στα για κάποιο παίκτη μια στρατηγική του είναι ξεκάθαρα καλύτερη από τις υπόλοιπες. Ας εξετάσουμε λοιπόν το παρακάτω παράδειγμα:

$$G_2 = \begin{bmatrix} (9,6) & (8,2) \\ (1,1) & (3,3) \end{bmatrix}$$

Στο παιχνίδι αυτό ο πρώτος παίκτης επιλέγει γραμμή (και έχει δύο στρατηγικές) και ο δεύτερος επιλέγει στήλη (και έχει επίσης δύο στρατηγικές). Από το παραπάνω σχήμα προκύπτει πως ότι και αν παίξει ο δεύτερος παίκτης, ο πρώτος παίκτης έχει μεγαλύτερο κέρδος αν παίξει την στρατηγική A^1. Πραγματικά λοιπόν σε αυτή την περίπτωση, η στρατηγική A^1 κυριαρχεί ανάμεσα σε όλες τις άλλες στρατηγικές του παίκτη αυτού. Συνεπώς ο πρώτος παίκτης θα παίξει A^1. Ο αντίπαλος παίκτης τώρα, που έχουμε υποθέσει ότι είναι λογικός (rational), ακολουθεί την πορεία της σκέψης του αντίπαλου παίκτη και ύστερα αποφασίζει να παίξει την στρατηγική B^1, η οποία μεγιστοποιεί το κέρδος του όταν ο πρώτος παίκτης παίζει A^1. Επομένως η λύση του παιχνιδιού είναι (A^1, B^1).

Ένα αντίστροφο σκεπτικό οδήγησε στην γένεση του σκεπτικού λύσης των κυριαρχούμενων στρατηγικών. Έτσι τώρα , αντί οι παίκτες να παίζουν τις κυρίαρχες (dominant) στρατηγικές τους ,επιλέγουν να μην παίζουν τις κυριαρχούμενες (dominated) στρατηγικές τους. Μια στρατηγική χαρακτηρίζεται ως κυριαρχούμενη (dominated) για τον παίκτη i, εάν ανεξάρτητα από ότι θα παίξουν οι υπόλοιποι παίκτες, ο παίκτης i έχει μικρότερο όφελος αν παίξει την στρατηγική αυτή από το αν παίξει κάποια άλλη. Η εφαρμογή της μεθόδου των κυριαρχούμενων στρατηγικών στο παραπάνω παράδειγμα θα οδηγούσε σε ακριβώς την ίδια ανάλυση. Δυστυχώς όμως τόσο η μέθοδος των κυρίαρχων όσο και των κυριαρχούμενων στρατηγικών δεν οδηγούν πάντα σε λύση του παιγνίου. Αυτό φαίνεται ξεκάθαρα στο παρακάτω παράδειγμα:

$$G_3 = \begin{bmatrix} (4,5) & (0,1) \\ (2,0) & (3,4) \end{bmatrix}$$

Πράγματι στο παραπάνω παίγνιο δεν υπάρχει κάποια στρατηγική που να κυριαρχεί των άλλων για κανέναν από τους δύο παίκτες, οπότε καθίσταται επιτακτική η ανάγκη εύρεσης ενός ισχυρότερου σκεπτικού λύσης, της ισορροπίας Nash. Όμως εάν προσέξουμε τα σημεία (4,5) (που αντιστοιχεί στις στρατηγικές (A^1, B^1) και (3,4) (που αντιστοιχεί στις στρατηγικές (A^2, B^2) παρατηρούμε ότι στα σημεία αυτά κανένας από τους δύο παίκτες δεν μπορεί να βελτιώσει μονομερώς την θέση του. Είναι λοιπόν, όπως θα ορίσουμε αυστηρά σε λίγο, ένα σημείο ισορροπίας Nash, κάθε παίκτης i, με δεδομένες τις στρατηγικές των υπόλοιπων παικτών, δεν έχει όφελος να αποκλίνει από την στρατηγική του. Το παιχνίδι, λοιπόν, του πίνακα G_3 έχει δύο σημεία Nash ισορροπίας.

Ανακεφαλαιώνοντας, λοιπόν, η έννοια της ισορροπίας Nash είναι πιο γενική από τις κυρίαρχες στρατηγικές γιατί αφενός βρίσκει εφαρμογή σε περισσότερες περιπτώσεις και αφετέρου οι λύσεις που προκύπτουν από τη κυριαρχία (dominance) είναι σημεία ισορροπίας Nash. Πράγματι, αν επιστρέψουμε στο παράδειγμα του πίνακα G_2, που με τη μέθοδο των κυρίαρχων στρατηγικών βρήκαμε ότι η λύση είναι (A^1, B^1), το ίδιο σημείο είναι και σημείο ισορροπίας Nash.

2.2.2 Ορισμός Ισορροπίας Nash

Η ισορροπία Nash αποτελεί το πιο ευρέως διαδεδομένο σκεπτικό λύσης της θεωρίας παιγνίων και συλλαμβάνει εκείνη την σταθερή κατάσταση κατά την οποία κάθε παίκτης έχει εκτιμήσει σωστά την συμπεριφορά των υπολοίπων παικτών και φέρεται ορθολογιστικά. Επίσης, είναι σημαντικό να ξεκαθαρίσουμε ότι δεν εξετάζει την διαδικασία με την οποία έχει επιτευχθεί η σταθερή αυτή κατάσταση.

Ορισμός 2.2. *Ορίζουμε ως ισορροπία Nash ενός στρατηγικού παιγνίου* $\langle N, (A_\imath), (\succ_i)\rangle$ *ένα διάνυσμα* $a^* \in A$ *κινήσεων που έχει την ιδιότητα ότι για κάθε παίκτη* $i \in N$ *ισχύει*

$$(a^*_{-i}) \succ_i (a^*_{-i}, a_i) \text{ για όλα τα } s_i \in A_i.^1$$

Επομένως για να είναι το a^* σημείο ισορροπίας Nash πρέπει να μην υπάρχει παίκτης i, τέτοιος ώστε να είναι σε θέση να επιλέξει μια ενέργεια που του αποφέρει όφελος μεγαλύτερο από ότι αποκομίζει από την ενέργεια (κίνηση)

[1] Όπου γενικά $x_{-i} = (x_j)_{j \in N\setminus\{i\}}$

a_i^*, δεδομένου ότι κάθε άλλος παίκτης j επιμένει στην κίνηση a_i^* που του επιβάλει το σημείο ισορροπίας. Με άλλα λόγια κανείς δεν είναι σε θέση να βελτιώσει την θέση του δρώντας μονομερώς.

Ένας εναλλακτικός ορισμός ο οποίος αποδεικνύεται αρκετά χρήσιμος είναι ο παρακάτω. Για κάθε $a_{-i} \in A_{-i}$ ορίζουμε ως $B_i(a_{-i})$ να είναι το σύνολο των καλύτερων αποκρίσεων του παίκτη i δεδομένου του a_{-i}:

$$B_i(a_{-i}) = \{a_i \in A_i : (a_{-i}, a_i) \succ_i (a_{-i}, a_i') \text{ για όλα τα } a_i' \in A_i\}.$$

Ονομάζουμε αυτήν την συνάρτηση B_i (η οποία επιστρέφει σύνολα) **συνάρτηση βέλτιστης απάντησης** του παίκτη i. Τώρα, ως ισορροπία Nash ορίζουμε το διάνυσμα a^* των κινήσεων για το οποίο ισχύει:

$$a_i^* \in B_i(a_{-i}^*) \text{ για όλα τα } i \in N.$$

Ο παραπάνω ορισμός ουσιαστικά μας υποδεικνύει μια μέθοδο για την εύρεση ισορροπιών Nash: πρώτα υπολογίζεις την συνάρτηση βέλτιστης απόκρισης για κάθε παίκτη, στην συνέχεια βρίσκεις ένα διάνυσμα A^* ενεργειών τέτοιο ώστε $a_i^* \in B_i(a_{-i}^*)$ για όλα τα $i \in N$.

2.2.3 Παραδείγματα Παιγνίων

Τα παρακάτω παραδείγματα παρουσιάζουν διάφορα χαρακτηριστικά είδη στρατηγικών αλληλεπιδράσεων¿ Τα παίγνια που θα παρουσιάσουμε είναι πολύ απλά: σε κάθε παιχνίδι συμμετέχουν ακριβώς δύο παίκτες και σε κάθε παίκτη αντιστοιχούν δύο στρατηγικές. Παρόλα αυτά όμως κάθε ένα από τα παίγνια αυτά συλλαμβάνει την ουσία κάποιου συγκεκριμένου τύπου στρατηγικής αλληλεπίδρασης που εμφανίζεται συχνά σε πιο πολύπλοκες αλληλεπιδράσεις.

Παράδειγμα 2.1 (Battle of The Sexes).

Ένα ζευγάρι επιθυμεί να πάει σε ένα μουσικό κονσέρτο. Όμως υπάρχουν δύο επιλογές, μπορούν να ακούσουν είτε Bach είτε Stravinsky. Η κύρια επιδίωξή τους είναι να βγουν έξω μαζί, αλλά ο άντρας προτιμά τον Bach και η γυναίκα τον Straninsky. Οι προτιμήσεις τους μάλιστα εκφράζονται από τον παρακάτω πίνακα.

$$G = \begin{array}{cc} & \begin{array}{cc} B & S \end{array} \\ \begin{array}{c} B \\ S \end{array} & \begin{bmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{bmatrix} \end{array}$$

Για μια αναδρομή στην ιστορία του παιγνίου της ”Μάχης των Φύλων” μπορεί να ανατρέξει κανείς στους Luce και Raiffa [9]

Το παίγνιο αυτό μοντελοποιεί μια κατάσταση κατά την οποία οι παίκτες θέλουν να συνεργαστούν μεταξύ τους αλλά τα συμφέροντά τους είναι σε σύγκρουση. "Η Μάχη των Φύλων " έχει δύο σημεία ισορροπίας Nash, τα (B,B) και (S,S), δηλαδή τα σημεία που οι παίκτες είτε συμφωνούν να παρακολουθήσουν Bach(B), είτε συμφωνούν να παρακολουθήσουν Stravinsky(S).

Παράδειγμα 2.2 (Ένα συνεργατικό παίγνιο).

Όπως και στο προηγούμενο παράδειγμα, έτσι και τώρα ένα ζευγάρι επιθυμεί να πάει σε ένα μουσικό κονσέρτο. Υπάρχουν δύο επιλογές, μπορούν να ακούσουν είτε Mahler είτε Bach. Η κύρια επιδίωξή τους είναι να βγουν έξω μαζί, αλλά ταυτόχρονα τόσο ο άντρας όσο και η γυναίκα προτιμούν τον Mahler. Οι προτιμήσεις τους μάλιστα εκφράζονται από τον παρακάτω πίνακα.

$$G = \begin{array}{c c} & \begin{array}{cc} M & B \end{array} \\ \begin{array}{c} M \\ B \end{array} & \left[\begin{array}{cc} (2,2) & (0,0) \\ (0,0) & (1,1) \end{array}\right] \end{array}$$

Όπως και η "Μάχη των Φύλων", έτσι και αυτό το παίγνιο έχει δύο σημεία ισορροπίας Nash, τα οποία αντιστοιχούν στις δύο περιπτώσεις ταύτισης των κινήσεων των δύο παικτών. Πράγματι, τόσο το (M,M) όσο και το (B,B) συνιστούν σημεία ισορροπίας Nash. Σε αντίθεση όμως με την "Μάχη των Φύλων", οι δύο παίκτες έχουν ένα αμοιβαίο συμφέρον να επιτύχουν μία συγκεκριμένη ισορροπία, την (M,M), παρόλα αυτά η έννοια της ισορροπίας Nash δεν αποκλείει την σταθερή εκείνη κατάσταση κατά την οποία η έκβαση του παιγνίου αντιστοιχεί στην "κατώτερη" ισορροπία (B,B).

Παράδειγμα 2.3 (Γεράκι Περιστέρι).

Δύο ζώα μονομαχούν για το ίδιο θήραμα. καθένα από αυτά μπορεί να συμπεριφερθεί είτε σαν γεράκι είτε σαν περιστέρι. Το καλύτερο αποτέλεσμα για κάθε ένα από τα δύο ζώα αντιστοιχεί στην περίπτωση που το ίδιο συμπεριφέρεται σαν γεράκι, ενώ το αντίπαλο σαν περιστέρι. Το χειρότερο αποτέλεσμα είναι όταν και τα δύο ζώα συμπεριφέρονται σαν γεράκια. Έτσι κάθε ζώο προτιμά να συμπεριφέρεται σαν γεράκι όταν ο αντίπαλος του συμπεριφέρεται σαν περιστέρι και σαν περιστέρι όταν ο αντίπαλος του συμπεριφέρεται σαν γεράκι. Οι προτιμήσεις τους αυτές εκφράζονται από τον παρακάτω πίνακα.

$$P = \begin{array}{c c} & \begin{array}{cc} \mathit{\Pi} & \mathit{\Gamma} \end{array} \\ \begin{array}{c} \mathit{\Pi} \\ \mathit{\Gamma} \end{array} & \left[\begin{array}{cc} (3,3) & (1,4) \\ (4,1) & (0,0) \end{array}\right] \end{array}$$

Το παίγνιο αυτό έχει δύο σημεία ισορροπίας, τα (Γ,Π) και (Π,Γ)τα οποία αντιστοιχούν στις δύο δυνατές περιπτώσεις που οι παίκτες συμφωνούν να διαφωνήσουν. Έτσι σε κάθε μία από αυτές τις περιπτώσεις κάποιος από τους δύο παίκτες ενδίδει και λειτουργεί σαν περιστέρι, την στιγμή που ο αντίπαλός του λειτουργεί σαν γεράκι.

Παράδειγμα 2.4 (Σύγκριση κερμάτων).

Κάθε ένας από τους δύο παίκτες επιλέγει είτε Κεφαλή είτε Γράμματα. Εάν οι επιλογές τους διαφέρουν τότε ο πρώτος παίκτης πληρώνει τον δεύτερο 1 Ευρώ. Στην αντίθετη περίπτωση, που οι επιλογές τους είναι ίδιες τότε ο δεύτερος παίκτης πληρώνει τον πρώτος1 Ευρώ. Το μοντέλο του παιγνίου αυτού απεικονίζεται στο παρακάτω σχήμα.

$$P = \begin{array}{cc} & \begin{array}{cc} K & \Gamma \end{array} \\ \begin{array}{c} K \\ \Gamma \end{array} & \begin{bmatrix} (1,-1) & (-1,1) \\ (-1,1) & (1,-1) \end{bmatrix} \end{array}$$

Το παίγνιο αυτό δεν έχει (γνήσιο) σημείο ισορροπίας Nash. Παρατηρούμε ότι το συνολικό όφελος είναι πάντα 0 (zero-sum), δηλαδή ότι κερδίζει ο ένας παίκτης το χάνει ο άλλος. Σε αυτά τα παιχνίδια αρκεί να έχουμε τον πίνακα με τα οφέλη του ενός παίκτη. (εάν επιτρέπαμε σε κάθε παίκτη να μην επιλέγει μια συγκεκριμένη κίνηση αλλά μια πιθανοτική κατανομή πάνω στις δυνατές του κινήσεις, τότε το παίγνιο αυτό θα είχε μεικτό σημείο ισορροπίας Nash όπου κάθε παίκτης επιλέγει κεφαλή ή γράμματα με πιθανότητα 1/2, αλλά δεν θα επεκταθούμε στο ζήτημα αυτό, διότι αποτελεί αντικείμενο της εργασίας αυτής).

2.3 Παράδοξα

Όπως είδαμε από την πληθώρα των παραδειγμάτων που προηγήθηκαν, η έννοια της ισορροπίας Nash ως σκεπτικό λύσης είναι ιδιαίτερα επιτυχημένη όχι μόνο εξαιτίας της γενικότητας της λύσης, αλλά και λόγω του γεγονότος ότι επιτυγχάνει τις περισσότερες φορές να υπολογίσει την πραγματική εκείνη λύση η οποία θα αναδυόταν στον πραγματικό κόσμο κατά την αλληλεπίδραση λογικών παικτών. Δυστυχώς όμως όπως θα δούμε από το πολύ γνωστό παράδοξο που θα ακολουθήσει δεν επιτυγχάνει πάντα.

Παράδειγμα 2.5 (Το Δίλημμα των Φυλακισμένων).

Το πιο γνωστό παράδοξο είναι το δίλημμα των φυλακισμένων (prisoner's dilemma), το οποίο προκύπτει από την παρακάτω ιστορία. Δύο κλέφτες έχουν συλληφθεί, αλλά δεν έχει αρκετά ενοχοποιητικά στοιχεία ώστε να φυλακιστούν. Αν υπήρχαν, θα φυλακίζονταν και οι δύο για 3 χρόνια, ενώ με τα υπάρχοντα στοιχεία θα φυλακιστούν για 1 χρόνο. Για να τους αναγκάσει να ομολογήσουν, η αστυνομία τους βάζει σε χωριστά δωμάτια και φανερώνει στον καθένα ότι αν ομολογήσει ενώ ο άλλος δεν ομολογήσει, θα χρησιμοποιήσει την κατάθεσή του για να ενοχοποιήσει τον άλλο, τον οποίο θα φυλακίσει για 4 χρόνια, ενώ αυτόν που ομολόγησε θα του δώσει χάρη και δε θα τον φυλακίσει καθόλου. Αν ομολογήσουν και οι δύο θα εκτίσουν ο καθένας 3 χρόνια.

Από τα παραπάνω προκύπτει ότι το προηγούμενο σενάριο μπορεί να μοντελοποιηθεί από ένα παίγνιο δύο παικτών, στο οποίο κάθε παίκτης έχει δύο στρατηγικές. Η μία στρατηγική του είναι να ομολογήσει (Ο) με την αστυνομία ενώ στην άλλη αρνείται να ομολογήσει (Α). Η πρώτη στρατηγική του κάθε παίκτη (πρώτη γραμμή και πρώτη στήλη αντίστοιχα για τους δύο παίκτες) είναι να ομολογήσει και η άλλη είναι να μη ομολογήσει. Στον πίνακα του παιχνιδιού βάζουμε τα χρόνια που θα πάει φυλακή κάθε παίκτης, τα οποία προφανώς θέλει κάθε παίκτης να ελαχιστοποιήσει (ως όφελος κάθε παίκτη μπορούμε να θεωρήσουμε το αντίθετο από τα χρόνια που θα κληθεί να εκτίσει).

$$P = \begin{array}{c c} & \begin{array}{cc} O & A \end{array} \\ \begin{array}{c} O \\ A \end{array} & \left[\begin{array}{cc} (3,3) & (0,4) \\ (4,0) & (1,1) \end{array} \right] \end{array}$$

Από τον παραπάνω πίνακα (P) βλέπουμε ότι το παιχνίδι έχει ένα γνήσιο σημείο ισορροπίας το (Ο,Ο), δηλαδή να συνεργαστούν και οι δύο παίκτες με την αστυνομία και να ομολογήσουν. Στο σημείο ισορροπίας φυλακίζονται και οι δύο παίκτες για 3 χρόνια, ενώ αν και οι δύο δεν ομολογούσαν θα φυλακίζονταν μόνο για ένα χρόνο. Αυτό ακριβώς είναι το παράδοξο.

2.4 Συμπεράσματα

Η αστοχία της αυτή της έννοιας της ισορροπίας Nash να υπολογίσει την πραγματική λύση δημιουργεί ερωτηματικά για την ικανότητα της ίδιας της έννοιας να προσεγγίσει τον τρόπο με τον οποίο σταθμίζουν τα γεγονότα και λαμβάνουν αποφάσεις οι λογικές οντότητες-παίκτες. Πραγματικά τέτοια παράδοξα έκρουσαν τον κώδωνα του κινδύνου σχετικά με τα παρεχόμενα από την θεωρία παιγνίων αποτελέσματα, αφού αποδεικνύουν με τον πιο κατηγορηματικό τρόπο την αδυναμία της να προσεγγίσει την συμπεριφορά των πραγματικών παικτών ακόμη και σε εντελώς στοιχειώδεις απλουστεύσεις σεναρίων αλληλεπίδρασης. Έτσι λοιπόν το παράδοξο αυτό αποτέλεσε την αφετηρία για μακροχρόνιες έρευνες που κατέληξαν σε σημαντικά συμπεράσματα, όπως θα δούμε αναλυτικά στα υπόλοιπα κεφάλαια.

Κεφάλαιο 3

Επαναλαμβανόμενα Παίγνια

Το μοντέλο ενός επαναλαμβανόμενου παιγνίου εξυπηρετεί την μελέτη και την ανάλυση μακροχρόνιων αλληλεπιδράσεων μεταξύ οντοτήτων. Χαρακτηριστικό των παιγνίων αυτών είναι ότι κάθε οντότητα-παίκτης λαμβάνει υπόψη του, το γεγονός ότι η συμπεριφορά του θα έχει αντίκτυπο στην διαμόρφωση των μελλοντικών αποκρίσεων και στρατηγικών των συμπαικτών του. Έτσι απώτερος στόχος των επαναλαμβανόμενων παιγνίων είναι η πρόβλεψη αλλά και η ανάλυση συμπεριφορών που εμφανίζονται σε βάθος χρόνου, όπως η εκδίκηση, η συνεργασία αλλά και οι απειλές.

3.1 Η κύρια ιδέα

Η κύρια ιδέα που κρύβεται πίσω από αυτήν την θεωρία φαίνεται καθαρά όταν παίζουμε το "δίλημμα των φυλακισμένων" για ν γύρους. Έτσι όπως έχουμε δει εάν αναλύσουμε κάθε γύρο απομονωμένο από τους υπόλοιπους έχουμε ότι το συγκεκριμένο παιχνίδι έχει μόνο ένα σημείο ισορροπίας Nash το (O,O) στο οποίο και οι δύο αυτομολούν και ομολογούν καταθέτοντας εναντίον του άλλου φυλακισμένου. Πράγματι για κάθε παίκτη, η στρατηγική O κυριαρχεί αυστηρά της στρατηγικής Δ (δεν ομολογούν), δηλαδή η O είναι καλύτερη από την Δ ανεξάρτητα από την επιλογή του άλλου παίκτη. Με απλά λόγια η θεωρία παιγνίων προβλέπει ότι στην περίπτωση αυτή δύο οποιοδήποτε λογικοί παίκτες θα επέλεγαν να δράσουν εγωιστικά παρόλο που αυτή η επιλογή τους, τους εξασφαλίζει μικρότερο κέρδος από το κέρδος τους στην περίπτωση της συνεργασίας Έτσι λοιπόν η κύρια ιδέα πίσω από την θεωρία των επαναλαμβανόμενων παιγνίων είναι ότι εάν το παιχνίδι παίζεται σε διαδοχικούς γύρους τότε μπορούμε να εξασφαλίσουμε την αμοιβαίως επιθυμητή έκβασή του, κατά την οποία σε κάθε γύρο του παιχνιδιού οι παίκτες συνεργάζονται μεταξύ τους και δεν ομολογούν (Δ,Δ), εφόσον κάθε παίκτης πιστεύει ότι μια αποστασία εκ μέρους του θα τερματίσει την μεταξύ τους συνεργασία και θα προκαλέσει μακροπρόθεσμα απώλειες που υπερκαλύπτουν το βραχυπρόθεσμο όφελος που θα αποκομίσει από την αποστασία του.

Η πρωταρχική επιδίωξη της θεωρίας είναι να επισημάνει τους τύπους εκείνους των στρατηγικών οι οποίες υποστηρίζουν αμοιβαίως αποδεκτές εκβάσεις των παιχνιδιών. Η θεωρία αυτή μας δίνει την δυνατότητα να μελετήσουμε συμπεριφορές, που προκύπτουν από την μακροχρόνια αλληλεπίδραση μεταξύ ατόμων και να επισημάνουμε πρότυπα τα οποία ουσιαστικά συνιστούν κοινωνικές φόρμες. Τα αποτελέσματα τα οποία περιγράφουμε στην συνέχεια, αποδεικνύουν ότι η ανάδειξη συνεργατικών προτύπων συμπεριφοράς προϋποθέτει την ύπαρξη κοινωνικών νορμών που τιμωρούν τις ανεπιθύμητες εγωιστικές συμπεριφορές.

Ένα σημαντικό τμήμα της δουλειάς που έχει πραγματοποιηθεί στο συγκεκριμένο πεδίο επικεντρώνεται στα σύνολα των αποδοχών, τα οποία μπορούν να εξασφαλιστούν μέσω κάποιας συνεργατικής ισορροπίας. Αυτά τα θεωρήματα τα οποία είναι γνωστά ως folk theorems καταλήγουν σε δύο σημαντικά συμπεράσματα. Έτσι, αφενός καταδεικνύουν ότι τα κοινωνικώς επιθυμητά αποτελέσματα μπορούν να συντηρηθούν μόνο εφόσον οι παίκτες δεν είναι κοντόφθαλμοι στις επιδιώξεις τους αλλά υποστηρίζουν μακροχρόνιες στρατηγικές, αφετέρου αποδεικνύουν ότι το σύνολο των στρατηγικών ισορροπίας των μακροχρόνιων παιγνίων είναι τόσο τεράστιο, που η έννοια της ισορροπίας αποδυναμώνεται. Με άλλα λόγια η εξασφάλιση της ισορροπίας δεν περιορίζει σε σημαντικό βαθμό το σύνολο των δυνατών εκβάσεων του παιχνιδιού, έτσι ώστε να είναι σε θέσει να μας δώσει μια ικανοποιητική πρόβλεψη για την έκβαση του παιχνιδιού.

3.2 Πεπερασμένα και μη πεπερασμένα παίγνια

Προτού προχωρήσουμε στην μεθοδική ανάλυση των κυριοτέρων αποτελεσμάτων της θεωρίας των επαναλαμβανόμενων παιγνίων θα προχωρήσουμε σε μια σημαντική διαμέρισή τους σε πεπερασμένα και μη.

Η ειδοποιός διαφορά ανάμεσα στα δύο είδη έγκειται στην ύπαρξη ή στην απουσία πεπερασμένου ορίζοντα, δηλαδή σημείου τερματισμού του παιγνίου. Όπως θα δούμε η ανάλυση των δύο παραλλαγών για κάποιο συγκεκριμένο παιχνίδι μπορεί να είναι διαφορετικές. Ένα κάπως ακραίο παράδειγμα διαφοράς εντοπίζεται στην περίπτωση του επαναλαμβανόμενου παιγνίου του διλήμματος των φυλακισμένων. Έτσι κάθε εκδοχή του παιχνιδιού με πεπερασμένο ορίζοντα, το μοναδικό του σημείο ισορροπίας είναι το μη επιθυμητό (O,O) στο οποίο και οι δύο παίκτες αυτομολούν σε κάθε περίοδο του παιχνιδιού. Στην περίπτωση τώρα του μη πεπερασμένου παιγνίου το σύνολο των προφίλ των απολαβών των εμπλεκομένων παικτών είναι τεράστιο.

Πάντως η έννοια της ύπαρξης ή όχι πεπερασμένου ορίζοντα καθώς και το πως αντιλαμβάνεται την έννοια του ορίζοντα ένας υπολογιστικά πεπερασμένος παίκτης είναι ένα ζήτημα το οποίο θα μας απασχολήσει κατά κόρον στην συνέχεια και για τον λόγο αυτό δεν θα επεκταθούμε σε άλλες λεπτομέρειες στο σημείο αυτό. Αντιθέτως, θα αρκεστούμε να επισημάνουμε ότι η επιλογή ενός μη πεπερασμένου μοντέλου για την προσομοίωση κάποιας αλληλεπίδρασης ατόμων αντικατοπτρίζει την περίπτωση που κάθε εμπλεκόμενο άτομο σε μια τυχαία περίοδο πιστεύει το παιχνίδι θα συνεχιστεί τουλάχιστον για μια περίοδο ακόμη, ενώ το πεπερασμένο μοντέλο απεικονίζει τις περιπτώσεις εκείνες όπου οι παίκτες αντιλαμβάνονται μια συγκεκριμένη περίοδο που σηματοδοτεί την λήξη του παιχνιδιού.

3.3 Μη πεπερασμένα παίγνια: Ορισμοί

Το μοντέλο ενός μη πεπερασμένου παιγνίου περιγράφει μια αλληλεπίδραση κατά την οποία ορισμένοι παίκτες συμμετέχουν επανειλημμένως σε ένα στρατηγικό παίγνιο G, το οποίο ονομάζουμε *παιχνίδι-βήμα*. Από εδώ και μπρος περιορίζουμε το ενδιαφέρον μας σε παίγνια στα οποία το σύνολο των στρατηγικών είναι συμπαγές και η σχέση προτίμησης κάθε παίκτη είναι συνεχής. Δεν υπάρχει όριο στον αριθμό των φορών που μπορεί να επαναληφθεί το G, ενώ οι παίκτες πραγματοποιούν τις κινήσεις τους ταυτόχρονα. Κατά την πραγματοποίηση μιας κίνησης του, κάθε παίκτης γνωρίζει όλες τις κινήσεις που έχουν πραγματοποιήσει κατά τις προηγούμενες περιόδους όλοι οι υπόλοιποι παίκτες. Η παραπάνω αλληλεπίδραση μοντελοποιείται από ένα εκτεταμένο παίγνιο με τέλεια πληροφορία (και ταυτόχρονες κινήσεις) ως εξής:

Έστω G= $\langle$ N, $(A_\imath)$, $(\succ_\imath)$ $\rangle$ ένα στρατηγικό παίγνιο, έστω A=$\times_{\imath \in N} A_\imath$. Ένα **μη πεπερασμένο επαναλαμβανόμενο παίγνιο του** G είναι ένα εκτεταμένο παίγνιο με τέλεια πληροφορία και ταυτόχρονες κινήσεις $\langle$ N, H, P, $(\succ_i^*)\rangle$ στο οποίο

- $H = \{\emptyset\} \bigcup (\bigcup_{t=1}^{\infty} A^t) \bigcup A^{\infty}$, όπου το $\emptyset$ είναι η αρχική ιστορία και A^∞ είναι το σύνολο των μη πεπερασμένων ακολουθιών $(a^t)_{t=1}^{\infty}$ των στρατηγικών προφίλ στο G)
- $P(h) = N$ για κάθε μη τερματική ιστορία $h \in H$
- $(\succ_i^*)$ είναι μια σχέση διάταξης πάνω στο A^∞ η οποία επεκτείνει την σχέση διάταξης $(\succ_\imath)$, με την έννοια ότι ικανοποιεί την ακόλουθη ιδιότητα της *ασθενούς διαχωρησιμότητας*: εάν $(a^t) \in A^\infty$, $a \in A$, $\acute{a} \in A$, και $a \succ_\imath \acute{a}$ τότε

$$(a^1, ..., a^{t-1}, a, a^{t+1}, ...) \succ_i^* (a^1, ..., a^{t-1}, á, a^{t+1}, ...)$$

για όλες τις τιμές του t.

Μια ιστορία είναι τερματική εάν και μόνο εάν είναι άπειρη. Μετά από μία οποιαδήποτε μη τερματική ιστορία κάθε παίκτης $i \in N$ επιλέγει μια κίνηση από το A_i. Επομένως η στρατηγική του i-οστού παίκτη είναι μια συνάρτηση που απεικονίζει σε κάθε πεπερασμένη ακολουθία εκβάσεων του G μια κίνηση στο A_i.

Στην συνέχεια θα επιβάλλουμε μερικούς επιπρόσθετους περιορισμούς πάνω στις σχέσεις προτίμησης (διάταξης) του κάθε παίκτη πέραν από τον περιορισμό της ασθενούς διαχωρισιμότητας. Έτσι λοιπόν από εδώ και μπρος θεωρούμε ότι η σχέση προτίμησης του i-οστού παίκτη $\succ_i^*$ βασίζεται πάνω σε μια συνάρτηση αποτίμησης (ή οφέλους) u_i στο G. Συγκεκριμένα υποθέτουμε ότι το εάν ισχύει το $(a^t) \succ_i^* (b^t)$ εξαρτάται αποκλειστικά και μόνο από την σχέση που έχουν μεταξύ τους οι ακολουθίες $(u_i(a^t))$ και $(u_i(b^t))$ των αποδοχών των παικτών στο G.

Στην συνέχεια παρουσιάζουμε τρία διαφορετικά είδη σχέσεων προτιμήσεις:

- **Ελαττούμενου βάρους**: Υπάρχει ένας αριθμός δ $\in$ (0,1) (ο παράγοντας ελάττωσης) τέτοιος ώστε η ακολουθία (u_i^t) των πραγματικών αριθμών είναι τουλάχιστον τόσο καλή όσο η ακολουθία (w_i^t) εάν και μόνο εάν $\sum_{t=1}^{\infty} \delta^{t-1}(u_i^t - w_i^t) \geq 0$. Στην περίπτωση αυτή ορίζουμε ως προφίλ οφέλους του i-οστού παίκτη την ποσότητα $((1-\delta)\sum_{t=1}^{T} \delta^{t-1} u_i^t)_{i \in N}$.

- **Όριο των μέσων**: Η ακολουθία (u_i^t) των πραγματικών αριθμών είναι προτιμητέα σε σχέση με την ακολουθία (w_i^t) εάν και μόνο εάν $\liminf \sum_{t=1}^{T}$ $(u_i^t - w_i^t)/T > 0$ (δηλαδή εάν και μόνο εάν υπάρχει $\epsilon > 0$ τέτοιο ώστε $\sum_{t=1}^{T}(u_i^t - w_i^t)/T > \epsilon$. Στην περίπτωση αυτή ορίζουμε ως προφίλ οφέλους του i-οστού παίκτη την ποσότητα $\lim_{T\to\infty}(\sum_{t=1}^{T} u_i^t/T)_{i\in N}$, εφόσον αυτή η ποσότητα ορίζεται.

- **Υπερκερασμένη**: Η ακολουθία (u_i^t) των πραγματικών αριθμών είναι προτιμητέα σε σχέση με την ακολουθία (w_i^t) εάν και μόνο εάν $\liminf \sum_{t=1}^{T}$ $(u_i^t - w_i^t) > 0$.

Με απλά λόγια όταν χρησιμοποιούμε το πρώτο κριτήριο για τις σχέσεις προτίμησης, τότε αντιστοιχούμε σε κάθε περίοδο ένα βάρος το οποίο μειώνεται εκθετικά με την πάροδο του χρόνου. Τα υπόλοιπα δύο κριτήρια αντιμετωπίζουν συμμετρικά όλες τις περιόδους του παιχνιδιού. Σύμφωνα με το δεύτερο κριτήριο, αυτό του ορίου των μέσων, κάθε παίκτης αποτιμά μία ακολουθία (u_i^t) απολαβών των παικτών ουσιαστικά από το όριο προς το οποίο τείνει ο μέσος όρος των απολαβών $\lim_{T\to\infty}\sum_{t=1}^{T} u_i^t/T$. Επίσης, όπως φαίνεται από τον παραπάνω ορισμό, το προφίλ οφέλους παραμένει ανεπηρέαστο από οποιεσδήποτε αλλαγές που πραγματοποιούνται σε ένα πεπερασμένο αριθμό περιόδων. Τέλος το τρίτο κριτήριο, αντιμετωπίζει όλες τις περιόδους συμμετρικά, δίνει έμφαση σε όλη την έκταση του εκτεταμένου παιγνίου, αλλά την ίδια στιγμή παραμένει ευαίσθητο σε αλλαγές στις απολαβές ακόμη και σε μια μόνο περίοδο.

Τα παρακάτω παραδείγματα επισημαίνουν τις διαφορές ανάμεσα στα τρία κριτήρια. Έτσι η ακολουθία (-1,1,0,0,...) είναι προτιμητέα για οποιοδήποτε $\delta \in (0,1)$ από το κριτήριο των ελαττούμενων βαρών από την ακολουθία (0,0,...) αλλά σύμφωνα με τα άλλα δύο κριτήρια οι δύο ακολουθίες είναι ισοδύναμες. Ανάλογα η ακολουθία (-1,2,0,0,...) είναι προτιμητέα από το κριτήριο της υπερκερασμένης από την ακολουθία (0,0,...) αλλά σύμφωνα με το κριτήριο του ορίου των μέσων οι δύο ακολουθίες είναι ισοδύναμες. Τέλος η ακολουθία (0,...,0,1,1,...) στην οποία M μηδενικά προηγούνται μιας άπειρης ακολουθίας άσσων είναι προτιμητέα σύμφωνα με το κριτήριο του ορίου των μέσων από την (1,0,0,...) για κάθε M, όμως για κάθε δ υπάρχει ένα M^* αρκετά μεγάλο τέτοιο ώστε για κάθε $M > M^*$, η δεύτερη ακολουθία να είναι προτιμητέα της πρώτης σύμφωνα με το κριτήριο των ελλατουμένων βαρών.

Έστω λοιπόν G= ⟨ N, $(A_\imath)$, $(\succ_\imath)$ ⟩ ένα στρατηγικό παίγνιο και για κάθε $i \in N$ έστω u_i η συνάρτηση οφέλους που αντιπροσωπεύει την $(\succ_\imath)$. Τώρα ορίζουμε το **δ-ελαττούμενο μη πεπερασμένο επαναλαμβανόμενο παίγνιο** $\langle N, (A_\imath), (u_i)\rangle$ να είναι το μη πεπερασμένο επαναλαμβανόμενο παίγνιο για το οποίο το παιχνίδι-βήμα είναι το G και η σχέση προτίμησης $\succ_i^*$ για κάθε παίκτη $i \in N$ προκύπτει από την συνάρτηση u_i χρησιμοποιώντας το κριτήριο των ελαττουμένων βαρών με παράγοντα ελάττωσης δ για κάθε παίκτη. Όμοια ορίζουμε το **το όριο του μέσου μη πεπερασμένο επαναλαμβανόμενο παίγνιο** και το **υπερκερασμένο μη πεπερασμένο επαναλαμβανόμενο παίγνιο**.

Τέλος έστω ότι συμβολίζουμε με $u(a)$ το προφίλ $(u_i(a))_{i\in N}$, τότε ορίζουμε ένα διάνυσμα $v \in \Re^N$ ως **προφίλ οφέλους** του $\langle N, (A_\imath), (u_i)\rangle$ εφόσον υπάρχει μια έκβαση $a \in A$ για την οποία $v = u(a)$. Ακόμη θα λέμε ότι ένα διάνυσμα $v \in \Re^N$ αποτελεί ένα εφικτό προφίλ οφέλους του $\langle N, (A_\imath), (u_i)\rangle$ εφόσον είναι κυρτός συνδυασμός προφίλ οφελών που αντιστοιχούν σε εκβάσεις του A, δηλαδή, εάν $v = \sum_{a\in A} \alpha_a u(a)$ για κάποιο σύνολο θετικών αριθμών $(\alpha_a)_{a\in A}$ με $\sum_{a\in A} \alpha_a = 1$.

3.4 Στρατηγικές ως αυτόματα

Όπως περιγράψαμε στην εισαγωγή του κεφαλαίου, το κύριο κατόρθωμα της θεωρίας των επαναλαμβανόμενων παιγνίων είναι ότι μας παρέχει ουσιαστικά μια ενόραση σχετικά με την δομή των συμπεριφορών που αναδύονται κατά την επαναλαμβανόμενη αλληλεπίδραση μεταξύ ατόμων. Σε αυτό το τμήμα του κεφαλαίου αναπτύσσουμε μια γλώσσα με την οποία θα μπορούμε να περιγράψουμε με ευκολία την δομή της ισορροπίας που θα βρίσκουμε. Έτσι λοιπόν ξεκινούμε ορίζοντας μια μηχανή, η οποία θα αποτελεί μια απλούστευση της διαδικασίας με την οποία ένας παίκτης υλοποιεί την στρατηγική του σε ένα επαναλαμβανόμενο παίγνιο. Μία **μηχανή** (ή **αυτόματο**)για έναν παίκτη i σε μη πεπερασμένο επαναλαμβανόμενο παίγνιο $\langle N, (A_\imath), (u_i)\rangle$ συνίσταται από τα εξής:

- Ένα σύνολο Q_i (το σύνολο των **καταστάσεων**)
- ένα στοιχείο $q_i^0 \in Q_i$ (η **αρχική κατάσταση**)
- Μια συνάρτηση $f_i : Q_i \to A_i$ η οποία αναθέτει μία ενέργεια (κίνηση) σε κάθε κατάσταση (η **συνάρτηση εξόδου**)
- Μια συνάρτηση $\tau_i : Q_i \times A \to Q_i$ η οποία αναθέτει μία κατάσταση σε κάθε ζεύγος που αποτελείται από μία κατάσταση και ένα προφίλ ενεργειών (η **συνάρτηση μετάβασης**)

Έτσι λοιπόν κάθε αυτόματο στην πρώτη περίοδο ξεκινά στην κατάσταση q_i^0 και το αυτόματο επιλέγει την ενέργεια $f_i(q_i^0)$. Στην συνέχεια οποτεδήποτε το αυτόματο βρίσκεται σε κάποια κατάσταση q_i, επιλέγει την ενέργεια $f_i(q_i)$ που αντιστοιχεί στην συγκεκριμένη κατάσταση. Η συνάρτηση μετάβασης τ_i καθορίζει πώς το αυτόματο μετακινείται από μία κατάσταση στην επόμενη. Πραγματικά εάν το αυτόματο είναι σε κατάσταση q_i και το προφίλ των ενεργειών που έχει επιλεγεί είναι το a, τότε το αυτόματο μεταβαίνει στην κατάσταση $\tau_i(q_i, a)$.

Προκειμένου να καταστήσουμε πιο κατανοητά τα παραπάνω θα παρουσιάσουμε μερικά παραδείγματα αυτομάτων, οι οποίες υλοποιούν συγκεκριμένες στρατηγικές για το επαναλαμβανόμενο *Δίλημμα των Φυλακισμένων*.

Παράδειγμα 3.1. *Ένα αυτόματο που υλοποιεί την "βλοσυρή" στρατηγική*

Το αυτόματο $(Q_i, q_i^0, f_i, \tau_i)$ που ορίζεται στην συνέχεια είναι η απλούστερη που υλοποιεί την "βλοσυρή" στρατηγική η οποία επιλέγει A για όσο χρονικό διάστημα και οι δύο παίκτες επιλέγουν A σε κάθε περίοδο στο παρελθόν διαφορετικά επιλέγει O. Με άλλα λόγια συνεργάζεται μέχρι την πρώτη αποστασία και στην συνέχεια "τιμωρεί" επ' αόριστον τον αντίπαλο παίκτη.

- $Q_i = q_\Delta, q_O$
- $q_i^0 = q_\Delta$

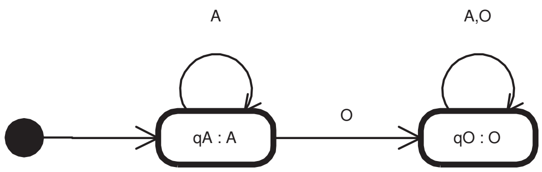

Figure 3.1: "βλοσυρή" στρατηγική

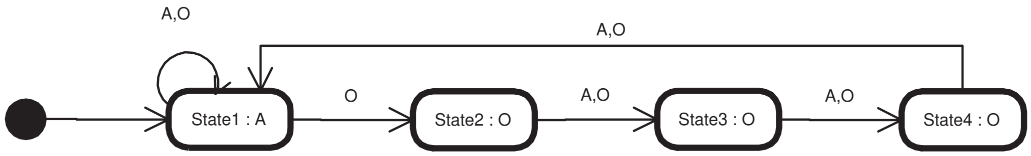

Figure 3.2: Το αυτόματο του παραδείγματος 2

- $f_i(q_\Delta) = \Delta$ και $f_i(q_O) = O$
- $\tau_i(q, a_j) = \begin{cases} q_\Delta & (q, a_j) = (q, \Delta) \\ q_O & \text{διαφορετικά} \end{cases}$

Παράδειγμα 3.2.

Το αυτόματο M_1 του παίκτη 1 που φαίνεται στο παρακάτω σχήμα, παίζει Α για όση διάρκεια ο παίκτης 2 παίζει Α, στην συνέχεια παίζει Ο για τρεις περιόδους και στην συνέχεια επιστρέφει στο να παίζει Α, εάν ο παίκτης 2 παίξει Ο την στιγμή που θα έπρεπε να παίξει Α. Πιο απλά μπορούμε να φανταστούμε ότι ο αντίπαλος παίκτης τιμωρείται για 3 περιόδους επειδή έπαιξε Ο και στην συνέχεια συγχωρείται. Επίσης είναι εύκολο να παρατηρήσει κανείς ότι το αυτόματο που υλοποιεί την στρατηγική αυτή πρέπει να έχει τουλάχιστον 4 καταστάσεις.

3.5 Στρατηγικές Σκανδάλης: Τα θεωρήματα Nash Folk

Στην συνέχεια θα μελετήσουμε τα σύνολα των ισορροπιών Nash για ένα μη πεπερασμένο επαναλαμβανόμενο παίγνιο. Θα δείξουμε ότι αυτό το σύνολο

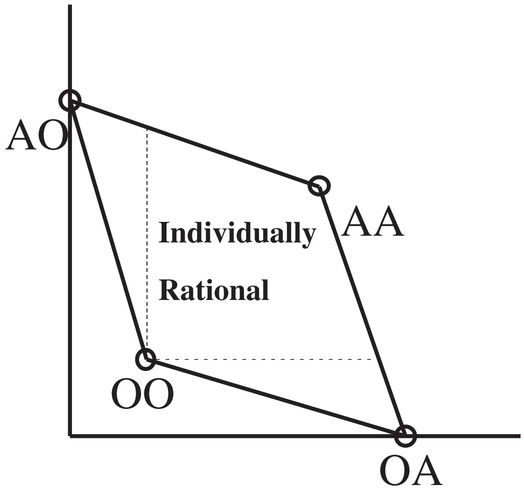

Figure 3.3: Η εφικτή περιοχή κέρδους στην περίπτωση του Διλήμματος των Φυλακισμένων

περιέχει εκβάσεις για το επαναλαμβανόμενο παίγνιο, οι οποίες δεν προέρχονται από την επανάληψη των ισορροπιών Nash του παιγνίου-βολής [24]. Έτσι, προκειμένου να υποστηρίξει μια τέτοια έκβαση κάθε παίκτης πρέπει να αποτρέπεται από την αποστασία υπό την απειλή τιμωρίας από τους συμπαίκτες του. Μια τέτοια τιμωρία μπορεί να πάρει πολλές μορφές. Μια πιθανότητα είναι κάθε παίκτης να χρησιμοποιεί μία στρατηγική σκανδάλης: με οποιαδήποτε λοιπόν αποστασία ο παίκτης ενεργοποιεί την στρατηγική σκανδάλης του και φέρει εις πέρας μια στρατηγική τιμωρίας που διαρκεί για πάντα. Συγκεκριμένα στην ισορροπία που θα μελετήσουμε εδώ κάθε παίκτης χρησιμοποιεί μια τέτοια στρατηγική.

Έστω λοιπόν G= $\langle N, (A_\imath), (\succ_\imath)\rangle$ ένα στρατηγικό παίγνιο και για κάθε $i \in N$ έστω u_i η συνάρτηση οφέλους που αντιπροσωπεύει την σχέση προτίμησης $(\succ_\imath)$. Επίσης αξίζει να υπενθυμίσουμε ότι λέμε ότι ένα διάνυσμα $v \in \Re^N$ αποτελεί ένα εφικτό προφίλ οφέλους του $\langle N, (A_\imath), (u_i)\rangle$ εφόσον είναι κυρτός συνδυασμός προφίλ οφελών που αντιστοιχούν σε εκβάσεις του A, δηλαδή, εάν $v = \sum_{a \in A} \alpha_a u(a)$ για κάποιο σύνολο θετικών αριθμών $(\alpha_a)_{a \in A}$ με $\sum_{a \in A} \alpha_a = 1$ (όπου θεωρούμε ότι οι αριθμοί α_a είναι ρητοί για ευκολία στις αποδείξεις). Έστω λοιπόν $w = \sum_{a \in A} \alpha_a u(a)$ ένα τέτοιο προφίλ και έστω $\alpha_a = \beta_a/\gamma$ για κάθε $a \in A$, όπου β_a είναι ένας ακέραιος και $\gamma = \sum_{a \in A} \beta_a$. Τότε η ακολουθία των εκβάσεων του επαναλαμβανόμενου παιγνίου αποτελείται από μία αόριστη επανάληψη ενός κύκλου μήκους γ στον οποίο κάθε $a \in A$ παίζεται για β_a περιόδους αποφέρει κατά μέσον όρο ένα προφίλ οφελών σε κάθε κύκλο και κατά συνέπεια σε ολόκληρο το παιχνίδι w.

Ορίζουμε ως minmax όφελος του i-οστού παίκτη στο G, το οποίο από εδώ και μπρος θα συμβολίζουμε ως v_i, να είναι το χαμηλότερο όφελος που οι υπόλοιποι παίκτες μπορούν να επιβάλλουν στον i-οστό παίκτη:
$u_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_{-i}, a_i)$

Ένα προφίλ οφελών w για το οποίο ισχύει $w_i \geq v_i$ για όλα τα $i \in N$ λέγεται **επιβλητέο ή ατομικά εύλογο (enforceable ή individually rational)**, ενώ έαν ισχύει $w_i > v_i$ για όλα τα $i \in N$ λέγεται λέγεται **αυστηρά επιβλητέο ή αυστηρά ατομικά εύλογο (strictly enforceable ή strictly individually rational)**. Εάν $a \in A$ είναι μια έκβαση του G για το οποίο $u(a)$ είναι (αυστηρώς) επιβλητέο στο G, τότε αναφερόμαστε στο a ως αυστηρά επιβλητέα έκβαση του G. Συμβολίζουμε με $p_{-i} \in A_{-i}$ μία από τις λύσεις του προβλήματος ελαχιστοποίησης που εμφανίζεται στο δεξιό σκέλος της παραπάνω σχέσης. Για κάθε προφίλ ενεργειών $A \in A$, έστω $b_i(a_{-i} \in A$ η πράξη του παίκτη i στο G που αποτελεί την καλύτερη απάντηση στο a_{-i} (π.χ. $b_i(a_{-i} \in B_i(a_{-i})$. Το σύνολο των πράξεων αυτών (p_{-i}) αποτελεί την πιο αυστηρή τίμωρία που μπορούν οι υπόλοιποι παίκτες να επιβάλλουν στον i στο G. Ακόμη αξίζει να αναφέρουμε ότι θεωρούμε ότι οι τιμωρίες που επιβάλλονται είναι ντετερμινιστικές. Έτσι οι "τιμωροί" δεν επιτρέπουμε να λειτουργούν με πιθανοτικό τρόπο, καθώς αυτό θα μετέβαλε το σύνολο των δυνατών προφίλ οφέλους και τα επιβλητέα οφέλη, αλλά όχι και την δομή του συνόλου των σημείων ισορροπιών του επαναλαμβανόμενου παιγνίου.

Στα επόμενα δύο αποτελέσματα δείχνουμε ότι το σύνολο των προφίλ οφέλους που αντιστοιχούν σε σημεία ισορροπιών Nash των απείρων επαναλαμβανόμενων παιγνίων στα οποία οι παίκτες αποτιμούν την ακολουθία των πληρωμών τους μέσω του κριτηρίου των ορίων των μέσων είναι το σύνολο *όλων* των δυνατών επιβλητέων προφίλ οφέλους του παιγνίου-βήματος [6]. Το τρίτο αποτέλεσμα αποδεικνύει ότι το ίδιο περίπου ισχύει και όταν οι παίκτες χρησιμοποιούν το κριτήριο των ελαττούμενων βαρών όταν οι παίκτες χρησιμοποιούν παράγοντα ελάττωσης δ που είναι κοντά στην μονάδα.

Θεώρημα 3.1. *Κάθε προφίλ οφέλους που αντιστοιχεί σε σημεία ισορροπιών του όριου του μέσου μη πεπερασμένου επαναλαμβανόμενου παιγνίου $G = \langle N, (A_\imath), (\succ_\imath) \rangle$ είναι ένα επιβλητέο προφίλ οφέλους του G. Το ίδιο ισχύει και για κάθε $\delta \in (0,1)$ κάθε προφίλ οφέλους που αντιστοιχεί σε σημεία ισορροπιών Nash του δ-ελαττουμένου μη πεπερασμένου επαναλαμβανόμενου παιγνίου του G.*

Απόδειξη Έσω w ένα προφίλ οφέλους του ορίου των μέσων μη πεπερασμένου παιγνίου του G το οποίο δεν είναι επιβλητέο. Έστω τώρα $w_i < v_i$, τότε το προφίλ οφέλους w δεν αντιστοιχεί σε σημείο ισορροπίας Nash του επαναλαμβανόμενου παιγνίου, καθώς κάθε προφίλ στρατηγικής s, η στρατηγική s'_i

του παίκτη i που ορίζεται ως $s'_i(h) = b_i(s_{-i}(h))$ για κάθε ιστορία h εξασφαλίζει στον παίκτη i όφελος τουλάχιστον u_i σε κάθε περίοδο. Το ίδιο επιχείρημα ισχύει και για οποιοδήποτε δ-ελαττούμενο μη πεπερασμένο επαναλαμβανόμενο παίγνιο του G. □

Θεώρημα 3.2 (*Nash folk θεώρημα για του κριτήριο του ορίου των μέσων*). *Κάθε δυνατό επιβλητέο προφίλ οφέλους του $G = \langle N, (A_\imath), (u_\imath)\rangle$ είναι προφίλ οφέλους που αντιστοιχεί σε σημεία ισορροπιών Nash του ορίου του μέσου μη πεπερασμένου επαναλαμβανόμενου παιγνίου G.*

Απόδειξη Έστω $w = \sum_{a \in A}(\beta_a/\gamma)u(a)$ ένα δυνατό επιβλητέο προφίλ οφέλους, όπου β_a για κάθε $a \in A$ είναι ένας ακέραιος και $\gamma = \sum_{a \in A}\beta_a$ και έστω (a^t) η κυκλική ακολουθία των προφίλ κινήσεων για τα οποία ο κύκλος (μήκους γ) περιέχει β_a επαναλήψεις του a για κάθε $A \in A$. Έστω s_i η στρατηγική του παίκτη i στο επαναλαμβανόμενο παίγνιο που επιλέγει a_i^t σε κάθε περίοδο t, εκτός εάν υπήρχε κάποια προηγούμενη περίοδο t' στην οποία κάποιος παίκτης (εκτός του i) παρεκτράπηκε από την $a^{t'}$, οπότε στην περίπτωση αυτή επιλέγει $(p_{-j})_i$, όπου j ο παίκτης που παρέκκλινε κατά την περίοδο t'. Το προφίλ στρατηγικών s είναι μία ισορροπία Nash του επαναλαμβανόμενου παιγνίου αφού ο παίκτης j που παρεκκλίνει λαμβάνει το πολύ όφελος ίσο με το minmax όφελος του v_j σε κάθε ακόλουθη περίοδο. Το προφίλ οφέλους που απορρέει από το s είναι w.□

Η στρατηγική s_i στην παραπάνω απόδειξη είναι μια στρατηγική σκανδάλης, καθώς τιμωρούν τον παίκτη ο οποίος παρέκκλινε από την αμοιβαίως αποδεκτή στρατηγική επ'αόριστον. Μια τέτοια τιμωρία είναι αχρείαστα σκληρή. Πραγματικά αρκεί η τιμωρία να επιβάλλεται για τόσες περιόδους (περιορίζοντας έτσι τις απολαβές του παίκτη στο minmax) όσες είναι αρκετές για να εξανεμίσουν το όφελος που απέκτησε ο παίκτης από την παρέκκλιση του.

Θεώρημα 3.3 (Nash folk θεώρημα για του κριτήριο των ελαττούμενων βαρών). *Έστω w ένα δυνατό και αυστηρά επιβλητέο προφίλ οφέλους του $G = \langle N, (A_\imath), (u_\imath)\rangle$. Για όλα τα $\varepsilon > 0$ υπάρχει ένα $\underline{\delta} < 1$ τέτοιο ώστε για εάν $\delta > \underline{\delta}$ τότε το ελαττούμενο μη πεπερασμένο παίγνιο του G έχει ισορροπία Nash, της οποίας το προφίλ οφέλους ικανοποιεί ότι $|w' - w| < \varepsilon$.*

Απόδειξη Ανάλογη με αυτήν του προηγουμένου θεωρήματος.□

3.6 Πεπερασμένα παίγνια: Ορισμοί

Το μοντέλο ενός πεπερασμένου παιγνίου περιγράφει μια αλληλεπίδραση κατά την οποία ορισμένοι παίκτες συμμετέχουν επανειλημμένως σε ένα στρατηγικό παίγνιο G, το οποίο ονομάζουμε *παιχνίδι-βήμα*. Από εδώ και μπρος περιορίζουμε το ενδιαφέρον μας σε παίγνια στα οποία το σύνολο των στρατηγικών είναι συμπαγές και η σχέση προτίμησης κάθε παίκτη είναι συνεχής. Εδώ όμως υπάρχει όριο στον αριθμό των φορών που μπορεί να επαναληφθεί το G (έστω T), παράλληλα οι παίκτες πραγματοποιούν τις κινήσεις τους ταυτόχρονα. Κατά την πραγματοποίηση μιας κίνησης του, κάθε παίκτης γνωρίζει όλες τις κινήσεις που έχουν πραγματοποιήσει κατά τις προηγούμενες περιόδους όλοι οι υπόλοιποι παίκτες. Η παραπάνω αλληλεπίδραση μοντελοποιείται από ένα εκτεταμένο παίγνιο με τέλεια πληροφορία (και ταυτόχρονες κινήσεις) ως εξής:

Έστω G= $\langle$ N, $(A_\imath)$, $(\succ_\imath)$ $\rangle$ ένα στρατηγικό παίγνιο, έστω A=$\times_{\imath \in N} A_\imath$. Ένα **πεπερασμένο επαναλαμβανόμενο παίγνιο του** G **T-περιόδων** είναι ένα εκτεταμένο παίγνιο με τέλεια πληροφορία και ταυτόχρονες κινήσεις $\langle$ N, H, P, $(\succ_\imath^*)\rangle$ στο οποίο

- $H = \{\emptyset\} \bigcup (\bigcup_{t=1}^{T} A^t)$, όπου το $\emptyset$ είναι η αρχική ιστορία
- $P(h) = N$ για κάθε μη τερματική ιστορία $h \in H$
- $(\succ_\imath^*)$ είναι μια σχέση διάταξης πάνω στο A^T η οποία επεκτείνει την σχέση διάταξης $(\succ_\imath)$, με την έννοια ότι ικανοποιεί την ακόλουθη ιδιότητα της *ασθενούς διαχωρησιμότητας*: εάν $(a^t) \in A^T$, $a \in A$, $\acute{a} \in A$, και $a \succ_\imath \acute{a}$ τότε

  $$(a^1, ..., a^{t-1}, a, a^{t+1}, ...) \succ_i^* (a^1, ..., a^{t-1}, \acute{a}, a^{t+1}, ...)$$

  για όλες τις τιμές του t.

Μια ιστορία είναι τερματική εάν και μόνο εάν είναι αποτελείται από T περιόδους. Μετά από μία οποιαδήποτε μη τερματική ιστορία κάθε παίκτης $i \in N$ επιλέγει μια κίνηση από το A_i. Επομένως η στρατηγική του i-οστού παίκτη είναι μια συνάρτηση που απεικονίζει σε κάθε πεπερασμένη ακολουθία εκβάσεων του G μια κίνηση στο A_i.

Στην συνέχεια θα επιβάλλουμε μερικούς επιπρόσθετους περιορισμούς πάνω στις σχέσεις προτίμησης (διάταξης) του κάθε παίκτη πέραν από τον περιορισμό της ασθενούς διαχωρησιμότητας. Έτσι λοιπόν από εδώ και μπρος θεωρούμε ότι η σχέση προτίμησης του i-οστού παίκτη $\succ_i^*$ βασίζεται πάνω σε μια συνάρτηση αποτίμησης (ή οφέλους) u_i στο G. Συγκεκριμένα υποθέτουμε ότι το εάν ισχύει το $(a^t) \succ_i^* (b^t)$ εξαρτάται αποκλειστικά και μόνο από την σχέση που έχουν μεταξύ τους οι ακολουθίες $(u_i(a^t))$ και $(u_i(b^t))$ των αποδοχών των παικτών στο G. Μάλιστα θα περιορίσουμε το ενδιαφέρον μας στην περίπτωση εκείνη, στην οποία η σχέση προτίμησης του i-οστού παίκτη $\succ_i^*$ στο πεπερασμένα επαναλαμβανόμενο παίγνιο αντικατοπτρίζεται από την συνάρτηση $\sum_{t=1}^{T} u_i(a^t)/T$, όπου u_i είναι η συνάρτηση οφέλους που αντιπροσωπεύει τις προτιμήσεις του i-οστού παίκτη στο παιχνίδι-βήμα. Θα αναφερόμαστε σε αυτό το παίγνιο ως το **Τ-περιόδων πεπερασμένο επαναλαμβανόμενο παίγνιο του** $\langle N, (A_\imath), (u_\imath)\rangle$.

3.7 Ισορροπία Nash σε πεπερασμένα παίγνια

Το προαίσθημα το οποίο βρίσκεται πίσω από τα θεωρήματα Nash folk είναι ότι στην περίπτωση των μη πεπερασμένων παιγνίων ένα αμοιβαίως επιθυμητό αποτέλεσμα μπορεί να συντηρείται, εφόσον κάθε παίκτης αποτρέπεται από το να αποκλίνει από τον κοινό στόχο υπό την απειλή ότι θα τιμωρηθεί εάν το επιχειρήσει. Το ίδιο επιχείρημα ισχύει, με ορισμένες τροποποιήσεις, σε μια μεγάλη κλάση των πεπερασμένων παιγνίων. Η ανάγκη για τροποποιήσεις προέρχεται από το γεγονός ότι το αποτέλεσμα στην τελευταία περίοδο μιας ισορροπίας Nash ενός πεπερασμένα επαναλαμβανόμενου παιγνίου πρέπει να είναι και ισορροπία Nash του παιγνίου-βήματος. Η απαίτηση αυτή ρίχνει μία σκιά πάνω στο υπόλοιπο παιχνίδι. Αυτή η σκιά είναι ακόμη πιο απειλητική στην ειδική περίπτωση που το όφελος που αποκομίζει κάθε παίκτης σε κάθε ισορροπία Nash του παιγνίου-βήματος είναι ίσο με το minmax όφελός του (όπως συμβαίνει στην περίπτωση του *διλήμματος των φυλακισμένων*. Στην περίπτωση αυτή το διαισθητικό επιχείρημα που βρισκόταν πίσω από τα θεωρήματα Nash folk αποτυγχάνει. Πραγματικά στην περίπτωση αυτή σε κάθε περίοδο η έκβαση του παιγνίου πρέπει να αποτελεί και σημείο ισορροπίας του παιγνίου βήματος, καθώς σε διαφορετική περίπτωση (εάν υπήρχε μια περίοδος στην οποία η έκβαση δεν ήταν σημείο ισορροπίας του παιγνίου-βήματος) τότε στην τελευταία περίοδο κάποιος παίκτης θα μπορούσε να αποκλίνει ατιμωρησία. Το παραπάνω επιχείρημα παρουσιάζεται με αυστηρότερο τρόπο παρακάτω:

Θεώρημα 3.4. *Εάν το προφίλ ισορροπίας κάθε ισορροπίας Nash του στρατηγικού παιγνίου G είναι το προφίλ (u_i) των minmax απολαβών στο G τότε για οποιαδήποτε τιμή του T η έκβαση $(a^1, ..., a^T)$ οποιαδήποτε ισορροπίας Nash του T-περιόδων επαναλαμβανόμενου παιγνίου του G έχει την ιδιότητα ότι το a^t είναι μια Nash ισορροπία του G για όλα τα $t = 1, ...T$.*

Απόδειξη Έστω $G = \langle N, (A_\imath), (u_\imath)\rangle$ και έστω $a = (a^1, ..., a^T)$ η έκβαση μιας ισορροπίας Nash s του T-περιόδων επαναλαμβανόμενου παιγνίου του G. Έστω τώρα ότι η a^τ δεν είναι ισορροπία Nash του G για κάποια περίοδο τ. Τότε υπάρχει κάποια $\tau \geq 1$ περίοδος, η οποία είναι η τελευταία για την οποία η a^t δεν είναι ισορροπία Nash του G. Ας υποθέσουμε ότι $u_i(a^t_{-i}, a_i) > u_i(a^t)$, τότε μπορούμε θεωρήσουμε μια στρατηγική $\hat{s}_i$ του i-οστού παίκτη, η οποία διαφέρει από την s_i μόνο στο ότι μετά από μία ιστορία $(a^1, ..., a^{t-1})$ επιλέγει

a_i και ύστερα από κάθε άλλη μακρύτερη ιστορία h επιλέγει μία ενέργεια (κίνηση), η οποία δεδομένου του προφίλ $s_{-i}(h)$ των ενεργειών που σχεδιάζουν οι υπόλοιποι παίκτες (μετά την ιστορία h), αποφέρει τουλάχιστον το minmax όφελος του i-οστού παίκτη. Η έκβαση του $(s_{-i}, \hat{s}_i)$ είναι μια τερματική ιστορία $\hat{a}$ η οποία είναι ακριβώς ίδια με την a μέχρι την περίοδο $t-1$ και ισχύει $u_i(\hat{a}^t) > u_i(a^t)$ και $u_i(\hat{a}^r) \geq u_i(a^r)$ για κάθε $r \geq t+1$. Επομένως ο i-οστός παίκτης προτιμά το $\hat{a}$ από το a, γεγονός που αντιτίθεται στην υπόθεσή μας ότι το s είναι ισορροπία Nash του επαναλαμβανόμενου παιγνίου. □

Αυτό το αποτέλεσμα ισχύει προφανώς για ένα πολύ μικρό ποσοστό του συνόλου των παιγνίων. Τώρα εάν το παίγνιο δεν ικανοποιεί τις προϋποθέσεις του παραπάνω θεωρήματος, δηλαδή το παίγνιο-βήμα έχει μία ισορροπία Nash a^* στην οποία το κέρδος κάποιου παίκτη υπερβαίνουν το minmax κέρδος του, τότε ο παίκτης μπορεί να τιμωρηθεί για την αποστασία του κατά την προτελευταία περίοδο του παιγνίου, οπότε η έκβαση του παιγνίου στην τελευταία περίοδο είναι a^*. Η τιμωρία όμως μπορεί να μην είναι αρκετή έτσι ώστε να αποτρέψει την αποστασία εάν η διαφορά ανάμεσα δύο minmax κέρδος του παίκτη και το κέρδος του στην a^* είναι μικρό. Παρόλα αυτά υπάρχει πάντα κάποιος ακέραιος L τέτοιος ώστε εάν η έκβαση του παιγνίου είναι a^* στις τελευταίες L περιόδους τότε κάθε απόκλιση από την αμοιβαία επιθυμητή συμπεριφορά σε οποιαδήποτε προηγούμενη περίοδο αποτρέπεται από την απειλή της τίμωρίας του παίκτη κατά τις τελευταίες L περιόδους (επιβάλλοντας του να εισπράττει μόνο το minmax κέρδος του). Επιπρόσθετα η τιμή του L είναι ανεξάρτητη του μήκους T του παιγνίου, έτσι ώστε εάν για κάθε παίκτη υπάρχει στο παιχνίδι μία ισορροπία Nash τέτοια ώστε το κέρδος του παίκτη να υπερβαίνει το minmax κέρδος του, τότε για αρκετά μεγάλο T όλα τα δυνατά αυστηρά επιβλητέα προφίλ κερδών μπορούν να επιτευχθούν (περίπου) από τον μέσο όρο του προφίλ κερδών σε μία ισορροπία Nash του T-περιόδων επαναλαμβανόμενου παιγνίου. Για χάριν ευκολίας θα αποδείξουμε τον παραπάνω ισχυρισμό μόνο στην περίπτωση που το παιχνίδι-βήμα έχει μία μόνο ισορροπία Nash στην οποία το κέρδος κάθε παίκτη υπερβαίνει το minmax κέρδος του. Επίσης θα περιοριστούμε σε μονοπάτια ισορροπίας τα οποία προέρχονται από επαναλήψεις μιας έκβασης του παιγνίου-βήματος.

Θεώρημα 3.5 (*Nash folk για πεπερασμένα επαναλαμβανόμενα παίγνια*). *Εάν το πεπερασμένο επαναλαμβανόμενο παίγνιο $\langle N, (A_\imath), (u_\imath)\rangle$ έχει μία Nash ισορροπία $\hat{a}$ στην οποία το κέρδος κάθε παίκτη υπερβαίνει το minmax κέρδος του u_i τότε για κάθε αυστηρά επιβλητέα έκβαση a^* του G και για κάθε $\varepsilon > 0$ τότε υπάρχει ένας ακέραιος $T*$ τέτοιο ώστε εάν $T > T^*$ το T-περιόδων επαναλαμβανόμενο παίγνιο του G έχει μια Nash ισορροπία στην οποία το κέρδος* *κάθε παίκτη i απέχει από το $u_i(a^*)$ το πολύ ε.*

Απόδειξη Θεωρούμε ότι η στρατηγική του i-οστού παίκτη διεκπεραιώνεται από το ακόλουθο αυτόματο. Το σύνολο των καταστάσεων αποτελείται από τις καταστάσεις $Norm^t$ για $t = 1, ..., T - L$ (όπου το L καθορίζεται αργότερα),Nash και $P(j)$ για κάθε $j \in N$. Κάθε παίκτης i επιλέγει a_i^* όταν βρίσκεται στην $Norm^t$ για όλες τις τιμές του t, $\hat{a}_i$ και τιμωρεί τον παίκτη j επιλέγοντας $(p_{-j})_i$ στην $P(j)$. Εάν ένας παίκτης j αποστατεί ενώ το αυτόματο βρίσκεται στην κατάσταση $Norm^t$ τότε μεταβαίνει στην $P(j)$, διαφορετικά μεταβαίνει στην $Norm^{t+1}$ εάν $t < T - L$ ή στην Nash εάν $t = T - L$. Μόλις το αυτόματο φθάσει στις καταστάσεις $P(j)$ ή $Nash$ τότε παραμένει σε αυτές μέχρι το τέλος. Έτσι έχουμε ως αποτέλεσμα ότι επιλέγεται η $a*$ στις πρώτες $T - L$ περιόδους και η $\hat{a}$ στις τελευταίες L περιόδους. Έτσι ανακεφαλαιώνοντας το αυτόματο του i-οστού παίκτη είναι η παρακάτω:

- Καταταστάσεις:$\{Norm^t : 1 \leq t \leq T - L\} \cup \{P(j) : j \in N\} \cup \{Nash\}$.
- Αρχική κατάσταση: $Norm^1$.
- Συνάρτηση εξόδου: Στην $Norm^t$ επιλέγει a_i^t, στην $P(j)$ επιλέγει $(p_{-j})_i$ και στην $Nash$ επιλέγει $\hat{a}_i$.
- Συνάρτηση μετάβασης:
  - Από $Norm^t$ μεταβαίνει στην $Norm^{t+1}$ εκτός εάν είτε $t = T - L$, οπότε στην περίπτωση αυτή μεταβαίνει στην $Nash$ είτε ακριβώς ένας παίκτης, έστω ο j απέκλινε από το $a*$, οπότε στην περίπτωση αυτή μεταβαίνει στην $P(j)$
  - Μόλις το αυτόματο φθάσει σε κάποια από τις καταστάσεις $P(j)$ (για κάθε $j \in N$) ή Nash παραμένει σε αυτές.

Επομένως απομένει να καθορίσουμε το L. Μία κερδοφόρος αποστασία είναι δυνατή μόνο σε μία από τις καταστάσεις $Norm^t$, έτσι λοιπόν προκειμένου να αποτρέψουμε μία αποστασία απαιτούμε το L να είναι αρκετά μεγάλο ώστε $max_{a_i \in A_i} u_i(a^*_{-i}, a_i) - u_i(a^*) \leq L(u_i(\hat{a}) - u_i)$ για όλα τα $i \in N$. Με άλλα λόγια ένα αρκετά μεγάλο L έτσι ώστε η αποστασία να καθίσταται ζημιογόνος για κάθε παίκτη. Τέλος, προκειμένου να επιτύχουμε ένα προφίλ κερδών που να απέχει το πολύ ε από το $u(a^*)$ επιλέγουμε T^* τέτοιο ώστε $|[(T^* - L)u_i(a^*) + Lu_i(\hat{a})]/T^* - u_i(a^*)| < \varepsilon$ για όλα τα $i \in N$. $\square$

Κεφάλαιο 4

Εισαγωγή Περιορισμών Πολυπλοκότητας

Στο κεφάλαιο αυτό, θα ερευνήσουμε την δομή των στρατηγικών ισορροπίας σε μη πεπερασμένα επαναλαμβανόμενα παίγνια στα οποία κάθε παίκτης λαμβάνει υπόψη του την πολυπλοκότητα της στρατηγικής του.

4.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο μέσω των θεωρημάτων Nash folk, δείξαμε ότι το σύνολο των δυνατών απολαβών που είναι συμβατά με την ισορροπία Nash σε ένα επαναλαμβανόμενο παίγνιο είναι τεράστιο. Πράγματι τα θεωρήματα εμπεριέχουν την κατασκευή κάποιων ισορροπιών χωρίς όμως να λαμβάνουν υπόψη καθόλου την πολυπλοκότητα των στρατηγικών αυτών. Έτσι παραδείγματος χάρη δεν προϋποθέτουν ότι οι στρατηγικές ισορροπίας πρέπει να είναι λογικές κάτω από οποιαδήποτε έννοια. Σε αυτό λοιπόν το κεφάλαιο θα επικεντρωθούμε στην δομή των στρατηγικών ισορροπίας, παρά στο σύνολο των απολαβών ισορροπίας χρησιμοποιώντας το εργαλείο της μηχανής (αυτομάτου) που εισαγάγαμε στο προηγούμενο κεφάλαιο.

Πραγματικά λοιπόν, στο επίκεντρο του κεφαλαίου αυτού βρίσκεται το γεγονός ότι κατά την επιλογή μιας συγκεκριμένης στρατηγικής κάποια οντότητα-παίκτης προσπαθεί να συγκεράσσει δύο συχνά αντιφατικές επιδιώξεις. Έτσι, από την μία πλευρά επιθυμεί η στρατηγική που επιλέγει να του επιφέρει όσο το δυνατόν μεγαλύτερο όφελος, από την άλλη όμως επιδιώκει η στρατηγική του να είναι όσο το δυνατόν απλούστερη. Οι λόγοι για τους οποίους ένας παίκτης δίνει τόσο μεγάλη έμφαση στην ελαχιστοποίηση της πολυπλοκότητας της στρατηγικής του μπορεί να είναι πολλοί. Έτσι παραδείγματος χάρη, ένα πολύπλοκο σχέδιο είναι πιο πιθανό να αποτύχει, είναι δυσκολότερο να κατανοηθεί από αυτούς που πρέπει να το φέρουν εις πέρας και μπορεί να απαιτεί για την υλοποίησή του περισσότερο χρόνο. Εμείς δεν θα εξετάσουμε τους λόγους αυτούς αλλά απλώς δεχόμαστε ότι η πολυπλοκότητα κοστίζει και είναι ένας παράγοντας που λαμβάνεται υπόψη από τους παίκτες.

Έτσι λοιπόν θα μελετήσουμε την αλληλεπίδραση παικτών που λαμβάνουν υπόψη τους την πολυπλοκότητα της συμπεριφοράς τους στα πλαίσια του συγκεκριμένου μοντέλου στο οποίο συμμετέχουν. Το μοντέλο των επαναλαμβανόμενων παιγνίων είναι ιδανικό για το σκοπό αυτό, καθώς σε αυτά η στρατηγική κάθε παίκτη καθορίζει τις ενέργειες του κάτω από πληθώρα διαφορετικών συνθηκών και επομένως μπορεί να θεωρηθεί ότι αντιστοιχεί σε ένα σύνολο κανόνων που διέπουν την συμπεριφορά τους και επιδέχεται ανάλυση ως προς την πολυπλοκότητα του.

Θα επιμείνουμε κυρίως στο μοντέλο των μη πεπερασμένων επαναλαμβανόμενων παιγνίων, τόσο επειδή επιτρέπουν την χρήση στρατηγικών που διαισθητικά μπορεί να κυμαίνονται από πολύ απλές μέχρι εξαιρετικά πολύπλοκες, όσο εξαιτίας της ευρείας χρήσης του συγκεκριμένου μοντέλου σε διάφορα γνωστικά πεδία.

4.2 Παίγνια Αυτομάτων

Στο κεφάλαιο αυτό θα εστιάσουμε την προσοχή μας, για λόγους απλότητας, σε ένα μη πεπερασμένο επαναλαμβανόμενο παίγνιο, στο οποίο οι σχέσεις προτίμησης υπακούουν τον κριτήριο της ελαττουμένης. Έτσι μελετούμε την συμπεριφορά παικτών στο δ-ελαττούμενο μη πεπερασμένα επαναλαμβανόμενο παίγνιο του $G = \langle \{1, 2\}, (A_i), (u_i) \rangle$. Θα μελετήσουμε τις συμπεριφορές τους αναλύοντας το *παίγνιο αυτομάτων* στο οποίο κάθε παίκτης επιλέγει ένα αυτόματο (μηχανή) για να παίξει το μη πεπερασμένα επαναλαμβανόμενο παίγνιο. Σε αυτό το κεφάλαιο ορίζουμε το αυτόματο κάθε παίκτη να είναι η ακόλουθη τετράδα $\langle Q_i, q_i^0, f_i, \tau_i \rangle$ στην οποία:

- Ένα σύνολο Q_i (το σύνολο των **καταστάσεων**)
- ένα στοιχείο $q_i^0 \in Q_i$ (η **αρχική κατάσταση**)
- Μια συνάρτηση $f_i : Q_i \to A_i$ η οποία αναθέτει μία ενέργεια (κίνηση) σε κάθε κατάσταση (η **συνάρτηση εξόδου**)
- Μια συνάρτηση $\tau_i : Q_i \times A_j \to Q_i$ (όπου $j \neq i$) η οποία αναθέτει μία κατάσταση σε κάθε ζεύγος που αποτελείται από μία κατάσταση και μια ενέργεια του αντιπάλου παίκτη (η **συνάρτηση μετάβασης**)

Ο ορισμός διαφέρει από αυτόν που δώσαμε στο προηγούμενο κεφάλαιο στο γεγονός ότι τώρα η συνάρτηση μετάβασης κάθε παίκτη εξαρτάται μόνο από την ενέργεια του αντιπάλου παίκτη και όχι από την έκβαση κάποιου στρατηγικού παιγνίου (δηλαδή εδώ από ένα ζεύγος πράξεων). Το αυτόματο αντιστοιχεί στην έννοια της στρατηγικής σε εκτεταμένα παίγνια, τα οποία απαιτούν να καθορίζεται μια ενέργεια-απάντηση του παίκτη για κάθε δυνατή ιστορία, ακόμη και αυτές που αποκλείονται από την ίδια την στρατηγική (π.χ. επειδή δεν συμφωνούν με το παρελθόν της). Εδώ επιθυμούμε το αυτόματο να αντιστοιχεί σε ένα σχέδιο δράσης με την συνήθη του έννοια, και επομένως η συνάρτηση μετάβασης κάθε παίκτη να δέχεται ως είσοδο μόνο την κίνηση του αντιπάλου.

Τώρα κάθε ζεύγος αυτομάτων (M_1, M_2) προκαλεί μία ακολουθία εξόδων $(a^t(M_1, M_2))_{t=1}^{\infty}$)του G και μία ακολουθία ζευγών καταστάσεων $(q^t(M_1, M_2))_{t=1}^{\infty})$ που ορίζεται ως εξής: για $i = 1, 2$ και $t \geq 1$ έχουμε

- $q_i^t(M_1, M_2) = q_i^0$
- $a_i^t(M_1, M_2) = f_i(q_i^t(M_1, M_2))$
- $q_i^{t+1}(M_1, M_2) = \tau_i(q_i^t(M_1, M_2), a_j^t(M_1, M_2))$ (όπου $j \neq i$)

Για μια πιο λεπτομερή συζήτηση πάνω στην έννοια του αυτομάτου προτείνεται το βιβλίο των Hopcroft και Ullman [7]. Περιορίζουμε κάθε παίκτη να επιλέγει ένα αυτόματο το οποίο έχει πεπερασμένο αριθμό καταστάσεων και συμβολίζουμε το σύνολο των αυτομάτων αυτών για τον παίκτη i ως M_i. Τέλος προκειμένου να συμπληρώσουμε την περιγραφή του παιγνίου αυτού πρέπει να περιγράψουμε τις προτιμήσεις των παικτών. Εάν υποθέσουμε ότι κάθε παίκτης i ενδιαφέρεται μόνο για το όφελός του $U_i(M_1, M_2) = (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_i(a^t(M_1, M_2))$ στο επαναλαμβανόμενο παίγνιο, τότε καταλήγουμε στο ίδιο συμπέρασμα με αυτό του επαναλαμβανόμενου παιγνίου αφού οι στρατηγικές σκανδάλης που είχαν χρησιμοποιηθεί κατά την απόδειξη μπορούν να υλοποιηθούν από πεπερασμένα αυτόματα. Εάν τώρα αντίθετα ο παίκτης δεν ενδιαφέρεται μόνο για το όφελος που αποκομίζει αλλά και για την πολυπλοκότητα της στρατηγικής που χρησιμοποιεί, τότε όπως θα δείξουμε τα αποτελέσματα που προκύπτουν είναι πολύ διαφορετικά από τα folk θεωρήματα.

Η πολυπλοκότητα ενός αυτομάτου μπορεί να οριστεί με διάφορους τρόπους. Εμείς θα επιλέξουμε μια κάπως απλοϊκή λύση επιλέγοντας ως *μέτρο της πολυπλοκότητας* $c(M)$ του αυτομάτου μας $\langle Q_i, q_i^0, f_i, \tau_i\rangle$ να είναι το πλήθος των καταστάσεών του (δηλαδή η πληθικότητα του Q).

Τέλος στον ορισμό που ακολουθεί θεωρούμε ότι στις σχέσεις προτίμησης των παικτών στο παίγνιο αυτομάτων παίζουν θετικό ρόλο οι απολαβές του παίκτη στο επαναλαμβανόμενο παίγνιο και αρνητικό ρόλο η πολυπλοκότητα του αυτομάτου του.

Ορισμός 4.1. *Ένα παίγνιο αυτομάτων ενός δ-ελαττούμενο μη πεπερασμένα επαναλαμβανόμενου παιγνίου του* $\langle\{1,2\},(A_i),(u_i)\rangle$ *είναι ένα στρατηγικό παίγνιο* $\langle\{1,2\},(M_i),(\succ_i)\rangle$ *στο οποίο για κάθε παίκτη i ισχύει*

- M_i *είναι το σύνολο όλων των πεπερασμένων αυτομάτων του παίκτη i στο μη πεπερασμένο παίγνιο*

- *η* $\succ_i$ *είναι μια σχέση διάταξης, η οποία είναι αύξουσα όσον αναφορά τις απολαβές του i-οστού παίκτη από το επαναλαμβανόμενο παίγνιο και φθίνουσα σε συνάρτηση με την πολυπλοκότητα του αυτομάτου: δηλαδή* $(M_1,M_2) \succ_i ((M_1',M_2')$ *εάν είτε* $U_i((M_1,M_2) > U_i((M_1',M_2')$ *και* $c(M_i)=c(M_i')$, *είτε* $U_i((M_1,M_2) = U_i((M_1',M_2')$ *και* $c(M_i) < c(M_i')$.

Δύο ειδικές περιπτώσεις για την σχέση προτίμησης που αξίζουν ειδικής μνείας. Η πρώτη είναι όταν οι επιδράσεις τόσο του οφέλους του παίκτη από το επαναλαμβανόμενο παίγνιο, όσο και η πολυπλοκότητα του αυτομάτου είναι αθροιστικές, δηλαδή η $\succ_i$ αναπαρίσταται από την $U_i((M_1,M_2) - \gamma c(M_i)$ για κάποιο $\gamma > 0$, όπου μπορούμε να θεωρήσουμε ότι το γ αντιπροσωπεύει το κόστος κάθε κατάστασης του αυτομάτου. Η δεύτερη περίπτωση είναι αυτή κατά την οποία εξετάζει τις προτιμήσεις του λεξικογραφικά, δηλαδή πρώτα τον ενδιαφέρουν οι απολαβές του και ύστερα η πολυπλοκότητα του αυτομάτου. Το ενδιαφέρον στοιχείο αυτού του μοντέλου είναι ότι πλησιάζει πολύ τις προτιμήσεις που χρησιμοποιούνται στο σύνηθες μοντέλο των επαναλαμβανόμενων παιγνίων, όπου δεν γίνεται λόγος για πολυπλοκότητα, οπότε το μοντέλο αυτό μπορεί να ειδωθεί ως άμεσος απόγονος του συνήθους μοντέλου.

Παράδειγμα 4.1.

$$G = \begin{array}{cc} & \begin{array}{cc} O & A \end{array} \\ \begin{array}{c} O \\ A \end{array} & \begin{bmatrix} (3,3) & (0,4) \\ (4,0) & (1,1) \end{bmatrix} \end{array}$$

Υποθέτουμε τώρα ότι το παιχνίδι G είναι το ''Δίλημμα των Φυλακισμένων'' με τις απολαβές που φαίνονται στον παραπάνω πίνακα. Αν θεωρήσουμε τώρα το αυτόματο που υλοποιεί την ''βλοσυρή'' στρατηγική όπως φαίνεται στο σχήμα, έχουμε ότι εφόσον ο κοινός παράγοντας ελάττωσης δ είναι αρκετά μεγάλος τότε το αυτόματο αυτό αποτελεί βέλτιστη απάντηση στον εαυτό του στο δ-ελλατούμενο. Ακόμη και χρησιμοποιώντας ένα πιο πολύπλοκο αυτόματο, ο παίκτης 1 δεν μπορεί να εξασφαλίσει μεγαλύτερο όφελος στο επαναλαμβανόμενο παίγνιο. Όμως υπάρχει αυτόματο, που δεδομένου ότι ο παίκτης 2 χρησιμοποιεί το αυτόματο M, επιτυγχάνει το ίδιο όφελος όντας απλούστερο. Αυτό το αυτόματο αποτελείται από μόνο μία κατάσταση και επιλέγει να παίζει

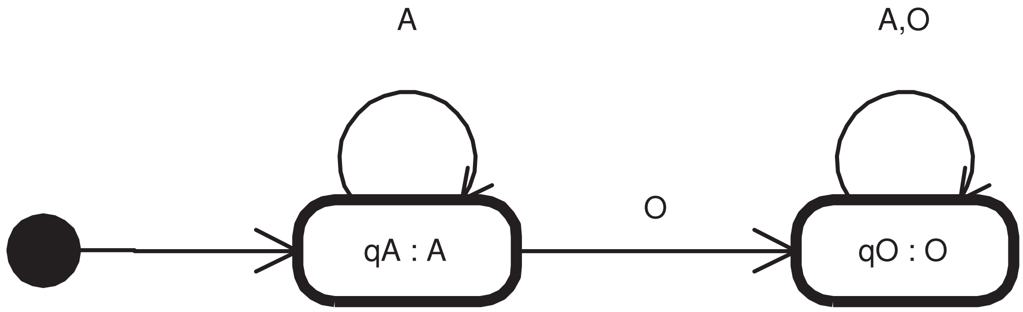

Figure 4.1: Το αυτόματο M που υλοποιεί την ”βλοσυρή” στρατηγική

πάντα A. Πραγματικά η κατάσταση qO έχει σχεδιαστεί με μοναδικό λόγο να δρα ως απειλή για τον αντίπαλο παίκτη, όμως στην ισορροπία η απειλή είναι περιττή αφού κάθε παίκτης παίζει πάντα A. Άρα κάθε παίκτης μπορεί να πετάξει την κατάσταση qO χωρίς να επηρεαστεί το αποτέλεσμα, άρα το (M, M) δεν είναι ισορροπία του παιγνίου αυτομάτων.

Βλέπουμε λοιπόν ότι στο παίγνιο αυτομάτων κάθε παίκτης πρέπει να ισοσταθμίσει την επιθυμία του για υψηλότερα κέρδη με την ανάγκη επιλογής μιας απλής στρατηγικής. Κατά μια έννοια λοιπόν αυτό το πρόβλημα είναι δυσκολότερο αφού ο παίκτης δεν προσπαθεί απλώς να βρει μια βέλτιστη στρατηγική στο επαναλαμβανόμενο παίγνιο αλλά ταυτόχρονα πρέπει να λαμβάνει υπόψη του την πολυπλοκότητα της στρατηγικής που ακολουθεί.

4.3 Η Δομή της Ισορροπίας στο Παίγνιο Αυτομάτων

Τώρα θα μελετήσουμε σε βάθος την δομή της ισορροπίας στην περίπτωση του παιγνίου αυτομάτων, όπως παρουσιάζεται από τους Rubinstein, Abreu και Piccione στα [20], [1] και [16]. Αρχικά γενικεύουμε την παρατήρηση που παρουσιάσαμε στην προηγούμενη παράγραφο ότι όταν ένα αυτόματο M_i κάποιου παίκτη i έχει μία κατάσταση που δεν χρησιμοποιείται όταν οι M_1 και M_2 συνεργάζονται, τότε το (M_1, M_2) δεν είναι σημείο ισορροπίας Nash, αφού η κατάσταση αυτή μπορεί να διαγραφεί χωρίς να επηρεάσει την έκβαση του παιγνίου, και ο παίκτης i επιλέγει το αυτόματο εκείνο στο οποίο η κατάσταση αυτή έχει διαγραφεί.

4.3.1 Ο αριθμός των καταστάσεων των δύο αυτομάτων είναι ο ίδιος και η ισορροπία του παιγνίου αυτομάτων είναι ισορροπία του επαναλαμβανόμενου παιγνίου

Λήμμα 4.2. *Εάν το (M_1^*, M_2^*) αποτελεί ισορροπία Nash του παιγνίου αυτομάτων τότε για κάθε κατάσταση q_i του αυτομάτου M_i^* υπάρχει μία περίοδος t τέτοια ώστε $q_i^t(M_1^*, M_2^*) = q_i$.*

Το επόμενο αποτέλεσμά δηλώνει ότι σε μία ισορροπία Nash, κάθε αυτόματο που συμμετέχει σ'αυτήν έχει τον ίδιο αριθμό καταστάσεων και οποιαδήποτε ισορροπία Nash του παιγνίου αυτομάτων αντιστοιχεί σε ισορροπία Nash του επαναλαμβανόμενου παιγνίου.

Πρόταση 4.3. *Εάν το (M_1^*, M_2^*) αποτελεί ισορροπία Nash του παιγνίου αυτομάτων τότε*

- $c(M_1^*) = c(M_2^*)$
- *το ζεύγος των στρατηγικών του επαναλαμβανόμενου παίγνιου που αντιστοιχούν στο ζεύγος αυτομάτων (M_1^*, M_2^*) είναι ισορροπία Nash του επαναλαμβανόμενου παιγνίου.*

Απόδειξη Καταρχάς υπενθυμίζουμε ότι στην ισορροπία όλες οι καταστάσεις του αυτομάτου χρησιμοποιούνται τουλάχιστον μία φορά. Σε διαφορετική περίπτωση ο παίκτης στον οποίο αντιστοιχεί το αυτόματο με τις άεργες καταστάσεις μπορεί να διαγράψει τις καταστάσεις αυτές και να επιλέξει έτσι να χρησιμοποιεί ένα απλούστερο αυτόματο χωρίς να ελαττώσει το όφελος που προσλαμβάνει στο επαναλαμβανόμενο παίγνιο. □

Τώρα, για κάθε στρατηγική s_j του παίκτη j στο επαναλαμβανόμενο παίγνιο συμβολίζουμε $U_j(M_i^*, s_j)$ το όφελος του παίκτη j στο επαναλαμβανόμενο παίγνιο όταν χρησιμοποιεί την s_j και ο παίκτης i χρησιμοποιεί την στρατηγική που αντιστοιχεί στο M_i^*. Αφού το αυτόματο M_i^* είναι πεπερασμένο, το πρόβλημα του παίκτη j της εύρεσης βέλτιστης απόκρισης στο αυτόματο M_i^*στο επαναλαμβανόμενο παίγνιο ($\max_{s_j} U_j(M_i^*, s_j)$), αγνοώντας την πολυπλοκότητα, έχει λύση (Derman [5] Θεώρημα 1, σελ 23). Μπορούμε όμως να περιγράψουμε την βέλτιστη απόκριση ως εξής: έχοντας $M_i^* = \langle Q_i, q_i^0, f_i, \tau_i \rangle$, για κάθε $q \in Q_i$, έστω $V_j(q) = \max_{s_j} U_j(M_i^*(q), s_j)$, τότε η $M_i^*(q)$ διαφέρει από την M_i^* μόνο στην αρχική κατάσταση, q. Για κάθε $q \in Q_i$, έστω $A_j(q)$ το σύνολο λύσεων στον πρόβλημα:

$$\max_{a_j \in A_j} \{u_j(f_i(q), a_j) + \delta V_j(\tau_i(q, a_j))\}.$$

Η στρατηγική του παίκτη j είναι βέλτιστη απόκριση, στο επαναλαμβανόμενο παίγνιο, στην στρατηγική που αντιστοιχεί στο M_i^*, εάν και μόνο εάν η ενέργεια

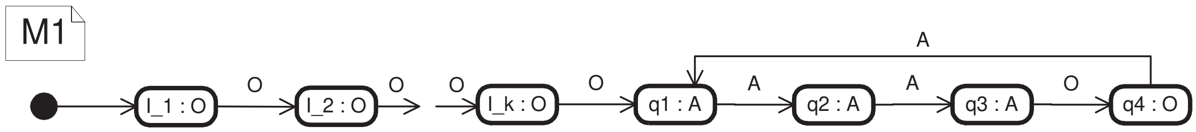

Figure 4.2: Το αυτόματο $M1$

που επιλέγει όταν το αυτόματο του παίκτη i βρίσκεται στην κατάσταση q ανήκει στο σύνολο $A_j(q)$.

Ας επιστρέψουμε τώρα στο πρόβλημα $\max_{M_j} U_j(M_i^*, M_j)$. Αυτό το πρόβλημα διαφέρει από το $\max_{s_j} U_j(M_i^*, s_j)$, μόνο στο γεγονός ότι εδώ ο παίκτης j αντί να επιλέγει στρατηγική επιλέγει αυτόματο. Όμως ακόμη και με τον περιορισμό αυτό δεν έχουμε πρόβλημα, καθώς ο παίκτης j μπορεί να εξασφαλίσει όφελος $\max_{s_j} U_j(M_i^*, s_j)$ χρησιμοποιώντας το παρακάτω αυτόματο:

- Σύνολο καταστάσεων είναι Q_i.
- Αρχική κατάσταση q_i
- Συνάρτηση εξόδου f_j που ορίζεται ως: $f_j(q) = a_q^* \in A_j(q)$
- Συνάρτηση μετάβασης $\tau_j(;, \chi) = t_i(q, f_i(q)$ για κάθε $x \in A_i$.

  Από τα παραπάνω προκύπτει ότι αφού το (M_1^*, M_2^*) είναι ισορροπία Nash του παιγνίου αυτομάτων έπεται ότι $c(M_J^*) \leq c(M_i^*)$ και επομένως $c(M_J^*) \leq c(M_i^*)$. Επιπλέον, αφού ο παίκτης j μπορεί να χρησιμοποιήσει το αυτόματό του προκειμένου να επιτύχει κέρδος $\max_{s_j} U_j(M_i^*, s_j)$ έπεται ότι οι στρατηγικές που προέρχονται από το ζεύγος των αυτομάτων (M_i^*, M_j) αποτελούν ισορροπία Nash του επαναλαμβανόμενου παιγνίου.

4.3.2 Ύπαρξη 1-1 αντιστοιχίας ανάμεσα στις ενέργειες στο μονοπάτι ισορροπίας

Τώρα θα παρουσιάσουμε ένα αποτέλεσμα το οποίο έχει πολύ σημαντικές επιπτώσεις για την δομή των ισορροπιών Nash στο παίγνιο αυτομάτων. Προκειμένου να διευκολύνουμε την κατανόηση του αποτελέσματος της παραγράφου αυτής θα ξεκινήσουμε με ένα παράδειγμα για την περίπτωση του μη πεπερασμένα επαναλαμβανόμενου παιγνίου του "Διλήμματος των Φυλακισμένων".

Παράδειγμα 4.2.

Το ζεύγος των αυτομάτων δημιουργεί ένα μονοπάτι το οποίο ξεκινά με $k \geq 2$ περιόδους, στις οποίες η έκβαση του παιγνίου είναι (O,O) (όπου οι παίκτες

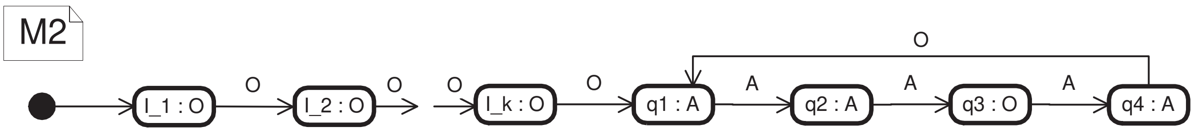

Figure 4.3: Το αυτόματο $M2$

επιδεικνύουν τις απειλές τους) και μετά αποτελείται από έναν κύκλο μήκους 4, στις οποίες η έκβαση του παιγνίου είναι (A,A),(A,A),(A,O) και (O,A), και ο οποίος επαναλαμβάνεται επ'αόριστον. Οποιαδήποτε απόκλιση παίκτη από την προαναφερθείσα συμπεριφορά στον κύκλο κάνει το αυτόματο του αντιπάλου να μεταβεί στην αρχική του κατάσταση τιμωρώντας έτσι τον αποστάτη για κ περιόδους. Το ζευγάρι των αυτομάτων αυτών είναι ισορροπία Nash για το επαναλαμβανόμενο παίγνιο όταν ο παράγοντας ελάττωσης δ είναι κοντά στην μονάδα. Όμως το ζεύγος των αυτομάτων δεν συνιστά ισορροπία στο παίγνιο αυτομάτων. Για να το επιβεβαιώσουμε αυτό ας εξετάσουμε το αυτόματο M_1. Σε κάθε μία από τις 3 καταστάσεις, q_1, q_2 και q_3, ο παίκτης 1 επιλέγει την ίδια κίνηση O. Όμως θα μπορούσε να χρησιμοποιήσει το αυτόματο M_1', στο οποίο οι 3 καταστάσεις έχουν αντικατασταθεί από μία την q, στην οποία επιλέγει την κίνηση A (αρνείται να ομολογήσει). Εάν το M_1', φθάσει στην q, τότε παραμένει σε αυτήν μέχρι ο παίκτης 2 να επιλέξει την κίνηση A οπότε μεταβαίνει στην q_4, οπότε ο παίκτης 2 επιλέγει O. Οι μεταβάσεις από την κατάσταση I_k, όταν ο παίκτης 2 παίζει O και από την q_4, εάν ο παίκτης 2 παίζει A, διορθώνονται στην q. Ακόμη το ζεύγος (M_1', M_2) παράγει την ίδια ακολουθία εκβάσεων όπως και το (M_1, M_2), οπότε ο παίκτης 1 μπορεί να επιλέξει το αυτόματο M_1' αποκομίζοντας κέρδος αφού ελαττώνει την πολυπλοκότητά της στρατηγικής του.

Γενικότερα θα δείξουμε ότι εάν ένα ζεύγος αυτομάτων που συνιστά ισορροπία Nash παράγει αποτελέσματα στα οποία ένας από τους δύο παίκτες πραγματοποιεί την ίδια κίνηση σε δύο διαφορετικές περιόδους, τότε και ο αντίπαλος παίκτης πραγματοποιεί την ίδια κίνηση στις συγκεκριμένες περιόδους (αντίθετα από ότι συνέβαινε στο παράδειγμα που μόλις περιγράψαμε στις περιόδους $k+1$ και $k+2$).

Πρόταση 4.4. *Εάν το ζεύγος (M_1^*, M_2^*) είναι ισορροπία Nash του παιγνίου αυτομάτων, τότε υπάρχει μια 1-1 αντιστοιχία μεταξύ των παικτών 1 και 2, όπως καθορίζονται από τα αυτόματα M_1^* και M_2^*, δηλαδή εάν $a_i^t(M_1^*, M_2^*) = a_i^s(M_1^*, M_2^*)$ για κάποια $t \neq s$, τότε $a_j^t(M_1^*, M_2^*) = a_i^s(M_1^*, M_2^*)$.*

Απόδειξη Έστω $M_i^* = (Q_i, q_i^0, f_i, \tau_i)$, σύμφωνα με την προηγούμενη πρόταση το M_j^* είναι βέλτιστη απάντηση στο M_i^* (υπό την έννοια του επαναλαμβανόμενου παιγνίου), οπότε $f_j(q_j^t(M_1^*, M_2^*)) \in A_j(q_j^t(M_1^*, M_2^*))$ για όλα τα t.

Έτσι, εάν υπάρχουν δύο περίοδοι t και s στις οποίες
$a_j^t(M_1^*, M_2^*) = a_j^s(M_1^*, M_2^*)$ και $a_i^t(M_1^*, M_2^*) \neq a_i^t(M_1^*, M_2^*)$, τότε

$q_i^t(M_1^*, M_2^*) \neq q_i^s(M_1^*, M_2^*)$,

$f_j(q_j^t(M_1^*, M_2^*) \in A_j(q_i^t(M_1^*, M_2^*))$ και

$f_j(q_j^s(M_1^*, M_2^*) \in A_j(q_i^s(M_1^*, M_2^*))$

Όπως έχουμε δει στην ισορροπία ισχύει $c(M_i^*) = c(M_j^*)$, όμως ακολούθως περιγράφουμε ένα αυτόματο που εκτελεί μία βέλτιστη στρατηγική για τον παίκτη j χρησιμοποιώντας μόνο $c(M_i^*) - 1$ καταστάσεις. Το αυτόματο αυτό έχει:

- Σύνολο καταστάσεων $Q_i - \{q_i^s\}$.
- Αρχική κατάσταση q_i^0 εάν $q_i^s \neq q_i^0$ αλλιώς q_i^t.
- Συνάρτηση εξόδου που ορίζεται από την f_j
- Η συνάρτηση μετάβασης που ορίζεται ως εξής:

$$\tau_i(q, x) = \begin{cases} \tau_i'(q, f_j(q)) & \text{εάν } q \neq q_j^t \\ \tau_i'(q_i^s, f_j(q_i^s)) & \text{εάν } q = q_j^t \text{ και } x = a_i^s(M_1^*, M_2^*) \\ \tau_i'(q_i^t, f_j(q_i^t)) & \text{εάν } q = qq_j^t \text{ και } x \neq a_i^s(M_1^*, M_2^*) \end{cases}$$

όπου η τ_i' είναι ακριβώς η ίδια με την τ_i εκτός από το ότι $\tau_i'(q, x) = q_i^t$ όπου $\tau_i(q, x) = q_i^s$. $\square$

Το αποτέλεσμα αυτό έχει άμεσες και σημαντικότατες επιπτώσεις στην ακολουθία των εκβάσεων ισορροπίας όπου το G είναι 2×2 παίγνιο. Για παράδειγμα, εάν το G είναι το "Δίλημμα των Φυλακισμένων", ότι οι εκβάσεις του παιγνίου σε ισορροπία πρέπει να ανήκουν είτε στο σύνολο $(O, O), (A, A)$ είτε στο σύνολο $(O, A), (A, O)$.

4.3.3 Ύπαρξη 1-1 αντιστοιχίας ανάμεσα στις καταστάσεις στο μονοπάτι ισορροπίας

Τώρα επιστρέφουμε στο ζήτημα της δομής των αυτομάτων ισορροπίας. Επειδή το αυτόματο κάθε παίκτη είναι πεπερασμένο, υπάρχει ένας ελάχιστος αριθμός t' τέτοιος ώστε για κάποιο $t > t'$, να έχουμε $q_i^t = q_i^{t'}$ για $i = 1, 2$. έστω τώρα t^* η μικρότερη (κατ' αύξων αριθμό) περίοδος t, τέτοια ώστε $t > t'$ και $q_i^t = q_i^{t'}$ για $i = 1, 2$. Η ακολουθία των ζευγών καταστάσεων που ξεκινά από την περίοδο t' είναι ένας κύκλος μήκους $t^* - t'$. Αναφερόμαστε σε αυτό το στάδιο ως *κυκλική φάση* και στο προηγούμενο στάδιο ως *εισαγωγική φάση.*

Στην συνέχεια θα δείξουμε ότι τα σύνολα των καταστάσεων που χρησιμοποιούν οι παίκτες κατά την εισαγωγική φάση και την κυκλική φάση είναι ξένα μεταξύ τους. Επιπλέον θα δείξουμε ότι κατά την εισαγωγική φάση, κάθε κατάσταση χρησιμοποιείται ακριβώς μία φορά και ότι κατά την κυκλική φάση, κάθε μία από τις καταστάσεις επαναλαμβάνεται ακριβώς μια φορά σε κάθε κύκλο. Επομένως στην ισορροπία υπάρχει μία 1-1 αντιστοιχία ανάμεσα στις καταστάσεις των αυτομάτων που χρησιμοποιούν οι παίκτες. Πρακτικά αυτό σημαίνει ότι σε κάθε περίοδο κάθε αυτόματο "γνωρίζει" την κατάσταση του άλλου αυτομάτου.

Πρόταση 4.5. *Εάν (M_1^*, M_2^*) είναι μια ισορροπία του παιγνίου αυτομάτων, τότε υπάρχει μία περίοδος t^* και ένας ακέραιος $l < T^*$ τέτοιος ώστε για i=1,2 οι καταστάσεις στην ακολουθία $(q_i^t(M_1^*, M_2^*)_{t=1,2,\ldots,t^*-1}$ είναι διαφορετικές και $q_i^t(M_1^*, M_2^*) = q_i^{t-l}(M_1^*, M_2^*)$ για $t \geq t^*$*

*Απόδειξη*Έστω t^* η πρώτη περίοδος στην οποία μία από τις καταστάσεις από κάποια από τα δύο αυτόματα εμφανίζεται για δεύτερη φορά. Ορίζουμε δηλαδή ως tj να είναι ο ελάχιστος χρόνος για τον οποίο υπάρχει παίκτης i και περίοδος $t_i < t^*$ τέτοιοι ώστε $q_i^{t*} = q_i^{t_i}$ και επομένως $a_i^{t*} = a_i^{t_i}$. Όμως τώρα από την προηγούμενη πρόταση έχουμε ότι αφού $a_i^{t*} = a_i^{t_i}$ έπεται έχουμε ότι $a_j^{t*} = a_j^{t_i}$. Από αυτό έπεται ότι για κάθε $k \geq 0$ $q_i^{t*+k} = q_i^{t_i+k}$, αφού κατά την ισορροπία όλες οι καταστάσεις εμφανίζονται τουλάχιστον μία φορά, έχουμε ότι $c(M_i^*) = t^* - 1$. Από τον ορισμό του t^* έχουμε ότι όλες οι καταστάσεις του M_j^* μέχρι την $t^* - 1$ περίοδο είναι διαφορετικές και επειδή κατά την ισορροπία ισχύει $c(M_i^*) = c(M_j^*)$ υπάρχει $t_j < t^*$ τέτοιο ώστε $q_i^{t_j} = q_i^{t^*}$. Επομένως απομένει να δείξουμε ότι $t_j = t_i$. Έστω ότι δεν ισχύει, έστω π.χ. ότι ισχύει $t_j > t_i$, τότε ο παίκτης j μπορεί να παράγει την ίδια ακολουθία εκβάσεων με ένα αυτόματο στο οποίο η κατάσταση $q_j^{t_i}$ έχει εξαφανιστεί δημιουργώντας την μετάβαση από την $q_j^{t_i-1}$ στην $q_j^{t^*}$, παραλείποντας την $q_j^{t_i}$. Αυτό ισχύει επειδή το αυτόματο αυτό έχει την ιδιότητα ότι οποτεδήποτε το αυτόματο M_i^* είναι στην κατάσταση q, ο παίκτης j επιλέγει μια ενέργεια από το $A_j(q)$, δηλαδή μια βέλτιστη στρατηγική για το επαναλαμβανόμενο παίγνιο εκτελείται από ένα αυτόματο που έχει λιγότερες καταστάσεις από το M_j^*, πράγμα που αντιτίθεται στην υπόθεση ότι το (M_1^*, M_2^*) είναι ισορροπία. □

Σχόλιο Τα αποτελέσματα της ενότητας αυτής αποδεικνύουν ότι το σύνολο ισορροπιών του παιγνίου αυτομάτων είναι πολύ μικρότερο από αυτό των επαναλαμβανόμενων παιγνίων. Φυσικά προκειμένου να καθορίσουμε το σύνολο των ισορροπιών απαιτείται να γνωρίζουμε τις σχέσεις προτιμήσεις του συγκεκριμένου προβλήματος. Έτσι ας θεωρήσουμε την περίπτωση του "Διλήμματος των

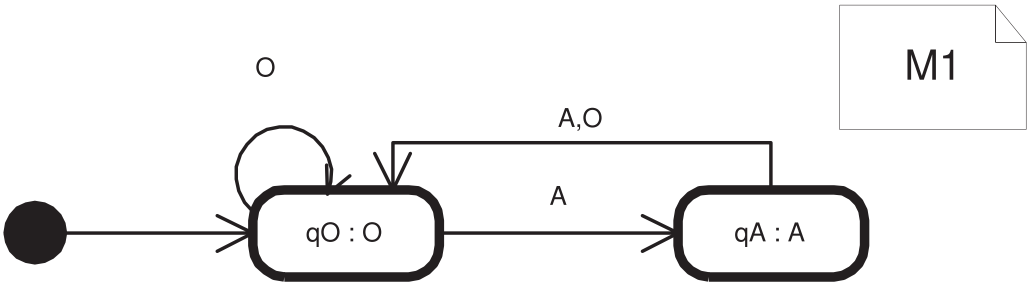

Figure 4.4: Το αυτόματο του παίκτη 1

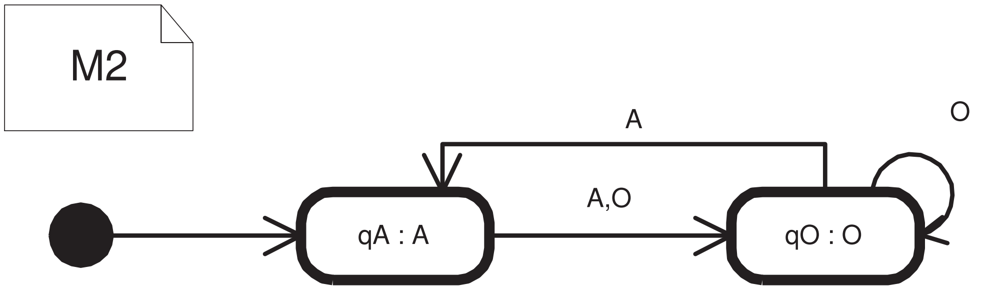

Figure 4.5: Το αυτόματο του παίκτη 2

Φυλακισμένων" που οι σχέσεις προτίμησης των δύο παικτών όσον αναφορά το προσδοκώμενο κέρδος και την πολυπλοκότητα είναι λεξικογραφικές. Όπως γνωρίζουμε ήδη το μονοπάτι ισορροπίας αποτελείται είτε από $\{(O,O),(A,A)\}$ είτε από $\{(O,A),(A,O)\}$. Έτσι στην περίπτωση που θέλουμε το μονοπάτι ισορροπίας να αποτελείται μόνο από $\{(A,O),(O,A)\}$, μπορούμε να το επιτύχουμε χωρίς εισαγωγική φάση. Πράγματι για κάθε ζεύγος θετικών ακεραίων n_1, n_2 που ικανοποιεί την σχέση $4n_i/(n_1+n_2) > 1$ για κάθε i και αρκετά μεγάλο δ, υπάρχει μία ισορροπία του παιγνίου αυτομάτων στην οποία ο κύκλος ξεκινά αμέσως και αποτελείται από n_1 γύρους στους οποίους εμφανίζεται το (A,O) και n_2 γύρους στους οποίους εμφανίζεται το (O,A), χωρίς εισαγωγική φάση. (Ο περιορισμός $4n_i/(n_1+n_2) > 1$ εξασφαλίζει για κάθε παίκτη ότι το αναμενόμενο κέρδος του υπερβαίνει το minmax κέρδος του δηλαδή το 1.)

Ένα ζεύγος αυτομάτων ισορροπίας στην περίπτωση που $n_1 = n_2 = 1$ φαίνεται στο παρακάτω σχήμα. Η ερμηνεία που μπορούμε να αποδώσουμε στην ισορροπία αυτή είναι ότι οι παίκτες εναλλάσσουν ρόλους, όντας εναλλάξ ευγενικοί ο ένας στον άλλο. Έτσι λοιπόν, μπορούμε να δούμε την έκβαση (A,O), ως το σημείο κατά το οποίο ο παίκτης 1 δίνει ένα δώρο στον παίκτη 2, ενώ την έκβαση (O, A), ως το σημείο κατά το οποίο ο παίκτης 2 δίνει ένα δώρο στον παίκτη 1. Στην ισορροπία ο παίκτης δεν ενδιαφέρεται εάν ο αντίπαλος απορρίπτει το δώρο του (π.χ. εάν επιλέγει Α ενώ θα έπρεπε να επιλέξει Ο και να αποδεχθεί το δώρο), αλλά επιμένει ο αντίπαλός του να του προσφέρει ένα δώρο (να παίξει Α) σε περιόδους που αναμένει δώρο. Εάν δεν δεχθεί το δώρο του, τότε δεν μεταβαίνει στην κατάσταση που είναι ευγενικός.

Κεφάλαιο 5

Επίλυση Παραδόξων

5.1 Εισαγωγή

Ένα ζήτημα το οποίο ανακύπτει συχνά στην θεωρία παιγνίων είναι αυτό του "παραδόξου του πεπερασμένου ορίζοντα" [9],[19]. Το πιο αναγνωρίσιμο παράδειγμα των παραδόξων αυτών είναι αυτό του πεπερασμένου διλήμματος των φυλακισμένων. Στο συγκεκριμένο παιχνίδι όπως έχουμε ήδη αναπτύξει, οι συνήθεις παιγνιο-θεωρητικές θεωρήσεις οδηγούν σε αποτελέσματα τα οποία απέχουν κατά πολύ από τα αναμενόμενα και που έρχονται σε σύγκρουση με την εμπειρία μας. Το γεγονός αυτό ώθησε πολλούς, που ασχολούνται με την θεωρία παιγνίων να αμφισβητήσουν την καταλληλότητα των βασικών σκεπτικών λύσης. Επίσης παρείχε το έναυσμα προς μια κατεύθυνση έρευνας που επιχειρεί να εξηγήσει αυτά τα παράδοξα αυτά ως αποτελέσματα της περιορισμένης "λογικότητας" του παίκτη.

Θα ξεκινήσουμε την ανάλυσή μας με ένα παίγνιο το οποίο είναι απλούστερο από τα συνήθη παίγνια που χρησιμοποιούνται για την ανάλυση των παραδόξων αυτών αλλά παρόλα αυτά εμπεριέχει όλα τα στοιχεία που είναι απαραίτητα για την ανάλυσή μας, έτσι θα διευκολύνουμε την κατανόηση της ανάλυσης του πεπερασμένου "Διλήμματος των Φυλακισμένων" που θα ακολουθήσει. Έτσι λοιπόν το παίγνιο μας μοιάζει με το πεπερασμένο "Δίλημμα των Φυλακισμένων", με μία μετατροπή, ότι μόλις ένας από τους δύο παίκτες αποστατήσει και πάψει να συνεργάζεται με τον άλλο το παίγνιο τερματίζεται.

Έστω $G_0(T)$, το εκτεταμένο παίγνιο με δύο παίκτες 1 και 2. Το παίγνιο διαρκεί T περιόδους, που αριθμούνται 1,2,...,T. Σε κάθε περίοδο, οι δύο παίκτες πρέπει να επιλέξουν ταυτόχρονα είτε να σταματήσουν το παίγνιο, είτε να το συνεχίσουν. Έτσι δηλαδή σε κάθε περίοδο κάθε παίκτης έχει δύο επιλογές: είτε να σταματήσει (Σtop) είτε να συνεχίσει (Cont). Το παίγνιο συνεχίζει (από την περίοδο t στην περίοδο $t+1$), μόνο εάν και οι δύο παίκτες επιλέξουν

Cont. Κάθε παίκτης κερδίζει οφέλη καθόλη την διάρκεια του παιγνίου. Στο τέλος κάθε περιόδου (συμπεριλαμβανομένης και της T) στην οποία και οι δύο παίκτες επιλέγουν Cont, κάθε παίκτης προσθέτει 3 στο συνολικό όφελός του. Εάν και οι δύο παίκτες επιλέξουν Stop, τότε κάθε παίκτης κερδίζει 1. Εάν ο ένας παίκτης επιλέξει Stop και ο άλλος Cont, τότε ο παίκτης που επιλέγει να σταματήσει το παιχνίδι κερδίζει 5, ενώ ο αντίπαλος δεν κερδίζει τίποτα (παίρνει 0). Το σημαντικό χαρακτηριστικό γνώρισμα των απολαβών αυτών είναι ότι το ιδανικό σενάριο για κάθε παίκτη i είναι να σταματήσει το παιχνίδι την περίοδο t, εάν και μόνο εάν πιστεύει ότι ο παίκτης j σκοπεύει να σταματήσει το παιχνίδι κατά την περίοδο t ή κατά την περίοδο $t+1$. Εάν ο παίκτης i πιστεύει ότι ο παίκτης j δεν πρόκειται να επιλέξει ποτέ Stop, τότε ο παίκτης i θα πετύχει το βέλτιστο συνολικό κέρδος εάν σταματήσει το παιχνίδι την περίοδο T.

Το σύνολο των μη τερματικών ιστοριών αποτελείται από ακολουθίες ζευγών (Cont,Cont) μήκους t όπου $0 \geq t < T$. Η στρατηγική ενός παίκτη καθορίζεται από την περίοδο στην οποία ο παίκτης επιλέγει να σταματήσει το παιχνίδι, εφόσον επιλέγει να το σταματήσει. Επομένως ένας παίκτης στο παίγνιο $G_0(T)$ έχει T στρατηγικές $S(1), ..., S(T)$, όπου η στρατηγική S(t) σημαίνει "σταμάτα την περίοδο t" και μία στρατηγική σύμφωνα με την οποία δεν επιλέγει ποτέ να σταματήσει το παιχνίδι και η οποία συμβολίζεται ως CONT.

Παρά τα ενδεχόμενα "τεράστια" κέρδη που οι παίκτες μπορεί να απεικονίσουν εάν συνεχίζουν το παίγνιο μέχρι το τέλος, τότε το παιχνίδι έχει μόνο μία ισορροπία Nash, κατά την οποία και οι δύο παίκτες σταματούν το παιχνίδι στην πρώτη περίοδο. Επίσης είναι εύκολο να επιβεβαιώσει κανείς ότι το παίγνιο δεν έχει κάποια άλλη μικτή ισορροπία.

Στο υπόλοιπο τμήμα του κεφαλαίου παρουσιάζουμε ορισμένες προσεγγίσεις που επιχειρούν να επιλύσουν το πρόβλημα του πεπερασμένου ορίζοντα. Παρέχουν κάποια ικανοποιητική λύση στην περίπτωση του παιγνίου μας $G_0(T)$.

5.2 Υλοποίηση των στρατηγικών με αυτόματα

Οι απαραίτητοι υπολογισμοί για την υλοποίηση της στρατηγικής $S(t)$ με $t > 1$ εμπεριέχει το μέτρημα μέχρι το t. Η υλοποίηση των στρατηγικών CONT και $S(1)$ δεν προϋποθέτουν κανένα υπολογισμό. Επομένως οι δυνατότητες να μοντελοποιήσουν την πολυπλοκότητα των στρατηγικών είναι αρκετά περιορισμένες. Το μόνο πιθανό θέμα πολυπλοκότητας αφορά την πολυπλοκότητα, σχετίζεται με την δυσκολία ενός παίκτη να αναγνωρίζει το χρόνο που πρέπει να σταματήσει. Αντίθετα με διάφορα ζητήματα πολυπλοκότητας που εξετάσαμε στο προηγούμενο κεφάλαιο, αυτά δεν εμφανίζονται στο παίγνιο $G_0(T)$. Ο παίκτης δεν μπορεί να παρακολουθήσει τον αντίπαλο, αφού εάν επιλέξει Stop, τότε το παίγνιο σταματά. Οι παίκτες δεν έχουν την δυνατότητα να φθάσουν σε ένα μακροπρόθεσμο συμβιβασμό, βασιζόμενοι στην επιλογή διαφορετικών κινήσεων καθώς υπάρχει μία μόνο κίνηση που δεν τερματίζει το παίγνιο. Με άλλα λόγια μια ιστορία περιγράφεται πλήρως από το μήκος της.

Όπως είδαμε στο προηγούμενο κεφάλαιο, περιγράφουμε την πολυπλοκότητα μιας στρατηγικής ως τον ελάχιστο δυνατό αριθμό καταστάσεων αυτομάτου που μπορεί να υλοποιήσει την στρατηγική. Χρησιμοποιούμε ένα πεπερασμένο αυτόματο, το οποίο δέχεται ως είσοδο ένα μήνυμα που σηματοδοτεί το πέρας μιας περιόδου και παράγει ως έξοδο μία ενέργεια (Stop ή Cont). Έχουμε δηλαδή ότι το αυτόματο είναι μια τετράδα (Q, q_0, f, τ) όπου Q είναι ένα πεπερασμένο σύνολο καταστάσεων, $q_0 \in Q$ είναι η αρχική κατάσταση, η συνάρτηση εξόδου $f : Q \to \{Stop, Cont\}$ καθορίζει μια κίνηση Stop ή Cont για κάθε κατάσταση και $\tau : Q \to Q$ είναι συνάρτηση μετάβασης.

Αξίζει να σημειώσουμε ότι αντίθετα με το προηγούμενο κεφάλαιο όπου η συνάρτηση μετάβασης εξαρτιόταν τόσο από την κατάσταση του αυτομάτου όσο και από τις αποκρίσεις του αντίπαλου παίκτη, εδώ εξαρτάται μόνο από την κατάσταση του αυτομάτου και αυτό ισχύει αφού το παίγνιο συνεχίζεται μόνο εφόσον η είσοδος που δέχεται είναι Cont. Επομένως η είσοδος της απόκρισης του αντίπαλου παίκτη μπορεί να παραληφθεί.

Φυσικά όλες οι στρατηγικές του $G_0(T)$ μπορούν να υλοποιηθούν από ένα αυτόματο. Υπάρχει ένα αυτόματο με $M(t)$ με t καταστάσεις (αλλά όχι λιγότερες) που εκτελεί την στρατηγική $S(t)$. Η στρατηγική CONT υλοποιείται από ένα αυτόματο $M(CONT)$ με μία κατάσταση.

Τα τρία ακόλουθα παίγνια αυτομάτων, τα οποία περιγράφουμε στις επόμενες 3 ενότητες είναι παραλλαγές του $G_0(T)$. Σε όλα αυτά τα παίγνια αυτομάτων, ο παίκτης παίρνει μία στρατηγική απόφαση πριν ξεκινήσει το παίγνιο, δηλαδή επιλέγει μία μηχανή που υλοποιεί την στρατηγική και λειτουργεί αυτόματα χωρίς επιπλέον παρέμβαση εκ μέρους του. Στο πρώτο παίγνιο, μεταβάλλουμε τις συναρτήσεις οφέλους έτσι ώστε να συσσωματώσουμε το κόστος της πολυπλοκότητας. Στο δεύτερο παίγνιο θέτουμε έναν περιορισμό στην πολυπλοκότητα των δυνατών αυτομάτων. Στο τρίτο μεταβάλλουμε την έννοια του αυτομάτου έτσι ώστε πέρα από το να παίζει το παίγνιο $G_0(T)$, μπορεί να στέλνει καθώς και να λαμβάνει μηνύματα.

5.3 Προσεγγίσεις για επίλυση των παραδόξων πεπερασμένου ορίζοντα

5.3.1 Παίγνιο με κόστος πολυπλοκότητας

Σε αυτήν την ενότητα, εφαρμόζουμε την προσέγγιση κατά την οποία η πολυπλοκότητα κοστίζει. Κάθε παίκτης προσπαθεί να αυξήσει το $G_0(T)$ όφελός του και να ελαττώσει την πολυπλοκότητα του αυτομάτου του. Έστω $c > 0$, στο παίγνιο αυτομάτων $G_1(T, c)$ κάθε παίκτης πρέπει να επιλέξει ένα αυτόματο και η συνάρτηση οφέλους του παίκτη είναι [κέρδος του παίκτη στο $G_0(T)$]-$c\times$[αριθμός καταστάσεων του αυτομάτου]

Πρόταση 5.1. *Εάν $c \geq 2/(T-1)$, τότε τα ζεύγη των αυτομάτων $(M(CONT), M(CONT))$ και $(M(1), M(1))$ είναι οι μόνες (γνήσιες) ισορροπίες του $G_1(T, c)$. Εάν $c < 2/(T-1)$, τότε το ζεύγος αυτομάτων $(M(1), M(1))$ είναι η μοναδική ισορροπία του $G_1(T, c)$.*

Απόδειξη Πρώτα παρατηρούμε ότι στην ισορροπία, εάν ένας από τους παίκτες χρησιμοποιεί το αυτόματο $M(1)$, τότε και ο άλλος χρησιμοποιεί το αυτόματο $M(1)$. Επομένως το ζεύγος $(M(1), M(1))$ είναι πάντα ισορροπία στο $G_1(T, c)$. Επίσης εάν σε ισορροπία ένας παίκτης χρησιμοποιεί $M(CONT)$, τότε ο άλλος παίκτης δεν χρησιμοποιεί κάποια $M(t)$ με $t < T$ επειδή ανταλλάσσοντας την με την $M(CONT)$ προκύπτει ότι αυξάνεται το κέρδος στο $G_0(T)$ και ελαττώνεται η πολυπλοκότητα του αυτομάτου.

Δεν υπάρχει ισορροπία στην οποία ένας από τους παίκτες, ας πούμε ο παίκτης 1, χρησιμοποιεί $M(t)$ με $1 < t < T$ επειδή η καλύτερη απάντηση του παίκτη 2 πρέπει να είναι $M(CONT), M(1)$ ή κάποιο $M(t')$ με $t' < t$. Έχουμε δει ότι $M(t), M(CONT)$ και $(M(t), M(1))$ $(t > 1)$ δεν είναι ισορροπία. Το ζεύγος $(M(t), M(t')$ με $1 < t' \leq t$ δεν είναι ισορροπία επειδή ο παίκτης 1 μπορεί να αλλάξει αυτόματο και να χρησιμοποιήσει το $M(t'-1)$ αποκομίζοντας κέρδος.

Το ζεύγος $(M(T), M(CONT))$ επίσης δεν είναι ισορροπία. Είτε

$c > 2/(T-1)$ είτε $2/(T-2) > c$. Εάν $c > 2/(T-1)$, ο παίκτης 1 θα ωφεληθεί περισσότερο εάν χρησιμοποιήσει το αυτόματο $M(CONT)$, αφού χάνει 2 από το κέρδος του στο $G_0(T)$ αλλά ταυτόχρονα "σώνει" $T-1$ καταστάσεις. Εάν $2/(T-2) > c$, τότε ο παίκτης θα ωφεληθεί περισσότερο εάν χρησιμοποιήσει το αυτόματο $M(T-1)$, αφού χρησιμοποιεί $T-2$ καταστάσεις περισσότερες αλλά κερδίζει 2 επιπλέον μονάδες οφέλους.

Τελικά το $(M(CONT), M(CONT))$ είναι ισορροπία μόνο εάν δεν συμφέρει κανένα παίκτη να χρησιμοποιήσει το M(T), δηλαδή εάν $c > 2/(T-1)$.
□

Στο προηγούμενο κεφάλαιο, αναθέτοντας κόστος στην πολυπλοκότητα ελαττώσαμε το σύνολο των εκβάσεων ισορροπίας. Εδώ σε αντίθεση, το σύνολο των ισορροπιών Nash του $G_1(T, c)$ δεν είναι απαραίτητα υποσύνολο των σημείων ισορροπίας του $G_0(T)$. Έτσι όταν το c είναι μεγάλο έχουμε μία νέα ισορροπία στην οποία το παίγνιο συνεχίζεται μέχρι τέλους.

5.3.2 Περιορισμένη Πολυπλοκότητα

Σε αυτήν την ενότητα, θα αλλάξουμε το παίγνιο $G_0(T)$ έτσι ώστε οι παίκτες να είναι περιορισμένοι όσον αναφορά την πολυπλοκότητα της στρατηγικής που μπορούν να εφαρμόσουν. έστω B ένας φυσικός αριθμός, τότε στο παίγνιο $G_2(T, B)$ κάθε παίκτης πρέπει να επιλέξει ένα αυτόματο για να παίξει το παίγνιο $G_0(T)$, όπου όμως το αυτόματο αυτό δεν πρέπει να έχει πάνω από B καταστάσεις. Τα κέρδη είναι τα ίδια όπως και στο $G_0(T)$. Επίσης αντίθετα από το $G_1(T, c)$, η χρήση των B καταστάσεων γίνεται χωρίς επιπλέον κόστος, αλλά τα αυτόματα δεν πρέπει να έχουν πάνω από B καταστάσεις.

Η ανάλυση του $G_2(T, B)$ είναι στοιχειώδης. Εάν $B < T$,τότε οι παίκτες δεν είναι σε θέση να μετρήσουν μέχρι το T, οπότε το επιθυμητό αποτέλεσμα συνεργασίας μπορεί να προκύψει στην ισορροπία. Αντίθετα εάν είναι σε θέση να μετρήσουν μέχρι το T, (δηλαδή εάν $B \geq T$), τότε η ύπαρξη ισορροπίας του περιορισμού πάνω στον αριθμό καταστάσεων δεν οδηγεί σε νέες ισορροπίες.

Πρόταση 5.2. *Στο $G_2(T, B)$:*

- *Εάν $B \geq T$, τότε η μοναδική ισορροπία Nash είναι η $(M(1), M(1))$.*
- *Εάν $B < T$, τότε οι μοναδικές γνήσιες ισορροπίες Nash είναι οι $(M(CONT), M(CONT))$ και $(M(1), M(1))$.*

Απόδειξη Εάν $B \geq T$, τότε το παίγνιο αυτομάτων $G_2(T, B)$ είναι ακριβώς ίδιο με το $G_0(T)$, αφού όλες οι στρατηγικές του $G_0(T)$ μπορούν να υλοποιηθούν από αυτόματα με το πολύ T καταστάσεις.
Ένα αυτόματο το οποίο, όταν παίζει εναντίον του $M(CONT)$, αποφέρει κέρδος στο $G_0(T)$ μεγαλύτερο από ότι επιτυγχάνει το $M(CONT)$πρέπει να έχει τουλάχιστον T καταστάσεις, οι οποίες δεν είναι διαθέσιμες όταν $B < T$.
$\square$

5.3.3 Τα αυτόματα στέλνουν μηνύματα

Σε αντίθεση με το παίγνιο $G_2(T, B)$, το παίγνιο που θα αναπτύξουμε στην παράγραφο αυτή επιτυγχάνει μια επιθυμητή ισορροπία ανεξαρτήτως του ορίου που θα θέσουμε πάνω στο πλήθος των καταστάσεών της. Παρόλα αυτά προκειμένου να επιτύχουμε το αποτέλεσμα αυτό θα πρέπει να μεταβάλλουμε λίγο τον ορισμό του αυτομάτου. Έτσι λοιπόν στην ενότητα αυτή κάθε αυτόματο πέρα από την συμμετοχή του στο παίγνιο $G_0(T)$, ανταλλάσσει μηνύματα. Στην επιθυμητή ισορροπία η ανάγκη της αποστολής ορθών μηνυμάτων απορροφά διαθέσιμους πόρους του συστήματος (καταστάσεις) από το να χρησιμοποιούνται για την μέτρηση μέχρι το T και την αναγνώριση της τερματικής περιόδου, οπότε έτσι αποτρέπεται κάθε επικερδής αποστασία. Αυτή είναι μια εύλογη στρατηγική ιδέα, καθώς παραπέμπει σε συνήθεις στρατηγικούς χειρισμούς που συναντάμε στην καθημερινή μας ζωή, κατά τους οποίους ένας παίκτης προσπαθεί να εκτρέψει την προσοχή του αντιπάλου του από το να πραγματοποιήσει κάποια ανεπιθύμητη κίνηση.

Έστω L το σύνολο των συμβόλων μηνυμάτων, τότε ένα αυτόματο μηνυμάτων αποτελείται από τα:

- ένα σύνολο καταστάσεων Q
- μία αρχική κατάσταση q_0,
- μία συνάρτηση εξόδου $f : Q \to \{Stop\} \bigcup \{(Cont, m) : m \in L\}$ και
- μία συνάρτηση μετάβασης $\tau : Q \times L \to Q$.

Η σημασία του $f(q) = (Cont, m)$ είναι ότι στην κατάσταση q, το αυτόματο επιλέγει $Cont$ και στέλνει μήνυμα m, ενώ το $f(q) = Stop$ σημαίνει ότι στην κατάσταση q, το αυτόματο επιλέγει $Stop$. Όμοια το $\tau(q, m) = q'$ σημαίνει ότι εφόσον το αυτόματο βρίσκεται στην κατάσταση q και δεχθεί ένα μήνυμα m, τότε μεταβαίνει στην κατάσταση q'. Στο παίγνιο $G_3(T, B, L)$ κάθε παίκτης πρέπει να επιλέξει ένα αυτόματο μηνυμάτων χρησιμοποιώντας το σύνολο των L μηνυμάτων, όπου κάθε αυτόματο-μηνυμάτων έχει το πολύ B καταστάσεις. Τα κέρδη είναι τα ίδια όπως ορίστηκαν για το $G_0(T)$.

Πρόταση 5.3. *Για κάθε $T \geq 2$ και για όλα τα B (ακόμη και για $B > T$), υπάρχει ένα σύνολο μηνυμάτων L, τέτοιο ώστε το παίγνιο $G_3(T, B, L)$ έχει μία μεικτή στρατηγική κατά την οποία κανένας παίκτης δεν σταματά το παίγνιο.*

Απόδειξη Έστω $L = L_1 \bigcup L_2$ όπου L_i είναι ένα σύνολο με $B - 2$ στοιχεία και όπου τα L_1 και L_2 είναι ξένα μεταξύ τους. Έστω s_i η μεικτή στρατηγική του i-οστού παίκτη, κατά την οποία επιλέγει με πιθανότητα $1/(B-2)$, ένα αυτόματο B καταστάσεων, το οποίο αρχικά στέλνει ένα μήνυμα $m^* \in L_i$ και στην συνέχεια επαναλαμβάνει το μήνυμα που του στέλνει ο παίκτης j. Εάν το αυτόματο θεωρήσει ότι το άλλο αυτόματο δεν απαντά με το ίδιο μήνυμα m^*, τότε τερματίζει το παίγνιο. Επομένως το σύνολο όλων των καταστάσεων είναι το $L_j \bigcup \{M^*, S\}$. Έτσι λοιπόν ακολούθως περιγράφουμε το αυτόματο αυτό:

Κατάσταση q	**Έξοδος f(q)**	**Μετάβαση τ(q,n)**
m^*(*αρχική*)	$(Cont, m^*)$	n
$n^* \in L_j$	$(Cont, n^*)$	$\begin{cases} n^* & \text{εάν } n = m^* \\ S & \text{διαφορετικά} \end{cases}$
S	*Stop*	

Δεν είναι δύσκολο να επιβεβαιώσει κανείς ότι το ζεύγος (s_1, s_2) όπως ορίστηκε παραπάνω αποτελεί μικτό σημείο ισορροπίας. Πράγματι προκειμένου ο παίκτης i να αυξήσει το όφελός του στο $G_0(T)$, στην περίπτωση που ο παίκτης j στείλει ένα μήνυμα m, πρέπει να αφιερώσει $T-1$ καταστάσεις στην πιθανότητα ότι το μήνυμα που θα στείλει ο j είναι το m. Έχουμε δηλαδή ότι προκειμένου να εξασφαλίσει ένα μικρό όφελος 2 με πιθανότητα $1/(B-2)$ πρέπει να υποστεί μία μεγάλη απώλεια εξαιτίας της αδυναμίας του να απατήσει ορθά στα υπόλοιπα $T-1$ μηνύματα τα οποία θα σταλούν με πιθανότητα $(T-1)/(B-2)$. □

5.3.4 Η ε-ισορροπία: Όταν η απόκλιση κοστίζει

Εδώ θα παρουσιάσουμε μία διαφορετική προσέγγιση, η οποία είναι η πρώτη που επιχειρεί να επιλύσει τα παράδοξα αυτά στηριζόμενη στην βάση της πεπερασμένης "λογικότητας". Σύμφωνα λοιπόν με αυτήν την προσέγγιση, ένας παίκτης ο οποίος έχει συνηθίσει να λειτουργεί με κάποιο συγκεκριμένο τρόπο, σύμφωνα με κάποια πρότυπα συμπεριφοράς (συμπεριφορά ισορροπίας) χρειάζεται έναν πολύ σημαντικό λόγο για να πραγματοποιήσει μία αλλαγή. Έτσι λοιπόν προκειμένου ο παίκτης να επιλέξει μια διαφορετική στρατηγική, θεωρούμε ότι δεν επαρκεί η πιθανή αποκόμιση κέρδος από την κίνησή του αυτή, αλλά επίσης πρέπει το κέρδος αυτό να υπερβαίνει το κόστος της αλλαγής. Αυτή η θεώρηση αντικατοπτρίζει ορισμένους πραγματικούς περιορισμούς που συνδέονται με την αλλαγή της στρατηγικής όπως το κόστος της επανεκπαίδευσης των ατόμων που πρέπει να την εκτελέσουν, ο χρόνος που μεσολαβεί για την προσαρμογή του παίκτη, το συναισθηματικό κόστος κ.α. Σύμφωνα με την προσέγγιση αυτή, οι παίκτες είναι σε θέση να υπολογίσουν το κέρδος που αντιστοιχεί σε μία αλλαγή και να το συγκρίνουν με το κόστος της αλλαγής $\varepsilon > 0$.

Σύμφωνα λοιπόν με τα παραπάνω, ο ορισμός της ε-ισορροπίας είναι ένα ζεύγος στρατηγικών (s_1, s_2), τέτοια ώστε για κάθε παίκτη i, η βέλτιστη απάντηση στην s_j δεν αποφέρει κέρδος που να είναι υψηλότερο κατά μία ποσότητα μεγαλύτερη του ε, από το κέρδος που αποφέρει η στρατηγική s_i. Προφανώς κάθε ισορροπία Nash είναι ε-ισορροπία, ανεξάρτητα από την τιμή του ε.

Επιστρέφοντας τώρα στο $G_0(T)$, ένα "μικρό" ε (μικρό σε σχέση με το μέγιστο κέρδος που μπορεί να συσσωρευτεί από το παίγνιο) επαρκεί για να διατηρήσει συμπεριφορές που δεν αντιστοιχούν σε ισορροπία. Έτσι ανεξάρτητα από το μήκος T του παιγνίου, κανείς δεν μπορεί να κερδίσει παραπάνω από 2 για οποιαδήποτε παρέκκλιση από το (CONT,CONT). Επομένως για κάθε $\varepsilon \geq 2$ και για όλα τα T, το ζεύγος (CONT,CONT) (όπως και αρκετά ζεύγη στρατηγικών) είναι μία ε-ισορροπία του $G_0(T)$.

Πάντως η συγκεκριμένη προσέγγιση δεν παρέχει μία τόσο πειστική εξήγηση για την παράδοξη παρατηρούμενη συμπεριφορά. Πραγματικά, μπορούμε να δούμε ότι το ίδιο ε εξασφαλίζει ε-ισορροπία $(CONT, CONT)$ ανεξαρτήτου του T. Το παίγνιο $G_0(1)$ είναι το "Δίλημμα των Φυλακισμένων", στο οποίο η κίνηση Stop είναι η επικρατούσα στρατηγική. Παρόλα αυτά το (CONT,CONT) είναι ε-ισορροπία (για $\varepsilon \geq 2$), όπως για οποιοδήποτε άλλο παίγνιο $G_0(T)$.

5.4 Το δίλημμα των φυλακισμένων

Στην συνέχεια θα παρουσιάσουμε μία επεξήγηση για το πεπερασμένο πρόβλημα των φυλακισμένων. Θα ακολουθήσουμε την μέθοδο που χρησιμοποίησαν οι Παπαδημητρίου και Γιαννακάκης [14] και [15] για να αποδείξουμε ότι όταν το επαναλαμβανόμενο παίγνιο του Διλήμματος των Φυλακισμένων παίζεται από πεπερασμένα αυτόματα με λιγότερο από εκθετικές (σε σχέση με τον αριθμό των γύρων) καταστάσεις, τότε μπορούμε να επιτύχουμε ως σημείο ισορροπίας την συνεργασία ανάμεσα στους παίκτες.

Πριν υπεισέρθουμε στην αυστηρή και μεθοδική ανάλυση του παιγνίου θα εκθέσουμε τα κύρια σημεία της απόδειξης. Καταρχάς η βασική ιδέα της απόδειξης όπως συνομολογείται και από τους Παπαδημητρίου και Γιαννακάκης, οφείλεται σε μία παρατήρηση του Neyman [10] πάνω στο ίδιο πρόβλημα (ο οποίος αργότερα το έλυσε και ο ίδιος και μάλιστα πρόσφερε και μια ενδιαφέρουσα γενίκευσή του [11]). Σύμφωνα λοιπόν με την παρατήρηση του Neyman, το πλήθος των ιστοριών του παιγνίου είναι εκθετική συνάρτηση του μήκους του παιγνίου. Πραγματικά σε κάθε περίοδο του παιγνίου αντιστοιχούν 4 δυνατά αποτελέσματα (A,A),(A,O),(O,A),(O,O) άρα σε για μία ιστορία μήκους t έχουμε 4^t δυνατές εξελίξεις. Οπότε τώρα είναι δυνατό να "γεμίσει" κανείς όλες τις θέσεις μνήμης (καταστάσεις) των αυτόματων απαιτώντας και οι δύο παίκτες να θυμούνται μικρές ιστορίες π.χ. απαντώντας με προσυμφωνημένους τρόπους και στην συνέχεια παίζουν συνεργαζόμενοι κατά την διάρκεια των εναπομεινάντων γύρων. Έτσι καμιά κατάσταση των αυτομάτων δεν θα παραμένει αχρησιμοποίητη προκειμένου να μετρηθούν οι γύροι, οπότε κατά την μεγαλύτερη διάρκεια του παιγνίου οι παίκτες συνεργάζονται.

Βασιζόμενοι πάνω στην παρατήρηση αυτή, προχωρούμε στην σκιαγράφηση των σημαντικότερων δυσκολιών του θεωρήματος αυτού. Έτσι λοιπόν η ισορροπία στην οποία καταλήγει το παίγνιο είναι μικτή ισορροπία, κατά την οποία τα αυτόματα που συμμετέχουν σε αυτή ανταλλάζουν μικρές ακολουθίες από (Α,Ο) (που από εδώ και πέρα θα ονομάζουμε ταυτότητες). Κάθε αυτόματο που συμμετέχει στην μεικτή στρατηγική έχει την δική του ταυτότητα. Μετά το πρώτο στάδιο, (και αντίθετα από την εισήγηση του Neyman) οι δύο παίκτες ανταλλάσσουν περιοδικά το αποκλειστικό ή των ταυτοτήτων τους καθόλη την διάρκεια των n γύρων, με μακριά ενδιάμεσα διαστήματα συνεργασίας (εδώ με αποκλειστικό ή εννοούμε ότι εννοούμε μια ακολουθία ενεργειών που η i-οστή θα είναι Α εάν οι i-οστές ενέργειες των δύο ακολουθιών ήταν ίσες ή Ο εάν ήταν διαφορετικές). Αυτό εξασφαλίζει ότι θυμούνται και οι δύο τις κάρτες των αντιπάλων τους και επομένως δεν έχουν ελεύθερη μνήμη για να σχεδιάσουν κάποια αποστασία.

Η απόδειξη παρουσιάζει δύο περαιτέρω δυσκολίες, τις οποίες θα αναφέρουμε για να αιτιολογήσουμε την κατασκευή που θα παρουσιάσουμε κατά την αναλυτική απόδειξη του θεωρήματος. Οι ταυτότητες που ανταλλάσσουν οι παίκτες κατά την εκκίνηση του παιγνίου είναι ακολουθίες της μορφής $\{A,O\}^d$ για κάποιο μικρό d. Όμως η ανταλλαγή των ταυτοτήτων συνεπάγεται ότι οι παίκτες θα παίξουν το "δίλημμα των φυλακισμένων" για γύρους. Επομένως οι παίκτες που έχουν ταυτότητες οι οποίες έχουν πολλά Ο, είναι σε πλεονεκτική θέση σε σχέση με αυτούς που έχουν πολλά Α. Για να υπερπηδηθεί αυτό το πρόβλημα χρησιμοποιούνται δύο τεχνάσματα: κατά πρώτο λόγο λοιπόν η ανταλλαγή των ταυτοτήτων ακολουθείται από μία φάση της οποίας ο σκοπός είναι να εξαλείψει τον τυχόν προβάδισμα που έχει αποκτήσει κάποιος παίκτης λόγο της ανταλλαγής ταυτοτήτων. Έτσι όμως το πρόβλημα παραμένει κατά τις περιοδικές επαναλήψεις των ανταλλαγών των ταυτοτήτων ανάμεσα στους παίκτες. Αυτή η διαφορά εξαλείφεται, καθώς επιβάλλουμε στους παίκτες να εκπέμπουν το αποκλειστικό ή των ταυτοτήτων τους.

Ένα άλλο πρόβλημα που συναντάται κατά την απόδειξη είναι και το ακόλουθο: οι περισσότερες καταστάσεις των αυτομάτων που χρησιμοποιούνται έχουν μία αναμενόμενη μετάβαση (η οποία αντιστοιχεί στην συνεργασία μεταξύ των αντιπάλων) και μία μετάβαση προς μία κατάσταση "τιμωρίας" στην οποία το αυτόματο αποστατεί από την συνεργασία επ'αόριστον. Εάν όμως κάποιος αποσκοπεί να ξεγελάσει τις μηχανές αυτές δεν χρειάζεται αυτές τις μεταβάσεις τιμωρίας). Πραγματικά αρκεί να χρησιμοποιεί τις αναμενόμενες μεταβάσεις για να ξεγελάσει τον αντίπαλο ότι σκοπεύει να συνεργαστεί μαζί του. Αλλά αυτό σημαίνει ότι η συντριπτική πλειοψηφία των καταστάσεων (όλες εκτός από τις αρχικές κατά τις οποίες πραγματοποιείται η ανταλλαγή των ταυτοτήτων) είναι χρησιμοποιημένες κατά το ήμισυ. Έχουμε δηλαδή ότι η μία από τις δύο μεταβάσεις είναι διαθέσιμο για επαναχρησιμοποίηση, έτσι με τον τρόπο αυτό διασώζουμε περίπου τις μισές καταστάσεις. Ας υποθέσουμε για παράδειγμα ότι το αυτόματο του παίκτη I έχει μία κατάσταση q, στην οποία ο ειλικρινής παίκτης I παίζει A και περιμένει ο παίκτης II να παίξει A και έστω ότι υπάρχει μία κατάσταση p στην οποία ο παίκτης I παίζει A και περιμένει O. Έχουμε στην περίπτωση αυτή ότι ο παίκτης που επιθυμεί να κοροϊδεύσει τον αντίπαλό του μπορεί να χρησιμοποιήσει μόνο μία κατάσταση, ας πούμε την q στην θέση των δύο προηγούμενων. Έτσι μπορεί να θέσει το $\tau(q, O)$ ίσο με το $\tau(p, O)$ και να επανακατευθύνει όλες τις μεταβάσεις που πήγαιναν στο p να πάνε στο q και να αφαιρέσει την κατάσταση p. Ξεκάθαρα, όσο ο παίκτης II παίζει τίμια (όπως αναμένεται), δεν θα παρατηρήσει καμία διαφορά ανάμεσα στο "μεταποιημένο" και το κανονικό αυτόματο. Εφαρμόζοντας την στρατηγική αυτή ο παίκτης I μπορεί να σώσει πολλές καταστάσεις, τις οποίες μπορεί να χρησιμοποιήσει για να μετρήσει και να αποστατήσει στο τέλος. Παρόλα αυτά η κατασκευή μας εξασφαλίζει ότι οι αναμενόμενες μεταβάσεις του παίκτη II πάντα συμπίπτουν με την στρατηγική που επέλεξε ο παίκτης I και αντίστροφα. Με άλλα λόγια μετά την αρχική περίοδο ανταλλαγής, οι δύο παίκτες εκτελούν την ίδια ακολουθία στρατηγικών, γεγονός που δεν επιτρέπει την επαναχρησιμοποίηση καταστάσεων.

Μετά από αυτήν εισαγωγική συζήτηση είμαστε σε θέση να προχωρήσουμε στην αυστηρή διατύπωση και απόδειξη του θεωρήματος.

Θεώρημα 5.4. *Για κάθε $\varepsilon > 0$, στο n-γύρων επαναλαμβανόμενο παίγνιο αυτομάτων του "Διλήμματος των Φυλακισμένων", με αυτόματα που τα μεγέθη τους είναι φραγμένα από τα $s_I(n), S_I I(n)$, εάν τουλάχιστον το ένα από αυτά είναι μικρότερο από $2^{c_\varepsilon n}$,τότε $c_\varepsilon = \varepsilon/(12(1+\varepsilon))$, τότε για αρκετά μεγάλο n υπάρχει μία (μικτή) ισορροπία με αναμενόμενο όφελος για κάθε παίκτη τουλάχιστον $3 - \varepsilon$.*

Απόδειξη Θα περιγράψουμε την κατασκευή στην περίπτωση που και τα δύο όρια είναι μικρότερα ή ίσα του $2^{c_\varepsilon n}$ και μάλιστα ίσα με $S = 2^d(d + l + 2)$ για κάποιους ακέραίους d και l, τέτοιους ώστε να ισχύουν οι παρακάτω ανισότητες:

$$3l + sd < n \text{ και } (2 + d)/l + 4d/n < \varepsilon$$

Στην συνέχεια θα γενικεύσουμε την βασική κατασκευή σε τυχαία όρια, το ένα από τα οποία είναι τουλάχιστον $2^{c_\varepsilon n}$, όπως αναφέρεται στο θεώρημα.

Η ισορροπία αποτελείται από μία μεικτή στρατηγική για τον παίκτη I, που περιγράφεται στην συνέχεια και μία μικτή στρατηγική για τον παίκτη II. Η μεικτή στρατηγική για τον I αποτελείται από 2^d αυτόματα, σε κάθε ένα αντιστοιχεί πιθανότητα 2^{-d} και κάθε ένα έχει S καταστάσεις. Τα αυτόματα ταυτοποιούνται από ακολουθίες $\{A, O\}^d$. Τα αυτόματα του παίκτη I συμβολίζονται A_x, ενώ του παίκτη II συμβολίζονται B_y, όπου τα $x, y \in \{A, O\}^d$ είναι οι ταυτότητες των δύο αυτομάτων. Η αρχική κατάσταση του A_x είναι η ρίζα ενός

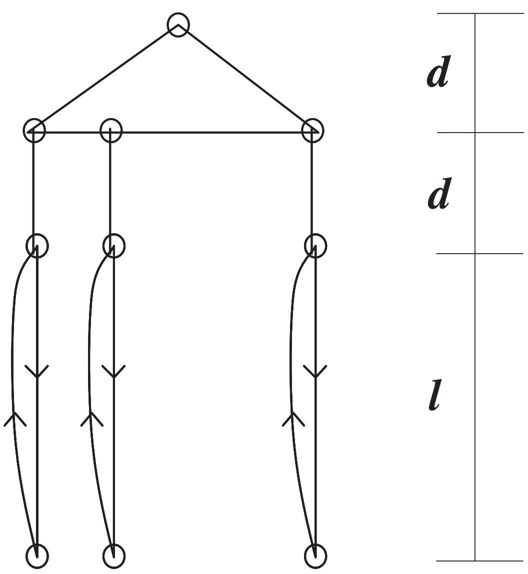

Figure 5.1: Το αυτόματο A_x

πλήρους δυαδικού δένδρου βάθους d, του οποίου οι καταστάσεις στο i-οστό επίπεδο είναι (όπου το πρώτο επίπεδο συνιστά την ρίζα) παίζει την στρατηγική του $\{A, O\}$ που αποτελούν το i-οστό γράμμα του x. Τα 2^d φύλλα του δένδρου σηματοδοτούνται από τις στρατηγικές κατά τις μεταβάσεις που οδήγησαν σε αυτό το φύλλο: Η ακολουθία y που αντιστοιχεί στο φύλλο y είναι η ταυτότητα του αντιπάλου που οδηγεί σε αυτήν την κατάσταση, δηλαδή του B_y.

Όταν οι δύο παίκτες A_x και B_y έχουν φθάσει σε ένα φύλλο του δυαδικού δένδρου, τότε λέμε ότι *έχουν ανταλλάξει ταυτότητες*. Αυτό περιλαμβάνει $2 \times 2^d - 1$ καταστάσεις. Υπάρχει ακόμη μία κατάσταση τιμωρίας, στην οποία αντιστοιχεί ενέργεια O και όλες της οι μεταβάσεις καταλήγουν σε αυτήν. Ο κύριος όγκος του χώρου καταστάσεων αποτελείται από διαχωρισμένα μεταξύ τους σύνολα $d+1+l$ καταστάσεων, τα οποία κρέμονται από τα 2^d φύλλα. Αυτές οι καταστάσεις είναι διευθετημένες ως μονοπάτια μήκους $d+1$ (η *δεύτερη φάση ή φάση διόρθωσης*), η οποία ακολουθείται από ένα βρόχο μήκους l (η *τρίτη φάση ή κύρια φάση* κατά την οποία πραγματοποιείται *η συνεργασία και το τσεκάρισμα*). Μπορούμε να παρατηρήσουμε ότι ο συνολικός αριθμός καταστάσεων είναι πράγματι S.

Από όλες αυτές τις $2^d(d+1+l)$ καταστάσεις, μία από τις δύο μεταβάσεις (την οποία ονομάζουμε *μετάβαση τιμωρίας*) η οποία μεταβαίνει στην κατάσταση τιμωρίας και η άλλη η οποία μεταβαίνει στην επόμενη κατάσταση μέσα στο μονοπάτι ή στο βρόχο και η οποία ονομάζεται αναμενόμενη μετάβαση (μόνο οι αναμενόμενες μεταβάσεις φαίνονται στο σχήμα).

Ο σκοπός των πρώτων d βημάτων είναι της φάσης διόρθωσης είναι να διορθώσουν τις διαφορές ανάμεσα στα διάφορα αυτόματα όσον αναφορά τις ταυτότητές τους. Στο ι-οστό βήμα της φάσης διόρθωσης , για $i = 1, ..., d$ και οι δύο παίκτες παίζουν A εάν και μόνο εάν το i-οστό γράμμα των ταυτοτήτων τους και για τους δύο, διαφορετικά (αν τουλάχιστον ο ένας από τους δύο έχει O στο i-οστό γράμμα της ταυτότητάς του) και οι δύο παίζουν O. Στο $d+1$-οστό βήμα της φάσης διόρθωσης και οι δύο παίκτες παίζουν O. Ο σκοπός του βήματος αυτού είναι να διαχωρίσει τις καταστάσεις της φάσης διόρθωσης από την κύρια φάση κατά την οποία και οι δύο παίκτες παίζουν A στο τελευταίο βήμα.

Μετά την φάση διόρθωσης τα αυτόματα εισέρχονται στην κύρια φάση, ένα βρόχο μήκους l. Αφού $l << n$, αυτός ο βρόχος θα επαναληφθεί πολλές φορές. Ο βρόχος αποτελείται από δύο τμήματα: ένα τμήμα συνεργασίας όπου οι παίκτες ακολουθούν την αμοιβαίως επιθυμητή συμπεριφορά (και το οποίο συνιστά το κύριο μέρος του βρόχου) και ένα τμήμα ελέγχου του οποίου ο σκοπός είναι να ελέγχει την τιμιότητα του άλλου παίκτη (συγκεκριμένα του παίκτη B_y, αφού συζητάμε για το τμήμα του A_x, το οποίο κρέμεται από το φύλλο που αντιστοιχεί στην ταυτότητα y) αλλά και να επιστρέψει στον άλλο παίκτη να ελένξει την τιμιότητα του A_x. Κατά την διάρκεια του κυρίου βρόχου τα δύο αυτόματα συνεργάζονται (παίζοντας A) συνέχεια, εκτός από μία ακολουθία $d+2$ βημάτων τα οποία αποτελούν το τμήμα ελέγχου. Αυτά τα $d+2$ βήματα είναι τοποθετημένα σε ένα βρόχο με τέτοιο τρόπο που n-οστός γύρος δεν πραγματοποιείται κατά την διάρκεια του τμήματος ελέγχου. Επομένως υποθέτοντας ότι $n - 2d \ mod \ l < l - d - 2$, μπορούμε να τοποθετήσουμε το σημείο ελέγχου στο τέλος του βρόχου, διαφορετικά αυτά τα βήματα μετακινούνται στα πρώτα $d+2$ βήματα του βρόχου (σημειώστε ότι d«l). Κατά την διάρκεια αυτών των $d+2$ βημάτων τα οποία αποτελούν το τμήμα ελέγχου, τα A_x και B_y επαναλαμβάνουν και τα δύο την ακολουθία των πράξεων $O(x \oplus y)A$, (δηλαδή πρώτα D μετά το αποκλειστικό ή των x και y και μετά A), όπου ορίζουμε $O \oplus O = A$, $A \oplus A = A$, $A \oplus O = O$ και $O \oplus A = O$. Αυτό έχει όπως θα δείξουμε τα επόμενα αποτελέσματα:

- Ο A_x εξασφαλίζει ότι ο αντίπαλός του θυμάται το x, δηλαδή την ταυτότητα του A_x.
- Όμοια, ο A_x πείθει τον B_y ότι πραγματικά θυμάται το y.
- Η αστάθεια που προέρχεται από τον αριθμό των O στο x εξαφανίζεται όταν συμψηφίζεται πάνω σε όλα τα B_y
- Μετά την περίοδο ανταλλαγής, και τα δύο αυτόματα εκπέμπουν ακριβώς την ίδια ακολουθία, δηλαδή σε κάθε κατάσταση στα τμήματα του

αυτομάτου που κρέμονται από τα φύλλα του δυαδικού δέντρου, η αναμενόμενη μετάβαση ταυτοποιείται από την ενέργεια της κατάστασης. (είναι σημαντικό για την απόδειξη του ελαχίστου των καταστάσεων).

Το αναμενόμενο κέρδος των δύο στρατηγικών είναι εύκολο να υπολογιστεί. Η τιμή του είναι $3n - \frac{5}{2}d - 2 - (d+2) \cdot \lfloor \frac{n-2d}{l} \rfloor$. Ο όρος $3n$ σημαίνει ότι τα αυτόματα αποκομίζουν ένα όφελος μεγέθους 3 κατά τα περισσότερα βήματα (με μερικές εξαιρέσεις που προκαλούνται από το πρωτόκολλο και αντιστοιχούν από τους υπόλοιπους όρους). Ο όρος $\frac{5}{2}d$ σημαίνει ότι κατά τα $2d$ πρώτα βήματα, το αυτόματο θα λάβει κατά μέσο όρο όφελος $\frac{7}{4}$ (κατά μέσο όρο εναντίον όλων των αντιπάλων B_y και κατά την διάρκεια όλων των βημάτων που αντιστοιχούν στην παρουσίαση των ταυτοτήτων και την φάση της διόρθωσης). Αυτό συμβαίνει επειδή, για κάθε $i \leq d$ εάν το i-οστό γράμμα του x είναι ένα A, τότε το κατά μέσο όρο όφελος κατά την φάση ανταλλαγής ταυτοτήτων είναι $\frac{3+0}{2} = \frac{3}{2}$, ενώ στην φάση διόρθωσης είναι $\frac{3+1}{2} = 2$. Όμοια, εάν το i-οστό γράμμα του x είναι ένα O, τότε το κατά μέσο όρο όφελος κατά την φάση ανταλλαγής ταυτοτήτων είναι $\frac{2+0}{1} = \frac{5}{2}$, ενώ στην φάση διόρθωσης είναι πάντα 1. Έτσι λοιπόν και στις δύο περιπτώσεις, ο μέσος όρος είναι $\frac{7}{4}$. Ο όρος -2 οφείλεται στο $d+1$-οστό βήμα της φάσης διόρθωσης. Τελικά, για κάθε μία από τις εκτελέσεις του βρόχου χάνουμε ένα κέρδος μεγέθους 2 στο O βήμα που σηματοδοτεί την εκκίνηση του τμήματος ελέγχου και μεγέθους 1 κατά μέσο όρο σε κάθε βήμα κατά το τμήμα του αποκλειστικού ή. (Στην περίπτωση που ισχύει $n - 2d \ mod \ l \geq l - d - 1$, τότε το $\lfloor \frac{n-2d}{l} \rfloor$ γίνεται $\lceil \frac{n-2d}{l} \rceil$.) Επίσης αξίζει να τονίσουμε ότι το κέρδος είναι ανεξάρτητο από το x, αυτό είναι πολύ σημαντικό για να εξασφαλίσουμε ότι το ζεύγος των μεικτών στρατηγικών που περιγράφεται από μία ισορροπία. Ακόμη αφού $\frac{2+d}{l} + \frac{4d}{n} < \varepsilon$ αυτό το κέρδος είναι μέσα στα όρια του εn από το ιδανικό όφελος συνεργασίας $3n$.

Παραμένει να αποδείξουμε ότι το ζεύγος των μεικτών στρατηγικών είναι μία ισορροπία, Ας υποθέσουμε ότι κάποιο αυτόματο-απατεώνας που μπορούμε να συμβολίσουμε A', το οποίο παίζει ενάντια στην παραπάνω μεικτή στρατηγική του παίκτη ΙΙ. Θα δείξουμε ότι το A' δεν μπορεί να κερδίσει μεγαλύτερο όφελος εφόσον δεν έχει πάνω από S καταστάσεις. Μάλιστα ακόμη και με $n - 2d - 3l - 3$ έξτρα καταστάσεις δεν μπορεί να κερδίσει μεγαλύτερο όφελος.

Ας υποθέσουμε ότι το A' δεν "αναγκάζει" κανένα από τα αυτόματα του αντιπάλου B_y να πραγματοποιήσουν μία μετάβαση στην κατάσταση τιμωρίας κατά την διάρκεια των πρώτων $2d + 1 + 2l$ βημάτων. Έστω x_y η ταυτότητα που το A' δίνει στο B_y (δηλαδή τα πρώτα d παίγνια-βήματα μεταξύ των A' και B_y).

Λήμμα 5.5. *Εάν $y \neq y'$, τότε $y \oplus x_y \neq y' \oplus x_{y'}$.*

Απόδειξη Έστω $i \geq 1$ να είναι το πρώτο γράμμα κατά το οποίο διαφέρουν τα y και y', τότε τα x_y και $x_{y'}$ συμφωνούν στα πρώτα $i-1$ γράμματα. Αφού όμως το A' αποφασίζει το i-οστό γράμμα του x_y βασιζόμενο μόνο στα πρώτα $i-1$ γράμματα του y, πρέπει να έχουμε ότι τα i-οστά γράμματα των x_y και $x_{y'}$ είναι τα ίδια. Επομένως, τα $y \oplus x_y$ και $y' \oplus x_{y'}$ διαφέρουν στο i-οστό γράμμα. $\square$

Έστω S_y το σύνολο των καταστάσεων που επισκέπτονται από το A' στους γύρους από το $d+1$ μέχρι $2d+2l$ όταν παίζει εναντίον του B_y.

Λήμμα 5.6. *Υπάρχουν τουλάχιστον $d+1+l$ διαφορετικές καταστάσεις στο S_y. Επιπλέον, εάν $y \neq y'$, τότε τα S_y και $S_{y'}$ είναι ξεχωριστά.*

Απόδειξη Ας θεωρήσουμε οποιαδήποτε κατάσταση $s \in S_y$, και έστω i το βήμα κατά το οποίο επισκέφθηκε η κατάσταση s. Ας ορίσουμε το *κανονικό μονοπάτι* από την s κατά το βήμα i να είναι το μονοπάτι μήκους $2d+2l-i$ που ξεκινά από το s και πάντα ακολουθεί την μετάβαση που αντιστοιχεί στην ενέργεια που είχε παρθεί στην συγκεκριμένη κατάσταση. Θα χρησιμοποιούμε τον όρο κανονικό μονοπάτι για να αναφερόμαστε συνήθως στην ακολουθία καταστάσεων κατά μήκος του μονοπατιού. Από τον ορισμό αυτό προκύπτει ότι εάν η ίδια κατάσταση επισκέπτεται πάνω από μία φορά, τότε τα διάφορα κανονικά μονοπάτια της λειτουργούν ως πρόθεμα για κάποιο άλλο.

Αφού υποθέτουμε ότι το A' δεν τιμωρείται από το B_y κατά την διάρκεια των πρώτων $2d+1+2l$ βημάτων, το κανονικό μονοπάτι που ξεκινά από την κατάσταση που επισκέπτεται κατά το βήμα $d+1$ από το αυτόματο A' πρέπει να είναι ακριβώς το ίδιο, όπως το μονοπάτι των αναμενόμενων μεταβάσεων που ξεκινά από το x_y φύλο του B_y. Αυτό το τελευταίο μονοπάτι αποτελείται από τις $d+1$ μεταβάσεις της φάσης διόρθωσης, ακολουθούμενο από $[C^{l-d-2}D_y \oplus x_y C]^2$. (Υποθέτουμε ότι το τμήμα $D_y \oplus x_y C$ συμβαίνει στο τέλος του κυρίου βρόχου. Το ίδιο επιχείρημα ισχύει εάν βρισκόταν στην αρχή του βρόχου.) Επομένως το S_y πρέπει να περιέχει αρκετές καταστάσεις ώστε να μπορεί να υλοποιήσει το κανονικό μονοπάτι. Είναι εύκολο να ελένξει κανείς ότι οι πρώτες $d+1+l$ καταστάσεις του μονοπατιού αυτού πρέπει να είναι διαφορετικές, αφού τα κανονικά μονοπάτια τους δεν αποτελούν πρόθεμα κάποιου άλλου, θεωρώντας για κάθε τέτοιο μονοπάτι την πρώτη εμφάνιση ενός μπλοκ με τουλάχιστον $l-d-2$ A ακολουθούμενα από ένα O. Οι καταστάσεις της διόρθωσης και της κύριας φάσης διακρίνονται μεταξύ τους από το μήκος και την θέση του μπλοκ των A. Επίσης αξίζει να σημειωθεί ότι αυτό ισχύει και στην περίπτωση που $x \oplus x_y = A^d$ εξαιτίας του O που υπάρχει πριν το $y \oplus x_y$. Τελικά, τα κανονικά μονοπάτια των πρώτων $d+1+l$ καταστάσεων στο S_y και $S_{y'}$ δεν είναι προθέματα κάποιου άλλου (εξαιτίας του παραπάνω λήμματος 5) και επομένως πρέπει να είναι όλα διαφορετικά. $\square$

Έχοντας λοιπόν "λογοδοτήσει" για όλες τις καταστάσεις του A_x, εκτός από τις $2^d - 1$ καταστάσεις που βρίσκονται στην κορυφή, και την κατάσταση τιμωρίας. Αντίθετα για κατάσταση τιμωρίας δεν μπορούμε να είμαστε σίγουροι καθώς ο απατεώνας μπορεί να επιλέξει να μην τιμωρήσει τον αντίπαλο του. Ακόμη έχουμε ότι οι καταστάσεις του δυαδικού δένδρου πρέπει να είναι όλες διαφορετικές μεταξύ τους και διαφορετικές από τις καταστάσεις του S_y:

Λήμμα 5.7. *Οι καταστάσεις στις οποίες μπορεί να μεταβεί το A' κατά την διάρκεια πρώτων d βημάτων είναι $2^d - 1$ καταστάσεις οι οποίες είναι διαφορετικές από τις καταστάσεις του S_y.*

Απόδειξη Ας θεωρήσουμε τις $2^d - 1$ ακολουθίες μήκους το πολύ $d - 1$, και τις καταστάσεις στις οποίες οδηγούν το A'. Θα δείξουμε ότι οι καταστάσεις αυτές είναι διακριτές και ξεχωριστές από αυτές του S_y. Ας θεωρήσουμε μία ακολουθία z μήκους $i < d$ και έστω $s(z)$ η κατάσταση στην οποία φθάνει το A' ως απάντηση στο z. Ακόμη ας υποθέσουμε ότι w είναι το κανονικό μονοπάτι (ακολουθία) που ξεκινά από το $s(z)$ και έχει μήκος $d - i$, έστω $y = zw$ και έστω $s(y)$ η τελική κατάσταση του μονοπατιού, τότε εξ' ορισμού έχουμε ότι $s(y)$ είναι η κατάσταση που επισκέπτεται το A' στο $d + 1$ βήμα όταν παίζει με το B_y, δηλαδή η πρώτη κατάσταση του S_y. Έχουμε λοιπόν τότε από το προηγούμενο λήμμα ότι οι καταστάσεις $s(z)$ για διαφορετικά z μήκους μικρότερα του d είναι διαφορετικά το ένα από το άλλο. Είναι εύκολο να διαπιστωθεί με ανάλογο τρόπο ότι αυτές οι καταστάσεις είναι διαφορετικές από όλες τις καταστάσεις όλων των συνόλων S_y. □

Αξίζει να σημειωθεί ότι το A' δεν χρειάζεται να είναι ακριβώς το ίδιο με κανένα από τα A_x επειδή μπορεί να δίνει διαφορετικές ταυτότητες x_y σε διαφορετικά B_y. Όμως, για κάθε $i = 1, ..., d$, το i-οστό σύμβολο που δίνει το A' είναι ανεξάρτητο του i-οστού συμβόλου του y. Εκ κατασκευής της φάσης διόρθωσης συνεπάγεται ότι το κέρδος που λαμβάνει το A' κατά τα πρώτα $2d+1$ είναι το ίδιο με το κέρδος που λαμβάνει οποιοδήποτε από τα A_x (δηλαδή ένα κέρδος κατά μέσο όρο ίσο με $\frac{7}{4}$ για καθένα από τα πρώτα $2d$ βήματα και ένα για το $2d+1$-οστό βήμα. Σύμφωνα με το πρώτο λήμμα της απόδειξης, η ακολουθία $y \oplus x_y$, για κάθε y, οδηγεί σε όλες τις δυνατές ακολουθίες μήκους d, οπότε το A' αποκομίζει το ίδιο κέρδος όπως και οποιοδήποτε A_x στα επόμενα $2l$ βήματα. Ο μόνος τρόπος ώστε το A' να αποκτήσει επιπλέον όφελος είναι να αποκλίνει από το κανονικό παίγνιο για κάποιο B_y. Έστω ότι ισχύει αυτή η περίπτωση, τότε μετά την απόκλιση το A' θα τιμωρηθεί για την υπόλοιπη διάρκεια του παιγνίου, επομένως μία αποστασία είναι επικερδής μόνο εφόσον συμβαίνει στο τελευταίο βήμα (π.χ. το A' παίζει O αντί για A. Επομένως το αυτόματο A' παίζοντας με το B_y ακολουθεί από το βήμα $2d+1$ μέχρι το n ένα κανονικό μονοπάτι της μορφής $[A^{l-d-2}O(y \oplus x_y)A]^k A^j O$ με $j < l-d-2$. Είναι εύκολο να ελένξει κανείς ότι όλες οι καταλήξεις με μήκος μεγαλύτερο του l δεν έχουν κοινά προθέματα, επομένως όλες εκτός από πιθανώς από τις τελευταίες l καταστάσεις είναι διακριτές. Με παρόμοια επιχειρήματα, δείχνουμε ότι είναι διαφορετικές από τις καταστάσεις άλλων $S_{y'}$ και από τις καταστάσεις που επισκέπτεται το A' κατά τα πρώτα d βήματα. Έχουμε ήδη μετρήσει στο S_y τις πρώτες l καταστάσεις του μονοπατιού. Επομένως, το A' χρειάζεται ένα επιπλέον σύνολο $n-2d-2l-2$ καταστάσεων προκειμένου να μπορεί να αποστατήσει κατά το τελευταίο βήμα και να κερδίσει ένα πλεονέκτημα. Επομένως, εάν ο A' δεν τιμωρείται από κανένα B_y κατά τα πρώτα $2d+1+2l$ βήματα, δεν μπορεί να βελτιώσει το κέρδος του A_x.

Ας υποθέσουμε ότι το A' τιμωρείται από κάποια B_y κατά τους πρώτους $2d+1+2l$ γύρους και έστω P να είναι το σύνολο των y για τα οποία αυτό συμβαίνει, τότε για κάθε $y \in P$, το όφελος του A' εναντίον του B_y είναι τουλάχιστον $(n-2d-1-2l)$ λιγότερο από το όφελος οποιουδήποτε A_x, (αφού το A_x και το B_y παίζουν AA σε περισσότερο από τους μισούς γύρους σε κάθε βρόχο), με αποτέλεσμα το συνολικό έλλειμμα να είναι $|P|(n-2d-1-2l)$. Το δεύτερο λήμμα της απόδειξης (λήμμα 6) ακόμα ισχύει για τα S_y, τα οποία αντιστοιχούν σε $y \notin P$. Όσο για την απόδειξη του λήμματος 7, ισχύει σε όλες τις ακολουθίες μήκους μικρότερου του d που δεν είναι προθέματα για κανένα $y \in P$. Με άλλα λόγια, για κάθε $y \in P$ μπορεί να αποτύχουμε να δικαιολογήσουμε τις περισσότερες από τις $(2d+1+l)$ καταστάσεις, που σχηματίζουν ένα σύνολο $|P|(2d+1+l)$ καταστάσεων. Ο μοναδικός τρόπος για το A' να κερδίσει από μία αποστασία είναι να κερδίσει ένα όφελος 1 κατά τον τελευταίο γύρο για ορισμένα y. Όμως προκειμένου να το επιτύχει αυτό, το A' πρέπει να ξετυλίξει τον $|l|$ κύκλο που αντιστοιχεί σε αυτά τα y και να μετρήσει μέχρι το n. Όπως εξηγήσαμε και παραπάνω, αυτό θα χρειαστεί τουλάχιστον $n-2d-1-2l$ επιπλέον καταστάσεις για κάθε y. Το συνολικό όφελος είναι $\frac{|P|(2d+1+l)}{n-2d-2l}$, λιγότερο από την απώλεια, γι' αυτό υποθέτουμε ότι $P=0$.

Έτσι ολοκληρώσαμε την απόδειξη της περίπτωσης της ύπαρξης του ορίου $S = 2^d(d+l+2)$. Το ζεύγος των στρατηγικών που έχουμε κατασκευάσει δεν μπορεί να βελτιωθεί από κανένα παίκτη (ακόμη και όταν χρησιμοποιεί αυτόματα με μέχρι $S+(n-2d-2l-3)$ καταστάσεις). Στην συνέχεια θα αλλάξουμε λίγο την βασική κατασκευή προκειμένου να μπορέσουμε να χαλαρώσουμε λίγο τους περιορισμούς πάνω στα όρια. Καταρχήν θα εξετάσουμε την εύκολη περίπτωση κατά την οποία ένα από τα όρια είναι μικρότερο του n.

Λήμμα 5.8. *Εάν $S_I(n) < n-1$ για κάποιο n, τότε υπάρχει μία ισορροπία με όφελος τουλάχιστον $3n-3$.*

Εάν $S_{II}(n) < n$, τότε τα δύο αυτόματα παίζουν ένα σου και ένα μου (titfor-tat)είναι σε ισορροπία. Διαφορετικά η ισορροπία αποτελείται από ένα σου και ένα μου, ενάντια σε ένα αυτόματο ν καταστάσεις που συνεργάζεται σε όλους τους γύρους εκτός από τον τελευταίο. □

Επομένως, μπορούμε να θεωρήσουμε ότι τα όρια και για τους δύο αντιπάλους είναι τουλάχιστον $n-1$. Αρχικά θα χαλαρώσουμε την υπόθεση για ότι το S, το κοινό όριο καταστάσεων, είναι της μορφής $S = 2^d(d+l+2)$. Έστω $S \leq 2^{c_\varepsilon n}$το δοσμένο όριο, έστω d ο μεγαλύτερος ακέραιος τέτοιος ώστε $2^d d(2+\frac{2}{\varepsilon}) \leq S$ και έστω l ο μεγαλύτερος ακέραιος τέτοιος ώστε $2^d(d+l+2) \leq S$. Από την επιλογή μας για το d έχουμε $d+\log d+\log(2+\frac{2}{\varepsilon}) \leq c_\varepsilon n$, επομένως $d+1 \leq c_\varepsilon n = \varepsilon n/12(1+\varepsilon)$. Από την άλλη $2^{d+1}(d+1)(2+\frac{2}{\varepsilon}) \geq S \geq n-1$. Από την επιλογή των l και d έχουμε ότι $d+l+2 \leq S/2^d < 2(d+1)(2+\frac{2}{\varepsilon})$, επομένως $3l+2d < (d+1)12(1+\varepsilon)/\varepsilon = (d+1)/c_e < n$, δηλαδή ικανοποιείται η πρώτη ανισότητα του θεωρήματος. Ακόμη, $d+l+3 > S/2^d \geq d(2+\frac{2}{\varepsilon}$, οπότε εάν το n είναι αρκετά μεγάλο έτσι ώστε $d \geq 6$ (π.χ. $n \geq 1+2^6 6(2+\frac{2}{\varepsilon})$), τότε η ανισότητα αυτή συνεπάγεται ότι $2l\varepsilon > 4d+2\varepsilon(d-3) \geq 3d+6$ επομένως $\frac{2+d}{l} < \frac{2\varepsilon}{3}$. Αφού $\frac{4d}{n} < \frac{\varepsilon}{3}$ έπεται ότι η δεύτερη ανισότητα του θεωρήματος, δηλαδή η $\frac{2+d}{l} + \frac{4d}{n} < \varepsilon$ ισχύει επίσης.

Η κατασκευή για αυτήν την επέκταση είναι μία μικρή μετατροπή της βασικής, η οποία όμως αιτιολογεί $R < 2^d$ επιπλέον καταστάσεις, όπου R είναι το υπόλοιπο της διαίρεσης του S με το 2^d. Το επιτυγχάνουμε αυτό ως εξής: Έστω T οποιοδήποτε σύνολο R ακολουθιών μήκους d, τότε για κάθε ζεύγος αυτομάτων A_x και B_y τέτοια ώστε $x \oplus y \in T$ ας θεωρήσουμε τις φάσεις διόρθωσης για αυτά τα αυτόματα που αντιστοιχούν στην αλληλεπίδραση αυτών των δύο. Αύξησε τα μήκη των φάσεων αυτών από $d+1$ σε $d+2$ εισάγοντας μία κατάσταση που παίζει A ακριβώς πριν από την τελευταία κατάσταση της φάσης (στην οποία παίζεται O). Είναι φανερό ότι αυτή η μετατροπή δεν επηρεάζει το όφελος αλλά αυξάνει τον αριθμό των καταστάσεων σε όλα τα αυτόματα σε S.

Τέλος επεκτείνουμε την κατασκευή στην γενική περίπτωση όπου τα όρια μπορεί να μην είναι ίσα και όπου στην ουσία μπορεί μόνο ένας παίκτης να είναι φραγμένος. Σύμφωνα με το λήμμα 8 μπορούμε να υποθέσουμε ότι και τα δύο όρια είναι τουλάχιστον $n-1$ Εάν και τα δύο είναι το πολύ $n+1$ μπορούμε να χρησιμοποιήσουμε την αρχική κατασκευή με $S = n-1$, γιατί όπως θυμόμαστε το ζεύγος των στρατηγικών μας δεν μπορεί να βελτιωθεί από κανένα αυτόματο-απατεώνα ακόμη και εάν έχει $n-2d-2l-3$ επιπλέον καταστάσεις. Έτσι, ας υποθέσουμε ότι ένα από τα όρια, ας πούμε το $S_I I(n)$ ισχύει $S_I I(n) \geq n-2$ και ότι για τον παίκτη I ισχύει $S_I(n) \leq 2^{c_\varepsilon n}$.

Στο ζεύγος ισορροπίας σας, ο παίκτης I χρησιμοποιεί μία στρατηγική μόνο δύο αυτομάτων ισορροπίας που αντιστοιχούν στις ταυτότητες C^d και $C^{d-1}D$ και ο II χρησιμοποιεί μία μεικτή στρατηγική που αποτελείται από 2^d αυτόματα, ένα για κάθε $y \in \{A, O\}^d$. Οι αναμενόμενες μεταβάσεις είναι οι ίδιες όπως και πριν: ανακοίνωση της ταυτότητας (d βήματα), φάση διόρθωσης και επανειλημμένες εκτελέσεις του κυρίου βρόχου μήκους l. Παρόλα αυτά, η φάση διόρθωσης και η κυρία φάση είναι τώρα διαφορετικές επειδή τα επιχειρήματα που είχαμε επικαλεστεί και βασίζονταν στην ποικιλία των x δεν ισχύουν πια. Η γενική κατασκευή κάθε αυτομάτου του I είναι παρόμοια με την βασική κατασκευή: ένα πλήρες δυαδικό δένδρο στην κορυφή το οποίο ακολουθείται από αλυσίδες και βρόχους που κρέμεται από τα φύλλα. Κάθε αυτόματο του II έχει $n+2$ καταστάσεις οι οποίες σχηματίζουν αλυσίδα ύψους n με μία κατάσταση σε κάθε επίπεδο εκτός από τα επίπεδα $d+1$ και $d+2$ (που αντιστοιχούν στο πρώτο και το δεύτερο βήμα της φάσης διόρθωσης) και έχουν δύο καταστάσεις. Η τελευταία κατάσταση (στο n-οστό επίπεδο) είναι η κατάσταση τιμωρίας.

Η φάση ανακοίνωσης των ταυτοτήτων, αποτελείται από τα πρώτα d βήματα είναι απλή: ο I παίζει την ταυτότητά του C^d ή $C^{d-1}D$ και ο II παίζει$y \in \{A, O\}^d$. Ο παίκτης I καταγράφει το y (π.χ. τα πρώτα d επίπεδα του αυτομάτου του I σχηματίζουν ένα πλήρες δυαδικό δένδρο, ενώ κάθε αυτόματο του II ξεκινά με ένα μονοπάτι μήκους d (με όλες τις μεταβάσεις πάνω στο C). Στο υπόλοιπο του παιγνίου και οι δύο παίκτες παίζουν ντετερμινιστικά και οι δύο είτε AA είτε OO.

Θα περιγράψουμε την κύρια (τρίτη) φάση προτού περιγράψουμε την φάση διόρθωσης, Κάθε αυτόματο του I έχει πάλι ένα βρόχο για κάθε $y \in \{A, O\}^d$ μήκους l. Έστω $\hat{d}$ ένας ζυγός αριθμός τέτοιος ώστε να είναι τουλάχιστον $d + \lceil \log d \rceil$ και έστω f μία οποιαδήποτε 1-1 συνάρτηση από το $\{A, O\}^d$ στο σύνολο των ακολουθιών μήκους $\hat{d}$ οι οποίες περιέχουν ίσο αριθμό A και O (αξίζει να σημειωθεί ότι το $\hat{d}$ είναι αρκετά μεγάλο έτσι ώστε να υπάρχουν τουλάχιστον 2^d ακολουθίες μήκους $\hat{d}$ με ίσο αριθμό A και O και επομένως υπάρχει τέτοια συνάρτηση f. Η κύρια ̃φάση δεν περιέχει τώρα το $\oplus$ των δύο ταυτοτήτων, αλλά αποτελείται από βρόχους $A^{l-d-2}Of(y)A$. Η εισαγωγή στον βρόχο γίνεται πάλι στην κατάλληλη κατάσταση έτσι ώστε το παίγνιο να τερματίζει με ένα μπλοκ από A. Αυτή η κατασκευή εξασφαλίζει ότι τα κέρδη είναι ισοκαταμερισμένα κατά την διάρκεια αυτής της φάσης. Σε κάθε αυτόματο του παίκτη II ο βρόχος ξετυλίγεται μέχρι το τελευταίο στάδιο n κατά το οποίο ο παίκτης II αποστατεί.

Η φάση διόρθωσης έχει $d+1$ βήματα εκτός από ένα σύνολο T από R ταυτότητες y του II, για τις οποίες προσθέτουμε ένα ακόμη βήμα στην φάση διόρθωσης όπως πριν έτσι ώστε ο χώρος που παρέχεται στον παίκτη I να χρησιμοποιείται πλήρως (θα δώσουμε την ακριβή τιμή του R στην συνέχεια). Οι παίκτες συμπεριφέρονται ακολούθως. Από εδώ και μπρος αναφερόμαστε στο τελευταίο βήμα της πρώτης ̃φάσης (το βήμα που ο I παίζει A ή O) ως το τυχαίο βήμα. Στο πρώτο βήμα της φάσης διόρθωσης και οι δύο παίκτες παίζουν A αν είχαν παίξει A στο τυχαίο βήμα, διαφορετικά παίζουν και οι δύο O. Αυτό το βήμα εξισορροπεί το κέρδος του τυχαίου βήματος και για τους δύο παίκτες. Η υπόλοιπη φάση διόρθωσης αποσκοπεί στην εξισορρόπηση του κέρδους του παίκτη ΙΙ κατά τα πρώτα $d-1$ βήματα από τις διαφορετικές ταυτότητες. Έστω λοιπόν ότι y, η ταυτότητα του ΙΙ, περιέχει στα πρώτα $d-1$ γράμματα έναν άρτιο αριθμό από O, ας πούμε $2a$. Τότε στο υπόλοιπο της φάσης διόρθωσης τόσο ο Ι όσο και ο ΙΙ εκπέμπουν την ακολουθία $O^a A^{d-a-1} O$ εάν $y \notin T$ και $O^a A^{d-a} O$ εάν $y \in T$ (δηλαδή εάν $y \in T$ προσθέτουμε ένα επιπλέον A βήμα πριν το τελικό O βήμα). Έστω τώρα ότι το y περιέχει ένα περιττό αριθμό από O, ας πούμε $2a+1$ στα πρώτα $d-1$ γράμματα, τότε εάν οι δύο παίκτες συμφώνησαν στο τυχαίο βήμα (δηλαδή έπαιξαν και οι δύο είτε A είτε O), τότε και οι δύο παίκτες στέλνουν την ακολουθία $AO^\alpha A^{d-a-2} O$ εάν $y \notin T$ ή την ακολουθία $AO^\alpha A^{d-a-1} O$ εάν $y \in T$. Στην αντίθετη περίπτωση (εάν διαφώνησαν κατά το τυχαίο βήμα), τότε και οι δύο παίκτες στέλνουν την ακολουθία $O^{a+1} A^{d-a-2} O$ εάν $y \notin T$ ή την ακολουθία $O^{a+1} A^{d-a-1} O$ εάν $y \in T$.

Η φάση διόρθωσης του αυτομάτου του παίκτη Ι κρέμεται από ένα φύλλο που αντιστοιχεί σε μία ταυτότητα y είναι απλώς ένα μονοπάτι. Το τμήμα της διόρθωσης του αυτομάτου του παίκτη ΙΙ Έχει δύο κόμβους στα πρώτα δύο επίπεδα και ένα κόμβο στα υπόλοιπα. Επίσης αξίζει να σημειωθεί ότι για κάθε ταυτότητα y του ΙΙ, η κίνηση του Ι κατά το τυχαίο βήμα επηρεάζει μόνο τα δύο πρώτα βήματα της φάσης διόρθωσης εάν ο αριθμός των O κατά τα πρώτα $d-1$ βήματα είναι περιττός και μόνο κατά το πρώτο βήμα εάν είναι άρτιος.

Είναι εύκολο να επιβεβαιωθεί ότι ο παίκτης Ι δεν έχει κάποιο όφελος από το να επιλέξει ένα σύμβολο αντί για κάποιο άλλο κατά το τυχαίο βήμα όταν παίζει ενάντια σε αυτήν την στρατηγική του ΙΙ. Πράγματι ας υποθέσουμε ότι η ταυτότητα του ΙΙ έχει άρτιο αριθμό από O κατά τα πρώτα $d-1$ βήματα. Στην συνέχεια η επιλογή του Ι κατά το τυχαίο βήμα επηρεάζει μόνο αυτό το βήμα και το επόμενο. Εάν επιλέξει A, τότε το αναμενόμενο όφελος του είναι $1/2(3+3)+1/2(0+1)=7/2$. Όμοια εάν επιλέξει O το αναμενόμενο όφελός του είναι $1/2(4+1)+1/2(1+1)=7/2$. Αν υποθέσουμε τώρα ότι η ταυτότητα του ΙΙ έχει περιττό αριθμό από O στα πρώτα $d-1$ γράμματα, τότε το αναμενόμενο κέρδος του Ι κατά το τυχαίο βήμα και το πρώτο βήμα της διόρθωσης είναι πάλι $7/2$ ανεξάρτητα από το εάν παίζει A ή O. Το αναμενόμενο κέρδος του κατά τα υπόλοιπα βήματα της φάσης διόρθωσης δεν εξαρτάται από την επιλογή του κατά το τυχαίο βήμα επειδή για κάθε επιλογή υπάρχει η ίδια πιθανότητα $(1/2)$ ότι θα συμφωνήσουν ή θα διαφωνήσουν με την επιλογή του παίκτη ΙΙ στο τυχαίο βήμα.

Ας θεωρήσουμε τώρα το αναμενόμενο όφελος του παίκτη ΙΙ από τις πρώτες δύο φάσεις για μία ταυτότητα $y \notin T$. Το τυχαίο βήμα και το πρώτο βήμα της φάσης διόρθωσης συνεισφέρουν όπως παραπάνω $7/2$ (ανεξάρτητα από την επιλογή του ΙΙ κατά το τυχαίο βήμα). Εάν το y έχει άρτιο αριθμό, $2a$, από O κατά τα πρώτα $d-1$ γράμματα, τότε τα εναπομείναντα βήματα των πρώτων δύο φάσεων συνεισφέρουν $(2a)4 + (d - 1 - 2a)3$ από την πρώτη φάση και $(a+1)1+(d-a-1)3$ από την δεύτερη φάση που αθροίζουν σε $6d-5$. Εάν το y έχει περιττό αριθμό, $2a+1$, από O κατά τα πρώτα $d-1$ γράμματα, τότε τα εναπομείναντα βήματα των πρώτων δύο φάσεων συνεισφέρουν $(2a+1)4+(d-1-2a-1)3$ από την πρώτη φάση και $1/2[(a+1)1+(d-a-1)3+(a+2)1+(d-a-2)3]$ από την δεύτερη φάση που αθροίζουν πάλι σε $6d-5$. Επομένως, εάν $y \notin T$, το αναμενόμενο όφελος του παίκτη II από την πρώτη από τις δύο φάσεις είναι $6d-3/2$, ανεξαρτήτως του y. Στην περίπτωση μίας ταυτότητας $y \in T$, το αναμενόμενο όφελος κατά την διάρκεια των πρώτων δύο φάσεων είναι $6d+3/2$ (ανεξάρτητα από το όφελος του y), δηλαδή έχει ένα επιπρόσθετο όφελος 3 εξαιτίας του επιπλέον βήματος (A, A) κατά την φάση διόρθωσης, το οποίο όμως ακυρώνεται εξαιτίας του ενός λιγότερου (A, A) βήματος κατά την διάρκεια της κύριας φάσης. Επομένως το ολικό αναμενόμενο όφελος του II είναι ανεξάρτητο της ταυτότητας y.

Εδώ αναφέρουμε ότι κάθε αυτόματο του παίκτη I έχει $2^d(d+l+2)+R$ καταστάσεις. Οι παράμετροι d, l και R επιλέγονται ως εξής: Έστω d ο μεγαλύτερος ακέραιος τέτοιος ώστε $2^d d(2+\frac{2}{\varepsilon}) \leq S_I(n)$. Έστω l ο μεγαλύτερος ακέραιος τέτοιος ώστε $2^d(d+l+2) \leq S_I(n)$. Τέλος θέτουμε $R = S_I(n) - 2^d(d+l+2)$ (δηλαδή r είναι το υπόλοιπο της διαίρεσης του $S_I(n)$ με το 2^d) και έστω T οποιοδήποτε σύνολο R ακολουθιών μήκους d. ΈΤσι ολοκληρώσαμε την περιγραφή των δύο στρατηγικών μας.

Είναι εύκολο να δείξουμε ότι κάθε αυτόματο του II παίζει βέλτιστα απέναντι στην στρατηγική του I: Η ακολουθία μήκους d που παίζεται από τον II στα πρώτα d βήματα καθορίζει το υπόλοιπο παίγνιο επειδή εάν ο II αποστατήσει πριν το τέλος, τότε θα τιμωρηθεί. Επιπλέον ο II αποστατεί στο τελευταίο βήμα, όπως τον συμφέρει. Από την κατασκευή όλες οι αρχικές ακολουθίες μήκους d αποφέρουν το ίδιο όφελος για τον II. Η απόδειξη ότι ο I δεν μπορεί να βελτιώσει την στρατηγική του με το δεδομένο όριο στις καταστάσεις είναι η ίδια με πριν. Έτσι ολοκληρώνεται η απόδειξη του θεωρήματος. □

5.5 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκαν διάφορες προσεγγίσεις που αποσκοπούσαν στην ερμηνεία των παραδόξων πεπερασμένων οριζόντων. Πάντως οι ερμηνείες που προσφέρουν δεν κρίνονται πάντα ικανοποιητικές καθώς ορισμένες φαίνονται αρκετά απλοϊκές (εάν ο παίκτης δεν μπορεί να μετρήσει μέχρι το περίοδο τέλους του παιγνίου τότε το παράδοξο εξαφανίζεται), ενώ άλλες (όπως η μέθοδος του Neyman) χρησιμοποιούν μεθόδους που κάθε άλλο παρά φυσικές θα μπορούσαν να χαρακτηρισθούν (επικοινωνία ανάμεσα στους παίκτες μέσω των κινήσεών τους). Ενδεχομένως το πρόβλημα να εντοπίζεται στο γεγονός ότι οι παίκτες κατά την συμμετοχή τους στα παίγνια αυτά δεν χρησιμοποιούν την ανάγωγη κατά την αξιολόγηση της τρέχουσας κατάστασής τους, αλλά άλλες διαδικασίες συλλογισμού. Ο καθορισμός τους, πάντως, αποτελεί ένα συναρπαστικό αντικείμενο για μελλοντική έρευνα.

Κεφάλαιο 6

Ζητήματα Υπολογισιμότητας

6.1 Εισαγωγή

Όπως έχουμε αναφέρει ήδη, η εργασία αυτή επικεντρώνεται στο πάντρεμα γνωστών ιδεών της θεωρίας παιγνίων με έννοιες της πολυπλοκότητας σε μία προσπάθεια να προσεγγίσουμε κατά το δυνατών την συμπεριφορά των "λογικών" παικτών. Πραγματικά μέχρι τώρα έχουμε επεκτείνει την κλασσική θεωρία παιγνίων, παρουσιάζοντας μοντέλα στα οποία υπεισέρχεται η έννοια του κόστους της πολυπλοκότητας της εφαρμοζόμενης στρατηγικής. Όμως όπως γνωρίζουμε, δεν μπορούμε να υπολογίσουμε κάθε διεργασία που μπορούμε να περιγράψουμε με την βοήθεια μίας υπολογιστικής μηχανής. Επομένως γεννάται το ερώτημα, μήπως ακόμη και εάν αγνοήσουμε τα υπολογιστικά κόστη, η προσπάθειά μας να υπερπηδήσουμε τους σκοπέλους της προσέγγισης της συμπεριφοράς πραγματικών παικτών (οι οποίοι αναδεικνύονται από διάφορα παράδοξα) προσκρούει σε θεμελιώδεις περιορισμούς υπολογισιμότητας; Κατά το κεφάλαιο αυτό θα εξετάσουμε, χωρίς να επεκταθούμε σε μεγάλο βάθος το ζήτημα αυτό.

Έτσι για να συγκεκριμενοποιήσουμε την συζήτηση αυτή ας επιστρέψουμε στο παίγνιο της Μάχης των Φύλων:

Παράδειγμα 6.1 (Battle of The Sexes).

Ένα ζευγάρι επιθυμεί να πάει σε ένα μουσικό κονσέρτο. Όμως υπάρχουν δύο επιλογές, μπορούν να ακούσουν είτε Bach είτε Stravinsky. Η κύρια επιδίωξή τους είναι να βγουν έξω μαζί, αλλά ο άντρας προτιμά τον Bach και η γυναίκα τον Straninsky. Οι προτιμήσεις τους μάλιστα εκφράζονται από τον παρακάτω πίνακα.

$$G = \begin{array}{c c} & \begin{array}{cc} B & S \end{array} \\ \begin{array}{c} B \\ S \end{array} & \begin{bmatrix} (2,1) & (0,0) \\ (0,0) & (1,2) \end{bmatrix} \end{array}$$

Για μια αναδρομή στην ιστορία του παιγνίου της "Μάχης των Φύλων" μπορεί να ανατρέξει κανείς στους Luce και Raiffa(1957, pp.90-91)

Το παίγνιο αυτό μοντελοποιεί μια κατάσταση κατά την οποία οι παίκτες θέλουν να συνεργαστούν μεταξύ τους αλλά τα συμφέροντά τους είναι σε σύγκρουση. "Η Μάχη των Φύλων " έχει δύο σημεία ισορροπίας Nash, τα (B,B) και (S,S), δηλαδή τα σημεία που οι παίκτες είτε συμφωνούν να παρακολουθήσουν Bach(B), είτε συμφωνούν να παρακολουθήσουν Stravinsky(S).

Βασιζόμενοι όμως αποκλειστικά στα δεδομένα που μας παρέχονται από τη δομή του στρατηγικού παιγνίου μπορούμε να πούμε πολύ λίγα για την κίνηση που θα αναμέναμε από τον παίκτη 1. Πραγματικά ο παίκτης 1 πρέπει να γνωρίζει πώς ο παίκτης 2 αναλύει το παίγνιο προκειμένου να καταλήξει στην απόφασή του. Έτσι για παράδειγμα εάν ο παίκτης 1 γνωρίζει ότι ο παίκτης 2 είναι ένας απαιτητικός παίκτης ο οποίος πάντα επιδιώκει ακόμη και με ρίσκο να αποκομίσει το μεγαλύτερο δυνατό όφελος, τότε θα ήταν εύλογο για τον παίκτη 1 να παίζει S. Σε ένα άλλο σενάριο, εάν ο παίκτης 1 γνωρίζει ότι ο παίκτης 2 είναι απαιτητικός μόνο εφόσον παίζει εναντίον ενός "αδύναμου" αντιπάλου, τότε η εύλογη κίνηση του παίκτη 1 θα εξαρτιόταν από την πεποίθηση που έχει σχετικά με το πώς τον αξιολογεί ο παίκτης 2. Η ιδέα ότι η κίνηση ενός "πραγματικού" παίκτη εξαρτάται από την γνώση ή την πεποίθηση που τρέφει για τον τύπο του αντιπάλου βρίσκει εφαρμογή ακόμη και στην περίπτωση παιγνίων με μοναδικό σημείο ισορροπίας. Έστω για παράδειγμα το παρακάτω παίγνιο:

$$G = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{bmatrix} (2,1) & (0,0) \\ (0,3) & (1,2) \end{bmatrix} \end{array}$$

Παρατηρούμε ότι το παίγνιο αυτό έχει μοναδικό σημείο ισορροπίας το (A, A), παρόλα αυτά για να παίξει ο παίκτης 1 πρέπει να ξέρει ότι ο παίκτης 2 δεν επιλέγει την κυριαρχούμενη στρατηγική B.

Οι περιορισμοί υπολογισιμότητας λοιπόν, θα μας απασχολήσουν στο πλαίσιο του στρατηγικού παιγνίου που ένας παίκτης γνωρίζει πλήρως τον άλλο. Χρησιμοποιώντας την καθιερωμένη παιγνιο-θεωρητική ορολογία, θα αναφερόμαστε στην ταυτότητα ενός παίκτη ως τον *τύπο* του. Λέγοντας τύπο εννοούμε νόρμες συμπεριφορών της μορφής "να είσαι ευγενικός με τους ευγενικούς παίκτες και επιθετικός με τους υπόλοιπους". Επίσης θα υποθέτουμε ότι κάθε παίκτης έχει στην διάθεσή του την πλήρη περιγραφή των αντίπαλων παικτών και είναι ελεύθερος να την χρησιμοποιήσει όπως αυτός κρίνει σκόπιμο. Έτσι μπορεί κατά τον συλλογισμό του να μην χρησιμοποιεί ολόκληρη την περιγραφή του αντιπάλου, αλλά έχει πρόσβαση σε όλες τις λεπτομέρειες. Αντίθετα, δεν υποθέτουμε ότι ο παίκτης έχει πρόσβαση στην περιγραφή του εαυτού του. Η παραπάνω υπόθεση δεν επηρεάζει την ουσία των παρατηρήσεων που θα πραγματοποιήσουμε αλλά μάλλον εκφράζει πιο πιστά την πραγματικότητα.

Ένα πολύ σημαντικό ερώτημα που απορρέει από την παραπάνω παράγραφο είναι πως ορίζουμε την πλήρη περιγραφή κάποιου παίκτη. Έτσι λοιπόν προφανώς θέλουμε να υπάρχει κάποιος τύπος t_x που παίζει x ανεξάρτητα από την γνώση που έχει για τον αντίπαλο παίκτη. Επίσης θέλουμε να υπάρχει ένας τύπος που παίζει y όποτε αντιμετωπίζει αντίπαλο τύπου t_x ενώ παίζει z στην περίπτωση που αντιμετωπίζει άλλο τύπο. Γρήγορα λοιπόν συνειδητοποιούμε ότι αυτός ο αφελής ορισμός του τύπου οδηγεί σε ένα τεράστιο σύνολο. Επιπλέον, δεν είναι δυνατό να αντιμετωπίσουμε τον τύπο σαν να επρόκειτο για μία οποιαδήποτε συνάρτηση απόκρισης που αναθέτει αυθαίρετα τιμές απέναντι σε κάθε τύπο εναντίον του οποίου παίζει. Εάν πραγματοποιούσαμε κάτι τέτοιο, τότε η πληθικότητα του συνόλου των τύπων θα ήταν δύναμη της ίδιας της πληθικότητας, το οποίο είναι προφανώς αδύνατο. Αντίθετα, θα ορίσουμε τον τύπο ως έναν αλγόριθμο που δέχεται ως είσοδο μια πεπερασμένη περιγραφή του αλγορίθμου του αντιπάλου και παράγει μία ακολουθία ενεργειών ως έξοδο.

Η υπόθεση ότι ένας παίκτης γνωρίζει τον τύπο του άλλου είναι πολύ ισχυρή. Πρέπει να σημειώσουμε ότι αυτού του είδους η γνώση δεν είναι η γνώση ισορροπίας, που έχει ένας παίκτης για τους υπόλοιπους βασιζόμενος πάνω στις παρατηρήσεις του κατά την "σταθερή" κατάσταση ισορροπίας. Εδώ, υποθέτουμε ότι ένας παίκτης γνωρίζει πλήρως τον αλγόριθμο, δηλαδή τον τρόπο σκέψης και συμπεριφοράς του άλλου.

Στην συνέχεια θα επιχειρήσουμε μια πολύ "χαλαρή" εισαγωγή στην απάντηση του τί σημαίνει υπολογίζω και τι είναι υπολογιστό. Αυτό το ζήτημα υπήρξε κεφαλαιώδους σημασίας στην έρευνα των μαθηματικών κατά την διάρκεια του εικοστού αιώνα. Μία αρκετά κοινή θεώρηση είναι η αντιμετώπιση του υπολογισμού ως μία ακολουθία από απλές πράξεις (λειτουργίες) που πραγματοποιούνται σύμφωνα με ένα πεπερασμένο σύνολο εντολών και χρησιμοποιώντας κάποιο "χώρο εργασίας" πάνω στον οποίο μπορεί να αποθηκευτεί μνήμη (π.χ. ενδιάμεσων αποτελεσμάτων) κατά την διάρκεια του υπολογισμού. Οι λειτουργίες είναι της παρακάτω μορφής: ανάγνωση συμβόλων, αντικατάσταση ενός συμβόλου από ένα άλλο και μετακίνηση από μία θέση του χώρου εργασίας σε κάποια άλλη. Οι εντολές καθορίζουν τί πρέπει να πράξει η υπολογιστική μηχανή και πώς πρέπει να κινηθεί πάνω στο χώρο μνήμης.

Παρακάτω δίνουμε την περιγραφή του πιο γνωστού φορμαλισμού της υπολογιστικής μηχανής: το μοντέλο της μηχανής Turing (TM). Η μηχανή αποτελείται από ένα μη πεπερασμένο αριθμό κελιών (κυττάρων) πάνω σε μία ταινία (η οποία έχει κάποια συγκεκριμένη αρχή). Σε κάθε ένα από τα κελιά ένα σύμβολο, από ένα πεπερασμένο αριθμό συμβόλων S, μπορεί να γραφεί. Η είσοδος ξεκινά από την αρχή (αριστερό άκρο της ταινίας) στο οποίο βρίσκεται η κεφαλή της μηχανής κατά την εκκίνηση. Η μηχανή λειτουργεί σαν πεπερασμένο αυτόματο: Έχει ένα πεπερασμένο αριθμό καταστάσεων Q q_0,q_1,..., όπου q_0 είναι η αρχική κατάσταση και $F \subseteq Q$ ένα σύνολο τερματικών καταστάσεων της μηχανής. Η συνάρτηση μετάβασης αναθέτει σε κάθε ζεύγος $(q, s) \in Q \times S$ μία τριάδα (q', s', d) όπου $q' \in Q$ είναι η επόμενη κατάσταση της μηχανής, $s' \in S$ είναι το σύμβολο που γράφει πάνω στο κελί στην θέση του s, και $d \in L, R$ είναι η κατεύθυνση κατά την οποία κινείται η κεφαλή πάνω στην ταινία. Η λειτουργία της μηχανής ξεκινά από την αρχική κατάσταση, ενώ η κεφαλή όπως αναφέραμε βρίσκεται στην αρχή της ταινίας. Η μηχανή σταματά εφόσον φθάσει σε κάποια τερματική κατάσταση ($\in F$). Τέλος εάν m μηχανή και x είσοδος τότε συμβολίζουμε με $m(x)$ την έξοδο του m εφόσον αυτό σταματά. Συγκεντρωτικά δηλαδή μπορούμε να πούμε ότι η μηχανή Turing αποτελείται από:

- Ένα σύνολο Q (το σύνολο των **καταστάσεων**)
- ένα στοιχείο $q_0 \in Q$ (η **αρχική κατάσταση**)
- ένα σύνολο **τερματικών καταστάσεων** $F \subseteq Q$
- Μια συνάρτηση $\tau : Q \times S \to Q \times S \times \{L, R\}$ η οποία αναθέτει σε κάθε ζεύγος $(q, s) \in Q \times S$ μία τριάδα (q', s', d) όπου $q' \in Q$ είναι η επόμενη κατάσταση της μηχανής, $s' \in S$ είναι το σύμβολο που γράφει πάνω στο κελί στην θέση του s, και $d \in \{L, R\}$ είναι η κατεύθυνση κατά την οποία κινείται η κεφαλή πάνω στην ταινία (η **συνάρτηση μετάβασης**)

Εδώ αξίζει να σημειώσουμε ότι το σύνολο των μηχανών Turing είναι αριθμήσιμο. Μάλιστα είναι αλγοριθμικά απαριθμήσιμο με την έννοια ότι μπορούμε να ορίσουμε ένα αλγόριθμο ο οποίος θα δίνει ως έξοδο την ακολουθία όλων των μηχανών Turing χωρίς επαναλήψεις.

Στην συνέχεια θα παρουσιάσουμε ορισμένα γνωστά αποτελέσματα της θεωρίας της υπολογισιμότητας (χωρίς να επιμένουμε σε ιδιαίτερα αυστηρές αποδείξεις) και θα εξετάσουμε πώς μπορούμε να αξιοποιήσουμε τα αποτελέσματα αυτά στο πλαίσιο της θεωρίας παιγνίων, όπως παρουσιάστηκαν από τους Binmore [3] και Anderlini [2].

6.2 Αποτελέσματα Υπολογισιμότητας

Ακολούθως παρουσιάζουμε λοιπόν διάφορα αποτελέσματα τα οποία είναι σημαντικά για την εφαρμογή περιορισμών υπολογισιμότητας σε παιγνιο-θεωρητικό πλαίσιο.

Πρόταση 6.1. *Υπάρχει μία μηχανή Turing που αναγνωρίζει τον εαυτό της*

Αναζητούμε μία μηχανή m^* που τερματίζει με έξοδο έστω C, όποτε η είσοδος είναι μία περιγραφή του εαυτού της και τερματίζει με διαφορετική έξοδο (έστω D) οπότε η είσοδος δεν είναι m^*. Είναι προφανές ότι η μηχανή δεν μπορεί να συγκρίνει την είσοδο με το m^* καθώς δεν δέχεται το m^* ως μέρος της εισόδου του. Ακόμη η μηχανή m^* δεν μπορεί να συγκρίνει την είσοδο με κάποιο κείμενο που είναι γραμμένο μέσα της επειδή η πλήρη περιγραφή της μηχανής θα περιέχει περισσότερα σύμβολα από το κείμενο αυτό. Επομένως χρειαζόμαστε κάποια καινοτόμο ιδέα. Πραγματικά ο παρακάτω αλγόριθμος πραγματοποιεί τη ζητούμενη διεργασία:

Εκτύπωσε C και σταμάτα εάν το παρακάτω γεγονός είναι αλήθεια και εκτύπωσε D και σταμάτα διαφορετικά: η είσοδος πριν και μετά την τρίτη εμφάνιση της λέξης "είναι" είναι Εκτύπωσε C και σταμάτα εάν το ακόλουθο γεγονός είναι αληθές και εκτύπωσε D και σταμάτα διαφορετικά: η είσοδος πριν και μετά την τρίτη εμφάνιση της λέξης "είναι".

Μία εναλλακτική περιγραφή του προγράμματος είναι η εξής:

1. Γράψε αυτό πάνω στην ταινία:

    - 2. Επιβεβαίωσε ότι η άλλη μηχανή έχει 4 εντολές, από τις οποίες οι τελευταίες 3 είναι ακριβώς ίδιες με το κείμενο πάνω στην ταινία.
    - 3. Έλενξε εάν η εντολή 1 ξεκινά με το κείμενο "Γράψε αυτό πάνω στην ταινία", ακολουθούμενο από το κείμενο πάνω στην ταινία.
    - 4. Εάν οι απαντήσεις και στις δύο εντολές 2 και 3 είναι θετικές τότε εκτύπωσε C, αλλιώς εκτύπωσε D.

2. Επιβεβαίωσε ότι η άλλη μηχανή έχει 4 εντολές, από τις οποίες οι τελευταίες 3 είναι ακριβώς ίδιες με το κείμενο πάνω στην ταινία.

3. Έλενξε εάν η εντολή 1 ξεκινά με το κείμενο "Γράψε αυτό πάνω στην ταινία", ακολουθούμενο από το κείμενο πάνω στην ταινία.

4. Εάν οι απαντήσεις και στις δύο εντολές 2 και 3 είναι θετικές τότε εκτύπωσε C, αλλιώς εκτύπωσε D.

Αναλύοντας προσεκτικά την λειτουργία των παραπάνω προγραμμάτων παρατηρούμε ότι πράγματι επιτυγχάνουν το επιθυμητό σκοπό. Η παρουσίαση του αυτομάτου που αναγνωρίζει τον ευατό του οφείλεται στον Howard [8].

Πρόταση 6.2. *Υπάρχει μία μηχανή Turing που μιμείται όλες τις άλλες*

Με άλλα λόγια υπάρχει μία μηχανή m^*, η οποία όποτε δέχεται ως είσοδο ένα ζεύγος (m, x), όπου m είναι η περιγραφή μιας μηχανής και x είναι μία τυχαία ακολουθία συμβόλων, τότε παράγει το ίδιο αποτέλεσμα με την μηχανή m όταν δέχεται ως είσοδο το x. Αναζητούμε δηλαδή μία μηχανή Turing m^* όπου $m^*(m, x)$ τερματίζει εάν και μόνο εάν η $m(x)$ τερματίζει και τότε $m^*(m, x) = m(x)$.

Μια τέτοια μηχανή m^* υπάρχει και λέγεται καθολική μηχανή Turing (universal Turing machine). Εδώ αξίζει να σημειωθεί η βασική ιδέα για την κατασκευή της καθολικής μηχανής Turing είναι ότι η UTM θα διαβάσει την περιγραφή της m και θα χρησιμοποιήσει τον αλγόριθμο που περιγράφει για να υπολογίσει την έξοδό της με είσοδο το x.

Πρόταση 6.3. *Υπάρχει μία μηχανή Turing που προβλέπει την έξοδο οποιασδήποτε μηχανής Turing που παίζει εναντίον της.*

Η καθολική μηχανή Turing m^*, όπως περιγράφθηκε παραπάνω, μπορεί να υπολογίσει για κάθε m, την έξοδο $m(m^*)$ εφόσον λαμβάνει ως είσοδο την περιγραφή του m^*. Τώρα θα δούμε ότι υπάρχει μία μηχανή που μπορεί να υπολογίσει την απόκριση οποιασδήποτε μηχανής στην δική της περιγραφή χωρίς να δέχεται την περιγραφή της ως είσοδο. Με άλλα λόγια ενδιαφερόμαστε για την κατασκευή μίας μηχανής m^* τέτοια ώστε για οποιαδήποτε μηχανή m θα σταματά εάν και μόνο εάν η $m(m^*)$ σταματά και θα δίνει το ίδιο αποτέλεσμα ως έξοδο. Η μηχανή δέχεται το m ως είσοδο. Στην συνέχεια διατρέχει την λίστα των μηχανών (αφού είναι αλγοριθμικά απαριθμητές) μέχρι (χρησιμοποιώντας την ιδέα του πρώτου αποτελέσματος) να αναγνωρίσει τον εαυτό της m^*. Τελικά υλοποιεί την καθολική μηχανή Turing προκειμένου να δώσει την έξοδο της μηχανής της εισόδου m όταν δέχεται είσοδο m^*.

Πρόταση 6.4. *Ύπαρξη μίας μηχανής που δεν προβλέπεται ποτέ σωστά*

Αναζητούμε μία μηχανή m' που θα έχει την δυνατότητα όποτε το $m(m')$ σταματά, τότε το $m'(m)$ θα σταματά με διαφορετική έξοδο. Η μηχανή θα χρησιμοποιεί την απαρίθμηση όλων των μηχανών. Η μηχανή m' διατρέχει όλες τις μηχανές μέχρι να αναγνωρίσει τον εαυτό της, και στην συνέχεια χρησιμοποιεί την καθολική μηχανή Turing m^* με είσοδο (m,m') προκειμένου να προβλέψει το $m'(m)$. Εάν ο υπολογισμός $m'(m)$ φθάσει σε μία τερματική κατάσταση τότε θα τερματίσει αλλά όχι πριν αλλάξει την έξοδο.

6.3 Υπάρχει λογικός παίκτης;

Τώρα είμαστε έτοιμοι να εφαρμόσουμε τα αποτελέσματα της προηγούμενης ενότητας και τα οποία προέρχονται από την θεωρία πολυπλοκότητας σε ένα παιγνιο-θεωρητικό πλαίσιο. Ανακεφαλαιώνοντας σκοπός του συγκεκριμένου κεφαλαίου είναι να αναγνωρίσουμε τυχόν όρια στην προσπάθειά μας να προσεγγίσουμε την συμπεριφορά ”πραγματικών παικτών”, τα οποία οφείλονται σε περιορισμούς υπολογισιμότητας. Κατά την προσπάθειά μας αυτή θα επικεντρωθούμε σε στρατηγικά παίγνια δύο παικτών. Μάλιστα θεωρούμε ότι η είσοδος κάθε μηχανής Turing είναι η περιγραφή της μηχανής του άλλου παίκτη, ενώ η έξοδος της μηχανής καθορίζει την κίνηση του παίκτη.

Θα μπορούσαμε να ισχυριστούμε (αφελώς), ότι προκειμένου κάποιος παίκτης (ο πρώτος) να φέρεται ”λογικά” πρέπει πρώτα να υπολογίσει την έξοδο της μηχανής του δεύτερου παίκτη και μετά να υπολογίσει την βέλτιστη απόκριση στην ενέργεια του παίκτη αυτού. Όμως, αν ο δεύτερος παίκτης πραγματοποιήσει το ίδιο τότε εισερχόμαστε σε ένα ατέρμονο βρόχο. Μπορεί όμως οι παίκτες να καταλήγουν στις στρατηγικές τους με κάποιο άλλο τρόπο, αφού δέχεται ολόκληρη την περιγραφή του άλλου παίκτη ως είσοδο.

Εδώ όμως οφείλουμε να διευκρινίσουμε μία κρίσιμη λεπτομέρεια. Όπως είπαμε μία μηχανή Turing μπορεί να μην τερματίζει η μπορεί να τερματίζει σε μία έξοδο που δεν αντιστοιχεί σε κάποια από τις δυνατές κινήσεις του παίκτη. Αυτό δημιουργεί ένα πρόβλημα καθώς το παίγνιο απαιτεί κάθε παίκτης να πραγματοποιεί μία κίνηση. όμως το κλασσικό μοντέλο του παιγνίου δεν καθορίζει κάποια εναλλακτική (default) κίνηση. Έτσι πρέπει να εμπλουτίσουμε το μοντέλο και να καθορίσουμε την κίνηση που πρέπει να πραγματοποιεί ο παίκτης στην περίπτωση που η μηχανή του δεν τερματίζει ή εάν τερματίζει δίνει κάποια έξοδο που δεν συμφωνεί με τις πιθανές κινήσεις.

Έτσι λοιπόν από εδώ και πέρα υποθέτουμε ότι κάθε παίκτης i έχει μία ενέργεια d_i, η οποία αποτελεί την default κίνηση του. Έτσι οποτεδήποτε μία μηχανή δεν παράγει κάποιο ”επιθυμητό” αποτέλεσμα, τότε ο παίκτης θα παίζει d_i. Αυτό βέβαια αποτελεί μία αυθαίρετη υπόθεση. Η ανάγκη της παραδοχής μίας τέτοιας υπόθεσης επιδεικνύει ότι το μοντέλο των παιγνίων δεν περιέχει λεπτομέρειες που είναι πρωταρχικής σημασίας σε σχέση με την φάση λήψη απόφασης κατά την ανάλυση.

Τα αποτελέσματα που περιγράψαμε κατά την πρώτη ενότητα δείχνουν ότι η ύπαρξη λογικού παίκτη εξαρτάται σε μεγάλο βαθμό από το υπό εξέταση παίγνιο. Έτσι εάν το παίγνιο έχει κάποια κυρίαρχη στρατηγική, τότε ο ”λογικός” παίκτης υφίσταται. Σε άλλα λιγότερο ”εύκολα” παίγνια η ύπαρξη λογικού παίκτη μπορεί να εξαρτάται επίσης από την επιλογή των default κινήσεων.

Έτσι λοιπόν στην Πάλη των Φύλων εάν η default κίνηση και των δύο παικτών είναι S, τότε η μηχανή m^* που περιγράφεται στην πρόταση 3, έχει την ιδιότητα ότι για οποιαδήποτε μηχανή m, παράγει μία ενέργεια που αποτελεί βέλτιστη απάντηση. Σταματά με έξοδο B εάν και μόνο εάν η άλλη μηχανή σταματά με B, ενώ σταματά με έξοδο S εάν το $m^*(m)$ σταματά με έξοδο S. Σε περίπτωση που το $m(m^*)$ δεν σταματά ή σταματά με έξοδο που δεν αντιστοιχεί σε ενέργεια, τότε το $m^*(m)$ κάνει το ίδιο και τα δύο παίζουν S.

Η γενίκευση του παραπάνω επιχειρήματος είναι η ακόλουθη παρατήρηση: Έστω G ένα στρατηγικό παίγνιο δύο παικτών με default ενέργειες d_1,d_2 τέτοιες ώστε το (d_1, d_2) να είναι ισορροπία Nash, τότε κάθε παίκτης i έχει μία μηχανή m_i^* τέτοια ώστε για κάθε μηχανή m_j, η ενέργεια που προκύπτει από τη $m_i^*(m_j)$ να αποτελεί απόκριση στην ενέργεια της $m_j^*(m_i)$.

Όμως μπορούμε να δούμε ότι στην περίπτωση της μάχης των Φύλων με Διαφορετικές (default) κινήσεις δεν υπάρχει μηχανή που υπολογίζει την βέλτιστη απόκριση απέναντι σε όλες τις πιθανές μηχανές του αντιπάλου, ενώ αντίθετα όταν οι δύο default κινήσεις συμπίπτουν τότε η βέλτιστη αυτή μηχανή υπάρχει. Μπορούμε εύκολα να καταλήξουμε σε άλλα τέτοια "κακά" παραδείγματα με την χρήση της μηχανής που περιγράφεται στην τελευταία από τις παραπάνω προτάσεις.

Σε κάθε περίπτωση πάντως η ύπαρξη ή όχι μιας τέτοιας μηχανής δεν μπορεί να εκφράσει πλήρως τους περιορισμούς που συναντάμε κατά την προσπάθεία μας να προσεγγίσουμε την συμπεριφορά του πραγματικού παίκτη. Οι απλούστεροι περιορισμοί του χρόνου, του μεγέθους της μηχανής και της ικανότητας να πραγματοποιεί παρατηρήσεις και υπολογισμούς όπως είδαμε αναλυτικά και στο προηγούμενο κεφάλαιο τις ουσιαστικότερες δυσκολίες κατά την προσέγγισή μας αυτή.

Βιβλιογραφία

[1] Abreu, D., and A. Rubinstein. (1988). "The Structure of Nash Equilibrium in Repeated Games with Finite Automata." *Econometrica* 56, 1259-1282.

[2] Anderlini, L. (1990). "Some Notes on Church's Thesis and the Theory of Games." Theory and Decision 29, 15-52.

[3] Binmore, K. (1987). "Modeling Rational Players, I." *Economics and Philosophy* 3, 179-214.

[4] Binmore, K. (1988). "Modeling Rational Players, II." *Economics and Philosophy* 4, 9-55.

[5] Derman, C. (1970). Finite State Markovian Decision Processes. New York: Academic Press.

[6] Friedman, J. W. (1971) "Repeated Games of Incomplete Information: Non-Zero-Sum". *Handbook of Game Theory and Economic Applications* Vol.1, 155-177.

[7] Hopcroft, J. E., and J. D. Ullman. (1979). *Introduction to Automata Theory: Languages and Computation*. Reading, Mass.: Addison Wesley

[8] Howard, J. (1988). "Cooperation in the Prisoner's Dilemma." *Theory and Decision* 24, 203-213.

[9] Luce, R. D., and H. Raiffa. (1957). *Games and Decisions*. New York: John Wiley.

[10] Neyman, A. (1985) "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma." *Economic Letters* 19, 227-229.

[11] Neyman, A. (1998) "Finitely Repeated Games with Finite Automata." *Mathematics of Operation Research* 3, 513-552.

[12] Osbourne, M., and A. Rubinstein. (1994). A course in Game Theory. Cambridge, Mass.: MIT Press.

[13] Papadimitriou C. H. (1992). "On Players with a Bounded Number of States." *Games and Economic Behavior* 4, 122-131

[14] Papadimitriou C. H., and Yannakakis M. (1994). "On Complexity as Bounded Rationality." *Proc. 26th ACM Symposium on the Theory of Computing* 726-733

[15] Papadimitriou C. H., and Yannakakis M. (1996). "On Bounded Rationality and Complexity." manuscript.

[16] Piccione, M. (1992). "Finite Automata with Discounting." *Journal of Economic Theory* 56, 180-193.

[17] Piccione, M., and A. Rubinstein. (1993). "Finite Automata Play a Repeated Extensive Game." *Journal of Economic Theory* 61, 160-168.

[18] Radner, R. (1986). "Can Bounded Rationality Resolve the Prisoner's Dilemma?" In A. Mas-Colell and W. Hildenbrand, eds. Essays in Honor of Gerard Debreu, 387-399. Amsterdam: North-Holland

[19] Rosenthal, R. (1982). "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox." *Journal of Economic Theory* 25, 92-100.

[20] Rubinstein, A. (1986). "Finite Automata Play a Repeated Prisoner's Dilemma" *Journal of Economic Theory* 39, 83-96.

[21] Rubinstein, A. (1987). "Is There a 'Finite Automata' Resolution to the Finitely Repeated Games Paradoxes?" Mimeo.

[22] A. Rubinstein. (1998). Modeling Bounded Rationality. Cambridge, Mass.: MIT Press.

[23] Simon, H. A. (1996). The Sciences of the Artificial. Third Edition. Cambridge, Mass.: MIT Press.

[24] Shubik, M. (1959). Strategy and Market Structure. New York: Wiley.

Document powered by $\LaTeX$