ΑΛΕΞΑΝΔΡΕΙΟ Τ.Ε.Ι. ΘΕΣΣΑΛΟΝΙΚΗΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πτυχιακή εργασία

# «Μάθηση και λήψη αποφάσεων στο παιχνίδι πέτρα, ψαλίδι, χαρτί»

Του φοιτητή
Τέντσου Παναγιώτη
Αρ. μητρώου: 01 16/84

Επιβλέπων καθηγητής
Παναγιώτης Αδαμίδης

Θεσσαλονίκη 2010

## *Περίληψη*

Σκοπός της πτυχιακής αυτής είναι η ανάδειξη της επίδρασης νίκης με διαφορετική αξία κέρδους στην ανάπτυξη στρατηγικής στο παίγνιο πέτρα ψαλίδι χαρτί. Αναπτύσσεται μια εφαρμογή, ένας πράκτορας μηχανικής μάθησης, η οποία έχει ως σκοπό να εντοπίζει την συμπεριφορά των παικτών όταν αυτή δεν είναι τυχαία και καθώς το παιχνίδι συνεχίζεται για αρκετούς γύρους να εκμεταλλεύεται την συμπεριφορά αυτή προς όφελος του. Η προσέγγιση του συγκεκριμένου θέματος γίνεται από την πλευρά της θεωρίας παιγνίων και της μηχανικής μάθησης χρησιμοποιώντας τεχνικές και από τις δύο αυτές επιστήμες για να επιτευχθεί το επιθυμητό αποτέλεσμα. Αρχικά περιγράφονται τα διάφορα είδη των παιγνίων και οι λύσεις που προσφέρει η θεωρία παιγνίων, σχετικά πάντα με το συγκεκριμένο παίγνιο που εξετάζεται. Στην συνέχεια γίνεται αναφορά στην μηχανική μάθηση και στους διαφόρους τρόπους, αλγορίθμους, με τους οποίους προσεγγίζει αυτή το πρόβλημα της μάθησης. Προχωρώντας γίνεται ανάλυση του παιγνίου πέτρα ψαλίδι χαρτί και εξηγείται η δυναμική του. Στην συνέχεια αναλύεται η εφαρμογή και εξηγούνται οι αλγόριθμοι που αναπτύχθηκαν για τον εντοπισμό της μη τυχαίας συμπεριφοράς του αντιπάλου και η τεχνικές επιλογής κίνησης από τον πράκτορα. Στο τέλος της πτυχιακής αυτής αναφέρονται τα αποτελέσματα της εφαρμογής που αναπτύχτηκε, εναντίον του παίκτη αλλά και μεταξύ των ίδιων των αλγορίθμων, καθώς και τα συμπεράσματα που προέκυψαν από την ενασχόληση με το θέμα αυτό.

## *Abstract*

Purpose of this paper is to highlight the impact in winning when different gain values are used while developing strategies in the game rock paper scissors. The application developed, a machine learning agent, aims to identify the players behavior when it is not random and as the game continues for several rounds to exploit this behavior to his own benefit. The approach used in this attempt is both from the side of game theory and also from that of machine learning, using techniques from both disciplines to achieve the desired result. Initially the different types of games and solutions offered by game theory are described, always concerning the game examined. Onwards the ways, algorithms, in which machine learning approaches the problem of learning are described. Moving on follows an analysis of the game rock paper scissors and its' dynamics explained. As this paper continues the application and the algorithms developed to detect the non random behavior of the opponent and the techniques used for selecting moves by the agent are analyzed and explained. At the end of this paper the results of this application are stated, both against the player and between the algorithms themselves, and also the conclusions drawn from dealing with this subject.

# Περιεχόμενα

# Ευρετήριο σχημάτων



# Ευρετήριο παραδειγμάτων



# Ευρετήριο πινάκων

# Κεφάλαιο 1

## ΕΙΣΑΓΩΓΗ

# 1. **Εισαγωγή**

Η θεωρία παιγνίων προσφέρει τα εργαλεία με τα οποία μπορεί κάποιος να αποφασίσει όταν δεν ξέρει τι να διαλέξει σε κάποια διαδικασία αποφάσεων όπως είναι ένα παιχνίδι, αλλά δεν περιορίζεται μόνο σε παίγνια. Τα παίγνια περιγράφονται με διάφορα χαρακτηριστικά και ανάλογα με αυτά η θεωρία παιγνίων προτείνει διάφορες λύσεις για αυτά. Ένα από τα παίγνια που έχει εξετάσει η θεωρία παιγνίων είναι και το πέτρα ψαλίδι χαρτί.

Η μηχανική μάθηση είναι η επιστήμη που μελετάει την κατασκευή συστημάτων υπολογιστών τα οποία βελτιώνονται με την εμπειρία. Με την χρήση αλγορίθμων και τεχνικών που βασίζονται στην μηχανική μάθηση σχεδιάζονται πράκτορες, προγράμματα, τα οποία με βάση την ανάλογη εμπειρία από το περιβάλλον στο οποίο κινείται ο πράκτορας βελτιώνει την απόδοσή του στην ανάλογη διεργασία.

Με τον συνδυασμό αυτών των δύο στην πτυχιακή αυτή αναπτύσσεται μια εφαρμογή η οποία με βάση το παίγνιο πέτρα ψαλίδι χαρτί, μια διαφοροποιημένη εκδοχή του με διαφορετικές αποδόσεις για κάθε κίνηση για την ακρίβεια, εξετάζει τις δυνατότητες μάθησης και προσαρμογής του συγκεκριμένου πράκτορα σε μη τυχαίες συμπεριφορές του αντιπάλου του. Η ανάπτυξη της εφαρμογής γίνεται πάντα στα πλαίσια της δυναμικής και τον συνθηκών που μπορούνε να επικρατήσουν στο συγκεκριμένο παίγνιο. Σκοπός της εφαρμογής δεν είναι η δημιουργία ενός πράκτορα που θα κερδίζει τον χρήστη, καθώς στο συγκεκριμένο παίγνιο ο παράγοντας τύχη παίζει μεγάλο ρόλο, αλλά η προσαρμογή του πράκτορα στις μη τυχαίες συμπεριφορές και η εκπαίδευση του από τις εμπειρίες αυτές, στον βαθμό δυνατό.

Στην πορεία της ανάπτυξης του πράκτορα αναπτύχθηκαν διάφοροι αλγόριθμοι και τεχνικές, άλλοι βασισμένοι στην θεωρία παιγνίων και σε έτοιμους αλγορίθμους της μηχανικής μάθησης και άλλοι από το μηδέν. Κύρια ενασχόληση της πτυχιακής αυτής ήταν η ανάπτυξη ενός πράκτορα που θα εντοπίζει τα μοτίβα στις κινήσεις του αντιπάλου και θα απαντάει ανάλογα σε αυτά.

Στα τελευταία κεφάλαια αναφέρονται τα αποτελέσματα της πτυχιακής, όχι μόνο ενάντια στον χρήστη αλλά και η σύγκριση των αναπτυχθέντων αλγορίθμων και τεχνικών μεταξύ τους. Δίνονται επίσης παρατηρήσεις που προέκυψαν κατά την διάρκεια της ανάπτυξης της εργασίας αυτής καθώς και συμπεράσματα και προτάσεις για περαιτέρω έρευνα πάνω στο θέμα αυτό.

# Κεφάλαιο 2

## ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ

Στο κεφάλαιο αυτό αναπτύσσονται οι βασικές έννοιες της θεωρίας παιγνίων και το τι πραγματεύεται αυτή. Παρουσιάζονται τα διάφορα είδη παιγνίων καθώς και οι διάφορες λύσεις ή έννοιες λύσεων που χρησιμοποιούνται από την θεωρία παιγνίων για την λύση τους. Γίνεται αναφορά στα διάφορα παίγνια που κατά καιρούς έχουνε εξεταστεί από αυτήν και σε ορισμένα κλασικά παραδείγματα της. Σκοπός του κεφαλαίου αυτού είναι να καλύψει το γενικό πλαίσιο της θεωρίας παιγνίων αλλά πάντα υπό το πρίσμα της συγκεκριμένης πτυχιακής εργασίας και του συγκεκριμένου παραδείγματος της.

# 2. Θεωρία παιγνίων

## *2.1 Ιστορική αναφορά*

Η πρώτη γνωστή αναφορά στην θεωρία παιγνίων γίνεται σε ένα γράμμα του James Waldegrave το 1713. Σε αυτό το γράμμα ο Waldegrave δίνει μια λύση ελαχιστοποίησης της μέγιστης ζημιάς μικτής στρατηγικής (minimax mixed strategy solution) για μια εκδοχή για δύο παίχτες του παιχνιδιού με χαρτιά le Her [19]. Η πρώτη φορά που επιδιώχθηκε μια γενική ανάλυση θεωρίας παιγνίων ήταν με την έκδοση του *Recherches sur les principes mathématiques de la théorie des richesses* (Έρευνες για τις μαθηματικές αρχές της θεωρίας του πλούτου) του Antoine Augustin Cournot το 1838. Σε αυτό ο Cournot θεωρεί ένα δυοπώλιο και παρουσιάζει μια λύση που είναι μια περιορισμένη εκδοχή της ισορροπίας Nash (Nash equilibrium) [8].

Παρόλο που η ανάλυση του Cournot είναι πιο γενική από αυτήν του Waldegrave η θεωρία παιγνίων ουσιαστικά δεν υπήρχε σαν ξεχωριστό πεδίο έρευνας μέχρις ότου ο John von Neumann εξέδωσε μια σειρά από έγγραφα το 1928. Μπορεί ο γάλλος μαθηματικός Emile Borel να ασχολήθηκε νωρίτερα με τα παίγνια ως ένα σημείο αλλά δίκαια χαρακτηρίζεται σαν δημιουργός της θεωρίας παιγνίων ο John von Neumann [3]. Ο Γεννημένος στην Ουγγαρία Αμερικανός μαθηματικός και ο συνάδελφος του στο πανεπιστήμιο του Princeton Oskar Morgenstern, γεννημένος στην Γερμανία Αμερικανός οικονομολόγος, ανέπτυξαν την θεωρία παιγνίων για να λύσουνε οικονομικά προβλήματα [48]. Ο Von Neumann ήταν ένας ιδιοφυής μαθηματικός, η δουλεία του οποίου εκτείνεται από την θεωρία συνόλων μέχρι τους υπολογισμούς του κλειδί για την ανάπτυξη της ατομικής βόμβας και της βόμβας υδρογόνου αλλά και της ανάπτυξης των υπολογιστών. Το αποκορύφωμα της δουλείας του Von Neumann όσον αφορά την θεωρία παιγνίων έγινε το 1944 με το βιβλίο του *The Theory of Games and Economic behavior* που έγραψε μαζί με τον Oskar Morgenstern [3]. Ο von Neumann και ο Morgenstern ισχυρίστηκαν ότι τα μαθηματικά που αναπτύχθηκαν για τις φυσικές επιστήμες, που περιγράφουν το πώς δουλεύει μια αμερόληπτη φύση, ήτανε ένα φτωχό μοντέλο για τα οικονομικά. Παρατήρησαν ότι τα οικονομικά μοιάζουνε πολύ με ένα παιχνίδι, στο οποίο οι παίχτες προσπαθούν να προβλέψουν ο ένας τις κινήσεις του άλλου, και συνεπώς απαιτείται ένα νέο είδος μαθηματικών, το οποίο και ονόμασαν θεωρία παιγνίων. Το όνομα φαίνεται να είναι λίγο εσφαλμένος χαρακτηρισμός καθώς η θεωρία παιγνίων γενικά δεν χαρακτηρίζεται από την διασκέδαση και την ελαφρότητα που συνήθως αποδίδεται στα παιχνίδια [48]. Σε αυτήν την περίοδο η δουλειά πάνω στην θεωρία παιγνίων αφορά κυρίως την θεωρία παιγνίων συνεργασίας, η οποία αναλύει τις βέλτιστες στρατηγικές για ομάδες ατόμων, υποθέτοντας ότι μπορούν να επιβληθούν οι μεταξύ τους συμφωνίες για τις κατάλληλες στρατηγικές [3].

Το 1950 εμφανίστηκε η πρώτη συζήτηση για το δίλημμα του κρατούμενου (prisoners dilemma), ένα είδος παιχνιδιού που αποτελεί ακόμα και σήμερα ίσως το πιο χαρακτηριστικό παράδειγμά της, και διεξήχθη ένα πείραμα πάνω σε αυτό στην εταιρία RAND, μια μη κερδοσκοπική εταιρία για έρευνα και ανάλυση. Την ίδια εποχή ο John Nash ανέπτυξε ένα κριτήριο για την αμοιβαία συνέπεια των στρατηγικών των παιχτών, γνωστή σαν ισορροπία Nash (Nash equilibrium), εφαρμόσιμη σε ένα ευρύτερο φάσμα παιγνίων από το κριτήριο που είχαν προτείνει οι Von Neumann και Morgenstern. Αυτή η ισορροπία είναι επαρκώς γενική ώστε να επιτρέπει την ανάλυση και παιγνίων μη συνεργασίας πέρα από τα συνεργασίας [29] [30]. Η θεωρία παιγνίων παρουσίασε μια άνθιση την δεκαετία του 50, κατά την οποία αναπτύχθηκαν οι έννοιες του πυρήνα (Core), τις εκτεταμένης μορφής παιγνίου (extensive form game), του φανταστικού παιχνιδιού (fictitious play), του επαναλαμβανόμενου παιγνίου (repeated game), και τις τιμής Shapley (Shapley value). Επιπρόσθετα εμφανίστηκαν οι πρώτες εφαρμογές της θεωρίας παιγνίων στην φιλοσοφία και στις πολιτικές επιστήμες την περίοδο αυτή [34].

Το 1965 ο Reinhart Selten εισήγαγε το σκεπτικό λύσης του για την υποπαιγνιακή τέλεια ισορροπία (subgame perfect equilibria), το οποίο βελτίωσε περαιτέρω την ισορροπία Nash (Nash equilibrium) (αργότερα επίσης θα εισήγαγε και την τέλεια ισορροπία τρεμάμενου χεριού (trembling hand perfection)) [39]. Το 1967 ο John Harsanyi ανέπτυξε τις έννοιες της πλήρης πληροφόρησης (complete information) και του Μπεϊζιανού παιγνίου (Bayesian game) [15]. Το 1994 οι Nash, Selten και Harsanyi βραβεύθηκαν με το βραβείο Νόμπελ οικονομικών για την προσφορά τους στην οικονομική θεωρία παιγνίων. Την δεκαετία του 1970 η θεωρία παιγνίων εφαρμόστηκε εκτενέστατα στην βιολογία, κυρίως λόγο της δουλείας του John Maynard Smith και της εξελικτικά σταθερής στρατηγικής του ( evolutionary stable strategy) [27]. Επιπρόσθετα την δεκαετία αυτή εισήχθηκαν και αναλύθηκαν οι έννοιες της συσχετισμένης ισορροπίας (correlated equilibrium), της τέλειας ισορροπίας τρεμάμενου χεριού (trembling hand perfection) και της κοινής γνώσης (common knowledge). Το 2005 οι εκφραστές της θεωρίας παιγνίων Thomas Schelling και Robert Aumann ακολούθησαν τους Nash, Selten και Harsanyi και βραβεύτηκαν και αυτοί με τα βραβείο νόμπελ οικονομικών. O Schelling εργάστηκε πάνω στα δυναμικά μοντέλα, πρώιμα παραδείγματα της εξελικτικής θεωρίας παιγνίων ( evolutionary game theory) . O Aumann συνείσφερε περισσότερο στις ισορροπίες εισάγοντας μια συσχετισμένη ισορροπία και αναπτύσσοντας μια εκτεταμένη επίσημη ανάλυση της υπόθεσης της κοινής γνώσης (common knowledge) και των συνεπειών της [31]. Το 2007 ο Roger Myerson μαζί με τον Leonid Hurwicz και τον Eric Maskin βραβεύτηκαν με το βραβείο νόμπελ οικονομικών «για την εισαγωγή των θεμελίων της θεωρίας της μηχανικής σχεδίασης» (mechanism design). Η συνεισφορά του Myerson συμπεριλαμβάνει την έννοια της γνήσιας ισορροπίας (proper equilibrium) και ένα σημαντικό κείμενο το: Game Theory, Analysis Of Conflict ( Myerson 1997). Οχτώ συνολικά εκφραστές της θεωρίας παιγνίων έχουνε βραβευτεί με το νόμπελ οικονομικών και ο John Maynard Smith βραβεύτηκε με το βραβείο Crafoord για την εφαρμογή της θεωρίας παιγνίων στην βιολογία.

## *2.2 Τι είναι η θεωρία παιγνίων*

Η θεωρία παιγνίων είναι μια διακλάδωση των εφαρμοσμένων μαθηματικών που χρησιμοποιείται στις κοινωνικές επιστήμες, κυρίως στα οικονομικά, καθώς και στην βιολογία, την μηχανική, τις πολιτικές επιστήμες, της διεθνής σχέσεις, την πληροφορική και την φιλοσοφία [5]. Η θεωρία παιγνίων είναι ένα σύνολο από αναλυτικά εργαλεία σχεδιασμένα να μας βοηθήσουνε να καταλάβουμε τα φαινόμενα που παρατηρούμε όταν αλληλεπιδρούν άτομα που παίρνουν αποφάσεις [25]. Τα εργαλεία αυτά χρησιμοποιούνται για την ανάλυση καταστάσεων στις οποίες οι συμμετέχοντες, που αποκαλούνται παίκτες, παίρνουν αποφάσεις που είναι αλληλοεξαρτούμενες. Αυτή η αλληλεξάρτηση κάνει τον κάθε παίχτη να λαμβάνει υπόψη τις πιθανές αποφάσεις, ή στρατηγικές, των άλλων παιχτών για να διαμορφώσει την δικιά του στρατηγική [48]. Οι βασικές παραδοχές στις οποίες βασίζεται η θεωρία παιγνίων είναι ότι τα άτομα που παίρνουν τις αποφάσεις επιδιώκουν σαφώς καθορισμένους εξωγενείς στόχους (είναι ορθολογικοί) και λαμβάνουν υπόψη την γνώση ή τις προσδοκίες της συμπεριφοράς των άλλων ατόμων που παίρνουν αποφάσεις (κρίνουν στρατηγικά δηλαδή) [25]. Η θεωρία παιγνίων επιδιώκει να συλλάβει μαθηματικά την συμπεριφορά σε *στρατηγικές καταστάσεις*, στις οποίες η επιτυχία των επιλογών ενός ατόμου εξαρτάται από τις επιλογές των άλλων. Μολονότι αρχικά η θεωρία παιγνίων είχε αναπτυχθεί για να αναλύσει διαγωνισμούς στους οποίους ένα άτομο έχει κέρδος σε βάρος ενός άλλου (παίγνια μηδενικού αθροίσματος, zero sum games), έχει επεκταθεί για να μεταχειρίζεται μια μεγάλη γκάμα αλληλεπιδράσεων, οι οποίες κατατάσσονται σύμφωνα με διάφορα κριτήρια [12].

Μια λύση σε ένα παίγνιο περιγράφει τις βέλτιστες αποφάσεις των παικτών, που μπορούνε να έχουν παρόμοια, αντίθετα ή μικτά ενδιαφέροντα, και τα αποτελέσματα που μπορούνε να προκύψουν από αυτές τις αποφάσεις [48]. Παραδοσιακές εφαρμογές της θεωρίας παιγνίων προσπαθούν να βρούνε ισορροπίες (equilibrium), ισορροπία είναι η κατάσταση ενός συστήματος στο οποίο εξισορροπούνται οι ανταγωνιστικές επιδράσεις, σε αυτά τα παίγνια. Σε μια ισορροπία κάθε παίκτης του παιχνιδιού έχει υιοθετήσει μια στρατηγική που είναι απίθανο να την αλλάξει. Πολλές έννοιες της ισορροπίας έχουν αναπτυχθεί, με ποιο διάσημη αυτή του Nash (Nash equilibrium), σε μια προσπάθεια να συλλάβουν αυτήν την ιδέα. Αυτές οι έννοιες της ισορροπίας προωθούνται διαφορετικά ανάλογα με το πεδίο εφαρμογής τους, παρόλο που συχνά επικαλύπτονται ή συμπίπτουν. Αυτή η μεθοδολογία έχει και τους πολέμιους της βέβαια και οι διαφωνίες συνεχίζονται για την καταλληλότητα συγκεκριμένων εννοιών ισορροπίας, των ισορροπιών σαν σύνολο ίσως, και την χρησιμότητα των μαθηματικών μοντέλων πιο γενικά [12].

Σήμερα, «η θεωρία παιγνίων είναι ένα είδος ομπρέλας ή ‘ενοποιημένου πεδίου θεωρίας για την λογική πλευρά της κοινωνικής επιστήμης, όπου το κοινωνική ερμηνεύεται γενικά ώστε να συμπεριλαμβάνει ανθρώπινους αλλά και μη ανθρώπινους παίκτες (υπολογιστές, ζώα, φυτά)» [5]. Θα ήτανε έκπληξη αν μία και μόνο θεωρία θα μπορούσε να απευθυνθεί σε ένα τόσο μεγάλο εύρος παιγνίων και στην πραγματικότητα δεν υπάρχει μία μόνο θεωρία παιγνίων. Ένα σύνολο θεωριών έχουνε προταθεί, κάθε μία από τις οποίες είναι εφαρμόσιμη σε διαφορετικές περιπτώσεις και με την δική της έννοια για το τι καθιστά λύση στην ανάλογη περίπτωση [48].

## *2.3 Εφαρμογές της θεωρίας παιγνίων*

Τα μοντέλα της θεωρίας παιγνίων είναι πολύ αφηρημένες αναπαραστάσεις κατηγοριών πραγματικών καταστάσεων. Η γενικότητά τους, τους επιτρέπει να μπορούν να χρησιμοποιηθούν για να μελετήσουν ένα μεγάλο εύρος φαινομένων. Για παράδειγμα, η θεωρία τις ισορροπίας Nash (Nash equilibrium) έχει χρησιμοποιηθεί για να μελετήσει ολιγοπωλιακό και πολιτικό ανταγωνισμό. Η θεωρία τις μικτής στρατηγικής ισορροπίας (mixed strategy equilibrium) έχει χρησιμοποιηθεί για να εξηγήσει τις διανομές των μηκών των γλωσσών στις μέλισσες και το μήκος του σωλήνα, της στεφάνης, στα άνθη. Η θεωρία των επαναλαμβανομένων παιγνίων (repeated games) έχει χρησιμοποιηθεί για να διαφωτίσει κοινωνικά φαινόμενα όπως είναι οι απειλές και οι υποσχέσεις. Η θεωρία του πυρήνα (Core) αποκαλύπτει μια έννοια κατά την οποία το αποτέλεσμα των συναλλαγών κάτω από ένα σύστημα τιμών είναι σταθερό σε μια οικονομία που περιέχει πολλά αυτόβουλα υποκείμενα, δρώντες [25].

Η θεωρία παιγνίων έχει εφαρμοστεί σε ένα μεγάλο εύρος περιπτώσεων στις οποίες οι επιλογές των παικτών αλληλεπιδρούν για να επηρεάσουν το αποτέλεσμα. Τονίζοντας τις στρατηγικές πτυχές της επιλογής αποφάσεων, ή της πτυχές που ελέγχονται από τους παίκτες αντί για την καθαρή τύχη, η θεωρία παιγνίων συμπληρώνει αλλά και πηγαίνει περαιτέρω από την θεωρία των πιθανοτήτων. Έχει χρησιμοποιηθεί για παράδειγμα για να προσδιορίσει τις πολιτικές συμμαχίες ή επιχειρηματικούς ομίλους που είναι πιθανό να σχηματιστούν, για την εύρεση της βέλτιστης τιμής πώλησης κάποιου προϊόντος ή υπηρεσίας απέναντι στον ανταγωνισμό, του υπολογισμού της δύναμης ενός ψηφοφόρου ή μιας ομάδας ψηφοφόρων, για το ποιος θα επιλεχθεί για ένορκος, για την εύρεση της καλύτερης τοποθεσίας για την δημιουργία μια κατασκευαστικής μονάδας αλλά και τον προσδιορισμό της συμπεριφοράς κάποιων ζώων και φυτών στην μάχη τους για την επιβίωση. Έχει ακόμα χρησιμοποιηθεί και για να προκαλέσει την νομιμότητα ορισμένων συστημάτων ψηφοφορίας [47].

Η θεωρία παιγνίων έχει χρησιμοποιηθεί για να μελετήσει μια πληθώρα ανθρώπινων αλλά και ζωικών συμπεριφορών. Αρχικά αναπτύχθηκε για τα οικονομικά, για να κατανοήσει ένα μεγάλο σύνολο οικονομικών συμπεριφορών, συμπεριλαμβανομένων και τον συμπεριφορών εταιριών, αγορών και καταναλωτών [26]. Η χρήση της στις κοινωνικές επιστήμες έχει επεκταθεί και έχει εφαρμοστεί και σε πολιτικές, κοινωνικές και ψυχολογικές συμπεριφορές.

Πέρα από το να χρησιμοποιείται για να προβλέψει και να εξηγήσει συμπεριφορές η θεωρία παιγνίων έχει επίσης προσπαθήσει να χρησιμοποιηθεί για να αναπτύξει θεωρίες ηθικής ή κανονιστικής συμπεριφοράς. Στα οικονομικά και στην φιλοσοφία οι ακαδημαϊκοί έχουνε εφαρμόσει την θεωρία παιγνίων για να βοηθήσουν στην κατανόηση της καλής ή αρμόζουσας συμπεριφοράς. Τέτοιου είδους επιχειρήματα της θεωρίας παιγνίων μπορούν εντοπιστούν ήδη από την εποχή του Πλάτωνα [37].

## *2.4 Παίγνια*

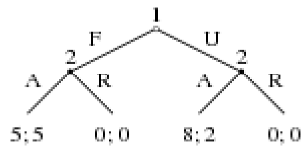
> Παίγνιο: Συμμετέχουν τουλάχιστον δύο παίκτες με τουλάχιστον δύο στρατηγικές ο καθένας και αντίθετα συμφέροντα. Το αποτέλεσμα για κάθε παίκτη καθορίζεται από τις συνδυασμένες επιλογές όλων των παικτών και δίνεται από τον πίνακα αποτελεσμάτων του παιγνίου (reward ή pay-off matrix) [2].

Ένα παιχνίδι είναι μια περιγραφή των στρατηγικών αλληλεπίδρασης που περιλαμβάνει τους περιορισμούς σχετικά με τις ενέργειες που οι παίκτες μπορούνε να κάνουν και τα συμφέροντά τους, αλλά δεν διευκρινίζει τις ενέργειες που οι παίκτες τελικά θα κάνουν. Μια λύση είναι μια συστηματική περιγραφή των αποτελεσμάτων που μπορεί να προκύψουν σε μια οικογένεια παιχνιδιών. Η θεωρία παιγνίων προτείνει λογικές λύσεις για κατηγορίες παιχνιδιών και εξετάζει τις ιδιότητές τους [25]. Τα παιχνίδια που μελετάει η θεωρία παιγνίων είναι καλώς καθορισμένα μαθηματικά αντικείμενα. Ένα παιχνίδι συνίσταται από ένα σύνολο παικτών, ένα σύνολο κινήσεων, ή αλλιώς στρατηγικών, διαθέσιμων σε αυτούς τους παίκτες, και μια προδιαγραφή των κερδών για κάθε συνδυασμό των στρατηγικών [35].

### 2.4.1 Αναπαράσταση παιγνίων

Ένα παιχνίδι μπορεί να περιγραφεί με ένα από τους τρεις ακόλουθους τρόπους: με εκτεταμένη μορφή (ή δενδροδιάγραμμα). με κανονική (ή στρατηγική) μορφή ή μορφή χαρακτηριστικής συνάρτησης. Μερικές φορές αυτές οι μορφές συνδυάζονται, όπως περιγράφεται στην Θεωρία των κινήσεων [47]. Τα περισσότερα παιχνίδια συνεργασίας αναπαριστώνται με την μορφή χαρακτηριστικής συνάρτησης ενώ η εκτεταμένη και κανονική μορφή χρησιμοποιούνται για να καθορίσουν τα παίγνια μη συνεργασίας.

**Εκτεταμένη μορφή (extensive form)**. Τα περισσότερα επιτραπέζια παιχνίδια, τα οποία προχωράνε βήμα βήμα, μια κίνηση την φορά, μπορούνε να αναπαρασταθούν σε εκτεταμένη μορφή. Τα παίγνια εκτεταμένης μορφής μπορούνε να περιγραφούν με ένα «δένδρο παιγνίου», στο οποίο κάθε σειρά, επιλογή παίκτη, είναι μια κορυφή του δένδρου, με κάθε διακλάδωση να υποδεικνύει της διαδοχικές επιλογές των παικτών.




**Σχ. 1 Ένα παίγνιο εκτεταμένης μορφής**

Η εκτεταμένη μορφή μπορεί να χρησιμοποιηθεί για να αναπαραστήσει παιχνίδια στα οποία υπάρχει κάποιου είδους σημαντική σειρά. Η εικόνα στα δεξιά δείχνει ένα δέντρο ενός τέτοιου παιχνιδιού με κάθε νούμερο στις κορυφές να προσδιορίζει το ποιος παίκτης κάνει την επιλογή και με τις αποδώσεις να καθορίζονται στο κάτω μέρος του δέντρου.

Στο παίγνιο που απεικονίζεται στο σχήμα 1 υπάρχουν δύο παίκτες. Ο παίκτης 1 παίζει πρώτος και διαλέγει είτε F είτε U. Ο παίκτης 2 παρατηρεί την κίνηση του παίκτη 1 και μετά επιλέγει A ή R. Υποθέτοντας ότι ο παίκτης 1 διαλέγει U και μετά ο παίκτης 2 διαλέγει A τότε ο παίκτης 1 παίρνει 8 και ό παίκτης 2 παίρνει 2.

Η εκτεταμένη μορφή μπορεί επίσης να αναπαραστήσει και παίγνια ταυτόχρονης κίνησης καθώς και παίγνια ελλιπής πληροφόρησης. Για να γίνει αυτό, είτε μια διακεκομμένη γραμμή ενώνει διαφορετικές κορυφές για να τις αναπαραστήσει σαν μέρος του ίδιου συνόλου πληροφορίας (information set), για παράδειγμα οι παίκτες δεν ξέρουν σε ποιο σημείο βρίσκονται, είτε μια κλειστεί γραμμή σχεδιάζεται γύρω από αυτές [25].

**Κανονική (στρατηγική) μορφή (normal form).** Η κανονική μορφή χρησιμοποιείται κυρίως για να περιγράψει παιχνίδια δύο παικτών. Σε αυτήν την μορφή ένα παίγνιο αναπαριστάται από έναν πίνακα αποδόσεων (payoff matrix), όπου κάθε σειρά περιγράφει την στρατηγική ενός παίκτη και κάθε στήλη περιγράφει την στρατηγική του άλλου παίκτη. Η καταχώρηση του πίνακα στην τομή κάθε γραμμής και στήλης δίνει το αποτέλεσμα της επιλογής των ανάλογων στρατηγικών από τους παίκτες. Οι αποδώσεις σε κάθε παίκτη που συνδέονται με αυτό το αποτέλεσμα είναι η βάση για τον καθορισμό του αν οι στρατηγικές είναι σε ισορροπία (equilibrium) ή σταθερές.

| | Ο παίκτης 2 επιλέγει αριστερά | Ο παίκτης 2 επιλέγει δεξιά |
|---|---|---|
| Ο παίκτης 1 επιλέγει πάνω | **4**, **3** | **–1**, **–1** |
| Ο παίκτης 1 επιλέγει κάτω | **0**, **0** | **3**, **4** |

**Σχ. 2 Κανονική μορφή ή πίνακας αποδόσεων ενός παιγνίου 2 ατόμων, 2 στρατηγικών**

Ποιο γενικά μπορεί να αναπαρασταθεί από οποιαδήποτε συνάρτηση συνδέει μια απόδοση για κάθε παίκτη με κάθε πιθανό συνδυασμό ενεργειών [10]. Στο σχήμα 2 υπάρχουν δύο παίκτες, ο ένας διαλέγει από τις γραμμές και ο άλλος από τις στήλες. Κάθε παίκτης έχει δύο στρατηγικές, οι οποίες καθορίζονται από το πλήθος των γραμμών και από το πλήθος των στηλών. Οι αποδόσεις παρέχονται στο εσωτερικό, ο πρώτος αριθμός είναι η απόδοση που λαμβάνει ο παίκτης των γραμμών, ο παίκτης 1 σε αυτήν την περίπτωση, και ο δεύτερος είναι η απόδοση για το παίκτη των στηλών, ο παίκτης 2. Υποθέτοντας ότι ο παίκτης 1 διαλέγει «πάνω» και ο παίκτης 2 διαλέγει «αριστερά» τότε ο 1 παίρνει σαν απόδοση 4 και ο 2 παίρνει 3.

Όταν ένα παίγνιο αναπαρίσταται σε κανονική μορφή υποτίθεται ότι ο κάθε παίκτης ενεργεί ταυτόχρονα ή, τουλάχιστον, χωρίς να ξέρει της ενέργειες του άλλου παίκτη. Εάν οι παίκτες έχουνε κάποια πληροφόρηση για τις επιλογές των άλλων παικτών τότε το παιχνίδι συνήθως αναπαρίσταται με εκτεταμένη μορφή [25].

**Μορφή χαρακτηριστικής συνάρτησης (Characteristic function form).** Η μορφή χαρακτηριστικής συνάρτησης γενικά χρησιμοποιείται για να αναλύσει παίγνια με περισσότερους από δύο παίκτες. Υποδηλώνει την ελάχιστη τιμή που κάθε συνασπισμός παικτών, συμπεριλαμβανομένων και των συνασπισμών ενός παίκτη, μπορεί να εγγυηθεί για τον εαυτό του όταν παίζει ενάντια σε έναν συνασπισμό που αποτελείται από όλους τους άλλους παίκτες [12].

Στα παίγνια συνεργασίας με μεταφέρσιμη ωφέλεια (transferable utility), δεν δίνονται ξεχωριστές αμοιβές. Αντίθετα, η χαρακτηριστική συνάρτηση καθορίζει την ωφέλεια του κάθε συνασπισμού. Η καθιερωμένη υπόθεση είναι ότι ο άδειος συνασπισμός έχει ωφέλεια 0.

Η καταγωγή αυτής της μορφής μπορεί να βρεθεί στο βιβλίο των von Neumann και Morgenstern οι οποίοι, όταν μελετούσαν κανονικής μορφής παίγνια συνασπισμού, υπέθεσαν ότι όταν σχηματίζεται ένας συνασπισμός C τότε αυτός παίζει ενάντια στον συμπληρωματικό συνασπισμό ( N/C ) σαν να έπαιζαν ένα παίγνιο 2 παικτών. Η ωφέλεια της ισορροπίας του C είναι χαρακτηριστική. Υπάρχουν διαφορετικά μοντέλα για να εξαχθούν οι τιμές συνασπισμών από τα παίγνια κανονικής μορφής, αλλά δεν μπορούν όλα τα παίγνια σε μορφή χαρακτηριστικής συνάρτησης να εξαχθούν από αυτά της κανονικής μορφής [46].

Επίσημα, ένα παίγνιο σε μορφή χαρακτηριστικής συνάρτησης (γνωστό και ως TU-game) δίνεται σαν ζεύγος (N,v) όπου το N δηλώνει ένα σύνολο παικτών και u : $2^N \to \mathbb{R}$ είναι μια χαρακτηριστική συνάρτηση [9].

Η χαρακτηριστική συνάρτηση έχει γενικευθεί για παιχνίδια χωρίς την υπόθεση της μεταφέρσιμης ωφέλειας.

**Διαχωριστική μορφή συνάρτησης (Partition function form).** Η χαρακτηριστική συνάρτηση αγνοεί τις πιθανές εξωτερικότητες, κάποιο αντίκτυπο μη άμεσα συσχετιζόμενο με την συνδιαλλαγή, των σχηματισμών συνασπισμών. Στην διαχωριστική μορφή συνάρτησης η ωφέλεια ενός συνασπισμού εξαρτάται όχι μόνο από μέλη του, αλλά και στον τρόπο με τον οποίο οι υπόλοιποι παίκτες είναι διαχωρισμένοι [42].

## 2.4.2 Είδη παιγνίων

Τα παίγνια ανάλογα με τις ιδιότητες και τους κανόνες που τα διέπουν μπορούνε να χωριστούν σε διάφορες κατηγορίες, ο συνδυασμός των οποίων μπορεί να τα χαρακτηρίσει. Διάφορες θεωρίες και προτάσεις λύσης αντιστοιχούν σε συγκεκριμένες κατηγορίες παιγνίων.

### 2.4.2.1 Συνεργασίας και μη συνεργασίας

Σε όλα τα μοντέλα της θεωρίας παιγνίων η βασική οντότητα είναι ένας παίκτης. Ο παίκτης μπορεί να ερμηνευτεί σαν ένα άτομο ή σαν μια ομάδα ατόμων που παίρνουν μια απόφαση. Μόλις καθοριστεί το σύνολο των παικτών είναι δυνατόν να διακριθούν δύο τύποι μοντέλων: αυτό στο οποίο τα σύνολα των πιθανών ενεργειών ανεξάρτητων παικτών αποτελούν θεμελιακά στοιχειά και εκείνο στο οποίο τα σύνολα των πιθανών κοινών ενεργειών ομάδων παικτών είναι θεμελιακά στοιχεία [25]. Το πρώτο μοντέλο μπορεί να αποκαλεστεί ως «μη συνεργασίας» ενώ το δεύτερο ως «συνεργασίας», παρόλο που αυτοί οι όροι δεν εκφράζουν καλά τις διαφορές μεταξύ των δύο μοντέλων.

Ένα παίγνιο είναι συνεργασίας εάν οι παίκτες μπορούν να σχηματίσουν δεσμευτικές υποσχέσεις. Για παράδειγμα το νομικό σύστημα τους επιβάλλει να τηρούν τις υποσχέσεις τους. Στα παίγνια μη συνεργασίας αυτό δεν είναι δυνατόν. Συχνά θεωρείται ότι η επικοινωνία ανάμεσα στους παίκτες επιτρέπεται στα παίγνια συνεργασίας αλλά όχι σε αυτά της μη συνεργασίας [16]. Ένας πωλητής αυτοκινήτων και κάποιος πιθανός πελάτης εμπλέκονται σε ένα παίγνιο συνεργασίας εάν καταλήξουν σε μια τιμή και υπογράψουν ένα συμβόλαιο.

Ωστόσο, το παζάρεμα που κάνουνε για να φτάσουνε σε αυτό το σημείο θα είναι μη συνεργασίας. Παρόμοια όταν οι άνθρωποι πλειοδοτούν σε μια δημοπρασία παίζουν ένα παίγνιο μη συνεργασίας, ακόμα και αν αυτός με την υψηλότερη προσφορά συμφωνήσει να ολοκληρώσει την αγορά [47]. Υβριδικά παιχνίδια περιέχουν στοιχεία και από τους δύο τύπους. Για παράδειγμα, ένας συνασπισμός παικτών σχηματίζεται σε ένα παίγνιο συνεργασίας αλλά αυτοί παίζουν με μη συνεργάσιμο τρόπο.

Από τους δύο αυτούς τύπους παιχνιδιών τα μη συνεργασίας είναι ικανά να αναπαραστήσουν καταστάσεις στις πιο μικρές λεπτομέρειες, παράγοντας ακριβή αποτελέσματα. Τα παίγνια συνεργασίας δίνουν έμφαση στο παιχνίδι γενικότερα. Έχουνε γίνει αξιοσημείωτες προσπάθειες να συνδέσουν τις δύο προσεγγίσεις. Το αποκαλούμενο πρόγραμμα Nash (Nash-programme) έχει ήδη καθιερώσει πολλές από τις λύσεις συνεργασίας σαν ισορροπίες μη συνεργασίας [6]. Τα τελευταία χρόνια οι περισσότερες έρευνες αφιερώνονται στα παίγνια μη συνεργασίας αλλά αυτό δε σημαίνει ότι η σχετική σημασία των παιγνίων συνεργασίας είναι μικρότερη ή λιγότερο βασική από τα άλλα.

#### 2.4.2.2 Συμμετρικά και ασυμμετρικά

Ένα συμμετρικό παίγνιο είναι ένα παιχνίδι στο οποίο η ωφέλεια για το παίξιμο μιας συγκεκριμένης στρατηγικής εξαρτώνται μόνο από τις άλλες στρατηγικές που υπάρχουν και όχι από το ποιος τις παίζει [24]. Εάν οι ταυτότητες των παιχτών μπορούν να αλλαχθούν χωρίς να επηρεαστούν τα οφέλη των στρατηγικών τότε το παιχνίδι είναι συμμετρικό. Πολλά από τα 2x2 παίγνια που συνήθως εξετάζονται είναι συμμετρικά. Οι καθιερωμένες αναπαραστάσεις των παιχνιδιών του δειλού (chicken), του διλήμματος του κρατουμένου (prisoners dilemma) και του κυνηγιού ελαφιού (stag hunt) είναι όλα συμμετρικά παίγνια. Μερικοί μελετητές θα θεωρούσαν ορισμένα ασυμμετρικά παίγνια σαν παραδείγματα αυτών των παιγνίων επίσης. Ωστόσο, τα πιο κοινά οφέλη για αυτά τα παιχνίδια είναι συμμετρικά.

| | E | F |
|---|---|---|
| E | 1, 2 | 0, 0 |
| F | 0, 0 | 1, 2 |

**Σχ. 3 Ένα ασυμμετρικό παίγνιο**

Τα ασυμμετρικά παίγνια που μελετούνται ποιο συχνά είναι παίγνια όπου δεν υπάρχουν πανομοιότυπα σύνολα στρατηγικών και για τους δύο παίκτες. Για παράδειγμα, το τελεσιγραφικό παίγνιο και παρόμοια το παιχνίδι του δικτάτορα έχουν διαφορετικές στρατηγικές για κάθε παίκτη. Είναι πιθανό παρόλα αυτά για ένα παίγνιο να έχει πανομοιότυπες στρατηγικές και για τους δύο παίκτες και να είναι ασυμμετρικό. Για παράδειγμα, το παίγνιο στο σχήμα 3 είναι ασυμμετρικό παρόλο που έχει πανομοιότυπα σύνολα στρατηγικών και για τους δύο παίκτες.

### 2.4.2.3 Μηδενικού και μη μηδενικού αθροίσματος

Ο βαθμός στον οποίο οι στόχοι των παικτών συμπίπτουν ή συγκρούονται είναι μια άλλη βάση με την οποία ταξινομούνται τα παίγνια. Τα παίγνια σταθερού αθροίσματος είναι παιχνίδια απόλυτης σύγκρουσης, ή όπως αλλιώς αποκαλούνται παίγνια καθαρού ανταγωνισμού [47]. Οι παίκτες στα παίγνια σταθερού αθροίσματος έχουν εντελώς αντίθετα συμφέροντα, ενώ στα παίγνια μεταβλητού αθροίσματος μπορεί όλοι να είναι νικητές ή χαμένοι. Σε μια διαφωνία υπαλλήλων και διοίκησης για παράδειγμα, τα δύο μέρη σίγουρα έχουνε συγκρουόμενα συμφέροντα αλλά και οι δύο ωφελούνται αν αποτραπεί μια απεργία.

Τα παίγνια μηδενικού αθροίσματος είναι μια ειδική περίπτωση των παιγνίων σταθερού αθροίσματος, στα οποία οι επιλογές των παικτών δεν μπορούνε ούτε να αυξήσουν ούτε να μειώσουν τους διαθέσιμους πόρους. Σε αυτά, το συνολικό όφελος όλων των παικτών στο παιχνίδι, για κάθε συνδυασμό στρατηγικών, πάντα αθροίζει μηδέν. Πιο απλά ένα παίκτης ωφελείται ισόποσα σε βάρος των άλλων. Το πόκερ είναι ένα χαρακτηριστικό παράδειγμα παιγνίου μηδενικού αθροίσματος, αγνοώντας την πιθανότητα να παίρνει μερίδιο το καζίνο για παράδειγμα, γιατί ο κερδισμένος παίρνει ακριβώς ότι χάνουν οι χαμένοι, το συνολικό ποσό δηλαδή παραμένει σταθερό παρόλο που αλλάζει χέρια κατά την διάρκεια του παιχνιδιού.

Άλλα παίγνια μηδενικού αθροίσματος είναι το παιχνίδι με τα κέρματα (matching pennies) και το σκάκι, γιατί το πλήθος των κομματιών δεν αυξάνεται ποτέ.

Πολλά παίγνια που εξετάζονται από τους ερευνητές της θεωρίας παιγνίων, συμπεριλαμβανομένου και του διασήμου διλήμματος του κρατουμένου, είναι παίγνια μη μηδενικού αθροίσματος γιατί μερικές εκβάσεις έχουνε καθαρά αποτελέσματα μεγαλύτερα ή μικρότερα από το μηδέν. Πιο απλά στα παίγνια μη μηδενικού αθροίσματος η ωφέλεια ενός παίκτη δεν συμπίπτει απαραίτητα με την απώλεια κάποιου άλλου παίκτη.

| | A | B |
|---|---|---|
| A | –1, 1 | 3, –3 |
| B | 0, 0 | –2, 2 |

**Σχ. 4 Παίγνιο μηδενικού αθροίσματος**

Παίγνια σταθερού αθροίσματος αντιστοιχούν σε δραστηριότητες όπως είναι η κλοπή και ο τζόγος, αλλά όχι στην θεμελιώδη οικονομική κατάσταση στην οποία υπάρχει ενδεχόμενο κέρδος από το εμπόριο [28]. Είναι δυνατό να μετατραπεί οποιοδήποτε παίγνιο σε (πιθανώς ασυμμετρικό) παίγνιο μηδενικού αθροίσματος προσθέτοντας έναν επιπρόσθετο εικονικό παίκτη οι απώλειες του οποίου αντισταθμίζουν τα καθαρά κέρδη των παικτών.

Υπάρχει μεγάλη διαφωνία μεταξύ των οικονομολόγων για το εάν η οικονομία σαν σύνολο είναι ένα παίγνιο μηδενικού αθροίσματος ή όχι. Εάν είναι, τότε ολόκληρη η οικονομία είναι μια συνεχής ανακατανομή του ίδιου μεγέθους πλούτου. Καθώς οι πλούσιοι γίνονται πλουσιότεροι τότε οι φτωχοί γίνονται φτωχότεροι σύμφωνα με αυτούς που συμφωνούν. Οι περισσότεροι ωστόσο πιστεύουν ότι η οικονομία δεν αποτελεί παίγνιο μηδενικού αθροίσματος. Καθώς δημιουργείται πλούτος, οι πλούσιοι γίνονται πλουσιότεροι αλλά και *οι φτωχοί γίνονται πλουσιότεροι*. Αυτή είναι μια θεμελιώδης διαφορά μεταξύ της σοσιαλιστικής θεωρίας και της καπιταλιστικής θεωρίας [50].

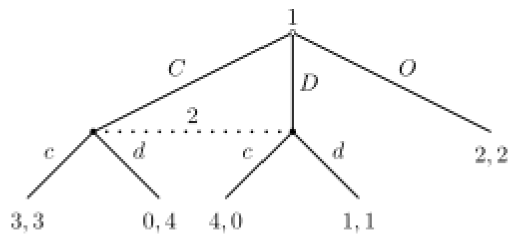
#### 2.4.2.4 Ταυτόχρονα (στατικά) και διαδοχικά (δυναμικά)

Τα ταυτόχρονα παίγνια είναι παιχνίδια στα οποία όλοι οι παίκτες κάνουν την κίνηση τους ταυτόχρονα ή αν δεν ισχύει αυτό οι παίκτες που παίζουν αργότερα δεν γνωρίζουν τις ενέργειες αυτών που έπαιξαν νωρίτερα, κάνοντας τα «θεωρητικά» ταυτόχρονα. Τα διαδοχικά παίγνια, ή αλλιώς δυναμικά παίγνια, είναι παιχνίδια στα οποία οι παίκτες που παίζουνε αργότερα έχουνε κάποια γνώση για τις ενέργειες που έγιναν νωρίτερα [49]. Αυτά δε χρειάζεται να είναι τέλειας πληροφόρησης, μια κατηγορία που εξηγείται παρακάτω, για κάθε ενέργεια των παικτών που έδρασαν νωρίτερα, θα μπορούσε να υπάρχει ελάχιστη γνώση. Για παράδειγμα, ένας παίκτης μπορεί να γνωρίζει ότι κάποιος προηγούμενος παίκτης δεν έκανε κάποια συγκεκριμένη ενέργεια ενώ δεν γνωρίζει ποια από τις υπόλοιπες διαθέσιμες ενέργειες έκανε πράγματι ο πρώτος παίκτης.

Συχνά η κανονική μορφή χρησιμοποιείται για να αναπαραστήσει ταυτόχρονα παίγνια και η εκτεταμένη μορφή για τα διαδοχικά, παρόλο που αυτός δεν είναι κάποιος απόλυτος κανόνας από τεχνικής άποψης.

#### 2.4.2.5 Τέλειας και ατελούς πληροφόρησης

Ένα σημαντικό υποσύνολο των διαδοχικών παιγνίων αποτελούν τα παίγνια τέλειας πληροφόρησης. Ένα παιχνίδι αποκαλείται τέλειας πληροφόρησης εάν όλοι παίκτες γνωρίζουν τις κινήσεις που έγιναν προηγουμένως από όλους τους άλλους παίκτες [47]. Είναι επόμενο, μόνο τα διαδοχικά παίγνια να μπορούν να είναι τέλειας πληροφόρησης, εφόσον στα ταυτόχρονα παίγνια δεν γνωρίζουν όλοι οι παίκτες τις ενέργειες των άλλων. Τα περισσότερα παίγνια που μελετώνται στην θεωρία παιγνίων είναι ατελούς πληροφόρησης, παρόλο που υπάρχουν ορισμένα ενδιαφέροντα παραδείγματα τέλειας πληροφόρησης όπως είναι το τελεσιγραφικό παίγνιο και το παιχνίδι τις εκατονταποδαρούσας [41]. Τα παίγνια τέλειας πληροφόρησης συμπεριλαμβάνουν το σκάκι, το go, το mancala και το arimaa. Το πόκερ για παράδειγμα, είναι ατελούς πληροφόρησης γιατί οι παίκτες δεν γνωρίζουν όλα τα χαρτιά των αντιπάλων τους.




**Σχ. 5 Ένα παίγνιο ατελούς πληροφόρησης (οι διακεκομμένες γραμμές αναπαριστούν την άγνοια από πλευράς του παίκτη 2)**

Η τέλεια πληροφόρηση συχνά συγχέεται με την πλήρης πληροφόρηση η οποία είναι μια παρόμοια έννοια. Η πλήρης πληροφόρηση απαιτεί κάθε παίκτης να γνωρίζει τις στρατηγικές και τα οφέλη των άλλων παικτών αλλά όχι απαραίτητα τις ενέργειες τους.

### 2.4.2.6 Άπειρης διάρκειας

Τα παίγνια, όπως μελετώνται από τους οικονομολόγους και τους παίκτες του πραγματικού κόσμου, γενικά ολοκληρώνονται σε πεπερασμένο αριθμό κινήσεων. Οι καθαροί μαθηματικοί δεν είναι τόσο περιοριστικοί και οι θεωρητικοί των συνόλων ιδιαίτερα μελετάνε παίγνια που διαρκούν άπειρο αριθμό κινήσεων στα οποία ο νικητής, ή οποιοσδήποτε θα είχε κάποιο όφελος, δεν γίνεται γνωστός μέχρις ότου όλες αυτές οι κινήσεις ολοκληρωθούν.

Η εστίαση της προσοχής είναι συνήθως όχι τόσο στο ποιος είναι ο καλύτερος τρόπος για να παιχτεί ένα παίγνιο, αλλά πιο απλά αν ο ένας ή ο άλλος παίχτης έχει στρατηγική νίκης (μπορεί να αποδειχθεί, χρησιμοποιώντας το αξίωμα της επιλογής, ότι υπάρχουν παίγνια, ακόμα και με τέλεια πληροφόρηση και όπου το μόνο αποτέλεσμα είναι νίκη ή ήττα, για τα οποία κανένας παίκτης δεν έχει στρατηγική νίκης) [47]. Η ύπαρξη τέτοιων στρατηγικών, για έξυπνα σχεδιασμένα παίγνια, έχει μεγάλες επιπτώσεις στην περιγραφική θεωρία συνόλων.

### 2.4.2.7 Διακριτά και συνεχή

Μεγάλο μέρος της θεωρίας παιγνίων ασχολείται με πεπερασμένα, διακριτά παίγνια που έχουν πεπερασμένο αριθμό παικτών, κινήσεων, συμβάντων, αποτελεσμάτων κτλ. Πολλές έννοιες μπορούνε να επεκταθούν όμως. Τα συνεχή παίγνια επιτρέπουν στους παίκτες να διαλέξουν στρατηγική από ένα συνεχές σύνολο στρατηγικών. Για παράδειγμα, ο διαγωνισμός του Cournot είναι συνήθως αναπαριστώμενος με τις στρατηγικές των παικτών να είναι μη αρνητικές ποσότητες, συμπεριλαμβανομένων και κλασματικών ποσοτήτων. Διαφορικά παίγνια όπως είναι το συνεχές παίγνιο της καταδίωξης και διαφυγής είναι συνεχή παίγνια.

Τέλος, ένα παίγνιο λέγεται διακριτό όταν κάθε παίκτης έχει διακριτό αριθμό επιλογών, ο αριθμός των παικτών είναι διακριτός και το παίγνιο δεν μπορεί να συνεχίζεται επ' άπειρον. Το σκάκι, το πόκερ και τα περισσότερα επιτραπέζια παίγνια είναι διακριτά.

#### 2.4.2.8 Ενός και πολλών παικτών

Τα προβλήματα μεμονωμένης απόφασης κάποιες φορές θεωρούνται σαν παίγνια ενός παίκτη. Αν και αυτές οι καταστάσεις δεν ανήκουν στην θεωρία παιγνίων περιγράφονται χρησιμοποιώντας πολλά από τα ίδια εργαλεία της θεωρίας αποφάσεων. Είναι μόνο όταν υπάρχουν δύο ή παραπάνω παίκτες που ένα πρόβλημα γίνεται πρόβλημα της θεωρίας παιγνίων. Ο παίκτης δεν χρειάζεται να είναι κάποιο άτομο, μπορεί να είναι ένα κράτος, μια εταιρία, μια ομάδα ατόμων με τα ίδια συμφέροντα κτλ. Ένας παίκτης ο οποίος ενεργεί τυχαία, κάνει τυχαίες κινήσεις γνωστές και ως «φυσικές κινήσεις», προστίθεται συχνά [25]. Αυτός ο παίκτης δεν θεωρείται συνήθως σαν ένας τρίτος παίκτης σε ότι αφορά το κατά τ' άλλα παίγνιο δύο παικτών, αλλά χρησιμεύει για να παρέχει το «ρίξιμο των ζαριών» όπου αυτό χρειάζεται στο παιχνίδι. Εάν οι παίκτες είναι παραπάνω από 2 τότε το παίγνιο ονομάζεται ν-παικτών. Τα παίγνια με άπειρο πλήθος παικτών συνήθως αποκαλούνται παίγνια ν-παικτών [20]. Τα μοναχικά παίγνια, π.χ. puzzles, μπορούνε να συμπεριλαμβάνουν και παίγνια συνεργασίας στα οποία όλοι οι παίκτες προσπαθούν να λύσουν το ίδιο πράγμα χωρίς ανταγωνισμό. Παίγνιο δύο παικτών είναι το σκάκι και τα διάφορα αθλήματα για παράδειγμα. Ν-παικτών είναι η μονόπολη, το πόκερ, το λαχείο ή το χρηματιστήριο.

#### 2.4.2.9 Μεταπαίγνια

Αυτά είναι παίγνια τα οποία είναι η ανάπτυξη των κανόνων για ένα άλλο παίγνιο, το παίγνιο στόχος ή θέμα. Τα μεταπαίγνια προσπαθούν να μεγιστοποιήσουν την αξία χρησιμότητας του κανόνα που αναπτύσσεται [17]. Η θεωρία των μεταπαιγνίων σχετίζεται με την θεωρία μηχανικού σχεδιασμού (mechanism design).

### 2.4.3 Λύσεις ή έννοιες λύσεων της θεωρίας παιγνίων

Στην συνέχεια παρουσιάζονται τα λογικά εργαλεία με τα οποία η θεωρία παιγνίων επιδιώκει να λύσει «παίγνια», δηλαδή να προβλέψει πώς θα συμπεριφερθούν οι ορθολογικά σκεπτόμενοι άνθρωποι σε καταστάσεις αλληλεπίδρασης στις οποίες το αποτέλεσμα καθορίζεται από το συνδυασμό των δράσεων κάθε ατόμου. Το πρώτο που πρέπει να γίνει, είναι να οριστούν δύο τύποι στρατηγικής μεταξύ των οποίων μπορεί να διαλέξει ο παίκτης.

#### 2.4.3.1 Μικτές και καθαρές στρατηγικές

Η πιο απλή στρατηγική που μπορεί να έχει ένας παίκτης, είναι να επιλέξει κάποιον πολύ συγκεκριμένο τρόπο να δράσει, να κινηθεί. Για παράδειγμα: «χτυπάω τον αντίπαλο» ή «βοηθώ κάποιον να διασχίσει το δρόμο». Αυτός ο τρόπος επιλογής δράσης ονομάζεται καθαρή στρατηγική. Ωστόσο, υπάρχουν στιγμές που μπορεί να μην είναι βέβαιος σχετικά με το ποια είναι η βέλτιστη καθαρή στρατηγική. Κάτω από τη σκιά της αβεβαιότητας, σχετικά με το ποιά είναι η βέλτιστη καθαρή στρατηγική, μπορεί να ενεργήσει ως εάν να διαλέγει στην τύχη μεταξύ δύο ή περισσότερων καθαρών στρατηγικών: π.χ., όταν δεν έχει αξιόπιστες μετεωρολογικές πληροφορίες, να αποφασίσει αν θα πάρει ή όχι μαζί του ομπρέλα στρίβοντας ένα κέρμα [1]. Ο τύπος αυτός στρατηγικής αποκαλείται μικτή στρατηγική, με την έννοια ότι επιλέγεται ένα «μίγμα πιθανοτήτων» ενός συνόλου καθαρών στρατηγικών. Στο παράδειγμά για το ότι θα πάρει κάποιος μαζί του ομπρέλα, η μικτή στρατηγική επιλογή του μπορεί να εκφραστεί ως: «Παίρνω ομπρέλα με πιθανότητα $p = ½$ και δεν παίρνω με πιθανότητα $1 – p = ½$».

Ακολουθεί ένα παράδειγμα μιας καθαρά στρατηγικής αλληλεπίδρασης, αντίθετα με το «παίγνιο» εναντίον του καιρού. Υποτίθεται ότι μια κοπέλα σκέπτεται να πάει σε κάποιο πάρτι και πρέπει να επιλέξει μεταξύ δύο καθαρών στρατηγικών: «Να φορέσει μαύρα παπούτσια» (Μ) ή «να φορέσει κόκκινα παπούτσια» (Κ). Ας υποθέσουμε, επίσης, ότι θα επέλεγε να φορέσει μαύρα παπούτσια αν είχε αξιόπιστες πληροφορίες ότι οι περισσότερες φίλες της θα φορούσαν κόκκινα παπούτσια. Αν, αντιθέτως, οι «πληροφοριοδότες» της την ενημέρωναν ότι οι περισσότερες φίλες της θα φορούσαν μαύρα παπούτσια, τότε θα επέλεγε να φορέσει κόκκινα παπούτσια, στην προσπάθειά της να είναι διαφορετική. Τι θα κάνει όμως όταν δεν έχει αξιόπιστες πληροφορίες όσον αφορά το ποσοστό αυτών που θα φορούν κόκκινα παπούτσια στο πάρτι; Σε αυτήν την περίπτωση, θα μπορούσε κάλλιστα να επιλέξει τυχαία μεταξύ των δύο, υιοθετώντας την ακόλουθη μικτή στρατηγική: να επιλέξει την καθαρή στρατηγική Μ με πιθανότητα p και την καθαρή στρατηγική Κ με πιθανότητα 1–p.

Γενικεύοντας, εάν ένας παίκτης μπορεί να επιλέξει από N καθαρές στρατηγικές (S1,S2,…..,SN), τότε μια μικτή στρατηγική ορίζεται από τις πιθανότητες (p1,p2,..…,pN) με βάση τις οποίες θα πρέπει να επιλέξει καθεμιά από τις καθαρές στρατηγικές του. Για να είναι καλώς καθορισμένη μια μικτή στρατηγική, κάθε μία από τις πιθανότητες πρέπει να βρίσκεται ανάμεσα στα 0 και 1, και το άθροισμα τους να είναι ίσο με 1. Η επιλογή επίσης της Μ (p1=0,p2=0…..pj=1….,pN=0) είναι ισοδύναμη με την επιλογή της καθαρής στρατηγικής Sj [26].

Εδώ η κοπέλα προτιμά μια μικτή στρατηγική εξαιτίας της αβεβαιότητάς της σχετικά με το τι θα κάνουν οι υπόλοιπες. Ένας άλλος λόγος για να επιλέξει κανείς μικτή στρατηγική είναι ότι μπορεί να θέλει να κρατήσει τους αντιπάλους του σε κατάσταση αβεβαιότητας. Γιατί αν κάποιος θέλει να κάνει τους άλλους να είναι αβέβαιοι όσον αφορά το τι πρόκειται να κάνει, ίσως ο καλύτερος τρόπος είναι να μείνει και ο ίδιος εξίσου αβέβαιος ως προς το τι θα κάνει. Αυτό ισοδυναμεί με το να πράττει σαν να διαλέγει τυχαία μεταξύ των καθαρών στρατηγικών, δηλαδή να χρησιμοποιεί μια μικτή στρατηγική. Για παράδειγμα, ένας παίκτης ετοιμάζεται να εκτελέσει πέναλτι στο ποδόσφαιρο. Ο τερματοφύλακας θα ήθελε να ξέρει αν θα σουτάρει προς τα αριστερά ή προς τα δεξιά του. Για να τον κρατήσει σε αβεβαιότητα, μπορεί να επιλέξει μια μικτή στρατηγική: «σουτάρει την μπάλα αριστερά» με πιθανότητα 40 τοις εκατό, «σουτάρει την μπάλα δεξιά», με πιθανότητα 40 τοις εκατό, και «σουτάρει την μπάλα στο κέντρο της εστίας» με πιθανότητα 20 τοις εκατό.

#### 2.4.3.2 Κανονική και εκτεταμένη μορφή και στατικά και δυναμικά παίγνια

Έχει ήδη αναφερθεί ότι οι δύο κύριες εκφράσεις του τρόπου με τον οποίο οι στρατηγικές των παιγνίων αλληλεπιδρούν για να παραγάγουν αποτελέσματα είναι η κανονική μορφή και η εκτεταμένη μορφή. Η κανονική μορφή παιγνίου μοιάζει συνήθως με πίνακα, και είναι επίσης γνωστή ως μορφή πίνακα ή στρατηγική μορφή, ο οποίος συνδέει συνδυασμούς καθαρών στρατηγικών με αποτελέσματα εκφρασμένα σε μονάδες ωφέλειας για τον κάθε παίκτη ( Παίγνιο 2.1). Δεδομένου ότι οι στήλες και οι γραμμές του πίνακα είναι καθαρές στρατηγικές το ακόλουθο κείμενο επικεντρώνεται, επί του παρόντος, σε αυτές.[1]

Για ευκολία στην περιγραφή, ο παίκτης που επιλέγει μεταξύ γραμμών θα είναι γυναίκα και θα αναφέρεται ως παίκτης R, από την αγγλική λέξη για τη γραμμή, Row. Αναμενόμενα, ο παίκτης που επιλέγει μεταξύ στηλών θα είναι άνδρας και θα αναφέρεται ως C, από την αγγλική λέξη για τη στήλη, Column. Η πρώτη στρατηγική επιλογή της παίκτριας R συμβολίζεται με R1 και είναι η πρώτη σειρά, η δεύτερη με R2 και είναι η δεύτερη κτλ. Ας υποθέσουμε τώρα ότι η R επιλέγει R1 και ο C επιλέγει C1. Το αντίστοιχο αποτέλεσμα είναι (R1,C1). Στο παράδειγμα αυτό η R αποκομίζει 8 μονάδες ωφέλειας και αντίστοιχα το ίδιο και ο C. Το πρώτο νούμερο σε κάθε κελί του πίνακα αποδόσεων είναι η απόδοση σε ωφέλεια για την παίκτρια R, ενώ το δεύτερο νούμερο ανήκει στον παίκτη C. Παραδείγματος χάρη, το αποτέλεσμα (R1,C2) δίνει 2 μονάδες ωφέλειας στον C και καμιά στην R.

| | C1 | C2 |
|---|---|---|
| R1 | 8, 8 | 0, 2 |
| R2 | 9, 3 | 1, 4 |

**Παίγνιο 2. 1 Η κανονική μορφή παρουσίασης του παιγνίου**

Από την κανονική μορφή όμως του παιγνίου δεν γίνεται αντιληπτό κάτι για τη διαδικασία, ή την ακολουθία, του παιγνίου που σημαίνει ότι οι παίκτες κάνουν τις επιλογές τους ταυτόχρονα. Όταν οι παίκτες επιλέγουν πράγματι ταυτόχρονα, η κανονική μορφή είναι αρκετή για να περιγράψει το παίγνιο. Ωστόσο, όταν ο ένας

[1] Από εδώ και μπρος όταν αναφέρονται «στρατηγικές» θα εννοούνται καθαρές στρατηγικές, οι μικτές στρατηγικές αναφέρονται στην συνέχεια και όταν αναφερθούν θα δηλώνεται ρητά.

παίκτης κάνει την επιλογή του πριν ο αντίπαλος έχει την ευκαιρία να κινηθεί, η κανονική μορφή δεν μεταβιβάζει αυτή τη στρατηγικά κρίσιμη πληροφορία όσον αφορά την ακολουθία των επιλογών [25]. Τότε χρειαζόμαστε μια διαφορετική απεικόνιση: την εκτεταμένη μορφή, που είναι επίσης γνωστή ως δυναμική μορφή ή μορφή δενδροδιαγράμματος. Το παίγνιο 2.1 θα μπορούσε να αναπαρασταθεί και σε εκτεταμένη μορφή, με τους κανόνες που αναφέρθηκαν στην αναπαράσταση παιγνίων, αλλά στην συνέχεια θα εξεταστούν παίγνια σε κανονική μορφή.

Όταν εξετάζονται παίγνια στην κανονική μορφή τους ξεκινάτε με την υπόθεση ότι οι παίκτες επιλέγουν στρατηγική, μεταξύ των διαθέσιμων στρατηγικών, μία φορά, ταυτόχρονα και χωρίς να επικοινωνούν μεταξύ τους. Εφόσον οι παίκτες παίζουν μία φορά, και βλέπουν τι διάλεξε ο αντίπαλος τους μόνο όταν το παίγνιο θα έχει τελειώσει, δεν γίνεται να θεωρηθεί το παίγνιο δυναμικό, ότι αναπτύσσεται δηλαδή καθώς περνάει ο χρόνος και επιτρέπεται κάποιος τρόπος μάθησης, όπως είναι η δοκιμή και το λάθος. Ο χρόνος δηλαδή περνάει χωρίς να γίνεται κάτι και μόνο για όση ώρα ο παίκτης αποφασίζει ποια θα είναι η πρώτη και μοναδική του επιλογή. Παραδείγματος χάρη, ο παίκτης σκέπτεται προσεκτικά: «Αν ο αντίπαλός μου κάνει την κίνηση Χ εγώ θα κάνω την κίνηση Υ. Στην περίπτωση αυτή, όμως, δεν γνωρίζει ο αντίπαλός μου ότι αυτή είναι η κίνηση που πρόκειται να κάνω; Μάλλον ναι. Γι’ αυτό, είναι ίσως καλύτερο να κάνω την κίνηση Ζ…» [1]. Ο τύπος αυτός σκέψης πριν την μοναδική κίνηση δεν αποτελεί μέρος δυναμικού παιγνίου, δεν αποτελεί κομμάτι ενός βρόχου ανάδρασης από παρατηρήσεις σε πεποιθήσεις και σε δράσεις, και πάλι σε παρατηρήσεις. Όταν απουσιάζει μια τέτοια δυναμική διαδικασία, λέγεται ότι το, στατικό, παίγνιο εκτυλίσσεται σε λογικό χρόνο. Στο λογικό χρόνο αντιπαρατίθεται ο ιστορικός, πραγματικός, χρόνος κατά τον οποίο υπάρχουν «γεγονότα», γίνονται πραγματικές στρατηγικές επιλογές, οι οποίες με χειροπιαστό πλέον τρόπο τροφοδοτούν τις σκέψεις των παικτών.

Τα στατικά παίγνια γεννούν στρατηγική σκέψη που πραγματοποιείται σε λογικό χρόνο. Απεναντίας, τα δυναμικά παίγνια αναπτύσσονται σε ιστορικό χρόνο και μπορούν να εκφραστούν κατάλληλα μόνο σε εκτεταμένη μορφή [12]. Στην συνέχεια εστιάζεται η προσοχή σε παίγνια μίας ταυτόχρονης κίνησης, δηλαδή σε στατικά παίγνια στα οποία οι παίκτες επιλέγουν μόνο μία φορά και ταυτόχρονα. Οι αλληλεπιδράσεις παρουσιάζονται σε κανονική μορφή και όλη η σκέψη εκτυλίσσεται σε λογικό χρόνο, επειδή όλος ο ιστορικός χρόνος συμπυκνώνεται σε μία, μοναδική στιγμή: τη στιγμή που γίνονται οι επιλογές.

### 2.4.3.3 Λύση και ισορροπίες παιγνίου

Όταν αναφέρεται η φράση «λύση» παιγνίου, καθώς και «έννοια λύσης», από τους θεωρητικούς εννοείται κάτι συγκεκριμένο. Ο σκοπός για τον οποίο μελετώνται τα παίγνια είναι για να επιτευχθεί πρόβλεψη για το πώς θα εκτυλιχθούν οι στρατηγικές αυτές αλληλεπιδράσεις. Η θεωρία παιγνίων εστιάζει στην πρόβλεψη των αποτελεσμάτων ορθολογικών συμπεριφορών. Για να μπορεί να ισχυριστεί η θεωρία παιγνίων ότι «έλυσε» ένα παίγνιο θα πρέπει να οδηγεί σε σαφείς προβλέψεις για το αποτέλεσμα του. Ο μεγαλύτερος στόχος της είναι να δώσει μια γενική θεωρία που να μπορεί να «λύνει» μια όσο το δυνατόν μεγαλύτερη τάξη παιγνίων.

Συχνά ο όρος «λύση» χρησιμοποιείται ως συνώνυμο του όρου «ισορροπία». Δανεισμένη από τη φυσική, η έννοια της ισορροπίας αναφέρεται σε μια σταθερή κατάσταση πραγμάτων δηλαδή σε μια κατάσταση στην οποία δεν υπάρχουν εσωτερικές δυνάμεις που να μπορούν να την μεταβάλουν. Σύμφωνα με αυτή, η λύση ενός παιγνίου για τους θεωρητικούς της θεωρίας παιγνίων είναι κάτι στο οποίο οι παίκτες, σκεπτόμενοι ορθολογικά, είναι υποχρεωμένοι να οδηγούνται.

Πράγματι αν η λύση που ανακαλύφθηκε είναι λογική και μοναδική, τότε οι ορθολογικοί παίκτες όχι απλά θα επιλέξουν τις συγκεκριμένες στρατηγικές, οι οποίες δίνουν τη λύση αυτή, αλλά και θα μείνουν πιστοί σε αυτές, ή δεν θα μετανιώσουν ότι τις υιοθέτησαν αργότερα. Τέτοιου είδους στρατηγικές θεωρείται πως συνθέτουν μια μοναδική ισορροπία. Από τη στιγμή που οι παίκτες θα τις έχουν κατανοήσει δεν θα έχουν κανένα λόγο να μην τις υιοθετήσουν.

### 2.4.3.4 Λύση Ελαχιστοποίησης της Μέγιστης Ζημίας (minimax)

Ο John von Neumann θεωρείται ιδρυτής της θεωρία παιγνίων επειδή, στο διάσημο δοκίμιό του το 1928, παρουσιάζει μια λύση που ισχύει για μια ευρεία τάξη παιγνίων. Η γενική θεωρία λύσης παιγνίων φιλοδοξούσε να βοηθήσει τους κοινωνικούς επιστήμονες στην ανάλυση, άρα και στην κατανόηση, πολλών και διάφορων κοινωνικοοικονομικών συγκρούσεων. Τα επιτεύγματα του von

Neumann ήταν σημαντικά ωστόσο αναφέρεται ότι στην σύγχρονη θεωρία παιγνίων η λύση του δεν αποτελεί βασικό σημείο [1]. Το γιατί εξηγείται στην πορεία.

Ακολουθεί το Παίγνιο 2.2. Είναι ένα παίγνιο μηδενικού αθροίσματος καθώς οι αποδώσεις των δύο αντιπάλων για κάθε αποτέλεσμα αθροίζουν μηδέν. Ο von Neumann μελέτησε τα παίγνια αυτά και απέδειξε ότι λύνονται με τον ίδιο τρόπο.

| | C1 | C2 | C3 |
|---|---|---|---|
| R1 | -6, 6 | -12, 12 | 0, 0 |
| R2 | -2, 2 | 9, -9 | 1, -1 |
| R3 | -1, 1 | 0, 0 | 2, -2 |

**Παίγνιο 2. 2 Παίγνιο μηδενικού αθροίσματος**

Ο Neumann ήθελε, για μηδενικού αθροίσματος πάντα παίγνια, να ανακαλύψει έναν τρόπο να αποφασίζουν οι παίκτες τις κινήσεις τους ώστε να έχουν τις μεγαλύτερες αποδόσεις, ανεξάρτητα από τι στρατηγικές θα ακολουθήσει ο αντίπαλος του [26].

Στο παίγνιο 2.2 είναι προφανές ότι η απόδοση για την R εξαρτάται σε μεγάλο βαθμό από την απάντηση του παίκτη C. Αν διαλέξει R2 μπορεί να χάσει 2 ή να πάρει 9 μονάδες εφόσον ο αντίπαλος διαλέξει C1 ή C2 αντίστοιχα. Σκοπός όμως είναι η R να έχει μια ορισμένη απόδοση ανεξάρτητα από τις επιλογές του C. Στην συνέχεια υποτίθεται ότι η παίκτρια R παίζει απαισιόδοξα. Η χειρότερη απόδοση για την R για την κίνηση R2 θα είναι η –2, έναντι 9 για C2 και 1 για C3 αντίστοιχα επιλογές του C. Συμβολίζεται η χαμηλότερη αυτή απόδοση ως Min(R2) = –2. Με το ίδιο σκεπτικό αν η R επέλεγε R1 η χειρότερη απόδοση είναι -12. Χρησιμοποιώντας τον ίδιο συμβολισμό, έχουμε Min(R1) = –12. Τέλος το χειρότερο αποτέλεσμα για την R αν επιλέξει την κίνηση R3 είναι η Min(R3) = –1. Το λιγότερο «κακό», από αυτά τα ελάχιστα, προφανώς είναι εκείνο που αντιστοιχεί στην κίνηση R3. Δηλαδή, αν η R παίξει R3, μπορεί να εξασφαλίσει την καλύτερη από τις χειρότερες αποδόσεις (δηλ. –1, αντί είτε –2 ή –12). Παίζοντας απαισιόδοξα η R έχει κάθε λόγο να επιλέξει τη στρατηγική με την οποία μεγιστοποιεί τη συνάρτηση Min(Ri), όπου i = 1, 2, 3. Με άλλα λόγια, ο απαισιόδοξός παίκτης θα επιλέξει τη στρατηγική Ri που αντιστοιχεί στη λεγόμενη maximin, ή στη στρατηγική που επιφέρει την μεγιστοποίηση του ελάχιστου οφέλους, του παίκτη.

Εξετάζονται οι επιλογές του παίκτη C παίζοντας και αυτός απαισιόδοξα, κάτω από την ίδια οπτική. Αν ο παίκτης C επιλέξει την κίνηση C1, το χειρότερο που μπορεί να του συμβεί είναι να επιλέξει ο αντίπαλός του την κίνηση R3, οπότε η απόδοση του θα είναι Min(C1) = +1. Ομοίως, τα ελάχιστα Min(C2) = –9 και Min(C3) = –2. Η maximin, επομένως, του παίκτη C αντιστοιχεί στη στρατηγική C1 και είναι ίση με +1, η καλύτερη από τις χειρότερες δυνατές αποδόσεις του.

Παρατηρείται ότι αν αθροιστούν τα maximin των δύο παικτών το αποτέλεσμα είναι ίσο με μηδέν, αφού το maximin της R είναι ίσο με –1 και το maximin C είναι ίσο με +1. Με την παρατήρηση αυτή ο von Neumann δημιούργησε ένα θεώρημα με το οποίο την γενικεύει για όλα τα παίγνια μηδενικού αθροίσματος. Έδειξε δηλαδή ότι στα παίγνια αυτά δύο παικτών το άθροισμα των maximin τους είναι πάντα μηδέν. Το θεώρημα αυτό ήταν το πρώτο που περιλάμβανε μια ολόκληρη κατηγορία παιγνίων και το οποίο ανακάλυπτε κάποια ιδιότητα που ίσχυε για όλα τα παίγνια αυτά. Γι’ αυτό και το θεώρημα αυτό λέγεται ότι ήταν και η αρχή της θεωρίας παιγνίων. Μια άλλη διατύπωση του θεωρήματος αυτού είναι ότι η ελάχιστη από τις μέγιστες απώλειες της παίκτριας R θα είναι πάντοτε ίση με το μεγαλύτερο από τα μικρότερα κέρδη του παίκτη C.

Θεωρείται ότι οι παίκτες επιλέγουν ορθολογικά απαισιόδοξα, κατά τον von Neumann, διότι σε ένα παίγνιο μηδενικού αθροίσματος το κέρδος του ενός είναι η ζημία του άλλου, και αντίστροφα. Αυτό είναι δυνατόν απλά να σημαίνει ότι ένας παίκτης καταλαβαίνει ότι ο αντίπαλός θα επιδιώκει πάντα να του επιφέρει όσο περισσότερη ζημιά γίνεται και αυτό διότι η δική του ζημία είναι το κέρδος του άλλου. Σύμφωνα με τον von Neumann υπάρχει ένα μοναδικό αποτέλεσμα που θα προσελκύει τους παίκτες, το αποτέλεσμα (R3,C1), που αντιστοιχεί στη στρατηγική maximin κάθε παίκτη, μεγιστοποιώντας το μικρότερο κέρδος τους, αλλά και για τον ίδιο λόγο ελαχιστοποιεί τη χειρότερή του ζημία δεδομένης της «απληστίας» του άλλου. Κοιτώντας ξανά το αποτέλεσμα (R3,C1) που δίνει και στους δύο παίκτες τις maximin αποδόσεις τους, γίνεται η σύγκριση με ένα άλλο αποτέλεσμα, όπως το (R1,C2). Με το αποτέλεσμα αυτό προκαλείται στην R πολύ μεγαλύτερη ζημιά από το maximin τη, –12 έναντι –1, ενώ στον C αποδίδεται πολύ μεγαλύτερο κέρδος από ότι το δικό του maximin, +12 έναντι +1. Σύμφωνα με τον von Neumann η ορθολογική R δεν θα δεχόταν κάτι λιγότερο από την maximin απόδοσή της. Οπότε αν και οι δύο παίκτες δεν θέλουν να εισπράξουν μια απόδοση κάτω από το maximin τους, τότε αντίστοιχα ούτε ο C θα περιμένει ότι θα καταφέρει να εισπράξει περισσότερα από τη δική του maximin απόδοση και ούτε θα το επιχειρήσει.

Συμπεραίνεται, ότι αν κανένας παίκτης δεν περιμένει να πάρει περισσότερα ή λιγότερα από τη maximin απόδοσή του άρα και οι δύο παίκτες θα έχουν στραμμένο το βλέμμα τους στη maximin απόδοσή τους. Έτσι το θεώρημα του von Neumann οδηγεί στη επίλυση των παιγνίων μηδενικού αθροίσματος δύο ατόμων [1]. Οι αντίπαλοι προσπαθούν να αποκτήσουν τα maximin οφέλη τους και όχι κάτι καλύτερο ή χειρότερο. Οι ορθολογικοί παίκτες είναι σαν να έλκονται από τη στρατηγική που τους εξασφαλίζει το μέγιστο εγγυημένο όφελος και είναι η πρώτη έννοια ισορροπίας στην θεωρία παιγνίων.

### 2.4.3.5 Η ισορροπία του Nash

Αναμφισβήτητα το θεώρημα minimax και η λύση που συνδέεται με αυτό ήταν έργο μιας μεγαλοφυΐας, το θεώρημα που γέννησε ίσως την θεωρίας παιγνίων [1]. Παραχώρησε τη θέση του όμως σε ένα άλλο αξιοσημείωτο συμπέρασμα που αποτέλεσε, και αποτελεί ως σήμερα, τον πυρήνα των λύσεων της Θεωρίας Παιγνίων, την ισορροπία John Nash.

Το πρόβλημα της θεωρίας του von Neumann είναι το εξής, σε παίγνια μη μηδενικού αθροίσματος δεν μπορούμε να υποθέσει κάνεις ότι οι ορθολογικά σκεπτόμενοι παίκτες θα ακολουθούσαν αυτήν την συμπεριφορά. Για να φανεί γιατί γίνεται αυτό ακολουθεί παίγνιο 2.3 από το οποίο γίνεται φανερό ότι στα παίγνια μη μηδενικού αθροίσματος, σε παίγνια στα οποία η ζημιά του ενός δεν σημαίνει υποχρεωτικά κέρδος του άλλου, η μεγιστοποίηση του ελάχιστου κέρδους δεν έχει βάση.

| | C1 | C2 |
|---|---|---|
| R1 | 500, 500 | 0, -2 |
| R2 | -2, 0 | 3, 3 |

**Παίγνιο 2. 3 Παίγνιο μη μηδενικού αθροίσματος**

Στο Παίγνιο 2.3 η πρώτη στρατηγική, R1 για την R και C1 για τον C, φαίνεται ελκυστική. Όμως η μέθοδος von Neumann υποδεικνύει την δεύτερη (R2, C2). Ωστόσο δεν υπάρχει βάσιμος λόγος για τον οποίο θα ήταν ορθολογικό για την παίκτρια R και τον παίκτη C να επιλέξουν το (R2,C2), παραβλέποντας τη δελεαστική προοπτική του αμοιβαίου οφέλους που τους προσφέρει το (R1,C1). Επί πλέον υπάρχει ένας άλλος λόγος που το θεώρημα του von Neumann δεν βοηθάει σε αυτήν την περίπτωση, πολύ απλά δεν ισχύει. Οι maximin αποδόσεις των δύο παικτών δεν αθροίζονται στο μηδέν.

Ακολουθεί ένα άλλο παράδειγμα αλληλεπίδρασης μη μηδενικού αθροίσματος, το Παίγνιο 2.1 που έχει ήδη αναφερθεί:

| | C1 | C2 |
|---|---|---|
| R1 | 8, 8 - | 0, 2 |
| R2 | + 9, 3 | + 1, 4 - |

**Παίγνιο 2. 1 Τα σύμβολα (+, -) πλάι στις αποδόσεις (pay-offs) δείχνουν τις στρατηγικές της «βέλτιστης αντίδρασης» που εξηγούνται στην συνέχεια.**

Στο παίγνιο αυτό, η maximin απόδοση της παίκτριας R είναι 1 και αντιστοιχεί στο R2, ενώ η maximin του C είναι 3 και αντιστοιχεί στο C1. Άρα η μέθοδος von Neumann θα έδειχνε το αποτέλεσμα (R2,C1). Παρατηρείται όπως και πριν ότι το θεώρημα του von Neumann δεν ισχύει, καθώς οι maximin αποδόσεις των δύο παικτών αθροίζουν 4, ο von Neumann βέβαια γνώριζε ότι το θεώρημά του δεν ισχύει για παίγνια μη μηδενικού αθροίσματος. Αν ο παίκτης C σκέπτεται ότι ο von Neumann έχει δίκιο προβλέποντας ότι η R θα επιλέξει τη στρατηγική R2, δεν θα επέλεγε το C1, σκεπτόμενος λογικά. Θα ήταν προτιμότερο να επιλέξει C2, εξασφαλίζοντας έτσι απόδοση 4 αντί μιας απόδοσης 3 μονάδων.

Το σκεπτικό αυτό αποτελεί το κέντρο της σύγχρονης θεωρίας παιγνίων με αποκορύφωμά της την έννοια ισορροπίας που διατυπώθηκε από τον John Nash το 1949. Σε αντίθεση με τη μέθοδο που διατυπώθηκε από τον von Neumann, που είχε σκοπό την υιοθέτηση μιας ορθολογικής στρατηγικής από κάθε άτομο ανεξάρτητα από τις αντιλήψεις του σχετικά με την πιθανότερη συμπεριφορά του αντιπάλου του, ο Nash υποθέτει ακριβώς το ανάποδο. Σκοπός ενός παίκτη είναι να έχει την μέγιστη δυνατή, προσδοκώμενη, ωφέλεια σε συνάρτηση με το τι περιμένει να διαλέξει ο αντίπαλος [1].

Κατά τον Nash ένας παίκτης πρέπει να έχει κάποιες προσδοκίες για την στρατηγική του αντιπάλου πριν διαμορφώσει την δικιά του. Παρατηρείται πόσο καλά ταιριάζει η ιδέα αυτή με την προηγούμενη επισήμανση ότι, στο Παίγνιο 2.1, αν ο παίκτης C προβλέπει ότι η παίκτρια R θα επιλέξει R2, η βέλτιστη επιλογή του είναι να επιλέξει C2. Αλλά ο παίκτης C πρέπει να έχει κάποιο καλό λόγο να θεωρεί ότι η αντίπαλός του θα επιλέξει πράγματι R2. Αν η R σκέπτεται λογικά, θα προσέξει το απλό γεγονός ότι η κίνηση R2 είναι η καλύτερη για την ίδια ό,τι και αν επιλέξει ο αντίπαλός της. Επομένως, αν ο C συμπεράνει ότι η R θα συνειδητοποιήσει το γεγονός αυτό, τότε θα προβλέψει ότι αυτή θα επιλέξει R2 οπότε η βέλτιστη δική του επιλογή είναι η στρατηγική κίνηση C2. Δεν έχει καμία σημασία το ότι η C2 δεν αποτελεί maximin στρατηγική. Το θεώρημα του von Neumann αποτελεί λύση για τα προβλήματα μηδενικού αθροίσματος για τον λόγο ότι συμπίπτει πάντα με τις στρατηγικές που οδηγείται η μέθοδος του Nash. Στα παίγνια μηδενικού αθροίσματος η απαισιοδοξία εκλογικεύεται από το γεγονός ότι ο ενός παίκτης επιδιώκει να μεγιστοποιήσει τις αποδόσεις του και αυτό ισοδυναμεί με προσπάθεια να ελαχιστοποιήσει τις αποδόσεις του άλλου.

Κάθε παίκτης προσπαθεί να βρει ποια είναι η καλύτερη αντίδρασή του σε κάθε στρατηγική του αντιπάλου του. Για να το πράξουν αυτό μπορεί να χρειαστεί επίσης να διακρίνουν ποια είναι η καλύτερη αντίδραση του αντιπάλου τους στις δικές τους στρατηγικές. Για να διεκπεραιωθεί η διαδικασία αυτή της εξαγωγής συμπεράσματος, προστίθενται τα σύμβολα (+) και (-) πλάι στις αποδόσεις τα οποία δείχνουν τις βέλτιστες αντιδράσεις των παικτών. Μια στρατηγική ενός παίκτη είναι βέλτιστη αντίδραση σε κάποια στρατηγική του αντιπάλου αν εξασφαλίζει την μέγιστη απόδοση για τη δεδομένη πάντα στρατηγική του αντιπάλου.

Αν ο C επέλεγε C1 στο παίγνιο 2.1 η R θα κέρδιζε 9 μονάδες ωφέλειας επιλέγοντας R2 και μόνο 8 εάν επέλεγε R1. Άρα η R2 είναι η καλύτερη απάντηση στην κίνηση C1. Ομοίως αν ο C επέλεγε C2 η R2 είναι πάλι η καλύτερη απάντηση με 1 έναντι 0. Με (+) λοιπόν στον πίνακα του παιγνίου 2.1 σημειώνονται οι υψηλότερες αποδώσεις της R για κάθε κίνηση του C. Σημειώνεται ότι τα σύμβολα συμπίπτουν στην στρατηγική R2. Με το ίδιο σκεπτικό σημειώνονται τα σύμβολα (-) για τον C.

Ο C παρατηρείται ότι έχει δύο διαφορετικές βέλτιστες αντιδράσεις, αντίθετα με την R που ήτανε η ίδια στρατηγική, τα σύμβολα (-) βρίσκονται σε διαφορετικές στήλες. Έτσι για τον C το τι θα πράξει εξαρτάται από το τι νομίζει ότι θα πράξει η R. Η R όμως έχει κάθε λόγο να επιμείνει στη στρατηγική R2 άρα το πιο πιθανό αποτέλεσμα είναι το (R2,C2), ο μόνος λόγος να παίξει C1 ο C είναι να πιστεύει ότι υπάρχουν πολλές πιθανότητες η R να μην είναι καλή στο παίγνιο αυτό.

Το κελί (R2,C2) είναι το μόνο στο οποίο γίνεται η ακόλουθη παρατήρηση. Είναι το μόνο κελί στο οποίο υπάρχει και το σύμβολο (+) και το σύμβολο (-). Η απόδοση 1 της R είναι η μεγαλύτερη που μπορεί να απομυζήσει μέσα στη δεύτερη στήλη (C2), δηλαδή, η R2 είναι η βέλτιστη αντίδραση της R στην κίνηση C2. Η απόδοση 4 του C είναι η μέγιστη απόδοση που μπορεί να κερδίσει ο C από την δεύτερη σειρά, δηλαδή η C2 είναι η βέλτιστη αντίδραση στην R2. Έτσι γίνεται η εξής παρατήρηση, συμπεραίνεται ότι η R2 είναι η βέλτιστη απάντηση της R στην κίνηση C2, ενώ ταυτόχρονα C2 είναι η βέλτιστη αντίδραση του C στην κίνηση R2. Με άλλα λόγια, το αποτέλεσμα (R2,C2) αντιστοιχεί σε δύο στρατηγικές που είναι οι βέλτιστες αντιδράσεις η μια στην άλλη. Αυτή ήταν και η ιδέα του Nash σχετικά με τη «λύση» των παιγνίων. Γεωμετρικά, στην κανονική μορφή αναπαράστασης των παιγνίων, η ισορροπία Nash σε καθαρές στρατηγικές αναδεικνύεται ως το κελί στο οποίο συμπίπτουν τα σύμβολα (+) και (-).

Η βασική ιδέα στην οποία βασίζεται η έννοια της ισορροπίας Nash είναι τόσο καλή λόγου της απλότητάς της. Άπαξ και οι παίκτες βρεθούν σε ισορροπία Nash, δεν μετανιώνουν για τη στρατηγική επιλογή τους, σε προσωπικό επίπεδο, με δεδομένη τη συμπεριφορά των άλλων. Υπό αυτήν την έννοια, η ιδέα του Nash αποτελεί μια ισορροπία παιγνίου βρίσκοντας ένα σημείο στο οποίο οι επιλογές και τα πιστεύω τον ορθολογικών παικτών ισορροπούν [25]. Όμως το γεγονός ότι δεν μετανιώνουν για τη στρατηγική επιλογή τους, σε προσωπικό επίπεδο με δεδομένη τη συμπεριφορά των άλλων, δεν σημαίνει ότι οι παίκτες θα κινηθούν προς την ισορροπία αυτή. Άλλωστε αυτή είναι η διαφορά ενός υπολογιστή από έναν άνθρωπο, ο άνθρωπος μπορεί και αναρωτιέται τι θα γίνει αν πάω ενάντια στην συμπεριφορά που υποτίθεται ότι είναι η καλύτερη για μένα.

#### 2.4.3.6 Αυστηρή και ασθενής κυριαρχία

Αναφέρθηκε μια λογική εξήγηση της ισορροπίας Nash στο παιγνίου 2.1. Ότι και να επιλέξει ο C η βέλτιστη απάντηση για την παίκτρια R είναι η R2. Αν ο C βέβαια επέλεγε C1 θα ήταν πολύ καλύτερο για την R, καθώς παίζοντας R2 θα είχε απόδοση 9 αντί 1 αλλά ότι και να επιλέξει ο C η R2 παραμένει βέλτιστη στρατηγική της R. Αν μια κίνηση είναι βέλτιστη απάντηση σε ότι κίνηση και να κάνει ο αντίπαλος τότε αυτή ονομάζεται αυστηρά κυρίαρχη στρατηγική ενώ οι στρατηγικές που οδηγούν με βεβαιότητα σε χειρότερα αποτελέσματα, σε σύγκριση με κάποια άλλη στρατηγική, όπως η R1 στο Παίγνιο 2.1, λέγονται αυστηρά κυριαρχούμενες στρατηγικές [49]. Εάν οι διαθέσιμες στρατηγικές είναι μόνο δύο και η μία από αυτές είναι αυστηρά κυρίαρχη, τότε η άλλη θα είναι υποχρεωτικά αυστηρά κυριαρχούμενη.

Αντίθετα, ο παίκτης C δεν έχει κυρίαρχη στρατηγική στο Παίγνιο 2.1. Η C1 είναι η βέλτιστη αντίδρασή του στην R1 και η C2 η βέλτιστη αντίδρασή του στην R2. Το γεγονός όμως ότι η R έχει κυρίαρχη στρατηγική επιτρέπει στον παίκτη C να προβλέψει ότι η R, καταλαβαίνοντας το συμφέρον της και παίζοντας ορθολογικά θα επιλέξει R2. Στην περίπτωση αυτή, η βέλτιστη απάντηση του C στην κυρίαρχη στρατηγική της R είναι C2. Με την έννοια αυτή η λογική κυριαρχίας εκλογικεύει τη μοναδική ισορροπία Nash στο παίγνιο αυτό, το αποτέλεσμα, την «λύση», (R2,C2). Στην συνέχεια ακολουθεί μια ασθενέστερη μορφή λογικής κυριαρχίας η οποία, εξαιτίας ακριβώς της «αδυναμίας» της, δεν μπορεί να εκλογικεύσει τόσο ισχυρά την ισορροπία Nash όσο η ισχυρή κυριαρχία στο Παίγνιο 2.1.

| | C1 | C2 |
|---|---|---|
| R1 | 0, 0 | +8, 4- |
| R2 | +4, 8- | +8, 8- |

**Παίγνιο 2. 4 Ασθενώς κυριαρχούμενες στρατηγικές**

Στο Παίγνιο 2.4 υπάρχουν τρία κελιά στα οποία συμπίπτουν τα σύμβολα (+) και (-) των βέλτιστων αντιδράσεων: (R1,C2), (R2,C1) και (R2,C2). Συνεπώς και τα τρία αυτά αποτελέσματα αποτελούν ισορροπίες Nash καθαρών στρατηγικών. Το μοναδικό κελί που δεν αντιστοιχεί σε ισορροπία Nash είναι το (R1,C1). Κοιτώντας τον πίνακα παρατηρείται ότι οι R1 και R2 είναι εξίσου καλές αντιδράσεις στην κίνηση C2. Ωστόσο η R έχει μια σαφή προτίμηση όταν περιμένει ότι ο C θα παίξει C1, η βέλτιστη αντίδρασή της στην κίνηση C1 είναι η R2. Σύμφωνα με τα παραπάνω η R2 ορίζεται ως μια ασθενώς κυρίαρχη στρατηγική. Πρόκειται δηλαδή για μια στρατηγική που έχει καλύτερο αποτέλεσμα στη μία από τις δύο στρατηγικές του αντιπάλου (C1), χωρίς να έχει χειρότερο αποτέλεσμα στην άλλη στρατηγική του αντιπάλου (C2). Προφανώς αφού στο παίγνιο αυτό υπάρχουν μόνο δύο στρατηγικές η ασθενής κυριαρχία της στρατηγικής R2 σημαίνει ότι η R1 είναι ασθενώς κυριαρχούμενη. Τα ίδια ισχύουν και για τον παίκτη C καθώς η στρατηγική C2 του είναι ασθενώς κυρίαρχη και η C1 ασθενώς κυριαρχούμενη.

Η ασθενής κυριαρχία εξηγεί το λόγο για τον οποίο ορισμένα αποτελέσματα δεν προσδιορίζονται ως ισορροπίες Nash του παιγνίου. Για παράδειγμα το αποτέλεσμα (R1,C1) επειδή αυτό προκύπτει μόνο αν οι παίκτες επιλέξουν τις ασθενώς κυριαρχούμενες στρατηγικές τους. Η αξία όμως της θεωρητικής αυτής έννοιας της ασθενούς κυριαρχίας είναι περιορισμένη. Ακόμη και αν είναι αληθές ότι το αποτέλεσμα (R1,C1) δεν θα προκύψει όταν η R προσδοκά ότι ο C θα επιλέξει C1 και ο C προσδοκά ότι η R θα επιλέξει R1, γιατί και για τους δύο παίκτες η βέλτιστη απάντηση είναι η δεύτερη στρατηγική τους δηλαδή το συγκεκριμένο αποτέλεσμα δεν θα προκύψει ως ισορροπία Nash, δε σημαίνει ότι δεν μπορεί να προκύψει και καθόλου. Θα μπορούσε κάλλιστα να προκύψει αν η R περιμένει ότι ο C θα διαλέξει την ασθενώς κυρίαρχη στρατηγική του C2 και ο C προσδοκά ότι η R θα επιλέξει την ασθενώς κυρίαρχη στρατηγική της R2. Αυτό γίνεται γιατί με δεδομένες τις προσδοκίες αυτές, για την R δεν έχει καμιά διαφορά το αν θα επιλέξει R1 ή R2 και αντίστοιχα για τον C δεν έχει επίσης καμιά διαφορά το αν θα επιλέξει C1 ή C2 καθώς και οι δύο στις αντίστοιχες περιπτώσεις θα κέρδιζαν 8 μονάδες ωφέλειας. Σύμφωνα με αυτά δηλαδή είναι δυνατόν οι παίκτες να καταλήξουν στο αποτέλεσμα (R1,C1).

Αυτό που συμπεραίνεται είναι το ακόλουθο: Ενώ μια ισορροπία Nash που υποστηρίζεται από τη λογική της αυστηρής κυριαρχίας είναι σταθερή, όπως είναι η (R2,C2) στο παίγνιο 2.1, οι ισορροπίες Nash που υποστηρίζονται μόνο από τη λογική της ασθενούς κυριαρχίας είναι ασταθείς, όπως είναι η (R2,C2) στο Παίγνιο 2.4 ή οι ισορροπίες Nash σε μικτές στρατηγικές.

### 2.4.3.7 Ισορροπία Nash σε μικτές στρατηγικές (ΙΝΜΣ)

Στην συνέχεια εξετάζονται παίγνια τα οποία, αντίθετα με αυτά που είδαμε ως τώρα, δεν έχουν μοναδικές λύσεις είτε γιατί δεν υπάρχουν καθόλου ισορροπίες σε αυτά, παίγνιο 2.9, είτε επειδή υπάρχουν πολλαπλές ισορροπίες Nash, σε καθαρές στρατηγικές όπως είναι τα 2.5 έως 2.8 στα οποία υπάρχουν δύο ισορροπίες στο καθένα. Σύμφωνα με αυτά που έχουνε αναφερθεί ως τώρα η ισορροπία Nash δεν προσφέρει συμβουλές στο τι πρέπει να επιλέξουν οι παίκτες. Ο Nash όμως έχει δώσει απάντηση και σε τέτοιου είδους προβλήματα, η οποία ονομάζεται ισορροπία Nash σε μικτές στρατηγικές (ΙΝΜΣ).

Έχουν αναφερθεί δύο τύποι στρατηγικής: η καθαρή και η μικτή στρατηγική. Η επιλογή καθαρής στρατηγικής αναφέρεται σε συγκεκριμένες κινήσεις ενώ η επιλογή μικτής στρατηγικής ισοδυναμεί με το να γίνει επιλογή τυχαία μεταξύ, συγκεκριμένων, καθαρών στρατηγικών, για παράδειγμα «η R επιλέγει την R1 με πιθανότητα p ή την R2 με πιθανότητα 1–p». Επιλέγοντας δηλαδή μικτή στρατηγική ένας παίκτης επιλέγει τις πιθανότητες κάθε μιας από τις καθαρές στρατηγικές που εμπεριέχονται στην συγκεκριμένη μικτή στρατηγική, αφήνοντας τα υπόλοιπα στην τύχη.

Οι παίκτες καλούνται να επιλέξουν στην τύχη σε δύο περιπτώσεις: όταν δεν ξέρουν τι πρέπει να επιλέξουν, οπότε είναι φυσικό να εμπεριέχεται μια δόση τύχης στην επιλογή τους, και όταν θέλουν να παραπλανήσουν τον αντίπαλό τους, καθώς εάν δεν γνωρίζει ο παίκτης τι θα επιλέξει δεν είναι δυνατόν να ξέρει ο αντίπαλός του. Ο Nash όμως, και η θεωρία παιγνίων κατ' επέκταση, κατάφερε να βρει τρόπο να υπολογίζει τις πιθανότητες με τις οποίες οι παίκτες που σκέφτονται ορθολογικά θα επιλέξουν μεταξύ των καθαρών στρατηγικών τους.

Το θεώρημα του Nash ήταν απλό: Σε κάθε πεπερασμένο παίγνιο υπάρχει μια ισορροπία Nash σε μικτές στρατηγικές, επιπλέον των οποιονδήποτε ισορροπιών Nash σε καθαρές στρατηγικές. Όταν λοιπόν οι παίκτες δεν έχουν κάποιο λόγο να επιλέξουνε μια από τις καθαρές στρατηγικές τους, γιατί όλες είναι ισορροπίες Nash είτε δεν υπάρχουν καθόλου, και είναι δηλαδή στρατηγικά αδιάφοροι μεταξύ αυτών τότε επιλέγουν τυχαία μεταξύ των στρατηγικών αυτών. Η Ισορροπία Nash σε Μικτές Στρατηγικές για καταστάσεις παιγνίων όπως αυτών που αναφέρθηκαν είναι μία ισορροπία στην οποία οι παίκτες επιλέγουν μεταξύ των καθαρών στρατηγικών τους, με συγκεκριμένες πιθανότητες οι οποίες εξαρτώνται αποκλειστικά από την στρατηγική δομή του κάθε παιγνίου, δηλαδή από τον πίνακα των αποδόσεων των παικτών.

Για να είναι ένα παίκτης αδιάφορος, η R για παράδειγμα, θα πρέπει οι προσδοκώμενες ωφέλειες της από τις δύο στρατηγικές, R1 και R2, να είναι ακριβώς ίσες. Δηλαδή EU(R1) = EU(R2), EU από το expected utility που είναι ο προσδοκώμενη ωφέλεια. Το ίδιο θα πρέπει αντίστοιχα να ισχύει και για το παίκτη C, οι προσδοκώμενες ωφέλειες και των δύο στρατηγικών του να είναι ίσες δηλαδή EU(C1) = EU(C2). Άρα εφόσον είναι αδιάφοροι προς της στρατηγικές πρέπει να επιλέξουν τυχαία. Έστω ότι η R επιλέγει R1 και R2 με πιθανότητες p και 1-p αντίστοιχα και ο C για C1 και C2 q και 1-q αντίστοιχα. Για να είναι συνεπής η τυχαία επιλογή οι πιθανότητες (p,q) πρέπει να είναι τέτοιες ώστε EU(R1) = EU(R2) και EU(C1) = EU(C2).

Στο παίγνιο 2.5 αν η R υιοθετήσει την στρατηγική R1 αποκομίζει είτε απόδοση 2 είτε απόδοση 0. Την πρώτη θα την αποκομίσει με πιθανότητα q, δηλαδή την πιθανότητα να επιλέξει C1 ο C, και την δεύτερη με 1-q. Αν πάλι επιλέξει R2 αναμένει απόδοση 0 με πιθανότητα q και 1 με 1-q. Οπότε:

| | C1 | C2 | |
|---|---|---|---|
| R1 | +2,1- | 0,0 | ⅓ |
| R2 | 0,0 | +1,2- | ⅔ |
| | ⅓ | ⅔ | ΙΝΜΣ |

**Παίγνιο 2.5**

ΜΤΦ: Μάχη των Φύλων

| | C1 | C2 | |
|---|---|---|---|
| R1 | 3,3 | +1,4- | ½ |
| R2 | +4,1- | 0,0 | ½ |
| | ½ | ½ | ΙΝΜΣ |

**Παίγνιο 2.6**

Γεράκι-Περιστέρι

| | C1 | C2 | |
|---|---|---|---|
| R1 | +2,2- | 0,0 | ⅓ |
| R2 | 0,0 | +1,1- | ⅔ |
| | ⅓ | ⅔ | ΙΝΜΣ |

**Παίγνιο 2.7**

Παίγνιο συντονισμού

| | C1 | C2 | |
|---|---|---|---|
| R1 | +1,1- | 2,0 | ½ |
| R2 | 0,2 | +3,3- | ½ |
| | ½ | ½ | ΙΝΜΣ |

**Παίγνιο 2.8**

Κυνήγι ελαφιού

| | C1 | C2 | |
|---|---|---|---|
| R1 | +1,0 | 0,1- | ½ |
| R2 | 0,1- | +1,0 | ½ |
| | ½ | ½ | ΙΝΜΣ |

**Παίγνιο 2.9**

Κρυφτούλι

| | C1 | C2 | |
|---|---|---|---|
| R1 | -1,-1 | -10,0- | 0 |
| R2 | +0,-10 | +-5,-5- | 1 |
| | 0 | 1 | ΙΝΜΣ |

**Παίγνιο 2.10**

Δίλημμα του κρατούμενου

**Παίγνιο 2. 5,6,7,8,9, 10**

EU(R1) = 2 × q + 0× (1 – q) = 2q

EU(R2) = 0 × q + 1 × (1 – q) = 1-q

Αφού EU(R1) = EU(R2) τότε 2q=1-q δηλαδή q = ⅓.

Ο C δηλαδή επιλέγει C1 με πιθανότητα q=⅓ και C2 με πιθανότητα 1-q=⅔. Η R, με τις ανάλογες πράξεις, βρίσκεται ότι επιλέγει R1 με p=⅓ και R2 1-p=⅔. Στο παίγνιο 2.10, που είναι το κλασικό δίλλημα του κρατουμένου p=1 καθώς έχει μια και μοναδική ισορροπία η οποία είναι αυστηρώς κυρίαρχη.

Συνοψίζοντας, επειδή ακριβώς οι παίκτες δεν έχουν κανέναν λόγο να επιλέξουν μεταξύ των καθαρών στρατηγικών τους, θα πρέπει να επιλέγουν την δεύτερη τους, καθαρή, στρατηγική (R2 η R και C2 ο C) με πιθανότητα ⅔. Αυτές οι πιθανότητες μάλιστα συνιστούν μια ισορροπία Nash.

### 2.4.3.8 Ισορροπία Nash σε μικτές στρατηγικές (ΙΝΜΣ) για παίγνια 3x3

Η βασική αρχή είναι απλή: Εξετάζονται οι καθαρές στρατηγικές των παικτών ανά δύο με σκοπό να ελεγχθεί αν αντιστοιχούν σε μια ΙΝΜΣ. Έτσι βρίσκονται όλες οι ΙΝΜΣ οι οποίες δίνουν θετική πιθανότητα σε δύο από τις τρεις καθαρές στρατηγικές του κάθε παίκτη. Στο τέλος ελέγχεται αν ορίζεται μια ΙΝΜΣ που να δίνει θετική πιθανότητα και στις τρεις καθαρές στρατηγικές του κάθε παίκτη.

| | C1 | C2 | C3 |
|---|---|---|---|
| R1 | $^{+}5, 5^{-}$ | 0, 0 | 0, 0 |
| R2 | 0, 0 | $^{+}50, 49$ | $49, 50^{-}$ |
| R3 | 0, 0 | $49, 50^{-}$ | $^{+}50, 49$ |

**Παίγνιο 2. 11**

Παρακάτω αναλύεται το παίγνιο 2.11. Ξεκινώντας με τις μικτές στρατηγικές που δίνουν μηδενική πιθανότητα στην τρίτη στρατηγική του κάθε παίκτη, υποθέτοντας δηλαδή ότι οι παίκτες επιλέγουν τις εξής μικτές στρατηγικές:

Μικτές στρατηγικές ΜΣ 12 (ένα και δύο)

P(R1)=p, P(R2)=1-p, P(R3)=0 και

P(C1)=q, P(C2)=1-q, P(C3)=0

Ας δούμε αν αυτές οι μικτές στρατηγικές, οι κατανομές πιθανοτήτων, της R [(p,1-p),0] και του C [(q,1-q),0] ορίζονται σε κατάσταση ισορροπίας Nash, δηλαδή, αν οι ΜΣ 12 αποτελούν ΙΝΜΣ. Αν ο C επιλέξει αυτή την μικτή του στρατηγική και η R την δική της τότε σε περίπτωση που η τελευταία οδηγήσει την R να επιλέξει την καθαρή στρατηγική R1, η προσδοκώμενη ωφέλειά της είναι EU(R1) = 5q. Αν παίξει R2, η προσδοκώμενη ωφέλειά της είναι EU(R2) = 50(1-q). Την R3 στο πλαίσιο της συγκεκριμένης μικτής στρατηγικής δεν μπορεί ποτέ να την επιλέξει καθώς Pr(R3)=0. Για να είναι αδιάφορη λοιπόν η R όταν ο C επιλέγει την μικτή στρατηγική του [(q,1-q),0], τότε EU(R1)=EU(R2) ή q=50/55. Αντίστοιχα για τον C μας οδηγεί στον υπολογισμό p=49/54.

Η πρώτη ΙΝΜΣ του Παιγνίου 2.11 λέει στην R να παίξει R1 με πιθανότητα p=49/54 και R2 με πιθανότητα 1-p=5/54, και στον C να παίξει C1 με πιθανότητα q=50/55 και C2 με πιθανότητα 1-q=5/55. Δηλαδή η πρώτη ΙΝΜΣ, ΜΣ 12, είναι για την R (49/54, 5/54,0) και για τον C (50/55, 5/55,0).

Πρέπει να εξεταστεί στην συνέχεια αν το παίγνιο 2.11 έχει και άλλη ΙΝΜΣ. Αυτή τη φορά θα εξεταστούν μικτές στρατηγικές που δίνουν μηδενική πιθανότητα στην πρώτη στρατηγική κάθε παίκτη, που είναι και η μοναδική ισορροπία Nash σε καθαρές στρατηγικές. Φαίνεται περίεργο να υπάρχει ΙΝΜΣ που να δίνει μηδενική πιθανότητα στη μοναδική ισορροπία Nash αλλά είναι δυνατόν όπως θα φανεί στην συνέχεια.

Μικτές στρατηγικές ΜΣ 23

P(R1)=0, P(R2)=p, P(R3)=1-p και

P(C1)=0, P(C2)=q, P(C3)=1-q.

Οι εξισώσεις αδιαφορίας είναι:

EU(R2) [= 50q+49(1-q)] = EU(R3) [= 49q+50(1-q)]

EU(C2) [= 49p+50(1-p)] = EU(C3) [= 50p+49(1-p)]

Η μοναδική λύση των παραπάνω δίνει: p=q=½. Σε αυτήν την ΙΝΜΣ οι παίκτες επιλέγουν τυχαία μεταξύ της δεύτερης και της τρίτης στρατηγικής και αγνοούν την πρώτη.

Στην συνέχεια πρέπει να εξεταστεί αν το Παίγνιο 2.11 διαθέτει και τρίτη ΙΝΜΣ η οποία αυτή τη φορά να δίνει μηδενική πιθανότητα στην δεύτερη στρατηγική του κάθε παίκτη.

Μικτές στρατηγικές ΜΣ 13

P(R1)=p, P(R2)= 0, P(R3)=1-p και

P(C1)=q, P(C2)= 0, P(C3)=1-q.

Οι εξισώσεις αδιαφορίας είναι:

EU(R1) [= 5q] = EU(R3) [= 50(1-q)]

EU(C1) [= 5p] = EU(C3) [= 49(1-p)]

Στην παραπάνω περίπτωση p=49/54 και q=50/55. Σε αυτήν την ΙΝΜΣ οι παίκτες αγνοούν την δεύτερη στρατηγική τους ενώ παράλληλα η R τείνει προς την R1, την διαλέγει με πιθανότητα 49/54, αφήνοντας μια μικρή πιθανότητα 5/54 με την οποία επιλέγει την R3. Αντίστοιχα ο C επιλέγει C1 με πιθανότητα 50/55 και με μικρή πιθανότητα 5/55 την C3.

Το τελευταίο βήμα είναι να δούμε αν υπάρχει μια μικτή στρατηγική που δίνει θετικές πιθανότητες και στις τρεις καθαρές στρατηγικές του κάθε παίκτη:

Μικτές στρατηγικές ΜΣ 123

P(R1)=p, P(R2)= r, P(R3)=1-p-r και

P(C1)=q, P(C2)= s, P(C3)=1-q-s

Οι εξισώσεις αδιαφορίας των παικτών αφορούν σε αυτή την περίπτωση και τις τρεις καθαρές στρατηγικές τους, μιας και οι εν λόγω μικτές στρατηγικές δίνουν θετική πιθανότητα σε κάθε μια καθαρή στρατηγική:

EU(R1) [=5q] = EU(R2) [= 50s+49(1-q-s) ]= EU(R3) [= 49s+50(1-q-s)]

EU(C1) [=5p] = EU(C2) [= 49r+50(1-p-r)] = EU(C3) [= 50r+49(1-p-r)]

Όμως σε αυτή την περίπτωση οι παραπάνω εξισώσεις αδιαφορίας είναι αδύνατον να ικανοποιηθούν ταυτόχρονα. Καταλήγουμε συνεπώς ότι μικτές στρατηγικές τύπου ΜΣ 123 δεν συνάδουν με ΙΝΜΣ. Συνοπτικά, το Παίγνιο 2.11 χαρακτηρίζεται από τέσσερεις ΙΝΜΣ: τις τρεις που βρέθηκαν παραπάνω και την ισορροπία Nash καθαρών στρατηγικών εκφρασμένη ως ΙΝΜΣ, δηλαδή, «παίξτε την δεύτερη στρατηγική σας με πιθανότητα 1, ή 100%»:

**Πίνακας 1 ΙΝΜΣ παιγνίου 2.11**

| **Παίγνιο 2.11** | **ΙΝΜΣ1** | **ΙΝΜΣ2** | **ΙΝΜΣ3** | **ΙΝΜΣ4** |
|---|---|---|---|---|
| Καθαρή στρατηγική | $ΜΣ_{12}$ | $ΜΣ_{23}$ | $ΜΣ_{13}$ | Καθ. Στρ. |
| R1 | 49/54 | 0 | 49/54 | 1 |
| R2 | 5/54 | ½ | 0 | 0 |
| R3 | 0 | ½ | 5/54 | 0 |
| C1 | 50/55 | 0 | 50/55 | 1 |
| C2 | 5/55 | ½ | 0 | 0 |
| C3 | 0 | ½ | 5/55 | 0 |

### 2.4.4 Επαναλαμβανόμενα παίγνια

Μια ειδική κατηγορία παιγνίων είναι τα επαναλαμβανόμενα παίγνια. Στην θεωρία παιγνίων ένα επαναλαμβανόμενο παίγνιο αποτελείται από κάποιο αριθμό επαναλήψεων κάποιου βασικού παιχνιδιού (αποκαλούμενου παίγνιο σκηνή) [12]. Αυτό συνήθως είναι ένα από τα καλά μελετημένα παίγνια δύο παικτών. Ένα επαναλαμβανόμενο παίγνιο αιχμαλωτίζει την ιδέα ότι ένας παίκτης θα πρέπει να λάβει υπόψη του τις επιπτώσεις της τρέχουσας ενέργειάς του στις μελλοντικές ενέργειες των άλλων παικτών, κάτι το οποίο αποκαλείται ενίοτε «φήμη». Η ύπαρξη διαφορετικών ιδιοτήτων ισορροπιών έγκειται στο ότι η απειλή για αντίποινα είναι αληθινή αφού ο παίκτης θα παίξει το ίδιο παίγνιο ξανά με τον ίδιο αντίπαλο [22]. Μπορεί να αποδειχθεί ότι κάθε στρατηγική που έχει κέρδος μεγαλύτερο από το minimax κέρδος, το λιγότερο καλό από όλα τα καλά οφέλη, μπορεί να είναι μια ισορροπία Nash που είναι ένα μεγάλο σύνολο στρατηγικών [25]. Τα μη επαναλαμβανόμενα παίγνια αποκαλούνται και παίγνια μίας σκηνής.

Τα επαναλαμβανόμενα παίγνια μπορούν να χωριστούν γενικά σε δύο κατηγορίες εξαρτώμενα από το εάν ο ορίζοντας, το πότε θα τελειώσουν, είναι πεπερασμένος ή άπειρος [25]. Τα αποτελέσματα σε αυτές τις δύο περιπτώσεις είναι πολύ διαφορετικά. Ακόμα και τα πεπερασμένα επαναλαμβανόμενα παίγνια δεν έχουν απαραίτητα πεπερασμένο ορίζοντα, ο παίκτης μπορεί να αντιληφθεί μια πιθανότητα ακόμα ενός γύρου και να πράξει ανάλογα. Το γεγονός ότι είναι πεπερασμένα δεν σημαίνει ότι είναι απαραίτητα συγκεκριμένος ο αριθμός των ενεργειών του κάθε παίκτη, για παράδειγμα στις τρεις κινήσεις ή νίκες. Θα μπορούσε ο ορίζοντας να είναι ένα συγκεκριμένο σκορ, άθροισμα των κερδών κάθε γύρου του παιγνίου σκηνή είτε αυτό είναι κάποιο νούμερο είτε όταν κάποιος παίκτης φτάσει στο μηδέν για παράδειγμα, κάτι το οποίο μπορεί να διαφέρει ανάλογα με τον πίνακα ωφέλειάς του. Δηλαδή να γνωρίζουν οι παίκτες πότε θα τελειώσει το παίγνιο αλλά όχι σε πόσες ενέργειες, κινήσεις, θα γίνει αυτό. Οι παίκτες επίσης μπορεί να ενεργούν διαφορετικά όταν ο ορίζοντας είναι μακριά συγκριτικά με το αν είναι κοντά, κάτι το οποίο μπορεί να θεωρηθεί σαν συνάρτηση τροποποιημένη από τον χρόνο, θα μπορούσε να μην είναι ο χρόνος σαν ώρα που περνάει αλλά σαν απόσταση από τον ορίζοντα σε κινήσεις για παράδειγμα, εφαρμοσμένη στα κέρδη [22]. Η στρατηγική τους δηλαδή μπορεί να διαφέρει για παράδειγμα στη μέση του παιχνιδιού και να είναι πιο ήπια αλλά προς το τέλος του παιγνίου να διαφοροποιηθεί. Αν έπεται στο εκάστοτε σκορ θα μπορούσε να προκαλέσει πιο επιθετικές ενέργειες με σκοπό την νίκη ή της λιγότερες απώλειες είτε αντίστοιχα αν προηγείται μια πιο συντηρητική συμπεριφορά για διατήρηση τις νίκης ή ίσως μεγιστοποίηση των κερδών. Αυτά φυσικά ισχύουν για όλα τα παίγνια με τη διαφορά ότι στα επαναλαμβανόμενα παίγνια είναι η επανάληψη του ίδιου παιγνίου σκηνή σε κάθε φάση, γύρο, του παιχνιδιού ανεξάρτητα από το σε ποιο σημείο βρίσκεται αυτό, στην αρχή ή στο τέλος. Δηλαδή όποια και να είναι η κατάσταση σε κάθε επανάληψη του παιγνίου σκηνή οι διαθέσιμες κινήσεις του εκάστοτε παίκτη και ο τυχόν πίνακας ωφέλειας, ή το κέρδος κάθε ενέργειας, για την συγκεκριμένη επανάληψη, είναι ο ίδιος. Αυτό που αλλάζει είναι το συνολικό κέρδος, αποτέλεσμα, και η απόσταση από τον ορίζοντα όπως είπαμε και αυτά είναι που διαφοροποιούν τις επιλογές του παίκτη. Η διαφοροποίηση των στρατηγικών των πεπερασμένων ενάντια στα μη πεπερασμένα παίγνια είναι ένα θέμα στο οποίο υπάρχει μεγάλη διαμάχη στον χώρο των μελετητών της θεωρίας παιγνίων και υπάρχουν πολλές διαφορετικές απόψεις.

Όπως αναφέρθηκε νωρίτερα τα πεπερασμένα παίγνια χωρίζονται σε δύο γενικές κατηγορίες. Στην πρώτη των πεπερασμένων επαναλαμβανόμενων παιγνίων όπου η διάρκεια είναι συγκεκριμένη και γνωστή, είναι βέλτιστο στον τελευταίο γύρο να παιχτεί η στρατηγική Nash, η καλύτερη επιλογή του παίκτη δηλαδή. Όταν η αμοιβή της ισορροπίας Nash είναι ίση με το minimax, το λιγότερο καλό από τα καλά αποτελέσματα, τότε ό παίκτης δεν έχει λόγο να επιμείνει σε μια κοινωνικά βέλτιστη στρατηγική και είναι ελεύθερος να παίξει μια εγωιστική στρατηγική, εφόσον η τιμωρία δεν μπορεί να τον επηρεάσει καθώς είναι ίση με την minimax απόδοση. Αυτή η απόκλιση σε εγωιστική στρατηγική ισορροπία Nash εξηγείται από το παράδοξο Chainstore. Η δεύτερη κατηγορία των πεπερασμένων επαναλαμβανόμενων παιγνίων συνήθως θεωρείται σαν επαναλαμβανόμενο παίγνιο άπειρης διάρκειας [22].

Τα πιο μελετημένα επαναλαμβανόμενα παίγνια είναι παίγνια τα οποία επαναλαμβάνονται πιθανώς άπειρες φορές. Σε πολλές περιπτώσεις η βέλτιστη μέθοδος για να παιχτεί ένα επαναλαμβανόμενο παίγνιο είναι να μην επαναλαμβάνεται η χρήση μιας στρατηγικής Nash  του παιγνίου σκηνή αλλά να γίνεται συνεργασία και να παίζεται μια κοινωνικά βέλτιστη στρατηγική. Αυτό μπορεί να μεταφραστεί σαν «κοινωνικά φυσικό» και ένα βασικό συστατικό των απείρως επαναλαμβανόμενων παιγνίων είναι η τιμωρία των παικτών που αποκλίνουν από αυτήν την στρατηγική συνεργασίας [22]. Η τιμωρία μπορεί να είναι κάτι σαν τη χρήση μιας στρατηγικής που οδηγεί σε μειωμένο κέρδος και στους δύο παίκτες για την υπόλοιπη διάρκεια του παιγνίου, αποκαλούμενη  στρατηγική ενεργοποίησης (trigger strategy). Υπάρχουν πολλά αποτελέσματα στα θεωρήματα που ασχολούνται με το πώς να επιτευχθεί και να διατηρηθεί μια κοινωνικά βέλτιστη ισορροπία στα επαναλαμβανόμενα παίγνια. Αυτά τα αποτελέσματα συλλογικά λέγονται «Παραδοσιακά θεωρήματα» (Folk Theorem) [25]. Ένα σημαντικό χαρακτηριστικό των επαναλαμβανόμενων παιγνίων είναι ο τρόπος με τον οποίο αναπαριστώνται οι προτιμήσεις κάποιου παίκτη. Υπάρχουν πολλοί διαφορετικοί τρόποι με τους οποίους μια σχέση προτιμήσεις μπορεί να αναπαρασταθεί σε ένα απείρως επαναλαμβανόμενο παίγνιο, οι κύριοι από αυτούς είναι:

- Προεξόφληση (Discounting) – η αποτίμηση του παιχνιδιού μειώνεται με τον χρόνο εξαρτώμενη από την παράμετρο προεξόφλησης δ [25]
- Όριο των μέσων (Limit of means) – μπορεί να θεωρηθεί ως μέσος όρος σε Τ περιόδους όπου το Τ τείνει στο άπειρο. $\lim_{T\to\infty} \sum_{t=1}^{T} \frac{1}{T}(v_i^t) > 0$ [25]

- Προσπέρασμα (Overtaking) – Η ακολουθία $v_i^t$ είναι ανώτερη της ακολουθίας $w_i^t$ εάν $\lim_{T\to\infty} \sum_{t=1}^{T} (v_i^t - w_i^t) > 0$ [25]

### 2.4.5 Λίστα παιγνίων της θεωρίας παιγνίων

Παρακάτω είναι μια λίστα των παιγνίων που συχνά μελετώνται στην θεωρία παιγνίων.

**Πίνακας 2 Λίστα παιγνίων**

| **Παίγνιο** | **Παίκ-τες** | **Στρατηγι-κές ανά παίκτη** | **Αριθμός ισορροπιών Nash καθαρής στρατηγικής** | **Διαδο-χικά** | **Τέλειας πληροφό-ρησης** | **Μηδενικού αθροίσματος** |
|---|---|---|---|---|---|---|
| Battle of the sexes | 2 | 2 | 2 | Όχι | Όχι | Όχι |
| Blotto games | 2 | Ποικίλουν | Ποικίλουν | Όχι | Όχι | Ναι |
| Cake cutting | Άπειροι | Άπειρες | Ποικίλουν [1] | Όχι | Ναι | Ναι |
| Centipede game | 2 | Ποικίλουν | 1 | Ναι | Ναι | Όχι |
| Chicken (aka hawk - dove) | 2 | 2 | 2 | Όχι | Όχι | Όχι |
| Coordination game | *N* | Ποικίλουν | >2 | Όχι | Όχι | Όχι |
| Cournot game | 2 | Άπειρες [2] | 1 | Όχι | Όχι | Όχι |
| Deadlock | 2 | 2 | 1 | Όχι | Όχι | Όχι |
| Dictator game | 2 | Άπειρες [2] | 1 | N/A[3] | N/A[3] | Ναι |

| **Παίγνιο** | **Παίκ-τες** | **Στρατηγι-κές ανά παίκτη** | **Αριθμός ισορροπιών Nash καθαρής στρατηγικής** | **Διαδο-χικά** | **Τέλειας πληροφό-ρησης** | **Μηδενικού αθροίσματος** |
|---|---|---|---|---|---|---|
| Diner's dilemma | *N* | 2 | 1 | Όχι | Όχι | Όχι |
| Dollar auction | 2 | 2 | 0 | Ναι | Ναι | Όχι |
| El Farol bar | *N* | 2 | Ποικίλουν | Όχι | Όχι | Όχι |
| Example of a game without a value | 2 | Άπειρες | 0 | Όχι | Ναι | Ναι |
| Guess 2/3 of the average | *N* | Άπειρες | 1 | Όχι | Όχι | Ναι |
| Kuhn poker | 2 | 27 & 64 | 0 | Ναι | Όχι | Ναι |
| Matching pennies | 2 | 2 | 0 | Όχι | Όχι | Ναι |
| Minority Game | *N* | 2 | Ποικίλουν | Όχι | Όχι | Όχι |
| Nash bargaining game | 2 | Άπειρες [2] | Άπειρες [2] | Όχι | Όχι | Όχι |
| Peace war game | *N* | Ποικίλουν | >2 | Ναι | Όχι | Όχι |
| Pirate game | *N* | Άπειρες [2] | Άπειρες [2] | Ναι | Ναι | Ναι |
| Prisoner's dilemma | 2 | 2 | 1 | Όχι | Όχι | Όχι |

| **Παίγνιο** | **Παίκ-τες** | **Στρατηγι-κές ανά παίκτη** | **Αριθμός ισορροπιών Nash καθαρής στρατηγικής** | **Διαδο-χικά** | **Τέλειας πληροφό-ρησης** | **Μηδενικού αθροίσματος** |
|---|---|---|---|---|---|---|
| Rock, Paper, Scissors | 2 | 3 | 0 | Όχι | Όχι | Ναι |
| Screening game | *N* | Ποικίλουν | Ποικίλουν | Ναι | Όχι | Όχι |
| Signaling game | *N* | Ποικίλουν | Ποικίλουν | Ναι | Όχι | Όχι |
| Stag hunt | 2 | 2 | 2 | Όχι | Όχι | Όχι |
| Traveler's dilemma | 2 | N >> 1 | 1 | Όχι | Όχι | Όχι |
| Trust game | 2 | Άπειρες | 1 | Ναι | Ναι | Όχι |
| Volunteer's dilemma | 2 | *N* | 2 | Όχι | Όχι | Όχι |
| War of attrition | 2 | 2 | 0 | Όχι | Όχι | Όχι |
| Ultimatum game | 2 | Άπειρες [2] | Άπειρες [2] | Ναι | Ναι | Όχι |
| Princess and monster game | 2 | Άπειρες | 0 | Όχι | Όχι | Ναι |

1. Για το πρόβλημα του μοιράσματος του κέικ (cake cutting) υπάρχει μια απλά λύση εάν το αντικείμενο που πρέπει να μοιραστεί είναι ομογενές, ένας παίκτης κόβει και ο άλλος διαλέγει ποιος παίρνει ποιο κομμάτι (συνεχιζόμενο για κάθε παίκτη). Με μη ομογενές αντικείμενο, όπως είναι ένα κέικ μισό σοκολάτα μισό βανίλια ή ένα κομμάτι γης με μια πηγή νερού, οι λύσεις είναι πολύ πιο περίπλοκες.
2. Μπορεί να υπάρχουν πεπερασμένες στρατηγικές ανάλογα με το πώς είναι μοιρασμένα τα αγαθά .

3. Εφόσον το παίγνιο του δικτάτορα (dictator game) περιλαμβάνει μόνο έναν παίκτη που στην ουσία διαλέγει στρατηγική (ο άλλος δεν κάνει τίποτα), δεν μπορεί να θεωρηθεί σαν διαδοχικό ούτε σαν τέλειας πληροφόρησης.

## 2.4.6 Παραδείγματα παιγνίων

### Μάχη των φύλων (Battle of the sexes)

Ένα ζευγάρι θέλει να πάει, μαζί, σε ένα κονσέρτο μουσικής του Bach ή του Stravinsky. Αυτό που τους ενδιαφέρει είναι να πάμε μαζί αλλά ο ένας προτιμάει τον Bach και ο άλλος τον Stravinsky. Αναπαριστώντας τις προτιμήσεις του κάθε ατόμου ακολουθεί το παίγνιο στον παρακάτω πίνακα.

| | Bach | Stravinsky |
|---|---|---|
| Bach | 2,1 | 0,0 |
| Stravinsky | 0,0 | 1,2 |

**Σχ. 6 Μάχη των φύλων**

Η μάχη των φύλων αναπαριστά μια κατάσταση στην οποία οι παίκτες επιθυμούν να συντονίσουν την συμπεριφορά τους αλλά έχουν αλληλοσυγκρουόμενα συμφέροντα. Το παίγνιο αυτό έχει δύο ισορροπίες Nash: (Bach , Bach) και (Stravinsky , Stravinsky). Δηλαδή υπάρχουν δύο σταθερές καταστάσεις, μία στην οποία και οι δύο παίκτες πάντα διαλέγουνε Bach και μία στην οποία διαλέγουνε πάντα Stravinsky.

### Παίγνιο συντονισμού (Coordination game)

Όπως και στην μάχη των φύλλων δύο άτομα επιθυμούν να βγούνε έξω μαζί αλλά σε αυτήν την περίπτωση συμφωνούν για το ποιο είναι το πιο επιθυμητό κονσέρτο. Το παίγνιο αναπαριστάται στον παρακάτω πίνακα.

| | Mozart | Mahler |
|---|---|---|
| Mozart | 2,2 | 0,0 |
| Mahler | 0,0 | 1,1 |

**Σχ. 7 Παίγνιο συντονισμού**

Όπως και η μάχη των φύλλων το παίγνιο αυτό έχει δύο ισορροπίες Nash: (Mozart, Mozart) και (Mahler, Mahler). Σε αντίθεση όμως με αυτό οι παίκτες εδώ έχουνε κοινό ενδιαφέρον να φτάσουνε σε μία από αυτές τις ισορροπίες, την (Mozart, Mozart). Παρόλα αυτά η έννοια της ισορροπίας Nash δεν αποκλείει μια σταθερή κατάσταση στην οποία το αποτέλεσμα είναι η κατώτερη ισορροπία (Mahler, Mahler).

### Παίγνιο Γεράκι –Περιστέρι (Hawk- Dove)

Δύο ζώα συναγωνίζονται για κάποιο θήραμα. Το καθένα μπορεί να συμπεριφέρεται είτε σαν γεράκι είτε σαν περιστέρι. Το καλύτερο αποτέλεσμα για κάθε ζώο είναι αυτό στο οποίο το ίδιο συμπεριφέρεται σαν γεράκι ενώ το άλλο σαν περιστέρι, το χειρότερο είναι αυτό στο οποίο και τα δύο ζώα συμπεριφέρονται σαν γεράκια.

| | Περιστέρι | Γεράκι |
|---|---|---|
| Περιστέρι | 3,3 | 1,4 |
| Γεράκι | 4,1 | 0,0 |

**Σχ. 8 Παίγνιο Γεράκι-Περιστέρι**

Κάθε ζώο προτιμάει να συμπεριφέρεται σαν γεράκι αν ο αντίπαλος του συμπεριφέρεται σαν περιστέρι και σαν περιστέρι αν ο αντίπαλος του συμπεριφέρεται σαν γεράκι. Το παίγνιο έχει δύο ισορροπίες Nash τις (Περιστέρι , Γεράκι) και (Γεράκι , Περιστέρι), που αντιστοιχούν στις δύο διαφορετικές συμβάσεις για τον παίκτη που υποχωρεί.

### Ταιριαστές δεκάρες (Matching Pennies)

Κάθε ένας από τους δύο παίκτες διαλέγει κορώνα ή γράμματα. Αν οι επιλογές τους διαφέρουν ο παίκτης 1 δίνει στον 2 ένα προσυμφωνημένο χρηματικό ποσό, εάν είναι ίδιες ο παίκτης 2 πληρώνει στον 1 το ποσό. Κάθε παίκτης νοιάζεται μόνο για το ποσό το οποίο θα εισπράξει.

| | Κορώνα | Γράμματα |
|---|---|---|
| Κορώνα | 1,-1 | -1,1 |
| Γράμματα | -1,1 | 1,1 |

**Σχ. 9 Ταιριαστές δεκάρες**

Ένα τέτοιο παίγνιο στο οποίο τα συμφέροντα των παικτών είναι εκ διαμέτρου αντίθετα αποκαλείται αυστηρά ανταγωνιστικό [25]. Το παίγνιο ταιριαστές δεκάρες δεν έχει καμιά ισορροπία Nash.

### 2.4.7 Το δίλλημα του κρατουμένου (The Prisoner's Dilemma)

Το δίλλημα του κρατουμένου είναι ένα θεμελιώδες πρόβλημα της θεωρίας παιγνίων που δείχνει γιατί δύο άτομα είναι πιθανό να μην συνεργαστούνε ακόμα και αν είναι προς το κοινό τους συμφέρον να το κάνουν. Αρχικά πλαισιώθηκε από τους Merrill Flood και Melvin Dresher εταιρεία RAND το 1950. Ο Albert W. Tucker επισημοποίησε το παίγνιο προσθέτοντας τις ποινές φυλάκισης στον πίνακα αποδόσεων και του έδωσε το όνομα δίλλημα του κρατουμένου (prisoner's dilemma) [33]. Στην κλασική του μορφή το παίγνιο παρουσιάζεται ως εξής:

> *Δύο ύποπτοι συλλαμβάνονται από την αστυνομία. Η αστυνομία δεν έχει επαρκεί στοιχεία για καταδίκη και έχοντας διαχωρίσει τους δύο κρατουμένους επισκέπτεται τον κάθε ένα από αυτούς και τους προσφέρει την ίδια συμφωνία. Εάν ο ένας καταθέσει, αποστατήσει (defects) από τον άλλο, για την δίωξη του άλλου και ό δεύτερος κρατήσει τη σιωπή του, συνεργαστεί με τον άλλο, ο προδότης αφήνεται ελεύθερος και αυτός που δεν κάρφωσε δέχεται ολόκληρη την ποινή των 10 ετών. Εάν και οι δύο δεν μιλήσουν καταδικάζονται σε μόνο 6 μήνες φυλάκιση για μικρό παράπτωμα. Εάν και οι δύο προδώσουν ο ένας τον άλλο τότε καταδικάζονται και οι δύο σε 5 χρόνια κάθειρξη ο καθένας. Κάθε κρατούμενος πρέπει να επιλέξει είτε να καρφώσει τον άλλο είτε να μείνει σιωπηλός. Ο καθένας διαβεβαιώνεται ότι ο άλλος δεν θα γνωρίζει για την προδοσία πριν το τέλος της έρευνας. Το ερώτημα είναι πως πρέπει να δράσουν οι κρατούμενοι.*

Εάν υποτεθεί ότι ο κάθε παίκτης νοιάζεται μόνο να ελαχιστοποιήσει τον χρόνο που θα καθίσει στην φυλακή τότε το δίλλημα του κρατουμένου αποτελεί ένα παίγνιο μη μηδενικού αθροίσματος στο οποίο οι δύο παίκτες μπορούνε είτε να συνεργαστούν, με τον άλλο παίκτη και όχι την αστυνομία, και να μην μιλήσουν, είτε να αποστατήσουν, προδώσουν, τον άλλο παίκτη. Σε αυτό το παίγνιο, όπως και στην περισσότερη θεωρία παιγνίων, η μόνη ανησυχία του κάθε παίκτη, κρατουμένου, είναι να μεγιστοποιήσει την δική του απόδοση χωρίς κανένα ενδιαφέρον για την ωφέλεια του άλλου. Στην μοναδική ισορροπία αυτού του παιγνίου η λογική επιλογή οδηγεί τον κάθε παίκτη να αποστατήσει, προδώσει, τον άλλο παίκτη παρόλο που η επιμέρους αμοιβή θα ήτανε μεγαλύτερη αν συνεργάζονταν μεταξύ τους, για να γινότανε αυτό βέβαια, καθώς δεν μπορούνε να επικοινωνήσουν μεταξύ τους, θα έπρεπε να έχει μεγάλη εμπιστοσύνη ο ένας στον άλλο.

| | **Συνεργασία** | **Προδοσία** |
|---|---|---|
| **Συνεργασία** | 6 μήνες,6μήνες | 10 χρόνια,0 |
| **Προδοσία** | 0 χρόνια,10 | 5 χρόνια,5 χρόνια |

**Σχ. 10 Δίλλημα του κρατουμένου (Οι αποδόσεις είναι ο καιρός κάθειρξης)**

Στην κλασική του μορφή στο παίγνιο αυτό η συνεργασία, με τον άλλο παίκτη πάντα, είναι αυστηρά κυριαρχούμενη από την προδοσία, έτσι ώστε η μόνη δυνατή ισορροπία είναι για όλους τους παίκτες να αποστατήσουν, προδώσουν. Ότι και να κάνει ο άλλος παίκτης ο ένας θα έχει πάντα μεγαλύτερη ωφέλεια παίζοντας προδοσία. Εφόσον σε οποιαδήποτε κατάσταση παίζοντας προδοσία υπάρχει μεγαλύτερη ωφέλεια από την συνεργασία όλοι οι λογικά σκεπτόμενη παίκτες θα προδώσουν [51].

Στο επαναλαμβανόμενο δίλλημα του κρατουμένου το παίγνιο παίζετε πολλές φορές. Έτσι κάθε παίκτης έχει την ευκαιρία να τιμωρήσει τον άλλο παίκτη για προηγούμενη με συνεργάσιμη συμπεριφορά [22]. Αν ο αριθμός των επαναλήψεων είναι γνωστός και στους δύο παίκτες από πριν η θεωρία λέει ότι οι παίκτες θα προδίδουν ξανά και ξανά χωρίς να έχει σημασία πόσες φορές θα παιχτεί το παίγνιο. Όμως αυτή η ανάλυση αποτυγχάνει να προβλέψει την ανθρώπινη συμπεριφορά σε μια πραγματική κατάσταση διλλήματος του κρατουμένου και επίσης αποτυγχάνει να προβλέψει τον βέλτιστο αλγόριθμο όταν προγράμματα υπολογιστών διαγωνίζονται μεταξύ τους. Μόνο όταν οι παίκτες παίξουν αόριστες ή τυχαίου πλήθους επαναλήψεις είναι που η συνεργασία μπορεί να είναι ισορροπία, στην ουσία μια υποπαιγνιακά τέλεια ισορροπία (subgame perfect equilibrium) που σημαίνει ότι και οι δύο παίκτες να προδίδουν πάντα παραμένει ισορροπία και υπάρχουν και πολλά άλλα αποτελέσματα ισορροπίας [22]. Σε αυτήν την περίπτωση το κίνητρο να προδώσει ένας παίκτης μπορεί να ξεπεραστεί με την απειλή της τιμωρίας, αντιποίνων.

Ένα παράδειγμα του επαναλαμβανόμενου διλλήματος του κρατουμένου είναι ο πόλεμος των χαρακωμάτων στον πρώτο παγκόσμιο. Σε αυτήν την περίπτωση παρόλο που αρχικά ήταν καλύτερο να προκαλέσει ο ένας όσο μεγαλύτερη ζημιά γινότανε στον αντίπαλο, καθώς περνούσε ο καιρός και οι δύο αντίπαλοι άρχισαν να «μαθαίνουν» ο ένας τον άλλο, κατάλαβαν ότι το να προκαλεί κάποιος όσο περισσότερη ζημιά γίνεται στον άλλο θα προκαλούσε απλά μια παρόμοια απάντηση, για παράδειγμα το να καταστρέψουν τα αποθέματα τροφής ο ένας του άλλου θα άφηνε και τους δύο να πεινάσουν. Μετά από ένα διάστημα οι αντιμαχόμενοι μάθανε ότι είναι επαρκές να δείχνουν για το τι είναι ικανοί αντί να πραγματοποιούνε την ίδια την απειλή [36].

## *2.5 Συμπεράσματα*

Η θεωρία παιγνίων προσφέρει μια απάντηση σε καταστάσεις στις οποίες ο παίκτης δεν γνωρίζει τι πρέπει να επιλέξει. Ανάλογα με το παίγνιο και τα χαρακτηριστικά του προτείνει μεθοδολογίες επίλυσης του αντίστοιχου προβλήματος. Ο πράκτορας όμως που θα αναπτυχθεί δεν φτάνει να μπορεί να λάβει αποφάσεις για την καλύτερη κίνηση, πρέπει να μπορεί να προσαρμόζεται και να μαθαίνει από τις κινήσεις του αντιπάλου του. Απάντηση σε αυτό δίνει μια άλλη επιστήμη, η μηχανική μάθηση.

# Κεφάλαιο 3

## ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Στο κεφάλαιο παρουσιάζεται το επιστημονικό πεδίο της μηχανικής μάθησης και αναπτύσσονται περιληπτικά ορισμένα θέματα για την περιγραφή της. Περιγράφονται οι κυριότεροι αλγόριθμοι μάθησης που χρησιμοποιούνται από τους πράκτορες για να καθορίσουν την συμπεριφορά τους στις διάφορες καταστάσεις που συναντώνται κατά την διαδικασία μάθησης. Αναπτύσσονται οι τρεις βασικές τεχνικές μάθησης με βάση την ανατροφοδότηση του συστήματος , η εποπτευόμενη, η μη εποπτευόμενη μάθηση και η ενισχυμένη μάθηση, και ορισμένοι συχνά χρησιμοποιούμενοι αλγόριθμοι για αυτές.

# 3. Μηχανική μάθηση

## *3.1 Τι είναι η μηχανική μάθηση*

Ένα επιστημονική πεδίο ορίζεται καλύτερα από την κεντρική ερώτηση την οποία μελετάει. Το πεδίο της μηχανικής μάθησης προσπαθεί να απαντήσει στην εξής ερώτηση:

> «Πως μπορούμε να κατασκευάσουμε συστήματα υπολογιστών τα οποία αυτόματα βελτιώνονται με την εμπειρία, και ποιοι είναι οι θεμελιώδεις νόμοι που διέπουν όλες της διαδικασίες μάθησης;»

Αυτή η ερώτηση καλύπτει ένα μεγάλο εύρος από εργασίες μάθησης. Όπως το πώς να σχεδιαστεί ένα αυτόνομο κινούμενο ρομπότ που μαθαίνει πώς να πλοηγείται σύμφωνα με την δικιά του εμπειρία, πώς να γίνεται εξόρυξη δεδομένων από ιστορικά ασθενών για βρεθεί ποιοι μελλοντικοί ασθενείς θα αποκριθούν καλύτερα και σε ποιες θεραπείες και πώς να φτιαχτούνε μηχανές αναζήτησης που αυτόματα ρυθμίζονται στα ενδιαφέροντα του χρήστη [44].

Η μηχανική μάθηση είναι μια φυσική απόφυση της τομής της επιστήμης των υπολογιστών και της στατιστικής. Θα μπορούσαμε να πούμε ότι η καθοριστική ερώτηση της επιστήμης των υπολογιστών είναι « Πως μπορούμε να κατασκευάσουμε μηχανές που λύνουν προβλήματα και ποια προβλήματα είναι εγγενώς επιλύσιμα/ δυσεπίλυτα; ». Η αντίστοιχη ερώτηση της στατιστικής θα μπορούσε να είναι «Τι μπορεί να συναχθεί από τα δεδομένα μαζί με μια σειρά υποθέσεων μοντελοποίησης και με τι αξιοπιστία;». Η καθοριστική ερώτηση της μηχανικής μάθησης βασίζεται και στα δύο, αλλά είναι μια ξεχωριστή ερώτηση. Η επιστήμη των υπολογιστών έχει εστιάσει κυρίως στο πώς να προγραμματίζει χειροκίνητα υπολογιστές ενώ η μηχανική μάθηση εστιάζει στην ερώτηση του πως μπορούμε να κάνουμε τους υπολογιστές να προγραμματίζουν τον εαυτό τους (από εμπειρία συν κάποια αρχική δομή). Η στατιστική από την άλλη εστιάζει κυρίως στο τι συμπεράσματα μπορούμε να αντλήσουμε από τα δεδομένα ενώ η μηχανική μάθηση ενσωματώνει επιπρόσθετες ερωτήσεις για το τι υπολογιστικές αρχιτεκτονικές και αλγορίθμους μπορούμε να χρησιμοποιήσουμε για να συλλάβουμε, αποθηκεύσουμε, κατατάξουμε, ανακτήσουμε και συγχωνεύσουμε τα δεδομένα αυτά, πως πολλαπλές υποδιεργασίες μάθησης μπορούνε να ενορχηστρωθούνε σε ένα μεγαλύτερο σύστημα και ερωτήσεις για την υπολογιστική επίλυση [44].

Ένα τρίτο πεδίο του οποίου η καθοριστική ερώτηση συνδέεται στενά με την μηχανική μάθηση είναι αυτό της μελέτης της μάθησης των ανθρώπων και των ζώων στην ψυχολογία, την νευροεπιστήμη και τα σχετικά πεδία. Οι ερωτήσεις τους πως οι υπολογιστές μπορούνε να μάθουνε και τους πως τα ζώα μπορούνε να κάνουνε το ίδιο πιθανότατα έχουν αλληλένδετες απαντήσεις. Μέχρι σήμερα ωστόσο η επίγνωση που η μηχανική μάθηση έχει αποκτήσει από μελέτες της ανθρώπινης μάθησης είναι πολύ πιο αδύναμες από εκείνες που έχει αποκτήσει από την στατιστική και την επιστήμη των υπολογιστών, κυρίως λόγου της χαμηλής κατανόησής μας της ανθρώπινης μάθησης. Ωστόσο, η συνέργεια μεταξύ των μελετών της μηχανικής και της ανθρώπινης μάθησης αυξάνεται, με τους αλγόριθμους της μηχανικής μάθησης όπως είναι της μάθησης χρονικής διαφοράς (temporal difference learning) που τώρα προτείνεται σαν εξήγηση για τα νευρωνικά σήματα που παρατηρούνται στα ζώα που μαθαίνουνε [44]. Μέσα στα επόμενα χρόνια είναι λογικό να περιμένετε η συνέργεια μεταξύ της μελέτης της ανθρώπινης και της μηχανικής μάθησης να μεγαλώσει ουσιαστικά, καθώς γειτονεύουν στο τοπίο βασικών επιστημονικών ερωτήσεων.

Άλλα πεδία, από την βιολογία στα οικονομικά και την θεωρία ελέγχου επίσης έχουν βασικό ενδιαφέρον στην ερώτηση του πως τα συστήματα μπορούνε αυτόματα να προσαρμοστούν ή να βελτιστοποιηθούν στο περιβάλλον τους, και η μηχανική μάθηση θα έχει μια αυξανόμενη ανταλλαγή ιδεών με αυτά τα πεδία τα επόμενα χρόνια. Για παράδειγμα , τα οικονομικά ενδιαφέρονται για ερωτήσεις όπως το πώς κατανεμημένες συλλογές από ιδιοτελή άτομα μπορούνε να σχηματίσουν ένα σύστημα, αγορά, που μαθαίνει τιμές που οδηγούν σε βέλτιστη κατά Παρέτο κατανομή για το κοινό καλό. Και η θεωρία ελέγχου, ιδίως η προσαρμοστική θεωρία ελέγχου, ενδιαφέρεται για ερωτήσεις όπως είναι το πώς ένα σερβοελεγχόμενο σύστημα μπορεί να βελτιώσει την στρατηγική ελέγχου του μέσω της εμπειρίας. Είναι ενδιαφέρον το ότι τα μαθηματικά μοντέλα για προσαρμογή σε αυτά τα άλλα πεδία είναι κάπως διαφορετικά από αυτά που χρησιμοποιούνται συνήθως στην μηχανική μάθηση, γεγονός που υποδηλώνει σημαντική δυνατότητα για διασταύρωση των μοντέλων και των θεωριών.

Η μηχανική μάθηση ερευνά το πώς μπορεί να αποκτηθεί γνώση, μάθηση, από παρατηρήσεις ή εμπειρίες ενός πράκτορα, αυτόβουλου υποκειμένου [23]. Είναι ένας επιστημονικός κλάδος που ασχολείται με τον σχεδιασμό και την ανάπτυξη αλγορίθμων που επιτρέπουν στους υπολογιστές να αλλάξουν συμπεριφορά βασισμένοι σε δεδομένα, που προέρχονται για παράδειγμα από αισθητήρες ή βάσης δεδομένων. Επίκεντρο της έρευνας της μηχανικής μάθησης είναι να μαθαίνεται αυτόματα η αναγνώριση περίπλοκων μοτίβων και το να παίρνονται ευφυής αποφάσεις με βάση κάποια δεδομένα. Συνεπώς, η μηχανική μάθηση συνδέεται στενά με τα πεδία της στατιστικής, της θεωρίας πιθανοτήτων, της εξόρυξης δεδομένων, της αναγνώρισης μοτίβων, της τεχνητής νοημοσύνης, του προσαρμοστικού ελέγχου και της θεωρητικής επιστήμης των υπολογιστών [11].

Προγραμματίζοντας έναν πράκτορα, δεν χρειάζεται να του παρέχουμε ολόκληρη την γνώση, που μπορεί να είναι αδύνατο σε περίπτωση που ο προγραμματιστής ή σχεδιαστής δεν έχει πλήρη γνώση έτσι και αλλιώς για το περιβάλλον. Αντί αυτού, με το να μαθαίνει την απαραίτητη γνώση παρέχουμε στον πράκτορα έναν επιπρόσθετο βαθμό αυτονομίας. Η συμπεριφορά ενός πράκτορα προσδιορίζεται εντελώς από τις εμπειρίες του [23].

Μπορούμε να ορίσουμε την μάθηση γενικά ώστε να περιλαμβάνει οποιοδήποτε πρόγραμμα υπολογιστή που βελτιώνει την απόδοσή του σε κάποια διεργασία μέσω της εμπειρίας. Για την ακρίβεια :

> Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία Ε σε σχέση με κάποια κατηγορία διεργασιών Δ και την μέτρηση απόδοσης Α, εάν η απόδοση του στις διεργασίες Δ, όπως μετριέται από το Α, βελτιώνεται με την εμπειρία Ε [43].

Γενικά, για να υπάρχει ένα καλώς καθορισμένο πρόβλημα μάθησης θα πρέπει να υπάρχουν αυτά τα τρία χαρακτηριστικά: η κατηγορία των διεργασιών, το μέτρο απόδοσης προς βελτίωση και η πηγή της εμπειρίας. Ανάλογα με το πώς καθορίζουμε τα Ε,Δ και Α θα μπορούσαμε να αποκαλέσουμε την εργασία για παράδειγμα εξόρυξη δεδομένων, αυτόνομη εύρεση, ενημέρωση βάσης δεδομένων, προγραμματισμό μέσω παραδείγματος κτλ [43].

Ο ορισμός αυτός της μάθησης είναι αρκετά γενικός ώστε να συμπεριλαμβάνει τις περισσότερες διεργασίες που θα αποκαλούσαμε συμβατικά διεργασίες μάθησης, όπως χρησιμοποιούμε τις λέξεις αυτές και στην καθομιλουμένη. Είναι επίσης αρκετά γενικός για να περικλείει προγράμματα υπολογιστών που βελτιώνονται μέσω της εμπειρίας με πολύ απλούς τρόπους [43]. Για παράδειγμα, ένα σύστημα βάσης δεδομένων που επιτρέπει στους χρήστες να ενημερώσουν τις εγγραφές δεδομένων θα ταίριαζε στον ορισμό αυτό του συστήματος μάθησης: βελτιώνει την απόδοσή του στο να απαντάει ερωτήματα βάσεων δεδομένων (queries), βασισμένο στην εμπειρία που απέκτησε από τις ενημερώσεις στη βάση.

## *3.2 Εφαρμογές της μηχανικής μάθησης*

Κατά τα τελευταία χρόνια η μηχανική μάθηση βρήκε τον δρόμο της στους τομείς της διοίκησης, του εμπορίου και της βιομηχανίας με εντυπωσιακό τρόπο. Η εξόρυξη δεδομένων είναι ίσως η ποιο ευρέως γνωστή επίδειξη αυτού του φαινομένου, συμπληρώνεται από λιγότερο δημοσιευμένες εφαρμογές της όπως είναι τα προσαρμοστικά συστήματα σε διάφορα βιομηχανικές εφαρμογές, οικονομικές προβλέψεις, ιατρικές διαγνώσεις και την κατασκευή προφίλ χρηστών για φυλλομετρητές (www-browsers). Αυτή η μεταφορά της μηχανικής μάθησης από τα ερευνητικά εργαστήρια στον «πραγματικό» κόσμο έχει προκαλέσει ένα αυξημένο ενδιαφέρον στις τεχνικές μάθησης, υπαγορεύοντας περαιτέρω προσπάθεια στην ενημέρωση των ανθρώπων άλλων κλάδων για τη εξέλιξη της τεχνολογίας στην μηχανική μάθηση και τις χρήσης της [14].

Εφαρμογές της μηχανικής μάθησης συμπεριλαμβάνουν την μηχανική αντίληψη, την μηχανική όραση, την επεξεργασίας φυσικής γλώσσας, την αναγνώριση συντακτικών μοτίβων, της μηχανές αναζήτησης, την βιοπληροφορική, τις διεπαφές μηχανικών εγκεφάλων και της χημειοπληροφορικής, την ανίχνευση της απάτης με πιστωτικές κάρτες, την χρηματιστηριακή ανάλυση, την ταξινόμηση ακολουθιών DNA, την αναγνώριση φωνής και γραφικού χαρακτήρα, τα παιχνίδια υπολογιστών, την μηχανική λογισμικού, τις προσαρμοστικές ιστοσελίδες και την μετακίνηση των ρομπότ [11].

Ένα μέτρο της προόδου στη μηχανική μάθηση είναι οι σημαντικές εφαρμογές της στον πραγματικό κόσμο, όπως αυτές που ακολουθούν. Παρόλο που τώρα θεωρούμε πολλές από αυτές ως δεδομένες, είναι αξιοσημείωτο να αναφέρουμε ότι μέχρι το 1985 δεν υπήρχαν σχεδόν καθόλου εμπορικές εφαρμογές της μηχανικής μάθησης [43].

*Αναγνώριση φωνής*. Όλα τα παρόν διαθέσιμα εμπορικά συστήματα για αναγνώριση φωνής χρησιμοποιούν μηχανική μάθηση σε κάποια μορφή της για να εκπαιδεύσουν το σύστημα να αναγνωρίζει την φωνή [44]. Ο λόγος είναι απλός: η ακρίβεια της αναγνώρισης φωνής είναι μεγαλύτερη αν το σύστημα εκπαιδευτεί, απ' ότι αν κάποιος αποπειραθεί να το προγραμματίσει με το χέρι. Στην πραγματικότητα, πολλά από τα εμπορικά συστήματα αναγνώρισης φωνής περιλαμβάνουν δύο ξεχωριστές φάσης μάθησης: μία πριν το λογισμικό αποσταλεί, η εκπαίδευση του γενικού συστήματος με τρόπο ανεξάρτητο από τον ομιλητή, και μια δεύτερη αφότου ο χρήστης έχει αγοράσει το λογισμικό, για να επιτευχθεί μεγαλύτερη ακρίβεια εκπαιδεύοντας το σύστημα με τρόπο εξαρτημένο από τον ομιλητή [43].

*Υπολογιστική όραση*. Πολλά από τα τρέχοντα συστήματα όρασης, από τα συστήματα αναγνώρισης προσώπου μέχρι τα συστήματα που αυτόματα κατατάσσουν εικόνες κυττάρων από μικροσκόπιο, αναπτύσσονται χρησιμοποιώντας μηχανική μάθηση, και πάλι επειδή το συστήματα που προκύπτουν είναι πιο ακριβή από αυτά που είναι γραμμένα με το χέρι. Μια μαζικής κλίμακας εφαρμογή της υπολογιστικής όρασης εκπαιδευμένη με την χρήση της μηχανικής μάθησης είναι αυτή που χρησιμοποιείται από το Αμερικάνικο ταχυδρομείο για να ταξινομήσει αυτόματα τα γράμματα που έχουνε χειρόγραφες διευθύνσεις. Πάνω από το 85% τις χειρόγραφης αλληλογραφίας στις Ηνωμένες Πολιτείες ταξινομούνται αυτόματα, χρησιμοποιώντας λογισμικό χειρόγραφης ανάλυσης εκπαιδευμένο με πολύ υψηλή ακρίβεια χρησιμοποιώντας μηχανική μάθηση σε ένα πολύ μεγάλο σύνολο δεδομένων [43].

*Βίο-επιτήρηση*. Μια πληθώρα κυβερνητικών προσπαθειών για τον εντοπισμό και την ανίχνευση ξεσπασμάτων ασθενειών τώρα χρησιμοποιούν μηχανική μάθηση. Για παράδειγμα το σχέδιο RODS περιλαμβάνει συλλογή σε πραγματικό χρόνο εκθέσεων εισαγωγών από τα επείγοντα περιστατικά όλης της δυτικής Πενσυλβάνιας, και την χρήση λογισμικού μηχανικής μάθησης για την μάθηση των προφίλ των τυπικών εισαγωγών έτσι ώστε να μπορούν να ανιχνευτούν ανώμαλα μοτίβα συμπτωμάτων και η γεωγραφική τους κατανομή. Τρέχουσες εργασίες περιλαμβάνουν την προσθήκη σε ένα πλούσιο σύνολο επιπρόσθετων δεδομένων, όπως είναι οι λιανικές αγορά over-the-counter φαρμάκων για την αύξηση της ροής πληροφορίας στο σύστημα, αυξάνοντας περαιτέρω την ανάγκη για αυτόματες μεθόδους μάθησης δεδομένου αυτού του ακόμα πιο περίπλοκου συνόλου δεδομένων [43].

*Έλεγχος ρομπότ*. Μέθοδοι μηχανικής μάθησης έχουν εφαρμοστεί με επιτυχία σε μια σειρά από ρομποτικά συστήματα. Για παράδειγμα, αρκετοί ερευνητές έχουνε επιδείξει την χρήση της μηχανικής μάθησης για την απόκτηση στρατηγικών ελέγχου για την σταθερή πτήση ελικοπτέρων και τους ελιγμούς με αυτά. Ο πρόσφατος χορηγούμενος διαγωνισμός από το DARPA, ένα τμήμα του Υπουργείου αμύνης των ΗΠΑ, σχετικός με ρομπότ που οδηγάνε αυτόνομα για πάνω από 100 μίλια μέσα σε έρημο κερδήθηκε από ένα ρομπότ που χρησιμοποιούσε μηχανική μάθηση για να τελειοποιήσει την δυνατότητά του να ανιχνεύει αντικείμενα που βρίσκονται μακριά (εκπαίδευε τον εαυτό του από ένα σύνολο δεδομένων που είχε συλλέξει μόνο του αποτελούμενο από έδαφος που είχε δει από απόσταση και αργότερα από κοντά) [43].

*Επιταχυνόμενες εμπειρικές επιστήμες*. Πολλές επιστήμες που βασίζονται κατά κόρων σε δεδομένα τώρα χρησιμοποιούν μεθόδους μηχανικής μάθησης για να βοηθήσουν στην διαδικασία της επιστημονικής ανακάλυψης. Η μηχανική μάθηση χρησιμοποιείται για την μάθηση μοντέλων γονιδιακής έκφρασης στα κύτταρα από υψηλής απόδοσης δεδομένα, για την εύρεση ασυνήθιστων αστρονομικών αντικειμένων από μαζικά δεδομένα από την έρευνα Sloan για τον ουρανό, με σκοπό την χαρτογράφηση του, και για τον χαρακτηρισμό περίπλοκων μοτίβων της ενεργοποίησης του εγκεφάλου που υποδεικνύουν διαφορετικές γνωστικές καταστάσεις των ανθρώπων σε fMRI τομογράφους [43]. Οι μέθοδοι της μηχανικής μάθησης αναμορφώνουν την πρακτική εφαρμογή πολλών εμπειρικών επιστημών που βασίζονται στα δεδομένα και πολλές από αυτές τις επιστήμες έχουν πλέον εργαστήρια πάνω στην μηχανική μάθηση στο πλαίσιο των διασκέψεων στον τομέα τους.

## *3.3 Αλγόριθμοι μάθησης*

Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται για διάφορες εφαρμογές. Ο σκοπός τους είναι να χρησιμοποιήσουν παρατηρήσεις ( εμπειρίες, δεδομένα, μοτίβα) για να βελτιώσουν ένα στοιχείο επίδοσης, το οποίο καθορίζει πως ο πράκτορας αντιδράει όταν του δοθούν συγκεκριμένες εισροές (inputs), είσοδοι, πληροφοριών. Το στοιχείο επίδοσης μπορεί να είναι ένας απλός ταξινομητής που προσπαθεί να ταξινομήσει μια περίπτωση εισόδου σε ένα σύνολο κατηγοριών ή μπορεί να είναι ένας πλήρες πράκτορας που δρα σε ένα άγνωστο περιβάλλον [23]. Δεχόμενος ανατροφοδότηση σχετικά με την απόδοση ο αλγόριθμος μάθησης προσαρμόζει το στοιχείο επίδοσης για να ενισχύσει τις δυνατότητες του.

Με βάση την ανατροφοδότηση μπορούμε να ξεχωρίσουμε τρεις βασικές μορφές μάθησης: την εποπτευόμενη μάθηση, την μη εποπτευόμενη μάθηση και την ενισχυμένη μάθηση. Στην εποπτευόμενη μάθηση οι αλγόριθμοι μάθησης δέχονται εισόδους και τις σωστές εξόδους και ψάχνουν για μια συνάρτηση που υπολογίζει την άγνωστη συνάρτηση στόχο. Στην μη εποπτευόμενη μάθηση ο πράκτορας δέχεται μόνο δεδομένα εισόδου και χρησιμοποιεί μια αντικειμενική συνάρτηση, όπως είναι μια συνάρτηση απόστασης, για να εξάγει συσπειρώσεις στα δεδομένα εισόδου ή συγκεκριμένα χαρακτηριστικά τα οποία είναι χρήσιμα για να περιγράψουν τα δεδομένα. Στην ενισχυμένη μάθηση ο πράκτορας δέχεται μια είσοδο και μια αξιολόγηση, ανταμοιβή, της δράσης που επιλέγει ο πράκτορας και ο αλγόριθμος μάθησης πρέπει να μάθει μια πολιτική η οποία χαρτογραφεί τις εισόδους σε δράσεις με αποτέλεσμα την καλύτερη απόδοση.

### 3.3.1 Εποπτευόμενη μάθηση (Supervised learning)

Η εποπτευόμενη μάθηση είναι μια τεχνική μηχανικής μάθησης για την εξαγωγή μιας συνάρτησης από δεδομένα εκπαίδευσης [43], είναι η αναζήτηση για αλγόριθμους που εκλογικεύουν από εξωτερικά παρεχόμενες περιπτώσεις για να παράγουν γενικές υποθέσεις οι οποίες στη συνέχεια κάνουν προβλέψεις γιε μελλοντικές περιπτώσεις [38]. Τα δεδομένα εκπαίδευσης αποτελούνται από ζεύγη αντικειμένων εισόδου, συνήθως διανύσματα, και επιθυμητών εξόδων. Η έξοδος της συνάρτησης μπορεί να είναι μια συνεχόμενη τιμή, αποκαλούμενη παλινδρόμηση, ή να μπορεί να προβλέψει μια ετικέτα κατηγορίας του αντικειμένου εισόδου, αποκαλούμενη κατάταξη [43]. Σκοπός του εποπτευόμενου μαθητή είναι να προβλέψει την τιμή της συνάρτησης για κάθε έγκυρο αντικείμενο εισόδου αφότου έχει δει έναν ορισμένο αριθμό παραδειγμάτων εκπαίδευσης, π.χ. ένα ζεύγος εισόδου και την έξοδο στόχο. Για να το πετύχει αυτό ο μαθητής πρέπει να γενικοποιήσει από τα παρουσιαζόμενα δεδομένα σε αθέατες καταστάσεις με ένα «λογικό» τρόπο (επαγωγική προκατάληψη). Με άλλα λόγια ο σκοπός της εποπτευόμενης μάθησης είναι να οικοδομήσει ένα συνοπτικό μοντέλο της κατανομής των ετικετών κατηγοριών από την άποψη των στοιχείων πρόβλεψης. Ο ταξινομητής αποτέλεσμα χρησιμοποιείται στην συνέχεια για να εκχωρήσει ετικέτες κατηγοριών στις περιπτώσεις δοκιμών όπου οι τιμές των στοιχείων πρόβλεψης είναι γνωστές αλλά η τιμή της ετικέτας κατηγορίας είναι άγνωστη [11].

Μερικοί από τους ποιο συχνά χρησιμοποιούμενους αλγόριθμους είναι:

- Artificial Neural Networks, Backpropagation
- Case-based reasoning
- Decision tree learning
- Instance-based learning, Nearest Neighbor Algorithm
- Naive bayes classifier
- Support vector machines

### 3.3.2 Μη εποπτευόμενη μάθηση (Unsupervised learning)

Στην μηχανική μάθηση η μη εποπτευόμενη μάθηση είναι μια κατηγορία προβλημάτων στην οποία προσπαθείτε να προσδιοριστεί το πώς είναι οργανωμένα τα δεδομένα [13]. Μελετάει το πώς τα συστήματα μπορούνε να μάθουν να αναπαριστούν συγκεκριμένα μοτίβα εισόδου με τρόπο ο οποίος αντανακλά την στατιστική δομή της συνολικής συλλογής των μοτίβων εισόδου. Ξεχωρίζει από την εποπτευόμενη μάθηση, και την ενισχυμένη, σε ότι ο μαθητής δίνεται μόνο μη κατηγοριοποιημένα παραδείγματα, δεν υπάρχουν σαφείς στόχοι έξοδοι ή περιβαλλοντικές αξιολογήσεις με την κάθε είσοδο. Στην μη εποπτευόμενη μάθηση ο μαθητής φέρει προηγούμενες τάσεις για το ποιες πτυχές της δομής της εισόδου θα πρέπει να λαμβάνονται στην έξοδο [32]. Τα μόνα πράγματα με τα οποία η μη εποπτευόμενη μάθηση έχει να δουλέψει είναι τα παρατηρούμενα μοτίβα τα οποία συνήθως υποτίθενται ότι είναι ανεξάρτητα δείγματα μιας άγνωστης κατανομής πιθανοτήτων και κάποιες σαφείς ή υπονοούμενες γνωστές εκ τον προτέρων πληροφορίες για το τι είναι σημαντικό [7]. Μια βασική ιδέα είναι ότι η είσοδος, όπως είναι μια εικόνα στην μηχανική όραση, έχει απομακρυσμένα ανεξάρτητα αίτια, όπως αντικείμενα σε δεδομένους χώρους που φωτίζονται από συγκεκριμένο φωτισμό. Εφόσον είναι σε αυτά τα ανεξάρτητα αίτια στα οποία πρέπει φυσιολογικά να δράσουμε, η καλύτερη αναπαράσταση για μια είσοδο είναι από την άποψη τους. Η μη εποπτευόμενη μάθηση συνδέεται στενά με το πρόβλημα της εκτίμησης πυκνότητας στην στατιστική. Ωστόσο επίσης περιλαμβάνει και πολλές άλλες τεχνικές που επιδιώκουν να συνοψίσουν και να εξηγήσουν βασικά χαρακτηριστικά των δεδομένων. Δύο κατηγορίες μεθόδων έχουνε προταθεί για την μη εποπτευόμενη μάθηση: Οι τεχνικές εκτίμησης πυκνότητας που κατασκευάζουν στατιστικά μοντέλα, όπως είναι τα Bayesian δίκτυα, του πως τα βαθύτερα αίτια θα μπορούσαν να δημιουργήσουν την είσοδο και οι τεχνικές εξόρυξης χαρακτηριστικών που προσπαθούν νε εξάγουν στατιστικές κανονικότητες, ή και παρατυπίες, κατευθείαν από τις εισόδους [32].

Μερικοί από τους ποιο συχνά χρησιμοποιούμενους αλγόριθμους είναι:

- Artificial neural network, Self-organizing map, Adaptive resonance theory
- Data clustering
- Expectation-maximization algorithm

### 3.3.3 Ενισχυμένη μάθηση (Reinforcement learning)

Εμπνευσμένη από την σχετική θεωρία ψυχολογίας στην επιστήμη τον υπολογιστών η ενισχυμένη μάθηση ασχολείται με το πώς ένας πράκτορας θα πρέπει να δράσει σε ένα περιβάλλον ώστε να μεγιστοποιήσει κάποια έννοια μιας μακροχρόνιας αμοιβής [40]. Η ενισχυμένη μάθηση ορίζεται όχι χαρακτηρίζοντας τις μεθόδους μάθησης αλλά χαρακτηρίζοντας ένα πρόβλημα μάθησης. Κάθε μέθοδος που είναι κατάλληλη για την επίλυση αυτού του προβλήματος θεωρείται ως μέθοδος ενισχυμένης μάθησης. Η βασική ιδέα είναι απλά να συλλάβει τις πιο σημαντικές πτυχές του αληθινού προβλήματος που αντιμετωπίζει ο πράκτορας μάθησης που αλληλεπιδρά με το περιβάλλον του για να πετύχει κάποιο στόχο. Κάθε φορά που ο πράκτορας δρα στο περιβάλλον του, ο εκπαιδευτής μπορεί να παρέχει μια ανταμοιβή ή ποινή για να δείξει το πόσο επιθυμητή είναι η κατάσταση που προέκυψε. Για παράδειγμα όταν εκπαιδεύεται ένας πράκτορας για να παίζει ένα παιχνίδι ο εκπαιδευτής μπορεί να παρέχει μια θετική ανταμοιβή όταν κερδίζεται το παιχνίδι, μια αρνητική όταν χάνεται και μηδενική σε όλες τις άλλες καταστάσεις. Το έργο του πράκτορα είναι να μάθει από αυτήν την έμμεση, καθυστερημένη ανταμοιβή, να διαλέγει ακολουθίες ενεργειών που παράγουν την μεγαλύτερη αθροιστικά ανταμοιβή.

Οι αλγόριθμοι της ενισχυμένης μάθησης προσπαθούν να βρούνε μια «πολιτική» που χαρτογραφεί τις καταστάσεις του κόσμου στις δράσεις που ο πράκτορας πρέπει να πάρει σε αυτές τις καταστάσεις [43]. Στα οικονομικά και στην θεωρία παιγνίων η ενισχυμένη μάθηση θεωρείται ως οριοθετημένα ορθολογική ερμηνεία για το πώς μπορούν να προκύψουν ισορροπίες. Το περιβάλλον εκφράζεται συνήθως ως μια πεπερασμένης κατάστασης διαδικασία αποφάσεων Markov (Markov decision process, MDP), και οι αλγόριθμοι ενισχυμένης μάθησης για αυτό το πλαίσιο είναι στενά συσχετιζόμενες με τεχνικές δυναμικού προγραμματισμού. Οι πιθανότητες κατάστασης μετάβασης και οι πιθανότητες ανταμοιβών στις MDPs είναι συνήθως στοχαστικές αλλά σταθερές κατά τη διάρκεια του προβλήματος [18]. Η ενισχυμένη μάθηση διαφέρει από το πρόβλημα της εποπτευόμενης μάθησης στο ότι οι σωστές είσοδοι/έξοδοι δεν δίνονται ποτέ και ούτε οι υπό-βέλτιστες δράσεις διορθώνονται ρητά. Ακόμη, επικεντρώνεται στην συνεχόμενη επίδοση, η οποία περιλαμβάνει την εξεύρεση ισορροπίας μεταξύ εξερεύνησης (exploration), ανεξερεύνητου εδάφους, και εκμετάλλευσης (exploitation), της τρέχουσας γνώσης. Ο συμβιβασμός μεταξύ εξερεύνησης και εκμετάλλευσης έχει μελετηθεί κυρίως μέσω του προβλήματος του πολύ-χερου ληστή (multi-armed bandit) [43].

Επισήμως, το βασικό μοντέλο ενισχυμένης μάθησης, όπως αυτό εφαρμόζεται στις MDPs, αποτελείται από:

1. Ένα σύνολο από περιβαλλοντικές καταστάσεις (states) S
2. Ένα σύνολο από δράσεις/ενέργειες (actions) A
3. Ένα σύνολο από κλιμακούμενες ανταμοιβές R

Σε κάθε στιγμή t, ο πράκτορας αντιλαμβάνεται την κατάστασή του $s_t \in S$ και το σύνολο των πιθανών δράσεων $A(s_t)$. Διαλέγει μια δράση $\alpha \in A(s_t)$ και δέχεται από το περιβάλλον την καινούργια κατάσταση st+1 και μια ανταμοιβή rt. Βασισμένος σε αυτές τις αλληλεπιδράσεις ο πράκτορας της μηχανικής μάθησης πρέπει να αναπτύξει μια πολιτική $\Pi : S \rightarrow A$ η οποία μεγιστοποιεί την ποσότητα $R= r_0 + r_1 + \ldots + r_n$ για τις MDPs που έχουν τελική κατάσταση ή την ποσότητα

$$R= \sum_t \gamma^t r_t$$

για MDPs χωρίς τελική κατάσταση, όπου $0 \leq \gamma \leq 1$ είναι κάποιος συντελεστής προεξόφλησης μελλοντικής ανταμοιβής [43]. Γι' αυτό η ενισχυμένη μάθηση είναι καλύτερη για προβλήματα που περιλαμβάνουν μακροχρόνιες έναντι βραχυχρόνιων ανταμοιβών. Έχει εφαρμοστεί επιτυχημένα σε διάφορα προβλήματα όπως είναι ο ρομποτικός έλεγχος, προγραμματισμός ανελκυστήρων, οι τηλεπικοινωνίες, το τάβλι και το σκάκι [40].

Μερικοί από τους ποιο συχνά χρησιμοποιούμενους αλγόριθμους στην ενισχυμένη μάθηση είναι:

- Temporal difference learning
- Q learning
- SARSA
- Fictitious play

## *3.4 Συμπεράσματα*

Με την χρήση των αλγορίθμων της μηχανικής μάθησης και των παραλλαγών τους ένας πράκτορας μπορεί να εκπαιδευτεί από την εμπειρία που αποκτά κατά την διάρκεια της διεργασίας που μελετάται και να βελτιώνεται κατά την διάρκεια αυτής. Μαζί με την θεωρεία παιγνίων θέτεται ένα πλαίσιο μέσα στο οποίο θα μπορούσε να κινείτε το σκεπτικό του πράκτορα. Το ποια εργαλεία από αυτές τις επιστήμες θα χρησιμοποιηθούν όμως εξαρτάται από την διεργασία, παίγνιο, που μελετάται.

# Κεφάλαιο 4

## ΤΟ ΠΑΙΓΝΙΟ ΠΕΤΡΑ ΨΑΛΙΔΙ ΧΑΡΤΙ

Στο κεφάλαιο αυτό παρουσιάζονται ορισμένα στοιχεία για το παίγνιο πέτρα ψαλίδι χαρτί αυτό καθαυτό και την σχέση του με την θεωρία παιγνίων. Περιγράφεται η δυναμική και η ισορροπία που το χαρακτηρίζουν, που είναι και οι λόγοι για τους οποίους ασχολήθηκαν με αυτό οι θεωρητικοί της επιστήμης αυτής. Στην συνέχεια αναπτύσσεται η λύση του με βάση την θεωρία για την εκδοχή του παιγνίου που εξετάζεται από την εργασία αυτή και πως αυτή η εκδοχή διαφοροποιεί το παίγνιο σε σχέση με την κλασική μορφή του.

# 4. Το παίγνιο πέτρα ψαλίδι χαρτί

## *4.1 Ιστορική αναφορά*

Η ακριβής προέλευση του παιγνίου είναι άγνωστη, παρόμοιες εκδοχές του υπήρχαν στην αρχαία Ιαπωνία και Κίνα απ' όπου και πιθανότατα προήλθε. Το παιχνίδι ήτανε δημοφιλές στην Ιαπωνία του 19ου αιώνα και απέκτησε παγκόσμια δημοτικότητα μέσα στον 20$^{ο}$ αιώνα. Είναι γνωστό με διάφορα ονόματα όπως Roshambeau, jan-ken-pon κτλ.

Παρόλο που το παίγνιο γενικά χρησιμοποιείται για την επίλυση μικροδιαφορών στην καθημερινή ζωή έχουν καταγραφεί περιπτώσεις που χρησιμοποιήθηκε για κάτι παραπάνω από αυτό. Το 2006 ένας ομοσπονδιακός δικαστής στην Φλόριντα των Ηνωμένων Πολιτειών διέταξε τις δύο αντιμαχόμενες πλευρές σε μια μακροσκελή δικαστική υπόθεση να λύσουν μια ασήμαντη διαφορά τους παίζοντας έναν γύρο του παιγνίου πέτρα ψαλίδι χαρτί, με προφανή σκοπό να τονίσει την έλλειψη σοβαρότητας της κατάστασης [4]. Στο πιο ακριβό ίσως παιχνίδι πέτρα ψαλίδι χαρτί που παίχτηκε ποτέ το 2005 ένας διευθύνων σύμβουλος μια ιαπωνικής κατασκευαστικής εταιρείας ηλεκτρονικών εξαρτημάτων αποφάσισε να πουλήσει την συλλογή πινάκων ιμπρεσιονιστών που κατείχε η εταιρεία του σε πλειστηριασμό, μεταξύ των οποίων έργα των Cézanne, Πικάσο και Βαν Γκονγκ. Ο διευθύνων σύμβουλος επικοινώνησε με δύο οίκους πλειστηριασμών στην Αμερική τους Christie's και Sotheby's. Και οι δύο εταιρείες έκαναν της προφορές τους αλλά καμία από τις δύο δεν κατάφερε να τον πείσει να τους αναθέσει την πώληση και καθώς ο Ιάπωνας δεν ήθελε να χωρίσει τη συλλογή τους ανέθεσε να αποφασίσουν αυτές ποια από τις δύο εταιρείες θα αναλάβει τον πλειστηριασμό. Οι δύο εταιρείες, προβλέψιμα ίσως, δεν κατάφεραν να αποφασίσουν ποια από τις δύο θα αναλάβει την δουλεία και ο διευθύνων σύμβουλος τους είπε να παίξουν έναν γύρο του παιγνίου πέτρα ψαλίδι χαρτί για να αποφασιστεί ποιος από τους δύο οίκους θα κερδίσει δηλώνοντας «Μπορεί να φαίνεται περίεργο στους τρίτους αλλά πιστεύω ότι αυτός είναι ο καλύτερος τρόπος για να αποφασίσει κανείς μεταξύ δύο πραγμάτων τα οποία είναι εξίσου καλά» [45]. Για την ιστορία το παιχνίδι κέρδισε ο οίκος Christie's κερδίζοντας εκατομμύρια δολάρια σε προμήθεια, με ποιο ακριβή πώληση τον πίνακα "Large Trees Under the Jas de Bouffan" του Cézanne που πουλήθηκε για 11.776.000 δολάρια.

Το παίγνιο βρήκε τον δρόμο του προς την θεωρία παιγνίων λόγο της δυναμικής και τις ισορροπίας που το διακατέχει όπως αναφέρεται στην συνέχεια.

## *4.2 Η θεωρία παιγνίων και το πέτρα ψαλίδι χαρτί*

Το παίγνιο πέτρα ψαλίδι χαρτί είναι ένα δημοφιλές παράδειγμα των εγχειριδίων της θεωρίας παιγνίων καθώς γίνεται αναφορά σε αυτό σχεδόν σε κάθε ένα από αυτά. Στην κλασική του μορφή είναι ένα παίγνιο μηδενικού αθροίσματος στο οποίο οι παίκτες διαλέγουν τις κινήσεις τους ταυτόχρονα και αναπαριστάται με την κανονική μορφή. Όπως είναι γνωστό οι παίκτες διαλέγουν ανάμεσα από τρεις στρατηγικές την πέτρα, Rock, το χαρτί, Paper, και το ψαλίδι, Scissors. Η πέτρα νικάει το ψαλίδι, το ψαλίδι νικάει το χαρτί και το χαρτί νικάει την πέτρα δημιουργώντας έναν κύκλο έτσι ώστε να υπάρχει μια απάντηση σε κάθε στρατηγική δημιουργώντας μια ισορροπία στο παίγνιο. Κάθε κίνηση μπορεί να κερδίσει μια άλλη αλλά μπορεί να ηττηθεί και από κάποια άλλη, αφήνοντας τους υπόλοιπους συνδυασμούς κινήσεων να επιφέρουν ουδέτερο αποτέλεσμα. Στην κλασική του μορφή η νίκη αποδίδει ένα βαθμό κέρδος και η ήττα επιφέρει ένα βαθμό απώλειας με τα ουδέτερα αποτελέσματα να μην αλλάζουν την βαθμολογία, όπως αναπαριστάται στον πίνακα αποδόσεων που ακολουθεί.

| | R | P | S |
|---|---|---|---|
| R | 0, 0 | -1, 1 | 1, -1 |
| P | 1, -1 | 0, 0 | -1 , 1 |
| S | -1, 1 | 1, -1 | 0, 0 |

**Σχ. 11 Κλασικό πέτρα ψαλίδι χαρτί.**

Το κύριο χαρακτηριστικό του παιγνίου αυτού είναι η δυναμική αυτή του κύκλου που το χαρακτηρίζει, που είναι ο κύριος λόγος που χρησιμοποιείται το παίγνιο αυτό. Υπάρχει μια μοναδική μικτή ισορροπία Nash στην κλασική μορφή του σύμφωνα με την οποία οι παίκτες διαλέγουν από τις διαθέσιμες στρατηγικές με ίσες πιθανότητες, ίσες με 1/3. Δεν παρατίθεται η λύση της κλασικής μορφής του παιγνίου αυτού καθώς θεωρείται κάτι δεδομένο και η διαδικασία είναι η ίδια με αυτήν που δίνεται παρακάτω για την λύση της εκδοχής του παιγνίου αυτού με την οποία ασχολείται η εργασία αυτή.

## *4.3 Εφαρμογές της δυναμικής του πέτρα ψαλίδι χαρτί*

Όπως έχει ήδη αναφερθεί η θεωρία παιγνίων έχει χρησιμοποιηθεί για να βοηθήσει στην ανάλυση και εξαγωγή συμπερασμάτων σε διάφορες επιστήμες. Το μοντέλο το οποίο χρησιμοποιείται στην κάθε περίπτωση εξαρτάται από το είδος του προβλήματος το οποίο εξετάζεται, προσαρμοζόμενο κάθε φορά στις εκάστοτε συνθήκες. Η θεωρία παιγνίων μπορεί να χρησιμοποιηθεί και εφαρμόζοντας αυτά που ήδη ξέρουμε από αυτήν ώστε να δημιουργηθεί κάποιο επιθυμητό μοντέλο αλλά και για να εξηγηθεί και να εξαχθούν συμπεράσματα από κάποιο ήδη υπάρχον μοντέλο. Η δυναμική του παιγνίου πέτρα ψαλίδι χαρτί που προαναφέρθηκε έχει χρησιμοποιηθεί σε διάφορες περιπτώσεις, από τα παιχνίδια στρατηγικής στους υπολογιστές μέχρι την βιολογία. Στα παιχνίδια στρατηγικής έχει εφαρμοστεί για να δημιουργήσει ισορροπία ανάμεσα στους αντιπάλους, χρησιμοποιώντας τον κύκλο που δημιουργεί το παίγνιο αυτό με την μία μονάδα να υπερτερεί κάποιας άλλης αλλά να υστερεί κάποιας τρίτης ώστε να υπάρχει μια αίσθηση ισορροπίας, που είναι και το κύριο χαρακτηριστικό της κλασικής μορφής του παιγνίου αυτού όπως αναφέρθηκε νωρίτερα, αφήνοντας το πλεονέκτημα στον παίκτη που κατανοεί και μπορεί να εκμεταλλευτεί την ισορροπία αυτή και παράλληλα ενθαρρύνοντας την ποικιλότητα στο παιχνίδι. Στην βιολογία η δυναμική του παιγνίου πέτρα ψαλίδι χαρτί έχει χρησιμοποιηθεί για να περιγράψει την ισορροπία σε κάποιο οικοσύστημα. Σε ένα ενδιαφέρον παράδειγμα της δυναμικής αυτής οι M. Frean και E. R. Abraham αναλύουν πως σε ένα οικοσύστημα η εξελικτική σταθερότητα δημιουργείται από το παράδοξο ότι η κυριαρχία του συστήματος που εξετάζουν επιτυγχάνεται με αποδυνάμωση του ρυθμού εισβολής του ενός είδους, μια επιλογή που δεν εξετάζεται από την εξέλιξη [21].

## *4.4 Ανάλυση παιγνίου*

Στην εκδοχή του παιγνίου που αναπτύσσεται στην εργασία αυτή γίνεται διαφοροποίηση των αποδόσεων από την κλασική μορφή του με αποτέλεσμα να επηρεάζεται η ισορροπία του, διαφοροποιώντας το από την κλασική μορφή. Παρόλο που η δυναμική του παραμένει η ίδια, μια στρατηγική υπερτερεί μιας άλλης και υστερεί μιας τρίτης, η απόδοση σε περίπτωση νίκης δεν είναι η ίδια για κάθε κίνηση. Το παίγνιο επίσης δεν αντιμετωπίζεται σαν μηδενικού αθροίσματος, βοηθώντας έτσι να είναι πιο ξεκάθαρο το αποτέλεσμα μετά από πολλούς γύρους καθώς το παίγνιο θα χρησιμοποιηθεί σαν παίγνιο σκηνή για πολλές επαναλήψεις. Η διαφοροποίηση στην απόδοση μιας κίνησης αυξάνοντας το κέρδος που μπορεί να προσφέρει την κάνει πιο «επιθυμητή» και η μείωση φέρει το ανάλογο αποτέλεσμα κάνοντας τη λιγότερο «επιθυμητή». Στην συγκεκριμένη εκδοχή του παιγνίου η πέτρα αποκομίζει 200 βαθμούς, το ψαλίδι 150 και το χαρτί 100, αλλάζοντας ανάλογα την ισορροπία Nash μικτής στρατηγικής όπως δείχνεται στην συνέχεια. Σε μια αναλογία με μια ακραία εκδοχή του παιγνίου πέτρα ψαλίδι χαρτί, η οποία χρησιμοποιείται σαν «παιχνίδι μεθυσμένων», στην οποία κάθε κίνηση με την οποία νικάει κάθε παίκτης γίνεται πραγματικότητα γίνεται άμεσα κατανοητό το κίνητρο με το οποίο επιλέγονται οι στρατηγικές σε κάθε γύρο. Στην ακραία αυτή εκδοχή όταν κάποιος νικάει παίζοντας πέτρα δίνει γροθιά στον αντίπαλο του, όταν νικάει με χαρτί δίνει χαστούκι και όταν νικάει με ψαλίδι «τσιμπάει» τον αντίπαλο. Η διαδικασία για οποιεσδήποτε αποδόσεις σε κάθε στρατηγική είναι η ίδια.

Ακολουθεί ο πίνακας αποδόσεων, ο παίκτης έχει το κάθετο σύνολο στρατηγικών και η μηχανή το οριζόντιο.

| | R | P | S |
|---|---|---|---|
| R | 0, 0 | 0, 100$^{-}$ | $^{+}$200, 0 |
| P | $^{+}$100, 0 | 0, 0 | 0, 150$^{-}$ |
| S | 0, 200$^{-}$ | $^{+}$150, 0 | 0, 0 |

**Σχ. 12 Διαφοροποιημένη εκδοχή πέτρα ψαλίδι χαρτί**

Στην συνέχεια επιλύεται η διαφοροποιημένη εκδοχή του παιγνίου βρίσκοντας την ισορροπία Nash μικτής στρατηγικής και για τις τρεις κινήσεις.

Μικτές στρατηγικές ΜΣ 123

Οι πιθανότητες του παίκτη P και της μηχανής M να διαλέξουν την κάθε στρατηγική:

PrP(R)=p, PrP(P)= r, PrP(S)=1-p-r και

PrM(R)=q, PrM(P)= s, PrM(S)=1-q-s

Εφόσον δεχόμαστε ότι οι παίκτες είναι αδιάφοροι για το ποια κίνηση θα επιλέξουνε, καθώς ξεκινάμε με την υπόθεση ότι δεν ξέρουν ποια κίνηση να διαλέξουν, η προσδοκώμενη ωφέλεια EU για κάθε κίνηση είναι η ίδια.

Εξίσωση αδιαφορίας για τον παίχτη:

EUP(R) [=0q+0s+200(1-q-s)] =

EUP(P) [=100q+0s+0(1-q-s)] =

EUP(S) [=0q+150s+0(1-q-s)]

EUP(R) [=200(1-q-s)] = EUP(P) [=100q] = EUP(S) [=150s]

Λύνουμε την παραπάνω εξίσωση: (απλοποιήσεις δεν γίνονται και τα νούμερα αφήνονται ως έχουν για να είναι ορατό το ότι αναπαριστούν κελιά του πίνακα αποδόσεων και ποιο συγκεκριμένα το πρώτο από τα δύο στοιχεία του κελιού καθώς είναι του παίχτη , αντίστοιχα για το δεύτερο στοιχείο του κελιού θα ήτανε της μηχανής).

200(1-q-s) = 100q → 200-200q-200s = 100q → 200-200s = 100q+200q→

200-200s = q(100+200) → q = (200-200s) / (100+200)

100q = 150s → 100(200-200s) / 100+200 = 150s →

100(200-200s) = (100+200)150s →

(100*200) – (100*200s) = (100+200)150s →

100*200 = (100+200)150s + (100*200)s →

100*200 = s((100+200)150+(100*200)) →

s = (100*200) / ((100+200)150+(100*200))

Άρα PrM(P) = s = 20000 / (45000 + 20000) = 20/65 = 4 / 13

Και PrM(R) = q = (200-200s) / (100+200) = (200 – 200*4/13) / 300 =

= (200 – 800/13) / 300 = (2600/13 – 800/13) / 300 =

= (1800/13) / 300 = 1800/13*300 = 1800 / 3900 = 18 / 39 = 6 / 13

Οπότε PrM(S) = 1-q-s = 1- (6/13) – (4/13) = 3 / 13

Εξίσωση αδιαφορίας για την μηχανή:

EUM(R) [=0p+0r+200(1-p-r)] =

EUM(P) [=100p+0r+0(1-p-r)] =

EUM(S) [=0p+150r+0(1-p-r)]

Παρατηρούμε ότι είναι ακριβώς οι ίδιες εφόσον οι αποδόσεις για τον παίχτη και την μηχανή είναι οι πανομοιότυπες. Οπότε τα PrP(R)=p, PrP(P)= r, PrP(S)=1-p-r είναι ίσα με τα PrM(R)=q, PrM(P)= s, PrM(S)=1-q-s αντίστοιχα. Για να είναι αδιάφοροι δηλαδή και οι δύο, εφόσον στην συγκεκριμένη περίπτωση οι αποδώσεις και για τους δύο είναι ίδιες, οι πιθανότητες να παίξει κάποιος πέτρα θα πρέπει να είναι Pr(R) = 6/13, χαρτί Pr(P) = 4/13 και ψαλίδι Pr(S) = 3/13.

Οπότε η εξίσωση αδιαφορίας είναι:

EU(R) = 200 * 3/13 = 600 / 13 =

EU(P) = 100 * 6/13 = 600 / 13 =

EU(S) = 150 * 4/13 = 600 / 13

Παρατηρείται ότι ενώ το χαρτί έχει απόδοση μικρότερη από το ψαλίδι σύμφωνα με την θεωρία παιγνίων θα πρέπει να παίζεται πιο συχνά από αυτό. Αυτό γίνεται διότι για να είναι οι παίκτες αδιάφοροι για το ποια κίνηση θα επιλέξουν στην εξίσωση αδιαφορίας η απόδοση της κάθε στρατηγικής πολλαπλασιάζεται με την πιθανότητα της κίνησης την οποία νικάει και πρέπει όλες οι προσδοκώμενες ωφέλειες να είναι ίσες όπως προαναφέρθηκε. Οπότε η πέτρα, με την μεγαλύτερη απόδοση, πρέπει να πολλαπλασιαστεί με την μικρότερη πιθανότητα, το χαρτί, με την μικρότερη απόδοση, με την μεγαλύτερη πιθανότητα και το ψαλίδι, με την μέτρια απόδοση, με την μέτρια πιθανότητα. Έτσι η θεωρία παιγνίων προσφέρει στους παίκτες μια απάντηση για το ποια στρατηγική, μικτή στρατηγική στην συγκεκριμένη περίπτωση, πρέπει να επιλέξουν σκεπτόμενη ορθολογικά.

Όπως φαίνεται αλλάζοντας τις αποδώσεις του παιγνίου αυτού για την εκδοχή που εξετάζουμε αλλάζει η ισορροπία Nash μικτών στρατηγικών που υπήρχε στην κλασική μορφή του, η οποία είναι να παίζονται οι στρατηγικές με την ίδια πιθανότητα, 1/3 δηλαδή. Στην κλασική μορφή του, όπως είναι προφανές, η ιδανική στρατηγική είναι επιλέγουν οι παίκτες τυχαία ανάμεσα στις στρατηγικές τους αλλά διαφοροποιώντας τις αποδώσεις αυτό αλλάζει και προσαρμόζεται σε αυτές, ανάλογα πάντα με τις τιμές που τους έχουμε δώσει.

Όμως οι παίκτες δεν σκέπτονται πάντα ορθολογικά και μπορεί να ακολουθήσουν την δική τους στρατηγική που πιστεύουν ότι μπορεί να τους προσφέρει την νίκη καθώς ο σκοπός του παιγνίου είναι να κερδίσουν. Μπορεί να προσπαθήσουν να ακολουθήσουν κάποιο μοτίβο κινήσεων πιστεύοντας ότι ο αντίπαλος δεν θα το προβλέψει, επηρεαζόμενοι πάντα από τις διαφοροποιημένες αποδώσεις της εκδοχής αυτής του παιγνίου που εξετάζεται.

## *4.5 Συμπεράσματα*

Η δυναμική του παιγνίου πέτρα ψαλίδι χαρτί το καθιστά ενδιαφέρον αντικείμενο για μελέτη, που όμως λόγο των κανόνων που το διέπουν δεν είναι δυνατό να παραληφθεί ο παράγοντας τύχη. Η διαφοροποιημένη εκδοχή του παιγνίου προσφέρει στους παίκτες την δυνατότητα να δοκιμάσουν διαφορετικές στρατηγικές απ’ ότι στην κλασική μορφή του και προσδίδει περαιτέρω ενδιαφέρων για το αποτέλεσμα του παιγνίου. Η προσέγγιση με την οποία θα αναπτυχθεί ο πράκτορας που επιλέγει κινήσεις και θα προσαρμόζεται στις κινήσεις του αντιπάλου, τυχαίες και μη, πρέπει να λαμβάνει υπόψη της όλα τα χαρακτηριστικά και αποτελέσματα που έχουν αναφερθεί ως τώρα.

# Κεφάλαιο 5

## Η ΕΦΑΡΜΟΓΗ

Στο κεφάλαιο αυτό αναλύεται η κεντρική ιδέα της εφαρμογής, οι δυνατότητες της και πως αυτή εξελίχτηκε κατά την πορεία της ανάπτυξής της. Περιγράφονται οι αλγόριθμοι και οι τεχνικές που χρησιμοποιούνται και αναπτύχτηκαν και πως αυτοί δουλεύουν επιγραμματικά. Στο τέλος του κεφαλαίου εξηγείται και το interface της εφαρμογής.

# 5. **Η εφαρμογή**

## *5.1 Η κεντρική ιδέα*

Σκοπός του πρακτικού τμήματος της πτυχιακής αυτής ήταν η ανάπτυξη μιας εφαρμογής στην οποία ο χρήστης, παίκτης, θα παίζει ενάντια στον υπολογιστή ένα αριθμό γύρων του παιγνίου σκηνή με στόχο όχι απαραίτητα την νίκη αλλά την προσαρμογή της μηχανής, στο επίπεδο δυνατό, στις κινήσεις του εκάστοτε παίκτη και με απώτερο σκοπό την εξαγωγή συμπερασμάτων από την αλληλεπίδραση αυτή. Αργότερα καθώς εξελίσσονταν η εφαρμογή και αναπτύχθηκαν παραπάνω από ένας αλγόριθμοι και τεχνικές αντιμετώπισης των κινήσεων του αντιπάλου κρίθηκε θεμιτό να είναι δυνατός ο συναγωνισμός αυτόν των δυνατοτήτων μεταξύ τους χωρίς την παρέμβαση του παίκτη σε ένα είδος εξομοίωσης του παιγνίου. Στην εξομοίωση αυτή ο χρήστης απλά διαλέγει τους αλγόριθμους που θα συναγωνιστούν και τις δυνατότητες του κάθε αντιπάλου/μηχανής.

Υπάρχουν τρεις εκδοχές του παιγνίου πέτρα ψαλίδι χαρτί στις οποίες ο παίκτης έχει την δυνατότητα να συναγωνιστεί τον πράκτορα, μηχανή. Η πρώτη είναι η κλασική μορφή του παιγνίου, όπως αυτή συναντάται συνήθως στην θεωρία παιγνίων, στην οποία κάθε νίκη ανταμείβεται με 1 βαθμό και κάθε ήττα τιμωρείται με απώλεια ενός βαθμού (παίγνιο μηδενικού αθροίσματος). Στην δυναμική του συγκεκριμένου παιγνίου βέβαια όταν ο ένας παίκτης κερδίζει ο άλλος χάνει όποτε οι δύο αυτές καταστάσεις συμβαίνουν ταυτόχρονα. Ο πίνακας αποδόσεων της κλασικής εκδοχής έχει αναφερθεί και νωρίτερα και παρατίθεται στην συνέχεια.

| | R | P | S |
|---|---|---|---|
| R | 0, 0 | -1, 1 | 1, -1 |
| P | 1, -1 | 0, 0 | -1 , 1 |
| S | -1, 1 | 1, -1 | 0, 0 |

**Σχ. 13 Κλασικό πέτρα ψαλίδι χαρτί.**

Η δεύτερη εκδοχή, που είναι και το κυρίως κομμάτι που ενδιαφέρει την πτυχιακή αυτή όσον αφορά το παιχνίδι ανθρώπου εναντίον πράκτορα, είναι η διαφοροποίηση των αποδόσεων που αναφέρθηκε στο προηγούμενο κεφάλαιο, ο πίνακας αποδόσεων της οποίας δίνεται για άλλη μια φορά στην συνέχεια.

| | R | P | S |
|---|---|---|---|
| R | 0, 0 | 0, 100$^-$ | $^+$200, 0 |
| P | $^+$100, 0 | 0, 0 | 0, 150$^-$ |
| S | 0, 200$^-$ | $^+$150, 0 | 0, 0 |

**Σχ. 14 Διαφοροποιημένη εκδοχή πέτρα ψαλίδι χαρτί**

Στην τρίτη και τελευταία εκδοχή δίνεται η δυνατότητα στον παίκτη να επιλέξει αυτός τις αποδώσεις της κάθε στρατηγικής δημιουργώντας έτσι την δική του εκδοχή του παιγνίου. Η εξέταση φυσικά των διαφόρων εκδοχών που μπορούν να προκύψουν δεν αποτελεί αντικείμενο της πτυχιακής αυτής καθώς οι δυνατότητες είναι απεριόριστες. Σαν ενδεικτική και προεπιλεγμένη μορφή της εκδοχής αυτής δίνεται η ακόλουθη, στην οποία κάθε νίκη επιβραβεύεται με έναν βαθμό και δεν υπάρχει απώλεια σε περίπτωση ήττας.

| | R | P | S |
|---|---|---|---|
| R | 0, 0 | 0, 1 | 1, 0 |
| P | 1, 0 | 0, 0 | 0, 1 |
| S | 0, 1 | 1, 0 | 0, 0 |

**Σχ. 15 Ενδεικτική εκδοχή πέτρα ψαλίδι χαρτί**

Οι συγκεκριμένες αποδόσεις κρίνονται ίσως πιο κατάλληλες και για την εξαγωγή συμπερασμάτων και αξιολόγησης των αλγορίθμων, καθώς είναι πιο εύκολο να έχει κανείς μια εικόνα για το αποτέλεσμα όταν απλά επιβραβεύεται με έναν βαθμό αυτός που νίκησε σε έναν γύρο. Αυτός είναι και ο λόγος που ο συγκεκριμένος πίνακας αποδόσεων χρησιμοποιείται και από το παιχνίδι πράκτορα εναντίον πράκτορα ώστε να είναι πιο ξεκάθαρη η εξέλιξη και το αποτέλεσμα του παιγνίου καθώς ιδίως όταν οι επαναλήψεις του παιγνίου σκηνή είναι πάρα πολλές δεν δίνεται καθαρή εικόνα του αποτελέσματος με διαφοροποιημένες από αυτήν αποδόσεις. Σκοπός του παιξίματος πράκτορα εναντίον πράκτορα είναι η αξιολόγηση και παρατήρηση των αλγορίθμων και δυνατοτήτων της εφαρμογής όποτε δεν δίνεται η δυνατότητα επιλογής διαφορετικών αποδόσεων, καθώς όπως αναφέρθηκε δεν βοηθάνε στην εξαγωγή συμπερασμάτων.

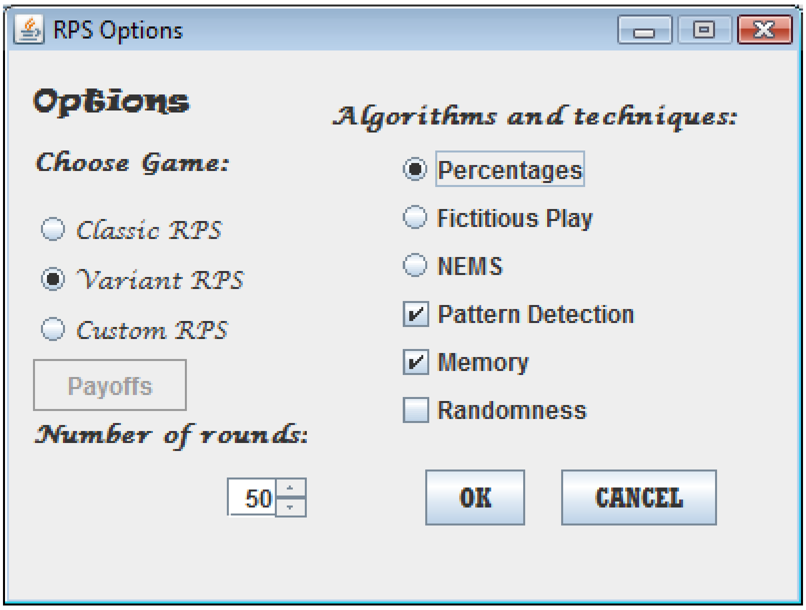
Οι κινήσεις της μηχανής γίνονται με βάση κάποιον αλγόριθμο ο οποίος ανάλογα με τις προηγούμενες κινήσεις του παίκτη, εμπειρία, αποφασίζει το πώς θα αντιδράσει. Τα πλαίσια στα οποία κινείται η αντίδραση της μηχανής εξαρτώνται από τον εκάστοτε αλγόριθμο και από το ποιες δυνατότητες παιξίματος είναι ενεργοποιημένες. Γενικά συναντώνται δύο καταστάσεις κατά την διάρκεια του παιχνιδιού, σαν παιχνίδι θα αναφέρεται η ολοκλήρωση του επιλεγμένου αριθμού επαναλήψεων του παιγνίου σκηνή. Η πρώτη είναι η κατάσταση στην οποία ο εκάστοτε αλγόριθμος δεν έχει επαρκή δεδομένα, είτε απλά καταλαβαίνει ότι έχει κάνει λάθος οπότε διαλέγει να μην χρησιμοποιήσει τα δεδομένα που έχει, ώστε να παρέχει κάποια συγκεκριμένη αντίδραση οπότε απαντάει με την χρήση της θεωρίας παιγνίων επιλέγοντας σύμφωνα με την ισορροπία της ανάλογης εκδοχής του παιγνίου είτε με βάση τον τρόπο λειτουργίας του, εξαρτάται από το ποιος είναι ο επιλεγμένος αλγόριθμος. Αρχικά η εφαρμογή ήταν σχεδιασμένη να χρησιμοποιεί αποκλειστικά και για όλη την διάρκεια του παιχνιδιού τον εκάστοτε επιλεγμένο αλγόριθμο και να υπάρχει ένας ξεχωριστός αλγόριθμος που να χρησιμοποιεί όλες τις τεχνικές αλλά στην πορεία έγινε δυνατόν να μπορούν να συνδυαστούν οι τεχνικές με οποιονδήποτε αλγόριθμο δημιουργώντας έτσι διάφορους πιθανούς «αντιπάλους» για τον παίκτη.

Η δεύτερη κατάσταση είναι ουσιαστικά η χρήση του αλγορίθμου, τεχνικής, εύρεσης μοτίβων στις κινήσεις του παίκτη, ο οποίος δεν υπάρχει σαν ξεχωριστός αλγόριθμος παρά μόνο σε συνδυασμό με κάποιον από τις βασικούς αλγορίθμους. Αυτός ο αλγόριθμος ήταν και η κυρίως ενασχόληση της πτυχιακής αυτής, μαζί με την επέκταση του, την χρήση μνήμης για την εύρεση μοτίβων που τυχόν συναντήθηκαν νωρίτερα κατά την διάρκεια του παιχνιδιού. Όταν ο πράκτορας δηλαδή συναντάει μοτίβα σύμφωνα με την εμπειρία του αντιδράει δίνοντας την ανάλογη απάντηση στο εκάστοτε μοτίβο. Αν αντιληφθεί ότι έχει κάνει λάθος επιστρέφει στην προηγούμενη, πρώτη, κατάσταση και είτε παίζει την αντίστοιχη με την εκδοχή του παιγνίου ισορροπία Nash είτε χρησιμοποιεί τον βασικό του αλγόριθμο για να επιλέξει την επόμενή του κίνηση.

## 5.2 Οι αλγόριθμοι

Η αρχική ιδέα ήταν να αναπτυχθεί σαν κυρίως αλγόριθμος κάποια από τις έτοιμες τεχνικές που υπάρχουν και είχε επιλεχθεί η ενισχυμένη μάθηση σαν κατηγορία αλγορίθμου. Αναπτύχθηκε ο αλγόριθμος fictitious play αλλά τελικά αποφασίστηκε να γραφεί ένας αλγόριθμος από την αρχή ο οποίος θα χρησίμευε για την εύρεση μοτίβων στις κινήσεις του παίκτη προσαρμοσμένος στο συγκεκριμένο παίγνιο.

Παρατίθεται το μενού επιλογών για το παιχνίδι χρήστη εναντίον μηχανής με τις δυνατές επιλογές αλγορίθμων και τεχνικών:




**Εικόνα 5.1 Μενού επιλογών χρήστη εναντίον μηχανής**

Στο μενού αυτό ο χρήστης μπορεί να διαλέξει τις επαναλήψεις του παιγνίου σκηνή και την εκδοχή του παιγνίου. Στο δεξί κομμάτι του μενού υπάρχουν οι διάφοροι αλγόριθμοι και τεχνικές από τις οποίες μπορεί να επιλέξει ο χρήστης. Στην τελική μορφή της εφαρμογής όπως αναφέρθηκε επιλέγεται ένας από τους τρεις βασικούς αλγόριθμους, percentages, fictitious play, NEMS και κάποιος συνδυασμός αλγορίθμων/τεχνικών τρόπου παιχνιδιού από τους υπόλοιπους, pattern detection, memory που ενεργοποιείται μόνο μαζί με το pattern detection, και randomness, όπως εξηγείται στην συνέχεια.

### 5.2.1 NEMS

Ο αλγόριθμος αυτός αναφέρεται πρώτος διότι χρησιμοποιείται και από τους άλλους δύο βασικούς αλγορίθμους στην πρώτη τους κίνηση. Στην πρώτη κίνηση κανένας αλγόριθμος δεν έχει κάποιο στοιχείο για να διαλέξει τι θα παίξει οπότε επιλέγει σύμφωνα με την ισορροπία Nash μικτής στρατηγικής ( Nash Equilibrium for Mixed Strategies, NEMS). Αυτό σημαίνει ότι διαλέγει τυχαία ανάμεσα στην κατανομή πιθανοτήτων που έχει δώσει σε κάθε κίνηση σύμφωνα με τον πίνακα αποδόσεων της ανάλογης εκδοχής. Ο NEMS υπολογίζει την ισορροπία σύμφωνα με τον πίνακα αποδόσεων χρησιμοποιώντας κελιά του πίνακα αντί για τιμές ώστε να δουλεύει για οποιεσδήποτε αποδόσεις και να δοθούν σε κάθε στρατηγική σύμφωνα με τους ακόλουθους τύπους που αναπτύχθηκαν για την διαδικασία αυτή. Τα νούμερα στις αγκύλες των εξισώσεων αντιπροσωπεύουν τα κελιά του πίνακα αποδόσεων, χωρισμένα στα δύο ώστε να μπορεί να αντιστοιχηθεί κάθε απόδοση σε κάθε παίχτη ξεχωριστά. Στον πίνακα τα νούμερα που φαίνονται δηλαδή δεν αντιστοιχούν σε αποδόσεις αλλά στο ποιο κελί είναι το καθένα, με το πρώτο νούμερο να είναι η γραμμή και το δεύτερο η στήλη.

| | R | | P | | S | |
|---|---|---|---|---|---|---|
| R | 0,0 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 |
| P | 1,0 | 1,1 | 1,2 | 1,3 | 1,4 | 1,5 |
| S | 2,0 | 2,1 | 2,2 | 2,3 | 2,4 | 2,5 |

**Σχ. 16 Πίνακας αντιστοίχισης κελιών στις εξισώσεις**

x=Payoffs[1][2]-Payoffs[1][4]-Payoffs[2][2]+Payoffs[2][4]

y=Payoffs[2][4]-Payoffs[1][4]

z=Payoffs[1][4]-Payoffs[0][4]

w=Payoffs[0][4]-Payoffs[0][2]+Payoffs[1][2]-Payoffs[1][4]

k=Payoffs[0][0]-Payoffs[0][4]-Payoffs[1][0]+Payoffs[1][4]

m=Payoffs[1][4]-Payoffs[1][0]+Payoffs[2][0]-Payoffs[2][4]

S: Πιθανότητα να παιχτεί χαρτί

PrP = (k*y+z*m)/(x*k-w*m)

q: Πιθανότητα να παιχτεί πέτρα

PrR = (z+PrP*w)/k

1-q-s: Πιθανότητα να παιχτεί ψαλίδι

PrS = 1-PrR-PrP

Ανάλογα με τα ποσοστά που βγάζουν οι εξισώσεις αυτές είναι και οι κινήσεις του πράκτορα. Ο αλγόριθμος αυτός χρησιμοποιεί τα ποσοστά αυτά για τις κινήσεις του σε οποιεσδήποτε κατάσταση και αν βρίσκεται. Καταρχήν στην αρχική κατάσταση, όπως και όλοι οι αλγόριθμοι, στην οποία δεν έχει άλλα κριτήρια για να επιλέξει κίνηση οπότε παίζει την ισορροπία Nash. Κατά την διάρκεια του παιχνιδιού όταν βρεθεί ξανά σε μια κατάσταση στην οποία δεν ξέρει τι να επιλέξει, είτε γιατί δεν έχει χρήσιμα δεδομένα είτε γιατί έχει κάνει λάθος όποτε σταματάει την τακτική που χρησιμοποιούσε, επιστρέφει πάλι στην ισορροπία της θεωρίας παιγνίων. Εάν δεν είναι ενεργοποιημένη καμιά άλλη ρύθμιση, όπως η εύρεση μοτίβων, τότε αυτή είναι και η μόνη συμπεριφορά του πράκτορα δηλαδή παίζει σε κάθε γύρω την ισορροπία, την βέλτιστη κίνηση σύμφωνα με την θεωρία παιγνίων, ανεξάρτητα από τις κινήσεις του αντιπάλου και βασιζόμενος μόνο στην θεωρία παιγνίων.

### 5.2.2 Percentages

Ο αλγόριθμος percentages αποτελεί ένα τμήμα του αλγορίθμου fictitious play που εξηγείται στην συνέχεια. Στον percentages ο πράκτορας καταγράφει τις κινήσεις του χρήστη και χρησιμοποιεί τα ποσοστά κάθε κίνησης για να αποφασίσει ποια θα είναι η δικιά του κίνηση. Στην πρώτη κίνηση, όπως αναφέρθηκε, χρησιμοποιεί την ισορροπία Nash που είναι και η μόνη φορά που θα παίξει την ισορροπία. Από κει πέρα σε όλες τις υπόλοιπες κινήσεις χρησιμοποιεί τα ποσοστά κινήσεων του αντιπάλου και παίζει τις ακριβώς αντίθετες κινήσεις με αυτά, δηλαδή αν ο παίκτης διαλέγει 30% πέτρα θα παίξει 30% χαρτί για παράδειγμα. Ο αλγόριθμος αυτός δεν καταλήγει να παίζει κάποια συγκεκριμένη κίνηση αλλά κατανέμει τις πιθανότητες επιλογής κίνησης του σύμφωνα με τις κινήσεις του αντιπάλου. Αντίθετα με τον NEMS βασίζεται εξολοκλήρου στις κινήσεις που επιλέγει ο παίκτης-αντίπαλος, πέρα από την πρώτη κίνηση φυσικά. Σε οποιαδήποτε κατάσταση και αν βρίσκεται το παιχνίδι έχει δεδομένα με βάση τα οποία μπορεί να επιλέξει καθώς αντλεί τα δεδομένα αυτά από το σύνολο τον κινήσεων του αντιπάλου οπότε δεν βρίσκεται ποτέ σε πλήρη σύγχυση για το τι να επιλεχθεί. Αν είναι επιλεγμένη η εύρεση μοτίβων, που εξηγείται παρακάτω, όπως και όλοι οι αλγόριθμοι βέβαια, και έχει βρεθεί κάποιο μοτίβο στις κινήσεις του παίκτη χρησιμοποιείται αυτό αντί του ανάλογου βασικού αλγόριθμου στον οποίο θα επιστρέψει μόνο όταν κάνει λάθος.

### 5.2.3 Fictitious play

Ο fictitious play είναι μια πιο ολοκληρωμένη μορφή του προηγούμενου αλγορίθμου. Όπως και ο percentages στην πρώτη κίνηση χρησιμοποιεί τον NEMS για επιλέξει μια κίνηση από την ισορροπία Nash και στους υπόλοιπους γύρους χρησιμοποιεί τα ποσοστά των κινήσεων του αντιπάλου. Η διαφορά με τον percentages είναι ότι δεν χρησιμοποιεί τα αντίθετα ποσοστά κινήσεων του αντιπάλου για να επιλέξει την κίνηση του. Χρησιμοποιεί τα ποσοστά μεν αλλά αυτό που κάνει είναι σύμφωνα με αυτά να βρίσκει την προσδοκώμενη ωφέλεια, σύμφωνα με τον πίνακα αποδόσεων, κάθε στρατηγικής και να παίζει την στρατηγική με την μέγιστη προσδοκώμενη ωφέλεια. Αν τυχόν οι μεγαλύτερες προσδοκώμενες ωφέλειες των στρατηγικών συμπίπτουν επιλέγει τυχαία μεταξύ αυτών. Το σκεπτικό είναι ότι αντί να διαλέξει από την κατανομή των ποσοστών των κινήσεων βρίσκει ποια από τις κινήσεις αυτές έχει την μεγαλύτερη προσδοκώμενη ωφέλεια, σύμφωνα με τα ποσοστά και τον πίνακα αποδόσεων πάντα, και παίζει μόνο αυτή καθώς θεωρείται καλύτερη από τις υπόλοιπες. Αυτό γίνεται μέχρι να αλλάξουν τα ποσοστά φυσικά και να επιλεγεί κάποια άλλη κίνηση. Όπως και με τον percentages αν δεν είναι ενεργοποιημένη καμία άλλη ρύθμιση τότε αυτή θα είναι η συμπεριφορά του πράκτορα σε όλη την διάρκεια του παιχνιδιού. Αν είναι ενεργοποιημένη η εύρεση μοτίβων και έχει βρεθεί κάποιο μοτίβο τότε, όπως και με τους άλλους βασικούς αλγορίθμους, θα χρησιμοποιηθεί αυτό και θα επιστρέψει στον βασικό αλγόριθμο όταν το μοτίβο αυτό δεν θα ισχύει ποια.

### 5.2.4 Pattern detection

Σκοπός του αλγορίθμου αυτού είναι η εύρεση μοτίβων στις κινήσεις του παίκτη και η απάντηση σε αυτές με τις ανάλογες κινήσεις νίκης απέναντι στο μοτίβο αυτό, πάντα με την προϋπόθεση ότι ο παίκτης συνεχίζει να παίζει το μοτίβο που έχει βρεθεί. Όταν ο αλγόριθμος διαπιστώσει ότι έκανε λάθος, είτε γιατί όντος έκανε λάθος είτε γιατί ο αντίπαλός του άλλαξε τρόπο παιχνιδιού, τότε επιστρέφει στον βασικό αλγόριθμο και συνεχίζει να προσπαθεί παράλληλα να βρει κάποιο μοτίβο στις κινήσεις του παίκτη. Οι καταστάσεις αυτές εναλλάσσονται μέχρι το πέρας των επαναλήψεων.

Για να θεωρείται ένα σύνολο κινήσεων μοτίβο επιλέχθηκε να πρέπει να επαναληφθεί τουλάχιστον δύο φορές. Δεδομένου ότι στο παιχνίδι χρήστη εναντίον μηχανής οι επαναλήψεις δεν θα μπορούσαν να είναι πάρα πολλές κρίθηκε ότι ο ελάχιστος αριθμός επαναλήψεων των κινήσεων για να θεωρηθούν αυτές μοτίβο να είναι αρκετός. Το περιβάλλον στο οποίο σχεδιάστηκε να δουλεύει ο πράκτορας είναι πολλαπλών πρακτόρων (multi agent), δύο, και πάνω σε αυτό βασίζεται η νοοτροπία του.

Ο αλγόριθμος ξεκινάει ελέγχοντας προς τα πίσω τις κινήσεις του παίκτη, οι οποίες αποθηκεύονται με την σειρά την οποία γίνονται για την διάρκεια του παιχνιδιού. Αναζητούνται μοτίβα οσονδήποτε κινήσεων, ξεκινώντας από μια κίνηση και αυξάνοντας κατά ένα σε κάθε έλεγχο μέχρι το μισό του συνόλου των κινήσεων, καθώς όπως αναφέρθηκε μοτίβο θεωρείται ένα σύνολο κινήσεων που επαναλαμβάνεται δύο φορές όποτε το μεγαλύτερο μοτίβο που μπορεί να υπάρξει στο σύνολο των κινήσεων του παίκτη είναι μήκους ίσο με το μισό του συνόλου των κινήσεων. Ο έλεγχος γίνεται με τη σειρά των κινήσεων όπως επιλέχθηκαν αυτές από τον παίκτη ξεκινώντας από το τέλος προς την αρχή.

Όταν βρεθεί κάποιο μοτίβο, όσες κινήσεις και να περιέχει, ενεργοποιείται το κομμάτι του αλγορίθμου που παίζει τις αντίστοιχες κινήσεις απέναντι στο μοτίβο αυτό. Αν για παράδειγμα το μοτίβο είναι πέτρα-πέτρα-ψαλίδι τότε ο αλγόριθμος ξεκινάει να επαναλάβει τις απαντήσεις στο μοτίβο αυτό οι οποίες είναι οι αντίθετες από αυτές του μοτίβου στην ανάλογη σειρά, δηλαδή χαρτί-χαρτί-πέτρα. Αρχικά ο αλγόριθμος ήταν σχεδιασμένος να παίζει τις κινήσεις αυτές από την αρχή μέχρι το τέλος του μήκους του μοτίβου αλλά γρήγορα έγινε αντιληπτό ότι όταν βρισκόταν κάποιο λάθος στις κινήσεις του έπρεπε να σταματάει την χρήση του μοτίβου αυτού κατευθείαν και να χρησιμοποιεί τον βασικό αλγόριθμο. Στην αρχική εκδοχή ο αλγόριθμος σταματούσε την ανίχνευση μοτίβου όταν ήδη είχε βρεθεί κάποιο και ξανά ξεκινούσε μόνο όταν έκανε λάθος, είτε άλλαζε ο παίκτης στρατηγική απλά. Η ανίχνευση του λάθους γίνεται ελέγχοντας τις κινήσεις του αντιπάλου εάν ακολουθεί ακόμα το μοτίβο αυτό, συγκρίνοντας τες με τις κινήσεις που χρησιμοποιούσε προηγουμένως. Στην πορεία της εργασίας ο τρόπος που δουλεύει ο πράκτορας άλλαξε καθώς διαπιστώθηκε ότι είναι καλύτερο να γίνεται η αναζήτηση μοτίβου σε κάθε κίνηση, ανεξάρτητα με το αν έχει ήδη ανιχνευτεί και εφαρμόζεται κάποιο μοτίβο, οπότε όταν είναι ενεργοποιημένη η ρύθμιση αυτή η ανίχνευση δεν σταματάει ποτέ. Στην τελική εκδοχή ο έλεγχος του μοτίβου γίνεται στο τέλος κάθε κίνησης καθώς όταν γινότανε στην αρχή της επόμενης, παρόλο που και στις δύο αυτές εκδοχές γινότανε πριν παίξει την κίνηση του ο πράκτορας, ο αλγόριθμος «έχανε» την τελευταία κίνηση, αργούσε δηλαδή μια κίνηση να αντιληφθεί ότι έχει κάνει λάθος.

Παρατηρήθηκε στην πορεία της εργασίας ότι κάποιες φορές ο αλγόριθμος της εύρεσης μοτίβου είναι δυνατόν να εντοπίσει μοτίβα τα οποία ο παίκτης δεν έχει αντιληφθεί ότι ακολουθεί μέσα στην αλληλουχία κινήσεών του, είτε γιατί αυτά σχηματίζονται μέσα από την επανάληψη των κινήσεων του μοτίβου είτε με κάποιες κινήσεις προγενέστερες της χρήσης του μοτίβου που, τυχαία, ταιριάζουν με το μοτίβο που ήθελε να ακολουθήσει. Για παράδειγμα, αν ο παίκτης αποφάσισε να ακολουθήσει το μοτίβο R (Rock) – P (Paper) – P (Paper) – S (Scissors) και η τελευταία κίνηση πριν την έναρξη της χρήσης του μοτίβου ήτανε S τότε σχηματίζεται η παρακάτω αλληλουχία κινήσεων.

…. S-R-P-P-S-R-P-P-S

Ενώ ο παίκτης έχει διαλέξει να χρησιμοποιεί το μοτίβο R-P-P-S η τελευταία του κίνηση πριν από την πρώτη επανάληψη των κινήσεων του δημιουργεί ένα άλλο μοτίβο για τον πράκτορα, το S-R-P-P. Αυτό βέβαια δεν δημιουργεί κάποια σύγχυση στον αλγόριθμο καθώς οι κινήσεις με τις οποίες απαντάει και στα δύο μοτίβα είναι ακριβώς οι ίδιες, απλά αλλάζει η σειρά τους ώστε να προσαρμοστεί στο ανάλογο μοτίβο που εντόπισε.

Κάτι ανάλογο παρατηρείται και όταν το μοτίβο επαναληφθεί πολλές φορές συνεχόμενα. Επειδή ο αλγόριθμος ελέγχει για νέα μοτίβα σε κάθε γύρο καθώς ο παίκτης προχωράει στην επομένη κίνηση και ο πράκτορας αναζητάει μοτίβα συμπεριλαμβάνοντας την νέα τελευταία κίνηση του παίκτη σε κάθε γύρο εντοπίζει ένα διαφορετικό μοτίβο καθώς η ανακύκλωση των κινήσεων του παίκτη προχωράει. Για παράδειγμα:

**Πίνακας 3 Παράδειγμα ανακύκλωσης κινήσεων**

| Γύρος | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Παίκτης | R | P | P | S | R | P | P | S | R | P | P | S |
| Πράκτορας | … | … | … | … | … | … | … | … | P | S | S | R |

Στον παραπάνω πίνακα (πίνακας 3) δίνονται οι κινήσεις του παίκτη σε μια τριπλή επανάληψη του μοτίβου, με το σημείο που τελειώνει το μοτίβο που επέλεξε ο παίκτης να σημειώνεται με μια πιο σκούρα γραμμή, η πρώτη του επανάληψη που εντοπίζει ο αλγόριθμος με ανοιχτό γκρι και η δεύτερη με σκούρο γκρι. Οι πρώτες οχτώ κινήσεις του πράκτορα δεν μας ενδιαφέρουν στο παράδειγμα αυτό. Στο τέλος του 8$^{ου}$ γύρου ο πράκτορας έχει εντοπίσει το μοτίβο R-P-P-S οπότε απαντάει στον 9$^{ο}$ με P σύμφωνα με το μοτίβο αυτό. Στο τέλος του ένατου γύρου ο πράκτορας ξανακάνει αναζήτηση για μοτίβα, συμπεριλαμβάνοντας την νέα κίνηση που χρησιμοποιήθηκε από τον παίκτη, και το νέο μοτίβο που εντοπίζει είναι το εξής: P-P-S-R (πίνακας 4).

**Πίνακας 4 Παράδειγμα ανακύκλωσης κινήσεων**

| Γύρος | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Παίκτης | R | P | P | S | R | P | P | S | R | P | P | S |
| Πράκτορας | … | … | … | … | … | … | … | … | P | S | S | R |

Όπως αναφέρθηκε νωρίτερα αυτό δεν προκαλεί καμιά σύγχυση στον πράκτορα καθώς είναι όντος μοτίβο, το οποίο προκαλείται από την αλληλουχία των κινήσεων του παίχτη καθώς επαναλαμβάνει το μοτίβο που είχε επιλέξει. Η απάντηση του πράκτορα θα ήταν ακριβώς η ίδια ακόμα και αν δεν έκανε έλεγχο για νέα μοτίβα και χρησιμοποιούσε το αρχικό που είχε εντοπιστεί στο τέλος του 8ου γύρου. Η μόνη διαφορά είναι ότι στην αρχική περίπτωση ο πράκτορας θα έπαιζε στον 10ο γύρο την δεύτερη κίνηση του μοτίβου που είχε βρεθεί στο τέλος του 8ου γύρου ενώ, εφόσον γίνεται αναζήτηση σε κάθε γύρο για νέο μοτίβο, αυτό που θα κάνει είναι να παίξει την πρώτη κίνηση του νέου μοτίβου που βρέθηκε στο τέλος του 9ου γύρου. Οι δύο όμως αυτές κινήσεις είναι ακριβώς οι ίδιες οπότε η απάντηση του πράκτορα είναι οι ίδια και στις δύο περιπτώσεις. Το ίδιο γίνεται και σε κάθε γύρο που ακολουθεί, μέχρι φυσικά να αλλάξει τακτική ο παίκτης κάτι που δεν γίνεται στο παράδειγμα αυτό.

**Πίνακας 5 Παράδειγμα ανακύκλωσης κινήσεων**

| Γύρος | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Παίκτης | R | P | P | S | R | P | P | S | R | P | P | S |
| Πράκτορας | … | … | … | … | … | … | … | … | P | S | S | R |

Σημειώνεται ξανά ότι η διαδικασία αυτή αποτελεί εσωτερικό μηχανισμό του αλγορίθμου και δεν γίνεται αντιληπτή από τον χρήστη ο οποίος έχει την εντύπωση ότι παίζει ένα συγκεκριμένο μοτίβο χωρίς να καταλαβαίνει ότι η αλληλουχία των κινήσεων που χρησιμοποιεί δημιουργεί και άλλα μοτίβα πέρα από αυτό που επέλεξε. Ο λόγος που αναφέρεται είναι σαν ενδιαφέρουσα παρατήρηση της διαδικασίας εύρεσης μοτίβων και δεν έχει καμία διαφορά στο τελικό αποτέλεσμα των κινήσεων που θα παίξει η μηχανή, σε όλες τις περιπτώσεις απαντάει όπως θα έπρεπε στις κινήσεις του αντιπάλου.

### 5.2.5 Memory

Σκοπός του αλγορίθμου αυτού είναι όταν ένα μοτίβο έχει χρησιμοποιηθεί ξανά κατά την διάρκεια του παιχνιδιού να μην χρειάζεται να επαναληφθεί ξανά ώστε να θεωρηθεί μοτίβο αλλά να ξεκινάει η απάντηση στο μοτίβο αυτό με την πρώτη επανάληψη. Είναι προφανές ότι ο αλγόριθμος αυτός μπορεί να χρησιμοποιηθεί μόνο όταν είναι ενεργοποιημένος ο αλγόριθμος αναζήτησης μοτίβων γι’ αυτό και δεν δίνεται η δυνατότητα χρήσης του όταν δεν είναι.

Κάθε φορά που εντοπίζεται ένα μοτίβο με την χρήση του pattern detection αυτό αποθηκεύεται στην μνήμη του αλγορίθμου. Κατά την διάρκεια του παιχνιδιού η αλληλουχία των κινήσεων του παίκτη ελέγχεται από τον αυτόν τον αλγόριθμο και αναζητείται ταυτοποίηση ενός συνόλου κινήσεων με κάποιο αποθηκευμένο μοτίβο στην μνήμη του πράκτορα. Εάν βρεθεί κάτι τέτοιο στην μνήμη, μια ακολουθία ίδιου μήκους και αλληλουχίας κινήσεων, τότε ο αλγόριθμος ενεργοποιεί το παίξιμο από μνήμης και ξεκινάει την απάντηση στο μοτίβο το οποίο συναντήθηκε. Η διαδικασία συνεχίζεται, όπως και με τον αλγόριθμο αναζήτησης μοτίβου, μέχρις ότου ο πράκτορας κάνει λάθος, είτε γιατί ο αντίπαλος δεν είχε σκοπό να επαναλάβει το μοτίβο και απλά έτυχε είτε γιατί άλλαξε τακτική. Ο αλγόριθμος συνεχίζει την αναζήτηση στην μνήμη για ταυτοποίηση μιας ακολουθίας κινήσεων ως αποθηκευμένο μοτίβο κατά την διάρκεια του παιξίματος από μνήμη και ενημερώνεται ανάλογα με την επόμενη κίνηση του αντιπάλου. Αν για παράδειγμα μια ακολουθία κινήσεων του παίκτη αναγνωριστεί ως μοτίβο αλλά ταυτόχρονα αποτελεί μέρος ενός μεγαλύτερου μοτίβου το οποίο ακολουθεί ο παίκτης στις επόμενες κινήσεις του, όταν γίνει η ταυτοποίηση του μεγαλύτερου μοτίβου τότε ο αλγόριθμος ενημερώνεται και προσαρμόζει της κινήσεις του στο νέο αυτό μοτίβο.

Κατά την διαδικασία του αλγορίθμου αυτού ο αλγόριθμος εύρεσης μοτίβων δεν σταματάει την αναζήτησή του και συνεχίζει να ελέγχει της κινήσεις του παίκτη σε κάθε γύρο. Στην πορεία ανάπτυξης του πράκτορα αποφασίστηκε ο αλγόριθμος μνήμης να έχει προτεραιότητα έναντι του αλγορίθμου εύρεσης μοτίβου. Όταν δηλαδή ο αλγόριθμος μνήμης εντοπίσει στις κινήσεις του παίκτη κάποιο μοτίβο που είχε επαναληφθεί προηγουμένως και έχει ξεκινήσει να το εφαρμόζει να μην μπορεί να διακοπεί από τον αλγόριθμο εύρεσης μοτίβου όταν αυτός εντοπίσει κάποιο μοτίβο καθώς εφαρμόζεται η μνήμη. Ο μόνος τρόπος να διακοπεί η εφαρμογή της μνήμης είναι να κάνει λάθος αυτή, οπότε και ο πράκτορας επιστρέφει στην ουδέτερη κατάσταση και αναζητάει νέο μοτίβο παίζοντας τον βασικό αλγόριθμο. Αυτό γίνεται διότι η ακολουθία κινήσεων που εφαρμόζεται από την μνήμη μπορεί να περιέχει μέσα της μικρότερα μοτίβα τα οποία αν είχε προτεραιότητα ο αλγόριθμος αναζήτησης μοτίβου θα εντόπιζε και θα εφάρμοζε οδηγώντας τον πράκτορα να κάνει λάθος. Εξάλλου, είναι λογικό ο πράκτορας να μην έχει λόγο να αλλάξει τακτική εφόσον δεν έχει κάνει ακόμα κάποιο λάθος στις κινήσεις του.

### 5.2.6 Randomness

Σε παίγνια όπως είναι το πέτρα ψαλίδι χαρτί είναι δυνατόν ο παίκτης να προσπαθήσει να ξεγελάσει τον αντίπαλο του και να τον κάνει να πιστέψει ότι θα παίξει μια συγκεκριμένη κίνηση, προσποιούμενος για παράδειγμα ότι συνεχίζει να ακολουθεί ένα μοτίβο αλλά τελικά παίζοντας την αντίθετη κίνηση από αυτή που περιμένει ο αντίπαλος κερδίζοντας τον.

Παρόλο που, όπως αναφέρθηκε, σκοπός της εφαρμογής αυτής δεν είναι να δημιουργηθεί ένας πράκτορας ο οποίος θα προσπαθεί να μεγιστοποιήσει τις πιθανότητες νίκης του αλλά η προσαρμογή της μηχανής, στο επίπεδο δυνατό, στις κινήσεις του εκάστοτε παίκτη με απώτερο σκοπό την εξαγωγή συμπερασμάτων από την αλληλεπίδραση αυτή, αναπτύχθηκε η τεχνική αυτή ώστε ο πράκτορας να γίνει λιγότερο προβλέψιμος. Στην κλασική εκδοχή του παιγνίου αυτού, είτε σε μια εκδοχή στην οποία οι αποδώσεις έχουν μικρή διαφορά μεταξύ τους, αυτό δεν αποτελεί ιδιαίτερο πρόβλημα καθώς το να θυσιάσει ο παίκτης μερικές κινήσεις για να κερδίσει μία δεν έχει ιδιαίτερο αποτέλεσμα. Σε μια εκδοχή όμως στην οποία η μία από της κινήσεις δίνει πολύ μεγαλύτερη ανταμοιβή από τις άλλες τότε κάποιος παίκτης θα μπορούσε να το εκμεταλλευτεί αυτό προσπαθώντας να ξεγελάσει τον πράκτορα να παίξει την κίνηση την οποία κερδίζει η πιο προσοδοφόρα αυτή στρατηγική.

Αυτό που κάνει η τεχνική αυτή για να αντιμετωπίσει το προαναφερθέν πρόβλημα και να γίνει λιγότερο προβλέψιμος ο πράκτορας είναι όταν έχει βρεθεί ένα μοτίβο, ή παίζεται από την μνήμη, να δίνεται 50% πιθανότητα να χρησιμοποιηθεί όντως η ανάλογη στο μοτίβο κίνηση αλλά και 50% πιθανότητα να παιχτεί από τον πράκτορα η κίνηση που έπαιξε ο ίδιος ο παίκτης στην πρώτη κίνηση του μοτίβου. Αυτό γίνεται γιατί η κίνηση αυτή είναι η αντίθετη, νικάει, της κίνησης που ο χρήστης θα χρησιμοποιήσει προσπαθώντας να ξεγελάσει το μηχάνημα, εξουδετερώνοντας ουσιαστικά την τακτική αυτή καθώς έχει 50% πιθανότητα ο χρήστης να κερδίσει και 50% να χάσει ακολουθώντας την τακτική αυτή. Η τεχνική αυτή βέβαια παρουσιάζει το πρόβλημα ότι μειώνει την χρησιμότητα του αλγορίθμου εύρεσης μοτίβων, κατά 50%, καθώς ακόμα και αν ο παίκτης δεν έχει σκοπό να ξεγελάσει τον πράκτορα ο πράκτορας δεν μπορεί να το γνωρίζει αυτό με αποτέλεσμα να έχει 50% πιθανότητα λάθους. Λάθος όμως το οποίο δεν χρεώνεται στον πράκτορα με ήττα στον γύρο αυτό αλλά με ισοπαλία διότι θα παίξει την ίδια ακριβώς κίνηση με τον παίκτη εάν αυτός εξακολουθεί να παίζει το μοτίβο. Δίνεται όμως η δυνατότητα χρήσης της σαν απάντηση στο πρόβλημα της προβλεψιμότητας των πρακτόρων. Λόγο της χρήσης των μοτίβων μια κίνησης, δηλαδή δύο ίδιες κινήσεις συνεχόμενες, από τους παίκτες στην προσπάθεια τους να ξεγελάσουν τον αλγόριθμο με την ενεργοποίηση του randomness οι κινήσεις αυτές δεν θεωρούνται πια μοτίβο, καθώς εξάλλου η πιθανότητα να δημιουργηθούν τυχαία είναι μεγάλη έτσι και αλλιώς. Δίνεται προσοχή στο γεγονός ότι η εφαρμογή αυτής της τεχνικής μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα για την λειτουργία του πράκτορα όταν προσπαθούν να εξαχθούν συμπεράσματα για την συμπεριφορά του οπότε καλό είναι να αποφεύγεται η χρήση της όταν γίνονται δοκιμές είτε σκοπός του παιχνιδιού είναι να φανεί η λειτουργία των υπολοίπων αλγορίθμων. Στην εκδοχή της εξομοίωσης του παιχνιδιού, πράκτορα εναντίον πράκτορα, η συγκεκριμένη τεχνική δεν έχει ιδιαίτερο νόημα καθώς δεν πρόκειται να συναντηθεί η συμπεριφορά την οποία σχεδιάστηκε να αντιμετωπίζει και απλά θα κάνει τα αποτελέσματα πιο τυχαία, παρόλο που δίνεται η δυνατότητα χρήσης της.

## *5.3 Συμπεράσματα*

Με βάση όλα αυτά που αναφέρθηκαν στα προηγούμενα κεφάλαια αναπτύχθηκε μια εφαρμογή, ένας πράκτορας, ο οποίος έχει την δυνατότητα να προσαρμόζεται στις κινήσεις του αντιπάλου και να μπορεί να εκμεταλλεύεται την μη τυχαία συμπεριφορά του. Οι τρεις βασικοί αλγόριθμοι προσεγγίζουν το πρόβλημα της επιλογής κίνησης με διαφορετική ματιά και οι διάφοροι αλγόριθμοι/τεχνικές βοηθούν στην προσαρμογή του πράκτορα στις εκάστοτε καταστάσεις. Η απόδοση του πράκτορα και το αποτέλεσμα του παιχνιδιού, ανάλογα και με τις επιλεγμένες ρυθμίσεις, διαφέρει ανάλογα με τον τρόπο παιξίματος του αντιπάλου, χρήστη ή και μηχανής καθώς δίνεται η δυνατότητα να συγκριθούν οι αλγόριθμοι μεταξύ τους σε ένα είδος εξομοίωσης του παιχνιδιού.

# Κεφάλαιο 6

## ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο κεφάλαιο αυτό παρουσιάζονται και σχολιάζονται τα αποτελέσματα των δύο ειδών παιχνιδιού, το παιχνίδι πράκτορα εναντίον πράκτορα και χρήστη εναντίον πράκτορα. Περιγράφονται ο τρόπος και η οπτική γωνία σύμφωνα με την οποία έγιναν οι δοκιμές των δύο ειδών για κάθε ενεργοποιημένο αλγόριθμο και τεχνική.

# 6. Αποτελέσματα

## *6.1 Γενικά*

Τα αποτελέσματα της εφαρμογής μπορούνε να χωριστούνε σε δύο γενικές κατηγορίες, στα αποτελέσματα του παιχνιδιού πράκτορα εναντίον παίκτη, που η συμπεριφορά είναι πιο απρόβλεπτη, και στα αποτελέσματα πράκτορα εναντίον πράκτορα, που κινούνται μέσα στα συγκεκριμένα πλαίσια και δυνατότητες των πρακτόρων. Κάθε κατηγορία παράγει διαφορετικά αποτελέσματα ανάλογα με την εκδοχή του παιχνιδιού που έχει επιλεχθεί, τους επιλεγμένους βασικούς αλγορίθμους και τους ενεργοποιημένους αλγορίθμους/τεχνικές παιξίματος.

## *6.2 Παιχνίδι πράκτορα εναντίον πράκτορα*

Κάθε πράκτορας μπορεί να χρησιμοποιεί διαφορετικές ρυθμίσεις επιλογής κίνησης και το μόνο σίγουρο κοινό που μπορεί να υπάρξει είναι οι κοινές αποδόσεις για κάθε στρατηγική, κάτι που δεν είναι απόλυτο καθώς δίνεται η δυνατότητα να προσαρμοστεί ο πίνακας αποδόσεων σε οποιεσδήποτε τιμές, ακόμα και διαφορετικές για κάθε παίκτη. Στην περίπτωση αυτή όμως η δυναμική του παιγνίου και η μορφή του μπορούν να αλλάξουν δραματικά, δημιουργώντας ένα τελείως διαφορετικό παίγνιο κάτι το οποίο δεν ήταν σκοπός και δεν εξετάζεται από αυτήν την πτυχιακή.

Θα εξεταστούν δύο εκδοχές του παιχνιδιού. Η πρώτη είναι με αποδόσεις ενός βαθμού σε κάθε νίκη, η προεπιλεγμένη προσαρμοσμένη εκδοχή, και η δεύτερη με τα ενδεικτικά νούμερα τις πτυχιακής αυτής, η διαφοροποιημένη εκδοχή (variant). Η περίπτωση της κλασικής μορφής του παιγνίου, σύμφωνα με την θεωρία παιγνίων, στην οποία δίνεται και ποινή ενός βαθμού στον παίκτη που χάνει, μηδενικού αθροίσματος, δεν εξετάζεται καθώς χαρακτηρίζεται από την ίδια ισορροπία που υπάρχει και στην προσαρμοσμένη εκδοχή και δεν προσφέρει κάτι παραπάνω για ανάλυση, η οποία ανάλυση γίνεται και πιο δύσκολη καθώς υπάρχει προσθαφαίρεση βαθμών και το τελικό αποτέλεσμα γίνεται πιο δυσνόητο. Αρχικά εξετάζεται η περίπτωση που και οι δύο πράκτορες χρησιμοποιούν τις ίδιες ρυθμίσεις επιλογών κίνησης (mirror match), δηλαδή οι ρυθμίσεις για machine 1 και machine 2 είναι ακριβώς οι ίδιες.

Ο αριθμός των επαναλήψεων έχει βαρύτητα καθώς οι αλγόριθμοι χρειάζονται κάποιο αριθμό κινήσεων για να δώσουν προφανή αποτελέσματα. Μετά από τις πρώτες δοκιμές αποφασίστηκε να γίνονται 200 επαναλήψεις για αρχή κάθε διαμάχης καθώς το νούμερο φάνηκε να είναι επαρκές για να δώσει μια εικόνα της συμπεριφοράς των αλγορίθμων. Όποτε κρίθηκε αναγκαίο οι επαναλήψεις μειώθηκαν ή αυξήθηκαν, εάν για παράδειγμα φαινόταν να υπάρχει κάποιο πιο συγκεκριμένο σημείο στο οποίο οι αλγόριθμοι έρχονται σε ισορροπία. Οι βασικοί αλγόριθμοι δοκιμάστηκαν και σε πολύ περισσότερες επαναλήψεις, δεν είναι τόσο χρονοβόροι καθώς κάθε κίνηση χρειάζεται συγκεκριμένο χρόνο για να εκτελεστεί. Δοκιμάστηκαν ακόμα και 50000 επαναλήψεις αλλά φάνηκε ότι ακόμα και πολύ μικρότερα νούμερα από αυτό είναι αρκετά καθώς μετά από ένα σημείο δεν υπάρχει καμία διαφορά στα αποτελέσματα. Η χρήση των αλγορίθμων αναζήτησης μοτίβων και της μνήμης αποδείχθηκαν πολύ πιο χρονοβόροι από την άλλη πλευρά, ιδίως η χρήση μνήμης, καθώς όσο αυξάνονται οι κινήσεις τόσο περισσότερο χρόνο χρειάζεται ο πράκτορας για να ελέγξει το ιστορικό του και την μνήμη του για να αποφασίσει την κίνηση του.

### 6.2.1 Percentages εναντίον Percentages

*Προεπιλογή αποδόσεων*

Οι δύο πράκτορες διαγωνίζονται με τις προεπιλεγμένες αποδόσεις, 1 βαθμό για νίκη, με μόνη ρύθμιση και για τους δύο να είναι ο ίδιος βασικός αλγόριθμος, ο percentages.

Όπως είναι λογικό οι δύο πράκτορες δεδομένου αρκετού αριθμού επαναλήψεων, ώστε να ισορροπήσουν ως ένα σημείο τα αποτελέσματα, βρίσκονται πάντα κοντά στην βαθμολογία. Εφόσον χρησιμοποιούν και οι δύο τις ίδιες ρυθμίσεις η νίκη εναλλάσσεται ανάμεσα στους δύο πράκτορες χωρίς κάποια ιδιαίτερη παρατήρηση. Ο συγκεκριμένος αλγόριθμος επιλέγει τις κινήσεις του βάση ποσοστών των κινήσεων του αντιπάλου, οπότε εφόσον ο αντίπαλος χρησιμοποιήσει και τις τρεις κινήσεις πάντα υπάρχει πιθανότητα, έστω και μικρή, να επιλεχθεί οποιαδήποτε κίνηση. Συνεπώς δεν επιλέγει πάντα μια συγκεκριμένη κίνηση, ακόμα και αν η μεγαλύτερη πιθανότητα είναι να χρησιμοποιηθεί μια συγκεκριμένη δεν είναι σίγουρο ότι θα επιλεχθεί αυτή. Αυτό διαφοροποιεί την διαφορά πόντων καθώς το αποτέλεσμα επηρεάζεται από την τύχη ως ένα βαθμό. Γενικά οι αλγόριθμοι με τον αριθμό επαναλήψεων να είναι 200 βρίσκονταν πάντα κοντά στην βαθμολογία με τους αριθμούς να κυμαίνονται από 55 έως 85 περίπου και με την διαφορά τους να κυμαίνεται συνήθως στο συν πλην 20.

Στην συνέχεια έγινε αύξηση των επαναλήψεων και παρατηρήθηκε ότι η αναλογία της βαθμολογίας παραμένει σταθερή με τις μικρότερες επαναλήψεις απλά αυτό που αλλάζει είναι ότι η ισορροπία πέρα από το σκορ έρχεται σιγά σιγά και στα ποσοστά των κινήσεων, δηλαδή προτιμούνται οι τρεις στρατηγικές περίπου με το ίδιο ποσοστό. Η εναλλαγή των κινήσεων μεταξύ τους γίνεται με κανονικούς ρυθμούς χωρίς κάτι ιδιαίτερο.

Ιδιαίτερη περίπτωση θεωρούνται τα παιχνίδια στα οποία και οι δύο πράκτορες ξεκινάνε με την ίδια κίνηση καθώς από πολύ νωρίς στις επαναλήψεις έρχεται ισορροπία κινήσεων καθώς οι αλγόριθμοι απαντάνε ο ένας στον άλλο με τις ίδιες σχεδόν κινήσεις, διαφοροποιημένες από το γεγονός ότι υπάρχει ο παράγοντας τύχη που προαναφέρθηκε στο συγκεκριμένο είδος.

*Διαφοροποιημένη εκδοχή*

Ο συγκεκριμένος αλγόριθμος δεν εξαρτάται από τις αποδόσεις των κινήσεων οπότε δεν παρατηρείται κάποια διαφορετική συμπεριφορά των κινήσεων καθώς η μηχανή δεν επηρεάζεται από τις τιμές αυτές στην επιλογή των κινήσεων της προσαρμοζόμενη απλά στις κινήσεις του αντιπάλου. Εφόσον ο αντίπαλος πράκτορας δεν επηρεάζεται, αντίθετα με τους χρήστες, η ισορροπίες στην βαθμολογία και τα ποσοστά των κινήσεων παραμένουν οι ίδιες με το προηγούμενο παράδειγμα.

### 6.2.2 Fictitious play εναντίον Fictitious play

*Προεπιλογή αποδόσεων*

Ο συγκεκριμένος αλγόριθμος λειτουργεί διαφορετικά από τον προηγούμενο, παρόλο που και αυτός είναι ποσοστιαίος, επιλέγοντας την κίνηση με την καλύτερη προσδοκώμενη ωφέλεια. Οι επιλογές του είναι πολύ πιο συγκεκριμένες καθώς σε κάθε κίνηση επιλέγεται μια συγκεκριμένη στρατηγική με την πιθανότητα των άλλων στρατηγικών να είναι μηδέν. Η μόνη περίπτωση που επιλέγει τυχαία είναι όταν οι προσδοκώμενες ωφέλειες δύο στρατηγικών είναι ακριβώς οι ίδιες κάτι που όμως συμβαίνει τόσο παροδικά και σπάνια που είναι αμελητέο.

Αυτό που έγινε γρήγορα αντιληπτό, με την παρατήρηση του γραφήματος, είναι ότι επιλέγονται κινήσεις με πολύ συγκεκριμένο τρόπο. Οι πράκτορες επαναλαμβάνουν την κίνηση με την μεγαλύτερη προσδοκώμενη ωφέλεια συνέχεια μέχρι να αλλάξουν τα ποσοστά και να βρεθεί άλλη κίνηση που να δίνει την μεγαλύτερη ωφέλεια, κάτι που ενίοτε μπορεί να αργήσει για πολλές κινήσεις.

Στον συγκεκριμένο αλγόριθμο παρατηρήθηκε ότι εφόσον οι δύο πράκτορες δεν ξεκινήσουν με την ίδια κίνηση το άθροισμα των βαθμολογιών τους είναι πάντα ίσο με το σύνολο των κινήσεων, ανεξάρτητα από το πλήθος των επαναλήψεων. Δηλαδή οι δύο πράκτορες δεν έρχονται ποτέ σε ισοπαλία. Αυτό που γίνεται είναι να εναλλάσσονται διαστήματα κατά τα οποία ο ένας πράκτορας κερδίζει συνέχεια μέχρι να αλλάξουν τα ποσοστά κινήσεων και να έρθει η σειρά του άλλου πράκτορα να κερδίζει για πολλές συνεχόμενες κινήσεις. Αυτό που γίνεται όσο αυξάνονται οι επαναλήψεις είναι να μεγαλώνει το διάστημα κατά το οποίο κερδίζονται οι συνεχόμενη πόντοι. Όσον αφορά τα ποσοστά των κινήσεων δεν παρατηρήθηκε κάποια συγκεκριμένη συμπεριφορά, απλά εναλλάσσονται μεταξύ τους χωρίς να ισορροπούνε.

Στην περίπτωση που και οι δύο πράκτορες ξεκινήσουν με την ίδια κίνηση τότε αυτό που γίνεται είναι να φέρνουν ισοπαλία για κάποιο μικρό αριθμό κινήσεων στην αρχή μέχρι να ανέβουν τα ποσοστά και να επαναληφθεί η συμπεριφορά που αναφέρθηκε προηγουμένως. Στις πρώτες κινήσεις θα φέρουν κάποιες συνεχόμενες ισοπαλίες αλλά αυτή θα είναι και η τελευταία φορά που θα έρθουνε σε ισοπαλία. Η διαφορά του αθροίσματος των βαθμολογιών των δύο πρακτόρων με τον αριθμό των επαναλήψεων είναι αυτός ακριβώς ο αριθμός των ισοπαλιών στην αρχή του παιχνιδιού.

Η διαφορά στην βαθμολογία των δύο πρακτόρων αυτήν την φορά μπορεί να είναι μεγαλύτερη αλλά απλά εξαρτάται από το πότε σταματάει το παίγνιο. Όπως αναφέρθηκε το σκορ εναλλάσσεται ανάμεσα στους δύο πράκτορες.

*Διαφοροποιημένη εκδοχή*

Οι διαφορετικές αποδόσεις σε αυτόν τον αλγόριθμο επηρεάζουν την προσδοκώμενη ωφέλεια των κινήσεων αλλά αυτό δεν επηρεάζει άμεσα το αποτέλεσμα της διαμάχης των δύο πρακτόρων καθώς είναι και για τους δύο το ίδιο. Αυτό που αλλάζει σε σχέση με την προηγούμενη εκδοχή είναι το ποια κίνηση θα επιλέγονταν στην αντίστοιχη περίπτωση καθώς το κριτήριο επιλογής έχει μεταβληθεί.

### 6.2.3 NEMS εναντίον NEMS

*Προεπιλογή αποδόσεων*

Ο αλγόριθμος αυτός παίζει την ισορροπία Nash που προκύπτει από τον πίνακα αποδόσεων χωρίς να εξαρτάται από τις κινήσεις του αντιπάλου. Η βαθμολογία των δύο πρακτόρων βρίσκεται πάντα κοντά στην ισοπαλία, δεδομένου ότι χρησιμοποιούνε και οι δύο τα ίδια ακριβώς ποσοστά επιλογής κινήσεων.

Από το γράφημα είναι προφανές και για τους δύο πράκτορες ότι παίζεται η ισορροπία, δεδομένου ότι δίνεται αρκετός αριθμός επαναλήψεων για να φανεί αυτό στατιστικά. Περίπου στις 500 επαναλήψεις γίνεται πια σχεδόν πάντα προφανές, πολλές φορές και νωρίτερα, ότι τα ποσοστά επιλογής κινήσεων και από τους δύο πράκτορες έρχονται σε ισορροπία, είναι περίπου ίσα μεταξύ τους. Όσο αυξάνονται οι κινήσεις τόσο πιο κοντά έρχονται στην ισορροπία.

*Διαφοροποιημένη εκδοχή*

Με την αλλαγή των αποδόσεων αυτό που αλλάζει είναι το ποια είναι η ισορροπία Nash και φαίνεται ότι τα ποσοστά των κινήσεων συγκλίνουν σε αυτήν την ισορροπία αυτή τη φορά, που είναι περίπου 46% πέτρα, 30% χαρτί και 23% ψαλίδι.

### 6.2.4 Pattern detection, memory, randomness

Η ενεργοποίηση των αλγορίθμων και τεχνικών pattern detection και memory δεν έχει τόσο μεγάλη επίπτωση σε καθαρά ποσοστιαίους αλγορίθμους, όπως είναι ο percentages και ο NEMS. Η χρησιμότητα του pattern detection βασίζεται στην επανάληψη μοτίβων, κάτι που παρόλο που μπορεί να δημιουργηθεί και τυχαία από αυτούς τους βασικούς αλγορίθμους είναι σπάνιο και δεν επηρεάζει ιδιαίτερα το αποτέλεσμα. Η χρησιμότητα του memory δεδομένου του προηγούμενου γεγονότος είναι ακόμα πιο περιορισμένη. Τα μόνα συχνά χρησιμοποιούμενα μοτίβα είναι η επανάληψη της ίδιας κίνησης που επιφέρει την αντίστοιχη απάντηση του αντίπαλου πράκτορα αλλά εφόσον ισχύει και για τους δύο παίκτες δεν επηρεάζει δραματικά το αποτέλεσμα. Οι ισορροπίες που αναφέρθηκαν νωρίτερα για τους βασικούς αυτούς αλγορίθμους εξακολουθούν να ισχύουν, ανεξάρτητα διαφοροποίησης πίνακα αποδόσεων.

Με την χρήση του pattern detection στην ιδιαίτερη περίπτωση του fictitious play παρατηρείται ότι το τελικό αποτέλεσμα επηρεάζεται παραπάνω από τις άλλες περιπτώσεις. Δεν ισχύει η παρατήρηση του αθροίσματος των βαθμολογιών που έγινε νωρίτερα, το άθροισμά τους δεν είναι ίσο με το σύνολο των επαναλήψεων. Οι δύο πράκτορες βρίσκονται σχεδόν πάντα πολύ κοντά στην βαθμολογία αλλά δεν υπάρχουν πια τα μεγάλα διαστήματα κατά τα οποία ο ένας πράκτορας κέρδιζε πολλούς συνεχόμενους γύρους παίζοντας την ίδια κίνηση και μετά αυτό εναλλάσσονταν με τον άλλο πράκτορα. Λόγο της εύρεσης του μοτίβου αυτού οι επαναλήψεις της ίδιας κίνησης συνεχόμενα είναι πολύ λιγότερες, 5-6 μέγιστο περίπου, από εκεί που ήτανε ακόμα και εκατοντάδες. αν οι συνολικές επαναλήψεις του παιγνίου σκηνή ήταν χιλιάδες, καθώς η απάντηση του αντίπαλου πράκτορα στο μοτίβο αυτό αλλάζει γρήγορα τα ποσοστά των κινήσεων και τα ανακυκλώνει πολύ πιο γρήγορα. Δεν παρατηρούνται επίσης όπως αναφέρθηκε τόσο μεγάλες διαφορές στην βαθμολογία όσο υπήρχανε χωρίς την χρήση του pattern detection. Για την ακρίβεια συνήθως η διαφορά είναι μικρότερη ή περίπου 5 βαθμοί μόνο, καθώς όπως προαναφέρθηκε οι πράκτορες δεν επαναλαμβάνουν ποια λόγο της εύρεσης του μοτίβου αυτού την ίδια νικηφόρα κίνηση πολλές φορές συνεχόμενες.

Η χρήση του memory δεν επηρεάζει ιδιαίτερα τα παραπάνω αποτελέσματα καθώς σκοπός του είναι να κερδίσει ορισμένες κινήσεις προλαβαίνοντας ένα μοτίβο που έχει επαναληφθεί και νωρίτερα αλλά στην περίπτωση αυτή τα μοτίβα είναι πολύ μικρά, 2-4 κινήσεων συνήθως, όποτε δεν επηρεάζει το αποτέλεσμα σε κάποιο σημαντικό βαθμό.

Μια ενδιαφέρουσα παρατήρηση είναι ότι ενώ χωρίς την χρήση του pattern detection στον fictitious play τα ποσοστά κινήσεων των πρακτόρων ήταν ακανόνιστα και απλά εναλλάσσονταν μεταξύ τους, με την χρήση του οι κινήσεις συγκλίνουν στην ισορροπία Nash της εκάστοτε εκδοχής του παιγνίου. Στην προεπιλεγμένη εκδοχή, με απόδοση 1 βαθμό στην νίκη, τα ποσοστά συγκλίνουν στην ισότητα, ισορροπία Nash 33% σε κάθε κίνηση. Αλλά και στην διαφοροποιημένη εκδοχή γίνεται ακριβώς το ίδιο με την ισορροπία Nash για τα ανάλογα ποσοστά, 46% πέτρα, 30% χαρτί και 23% ψαλίδι.

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο η χρήση της τεχνικής randomness δεν ενδείκνυται για το παιχνίδι πράκτορα εναντίον πράκτορα καθώς σχεδιάστηκε για να αντιμετωπίζει μια περίπτωση που τυχαίνει μόνο εναντίον των χρηστών καθώς ο πράκτορας δεν μπορεί να «σκεφτεί» με αυτόν τον τρόπο και να προσπαθήσει να ξεγελάσει τον αντίπαλο με αυτό το σκεπτικό. Το μόνο που επιτυγχάνεται με την ενεργοποίηση του είναι να γίνουν πιο τυχαία τα αποτελέσματα.

### 6.2.5 Percentages εναντίον Fictitious play

*Προεπιλογή αποδόσεων*

Σε 100 παιχνίδια των 200 γύρων μόνο με την χρήση των βασικών αλγορίθμων ο percentages φάνηκε να υπερτερεί του fictitious play κερδίζοντας τα 58 παιχνίδια έναντι 40 με τους δύο αλγορίθμους να έρχονται σε ισοπαλία 2 φορές. Ο percentages κατά μέσο όρο κέρδιζε 89.9 βαθμούς ενώ ο fictitious play 85.4 . Η επανάληψη των ίδιων κινήσεων από τον fictitious play βοήθησε τον percentages να κερδίζει περισσότερους γύρους.

Με την ενεργοποίηση του αλγορίθμου εύρεσης μοτίβων το αποτέλεσμα έγινε πιο προφανές καθώς ο percentages κέρδισε και τα 100 παιχνίδια με μέσο όρο βαθμολογίας τους 85.9 βαθμούς έναντι 69.5 του fictitious play.

Η χρήση του αλγορίθμου μνήμης και από τους δύο πράκτορες δεν επηρέασε ιδιαίτερα το αρχικό αποτέλεσμα παρόλο που έδωσε στον fictitious play ορισμένα παιχνίδια, είτε έκανε τον percentages να χάσει κάποια μόνος του. Ο percentages κέρδισε 76 έναντι 22 του fictitious play με 2 ισοπαλίες και μέσο όρο 82.2 έναντι 75.3 αντίστοιχα. Η συνεχής επανάληψη των μικρών μοτίβων, 3-4 κινήσεων, έδωσε στον fictitious play ορισμένες νίκες αλλά δεν άλλαξε την ισορροπία του παιχνιδιού.

*Διαφοροποιημένη εκδοχή*

Με την αλλαγή των αποδόσεων και την χρήση μόνο των βασικών αλγορίθμων δεν επηρεάστηκε το αποτέλεσμα, ο percentages κέρδισε για άλλη μια φορά με 59 έναντι 40 και 1 ισοπαλία. Ο μέσος όρος ήταν 14288.0 και του FP 12077.5 .

Η ενεργοποίηση της εύρεσης μοτίβων είχε το αναμενόμενο αποτέλεσμα καθώς η αλλαγή των αποδόσεων δεν έχει μεγάλη επίπτωση όταν υπάρχουν συνεχή μοτίβα στις κινήσεις ενός παίκτη. Ο percentages κέρδισε και τα 100 παιχνίδια με μέσο όρο 12842.0 έναντι 9469. 5 του FP.

Με την χρήση μνήμης και για τους δύο πράκτορες η απόδοση του FP βελτιώθηκε ως ένα σημείο αλλά φάνηκε αρκετά κατώτερος του percentages κερδίζοντας μόνο 14 παιχνίδια έναντι 86 του percentages. Οι μέσοι όροι βαθμολογίας ήταν 10820.0 και 12474.0 αντίστοιχα. Η αλλαγή των αποδόσεων φαίνεται να κάνει τον FP να προτιμάει ακόμα ποιο συγκεκριμένες κινήσεις κάνοντας τον πιο ευάλωτο και στον αλγόριθμο μνήμης.

### 6.2.6 Percentages εναντίον NEMS

*Προεπιλογή αποδόσεων*

Οι δύο αλγόριθμοι βρίσκονταν πάντα πολύ κοντά στο σκορ καθώς ο NEMS χρησιμοποιεί την ισορροπία Nash και ο percentages χρησιμοποιεί τα ποσοστά αυτά για να απαντήσει οπότε οι δύο πράκτορες φαίνονται σχεδόν ισοδύναμοι. Ο percentages υπερτερούσε οριακά με 49 νίκες έναντι 46 του αντιπάλου και 5 ισοπαλίες. Οι μέσοι όροι βαθμολογίας ήτανε 67.3 έναντι 66.7 αντίστοιχα, που είναι αρκετά κοντά ώστε οι αλγόριθμοι να θεωρούνται ισοδύναμοι.

Η χρήση του pattern detection λειτούργησε οριακά υπέρ του NEMS δίνοντας του το προβάδισμα με 48 νίκες έναντι 46 του percentages και 6 ισοπαλίες. Οι μέσοι όροι ήταν 70.6 και 69.7 αντίστοιχα.

Με την χρήση της μνήμης δεν άλλαξε κάτι στην ισορροπία των δύο πρακτόρων, η οριακή υπεροχή του NEMS διατηρήθηκε με 52 νίκες έναντι 43 του percentages και 5 ισοπαλίες. Όπως και πριν οι μέσοι όροι βρίσκονταν πολύ κοντά με 74.7 και 73.0 αντίστοιχα.

*Διαφοροποιημένη εκδοχή*

Η αλλαγή των αποδόσεων λειτούργησε υπέρ του NEMS κερδίζοντας του ακόμα περισσότερα παιχνίδια απ’ ότι η προκαθορισμένη προσαρμοσμένη εκδοχή. Κέρδισε 76 έναντι 24 του percentages με μέσο όρο 10412.5 έναντι 9193.0 αντίστοιχα. Οι ποσοστιαίοι αλγόριθμοι διαλέγουν κινήσεις, όπως έχει αναφερθεί, διαλέγοντας με συγκεκριμένη κατανομή ποσοστών σύμφωνα με τις κινήσεις του αντιπάλου αλλά ο NEMS παίζει πάντα με τα ποσοστά της ισορροπίας NASH τα οποία είναι, θεωρητικά, η καλύτερη κατανομή ποσοστών αν είναι να παίξει κανείς τυχαία.

Η διαφοροποίηση των αποδόσεων συνέχισε να ευνοεί τον NEMS και με την χρήση της αναζήτησης μοτίβων. Ο NEMS κέρδισε 71 και έχασε 28 με 1 ισοπαλία. Ο pattern detection βοήθησε τον percentages να ανεβάσει τον μέσο όρο του αρκετά στο 10079.5 ενώ ο μέσος όρος του NEMS αυξήθηκε λιγότερο στο 10822.0 Οι μέσοι όροι και των δύο αυξήθηκαν γιατί υπήρξαν περισσότερες νίκες στους γύρους με την εύρεση μοτίβων στις κινήσεις του αντιπάλου με προφανή παραπάνω βελτίωση του percentages αλλά όχι αρκετή για να κάνει την διαφορά.

Η ενεργοποίηση της μνήμης βοήθησε τον percentages να κερδίσει ορισμένα παιχνίδια παραπάνω, 36 έναντι 64 του NEMS. Οι μέσοι όροι ήταν 10888.5 και 11314.5 βελτιώνοντας τις αποδόσεις και των δύο παικτών, καθώς η μνήμη βοηθάει στον εντοπισμό των επαναλαμβανόμενων μοτίβων νωρίτερα.

### 6.2.7 Fictitious play εναντίον NEMS

*Προεπιλογή αποδόσεων*

Στα 100 παιχνίδια των 200 γύρων οι πράκτορες ήταν για άλλη μια φορά αρκετά κοντά με τον NEMS να κερδίζει τα 49 έναντι 46 του FP (Fictitious Play) με 5 ισοπαλίες. Οι μέσοι όροι ήταν 67.1 με 66.7 αντίστοιχα.

Με ενεργοποιημένη την εύρεση μοτίβων ο NEMS κέρδισε και τα 100 παιχνίδια, καθώς όπως έχει φανεί ως τώρα η επανάληψη των ίδιων κινήσεων από τον fictitious play τον καθιστά πολύ ευάλωτο στον αλγόριθμο εύρεσης μοτίβων. Ο μέρος όρος βαθμολογίας του NEMS ήταν 85.0 και του fictitious play 68.7 .

Η μνήμη διατήρησε την διαφορά των δύο αλγορίθμων σε υψηλά επίπεδα με τον NEMS να κερδίζει 75 παιχνίδια και να χάνει 23 με 2 ισοπαλίες, με μέσο όρο 82.5 έναντι 75.9 του fictitious play. Η λειτουργία της μνήμης όταν ενεργοποιήθηκε για κάθε αλγόριθμο ξεχωριστά δεν αλλοίωσε την υπεροχή του NEMS καθώς η επανάληψη των ίδιων κινήσεων του fictitious play τον καθιστούσε πολύ ευάλωτο στην εύρεση μοτίβων. Η ενεργοποίηση της μνήμης και για τους δύο πράκτορες όμως τελικά οδήγησε στην απόκτηση ορισμένων βαθμών υπέρ του fictitious play καθώς λειτούργησε υπέρ της ποικιλότητας των κινήσεων του κερδίζοντας τα 23 αυτά παιχνίδια, ανεβάζοντας τον μέσο όρο του στο 75.9 από 68.7 και παράλληλα μειώνοντας τον μέσο όρο του NEMS σε 82.5 από 85. Δεδομένου ότι τα μοτίβα που τυχόν χρησιμοποιούνται και από τους δύο πράκτορες είναι μικρά οπότε και πιο εύκολο να επαναληφθούν η χρήση μνήμης επηρεάζει το αποτέλεσμα αλλά πολύ περισσότερο όταν είναι ενεργοποιημένη και από τους δύο πράκτορες. Στην συγκεκριμένη περίπτωση ο NEMS με την χρήση της μνήμης γίνεται λίγο πιο προβλέψιμος καθώς επαναλαμβάνει τις ίδιες κινήσεις σαν απάντηση στα προηγούμενα μοτίβα του fictitious play, δηλαδή όταν έχει ενεργοποιημένη την μνήμη λειτουργεί κατά του, ενάντια στον fictitious play πάντα. Παράλληλα, όπως αναφέρθηκε, η μνήμη βοηθάει τον fictitious play να είναι λιγότερο προβλέψιμος καθώς η επανάληψη των ίδιων κινήσεων που χρησιμοποιεί γίνεται πιο αραιή με τις κινήσεις της μνήμης ενδιάμεσα. Οι δύο αυτοί παράγοντες είναι που ανέβασαν την απόδοση του fictitious play όταν η μνήμη ενεργοποιήθηκε και για τους δύο πράκτορες.

*Διαφοροποιημένη εκδοχή*

Ο NEMS μεγάλωσε την διαφορά του από τον FP στην διαφοροποιημένη εκδοχή σε σχέση με την προκαθορισμένη προσαρμοσμένη κερδίζοντας 56 παιχνίδια έναντι 41 του αντιπάλου του με 3 ισοπαλίες. Οι μέσοι όροι ήταν 10321.5 και 9145.0 αντίστοιχα.

Η χρήση του pattern detection είχε το αναμενόμενο, τώρα πια, αποτέλεσμα καθώς το μειονέκτημα του fictitious play έναντι στην εύρεση μοτίβων θεωρείται δεδομένο και η αλλαγή των αποδόσεων σε τέτοιες περιπτώσεις μικρής σημασίας. Ο NEMS κέρδισε και τα 100 παιχνίδια με 13331.5 μέσο όρο βαθμολογίας έναντι 9490.0 του FP.

Η μνήμη, όπως και σε όλες τις άλλες περιπτώσεις, όταν ενεργοποιήθηκε και για τους δύο πράκτορες βελτίωσε την θέση του FP ελαφρά κερδίζοντας 10 παιχνίδια και χάνοντας τα άλλα 90. Οι μέσοι όροι ήταν 10422.0 για τον FP και 12445.0 για τον NEMS. Η βελτίωση του FP στην βαθμολογία και η επιδείνωση του NEMS έφεραν τους δύο πράκτορες πιο κοντά στην βαθμολογία αλλά η μεγάλη διαφορά είναι ακόμα προφανής.

### 6.2.8 Συγκεντρωτικός πίνακας αποτελεσμάτων διαμάχης διαφορετικών αλγορίθμων

**Πίνακας 6 Συγκεντρωτικός πίνακας αποτελεσμάτων διαμάχης διαφορετικών αλγορίθμων**

| | | | PER | FP | TIE | PER | NEMS | TIE | FP | NEMS | TIE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CUSTOM | ΒΑΣΙΚΟΣ | ΣΚΟΡ | 58 | 40 | 2 | 49 | 46 | 5 | 46 | 49 | 5 |
| | | Μ.Ο. | 89.9 | 85.4 | | 67.3 | 66.7 | | 66.7 | 67.1 | |
| CUSTOM | Pattern detection | ΣΚΟΡ | 100 | 0 | 0 | 46 | 48 | 6 | 0 | 100 | 0 |
| | | Μ.Ο. | 85.9 | 69.5 | | 69.7 | 70.6 | | 68.7 | 85.0 | |
| CUSTOM | memory | ΣΚΟΡ | 76 | 22 | 2 | 43 | 52 | 5 | 23 | 75 | 2 |
| | | Μ.Ο. | 82.2 | 75.3 | | 73.0 | 74.7 | | 75.9 | 82.5 | |
| VARIANT | ΒΑΣΙΚΟΣ | ΣΚΟΡ | 59 | 40 | 1 | 24 | 76 | 0 | 41 | 56 | 3 |
| | | Μ.Ο. | 14288.0 | 12077.5 | | 9193.0 | 10412.5 | | 9145.0 | 10321.5 | |
| VARIANT | Pattern detection | ΣΚΟΡ | 100 | 0 | 0 | 28 | 71 | 1 | 0 | 100 | 0 |
| | | Μ.Ο. | 12852.0 | 9469.5 | | 10079.5 | 10822.0 | | 9449.0 | 13331.5 | |
| VARIANT | memory | ΣΚΟΡ | 86 | 14 | 0 | 36 | 64 | 0 | 10 | 90 | 0 |
| | | Μ.Ο. | 12474.0 | 10820.0 | | 10888.5 | 11314.5 | | 10422.0 | 12445.0 | |

## *6.3 Παιχνίδι χρήστη εναντίον πράκτορα*

Οι αλγόριθμοι δοκιμάστηκαν ενάντια σε διάφορους παίκτες ώστε να εξαχθούν πιο αντικειμενικά συμπεράσματα. Χρησιμοποιήθηκαν οι διαφοροποιημένη εκδοχή και η προκαθορισμένη προσαρμοσμένη εκδοχή, όπως έγινε και στο παιχνίδι πράκτορα εναντίον πράκτορα.

Δοκιμάστηκαν διάφοροι αριθμοί επαναλήψεων, πολύ λιγότεροι απ’ ότι στην περίπτωση πράκτορα εναντίον πράκτορα καθώς όπως προαναφέρθηκε δεν είναι πρακτική η χρήση μεγάλου αριθμού επαναλήψεων σε αυτή την περίπτωση, παρόλο που δοκιμάστηκαν και λίγο μεγαλύτερες επαναλήψεις σε κάποιες περιπτώσεις. Οι διάφοροι αλγόριθμοι δοκιμάζονταν συνήθως σε 20 ή 40 επαναλήψεις. Σκοπός της εφαρμογής ήταν να ανταποκρίνονται οι αλγόριθμοι σε λίγες επαναλήψεις καθώς ο αρχικός στόχος ήταν να χρησιμοποιηθούν μόνο εναντίον χρηστών. Όπως έχει αναφερθεί βέβαια δεν είναι δυνατόν να εξαχθούν συμπεράσματα για την λειτουργία των αλγορίθμων με πολύ λίγες επαναλήψεις διότι δεν προλαβαίνουν όλοι οι αλγόριθμοι να εξάγουν συγκεκριμένα συμπεράσματα αλλά ούτε οι παίκτες να αναπτύξουν πιο περίπλοκα σκεπτικά κινήσεων. Για τους λόγους αυτούς αποφασίστηκε οι ελάχιστες επαναλήψεις να είναι οι 20.

Η μεγαλύτερη διαφορά με το παιχνίδι πράκτορα εναντίον πράκτορα είναι ότι οι παίκτες είναι λιγότερο προβλέψιμοι απ’ ότι οι πράκτορες και μπορούνε να αλλάζουν στρατηγική στην μέση του παιχνιδιού, αντίθετα με τους πράκτορες οι οποίοι πάντα χρησιμοποιούν την ίδια κεντρική ιδέα για την επιλογή των κινήσεων τους, ανάλογα με τις ενεργοποιημένες ρυθμίσεις, ακόμα και αν μπορούν να προσαρμοστούν όταν αντιλαμβάνονται συγκεκριμένες συμπεριφορές. Σημειώνεται επίσης ότι λόγο της φύσης του παιχνιδιού οι παίκτες τείνουν να χρησιμοποιούν συνήθως τυχαίες επιλογές κινήσεων, τουλάχιστον στην αρχή του παιχνιδιού.

Η προσέγγιση των αποτελεσμάτων εναντίον του χρήστη είναι διαφορετική από αυτήν εναντίον της μηχανής καθώς συγκεκριμένοι αλγόριθμοι λειτουργούν καλύτερα ενάντια σε συγκεκριμένες συμπεριφορές οπότε εξετάζεται η απόδοση της εφαρμογής σε τέτοιες καταστάσεις και όχι καθαρά αριθμητικά αποτελέσματα, τα οποία δεν έχουν πολλά να προσφέρουν στην αξιολόγηση των καταστάσεων αυτών.

Όπως έχει προαναφερθεί σκοπός της πτυχιακής αυτής δεν ήταν να δημιουργηθεί μια εφαρμογή που να κερδίζει απαραίτητα τον χρήστη αλλά να προσαρμόζεται, στο επίπεδο δυνατό, στις κινήσεις του εκάστοτε παίκτη και με απώτερο σκοπό την εξαγωγή συμπερασμάτων από την αλληλεπίδραση αυτή.

### 6.3.1 Προκαθορισμένη προσαρμοσμένη εκδοχή

Οι παίκτες φαίνεται να ξεκινάνε χωρίς κάποια ιδιαίτερη προτίμηση κινήσεων, στην αρχή οι κινήσεις τους είναι λίγο πολύ τυχαίες, που είναι και η ενδεδειγμένη τακτική για το παίγνιο αυτό με αυτές της αποδοχές. Οι περισσότεροι παίκτες στην πορεία του παιχνιδιού φάνηκε να προσπαθούν να προβλέψουν τις κινήσεις του υπολογιστή πιστεύοντας ότι ακολουθεί κάποια συγκεκριμένη στρατηγική και πιθανός χρησιμοποιεί κάποια συγκεκριμένα μοτίβα ανά περιόδους, που είναι λίγο πολύ και το σκεπτικό το οποίο χρησιμοποιεί ο πράκτορας για τους ίδιους τους παίκτες. Η φύση του παιχνιδιού είναι τέτοια που να ευνοεί την τυχαία επιλογή κινήσεων, που είναι και η ενδεδειγμένη ισορροπία Nash, κάτι στο οποίο το μόνο που μπορούσε να κάνει η εφαρμογή είναι να απαντάει στις κινήσεις του παίχτη με αυτόν τον τρόπο.

Ενάντια στις τρεις βασικές μορφές του αλγορίθμου το αποτέλεσμα κινούνταν συνήθως κοντά στην ισορροπία ανάμεσα στους δύο αντιπάλους. Οι ποσοστιαίοι αλγόριθμοι που απαντάνε με τα ανάλογα ποσοστά εάν ο χρήστης χρησιμοποιεί κάποιες κινήσεις παραπάνω από κάποιες άλλες έφεραν τα αναμενόμενα αποτελέσματα καθώς μετά από ορισμένες κινήσεις αυτή η συμπεριφορά γίνονταν αντιληπτή από τον αλγόριθμο και απαντούσε ανάλογα. Ο fictitious play απέδιδε καλύτερα με λιγότερες επαναλήψεις καθώς οι παίκτες μετά από ορισμένα παιχνίδια αντιλαμβάνονταν ότι ο αλγόριθμος είναι πιθανό να επαναλάβει κάποιο μοτίβο όποτε προσπαθούσαν να προβλέψουν αυτές τις κινήσεις μειώνοντας την απόδοση του καθώς, όπως έχει αναφερθεί αρκετές φορές, ο αλγόριθμος γίνεται όντως προβλέψιμος μετά από ένα σημείο διότι έχει συγκεκριμένη αντίληψη της «καλύτερης κίνησης». Ο NEMS, με την χρήση της ισορροπίας Nash, δεν γίνεται ποτέ προβλέψιμος καθώς χρησιμοποιώντας τα ποσοστά της ισορροπίας για την εκδοχή αυτή του παιγνίου, 33% για κάθε κίνηση, ανακύκλωνε τις κινήσεις του χωρίς να μπορούν οι χρήστες να προβλέψουν τι θα παίξει, καθώς όπως έχει αναφερθεί νωρίτερα ο καλύτερος τρόπος να μην γνωρίζουν οι αντίπαλοι τι θα παίξει ένας παίχτης είναι να μην γνωρίζει ούτε ο ίδιος τι θα παίξει. Σε τυχαίες συμπεριφορές και οι τρεις αλγόριθμοι ανταποκρίνονταν ανάλογα, με τον νικητή όπως είναι επόμενο να κρίνεται τυχαία. Όταν οι παίκτες προτιμούσαν συγκεκριμένες κινήσεις η απόδοση των ποσοστιαίων αλγορίθμων ανέβαινε δραματικά ενώ του NEMS παρέμενε σταθερή με την εξής όμως παρατήρηση: η συντριπτική πλειοψηφία των παικτών αντιλαμβανόταν ότι οι αλγόριθμοι προσαρμόζονταν σε αυτό το φαινόμενο και σταματούσαν να επαναλαμβάνουν τις προηγούμενες προτιμήσεις τους είτε απαντούσαν στον αλγόριθμο, στους ποσοστιαίους γιατί στον NEMS δεν συνέβαινε ποτέ αυτό το φαινόμενο, με την ανάλογη κίνηση κερδίζοντας αρκετές κινήσεις σε βάρος του, μέχρι να αλλάξουν τα ποσοστά και να προσαρμοστεί ο αλγόριθμος.

Η ενεργοποίηση της εύρεσης μοτίβων, κυρίως στα πρώτα παιχνίδια που έπαιζε κάποιος παίκτης εναντία στον πράκτορα, άλλαξε την ισορροπία υπέρ της μηχανής καθώς οποιοδήποτε μοτίβο επαναλαμβανόταν αναγνωρίζονταν και δίνονταν η ανάλογη απάντηση από τον αλγόριθμο. Η τεχνική αυτή είναι ανεξάρτητη από τον βασικό αλγόριθμο οπότε όταν υπάρχει κάποιο μοτίβο δεν επηρεάζεται καθόλου από το ποιος είναι ενεργοποιημένος. Οι παίκτες μετά από ένα σημείο όμως αντιλαμβάνονταν ότι ο πράκτορας καταλαβαίνει τους επαναλαμβανόμενους συνδυασμούς και διέκοπταν γρήγορα το συγκεκριμένο μοτίβο. Ιδίως όταν οι παίκτες έπαιζαν πολλά παιχνίδια εναντίων του αλγόριθμου αποφεύγονταν η χρήση μοτίβων και χρησιμοποιούνταν τυχαία επιλογή κινήσεων. Η εμπειρία ενάντια στην εύρεση μοτίβων έκανε του παίκτες να προσπαθούν να προβλέψουν τις κινήσεις του πράκτορα και τα κατάφερναν ως ένα σημείο καθώς η επόμενη κίνηση του πράκτορα αν ο παίκτης χρησιμοποιήσει κάποιο μοτίβο είναι συγκεκριμένη. Αυτός βέβαια είναι και ο στόχος του αλγορίθμου, και της πτυχιακής αυτής κατ επέκταση. Παρόλο που η εύρεση μοτίβων ήταν επιτυχείς στις κινήσεις των παικτών αυτό δημιούργησε το πρόβλημα ότι αλγόριθμος γινόταν προβλέψιμος και οι χρήστες που έπαιζαν πολλές φορές εναντίον του το αντιλαμβάνονταν αυτό και θυσιάζοντας ορισμένες κινήσεις μπορούσαν να «κλέψουν» την νίκη για έναν γύρο από τον πράκτορα. Αυτό βέβαια γινόταν μια φορά σε κάθε μοτίβο καθώς ο αλγόριθμος αντιλαμβανόταν το λάθος του και διέκοπτε την χρήση του μοτίβου απάντηση αμέσως. Παίκτες οι οποίοι ακολουθούσαν την συμπεριφορά αυτή φάνηκε ότι είχαν περισσότερες πιθανότητες νίκης απέναντι στις ρυθμίσεις αυτές.

Η χρήση μνήμης βοήθησε ως ένα βαθμό στην αντιμετώπιση του προηγούμενου φαινομένου, παρόλο που δεν ήταν πότε αυτός ο σκοπός της. Με την χρήση της τεχνικής αυτής εντοπίζονταν μοτίβα που είχανε χρησιμοποιηθεί από τους παίκτες νωρίτερα στο παιχνίδι και δινόταν απάντηση σε αυτά χωρίς την ανάγκη της δεύτερης επανάληψης του μοτίβου για την εύρεση του, κερδίζοντας έτσι ορισμένες παραπάνω κινήσεις. Με την χρήση του αλγορίθμου αυτού οι παίκτες χρειάζονταν να αλλάζουν το μοτίβο που τυχόν χρησιμοποιούσαν για αναγκάσουν τον αλγόριθμο να παίξει μια συγκεκριμένη κίνηση δυσκολεύοντας έτσι ως ένα βαθμό την διαδικασία αυτή.

Το πρόβλημα της προβλεψιμότητας όμως συνέχισε να υφίσταται και είναι και ο λόγος για το οποίο δημιουργήθηκε η τεχνική randomness. Με την χρήση της ο πράκτορας γίνεται πιο απρόβλεπτος, προσπαθώντας παράλληλα να «τιμωρήσει» τους παίκτες οι οποίοι προσπαθούν να τον ξεγελάσουν. Η τεχνική φαίνεται να δουλεύει στην πράξη καθώς αποθαρρύνει τους παίκτες να προσπαθήσουν να εκμεταλλευτούν την προβλεψιμότητα του pattern detection καθώς εξουδετερώνει τα ποσοστά επιτυχίας της στρατηγικής αυτής. Η απόδοση του πράκτορα εναντίον χρηστών που χρησιμοποιούν τέτοιες μεθόδους ανέβηκε αρκετά κάνοντας όμως το αποτέλεσμα να βασίζεται περισσότερο στη τύχη, που είναι και λογικό λόγο της φύσης του παιχνιδιού. Όταν βέβαια ο χρήστης όντως χρησιμοποιεί μοτίβα στις κινήσεις του η τεχνική αυτή έχει 50% πιθανότητα να κάνει λάθος, για την ακρίβεια ακόμα και όταν δεν χρησιμοποιηθεί το μοτίβο που έχει εντοπιστεί ο πράκτορας δεν θα χάσει στον γύρο αυτό απλά θα φέρει ισοπαλία διότι η 50% πιθανότητα να μην χρησιμοποιήσει το μοτίβο που βρήκε είναι να αντιγράψει ουσιαστικά της κινήσεις του αντιπάλου, κάτι σχετικά ανώδυνο για τον πράκτορα καθώς δεν μεγαλώνει η διαφορά στην βαθμολογία αλλά απλά μένει σταθερή. Ο πράκτορας όμως θα χάσει του βαθμούς που θα μπορούσε να είχε πάρει στην περίπτωση αυτή.

### 6.3.2. Διαφοροποιημένη εκδοχή

Στην διαφοροποιημένη εκδοχή η νίκη με πέτρα δίνει 200 βαθμούς, με ψαλίδι 150 και με χαρτί 100. Με την αλλαγή των αποδόσεων υπήρχαν διάφορες συμπεριφορές από τους παίκτες. Οι περισσότεροι παίκτες αύξησαν τα ποσοστά των κινήσεων με μεγαλύτερες αποδόσεις στις επιλογές τους αλλά υπήρχαν και παίκτες που έμειναν αδιάφοροι στην αλλαγή των αποδόσεων, εξακολουθώντας να επιλέγουν κινήσεις με τα ίδια ποσοστά. Υπήρχαν ακόμα και παίκτες, στο πνεύμα του ότι ο πράκτορας προσπαθεί να προβλέψει τις κινήσεις τους, που πίστευαν ότι επειδή ο υπολογιστής θα περιμένει να αυξήσουν τα ποσοστά των κινήσεων με τις μεγαλύτερες αποδόσεις έκαναν ακριβώς το ανάποδο. Δηλαδή μείωσαν τα ποσοστά των κινήσεων με μεγαλύτερες αποδόσεις, πιστεύοντας ότι ο πράκτορας θα έπαιζε περισσότερο χαρτί για παράδειγμα για να ανταποκριθεί στην αύξηση του ποσοστού πέτρας των παικτών, πιστεύοντας ότι θα ξεγελάσουν τον αλγόριθμο με αυτόν τον τρόπο.

Οι βασικοί αλγόριθμοι προσαρμόζονται με συγκεκριμένο τρόπο στην αλλαγή αυτή των αποδόσεων και τις συμπεριφοράς των παικτών. O percentages δεν επηρεάζεται άμεσα από την αλλαγή των αποδόσεων, αυτό που κάνει απλά είναι να προσαρμόζεται στα ποσοστά κινήσεων του παίκτη ,όποια και αν είναι αυτά, οπότε δεν υπάρχει κάποια αλλαγή στην γενική εικόνα του αλγορίθμου. O fictitious play από την άλλη χρησιμοποιεί και τα ποσοστά του αντιπάλου αλλά και τον πίνακα αποδόσεων για να επιλέξει της κινήσεις του. Όσο μεγαλώνει το ποσοστό που ο αντίπαλος παίζει πέτρα για παράδειγμα μεγαλώνει η πιθανότητα να παίξει χαρτί αλλά και αν μεγαλώσει η απόδοση της πέτρας αυξάνεται πάλι η πιθανότητα να παίξει χαρτί, κάτι που είναι σταθερό κατά την διάρκεια του παιχνιδιού αλλά αλλάζει σε σχέση με τις αποδόσεις της προκαθορισμένης προσαρμοσμένης εκδοχής. O FP δηλαδή θεωρεί της κινήσεις με μεγαλύτερες αποδόσεις πιο κερδοφόρες ακόμα και αν ο παίκτης παίζει με τα ίδια ποσοστά που έπαιζε και στην προηγούμενη εκδοχή. O NEMS παίζει την ισορροπία Nash σε κάθε γύρο με τις ανάλογες πιθανότητες που έχουν αναφερθεί. Η γενική εικόνα των βασικών αλγορίθμων σε σχέση με την διαφοροποιημένη εκδοχή είναι βελτίωση ενάντια στον παίκτη καθώς ο αλγόριθμος προσαρμόζεται αυτόματα στις νέες συνθήκες ενώ οι παίκτες παίζουν κινήσεις διαφοροποιημένες μεν με την προηγούμενη εκδοχή αλλά χωρίς κάποια συγκεκριμένα ποσοστά στο μυαλό τους, κάτι που φαίνεται να ευνοεί τον πράκτορα. Μεγάλη επιδείνωση στην απόδοση του έδειξε ο FP καθώς οι κινήσεις έγινα ακόμα πιο συγκεκριμένες και απόλυτα προβλέψιμες ανά περιόδους. Οι percentages και NEMS από την άλλη ανέβασαν την απόδοσή τους.

Η χρήση της εύρεσης μοτίβων δεν έχει καμιά διαφορά από την προηγούμενη εκδοχή, καθώς δεν επηρεάζεται από αποδόσεις, όσον αφορά τα μοτίβα που εντοπίζονται και συνεχίζει να τα εντοπίζει όπου αυτά υπάρχουν. Η γενική εικόνα των αλγορίθμων είναι ίδια με αυτήν των βασικών αλγορίθμων όταν οι παίκτες αποφεύγουν τα μοτίβα, προστατεύοντας όμως ως έναν βαθμό τους αλγορίθμους που επαναλαμβάνουν την ίδια κίνηση καθώς η απάντηση σε αυτούς αποτελεί μοτίβο οπότε ανέβασαν την απόδοσή τους ως ένα σημείο. Όποτε χρησιμοποιούνται μοτίβα λειτουργεί απόλυτα υπέρ του πράκτορα, πάντα με το γνωστό πλέον πρόβλημα της προβλεψιμότητας. Στην συγκεκριμένη περίπτωση μάλιστα, που κάποιες κινήσεις δίνουν παραπάνω βαθμούς όπως η νίκη με πέτρα, οι παίκτες με αυτό το σκεπτικό προσπαθούσαν να εκβιάσουν κινήσεις από τον πράκτορα επιδιώκοντας να τον κάνουν να παίξει ψαλίδι ώστε να αποκομίσουν περισσότερους βαθμούς ξεγελώντας τον, μέθοδος που φάνηκε να δουλεύει επιτυχώς πολλές φορές αν ακολουθούνταν σωστά.

Η χρήση μνήμης επίσης δεν επηρεάζεται από την αλλαγή αποδόσεων όταν δουλεύει όπως προβλέπεται. Γενικά ισχύει η εικόνα του παιχνιδιού που αναφέρθηκε και στην περίπτωση της χρήσης μόνο του αλγορίθμου εύρεσης μοτίβων. Ο πράκτορας βελτιώνεται όπως και στην προσαρμοσμένη εκδοχή με την χρήση της μνήμης και έχει την αναμενόμενη βελτίωση που αναφέρθηκε σε αυτήν και για το πρόβλημα της προβλεψιμότητας.

Η τεχνική randomness φαίνεται ακόμα πιο χρήσιμη στην εκδοχή αυτή εναντίον των παικτών καθώς όταν ο πράκτορας γίνεται προβλέψιμος και γίνεται δυνατό να εκβιαστούν μεγαλύτερες βαθμολογίες με αυτές τις αποδόσεις. Η βελτίωση είναι αισθητή με την ενεργοποίησή της παρόλο που δεν εξαλείφει το πρόβλημα τελείως καθώς το αποτέλεσμα βασίζεται στην τύχη.

# Κεφάλαιο 7

## ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό παρουσιάζονται τα συμπεράσματα του κεφαλαίου των αποτελεσμάτων, αλλά και ολόκληρης της εργασίας γενικότερα. Γίνονται επίσης παρατηρήσεις για τους σημαντικότερους παράγοντες που επηρέασαν τα αποτελέσματα των πρακτόρων. Τέλος, γίνονται προτάσεις για περαιτέρω ερεύνα με βάση την εμπειρία που αποκτήθηκε κατά την ανάπτυξη της εφαρμογής.

# 7. Συμπεράσματα

## *7.1 Συμπεράσματα εφαρμογής*

Υπάρχουν διάφοροι παράγοντες που επηρεάζουν το αποτέλεσμα και την εξέλιξη του παιγνίου ανάλογα με το είδος που έχει επιλεχθεί και φυσικά της επιλεγμένες και ενεργοποιημένες ρυθμίσεις. Το πλήθος των γύρων είναι ένα από αυτά καθώς το αποτέλεσμα ορισμένων αλγορίθμων διαφοροποιείται, λιγότερο ή περισσότερο, ανάλογα με το αν υπάρχουν πολλές ή λίγες επαναλήψεις. Λογικό είναι επίσης να δύνεται και στον παίκτη δυνατότητα να χρησιμοποιήσει πιο περίπλοκα μοτίβα όταν ο αριθμός των επαναλήψεων είναι επαρκής. Ο αριθμός τον επαναλήψεων στο παιχνίδι πράκτορα εναντίον πράκτορα μπορεί να είναι κατά πολύ μεγαλύτερος και δεν περιορίζεται από παράγοντες που περιορίζεται το παιχνίδι χρήστη εναντίον πράκτορα καθώς καλό είναι να αποφεύγονται οι εκατοντάδες επαναλήψεις σε μια τέτοια περίπτωση γιατί απλά καταλήγουν οι παίκτες να επιλέγουν τυχαία.

Ο πίνακας αποδόσεων είναι ένας άλλος παράγοντας που επηρεάζει διότι μπορεί να προκαλέσει δραματική αλλαγή στην επιλογή κινήσεων των αλγορίθμων που επηρεάζονται από αυτόν, κυρίως στον fictitious play αλλά και στον NEMS. Οι διαφοροποιημένες αποδόσεις επηρεάζουν του παίκτες με διαφορετικό τρόπο και ενίοτε η σύγχυση που προκαλούν λειτουργεί υπέρ του πράκτορα. Παρατηρήθηκε ότι είναι δυνατό να εκμεταλλευτεί κάποιος παίκτης την προβλεψιμότητα τη εύρεσης μοτίβων και να χρησιμοποιήσει τις αλλαγές στις αποδόσεις των κινήσεων για να κερδίσει περισσότερους πόντους όταν καταφέρνει να ξεγελάσει των πράκτορα.

Όσον αφορά στο παιχνίδι πράκτορα εναντίον πράκτορα όπου διαγωνίζονται ίδιοι αλγόριθμοι μεταξύ τους στους ποσοστιαίους αλγόριθμους έγινε γρήγορα αντιληπτό ότι για το αποτέλεσμα και την εξέλιξη του παιγνίου είχε σημασία η πρώτη κίνηση των δύο αλγορίθμων. Αν δηλαδή και οι δύο πράκτορες ξεκινούσαν με την ίδια κίνηση το αποτέλεσμα και οι κινήσεις που επιλέγονταν από αυτούς βρίσκονταν μέσα σε ποιο συγκεκριμένα πλαίσια. Ανάλογα με το εάν ο αλγόριθμος περιέχει κάποιο βαθμό τυχαίας επιλογής, percentages, ή επέλεγε συγκεκριμένη κίνηση, fictitious play, το αποτέλεσμα σύγκλινε λιγότερο ή περισσότερο σε ισορροπία, με πάντα όμως την τάση να συγκλίνει να είναι προφανής. Οι βασικοί αλγόριθμοι βρίσκονται πάντα κοντά στην βαθμολογία και σύμφωνα με τα αποτελέσματα ο μέρος όρος βαθμολογίας είναι πιο χαρακτηριστικός απ' ότι το αποτέλεσμα των νικών, κάτι που ισχύει με οποιαδήποτε ρύθμιση ενεργοποιημένη.

Στα παιχνίδια των πρακτόρων μεταξύ τους φάνηκε να υπερτερεί ο αλγόριθμος NEMS με τον percentages να έρχεται δεύτερος και τον fictitious play τελευταίο λόγο του προβλήματος επανάληψης των ίδιων κινήσεων. Αυτό βέβαια δεν αποτελεί γενικό μειονέκτημα του αλγορίθμου fictitious play απλά τον καθιστά ακατάλληλο για τέτοιες συνθήκες καθώς είναι σχεδιασμένος για single agent περιβάλλοντα, περιβάλλοντα δηλαδή στα οποία υπάρχει μόνο ένας πράκτορας. Αυτό σημαίνει ότι ο πράκτορας σε αυτά δεν χρειάζεται να λαμβάνει υπόψη του άλλους πράκτορες στο περιβάλλον και τον απασχολεί μόνο το πώς οι δικές τους πράξεις επηρεάζουν το περιβάλλον, π.χ. ένα ρομπότ εξερεύνησης. Η προσαρμογή του αλγορίθμου αυτού στο multi agent ανταγωνιστικό περιβάλλον που αποτελεί το επιλεγμένο παίγνιο δουλεύει καλύτερα σε λίγες επαναλήψεις καθώς με πολλές επαναλήψεις ο αλγόριθμος κλειδώνει σε συγκεκριμένη κίνηση και αργεί να αλλάξει τακτική λόγο του αριθμού των κινήσεων που αργούν να αλλάξουν τα ποσοστά που χρησιμοποιεί. Η χρήση της ισορροπίας Nash από την άλλη φάνηκε να είναι η ενδεδειγμένη λύση καθώς είναι ο λιγότερο προβλέψιμος αλγόριθμος. Όπως γίνεται κατανοητό το αποτέλεσμα κάθε αλγορίθμου δεν γίνεται να μεταφραστεί με το ίδιο τρόπο καθώς χρησιμοποιούν διαφορετική νοοτροπία και λειτουργούν καλύτερα σε συγκεκριμένες συνθήκες ο καθένας.

Στο παιχνίδι χρήστη εναντίον πράκτορα τα συμπεράσματα εξετάζονται με τις πλήρεις δυνατότητες της εφαρμογής ενεργοποιημένες. Ο βασικός αλγόριθμος με την καλύτερη γενική επίδοση, όπως και στο παιχνίδι πράκτορα εναντίον πράκτορα, ήτανε ο NEMS με τους υπόλοιπους αλγορίθμους ενεργοποιημένους να βοηθάνε ανάλογα με τον αντίπαλο. Ο αλγόριθμος εύρεσης μοτίβων εντόπιζε κάθε μοτίβο που χρησιμοποιούσαν οι παίκτες και απαντούσε με τις ανάλογες κινήσεις, ακόμα και σε πιο περίπλοκα μοτίβα. Η χρήση μνήμης βοήθησε στον να κερδίζονται γύροι νωρίτερα όταν βρίσκονταν προηγουμένως χρησιμοποιούμενα μοτίβα στο ιστορικό της. Ο συνδυασμός των δύο αυτών αλγορίθμων έκανε τους παίκτες να σταματούνε την χρήση μοτίβων, κάτι που ίσως δείχνει την επιτυχία του αλγορίθμου στον τομέα αυτό, μόλις αντιλαμβάνονταν την δυνατότητα αυτή του πράκτορα. Στην προσπάθεια του όμως ο πράκτορας να εκμεταλλευτεί τις μη τυχαίες συμπεριφορές των αντιπάλων του έγινε προβλέψιμος ο ίδιος, για μία κίνηση σε κάθε μοτίβο, κάτι το οποίο φαίνεται αναπόφευκτο και η περαιτέρω ανάπτυξη του για προσπάθεια βελτίωσης των ποσοστών νίκης πρέπει να ξεκινάει με δεδομένο αυτό το φαινόμενο.

Οι ελάχιστες δυνατές κινήσεις που μπορούσαν να χρησιμοποιήσουν οι παίκτες για να ξεγελάσουν τον πράκτορα ήταν τρεις, δύο για το μοτίβο μιας κίνησης, πέτρα-πέτρα για παράδειγμα, και η τρίτη με την οποία προσπαθούν να κερδίσουν τον πράκτορα παίζοντας το αντίθετο από αυτό που περιμένουν να παίξει αυτός. Δηλαδή στο παράδειγμα πέτρα-πέτρα ο αλγόριθμος εντοπίζοντας το μοτίβο μιας κίνησης απαντάει με χαρτί και οι παίκτες που είχαν εντοπίσει το φαινόμενο έπαιζαν ψαλίδι. Η εύρεση μοτίβων και η μνήμη αναγκάζουν τον παίκτη να ανακυκλώνει τα μικρά αυτά μοτίβα ώστε να μην γίνει προβλέψιμος αλλά απλά καθυστερούν την τακτική του παίκτη. Σκοπός της πτυχιακής αυτής δεν ήτανε η ανάπτυξη μιας εφαρμογής που να νικάει τον παίκτη αλλά η αντιμετώπιση μη τυχαίων συμπεριφορών, που έγινε με τους αλγορίθμους εύρεσης μοτίβων και μνήμης. Παρόλα αυτά αναπτύχθηκε η τεχνική randomness για να προσπαθήσει να δώσει απάντηση στο πρόβλημα της προβλεψιμότητας. Με την τεχνική αυτή δόθηκε μιας κάποιας μορφής λύση στο πρόβλημα αυτό με 50% πιθανότητα να απαντήσει στην τακτική αυτή ο πράκτορας κάνοντας τον όμως να δουλεύει με ποιο τυχαία επιλογή κινήσεων. Με την χρήση της επίσης αποκλείστηκε από τα μοτίβα η διπλή επανάληψη μιας κίνησης, η οποία έτσι και αλλιώς αποτέλεσε ιδιάζουσα περίπτωση στην ανάπτυξη των αλγορίθμων εύρεσης μοτίβου καθώς είναι πολύ εύκολο να δημιουργηθεί τυχαία και θεωρήθηκε αμφιλεγόμενο τον αν θα έπρεπε να θεωρείται μοτίβο. Με την τελευταία τεχνική οι ελάχιστες κινήσεις που χρειάζονται για να ξεγελαστεί ο πράκτορας από τρεις ανέβηκαν σε πέντε κάνοντας πιο περίπλοκη την τακτική αυτή για τους παίκτες και βελτιώνοντας την επίδοση του πράκτορα.

Η ισορροπία μεταξύ του να σχεδιαστεί ο αλγόριθμος να προβλέπει κινήσεις αλλά να μην γίνεται προβλέψιμος ο ίδιος στην προσπάθειά του αυτή ήταν βασικό πρόβλημα στην προσπάθεια βελτίωσης του πράκτορα αλλά δεν αποτελούσε ποτέ σκοπό της πτυχιακής αυτής και αναπτύχθηκε ενδεικτικά.

Οι πρώτες κινήσεις των παικτών ήτανε πάντα πιο τυχαίες και φάνηκε να έχει σημασία ποιος κερδίζει μετά τις πρώτες 4-5 κινήσεις και με πόση διαφορά, ιδίως στους ποσοστιαίους αλγορίθμους οι οποίοι αποκτούνε ουσιαστικά πιο διαμορφωμένη στρατηγική μετά από τις κινήσεις αυτές. Η διαφορά σε αυτές της πρώτες κινήσεις είναι η βάση με την οποία ξεκινούσε η δεύτερη φάση του παιχνιδιού την οποία ο εκάστοτε παίκτης προσπαθούσε να καλύψει όταν πια το παιχνίδι έχει ξεκινήσει κανονικά και οι δύο πλευρές αρχίζουν να δείχνουν σημάδια της τακτικής τους. Λόγο της φύσης του παιχνιδιού που ευνοεί την τύχη το αποτέλεσμα των τυχαίων γύρων, είτε το γύρων που μεσολαβούν μέχρι να βρεθεί ένα μοτίβο, παίζει ρόλο στο τελικό αποτέλεσμα. Αυτό ισχύει και για τον χρήστη και για την μηχανή καθώς ο πράκτορας χρειάζεται μερικές κινήσεις για να εκτελεστεί το μοτίβο από τον παίκτη και να το εντοπίσει στις οποίες παίζει τυχαία εφόσον δεν έχει δεδομένα για κάτι άλλο είτε είναι ποσοστιαίος και χρησιμοποιεί τυχαία ποσοστά έτσι και αλλιώς. Για τον χρήστη ισχύει το ίδιο αν ακολουθεί την τακτική να προσπαθεί να ξεγελάσει τον πράκτορα καθώς το να κερδίσει έναν γύρο μπορεί να είναι ώφελος αλλά στις 4 υπόλοιπες, με την χρήση του randomness, κινήσεις το αποτέλεσμα είναι τυχαίο.

Η διαφορά κέρδους των δύο αντιπάλων δεν φαίνεται να έχει ιδιαίτερη επίπτωση λόγο της φύσης του παιχνιδιού, πέρα από ίσως μια αύξηση στις κινήσεις που δίνουν παραπάνω βαθμούς όταν ο ορίζοντας του παιγνίου σκηνή είναι κοντά, καθώς έτσι και αλλιώς οι παίκτες δεν μπορούν να κάνουν κάτι παραπάνω από το να προσαρμόζονται στην στρατηγική του αντιπάλου τους είτε να προσπαθούν να γίνονται όσο λιγότερο προβλέψιμοι γίνεται. Η επίδοση και των δύο αντιπάλων, χρήστη και πράκτορα, δεν επηρεάζεται, πέρα από τις πρώτες κινήσεις όπως προαναφέρθηκε, από τον αριθμό του γύρο στον οποίο βρίσκεται το παιχνίδι αλλά εξαρτάται από τις εκάστοτε συνθήκες και τον εντοπισμό αδυναμίας στις κινήσεις του αντιπάλου.

## *7.2 Περαιτέρω έρευνα*

Η χρήση αλγορίθμων εύρεσης μοτίβων και εντοπισμού μη τυχαίων συμπεριφορών θα είχε μεγαλύτερο αποτέλεσμα σε κάποιο παίγνιο στο οποίο οι παίκτες θα είχαν κίνητρο να ακολουθήσουν κάποιο μοτίβο καθώς στο πέτρα-ψαλίδι-χαρτί δεν υπάρχει κίνητρο για κάτι τέτοιο και η συνήθης, και ενδεικτική, συμπεριφορά είναι να επιλέγονται τυχαία κινήσεις. Ενδιαφέρον θα παρουσίαζε η ανάπτυξη ή μεταφορά του συγκεκριμένου αλγορίθμου σε ένα τέτοιο περιβάλλον ώστε να φανεί καλύτερα η απόδοση του σε τέτοιες συνθήκες καθώς μόνο για λόγους έλεγχου συναντήθηκαν πιο περίπλοκα μοτίβα από τους πραγματικούς παίκτες.

Περαιτέρω ανάπτυξη των αλγορίθμων θα μπορούσε να είναι η μετατροπή τους για νευρωνικά δίκτυα, και η σύγκριση τους με τον υπάρχον αλγόριθμο. Θα μπορούσαν επίσης να εξεταστούν οι διάφοροι μέθοδοι εύρεσης μοτίβων και μνήμης με την χρήση άλλων τεχνικών όπως είναι η χρήση λίστας ή πινάκων αναζήτησης καθώς οι συγκεκριμένοι αλγόριθμοι χρησιμοποιούνε διανύσματα.

Ένα άλλο κομμάτι των αλγορίθμων που παρουσιάζει ενδιαφέρον είναι η ταχύτητα αναζήτησης. Στο παιχνίδι πράκτορα εναντίον πράκτορα που η χρήση εκατοντάδων ή και χιλιάδων επαναλήψεων ήταν δυνατή ο αλγόριθμος της μνήμης όσο περνούσαν οι γύροι και η μνήμη μεγάλωνε γινόταν αρκετά χρονοβόρος με αποτέλεσμα όταν οι επαναλήψεις γίνονταν πάρα πολλές να χρειάζονται ώρες για την ολοκλήρωση του παιχνιδιού. Με την χρήση των τεχνικών που αναφέρθηκαν παραπάνω θα μπορούσε να εξεταστεί η βελτίωση της ταχύτητας του αλγορίθμου και η εύρεση βέλτιστης τεχνικής για τις διαδικασίας αυτές.

# Βιβλιογραφία

[1] Γ. Βαρουφάκης, (2006), Θεωρία παιγνίων, Η θεωρία που φιλοδοξεί να ενοποιήσει τις Κοινωνικές Επιστήμες, Δάρδανος-Τυπωθητώ, Αθήνα

[2] Π. Μηλιώτης, (Μάιος 2002), Θεωρία Παιγνίων, Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα

[3] A. W. Tucker and R. D. Luce (1959), Contributions to the Theory of Games, Princeton University press, Princeton, New Jersey

[4] Associated Press. (2006-06-26), "Exasperated judge resorts to child's game".

[5] Aumann, Robert J. (1987), "Game theory", Macmillan Publishers Limited, Houndmills, Basingstoke, Hampshire, England

[6] Avinash Dixit (June 2000), Princeton University, John Nash - Founder of Modern Game Theory

[7] Barlow, HB (1989), Unsupervised learning, MIT Press, Cambridge, England

[8] Cournot, Augustin A. (1838), Recherches sur les Principes Mathematiquesde la Theorie des Richesses, Paris

[9] Davis, M. Maschler, M. (1965), "The kernel of a cooperative game", Princeton University press, Princeton, New Jersey

[10] Dutta, Prajit K. (1999), Strategies and games, MIT Press, Cambridge, Massachusetts, USA

[11] Ethem Alpaydın (2004), Introduction to Machine Learning (Adaptive Computation and Machine Learning), MIT Press, Cambridge, Massachusetts, USA

[12] Fudenberg Drew, Tirole Jean (1991), Game theory, MIT Press, Cambridge, Massachusetts, USA

[13] Geoffrey Hinton, Terrence J. Sejnowski (editors) (1999): Unsupervised Learning: Foundations of Neural Computation, MIT Press, Cambridge, Massachusetts, USA

[14] Georgios Paliouras, Vangelis Karkaletsis, Constantine D. Spyropoulos (Eds.) (2001), Machine learning and it's applications, Springer, Berlin

[15] Harsanyi, J. C. (1967-8), Games with Incomplete Information Played by 'Bayesian' Players, University of California, Berkeley

[16] Harsanyi, John C. (1974), "An equilibrium point interpretation of stable sets", University of California, Berkeley

[17] Howard, Nigel (1971). Paradoxes of Rationality: Games, Metagames, and Political Behavior, MIT Press, Cambridge, Massachusetts, USA

[18] Kaelbling, Leslie P. Michael L. Littman Andrew W. Moore (1996). "Reinforcement Learning: A Survey", AI Access Foundation and Morgan Kaufmann Publishers, USA

[19] Kuhn, H.W. (1968). Simplicial approximation of fixed points, Princeton University press, Princeton, New Jersey

[20] Luce R D & Raiffa H. (1957). Games and decisions: introduction and critical survey, New York: Wiley, Harvard University, Cambridge

[21] M Frean, E R Abraham (2001), Rock-scissors-paper and the survival of the weakest, Victoria University, Wellington, New Zealand.

[22] Mailath, G. and Samuelson, L. (2006) Repeated games and reputations: long-run relationships, Oxford University Press, US

[23] Marco Wiering,(1999) Explorations in Efficient Reinforcement Learning, Institute of Information & Computing Sciences

[24] Martin Beckenkamp, Heike Hennig-Schmidt, Frank P. Maier-Rigaud, (2007) Cooperation in Symmetric and Asymmetric Prisoner's Dilemma Games, Max Planck Institute for Research on Collective Goods, Bonn

[25] Martin J. Osborne and Ariel Rubinstein (1994), A course in game theory, MIT Press, Cambridge, Massachusetts, USA

[26] Martin J. Osborne, (2004), An introduction to game theory, Oxford University Press, USA

[27] Maynard Smith, John (1972), Game Theory and the Evolution of Fighting, Edinburgh University Press, Edinburgh

[28] Michael Bowling, Manuela Veloso (2001), Rational and Convergent Learning in Stochastic Games, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA

[29] Nash, J. F. (1951), Non-Cooperative Games, Princeton University Press, Princeton NJ

[30] Nash, J. F. (1953), Two Person Cooperative Games, Princeton University Press, Princeton NJ

[31] Paul Walker (2005), “An Outline of the History of Game Theory", Department of Economics, University of Canterbury, Christchurch, New Zealand

[32] Peter Dayan (1999b), Unsupervised learning, MIT Press, Cambridge, Massachusetts, USA

[33] Poundstone, W. (1992) Prisoner's Dilemma, Anchor Books, Doubleday, NY

[34] Poundstone, William (1992), Prisoner's Dilemma: John von Neumann, Game Theory and the Puzzle of the Bomb, Anchor Books, Doubleday, NY

[35] Rasmusen, Eric (2006), Games and Information: An Introduction to Game Theory, Blackwell Publishers, Indiana

[36] Robert Axelrod (1984), The Evolution of Cooperation, Basic Books, NY

[37] Ross, Don. (2008), "Game Theory", The Stanford Encyclopedia of Philosophy, MIT Press, Cambridge, MA

[38] S. Kotsiantis (2007), Supervised Machine Learning: A Review of Classification Techniques, University of Peloponnese, Greece

[39] Selten, R. (1965), Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit, Zeitschrift fur die gesamte Staatswissenschaft

[40] Sutton, Richard S. Andrew G. Barto (1998). Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA

[41] Thomas, L. C. (2003). Games, Theory and Applications, Ellis Horwood Ltd, Dover Publications, NY

[42] Thrall Robert M., Lucas, William F. (1963), "n-person games in partition function form", Wiley Periodicals

[43] Tom M. Mitchell (1997), Machine learning, McGraw-Hill Science/Engineering/Math, London, UK

[44] Tom M. Mitchell (2006), The discipline of Machine learning, School of Computer Science, Carnegie Mellon University, Pittsburgh

[45] Vogel, Carol (April 29, 2005), "Rock, Paper, Payoff: Child's Play Wins Auction House an Art Sale", The New York Times

[46] Von Neumann, John; Morgenstern, Oskar (1944), Theory of Games and Economic Behavior, Princeton University Press, Princeton NJ

[47] Britannica Online Encyclopedia, http://www.britannica.com, Science & Technology, Game-theory, Classification of games, 28-12-2009

[48] Britannica Online Encyclopedia, http://www.britannica.com, Science & Technology, Game-theory, 28-12-2009

[49] Game Theory net http://www.gametheory.net/dictionary, 6-1-2010

[50] Game theory http://members.cox.net/mathmistakes/game_theory.htm, 27-12-2009

[51] Stanford Encyclopedia of Philosophy,http://plato.stanford.edu/entries/prisoner-dilemma/, 14/11/2009

# Παράρτημα

I. Λεξικό όρων

II. Περιγραφή της εφαρμογής

# I. Λεξικό όρων

# I. Λεξικό όρων

**Bayesian game (Μπεϊζιανό παίγνιο).** Στην θεωρία παιγνίων, μπεϊζιανό παίγνιο είναι αυτό στο οποίο οι πληροφορίες για τα χαρακτηριστικά των άλλων παικτών (π.χ. αποδώσεις) δεν είναι πλήρης. Ακολουθώντας το πλαίσιο του John C. Harsanyi, ένα μπεϊζιανό παίγνιο μπορεί να μοντελοποιηθεί εισάγοντας την φύση σαν ένα παίχτη στο παίγνιο. Η φύση αναθέτει μια τυχαία μεταβλητή σε κάθε παίκτη η οποία μπορεί να πάρει τιμές τύπου κάθε παίκτη και συσχετιζόμενων πιθανοτήτων ή μιας συνάρτησης πυκνότητας πιθανότητας αυτών των τύπων (στην πορεία του παιγνίου η φύση διαλέγει τυχαία έναν τύπο για κάθε παίκτη σύμφωνα με την κατανομή πιθανοτήτων στον χώρο τύπων κάθε παίκτη). Η προσέγγιση του Harsanyi στην μοντελοποίηση ενός μπεϊζιανού παιγνίου με αυτόν τον τρόπο επιτρέπει στα παίγνια μη πλήρης πληροφόρησης να γίνουν παίγνια ατελούς πληροφόρησης (στα οποία το ιστορικό του παιγνίου δεν είναι διαθέσιμο στους παίκτες). Ο τύπος ενός παίκτη καθορίζει την συνάρτηση απόδοσης του παίκτη αυτού και η συσχετισμένη πιθανότητα με τον τύπο είναι η πιθανότητα ότι ο παίκτης για τον οποίο ο τύπος καθορίζεται είναι αυτού του τύπου. Σε ένα μπεϊζιανό παίγνιο το να μην είναι πλήρης πληροφόρησης σημαίνει ότι τουλάχιστον ένας παίκτης δεν είναι σίγουρος για τον τύπο, οπότε και για την συνάρτηση απόδοσης, ενός άλλου παίκτη.

Τέτοιους είδους παίγνια αποκαλούνται μπεϊζιανά λόγου της πιθανοτικής ανάλυσης που συνδέεται με αυτά. Οι παίκτες έχουν αρχικές πεποιθήσεις για τον τύπο του κάθε παίκτη (όπου πεποίθηση είναι μια κατανομή πιθανοτήτων των πιθανών τύπων για έναν παίκτη) και μπορούν να αναθεωρήσουν τις πεποιθήσεις τους σύμφωνα με τον κανόνα του Bayes καθώς το παιχνίδι εξελίσσεται. Για παράδειγμα η πεποίθηση που έχει ένας παίκτης για τον τύπο ενός άλλου παίκτη μπορεί να αλλάξει με βάση τις κινήσεις που έχουνε παίξει. Η έλλειψη πληροφόρησης των παικτών και η μοντελοποίηση των πεποιθήσεων σημαίνει ότι τέτοιου είδους παίγνια χρησιμοποιούνται επίσης για να αναλύσουν και σενάρια ατελούς πληροφόρησης.

**Common knowledge (κοινή γνώση).** Η κοινή γνώση είναι ειδικό είδος γνώσης ενός συνόλου πρακτόρων. Υπάρχει κοινή γνώση του p σε ένα σύνολο πρακτόρων G όταν όλοι οι πράκτορες στο G γνωρίζουν το p, όλοι ξέρουν ότι γνωρίζουν το p, όλοι ξέρουν ότι ξέρουν ότι γνωρίζουν το p, κτλ συνεχιζόμενο στο άπειρο.

Η έννοια εμφανίστηκε για πρώτη φορά στην φιλοσοφική λογοτεχνία από τον David Kellogg Lewis στην μελέτη του Convention (1969). Της δόθηκε για πρώτη φορά μαθηματική μορφή σε ένα πλαίσιο της θεωρίας συνόλων από τον Robert Aumann (1976). Το ενδιαφέρον των επιστημόνων των υπολογιστών για την γνωσιοθεωρητική λογική (epistemic logic) μεγάλωσε γενικά, αλλά και συγκεκριμένα για την κοινή γνώση, στις αρχές της δεκαετίας του 80. Από τότε υπάρχουν διάφορα προβλήματα τα οποία βασίζονται στην έννοια αυτή και έχουν μελετηθεί από τους μαθηματικούς.

**Complete information (πλήρης πληροφόρηση).** Η πλήρης πληροφόρηση είναι μια έννοια που χρησιμοποιείται στα οικονομικά και στην θεωρία παιγνίων για να περιγράψει μια οικονομική κατάσταση ή ένα παίγνιο στο οποίο η γνώση για τους άλλους συμμετέχοντες στην αγορά ή παίκτες είναι διαθέσιμη σε όλους τους συμμετέχοντες. Κάθε παίκτης γνωρίζει τις αποδόσεις και τις στρατηγικές διαθέσιμες στους άλλους παίκτες.

Η πλήρης πληροφόρηση είναι ένα από τα θεωρητικά προαπαιτούμενα μιας αποτελεσματικής απόλυτα ανταγωνιστικής αγοράς. Υπό μία έννοια είναι μια απαίτηση της υπόθεσης που έχει γίνει επίσης στην οικονομική θεωρία ότι οι συμμετέχοντας σε μια αγορά δρουν λογικά. Εάν ένα παίγνιο δεν είναι πλήρους πληροφόρησης, τότε ο κάθε παίκτης ξεχωριστά δεν θα ήταν δυνατό να προβλέψει το αποτέλεσμα της πράξης του στους άλλους παίκτες, ακόμα και αν ο δρώντας υπέθεται ότι οι άλλοι παίκτες πράττουν λογικά.

**Core (η θεωρία του πυρήνα).** Ο πυρήνας είναι ένα σύνολο εφικτών κατανομών που δεν είναι δυνατόν να βελτιωθούν στην πορεία από ένα υποσύνολο, έναν συνασπισμό, των καταναλωτών της οικονομίας. Ένας συνασπισμός λέγεται ότι βελτιώνεται ή εμποδίζει μια εφικτή κατανομή εάν τα μέλη αυτού του συνασπισμού βρίσκονται σε καλύτερη θέση εφαρμόζοντας μια άλλη εφικτή κατανομή η οποία είναι πανομοιότυπη της πρώτης εκτός από το ότι κάθε μέλος του συνασπισμού έχει διαφορετική δέσμη κατανάλωσης η οποία είναι μέρος μιας συνολικής δέσμης κατανάλωσης που μπορεί να κατασκευαστεί από την κοινά διαθέσιμη τεχνολογία και τις αρχικές προικοδοτήσεις του κάθε καταναλωτή στον συνασπισμό.

Μια κατανομή λέγεται ότι έχει την ιδιότητα του πυρήνα όταν εάν υπάρχει ένας συνασπισμός που μπορεί να βελτιώσει. Ο πυρήνας είναι το σύνολο των εφικτών κατανομών με την ιδιότητα του πυρήνα.

**Correlated equilibrium (συσχετισμένη ισορροπία).** Στην θεωρία παιγνίων η συσχετισμένη ισορροπία είναι μια έννοια λύσης που είναι πιο γενική από την ευρέως γνωστή ισορροπία Nash. Πρωτοαναφέρθηκε από τον μαθηματικό Robert Aumann (1974). Η ιδέα είναι ότι κάθε παίκτης διαλέγει την δράση του σύμφωνα με την παρατήρησή του για την αξία του ίδιου κοινού σήματος. Μια στρατηγική αντιστοιχεί μια ενέργεια σε κάθε πιθανή παρατήρηση που μπορεί ένας παίκτης να κάνει. Εάν κανένας παίκτης δεν θέλει να αποκλίνει από την προτεινόμενη στρατηγική, υποθέτοντας ότι οι υπόλοιποι δεν θα αποκλίνουν, η κατανομή ονομάζεται συσχετισμένη ισορροπία.

**Evolutionary game theory (εξελικτική θεωρία παιγνίων, ΕΘΠ) .** Η ΕΘΠ είναι η εφαρμογή της θεωρίας παιγνίων στην εξαρτώμενη από την αλληλεπίδραση στρατηγική εξέλιξης των πληθυσμών. Η ΕΘΠ είναι χρήσιμη σε ένα βιολογικό πλαίσιο ορίζοντας ένα πλαίσιο στρατηγικών στον οποίο προσαρμοστικά χαρακτηριστικά μπορούν να διαμορφωθούν. Προέκυψε το 1973 με την επισημοποίηση της εξελικτικά σταθερής στρατηγικής των John Maynard Smith και τον George R. Price σαν εφαρμογή της μαθηματικής θεωρίας των παιγνίων σε βιολογικά περιεχόμενο, που προέκυψε από την συνειδητοποίηση ότι η εξαρτώμενη από την συχνότητα προσαρμοστικότητα εισάγει μια στρατηγική πτυχή στην εξέλιξη . Η ΕΘΠ διαφέρει από την κλασική θεωρία παιγνίων εστιάζοντας στην δυναμική της αλλαγής στρατηγικής περισσότερο από τις ιδιότητες της ισορροπίας των στρατηγικών. Παρά το όνομά της, η εξελικτική θεωρία παιγνίων έχει γίνει αυξημένου ενδιαφέροντος για τους οικονομολόγους, κοινωνιολόγους, ανθρωπολόγους και φιλοσόφους.

**Evolutionarily stable strategy (εξελικτικά σταθερή στρατηγική) .** Στην θεωρία παιγνίων και στην οικολογία της συμπεριφοράς, η εξελικτικά σταθερή στρατηγική (ΕΣΣ) είναι μια στρατηγική η οποία, ένα υιοθετηθεί από έναν πληθυσμό παικτών, δεν μπορεί να προσβληθεί από οποιαδήποτε εναλλακτική στρατηγική που είναι αρχικά σπάνια. Μια ΕΣΣ είναι μια βελτίωση της ισορροπίας Nash, είναι μια ισορροπία Nash η οποία είναι εξελικτικά σταθερή υπό την έννοια ότι όταν καθοριστεί σε έναν πληθυσμό η φυσική επιλογή και μόνο αρκεί για να αποτρέψει εναλλακτικές, μεταλλαγμένες, στρατηγικές από το να εισβάλουν.

Η ΕΣΣ αναπτύχθηκε προκειμένου να καθοριστεί μια κατηγορία λύσεων για παιγνιοθεωρητικά προβλήματα, όπως και η ισορροπία Nash, αλλά που θα μπορούσε να εφαρμοστεί στην εξέλιξη της κοινωνικής συμπεριφοράς των ζώων. Η ισορροπία Nash μπορεί μερικές φορές να υπάρχει λόγω της εφαρμογής της ορθολογικής πρόβλεψης, η οποία θα ήταν όμως ανάρμοστη σε ένα εξελικτικό πλαίσιο. Τελολογικές δυνάμεις όπως η ορθολογική πρόβλεψη δεν μπορούν να εξηγήσουν τα αποτελέσματα διαδικασιών δοκιμής και λάθους, όπως είναι η εξέλιξη, και ως εκ τούτο δεν έχουν θέση στις βιολογικές εφαρμογές. Ο ορισμός μιας ΕΣΣ αποκλείει τέτοιου είδους ισορροπίες Nash.

Αρχικά αναπτύχθηκε το 1973 και από τότε η ΕΣΣ έχει χρησιμοποιηθεί ευρέως στην οικολογία της συμπεριφοράς και στα οικονομικά όπως επίσης χρησιμοποιείται στην ανθρωπολογία, στην εξελικτική ψυχολογία, την φιλοσοφία και τις πολιτικές επιστήμες.

**Fictitious play (εικονικό παιχνίδι)**. Στην θεωρία παιγνίων το εικονικό παιχνίδι είναι ένας κανόνας μάθησης που αναφέρθηκε για πρώτη φορά από τον G.W. Brown (1951). Σε αυτό κάθε παίκτης εικάζει ότι ο αντίπαλός του παίζει σταθερές, πιθανότατα μικτές, στρατηγικές. Έτσι, σε κάθε γύρο κάθε παίκτης ανταποκρίνεται με την βέλτιστη απάντηση στην εμπειρική συχνότητα του παιχνιδιού του αντιπάλου του. Μια τέτοια μέθοδος είναι βεβαίως επαρκής εάν ο αντίπαλος όντως χρησιμοποιεί μια σταθερή στρατηγική, καθώς είναι ελαττωματική εάν η στρατηγική του αντιπάλου δεν είναι σταθερή. Η στρατηγική του αντιπάλου μπορεί για παράδειγμα να εξαρτάται από την τελευταία κίνηση του παίκτη που χρησιμοποιεί το εικονικό παιχνίδι.

**Markov decision processes (MDP) (διαδικασίες αποφάσεων Markov, ΔΑΜ).** Πήραν το όνομά τους από τον Andrey Markov και παρέχουν ένα μαθηματικό πλαίσιο για την μοντελοποίηση καταστάσεων λήψης απόφασης όπου τα αποτελέσματα είναι εν μέρει τυχαία και εν μέρει υπό τον έλεγχο του λήπτη αποφάσεων. Οι ΔΑΜ είναι χρήσιμες για την μελέτη ενός ευρέος φάσματος προβλημάτων βελτιστοποίησης, επιλυόμενων με την χρήση του δυναμικού προγραμματισμού και της ενισχυμένης μάθησης. Οι ΔΑΜ ήταν γνωστές τουλάχιστον ήδη από την δεκαετία του 50. Μεγάλο μέρος της έρευνας στον τομέα αυτό ξεκίνησε λόγο του βιβλίου του Ronald A. Howard «Δυναμικός προγραμματισμός και διαδικασίες Markov» το 1960. Σήμερα χρησιμοποιούνται σε διάφορους τομείς, όπως η ρομποτική, ο αυτόματος έλεγχος, τα οικονομικά και η κατασκευή.

Πιο συγκεκριμένα, μια ΔΑΜ είναι μια στοχαστική διαδικασία ελέγχου διακριτού χρόνου. Σε κάθε χρονικό βήμα η διαδικασία βρίσκεται σε κάποια κατάσταση s και ο λήπτης αποφάσεων μπορεί να επιλέξει οποιαδήποτε δράση a που είναι διαθέσιμη στην κατάσταση s. Η διαδικασία ανταποκρίνεται στο επόμενο χρονικό βήμα με τυχαία μετακίνηση σε μια νέα κατάσταση s’ και δίνοντας στον ιθύνοντα μια αντίστοιχή ανταμοιβή $R_a(s,s')$.

Η πιθανότητα ότι η διαδικασία θα επιλέξει την s’ σαν την νέα κατάσταση της επηρεάζεται από την επιλεγμένη δράση. Συγκεκριμένα, δίνεται από την συνάρτηση μετάβασης κατάστασης $P_a(s,s')$. Έτσι, η επόμενη κατάσταση s' εξαρτάται από την τρέχουσα κατάσταση s και την δράση a του λήπτη αποφάσεων. Αλλά, με δεδομένα το s και a, είναι υπό όρους ανεξάρτητη όλων των προηγούμενων καταστάσεων και δράσεων, με άλλα λόγια οι μεταβάσεις κατάστασης μιας ΔΑΜ διαθέτουν την ιδιότητα Markov.

Οι ΔΑΜ είναι μια επέκταση των αλυσίδων Markov, η διαφορά είναι η προσθήκη δράσεων, επιτρέποντας την επιλογή, και της επιβράβευσης, δίνοντας κίνητρο. Αν υπήρχε μόνο μια δράση, ή αν η χρησιμοποιούμενη δράση ήταν προκαθορισμένη σε κάθε κατάσταση, μια διαδικασία απόφασης Markov μπορεί να θεωρηθεί σαν αλυσίδα Markov.

**Mechanism design (μηχανική σχεδίαση).** Η μηχανική σχεδίαση, μερικές φορές αποκαλείται και αντίστροφη θεωρία παιγνίων, είναι ένα πεδίο της θεωρίας παιγνίων που μελετάει της έννοιες λύσεων για μια κατηγορία παιγνίων ιδιωτικών πληροφορίων. Τα διακριτά χαρακτηριστικά των παιγνίων αυτών είναι:

- Ότι ο σχεδιαστής του παιγνίου επιλέγει την δομή του παιγνίου αντί να την κληρονομεί
- Ότι ο σχεδιαστής ενδιαφέρεται για το αποτέλεσμα του παιγνίου

Ένα τέτοιο παιχνίδι αποκαλείται «παίγνιο μηχανικού σχεδιασμού» και συνήθως λύνεται με την παροχή κινήτρων στους πράκτορες να αποκαλύψουν τις ιδιωτικές τους πληροφορίες. Το βραβείο νόμπελ στις οικονομικές επιστήμες το 2007 απονεμήθηκε στους Leonid Hurwicz, Eric Maskin, and Roger Myerson «για την εισαγωγή των θεμελίων της θεωρίας της μηχανικής σχεδίασης».

**Monte Carlo methods (μέθοδοι Monte Carlo).** Οι μέθοδοι Monte Carlo είναι μια κατηγορία υπολογιστικών αλγορίθμων που βασίζονται στην επανάληψη τυχαίων δειγματοληψιών για τον υπολογισμό των αποτελεσμάτων τους. Συχνά χρησιμοποιούνται στην προσομοίωση φυσικών και μαθηματικών συστημάτων. Λόγου της εξάρτησης τους από επαναλαμβανόμενους υπολογιστικούς τυχαίους ή ψευδό-τυχαίους αριθμούς αυτές οι μέθοδοι είναι καταλληλότερες για τον υπολογισμό με την χρήση υπολογιστή και συνήθως χρησιμοποιούνται όταν είναι ανέφικτο να υπολογιστεί ένα ακριβές αποτέλεσμα με ένα ντετερμινιστικό αλγόριθμο.

Οι μέθοδοι εξομοίωσης Monte Carlo είναι ιδιαίτερα χρήσιμες στην μελέτη συστημάτων με μεγάλο αριθμό συνδεδεμένων βαθμών ελευθερίας, όπως είναι τα τα υγρά, διαταραγμένα υλικά, συνδεδεμένα στερεά και κυτταρικές δομές. Γενικότερα οι μέθοδοι Monte Carlo είναι χρήσιμες για την μοντελοποίηση φαινομένων με σημαντική αβεβαιότητα εισροών, όπως είναι ο υπολογισμός του ρίσκου στις επιχειρήσεις. Αυτές οι μέθοδοι χρησιμοποιούνται ευρέως και από του μαθηματικούς , ένα κλασικό παράδειγμα της χρήσης τους είναι για την αξιολόγηση οριστικών ολοκληρωμάτων , ιδίως πολυδιάστατων ολοκληρωμάτων με πολύπλοκες οριακές συνθήκες. Πρόκειται για μια πολύ επιτυχημένη μέθοδο ανάλυσης κινδύνου σε σύγκριση με εναλλακτικές μεθόδους ή την ανθρώπινη διαίσθηση. Όταν οι προσομοιώσεις Monte Carlo εφαρμόστηκαν στην εξερεύνηση του διαστήματος και της έρευνας για πετρέλαιο, οι πραγματικές παρατηρήσεις αποτυχιών, υπέρβασης κόστους και χρονοδιαγραμμάτων ήταν συνήθως καλύτερα προβλεπόμενες από την προσομοίωση από ότι με την ανθρώπινη διαίσθηση ή εναλλακτικά «ήπιες» μεθόδους.

Ο όρος «μέθοδος Monte Carlo» επινοήθηκε το 1940 από φυσικούς που εργάζονταν σε πυρηνικά προγράμματα όπλων στο Los Alamos National Laboratory.

**Multi-armed bandit problem (προβλήμα του πολύ-χερου ληστή).** Ο πολύχερος ληστής, αποκαλούμενος και ν-χερος ληστής, είναι ένα απλό πρόβλημα μηχανικής μάθησης βασισμένο στην αναλογία ενός παραδοσιακού κουλοχέρη, μονόχερος ληστής, αλλά με περισσότερο από έναν μοχλούς. Όταν χρησιμοποιείται κάθε μοχλός δίνει μια αμοιβή σύμφωνα με μια κατανομή συσχετισμένη με τον μοχλό αυτό. Ο σκοπός του τζογαδόρου είναι να μεγιστοποιήσει το σύνολο των αμοιβών του μέσα από επαναλαμβανόμενες χρήσης των μοχλών. Θεωρείται κλασικά δεδομένο ότι ο τζογαδόρος δεν έχει αρχική γνώση σχετικά με τους μοχλούς. Η κρίσιμη εξισορρόπηση που πρέπει να κάνει ο χρήστης είναι ανάμεσα στην «εκμετάλλευση» του μοχλού που δίνει την μεγαλύτερη προσδοκώμενη ανταμοιβή και της «εξερεύνησης» για να μάθει περισσότερες πληροφορίες για τις προσδοκώμενες ωφέλειες των άλλων μοχλών.

**Proper equilibrium (γνήσια ισορροπία)** . Είναι μια βελτίωση της ισορροπίας Nash που αποδίδεται στον Roger B. Myerson. Η γνήσια ισορροπία βελτιώνει περεταίρω την έννοια της τέλειας ισορροπίας τρεμάμενου χεριού με την παραδοχή ότι πιο δαπανηρά «τρέμουλα» γίνονται με σημαντικά μικρότερες πιθανότητες από ότι αυτά με χαμηλότερο κόστος.

**Shapley value (τιμή Shapley).** Στην θεωρία παιγνίων η τιμή Shapley, που πήρε το όνομά της προς τιμήν του Lloyd Shapley που την εισήγαγε το 1953, περιγράφει μια προσέγγιση για την δίκαιη κατανομή των κερδών που λαμβάνονται με την συνεργασία μεταξύ διαφόρων παραγόντων.

Το σενάριο έχει ως εξής: ένας συνασπισμός παραγόντων συνεργάζεται και αποκτά ένα ορισμένο συνολικό όφελος από αυτήν την συνεργασία. Δεδομένου ότι ορισμένοι παράγοντες μπορεί να συμβάλλουν περισσότερο στον συνασπισμό από ότι κάποια άλλοι, η ερώτηση που προκύπτει είναι πώς να μοιραστούν δίκαια τα κέρδη ανάμεσα στους παράγοντες, ή «πόσο σημαντικός είναι κάθε παράγοντας στο συνολικό εγχείρημα και μια ανταμοιβή θα πρέπει λογικά να περιμένουν».

**Stochastic game (στοχαστικό παίγνιο)** . Στην θεωρία παιγνίων ένα στοχαστικό παίγνιο, η έννοια του οποίου αναπτύχτηκε από τον Lloyd Shapley στις αρχές του 50, είναι ένα δυναμικό παίγνιο με πιθανοτικές μεταβάσεις που παίζεται από έναν η περισσότερους παίκτες. Το παίγνιο παίζεται σε μία σειρά από στάδια. Στην αρχή κάθε γύρου το παιχνίδι βρίσκεται σε κάποια κατάσταση. Οι παίκτες διαλέγουν ενέργειες και κάθε παίκτης λαμβάνει μια απόδοση που εξαρτάται από την κατάσταση στην οποία βρίσκεται το παίγνιο και τις επιλεγμένες ενέργειες. Το παίγνιο τότε προχωράει σε μια νέα τυχαία κατάσταση της οποίας οι κατανομές εξαρτώνται από την προηγούμενη κατάσταση και τις ενέργειες που επιλέχθηκαν από τους παίκτες. Η διαδικασία επαναλαμβάνεται για την καινούργια κατάσταση και το παιχνίδι συνεχίζεται για ένα πεπερασμένο ή μη αριθμό σταδίων-γύρων. Η συνολική ανταμοιβή για έναν παίκτη θεωρείται συχνά να είναι το προεξοφλημένο άθροισμα των αποδόσεων του γύρου ή το κάτω όριο του μέσου όρου των αμοιβών του γύρου.

**Subgame perfect equilibrium (υποπαιγνιακά τέλεια ισορροπία).** Στην θεωρία παιγνίων η υποπαιγνιακά τέλεια ισορροπία, ή υποπαιγνιακά τέλεια ισορροπία Nash, είναι μια βελτίωση της ισορροπίας Nash σε δυναμικά παίγνια. Ένα στρατηγικό προφίλ αποτελεί υποπαιγνιακά τέλεια ισορροπία εάν αναπαριστά μια ισορροπία Nash κάθε υποπαιγνίου του αυθεντικού παιχνιδιού. Αυτό σημαίνει, εάν πρώτον οι παίκτες έπαιξαν ένα μικρότερο παίγνιο που αποτελούνταν από ένα κομμάτι μόνο του μεγαλύτερου παιγνίου και δεύτερον η συμπεριφορά τους αναπαριστά μια ισορροπία Nash αυτού του μικρότερου παιγνίου, τότε η συμπεριφορά τους είναι υποπαιγνιακά τέλεια ισορροπία του μεγαλύτερου παιγνίου.

Μια κοινή μέθοδος για τον εντοπισμό υποπαιγνιακών τέλειων ισορροπιών στην περίπτωση ενός πεπερασμένου παιγνίου είναι η επαγωγή προς τα πίσω. Σε αυτήν οι αποφάσεις παίρνονται με βάση τις τελευταίες ενέργειες του παιχνιδιού και σύμφωνα με αυτές αποφασίζεται ποιες ενέργειες ο τελευταίος παίκτης θα πρέπει να επιλέξει σε κάθε πιθανή κατάσταση για να μεγιστοποιήσει την ωφέλειά του. Τότε υποθέτεται ότι ο τελευταίος δρώντας θα κάνει αυτές τις ενέργειες και λαμβάνονται υπόψη οι προτελευταίες ενέργειες, διαλέγοντας για άλλη μια φορά της ενέργειες που μεγιστοποιούν την ωφέλεια του δρώντα. Η διαδικασία συνεχίζεται μέχρι την πρώτη κίνηση του παιγνίου. Οι στρατηγικές που παραμένουν αποτελούν όλες υποπαιγνιακά τέλειες ισορροπίες. Παρόλα αυτά η επαγωγή προς τα πίσω δεν μπορεί να εφαρμοστεί σε παιχνίδια ατελούς ή ελλιπής πληροφόρησης διότι αυτό συνεπάγεται ότι θα εμπλέκονταν μη μονήρεις σύνολα πληροφοριών. Η επαγωγή προς τα πίσω επίσης απαιτεί να υπάρχει πεπερασμένος αριθμός κινήσεων, δεν μπορεί δηλαδή να εφαρμοστεί σε παίγνια άπειρης διάρκειας.

Το τελεσιγραφικό παίγνιο είναι ένα χαρακτηριστικό παράδειγμα παιγνίου με λιγότερες υποπαιγνιακά τέλειες ισορροπίες από ότι ισορροπίες Nash.

**Temporal difference (TD) learning** (**μάθηση χρονικής διαφοράς**) . Η μάθηση χρονικής διαφοράς είναι μια μέθοδος πρόβλεψης. Έχει χρησιμοποιηθεί κυρίως για να λύσει το πρόβλημα της ενισχυμένης μάθησης. Είναι ένας συνδυασμός των ιδεών Monte Carlo και του δυναμικού προγραμματισμού. Η μάθηση χρονικής διαφοράς μοιάζει με μέθοδο Monte Carlo λόγω του ότι μαθαίνει κάνοντας δειγματοληψία του περιβάλλοντος σύμφωνα με κάποια πολιτική. Σχετίζεται με τις τεχνικές δυναμικού προγραμματισμού γιατί υπολογίζει την τρέχουσα πρόβλεψη με βάση προηγούμενα μαθημένες προβλέψεις, μια διαδικασία γνωστή και ως bootstrapping. Ο αλγόριθμος μάθησης χρονικής διαφοράς σχετίζεται με το μοντέλο χρονικής διαφοράς της μάθησης των ζώων.

Ως μέθοδος πρόβλεψης η μάθηση χρονικής διαφοράς λαμβάνει υπόψη το γεγονός ότι οι επόμενες προβλέψεις είναι συχνά συσχετισμένες υπό κάποια έννοια. Στην κλασική εποπτευόμενη μάθηση μαθαίνει κανείς μόνο από τις πράγματι παρατηρούμενες τιμές: γίνεται μια πρόβλεψη, και όταν η παρατήρηση είναι διαθέσιμη, η πρόβλεψη προσαρμόζεται να ταιριάζει καλύτερα στην παρατήρηση. Η κεντρική ιδέα της μάθησης χρονικής διαφοράς είναι ότι οι προβλέψεις προσαρμόζονται να ταιριάζουν άλλες, πιο ακριβείς, προβλέψεις για το χαρακτηριστικό. Η διαδικασία είναι μια μορφής bootstrapping, όπως φαίνεται από το ακόλουθο παράδειγμα:

> .Ας υποθέσουμε ότι θέλουμε να προβλέψουμε τον καιρό για το Σάββατο, και έχουμε κάποιο μοντέλο που προβλέπει τον καιρό του Σαββάτου, δεδομένου του καιρού κάθε ημέρας της εβδομάδας. Στην τυπική περίπτωση θα περιμέναμε μέχρι το Σάββατο και θα προσαρμόζαμε τότε όλα τα μοντέλα. Ωστόσο όταν για παράδειγμα είναι Παρασκευή θα έχουμε μια πολύ καλή ιδέα του τι καιρός θα επικρατεί το Σάββατο, και έτσι θα μπορούμε να αλλάξουμε, για παράδειγμα, το μοντέλο της Δευτέρας πριν φτάσει το Σάββατο.

Μαθηματικά μιλώντας, τόσο σε μια κλασική όσο και σε μάθηση χρονικής διαφοράς προσέγγιση, θα γινόταν προσπάθεια να βελτιστοποιηθεί κάποια συνάρτηση κόστους, σχετιζόμενη με το λάθος στις προβλέψεις της προσδοκίας κάποιας τυχαίας μεταβλητής E[z]. Ωστόσο ενώ στην κλασική προσέγγιση κάποιες φορές υποθέτεται ότι E[z] = z, η πραγματικά παρατηρούμενη τιμή, στην προσέγγιση της μάθησης χρονικής διαφοράς χρησιμοποιείται ένα μοντέλο. Για την συγκεκριμένη περίπτωση της ενισχυμένης μάθησης, η οποία είναι η κύρια εφαρμογή των μεθόδων μάθησης χρονικής διαφοράς, το z είναι η συνολική επιστροφή και το E[z] δίνεται από την εξίσωση Bellman της επιστροφής.

**Transferable utility (μεταφέρσιμη ωφέλεια).** Είναι ένας όρος που χρησιμοποιείται στην συνεργατική θεωρία παιγνίων και στα οικονομικά. Η ωφέλεια είναι μεταφέρσιμη εάν ένας παίκτης μπορεί να μεταβιβάσουν μέρος της ωφέλειας του χωρίς απώλειες σε έναν άλλο παίκτη. Τέτοιου είδους μεταβιβάσεις είναι δυνατές εάν οι παίκτες χρησιμοποιούν κοινό «νόμισμα» που αποτιμάται ίσα από όλους. Σημειώνεται ότι το να είναι δυνατό να μεταβιβαστούν απολαβές σε μετρητά δεν σημαίνει ότι η ωφέλεια είναι μεταφέρσιμη: πλούσιοι και φτωχοί παίκτες μπορούν να αποκομίσουν διαφορετική ωφέλεια από το ίδιο χρηματικό ποσό.

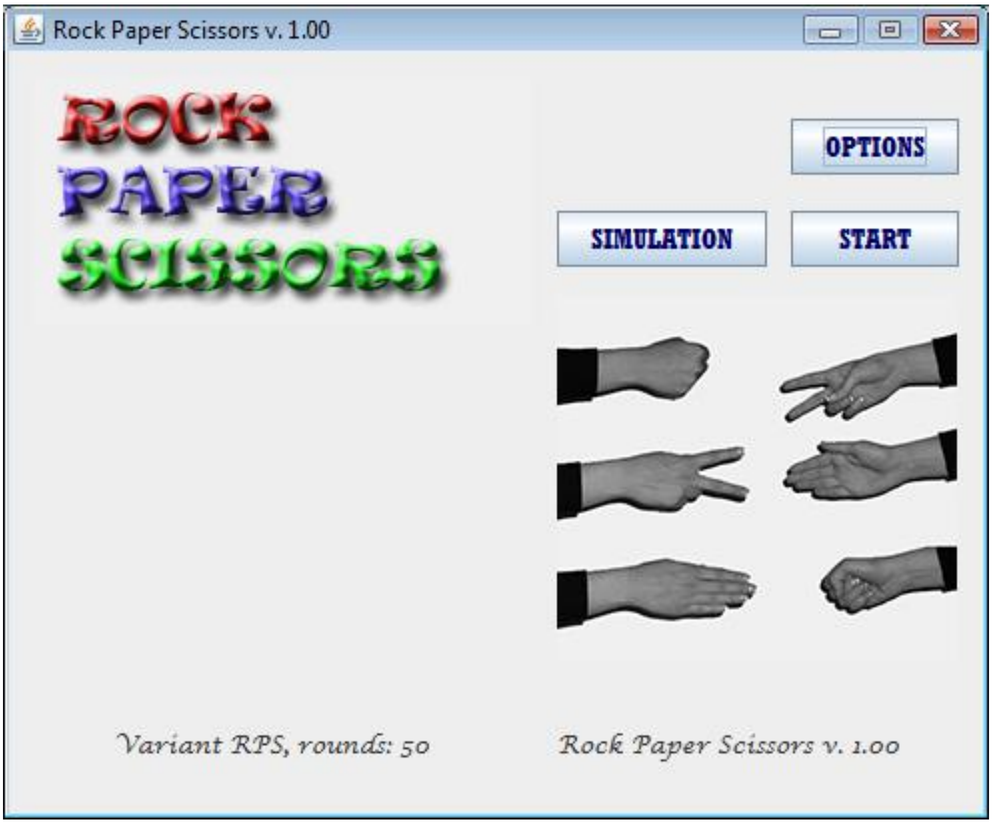
Η μεταφέρσιμη ωφέλεια θεωρείται σε πολλά παίγνια συνεργασίας, όπου οι απολαβές δεν δίνονται σε ξεχωριστούς παίκτες αλλά σε συνασπισμούς. Σε αυτήν την περίπτωση η υπόθεση συνεπάγεται ότι ανεξάρτητα από την διαίρεση των απολαβών στον συνασπισμό, τα μέλη του συνασπισμού απολαμβάνουν την ίδια συνολική ωφέλεια.

**Trembling hand perfect equilibrium (τέλεια ισορροπία τρεμάμενου χεριού).** Είναι μια βελτίωση της ισορροπίας Nash που αποδίδεται στον Reinhard Selten. Μια τέλεια ισορροπία τρεμάμενου χεριού είναι μια ισορροπία που λαμβάνει υπόψη της την πιθανότητα εφαρμογής στρατηγικών διαφορετικών των ισορροπιών υποθέτοντας ότι οι παίκτες , με ένα γλίστρημα του χεριού ή «τρέμουλο» μπορει να επιλέξουν ακούσια στρατηγικές, αν και με αμελητέα πιθανότητα.

# II. Περιγραφή της εφαρμογής

# II.Περιγραφή της εφαρμογής

Η διεπαφή της εφαρμογής χωρίζεται σε δύο γενικές κατηγορίες μια για κάθε δυνατό είδος παιχνιδιού, το παιχνίδι χρήστη εναντίον πράκτορα και πράκτορα εναντίον πράκτορα. Ακολουθεί η εισαγωγική οθόνη της εφαρμογής

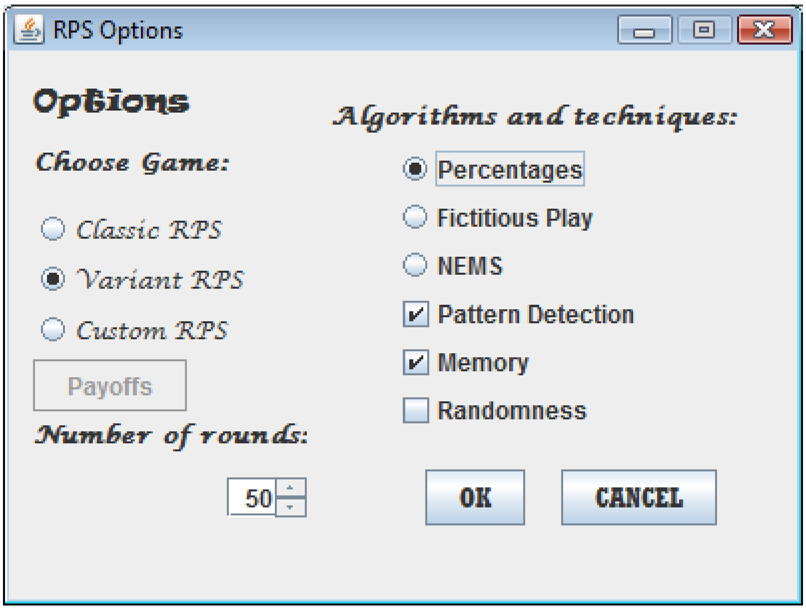



**Εικόνα II.2 Εισαγωγική οθόνη εφαρμογής.**

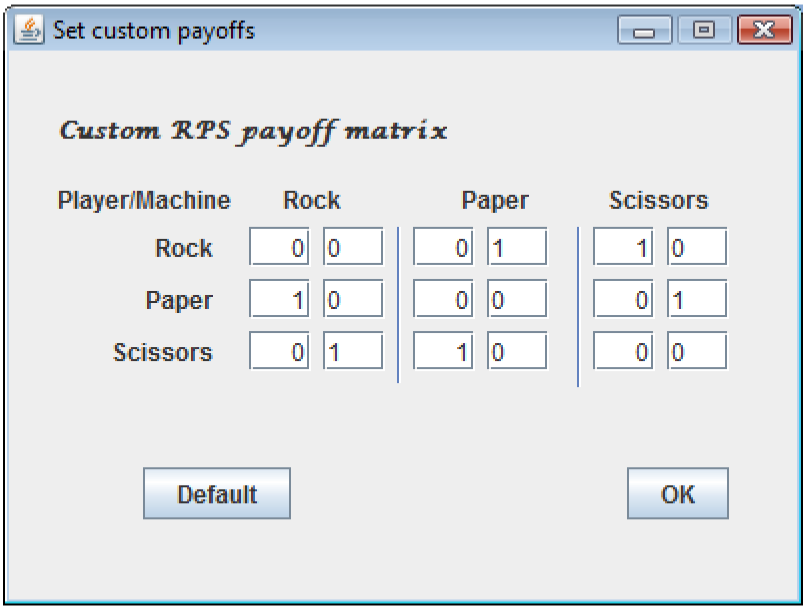
Εδώ δίνεται η δυνατότητα στον παίκτη να επιλέξει ποιος είδος παιχνιδιού θέλει. Με το κουμπί simulation διαλέγει να προχωρήσει στο παιχνίδι πράκτορα εναντίον πράκτορα και με το start να ξεκινήσει το παιχνίδι χρήστη εναντίον μηχανής. Οι ρυθμίσεις του δεύτερου είναι προσβάσιμες με το κουμπί options.

## II.I Χρήστης εναντίον πράκτορα

Στο μενού του κουμπιού options, που δήχθηκε και νωρίτερα, δίνεται στον χρήστη η δυνατότητα να επιλέξει την εκδοχή του παιχνιδιού που θέλει να παίξει, Classic, Variant και Custom, καθώς και να ρυθμίσει της αποδόσεις της προσαρμοσμένης εκδοχής (Custom), αν επιλέξει αυτή με το κουμπί Payoffs.




**Εικόνα II.3 Μενού επιλογών χρήστη εναντίον μηχανής**

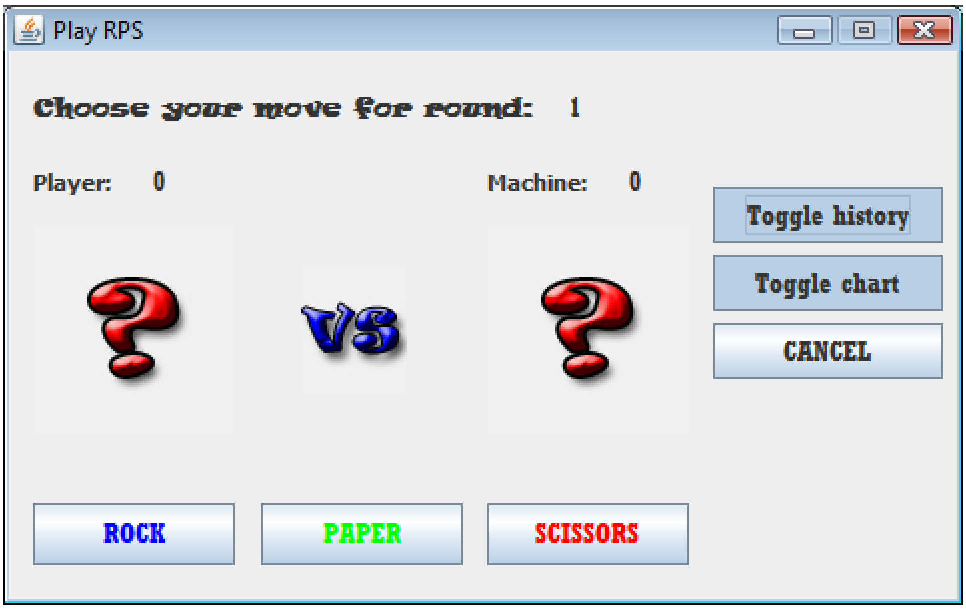



**Εικόνα II.4 Ρύθμιση αποδόσεων προσαρμοσμένης εκδοχής**

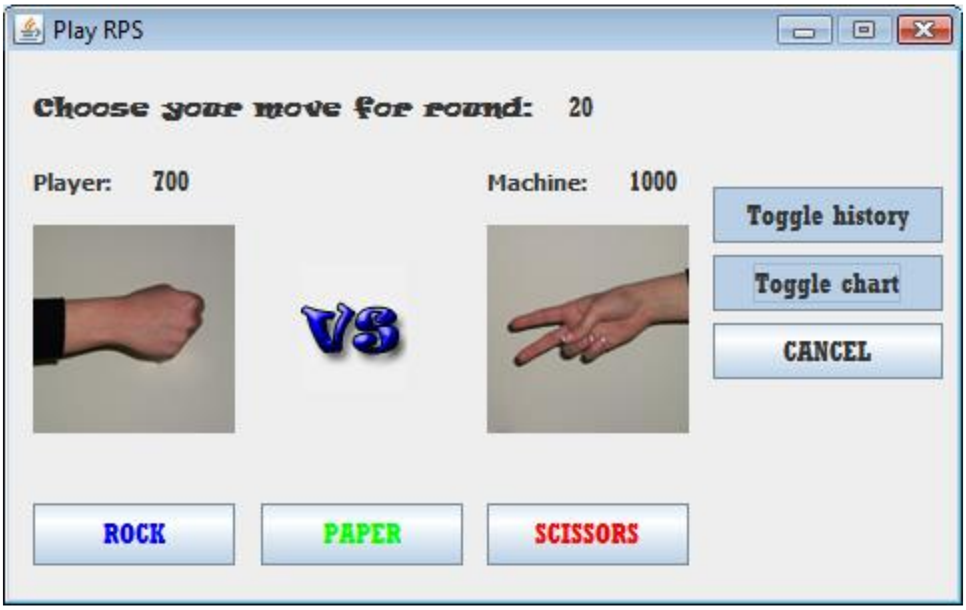
Από το μενού επιλογών μπορεί ο χρήστης να επιλέξει ποιον βασικό αλγόριθμο θέλει να χρησιμοποιεί ο πράκτορας (percentages, fictitious play, NEMS) και ποιους από τους αλγορίθμους/τεχνίκες επιθυμεί να είναι ενεργοί κατά την διάρκεια του παιχνιδιού (Pattern detection, memory, randomness). Στο σημείο αυτό από το πεδίο number of rounds μπορεί να επιλέξει της επαναλήψεις του παιγνίου σκηνή.

Πατώντας οκ δέχεται τις ρυθμίσεις και με cancel τις ακυρώνει. Επιστρέφοντας στην εισαγωγική οθόνη μπορεί να πατήσει το κουμπί start και να ξεκινήσει το παίγνιο με τις επιλεγμένες ρυθμίσεις. Πατώντας το start εμφανίζονται τρία παράθυρα, το παράθυρο στο οποίο ο παίκτης αλληλεπιδρά με την εφαρμογή και βλέπει το σκορ και δύο παράθυρα με στατιστικά στοιχεία, το ένα με την ακολουθία των κινήσεων των δύο αντιπάλων σε κάθε γύρο (ιστορικό) και το άλλο με την αντιστοιχία σε ποσοστά ή καταμέτρηση των συνολικών επιλεγμένων κινήσεων και από τους δύο παίκτες σε κάθε γύρο σε μορφή γραφήματος (γράφημα).

Στο πρώτο παράθυρο (εικόνα 5.4) ο παίκτης επιλέγει την κίνηση που θέλει να ακολουθήσει στον ανάλογο γύρο βλέποντας μια γραφική απεικόνιση των κινήσεων. Με τα κουμπιά toggle stats και toggle chart μπορεί να ανοιγοκλείνει το ιστορικό και το γράφημα αντίστοιχα.

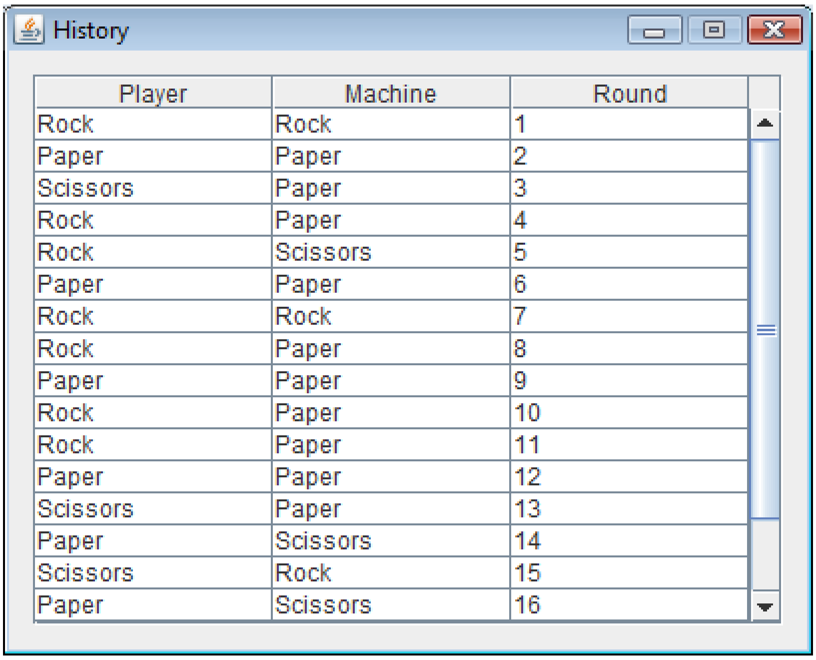



**Εικόνα II.5 Αρχικό παράθυρο παιχνιδιού.**




**Εικόνα ΙΙ.6 Παράθυρο παιχνιδιού σε εξέλιξη.**

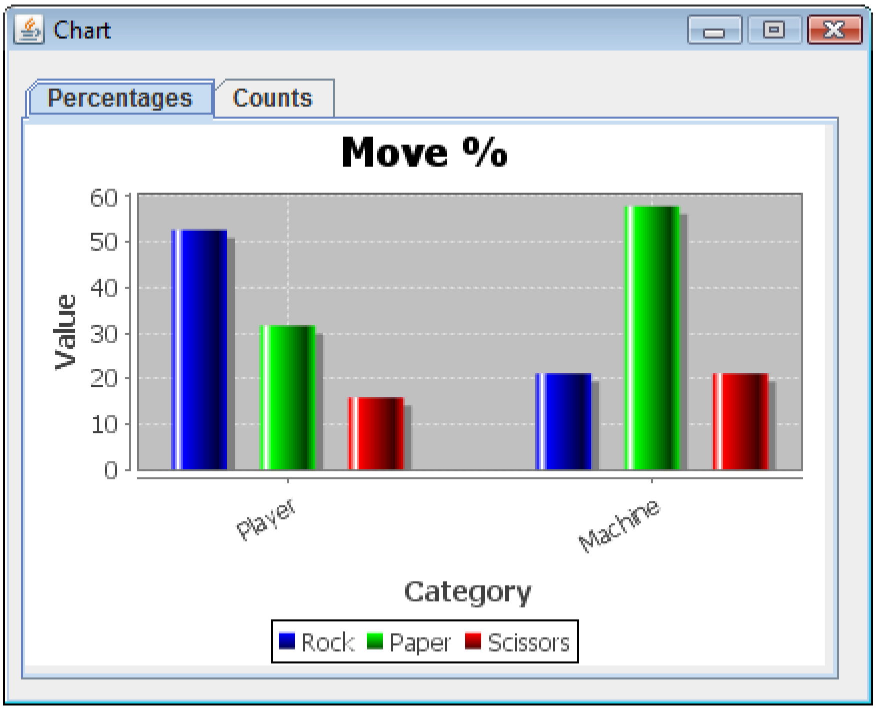
Στο παράθυρο του ιστορικού ο χρήστης μπορεί να δει την εξέλιξη του παιχνιδιού μέχρι τον γύρο στον οποίο βρίσκεται (εικόνα 5.6).


History

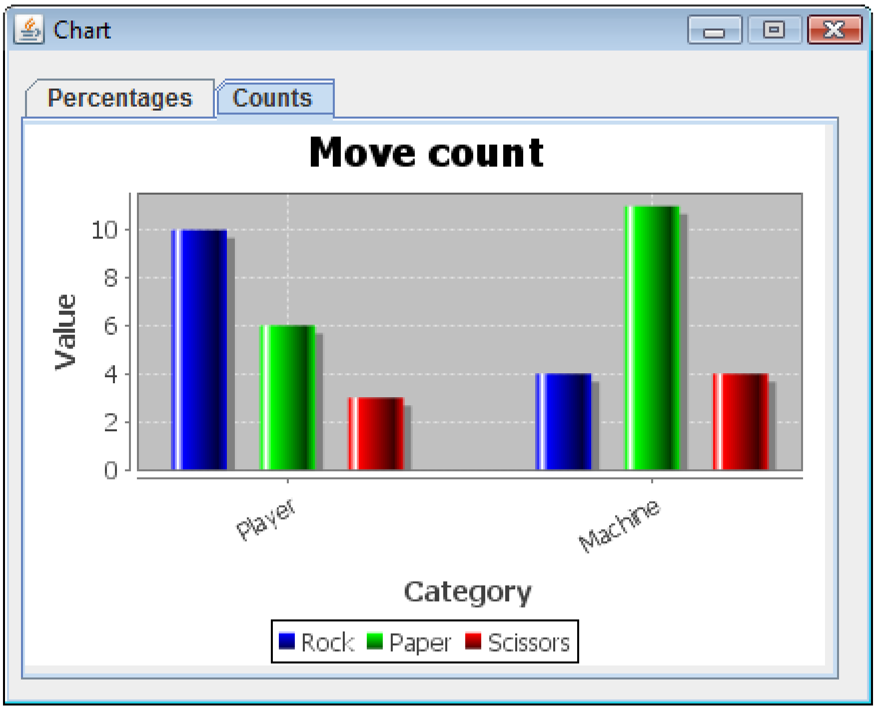
| Player | Machine | Round |
|---|---|---|
| Rock | Rock | 1 |
| Paper | Paper | 2 |
| Scissors | Paper | 3 |
| Rock | Paper | 4 |
| Rock | Scissors | 5 |
| Paper | Paper | 6 |
| Rock | Rock | 7 |
| Rock | Paper | 8 |
| Paper | Paper | 9 |
| Rock | Paper | 10 |
| Rock | Paper | 11 |
| Paper | Paper | 12 |
| Scissors | Paper | 13 |
| Paper | Scissors | 14 |
| Scissors | Rock | 15 |
| Paper | Scissors | 16 |

**Εικόνα ΙΙ.7 Ιστορικό.**

Στο παράθυρο του γραφήματος καταγράφεται η κατανομή των κινήσεων κάθε παίκτη, σε ποσοστό (εικόνα 5.7). ή καταμέτρηση (εικόνα 5.8) ανάλογα με την επιλεγμένη καρτέλα, στον τρέχοντα γύρο. Για την ανάπτυξη των γραφημάτων της εφαρμογής χρησιμοποιήθηκε η βιβλιοθήκη Jfreechart.

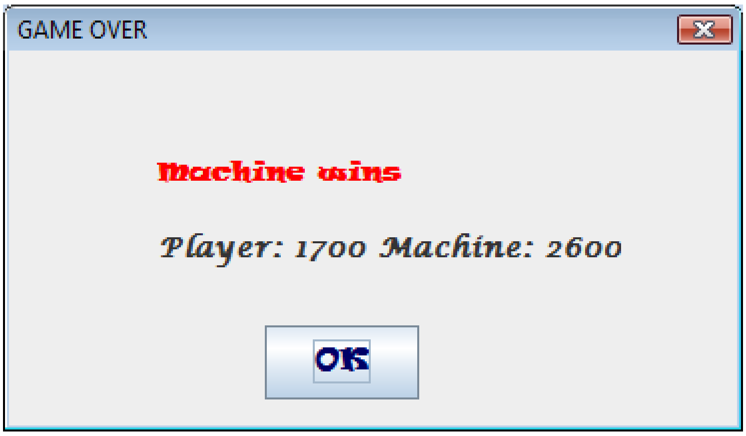



**Εικόνα ΙΙ.8 Γράφημα ποσοστών.**




**Εικόνα ΙΙ.9 Γράφημα καταμέτρησης κινήσεων.**

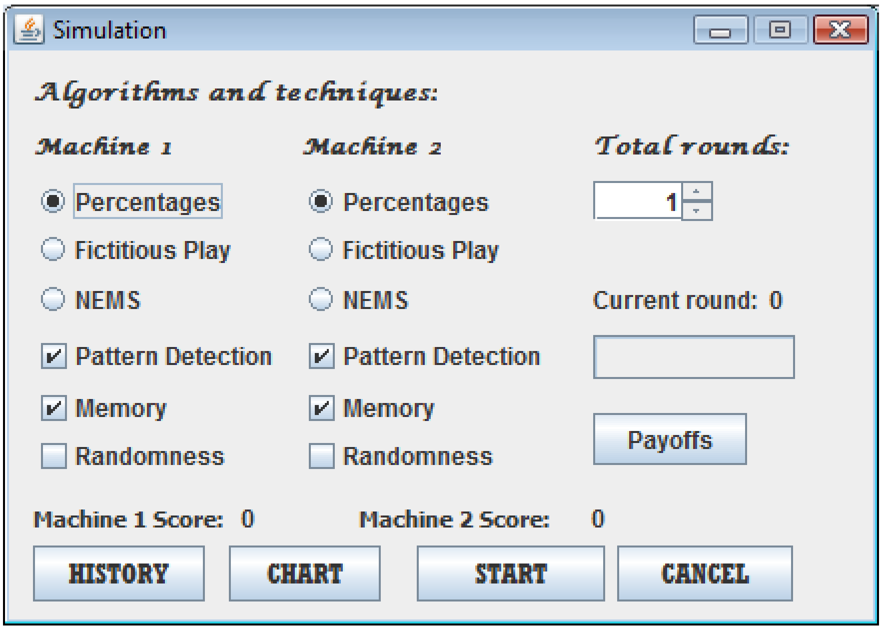
Όταν ολοκληρωθεί ο αριθμός των επιλεγμένων επαναλήψεων του παιγνίου σκηνή εμφανίζεται ένα παράθυρο που ενημερώνει τον χρήστη για τον νικητή του παιχνιδιού και την βαθμολογία.



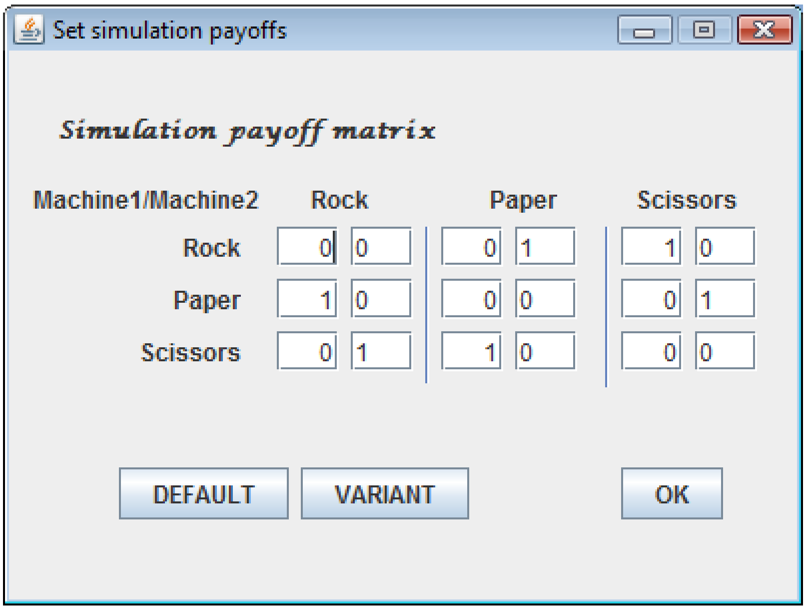

**Εικόνα II.10 Παράθυρο βαθμολογίας**

## II.II Πράκτορας εναντίον πράκτορα

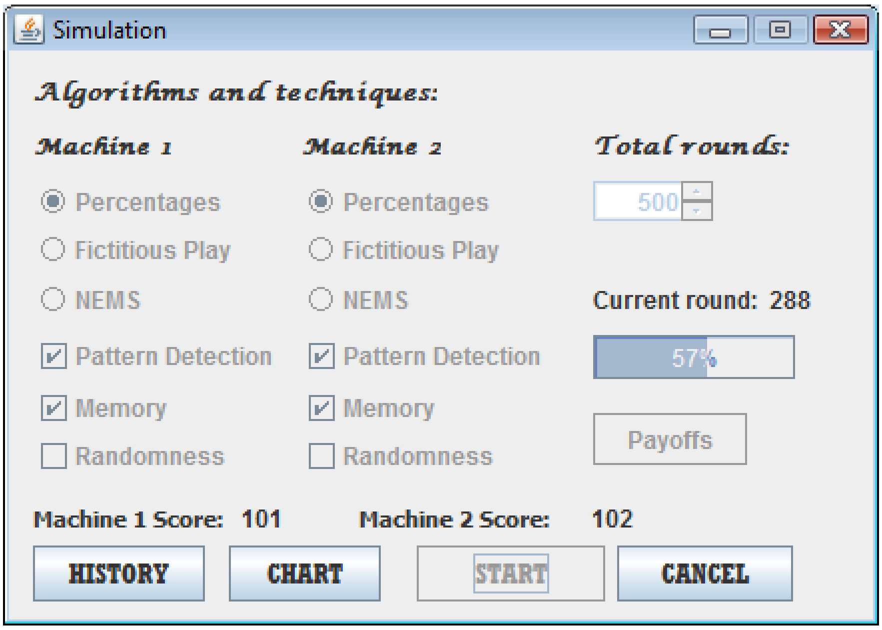
Πατώντας το κουμπί simulation από την εισαγωγική οθόνη ανοίγει το παράθυρο της εξομοίωσης πράκτορα εναντίον πράκτορα (εικόνα 5.10) και τα δύο στατιστικά παράθυρα που αναφέρθηκαν και για το παιχνίδι χρήστη εναντίον μηχανής.




**Εικόνα II.11 Αρχικό παράθυρο εξομοίωσης.**




Εικόνα II.12 Προσαρμογή πίνακα αποδόσεων πράκτορα εναντίον πράκτορα




Εικόνα II.13 Παράθυρο εξομοίωσης σε εξέλιξη.

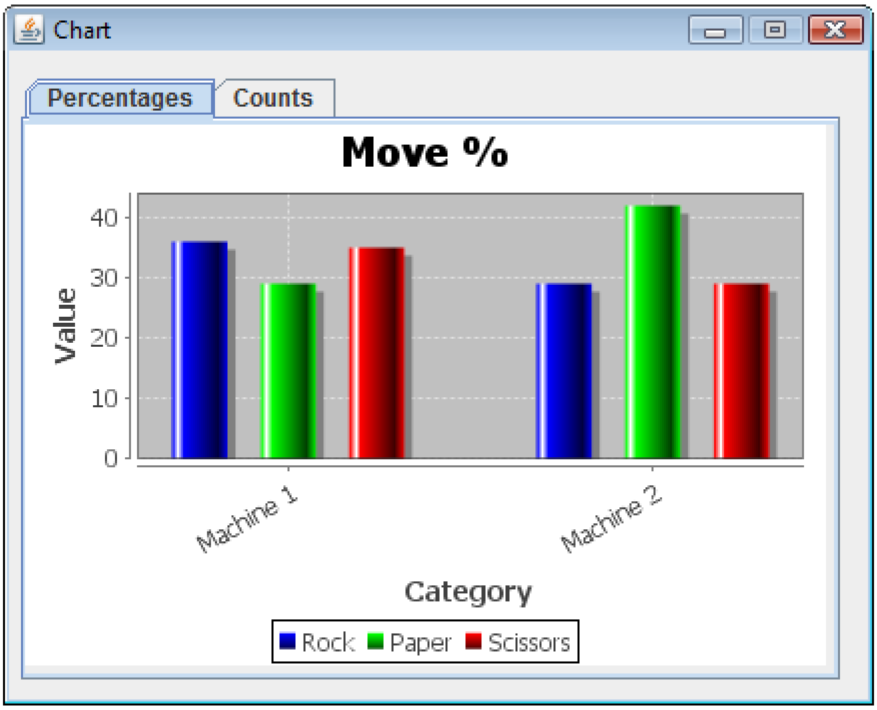
Στο παράθυρο της εξομοίωσης ο χρήστης μπορεί να διαλέξει από τους βασικούς αλγορίθμους και τους αλγορίθμους/τεχνικές που υπήρχαν διαθέσιμοι και στο παιχνίδι χρήστη εναντίον μηχανής, αυτή τη φορά και για τους δύο πράκτορες. Με το κουμπί Payoffs μπορεί να προσαρμόσει τον πίνακα αποδόσεων και με τα κουμπιά history και chart να ανοιγοκλείνει τα παράθυρα του ιστορικού και του γραφήματος, όπως και στο άλλος είδος παιχνιδιού που αναφέρθηκε. Στο πεδίο total rounds δίνεται το πλήθος των επαναλήψεων του παιγνίου σκηνή. Στο παράθυρο αυτό ο χρήστης παρατηρεί την πρόοδο του παιχνιδιού την βαθμολογία των δύο πρακτόρων , σε ποιον γύρο βρίσκεται καθώς και μια μπάρα προόδου που βρίσκεται κάτω από αυτό και δείχνει ποσοστιαία της πρόοδο. Με το κουμπί start ξεκινάει η εξομοίωση.

Τα παράθυρα του ιστορικού και του γραφήματος δείχνουν ακριβώς τα ίδια πράγματα με πριν, αυτή τη φορά για τις δύο μηχανές.

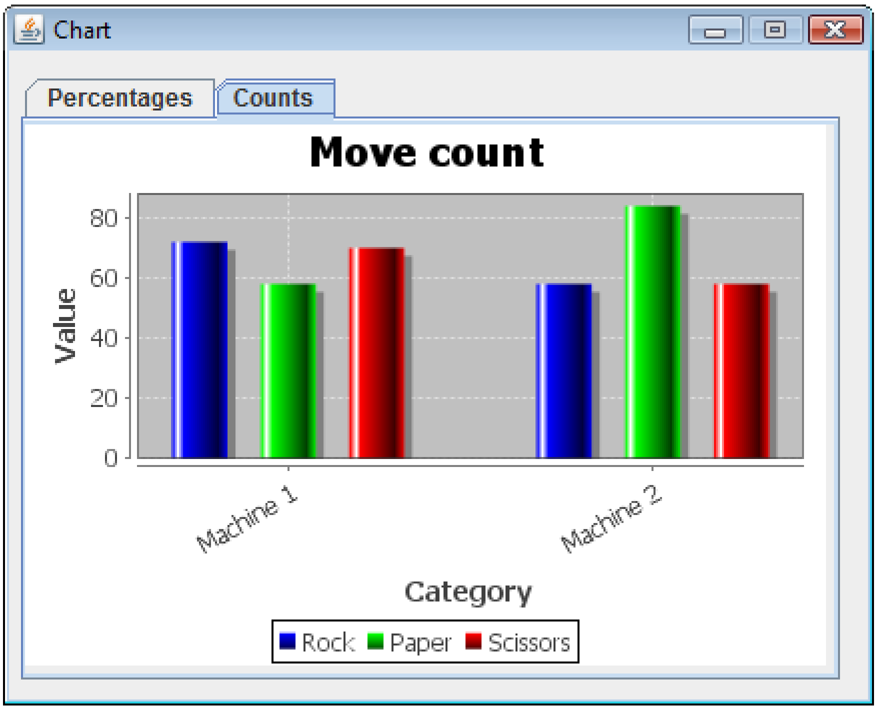
History

| Machine 1 | Machine 2 | Round |
|---|---|---|
| Rock | Rock | 1 |
| Paper | Paper | 2 |
| Paper | Paper | 3 |
| Scissors | Scissors | 4 |
| Rock | Scissors | 5 |
| Rock | Rock | 6 |
| Rock | Paper | 7 |
| Scissors | Paper | 8 |
| Scissors | Paper | 9 |
| Scissors | Rock | 10 |
| Paper | Rock | 11 |
| Paper | Scissors | 12 |
| Rock | Scissors | 13 |
| Rock | Paper | 14 |
| Paper | Paper | 15 |
| Scissors | Paper | 16 |

**Εικόνα ΙΙ.14 Ιστορικό πράκτορα εναντίον πράκτορα.**




**Εικόνα ΙΙ.15 Γράφημα ποσοστών πράκτορα εναντίον πράκτορα.**




**Εικόνα ΙΙ.16 Γράφημα καταμέτρησης πράκτορα εναντίον πράκτορα**