



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ &  
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

---

# Η Αξιωματοποίηση Συναισθήματος και η Σχέση του με την Έννοια της Εμπιστοσύνης για Πολυπρακτορικά Συστήματα

---

Μεταπτυχιακή Διατριβή

**Ράπτη Έλλη**

Επιβλέπουσα: **Δασκαλοπούλου Ασπασία, Επίκουρη Καθηγήτρια**

Μέλη Επιτροπής: **Αντωνόπουλος Χρήστος, Επίκουρος Καθηγητής**  
**Κουτσόπουλος Ιορδάνης, Επίκουρος Καθηγητής**

Βόλος, Φεβρουάριος 2014

UNIVERSITY OF THESSALY  
FACULTY OF ENGINEERING  
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

---

# **Formalization of Emotion and its Relevance to the Concept of Trust in Multi-agent Systems**

---

Master Thesis

**Rapti Elli**

Supervisor: **Daskalopoulou Aspasia, Assistant Professor**

Committee Members: **Antonopoulos Christos, Assistant Professor**  
**Koutsopoulos Iordanis, Assistant Professor**

Volos, February 2014

---

## Ευχαριστίες

Με την εκπόνηση της μεταπτυχιακής διατριβής μου θα ήθελα να ευχαριστήσω αρχικά κυρία Ασπασία Δασκαλοπούλου, επίκουρη καθηγήτρια του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας, για την εμπιστοσύνη που έδειξε για την εκπόνηση της παρούσας εργασίας. Οι συμβουλές της ήταν πολύτιμες, τόσο για την ολοκλήρωση της, όσο και για την ποιότητα των αποτελεσμάτων.

Επιπλέον, θα ήθελα να ευχαριστήσω τον κύριο Αντώνη Καραγεώργο, επίκουρο καθηγητή του ΤΕΙ Θεσσαλίας, για την γνώση που μου προσέφερε καθώς και για την άδειά του για τη χρήση του μοντέλου προσομοίωσης που χρησιμοποιήθηκε στα πλαίσια της παρούσας εργασίας.

---

## Πρόλογος

Η εμπιστοσύνη και η αξιοπιστία, όπως και η προστασία της ιδιωτικής ζωής, η φήμη, η ασφάλεια και ο έλεγχος, αποτελούν τα μείζονα ερευνητικά θέματα της επιστήμης των υπολογιστών. Η εμπιστοσύνη είναι ιδιαίτερα σημαντική σε εφαρμογές όπως είναι η αλληλεπίδραση ανθρώπου-μηχανής για την μοντελοποίηση των σχέσεων μεταξύ των χρηστών και των βοηθητικών μηχανημάτων τους. Όμως, η έννοια της εμπιστοσύνης δεν αφορά απλά την ασφαλή επικοινωνία, αλλά επεκτείνεται στην αξιολόγηση της αξιοπιστίας της πληροφορίας σύμφωνα με την κατάσταση του ιδιοκτήτη. Έτσι, διακρίνονται και διαφορετικά ήδη εμπιστοσύνης, όπως είναι η εμπιστοσύνη στο περιβάλλον και στις υποδομές (κοινωνικοτεχνικό σύστημα) που συμπεριλαμβάνει και την εμπιστοσύνη στον προσωπικό πράκτορα αλλά και στους μεσάζοντες πράκτορες, εμπιστοσύνη στους πιθανούς συνεργάτες καθώς και εμπιστοσύνη στις αρχές (αν υπάρχουν).

Όσον αφορά τις πολυπρακτορικές κοινωνίες, η εμπιστοσύνη αποτελεί ένα από τα πιο βασικά ζητήματα, ειδικά σε μεγάλης κλίμακας κατανομημένα συστήματα. Βρίσκεται στον πυρήνα όλων των αλληλεπιδράσεων μεταξύ των οντοτήτων που πρέπει να λειτουργούν σε ένα τέτοιο αβέβαιο και συνεχώς μεταβαλλόμενο περιβάλλον. Στην παρούσα διατριβή γίνεται η εφαρμογή των τυπικολογικών (modal-logic) χαρακτηρισμών ορισμένων συναισθημάτων, όπως προτάθηκαν από τον Ingmar Pörn και παρουσιάστηκαν από τους Andrew Jones και Jeremy Pitt. Περίπλοκα συναισθήματα, όπως είναι η εμπιστοσύνη, η ντροπή και η μετάνοια, περιγράφονται με χρήση modal logic τα οποία στη συνέχεια εφαρμόζονται σε ένα πολυπρακτορικό σύστημα, το οποίο μοντελοποιεί ένα peer-to-peer σύστημα για τη σύνθεση υπηρεσιών διαδικτύου.

---

## Abstract

Trust and trustworthiness, along with related concepts such as privacy, reputation, security and control, have become some of the most important research topics in computer science. Trust is particularly important in applications such as human-computer interaction in order to model the relationship between the users and their personal assistants. The notion of trust spans further than just the secure communication between two parties, including the notion of reliability of the information exchanged depending on the status of the partner. Thereby, different kinds of trust are discerned, including trust in the environment and in the infrastructure (socio-technical system) which further involves trust in one's personal agent and in other mediating agents, trust in the potential partners and trust in the authorities (if any).

In the case of multi-agent societies, trust comprises one of the most important issues, especially in large scale distributed systems. Trust is the core of all interactions between the different entities that have to operate within such an uncertain and constantly changing environment. This thesis is an implementation of the modal-logical characterization of certain emotions, as introduced by Ingmar Pörn and further presented by Andrew Jones and Jeremy Pitt. Complex emotions, such as trust, regret and shame, described through modal logic are applied in a multi-agent scenario, concerning a Peer-to-Peer model for web service composition.

---

## Περιεχόμενα

<b>Ευχαριστίες</b> .....	3
<b>Πρόλογος</b> .....	4
<b>Abstract</b> .....	5
<b>Εισαγωγή</b> .....	9
<b>Κεφάλαιο 1</b> .....	11
<b>Ευφυείς Πράκτορες</b> .....	11
1.1 Τεχνητή Νοημοσύνη .....	11
1.2 Τι εννοούμε με τον όρο ‘Πράκτορας’; .....	12
1.3 Κατηγορίες Πρακτόρων .....	13
1.4 Το Περιβάλλον ενός Πράκτορα .....	15
1.5 Ευφυείς Πράκτορες.....	15
1.6 Αφηρημένες Αρχιτεκτονικές Ευφύων Πρακτόρων .....	17
1.6.1 Καθαρά Αντιδραστικοί Πράκτορες .....	20
1.6.2 Πράκτορες με Αντίληψη .....	20
1.6.3 Πράκτορες με Εσωτερική Κατάσταση .....	21
1.7 Αρχιτεκτονικές Ευφύων Πρακτόρων με Εσωτερική Κατάσταση .....	22
1.7.1 Ορθολογικοί Πράκτορες.....	23
1.7.2 Αντιδραστικοί Πράκτορες .....	24
1.7.3 Πράκτορες που Βασίζονται σε Πεποιθήσεις – Επιθυμίες – Προθέσεις.....	25
1.7.4 Υβριδικοί Πράκτορες.....	28
1.7.5 Πολυεπίπεδη Αρχιτεκτονική .....	29
1.7.6 Κινητοί Πράκτορες .....	30
<b>Κεφάλαιο 2</b> .....	32
<b>Πολυπρακτορικά Συστήματα</b> .....	32
2.1 Εισαγωγή.....	32
2.2 Επικοινωνία Πρακτόρων .....	34
2.2.1 Πρωτόκολλα Αλληλεπίδρασης .....	35
2.2.2 Γλώσσες Επικοινωνίας .....	36
2.3 Μεθοδολογίες Ανάπτυξης Πολυπρακτορικών Συστημάτων .....	39
2.3.1 Ρόλοι Πρακτόρων.....	41
2.3.2 Gaia.....	41
2.3.3 Tropos.....	42

---

<b>Κεφάλαιο 3</b> .....	44
<b>Εμπιστοσύνη και Πολυπρακτορικά Συστήματα</b> .....	44
3.1 Η Έννοια της Εμπιστοσύνης.....	44
3.2 Εμπιστοσύνη στα Πολυπρακτορικά Συστήματα.....	45
3.3 Εμπιστοσύνη Ατομικού Επιπέδου.....	48
3.3.1 Μαθησιακή και Εξελικτική Εμπιστοσύνη.....	49
3.3.3 Κοινωνικογνωστικά Μοντέλα Εμπιστοσύνης.....	52
3.4.1 Πρωτόκολλα Αλληλεπίδρασης.....	55
3.5 Σύνοψη Μοντέλων Εμπιστοσύνης.....	58
<b>Κεφάλαιο 4</b> .....	60
<b>Λογική στα Πολυπρακτορικά Συστήματα</b> .....	60
4.1 Εισαγωγή.....	60
4.2 Modal Logic.....	60
4.3 Σημασιολογία της Modal Logic.....	61
4.4 Normal Modal Logic.....	61
4.4.1 Θεωρία Αντιστοιχίας.....	63
4.4.2 Normal Modal Logic ως Επιστημική Λογική.....	65
4.4.3 Αξιώματα Γνώσης και Πεποίθησης.....	66
4.5 Επιστημική Γνώση για Πολυπρακτορικά Συστήματα.....	66
4.6 Στόχοι και Επιθυμίες.....	67
4.7 Κοινή και Καταναεμημένη Γνώση.....	68
<b>Κεφάλαιο 5</b> .....	69
<b>Συναισθήματα στα Πολυπρακτορικά Συστήματα</b> .....	69
5.1 Τυπικός Χαρακτηρισμός των Συναισθημάτων.....	69
5.1.1 Επιστημικό Στοιχείο.....	69
5.1.2 Στοιχείο Βούλησης.....	71
5.1.3 Πολύπλοκα Συναισθήματα: Εμπιστοσύνη και Ντροπή.....	72
5.2 Ψυχοφυσιολογική Ταξινόμηση του Συναισθήματος.....	74
5.3 Συγχώρεση.....	76
<b>Κεφάλαιο 6</b> .....	78
<b>Μοντελοποίηση της Εμπιστοσύνης ενός Πολυπρακτορικού Συστήματος μέσω του MASON</b> .....	78
6.1 Εισαγωγή.....	78
6.2 Εύρεση και Σύνθεση Υπηρεσιών Διαδικτύου.....	78
6.2.1 Προσεγγίσεις.....	79

---

---

6.2.2	Peer-to-peer Μοντέλο Εύρεσης και Σύνθεσης Υπηρεσιών.....	79
6.3	MASON.....	84
6.3.1	Αρχιτεκτονική.....	84
6.3.2	Κονσόλα Διαχείρισης.....	86
6.3.3	Οθόνη Απεικόνισης.....	87
6.4	Αρχιτεκτονική Συναισθηματικών Πρακτόρων.....	87
6.5	Μοντελοποίηση της Εμπιστοσύνης ενός Πολυπρακτορικού Συστήματος.....	89
6.5.1	Σενάριο Υλοποίησης.....	90
6.5.2	Πειραματικά Αποτελέσματα.....	91
6.6	Μελλοντικές Προεκτάσεις.....	95
	<b>Βιβλιογραφία.....</b>	<b>97</b>



---

## Εισαγωγή

Τα τελευταία χρόνια, η χρήση των ευφυών πρακτόρων αποδεικνύεται ολοένα και πιο σημαντική για διάφορα ερευνητικά πεδία, κυρίως στην περιοχή της Επιστήμης Υπολογιστών και ειδικότερα στον τομέα της Τεχνητής Νοημοσύνης. Τα πολυπρακτορικά συστήματα, ή αλλιώς τα συστήματα πολλαπλών πρακτόρων, είναι συστήματα που έχουν σχεδιαστεί και υλοποιηθεί ως ένα σύνολο πρακτόρων που αλληλεπιδρούν, δηλαδή συνεργάζονται, συντονίζονται, διαπραγματεύονται, με στόχο την επίλυση πολύπλοκων προβλημάτων προσομοιώνοντας την ανθρώπινη συμπεριφορά.

Το πρώτο κεφάλαιο της παρούσας διατριβής αποτελεί μια εισαγωγή στον τομέα των ευφυών πρακτόρων, όπου γίνεται μια λεπτομερής παρουσίαση των ειδών, των χαρακτηριστικών και του τρόπου λειτουργίας των πρακτόρων. Οι ευφυείς πράκτορες μελετώνται με βάση τα δομικά και λειτουργικά τους χαρακτηριστικά. Εξετάζονται τα είδη του περιβάλλοντος μέσα στα οποία δρουν καθώς και τους διαφορετικούς τρόπους με τους οποίους αντιδρά η κάθε αρχιτεκτονική σε κάθε περιβάλλον.

Στο δεύτερο κεφάλαιο γίνεται μια λεπτομερής περιγραφή των βασικών εννοιών που αφορούν τα πολυπρακτορικά συστήματα. Αναλύεται η αλληλεπίδραση των πρακτόρων μέσω των πρωτοκόλλων επικοινωνίας που εφαρμόζονται καθώς και οι γλώσσες οι οποίες χρησιμοποιούνται για την επίτευξη της επικοινωνίας μεταξύ των πρακτόρων ενός συστήματος (KQML, FIPA ACL). Τέλος, παρουσιάζονται δύο από τις σημαντικότερες και πιο διαδεδομένες μεθοδολογίες σχεδιασμού και ανάπτυξης πολυπρακτορικών συστημάτων (Gaia, Tropos).

Στο τρίτο κεφάλαιο γίνεται μια αναλυτική περιγραφή της σημασίας της εμπιστοσύνης στα υπολογιστικά συστήματα και συγκεκριμένα σε μεγάλης κλίμακας καταναμημένα πολυπρακτορικά συστήματα. Παρουσιάζονται διάφορα μοντέλα εμπιστοσύνης ατομικού επιπέδου, τα οποία καθιστούν έναν πράκτορα ικανό να κρίνει το επίπεδο εμπιστοσύνης των αντιπάλων του, καθώς και μηχανισμοί και πρωτόκολλα για την εξασφάλιση της εμπιστοσύνης σε επίπεδο συστήματος και τα οποία έχουν ως στόχο να εξαναγκάσουν τους πράκτορες να δρουν και να αλληλεπιδρούν με ειλικρίνεια.

Στο τέταρτο κεφάλαιο παρουσιάζεται η εφαρμογή της λογικής στα πολυπρακτορικά συστήματα. Δίδεται μια λεπτομερής ανάλυση της κυρίαρχη προσέγγισης των πολυπρακτορικών συστημάτων η οποία καλείται modal logic. Η κύρια ιδέα πίσω από την modal logic είναι η ανάπτυξη ενός λογικού μοντέλου που θα μπορεί να χρησιμοποιηθεί για τον χαρακτηρισμό της διανοητικής κατάστασης των πρακτόρων όπως αυτοί δρουν και αλληλεπιδρούν.

Στο πέμπτο κεφάλαιο δίδεται η περιγραφή πολύπλοκων συναισθημάτων, όπως είναι η εμπιστοσύνη, η μετάνοια και η ντροπή, χρησιμοποιώντας modal-logical χαρακτηρισμούς, όπως αρχικά προτάθηκαν από τον Ingmar Pörn και στη συνέχεια αναλύθηκαν από τους Andrew Jones και Jeremy Pitt. Συγκεκριμένα, η προσέγγιση

---

βασίζεται στην παραδοχή ότι κάθε συναίσθημα παρουσιάζει διπλή σκοπιμότητα, τη βουλητική και επιστημική άποψη της σκοπιμότητας.

Τέλος, στο έκτο κεφάλαιο γίνεται μια λεπτομερής περιγραφή του peer-to-peer μοντέλου σύνθεσης υπηρεσιών διαδικτύου που αναπτύχθηκε στο MASON σύμφωνα με τους τυπολογικούς χαρακτηρισμούς πολύπλοκων συναισθημάτων, όπως είναι η εμπιστοσύνη, η μετάνοια και η ντροπή, και οι οποίοι παρουσιάστηκαν λεπτομερώς στο πέμπτο κεφάλαιο. Παρουσιάζονται και σχολιάζονται τα αποτελέσματα που λήφθηκαν από την εκτέλεση του αλγόριθμου και δίδονται προτάσεις μελλοντικές επεκτάσεις του μοντέλου.

---

# Κεφάλαιο 1

## Ευφυείς Πράκτορες

---

### 1.1 Τεχνητή Νοημοσύνη

Ο τομέας της Τεχνητής Νοημοσύνης επιχειρεί να κατανοήσει τις ευφυείς οντότητες, όμως, αντίθετα με την ψυχολογία και την φιλοσοφία που επίσης ασχολούνται με την ευφυΐα, η Τεχνητή Νοημοσύνη εκτός από το να τις κατανοήσει προσπαθεί και να χτίσει τις ευφυείς οντότητες. Αν και κανείς δεν μπορεί να προβλέψει το μέλλον, είναι προφανές ότι τα υπολογιστικά συστήματα με νοημοσύνη ανθρώπινου επιπέδου (ή και περισσότερη) θα έχουν σημαντικές επιπτώσεις στην καθημερινή μας ζωή καθώς και στην μελλοντική πορεία του ανθρώπινου πολιτισμού.

Η Τεχνητή Νοημοσύνη αποτελεί ένα από τους νεότερους επιστημονικούς κλάδους καθώς ξεκίνησε επίσημα το 1956 αν και η Νοημοσύνη μελετάται πάνω από 2000 χρόνια. Περιλαμβάνει μια μεγάλη ποικιλία υποπεδίων, από περιοχές γενικού ενδιαφέροντος όπως η αντίληψη και λογικό συλλογισμό, σε συγκεκριμένα έργα όπως το σκάκι, η απόδειξη μαθηματικών θεωρημάτων, η συγγραφή ποιημάτων και η διάγνωση ασθενειών. Συνήθως, επιστήμονες άλλων τομέων κινούνται προς την Τεχνητή Νοημοσύνη, όπου μπορούν να βρουν τα κατάλληλα εργαλεία και το λεξιλόγιο έτσι ώστε να συστηματοποιήσουν και να αυτοματοποιήσουν τα πνευματικά καθήκοντα για τα οποία έχουν εργαστεί σε όλη τους τη ζωή. Με παρόμοιο τρόπο, οι επιστήμονες που ασχολούνται με την Τεχνητή Νοημοσύνη επιλέγουν διάφορους τομείς πάνω στους οποίους επιχειρούν να εφαρμόσουν τις μεθόδους τους. Με αυτό τον τρόπο αποδεικνύεται ότι είναι ουσιαστικά ένας καθολικός τομέας [1].

## 1.2 Τι εννοούμε με τον όρο ‘Πράκτορας’;

Τα τελευταία χρόνια η χρήση των πρακτόρων, και ειδικότερα του ευφυών πρακτόρων, αποδεικνύεται σημαντική σε διάφορα ερευνητικά πεδία, αλλά κυρίως στην επιστήμη των υπολογιστών και ειδικότερα στον τομέα της τεχνητής νοημοσύνης.

Ο τομέας των πρακτόρων είναι αρκετά νέος με αποτέλεσμα να μην υπάρχει ένα μοναδικός και διεθνώς αποδεκτός ορισμός του όρου ‘πράκτορας’. Αντί ενός καθολικού ορισμού, μέχρι σήμερα, η παγκόσμια επιστημονική κοινότητα έχει καταλήξει στο γεγονός ότι η έννοια της αυτονομίας αποτελεί έναν καθοριστικό παράγοντα στον ορισμό ενός πράκτορα. Η δυσκολία στην υιοθέτηση ενός καθολικού ορισμού για την έννοια του όρου ‘πράκτορας’ είναι ότι τα διάφορα χαρακτηριστικά που συσχετίζονται με έναν πράκτορα και κατ’ επέκταση με τις λειτουργίες του έχουν εντελώς διαφορετική σημασία για τους διαφορετικούς επιστημονικούς τομείς [2]. Επομένως, ενώ υπάρχει μια γενική συναίνεση ότι η αυτονομία είναι κεντρικό στοιχείο στον καθορισμό της έννοιας του πράκτορα, πέρα από αυτό παρατηρείται ασυμφωνία. Έτσι, για παράδειγμα, ενώ σε ορισμένες εφαρμογές η δυνατότητα εκμάθησης των πρακτόρων είναι υψίστης σημασίας, σε άλλες εφαρμογές η εκμάθηση μπορεί να είναι όχι μόνο ασήμαντη αλλά και μια ανεπιθύμητη δυνατότητα.

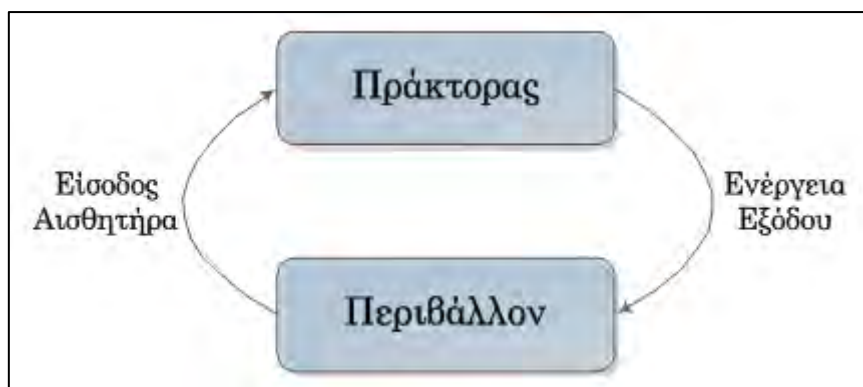
Κάποιοι από τους ορισμούς που έχουν δοθεί για τον όρο ‘πράκτορας’ είναι οι εξής:

«Ένας πράκτορας είναι ένα υπολογιστικό σύστημα που είναι τοποθετημένο σε κάποιο περιβάλλον και είναι σε θέση να δρα αυτόνομα σε αυτό το περιβάλλον έτσι ώστε να επιτύχει του στόχους του σχεδιασμού του» [3].

«Οι πράκτορες είναι προγράμματα που συμμετέχουν σε διαλόγους και διαπραγματεύονται και συντονίζουν την μεταφορά των πληροφοριών» [4].

«Ένας πράκτορας είναι μια οντότητα που μπορεί να θεωρηθεί ότι αντιλαμβάνεται το περιβάλλον του μέσω αισθητήρων και ενεργεί πάνω στο περιβάλλον του μέσω μηχανισμών ενεργειών» [1].

Η εικόνα 1.1 μας δίνει μια αφηρημένη άποψη ενός πράκτορα. Στο συγκεκριμένο διάγραμμα, φαίνεται η ενέργεια που παράγεται ως έξοδος από τον πράκτορα η οποία θα έχει ως αποτέλεσμα να επηρεάσει το περιβάλλον του. Στους περισσότερους τομείς που χαρακτηρίζονται από πολυπλοκότητα, ένας πράκτορας δεν μπορεί να έχει καθολικό έλεγχο πάνω στο περιβάλλον του. Στην καλύτερη περίπτωση, θα έχει μερικό έλεγχο, με την έννοια ότι θα μπορεί απλά να το επηρεάσει με τις πράξεις του. Από την πλευρά του πράκτορα, αυτό σημαίνει ότι η ίδια ενέργεια που πραγματοποιείται δύο φορές σε προφανώς όμοιες καταστάσεις μπορεί να φαίνεται ότι έχουν εντελώς διαφορετικά αποτελέσματα και, ειδικότερα, ενδέχεται να μην φέρουν το επιθυμητό αποτέλεσμα. Έτσι οι πράκτορες θα πρέπει να είναι προετοιμασμένοι για το ενδεχόμενο της αποτυχίας. Συνοψίζοντας τα παραπάνω, γενικά θεωρείται ότι το περιβάλλον ενός πράκτορα είναι μη ντετερμινιστικό.



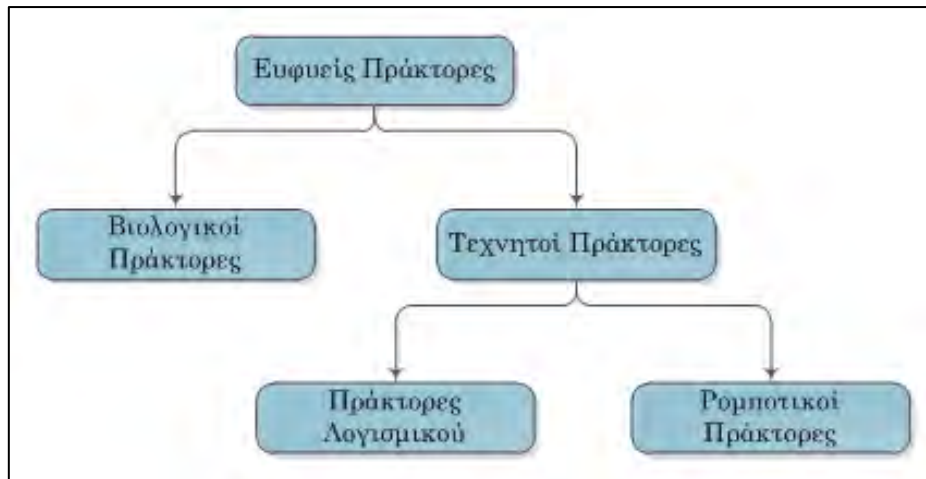
**Εικόνα 1.1** Ένας πράκτορας στο περιβάλλον του. Ο πράκτορας λαμβάνει ως είσοδο ένα ερέθισμα από το περιβάλλον και παράγει ως εξόδους ενέργειες που το επηρεάζουν. Η αλληλεπίδραση είναι συνήθως συνεχής και ατέρμονη.

Συνήθως, ένας πράκτορας έχει διαθέσιμο ένα ρεπερτόριο ενεργειών. Αυτό το σύνολο πιθανών ενεργειών αντιπροσωπεύει την ικανότητα του πράκτορα να τροποποιεί το περιβάλλον του (effectoric capability). Δεν μπορούν να εφαρμοσθούν όλες οι ενέργειες ενός πράκτορα σε όλες τις καταστάσεις. Για παράδειγμα, μια ενέργεια 'σήκωσε το τραπέζι' είναι εφαρμόσιμη μόνο στην περίπτωση όπου το βάρος του τραπεζιού είναι αρκετά μικρό έτσι ώστε ο πράκτορας να είναι σε θέση να το σηκώσει. Από αυτό μπορούμε να συμπεράνουμε λοιπόν ότι οι ενέργειες έχουν προϋποθέσεις που προσδιορίζουν σε ποιες περιπτώσεις μπορούν να εφαρμοσθούν.

Οι πράκτορες μπορεί να 'κατοικούν' είτε σε ένα φυσικό περιβάλλον, δηλαδή στον φυσικό κόσμο, είτε να είναι ένα πρόγραμμα ενός περιβάλλοντος λογισμικού. Για παράδειγμα, ένας θερμοστάτης μπορεί να θεωρηθεί ως ένας πράκτορας φυσικού περιβάλλοντος, ο οποίος λαμβάνοντας ως ερέθισμα την θερμοκρασία του δωματίου στο οποίο βρίσκεται ανοίγει και κλείνει ανάλογα την θέρμανση. Αντίστοιχα, ένας πράκτορας σε ένα περιβάλλον λογισμικού λαμβάνει πληροφορίες για το περιβάλλον του εκτελώντας κάποιες συναρτήσεις και παράγει ως έξοδο ενέργειες λογισμικού, όπως είναι η αλλαγή ενός εικονιδίου στην οθόνη ή η εκτέλεση ενός προγράμματος [5].

### 1.3 Κατηγορίες Πρακτόρων

Σε ένα πιο αφηρημένο επίπεδο διακρίνονται δύο μεγάλες κατηγορίες πρακτόρων, οι βιολογικοί και οι τεχνητοί πράκτορες (Εικόνα 1.2). Οι βιολογικοί πράκτορες χρησιμοποιώντας τις αισθήσεις τους αντιλαμβάνονται το περιβάλλον τους και εκμεταλλευόμενοι τις γνώσεις τους βγάζουν συμπεράσματα για αυτό. Έτσι, χρησιμοποιώντας τους μηχανισμούς δράσης τους αντίστοιχα εφαρμόζουν τις ενέργειες που προκύπτουν ως αποτέλεσμα του συλλογισμού τους. Οι τεχνητοί πράκτορες λειτουργούν με παρόμοιο τρόπο ενώ διαχωρίζονται σε δύο επιμέρους κατηγορίες, τους ρομποτικούς και τους πράκτορες λογισμικού. Οι ρομποτικοί πράκτορες διαθέτουν μηχανικούς ή ηλεκτρονικούς αισθητήρες και μηχανισμούς δράσης με τα οποία μπορούν να δρουν στον πραγματικό κόσμο (Εικόνα 1.3 (α)). Οι πράκτορες λογισμικού από την άλλη πλευρά είναι προγράμματα λογισμικού και οι δράσεις τους εφαρμόζονται σε ένα υπολογιστικό σύστημα (Εικόνα 1.3 (β)).

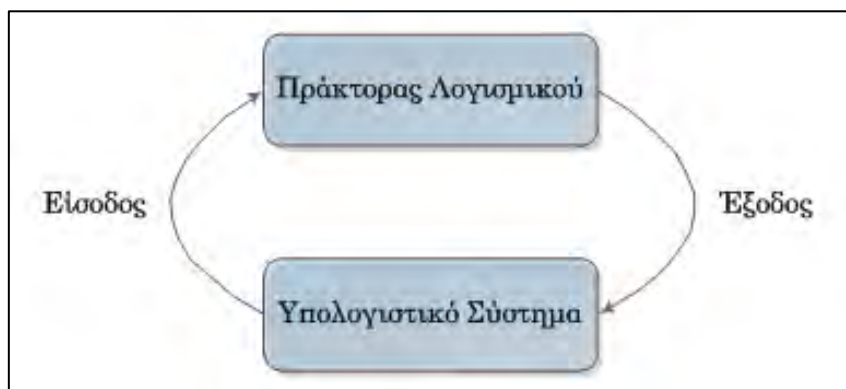


**Εικόνα 1.2** Ιεραρχική ταξινόμηση πρακτόρων

Τόσο οι βιολογικοί, όσο και οι τεχνητοί πράκτορες διαθέτουν μία διαδικασία συλλογισμού, μέσω της οποίας επεξεργάζονται τα ερεθίσματα που λαμβάνουν από το περιβάλλον τους και εφαρμόζουν τα αποτελέσματα που προκύπτουν από αυτόν τον συλλογισμό στο περιβάλλον, τροποποιώντας έτσι την κατάστασή του [2].



**(α)**



**(β)**

**Σχήμα 1.3** Πράκτορες **(α)** ρομποτικοί και **(β)** λογισμικού

## 1.4 Το Περιβάλλον ενός Πράκτορα

Σύμφωνα με τους Russell και Norvig (1995) [1] το περιβάλλον ενός πράκτορα έχει τις παρακάτω ιδιότητες:

- **Προσβάσιμο ή Δυσπρόσιτο** – Ένα προσβάσιμο περιβάλλον είναι εκείνο στο οποίο ο πράκτορας μπορεί να λάβει πλήρεις, ακριβείς, και ενημερωμένες πληροφορίες σχετικά με την κατάσταση του περιβάλλοντος. Στον πραγματικό κόσμο οι περισσότερες περιπτώσεις περιβάλλοντος (συμπεριλαμβανομένου του καθημερινού φυσικού κόσμου καθώς και το διαδίκτυο) δεν είναι προσβάσιμες με αυτήν την έννοια.
- **Ντετερμινιστικό ή Μη Ντετερμινιστικό** – Ένα περιβάλλον θεωρείται ντετερμινιστικό όταν για κάθε ενέργεια του πράκτορα επέρχεται ένα μοναδικό και προκαθορισμένο αποτέλεσμα, δεν υπάρχει δηλαδή αμφιβολία για την κατάσταση η οποία θα προκύψει από την εκτέλεση αυτής της ενέργειας.
- **Στατικό ή Δυναμικό** – Ένα στατικό περιβάλλον είναι αυτό που παραμένει αμετάβλητο και μεταβάλλεται μόνο από τις ενέργειες του πράκτορα. Σε αντίθεση, σε ένα δυναμικό περιβάλλον υπάρχουν και άλλες διεργασίες που λειτουργούν σε αυτό, και ως εκ τούτου αλλάζει με τρόπους που είναι πέρα από τον έλεγχο του πράκτορα. Ο φυσικός κόσμος είναι ένα άκρως δυναμικό περιβάλλον όπως και το διαδίκτυο.
- **Διακριτό ή Συνεχές** – Ένα περιβάλλον είναι διακριτό όταν υπάρχει ένας σταθερός και πεπερασμένος αριθμός ενεργειών και αντιλήψεων για αυτό.

Γενικά, όσον αφορά τις παραπάνω ιδιότητες ενός περιβάλλοντος μπορούμε να πούμε ότι όσο πιο προσβάσιμο είναι ένα περιβάλλον, τόσο πιο απλό είναι να δημιουργήσεις πράκτορες που θα λειτουργούν αποτελεσματικά μέσα σε αυτό. Η ποιότητα των αποφάσεων που λαμβάνει ένας πράκτορας εξαρτάται άμεσα από την ποιότητα των πληροφοριών που διαθέτει. Επιπλέον, είναι προτιμότερο το περιβάλλον να είναι ντετερμινιστικό καθώς ο σχεδιαστής του πράκτορα μπορεί να είναι σίγουρος ότι οι ενέργειες του πράκτορα θα είναι πάντα επιτυχείς, δηλαδή θα οδηγούν πάντα στο επιθυμητό αποτέλεσμα. Τέλος, να είναι στατικό, καθώς θα είναι αρκετό για τον πράκτορα να συλλέξει πληροφορίες μόνο μια φορά εφόσον μόνο αυτός αλλάζει τις συνθήκες του, και διακριτό, καθώς είναι πιο εύκολο να επιλέξει την κατάλληλη ενέργεια για κάθε κατάσταση. Από την άλλη πλευρά, το πιο πολύπλοκο περιβάλλον είναι αυτό που είναι απρόσιτο, μη ντετερμινιστικό, δυναμικό και συνεχές, το οποίο είναι γνωστό και ως ανοικτό [5].

## 1.5 Ευφυείς Πράκτορες

Δεν θεωρούνται όλοι οι πράκτορες, όπως οι θερμοστάτες, ευφυείς και προκειμένου να τους ξεχωρίσουμε θα πρέπει να ορίσουμε τις ικανότητες που θα πρέπει να έχει ένας πράκτορας έτσι ώστε να θεωρείται ευφυής. Αυτές οι ιδιότητες σύμφωνα με τους Woolridge και Jennings (1995) [3] είναι οι εξής:

- **Αυτονομία** – Οι ευφυείς πράκτορες λειτουργούν χωρίς άμεση παρέμβαση από εξωτερικούς χρήστες, ενώ διαθέτουν κάποιου είδους έλεγχο στις πράξεις τους και στην εσωτερική τους κατάσταση.
- **Αντιδραστικότητα** – Οι ευφυείς πράκτορες είναι σε θέση να αντιλαμβάνονται το περιβάλλον και να αντιδρούν άμεσα στις αλλαγές που συμβαίνουν έτσι ώστε να ικανοποιήσουν τους σχεδιαστικούς τους στόχους.
- **Προνοητικότητα** – Οι ευφυείς πράκτορες είναι σε θέση να επιδεικνύουν μια στοχευμένη συμπεριφορά παίρνοντας πρωτοβουλίες έτσι ώστε να επιτύχουν τους στόχους τους.
- **Κοινωνικότητα** – Οι ευφυείς πράκτορες έχουν την ικανότητα να αλληλεπιδρούν με άλλους πράκτορες, αλλά πιθανώς και ανθρώπους, έτσι ώστε να επιτύχουν τους στόχους τους.

Υπάρχουν και επιμέρους δευτερεύοντα χαρακτηριστικά που αποδίδονται στους ευφυείς πράκτορες και τα οποία είναι τα εξής:

- **Κινητικότητα** – Η ικανότητά τους να κινούνται σε ένα ηλεκτρονικό περιβάλλον.
- **Ειλικρίνεια** – Ένας ευφυής πράκτορας δεν μεταδίδει εις γνώση του λανθασμένες πληροφορίες.
- **Αγαθή Προαίρεση** – Οι στόχοι του κάθε πράκτορα δεν συγκρούονται και κάθε πράκτορας επομένως πάντα προσπαθεί να κάνει αυτό που του ζητείται.
- **Ορθολογισμός** – Ένας πράκτορας θα ενεργήσει έτσι ώστε να επιτευχθούν οι στόχοι του εφόσον το επιτρέπουν οι πεποιθήσεις του.
- **Μάθηση/Προσαρμοστικότητα** – Οι πράκτορες βελτιώνουν την απόδοσή τους με τον χρόνο [3].

Αυτές οι ιδιότητες, που αποδίδονται στους ευφυείς πράκτορες, είναι πολύ πιο απαιτητικές απ' ό τι φαίνεται. Ας πάρουμε ως παράδειγμα την προνοητικότητα, που αναφέρεται σε μια στοχευμένη συμπεριφορά. Δεν είναι τόσο δύσκολο να χτίσει κάποιος ένα σύστημα που προβάλλει μια στοχευμένη συμπεριφορά – αυτό συμβαίνει κάθε φορά που γράφουμε μια συνάρτηση σε C ή μια μέθοδο σε Java. Όταν γράφουμε μια τέτοια διαδικασία, την περιγράφουμε σύμφωνα με τις υποθέσεις πάνω στις οποίες στηρίζεται (τις προϋποθέσεις) και στα αποτελέσματα που θα πάρουμε αν αυτές οι υποθέσεις ισχύουν (τις επόμενες καταστάσεις). Τα αποτελέσματα της διαδικασίας είναι και ο στόχος της, αυτό που ο προγραμματιστής επιθυμεί να επιτύχει με την συγγραφή του συγκεκριμένου κώδικα. Αν η προϋπόθεση ισχύει όταν καλείται η διαδικασία, τότε αναμένεται ότι θα εκτελεστεί με επιτυχία, ότι δηλαδή θα τερματίσει και στον τερματισμό της θα ισχύει η επόμενη κατάσταση (ο στόχος έχει επιτευχθεί). Αυτή είναι μια στοχευμένη συμπεριφορά, όταν η διαδικασία είναι ένα σχέδιο ή μια συνταγή για την επίτευξη ενός στόχου. Αυτό το μοντέλο λειτουργεί αποτελεσματικά σε ένα λειτουργικό περιβάλλον, όμως όταν αναφερόμαστε σε ένα μη-λειτουργικό περιβάλλον αυτό το απλό μοντέλο στοχευμένης συμπεριφοράς δεν είναι αποδεκτό καθώς κάνει κάποιες σημαντικές οριοθετημένες υποθέσεις. Συγκεκριμένα, θεωρεί δεδομένο ότι το περιβάλλον παραμένει αμετάβλητο κατά την διάρκεια εκτέλεσης της διαδικασίας. Στην περίπτωση όμως που το περιβάλλον αλλάξει και, συγκεκριμένα, αν οι υποθέσεις που αποτελούν προϋπόθεση της διαδικασίας γίνουν εσφαλμένες κατά την διάρκεια



εκτέλεσης της διαδικασίας, τότε η συμπεριφορά της διαδικασίας δεν μπορεί να οριστεί και συνήθως απλά τερματίζει εσφαλμένα. Επιπλέον, υποθέτουμε ότι ο στόχος, που είναι ο λόγος για τον οποίο εκτελείται η διαδικασία, παραμένει σε ισχύ τουλάχιστον μέχρι να τερματιστεί η διαδικασία. Αν ο στόχος δεν παραμένει σε ισχύ, τότε απλά δεν υπάρχει πλέον λόγος να συνεχιστεί η εκτέλεση της διαδικασίας.

Σε πολλά περιβάλλοντα, καμία από αυτές τις υποθέσεις δεν ισχύουν. Συγκεκριμένα, σε περιοχές που είναι υπερβολικά πολύπλοκες έτσι ώστε να τις παρατηρήσει εξ' ολοκλήρου ένας μόνος πράκτορας, είναι πολυπρακτορικές (δηλαδή περιοχές στις οποίες κατοικούν και μεταβάλλουν πάνω από ένας πράκτορες) ή όταν υπάρχει κάποια αμφιβολία σε ένα περιβάλλον, αυτές οι υποθέσεις δεν είναι λογικές. Σε τέτοιου είδους περιβάλλοντα, μια τυφλή εκτέλεση μιας διαδικασίας χωρίς τον έλεγχο του κατά πόσο ισχύουν οι προϋποθέσεις χαρακτηρίζουν μια κακή στρατηγική. Σε ένα δυναμικό περιβάλλον τέτοιου είδους, ένας πράκτορας θα πρέπει να είναι αντιδραστικός όπως περιγράψαμε. Με αυτό τον τρόπο, είναι σε θέση να αντιδρά ανάλογα στα γεγονότα που συμβαίνουν στο περιβάλλον του και επιδρούν είτε τους στόχους του είτε τις υποθέσεις που αφορούν τις διαδικασίες που εκτελεί ο συγκεκριμένος πράκτορας έτσι ώστε να επιτύχει τους στόχους του.

Η δημιουργία ενός καθαρά στοχευμένου συστήματος δεν είναι δύσκολη όπως και η δημιουργία ενός καθαρά αντιδραστικού που ανταποκρίνεται συνεχώς στο περιβάλλον του. Παρόλα αυτά, η δημιουργία ενός συστήματος που επιτυγχάνει την απόλυτη ισορροπία μεταξύ στοχευμένης και αντιδραστικής συμπεριφοράς είναι αρκετά δύσκολη υπόθεση. Θέλουμε πράκτορες που θα επιδιώκουν συστηματικά την επίτευξη των στόχων τους χρησιμοποιώντας πολύπλοκους μηχανισμούς που θα τους επιτρέπουν να αντιδρούν στις διάφορες μεταβολές. Από την άλλη πλευρά, δεν θέλουμε πράκτορες που θα συνεχίζουν την εκτέλεση των διαδικασιών τους σε μια προσπάθεια επίτευξης των στόχων τους όταν είναι προφανές ότι η διαδικασία θα αποτύχει ή όταν, για κάποιον λόγο, ο στόχος δεν ισχύει πλέον. Σε αυτές τις περιπτώσεις χρειαζόμαστε πράκτορες που θα είναι σε θέση να αντιδρούν εγκαίρως σε μια νέα κατάσταση. Όμως, είναι ανεπιθύμητο όταν οι πράκτορες αντιδρούν συνεχώς στα νέα δεδομένα και αδυνατούν να επικεντρωθούν αρκετά στο στόχο έτσι ώστε τελικά να τον επιτύχουν.

Είναι πλέον προφανές ότι η επίτευξη της ισορροπίας μεταξύ μιας στοχευμένης και μιας αντιδραστικής συμπεριφοράς είναι πολύ δύσκολη καθώς είναι σπάνια ακόμα και στους ανθρώπους. Αυτό είναι ένα από τα σημαντικά προβλήματα που αντιμετωπίζει ένας σχεδιαστής πρακτόρων [5].

## 1.6 Αφηρημένες Αρχιτεκτονικές Ευφύων Πρακτόρων

Μπορούμε εύκολα να εξάγουμε μια αφηρημένη άποψη των πρακτόρων. Ας υποθέσουμε αρχικά ότι το περιβάλλον μπορεί να είναι σε οποιαδήποτε κατάσταση από το πεπερασμένο σύνολο  $E$  των διακριτών, στιγμιαίων καταστάσεων:

$$E = \{e, e', \dots\}$$

Ας τονίσουμε ότι το αν το περιβάλλον είναι πραγματικά διακριτό δεν είναι σημαντικό στην συγκεκριμένη περίπτωση. Είναι μια υπόθεση που μπορεί να δικαιολογηθεί απόλυτα αν πούμε ότι κάθε συνεχές περιβάλλον μπορεί να μοντελοποιηθεί από ένα διακριτό περιβάλλον σε οποιονδήποτε επιθυμητό βαθμό ακρίβειας.

Οι πράκτορες έχουν στην διάθεσή τους ένα ρεπερτόριο ενεργειών οι οποίες μπορούν να μεταβάλλουν την κατάσταση του περιβάλλοντος.

$$A_c = \{\alpha, \alpha', \dots\}$$

Όπου  $A_c$  είναι το πεπερασμένο σύνολο των ενεργειών του πράκτορα.

Το βασικό μοντέλο αλληλεπίδρασης ενός πράκτορα με το περιβάλλον του είναι το ακόλουθο. Το περιβάλλον βρίσκεται αρχικά σε μια οποιαδήποτε κατάσταση και ο πράκτορας ξεκινά επιλέγοντας μια ενέργεια που θα πραγματοποιήσει σε αυτήν την κατάσταση. Ως αποτέλεσμα αυτής της ενέργειας, το περιβάλλον ανταποκρίνεται μεταβάλλοντας την κατάστασή του χωρίς ο πράκτορας να γνωρίζει ποια θα είναι αυτή εκ των προτέρων. Με βάση την νέα κατάσταση στην οποία βρίσκεται τώρα το περιβάλλον, ο πράκτορας βρίσκεται ξανά στην φάση όπου πρέπει να επιλέξει να πραγματοποιήσει μια ενέργεια. Το περιβάλλον ξανά ανταποκρίνεται με μια νέα κατάσταση όπου ο πράκτορας καλείται ξανά να επιλέξει να πραγματοποιήσει μια ενέργεια και ούτω καθεξής.

Μία διαδρομή,  $r$ , ενός πράκτορα σε ένα περιβάλλον είναι συνεπώς μια σειρά από διαστρωμένες καταστάσεις περιβάλλοντος και ενέργειες:

$$r: e_0 \xrightarrow{a_0} e_1 \xrightarrow{a_1} e_2 \xrightarrow{a_2} e_3 \xrightarrow{a_3} \dots \xrightarrow{a_{u-1}} e_u.$$

Όπου:

- $R$  είναι το σύνολο όλων των πιθανών πεπερασμένων ακολουθιών (του  $E$  και του  $A_c$ )
- $R^{A_c}$  είναι το υποσύνολο αυτών που τερματίζουν με μια ενέργεια
- $R^E$  είναι το υποσύνολο αυτών τερματίζουν σε μια κατάσταση περιβάλλοντος

και τα  $r, r', \dots$  αντιπροσωπεύουν τα μέλη του  $R$ .

Για να συμβολίσουμε την επίδραση ενός πράκτορα στο περιβάλλον του, εισάγεται μια συνάρτηση μετασχηματισμού κατάστασης:

$$\tau: R^{A_c} \rightarrow \wp(E).$$

Επομένως μια συνάρτηση μετασχηματισμού κατάστασης χαρτογραφεί μια διαδρομή (που υποτίθεται ότι λήγει με μια ενέργεια του πράκτορα) σε ένα σύνολο πιθανών καταστάσεων του περιβάλλοντος – αυτών που μπορεί να προκύψουν από τις ενέργειες του πράκτορα.

Σε αυτόν τον ορισμό υπάρχουν δύο σημεία που αξίζει να τονιστούν. Πρώτον, το περιβάλλον υποτίθεται ότι εξαρτάται από το ιστορικό, δηλαδή κάθε επόμενη κατάσταση του περιβάλλοντος δεν καθορίζεται αποκλειστικά από την ενέργεια του

πράκτορα και την τρέχουσα κατάστασή του. Οι ενέργειες που έχουν πραγματοποιηθεί στο παρελθόν από τον πράκτορα παίζουν επίσης ρόλο στον καθορισμό της τρέχουσας κατάστασης. Δεύτερον, ο ορισμός αυτός επιτρέπει μη-ντετερμινισμό στο περιβάλλον. Επομένως υπάρχει αβεβαιότητα του αποτελέσματος που θα επιφέρει η ενέργεια του πράκτορα σε μια κατάσταση.

Αν υποθέσουμε ότι  $\tau(r) = \emptyset$  (όπου το  $r$  υποτίθεται ότι λήγει με μια ενέργεια), τότε δεν υπάρχουν πιθανές διαδοχικές καταστάσεις του  $r$ . Σε αυτήν την περίπτωση το σύστημα τερματίζει την διαδρομή και είναι συνετό να ειπωθεί ότι όλες οι διαδρομές τελικά τερματίζουν.

Επίσημα, μπορούμε να πούμε ότι ένα περιβάλλον  $Env$  είναι ένα τριπλό  $Env = \langle E, e_0, \tau \rangle$ , όπου  $E$  είναι ένα σύνολο των καταστάσεων του περιβάλλοντος,  $e_0 \in E$  είναι η αρχική κατάσταση και  $\tau$  είναι μια συνάρτηση μετασχηματισμού καταστάσεων.

Στην συνέχεια οι πράκτορες μοντελοποιούνται ως συναρτήσεις που χαρτογραφούν διαδρομές (που τερματίζουν με μια κατάσταση του περιβάλλοντος) σε ενέργειες:

$$Ag : R^E \rightarrow Ac.$$

Επομένως, ένας πράκτορας παίρνει αποφάσεις για τις ενέργειες που θα επιλέξει να πραγματοποιήσει βασιζόμενος στο ιστορικό του συστήματος που έχει παρατηρήσει μέχρι εκείνη την στιγμή.

Αν και έχει σημειωθεί πιο πάνω ότι το περιβάλλον είναι έμμεσα μη-ντετερμινιστικό, οι πράκτορες υποτίθεται ότι είναι ντετερμινιστικοί. Έστω ότι  $AG$  είναι το σύνολο όλων των πρακτόρων. Αναφερόμαστε σε ένα σύστημα ως ένα ζευγάρι ενός πράκτορα με ένα περιβάλλον. Κάθε σύστημα θα έχει συσχετίσει με αυτό ένα σύνολο πιθανών διαδρομών και ορίζουμε το σύνολο των διαδρομών ενός πράκτορα  $Ag$  στο περιβάλλον  $Env$  με  $R(Ag, Env)$ . Για λόγους ευκολίας υποθέτουμε ότι το  $R(Ag, Env)$  περιέχει μόνο διαδρομές που τερματίζουν, δηλαδή διαδρομές  $r$  τέτοιες ώστε το  $r$  να μην έχει πιθανές διαδοχικές καταστάσεις:  $\tau(r) = \emptyset$ .

Η ακολουθία:

$$(e_0, \alpha_0, e_1, \alpha_1, e_2, \dots)$$

αντιπροσωπεύει μια διαδρομή ενός πράκτορα  $Ag$  στο περιβάλλον  $Env = \langle E, e_0, \tau \rangle$  αν

1.  $e_0$  είναι η αρχική κατάσταση του  $Env$
2.  $\alpha_0 = Ag(e_0)$  και
3. Για  $u > 0$

$$e_u \in \tau((e_0, \alpha_0, \dots, \alpha_{u-1})),$$

όπου

$$\alpha_u = Ag((e_0, \alpha_0, \dots, e_u)).$$

Δύο πράκτορες  $Ag_1$  και  $Ag_2$  ονομάζονται συμπεριφοριστικά ισοδύναμοι σε σχέση με το περιβάλλον  $Env$  αν και μόνο αν  $R(Ag_1, Env) = R(Ag_2, Env)$ , και απλώς στην

συμπεριφορά ισοδύναμοι αν και μόνο αν είναι συμπεριφοριστικά ισοδύναμοι σε κάθε περιβάλλον [5].

### 1.6.1 Καθαρά Αντιδραστικοί Πράκτορες

Κάποια είδη πρακτόρων αποφασίζουν για τις ενέργειες που θα πραγματοποιήσουν χωρίς να κάνουν αναφορά στο ιστορικό τους. Βασίζουν εξολοκλήρου την ανάληψη αποφάσεων στο παρόν, χωρίς να λαμβάνουν υπόψη το παρελθόν. Αυτοί οι πράκτορες καλούνται *καθαρά αντιδραστικοί πράκτορες*, καθώς απλά απαντούν άμεσα στο περιβάλλον τους. Μια άλλη ονομασία που έχει δοθεί σε αυτό το είδος πρακτόρων από τους Genesereth και Nilsson (1987) [6] είναι *τροπιστικοί πράκτορες*, καθώς ο όρος τροπισμός αναφέρεται στην τάση των φυτών και των ζώων να αντιδρούν σε συγκεκριμένα ερεθίσματα.

Η συμπεριφορά ενός καθαρά αντιδραστικού πράκτορα μπορεί να αναπαρασταθεί από την παρακάτω συνάρτηση:

$$Ag : E \rightarrow Ac.$$

Είναι προφανές το γεγονός ότι για κάθε καθαρά αντιδραστικό πράκτορα υπάρχει ένας αντίστοιχος 'κανονικός' πράκτορας. Παρ' όλα αυτά, το αντίστροφο δεν ισχύει.

Ο πράκτορας θερμοστάτης είναι ένα παράδειγμα καθαρά αντιδραστικού πράκτορα. Υποθέτοντας ότι το περιβάλλον ενός θερμοστάτη μπορεί να βρίσκεται σε μία από τις δύο καταστάσεις (είτε πολύ κρύο ή κανονική θερμοκρασία), τότε ο θερμοστάτης μπορεί απλά να οριστεί ως εξής [5]:

$$Ag(e) = \begin{cases} \text{θέρμανση κλειστή} & e = \text{κανονική θερμοκρασία} \\ \text{θέρμανση ανοικτή} & \text{άλλη περίπτωση} \end{cases}$$

### 1.6.2 Πράκτορες με Αντίληψη

Η αρχιτεκτονική ενός πράκτορα ουσιαστικά αναφέρεται στην χαρτογράφηση των εσωτερικών στοιχείων του – τις δομές δεδομένων, τις λειτουργίες που μπορούν να εφαρμοστούν πάνω σε αυτές τις δομές και την ροή ελέγχου μεταξύ των δομών. Στην Εικόνα 1.4 βλέπουμε τον διαχωρισμό της λειτουργίας λήψης αποφάσεων στα υποσυστήματα της αντίληψης και της ενέργειας.

Η λειτουργία της *όρασης* αποτυπώνει την ικανότητα του πράκτορα να παρατηρεί το περιβάλλον του, ενώ η λειτουργία της *ενέργειας* αντιπροσωπεύει την διαδικασία λήψης αποφάσεων του πράκτορα. Η λειτουργία της *όρασης* μπορεί να υλοποιηθεί με υλικό εξοπλισμό (hardware) όταν ο πράκτορας βρίσκεται στον φυσικό κόσμο, για παράδειγμα μπορεί να είναι μια κάμερα ή ένας αισθητήρας υπέρυθρων σε έναν κινητό ρομπότ. Όταν αναφερόμαστε σε έναν πράκτορα λογισμικού, ο αισθητήρας μπορεί να είναι εντολές του συστήματος που λαμβάνουν πληροφορίες σχετικά με το περιβάλλον του λογισμικού. Η έξοδος της λειτουργίας της *όρασης* είναι μια αντίληψη – μια

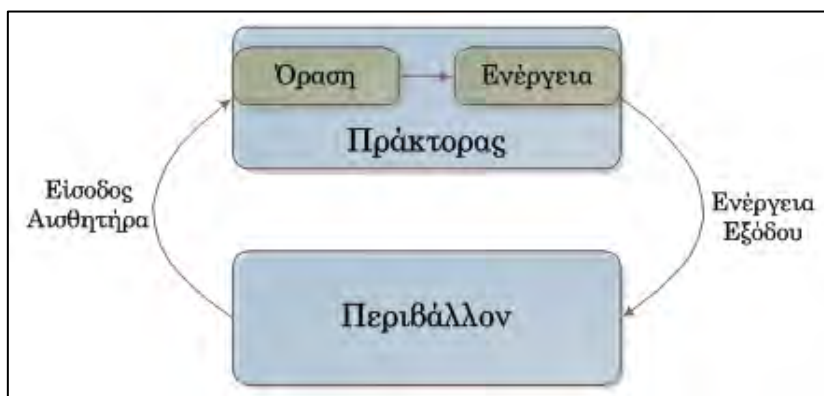
αντιληπτική είσοδος. Έστω ότι  $Per$  είναι ένα (μη κενό) σύνολο αντιλήψεων. Τότε η όραση είναι μια συνάρτηση

$$\text{Όραση: } E \rightarrow Per$$

που χαρτογραφεί τις καταστάσεις του περιβάλλοντος σε αντιλήψεις, και η ενέργεια είναι μια συνάρτηση

$$\text{Ενέργεια: } Per^* \rightarrow Ac$$

που χαρτογραφεί τις ακολουθίες των αντιλήψεων σε ενέργειες. Ένας πράκτορας  $Ag$  θεωρείται πλέον ότι είναι ένα ζευγάρι  $Ag = \langle \text{όραση, ενέργεια} \rangle$ , που αποτελείται από μια λειτουργία όρασης και μια λειτουργία ενέργειας [5].



Εικόνα 1.4 Υποσυστήματα αντίληψης και ενέργειας

### 1.6.3 Πράκτορες με Εσωτερική Κατάσταση

Μέχρι στιγμής έχουμε μοντελοποιήσει την λειτουργία της λήψης αποφάσεων ενός πράκτορα ως ακολουθίες καταστάσεων του περιβάλλοντος ή αντιλήψεις σε ενέργειες. Αυτό μας επιτρέπει να παριστάνουμε πράκτορες που η λήψη αποφάσεών τους επηρεάζεται από το ιστορικό. Παρ' όλα αυτά, είναι μια μη ενορατική αναπαράσταση που θα αντικατασταθεί με ένα ισοδύναμο, αλλά πιο φυσικό, σχέδιο. Εισάγουμε πλέον την έννοια των πρακτόρων που έχουν τη δυνατότητα να διατηρούν μια εσωτερική κατάσταση (Εικόνα 1.5).

Η λειτουργία της επιλογής ενέργειας προσδιορίζεται ως η χαρτογράφηση

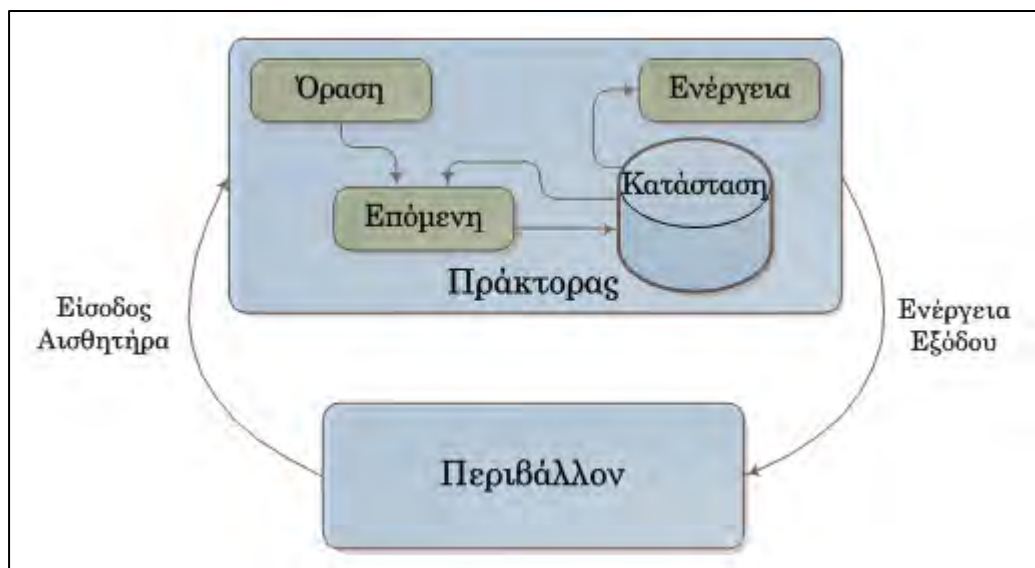
$$\text{Ενέργεια: } I \rightarrow Ac$$

από εσωτερικές καταστάσεις σε ενέργειες. Επιπλέον μια νέα λειτουργία εισάγεται, η *επόμενη*, που χαρτογραφεί μια εσωτερική κατάσταση και την αντίληψη γι' αυτήν την εσωτερική κατάσταση

$$\text{Επόμενη: } I \times Per \rightarrow I.$$

Είναι αξιοσημείωτο να τονιστεί το γεγονός ότι οι πράκτορες που βασίζονται στην εσωτερική κατάσταση δεν είναι πιο ισχυροί από τους τυπικούς πράκτορες που

παρουσιάστηκαν στην προηγούμενη παράγραφο. Στην πραγματικότητα, είναι πανομοιότυποι στην εκφραστική τους δύναμη – κάθε πράκτορας που βασίζεται στην εσωτερική κατάσταση μπορεί να μετατραπεί σε έναν τυπικό πράκτορα που είναι συμπεριφοριστικά ισοδύναμος [5].



Εικόνα 1.5 Πράκτορες που διατηρούν εσωτερική κατάσταση

## 1.7 Αρχιτεκτονικές Ευφύων Πρακτόρων με Εσωτερική Κατάσταση

Η αρχιτεκτονική των πρακτόρων είναι οι θεμελιώδεις μηχανισμοί που διέπουν τα αυτόνομα τμήμα που υποστηρίζουν την αποτελεσματική συμπεριφορά σε ένα περιβάλλον πραγματικού κόσμου, δυναμικό και ανοικτό. Στην πραγματικότητα, οι αρχικές προσεγγίσεις στον τομέα των υπολογιστών που βασίζονται σε τεχνολογία πρακτόρων επικεντρώθηκε στην ανάπτυξη αρχιτεκτονικών ευφύων πρακτόρων, και τα αρχικά βήματα καθόρισαν αρκετά στυλ αρχιτεκτονικής που διαρκούν μέχρι σήμερα. Κυμαίνονται από καθαρά αντιδραστικές (η συμπεριφοριστικές) αρχιτεκτονικές που λειτουργούν με έναν απλό τρόπο ερεθίσματος-απάντησης, όπως αυτά που βασίζονται στην αρχιτεκτονική υπαγωγής [7] στο ένα άκρο, σε πιο διαβουλευτικές αρχιτεκτονικές που λογικεύουν τις πράξεις τους, όπως αυτές που βασίζονται στο τρίπτυχο μοντέλο Πεποιθήσεις-Επιθυμίες-Προθέσεις (BDI, Belief-Desire-Intention) [8], στο άλλο άκρο. Μεταξύ αυτών των δύο βρίσκονται υβριδικές αρχιτεκτονικές, ή πολυεπίπεδες αρχιτεκτονικές, οι οποίες επιδιώκουν να ενώσουν την αντίδραση και την διαβούλευση έτσι ώστε να υιοθετήσουν τα πλεονεκτήματα και των δύο [9].

Οι κυριότερες προσεγγίσεις αρχιτεκτονικής για την σχεδίαση και υλοποίηση ευφύων πρακτόρων είναι οι εξής:

- **Ορθολογικοί Πράκτορες (Rational Agents):** Η λήψη των αποφάσεων ενεργειών γίνεται μέσω της λογικής επαγωγής.

- **Αντιδραστικοί Πράκτορες (Reactive Agents):** Οι αποφάσεις λαμβάνονται μέσω της άμεσης αντιστοιχίας μιας κατάστασης σε μια αντίστοιχη ενέργεια.
- **Πράκτορες που Βασίζονται σε Πεποιθήσεις-Επιθυμίες-Προθέσεις ή BDI Πράκτορες (Belief-Desire-Intention Agents):** Η λήψη αποφάσεων γίνεται μέσω ειδικών δομών που αναπαριστούν τις πεποιθήσεις, τις επιθυμίες και τις προθέσεις του κάθε πράκτορα.
- **Υβριδικό Πράκτορες (Hybrid Agents):** Αποτελούν ένα συνδυασμό αντιδραστικών πρακτόρων οι οποίοι διαθέτουν εσωτερική κατάσταση.
- **Πολυεπίπεδη Αρχιτεκτονική (Layered Agents):** Η λήψη αποφάσεων βασίζεται σε διαφορετικά επίπεδα λογισμικού, κάθε ένα εκ των οποίων αποτελεί μια άποψη του περιβάλλοντος σε ένα διαφορετικό επίπεδο αφαίρεσης.
- **Κινητοί Πράκτορες (Mobile Agents):** Οι πράκτορες έχουν τη δυνατότητα να μετακινούνται είτε στον φυσικό χώρο είτε μεταξύ υπολογιστικών συστημάτων ενός δικτύου [2].

### 1.7.1 Ορθολογικοί Πράκτορες

Σύμφωνα με την παραδοσιακή προσέγγιση της ανάπτυξης τεχνητών ευφυών συστημάτων, η ευφυής συμπεριφορά ενός συστήματος προκύπτει όταν τροφοδοτηθεί με τα εξής:

- μια συμβολική αναπαράσταση του περιβάλλοντός του
- μια συμβολική αναπαράσταση της επιθυμητής συμπεριφοράς του
- ένα σύστημα συντακτικής διαχείρισης των προαναφερθέντων αναπαραστάσεων.

Η μοντελοποίηση τέτοιου είδους συστημάτων γίνεται σύμφωνα με παραδοσιακές μαθηματικές τεχνικές, όπως είναι οι λογικές φόρμουλες για τη συμβολική αναπαράσταση και τα λογικά συμπεράσματα ή οι αποδείξεις θεωρημάτων για την συντακτική διαχείριση. Η ανάπτυξη πρακτόρων που ενεργούν σύμφωνα με αποδείξεις θεωρημάτων υπήρξε θετική, με αποτέλεσμα να δημιουργηθούν φορμαλιστικές θεωρίες που περιγράφουν τη συμπεριφορά ενός ευφυή πράκτορα. Μια τέτοια θεωρία, για παράδειγμα, περιγράφει τη διαδικασία δημιουργίας στόχων ενός πράκτορα έτσι ώστε να ανταποκριθεί στο σκοπό σχεδίασης του και τη διαδικασία λήψης αποφάσεων προκειμένου να πετύχει αυτούς τους στόχους. Σύμφωνα με την παραπάνω θεωρία επομένως καθορίζονται οι προδιαγραφές λειτουργίας (συμπεριφοράς) ενός πράκτορα. Σύμφωνα με την παραδοσιακή προσέγγιση υλοποίησης ενός τέτοιου συστήματος συμπεριλαμβάνει διαδικασία βελτίωσης, σύμφωνα με την οποία οι προδιαγραφές βελτιώνονται μέσω προοδευτικών και διακριτών σταδίων μέχρι την επίτευξη ενός συνόλου υλοποιήσιμων προδιαγραφών. Από την άλλη πλευρά, η υλοποίηση πρακτόρων που ενεργούν σύμφωνα με αποδείξεις θεωρημάτων, δεν περιλαμβάνει κάποια διαδικασία βελτίωσης, αλλά η θεωρία αντιμετωπίζεται ως ένα σύνολο εκτελέσιμων προδιαγραφών, οι οποίες εκτελούνται άμεσα και έχουν ως αποτέλεσμα την ανάπτυξη της συμπεριφοράς του πράκτορα [2].

Το πλεονέκτημα αυτής της προσέγγισης είναι ότι η ανθρώπινη γνώση είναι συμβολική επομένως η κωδικοποίηση είναι πιο εύκολη καθώς και ότι μπορεί να κατασκευαστεί έτσι ώστε να είναι υπολογιστικά ολοκληρωμένη, γεγονός που διευκολύνει του ανθρώπους στην κατανόηση της λογικής. Τα μειονεκτήματα είναι ότι είναι δύσκολο να μεταφραστεί ο πραγματικός κόσμος σε μια ακριβή, ανάλογα συμβολική αναπαράσταση, και το ότι η συμβολική αναπαράσταση και διαχείριση μπορεί να χρειαστεί σημαντικό χρόνο για να εκτελεσθεί με αποτελέσματα που είναι συχνά διαθέσιμα πολύ αργά ώστε να είναι χρήσιμα [9].

## 1.7.2 Αντιδραστικοί Πράκτορες

Εξαιτίας των μειονεκτημάτων που παρουσιάζουν οι πράκτορες με εσωτερική κατάσταση αναπτύχθηκαν οι αντιδραστικοί πράκτορες. Κυρίαρχο χαρακτηριστικό αυτών των πρακτόρων είναι η απουσία μιας εσωτερικής συμβολικής αναπαράστασης του περιβάλλοντός τους, με αποτέλεσμα η συμπεριφορά τους να βασίζεται μόνο σε αντιδράσεις στα ερεθίσματα που λαμβάνουν από το περιβάλλον τους. Οι αντιδραστικοί πράκτορες αποτελούν μια εναλλακτική προσέγγιση την ανάπτυξη ευφυών πρακτόρων, σύμφωνα με την οποία η ευφυής συμπεριφορά των τεχνητών συστημάτων προκύπτει από το συνδυασμό απλών μερών που αλληλεπιδρούν μεταξύ τους και όχι από πολύπλοκους χειρισμούς συμβόλων.

Η αρχιτεκτονική των αντιδραστικών πρακτόρων διαθέτει δύο κύρια χαρακτηριστικά, ότι οι αποφάσεις για τις ενέργειες του πράκτορα λαμβάνονται υπό τη συνεχή διερεύνηση του περιβάλλοντος και ότι πολλές συμπεριφορές μπορεί να ενεργοποιηθούν ταυτόχρονα. Στην πρώτη περίπτωση ο πράκτορας λαμβάνει ως είσοδο τα συνεχή ερεθίσματα του περιβάλλοντος και η ενέργεια εξόδου είναι αποτέλεσμα μιας συνάρτησης που αντιστοιχίζει την τρέχουσα είσοδο σε μια ενέργεια:

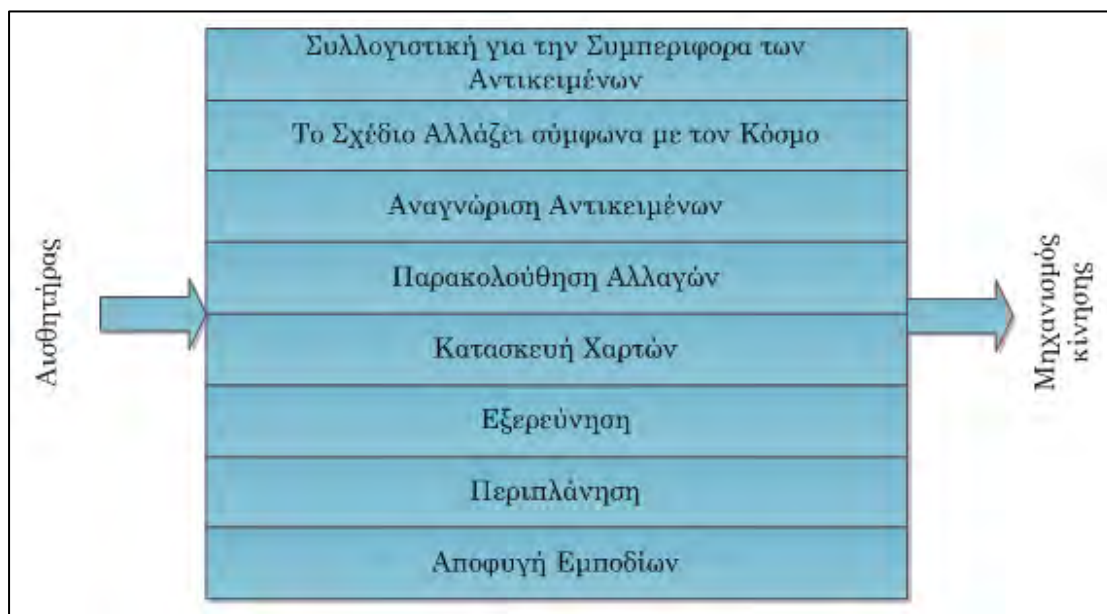
κατάσταση → ενέργεια

Στην δεύτερη περίπτωση μπορεί να προκύψουν περισσότερες από μια συναρτήσεις σε μια κατάσταση οι οποίες θα δώσουν την επόμενη ενέργεια. Έτσι προκύπτει η ανάγκη ύπαρξης ενός μηχανισμού επιλογής ενεργειών, ο οποίος μπορεί να λαμβάνει υπόψη κάποια κριτήρια, όπως είναι το κόστος, η ταχύτητα ή το ποσοστό κέρδους, ή γίνεται σύμφωνα με κάποιο ιεραρχικό μοντέλο [2].

Ίσως η πιο γνωστή αρχιτεκτονική αντιδραστικών πρακτόρων είναι η αρχιτεκτονική υπαγωγής του Brooks (1991) [7]. Η βασική ιδέα, πάνω στην οποία πραγματοποιήθηκε η συγκεκριμένη αρχιτεκτονική, είναι ότι μια ευφυής συμπεριφορά μπορεί να δημιουργηθεί χωρίς ρητές αναπαραστάσεις και αφηρημένη λογική που παρέχεται από τις συμβολικές τεχνικές της τεχνητής νοημοσύνης και ότι η ευφυΐα είναι μια αναδυόμενη ιδιότητα συγκεκριμένων πολύπλοκων συστημάτων. Η αρχιτεκτονική της υπαγωγής ορίζει επίπεδα πεπερασμένων μηχανών κατάστασης που είναι συνδεδεμένα με αισθητήρες που μεταδίδουν πληροφορίες σε πραγματικό χρόνο. Στην Εικόνα 1.6 βλέπουμε ένα παράδειγμα αυτής της αρχιτεκτονικής. Αυτά τα επίπεδα σχηματίζουν μια ιεραρχία συμπεριφορών στην οποία τα χαμηλότερα επίπεδα έχουν λιγότερο έλεγχο σε σχέση με αυτές που βρίσκονται πιο πάνω στην στοίβα, με αποτέλεσμα η λήψη



αποφάσεων να επιτυγχάνεται μέσω στοχευμένων συμπεριφορών. Οι πράκτορες που είναι σχεδιασμένοι με αυτήν την αρχιτεκτονική αντιλαμβάνονται τις συνθήκες και ενεργούν, αλλά δεν σχεδιάζουν.



**Εικόνα 1.6** Αρχιτεκτονική υπαγωγής για την καθοδήγηση ενός ρομπότ

Το πλεονέκτημα αυτής της προσέγγισης είναι η καλύτερη απόδοση τους σε ένα δυναμικό περιβάλλον, δηλαδή αντιδρούν πιο γρήγορα, αλλά δεν αιτιολογούν καλύτερα. Επιπλέον, ο σχεδιασμός τους είναι πιο απλός σε σχέση με τους πράκτορες που βασίζονται λογική. Παρόλα αυτά, το γεγονός ότι αυτού του είδους οι πράκτορες δεν χρησιμοποιούν μοντέλα του περιβάλλοντός τους οδηγεί σε κάποια μειονεκτήματα. Στην πραγματικότητα, τα δεδομένα που λαμβάνουν οι πράκτορες μέσω των αισθητήρων τους μπορεί να μην είναι αρκετά έτσι ώστε να προσδιορίσουν την κατάλληλη ενέργεια και η έλλειψη της κατάστασης του πράκτορα καθιστά αδύνατο τον σχεδιασμό πρακτόρων που μαθαίνουν από τις εμπειρίες τους. Ακόμη, δεδομένου ότι οι πράκτορες ενεργούν με βάση την αλληλεπίδραση μεταξύ των διαφορετικών συμπεριφορών, είναι πολύ δύσκολο να κατασκευαστούν αντιδραστικοί πράκτορες που θα ολοκληρώνουν συγκεκριμένα καθήκοντα, και συγκεκριμένα, όταν τέτοιου είδους πράκτορες πρέπει να υλοποιηθούν μέσω ενός μεγάλου αριθμού συμπεριφορών [9].

### 1.7.3 Πράκτορες που Βασίζονται σε Πεποιθήσεις – Επιθυμίες – Προθέσεις

Η αρχιτεκτονική πεποιθήσεις – επιθυμίες – προθέσεις (Belief-Desire-Intention, BDI) [8], είναι η πιο δημοφιλής αρχιτεκτονική πρακτόρων η οποία προέρχεται από την φιλοσοφία. Η νοητική κατάσταση των πεποιθήσεων, των επιθυμιών και των προθέσεων αναπαρίσταται χρησιμοποιώντας τροπική λογική. Υπάρχουν πολλά συστήματα πρακτόρων που έχουν υλοποιηθεί με βάση το μοντέλο BDI (π.χ. PRS [10],

JAM [11], JACK [12], dMARS [13] και Jadex [14]) με μια μεγάλη ποικιλία εφαρμογών που αποδεικνύουν την βιωσιμότητά του [9].

Οι πράκτορες BDI διαθέτουν μια πολύπλοκη αναπαράσταση του περιβάλλοντος με βάση την οποία σχεδιάζουν την επίτευξη των στόχων τους. Η εσωτερική τους κατάσταση αποτελείται από πεποιθήσεις, επιθυμίες, προθέσεις, στόχους και σχέδια. Συγκεκριμένα:

- Οι **πεποιθήσεις** αφορούν την άποψη και τη γνώση που διαθέτουν οι πράκτορες για το περιβάλλον τους και οι οποίες ενδέχεται να είναι εσφαλμένες.
- Οι **επιθυμίες** εκφράζουν την άποψη του πράκτορα για τις μελλοντικές καταστάσεις του περιβάλλοντος του, δηλαδή αν μια συγκεκριμένη κατάσταση θα είναι επιθυμητή ή όχι. Δεν αφορά τον καθορισμό αν μια επιθυμητή κατάσταση είναι εφικτό να επιτευχθεί, ενώ υπάρχει και η δυνατότητα σύγκρουσης μεταξύ των επιθυμητών καταστάσεων.
- Οι **στόχοι** είναι ένα υποσύνολο των επιθυμιών, για την επίτευξη των οποίων ενεργεί ο πράκτορας. Οι στόχοι σε αντίθεση πρέπει να είναι εφικτοί και να μην παρατηρούνται συγκρούσεις μεταξύ τους.
- Οι **προθέσεις** αποτελούν ένα υποσύνολο των στόχων, τις οποίες ο πράκτορας προσπαθεί να επιτύχει τη δεδομένη χρονική στιγμή. Καθώς δεν είναι δυνατό να επιτευχθούν όλοι οι στόχοι ενός πράκτορα ταυτόχρονα, σύμφωνα με κάποια κριτήρια ιεράρχησης επιλέγεται ένα υποσύνολό τους το οποίο και αποτελεί τις προθέσεις.
- Τα **σχέδια** αποτελούν τους διάφορους τρόπους μέσω των οποίων ο πράκτορας θα επιτύχει τις προθέσεις του [2].

Σύμφωνα με τις προσλαμβάνουσες αντιλήψεις από τον κόσμο (P) και το σύνολο των ενεργειών (A), η αρχιτεκτονική των πρακτόρων BDI διαθέτει τα εξής μέρη:

- B: σύνολο πεποιθήσεων
- D: σύνολο επιθυμιών
- I: σύνολο προθέσεων
- Συνάρτηση αναθεώρησης πεποιθήσεων:

$$B \times P \rightarrow B$$

- Συνάρτηση αναθεώρησης επιθυμιών:

$$B \times I \rightarrow D$$

- Συνάρτηση αναθεώρησης προθέσεων:

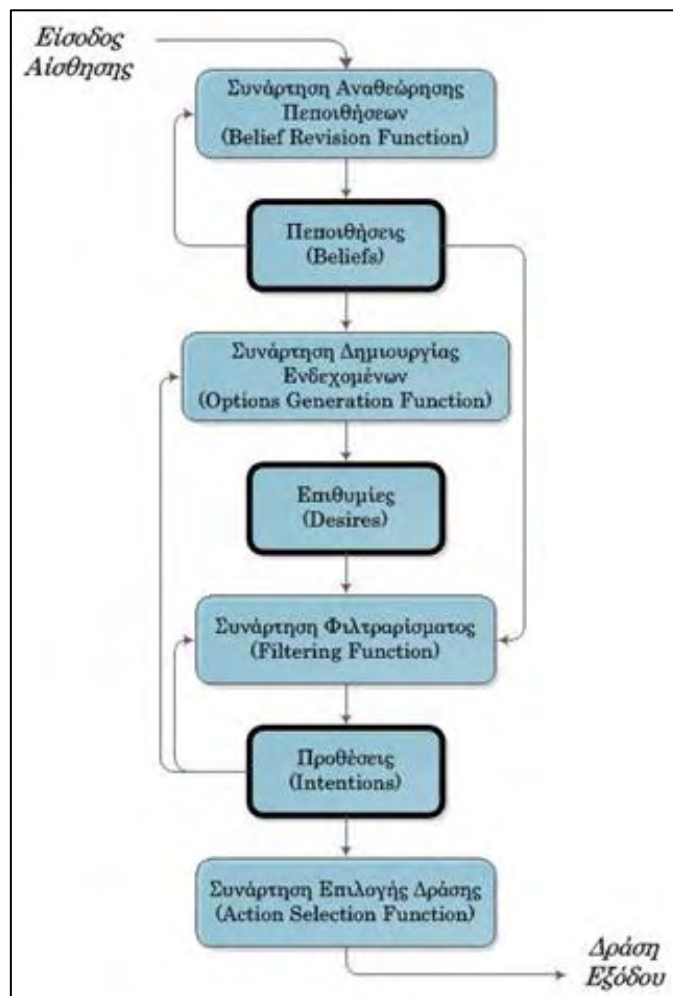
$$B \times D \times I \rightarrow I$$

- Συνάρτηση επιλογής ενεργειών:

$$I \rightarrow A$$

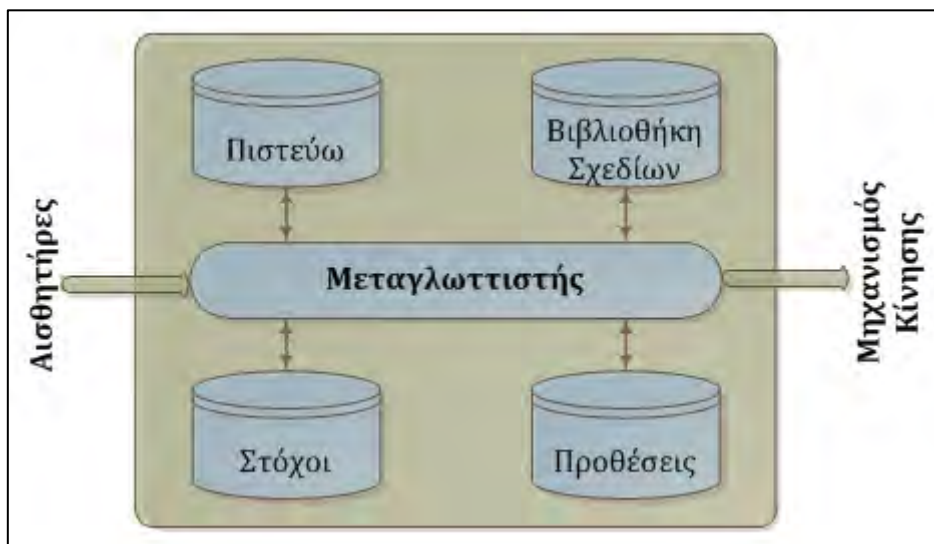
Η διαδικασία πρακτικού συλλογισμού των πρακτόρων BDI διαθέτει επτά κύριες διαδικασίες, όπως απεικονίζεται στην Εικόνα 1.7, οι οποίες είναι η εξής:

- I. Ένα σύνολο προσωρινών πεποιθήσεων, οι οποίες αντιπροσωπεύουν τις διαθέσιμες πληροφορίες για την τρέχουσα κατάσταση του περιβάλλοντος.
- II. Η Συνάρτηση Αναθεώρησης Πεποιθήσεων λαμβάνει ως είσοδο τις αντιλήψεις από το περιβάλλον καθώς και το σύνολο των προσωρινών πεποιθήσεων με βάση τα οποία παράγει ένα νέο σύνολο πεποιθήσεων.
- III. Η Συνάρτηση Δημιουργίας Ενδεχομένων εξάγει τις επιθυμίες του πράκτορα σύμφωνα με τις προσωρινές πεποιθήσεις αλλά και τις προθέσεις του πράκτορα.
- IV. Το σύνολο προσωρινών ενδεχομένων, που διαθέτει όλες τις πιθανές ακολουθίες ενεργειών που έχει διαθέσιμες ο πράκτορας.
- V. Η Συνάρτηση Φιλτραρίσματος αναπαριστά την διαδικασία συλλογισμού του πράκτορα σύμφωνα με την οποία αποφασίζει τις νέες προθέσεις του. Η διαδικασία συλλογισμού λαμβάνει υπόψη τις προσωρινές επιθυμίες, τις προθέσεις αλλά και τις πεποιθήσεις του πράκτορα.
- VI. Το σύνολο προσωρινών προθέσεων περιλαμβάνει τους τρέχοντες στόχους του πράκτορα.
- VII. Η Συνάρτηση Επιλογής Δράσης αποφασίζει τις ενέργειες που θα εκτελέσει ο πράκτορας με βάση τρέχουσες προθέσεις [2].



**Εικόνα 1.7** Αρχιτεκτονική BDI (Πεποιθήσεις-Επιθυμίες-Προθέσεις)

Μια από τις πιο γνωστές αρχιτεκτονικές BDI είναι το σύστημα PRS (Procedural Reasoning System) που βασίζεται στις τέσσερις δομές δεδομένων κλειδιά: πεποιθήσεις, επιθυμίες, προθέσεις και σχέδια, και σε έναν μεταγλωττιστή (Εικόνα 1.8). Στο σύστημα PRS, οι πεποιθήσεις αντιπροσωπεύουν τις πληροφορίες που έχει ένας πράκτορας για το περιβάλλον του, που μπορεί να είναι είτε ατελείς είτε λανθασμένες. Οι επιθυμίες αντιπροσωπεύουν τα καθήκοντα που έχουν ανατεθεί στο συγκεκριμένο πράκτορα και επομένως αντιστοιχούν στους στόχους που θα πρέπει να επιτύχει. Οι προθέσεις αντιπροσωπεύουν τις επιθυμίες που ο πράκτορας έχει δεσμευθεί να πραγματοποιήσει. Τα σχέδια προσδιορίζουν την σειρά ενεργειών που μπορεί να ακολουθήσει ο πράκτορας έτσι ώστε να επιτύχει τις προθέσεις του. Αυτές οι τέσσερις δομές δεδομένων διαχειρίζονται από τον μεταγλωττιστή του πράκτορα που είναι υπεύθυνος για την ενημέρωση των πεποιθήσεων του πράκτορα από τις παρατηρήσεις που έγιναν στο περιβάλλον, την παραγωγή νέων επιθυμιών (καθηκόντων) στην βάση των νέων πεποιθήσεων, και την επιλογή από την συλλογή των τωρινών ενεργών επιθυμιών ένα υποσύνολο που θα λειτουργήσει ως προθέσεις. Τέλος, ο μεταγλωττιστής πρέπει να επιλέξει να πραγματοποιήσει μια ενέργεια από την βάση των τρεχόντων προθέσεων του πράκτορα και την διαδικαστική του γνώση [9].



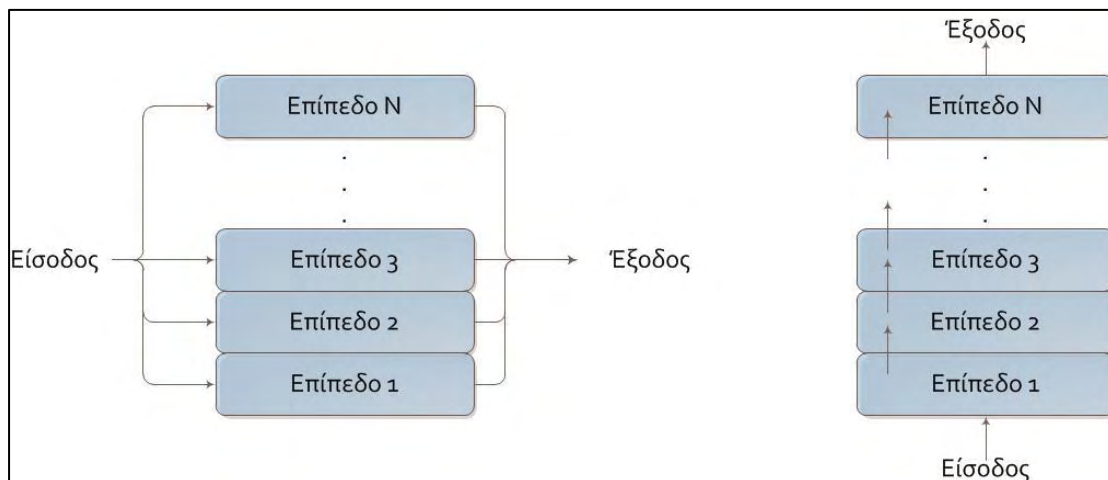
Εικόνα 1.8 Αρχιτεκτονική PRS

### 1.7.4 Υβριδικοί Πράκτορες

Η αρχιτεκτονική των υβριδικών πρακτόρων βασίζεται στον συνδυασμό των αρχιτεκτονικών των πρακτόρων με εσωτερική κατάσταση και των αντιδραστικών πρακτόρων. Η υβριδική αρχιτεκτονική διαθέτει δύο επίπεδα, όπου το ένα είναι υπεύθυνο για την αντιδραστική συμπεριφορά του πράκτορα και το άλλο για τη συμπεριφορά με εσωτερική κατάσταση.

Στους υβριδικούς πράκτορες η ροή ελέγχου είναι είτε οριζόντια ή κάθετη. Στην πρώτη περίπτωση όλα τα επίπεδα είναι συνδεδεμένα και στους αισθητήρες εισόδου και στους μηχανισμούς ενεργειών, ενώ στη δεύτερη περίπτωση μόνο ένα επίπεδο είναι

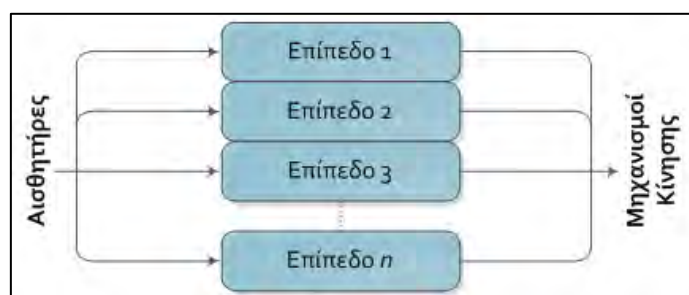
συνδεδεμένο στους αισθητήρες και ένα άλλο στους μηχανισμούς ενεργειών (Εικόνα 1.9) [2].



**Εικόνα 1.9** Οριζόντια (αριστερά) και κάθετη (δεξιά) ροή ελέγχου σε υβριδικούς πράκτορες

### 1.7.5 Πολυεπίπεδη Αρχιτεκτονική

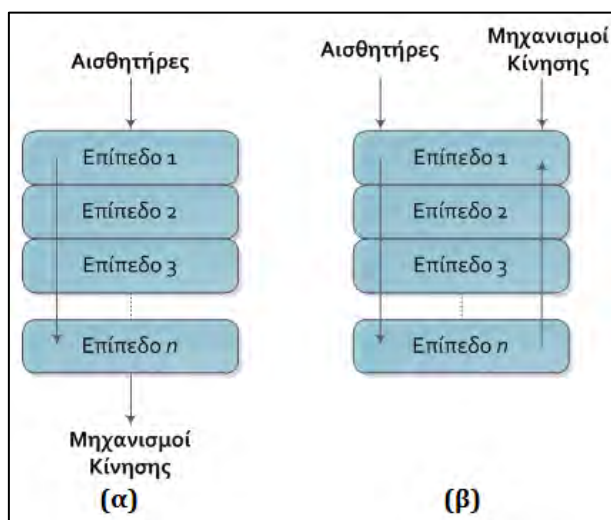
Οι πολυεπίπεδες αρχιτεκτονικές επιτρέπουν και την αντιδραστική και την διαβουλευτική συμπεριφορά. Για να διευκολύνουν αυτήν την ελαστικότητα, τα υποσυστήματα που τοποθετούνται ως τα επίπεδα μιας ιεραρχίας χρησιμοποιούνται για να φιλοξενήσουν και τους δύο αυτούς τύπους πρακτόρων. Υπάρχουν δύο είδη ροής ελέγχου μέσα στην πολυεπίπεδη αρχιτεκτονική: η οριζόντια (Ferguson, 1991) και η κάθετη διαστρωμάτωση (Muller et al., 1995). Στην οριζόντια, τα επίπεδα είναι άμεσα συνδεδεμένα με τον αισθητήρα εισόδου και την έξοδο ενεργειών (Εικόνα 1.10), και στην οποία ουσιαστικά το κάθε επίπεδο λειτουργεί ως ένας πράκτορας. Το κύριο πλεονέκτημα αυτής της αρχιτεκτονικής είναι η απλότητα του σχεδιασμού καθώς αν ο πράκτορας χρειαστεί  $n$  διαφορετικούς τύπους συμπεριφορών, τότε η αρχιτεκτονική χρειάζεται μόνο  $n$  επίπεδα.



**Εικόνα 1.10** Αρχιτεκτονική οριζόντιας διαστρωμάτωσης

Ωστόσο, δεδομένου ότι κάθε επίπεδο είναι ένας πράκτορας, οι ενέργειες τους μπορεί να μην συνάδουν δημιουργώντας την ανάγκη μιας λειτουργίας μεσολαβητή για τον έλεγχο αυτών των ενεργειών. Μια άλλη πολυπλοκότητα είναι ο μεγάλος αριθμός των πιθανών αλληλεπιδράσεων μεταξύ των οριζόντιων επιπέδων -  $m^n$  (όπου  $m$  είναι ο αριθμός των

ενεργειών ανά επίπεδο). Μια κάθετη αρχιτεκτονική διαστρωμάτωση απαλείφει κάποια από τα προαναφερθέντα προβλήματα καθώς ο αισθητήρας εισόδου και η έξοδος ενεργειών αντιμετωπίζονται από το πολύ ένα επίπεδο ο καθένας (δημιουργώντας μη ασυμβίβαστες προτάσεις ενεργειών). Η κάθετη αρχιτεκτονική διαστρωμάτωσης μπορεί να υποδιαιρεθεί σε αρχιτεκτονικές ελέγχου μονού περάσματος (one-pass) και διπλού περάσματος (two-pass). Στην αρχιτεκτονική one-pass, ο έλεγχος ρέει από το αρχικό επίπεδο που δέχεται τα δεδομένα εισόδου του αισθητήρα στο τελικό επίπεδο που παράγει την ενέργεια εξόδου (Εικόνα 1.11α). Στην αρχιτεκτονική two-pass, τα δεδομένα διαρρέουν όλη την ακολουθία των επιπέδων και, στην συνέχεια, ο έλεγχος διαρρέει τα επίπεδα προς τα πίσω (Εικόνα 1.11β). Τα κύριο πλεονέκτημα της κάθετης διαστρωμάτωσης είναι ότι η αλληλεπίδραση μεταξύ των επιπέδων μειώνεται σημαντικά στο  $m^2(n-1)$ . Το κύριο μειονέκτημα είναι ότι η αρχιτεκτονική εξαρτάται από όλα τα επίπεδα και επομένως δεν είναι ανεκτική σε σφάλματα, έτσι αν ένα επίπεδο αποτύχει, τότε ολόκληρο το σύστημα αποτυγχάνει [9].



**Εικόνα 1.11** Ροή δεδομένων και ελέγχου στην κάθετη διαστρωμάτωση **(α)** απλού φιλτραρίσματος και **(β)** διπλού φιλτραρίσματος

### 1.7.6 Κινητοί Πράκτορες

Οι κινητοί πράκτορες αποτελούν πρόγραμμα λογισμικού και έχουν τη δυνατότητα να μετακινούνται αυτόνομα από κόμβο σε κόμβο σε ένα ετερογενές δίκτυο έτσι ώστε να πραγματοποιήσουν κάποιον υπολογισμό εκ μέρους ενός χρήστη διαθέτοντας χαρακτηριστικά όπως η αυτονομία, η κοινωνικές δεξιότητες, η προσαρμοστικότητα και η αντιδραστικότητα. Ορίζονται ως αντικείμενα που διαθέτουν συμπεριφορά, κατάσταση και τοποθεσία. Οι ενέργειές τους καθορίζονται από το είδος της εφαρμογής κυμαίνονται από την πραγματοποίηση online αγορών μέχρι και τον έλεγχο συσκευών πραγματικού χρόνου. Μπορούν να εκτελεστούν οποιαδήποτε στιγμή ή να αναστείλουν τη λειτουργία τους, να μετακινηθούν εύκολα κατά μήκος του δικτύου και να συνεχίσουν την εκτέλεσή τους σε έναν άλλο host. Όταν ένας κινητός πράκτορας αποφασίζει να μετακινηθεί, αποθηκεύει την κατάστασή του, μεταφέρει αυτά τα δεδομένα στον νέο host και αρχίζει ξανά την εκτέλεση από το σημείο που σταμάτησε.

Οι κινητοί πράκτορες χρησιμοποιούνται σε μια μεγάλη περιοχή εφαρμογών όπως είναι η διαχείριση και παρακολούθηση δικτύων, η αναζήτηση και το φιλτράρισμα πληροφοριών και η ανίχνευση εισβολών.

Ένας κινητός πράκτορας αποτελείται από τον κώδικα και την κατάσταση εκτέλεσης του προγράμματος (τις τρέχουσες τιμές των μεταβλητών, την επόμενη εντολή που θα εκτελεσθεί). Αρχικά, ένας κινητός πράκτορας βρίσκεται σε ένα υπολογιστικό σύστημα που καλείται μητρικό σύστημα. Ο πράκτορας στη συνέχεια αποστέλλεται για εκτέλεση σε ένα απομακρυσμένο υπολογιστικό σύστημα που καλείται host (ή πλατφόρμα ή server) κινητών πρακτόρων. Όταν ο κινητός πράκτορας αποστέλλεται τότε ολόκληρος ο κώδικας του πράκτορα και η κατάσταση εκτέλεσης μεταφέρεται στον επόμενο host. Ο κάθε host παρέχει ένα κατάλληλο περιβάλλον εκτέλεσης για τον κινητό πράκτορα και του διαθέτει πόρους, όπως είναι η μνήμη και η επεξεργαστική ισχύ, για να εκτελέσει τις λειτουργίες του. Εφόσον ο κινητός πράκτορας εκτελέσει τις λειτουργίες του σε κάποιον host, τότε ο κινητός πράκτορας μετακινείται σε ένα άλλο υπολογιστικό σύστημα. Εφόσον μαζί του μεταφέρονται και οι πληροφορίες εκτέλεσης, οι κινητοί πράκτορες ανακτούν την εκτέλεση από το σημείο που σταμάτησαν στον προηγούμενο host χωρίς να χρειάζεται να ξεκινήσουν την εκτέλεση από την αρχή. Αυτή η διαδικασία συνεχίζεται μέχρι ο κινητός πράκτορας να επιστρέψει στο μητρικό σύστημα εφόσον έχει ολοκληρώσει την εκτέλεση στο τελευταίο σύστημα της διαδρομής του.

Η βασική αρχιτεκτονική των κινητών πρακτόρων αποτελείται από έναν πράκτορα και την πλατφόρμα εκτέλεσής του. Κάθε πράκτορας μπορεί να χρειαστεί να μετακινηθεί σε διάφορες πλατφόρμες μέχρι να ολοκληρώσει την εργασία του [15].

---

# Κεφάλαιο 2

## Πολυπρακτορικά Συστήματα

---

### 2.1 Εισαγωγή

Υπάρχει ένα δημοφιλές ρητό στην κοινότητα των πολυπρακτορικών συστημάτων:

«Δεν υπάρχει τέτοιο πράγμα όπως ένα σύστημα ενός μοναδικού πράκτορα».

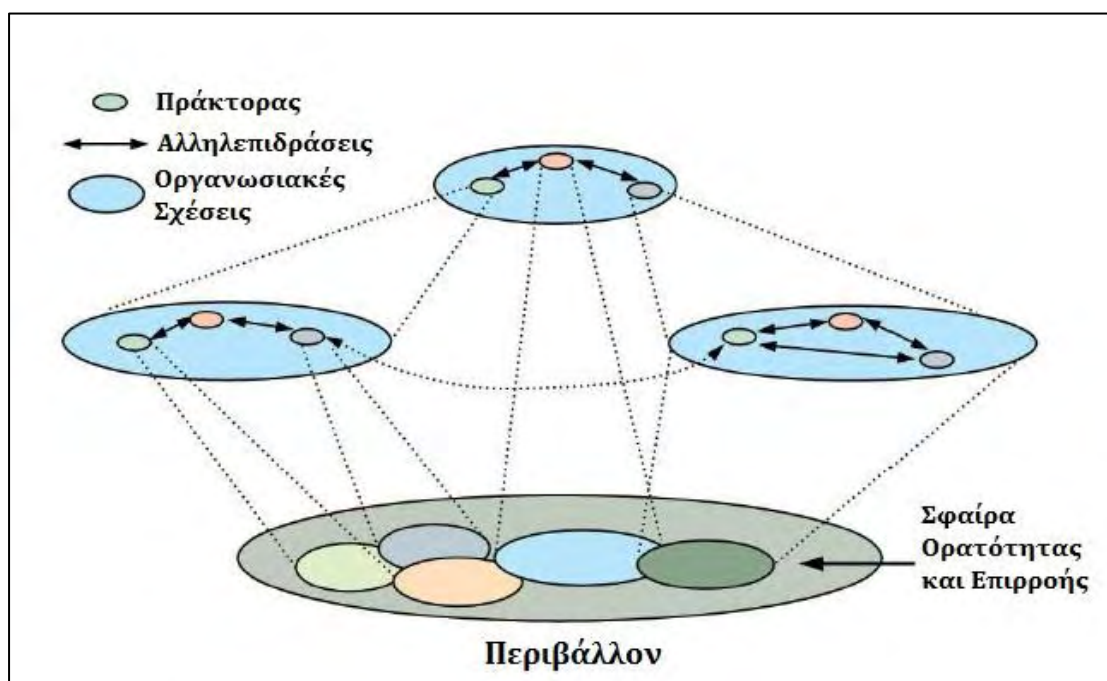
Η παραπάνω ρήση τονίζει, ίσως με περισσή έμφαση, ότι είναι πλέον συχνό φαινόμενο να υλοποιούνται συστήματα τα οποία αποτελούνται από πράκτορες που αλληλεπιδρούν. Το να απαρτίζεται ένα υπολογιστικό σύστημα από αλληλεπιδρώντα υποσυστήματα δεν αποτελεί καινοτομία στο χώρο της επιστήμης υπολογιστών. Εκείνο που είναι πραγματικά νέο είναι ότι η αλληλεπίδραση στα πολυπρακτορικά συστήματα είναι δυναμική και όχι στατικά καθορισμένη, δίνοντας τη δυνατότητα στο σύστημα να συμπεριφερθεί με σωστό τρόπο ακόμη και σε περιπτώσεις που δεν είχαν αρχικά προβλεφθεί. Έτσι η τεχνολογία πρακτόρων αποτελεί ουσιαστικά μια νέα σχολή προγραμματισμού με την έννοια ότι ο πράκτορας αποτελεί τη βασική μονάδα τόσο στο σχεδιασμό όσο και στην υλοποίηση του συστήματος. Αν και η δημιουργία γλωσσών προγραμματισμού αποκλειστικά για χρήση σε πράκτορες είναι ακόμη σε ερευνητικό στάδιο, εντούτοις φαίνεται να μπορεί να αποτελέσει ικανοποιητικό εργαλείο για τη διαχείριση της πολυπλοκότητας των συστημάτων λογισμικού [5].

Ένα πολυπρακτορικό σύστημα αποτελεί ένα σύστημα που σχεδιάστηκε και υλοποιήθηκε ως ένα σύνολο πολλών πρακτόρων που αλληλεπιδρούν, δηλαδή συνεργάζονται, συντονίζονται και διαπραγματεύονται. Τα πολυπρακτορικά συστήματα σχηματίζουν ένα δίκτυο από 'χαλαρά' συνδεδεμένους πράκτορες οι οποίοι ενεργούν



μαζί έτσι ώστε να επιλύσουν προβλήματα που δεν είναι δυνατό να επιλυθούν με τη γνώση ενός μόνο πράκτορα (Εικόνα 2.1). Επομένως, ένα πολυπρακτορικό σύστημα μπορεί να διαθέτει τους εξής στόχους:

- Επίλυση πολύπλοκων προβλημάτων που απαιτούν τη συνεργασία πολλών πρακτόρων.
- Επίλυση καταναμημένων προβλημάτων στα οποία είναι απαραίτητη η συλλογή πληροφοριών από πολλές καταναμημένες πηγές, όπως για παράδειγμα στα δίκτυα αισθητήρων.
- Επίλυση προβλημάτων με καταναμημένη γνώση.
- Διασύνδεση και λειτουργία ήδη υπάρχοντων συστημάτων.



**Εικόνα 2.1** Τυπική δομή ενός πολυπρακτορικού συστήματος

Σε ένα πολυπρακτορικό σύστημα οι ατομικοί πράκτορες έχουν τη δυνατότητα να εργάζονται αυτόνομα ή να συνεργάζονται. Στην πρώτη περίπτωση λειτουργούν ανταλλάσσοντας πληροφορίες ή και υπηρεσίες, προσπαθώντας έτσι να επιτύχουν τους ατομικούς τους στόχους. Στη δεύτερη περίπτωση, οι πράκτορες συνεργάζονται έτσι ώστε επιλύοντας επιμέρους προβλήματα να επιτευχθεί ο απώτερος στόχος ολόκληρου του συστήματος. Επομένως, το περιβάλλον ενός πολυπρακτορικού συστήματος είναι, στη γενική περίπτωση, ανοικτό όπου οι πράκτορες μπορούν να δρουν, να συνεργάζονται και να διαπραγματεύονται χωρίς κεντρικό σχεδιασμό για την επίτευξη των στόχων τους.

Για τη συνεργασία, τον συντονισμό και την πραγματοποίηση διαπραγματεύσεων μεταξύ των πρακτόρων απαιτείται η επικοινωνία μεταξύ των πρακτόρων μέσω της χρήσης κάποιας γλώσσας επικοινωνίας. Σύμφωνα με τους Huhns και Stephens [16] «ο συντονισμός είναι η δυνατότητα ενός συστήματος πρακτόρων να φέρουν εις πέρας ενέργειες μέσα σε ένα κοινό περιβάλλον». Έτσι, οι πράκτορες έχουν τη δυνατότητα να συντονίζουν τις ενέργειές τους είτε για να επιτύχουν έναν κοινό στόχο ή για να

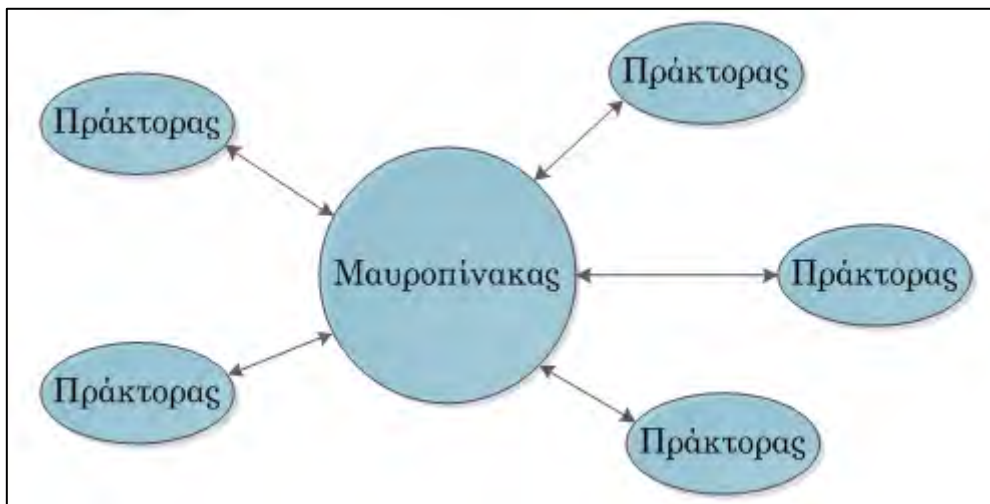
επιτύχουν τους προσωπικούς τους ανεξάρτητους στόχους. Κάποια χαρακτηριστικά ενός πολυπρακτορικού συστήματος είναι τα εξής:

- Κανένας πράκτορας δεν έχει πλήρη γνώση και πληροφορία του περιβάλλοντος.
- Το σύστημα λειτουργεί χωρίς την ύπαρξη ενός κεντρικού ελέγχου.
- Τα δεδομένα είναι κατά βάση κατανεμημένα.
- Οι υπολογισμοί πραγματοποιούνται με τρόπο ασύγχρονο [2].

## 2.2 Επικοινωνία Πρακτόρων

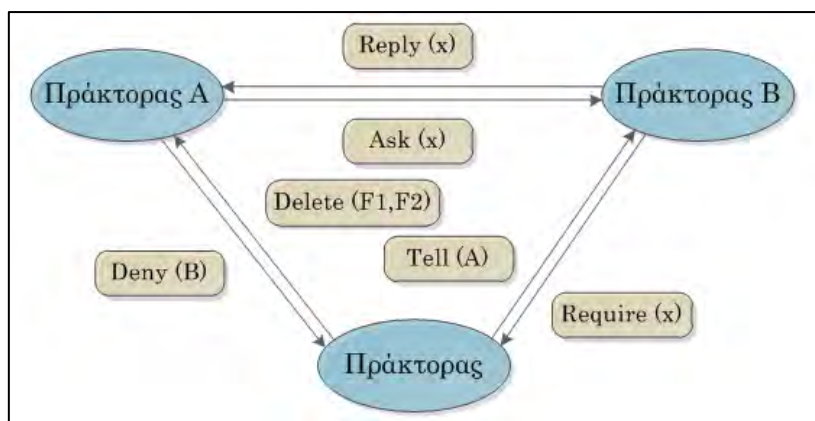
Ως επικοινωνία ορίζεται η δυνατότητα των πρακτόρων ενός πολυπρακτορικού συστήματος να ανταλλάξουν πληροφορίες με απώτερο σκοπό τον συντονισμό των ενεργειών τους. Η επικοινωνία μεταξύ των πρακτόρων μπορεί να επιτευχθεί με δύο τρόπους, με την άμεση ανταλλαγή μηνυμάτων και με τη χρήση ενός κοινού σημείου συγκέντρωσης δεδομένων που καλείται μαυροπίνακας. Η υλοποίηση της επικοινωνίας διαχωρίζεται σε τρία διαφορετικά επίπεδα, το επίπεδο της διασύνδεσης, το επίπεδο της σύνταξης μηνυμάτων και το επίπεδο της σημασιολογίας.

Η επικοινωνία μέσω μαυροπίνακα αποτελεί μια μορφή κοινής μνήμης [17] όπου όλοι οι πράκτορες έχουν τη δυνατότητα να γράφουν ή να διαβάζουν μηνύματα (Εικόνα 2.2). Έτσι, στα συστήματα μαυροπίνακα υπάρχει ένας κοινός χώρος όπου όλοι οι πράκτορες του συστήματος έχουν τη δυνατότητα πρόσβασης και μέσω του οποίου είτε ανταλλάσσουν αποτελέσματα ή διαμοιράζονται εργασίες με τη μορφή μηνυμάτων.



**Εικόνα 2.2** Αρχιτεκτονική επικοινωνίας με χρήση μαυροπίνακα

Παρ' όλα αυτά, στα σύγχρονα υπολογιστικά συστήματα ο πιο διαδεδομένος τρόπος επικοινωνίας είναι η άμεση ανταλλαγή μηνυμάτων μεταξύ των πρακτόρων (Εικόνα 2.3). Καθώς όμως σε αυτή την περίπτωση επικοινωνίας το περιεχόμενο των μηνυμάτων διαφοροποιείται σημαντικά στα διάφορα συστήματα, απαιτείται ο καθορισμός ενός πρωτοκόλλου, μιας γλώσσας σύνταξης και μεταφοράς του μηνύματος αλλά και ο προσδιορισμός των διαφορετικών τύπων μηνυμάτων.



**Εικόνα 2.3** Παράδειγμα ανταλλαγής μηνυμάτων μεταξύ πρακτόρων

Άλλοι παράγοντες που επηρεάζουν την επικοινωνία των πρακτόρων είναι ο τύπος της επικοινωνίας αλλά και ο βαθμός της επικοινωνίας. Ο τύπος επικοινωνίας μπορεί να είναι σύγχρονος ή ασύγχρονος, έτσι ο πράκτορας που αποστέλλει ένα μήνυμα στην πρώτη περίπτωση αναστέλλει τη λειτουργία του μέχρι να λάβει κάποια απάντηση, ενώ στη δεύτερη περίπτωση η απάντηση μπορεί να ληφθεί οποιαδήποτε στιγμή χωρίς να παρεμποδίσει τη λειτουργία του. Ο βαθμός επικοινωνίας αναφέρεται στον αριθμό των πρακτόρων που συμμετέχουν στην επικοινωνία. Έτσι, η ανταλλαγή μηνυμάτων μπορεί να είναι 1 προς 1 όταν μόνο 2 πράκτορες συμμετέχουν στην επικοινωνία, 1 προς N όταν ένας πράκτορας στέλνει σε πολλούς παραλήπτες ένα μήνυμα ή N προς N όταν οι αποστολείς και οι αποδέκτες ενός μηνύματος είναι πολλοί.

Ανεξάρτητα από το μοντέλο διασύνδεσης ενός πολυπρακτορικού συστήματος, είναι απαραίτητη προϋπόθεση η ύπαρξη δύο ειδών πρωτοκόλλων, των πρωτοκόλλων αλληλεπίδρασης και των πρωτοκόλλων επικοινωνίας [2].

### 2.2.1 Πρωτόκολλα Αλληλεπίδρασης

Τα πρωτόκολλα αλληλεπίδρασης προσδιορίζουν τις διαφορετικές αλληλεπιδράσεις των πρακτόρων κατά τη διάρκεια της επικοινωνίας ενός με τους άλλους, σχετικά με τα προβλήματα που λύνουν. Ένα πρωτόκολλο αλληλεπίδρασης ορίζει εκ των προτέρων ένα σύνολο πιθανών μεταβάσεων οι οποίες συνδέουν κάποιες καταστάσεις που οι πράκτορες μπορούν εκ περιτροπής να ασχολούνται έτσι ώστε να επιτυγχάνουν αποτελεσματικές ενέργειες αλληλεπίδρασης.

Τα πρωτόκολλα αλληλεπίδρασης τα οποία καλούνται και ως ‘πολιτικές συνομιλίας’ στο συγκεκριμένο έργο αναφέρονται σε πρότυπα συμπεριφοράς που οι πράκτορες πρέπει να ακολουθούν για να συμμετάσχουν σε μια επικοινωνία αλληλεπίδρασης με άλλους πράκτορες σε ένα σύστημα πολλαπλών πρακτόρων. Τα πρωτόκολλα αλληλεπίδρασης πρακτόρων είναι το πρότυπο ή η ακολουθία από τα μηνύματα των πρακτόρων που χρειάζονται για να επιτευχθεί ένα ουσιαστικό αποτέλεσμα στην επικοινωνία των πρακτόρων. Τα πρωτόκολλα αλληλεπίδρασης πρακτόρων παρέχουν τη δομή στην αλληλεπίδραση των πρακτόρων και πρόσθετες πληροφορίες που παρέχονται από το μήνυμα της αλληλεπίδρασης. Αυτές οι πληροφορίες περιλαμβάνουν στο ελάχιστο κοινές

προσδοκίες για τις πιθανές και απαιτούμενες κινήσεις συνδιάλεξης μεταξύ των πρακτόρων που αλληλεπιδρούν. Μια ομάδα πρακτόρων συμμορφώνεται με το πρωτόκολλο αλληλεπίδρασης προκειμένου να συμμετάσχει στις ανταλλαγές των μηνυμάτων. Επιπλέον όταν ένας πράκτορας στέλνει ένα μήνυμα, μπορεί να αναμένει μια απάντηση του παραλήπτη να είναι ανάμεσα σε μια σειρά από μηνύματα που υποδεικνύεται από το πρωτόκολλο και το ιστορικό αλληλεπίδρασης. Ερμηνεύοντας το πρωτόκολλο, κάθε μέλος της ομάδας μπορεί επίσης να χρησιμοποιήσει τους «κανόνες» της αλληλεπίδρασης με σκοπό να ικανοποιεί τους δικούς τους στόχους. Προκειμένου να επιτευχθεί μια έμμεση συναίνεση σχετικά με τις δυνατές καταστάσεις και ενέργειες σε μια αλληλεπίδραση επιδιώχθηκε το ίδιο το πρωτόκολλο να είναι 'σωστό' με την έννοια του να μην υπάρχουν αντιφατικές καταστάσεις, 'σαφές' χωρίς αόριστες ενέργειες, 'πλήρες' δηλαδή να μην υπάρχουν απροσδιόριστες καταστάσεις και 'επαληθεύσιμο', οι ιδιότητες που προσδιορίζουν την ορθότητα του να μπορούν να ελεγχθούν. Ένα πρωτόκολλο που δεν εμφανίζει αυτά τα χαρακτηριστικά τότε και με τη συμβολή της διαφορετικότητας των συμμετεχόντων οι πράκτορες πιθανόν να εκτελούν αντιφατικές και απροσδόκητες ενέργειες που οδηγούν σε κατάρρευση της αλληλεπίδρασης και προκαλούν δυσαρέσκεια [18].

## 2.2.2 Γλώσσες Επικοινωνίας

Η επικοινωνία αποτελεί ένα από τα πιο σημαντικά στοιχεία των πολυπρακτορικών συστημάτων και γι' αυτό το λόγω αναπτύχθηκαν οι γλώσσες επικοινωνίας πρακτόρων (Agent Communication Language, ACL) και οι οποίες αφορούν την παροχή μέσων για την ανταλλαγή πληροφοριών και γνώσης μεταξύ των πρακτόρων.

Οι ACL θα πρέπει να πληρούν κάποιες απαιτήσεις και οι οποίες χωρίζονται σε επτά διαφορετικές κατηγορίες ως εξής:

- **Φόρμα** – Μια καλή ACL θα πρέπει να είναι συντακτικά απλή, δηλωτική και εύκολα αναγνώσιμη από τους ανθρώπους. Θα πρέπει να είναι περιεκτική, αλλά ταυτόχρονα εύκολη στο να αναλυθεί και να παραχθεί. Για τη μετάδοση μιας δήλωση της γλώσσας σε κάποιον άλλο πράκτορα, η δήλωση θα πρέπει να περάσει μέσω της ροής bit του υποκείμενου μηχανισμού μεταφοράς. Έτσι, η γλώσσα θα πρέπει να είναι γραμμική ή σε άλλη περίπτωση θα πρέπει να είναι εύκολο να μεταφραστεί σε μια γραμμική ροή χαρακτήρων.
- **Περιεχόμενο** – Μια ACL θα πρέπει να διαχωρίζεται σε επίπεδα έτσι ώστε να είναι εύκολο να ενσωματωθεί με άλλα συστήματα. Συγκεκριμένα, θα πρέπει να είναι σαφής ο διαχωρισμός μεταξύ της γλώσσας επικοινωνίας, η οποία εκφράζει επικοινωνιακές πράξεις, και στη γλώσσα του περιεχομένου, ή οποία εκφράζει τα γεγονότα του συγκεκριμένου πεδίου εφαρμογής. Η γλώσσα θα πρέπει να είναι δεσμευμένη σε ένα καλά προκαθορισμένο σύνολο επικοινωνιακών πράξεων (αρχετύπων).
- **Σημασιολογία** – Η σημασιολογία είναι ένα θέμα που συχνά παραμελείται στο σχεδιασμό γλωσσών επικοινωνίας πρακτόρων. Αυτή η αμέλεια αντανακλά την ασάφεια που περιβάλλει τον σκοπό και τα επιθυμητά χαρακτηριστικά της γλώσσας. Αν και η σημασιολογική περιγραφή των γλωσσών και των αρχετύπων

τους συχνά περιορίζονται στην περιγραφή των φυσικών γλωσσών, μια καλά προκαθορισμένη σημασιολογική περιγραφή αποτελεί μια πολυτέλεια. Η σημασιολογική περιγραφή είναι ιδιαίτερα σημαντική όταν η γλώσσα επικοινωνίας προορίζεται για ένα ευρύ φάσμα εφαρμογών. Οι σχεδιαστές των εφαρμογών θα πρέπει να έχουν μια κοινή κατανόηση της γλώσσας, των αρχτύπων και των πρωτοκόλλων που σχετίζονται με τη χρήση τους, και να τηρούν αυτή την κοινή αντίληψη.

- **Υλοποίηση** – Η υλοποίηση θα πρέπει να είναι αποτελεσματική τόσο από άποψη ταχύτητας όσο και τη βέλτιστη χρήση του εύρους ζώνης. Θα πρέπει να παρέχουν καλή συναρμογή με τις υπάρχουσες εφαρμογές λογισμικού. Η διασύνδεση με το χρήστη πρέπει να είναι εύκολη στη χρήση, καθώς οι λεπτομέρειες που αφορούν το επικοινωνιακό υπόβαθρο της γλώσσας θα πρέπει να μην προβάλλονται στο χρήστη. Θα πρέπει να είναι εύκολη η ένταξη και η δημιουργία διασυνδέσεων εφαρμογών για ένα ευρύ φάσμα γλωσσών προγραμματισμού, στις οποίες περιλαμβάνονται οι διαδικαστικές γλώσσες, οι γλώσσες scripting, οι αντικειμενοστραφείς γλώσσες, καθώς και οι γλώσσες λογικού προγραμματισμού.
- **Τεχνολογία Δικτύων** – Η ACL θα πρέπει να σχεδιαστεί με τέτοιο τρόπο ώστε να προσαρμόζεται στη σύγχρονη τεχνολογία δικτύων. Αυτό είναι ιδιαίτερα σημαντικό καθώς κάποιο κομμάτι της επικοινωνίας θα αναφέρεται σε έννοιες που αφορούν δικτυωμένες επικοινωνίες. Η γλώσσα θα πρέπει να υποστηρίζει όλες τις βασικές συνδέσεις, όπως είναι οι point-to-point, multicast και broadcast, όπως και σύγχρονες και ασύγχρονες συνδέσεις. Επιπλέον, η ACL θα πρέπει να περιλαμβάνει ένα πλούσιο σύνολο αρχτύπων, το οποίο θα μπορεί στη συνέχεια να χρησιμεύσει ως ένα υπόστρωμα πάνω στο οποίο θα αναπτυχθούν γλώσσες υψηλότερου επιπέδου και πρωτόκολλα. Τέλος, τα συγκεκριμένα πρωτόκολλα θα πρέπει να είναι ανεξάρτητα από το μηχανισμό μεταφοράς που χρησιμοποιείται στο δίκτυο όπως είναι για παράδειγμα το TCP/IP.
- **Περιβάλλον** – Το περιβάλλον ανάπτυξης των πολυπρακτορικών συστημάτων είναι κατανοημένο, ετερογενές και δυναμικό. Κάτω από αυτές τις συνθήκες, η ACL θα πρέπει να παρέχει τα κατάλληλα εργαλεία για την αντιμετώπιση της ετερογένειας και του δυναμισμού. Επομένως, θα πρέπει να υποστηρίζει τη διαλειτουργικότητα με άλλες γλώσσες ή πρωτόκολλα. Τέλος, θα πρέπει να υποστηρίζει την αναζήτηση γνώσης σε μεγάλα δίκτυα όπως και να είναι εύκολη η προσάρτησή της σε κληροδοτημένα συστήματα.
- **Αξιοπιστία** – Γενικά, κάθε γλώσσα επικοινωνίας θα πρέπει να υποστηρίζει την αξιόπιστη και ασφαλή επικοινωνία μεταξύ των πρακτόρων. Θα πρέπει να υποστηρίζεται η ασφαλής και ιδιωτική παροχή ανταλλαγή μηνυμάτων μεταξύ δύο πρακτόρων. Επιπλέον, θα πρέπει να παρέχεται ένας τρόπος για την εγγύηση της ταυτότητας των πρακτόρων. Καθώς, ούτε οι πράκτορες ούτε και οι υποκείμενοι μηχανισμοί είναι αλάνθαστοι, η γλώσσα θα πρέπει να προστατεύεται από τα ακατάλληλα μηνύματα, ενώ ταυτόχρονα να υποστηρίζει μηχανισμούς για τον εντοπισμό και την επισήμανση σφαλμάτων και προειδοποιήσεων [64][2].

### 2.2.2.1 FIPA ACL

Η FIPA ACL αναπτύχθηκε το 1995 από το Foundation for Intelligent Physical Agents (FIPA) και καθορίζει την μορφή των μηνυμάτων που ανταλλάσσονται μεταξύ πρακτόρων. Ορίζει μια 'εξωτερική' γλώσσα για τα μηνύματα, καθορίζει 20 διαφορετικές δηλώσεις (performatives), όπως είναι η `inform`, για τον προσδιορισμό των απεσταλμένων μηνυμάτων, ενώ δεν επιβάλλει κάποια συγκεκριμένη γλώσσα για το περιεχόμενο των μηνυμάτων.

Η δομή ενός μηνύματος FIPA ACL έχει ως εξής:

(`inform`

`:sender agent1`

`:receiver agent2`

`:content (price good2 150)`

`:language sl`

`:ontology hpl -auction`

)

Στο παραπάνω μήνυμα ο αποστολέας ενημερώνει τον παραλήπτη ότι η τιμή του `good2` είναι 150. Συγκεκριμένα, με τη δήλωση `inform` ο αποστολέας θέλει να γνωστοποιήσει το περιεχόμενο του συγκεκριμένου μηνύματος στον παραλήπτη. Το στοιχείο `:sender` αναφέρεται στο ποιος είναι ο αποστολέας του συγκεκριμένου μηνύματος, το στοιχείο `:receiver` αναφέρεται στο ποιος είναι ο παραλήπτης του συγκεκριμένου μηνύματος, ενώ το στοιχείο `:content` αναφέρει το περιεχόμενο του μηνύματος. Τα στοιχεία `:language` και `:ontology` αναφέρονται στο περιεχόμενο και προσδιορίζουν τη γλώσσα του και την ορολογία του αντίστοιχα [5].

### 2.2.2.2 KQML

Η KQML (Knowledge and Query Manipulation Language) αποτελεί μια γλώσσα επικοινωνίας μεταξύ πρακτόρων, αλλά ταυτόχρονα λειτουργεί και ως ένα πρωτόκολλο, καθιστώντας την μια από τις πιο δημοφιλείς και σημαντικές ACL. Διαχωρίζεται σε τρία επιμέρους επίπεδα ως εξής:

- Επίπεδο περιεχομένου (`content layer`): αναφέρεται στο περιεχόμενο του μηνύματος το οποίο καθορίζεται από την εκάστοτε εφαρμογή.
- Επίπεδο μηνύματος (`message layer`): καθορίζει τον τύπο του μηνύματος με χρήση των δηλώσεων (`performatives`) που διαθέτει η KQML από το σύνολο των προκαθορισμένων, όπως είναι για παράδειγμα οι `ask`, `tell`, `reply` και `inform`. Επιπλέον, ορίζει το πρωτόκολλο επικοινωνίας που θα υιοθετηθεί και που αφορά τον τρόπο αποστολής και λήψης μηνυμάτων όπως περιγράφηκαν στην παράγραφο 2.2.1.

- Επίπεδο επικοινωνίας (communication layer): αφορά τον προσδιορισμό λεπτομερειών χαμηλού επιπέδου για τον τρόπο υλοποίησης της επικοινωνίας μεταξύ των πρακτόρων [19].

Ένα μήνυμα σε KQML έχει την ακόλουθη δομή:

```
(ask-one
  :content (PRICE IBM ?price)
  :receiver stock-server
  :language LPROLOG
  :ontology NYSE-TICKS
)
```

Το παραπάνω μήνυμα αφορά την ερώτηση του αποστολέα για την τιμή της μετοχής IBM. Η έκφραση ask-one αφορά τη δήλωση που χρησιμοποιεί ένας πράκτορας όταν χρειάζεται να κάνει μια ερώτηση σε έναν άλλο πράκτορα απαιτώντας μόνο μια απάντηση. Τα υπόλοιπα στοιχεία του μηνύματος αφορούν τα χαρακτηριστικά του, το πιο σημαντικό εκ των οποίων είναι το :content και το οποίο προσδιορίζει το περιεχόμενο του μηνύματος. Το στοιχείο :receiver αφορά τον παραλήπτη του συγκεκριμένου μηνύματος, ενώ το στοιχείο :language προσδιορίζει την γλώσσα στην οποία εκφράζεται το μήνυμα του περιεχομένου. Τέλος, το στοιχείο :ontology αφορά την ορολογία του συγκεκριμένου μηνύματος [5].

Καθώς η KQML αποτελεί μια βιβλιοθήκη επικοινωνιακών ενεργειών, όταν οι πράκτορες επικοινωνούν με χρήση της KQML τότε χρησιμοποιούν τη βιβλιοθήκη των προκαθορισμένων εννοιών της. Στη συνέχεια με χρήση του κατάλληλου κώδικα γίνεται η επεξεργασία των στοιχείων των μηνυμάτων που αφορούν την εκάστοτε εφαρμογή. Η γλώσσα KQML έχει εφαρμοσθεί σε μια πληθώρα εφαρμογών, όπως είναι τα συστήματα επίδειξης, το ηλεκτρονικό εμπόριο και τα πληροφοριακά συστήματα υγείας [2].

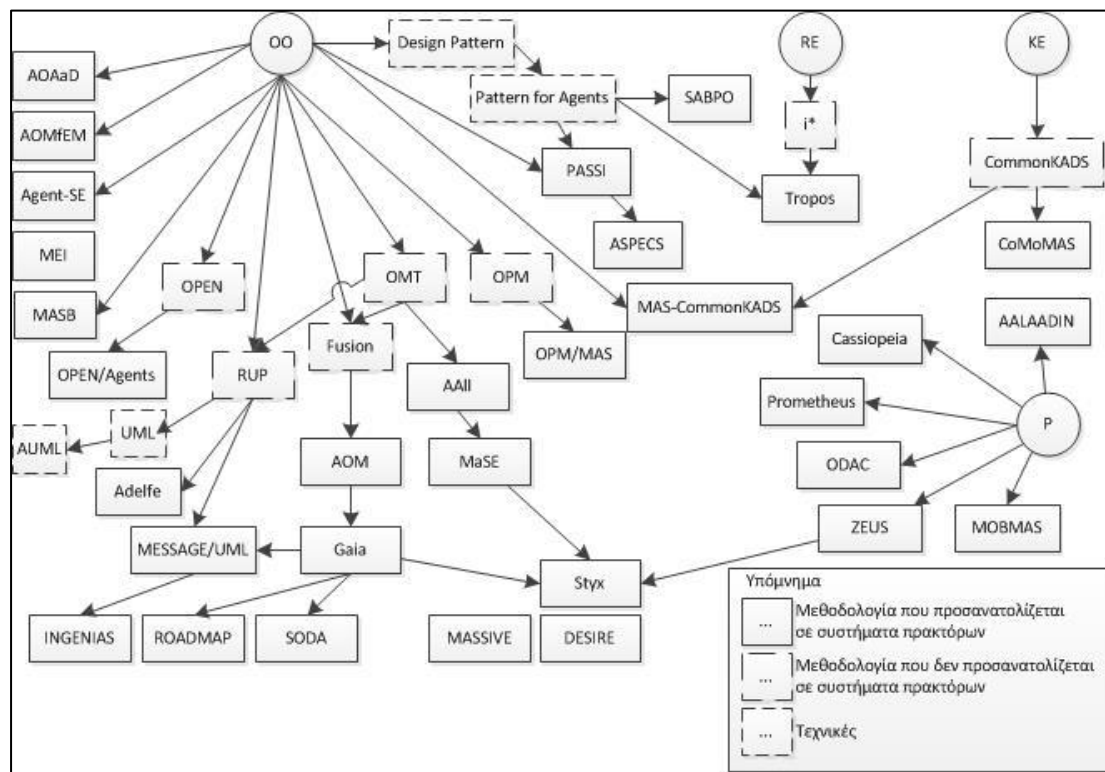
## 2.3 Μεθοδολογίες Ανάπτυξης Πολυπρακτορικών Συστημάτων

Η μεθοδολογία αποτελεί ένα σύνολο μεθόδων, κανόνων και αξιωμάτων που χρησιμοποιούνται από έναν επιστημονικό κλάδο και η ανάλυση των αρχών και των διαδικασιών έρευνας ενός συγκεκριμένου τομέα [20]. Ο σκοπός μιας μεθοδολογίας είναι να παρέχει μοντέλα, μεθόδους, εργαλεία και τεχνικές έτσι ώστε η ανάπτυξη του λογισμικού να μπορεί να επιτευχθεί με έναν συστηματικό τρόπο [21]. Στον τομέα της ανάπτυξης λογισμικού, μια μεθοδολογία προσφέρει τη σωστή κατευθυντήρια γραμμή για την κάλυψη ολόκληρου του κύκλου ζωής της ανάπτυξης ενός συστήματος, τόσο σε τεχνικό επίπεδο όσο και σε επίπεδο διαχείρισης [20].

Η προσέγγιση για την ανάπτυξη ενός λογισμικού που βασίζεται σε πράκτορες διαφέρει από την προσέγγιση της κλασικής ανάπτυξης λογισμικού καθώς υπάρχει μια ουσιαστική αναντιστοιχία μεταξύ των όρων που χρησιμοποιούνται στον αντικειμενοστραφή προγραμματισμό και σε αυτών που χρησιμοποιούνται στα πολυπρακτορικά συστήματα. Συγκεκριμένα, οι προσεγγίσεις του κλασικού προγραμματισμού αποτυγχάνουν να συλλάβουν επαρκώς την ευέλικτη και αυτόνομη συμπεριφορά των πρακτόρων καθώς και την πολυπλοκότητα των οργανωτικών δομών ενός πολυπρακτορικού συστήματος [22].

Οι τρέχουσες μεθοδολογίες για την ανάπτυξη συστημάτων που βασίζονται σε πράκτορες διακρίνονται σε τέσσερις κύριες κατηγορίες ως εξής:

- Αντικειμενοστραφείς μεθοδολογίες (Object-oriented, OO)
- Μεθοδολογίες που βασίζονται στη γνωστική μηχανική (Knowledge Engineering, KE)
- Μεθοδολογίες που βασίζονται στη μηχανική των απαιτήσεων (Requirements Engineering, RE)
- Μεθοδολογίες που βασίζονται σε διάφορες πηγές (P)



**Εικόνα 2.4** Γενεαλογία των μεθοδολογιών για την ανάπτυξη λογισμικού που βασίζεται σε πράκτορες

Στο διάγραμμα της Εικόνας 2.4 απεικονίζονται 22 αντικειμενοστραφείς μεθοδολογίες, 2 μεθοδολογίες που βασίζονται στη γνωστική μηχανική, μία μεθοδολογία που βασίζεται στη μηχανική των απαιτήσεων και 7 που βασίζονται σε διάφορες πηγές και που έχουν προταθεί αποκλειστικά για την ανάπτυξη πολυπρακτορικών συστημάτων, ενώ κάποιες από αυτές, όπως είναι η MAS CommonKADS [23], εκμεταλλεύονται γνώρισμα από πολλές προσεγγίσεις. Παρ' όλα αυτά, λίγες εξ αυτών είναι αρκετά ώριμες ή



περιγράφονται με αρκετή λεπτομέρεια ώστε να μπορούν να χρησιμοποιηθούν στην πράξη για την ανάπτυξη λογισμικού [24].

### 2.3.1 Ρόλοι Πρακτόρων

Ο φορμαλισμός των ρόλων των πρακτόρων επιτρέπει την εύκολη μοντελοποίηση σχεδιασμό και υλοποίησή τους. Το μοντέλο ρόλων προβλέπεται στον αντικειμενοστραφή τρόπο σχεδίασης και μοντελοποίησης συστημάτων. Οι ρόλοι πρέπει να έχουν τις παρακάτω ιδιότητες:

- Είναι **τμηματοποιημένοι**: Ένας ρόλος ορίζει ένα σύνολο οντοτήτων όλοι εκ των οποίων μπορούν να πάρουν την ίδια θέση σε μια επαναλαμβανόμενη δομή. Έτσι όλοι οι πράκτορες που μπορούν να αναλάβουν ένα συγκεκριμένο ρόλο θα μπορούσαν να εναλλαχθούν σε πολλαπλές λειτουργίες αυτού του ρόλου.
- Έχουν **υψηλή συνοχή**: Για να προαγάγει την τμηματικότητα ένας ρόλος πρέπει να περιλαμβάνει ένα καλά ορισμένο σύνολο από αρμοδιότητες και λειτουργίες.
- Είναι **συγκεκριμένοι**: Ένας ρόλος δεν μπορεί να έχει αυξημένες-πρόσθετες αρμοδιότητες.
- Είναι **ολοκληρωμένοι**: Ένας ρόλος δεν πρέπει να είναι τετριμμένος. Οι τετριμμένοι ρόλοι πρέπει να συγχωνεύονται με άλλους ρόλους.
- **Συνδέονται ελάχιστα**: Οι εξαρτήσεις μεταξύ ρόλων πρέπει να ελαχιστοποιούνται.

Στην περίπτωση κοινωνικών πρακτόρων κάποιοι πράκτορες μπορεί να αναλαμβάνουν περισσότερους από έναν ρόλους. Για παράδειγμα, ο πράκτορας που αναλαμβάνει τον ρόλο 'διαχειριστής' μπορεί να διαθέτει και τον ρόλο 'προγραμματιστής' [25].

### 2.3.2 Gaia

Η μεθοδολογία Gaia είναι μια αντικειμενοστραφής μεθοδολογία που στοχεύει στη καθοδήγηση του προγραμματιστή από τη δήλωση των απαιτήσεων σε μια λεπτομερή σχεδίαση που μπορεί να υλοποιηθεί άμεσα. Αποτελεί μια γενική και ολοκληρωμένη μεθοδολογία, δηλαδή μπορεί να εφαρμοστεί σε ένα μεγάλο φάσμα πολυπρακτορικών συστημάτων ενώ καλύπτει τόσο τις μακρο (κοινωνικές) όσο και τις μικρο (πράκτορες) πτυχές των συστημάτων. Η Gaia πρόσφατα επεκτάθηκε έτσι ώστε να περιλαμβάνει και νέες οργανωτικές απολήψεις που επιτρέπουν την υποστήριξη της ανάπτυξης ανοικτών πολυπρακτορικών συστημάτων.

Η Gaia βασίζεται στην ιδέα ότι ένα πολυπρακτορικό σύστημα, ως ένα υπολογιστικός οργανισμός, αποτελείται από διάφορους ρόλους που αλληλεπιδρούν. Διακρίνεται σε δύο φάσεις, τη φάση της ανάλυσης και τη φάση του σχεδιασμού. Κατά τη φάση της ανάλυσης ερευνάται αρχικά η πιθανή ύπαρξη πολλαπλών υπό-οργανισμών στο αναπτυσσόμενο σύστημα. Αν υπάρχουν, τότε αναλύονται και σχεδιάζονται σαν αυτόνομα αλληλεπιδρώντα πολυπρακτορικά συστήματα. Στη φάση του σχεδιασμού να μοντέλα που προκύπτουν από τη φάση της ανάλυσης μετατρέπονται σε αρκετά

χαμηλού επιπέδου απολήψεις έτσι ώστε να είναι δυνατή η εφαρμογή παραδοσιακών τεχνικών σχεδιασμού, όπως είναι οι αντικειμενοστραφείς τεχνικές.

Παρ' όλα αυτά, διαθέτει τους εξής περιορισμούς:

- Είναι αδύναμη στον εσωτερικό σχεδιασμό των πρακτόρων. Το μοντέλο των πρακτόρων μόνο καθορίζει τους ρόλους για κάθε πράκτορα, χωρίς να προσδιορίζει την εσωτερική αρχιτεκτονική των πρακτόρων και τις πνευματικές δομές, όπως είναι οι πεποιθήσεις, οι στόχοι, τα σχέδια και οι ενέργειες των πρακτόρων.
- Η υποστήριξη για την αλληλεπίδραση των πρακτόρων είναι περιορισμένη. Το μοντέλο γνωριμιών απλά προσδιορίζει τις συνδέσεις επικοινωνίας μεταξύ των κλάσεων των πρακτόρων και το μοντέλο αλληλεπιδράσεων απλά δείχνει τα 'θεσμοθετημένα μοτίβα' των αλληλεπιδράσεων. Έτσι, δεν παρέχεται ένα αναλυτικό σχέδιο της επικοινωνίας των πρακτόρων.
- Δεν αναγνωρίζει και δεν εφαρμόζει τα οφέλη των οντολογιών.
- Η Gaia δεν διαχειρίζεται άμεσα συγκεκριμένες τεχνικές μοντελοποίησης. Προτείνει, αλλά δεν δεσμεύεται σε συγκεκριμένες τεχνικές για την μοντελοποίηση, όπως για παράδειγμα, ρόλους, περιβάλλον και αλληλεπιδράσεις μεταξύ των πρακτόρων.
- Δεν ασχολείται άμεσα με θέματα υλοποίησης. Το αποτέλεσμα από τη χρήση της μεθόδου είναι λεπτομερείς αλλά τεχνολογικά ουδέτερες προδιαγραφές που θα πρέπει να υλοποιηθούν χρησιμοποιώντας το κατάλληλο λογισμικό προγραμματισμού πρακτόρων.
- Τέλος, δεν ασχολείται ρητά με τις δραστηριότητες της σύλληψης και μοντελοποίησης των απαιτήσεων και συγκεκριμένα των πρώιμων απαιτήσεων, αλλά μπορεί εύκολα να προσαρμοστεί και να ενσωματωθεί στις σύγχρονες προσεγγίσεις της μηχανικής των απαιτήσεων που προσανατολίζονται στους στόχους και των οποίων οι απολήψεις μοιάζουν αρκετά με αυτές της πληροφορικής που προσανατολίζεται στους πράκτορες [22].

### 2.3.3 Tropos

Η μεθοδολογία Tropos βασίζεται στη μηχανική των απαιτήσεων και είναι εμπνευσμένη από τη μέθοδο  $i^*$  [26] για την ανάλυση των πρώιμων απαιτήσεων ενός συστήματος λογισμικού. Αντιμετωπίζει τα συστήματα που βασίζονται σε πράκτορες ως έναν οργανισμό φορέων, οι οποίοι προσπαθούν να επιτύχουν τους στόχους τους μέσω σχεδίων και οι οποίοι έχουν εξαρτήσεις με άλλους φορείς. Η μεθοδολογία Tropos βασίζεται σε δύο κυρίαρχες ιδέες. Πρώτον, η έννοια του πράκτορα και οι σχετικές πνευματικές έννοιες, όπως τα σχέδια και οι στόχοι, χρησιμοποιούνται σε όλες τις φάσεις της ανάπτυξης λογισμικού, από την αρχική ανάλυση μέχρι την πραγματική υλοποίηση. Δεύτερον, η Tropos καλύπτει και τα πρώιμα στάδια της ανάλυσης των απαιτήσεων, παρέχοντας μια εκ βαθέων κατανόηση του περιβάλλοντος στο οποίο πρέπει να λειτουργήσουν το λογισμικό καθώς και των αλληλεπιδράσεων που λαμβάνουν χώρα μεταξύ του λογισμικού και των πρακτόρων των χρηστών.

Η διαδικασία ανάπτυξης διακρίνεται σε τέσσερις φάσεις. Στη φάση των πρώιμων απαιτήσεων διερευνώνται οι προθέσεις των ενδιαφερομένων μέσω δύο μοντέλων, του μοντέλου στρατηγικής εξάρτησης και του μοντέλου στρατηγικού συλλογισμού. Στη φάση των όψιμων απαιτήσεων αναγνωρίζονται οι λειτουργικές και μη λειτουργικές απαιτήσεις του συστήματος επεκτείνοντας το μοντέλο στρατηγικής εξάρτησης και το μοντέλο στρατηγικού συλλογισμού. Στη φάση του αρχιτεκτονικού σχεδιασμού επιλέγεται η κατάλληλη οργανωτική δομή για το υπό ανάπτυξη πολυπρακτορικό σύστημα αξιολογώντας τα ποιοτικά χαρακτηριστικά των υποψήφιων οργανωτικών δομών έναντι των μαλακών στόχων του συστήματος. Τέλος, η φάση του αναλυτικού σχεδιασμού περιλαμβάνει τις αλληλεπιδράσεις των πρακτόρων και τον εσωτερικό σχεδιασμό των πρακτόρων.

Παρ' όλα αυτά παρουσιάζει τους παρακάτω περιορισμούς:

- Η μεθοδολογία Tropos δεν διαθέτει δομημένες και λεπτομερείς τεχνικές για τη φάση του λεπτομερή σχεδιασμού. Συγκεκριμένα, δεν είναι ξεκάθαρο πως μπορούν να εξαχθούν τα πρωτόκολλα αλληλεπίδρασης των πρακτόρων και πως μπορούν να κατασκευαστούν τα σχέδια των πρακτόρων.
- Η μεθοδολογία Tropos δεν προορίζεται για οποιοδήποτε είδος λογισμικού. Για λογισμικό συστημάτων, όπως είναι οι μεταγλωττιστές ή το ενσωματωμένο λογισμικό, το περιβάλλον λειτουργίας του προς ανάπτυξη συστήματος είναι ένα αντικείμενο μηχανικής χωρίς αναγνωρίσιμους φορείς.
- Η τρέχουσα μορφή της μεθοδολογίας δεν είναι κατάλληλη για εξεζητημένους πράκτορες λογισμικού με ανεπτυγμένους μηχανισμούς συλλογισμού για σχέδια, στόχους και διαπραγματεύσεις.
- Δεν αναγνωρίζει και δεν εφαρμόζει τα οφέλη των οντολογιών [27].

---

## Κεφάλαιο 3

---

### Εμπιστοσύνη και Πολυπρακτορικά Συστήματα

---

#### 3.1 Η Έννοια της Εμπιστοσύνης

Η εμπιστοσύνη έχει οριστεί με πολλούς τρόπους στους διάφορους τομείς, παρ' όλα αυτά, ένας ορισμός που έχει δοθεί από τον Drasgurta [28] καλύπτει τις ανάγκες μας:

‘Η εμπιστοσύνη έγκειται στην πίστη ότι το άλλο συμβαλλόμενο μέρος θα κάνει αυτό που έχει δηλώσει (καθώς είναι ειλικρινές και αξιόπιστο) ή θα ανταποδώσει (ανταποδίδοντας για το κοινό καλό και των δύο), ενώ του δίδεται η ευκαιρία να αποσκιρτήσει για να έχει μεγαλύτερες απολαβές.’

Η εμπιστοσύνη ως έννοια συναντάται σε ένα ευρύ φάσμα εφαρμογών και απόψεων με επίπεδα ασυνεπούς επιρροής. Δεν υπάρχει ένας ενοποιημένος ορισμός καθώς είναι σε συγκεκριμένα πλαίσια, παρ' όλα αυτά ένα πλήθος εννοιών εμφανίζονται συχνά στην βιβλιογραφία της εμπιστοσύνης, όπως είναι η πολυπλοκότητα, η αξιοπιστία, η ευελιξία, η πρόβλεψη, η πειστικότητα, η επανάπαυση, η συνέπεια, τα βιώματα, η ασφάλεια, η ακρίβεια, η εξάρτηση, η υπευθυνότητα και η αβεβαιότητα. Οι έννοιες καλύπτουν όλο το φάσμα, απ' το επίπεδο συστήματος (τοπικό), όπως είναι η χρησιμότητα ή η αξιοπιστία ενός συγκεκριμένου τμήματος λογισμικού, μέχρι το μεγάλο κοινωνικό επίπεδο, όπως η υπευθυνότητα και η εξάρτηση.

Οι Kelly et al. [29] προσδιορίζουν την εμπιστοσύνη απλά ως την πίστη που δίνεται σε ένα άτομο ή αντικείμενο, και πιο συγκεκριμένα, τον βαθμό της πίστης στην δύναμη, στην ικανότητα, στη αλήθεια ή στην αξιοπιστία ενός ατόμου ή ενός αντικειμένου. Επιπλέον, αριθμούν και μια λίστα στοιχείων της εμπιστοσύνης που αναγνωρίζονται από

την βιβλιογραφία της έρευνας και συμπεριλαμβάνει την πίστη, την δύναμη, την οικειότητα, την κατανόηση, την χρησιμότητα, την αυτοεκτίμηση, την φήμη και την επεξήγηση της πρόθεσης. Όπως είναι προφανές, οι συνιστώσες της εμπιστοσύνης είναι πολλές.

Ίσως ο πιο σημαντικός παράγοντας της εμπιστοσύνης είναι το ρίσκο. Η ικανότητα να συζητούν για την εμπιστοσύνη και το ρίσκο επιτρέπει στους ανθρώπους να αλληλεπιδρούν ακόμη και σε καταστάσεις όπου διαθέτουν μόνο ένα μέρος των πληροφοριών. Αν το ρίσκο πιστεύεται ότι είναι πολύ μεγάλο, τότε η αλληλεπίδραση μπορεί να μην πραγματοποιηθεί. Τρεις διακριτές συνιστώσες ή επίπεδα της εμπιστοσύνης παρουσιάζονται στην συνέχεια, και στα οποία υπάρχει μια αυξανόμενη τάση στο να πιστεύει κανείς ότι το άτομο είναι άξιο εμπιστοσύνης. Τα επίπεδα αυτά είναι:

- Πρόβλεψη – χρησιμοποιείται στα πρώιμα στάδια ως βάση της εμπιστοσύνης
- Εξάρτηση – αντιστοιχεί στην εμπιστοσύνη που πηγάζει από τις ιδιότητες που αποδίδονται στην άλλη οντότητα
- Πίστη – αντανακλά την συναισθηματική ασφάλεια.

Η δυναμική φύση της εμπιστοσύνης είναι αυτοσυντηρούμενη και αυτοαναπτυσσόμενη. Αυξάνεται μέσω των αλληλεπιδράσεων και υποβιβάζεται από τα αποτυχημένα αποτελέσματα. Κάποιο επίπεδο εμπιστοσύνης επιτυγχάνεται όταν ένα σύστημα συναντά κάποιες συγκεκριμένες προσδοκίες που μπορεί να διαφέρουν μεταξύ των ατόμων καθώς και εξαιτίας των καταστάσεων [30].

### **3.2 Εμπιστοσύνη στα Πολυπρακτορικά Συστήματα**

Η εμπιστοσύνη είναι ένα βασικό μέλημα όσον αφορά μεγάλης κλίμακας καταναμημένα συστήματα. Βρίσκεται στον πυρήνα όλων των αλληλεπιδράσεων μεταξύ των οντοτήτων που πρέπει να λειτουργούν σε ένα τέτοιο αβέβαιο και συνεχώς μεταβαλλόμενο περιβάλλον.

Πιο συγκεκριμένα, τα ανοιχτά καταναμημένα συστήματα μπορούν να μοντελοποιηθούν ως ανοιχτά πολυπρακτορικά που συνθέτονται από αυτόνομους πράκτορες που αλληλεπιδρούν μεταξύ τους με την χρήση συγκεκριμένων μηχανισμών και πρωτοκόλλων. Με αυτήν την έννοια, οι αλληλεπιδράσεις συνθέτουν τον πυρήνα των πολυπρακτορικών συστημάτων. Έτσι, η ερευνητική κοινότητα των πρακτόρων έχει αναπτύξει ένα αριθμό μοντέλων αλληλεπιδράσεων που περιλαμβάνουν τον συντονισμό, την συνεργασία και την διαπραγμάτευση. Παρ' όλα αυτά, η εφαρμογή τους σε μεγάλης κλίμακας ανοιχτά καταναμημένα συστήματα παρουσιάζει πολλές προκλήσεις. Πρώτον, οι πράκτορες πιθανόν να αντιπροσωπεύουν διαφορετικούς ενδιαφερόμενους, όπου ο καθένας έχει τους δικούς του στόχους και επιδιώξεις. Αυτό σημαίνει ότι η πιο αληθοφανής στρατηγική σχεδιασμού ενός πράκτορα είναι να μεγιστοποιεί την ατομική του χρησιμότητα. Δεύτερον, δεδομένου ότι το σύστημα είναι ανοιχτό, οι πράκτορες μπορούν να συνδέονται και να φεύγουν οποιαδήποτε στιγμή. Αυτό σημαίνει ότι ένας πράκτορας μπορεί να αλλάζει ταυτότητα καθώς επανασυνδέεται στο σύστημα και

συνεπώς να αποφεύγει την τιμωρία για οποιαδήποτε κακή ενέργεια. Για παράδειγμα, ένας πράκτορας θα μπορούσε να πουλά προϊόντα χαμηλής ποιότητας, να εγκαταλείψει το σύστημα μόλις λάβει τις πληρωμές (έτσι να αποφεύγει την ανταπόδοση από τους αγοραστές ή τους διαχειριστές του συστήματος), και στην συνέχεια να συνδέονται ξανά στο σύστημα άθικτοι. Τρίτον, ένα ανοικτό καταναμημένο σύστημα επιτρέπει σε πράκτορες με διαφορετικά χαρακτηριστικά (π.χ. πολιτική, ικανότητες, ρόλους) να εισέρχονται σε αυτό και να αλληλεπιδρούν μεταξύ τους. Δεδομένου αυτό, οι πράκτορες είναι πιθανό να αντιμετωπίσουν έναν αριθμό πιθανών παρτενέρ με διάφορες ιδιότητες. Για παράδειγμα, κάποιος πράκτορας μπορεί να προσφέρουν τον ίδιο τύπο διαδικτυακής υπηρεσίας, όμως με διαφορετική αποδοτικότητα (π.χ. ταχύτητα εκτέλεσης) ή διαφορετικό βαθμό αποτελεσματικότητας (π.χ. παροχή πιο πλούσιων αποτελεσμάτων). Τέταρτον, ένα ανοικτό καταναμημένο σύστημα επιτρέπει στους πράκτορες να ανταλλάσουν προϊόντα ή υπηρεσίες (π.χ. μέσω διαφόρων ειδών δημοπρασιών ή μηχανισμών αγοραπωλησίας) και να συνεργάζονται (π.χ. με το σχηματισμό συνασπισμών ή εικονικών οργανισμών) με πολλούς τρόπους. Επομένως, οι σχεδιαστές πρακτόρων έρχονται αντιμέτωποι με μια επιλογή ενός αριθμού πιθανών πρωτοκόλλων αλληλεπιδράσεων που θα μπορούσαν να τους βοηθήσουν να επιτύχουν τους στόχους τους. Επιπλέον, η επιλογή ποιού πρωτοκόλλου (ή μηχανισμού) αλληλεπίδρασης θα υιοθετηθεί από είναι σημαντική καθώς κάθε πρωτόκολλο μπορεί να ενισχύσει ένα διαφορετικό σύνολο κανόνων αντιπαραθέσεων και κάθε πρωτόκολλο μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα για τους πράκτορες που παίρνουν μέρος (π.χ. η δημοπρασία του eBay επιτρέπει στον τελευταίο υψηλότερο πλειοδότη να κερδίσει, ενώ η δημοπρασία Vickrey προορίζει ως νικητή τον πράκτορα με την υψηλότερη αποτίμηση).

Κατά την δημιουργία ενός πρωτοκόλλου, προτίθεται η συνέχεια των κινήσεων των πρακτόρων και η κατανομή των πόρων, που προκαλούνται από την εφαρμογή του, να είναι φτιαγμένα έτσι ώστε να αποτρέπουν τους πράκτορες να εκμεταλλεύονται ο ένας τον άλλο (π.χ. με ψέματα ή συνομοσίες) για να ικανοποιήσουν τα εγωιστικά τους συμφέροντα. Επομένως, η ύπαρξη τέτοιων πρωτοκόλλων προσφέρει εγγυήσεις που διευκολύνουν την επιλογή παρτενέρ οποιαδήποτε στιγμή. Παρ' όλα αυτά, τα πρωτόκολλα μπορεί μερικές φορές να υπόκεινται σε ανταλλαγές, μεταξύ των κανόνων που επιβάλλουν, στην προσπάθειά τους να επιτύχουν τους στόχους τους (π.χ. στην ψηφοφορία, ενώ κάποιος περιορισμοί επιβάλλουν την αποκάλυψη της αλήθειας, μπορεί να οδηγήσουν σε ανυπακοή). Σε αυτές τις περιπτώσεις, οι πράκτορες έχουν την ευθύνη της επιλογής του τρόπου, του τόπου και με ποιόν θα αλληλεπιδράσουν χωρίς εγγυήσεις ότι η αλληλεπίδραση θα επιφέρει τελικά τα επιθυμητά αποτελέσματα. Η λήψη τέτοιων αποφάσεων θα απαιτούσε οι πράκτορες να έχουν πλήρη γνώση για τους αντιπάλους τους, το περιβάλλον και το αντικείμενο που διακυβεύεται. Αυτού του είδους οι πληροφορίες θα επέτρεπαν στον πράκτορα να σκαρώσει τα ενδεχόμενα για συγκεκριμένα γεγονότα και επομένως να δράσει με τέτοιο τρόπο ώστε να μεγιστοποιήσει την αναμενόμενη χρησιμότητά του. Επιπλέον, έχοντας τέτοιου είδους πληροφορίες, οι πράκτορες θα μπορούσαν να ενεργούν στρατηγικά υπολογίζοντας την καλύτερη αντίδρασή τους δεδομένου των πιθανών κινήσεων των αντιπάλων τους κατά την διάρκεια της αλληλεπίδρασης.

Παρ' όλα αυτά, τόσο το σύστημα (που επιβάλλει το πρωτόκολλο) όσο και οι πράκτορες μπορεί να έχουν περιορισμένες υπολογιστικές και αποθηκευτικές ικανότητες που περιορίζουν τον έλεγχο τους πάνω στις αλληλεπιδράσεις. Επιπλέον, το περιορισμένο εύρος και η ταχύτητα των καναλιών επικοινωνίας περιορίζουν τις αισθητήριες ικανότητες των πρακτόρων σε εφαρμογές πραγματικού κόσμου. Έτσι, σε λειτουργικό πλαίσιο είναι συνήθως ακατόρθωτο να φτάσει σε μια κατάσταση πλήρους γνώσης του περιβάλλοντος και των ιδιοτήτων των υπολοίπων πρακτόρων, των στρατηγικών και των συμφερόντων. Οι πράκτορες επομένως αναγκαστικά έρχονται αντιμέτωποι με έναν σημαντικό βαθμό αβεβαιότητας κατά την λήψη αποφάσεων (π.χ. μπορεί να είναι δύσκολο ή και ακατόρθωτο να υπολογιστούν οι πιθανότητες για κάποιον ενδεχόμενο). Σε αυτές τις περιστάσεις, οι πράκτορες θα πρέπει να εμπιστεύονται ο ένας τον άλλο έτσι ώστε να ελαχιστοποιείται η αβεβαιότητα που σχετίζεται με τις αλληλεπιδράσεις στα ανοιχτά καταναμημένα συστήματα.

Υπάρχουν δύο κύριες προσεγγίσεις όσον αφορά την έννοια της εμπιστοσύνης στα πολυπρακτορικά συστήματα. Κατ' αρχήν, για να επιτραπεί στους πράκτορες να εμπιστεύονται ο ένας τον άλλο, πρέπει να τους δοθεί η ικανότητα να επιχειρηματολογούν για την ανταποδοτική φύση, την αξιοπιστία ή την ειλικρίνεια των ομολόγων τους. Αυτή η ικανότητα συλλαμβάνεται μέσω μοντέλων εμπιστοσύνης. Το τελευταίο στοχεύει στο να καθίστανται ικανοί οι πράκτορες να υπολογίζουν την ποσότητα της εμπιστοσύνης που μπορούν να δώσουν στους παρτενέρ τους. Ένας υψηλός βαθμός εμπιστοσύνης σε έναν πράκτορα θα σήμαινε ότι έχει μεγαλύτερη πιθανότητα να επιλεχθεί ως παρτενέρ και πιθανόν να χρησιμοποιηθεί μια ανταποδοτική στρατηγική σε πολλαπλές αλληλεπιδράσεις. Αντίθετα, ένας χαμηλός βαθμός εμπιστοσύνης θα έχει ως αποτέλεσμα να μην επιλεχθεί (αν είναι διαθέσιμοι άλλοι πράκτορες με υψηλότερο βαθμό εμπιστοσύνης είναι διαθέσιμοι) και μια μη-ανταποδοτική στρατηγική υιοθετείται εναντίον του σε πολλαπλές αλληλεπιδράσεις (αν δεν υπάρχει καλύτερη εναλλακτική λύση). Με αυτόν τον τρόπο, τα μοντέλα εμπιστοσύνης στοχεύουν στην καθοδήγηση της λήψης αποφάσεων ενός πράκτορα όσον αφορά τον τρόπο, το τόπο καθώς και με ποίον να αλληλεπιδράσουν. Ωστόσο, για να το επιτύχουν, τα μοντέλα εμπιστοσύνης αρχικά απαιτούν από τους πράκτορες να συλλέξουν πληροφορίες για τα χαρακτηριστικά των αντιπάλων τους. Αυτό μπορεί να επιτευχθεί με πολλούς διαφορετικούς τρόπους, μέσω των συμπερασμάτων που εξάγονται από τα αποτελέσματα πολλαπλών άμεσων αλληλεπιδράσεων με τους συγκεκριμένους παρτενέρ ή μέσω έμμεσων πληροφοριών που παρέχονται από άλλους πράκτορες. Οι περίπτωση της άμεσης αλληλεπίδρασης οδηγεί στην μελέτη μεθόδων μέσω των οποίων οι πράκτορες μπορούν να μάθουν και να εξελίξουν καλύτερες στρατηγικές για την αντιμετώπιση έντιμων και ανέντιμων πρακτόρων τέτοιες ώστε οι απολαβές να μεγιστοποιηθούν στο μακροπρόθεσμο. Η περίπτωση των έμμεσων αλληλεπιδράσεων απαιτεί από τους πράκτορες να έχουν την ικανότητα να αναπτύξουν μεθόδους έγκυρης απόκτησης και επιχειρηματολόγησης για τις πληροφορίες που συλλέχθηκαν από τους άλλους πράκτορες.

Ενώ τα μοντέλα εμπιστοσύνης σχετίζονται με την ικανότητα επιχειρηματολόγησης και συλλογής πληροφοριών των πρακτόρων, η δεύτερη κύρια προσέγγιση της εμπιστοσύνης αφορά τον σχεδιασμό πρωτοκόλλων και μηχανισμών αλληλεπιδράσεων, δηλαδή των κανόνων αντιμετώπισης. Αυτοί οι μηχανισμοί αλληλεπίδρασης πρέπει να

σχεδιαστούν έτσι ώστε να εξασφαλιστεί ότι οι εμπλεκόμενοι μπορεί να είναι σίγουροι ότι θα κερδίσουν κάποια χρησιμότητα αν το αξίζουν (π.χ. ένας κακόβουλος πράκτορας δεν μπορεί να αλλάξει την σωστή κατανομή των απολαβών του μηχανισμού). Επομένως αναμένεται οι πράκτορες να αλληλεπιδρούν χρησιμοποιώντας ένα συγκεκριμένο μηχανισμό μόνο αν είναι άξιος εμπιστοσύνης). Για παράδειγμα, μια Αγγλική δημοπρασία μπορεί να είναι άξια εμπιστοσύνης (από τους πλειοδότες) μέχρι ένα σημείο καθώς εξασφαλίζει ότι ο εκπλειστηριαστής δεν μπορεί να πειράξει τις προσφορές καθώς διατυπώνονται δημόσια. Όμως, η ίδια δημοπρασία δεν είναι άξια εμπιστοσύνης (από τον εκπλειστηριαστή) όσον αφορά την απόσπαση της πραγματικής αξίας των αγαθών από τους πλειοδότες, καθώς η κυρίαρχη στρατηγική αυτού του μηχανισμού είναι η χαμηλότερη προσφορά από την πραγματική αξία και λίγο υψηλότερη από την τωρινή προσφορά. Αυτό τονίζει την ανάγκη πρωτοκόλλων που θα διασφαλίζουν ότι οι συμμετέχοντες δεν θα βρουν καλύτερη επιλογή από την ειλικρίνεια και την έντιμη αλληλεπίδραση μεταξύ τους.

Είναι προφανές ότι η εμπιστοσύνη διέπει τις πολυπρακτορικές αλληλεπιδράσεις σε όλα τα επίπεδα. Όσον αφορά τον σχεδιασμό των πρακτόρων και των ανοιχτών πολυπρακτορικών συστημάτων η εμπιστοσύνη μπορεί να συλληφθεί με τους ακόλουθους τρόπους:

- **Εμπιστοσύνη ατομικού επιπέδου** – οι πράκτορες έχουν κάποιες πεποιθήσεις για την εντιμότητα ή την ανταποδοτική φύση των παρτενέρ τους
- **Εμπιστοσύνη στο επίπεδο του συστήματος** – οι συμμετέχοντες στο σύστημα είναι αναγκασμένοι να είναι άξιοι εμπιστοσύνης από τους κανόνες αντιμετώπισης (δηλαδή τα πρωτόκολλα και τους μηχανισμούς) που ρυθμίζουν το σύστημα.

Οι παραπάνω προσεγγίσεις είναι συμπληρωματικές. Συνεπώς, ενώ το πρωτόκολλα στοχεύουν στην εξασφάλιση της φερεγγυότητας των πρακτόρων σε επίπεδο συστήματος, δεν μπορούν πάντα να το επιτύχουν χωρίς κάποια απώλεια στη απόδοση, και σε αυτές τις περιπτώσεις τα μοντέλα εμπιστοσύνης σε ατομικό επίπεδο είναι σημαντικά στην καθοδήγηση της λήψης αποφάσεων ενός πράκτορα. Με τον ίδιο τρόπο, όπου τα μοντέλα εμπιστοσύνης σε ατομικό επίπεδο δεν μπορούν να ανταπεξέλθουν με την ακατανίκητη αβεβαιότητα του περιβάλλοντος, τα μοντέλα εμπιστοσύνης τους επιπέδου του συστήματος, μέσω συγκεκριμένων μηχανισμών, στοχεύουν στον περιορισμό των αλληλεπιδράσεων και στην μείωση αυτής της αβεβαιότητας [31].

### 3.3 Εμπιστοσύνη Ατομικού Επιπέδου

Η εμπιστοσύνη ατομικού επιπέδου αναφέρεται στην οπτική πλευρά ενός πράκτορα που είναι τοποθετημένος σε ένα ανοιχτό περιβάλλον και προσπαθεί να επιλέξει το πιο αξιόπιστο παρτενέρ για αλληλεπίδραση από μια ομάδα πρακτόρων και να αποφασίσει την στρατηγική που θα εφαρμόσει με αυτόν. Υπάρχουν διάφοροι τρόποι που μπορεί να το επιτύχει. Πρώτον, θα μπορούσαν να αλληλεπιδράσουν όλοι μεταξύ τους και να μάθουν την συμπεριφορά του καθενός μέσω ενός αριθμού αντιπαραθέσεων. Τελικά, θα πρέπει να είναι ικανοί να επιλέξουν τον πιο αξιόπιστο ή έντιμο πράκτορα από την



ομάδα ή να καταστρώσουν μια κατάλληλη στρατηγική για να αντιμετωπίσουν του λιγότερο (η περισσότερο) αξιόπιστους. Σε αυτήν την περίπτωση, ο πράκτορας επιχειρηματολογεί για το αποτέλεσμα την άμεσης αλληλεπίδρασης με τους άλλους. Δεύτερον, ο πράκτορας θα μπορούσε να ρωτήσει τους άλλους πράκτορες για την αντίληψή τους για τους πιθανούς παρτενέρ. Σε αυτήν την περίπτωση, ο πράκτορας επιχειρηματολογεί για τις αλληλεπιδράσεις που είχαν οι άλλοι πράκτορες με τους πιθανούς παρτενέρ (έμμεσες αλληλεπιδράσεις). Τρίτον, ο πράκτορας θα μπορούσε να χαρακτηρίσει τα γνωστά κίνητρα των άλλων πρακτόρων. Αυτό συμπεριλαμβάνει την διαμόρφωση λογικών πεποιθήσεων για διαφορετικά χαρακτηριστικά των πρακτόρων και επιχειρηματολόγηση για αυτές τις πεποιθήσεις έτσι ώστε να αποφασίσει πόση εμπιστοσύνη θα πρέπει να δοθεί σε αυτές.

Σύμφωνα με όσα ειπώθηκα παραπάνω, τα μοντέλα εμπιστοσύνης ατομικού επιπέδου μπορούν να ταξινομηθούν σε κατηγορίες, τα μοντέλα που βασίζονται στην μάθηση (και στην εξέλιξη), τα μοντέλα που βασίζονται στην υπόληψη και στα κοινωνιογνωσιακά μοντέλα. Ενώ τα μαθησιακά και εξελικτικά μοντέλα στοχεύουν να προικίσουν τους πράκτορες με στρατηγικές που θα μπορούν να αντιμετωπίσουν τους πράκτορες που ψεύδονται και δεν ανταποδίδουν, τα μοντέλα υπόληψης καθιστούν ικανούς τους πράκτορες να συλλέγουν πιο πλούσιες πληροφορίες από το περιβάλλον τους και να βγάζουν ορθολογικά συμπεράσματα από τις πληροφορίες που λαμβάνονται από τους ομολόγους τους. Κοινωνικογνωστικά μοντέλα υιοθετούν ένα σχετικά υψηλό επίπεδο άποψης της εμπιστοσύνης που λαμβάνει ως δεδομένη την γνώση των κινήτρων των άλλων πρακτόρων και προτείνει τρόπους επιχειρηματολόγησης για αυτά τα κίνητρα [31].

### 3.3.1 Μαθησιακή και Εξελικτική Εμπιστοσύνη

Η εμπιστοσύνη είναι ένα αναδυόμενη ιδιότητα άμεσων αλληλεπιδράσεων μεταξύ ιδιοτελών πρακτόρων. Οι πράκτορες αλληλεπιδρούν πολλές φορές καθώς αυτό συμπίπτει με την αντίληψη της εμπιστοσύνης ως ένα κοινωνικό φαινόμενο που βασίζεται έμφυτα στις πολλαπλές αλληλεπιδράσεις μεταξύ δύο ατόμων. Επιπλέον, οι πράκτορες διαθέτουν ένα κίνητρο να αποστατήσουν. Για παράδειγμα, η αποστάτηση από μια αλληλεπίδραση θα μπορούσε να σημαίνει ότι ο πράκτορας δεν ικανοποιεί τους όρους ενός συμβολαίου, πουλά αγαθά χαμηλής ποιότητας, παραδίδει αργά ή δεν πληρώνει το απαιτούμενο χρηματικό ποσό σε έναν πωλητή. Στα συγκεκριμένα παραδείγματα, η αποστάτηση θα μπορούσε να επιφέρει υψηλότερες απολαβές για τον πράκτορα (π.χ. ο πωλητής λαμβάνει περισσότερα χρήματα από την πραγματική αξία των αγαθών που πουλήθηκαν) και να προκαλέσει μείωση της χρησιμότητας (π.χ. ο αγοραστής χάνει χρησιμότητα αγοράζοντας χαμηλής ποιότητας προϊόντα σε υψηλή τιμή). Παρ' όλα αυτά, η αποστάτηση μπορεί να μειώσει την πιθανότητα μελλοντικών αλληλεπιδράσεων καθώς ο χαμένος πράκτορας τυπικά θα αποφύγει να ρισκάρει στο μέλλον έχει απώλειες χρησιμότητας. Αντιθέτως, αν και οι δύο συμμετέχοντες μιας αλληλεπίδρασης συνεργαστούν, υποθέτουμε ότι θα λάβουν μεγαλύτερες απολαβές μακροπρόθεσμα. Για παράδειγμα, ένας πράκτορας που παραδίδει αγαθά εμπρόθεσμα ή πουλά αγαθά υψηλής ποιότητας μπορεί να οδηγήσει σε μελλοντικές αγορές από τον

αγοραστή. Σε όλες αυτές τις περιπτώσεις, υποθέτουμε ότι οι πράκτορες γνωρίζουν εκ των προτέρων τις απολαβές που σχετίζονται με κάθε μια από τις πράξεις τους.

Αν οι απολαβές για κάθε αναμέτρηση είναι γνωστή, οι πράκτορες μπορούν να επιχειρηματολογήσουν στρατηγικά εκτιμώντας την καλύτερη πιθανή κίνηση του αντιπάλου τους και συνεπώς καταστρώνοντας την καλύτερη απάντηση. Αυτή η ανάλυση εμπίπτει στην σφαίρα της θεωρίας παιγνίων η οποία συσχετίζει όλες τις αλληλεπιδράσεις ως παιχνίδια με διαφορετικές απολαβές (π.χ. κερδίζει ή χάνει την παρτίδα) για κάθε παίκτη ξεχωριστά. Στα περισσότερα παιχνίδια η κίνηση του αντιπάλου δεν θεωρείται γνωστή και σε τέτοια παιχνίδια μιας ευκαιρίας η πιο ασφαλής κίνηση (αυτός που ελαχιστοποιεί πιθανές απώλειες), αλλά όχι απαραίτητα και η πιο αποδοτική, θα επιλεγεί εκτός αν υπάρχει κάποιος τρόπος να επιβεβαιωθεί ότι το άλλο μέλος είναι αξιόπιστο. Συνεπώς, αν ένας πράκτορας πιστεύει ότι ο ομόλογός τους είναι ανταποδοτικός, τότε δεν θα αποστατήση ποτέ, ενώ στην αντίθετη περίπτωση θα αποστατήση με αποτέλεσμα και οι δύο να καταλήξουν με χαμηλότερες απολαβές απ' ό,τι αν εμπιστεύονταν ο ένας τον άλλο ή αν μάθαιναν να εμπιστεύονται ο ένας τον άλλο. Αυτή η πεποίθηση μπορεί να αποκτηθεί μόνο στην περίπτωση όπου το παιχνίδι επαναλαμβάνεται πολλές φορές έτσι ώστε να υπάρχει η ευκαιρία για τους πράκτορες να μάθουν την στρατηγική του αντιπάλου τους ή να προσαρμοστούν ο ένας στην στρατηγική του άλλου [31].

### 3.3.1.1 Στρατηγικές Μάθησης και Εξέλιξης

Είναι απαραίτητο ένας πράκτορας να μπορεί να προσαρμόζει την στρατηγική τους ανάλογα με τον τύπο του περιβάλλοντος που αντιμετωπίζουν έτσι ώστε να ελαχιστοποιήσει τις απώλειες και να προάγει την συνεργασία. Δίνοντας την δυνατότητα στους πράκτορες να προσαρμόζονται, οι Wu και Sun [32] απέδειξαν ότι η εμπιστοσύνη μπορεί στην πραγματικότητα να προκύψει μεταξύ τους. Αυτό σημαίνει ότι οι πράκτορες εξελίσσονται σε μια σχέση εμπιστοσύνης εκτιμώντας τα οφέλη κάθε πιθανής στρατηγικής σε πολλαπλές αλληλεπιδράσεις. Ένας πολυπρακτορικός πλειστηριασμός, στον οποίο ένας αριθμός πρακτόρων πλειοδοτεί για συμβόλαια σε μια ηλεκτρονική αγορά, χρησιμοποιείται ως παράδειγμα. Αρχικά, οι πράκτορες είναι όλοι συνεπείς μεταξύ τους (συνεργάζονται πάντα), και οι πωλητές μαθαίνουν να τους εκμεταλλεύονται. Για να το αντιμετωπίσουν αυτό, οι συνεπείς πράκτορες μαθαίνουν να χρησιμοποιούν την τεχνική 'μία σου και μία μου' (tit-for-tat) για να ελαχιστοποιήσουν τις απώλειές τους. Αυτό έχει ως αποτέλεσμα, οι κακοπροαίρετοι πωλητές (οι πράκτορες που εκμεταλλεύονται) μαθαίνουν να είναι ανταποδοτικοί καθώς η συνεργασία μπορεί να επιφέρει περισσότερα οφέλη μακροπρόθεσμα. Επομένως, η εμπιστοσύνη αναπτύσσεται ως αποτέλεσμα της εξέλιξης της στρατηγικής σε πολλαπλές αλληλεπιδράσεις. Αυτό το παράδειγμα δείχνει επιπλέον ότι η εξέλιξη της στρατηγικής επιτρέπει στους έντιμους πράκτορες να κερδίζουν τους κακοπροαίρετους μακροπρόθεσμα. Παρ' όλα αυτά, το συγκεκριμένο παράδειγμα δεν λαμβάνει υπόψη τις απώλειες στην χρησιμότητα βραχυπρόθεσμα όταν το ένα μέλος συνεργάζεται με το άλλο (π.χ. δίνοντας κάποιους πόρους).

Από αυτή την πλευρά, αναγνωρίζοντας το κόστος της συνεργασίας, ο Sen [33] αποδεικνύει πως η αμοιβαιότητα μπορεί να προκύψει όταν οι πράκτορες μαθαίνουν να προβλέπουν ότι θα μελλοντικά θα επωφεληθούν από την συνεργασία. Οι Mukherjee *et*

*al.* [34] απέδειξαν ότι η εμπιστοσύνη μπορεί να αποκτηθεί αν οι πράκτορες γνωρίζουν εκ των προτέρων την κίνηση του αντιπάλου τους.

Μέχρι αυτό το σημείο, όλα τα προαναφερθέντα μοντέλα ασχολούνται αυστηρά με το πρόβλημα συνεργασίας μεταξύ ιδιοτελών μερών. Όμως, δεν είναι όλες οι πολυπρακτορικές αλληλεπιδράσεις ανταγωνιστικές. Για παράδειγμα, οι πράκτορες μπορεί να είναι ιδιοτελείς, αλλά ταυτόχρονα να χρειάζεται να επιτευχθεί η μέγιστη απόδοση ως ομάδα ή κοινωνία καθώς αυτή προσδιορίζει την ατομική απόδοση του καθενός. Σύμφωνα με τον Birk [35, 36], η εμπιστοσύνη δεν προκύπτει αποκλειστικά μόνο μέσω στρατηγικών εξέλιξης, αλλά μπορεί να προκύψει αυστηρά μέσω της μάθησης. Χρησιμοποιώντας το δίλημμα του φυλακισμένου φτάνει στο συμπέρασμα ότι μια στρατηγική συνεργασίας κυριαρχεί σταδιακά σε ένα περιβάλλον όπου οι κακοπροαίρετοι πράκτορες (αυτοί που συμβάλουν λιγότερο) πλειοψηφούν. Αυτό συμβαίνει επειδή η χαμηλή επένδυση που παρέχεται από την κοινωνία επιδρά αρνητικά και στην χρησιμότητα του κάθε μέλους, αναγκάζοντας τους τελευταίους να μάθουν να συνεργάζονται για να έχουν μεγαλύτερη απόδοση. Καθώς όμως ο αριθμός των πρακτόρων που συνεργάζονται αυξάνεται, οι πράκτορες μαθαίνουν ξανά να αποστατούν έτσι ώστε να έχουν καλύτερη απόδοση. Επιπλέον, τα αποτελέσματα της έρευνας του Birk δείχνουν ότι η κοινωνία φτάνει σε μια ισορροπία με μεγάλο επίπεδο εμπιστοσύνης (ή συνεργασίας) μεταξύ των μελών της.

Τα παραπάνω μοντέλα μάθησης και εξέλιξης των πολυπρακτορικών στρατηγικών αλληλεπιδράσεων θεωρούν ότι οι αλγόριθμοι μάθησης των πολυπρακτορικών συστημάτων δουλεύουν έχοντας όλες τις πληροφορίες (π.χ. στρατηγικές, πίνακα αποδόσεων). Αυτά τα αποτελέσματα συνήθως προκύπτουν μέσω προσομοιώσεων χρησιμοποιώντας αυστηρές υποθέσεις και σκηνικά και όχι πραγματικά σενάρια όπου η κυρίαρχη προϋπόθεση της πλήρους γνώσεως των αποδόσεων δεν ισχύει. Επίσης, τα περισσότερα μοντέλα μάθησης αντιλαμβάνονται την έκβαση των αλληλεπιδράσεων ως δισταθή, δηλαδή είτε οι πράκτορες αποστατούν είτε συνεργάζονται. Στην πραγματικότητα, οι πράκτορες θα πρέπει να συμπεραίνουν, από τις πληροφορίες που έχουν συλλέξει μέσω των άμεσων αλληλεπιδράσεών τους, την απόδοση των αντιπάλων τους και το πώς η απόδοσή τους επηρεάζει τους στόχους τους. Αυτό οδηγεί στον σχεδιασμό ρεαλιστικών μέτρων εμπιστοσύνης [31].

### 3.3.1.2 Μέτρο Εμπιστοσύνης

Για να μπορέσει ένας πράκτορας να διαμορφώσει υπολογιστικά την εμπιστοσύνη στον αντίπαλό του, θα πρέπει αρχικά να αποδώσει μια εκτίμηση του επιπέδου της απόδοσης του αντιπάλου του. Η απόδοση τότε του αντιπάλου έπειτα από πολλαπλές αλληλεπιδράσεις μπορεί να αξιολογηθεί έτσι ώστε να ελεγχθεί το κατά πόσο είναι καλός και συνεπής στις πράξεις του. Επομένως, εκτός από την απόδοση, ένα πράκτορας χρειάζεται επίσης ένα μέσο παρακολούθησης της απόδοσης των πρακτόρων στις άμεσες αλληλεπιδράσεις τους. Για τον σκοπό αυτό οι Witkowski *et al.* [37] προτείνουν ένα μοντέλο όπου η εμπιστοσύνη ενός πράκτορα υπολογίζεται σύμφωνα με την απόδοσή τους στις προηγούμενες αλληλεπιδράσεις τους. Το συγκεκριμένο μοντέλο απλοποιεί τον υπολογισμό της εμπιστοσύνης μέσω εξισώσεων που χρησιμοποιούν μετρήσιμες ποσότητες όπως την κατανομή και την χρήση του εύρους ζώνης.

Άλλα μοντέλα ([38-40]) θεωρούν την απόδοση ενός πράκτορα ως μια δισταθή τιμή (καλοπροαίρετος – κακοπροαίρετος). Ενώ αυτά τα μοντέλα επιτυγχάνουν σε ένα προκαθορισμένο περιβάλλον προσομοίωσης, δεν μπορούν να χρησιμοποιηθούν ευρέως, καθώς οι πραγματικές αλληλεπιδράσεις σε ένα ανοιχτό καταναμημένο σύστημα περιλαμβάνουν πιο πλούσια αποτελέσματα. Γι' αυτό το σκοπό, οι Sabater και Sierra [41] ανέπτυξαν το σύστημα REGRET (σύστημα μετάνοιας) το οποίο, όχι μόνο περιορίζει την συνολική απόδοση σε μια δισταθή τιμή ή σε ένα αποτελεσματικό μέτρο, αλλά επιπλέον αποδίδει κάποια ασάφεια στον όρο της απόδοσης. Αντίθετα με το μοντέλο των Witkowski *et al.*, η εκτίμηση της εμπιστοσύνης με το σύστημα REGRET δεν βασίζεται μόνο στην άμεση αντίληψη ενός πράκτορα για την αξιοπιστία του αντιπάλου του, αλλά επιπλέον αξιολογεί την συμπεριφορά του με άλλους πράκτορες του συστήματος [31].

### 3.3.2 Μοντέλα Φήμης

Ο όρος φήμη αναφέρεται στην γνώμη ή την άποψη κάποιου ατόμου για κάτι.[41] Όσον αφορά τα πολυπρακτορικά συστήματα, θεωρούμε ότι αυτήν η άποψη μπορεί να προέρχεται κυρίως από την συγκέντρωση των απόψεων των μελών μιας κοινωνίας σχετικά με ένα από αυτά. Η φήμη μπορεί να είναι χρήσιμη σε συστήματα όπου αλληλεπιδρά ένας μεγάλος αριθμός πρακτόρων.

Οι διαφορετικές πτυχές της φήμης χωρίζουν το πεδίο στις παρακάτω ερευνητικές περιοχές:

- Σχεδιασμό μεθόδων για την συλλογή αξιολογήσεων που καθορίζουν την αξιοπιστία ενός πράκτορα χρησιμοποιώντας τις υπάρχουσες σχέσεις μεταξύ των μελών της κοινωνίας.
- Σχεδιασμό αξιόπιστων λογικών μεθόδων για την συλλογή όσο περισσότερων πληροφοριών από την ομάδα των αξιολογήσεων που έχουν ανακτηθεί από την κοινωνία.
- Σχεδιασμό μηχανισμών για την προώθηση αξιολογήσεων που περιγράφουν πραγματικά την αξιοπιστία του πράκτορα.

Προκειμένου να οργανώσουν την ανάκτηση και την ομαδοποίηση των αξιολογήσεων από άλλους πράκτορες, τα περισσότερα μοντέλα φήμης χρησιμοποιούν την ιδέα του κοινωνικού δικτύου. Σύμφωνα με αυτό, οι πράκτορες σχετίζονται μεταξύ τους κάθε φορά που αναλαμβάνουν ρόλους που τους διασυνδέουν ή όταν δημιουργούν δεσμούς επικοινωνίας μεταξύ τους. Μέσω αυτού του δικτύου κοινωνικών σχέσεων, οι πράκτορες λειτουργούν ως 'μάρτυρες' αλληλεπιδράσεων και μεταδίδουν πληροφορίες ο ένας για τον άλλο. Με αυτόν τον τρόπο οι πληροφορίες παίρνουν την μορφή αξιολόγησης των επιδόσεων, που μπορεί να διαμοιράζεται στους διάφορους κόμβους του κοινωνικού δικτύου δίνοντας την γενική ιδέα της φήμης [31].

### 3.3.3 Κοινωνικογνωστικά Μοντέλα Εμπιστοσύνης

Οι προσεγγίσεις των μοντέλων εμπιστοσύνης ατομικού επιπέδου που περιγράφηκαν στις προηγούμενες παραγράφους βασίζονται στην εκτίμηση των αποτελεσμάτων των

αλληλεπιδράσεων. Ωστόσο, για την εκτίμηση της αξιοπιστίας ενός αντιπάλου, μπορεί να είναι εξίσου σημαντικό να ληφθεί υπόψη η υποκειμενική αντίληψη για τον τελευταίο, καθώς επιτρέπει μια πιο ολοκληρωμένη ανάλυση των χαρακτηριστικών του αντιπάλου[42]. Αυτού του είδους οι πεποιθήσεις ή αντιλήψεις συνήθως αποθηκεύονται στην νοητική κατάσταση ενός πράκτορα και είναι σημαντική στην αξιολόγηση της αξιοπιστίας ενός πράκτορα για το ότι θα τηρήσει την συμφωνία.

Η συγκεκριμένη προσέγγιση επιλέγει την ανάθεση εργασιών, όπου ένα πράκτορας  $x$  επιθυμεί να αναθέσει μια εργασία σε έναν πράκτορα  $y$ . Για να το κάνει αυτό, ο πράκτορας  $x$  πρέπει να αξιολογήσει την εμπιστοσύνη που μπορεί να θέσει στον  $y$  λαμβάνοντας υπόψη τις διάφορες πεποιθήσεις που έχει για τα κίνητρα του πράκτορα  $y$ . Οι παρακάτω πεποιθήσεις είναι σημαντικές για τον καθορισμό του ύψους της εμπιστοσύνης που πρέπει να τεθεί στον πράκτορα  $y$  από τον πράκτορα  $x$ :

- **Ικανότητα:** Μια θετική αξιολόγηση του  $y$  από τον  $x$  λέγοντας ότι ο  $y$  είναι ικανός να διεξάγει την εργασία που του ανατέθηκε όπως ήταν αναμενόμενο. Αν ο πράκτορας  $y$  δεν είναι ικανός, δεν υπάρχει λόγος να τον εμπιστευθούν όσον αφορά την ολοκλήρωση μιας εργασίας.
- **Θέληση:** Ο  $x$  πιστεύει ότι ο  $y$  έχει αποφασίσει και πρόκειται να κάνει αυτό που του έχουν προτείνει. Αν ο πράκτορας  $y$  πιστεύεται ότι δεν είναι πρόθυμος να κάνει την εργασία, μπορεί να ψεύδεται λέγοντας ότι θέλει να την κάνει. Αυτό επομένως μειώνει την εμπιστοσύνη του  $x$  στον  $y$ .
- **Επιμονή:** Ο  $x$  πιστεύει ότι ο  $y$  είναι αρκετά σταθερός όσον αφορά την πρόθεσή του να κάνει αυτό που του έχουν προτείνει. Αν ο  $y$  είναι γνωστό ότι είναι ασταθής, τότε υπάρχει επιπλέον ρίσκο στη αλληλεπίδραση με τον  $y$ , και ως εκ τούτου στον  $y$  θα τεθεί χαμηλή εμπιστοσύνη αν και μπορεί να είναι πρόθυμος να κάνει την εργασία την στιγμή που του ανατίθεται η εργασία.
- **Κίνητρο:** Ο  $x$  πιστεύει ότι ο  $y$  έχει κίνητρο να βοηθήσει τον  $x$ , και ότι αυτό το κίνητρο πιθανόν θα υπερισχύσει έναντι άλλων κινήτρων που είναι εναντίον του  $x$  σε περίπτωση διαμάχης. Τα κίνητρα που αναφέρονται εδώ είναι τα ίδια με τα μακροχρόνια κέρδη που αποκτούνται βοηθώντας τον  $x$  να επιτύχει τους στόχους του. Αν ο  $y$  πιστεύεται ότι έχει κίνητρα, τότε ο  $x$  τείνει να τον εμπιστευτεί.

Για να υπολογίσει το επίπεδο εμπιστοσύνης που μπορεί να θέσει ένας πράκτορας  $x$  σε έναν πράκτορα  $y$ , ο πράκτορας  $x$  θα πρέπει να λάβει υπόψη κάθε μία από τις παραπάνω πεποιθήσεις (και πιθανώς και άλλες). Αυτές οι πεποιθήσεις επιδρούν στην εμπιστοσύνη, κάθε μία με τον δικό της τρόπο, και θα πρέπει να λαμβάνονται υπόψη σε μια ολοκληρωμένη αξιολόγηση όλων των ενδιαφερόμενων πεποιθήσεων.

Η κοινωνικογνωστική προσέγγιση για την μοντελοποίηση της εμπιστοσύνης, αν και λαμβάνει μια υψηλού επιπέδου άποψη του θέματος, στερείται την ορθολογική βάση των ορθολογικών μηχανισμών που παρέχουν τα μοντέλα και οι μηχανισμοί μάθησης και φήμης. Στην πραγματικότητα, η κοινωνικογνωστική προσέγγιση θα μπορούσε να εκμεταλλευτεί την αξιολόγηση που πραγματοποιείται από αυτά τα μοντέλα έτσι ώστε να διαμορφώσει τις κεντρικές πεποιθήσεις που παρουσιάζονται παραπάνω. Έτσι, όλα τα ατομικά μοντέλα εμπιστοσύνης θα μπορούσαν να συνεισφέρουν σε μια ολοκληρωμένη εκτίμηση της εμπιστοσύνης σε ατομικό επίπεδο. Αυτό θα λάμβανε

υπόψη στρατηγικές που έχουν αντληθεί έπειτα από πολλαπλές αλληλεπιδράσεις, την φήμη πιθανών παρτενέρ αλληλεπίδρασης, και τέλος τα πιθανά κίνητρα και τις ικανότητες του παρτενέρ σχετικά με την αλληλεπίδραση. Ωστόσο, μπορεί να είναι υπολογιστικά δαπανηρό για έναν πράκτορα να κρίνει όλους τους διαφορετικούς παράγοντες που επηρεάζουν την εμπιστοσύνη στους αντιπάλους του. Επιπλέον, δεδομένου ότι οι πράκτορες διαθέτουν μειωμένη ικανότητα συλλογής πληροφοριών από τις διάφορες πηγές του περιβάλλοντός τους, αντί να επιβάλλουν την ανάγκη ανάπτυξης εμπιστοσύνης σε ατομικό επίπεδο, μπορούν επικεντρωθούν στους κανόνες συνάντησης έτσι ώστε να διασφαλίσουν την αξιοπιστία των παρτενέρ τους. Με αυτόν τον τρόπο, οι κανόνες συνάντησης μπορούν να αντισταθμίσουν την περιορισμένη εφαρμογή των μοντέλων εμπιστοσύνης ατομικού επιπέδου [31].

### 3.4 Εμπιστοσύνη σε Επίπεδο Συστήματος

Στα ανοικτά πολυπρακτορικά συστήματα, θεωρούμε ότι οι πράκτορες αλληλεπιδρούν μέσω ενός αριθμού μηχανισμών ή πρωτοκόλλων που υπαγορεύουν τους κανόνες συνάντησης. Παραδείγματα τέτοιων μηχανισμών είναι οι δημοπρασίες, η ψηφοφορία, οι συμβάσεις εντός δικτύου, οι μηχανισμοί αγοραπωλησίας και οι διαπραγματεύσεις. Αυτοί οι μηχανισμοί θεωρούν ότι οι πράκτορες είναι εξολοκλήρου ιδιοτελείς και ως εκ τούτου πρέπει να βεβαιωθούν ότι οι κανόνες συναντήσεων θα προλαμβάνουν τα ψέματα και τις συμπαιγνίες μεταξύ των συμμετεχόντων. Αν και η θέσπιση αυτών των κανόνων δημιουργεί μια δυσκαμψία στο σύστημα, από την άλλη πλευρά δημιουργούν ένα κλίμα εμπιστοσύνης μεταξύ των πρακτόρων λόγω των διαφόρων περιορισμών που επιβάλλονται από το σύστημα. Αυτοί οι περιορισμοί μπορούν να εφαρμοστούν με διάφορους τρόπους. Πρώτον, είναι μερικές φορές δυνατόν να κατασκευαστεί το πρωτόκολλο των αλληλεπιδράσεων με τέτοιο τρόπο ώστε οι πράκτορες που συμμετέχουν να μην επωφελούνται σε χρησιμότητα με τα ψέματα και τις συμπαιγνίες. Δεύτερον, η φήμη ενός πράκτορα ως ψεύτης (ή ειλικρινής) μπορεί να διαδοθεί από το σύστημα. Επομένως, γνωρίζοντας ότι οι μελλοντικές τους αλληλεπιδράσεις τίθενται σε κίνδυνο αν έχουν την φήμη ότι είναι ψεύτες, οι πράκτορες εμμέσως αναγκάζονται να φέρονται σωστά (μέχρι να εγκαταλείψουν το σύστημα). Τρίτον, οι πράκτορες μπορεί να υποβληθούν σε έλεγχο κατά την είσοδό τους στο σύστημα παρέχοντας αποδείξεις για την αξιοπιστία τους μέσω συστάσεων από μια έμπιστη τρίτη πηγή.

Με βάση τα παραπάνω, η εμπιστοσύνη σε επίπεδο συστήματος μπορεί να χωριστεί στις εξής κατηγορίες:

- Σχεδίαση πρωτοκόλλων αλληλεπίδρασης που εκμαιεύουν την αλήθεια.
- Ανάπτυξη μηχανισμών φήμης που ενθαρρύνουν την ειλικρινή συμπεριφορά.
- Ανάπτυξη μηχανισμών ασφάλειας που διασφαλίζουν την αξιοπιστία των νεοεισερχόμενων πρακτόρων [31].

### 3.4.1 Πρωτόκολλα Αλληλεπίδρασης

Τα τελευταία χρόνια έχουν σχεδιαστεί αρκετά πρωτόκολλα και μηχανισμοί που διασφαλίζουν την ειλικρίνεια μεταξύ των πρακτόρων σε μια αλληλεπίδραση. Αυτά τα πρωτόκολλα στοχεύουν στην πρόληψη του ψεύδους και της κερδοσκοπίας των πρακτόρων σε μια αλληλεπίδραση. Αυτό το επιτυγχάνουν επιβάλλοντας κανόνες που υπαγορεύουν τα επιμέρους βήματα μιας αλληλεπίδρασης και των πληροφοριών που αποκαλύπτονται από τους πράκτορες κατά την διάρκειά της. Συνεπώς, με την τήρηση των συγκεκριμένων πρωτοκόλλων αναμένεται ότι οι πράκτορες δεν θα έχουν καμιά καλύτερη επιλογή πέρα της ειλικρίνειας. Το πιο δημοφιλές πρωτόκολλο αλληλεπίδρασης στις εφαρμογές πολυπρακτορικών συστημάτων είναι οι δημοπρασίες.

Υπάρχουν τέσσερις κύριοι τύποι μονομερών πλειστηριασμών, οι πλειστηριασμοί Αγγλικού τύπου, οι πλειστηριασμοί Ολλανδικού τύπου, οι πλειστηριασμοί κλειστών προσφορών ενός κύκλου και οι πλειστηριασμοί Vickrey, όπως περιγράφηκαν και στην παράγραφο 2.3.1.3.1. Οι πλειστηριασμοί Αγγλικού και Ολλανδικού τύπου επιβάλλουν την ειλικρίνεια από την πλευρά του πλειστηριαστή καθώς οι προσφορές γίνονται δημόσια. Παρ' όλα αυτά, οι πλειστηριασμοί Αγγλικού τύπου, Ολλανδικού τύπου και κλειστών προσφορών ενός κύκλου δεν διασφαλίζουν την πραγματική αποτίμηση των αγαθών που διακυβεύονται, καθώς η κυρίαρχη στρατηγική σε τέτοιου είδους πλειστηριασμούς είναι είτε η αποκάλυψη χαμηλότερης αξιολόγησης είτε η υποβολή προσφοράς ενός μικρού ποσού περισσότερο από την τρέχουσα υψηλότερη προσφορά από την πραγματική αποτίμηση κάποιου. Αντίθετα, οι πλειστηριασμοί Vickrey επιβάλλουν την ειλικρίνεια από τους πλειοδότες καθώς η κυρίαρχη στρατηγική είναι η υποβολή προσφοράς με βάση την πραγματική αποτίμηση καθώς σε οποιαδήποτε άλλη περίπτωση, λόγω της αβεβαιότητας για τις άλλες προσφορές και την τελική τιμή, θα μπορούσε να οδηγήσει σε κάποια απώλεια χρησιμότητας. Η υποβολή μεγαλύτερης προσφοράς από την πραγματική αποτίμηση θα μπορούσε να αναγκάσει τον πράκτορα να πληρώσει περισσότερα από την εκτίμησή του, ενώ η υποβολή χαμηλότερης προσφοράς θα μπορούσε να καταλήξει στο να χάσει ο πράκτορας στον πλειστηριασμό. Παρ' όλα αυτά, το κύριο μειονέκτημα του πλειστηριασμού Vickrey είναι ότι δεν μπορεί να διασφαλίσει την ειλικρίνεια εκ μέρους του πλειστηριαστή εφόσον οι προσφορές είναι κρυφές και γνωστές μόνο σε αυτόν.

Ωστόσο, όλοι οι μηχανισμοί που αναφέρθηκαν πιο πάνω δεν εξασφαλίζουν ότι στο σύστημα δε θα προκύψουν συμπαιγνίες, που σημαίνει ότι οι πράκτορες μπορούν να συνεργαστούν έτσι ώστε να ξεγελάσουν τον μηχανισμό ανταλλάσσοντας πληροφορίες για τις προσφορές τους. Οι συμπαιγνίες προϋποθέτουν ότι οι πράκτορες γνωρίζονται εκ των προτέρων και επομένως υποβάλουν προσφορές που δεν εκφράζουν τις πραγματικές τους προτιμήσεις. Επομένως, οι περισσότεροι πλειστηριασμοί δεν είναι ανθεκτικοί στο ψέμα και στις συμπαιγνίες αν δεν χρησιμοποιηθεί επιπλέον κάποιος μηχανισμός ασφάλειας [31].

### 3.4.2 Μηχανισμοί Φήμης

Τα μοντέλα φήμης, που περιγράφηκαν στην ενότητα 3.3.2, δεν λαμβάνουν υπόψη το γεγονός ότι οι πράκτορες είναι εγωιστές και επομένως δεν ανταλλάσσουν πληροφορίες χωρίς κάποιο αντάλλαγμα. Επιπλέον, τα συγκεκριμένα μοντέλα δεν προωθούν την χρήση της φήμης από κάποιους πράκτορες για να εκμαιεύσουν μια καλή συμπεριφορά από άλλους πράκτορες του συστήματος. Αυτά τα μοντέλα στοχεύουν στο να προσδώσουν στους πράκτορες μια καλύτερη αντίληψη για τους αντιπάλους τους και δεν λαμβάνουν υπόψη το αποτέλεσμα που μπορεί να έχει σε έναν αντίπαλο αν ο τελευταίος το γνωρίζει. Δεδομένων των ελλείψεων των μοντέλων φήμης, οι μηχανισμοί φήμης εξετάζουν το πρόβλημα τις επιβολής αξιόπιστης συμπεριφοράς και μοντελοποίησης της φήμης των πρακτόρων σε επίπεδο συστήματος. Οι μηχανισμοί αυτοί μπορούν να λειτουργούν είτε μέσω κεντρικών ή κατανεμημένων οντοτήτων που αποθηκεύουν τις αξιολογήσεις που παρέχονται από τους πράκτορες για τις αλληλεπιδράσεις τους με τους άλλους πράκτορες και στην συνέχεια τις δημοσιοποιούν, έτσι ώστε να έχουν πρόσβαση σε αυτές όλοι οι πράκτορες του συστήματος. Σε αυτήν την περίπτωση, το ίδιο το σύστημα διαχειρίζεται την συγκέντρωση και την ανάκτηση των αξιολογήσεων σε αντίθεση με τα μοντέλα φήμης που αφήνουν το έργο αυτό στους ίδιους τους πράκτορες. Με τον τρόπο αυτό οι μηχανισμοί φήμης μπορούν να χρησιμοποιηθούν για να αποτρέψουν το ψέμα και την κακή συμπεριφορά από την πλευρά των πρακτόρων. Επιπλέον, οι μηχανισμοί φήμης έχουν ως στόχο να προκαλέσουν ειλικρινείς αξιολογήσεις από μάρτυρες και στην πραγματικότητα να το κάνουν λογικό για τους πράκτορες να δίνουν αξιολογήσεις για τους άλλους στο σύστημα.

Τόσο οι μηχανισμοί φήμης όσο και τα πρωτόκολλα αλληλεπίδρασης προσπαθούν να επιβάλλουν την αξιόπιστη συμπεριφορά ελαχιστοποιώντας την ευκαιρία για τους πράκτορες να αποστατήσουν έτσι ώστε να έχουν μεγαλύτερες απολαβές. Στην περίπτωση όπου τα πρωτόκολλα αλληλεπίδρασης και οι μηχανισμοί φήμης δεν μπορούν να εγγηθούν μια αξιόπιστη συμπεριφορά εκ μέρους όλων των πρακτόρων του συστήματος, εξακολουθεί να υπάρχει ανάγκη να δοθεί στους πράκτορες ενός ανοικτού συστήματος η δυνατότητα να αποδείξουν την αξιοπιστία τους έτσι ώστε να καταστεί δυνατή η αναγνώρισή τους από τους άλλους πράκτορες ως αξιόπιστοι παρτενέρ αλληλεπίδρασης. Αυτό μπορεί να πραγματοποιηθεί μέσω συστάσεων από μια αναγνωρισμένη πηγή [31].

### 3.4.3 Μηχανισμοί Ασφάλειας

Στον τομέα της ασφάλειας των δικτύων, η εμπιστοσύνη χρησιμοποιείται για να περιγράψει την περίπτωση όπου ένας χρήστης μπορεί να αποδείξει αυτό που λέει ότι είναι [43]. Αυτό συνήθως συνεπάγεται ότι η ταυτότητά τους μπορεί να επικυρωθεί από αξιόπιστους τρίτους στο σύστημα. Αν και σε πρώτη ματιά αντιτίθεται στον ορισμό της εμπιστοσύνης, είναι μια βασική απαίτηση των μοντέλων και μηχανισμών εμπιστοσύνης γιατί βασίζονται στο γεγονός ότι οι πράκτορες αναγνωρίζονται από την ταυτότητά τους και επομένως απαιτείται η υλοποίηση πρωτοκόλλων ελέγχου ταυτότητας.



Για τον σκοπό αυτό, οι Poslad *et al.* [44] έχουν προτείνει έναν αριθμό προϋποθέσεων ασφαλείας που ισχυρίζονται ότι είναι απαραίτητα έτσι ώστε οι πράκτορες να εμπιστευονται ο ένας τον άλλο και τα μηνύματα που ανταλλάσσουν μεταξύ τους.

- **Ταυτότητα:** η ικανότητα προσδιορισμού της ταυτότητας μιας οντότητας. Αυτό μπορεί επίσης να περιλαμβάνει την ικανότητα προσδιορισμού της ταυτότητας του ιδιοκτήτη ενός πράκτορα.
- **Δικαιώματα Πρόσβασης:** η ικανότητα προσδιορισμού του είδους των δικαιωμάτων πρόσβασης που πρέπει να δοθούν σε έναν πράκτορα στο σύστημα, σύμφωνα με την ταυτότητά του.
- **Ακεραιότητα Περιεχομένων:** η ικανότητα προσδιορισμού αν ένα τμήμα λογισμικού, μηνύματος, ή άλλου είδους δεδομένων έχει τροποποιηθεί από την στιγμή που αποσταλεί από την πηγή προέλευσής του.
- **Απόρρητο Περιεχομένων:** η ικανότητα εξασφάλισης ότι μόνο οι προκαθορισμένες ταυτότητες μπορούν να εξετάσουν ένα μήνυμα ή κάποιο άλλο είδος δεδομένων. Στους υπόλοιπους οι πληροφορίες δεν είναι φανερές.

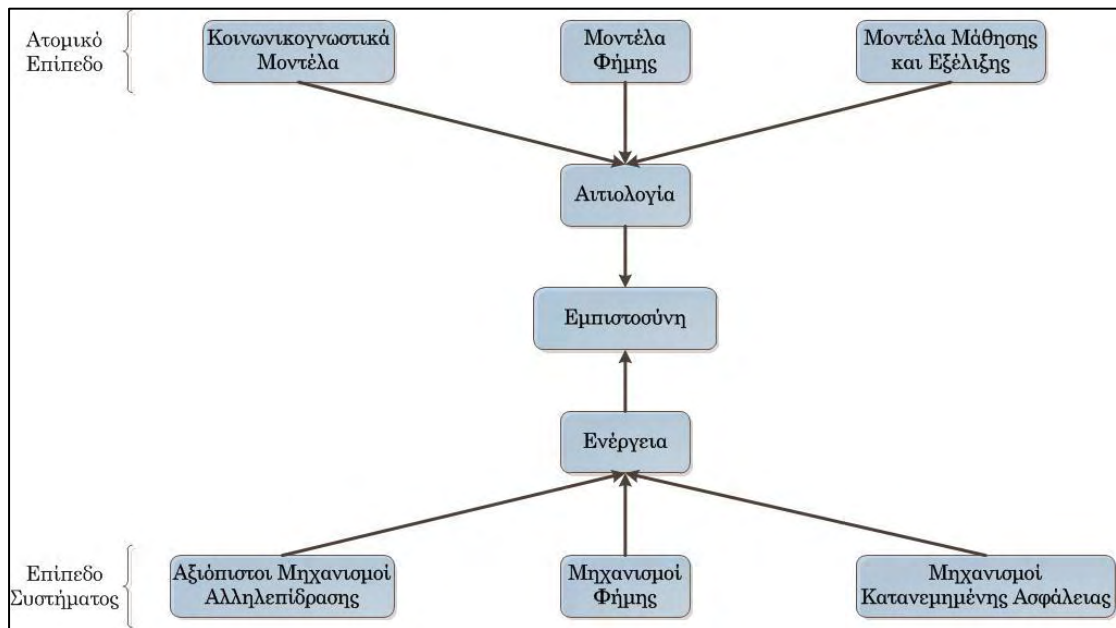
Οι συγκεκριμένες απαιτήσεις προσδιορίζονται και για την αφηρημένη αρχιτεκτονική FIPA (Foundation for Intelligent Physical Agents) και μπορούν να υλοποιηθούν με κρυπτογράφηση δημόσιου κλειδιού και μια υποδομή πιστοποιητικών. Ένα ψηφιακό πιστοποιητικό εκδίδεται από μια αρχή πιστοποίησης και επιβεβαιώνει ότι ένα δημόσιο κλειδί ανήκει από μια συγκεκριμένη οντότητα. Το δημόσιο κλειδί σε ένα πιστοποιητικό χρησιμοποιείται και για να κρυπτογραφηθεί και να υπογράψει ένα μήνυμα με τέτοιο τρόπο ώστε μόνο ο ιδιοκτήτης του να μπορεί να εξετάσει το περιεχόμενο και να βεβαιωθεί για την ακεραιότητά του. Τα δύο πιο δημοφιλή μοντέλα δημόσιου κλειδιού είναι το PGP (Pretty Good Privacy) και το μοντέλο εμπιστοσύνης X.509. Το πρώτο μοντέλο απευθύνεται σε ένα δίκτυο εμπιστοσύνης με την έννοια του ότι δεν υπάρχει κάποια κεντρική ή ιεραρχική σχέση μεταξύ των αρχών πιστοποίησης, ενώ το δεύτερο αποτελεί ένα αυστηρά ιεραρχικό μοντέλο εμπιστοσύνης για την ταυτοποίηση. Παρ' όλα αυτά, αυτά τα μέτρα πιστοποίησης δεν επαρκούν για την εξασφάλιση της ειλικρινούς και αξιόπιστης συμπεριφοράς των πρακτόρων στα ανοικτά πολυπρακτορικά συστήματα. Αποτελούν απλά ένα φράγμα κατά των πρακτόρων που δεν επιτρέπονται στο σύστημα ή μόνο επιτρέπουν την αναγνώρισή τους στο σύστημα. Προκειμένου να επιβληθεί η καλή συμπεριφορά σε ένα σύστημα, θα πρέπει να χορηγούνται πιστοποιητικά στους πράκτορες μόνο αν αυτοί πληρούν συγκεκριμένες προδιαγραφές που τους καθιστούν αξιόπιστους.

Για να επιτευχθεί αυτό, έμπιστες τρίτες πηγές είναι απαραίτητες για την έκδοση πιστοποιητικών σε πράκτορες που πληρούν τις προδιαγραφές της αξιοπιστίας (π.χ. είναι ανταποδοτικοί, αξιόπιστοι, ειλικρινείς). Για τον σκοπό αυτό, οι Herzberg *et al.* [45] παρουσιάζουν ένα μηχανισμό που βασίζεται σε πολιτικές και πιστοποιητικά, και ο οποίος μπορεί να αναθέσει ρόλους στους νεοεισερχόμενους. Το πιστοποιητικό υπογράφεται από κάποιον εκδότη και περιέχει κάποιες αξιώσεις για τον συγκεκριμένο πράκτορα. Δεν υπάρχουν περιορισμοί στο είδος των αξιώσεων και μπορεί να αφορούν την ιδιότητά του (π.χ. υπάλληλος εταιρίας), τις ικανότητές του, ακόμη και την αξιοπιστία του κατά την άποψη του εκδότη. Οι Mass και Shehory [43] επέκτειναν το έργο αυτό σε ανοικτά πολυπρακτορικά συστήματα [31].

### 3.5 Σύνοψη Μοντέλων Εμπιστοσύνης

Όπως παρουσιάζεται στην Εικόνα 3.1, ενώ τα μοντέλα εμπιστοσύνης ατομικού επιπέδου καθιστούν έναν πράκτορα ικανό να κρίνει το επίπεδο εμπιστοσύνης των αντιπάλων του, οι μηχανισμοί του επιπέδου του συστήματος στοχεύουν στην εξασφάλιση του ότι αυτές οι ενέργειες των αντιπάλων τους είναι στην πραγματικότητα αξιόπιστες. Πιο αναλυτικά, χρησιμοποιώντας τα μοντέλα εμπιστοσύνης τους, οι πράκτορες μπορούν:

- Να κρίνουν την στρατηγική που θα χρησιμοποιήσουν απέναντι σε αξιόπιστους και μη αξιόπιστους πράκτορες με τους οποίους θα αλληλεπιδράσουν (π.χ. να είναι ανταποδοτικοί ή εγωιστές απέναντί τους) δεδομένου του υπολογισμού των απολαβών από μελλοντικές αλληλεπιδράσεις.
- Να αξιολογήσουν τις πληροφορίες που συλλέγουν με διάφορα μέσα (π.χ. είτε άμεσα είτε έμμεσα μέσω των μοντέλων φήμης) για τους πιθανούς παρτενέρ σε μια αλληλεπίδραση.
- Να κρίνουν τα κίνητρα και τις ικανότητες των παρτενέρ τους σε κάθε αλληλεπίδραση έτσι ώστε να αποφασίζουν για τον αν θα πιστέψουν της αξιοπιστία τους.



**Εικόνα 3.1** Ταξινόμηση των προσεγγίσεων της εμπιστοσύνης σε πολυπρακτορικά συστήματα

Αντίθετα, οι μηχανισμοί και τα πρωτόκολλα που περιγράφηκαν για την εξασφάλιση της εμπιστοσύνης σε επίπεδο συστήματος, έχουν ως στόχο να εξαναγκάσουν τους πράκτορες να δρουν και να αλληλεπιδρούν με ειλικρίνεια με τους εξής τρόπους:

- Με την επιβολή όρων που θα τους αναγκάζει να χάνουν χρησιμότητα αν δεν τους ακολουθούν.
- Χρησιμοποιώντας την φήμη τους είτε για την προώθηση των μελλοντικών τους αλληλεπιδράσεων με άλλους πράκτορες στην κοινωνία ή για την αποφυγή μελλοντικών αλληλεπιδράσεων αν δεν συμπεριφέρονται σωστά.

- Επιβάλλοντας καθορισμένα πρότυπα καλής συμπεριφοράς που πρέπει να πληρούν και να διατηρούν έτσι ώστε να είναι αποδεκτοί στο σύστημα [31].

---

# Κεφάλαιο 4

## Λογική στα Πολυπρακτορικά Συστήματα

---

### 4.1 Εισαγωγή

Η λογική αποτελεί ένα πρότυπο για την μοντελοποίηση και την εκλογίκευση του κόσμου με έναν ακριβή τρόπο. Παρέχει λεξιλόγιο, εννοιολογικές δομές και μεθοδολογία για τις προδιαγραφές και τον έλεγχο.

Η κυρίαρχη προσέγγιση που χρησιμοποιείται στα πολυπρακτορικά συστήματα καλείται modal logic. Η κύρια ιδέα πίσω από την modal logic είναι η ανάπτυξη ενός λογικού μοντέλου που θα μπορεί να χρησιμοποιηθεί για τον χαρακτηρισμό της διανοητικής κατάστασης των πρακτόρων όπως αυτοί δρουν και αλληλεπιδρούν.

### 4.2 Modal Logic

Η Modal Logic είναι ένα είδος τυπικής λογικής, που αναπτύχθηκε αρχικά στην δεκαετία του 1960, και επεκτείνει την κλασική προτασιακή και κατηγορηματική λογική έτσι ώστε να περιλαμβάνει τελεστές που εκφράζουν τροπικότητα. Τα Modals, δηλαδή οι λέξεις που εκφράζουν τροπικότητα, χαρακτηρίζουν μια δήλωση. Για παράδειγμα η δήλωση «Ο John είναι χαρούμενος» θα μπορούσε να χαρακτηριστεί λέγοντας «Ο John είναι *συνήθως* χαρούμενος», όπου το η λέξη *συνήθως* αντιπροσωπεύει την λέξη modal. Τα παραδοσιακά Alethic Modalities, δηλαδή οι όροι της αλήθειας, περιλαμβάνουν την Πιθανότητα («Πιθανά,  $p$ », «Είναι πιθανό ότι  $p$ »), Ανάγκη («Αναγκαίως,  $p$ », «Είναι αναγκαίο ότι  $p$ ») και Αδυναμία («Είναι αδύνατο ότι  $p$ »). Άλλοι όροι που έχουν

τυποποιηθεί σε modal logic περιλαμβάνουν χρονικούς όρους («Ήταν η περίπτωση ότι  $p$ », «Ήταν πάντα ότι  $p$ », «Θα είναι  $p$ », «Θα είναι πάντα  $p$ »), δεοντολογικούς όρους («Είναι υποχρεωτικό ότι  $p$ », «Είναι επιτρεπτό ότι  $p$ »), επιστημικούς όρους ή όρους γνώσης («Είναι γνωστό ότι  $p$ ») και δοξαστικούς όρους ή όρους πεποιθήσεων («Πιστεύεται ότι  $p$ ») [46].

### 4.3 Σημασιολογία της Modal Logic

Η σημασιολογία της modal logic δίνεται συνήθως ως εξής: αρχικά ορίζουμε ένα πλαίσιο, το οποίο αποτελείται από ένα μη κενό σύνολο  $G$ , τα μέλη του οποίου γενικά ονομάζονται πιθανοί κόσμοι, και μια δυαδική σχέση  $R$ , που συνδέει (ή όχι) τους πιθανούς κόσμους του  $G$ . Αυτή η δυαδική σχέση καλείται *σχέση προσβασιμότητας*. Για παράδειγμα,  $w R v$  σημαίνει ότι ο κόσμος  $v$  είναι προσπελάσιμος μέσω του κόσμου  $w$ . Δηλαδή, η κατάσταση των πραγμάτων που είναι γνωστή ως  $v$  είναι ένα ζωντανό ενδεχόμενο για το  $w$ . Αυτό δίνει ένα ζευγάρι  $\langle G, R \rangle$ .

Στην συνέχεια, το πλαίσιο επεκτείνεται σε ένα μοντέλο προσδιορίζοντας τις τιμές αληθείας για όλες τις προτάσεις για κάθε έναν από τους κόσμους του  $G$ . Αυτό επιτυγχάνεται με τον ορισμό μιας σχέσης  $\models$  μεταξύ των πιθανών κόσμων και των προτασιακών γραμμάτων. Αν υπάρχει ένας κόσμος  $w$  τέτοιος ώστε  $w \models P$ , τότε το  $P$  είναι αληθές στο  $w$ . Ένα μοντέλο είναι επομένως μια ταξινομημένη τριάδα  $\langle G, R, \models \rangle$ .

Έτσι, ορίζεται αναδρομικά η αλήθεια ενός τύπου σε ένα μοντέλο ως εξής:

- $w \models \neg P$  αν και μόνο αν  $w \not\models P$
- $w \models (P \wedge Q)$  αν και μόνο αν  $w \models P$  και  $w \models Q$
- $w \models \Box P$  αν και μόνο αν για κάθε στοιχείο  $v$  του  $G$ , αν  $w R v$  τότε  $v \models P$
- $w \models \Diamond P$  αν και μόνο αν για κάθε στοιχείο  $v$  του  $G$ , ισχύει ότι  $w R v$  και  $v \models P$

Σύμφωνα με αυτήν την σημασιολογία, η αλήθεια είναι απαραίτητη σχετικά με ένα πιθανό κόσμο  $w$  αν είναι αλήθεια σε κάθε κόσμο που είναι προσπελάσιμος στον  $w$ , και δυνατή αν είναι αλήθεια σε κάποιον κόσμο που είναι προσπελάσιμος στον  $w$ . Η δυνατότητα επομένως εξαρτάται από την σχέση προσβασιμότητας  $R$ , που επιτρέπει την έκφραση της σχετικής φύσης της δυνατότητας [46].

### 4.4 Normal Modal Logic

Η επιστημική λογική συνήθως διατυπώνεται ως normal modal logic χρησιμοποιώντας την σημασιολογία που αναπτύχθηκε από τον Kripke [47]. Οι επιστημικοί όροι ασχολούνται με την βεβαιότητα των προτάσεων. Ο τελεστής  $\Box$  μεταφράζεται ως «Ο  $\chi$  γνωρίζει ότι...», και ο τελεστής  $\Diamond$  μεταφράζεται ως «Από όσα γνωρίζει ο  $\chi$ , μπορεί να είναι αλήθεια ότι...».

Για την ανάδειξη των αρχών της επιστημικής modal logic, αρκεί να οριστεί ένα απλό μοντέλο προτασιακής modal logic. Αυτή η λογική είναι ουσιαστικά η κλασική

προτασιακή λογική, με την προσθήκη δύο επιπλέον τελεστών: '□' (απαραιτήτως) και '◇' (πιθανώς).

Όσον αφορά την σύνταξη, έστω ότι  $Prop = \{p, q, \dots\}$  είναι ένα μετρήσιμο σύνολο ατομικών προτάσεων. Η σύνταξη του προτασιακού normal modal logic ορίζεται σύμφωνα με τους εξής κανόνες:

- Αν  $p \in Prop$ , τότε το  $p$  είναι ένας τύπος.
- Αν τα  $\varphi, \psi$  είναι τύποι, τότε το ίδιο ισχύει και για
 
$$\neg\varphi \quad \varphi \vee \psi$$
- Αν το  $\varphi$  είναι τύπος, τότε είναι και τα
 
$$\Box\varphi \quad \Diamond\varphi.$$

Οι τελεστές  $\neg$  (λογικό όχι) και  $\vee$  (λογικό ή) έχουν την καθορισμένη τους σημασία, το αληθές είναι μια λογική σταθερά που είναι πάντα αληθής. Οι υπόλοιποι σύνδεσμοι της προτασιακής λογικής μπορούν να οριστούν ως συντομογραφίες με τον συνηθισμένο τρόπο. Ο τύπος  $\Box\varphi$  διαβάζεται ως 'απαραιτήτως  $\varphi$ ', ενώ ο τύπος  $\Diamond\varphi$  διαβάζεται 'πιθανώς  $\varphi$ '.

Η Normal Modal Logic ασχολείται με την αλήθεια στους κόσμους, συνεπώς τα μοντέλα για τέτοια λογική περιέχουν ένα σύνολο κόσμων,  $W$ , και μια δυαδική σχέση,  $R$ , στο  $W$ , που καθορίζει τις σχέσεις μεταξύ των κόσμων. Επιπλέον, είναι απαραίτητη μια συνάρτηση αξιολόγησης  $\pi$ , που επισημαίνει τις προτάσεις που είναι αληθείς σε κάθε κόσμο.

Ένα μοντέλο ενός προτασιακού normal modal logic είναι μια τριάδα  $\langle W, R, \pi \rangle$ , όπου το  $W$  αντιπροσωπεύει ένα μη κενό σύνολο κόσμων,  $R \subseteq W \times W$ , και

$$\pi : W \rightarrow p(Prop)$$

είναι μια συνάρτηση αξιολόγησης, η οποία προσδιορίζει για κάθε κόσμο  $w \in W$  ποια ατομική πρόταση είναι αληθής στον  $w$ . Μια εναλλακτική, ισοδύναμη τεχνική θα ήταν να οριστεί το  $\pi$  ως εξής:

$$\pi : W \times Prop \rightarrow \{\text{true}, \text{false}\}$$

αν και οι κανονισμοί που ορίζουν την σημασιολογία της γλώσσας θα έπρεπε να αλλαχθούν ελαφρώς.

Η σημασιολογία της γλώσσας δίνεται μέσω της σχέσης ικανοποίησης, '⊨', που ισχύει μεταξύ ζευγών της μορφής  $\langle M, w \rangle$  (όπου το  $M$  είναι ένα μοντέλο και  $w$  είναι ο αναφερόμενος κόσμος), και τύπων της γλώσσας. Οι σημασιολογικοί κανόνες που ορίζουν αυτή την σχέση δίνονται στον πίνακα 4.1.

Ο ορισμός της ικανοποίησης των ατομικών προτάσεων συνεπώς συλλαμβάνει την ιδέα της αλήθειας στον 'τρέχον' κόσμο (ο οποίο εμφανίζεται στο αριστερό μέρος του '⊨'). Οι σημασιολογικοί κανόνες του 'true', '¬' και '∨' είναι καθιερωμένοι. Ο κανόνας του '□' συλλαμβάνει την ιδέα της αλήθειας σε όλους τους προσπελάσιμους κόσμους, ενώ ο κανόνας του '◇' συλλαμβάνει την ιδέα της αλήθειας σε τουλάχιστον έναν πιθανό κόσμο.

**Πίνακας 4.1** Σημασιολογία Normal Modal logic

$\langle M, w \rangle \models \text{true}$	
$\langle M, w \rangle \models p$	Όπου $p \in Prop$ , αν και μόνο αν $p \in \pi(w)$
$\langle M, w \rangle \models \neg\varphi$	Αν και μόνο αν $\langle M, w \rangle \not\models \varphi$
$\langle M, w \rangle \models \varphi \vee \psi$	Αν και μόνο αν $\langle M, w \rangle \models \varphi$ ή $\langle M, w \rangle \models \psi$
$\langle M, w \rangle \models \Box\varphi$	Αν και μόνο αν $\forall w' \in W \cdot \text{αν } (w, w') \in R \text{ τότε } \langle M, w' \rangle \models \varphi$
$\langle M, w \rangle \models \Diamond\varphi$	Αν και μόνο αν $\exists w' \in W \cdot (w, w') \in R \text{ και } \langle M, w' \rangle \models \varphi$

Ας σημειωθεί ότι οι δύο τροπικοί τελεστές είναι διπλοί με την έννοια ότι οι καθολικοί και υπαρξιακοί ποσοδείκτες της λογικής πρώτης τάξης είναι διπλοί:

$$\Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi.$$

Θα ήταν έτσι δυνατό να λαμβάνεται ο ένας ως αρχικός και ο άλλος ως συμπληρωματικός τελεστής [5].

### 4.4.1 Θεωρία Αντιστοιχίας

Για να γίνουν κατανοητές οι εξαιρετικές ιδιότητες αυτής της απλής λογικής, είναι αρχικά απαραίτητο να εισαχθούν οι έννοιες της εγκυρότητας και της ικανοποίησης. Έτσι ένας τύπος είναι:

- **Ικανοποιήσιμος** αν ικανοποιείται για κάποιο ζεύγος μοντέλου/κόσμου.
- **Μη ικανοποιήσιμος** αν δεν ικανοποιείται από κανένα ζεύγος μοντέλου/κόσμου.
- **Αληθής για ένα μοντέλο** αν ικανοποιείται για κάθε κόσμο στο μοντέλο.
- **Έγκυρος σε μια κατηγορία μοντέλων** αν είναι αληθής για κάθε μοντέλο της κατηγορίας.
- **Έγκυρος** αν είναι αληθής στην κατηγορία όλων των μοντέλων.

Αν το  $\varphi$  είναι έγκυρος, υποδεικνύεται γράφοντας  $\models\varphi$ . Ας σημειωθεί ότι η εγκυρότητα είναι ουσιαστικά ίδια με την έννοια της ‘ταυτολογίας’ στην κλασική προτασιακή λογική – όλες οι ταυτολογίες είναι έγκυρες.

Η συγκεκριμένη λογική έχει δύο βασικές ιδιότητες. Πρώτον, η ακόλουθη αξιωματική παράσταση είναι έγκυρη:

$$\models\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi).$$

Αυτό το αξίωμα ονομάζεται *K*, προς τιμήν του Kripke. Η δεύτερη ιδιότητα είναι η εξής:

$$\text{Αν } \models\varphi, \text{ τότε } \models\Box\varphi.$$

Εφόσον το *K* ισχύει, πλέον είναι ένα θεώρημα για κάθε ολοκληρωμένου αξιώματος της normal modal logic. Με παρόμοιο τρόπο, η δεύτερη ιδιότητα θα εμφανίζεται ως κανόνας αναφοράς σε κάθε αξίωμα της normal modal logic και γενικά ονομάζεται *κανόνας της αναγκαιότητας*.

Οι πιο ενδιαφέρουσες ιδιότητες της normal modal logic προέρχονται από τις ιδιότητες της σχέσης προσβασιμότητας,  $R$ , στα μοντέλα. Έστω ότι:

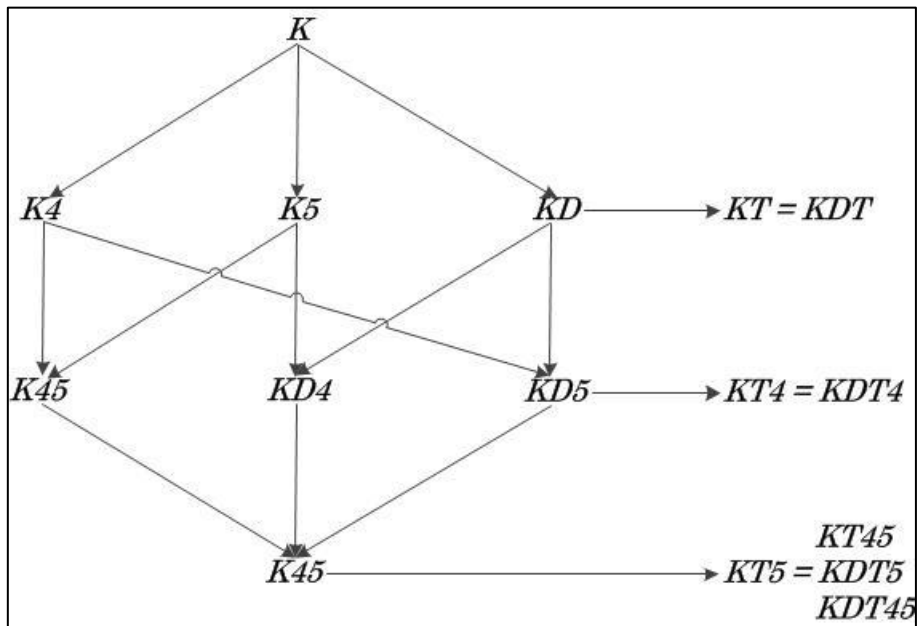
$$\Box\varphi \Rightarrow \varphi.$$

Προκύπτει τελικά ότι αυτό το αξίωμα είναι χαρακτηριστικό της τάξης των μοντέλων με αντανακλαστική σχέση προσβασιμότητας. Υπάρχει ένα πλήθος αξιωμάτων που αντιστοιχούν σε συγκεκριμένες ιδιότητες της  $R$ : η μελέτη του τόπου αντιστοίχισης των ιδιοτήτων του  $R$  με τα αξιώματα ονομάζεται *θεωρία αντιστοίχισης*. Στον πίνακα 4.2 παρουσιάζεται μια λίστα αξιωμάτων με την χαρακτηριστική τους ιδιότητα στην  $R$ , και έναν τύπο πρώτης τάξης που περιγράφει την ιδιότητα.

**Πίνακας 4.2** Θεωρία αντιστοίχισης

Ονομασία	Αξίωμα	Κατάσταση στην $R$	Χαρακτηρισμός Πρώτης Τάξης
T	$\Box\varphi \Rightarrow \varphi$	Αντανακλαστικό	$\forall w \in W \cdot (w, w) \in R$
D	$\Box\varphi \Rightarrow \Diamond\varphi$	Σειριακό	$\forall w \in W \cdot \exists w' \in R \cdot (w, w') \in R$
4	$\Box\varphi \Rightarrow \Box\Box\varphi$	Μεταβατικό	$\forall w, w', w'' \in W \cdot (w, w') \in R \wedge (w', w'') \in R \Rightarrow (w, w'') \in R$
5	$\Diamond\varphi \Rightarrow \Box\Diamond\varphi$	Ευκλείδειο	$\forall w, w', w'' \in W \cdot (w, w') \in R \wedge (w, w'') \in R \Rightarrow (w', w'') \in R$

Ένα λογικό σύστημα μπορεί να θεωρηθεί ως ένα σύστημα τύπων που είναι έγκυροι σε κάποια τάξη μοντέλων. Ένα μέλος αυτού του μοντέλου καλείται *θεώρημα* της λογικής (αν το  $\varphi$  είναι ένα θεώρημα, αυτό συνήθως δηλώνεται ως  $\vdash\varphi$ ). Η σημειογραφία  $K\Sigma_1 \dots K\Sigma_n$  χρησιμοποιείται συνήθως για να δηλώσει μικρότερα normal modal logic που περιέχουν αξιώματα  $\Sigma_1 \dots \Sigma_n$ .



**Εικόνα 4.1** Τα συστήματα modal βασισμένα στα αξιώματα T, D, 4 και 5



Για τα αξιώματα  $T, D, 4, 5$  του πίνακα 4.2 φαίνεται να υπάρχουν 16 διακριτά λογικά συστήματα (καθώς  $2^4 = 16$ ). Όμως, κάποια από αυτά τα συστήματα αποδεικνύεται ότι είναι ισοδύναμα καθώς περιέχουν τα ίδια θεωρήματα, με αποτέλεσμα να προκύπτουν μόνο 11 διακριτά λογικά συστήματα. Οι σχέσεις μεταξύ των συστημάτων περιγράφονται στην εικόνα 4.1. Το τόξο μεταξύ των  $A$  και  $B$  σημαίνει ότι το  $B$  είναι ένα αυστηρό υπερσύνολο του  $A$ : κάθε θεώρημα του  $A$  είναι και θεώρημα του  $B$ , αλλά το αντίστροφο δεν ισχύει. Αν  $A = B$  τότε τα  $A$  και  $B$  περιέχουν ακριβώς τα ίδια θεωρήματα [5].

#### 4.4.2 Normal Modal Logic ως Επιστημική Λογική

Για να χρησιμοποιηθεί η λογική που περιγράφηκε πιο πάνω ως επιστημική λογική, ο τύπος  $\Box\varphi$  θα διαβάζεται ως 'είναι γνωστό ότι  $\varphi$ '. Οι κόσμοι του μοντέλου ερμηνεύονται εναλλακτικά ως επιστημικοί και η σχέση προσβασιμότητας ορίζει ποιες είναι οι εναλλακτικές για κάθε δεδομένο κόσμο. Η συγκεκριμένη λογική ασχολείται με την γνώση ενός μόνο πράκτορα, ενώ για πολυπρακτορικά συστήματα θα πρέπει να προστεθεί ένα δεικτοδοτημένο σύνολο σχέσεων προσβασιμότητας, ένα για κάθε πράκτορα. Το μοντέλο τότε γίνεται μια δομή της μορφής:

$$\langle W, R_1, \dots, R_n, \pi \rangle,$$

Όπου το  $R_i$  αντιπροσωπεύει της σχέση προσβασιμότητας γνώσης ενός πράκτορα  $i$ . Η απλή γλώσσα που περιγράφηκε προηγουμένως επεκτείνεται αντικαθιστώντας τον μονό modal τελεστή ' $\Box$ ' με ένα δεικτοδοτημένο σύνολο μοναδιαίων modal τελεστών  $\{K_i\}$ , όπου  $i \in \{1, \dots, n\}$ . Επομένως, ο τύπος  $K_i\varphi$  διαβάζεται ως 'ο  $i$  γνωρίζει ότι  $\varphi$ '. Ο σημασιολογικός κανόνας του ' $\Box$ ' αντικαθίσταται από τον εξής:

$$\langle M, w \rangle \models K_i\varphi \text{ αν και μόνο αν } \forall w' \in W \cdot \text{αν } (w, w') \in R_i \text{ τότε } \langle M, w' \rangle \models \varphi.$$

Συνεπώς, κάθε τελεστής  $K_i$  έχει ακριβώς τις ίδιες ιδιότητες με τον ' $\Box$ '. Αντιστοιχίζοντας κάθε ένα από τα modal συστήματα  $\Sigma$ , ένα αντίστοιχο σύστημα  $\Sigma_n$  ορίζεται για την πολυπρακτορική λογική. Έτσι,  $K_n$  είναι η μικρότερη πολυπρακτορική επιστημική λογική και  $KT5_n$  η μεγαλύτερη.

Στην συνέχεια χρειάζεται να αξιολογηθεί το κατά πόσο αποτελεσματική αποδεικνύεται η normal modal logic ως λογική γνώσης πεποίθησης. Ο κανόνας αναγκαιότητας δηλώνει ότι ο πράκτορας γνωρίζει όλους του έγκυρους τύπους. Μεταξύ άλλων, αυτό σημαίνει ότι ο πράκτορας γνωρίζει όλες τις προτασιακές ταυτολογίες. Καθώς ο αριθμός τους είναι απεριόριστος, ο πράκτορας θα πρέπει να έχει έναν άπειρο αριθμό στοιχείων γνώσης.

Έστω το αξίωμα  $K$  που μας λέει ότι η γνώση ενός πράκτορα είναι κλειστή κάτω από επιπτώσεις. Υποθέτοντας ότι το  $\varphi$  είναι ένα λογικό επακόλουθο του συνόλου  $\Phi = \{\varphi_1, \dots, \varphi_n\}$ , τότε σε κάθε κόσμο όπου όλα τα  $\Phi$  είναι αληθή, το  $\varphi$  θα πρέπει να είναι επίσης αληθές, οπότε η παράσταση

$$\varphi_1 \wedge \dots \wedge \varphi_n \Rightarrow \varphi$$

πρέπει να είναι έγκυρη. Από αναγκαιότητα, ο τύπος αυτός πιστεύεται από τον πράκτορα καθώς οι γνώσεις του είναι κλειστές κάτω από επιπτώσεις, και επομένως αν πιστεύει ότι  $\Phi$ , τότε θα πρέπει επίσης να πιστεύει ότι  $\varphi$ . Συνεπώς, η γνώση ενός πράκτορα είναι κλειστή κάτω από λογικά επακόλουθα [5].

### 4.4.3 Αξιώματα Γνώσης και Πεποίθησης

Το αξίωμα  $D_n$  λέει ότι οι πεποιθήσεις ενός πράκτορα είναι μη αντιφατικές και μπορεί να γραφεί με τον ακόλουθο τύπο:

$$K_i \varphi \Rightarrow \neg K_i \neg \varphi,$$

που διαβάζεται ως 'αν ο  $i$  γνωρίζει ότι  $\varphi$ , τότε ο  $i$  δεν γνωρίζει ότι  $\neg \varphi$ '.

Το αξίωμα  $T_n$  συνήθως καλείται *αξίωμα γνώσης*, καθώς επιβεβαιώνει την εγκυρότητα της γνώσης. Είναι δηλαδή το αξίωμα που διαχωρίζει την γνώση από την πεποίθηση: κάποιος μπορεί να πιστεύει κάτι που να μην είναι αληθές, αλλά δίσταζε να πει ότι κάποιος γνωρίζει κάτι ψευδές. Η γνώση επομένως συνήθως ορίζεται ως η αληθινή πεποίθηση: ο  $i$  ξέρει ότι  $\varphi$  αν ο  $i$  πιστεύει ότι  $\varphi$  και το  $\varphi$  είναι αληθές.

Το αξίωμα  $4_n$  καλείται και *αξίωμα θετικής αυτοεξέτασης*, όπου αυτοεξέταση είναι η διαδικασία τις εξέτασης των προσωπικών πεποιθήσεων. Το αξίωμα θετικής αυτοεξέτασης λέει ότι ένα πράκτορας γνωρίζει αυτά που ξέρει. Με παρόμοιο τρόπο, το αξίωμα  $5_n$  είναι το *αξίωμα αρνητικής αυτοεξέτασης*, το οποίο λέει ότι ο πράκτορας έχει επίγνωση αυτών που δεν γνωρίζει. Αυτά τα δύο αξιώματα, της θετικής και αρνητικής αυτοεξέτασης υποδηλώνουν ότι ένας πράκτορας έχει πλήρη επίγνωση αυτών που γνωρίζει και αυτών που δεν γνωρίζει [5].

## 4.5 Επιστημική Γνώση για Πολυπρακτορικά Συστήματα

Ένα σύστημα περιλαμβάνει ένα περιβάλλον, το οποίο μπορεί να βρίσκεται σε οποιαδήποτε κατάσταση ενός συνόλου  $E$ , και ένα σύνολο  $n$  διαδικασιών  $\{1, \dots, n\}$ , κάθε μια από τις οποίες μπορεί να βρίσκεται σε οποιαδήποτε 'τοπική' κατάσταση από ένα σύνολο  $L$ . Επομένως, σε οποιαδήποτε στιγμή, ένα σύστημα μπορεί να είναι σε οποιαδήποτε καθολική κατάσταση ενός συνόλου  $G$ :

$$G = E \times L \times \dots \times L.$$

Μια *διαδρομή* ενός συστήματος είναι μια συνάρτηση που αναθέτει σε κάθε χρονική στιγμή μια καθολική κατάσταση: τα χρονικά σημεία είναι ισόμορφα με τους φυσικούς αριθμούς (συνεπώς ο χρόνος είναι διακριτός, δεσμευμένος στο παρελθόν και άπειρος στο μέλλον). Επομένως, μια διαδρομή  $r$  είναι μια συνάρτηση

$$r : \mathbb{N} \rightarrow G.$$

Ένα *σημείο* είναι μια διαδρομή μαζί με τον χρόνο:

$$\text{Σημείο} = \text{Διαδρομή} \times \mathbb{N}.$$

Ένα σημείο προσδιορίζει έμμεσα μια καθολική κατάσταση. Τα σημεία θα αποτελέσουν τους κόσμους στην λογική της γνώσης που θα αναπτυχθεί. Ένα σύστημα είναι ένα σύνολο διαδρομών.

Έστω ότι οι  $s$  και  $s'$  είναι δύο καθολικές καταστάσεις,

$$s = \langle e, l_1, \dots, l_n \rangle,$$

$$s' = \langle e', l_1', \dots, l_n' \rangle.$$

Ορίζεται μια σχέση  $\sim_i$  στις καταστάσεις, για κάθε διαδικασία  $i$ ,

$$s \sim_i s' \text{ αν και μόνο αν } (l_i = l_i').$$

Η σημασιολογία της γλώσσας προβάλλεται μέσω της σχέσης ικανοποίησης '⊨', η οποία ισχύει μεταξύ τριάδων της μορφής

$$\langle M, r, u \rangle$$

και τύπων της γλώσσας. Το  $\langle r, u \rangle$  είναι ένα σημείο και το  $M$  αντιπροσωπεύει μια δομή

$$\langle R, \pi \rangle,$$

όπου  $R$  είναι ένα σύστημα και

$$\pi: \text{Σημείο} \rightarrow \wp(\text{Prop})$$

επιστρέφει το σύνολο των ατομικών προτάσεων αληθές σε ένα σημείο. Η δομή  $\langle R, \pi \rangle$  καλείται *ερμηνευμένο σύστημα*. Οι μοναδικοί μη τυποποιημένοι σημασιολογικοί κανόνες είναι για προτάσεις και modal τύπους:

$$\langle M, r, u \rangle \models p \text{ όπου } p \in \text{Prop}, \text{ αν και μόνο αν } p \in \pi((r, u))$$

$$\langle M, r, u \rangle \models K_i \varphi \text{ αν και μόνο αν } \langle M, r', u' \rangle \models \varphi \text{ για κάθε } r' \in R$$

$$\text{και } u' \in N \text{ τέτοια ώστε } r(u) \sim_i r'(u') \text{ [5].}$$

## 4.6 Στόχοι και Επιθυμίες

Μια προσέγγιση για την ανάπτυξη λογικής στόχων ή επιθυμιών είναι η υιοθέτηση σημασιολογίας πιθανών κόσμων. Σύμφωνα με αυτό, κάθε στόχος/πιθανός κόσμος αντιπροσωπεύει μια πλευρά του κόσμου που θα προκύψει αν πραγματοποιηθούν οι στόχοι του πράκτορα. Όμως, αυτήν η προσέγγιση φέρει και τα προβλήματα των συνεπειών, με την έννοια ότι προβλέπει ότι οι πράκτορες έχουν ως στόχο τις λογικές συνέπειες των στόχων τους [5].

## 4.7 Κοινή και Κατανεμημένη Γνώση

*Κοινή γνώση* είναι η ‘πολιτισμική’ γνώση, δηλαδή τα πράγματα που ξέρουν όλοι και όλοι γνωρίζουν ότι όλοι ξέρουν. Αρχικά ορίζεται ένας τελεστής που αντιπροσωπεύει τα πράγματα που όλοι γνωρίζουν και ούτω καθεξής. Ένας μοναδιαίος modal τελεστής  $E$  προστίθεται στην modal γλώσσα και ο τύπος  $E\varphi$  σημαίνει ‘όλοι γνωρίζουν ότι  $\varphi$ ’. Μπορεί να οριστεί ως μια συντομογραφία:

$$E\varphi \triangleq K_1\varphi \wedge \dots \wedge K_n\varphi.$$

Ο τύπος  $E^k$  διαβάζεται ως ‘όλοι γνωρίζουν ότι  $\varphi$  στον βαθμό  $k$ ’, και ορίζεται ως εξής:

$$E^1\varphi \triangleq E\varphi$$

$$E^{k+1}\varphi \triangleq E(E^k\varphi).$$

Επειδή από μόνος του ο τελεστής  $E$  δεν αποτυπώνει επαρκώς την ιδέα της κοινής γνώσης, προστίθεται ένας συμπληρωματικός τελεστής  $C$  ο οποίος ορίζεται ως εξής:

$$C\varphi \triangleq E\varphi \wedge E^2\varphi \wedge \dots \wedge E^k\varphi \wedge \dots$$

Επομένως, κοινή γνώση είναι ο επ’ άπειρων συνδυασμός των ‘όλοι γνωρίζουν’, και ‘όλοι ξέρουν ότι όλοι ξέρουν’ και ούτω καθεξής.

Η *κατανεμημένη*, ή έμμεση, γνώση σχετίζεται άμεσα με την κοινή γνώση. Υποθέτοντας ότι υπάρχει ένας παντογνώστης παρατηρητής μιας ομάδας πρακτόρων, ο οποίος έχει την ικανότητα να ‘διαβάζει’ τις πεποιθήσεις/γνώσεις του κάθε πράκτορα. Επομένως, αυτός ο πράκτορας θα ήταν ικανός να συγκεντρώσει την συλλογική γνώση της ομάδας και θα ήταν γενικά σε θέση να συμπεράνει περισσότερα από οποιονδήποτε πράκτορα της ομάδας. Για παράδειγμα, υποθέτοντας ότι σε μια ομάδα δύο πρακτόρων ο πράκτορας 1 γνωρίζει μόνο ότι  $\varphi$  και ο πράκτορας 2 γνωρίζει μόνο ότι  $\varphi \Rightarrow \psi$ . Τότε, θα υπήρχε κατανεμημένη γνώση του  $\psi$ , ακόμη και αν κανένας πράκτορας δεν γνώριζε ρητά ότι  $\psi$ . Η κατανεμημένη γνώση δεν μπορεί να αποδοθεί με κανέναν από τους τελεστές που αναφέρθηκαν μέχρι στιγμής, γι’ αυτό ορίζεται ένας νέος τελεστής  $D$  που έχει τον ακόλουθο σημασιολογικό κανόνα:

$$\langle M, w \rangle \vdash D_\varphi \text{ αν και μόνο αν } \langle M, w' \rangle \models \varphi \text{ για όλα τα } w'$$

$$\text{έτσι ώστε } (w, w') \in (R_1 \cap \dots \cap R_n).$$

Οι διάφοροι τελεστές συλλογικής γνώσης σχηματίζουν μια ιεραρχία της μορφής:

$$C\varphi \Rightarrow E^k\varphi \Rightarrow \dots \Rightarrow E\varphi \Rightarrow K_i\varphi \Rightarrow D\varphi [5].$$

---

## Κεφάλαιο 5

---

### Συναίσθημα στα Πολυπρακτορικά Συστήματα

---

#### 5.1 Τυπικός Χαρακτηρισμός των Συναίσθημάτων

Οι Andrew Jones και Jeremy Pitt στο [48] λαμβάνοντας ως έναυσμα μια δημοσίευση του Ingmar Pörn [49], ο οποίος παρουσίασε έναν τυπολογικό χαρακτηρισμό ορισμένων συναίσθημάτων, προχωρούν σε μια τυπολογική περιγραφή πολύπλοκων συναίσθημάτων, όπως είναι η εμπιστοσύνη, η μετάνοια και η ντροπή. Η βασική παραδοχή της προσέγγισής του Pörn είναι το ότι κάθε συναίσθημα παρουσιάζει *διπλή σκοπιμότητα*. Για παράδειγμα, αν ο  $x$  ελπίζει ότι το τρένο έφτασε αργά, ο  $x$  '...εύχεται να είναι η περίπτωση όπου το τρένο έφτασε αργά και πιστεύει ότι είναι πιθανό να το έκανε'. Το συναίσθημα είναι η ελπίδα και το αντικείμενο του συναίσθηματος, στο συγκεκριμένο παράδειγμα, είναι η κατάσταση που το τρένο έφτασε αργά. Η διπλή σκοπιμότητα αναφέρεται στο τι εύχεται ο  $x$  και τι πιστεύει ο  $x$  - η *βουλητική* και *επιστημική* άποψη της σκοπιμότητας αντίστοιχα. Ο Pörn επεδίωξε να προσδιορίσει ένα σύνολο πρωτογενών ή ατομικών συναισθηματικών τύπων όσον αφορά τα αντίστοιχα επιστημικά στοιχεία και στοιχεία βούλησής τους [48].

##### 5.1.1 Επιστημικό Στοιχείο

Όσον αφορά το επιστημικό στοιχείο, ο Pörn έκανε την περεταίρω υπόθεση ότι η βασική ιδέα είναι αυτή της βεβαιότητας. Παρατήρησε ότι τέτοιου είδους συναίσθημα όπως η ελπίδα, ο φόβος και ανησυχία φαίνεται να μην συμβαδίζουν με την βεβαιότητα, ενώ κάποια άλλα συναίσθημα, όπως η λύπη, η απόγνωση, ο θυμός και η ντροπή, φαίνεται

να το απαιτούν. Υιοθετώντας την ιδέα ότι 'ο  $x$  είναι βέβαιος ότι  $p$ ' μπορεί να ερμηνευθεί ως 'ο  $x$  πιστεύει ότι ξέρει ότι  $p$ ', συμβόλισε την φράση 'ο  $x$  είναι βέβαιος ότι  $p$ ' με την modal logical έκφραση  $B_x K_x p$  και στην συνέχεια το απλοποίησε σε  $BKp$  με βάση την υπόθεση ότι θεωρούμε το ίδιο άτομο και στις δύο θέσεις δείκτη.

Θεωρώντας ότι οι όροι της πεποίθησης και της γνώσης είναι κανονικοί και ότι η λογική τους είναι τουλάχιστον αυτή των modal συστημάτων  $KD$  και  $KT$  αντίστοιχα, αρχικά παίρνουμε τις θετικές εκφράσεις  $BKp$ ,  $BK \neg p$ ,  $B \neg Kp$ , και  $B \neg K \neg p$  και στην συνέχεια τις αντίστοιχες αρνητικές εκφράσεις  $\neg BKp$ ,  $\neg BK \neg p$ ,  $\neg B \neg Kp$ , και  $\neg B \neg K \neg p$ . Αυτές οι οκτώ εκφράσεις μπορούν να οργανωθούν ως τέσσερις ταυτολογίες ως εξής:

$$T1. \quad BKp \vee \neg BKp$$

$$T2. \quad BK \neg p \vee \neg BK \neg p$$

$$T3. \quad B \neg Kp \vee \neg B \neg Kp$$

$$T4. \quad B \neg K \neg p \vee \neg B \neg K \neg p$$

Υπάρχουν 16 τρόποι για ακριβή επιλογή μιας διάζευξης εκ των  $T1. - T4.$  Για τον σχηματισμό ενός συνδέσμου τεσσάρων συζυγίων. Από αυτούς τους 16 συνδέσμους, μόνο 6 είναι λογικά συνεπείς, λαμβάνοντας υπόψη τις λογικές ιδιότητες που υιοθετήθηκαν για τους δύο modal τελεστές. Αυτές παρατίθενται από τον Pörn ως οι 'ατομικές επιστημικές θέσεις':

$$EP1. \quad BKp \wedge \neg B \neg Kp \wedge \neg BK \neg p \wedge B \neg K \neg p$$

$$EP2. \quad \neg BKp \wedge B \neg Kp \wedge BK \neg p \wedge \neg B \neg K \neg p$$

$$EP3. \quad \neg BKp \wedge B \neg Kp \wedge \neg BK \neg p \wedge B \neg K \neg p$$

$$EP4. \quad \neg BKp \wedge B \neg Kp \wedge \neg BK \neg p \wedge \neg B \neg K \neg p$$

$$EP5. \quad \neg BKp \wedge \neg B \neg Kp \wedge \neg BK \neg p \wedge B \neg K \neg p$$

$$EP6. \quad \neg BKp \wedge \neg B \neg Kp \wedge \neg BK \neg p \wedge \neg B \neg K \neg p$$

Αυτά τα έξι σημεία αλληλοαναιρούνται, και η διάζευξή τους είναι η λογική αλήθεια. Επομένως, ακριβώς μια εκ των  $EP1. - EP6.$  πρέπει να ισχύει για κάθε δεδομένη πρόταση  $p$ .

Ο Pörn περιγράφει την  $EP6.$  ως την επιστημική μηδενική θέση, και αμφιβάλλει για το αν είναι χαρακτηριστική επιστημική θέση κάποιου συναισθήματος καθώς σε αυτήν την θέση το υποκείμενο δεν έχει καμιά πεποίθηση όσον αφορά την  $p$ .

Οι εναπομείναντες πέντε σύνδεσμοι μπορούν να απλοποιηθούν με την αφαίρεση των συζεύξεων που λογικά υπονοούνται από μια ή περισσότερες άλλες συζεύξεις. Το αποτέλεσμα είναι η ακόλουθη λίστα των απλοποιημένων ατομικών επιστημικών θέσεων:

<i>SEP1.</i>	$BKp$
<i>SEP2.</i>	$BK \neg p$
<i>SEP3.</i>	$B \neg Kp \wedge B \neg K \neg p$
<i>SEP4.</i>	$B \neg Kp \wedge \neg BK \neg p \wedge \neg B \neg K \neg p$
<i>SEP5.</i>	$B \neg K \neg p \wedge \neg BKp \wedge \neg B \neg Kp. [48]$

### 5.1.2 Στοιχείο Βούλησης

Αν και ο Rörn δεν προσέφερε μια καθαρά modal logical αναπαράσταση του στοιχείου βούλησης, προσδιόρισε τρεις διακριτές θέσεις βούλησης που αντιστοιχούν ακριβώς στους τρεις στοιχειώδεις κανονιστικές θέσεις της Πρότυπης Δεοντικής Λογικής (ένα modal logical σύστημα του τύπου KD). Αυτές οι τρεις θέσεις είναι, αντίστοιχα: μια επιθυμία ότι  $p$ , μια επιθυμία ότι  $\neg p$  και η θέση αδιαφορίας στην οποία δεν υπάρχει επιθυμία ούτε για  $p$  ούτε για  $\neg p$ . Εισάγοντας τον τελεστή  $D$ , είναι:

<i>VOL1.</i>	$Dp$
<i>VOL2.</i>	$D \neg p$
<i>VOL3.</i>	$\neg Dp \wedge \neg D \neg p$

Όπως και για την επιστημική μηδενική θέση, έτσι και για την θέση βουλητικής αδιαφορίας, ο Rörn αμφιβάλει για τον ρόλο που παίζει στην ανάλυση των συναισθημάτων. Έτσι, ενώνοντας τα πέντε επιστημικά στοιχεία με τα δύο στοιχεία βούλησης παράγεται ένα σύνολο δέκα ατομικών συναισθηματικών τύπων. Έτσι προκύπτει η εξής λίστα:

<i>EM1.</i>	$BKp \wedge Dp$
<i>EM2.</i>	$BKp \wedge D \neg p$
<i>EM3.</i>	$B \neg Kp \wedge Dp$
<i>EM4.</i>	$B \neg Kp \wedge D \neg p$
<i>EM5.</i>	$B \neg Kp \wedge B \neg K \neg p \wedge Dp$
<i>EM6.</i>	$B \neg Kp \wedge B \neg K \neg p \wedge D \neg p$
<i>EM7.</i>	$B \neg Kp \wedge \neg BK \neg p \wedge \neg B \neg K \neg p \wedge Dp$

$$EM8. \quad B \neg Kp \wedge \neg BK \neg p \wedge \neg B \neg K \neg p \wedge D \neg p$$

$$EM9. \quad B \neg K \neg p \wedge \neg BKp \wedge \neg B \neg Kp \wedge Dp$$

$$EM10. \quad B \neg K \neg p \wedge \neg BKp \wedge \neg B \neg Kp \wedge D \neg p$$

Οι συναισθηματικοί τύποι που αντιπροσωπεύονται από τους *EM2.* και *EM3.* είναι και οι δύο δείγματα λύπης, καθώς, και στις δύο περιπτώσεις, ο *x* είναι σίγουρος ότι η κατάσταση είναι αντίθετη με την επιθυμία του. Αντιθέτως, οι *EM1.* και *EM4.* αντιπροσωπεύουν καταστάσεις στις οποίες η βεβαιότητά του *x* ταιριάζει με τις επιθυμίες του. Οι *EM5.* και *EM6.* αντιπροσωπεύουν ανησυχία καθώς ο *x* πιστεύει ότι, από όσο γνωρίζει, θα μπορούσε να ισχύει ότι *p* και, από όσο γνωρίζει, μπορεί να μην ισχύει *p*, και στην μια περίπτωση επιθυμεί *p* ενώ στην άλλη  $\neg p$ . Οι θέσεις *EM8.* και *EM9.* αντιπροσωπεύουν την ελπίδα: αν και ο *x* δεν είναι σίγουρος για το ποιές είναι οι επιθυμίες του, παρ' όλα αυτά πιστεύει ότι η πραγματοποίηση των επιθυμιών του συμβαδίζει με όσα ξέρει. Με παράλληλο συλλογισμό, θα μπορούσε κανείς να ερμηνεύσει τις θέσεις *EM7.* και *EM10.* ως ένα είδος φόβου [48].

### 5.1.3 Πολύπλοκα Συναισθήματα: Εμπιστοσύνη και Ντροπή

Ο Rörn πραγματεύεται αρκετούς τρόπους με τους οποίους, οι ατομικοί τύποι που αναφέρθηκαν πιο πάνω, μπορούν να παρέχουν την βάση για την διατύπωση και άλλων συναισθημάτων. Ένας από αυτούς του τρόπους περιλαμβάνει των περιορισμό του συναισθήματος του αντικειμένου, που αντιπροσωπεύεται από το *p* στους παραπάνω τύπους, σε συγκεκριμένους τύπους καταστάσεων των πραγμάτων.

Ο Jones [50] πρότεινε ότι το αντικείμενο ή το περιεχόμενο μιας *έμπιστης συμπεριφοράς* αποτελείται από δύο στοιχεία: πρώτον, ότι υπήρχε ένας κανόνας – για παράδειγμα, ένας κανόνας που υποβάλει μια υποχρέωση σε έναν πράκτορα – και δεύτερον, ότι θα τηρούνταν ο κανόνας – για την περίπτωση που αναφέρθηκε, η συμμόρφωση έγκειται στην εκπλήρωση των υποχρεώσεων του πράκτορα. Η συμπεριφορά του εμπιστευόμενου παριστάνεται απλά ως πεποίθηση, παρά ως βεβαιότητα του τύπου *BK*. Αντίστοιχα, ο εμπιστευόμενος χαρακτηριζόταν με βάση την πίστη του στους κανόνες και την πίστη του στην συμμόρφωση. Υποθέτοντας ότι στις προτάσεις *SEP1.* – *SEP5.*, που αναφέρθηκαν προηγουμένως, η πρόταση *p* περιγράφει την κατάσταση των πραγμάτων όπου υπάρχει ένας κανόνας και ότι ο κανόνας αυτός τηρείται. Τρεις εκ των προτάσεων *SEP1.* – *SEP5.* συμβιβάζονται με την αλήθεια *Bp*: η *SEP1.* λογικά συνεπάγεται *Bp*, καθώς ο τελεστής πεποίθησης είναι κλειστός κάτω από λογική συνέπεια και ο *Kp* λογικά συνεπάγεται *p*. Ο *Bp* μπορεί επίσης να προστεθεί σταθερά ως σύνδεσμος στην *SEP3.*, και ως αποτέλεσμα να απλοποιηθεί σε  $Bp \wedge B \neg Kp$ . Τέλος, ο *Bp* μπορεί επίσης να προστεθεί σταθερά ως σύνδεσμος στην *SEP5.*, και ως αποτέλεσμα να απλοποιηθεί αυτήν την φορά σε  $Bp \wedge \neg BKp \wedge \neg B \neg Kp$ . Επομένως, αυτοί οι ισχυρισμοί θα πρότειναν τρεις αμοιβαία αποκλειόμενους υποψήφιους για το επιστημικό/δοξαστικό στοιχείο της εμπιστοσύνης, όπου η πρόταση *p* φέρει την συγκεκριμένη, περιορισμένη ερμηνεία που δόθηκε πιο πάνω:



$$E/DTRUST1. \quad BKp$$

$$E/DTRUST2. \quad Bp \wedge B \neg Kp$$

$$E/DTRUST3. \quad Bp \wedge \neg BKp \wedge \neg B \neg Kp$$

Αν η εμπιστοσύνη είναι συναίσθημα, σύμφωνα με τον Pörn δεν μπορεί να χαρακτηριστεί μόνο με όρους επιστημικούς. Επομένως, η εμπιστοσύνη μπορεί να περιέχει και όρους βούλησης.

Ο Jones αναφέρει ότι ένας πράκτορας  $x$  είχε πεποιθήσεις σύμφωνα με τις οποίες τηρήθηκε ένας κανόνας, και θα συμμορφωνόταν με αυτόν, ακόμη και αν ο  $x$  δεν ενδιαφερόταν για το αν θα υπήρχε συμμόρφωση ως προς τον κανόνα, και πιθανώς να έμενε αδιάφορος στην ύπαρξη του ίδιου του κανόνα. Για παράδειγμα, ο  $x$  πιστεύει ότι οι Νορβηγοί γραφειοκράτες υποχρεούνται να συλλέγουν τα επιδόματα των αλιέων βακαλάου, και πιστεύει ότι θα πραγματοποιήσουν αυτήν τους την υποχρέωση, αλλά δεν έχει καμία επιθυμία για το αντικείμενο για το οποίο επιδεικνύει αυτήν την έμπιστη συμπεριφορά. Αυτό όμως αποτελεί ένα εκκεντρικό σενάριο όσον αφορά το κατά πόσο το πλαίσιο που προκαλεί το ενδιαφέρον σε θέματα εμπιστοσύνης είναι αυτό στο οποίο η συμμόρφωση ενδιαφέρει τον εμπιστευόμενο. Ίσως ο  $x$  να χάνει σε περίπτωση που η συμμόρφωση δεν είναι επικείμενη, και είναι εν μέρει επειδή συνήθως συσχετίζεται η εμπιστοσύνη με καταστάσεις στις οποίες ο εμπιστευόμενος δεν είναι αδιάφορος που τείνετε η σύνδεση της εμπιστοσύνης και του ρίσκου. Όμως, σύμφωνα με τον Pörn, όταν οι πεποιθήσεις εμπιστοσύνης συνδυάζονται με την αδιαφορία, τότε ο εμπιστευόμενος στερείται συμμετοχής, στερείται δέσμευσης, δεν υπάρχει συναισθηματική χροιά στην εμπιστοσύνη. Επομένως, ενώνοντας κάθε πρόταση  $E/DTRUST1. - E/DTRUST3.$  με το  $Dp$ , προκύπτουν οι τρεις υποψήφιοι τύποι για τον συμβολισμό της εμπιστοσύνης ως συναίσθημα:

$$EMTRUST1. \quad BKp \wedge Dp$$

$$EMTRUST2. \quad Bp \wedge B \neg Kp \wedge Dp$$

$$EMTRUST3. \quad Bp \wedge \neg BKp \wedge \neg B \neg Kp \wedge Dp$$

Με αυτούς τους τύπους είναι εύκολο να γίνει η συσχέτιση μεταξύ εμπιστοσύνης και ελπίδας. Ενώ μπορεί να είναι προφανές ότι ο  $EMTRUST1.$  ταιριάζει διαισθητικά με την έννοια της εμπιστοσύνης, ο  $EMTRUST2.$ , λόγω της αβεβαιότητας που εκφράζει ο δεύτερος όρος του, συγγενεύει περισσότερο με την ελπίδα, και ο  $EMTRUST3.$  ίσως παρουσιάζει μια 'δύναμη' που υπάγεται μεταξύ της εμπιστοσύνης και της ελπίδας.

Ο Pörn έχει υποδείξει ένα δεύτερο τρόπο με τον οποίο άλλοι συναισθηματικοί τύποι θα μπορούσαν να παραχθούν από τους ατομικούς τύπους, παίρνοντας τα συναισθήματα ως τα ίδια να είναι τα αντικείμενα των συναισθημάτων. Αυτός ο τρόπος χρησιμοποιείται για την ανάλυση της ερμηνείας της ντροπής.

Η βασική ιδέα είναι η κατανόηση της ντροπής σε σχέση με τις ενέργειες ενός αντικειμένου, ή την αποτυχία να ενεργήσει, ως την λύπη ότι κάποιο άλλο πρόσωπο μετανιώνει γι' αυτήν την ενέργεια/αποτυχία ενέργειας. Υποθέτοντας, αρχικά, ότι ο  $y$

λυπάται που ο  $x$  δεν έκανε το  $q$ . Το αντικείμενο του συναισθήματος μπορεί να συμβολιστεί ως  $\neg E_x q$ , όπου  $E_x$  είναι ένας σχετικοποιημένος modal τελεστής ενέργειας. Η μετάνοια του  $y$  που ο  $x$  δεν έχει κάνει το  $q$  μπορεί να ερμηνευθεί ως την βεβαιότητα του  $y$  ότι ο  $x$  δεν έχει κάνει το  $q$  σε συνδυασμό με την επιθυμία του  $y$  να κάνει ο  $x$  το  $q$ :

$$(REG1) \quad B_y K_y \neg E_x q \wedge D_y E_x q.$$

Έστω ότι ο  $x$  νιώθει ντροπή έναντι του  $y$ , καθώς (ο  $x$ ) δεν έχει κάνει το  $q$ , αν ο τύπος (REG1) είναι από μόνος του το αντικείμενο μιας συναισθηματικής κατάστασης του τύπου λύπης από την πλευρά του  $x$ :

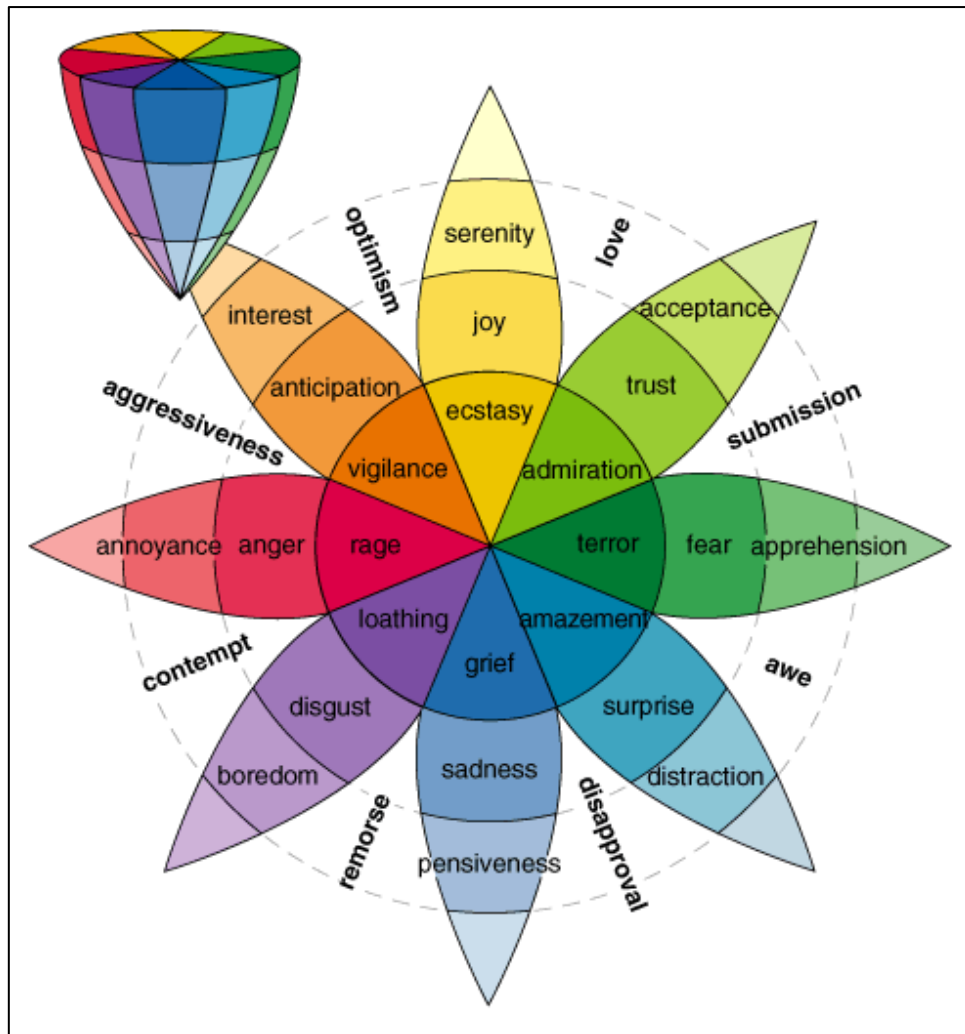
$$(ASH1) \quad B_x K_x (B_y K_y \neg E_x q \wedge D_y E_x q) \wedge D_x \neg (B_y K_y \neg E_x q \wedge D_y E_x q).$$

Ένα επιπλέον παράδειγμα αυτού του τύπου ντροπής θα ήταν η περίπτωση στην οποία ο  $x$  όχι μόνο αποτυγχάνει να κάνει το  $q$ , αλλά επιπλέον ο  $x$  είχε την υποχρέωση να κάνει το  $q$ , δηλαδή η ντροπή του  $x$  έναντι του  $y$  που (ο  $x$ ) παραβίασε την υποχρέωσή του. Θα ήταν δυνατό να διατυπωθεί επίσημα αυτού του είδους η περίπτωση αν η modal logical γλώσσα εμπλουτιζόταν εισάγοντας μια *διευθυντική normative modality* οδηγία – αντίθετα με μια αξιολογική, της οποίας η modality επιθυμίας είναι παράδειγμα [48].

## 5.2 Ψυχοφυσιολογική Ταξινόμηση του Συναισθήματος

Είναι πλέον αναγνωρισμένο το γεγονός ότι τα συναισθήματα παίζουν σημαντικό λόγο στις γνωστικές διεργασίες των ανθρώπων, επομένως είναι ενδιαφέρον να είναι κάποιος σε θέση να αναγνωρίσει και να ταξινομήσει αυτές τις συναισθηματικές καταστάσεις. Μια από τις πιο σημαντικές θεωρίες συναισθημάτων αποτελεί το ψυχοεξελικτικό μοντέλο του Plutchik [51]. Ο Plutchik υποστήριξε ότι τα συναισθήματα προσφέρουν ένα εξελικτικό πλεονέκτημα είτε στην αναπαραγωγή είτε στην επιβίωση, και διαπίστωσε ότι υπάρχουν οκτώ από αυτά σε τέσσερα διπολικά ζεύγη: η χαρά έναντι της λύπης, ο θυμός έναντι του φόβου, η εμπιστοσύνη έναντι της αηδίας και η έκπληξη έναντι της αναμονής. Η Εικόνα 5.1 αντιπροσωπεύει το κωνικό μοντέλο του Plutchik, προσδιορίζοντας τα ζεύγη, τις διαφορετικές ένταση και τα μίγματα που παράγουν τα διαφορετικά συναισθήματα.

Μια από τις ιδέες της συναισθηματικής πληροφορικής είναι ότι τα φυσιολογικά σήματα είναι μια ένδειξη της ψυχολογικής (συναισθηματικής) κατάστασης, και ότι μπορούν να χρησιμοποιηθούν τεχνικές επεξεργασίας σημάτων σε συνδυασμό με την θεωρία των συναισθημάτων για την συναγωγή της συναισθηματικής κατάστασης από καταγεγραμμένα σήματα. Σύμφωνα με πειράματα, είναι δυνατόν να προσδιοριστεί μια αλλαγή ενός φυσιολογικού μεγέθους (γαλβανική αντίδραση του δέρματος) που να συνδέεται με μια αλλαγή στην συναισθηματική κατάσταση. Επομένως, είναι δυνατόν να συναχθεί ένα σήμα φυσικής γλώσσας και με μια απεικόνιση του προσώπου και με ένα φυσικό σήμα. Αυτό σημαίνει ότι τα συναισθήματα έχουν μια σαφή οντολογική κατάσταση και επομένως μπορούν να κατηγοριοποιηθούν με μια τυπική αναπαράσταση.



Εικόνα 5.1 Το κωνικό μοντέλο συναισθημάτων του Plutchik (3D και 2D) [52]

Πρώτον, είναι χρήσιμο να παρατηρηθεί το που ευθυγραμμίζεται η ψυχοφυσιολογική ταξινόμηση του Plutchik με την modal logical ταξινόμηση του Rörn και που όχι. Για παράδειγμα, οι ατομικοί συναισθηματικοί τύποι του Rörn ήταν η ασφάλεια, η απόγνωση (λύπη), η ανησυχία, η ελπίδα και ο φόβος. Με βάση το ότι 'ο κόσμος' είναι, και δεν είναι, όπως τον θέλει ο πράκτορας, η ασφάλεια και η απόγνωση φαίνεται να αντιστοιχούν στην χαρά και στην θλίψη, η ελπίδα στην αναμονή και ο φόβος στον φόβο. Οι δύο πτυχές της ανησυχίας φαίνεται να αντιστοιχούν στον ενδιαφέρον και στην αντίληψη, δηλαδή στις αδύναμες (σε ένταση) εκδόσεις της αναμονής και του φόβου του Plutchik. Τέλος, τα ατομικά συναισθήματα του Plutchik, η εμπιστοσύνη και η αηδία, φαίνεται να αντιστοιχούν σε πολύπλοκους τύπους συναισθημάτων του Rörn, όπως τα *EMTRUST1*, *EMTRUST3*, και *ASH - REG1*, αντίστοιχα.

Δεύτερον, η βασική διαίσθηση του Rörn ότι υπάρχουν κάποια συναισθήματα που συμβαδίζουν με την βεβαιότητα, και κάποια όχι, είναι σύμφωνη με τον διπολικό διαχωρισμό του Plutchik, στο ότι η έκπληξη και ο θυμός απαιτούν βεβαιότητα, ενώ ο φόβος και η αναμονή όχι. Ωστόσο, τόσο η χαρά όσο και η θλίψη φαίνεται να απαιτούν βεβαιότητα, ενώ το άλλο διπολικό ζεύγος, η εμπιστοσύνη και η αηδία, παρουσιάζει μια σειρά θέσεων συμβατών και ασύμβατων με την βεβαιότητα [48].

### 5.3 Συγχώρεση

Για να χαρακτηριστεί μια συμπεριφορά εμπιστοσύνης, όπως αναφέρθηκε παραπάνω, ο εμπιστευόμενος θα πρέπει να κατέχει δύο πεποιθήσεις: ότι υπάρχει ένας κανόνας, και ότι η συμπεριφορά κάποιου άλλου θα συμμορφωθεί στον κανόνα αυτό. Για το χαρακτηρισμό μιας απόφασης εμπιστοσύνης, ο εμπιστευόμενος θα πρέπει να πραγματοποιήσει έναν υπολογισμό: ποια είναι η πιθανότητα συμμόρφωσης της συμπεριφοράς κάποιου στον κανόνα αυτό, και ποιο είναι το όφελος/κόστος αν κάποιος συμμορφωθεί/δεν συμμορφωθεί στον κανόνα αυτό. Αυτό ποικίλει ανάλογα με το πλαίσιο και, επομένως, η απόφαση εμπιστοσύνης κυμαίνεται από *δυνατή* εμπιστοσύνη, όπως χαρακτηρίζεται από τον *EMTRUST1.*, που είναι ένα είδος συντομευμένου τρόπου για την αντιστάθμιση του πιο ‘ακριβού’ υπολογισμού που απαιτείται για την εκτίμηση της *αδύναμης* εμπιστοσύνης, όπως χαρακτηρίζεται από τα *EMTRUST2.* – *EMTRUST3.*

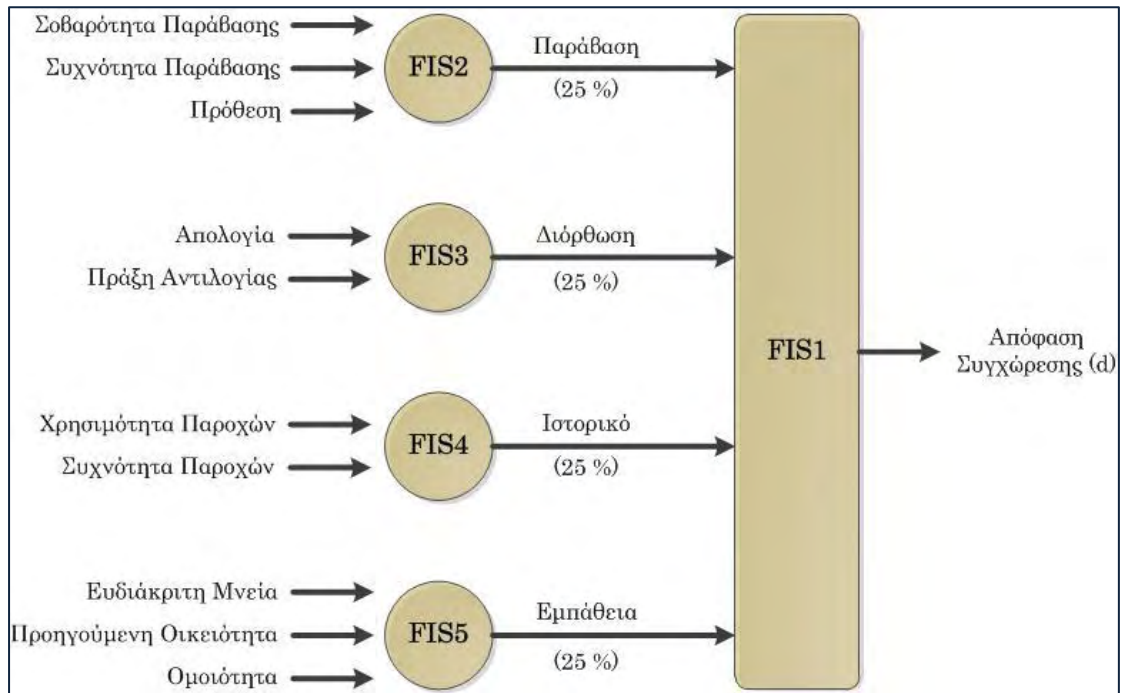
Για να υπάρξει οποιαδήποτε απόφαση εμπιστοσύνης, θα πρέπει (αντικειμενικά) να υπάρχει μια πεπερασμένη πιθανότητα λάθους. Ως αποτέλεσμα, είτε στην *δυνατή* είτε στην *αδύναμη* εμπιστοσύνη, υπάρχει πάντα η πιθανότητα η απόφαση να είναι λανθασμένη, και ένα σημαντικό στοιχείο της εμπιστοσύνης, που συχνά παραβλέπεται, είναι ποια θα είναι η ενέργεια όταν η απόφαση εμπιστοσύνης αποδειχθεί λανθασμένη.

Οι Vasalou et al. [53, 54] πρότειναν έναν μηχανισμό συγχώρεσης για την λήψη αποφάσεων σχετικά με την παραβίαση των κανόνων. Δεν βασίζεται στην φήμη, που είναι ένα ποσοτικός μηχανισμός τιμωρίας, αλλά αντιθέτως βασίζεται στην συγχώρεση. Αποτελεί έναν ποιοτικό μηχανισμό επιδιόρθωσης, που είναι γνωστό από ψυχολογικές μελέτες ότι τονώνει τις εθελοντικές πράξεις αποζημίωσης, μειώνει την αρνητική προδιάθεση προς έναν παραβάτη και τονίζει ένα θετικό κίνητρο για αυτοεπιδιόρθωση.

Ο συγκεκριμένος μηχανισμός συγχώρεσης αποτελείται από έντεκα συστατικά στοιχεία (την σοβαρότητα, την συχνότητα και την πρόθεση του παραβάτη, την απολογία και την αποκατάσταση, την χρησιμότητα και την συχνότητα της ωφέλιμης σχέσης, και την οικειότητα, την ομοιότητα και την ντροπή ή την αμηχανία) που διέπουν τα τέσσερα θετικά κίνητρα που σχετίζονται με την φύση του παραβάτη, μέτρα επιδιόρθωσης, τα ιστορικά στοιχεία και την ενσυναισθητική σχέση. Αυτό αναπτύχθηκε χρησιμοποιώντας ένα ασαφές σύστημα συμπερασμού (fuzzy inference system, FIS) το οποίο χρησιμοποιεί ασαφείς κανόνες για τον υπολογισμό μιας ασαφούς τιμής για κάθε ένα από τα τέσσερα θετικά κίνητρα των αντίστοιχων στοιχείων, που έχουν τα ίδια συνδυαστεί με ένα πέμπτο FIS για την εξαγωγή μιας απόφασης συγχώρεσης (εικόνα 5.2).

Είναι φανερό το πώς η διαφοροποιημένη ανάλυση του ‘φάσματος’ των συμπεριφορών εμπιστοσύνης *EMTRUST1.* – *EMTRUST3.* θα μπορούσαν να είναι είσοδοι για την απόφαση συγχώρεσης. Για παράδειγμα, για τον παραβάτη, η ‘δύναμη’ της εμπιστοσύνης θα συσχετίζεται με την αντιληπτή σοβαρότητα της αθέτησης της εμπιστοσύνης. Αυτό, σε συνδυασμό με τα αισθήματα μετάνοιας (REG1), σύμφωνα με την αυτό-αξιολόγηση του καθενός, και ένα αίσθημα ντροπής (ASH), σύμφωνα με την αυτό-αξιολόγηση του καθενός όπως φαίνεται στους άλλους, μπορεί να είναι αυτό που προκαλεί την απολογία ή άλλες πράξεις διόρθωσης, καθώς και ακούσιες αντανάκλαστικές ενέργειες όπως είναι το κοκκίνισμα. Αυτές είναι είσοδοι του μοντέλου συγχώρεσης του παραβάτη (απολογία ή πράξεις αντιλογίας στο FIS3, και ευδιάκριτη

μνεία στο FIS5), ενώ το στοιχείο ομοιότητας θα μπορούσε να χρησιμοποιεί την ίδια modal logical αιτιολογία ('αν ήμουν εγώ...'). Αυτός ο συσχετισμός παρουσιάζει ένα σημείο που τονίζεται από τους Castelfranchi και Lorini [55], οι οποίοι παρατήρησαν ότι τα συναισθήματα περιλαμβάνουν και ένα στοιχείο γνωστικής αξιολόγησης το οποίο οδηγεί το σύστημα σε μια ομοιοστατική ισορροπία (όπου το 'σύστημα' σε αυτήν την περίπτωση, αποτελεί δύο ανεξάρτητα αλλά αλληλοσυνδεόμενα στοιχεία, τον παραβάτη και τον προσβεβλημένο, και η ισορροπία προσμετρείται με την 'ένταση' των συναισθημάτων τους [48]).



Εικόνα 5.2 Μηχανισμός συγχώρεσης

---

## Κεφάλαιο 6

# Μοντελοποίηση της Εμπιστοσύνης ενός Πολυπρακτορικού Συστήματος μέσω του MASON

---

### 6.1 Εισαγωγή

Στο παρόν κεφάλαιο καταγράφεται και σχολιάζεται η συμπεριφορά ενός συστήματος πρακτόρων σε ένα peer-to-peer μοντέλο για την εύρεση και σύνθεση υπηρεσιών διαδικτύου, όπως περιγράφεται από τους Gharzouli και Boufaïda [56], το οποίο αποτελεί ένα από τα πιο σημαντικά ζητήματα των συστημάτων που βασίζονται σε υπηρεσίες. Το μοντέλο αναπτύχθηκε στην πλατφόρμα MASON [57], η οποία αποτελεί ένα εργαλείο προσομοίωσης πολυπρακτορικών συστημάτων σε Java και το οποίο έχει σχεδιαστεί με σκοπό την αποτελεσματική υποστήριξη συστημάτων με μεγάλο αριθμό πρακτόρων.

### 6.2 Εύρεση και Σύνθεση Υπηρεσιών Διαδικτύου

Οι υπηρεσίες διαδικτύου αποτελούν ένα σύστημα λογισμικού που σχεδιάστηκαν με σκοπό την υποστήριξη της διαλειτουργικής αλληλεπίδρασης μεταξύ των μηχανημάτων (machine-to-machine) μέσω του διαδικτύου. Διαθέτουν μια διασύνδεση που περιγράφεται σε μορφή επεξεργάσιμη από τα μηχανήματα, όπως είναι η WSDL (Web Service Description Language) [58]. Τα άλλα συστήματα αλληλεπιδρούν με τις υπηρεσίες διαδικτύου με τον τρόπο που προσδιορίζεται από την περιγραφή τους χρησιμοποιώντας μηνύματα SOAP (Simple Object Access Protocol) [59], τα οποία

συνήθως μεταφέρονται μέσω HTTP με XML σειριοποίηση σε συνδυασμό με άλλα πρότυπα του παγκόσμιου ιστού [60].

Ένα από τα πιο σημαντικά γνωρίσματα των υπηρεσιών διαδικτύου είναι η ικανότητά τους να δημιουργούν νέες (σύνθετες) υπηρεσίες συνδυάζοντας τις δυνατότητες των ήδη υπάρχοντων ατομικών υπηρεσιών [61]. Το πρόβλημα της σύνθεσης υπηρεσιών μπορεί να διαχωριστεί σε τρία βασικά υπο-προβλήματα. Το πρώτο είναι η δημιουργία ενός σχεδίου που περιγράφει πως αλληλεπιδρούν οι υπηρεσίες και πως η λειτουργικότητα που παρέχουν μπορεί ενοποιηθεί ώστε να δώσουν τη λύση ενός προβλήματος. Το δεύτερο αφορά την εύρεση των κατάλληλων υπηρεσιών που πραγματοποιούν τις εργασίες που ορίζονται από το σχέδιο. Τέλος, το τρίτο είναι η διαχείριση της αλληλεπίδρασης των εν λόγω υπηρεσιών [62].

### 6.2.1 Προσεγγίσεις

Η πλειοψηφία των προσεγγίσεων που απευθύνονται στην αρχιτεκτονική των υπηρεσιών διαδικτύου βασίζονται σε κεντρικοποιημένες μεθόδους εύρεσης, όπως είναι η UDDI όπου οι υπηρεσίες περιγράφονται από λειτουργίες διεπαφής υπηρεσιών και δημοσιεύουν τις ικανότητες και τις λειτουργίες τους μέσω ενός μητρώου.

Παρ' όλα αυτά, η κεντρικοποιημένη προσέγγιση στην εύρεση υπηρεσιών παρουσιάζει δύο μεγάλα προβλήματα. Το πρώτο είναι η ύπαρξη ενός κεντρικού οργάνου για τη δημοσίευση και εύρεση υπηρεσιών. Αυτές οι μέθοδοι δεν δίνουν τη δυνατότητα προσαρμογής σε δυναμικές αλληλεπιδράσεις, περιορίζουν την επεκτασιμότητα ενός ευέλικτου και δυναμικού περιβάλλοντος, επιφέρουν συμφόρηση των επιδόσεων και μπορεί να αποτελέσουν μοναδικά σημεία σφαλμάτων. Επιπλέον, ο κεντρικοποιημένος έλεγχος των δημοσιευμένων υπηρεσιών υποφέρει από προβλήματα όπως το υψηλό κόστος λειτουργίας και συντήρησης. Τέλος, ο καθολικός χώρος αποθήκευσης UDDI υποφέρει από την έλλειψη μέτρου, όπου πολλές από τις δημοσιευμένες υπηρεσίες διαδικτύου δεν είναι διαθέσιμες ή δεν υλοποιούνται από τους παρόχους. Το δεύτερο πρόβλημα είναι ότι η περιγραφή γίνεται συνήθως σε WSDL, ή οποία παρέχει μόνο μια τεχνική περιγραφή. Όμως, η αυτοματοποιημένη εύρεση υπηρεσιών απαιτεί μια πιο κατανοητή περιγραφή που θα είναι πιο ξεκάθαρη στα μηχανήματα [56].

### 6.2.2 Peer-to-peer Μοντέλο Εύρεσης και Σύνθεσης Υπηρεσιών

Το μοντέλο εύρεσης και σύνθεσης υπηρεσιών Pm4sws [56] υποστηρίζει την αυτοματοποιημένη εύρεση και σύνθεση σημασιολογικών υπηρεσιών διαδικτύου που είναι καταναμημένες σε ένα μη δομημένο δίκτυο Peer-to-peer. Ο κύριος στόχος του συγκεκριμένου μοντέλου είναι η σύνθεση υπηρεσιών διαδικτύου για τη δημιουργία μιας σύνθετης υπηρεσίας που ανταποκρίνεται σε ένα συγκεκριμένο αίτημα. Η υπηρεσία μπορεί να συντεθεί από ένα σύνολο υπηρεσιών διαδικτύου που είναι καταναμημένες σε διάφορους κόμβους του δικτύου.

Για την εύρεση των κατάλληλων υπηρεσιών διαδικτύου αναπτύχθηκε ένας επιδημικός αλγόριθμος που βασίζεται στην αντιστοίχιση της εισόδου/εξόδου και οδηγείται από τον στόχο που πρόκειται να επιτευχθεί. Επιπλέον, όλοι οι συμμετέχοντες κόμβοι αποθηκεύουν σε έναν κατανεμημένο πίνακα όλων των προηγούμενων σύνθετων υπηρεσιών στις οποίες πήραν μέρος. Αυτό το ιστορικό μπορεί να χρησιμοποιηθεί ώστε να δίδονται άμεσες απαντήσεις σε μελλοντικά παρόμοια αιτήματα.

### 6.2.2.1 Δομή Πίνακα Σύνθετων Υπηρεσιών

Κάθε κόμβος δημιουργεί έναν πίνακα για την αποθήκευση όλων των σύνθετων υπηρεσιών στις οποίες έχει συμμετάσχει. Τα διάφορα χαρακτηριστικά του πίνακα σύνθετων υπηρεσιών ορίζονται ως εξής:

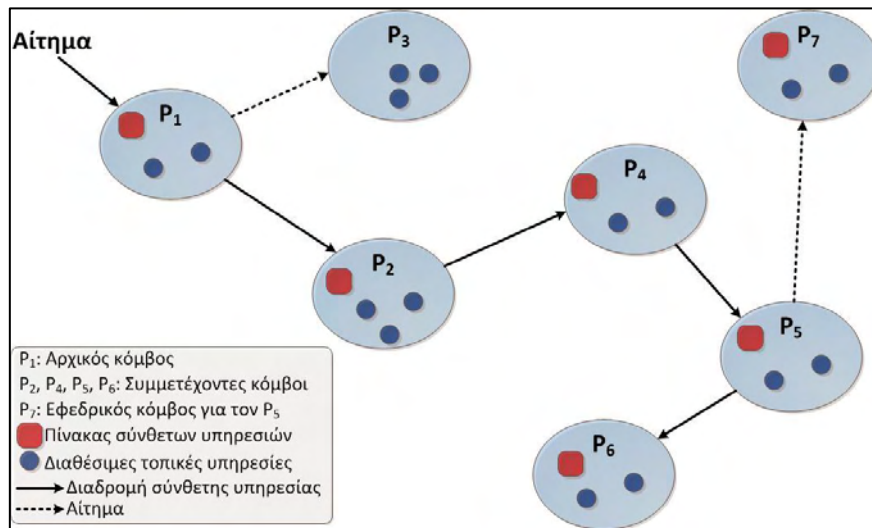
- **Αρχικός κόμβος (InitiatorPeer):** ο κόμβος που ξεκινά τη διαδικασία εύρεσης και σύνθεσης υπηρεσιών διαδικτύου
- **Κωδικός σύνθετης υπηρεσίας (CompID):** ο αναγνωριστικός κωδικός κάθε σύνθετης υπηρεσίας
- **Αρχική είσοδος (Init-Input):** η αρχική είσοδος της σύνθετης υπηρεσίας
- **Αρχική έξοδος (Init-Output):** η αρχική έξοδος της σύνθετης υπηρεσίας
- **Στόχος (Goal):** ο στόχος της σύνθετης υπηρεσίας
- **Εκτελέσιμες υπηρεσίες (Executed services):** οι υπηρεσίες που εκτελούνται τοπικά για την σύνθεση της επιθυμητής υπηρεσίας
- **Εφεδρικές υπηρεσίες (Reserve services):** τοπικές υπηρεσίες που μπορούν να χρησιμοποιηθούν στην ίδια σύνθεση
- **Προηγούμενοι κόμβοι (Precedent Peers):** κόμβοι οι οποίοι εκτελούν κάποιες από τις προηγούμενες υπηρεσίες της σύνθετης υπηρεσίας
- **Επόμενοι κόμβοι (Next Peers):** κόμβοι που εκτελούν τις επόμενες υπηρεσίες της σύνθετης υπηρεσίας
- **Εφεδρικοί κόμβοι (Reserve Peers):** κόμβοι που μπορούν να αντικαταστήσουν τους κόμβους που αποτυγχάνουν να καλύψουν τις ανάγκες της σύνθετης υπηρεσίας
- **Κατάσταση της σύνθετης υπηρεσίας (State of the composite service):** η σύνθετη υπηρεσία είναι ενεργή αν όλοι οι συμμετέχοντες κόμβοι που τη συνθέτουν είναι συνδεδεμένοι στο δίκτυο

Στο παράδειγμα της εικόνας 6.1 αναπαρίστανται οι παραπάνω έννοιες. Στο συγκεκριμένο παράδειγμα παρουσιάζεται η σύνθεση μιας υπηρεσία σε ένα μη δομημένο peer-to-peer δίκτυο, όπου οι ανεξάρτητες υπηρεσίες είναι κατανεμημένες μεταξύ πέντε διαφορετικών κόμβων, τους P<sub>2</sub>, P<sub>4</sub>, P<sub>5</sub> και P<sub>6</sub>. Στη συγκεκριμένη σύνθετη υπηρεσία ο κόμβος P<sub>7</sub> είναι εφεδρικός του P<sub>5</sub> και μπορεί να τον αντικαταστήσει σε περίπτωση που αποτύχει. Για τον P<sub>5</sub>, οι κόμβοι P<sub>2</sub> και P<sub>4</sub> είναι προηγούμενοι κόμβοι ενώ ο P<sub>6</sub> είναι ο επόμενος κόμβος.

### 6.2.2.2 Επιδημικός Αλγόριθμος Εύρεσης Υπηρεσιών που Βασίζεται στην Αντιστοίχιση Εισόδου/Εξόδου

**Ορισμός 1:** Μια υπηρεσία είναι μια τριάδα που ορίζεται ως εξής  $WS=(WS\text{-input}, WS\text{-output}, WS\text{-goal})$





Εικόνα 6.1 Παράδειγμα μιας μη δομημένης peer-to-peer σύνθετης υπηρεσίας

**Ορισμός 2:** Ο στόχος είναι ο εννοιολογικός προσδιορισμός ενός πεδίου υπηρεσιών των οποίων ο απώτερος σκοπός είναι παρόμοιος ή ίδιος. Ο στόχος προσδιορίζεται στη σημασιολογική περιγραφή μιας υπηρεσίας.

**Ορισμός 3:** Μια σύνθεση υπηρεσιών είναι μια διαδικασία που αποτελείται από ένα σύνολο βασικών υπηρεσιών. Στο πλαίσιο του συγκεκριμένου αλγόριθμου διακρίνονται δύο είδη συνθέσεων: τοπική και peer-to-peer σύνθεση υπηρεσιών.

**Ορισμός 4:** Μια τοπική σύνθετη υπηρεσία (LocCWS) είναι κατασκευασμένη από ένα σύνολο υπηρεσιών που ανήκουν στον ίδιο κόμβο.

**Ορισμός 5:** Μια peer-to-peer σύνθετη υπηρεσία (P2PCWS) είναι μια σύνθεση υπηρεσιών όπου οι βασικές υπηρεσίες είναι διανεμημένες σε διαφορετικούς κόμβους του δικτύου. Μια peer-to-peer σύνθετη υπηρεσία είναι μια τετράδα που ορίζεται ως  $P2PCWS=(InitiatorPeer, CompID, Init-Input, Init-Output, Goal)$ .

**Ορισμός 6:** Συμμετέχοντες κόμβοι είναι όλοι οι κόμβοι που συμμετείχαν (συνεργάστηκαν) στη δημιουργία μιας peer-to-peer σύνθετης υπηρεσίας, συμπεριλαμβανομένου και του αρχικού κόμβου.

#### 6.2.2.2.1 Βασικός αλγόριθμος

Κάθε κόμβος του δικτύου εκτελεί τον παρακάτω αλγόριθμο.

##### Βασικός Αλγόριθμος

1. Έναρξη
2. Λήψη του αιτήματος (WS-input, WS-output, WS-goal)
3. Αναζήτηση στις τοπικές υπηρεσίες όπου [(WS-input=Init-Input) και (WS-output=Init-Output) και (WS-goal=Goal)]
4. **Αν** (υπάρχει τοπική υπηρεσία) **τότε** στείλε θετική απάντηση και πήγαινε στο τέλος

5. **Av** (δεν υπάρχει τοπική υπηρεσία) **τότε** συνέθεσε τοπική υπηρεσία όπου [(LocCWS-input=Init-Input) και (LocCWS-output=Init-Output) και (LocCWS-goal=Goal)]
  6. **Av** (υπάρχει LocCWS) **τότε** στείλε θετική απάντηση και πήγαινε στο τέλος
  7. **Av** (δεν υπάρχει LocCWS) **τότε** αναζήτησε στον πίνακα σύνθετων υπηρεσιών (Init-Input, Init-Output, Goal)
  8. **Av** (δεν υπάρχει P2PCWS στον πίνακα σύνθετων υπηρεσιών) **τότε** ξεκινά μια νέα peer-to-peer αναζήτηση
  9. **Τέλος**
- 

Αρχικά, κάθε κόμβος εκτελεί τα βήματα 1 έως 6 του βασικό αλγόριθμου όπου ο κόμβος προσπαθεί να απαντήσει τοπικά στο αίτημα (τοπική βασική υπηρεσία ή τοπική σύνθεση υπηρεσιών). Στο βήμα 7 ο κόμβος αναζητά στον πίνακα σύνθετων υπηρεσιών για κάποια προηγούμενη σύνθετη υπηρεσία που δημιουργήθηκε μεταξύ των κόμβων του δικτύου και θα μπορούσε να καλύψει το συγκεκριμένο αίτημα.

#### 6.2.2.2.2 Αναζήτηση στον πίνακα συνθέσεων

Η διαδικασία αναζήτησης στον πίνακα σύνθετων υπηρεσιών είναι η εξής:

---

#### Αναζήτηση στον πίνακα σύνθετων υπηρεσιών

---

1. Έναρξη
  2. **Επιλογή** (InitiatorPeer, CompID) **από** τον πίνακα σύνθετων υπηρεσιών **όπου** [(P2PCWS-Input=Init-Input) και (P2PCWS-Output=Init-Output) και (P2PCWS-goal=Goal) και (state of the composite service=active)]/\* Δημιουργία λίστας επιλογής \*/
  3. **Av** (λίστα επιλογής δεν είναι άδεια) **τότε όσο** [(λίστα επιλογής δεν είναι άδεια) και (P2P σύνθεση υπηρεσιών δεν είναι εφικτή)] **κάνε**
    - a) Επιλογή InitiatorPeer από τη λίστα επιλογής
    - b) Αποστολή μηνύματος: **αναζήτηση** (CompID, Init-Input, Init-Output) από τον αρχικό κόμβο
    - c) **Av** (επιτυχής P2P εύρεση) **τότε** αποστολή απάντησης και πήγαινε στο τέλος /\* τέλος του βασικού προγράμματος \*/
  4. **Τέλος όσο**
  5. **Τέλος**
- 

Αν δεν υπάρχει πιθανότητα να απαντηθεί το αίτημα τοπικά η χρησιμοποιώντας τον πίνακα σύνθετων υπηρεσιών, ο κόμβος ξεκινά μια νέα P2P αναζήτηση (βήμα 7 του βασικού αλγόριθμου). Γι' αυτό το λόγο χρησιμοποιούνται δύο επιδημικούς αλγορίθμους εύρεσης που βασίζονται σε μεθόδους σχεδιασμού λογικής, τον αλγόριθμο before-chaining και τον αλγόριθμο back-chaining.

#### 6.2.2.2.3 Αλγόριθμος before-chaining

Με τον αλγόριθμο before-chaining, κάθε κόμβος αναζητά την είσοδο μιας υπηρεσίας, που μπορεί να προχωρήσει τη διαδικασία σύνθεσης υπηρεσιών. Για την έναρξη μιας P2P αναζήτησης σύμφωνα με αυτόν τον αλγόριθμο, ο κόμβος

προετοιμάζει την λίστα List-WS-Init-Input των υπηρεσιών που έχουν είσοδο ίδια με την αρχική είσοδο (Init-Input). Αυτές οι τοπικές υπηρεσίες μπορεί να είναι είτε βασικές είτε μερικώς τοπικά συντεθειμένες.

---

### Αλγόριθμος before-chaining

---

1. **Έναρξη**
  2. Δημιουργία List-WS-Init-Input (η λίστα των υπηρεσιών που έχουν την ίδια είσοδο)
  3. **Όσο** (List-WS-Init-Input δεν είναι άδεια) κάνε
    - a) Επιλογή του πρώτου στοιχείου του List-WS-Init-Input
    - b) Init-Input:=έξοδος του πρώτου στοιχείου του List-WS-Init-Input και υπολογισμός του TTL
    - c) Αποστολή μηνύματος αναζήτησης (InitiatorPeer, Init-Input, Init-Output, goal, TTL)
    - d) **Αν** (χρόνος  $\leq$  TTL και απάντηση επιτυχούς αναζήτησης) **τότε** τερματισμός αναζήτησης, αποστολή απάντησης επιτυχίας και πήγαινε στο τέλος
    - e) **Αλλιώς αν** (χρόνος  $\leq$  TTL και απάντηση μη επιτυχούς αναζήτησης) **τότε** επιλογή μιας εφεδρικής υπηρεσίας από τη List-WS-Init-Input
    - f) **Τέλος αν**
  4. **Τέλος**
- 

#### 6.2.2.2.4 Αλγόριθμος back-chaining

Στον αλγόριθμο αναζήτησης back-chaining κάθε κόμβος αναζητά την έξοδο της υπηρεσίας που θα συνεχίσει την διαδικασία σύνθεσης υπηρεσιών. Αυτός ο αλγόριθμος χρησιμοποιεί τη List-WS-Init-Output, που είναι η λίστα των υπηρεσιών που έχουν ίδια έξοδο με την αρχική έξοδο (Init-Output).

---

### Αλγόριθμος back-chaining

---

5. **Έναρξη**
  6. Δημιουργία List-WS-Init-Output (η λίστα των υπηρεσιών που έχουν την ίδια έξοδο)
  7. **Όσο** (List-WS-Init-Output δεν είναι άδεια) κάνε
    - g) Επιλογή του πρώτου στοιχείου του List-WS-Init-Output
    - h) Init-Output:=έξοδος του πρώτου στοιχείου του List-WS-Init-Output και υπολογισμός του TTL
    - i) Αποστολή μηνύματος αναζήτησης (InitiatorPeer, Init-Input, Init-Output, goal, TTL)
    - j) **Αν** (χρόνος  $\leq$  TTL και απάντηση επιτυχούς αναζήτησης) **τότε** τερματισμός αναζήτησης, αποστολή απάντησης επιτυχίας και πήγαινε στο τέλος
-

k) **Αλλιώς αν** (χρόνος  $\leq$  TTL και απάντηση μη επιτυχούς αναζήτησης) **τότε** επιλογή μιας εφεδρικής υπηρεσίας από τη List-WS-Init-Output

l) **Τέλος αν**

## 8. Τέλος

---

Αν η αναζήτηση είναι επιτυχής ξεκινώντας την αναζήτηση με τον αλγόριθμο back-chaining, η σύνθετη υπηρεσία αποθηκεύεται στον πίνακα σύνθετων υπηρεσιών ως εξής: ο αρχικός κόμβος είναι τελευταίος συμμετέχων κόμβος και κάθε συμμετέχων κόμβος αλλάζει τον προηγούμενο και τον επόμενο κόμβο.

## 6.3 MASON

Το MASON είναι το αποτέλεσμα μιας κοινής προσπάθειας μεταξύ του εργαστηρίου εξελικτικής υπολογιστικής του πανεπιστημίου George Mason και του κέντρου κοινωνικής πολυπλοκότητας GMU. Είναι ένα εργαλείο προσομοίωσης πολυπρακτορικών συστημάτων που έχει σχεδιαστεί για να υποστηρίζει αποτελεσματικά έναν μεγάλο αριθμό πρακτόρων σε ένα μοναδικό μηχάνημα.

### 6.3.1 Αρχιτεκτονική

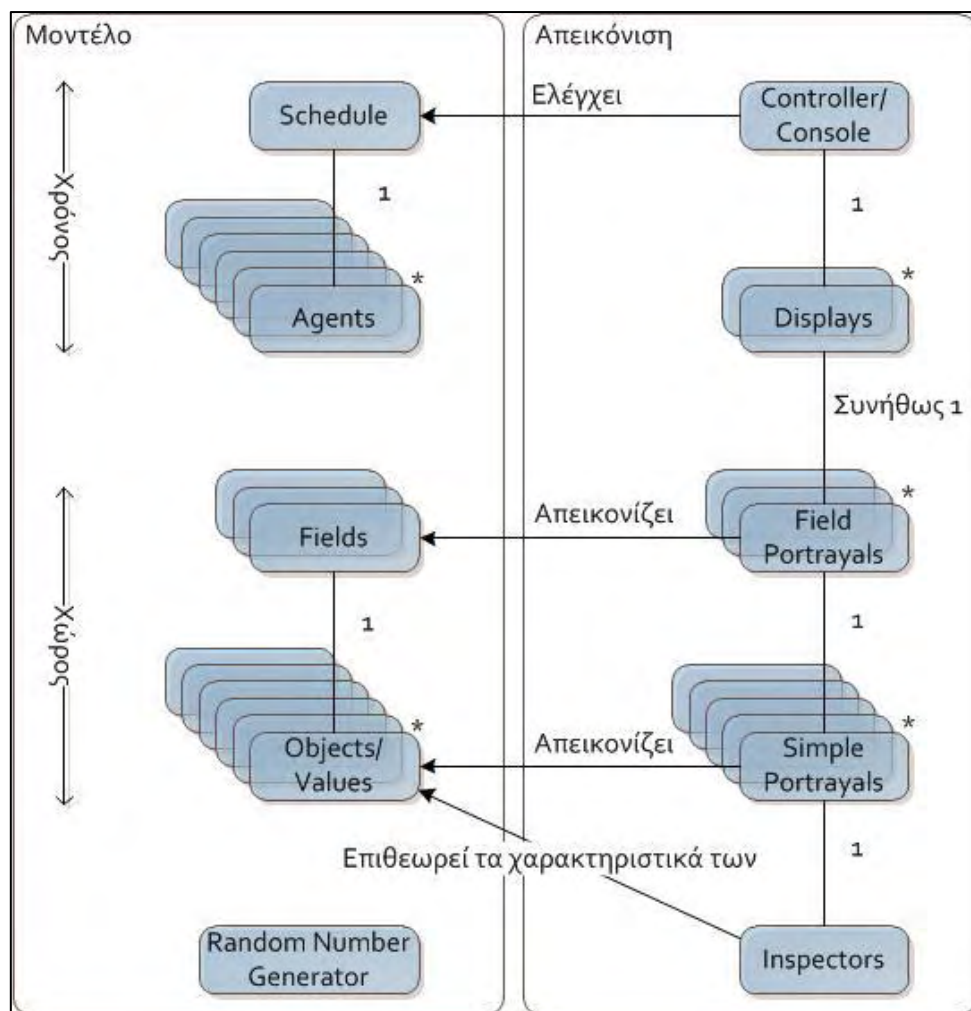
Το MASON αποτελείται από δύο ξεχωριστά τμήματα, το μοντέλο προσομοίωσης και την απεικόνιση (2D ή 3D). Αυτό σημαίνει ότι κάθε προσομοίωση είναι ανεξάρτητη από την απεικόνιση και μπορεί να πραγματοποιηθεί με ή χωρίς αυτή. Το μοντέλο μπορεί επίσης να σημειοελεγχθεί (checkpointed), που σημαίνει ότι μπορεί να σταματήσει και να εγγραφεί στο δίσκο, για να συνεχισθεί οποιαδήποτε άλλη στιγμή ακόμα και σε ένα άλλο είδος υπολογιστή. Στην εικόνα 6.2 απεικονίζεται η αρχιτεκτονική του MASON.

Το μοντέλο του MASON περιλαμβάνεται σε ένα ειδικό αντικείμενο που καλείται `sim.engine.SimState`. Περιέχει ένα πρόγραμμα (schedule) διακριτού χρόνου στο οποίο προγραμματίζονται οι διάφοροι πράκτορες ώστε να ανακληθούν κάποια στιγμή στο μέλλον. Ουσιαστικά αποτελεί την αναπαράσταση του μοντέλου του MASON στο χρόνο. Επιπλέον, το μοντέλο περιλαμβάνει ένα ή περισσότερα πεδία (fields) που αντιπροσωπεύουν τον χώρο. Ένα πεδίο είναι ουσιαστικά μια αυθαίρετη δομή δεδομένων που σχετίζουν διάφορα αντικείμενα ή τιμές μεταξύ τους. Το MASON παρέχει έναν αριθμό ενσωματωμένων πεδίων, όπως δίκτυα, συνεχή χώρο και πλέγματα. Τέλος, το μοντέλο περιλαμβάνει και μια υψηλής ποιότητας γεννήτρια τυχαίων αριθμών.

Το MASON παρέχει εργαλεία απεικόνισης που δίνουν τη δυνατότητα δισδιάστατης (2D) και τρισδιάστατης (3D) απεικόνισης, ενώ με τη δυνατότητα χρήσης plug-in παρέχουν επιπλέον δυνατότητες, όπως για GIS. Η απεικόνιση περιλαμβάνεται σε ένα ειδικό αντικείμενο που καλείται `sim.display.GUIState`. Μέσω ενός ελεγκτή (controller) δίνεται η δυνατότητα εκκίνησης, παύσης και διαχείρισης του προγράμματος (schedule). Ο πιο κοινός ελεγκτής είναι ένα παράθυρο που ονομάζεται κονσόλα (console). Ο ελεγκτής ή η

κονσόλα διαχειρίζονται επίσης έναν αριθμό παραθύρων που ονομάζονται οθόνες (displays), και οι οποίες χειρίζονται τις 2D ή 3D απεικονίσεις. Οι οθόνες απεικονίζουν πεδία χρησιμοποιώντας ένα ή περισσότερα πορτραίτα πεδίων (field portrayals), που συνήθως απεικονίζουν ανεξάρτητα αντικείμενα ή τιμές σε πεδία καλώντας ένα απλό πορτραίτο (simple portrayal) που είναι σχεδιασμένο να απεικονίζει αυτό το συγκεκριμένο αντικείμενο ή την τιμή. Τέλος, επιλέγοντας τα αντικείμενα στην οθόνη, τα απλά πορτραίτα δημιουργούν επιθεωρητές (inspector) που παρέχουν περαιτέρω επιθεώρηση των λεπτομερειών του μοντέλου.

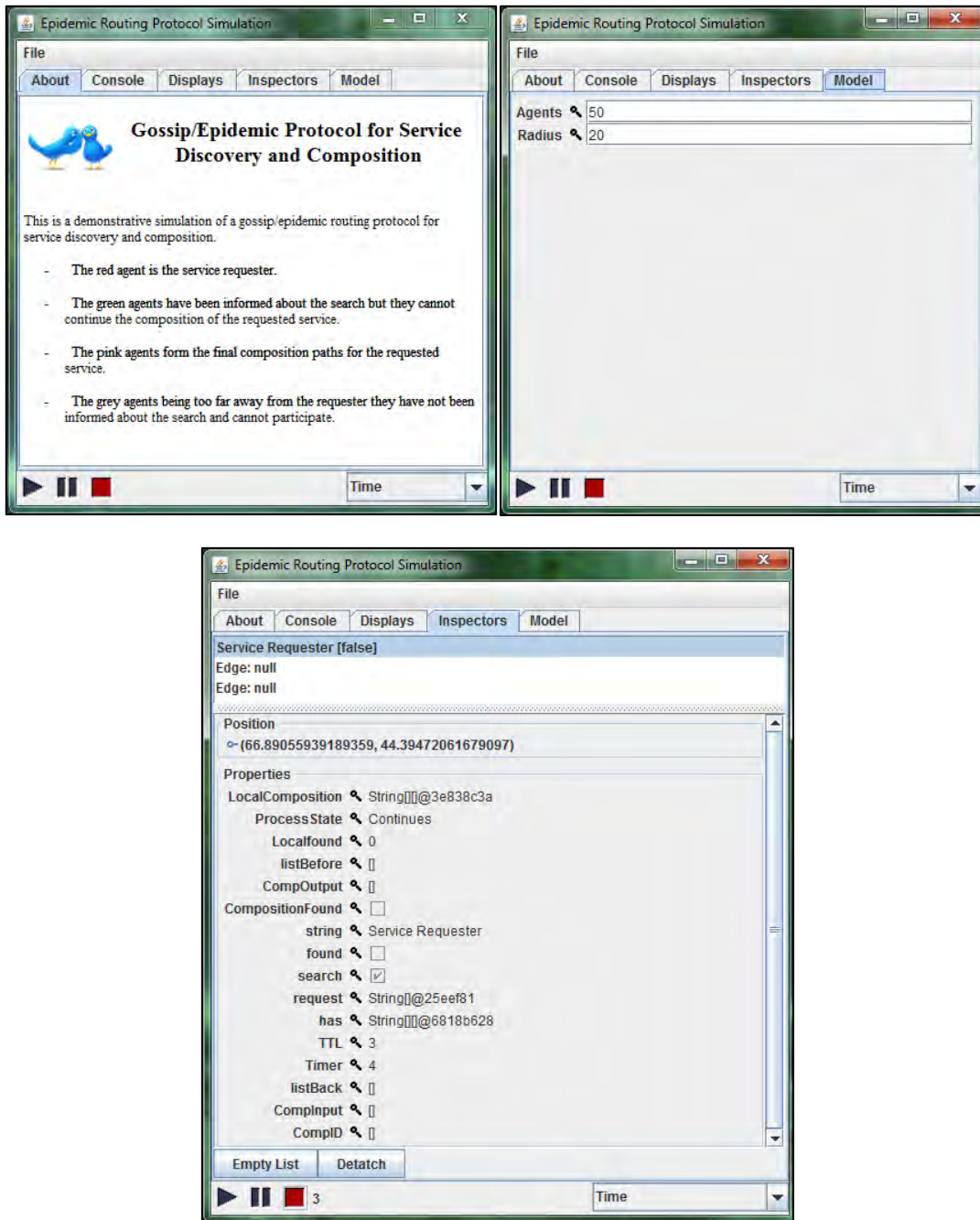
Επιπλέον, το MASON διαθέτει έναν μεγάλο αριθμό βοηθητικών προγραμμάτων για την υποστήριξη του μοντέλου σχεδιασμού, όπως είναι οι κατανομές τυχαίων αριθμών, διάφορες συλλογές αντικειμένων, αντικείμενα για τον έλεγχο ιδιοτήτων Java Bean, σημεία στο χώρο, ένα πλήθος GUI widgets, δημιουργία εικόνων, γραφημάτων και ταινιών.



Εικόνα 6.2 Αρχιτεκτονική πλατφόρμας MASON

### 6.3.2 Κονσόλα Διαχείρισης

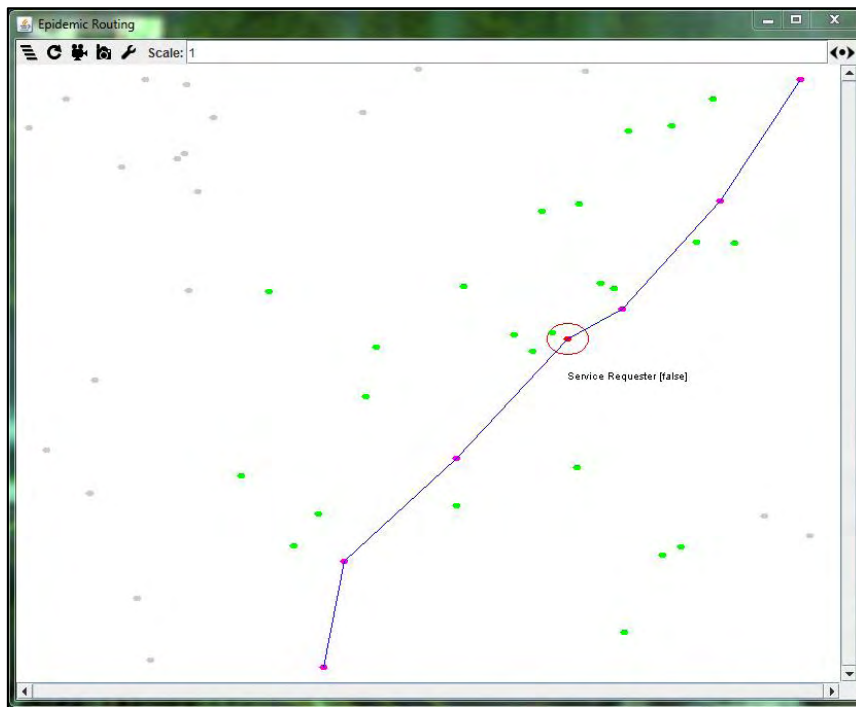
Η κονσόλα διαχείρισης αποτελεί το κέντρο εντολών του περιβάλλοντος προσομοίωσης (εικόνα 6.3). Ελέγχει την έναρξη, την παύση και τον τερματισμό της προσομοίωσης ενώ προβάλλει και κρύβει τις οθόνες και το πάνελ του ελεγκτή (inspector panel), μέσω του οποίου ελέγχουμε τις παραμέτρους της προσομοίωσης. Επιπλέον, μέσω της κονσόλας διαχείρισης γίνεται η αποθήκευση και η φόρτωση της προσομοίωσης από τα σημεία ελέγχου ενώ δίνεται και η δυνατότητα τερματισμού της προσομοίωσης.



Εικόνα 6.3 Στιγμιότυπα της κονσόλας διαχείρισης

### 6.3.3 Οθόνη Απεικόνισης

Στις προκαθορισμένες ρυθμίσεις η οθόνη απεικονίζει μια περιοχή 100x100. Χρησιμοποιώντας το κουμπί έναρξης της κεντρικής κονσόλας ξεκινά η προσομοίωση και προβάλλεται στον χρήστη μέσω της οθόνης (Εικόνα 6.4). Διαθέτει τις επιλογές λήψης εικόνας και βίντεο της προσομοίωσης ενώ η επιλογή των ρυθμίσεων δίνει επιπλέον δυνατότητες όπως είναι η δημιουργία γραφημάτων.



Εικόνα 6.4 Στιγμιότυπο της οθόνης απεικόνισης δύο διαστάσεων της προσομοίωσης

## 6.4 Αρχιτεκτονική Συναισθηματικών Πρακτόρων

Για τη δημιουργία ορθολογικών πρακτόρων το μοντέλο BDI (πεποιθήσεις, επιθυμίες και προθέσεις) αποτελεί την καλύτερη επιλογή. Παρ' όλα αυτά οι επιλογές των ανθρώπων δεν λαμβάνονται μόνο με βάση τις νοητικές τους λειτουργίες, αλλά σημαντικό ρόλο παίζουν και τα συναισθήματα. Αυτό έχει ως αποτέλεσμα την ανάπτυξη συναισθηματικών πρακτόρων, που στοχεύουν στην παροχή μιας πιο ρεαλιστικής εμπειρίας αλληλεπίδρασης ανθρώπων-μηχανής (Human Compute Interaction) αλλά και στην βελτίωση της απόδοσης των ορθολογικών πρακτόρων. Έτσι, δημιουργούνται πράκτορες οι οποίοι λαμβάνουν τις αποφάσεις τους σύμφωνα με τις πεποιθήσεις, τις επιθυμίες, τις προθέσεις και τα συναισθήματά τους.

Λαμβάνουμε ως παράδειγμα την αρχιτεκτονική των πρακτόρων που προτείνεται από τους Ruică and Florea [63] και η οποία επικεντρώνεται στην επιρροή των συναισθημάτων στη συμπεριφορά και τον τρόπο που χρησιμοποιούνται οι πόροι ανάλογα με τη συναισθηματική κατάσταση. Έτσι, ενσωματώθηκαν στην αρχιτεκτονική των πρακτόρων οι εξής έννοιες:

- **Αντιλήψεις (Percepts, P):**
  - Οτιδήποτε προέρχεται από το περιβάλλον, όπως ερεθίσματα και μηνύματα από άλλους πράκτορες
  - Επηρεάζονται από τα συναισθήματα
- **Πεποιθήσεις (Beliefs, B):**
  - Αποκτώνται από τις αντιλήψεις
  - Αναθεωρούνται για να λάβουν υπόψη τις τρέχουσες πεποιθήσεις και τις νέες αντιλήψεις
- **Επιθυμίες (Desires, D):**
  - Οι στόχοι που ορίζονται για κάθε πράκτορα κατά το σχεδιασμό
  - Παραμένουν σταθερές
- **Επιλογές (Options, O):**
  - Οι διαφορετικές εναλλακτικές για την επίτευξη των επιθυμιών
  - Δημιουργούνται με βάση τις τρέχουσες πεποιθήσεις και προθέσεις
  - Επηρεάζονται από τα συναισθήματα
- **Προθέσεις (Intentions, I):**
  - Επιλογές για τις οποίες ο πράκτορας έχει δεσμευτεί
  - Αναθεωρούνται με βάση τις τρέχουσες προθέσεις, πεποιθήσεις, επιλογές, συναισθήματα και διαθέσιμων πόρων
- **Συναισθήματα (Emotions, E):**
  - Επηρεάζουν τη συμπεριφορά, τις νοητικές διεργασίες και τους διαθέσιμους πόρους
  - Προκαθορισμένο σύνολο συναισθημάτων για κάθε σενάριο
- **Πόροι (Resources, R):**
  - Κάθε πράκτορας διαθέτει ένα σύνολο διαθέσιμων πόρων
  - Οι διαθέσιμοι πόροι είναι προκαθορισμένοι για κάθε σενάριο και για κάθε πράκτορα

Οι συναρτήσεις σύμφωνα με τις οποίες ορίζονται οι διεργασίες που λαμβάνουν χώρα στην εσωτερικούς μηχανισμούς των πρακτόρων είναι οι εξής:

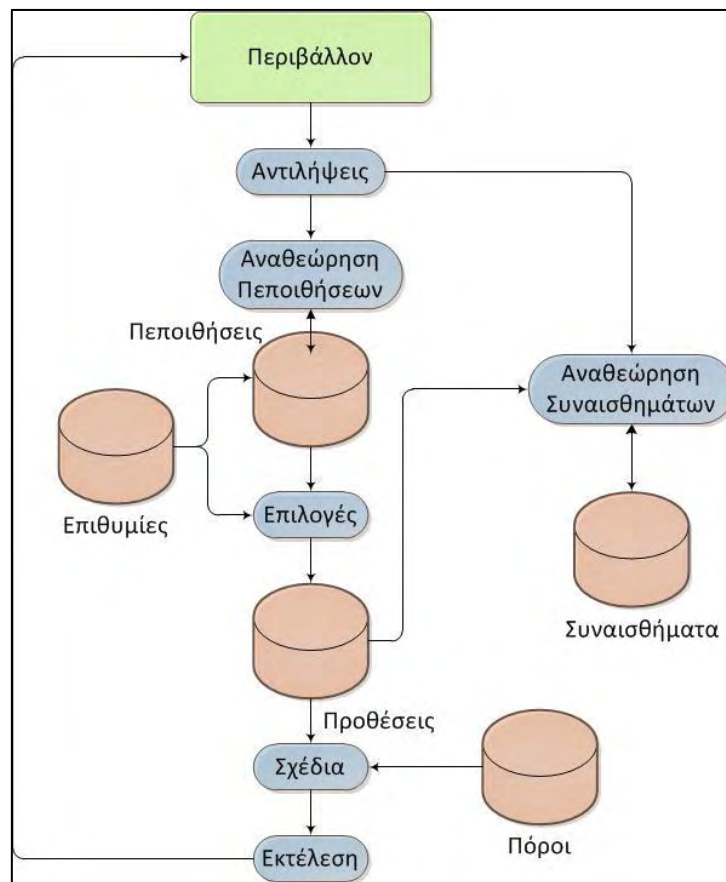
- Αντιλήψεις:  $Env \rightarrow P$  – λαμβάνοντας τα διάφορα ερεθίσματα από το περιβάλλον ( $Env$ ), και σύμφωνα με τα συναισθήματα ( $E$ ) δίδονται οι αντιλήψεις.
- Αναθεώρηση συναισθημάτων:  $P \times E \times I \rightarrow E$  – οι νέες αντιλήψεις, επηρεαζόμενες από τις προθέσεις και τα τρέχοντα συναισθήματα, αναθεωρούν τα συναισθήματα.
- Αναθεώρηση πεποιθήσεων:  $P \times B \rightarrow B$  – αναθεώρηση των τρεχόντων πεποιθήσεων σύμφωνα με τις αντιλήψεις.
- Σχέδια:  $I \times R \rightarrow \pi$  – διαρθρώνει τις προθέσεις σε σχέδια σύμφωνα με τους διαθέσιμους πόρους.

Στο σχήμα της εικόνας 6.5 απεικονίζεται η εσωτερική αρχιτεκτονική των πρακτόρων. Αρχικά ο πράκτορας αντιλαμβάνεται τα ερεθίσματα του περιβάλλοντος, τα οποία μπορεί να είναι το αποτέλεσμα μιας ενέργειας, μια αλλαγή στην κατάσταση του κόσμου ή ένα μήνυμα από έναν άλλο πράκτορα. Οι αντιλήψεις με τη σειρά τους καθορίζουν και τη διαμόρφωση των συναισθημάτων, όπως είναι η εμπιστοσύνη. Οι αντιλήψεις σε



συνδυασμό με τους διαθέσιμους πόρους προσδιορίζουν τις ενέργειες του κάθε πράκτορα σύμφωνα με τα σχέδιά του.

Οι αντιλήψεις επηρεάζουν την αναθεώρηση των πεποιθήσεων. Αυτή η διαδικασία λαμβάνει υπόψη τόσο τις τρέχουσες πεποιθήσεις όσο και τις νέες αντιλήψεις, το οποίο λειτουργεί ως ένα σύστημα διατήρησης της αλήθειας το οποίο εξαγει συνεπείς πεποιθήσεις. Στη συνέχεια προσδιορίζονται οι διαθέσιμες επιλογές ώστε ο πράκτορας να επιτύχει τις επιθυμίες του. Οι επιλογή που επιλέγεται στην κάθε περίπτωση αποτελεί τις προθέσεις του πράκτορα. Εφόσον προσδιοριστούν οι προθέσεις, ο πράκτορας δημιουργεί το σχέδιο σύμφωνα με το οποίο θα τις επιτύχει, λαμβάνοντας υπόψη τους διαθέσιμους πόρους. Έτσι, στο τέλος το σχέδιο είναι έτοιμο να εκτελεσθεί επιφέροντας τις αντίστοιχες αλλαγές στο περιβάλλον.



Εικόνα 6.5 Εσωτερική αρχιτεκτονική των πρακτόρων

## 6.5 Μοντελοποίηση της Εμπιστοσύνης ενός Πολυπρακτορικού Συστήματος

Για τη μοντελοποίηση της εμπιστοσύνης ενός πολυπρακτορικού συστήματος υλοποιήθηκε ο αλγόριθμος που παρουσιάστηκε στην παράγραφο 6.2.2. Οι έννοιες της εμπιστοσύνης, της μετάνοιας και της ντροπής υλοποιήθηκαν στο μοντέλο σύμφωνα με τις προτάσεις του κεφαλαίου 5.

### 6.5.1 Σενάριο Υλοποίησης

Το σενάριο που υλοποιήθηκε δημιουργεί μια σύνθετη υπηρεσία από τρεις βασικές υπηρεσίες:

- Search-ISBN: επιστρέφει τον κωδικό ISBN ενός βιβλίου
- Search-Price: η υπηρεσία λαμβάνει ως είσοδο τον κωδικό ISBN ενός βιβλίου και επιστρέφει την τιμή του σε αμερικανικό δολάριο
- Price-Converter: αποτελεί έναν μετατροπέα συναλλάγματος από αμερικανικό δολάριο σε ευρώ

Στον πίνακα 6.1 παρουσιάζεται επιγραμματικά η είσοδος και η έξοδος της κάθε ατομικής υπηρεσίας που χρησιμοποιήθηκε για την υλοποίηση του σεναρίου.

Πίνακας 6.1 Η είσοδος και η έξοδος των ατομικών υπηρεσιών

Υπηρεσία	Είσοδος	Έξοδος
Search-ISBN	Book-Title	ISBN
Search-Price	ISBN	Price-USD
Price-Converter	Price-USD	Price-EUR

Η σύνθεση των υπηρεσιών του πίνακα 6.1 δίνει μια σύνθετη υπηρεσία της παρακάτω μορφής:

Είσοδος: Book-Title

Έξοδος: Price-EUR

Στόχος: \*input:=Book-Title\*output:=Price-EUR

Το περιβάλλον εκτέλεσης μιας σύνθετης υπηρεσίας είναι εξαιρετικά δυναμικό γεγονός που σημαίνει ότι υπάρχουν περιπτώσεις όπου η εκτέλεση της σύνθετης υπηρεσίας δεν μπορεί να συνεχιστεί. Στον πίνακα 6.2 παρουσιάζονται δύο συμβάντα (events) που μπορεί να παρουσιαστούν κατά την εκτέλεση μιας σύνθετης υπηρεσίας και τα οποία χρησιμοποιήθηκαν στην υλοποίηση του αλγόριθμου.

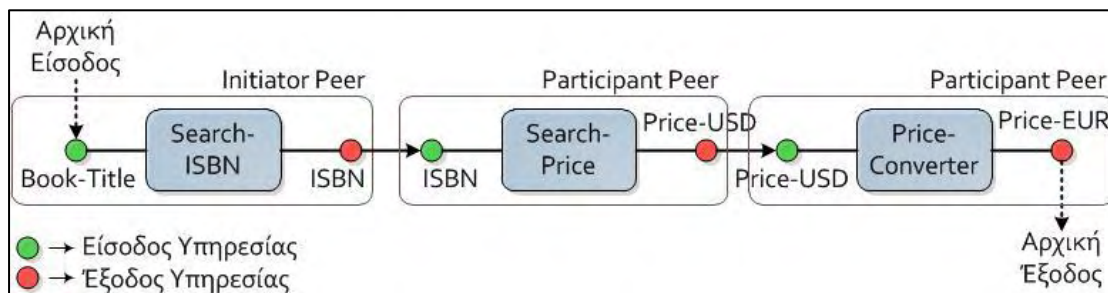
Πίνακας 6.2 Προϋποθέσεις συμβάντων

Event	Προϋποθέσεις
Deadline miss	Όταν μια υπηρεσία δεν έχει ολοκληρωθεί πριν την προκαθορισμένη προθεσμία
Execution timeout	Όταν μια υπηρεσία δεν έχει ξεκινήσει την εκτέλεσή της εντός ενός προκαθορισμένου χρονικού περιθωρίου

Για απλοποίηση, το σενάριο επιλέχθηκε να αναπτυχθεί σε τρεις κόμβους (peers) του δικτύου. Ο πρώτος κόμβος (Initiator Peer, Service Requester) υλοποιεί την υπηρεσία 'Search-ISBN', ο δεύτερος κόμβος (Participant Peer, Service Provider 1) υλοποιεί την υπηρεσία 'Search-Price' και ο τρίτος (Participant Peer, Service Provider 2) την υπηρεσία 'Price-Converter', όπως απεικονίζεται στην εικόνα 6.6. Στον πίνακα 6.3 παρουσιάζεται η αρχικοποίηση των στοιχείων του κάθε πράκτορα.

Πίνακας 6.3 Αρχικοποίηση των στοιχείων των πρακτόρων

<b>Αντιλήψεις</b>	- Εισαγωγή στοιχείων αναζήτησης από τον χρήστη - Μηνύματα από τους άλλους πράκτορες
<b>Πεποιθήσεις</b>	- Υπηρεσίες που διαθέτουν οι γειτονικοί πράκτορες
<b>Επιθυμίες</b>	- Δημιουργία σύνθετης υπηρεσίας
<b>Προθέσεις</b>	- Να εκτελέσει την υπηρεσία για την οποία έχει δεσμευθεί
<b>Συναισθήματα</b>	- Ισχυρή εμπιστοσύνη: η σύνθετη υπηρεσία εκτελέστηκε με επιτυχία - Ασθενής εμπιστοσύνη: ο πράκτορας λαμβάνει ένα deadline miss event - Ρήξη εμπιστοσύνης: ο πράκτορας λαμβάνει ένα execution timeout event - Μετάνοια: όταν η σύνθετη υπηρεσία δεν εκτελέστηκε από τον άλλο πράκτορα όπως είχε συμφωνηθεί (έχει ληφθεί execution timeout event) - Ντροπή: όταν ο πράκτορας αντιλαμβάνεται ότι απέτυχε να ικανοποιήσει τις προθέσεις του (έχει ληφθεί execution timeout event)
<b>Πόροι</b>	- Οι υπηρεσίες που διαθέτει ο πράκτορας



Εικόνα 6.6 Παράδειγμα μιας Peer-to-peer σύνθετης υπηρεσίας

## 6.5.2 Πειραματικά Αποτελέσματα

Πραγματοποιήθηκαν ενδεικτικά 10 διαφορετικές εκτελέσεις του αλγόριθμου του μοντέλου (Εικόνα 6.7) όπου λήφθηκαν τα αποτελέσματα που παρουσιάζονται στον πίνακα 6.4.

Πίνακας 6.4 Αποτελέσματα εκτέλεσης μοντέλου

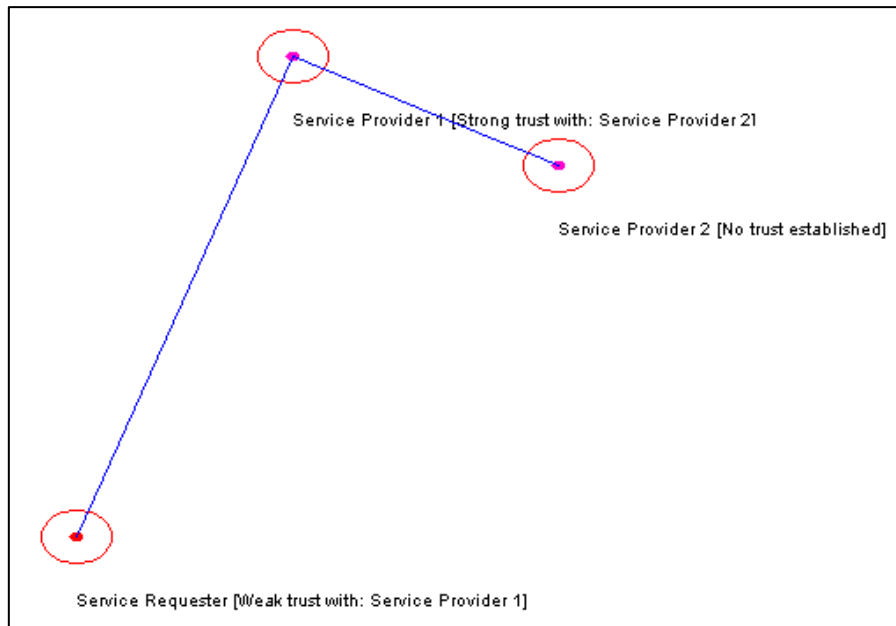
A/A	Πράκτορας	Εκτέλεση Υπηρεσίας	Σχέση Εμπιστοσύνης	Μετάνοια	Ντροπή
1	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ρήξη εμπιστοσύνης με τον Service Provider 2	Μετάνοια για τον Service Provider 2	Όχι
	Service Provider 2	Execution timeout	-	Όχι	Ντροπή προς τον Service

		event			Provider 1
2	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ασθενής εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Deadline miss event	-	Όχι	Όχι
3	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ασθενής εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Deadline miss event	-	Όχι	Όχι
4	Service Requester	Ναι	Ρήξη εμπιστοσύνης με τον Service Provider 1	Μετάνοια για τον Service Provider 1	Όχι
	Service Provider 1	Execution timeout event	Ρήξη εμπιστοσύνης με τον Service Provider 2	Μετάνοια για τον Service Provider 2	Ντροπή προς τον Service Requester
	Service Provider 2	Execution timeout event	-	Όχι	Ντροπή προς τον Service Provider 1
5	Service Requester	Ναι	Ρήξη εμπιστοσύνης με τον Service Provider 1	Μετάνοια για τον Service Provider 1	Όχι
	Service Provider 1	Execution timeout event	Ισχυρή εμπιστοσύνη με τον Service Provider 2	Όχι	Ντροπή προς τον Service Requester
	Service Provider 2	Ναι	-	Όχι	Όχι
6	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Ναι	-	Όχι	Όχι
7	Service Requester	Ναι	Ρήξη εμπιστοσύνης με τον Service Provider 1	Μετάνοια για τον Service Provider 1	Όχι
	Service	Execution	Ισχυρή	Όχι	Ντροπή προς

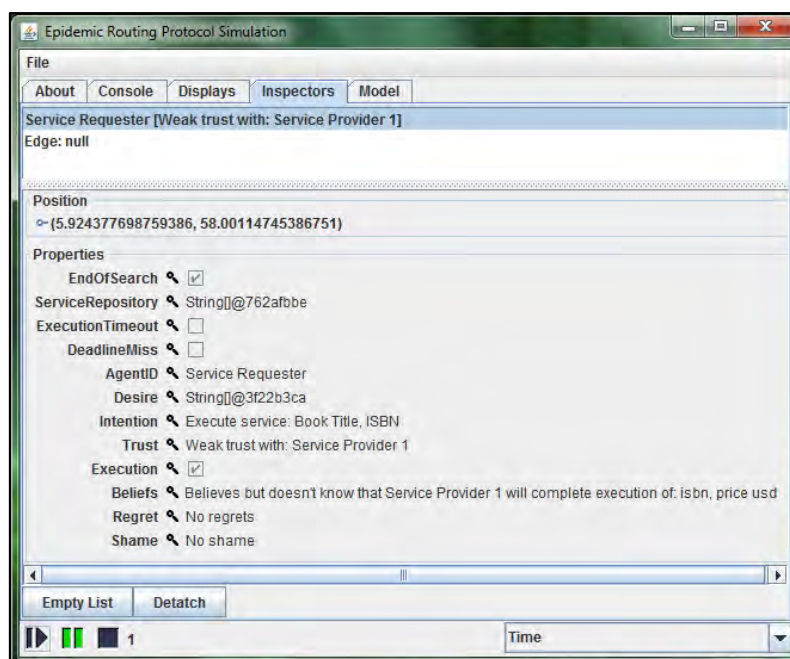
	Provider 1	timeout event	εμπιστοσύνη με τον Service Provider 2		τον Service Requester
	Service Provider 2	Ναι	-	Όχι	Όχι
8	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ασθενής εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Deadline miss event	-	Όχι	Όχι
9	Service Requester	Ναι	Ισχυρή εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Ναι	Ασθενής εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Deadline miss event	-	Όχι	Όχι
10	Service Requester	Ναι	Ασθενής εμπιστοσύνη με τον Service Provider 1	Όχι	Όχι
	Service Provider 1	Deadline miss event	Ισχυρή εμπιστοσύνη με τον Service Provider 2	Όχι	Όχι
	Service Provider 2	Ναι	-	Όχι	Όχι

Στο παρόν μοντέλο εφαρμόστηκαν οι τυπικολογικοί (modal-logic) χαρακτηρισμοί ορισμένων συναισθημάτων όπως περιγράφηκαν από τους Andrew Jones και Jeremy Pitt [48], οι οποίοι δίνουν τη δυνατότητα να περιγραφούν περίπλοκα συναισθήματα, όπως είναι η εμπιστοσύνη και η ντροπή, σε πολυπρακτορικά συστήματα. Σύμφωνα με τα αποτελέσματα του πίνακα 6.3, παρατηρούμε ότι δημιουργείται ρήξη στην εμπιστοσύνη μεταξύ δύο πρακτόρων όταν, για οποιονδήποτε λόγο ένας εκ των δύο αποτύχει να εκπληρώσει τις προθέσεις του και λαμβάνεται μήνυμα execution timeout. Το μήνυμα execution timeout μπορεί να ληφθεί είτε αν ο πράκτορας είναι κακοπροαίρετος, όπου έχει σκόπιμα αποφασίσει να μην τηρήσει τις δεσμεύσεις του, είτε εξαιτίας κάποιας βλάβης που τον καθιστά ανίκανο να ξεκινήσει την εκτέλεση της υπηρεσίας, όπως είναι μια βλάβη στο σύστημα ανταλλαγής μηνυμάτων. Αντίθετα, το μήνυμα deadline miss λαμβάνεται μόνο όταν κάποιος πράκτορας αδυνατεί να ολοκληρώσει την εκτέλεση της υπηρεσίας εντός των χρονικών περιθωρίων είτε εξαιτίας κάποιας βλάβης είτε λόγω κάποιου άλλου προβλήματος που παρουσιάστηκε κατά την διαδικασία εκτέλεσης, όπως είναι ο λάθος υπολογισμός χρονικής διάρκειας.

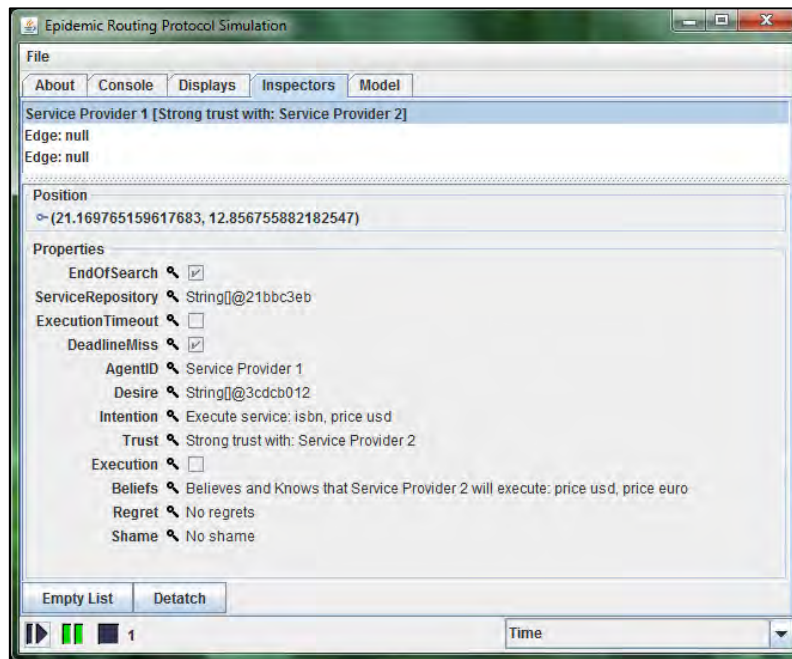
Επιπλέον, σύμφωνα με τη ροή της εκτέλεσης μιας σύνθετης υπηρεσίας δημιουργούνται συναισθήματα μετάνοιας προς τους πράκτορες που έχουν αποτύχει να εκτελέσουν τις υπηρεσίες για τις οποίες δεσμεύτηκαν, ενώ αντίστοιχα οι συγκεκριμένοι πράκτορες εμφανίζουν σημάδια ντροπής που δεν φάνηκαν ισάξιοι των προσδοκιών των άλλων πρακτόρων του συστήματος. Η αναπαράσταση των συγκεκριμένων συναισθημάτων είναι ιδιαίτερα χρήσιμη σε ανοικτά πολυπρακτορικά συστήματα μεγάλης κλίμακας καθώς δίνει τη δυνατότητα δημιουργίας ενός ασφαλούς περιβάλλοντος συνεργασίας, όπου κακόβουλοι πράκτορες θα μπορούσαν να απομονωθούν αυτόματα από το υπόλοιπο σύστημα.



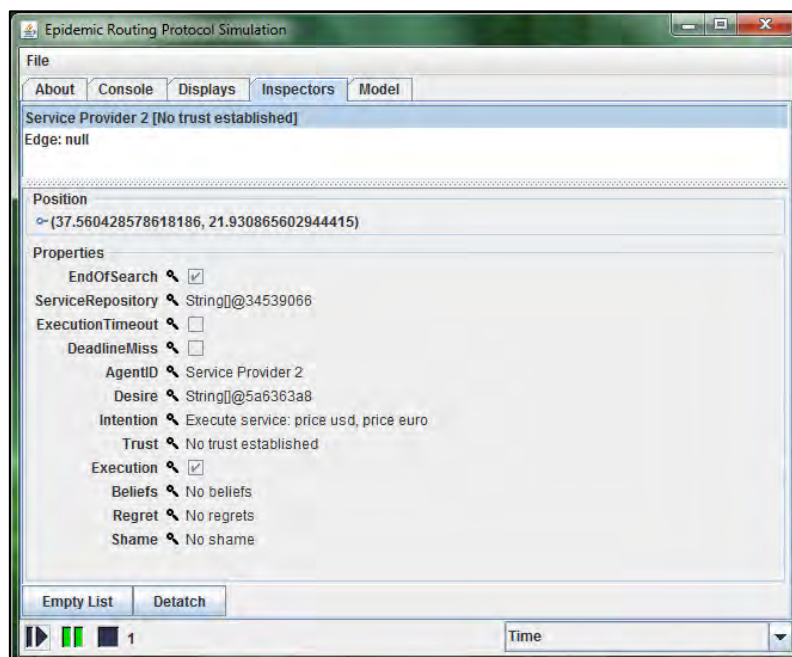
Εικόνα 6.7 Εμπιστοσύνη στο μοντέλο τριών κόμβων Peer-to-peer σύνθεσης υπηρεσιών



(α)



(β)



(γ)

Εικόνα 6.8 Οι παράμετροι των πρακτόρων του μοντέλου τις εικόνες 6.7 για τους (α) service requester, (β) service provider 1 και (γ) service provider 2

## 6.6 Μελλοντικές Προεκτάσεις

Η παρούσα μεταπτυχιακή διατριβή αποτελεί μια αρχική προσπάθεια για την εφαρμογή των τυπικολογικών χαρακτηρισμών των για την περιγραφή πολύπλοκων συναισθημάτων, όπως

είναι η εμπιστοσύνη, στα πολυπρακτορικά συστήματα. Καθώς η εμπιστοσύνη αποτελεί ένα από τα πιο σημαντικά ζητήματα όσον αφορά μεγάλης κλίμακας καταναμημένα συστήματα, προτείνονται οι εξής κατευθύνσεις για μελλοντική έρευνα:

- Η λεπτομερής και συστηματική σύγκριση της τυπολογικής προσέγγισης για την ταξινόμηση των συναισθημάτων με την ψυχοσωματική ταξινόμηση.
- Λεπτομερής διερεύνηση του μηχανισμού της συγχώρεσης, της παραβίασης της εμπιστοσύνης και του αισθήματος της ντροπής.
- Επέκταση του μοντέλου για την περιγραφή περαιτέρω συναισθημάτων, όπως είναι η απογοήτευση.
- Ανάπτυξη ενός εκτεταμένου πολυπρακτορικού μοντέλου για τη μελέτη των συναισθημάτων της εμπιστοσύνης, της ντροπής και της συγχώρεσης.
- Ανάπτυξη ενός μετρικού συστήματος για την αξιολόγηση των συναισθημάτων της εμπιστοσύνης, της ντροπής και της συγχώρεσης.
- Δημιουργία ενός συστήματος αξιολόγησης των πρακτόρων ενός πολυπρακτορικού συστήματος με βάση το ιστορικό της εμπιστοσύνης τους.



## Βιβλιογραφία

1. Russell, S. and P. Norvig, *Artificial Intelligence: A Modern Approach* 1995, Englewood Cliffs, New Jersey: Prentice-Hall.
2. Αθανάσογλου, Ι., *Πολυπρακτορικά Συστήματα: Υφιστάμενη Κατάσταση & Συγκριτική Μελέτη Εφαρμογών*, 2009, Πανεπιστήμιο Μακεδονίας: Θεσσαλονίκη. p. 128.
3. Wooldridge, M. and N.R. Jennings, *Intelligent Agents: Theory and Practice*. The Knowledge Engineering Review, 1995. **10**(2): p. 38.
4. Coen, M. *SodaBot: A Software Agent Construction System*. in *Proceedings of the 1994 Conference on Information and Knowledge Management Workshop on Intelligent Information Agents*. 1995. MIT AI.
5. Wooldridge, M., *An Introduction to Multiagent Systems* 2002: John Wiley & Sons Ltd.
6. Genesereth, M.R. and N.J. Nilsson, *Logical foundations of artificial intelligence*. Vol. 9. 1987: Morgan Kaufmann Los Altos, CA.
7. Brooks, R.A., *Intelligence without representation*. Artificial Intelligence, 1991. **47**(1): p. 139-159.
8. Rao, A.S. and M.P. Georgeff. *BDI Agents: From Theory to Practice*. in *ICMAS*. 1995.
9. Fabio, B., C. Giovanni, and G. Dominic, *Developing Multi-Agent Systems with JADE* 2007: Wiley. 303.
10. Georgeff, M.P. and A.L. Lansky. *Reactive reasoning and planning*. in *AAAI*. 1987.
11. Huber, M.J. *JAM: A BDI-theoretic mobile agent architecture*. in *Proceedings of the third annual conference on Autonomous Agents*. 1999. ACM.
12. Howden, N., et al. *JACK intelligent agents-summary of an agent infrastructure*. in *5th International conference on autonomous agents*. 2001.
13. d'Inverno, M., et al., *A formal specification of dMARS*, in *Intelligent Agents IV Agent Theories, Architectures, and Languages* 1998, Springer. p. 155-176.
14. Pokahr, A., L. Braubach, and W. Lamersdorf, *Jadex: A BDI reasoning engine*, in *Multi-agent programming* 2005, Springer. p. 149-174.
15. Virmani, C., *A Comparison of Communication Protocols for Mobile Agents*. International Journal of Advancements in Technology, 2012. **3**(2): p. 114-122.
16. Huhns, M.N. and L.M. Stephens, *Semantic Bridging of Independent Enterprise Ontologies*. Inter- and Intra-Organizational Integration: Building International Consensus, 2002.
17. Nii, H.P., *Blackboard Systems Part One*. AI Magazine, 1986a. **7**(2): p. 16.
18. Demazeau, Y. *From interactions to collective behaviour in agent-based systems*. in *In: Proceedings of the 1st. European Conference on Cognitive Science. Saint-Malo*. 1995. Citeseer.
19. <http://aibook.csd.auth.gr/include/ch28.pdf>, Κεφάλαιο 28, Πολυπρακτορικά συστήματα.
20. Silaghi, G.C., *Contributions to Conception, Design and Development of Collaborative Multi-Agent Systems*, 2005, PhD thesis, Babes-Bolyai University Cluj-Napoca, Romania.
21. Prachi, S.S. and R. Setiya, *Evaluation of Agent Oriented Software Engineering (AOSE) Methodologies - A review*. International Journal of Latest Research in Science and Technology, 2012. **1**(2): p. 4.
22. Wooldridge, M., N.R. Jennings, and D. Kinny, *The Gaia methodology for agent-oriented analysis and design*. Autonomous Agents and Multi-Agent Systems, 2000. **3**(3): p. 285-312.

23. Iglesias, C.A. and M. Garijo, *The agent-oriented methodology MAS-CommonKADS*. Agent-oriented methodologies, 2005. **3**: p. 46-78.
24. Sudeikat, J., et al., *Evaluation of agent-oriented software methodologies—examination of the gap between modeling and platform*, in *Agent-Oriented Software Engineering V2005*, Springer. p. 126-141.
25. Kristensen, B.B. and K. Østerbye, *Roles: Conceptual Abstraction Theory and Practical Language Issues*. Theory and Practice of Object Systems, 1996. **2**(3): p. 18.
26. Yu, E., *Modelling strategic relationships for process reengineering*. Social Modeling for Requirements Engineering, 2011. **11**: p. 2011.
27. Bresciani, P., et al., *Tropos: An agent-oriented software development methodology*. Autonomous Agents and Multi-Agent Systems, 2004. **8**(3): p. 203-236.
28. Drasgupta, P., *Trust as a commodity*. Trust: Making and Breaking Cooperative Relations, 1998: p. 25.
29. Kelly, C., et al., *Guidelines for trust in future ATM systems: A literature review.*, 2003, European Organisation for the Safety of Air Navigation. p. 7.
30. Tweedale, J. and P. Cutler, *Trust in Multi-Agent Systems*. KES, 2006: p. 7.
31. Ramchurn, S.D., D. Huynh, and N.R. Jennings, *Trust in Multi-Agent Systems*. The Knowledge Engineering Review, 2004. **19**(1): p. 25.
32. Wu, D.J. and Y. Sun. *The Emergence of Trust in Multi-Agent Bidding: a Computational Approach*. in *Proceedings of the 34th Hawaii International Conference on System Sciences*. 2001. Hawaii: IEEE Computer Society Press.
33. Sen, S. *Reciprocity: a Foundational Principle for Promoting Cooperative Behavior Among Self-interested Agents*. in *Proceedings of the Second International Conference on Multi-Agent Systems*. 1996. Menlo Park, CA: AAAI Press.
34. Mukherjee, R., B. Banerjee, and S. Sen, *Learning Mutual Trust*. Trust in Cyber-societies, 2001: p. 14.
35. Birk, A., *Boosting Cooperation by Evolving Trust* Applied Artificial Intelligence, 2000. **14**(8): p. 16.
36. Birk, A., *Learning to Trust*. In Falcone, R., Singh, M. & Tan, Y.-H. (eds.), *Trust in Cyber-societies*. Berlin: Springer-Verlag, 2001: p. 12.
37. Witkowski, M., A. Artikis, and J. Pitt, *Experiments in Building Experiential Trust in a Society of Objective Trust-Based Agents*. In Falcone, R., Singh, M. & Tan, Y.-H. (eds.), *Trust in Cyber-societies*. Berlin: Springer-Verlag, 2001: p. 22.
38. Mui, L., M. Mohtashemi, and A. Halberstadt. *A Computational Model of Trust and Reputation for e-Business*. in *35th Hawaii International Conference on System Science (HICSS 35 CDROM)*. 2002. Hawaii: IEEE Computer Society (online publication).
39. Schillo, M., P. Funk, and M. Rovatsos, *Using Trust for Detecting Deceptive Agents in Artificial Societies*. Applied Artificial Intelligence, Special Issue on Trust, Deception, and Fraud in Agent Societies, 2000. **14**(8): p. 24.
40. Sen, S. and N. Sajja, *Robustness of Reputation-Based Trust: Boolean Case*. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2002. **1**: p. 6.
41. Sabater, J. and C. Sierra, *REGRET: a Reputation Model for Gregarious Societies*. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2002: p. 8.
42. Castelfranchi, C., R. Falcone, and G. Pezzulo. *Trust in Information Sources as a Source for Trust: A Fuzzy Approach*. in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*. 2003. ACM Press.

43. Mass, Y. and O. Shehory, *Distributed Trust in Open Multi-Agent Systems*. In Falcone, R., Singh, M. & Tan, Y.-H. (eds.), *Trust in Cyber-societies*. Berlin: Springer-Verlag, 2001: p. 15.
44. Poslad, S., M. Calisti, and P. Charlton, *Specifying Standard Security Mechanisms in Multi-Agent Systems*. Workshop on Deception, Fraud and Trust in Agent Societies, AAMAS 2002, 2002: p. 6.
45. Herzberg, A., et al. *Access Control Meets Public Key Infrastructure, or: Assigning Roles to Strangers*. in *IEEE Symposium on Security and Privacy*. 2000. Oakland: IEEE Computer Society.
46. *Modal Logic*. 2008; Available from: [http://en.wikipedia.org/wiki/Modal\\_logic#Formalizations](http://en.wikipedia.org/wiki/Modal_logic#Formalizations).
47. Kripke, S., *Semantical Analysis of Modal Logic*. Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik, 1963. **9**: p. 30.
48. Jones, A.J.I. and J. Pitt. *On the Classification of Emotions, and its Relevance to the Understanding of Trust*. in *Trust in Agent Societies, Autonomous Agents & Multi-Agent Systems Conference, AAMAS 2011*. 2011. Taipei, Taiwan.
49. Pörn, I., *On the Nature of Emotions*. Changing Positions, Philosophical Studies, Department of Philosophy, University of Uppsala, 1986. **38**: p. 10.
50. Jones, A.J.I., *On the Concept of Trust*. Decision Support Systems, 2002. **33**(3): p. 8.
51. Plutchik, R., *Emotion: A Psychoevolutionary Synthesis*. Harper & Row, 1980.
52. Plutchik, R., *The Nature of Emotions*. <http://www.fractal.org/Bewustzijns-Besturings-Model/Nature-of-emotions.htm>.
53. Vasalou, A., J. Pitt, and G. Piolle, *From Theory to Practice: Forgiveness as a Mechanism to Repair Conflicts in CMC*. Proceedings iTrust IV, 2006: p. 15.
54. Vasalou, A., A. Hopfensitz, and J. Pitt, *In Praise of Forgiveness: Ways for Repairing Trust Breakdowns in One-Off Online Interactions*. International Journal of Human-Computer Studies, 2008. **66**: p. 15.
55. Castelfranchi, C. and E. Lorini, *Cognitive Anatomy and Functions of Expectations*. IJCAI'03 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions, 2003: p. 7.
56. Gharzouli, M. and M. Boufaida, *Pm4sws: A p2p model for semantic web services discovery and composition*. Journal of Advances in Information Technology, 2011. **2**(1): p. 15-26.
57. <http://cs.gmu.edu/~eclab/projects/mason/>.
58. W3C, *Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language*. <http://www.w3.org/TR/wsdl20/>, 2007.
59. W3C, *SOAP Version 1.2 Part 1: Messaging Framework (Second Edition)*. <http://www.w3.org/TR/soap12-part1/>, 2007.
60. W3C, *Web Services Glossary*. <http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice>, 2004.
61. Biswas, D., *Web services discovery and constraints composition*, in *Web Reasoning and Rule Systems 2007*, Springer. p. 73-87.
62. Sycara, K., et al., *Automated discovery, interaction and composition of semantic web services*. Web Semantics: Science, Services and Agents on the World Wide Web, 2003. **1**(1): p. 27-46.
63. Puică, M.-A. and A.-M. Florea, *Emotional Belief-Desire-Intention Agent Model: Previous Work and Proposed Architecture*. International Journal of Advanced Research in Artificial Intelligence (IJARAI), 2013. **2**(2).
64. Vaniya, S., B. Lad, and S. Bhavsar, *A Survey on Agent Communication Languages*. International Proceedings of Economics Development & Research, vol. 14, 2011.

