ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Τεχνητή νοημοσύνη: μπορεί να έχει συνείδηση και να ποσοτικοποιηθεί;

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικόλαου Βούλγαρη

Επιβλέπων: Δημήτριος-Διονύσιος Κουτσούρης
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2019

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ & ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Τεχνητή νοημοσύνη: μπορεί να έχει συνείδηση και να ποσοτικοποιηθεί;

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικόλαου Βούλγαρη

Επιβλέπων: Δημήτριος-Διονύσιος Κουτσούρης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την η Ιουνίου 2019.

(Υπογραφή)	(Υπογραφή)	(Υπογραφή)
.....	.....	.....
Δημήτριος-Διονύσιος Κουτσούρης Καθηγητής Ε.Μ.Π.	Γεώργιος Ματσόπουλος Καθηγητής Ε.Μ.Π.	Παναγιώτης Τσανάκας Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2019

……..
Νικόλαος Βούλγαρης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Η ραγδαία αύξηση της βιοϊατρικής έρευνας τα τελευταία χρόνια έχει οδηγήσει στην προσέγγιση θεμάτων, που μέχρι πρόσφατα ήταν περισσότερο φιλοσοφικά, όπως αυτό της συνείδησης. Τι είναι η συνείδηση, άλλωστε;

Στην παρούσα διπλωματική εργασία λοιπόν, γίνεται προσπάθεια κατανόησης αυτού του θέματος, μέσω της παρουσίασης ορισμένων ερευνών που επιχειρούν να αναλύσουν και να ποσοτικοποιήσουν την έννοια της συνείδησης. Ακολούθως, εξετάζεται το ενδεχόμενο οι μηχανές, και γενικότερα η τεχνητή νοημοσύνη, να αποκτήσουν κάποια στιγμή συνείδηση. Τέλος, παρέχεται και η επίδειξη λειτουργίας της εφαρμογής PyPhi, η οποία «μετράει» τη συνείδηση ενός κυκλώματος, βασισμένη σε μία αντίστοιχη θεωρία.

Λέξεις Κλειδιά: συνείδηση, συνείδηση και τεχνητή νοημοσύνη, θεωρία ενσωματωμένων πληροφοριών, αιτιατή πυκνότητα, αρχή ελεύθερης ενέργειας, PyPhi

Abstract

In the last years, the rapid growth of biomedical research has led to approaching issues, which until recently had only a theoretical base, like the issue of consciousness. So, what is consciousness exactly?

In this diploma thesis takes place a try of understanding this matter, through the presentation of some researches, which attempt to analyze and also quantify consciousness. Furthermore, it is examined whether machines, and artificial intelligence in general, can at some point obtain consciousness. At last, the function of the application PyPhi is presented, which "measures" the consciousness of a circuit, based on a corresponding theory.

Keywords: consciousness, consciousness in artificial intelligence, integrated information theory, causal density, free energy principle, PyPhi

Ευχαριστίες

Καταρχάς θέλω να ευχαριστήσω τον καθηγητή κ. Δημήτριο-Διονύσιο Κουτσούρη που μου έδωσε τη δυνατότητα να εργαστώ πάνω σε ένα θέμα ιδιαίτερο και πολύ ενδιαφέρον και να ολοκληρώσω τη διπλωματική μου εργασία.

Ιδιαίτερες ευχαριστίες θέλω να εκφράσω στη Βασιλεία Κωσταρίδη που με βοήθησε από την πρώτη στιγμή, με καθοδήγησε και αφιέρωσε χρόνο καθ' όλη τη διάρκεια της διπλωματικής εργασίας. Τέλος, ευχαριστώ θερμά την οικογένεια μου και τους φίλους μου για όλη τη στήριξη που είχα κατά τη διάρκεια των σπουδών μου.

Πίνακας Περιεχομένων

Πίνακας Σχημάτων

Πίνακας Πινάκων

1. Εισαγωγή

1.1 Η αφετηρία

Αν ρωτηθεί κάποιος για το πώς πέρασε το χθεσινό βράδυ μια πιθανή απάντηση θα ήταν πως έκατσε να φάει το φαγητό του στο ζεστό του δωμάτιο και εφόσον χόρτασε το άφησε. Η μυρωδιά του φαγητού επικρατούσε στο δωμάτιο, έτσι άνοιξε το παράθυρο. Από τη στιγμή όμως που αποκοιμήθηκε αυτή η μυρωδιά χάθηκε τελείως, ενώ έπρεπε να ξυπνήσει για να καταλάβει ότι κρύωνε και έπρεπε να κλείσει το παράθυρο, κι ας είχε ξυπνήσει από το κρύο. Αυτό πρακτικά καλείται συνείδηση, όλες οι εμπειρίες τις οποίες καθημερινά αποκτά κάποιος, μαζί με τις σκέψεις και τις προθέσεις που τις συνοδεύουν.

Το φαινόμενο αυτό καθίσταται από μόνο του συναρπαστικό κυρίως λόγω της υποκειμενικότητας από την οποία διακρίνεται. Για να καταλάβει κανείς πόσο βαθιά υποκειμενικό είναι αυτό το θέμα αρκεί να σκεφτεί ένα θηλαστικό ζώο, τη νυχτερίδα. Είναι εύκολα αποδεκτό πως, όπως οι σκύλοι και τα δελφίνια, έτσι και οι νυχτερίδες αποκτούν εμπειρίες. Είναι γνωστό επίσης ότι προκειμένου να προσανατολιστούν οι νυχτερίδες χρησιμοποιούν τον ηχοεντοπισμό. Εκπέμπουν, δηλαδή, ήχους σε υψηλή συχνότητα και στη συνέχεια χρησιμοποιούν την ηχώ που παράγεται από την αντανάκλαση των ηχητικών κυμάτων πάνω σε επιφάνειες και άλλα ζώα, ώστε να αποφύγουν εμπόδια και να εντοπίσουν τη λεία τους [1]. Από την πλευρά του ο άνθρωπος αυτό το καταλαβαίνει σαν διαδικασία, σε καμία όμως περίπτωση δεν μπορεί να συλλάβει το πρακτικό αποτέλεσμα της διαδικασίας. Πώς λειτουργεί η νυχτερίδα ξέρει, αλλά όχι πώς είναι να είσαι νυχτερίδα.

1.2 Ορισμοί της συνείδησης

Εντελώς τυπικά συνείδηση είναι η νοητική δυνατότητα ενός οργανισμού η οποία του επιτρέπει, σε προέκταση των αισθήσεών του, να γνωρίζει και να κατανοεί τον εαυτό του, το περιβάλλον του, τα συμβαίνοντα γύρω του και μέσα του και να έχει το δυνατόν την αίσθηση της «θέσης» και της σημασίας του στον κόσμο καθώς και του αντίκτυπου των πράξεων του. Για ένα άτομο χωρίς συνείδηση δεν υπάρχει ο κόσμος γύρω του, δεν υπάρχει καν ο ίδιος, δεν υπάρχει τίποτα [2].

Βλέπουμε λοιπόν πως κατά την επικρατούσα άποψη η συνείδηση θα μπορούσε να χαρακτηριστεί μια γενική κατάσταση. Ωστόσο, με την πάροδο των χρόνων συνδέθηκε με τη λειτουργία ορισμένων περιοχών του εγκεφάλου. Έτσι φιλοσοφικές έννοιες που την απέδιδαν ως μια διεργασία ή επιφαινόμενο ύπαρξης ακολουθήθηκαν από όρους όπως «συνειδητό νοητικό πεδίο». Άλλωστε σε αυτόν τον τομέα θα σταθούμε, στον τομέα της έρευνας που ξεπερνάει τα φιλοσοφικά ζητήματα και ασχολείται με θέματα όπως τα αίτια που συγκεκριμένα μέρη του εγκεφάλου φαίνεται να συνδράμουν περισσότερο στην απόκτηση και καταγραφή συνειδητών εμπειριών όντας το ίδιο ή και λιγότερο πλούσια σε νευρώνες και νευρωνικές

συνδέσεις. Ταυτόχρονα γίνεται πιο δύσκολα αποδεκτό ότι η συνείδηση είναι απλώς μια δυαδική κατάσταση, δηλαδή είτε υπάρχει είτε όχι, αφού για παράδειγμα παρατηρούνται διαφορές μεταξύ των καταστάσεων ύπνου δίχως όνειρα και ύπνου με όνειρα.

Με βάση τα παραπάνω άνοιξε ο δρόμος τον ερευνών σχετικά με την κατανόηση του θέματος. Αυτό οδήγησε στην ανάπτυξη διάφορων θεωριών σχετικά με τη συνείδηση, με την καθεμία, χωρίς υπερβολή, να χρησιμοποιεί και κάτι διαφορετικό στους όρους για να θέσει τη συνείδηση. Φυσικά στόχος όλων είναι να αποδώσουν ένα γενικά αποδεκτό πλαίσιο που να εξηγεί επιστημονικά τη συνείδηση και τα προβλήματα που έχουν αναπτυχθεί γύρω από αυτή.

1.3 Τα προβλήματα της συνείδησης

Γιατί όμως είναι σημαντικό να συνεχιστεί αυτή η προσπάθεια; Τι εμπόδια δημιουργεί το ότι δεν υπάρχουν ουσιαστικές γνώσεις επί του θέματος; Μέχρι πριν κάποια χρόνια αυτά τα ερωτήματα δεν ήταν καν προς συζήτηση, αφού υπήρχε ξεκάθαρη διάκριση μεταξύ των θεμάτων ύλης/σώματος που ήταν υπαρκτά και αντιμετωπίσιμα και των θεμάτων μυαλού που μόνο στη θεωρία μπορούσε κάποιος να αναζητήσει αν και όποιες απαντήσεις ήθελε. Με τον καιρό, ωστόσο, ήρθε η ανάπτυξη της νευροεπιστήμης που βοήθησε για μια πιο πραγματολογική προσέγγιση στο «τι είναι η συνείδηση;» και δημιούργησε τις γέφυρες ώστε αυτή η απολύτως υποκειμενική και φιλοσοφική κατάσταση να μπορέσει να εξηγηθεί αντικειμενικά, ακόμα να γίνει και μετρήσιμη.

Αυτή τη στιγμή, και εδώ και τα τελευταία λίγα χρόνια, στις έρευνες που διεξάγονται συμμετέχουν νευροεπιστήμονες, ψυχίατροι, προγραμματιστές, μαθηματικοί, ακόμα και φιλόσοφοι. Οι βασικοί άξονες των ερευνών είναι δύο, το «εύκολο» και το «δύσκολο» πρόβλημα της συνείδησης. Για αυτή τη διάκριση πρώτος έκανε λόγο ο David Chalmers, με το «εύκολο πρόβλημα» να εκφράζεται ως η κατανόηση του πώς ο εγκέφαλος δίνει τη δυνατότητα της αντίληψης, της αναγνώρισης, της μάθησης και της συμπεριφοράς. Εν συνεχεία έρχεται το «δύσκολο πρόβλημα» που είναι το πώς και γιατί τα παραπάνω χαρακτηριστικά που διαθέτει ο κάθε άνθρωπος σχετίζονται με τη συνείδηση του. Φυσικά χρειάζεται να τονιστεί πως σε καμία περίπτωση δεν υπάρχει η εγγύηση ότι μια υποτιθέμενη λύση του πρώτου θα μας δώσει και ένα υπόβαθρο για τη λύση του δεύτερου, αφού πολύ άνετα μπορεί να μην υπάρχει καμία συσχέτιση μεταξύ τους [2].

Υπάρχουν ερευνητές που έχουν χαρακτηρίσει τη βιολογία πίσω από τη συνείδηση το παραμεθόριο της επιστήμης. Η διαφορετικές αντιδράσεις του εγκεφάλου όταν βρίσκεται σε συνειδητή και μη κατάσταση είναι αυτές που έχουν δώσει ένα επιπλέον πάτημα στις διάφορες θεωρίες να αναπτυχθούν. Προκύπτουν ελπίδες για δυνατότητα μέτρησης της συνείδησης, κάτι πολύ ιδιαίτερο έτσι και επιτευχθεί. Το να είμαστε σε θέση να μπορούμε να ξέρουμε πόση συνείδηση έχει κάποιος όταν κοιμάται, όταν είναι σε κώμα ή όταν έχει λάβει δόση αναισθητικού μπορεί να αποδειχθεί από μικρός έως κρίσιμος παράγοντας σχετικά με το ζήτημα που χρήζει αντιμετώπισης. Όλα αυτά τα ερεθίσματα έχουν οδηγήσει σε σημαντική αύξηση του αριθμού των ερευνών. Χαρακτηριστικό είναι πως μέχρι και πριν λίγα χρόνια το δύσκολο πρόβλημα του Chalmers θεωρείτο πως ήταν άλυτο. Από την άλλη πλευρά βέβαια η

επιστήμη δε βρίσκεται σε θέση να ισχυριστεί πως οδεύει προς τη λύση του, απλά ότι γίνονται προσπάθειες που μπορεί να οδηγήσουν σε αυτή.

1.4 Επικρατούσες θεωρίες

Ένας από τους στόχους αυτής της εργασίας είναι να αναλυθούν και να κατανοηθούν κάποιες από αυτές τις προσπάθειες, ώστε να υπάρχει μια σαφή άποψη για το σημείο στο οποίο βρισκόμαστε. Φυσικά δεν είμαστε σε θέση να τις κρίνουμε, αλλά να παρουσιάσουμε όλα τα γνωστά στοιχεία γύρω από αυτές. Σε αυτό το σημείο θα αναφερθούν εν συντομία, χωρίς να αναλυθούν τα βασικά τους χαρακτηριστικά τα οποία ακολουθούν στη συνέχεια.

Μία λοιπόν από τις θεωρίες-ερευνητικές προσπάθειες που με την πάροδο του χρόνου έφτασε να έχει κερδίσει ένα μεγάλο κομμάτι του επιστημονικού κόσμου σχετικά με την εγκυρότητα της είναι η Θεωρία των Ενσωματωμένων Πληροφοριών, ή κανονικά Integrated Information Theory (IIT). Η IIT επιχειρεί να εξηγήσει τι είναι συνείδηση και γιατί αυτή φέρεται να συνδέεται με ορισμένα φυσιολογικά συστήματα του οργανισμού. Δοθέντων των συστημάτων, σύμφωνα με τη θεωρία δύναται να προβλεφθεί αν αυτά μπορούν να διαμορφώσουν συνείδηση, σε ποιο βαθμό είναι συνειδητά και ποιες συγκεκριμένες εμπειρίες λαμβάνουν. Η αρχική πρόταση της θεωρίας έγινε το 2004 από τον νευροεπιστήμονα και ψυχίατρο Giulio Tononi το 2004 και μέχρι στιγμής συνεχώς βρίσκεται υπό ανάπτυξη. Η τελευταία ολοκληρωμένη της μορφή δημοσιεύτηκε το 2014, με τίτλο IIT 3.0. Καλούμενη να αντιμετωπίσει το «δύσκολο πρόβλημα» του Chalmers, η IIT δεν προσπαθεί μέσω των διάφορων χαρακτηριστικών της λειτουργίας του εγκεφάλου να φτάσει στη συνείδηση, αλλά ξεκινά από τη συνείδηση, δεχόμενη την ύπαρξη της ως βέβαιη, και αιτιολογεί τις ιδιότητες που θα έπρεπε να έχει ένα υποτιθέμενο φυσιολογικό υποσύστημα, ώστε αυτό να μπορεί να θεωρηθεί μέρος της συνείδησης. Η δυνατότητα να πραγματοποιηθεί αυτό το άλμα από τη φαινομενολογία σε μηχανιστικά πρότυπα μας δίνεται από μία υπόθεση της IIT, ότι εφόσον μια συνειδητή εμπειρία μπορεί πλήρως να αποκωδικοποιηθεί από ένα βαθύτερο φυσιολογικό σύστημα του εγκεφάλου, τότε υπάρχουν υποσυστήματα αυτού που διαχειρίζονται τις επί μέρους ιδιότητες της εμπειρίας.

Έπειτα είναι απαραίτητο να αναφερθεί και να μελετηθεί η δουλειά που έχει γίνει από τον νευροεπιστήμονα και καθηγητή του πανεπιστημίου του Sussex Anil Seth. Είναι ίσως μαζί με την IIT η θεωρία που της αποδίδονται το μεγαλύτερο μέρος υποστηρικτών του επιστημονικού κόσμου. Δεν έχουν λείψει μάλιστα και ορισμένα debates μεταξύ των δύο βασικών εκπροσώπων τους. Χωρίς λοιπόν να έχει επικρατήσει κάποιο όνομα ως αντιπροσωπευτικό, η συγκεκριμένη θεωρία επιχειρεί να κατανοήσει τη βιολογική βάση της συνείδησης με τη βοήθεια των επιστημών της νευρολογίας, των μαθηματικών, της πληροφορικής, της ψυχολογίας, της φιλοσοφίας και της ψυχιατρικής. Στόχος της είναι η πρακτική μετάφραση, μετά από ουσιαστική και εκ βαθέως αποκωδικοποίηση, των περίπλοκων κυκλωμάτων του εγκεφάλου που θεμελιώνουν τη συνείδηση, για την αξιοποίηση της σε κλινικές περιπτώσεις ψυχιατρικών και νευρολογικών ανωμαλιών/ασθενειών. Με συμμετοχή σε πάνω από 130 δημοσιεύσεις από το 2004 και μετά το έργο του Anil Seth έχει αναγνωριστεί και εκτιμηθεί

παγκοσμίως, όπως και η συνεισφορά του στην έρευνα πάνω στο θέμα, οπότε δεν γίνεται να παραλειφθεί.

Καλύπτοντας αυτά τα δύο πεδία το επόμενο είναι η ανάλυση μίας επιπλέον «δυνατής», στο επιστημονικό περιβάλλον, θεωρίας. Ούτε αυτή έχει μπει υπό την ομπρέλα κάποιου τίτλου, ωστόσο η βάση της είναι η Αρχή Ελεύθερης Ενέργειας (Free Energy Principle) και αυτός που κατά κόρον τη διαμόρφωσε είναι ο Βρετανός νευροεπιστήμονας Karl Friston. Πολύ σύντομα, η θεμελιώδης προσπάθεια της Αρχής Ελεύθερης Ενέργειας γίνεται για να εξηγήσει πώς και γιατί τα βιολογικά συστήματα διατηρούν τη σειρά δράσης τους μέσω ενός πεπερασμένου αριθμού μη ισορροπημένων καταστάσεων. Σύμφωνα με αυτή τα βιολογικά συστήματα ελαχιστοποιούν τη λειτουργικότητα ελεύθερης ενέργειας των εσωτερικών τους καταστάσεων, η οποία συνεπάγεται πεποιθήσεις για απόκρυφες καταστάσεις στο περιβάλλον τους. Η υπονοούμενη ελαχιστοποίηση της μεταβλητής ελεύθερης ενέργειας συσχετισμένη με τις Μπεϋζιανές μεθόδους αποτελεί την εξήγηση της ενσωματωμένης αντίληψης για τη νευροεπιστήμη. Πρόκειται, δηλαδή, για ένα απλό αξίωμα που οδηγεί σε περίπλοκα συμπεράσματα, ξεκινώντας από την τάση οποιουδήποτε βιολογικού συστήματος να αντιστέκεται σε οποιασδήποτε μορφής διαταραχή.

Από εκεί και πέρα η συνεχής εργασία και έρευνα επί του θέματος της συνείδησης έχει οδηγήσει σε αρκετές επιπλέον προσεγγίσεις, οι οποίες είτε στέκονται μόνες τους, απλά αυτή τη χρονική στιγμή είναι λιγότερο αποδεκτές ή γνωστές, είτε έχουν μερικώς αξιοποιηθεί για την ανάπτυξη των παραπάνω ολοκληρωμένων θεωριών. Στο παρόν κείμενο θα γίνει μια σαφή παρουσίαση ορισμένων από αυτές, χωρίς ιδιαίτερη ανάλυση, οι οποίες ονομαστικά είναι η θεωρία του Καθολικού Χώρου Εργασίας του Bernard Baars, η θεωρία του Νευρωνικού Καθολικού Χώρου Εργασίας του Stanislas Dehaene, η θεωρητική προσέγγιση της Εστιακής Ανατροφοδότησης του Victor Lamme και η θεωρία των Μικροσυνειδήσεων του Semir Zeki.

1.5 Συνείδηση και Τεχνητή Νοημοσύνη

Ακολουθεί έτσι το κομμάτι που πρακτικά και μόνο αποτελεί μια πραγματική πρόκληση. Γίνεται η τεχνητή νοημοσύνη, οι μηχανές, οτιδήποτε αποτελεί μέρος της τεχνολογίας να αποκτήσει συνείδηση; Όχι αν είναι θεμιτό, αν είναι ιδεολογικά σωστό, αλλά αν γίνεται. Μέχρι πριν λίγα χρόνια η απάντηση ήταν όχι και τέλος. Πλέον όμως η τεχνολογία βρίσκεται στο σημείο που και αυτό το όχι χρειάζεται πολλά και εκτενή επιχειρήματα για να στηριχθεί. Για την εξέλιξη της τεχνητής νοημοσύνης δε χρειάζεται να γίνει λόγος καθώς όλοι τη γνωρίζουν. Δε λείπουν, μάλιστα, οι υπερβολές από μελλοντολόγους που κάνουν λόγο για το ότι ο άνθρωπος θα πρέπει να φοβάται αυτή την ανάπτυξη πολύ. Τομείς όπως η μηχανική μάθηση και η αναγνώριση φωνής έφεραν ήδη στη ζωή μας την Alexa της Amazon, την Siri της Apple, το Google Now και την Cortana της Microsoft, όλα τους εικονικά βοηθήματα, τίποτα παραπάνω, δίχως να ισχυρίζεται κανείς πως έχουν συνείδηση, δίχως όμως και να αρνείται πως σε λίγα χρόνια θα είναι δύσκολο να διακρίνεις τη συμπεριφορά τους από αυτή κανονικών ανθρώπων, εκτός βέβαια από το γεγονός ότι δε θα υπάρχει κάποιο αρνητικό χαρακτηριστικό σε αυτή [3].

Υπάρχει, λοιπόν, κάποιας μορφής κέρδος από την είσοδο της συνείδησης στα πλαίσια της τεχνητής νοημοσύνης; Με μια πολύ γρήγορη ματιά γίνεται κατανοητό πως πολλοί είναι αυτοί που θεωρούν πως όχι, κυρίως λόγω των παραγόντων που αναφέραμε προηγουμένως. Αυτό βέβαια δεν έχει εμποδίσει ένα ξεχωριστό πεδίο, αυτό της τεχνητής ή μηχανικής συνείδησης, να αναπτύσσεται μεμονωμένα τα τελευταία χρόνια. Αυτή η δουλειά προκύπτει μέσα από τις διάφορες μεθόδους που χρησιμοποιούνται μέχρι τώρα για τη μέτρηση της ανθρώπινης συνείδησης και επικεντρώνεται στην ανάπτυξη μοντέλων συνείδησης σε λογισμικά υπολογιστών και ρομποτικών συστημάτων. Στόχος είναι η παροχή βοήθειας στην ακόμα βαθύτερη κατανόηση της ανθρώπινης συνείδησης και της γνωστικής διαδικασίας και δυνατότητας του ανθρώπου, αλλά και η δημιουργία των προϋποθέσεων ώστε μελλοντικά να φτιαχτεί, γιατί όχι, και μια, έστω φαινομενικά, μηχανή με συνείδηση.

1.6 Η εφαρμογή

Για την ολοκλήρωση αυτής της εργασίας στο τέλος θα παρατεθούν τα αποτελέσματα που προέκυψαν μετά από προσομοιώσεις της εφαρμογής PyPhi. Η συγκεκριμένη εφαρμογή έχει αναπτυχθεί από μία ομάδα του νευροεπιστήμονα Giulio Tononi και, προφανώς βασισμένη πάνω στη Θεωρία των Ενσωματωμένων Πληροφοριών, αποδίδει ένα μέτρο στη συνείδηση που έχουν τα κυκλώματα που λαμβάνει ως είσοδο. Φυσικά βρίσκεται υπό διαρκή βελτίωση, αλλά είναι ένα χρήσιμο εργαλείο για την καλύτερη κατανόηση των αξιών και των εννοιών που χρησιμοποιεί η αντίστοιχη θεωρία.

2. Θεωρία Ενσωματωμένων Πληροφοριών (IIT)

2.1 Εισαγωγή

Καταρχάς θα γίνει προσπάθεια για την εξέταση της δουλειάς του Giulio Tononi και της Θεωρίας Ενσωματωμένων Πληροφοριών [4]. Μπορεί, λοιπόν, η συνείδηση να θεωρηθεί ως αυτό που ενεργοποιείται μέσα στον άνθρωπο με το που ξυπνάει και χάνεται όταν τον παίρνει ο ύπνος. Κάποιες φορές ίσως χαθεί και όταν κάποιος δεχθεί ένα δυνατό χτύπημα στο κεφάλι. Έτσι, η καθημερινή εμπειρία υποδεικνύει ότι η συνείδηση έχει ένα φυσιολογικό υπόστρωμα και ότι αυτό πρέπει να δουλεύει με έναν συγκεκριμένο τρόπο, ώστε να είναι κάποιος πλήρως ενσυνείδητος. Προκύπτει κατά συνέπεια το ερώτημα, ποιες είναι οι συνθήκες που καθορίζουν τον βαθμό στον οποίο η συνείδηση είναι παρούσα ανά πάσα στιγμή ή κατάσταση; Για παράδειγμα, είναι τα μωρά ή τα ζώα ενσυνείδητα και σε τι βαθμό; Και αν ναι, είναι ο βαθμός αυτός συγκρίσιμος με τους υπόλοιπους; Πόσο ενσυνείδητος είναι κάποιος που υπνοβατεί ή που περνάει μια ψυχοκινητική κρίση; Για να υπάρξει η δυνατότητα απάντησης σε τέτοιου είδους ερωτήματα πρέπει να επιτευχθεί η κατανόηση της συνείδησης τόσο μέσω πρακτικών μελετών όσο και θεωρητικής ανάλυσης.

Δεν είναι λίγοι βέβαια αυτοί που θεωρούν πως κάτι τέτοιο είναι εκτός των δυνατοτήτων του ανθρώπου. Γίνεται συχνά λόγος πως το καλύτερο που μπορεί να γίνει είναι να μαζεύονται συνεχώς περισσότερα δεδομένα για τις νευρικές συσχετίσεις της συνείδησης -δηλαδή τις πτυχές της εγκεφαλικής λειτουργίας που αλλάζουν όταν πτυχές της συνείδησης αλλάζουν- με την ελπίδα απλώς ότι κάποια μέρα θα καταλήξει κάποιος σε ένα συμπέρασμα. Μάλιστα υπάρχει και η άποψη ότι ακόμα και αν γίνουν γνωστά τα πάντα για τις νευρικές συσχετίσεις της συνείδησης, δε θα οδηγήσουν σε απαντήσεις στο γιατί ορισμένες φυσιολογικές διεργασίες παράγουν εμπειρίες ενώ άλλες όχι [5].

Φυσικά υπάρχουν αρκετές από αυτές τις συσχετίσεις οι οποίες είναι γνωστές. Για παράδειγμα, έχει αναγνωριστεί ότι η ευρεία καταστροφή του εγκεφαλικού φλοιού αφήνει τον άνθρωπο μόνιμα σε φυτική κατάσταση, ενώ η ολική αφαίρεση της παρεγκεφαλίτιδας, που είναι ιδιαίτερα πιο πλούσια σε νευρικά κύτταρα, δεν επηρεάζει σχεδόν καθόλου τη συνείδηση. Επίσης είναι γνωστό ότι οι νευρώνες στον εγκεφαλικό φλοιό παραμένουν ενεργοί κατά τη διάρκεια του ύπνου, είτε αυτός που κοιμάται ονειρεύεται είτε όχι. Τέλος τα διαφορετικά μέρη του φλοιού επηρεάζουν διαφορετικές ποιοτικά πτυχές της συνείδησης: κάποια ζημιά σε συγκεκριμένα μέρη του φλοιού μπορεί να καταστρέψει τη δυνατότητα αναγνώρισης χρωμάτων, ενώ μια άλλη κάκωση μπορεί να επιδράσει πάνω στην αντίληψη των σχημάτων. Μάλιστα, η συνεχής βελτίωση των εξειδικευμένων νευροεπιστημονικών εργαλείων οδηγεί και σε ακριβέστερες απόψεις για τις νευρικές συσχετίσεις της συνείδησης. Παρά όλα αυτά, όταν πρέπει να εξηγήσουμε γιατί στην απόκτηση εμπειριών κυριαρχεί ο φλοιός και όχι η παρεγκεφαλίτιδα ή γιατί κάποιες περιοχές του φλοιού αξιοποιούνται για τις εμπειρίες των χρωμάτων και άλλες για τις εμπειρίες των σχημάτων, δεν υπάρχει κάποια απάντηση.

Παρατίθεται, λοιπόν, μια προσέγγιση της Θεωρία Ενσωματωμένων Πληροφοριών, σε μια προσπάθεια κατανόησης της συνείδησης σε βάθος.

2.2 Η συνείδηση ως ενσωματωμένη πληροφορία

Η θεωρία ενσωματωμένων πληροφοριών (IIT) ισχυρίζεται ότι, σε θεμελιώδες επίπεδο, η συνείδηση είναι ενσωματωμένη πληροφορία, και ότι η ποιότητα της προκύπτει από τις ενημερωτικές σχέσεις ενός συγκροτήματος στοιχείων. Αυτοί οι ισχυρισμοί προέρχονται από την αντίληψη ότι πληροφορία και ενσωμάτωση αποτελούν βασικές ιδιότητες των εμπειριών μας. Αυτό ίσως να μη μας γίνεται άμεσα αποδεκτό, μάλλον διότι ο άνθρωπος είναι προικισμένος με συνείδηση «από πάντα», είναι κάτι δεδομένο για αυτόν. Για να αποκτηθεί μια οπτική, χρήσιμο είναι να σκεφτεί κανείς δύο συγκεκριμένα παραδείγματα.

Έστω, λοιπόν, το εξής: Βρισκόμαστε μπροστά σε μια κενή οθόνη, η οποία μπορεί να είναι είτε ανοιχτή είτε κλειστή, και μας έχει δοθεί η οδηγία να φωνάζουμε «Φως» όταν η οθόνη ανοίγει και «Σκοτάδι» όταν αυτή κλείνει. Μπροστά από την οθόνη έχει επίσης τοποθετηθεί μια φωτοδίοδος. Αποτελείται από έναν αισθητήρα που αντιδρά με το φως και την ένταση του και έναν ανιχνευτή συνδεδεμένο με τον αισθητήρα, ώστε να παίρνουμε το μήνυμα «Φως» όταν η φωτεινότητα ξεπερνάει ένα κατώφλι, αλλιώς να παίρνουμε το μήνυμα «Σκοτάδι». Προκύπτει λοιπόν το πρώτο ερώτημα σχετικά με τη συνείδηση. Εμείς, μέσω της υποκειμενικής μας εμπειρίας, μπορούμε και ξεχωρίζουμε το φως από το σκοτάδι. Και η φωτοδίοδος μπορεί να ξεχωρίσει το φως από το σκοτάδι, αλλά προφανώς δεν έχει κάποια υποκειμενική εμπειρία πάνω στις δύο καταστάσεις. Ποια είναι, έτσι, η βασική διαφορά της φωτοδιόδου με εμάς;

Σύμφωνα με την IIT η διαφορά έχει να κάνει με το πόση πληροφορία παράγεται όταν γίνεται η εκάστοτε διάκριση φωτός με σκοτάδι. Ένας από τους κλασικούς ορισμούς της πληροφορίας λέει πως πληροφορία είναι αυτό που προκύπτει από τη μείωση της αβεβαιότητας, δηλαδή ο μεγαλύτερος αριθμός εναλλακτικών που αποκλείονται φέρνει μεγαλύτερη μείωση της αβεβαιότητας, άρα και πιο ισχυρή πληροφορία. Συνήθως μετριέται μέσω της συνάρτησης εντροπίας, που είναι ο λογάριθμος με βάση το 2 του αριθμού των εναλλακτικών. Για παράδειγμα, το να ρίξουμε ένα νόμισμα αντιστοιχεί σε $\log_2(2) = 1$ bit πληροφορίας, ενώ το να ρίξουμε ένα ζάρι σε $\log_2(6) = 2.59$ bits πληροφορίας.

Καθίσταται πλέον αναγκαία η σύγκριση μας με τη φωτοδίοδο. Όταν η οθόνη είναι ανοιχτή ο μηχανισμός της φωτοδιόδου μας δίνει την αναφορά «Φως». Με τον τρόπο αυτό η φωτοδίοδος παράγει $\log_2(2) = 1$ bit πληροφορίας. Από την άλλη πλευρά, όταν εμείς βλέπουμε την οθόνη να είναι ανοιχτή η κατάσταση είναι αρκετά διαφορετική. Αυτό, γιατί εμείς στην ουσία δεν ξεχωρίζουμε απλώς το φως από το σκοτάδι, αλλά από έναν πολύ μεγαλύτερο αριθμό εναλλακτικών, παράγοντας ταυτόχρονα και μεγαλύτερο αριθμό bits πληροφορίας [5].

Για να γίνει αυτό κατανοητό αρκεί να σκεφτεί κανείς την οθόνη να γίνεται κόκκινη, μετά πράσινη, μετά μπλε και μετά να περνάει ένα προς ένα όλα τα καρέ κάθε ταινίας που έχει γυριστεί ποτέ. Για όλο αυτό το σύνολο εικόνων το μόνο που απασχολεί τη φωτοδίοδο είναι το ζήτημα της απόφασης του αν κάθε φορά ξεπερνιέται το κατώφλι φωτεινότητας. Για εμάς, ωστόσο, μια φωτεινή οθόνη δεν είναι διαφορετική μόνο από μια σκοτεινή οθόνη, αλλά και από έναν μεγάλο αριθμό άλλων εικόνων και έτσι όταν λέμε «Φως» στην ουσία εννοούμε το συγκεκριμένο φως έναντι αμέτρητων άλλων εναλλακτικών, η κόκκινη οθόνη, η μπλε οθόνη, κάποιο καρέ ταινίας κλπ. Επομένως εμείς δημιουργούμε έναν πολύ μεγάλο αριθμό bits πληροφορίας.

Η πληροφορία, η δυνατότητα επιλογής μέσα από μεγάλο αριθμό εναλλακτικών, μπορεί επομένως να αποτελεί ένα κύριο κομμάτι της συνείδησης. Ωστόσο, η πληροφορία υπονοεί και την ύπαρξη μιας οπτικής γωνίας και επιβάλλεται προσοχή για το ποια είναι αυτή. Για να γίνει αντιληπτό γιατί, αρκεί να σκεφτεί κάποιος το παράδειγμα μιας ψηφιακής κάμερας με ένα τσιπάκι αισθητήρα, αποτελούμενο από ένα εκατομμύριο δυαδικές φωτοδιόδους όμοιες με αυτή του παραπάνω παραδείγματος. Προφανώς ως σύνολο η κάμερα μπορεί να διαχωρίσει μεταξύ $2^{1000000}$ πιθανών καταστάσεων, κάτι που αντιστοιχεί σε 1.000.000 bits πληροφορίας. Προφανώς και πάλι δεν μπορεί να θεωρηθεί ότι η κάμερα έχει συνείδηση. Ποια είναι, λοιπόν, η διαφορά μας από αυτή;

Σύμφωνα με την ΙΙΤ η διαφορά μας έχει να κάνει με την ενσωματωμένη πληροφορία. Από την οπτική ενός εξωτερικού παρατηρητή, η κάμερα μπορεί να θεωρηθεί ως ένα σύστημα με εύρος 1.000.000 καταστάσεις. Στην ουσία όμως η κάμερα δεν αποτελεί μια ολοκληρωμένη οντότητα, από τη στιγμή που οι φωτοδίοδοι δεν αλληλοεπιδρούν μεταξύ τους, παρά η καθεμία εκτελεί τη δική της διάκριση με βάση κάποιο όριο φωτεινότητας ανεξάρτητα με το τι μπορεί να κάνουν όλες οι υπόλοιπες. Με άλλα λόγια, δεν υπάρχει κάποια εσωτερική οπτική που να μπορεί να θεωρήσει την κάμερα ως ένα αναπόσπαστο σύνολο, καθώς αν την αποσυνθέσουμε κανονικά η κάθε φωτοδίοδος θα συνεχίσει να κάνει την ίδια δουλειά με πριν.

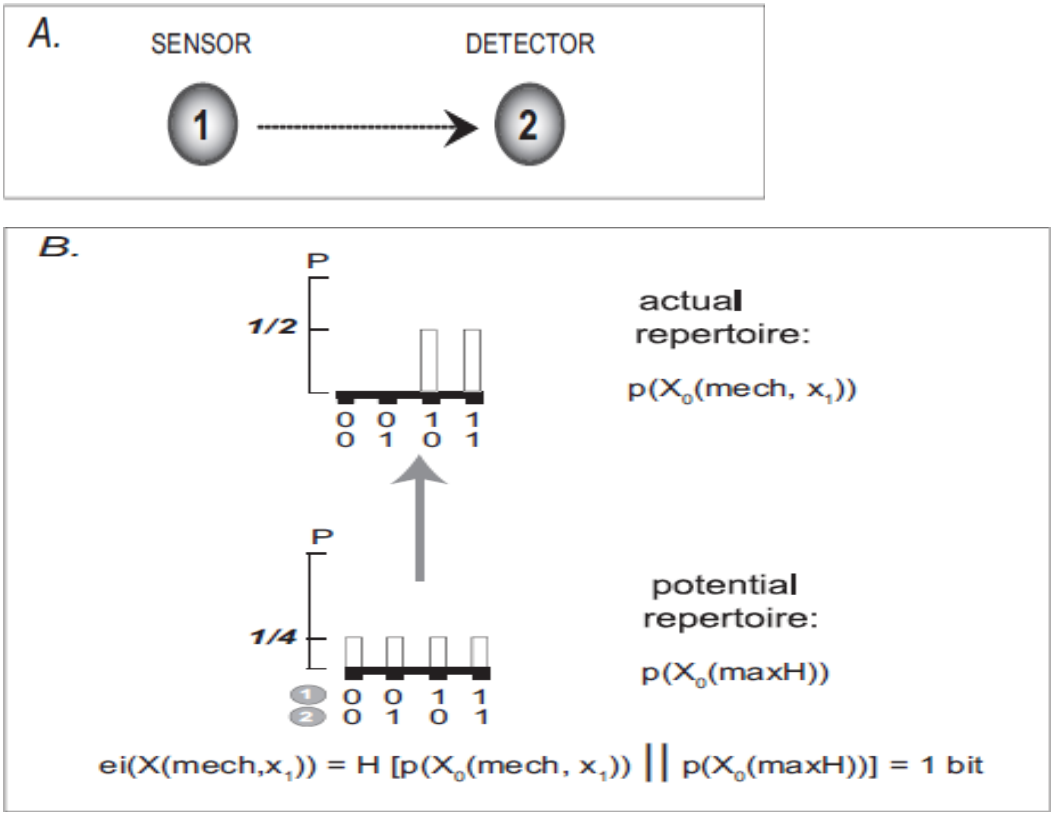
Αντίθετα ο άνθρωπος μπορεί να ξεχωρίσει μεταξύ ενός τεράστιου συνόλου εναλλακτικών καταστάσεων ως ένα ολοκληρωμένο σύστημα, το οποίο δεν μπορεί να διασπαστεί και αποτελείται από πολλά στοιχεία με διαφορετικές λειτουργίες. Φαινομενικά κάθε εμπειρία είναι ένα ολοκληρωμένο σύνολο, το οποίο έχει σημασία επειδή ακριβώς δεν μπορεί να διασπαστεί και γίνεται αντιληπτό από μία οπτική γωνία. Για παράδειγμα, η εμπειρία ενός κόκκινου τετραγώνου δεν μπορεί να αναλυθεί στην ξεχωριστή εμπειρία του χρώματος κόκκινου και σε αυτή του σχήματος τετραγώνου. Αντίστοιχα η εμπειρία του οπτικού πεδίου ενός ατόμου κάθε στιγμή δεν μπορεί να δεχθεί κάποια ανάλυση, όπως αυτά που βλέπει το δεξιό μάτι και αυτά που βλέπει το αριστερό, κάτι που δεν βγάζει καν νόημα για κάποιον, αφού πάντα η εμπειρία του είναι κάτι ολοκληρωμένο. Ο μόνος τρόπος που φαίνεται να μπορεί αυτό να αλλάξει είναι αν κάπως χωριστεί ο εγκέφαλος στα δύο, όπως με ασθενείς που δέχθηκαν τομή στο μεσολόβιο ώστε να αντιμετωπίσουν δριμύτατη επιληψία. Όντως αυτοί οι άνθρωποι αντιλαμβάνονται ξεχωριστά το δεξιό και το αριστερό οπτικό τους πεδίο, ουσιαστικά όμως δέχονται και δύο εμπειρίες ταυτόχρονα αντί για μία. Μηχανιστικά λοιπόν, αυτή η υποκείμενη ενότητα της εμπειρίας πρέπει να προκύπτει από αιτιατές αλληλεπιδράσεις μεταξύ συγκεκριμένων στοιχείων του εγκεφάλου. Αυτό σημαίνει πως αυτά τα στοιχεία που δουλεύουν ως ένα σύστημα, αν μπορούσαν να διαχωριστούν δε θα ήταν σε θέση να πραγματοποιήσουν τις ίδιες εργασίες.

Τέλος, είναι σημαντικό να αναφέρουμε ορισμένα χαρακτηριστικά της πληροφορίας-εμπειρίας στο χωροχρονικό επίπεδο. Για παράδειγμα η πληροφορία ρέει στο χρόνο με μια πεπερασμένη ταχύτητα. Έρευνες σχετικά με την αντίληψη μιας εμπειρίας κάνουν λόγο για τη διαδικασία της μικρογένεσης, κατά την οποία χρειάζονται 100-200 milliseconds για να αναπτυχθεί μέσα μας μια ολοκληρωμένη και αποτελούμενη από όλες τις αισθήσεις εμπειρία, ενώ χρειάζεται ακόμα περισσότερος χρόνος για να εκφραστεί πρακτικά μια συνειδητή σκέψη. Μάλιστα, επικρατεί η άποψη πως ακολουθείται μια συγκεκριμένη σειρά για την απόλυτη αντίληψη μιας εμπειρίας, αφού πρώτα γίνεται αντιληπτό πως κάτι στο περιβάλλον έχει αλλάξει, στη συνέχεια ποια ή ποιες

αισθήσεις αφορά (όραση, ακοή, όσφρηση κλπ.) και στη συνέχεια γίνεται κατανοητό τι ακριβώς κινείται ή ποιος μιλάει ή τι μυρίζει. Επιπλέον έρευνες υποστηρίζουν πως μία συνειδητή στιγμή δεν υπερβαίνει τα 2-3 δευτερόλεπτα. Βέβαια είναι ακόμα υπό εξέταση το αν μια συνειδητή εμπειρία μπορεί να παρομοιαστεί με ένα σύνολο συνεχόμενων στιγμιότυπων ή έχει μια συνεχόμενη ροή. Σε μια τελική φαινομενολογική ανάλυση προκύπτει ότι η συνείδηση προέρχεται από την ικανότητα να ενσωματώνουμε μεγάλο αριθμό πληροφοριών, η οποία ενσωμάτωση έχει χαρακτηριστική χωροχρονική κλίμακα.

2.3 Ποσοτικοποίηση ενσωματωμένης πληροφορίας

Πώς, λοιπόν, μπορεί να μετρηθεί η ενσωματωμένη πληροφορία; Πρώτα πρέπει να εκτιμηθεί πόση πληροφορία παράγεται από το σύστημα.



Σχήμα 1: (A): Μια «φωτοδίοδος» αποτελούμενη από αισθητήρα και ανιχνευτή. Ο μηχανισμός της φωτοδιόδου είναι τέτοιος ώστε ο ανιχνευτής να ανάβει όταν η τιμή του αισθητήρα ξεπερνάει ένα κατώφλι. (B): Όλο το σύστημα (αισθητήρας, ανιχνευτής) έχει τέσσερις πιθανές καταστάσεις (00, 01, 10, 11). Η διανομή πιθανότητας p(X_0(maxH)) = (1/4, 1/4, 1/4, 1/4) είναι η μέγιστη διανομή εντροπίας στις 4 καταστάσεις. Με δεδομένα το μηχανισμό της φωτοδιόδου και ότι ο ανιχνευτής είναι ανοιχτός, τότε κι ο αισθητήρας πρέπει να ήταν ανοιχτός. Έτσι από τις 4 πιθανές καταστάσεις απορρίπτονται οι 2 (00, 01), ενώ οι άλλες 2 (10,11) είναι ισοπίθανες και δυσδιάκριτες για το μηχανισμό. Έτσι η διανομή πιθανότητας γίνεται p(X_0(maxH)) = (0, 0, 1/2, 1/2). Η σχετική εντροπία (απόκλιση Kullback-Leibler) μεταξύ δύο κατανομών πιθανότητας p και q είναι H[p|q] = $p_i \log_2(p_i/q_i)$, οπότε η ενεργός πληροφορία ei(X(mech, x1)) με έξοδο x1=11 είναι 1 bit. (Ενεργός πληροφορία είναι η εντροπία της τελικής προς την αρχική κατανομή πιθανότητας) [5].

Ας θεωρήσουμε το σύστημα δύο δυαδικών μονάδων του **Σχήματος 1**, που ιδανικά αποτελεί μια φωτοδίοδο με έναν αισθητήρα S και έναν ανιχνευτή D. Το σύστημα χαρακτηρίζεται από την κατάσταση στην οποία βρίσκεται, που στην προκειμένη περίπτωση έστω ότι είναι 11 (με το πρώτο στοιχείο να αντιστοιχεί στον αισθητήρα και το δεύτερο στον ανιχνευτή), και από έναν μηχανισμό. Ο μηχανισμός προκύπτει από τη σύνδεση αισθητήρα-ανιχνευτή (βέλος) και δηλώνει μια αιτιατή αλληλεπίδραση, που στη συγκεκριμένη περίπτωση είναι ότι ο ανιχνευτής ελέγχει την κατάσταση του αισθητήρα και αν αυτός είναι ανοιχτός ανοίγει και ο ίδιος ενώ αν είναι κλειστός κλείνει και ο ίδιος [5].

Ενδεχομένως, το σύστημα σε μια τυχαία χρονική στιγμή θα μπορούσε να είναι σε μία από τέσσερις καταστάσεις (00, 01, 10, 11) με την ίδια πιθανότητα: p = (1/4, 1/4, 1/4, 1/4). Τυπικά, αυτό το εύρος των *a priori* πιθανοτήτων αντιπροσωπεύεται από τη μέγιστη εντροπία ή ομοιόμορφη κατανομή των πιθανών καταστάσεων του συστήματος σε χρόνο t=0, κάτι που εκφράζει απόλυτη αβεβαιότητα (p(X_0(maxH))). Θεωρώντας το παραπάνω εύρος ως ένα σετ όλων των πιθανών καταστάσεων εισόδου, ο συγκεκριμένος μηχανισμός X(mech) του συστήματος μπορεί να θεωρηθεί ότι προσδιορίζει ένα προηγμένο εύρος- την κατανομή πιθανότητας των καταστάσεων εξόδου που παράγεται όταν στο σύστημα δοθούν όλες οι πιθανές είσοδοι. Στην περίπτωση μας βέβαια το σύστημα έχει ήδη κατάσταση εξόδου (για t=1 έχουμε x_1=11). Προκύπτει, έτσι, ότι στην είσοδο x_0 είχαμε κατάσταση 10 ή 11 και η κατανομή των πιθανοτήτων γίνεται p = (0, 0, 1/2, 1/2). Η έξοδος και ο μηχανισμός δηλαδή μας δίνουν ένα εύρος *a posteriori* πιθανοτήτων για τις καταστάσεις του συστήματος (p(X_0(mech, x_1))) όταν t=0. Με τον τρόπο αυτό έχουμε σαφή μείωση της αβεβαιότητας/άγνοιας. Πιο συγκεκριμένα, ο μηχανισμός και η κατάσταση του συστήματος παράγουν 1 bit πληροφορίας έχοντας διακρίνει όσο είναι δυνατό τις πιθανές καταστάσεις.

Γενικά, η πληροφορία που παράγεται όταν ένα σύστημα χαρακτηρίζεται από έναν συγκεκριμένο μηχανισμό σε μια συγκεκριμένη κατάσταση μπορεί να μετρηθεί από τη *σχετική εντροπία* μεταξύ του πραγματικού (a posteriori) και του πιθανού (a priori) ρεπερτορίου πιθανών καταστάσεων, δίνοντας ως αποτέλεσμα τον όρο *ενεργός πληροφορία (ei):*

$$ei(X(mech, x_1)) = H[\ p(X_0(mech, x_1)) \ ||\ p(X_0(mech, x_1))\ ]$$

Η σχετική εντροπία, γνωστή και ως απόκλιση Kullback-Leibler, είναι η διαφορά μεταξύ κατανομών πιθανότητας: αν οι κατανομές είναι οι ίδιες, η σχετική εντροπία είναι μηδέν· όσο πιο διαφορετικές είναι τόσο μεγαλύτερη είναι η τιμή της (πάντα θετική). Ουσιαστικά παράγεται όλο και περισσότερη πληροφορία, όσο μια ομοιόμορφη κατανομή γίνεται συνεχώς πιο ανομοιόμορφη, μειώνοντας έτσι την αβεβαιότητα.

Από τη στιγμή που η ενεργός πληροφορία καθορίζεται αφού καθοριστεί ο μηχανισμός και η κατάσταση, μπορεί να θεωρηθεί ότι αποτελεί μια «εσωτερική» ιδιότητα του συστήματος. Για να την υπολογίσει κάποιος εξωτερικά, πρέπει να υποβάλλει το σύστημα σε όλα τα πιθανά ερεθίσματα για να αποκτήσει όλο το προηγμένο εύρος των εξόδων.

Στη συνέχεια πρέπει να εξεταστεί πόση από την πληροφορία που παράγεται από το σύστημα είναι ολοκληρωμένη, δηλαδή παράγεται από ένα σύστημα-οντότητα και όχι συλλεκτικά από μικρότερα τμήματα. Η ιδέα εδώ είναι να ασχοληθεί κανείς με τα μέρη του συστήματος ανεξάρτητα, να δει πόση πληροφορία παράγουν από μόνα τους και να τη συγκρίνει με τη συνολική πληροφορία που παράγεται από το σύστημα.

Αυτό μπορεί να επιτευχθεί προσφεύγοντας ξανά στη σχετική εντροπία για να μετρηθεί η διαφορά μεταξύ της κατανομής πιθανότητας που προκύπτει από όλο το σύστημα (p(X_0(mech,x_1))) με την κατανομή πιθανότητας που παράγεται ξεχωριστά από το κάθε μέρος (Πp($^kM_0$(mech,μ_1))), που είναι το πραγματικό εύρος των μερών ^kM. Η ενσωματωμένη πληροφορία συμβολίζεται με το γράμμα Φ. Έτσι έχουμε:

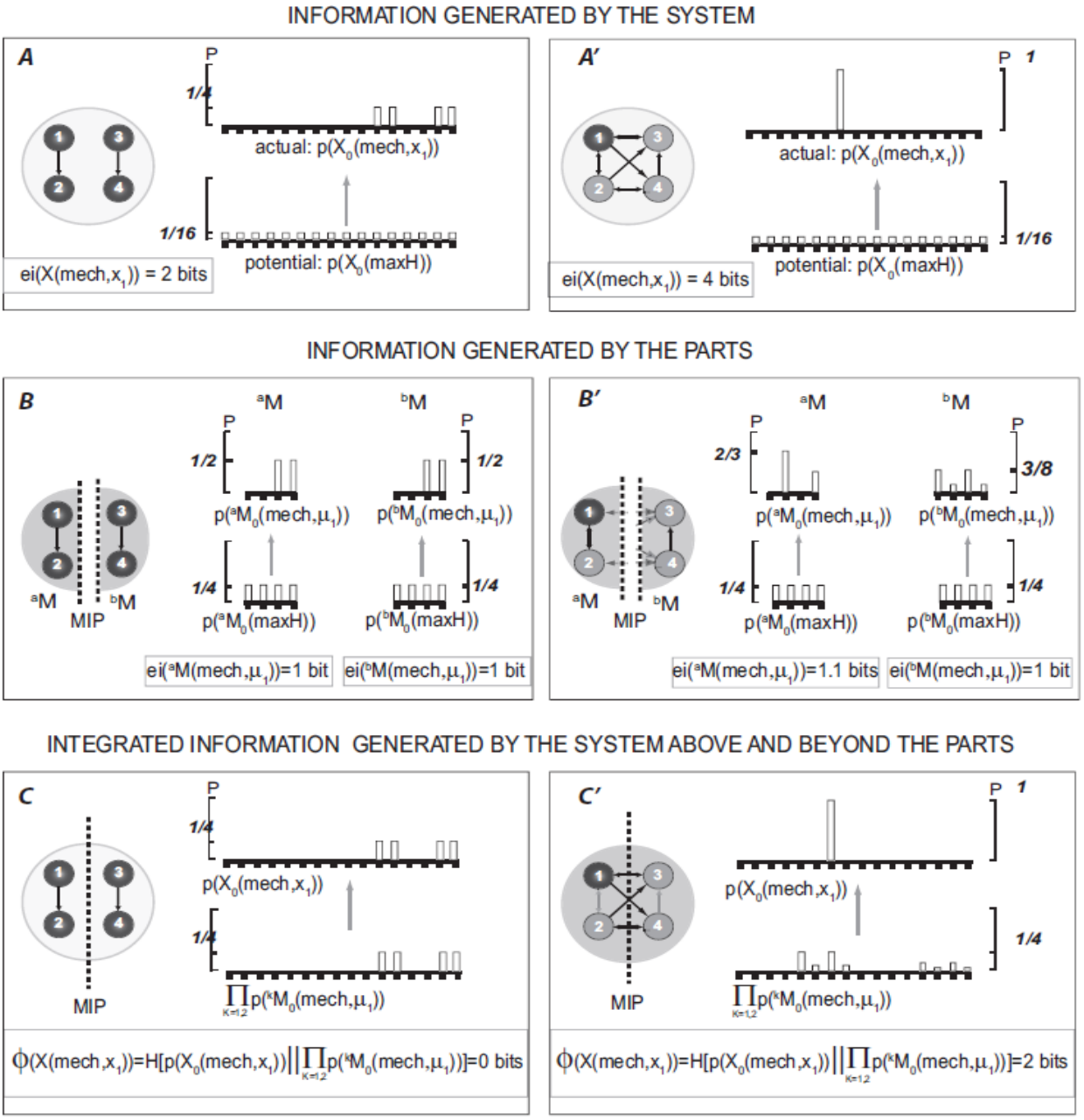
$$\Phi(\mathrm{X}(\mathrm{mech},x_1) = \mathrm{H}[\ \mathrm{p}(X_0(\mathrm{mech},\ x_1))\ ||\ \Pi\mathrm{p}(^kM_0(\mathrm{mech},\mu_1))\ ]\ \text{for}\ {}^kM_0 \in \mathrm{MIP}$$

Έτσι προκύπτει το πραγματικό εύρος που οφείλεται στις εσωτερικές αιτιατές αλληλεπιδράσεις κάθε κομματιού, το οποίο θεωρείται αυτόνομο σύστημα, ενώ οι είσοδοι που έρχονται από το εξωτερικό περιβάλλον αντιμετωπίζονται ως θόρυβοι. Η σύγκριση γίνεται με συγκεκριμένη αποσύνθεση του αρχικού συστήματος σε τέτοια μέρη, ώστε να μένει χωρίς να καταγράφεται η μικρότερη δυνατή ποσότητα πληροφορίας. Αυτή η *διαμέριση ελάχιστης πληροφορίας (minimum information partition-MIP)* αποσυνθέτει το σύστημα στα ελάχιστα κομμάτια του.

Για να γίνει κατανοητό πώς δουλεύει αρκεί να θεωρήσουμε δύο από τις εκατομμύρια φωτοδιόδους της ψηφιακής κάμερας (στο **Σχήμα 2**, αριστερά). Αναφέρθηκε πριν πως η κάθε φωτοδίοδος ανάλογα την κατάσταση στην οποία βρίσκεται παράγει 1 bit πληροφορίας. Θεωρώντας τις ανεξάρτητες στην περίπτωση μας έχουμε 1+1, άρα 2 bits πληροφορίας, επομένως από όλη την κάμερα παίρνουμε 1 εκατομμύριο bits πληροφορίας. Ωστόσο, όπως φαίνεται και στην εικόνα, το αποτέλεσμα των πραγματικών κατανομών που παράγεται ανεξάρτητα από τα μέρη της κάμερας είναι ίδιο με την πραγματική κατανομή του συστήματος ολόκληρου. Επομένως η σχετική εντροπία μεταξύ τους είναι μηδέν, δηλαδή το σύστημα δεν παράγει ολοκληρωμένη πληροφορία, παρά μόνο αυτή που προκύπτει από τα κομμάτια του.

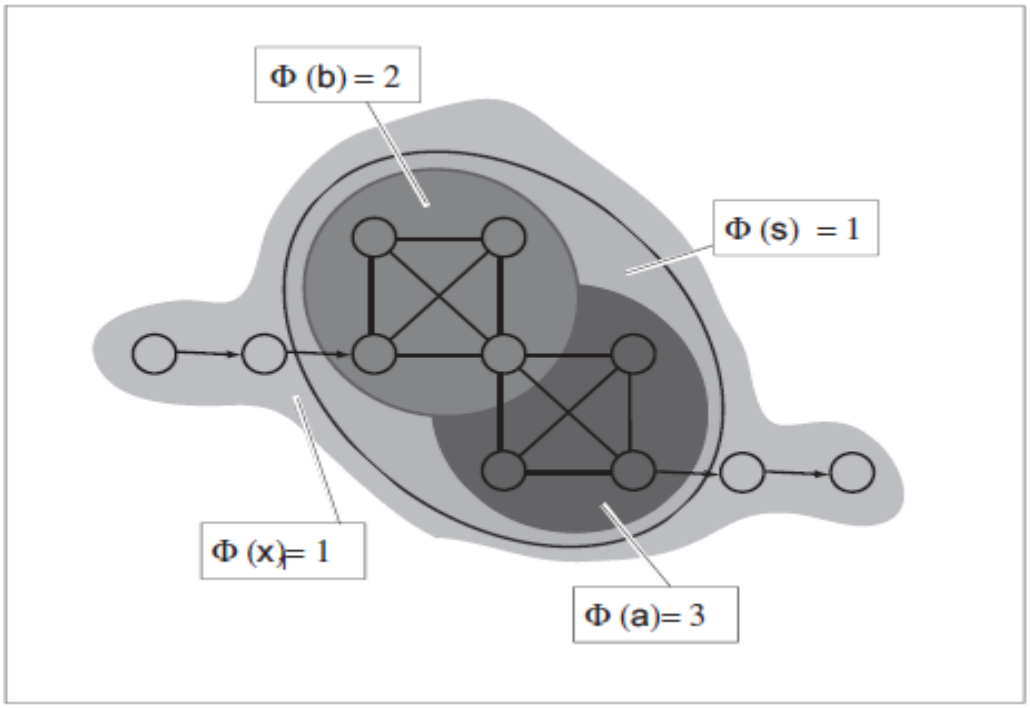
Προφανώς, για να είναι υψηλός ο δείκτης της ενσωματωμένης πληροφορίας το σύστημα πρέπει να είναι συνδεδεμένο με τέτοιο τρόπο, ώστε η πληροφορία να προκύπτει από αιτιατές αλληλεπιδράσεις μεταξύ, και όχι στο εσωτερικό, των επιμέρους μερών του. Ένα απλό παράδειγμα τέτοιου συστήματος παρουσιάζεται στο **Σχήμα 2** (δεξιά). Σε αυτή την περίπτωση η αλληλεπίδραση μεταξύ των ελάχιστων μερών του συστήματος παράγει περισσότερη πληροφορία από ότι τα μέρη ξεχωριστά ($\Phi(\mathrm{X}(\mathrm{mech},x_1) > 0$).

Τελικά μετρώντας το Φ για όλα τα υποσυστήματα, μπορούμε να αποφανθούμε ποια από αυτά αποτελούν συμπλέγματα. Συγκεκριμένα ένα σύμπλεγμα Χ είναι μια ομάδα στοιχείων που



Σχήμα 2: *Αριστερά:* Δύο φωτοδίοδοι ψηφιακής κάμερας. (A): Η πληροφορία που παράγεται από το σύστημα ως οντότητα. Το σύστημα παράγει 2 bits ενεργούς πληροφορίας θεωρώντας πως τα n_1 και n_3 ήταν ανοιχτά. (B): Η πληροφορία που παράγεται από τα μέρη. Η διαμέριση ελάχιστης πληροφορίας (MIP) είναι η αποσύνθεση του συστήματος σε ελαχιστοποιημένα μέρη, ώστε να μην καταγράφεται η μικρότερη δυνατή ποσότητα πληροφορίας. Εδώ η αποσύνθεση δίνει δύο φωτοδιόδους. (C): Η πληροφορία που παράγεται από ολόκληρο το σύστημα καταγράφεται και από τα μέρη του ξεχωριστά. Έτσι το πραγματικό εύρος του συστήματος σε σχέση με αυτό των μερών του έχουν σχετική εντροπία μηδέν. Επομένως το σύστημα δεν μπορεί να θεωρηθεί οντότητα. *Δεξιά:* Ενσωματωμένο σύστημα. Το κάθε στοιχείο είναι ανοιχτό όταν δέχεται 2 ή περισσότερες ανοιχτές καταστάσεις. Έστω ότι βρίσκεται στην κατάσταση x_1=1000. (A'): Λόγω μηχανισμού και της κατάστασης που βρισκόμαστε γνωρίζουμε ότι αυτό μπορεί να έχει προκύψει μόνο από μια συγκεκριμένη προηγούμενη κατάσταση, άρα η ενεργός πληροφορία που παίρνουμε είναι 4 bits. (B'): Η ενεργός πληροφορία που παράγεται από δύο ελάχιστα μέρη θεωρώντας τα ως αυτόνομα συστήματα και τις εξωτερικές εισόδους ως θορύβους. (C'): Ενσωματωμένη πληροφορία είναι αυτή που παράγεται από την αλληλεπίδραση μεταξύ των μερών (μαύρα βέλη). Εδώ, το πραγματικό ρεπερτόριο του συστήματος είναι διαφορετικό από αυτό των μερών και η μεταξύ τους σχετική εντροπία είναι 2 bit. Επομένως το σύστημα μπορεί να θεωρηθεί οντότητα (σύμπλεγμα) [5].

παράγουν ενσωματωμένη πληροφορία, η οποία δεν περιέχεται σε κάποια μεγαλύτερη ομάδα με υψηλότερο δείκτη Φ (**Σχήμα 3).** Ένα σύμπλεγμα λοιπόν μπορεί να θεωρηθεί ότι σχηματίζει μια οντότητα εφόσον έχει τη δική του εσωτερική «οπτική γωνία». Εφόσον η ενσωματωμένη πληροφορία παράγεται μέσα στα όρια του συμπλέγματος η εμπειρία είναι απαραίτητα ιδιωτική και σχετίζεται με την ατομική οπτική γωνία. Ένα φυσιολογικό σύστημα όπως ο εγκέφαλος είναι λογικό να περιέχει περισσότερα από ένα συμπλέγματα, ίσως πολλά μικρά και κάποια λίγο μεγαλύτερα. Μάλιστα ίσως και να υπάρχει κάθε στιγμή ένα βασικό σύμπλεγμα με αρκετά μεγαλύτερο Φ, στο οποίο υπόκειται η κυρίαρχη εμπειρία. Όπως φαίνεται και στο **Σχήμα 3** το κύριο σύμπλεγμα μπορεί να αποτελείται από στοιχεία άλλων συμπλεγμάτων με μικρότερο Φ. Έτσι, κάθε σύμπλεγμα μπορεί να είναι αιτιατά συνδεδεμένο μέσω πυλών με στοιχεία που δεν ανήκουν σε αυτό. Κατά την IIT αυτά τα στοιχεία μπορούν να επηρεάσουν εμμέσως την κατάσταση του βασικού συμπλέγματος χωρίς να συνεισφέρουν άμεσα στη συνειδητή εμπειρία που δημιουργεί.



Σχήμα 3: Σε αυτό το σύστημα, ο μηχανισμός είναι ότι τα στοιχεία δίνουν σήμα ως απόκριση σε έναν περιττό αριθμό ερεθισμάτων στις εισερχόμενες συνδέσεις (οι συνδέσεις χωρίς βέλη είναι αμφίδρομες). Αναλύοντας το σύστημα με βάση τη θεωρία μας προκύπτει ότι το σύστημα αποτελεί και ένα σύμπλεγμα (x, το απαλό γκρι) που περιέχει τρία επιπλέον συμπλέγματα (s,a,b). Παρατηρούμε ότι (i) τα συμπλέγματα επικαλύπτονται, (ii) ένα σύμπλεγμα μπορεί να αλληλοεπιδρά αιτιατά με στοιχεία που δεν είναι μέρος του, (iii) ομάδες στοιχείων με όμοια αρχιτεκτονική (a,b) παράγουν διαφορετική ποσότητα ενσωματωμένης πληροφορίας ανάλογα τις πύλες εισόδου και εξόδου [5].

2.4 Σύνδεση νευροεπιστήμης και εμπειρικών παρατηρήσεων

Μπορεί λοιπόν η παραπάνω προσέγγιση, έστω η βάση της, να βρει εφαρμογή πάνω στα δεδομένα που υφίστανται μέχρι σήμερα σχετικά με τη συνείδηση, ύστερα από δεκαετίες κλινικών και νευροβιολογικών παρατηρήσεων; Η εύρεση του Φ και των συμπλεγμάτων δεν είναι τόσο εύκολη

για ρεαλιστικά συστήματα, όμως μπορεί να επιτευχθεί με τη χρήση απλών δικτύων που έχουν δομική ομοιότητα με διάφορα μέρη του εγκεφάλου.

Για παράδειγμα, μέσω προσομοιώσεων έχει δειχθεί ότι το υψηλό Φ απαιτεί δίκτυα που συνδυάζουν τη λειτουργική ειδίκευση (κάθε στοιχείο του δικτύου έχει μοναδικό λειτουργικό ρόλο) με τη λειτουργική ενσωμάτωση (υπάρχουν πολλαπλά μονοπάτια για την αλληλεπίδραση των στοιχείων, **Σχήμα 4A**). Με άλλα λόγια, αυτού του είδους η αρχιτεκτονική είναι χαρακτηριστικό του κορτικοθαλαμικού συστήματος των θηλαστικών: διαφορετικά μέρη του εγκεφαλικού φλοιού ειδικεύονται σε διαφορετικές λειτουργίες και μέσω ενός δικτύου με πάρα πολλές συνδέσεις καταφέρνουν να αλληλοεπιδρούν μεταξύ τους. Μάλιστα, νευρολογικές έρευνες δείχνουν ότι το κορτικοθαλαμικό σύστημα είναι ακριβώς το μέρος του εγκεφάλου που οποιαδήποτε μορφή καταστροφής οδηγεί και σε απώλεια της συνείδησης.

Αντίθετα, το Φ είναι χαμηλό για συστήματα που αποτελούνται από μικρές, σχεδόν ανεξάρτητες ενότητες (**Σχήμα 4B**). Ίσως για αυτό άλλωστε η παρεγκεφαλίτιδα, παρά το μεγάλο αριθμό νευρώνων που διαθέτει, δε συνεισφέρει πολύ στη συνείδηση. Η οργάνωση της είναι σε ανεξάρτητα τμήματα παρεγκεφαλιδικού φλοιού που ενεργοποιούνται έχοντας ελάχιστη αλληλεπίδραση μεταξύ τους.

Προσομοιώσεις επίσης δείχνουν ότι μονάδες κατά μήκος πολλαπλών, διαχωρισμένων μονοπατιών εισόδου ή εξόδου δεν ενσωματώνονται στο ρεπερτόριο του κύριου συμπλέγματος (**Σχήμα 4C**). Για το λόγο αυτό, λογικά, η νευρική δραστηριότητα σε μονοπάτια εισόδου, αν και κρίσιμη για την ενεργοποίηση κάποια συνειδητής εμπειρίας, δε συνεισφέρει άμεσα στην εμπειρία αυτή, ακριβώς όπως και η δραστηριότητα σε μονοπάτια εξόδου, αν και κρίσιμη για την αναφορά της εμπειρίας.

Η προσθήκη πολλών παράλληλων κύκλων γενικά δεν αλλάζει τη σύνθεση του βασικού συμπλέγματος, αν και το Φ μπορεί να αλλάξει (**Σχήμα 4D**). Αντίθετα, κύκλοι ή βρόχοι στο φλοιό αποτελούν ειδικές υπορουτίνες, ικανές να επηρεάσουν την κατάσταση του βασικού κορτικοθαλαμικού συμπλέγματος χωρίς να είναι μέρος του. Τέτοιοι πληροφοριακά μονωμένοι βρόχοι θα μπορούσαν να απαρτίζουν τα νευρικά υποστρώματα για πολλές μη συνειδητές διαδικασίες που επηρεάζονται όμως από τη συνειδητή εμπειρία, όπως η αναγνώριση αντικειμένων ή η μετάφραση αυτών που θέλουμε να πούμε σε λόγια. Στο στάδιο που βρίσκεται η έρευνα, ωστόσο, είναι δύσκολο να πούμε ακριβώς ποια κυκλώματα του φλοιού μπορεί να αποτελούν ένα μεγάλο σύμπλεγμα με υψηλό Φ και ποια μπορεί να είναι πληροφοριακά μονωμένα.

Άλλες προσομοιώσεις δείχνουν ότι τα αποτελέσματα αποσυνδέσεων στο φλοιό γίνονται αμέσως εμφανή όσον αφορά την ενσωματωμένη πληροφορία. Ένα «κολοσσιαίο» κόψιμο ενός μεγάλου συμπλέγματος που βρίσκεται σε ανταπόκριση με το κορτικοθαλαμικό σύστημα, έχει απλώς σαν αποτέλεσμα τη δημιουργία δύο ξεχωριστών συμπλεγμάτων, κάτι που στηρίζεται και από τις έρευνες πάνω σε ασθενείς με χωρισμένο εγκέφαλο. Στη συγκεκριμένη μάλιστα περίπτωση το Φ των δύο συμπλεγμάτων δεν είναι ιδιαίτερα μειωμένο σε σχέση με το Φ του αρχικού συμπλέγματος. Λειτουργικές αποσυνδέσεις μπορούν επίσης να οδηγήσουν σε περιορισμό των νευρικών υποστρωμάτων της συνείδησης, όπως έχει παρατηρηθεί σε ψυχιατρικές περιπτώσεις και σε καταστάσεις ονειροπόλησης και ύπνωσης. Σε αυτή την περίπτωση μειώνεται τόσο το μέγεθος του συμπλέγματος όσο και η χωρητικότητα του σε ενσωματωμένη πληροφορία. Έτσι, λοιπόν, αν και

δεν είμαστε σε θέση να αποφασίσουμε ποιες ομάδες νευρώνων συμμετέχουν ή αποκλείονται από το βασικό σύμπλεγμα ανάλογα την περίσταση, ξέρουμε ότι το σύνολο των στοιχείων που αποτελούν το υπόβαθρο της συνείδησης σχηματίζουν ένα «δυναμικό σύμπλεγμα» ή «δυναμικό κορμό».

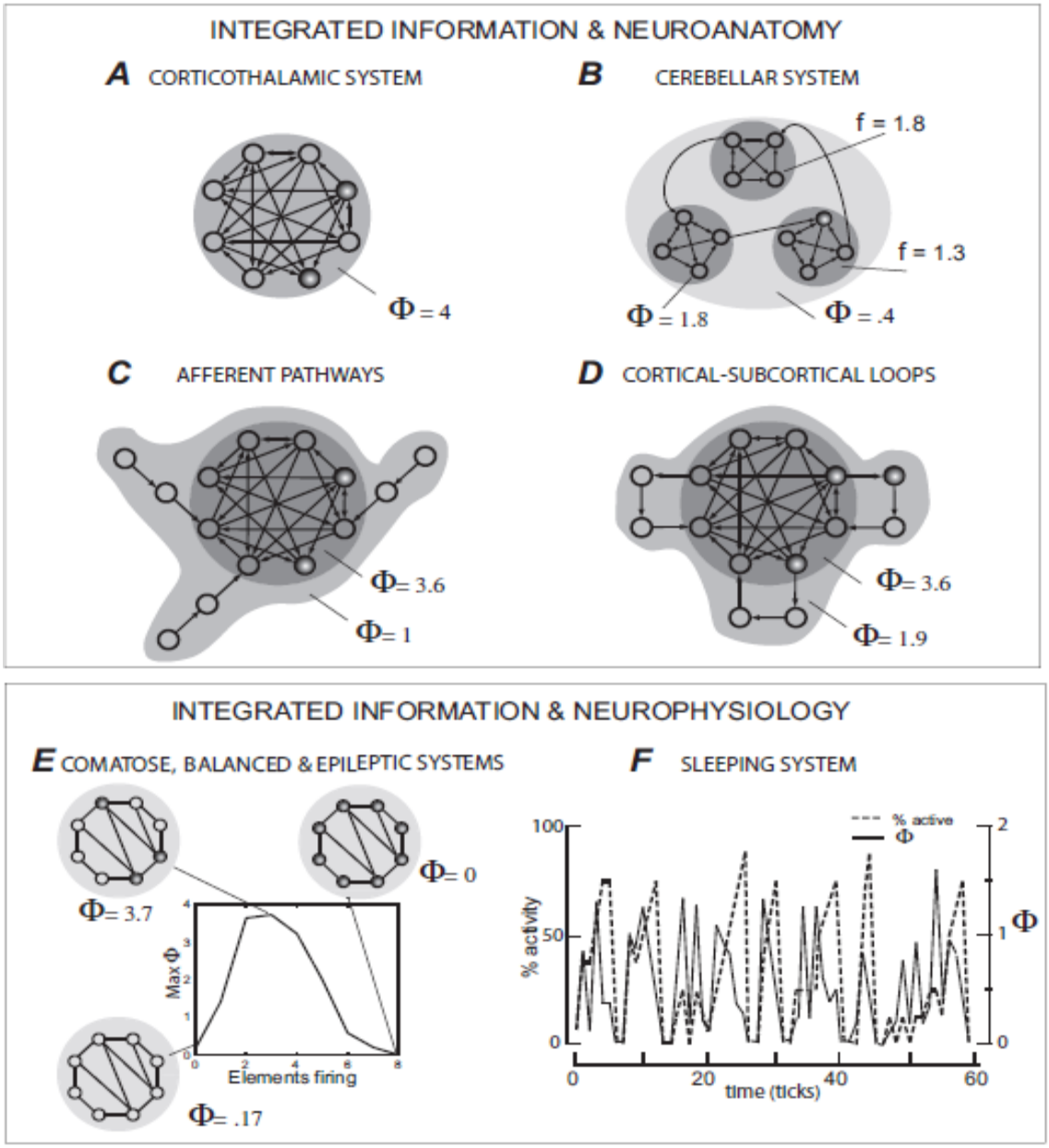
Επιπλέον προσομοιώσεις έχουν δείξει ότι η χωρητικότητα σε ενσωματωμένη πληροφορία μειώνεται όταν η νευρική δραστηριότητα είναι ιδιαίτερα υψηλή και κοντά σε συγχρονισμό, λόγω της δραματικής μείωσης του εύρους διακριτών καταστάσεων (**Σχήμα 4Ε**). Ίσως μάλιστα αυτός να είναι ο λόγος που παρουσιάζεται μείωση ή και πλήρης απουσία συνείδησης σε επιληπτικά επεισόδια.

Ένα ακόμα αρκετά αντιπροσωπευτικό παράδειγμα που παρατηρείται «ξεθώριασμα» της συνείδησης συμβαίνει σε ορισμένα στάδια ύπνου. Συμμετέχοντες σε έρευνα αφυπνίστηκαν κατά τη διάρκεια σταδίου του ύπνου που δεν κουνούσαν καθόλου τα μάτια τους (non-rapid eye movement – NREM), κατά βάση νωρίς τη νύχτα, και δήλωσαν ότι δεν είχαν καμία αίσθηση του εαυτού τους ή του περιβάλλοντος τους, παρά το ότι οι νευρώνες του κορτικοθαλαμικού συστήματος παρέμεναν ενεργοί. Όταν αφυπνίστηκαν σε στάδιο που κουνούσαν τα μάτια τους (REM) ή σε μια πιο ελαφρά φάση του NREM ύπνου, ανέφεραν ότι έβλεπαν όνειρα που τα χαρακτήριζαν ζωηρές εικόνες. Από την οπτική της ενσωματωμένης πληροφορίας, η μείωση της συνείδησης κατά τα πρώτα στάδια του ύπνου έρχεται ως συνέπεια της αστάθειας των κυκλωμάτων του φλοιού κατά τον NREM ύπνο. Εξαιτίας των εσωτερικών αλλαγών στην αγωγιμότητα, που προκαλείται από νευροδιαμορφωτικές αλλαγές (π.χ. χαμηλή ακετυλοχολίνη), οι νευρώνες του φλοιού δεν μπορούν να στέλνουν διαρκώς σήμα για περισσότερα από μερικά εκατοντάδες milliseconds και μπαίνουν σε μια υποβαθμισμένη-υπερπολωμένη κατάσταση. Είναι γεγονός, πως οι προσομοιώσεις επιβεβαιώνουν τα παραπάνω (**Σχήμα 4F**).

Τελικά, η συνείδηση δεν απαιτεί μόνο ένα νευρικό υπόστρωμα με κατάλληλη ανατομία και φυσιολογικές παραμέτρους, αλλά και χρόνο. Η IIT προβλέπει πως ο χρόνος που χρειάζεται για την παραγωγή μιας συνειδητής εμπειρίας στον εγκέφαλο αναδύεται άμεσα από το χρόνο που χρειάζεται να σχηματιστεί ένα ενσωματωμένο εύρος μεταξύ των στοιχείων του κορτικοθαλαμικού κύριου συμπλέγματος ώστε η κάθε διαφοροποίηση να είναι υψηλά «πληροφοριακή». Έτσι μία διαταραχή σε αυτό το σύμπλεγμα μικρότερη του ενός millisecond κατά πάσα πιθανότητα δε θα προκαλέσει κανένα αποτέλεσμα, οπότε θα έχουμε και Φ ίσο με μηδέν, ενώ μία των 100 milliseconds θα φέρει κάποιο αποτέλεσμα, άρα και αυξημένο Φ.

2.5 Η ποιότητα της συνείδησης

Αν η ποσότητα της ενσωματωμένης πληροφορίας που παράγεται από διαφορετικές δομές του εγκεφάλου (ή από την ίδια δομή με διαφορετικούς τρόπους) είναι η βάση μας για τις αλλαγές στο επίπεδο της συνείδησης, τότε τι είναι υπεύθυνο για την ποιότητα της κάθε εμπειρίας; Τι καθορίζει ότι τα χρώματα έχουν την όψη τους και ταυτόχρονα διαφέρουν στο πώς ακούγεται η μουσική; Για μία ακόμα φορά, η εμπειρία μας δείχνει ότι διαφορετικές ποιοτικά εμπειρίες προκύπτουν από διαφορετικές περιοχές του φλοιού. Έτσι, η καταστροφή ορισμένων μερών του εγκεφαλικού



Σχήμα 4: Συσχέτιση ενσωματωμένης πληροφορίας με νευροανατομία και νευροφυσιολογία. Τα στοιχεία στέλνουν σήμα ως ανταπόκριση σε δύο ή περισσότερα ερεθίσματα (εκτός από αυτά που έχουν μόνο μια είσοδο οπότε και την αντιγράφουν), ενώ όλες οι συνδέσεις είναι αμφίδρομες. (A): Ο υπολογισμός του Φ σε απλά μοντέλα ανατομίας υποδεικνύει ότι ένα λειτουργικά ενσωματωμένο και ειδικευμένο δίκτυο -όπως το κορτικοθαλαμικό σύστημα- είναι ικανό να παράγει υψηλές τιμές του Φ. (B, C, D): Αρχιτεκτονικές βασισμένες στην παρεγκεφαλίτιδα, με συγγενής διαδρόμους εισόδου-εξόδου και βρόχους επιτρέπουν στα συμπλέγματα να έχουν επιπλέον στοιχεία, αλλά με μειωμένο Φ σε σχέση με το βασικό κορτικοθαλαμικό σύμπλεγμα. (E): Το Φ μεγιστοποιείται σε ισορροπημένες καταστάσεις, ενώ αν πολύ λίγα ή πάρα πολλά στοιχεία είναι ενεργά το Φ καταρρέει. (F): Σε ένα σύστημα σαν το (E), το Φ καταρρέει αν ο αριθμός των στοιχείων που εκπέμπουν σήμα (διακεκομμένη γραμμή) είναι πολύ υψηλός, παραμένει σε χαμηλές τιμές όσο πέφτει η δραστηριότητα, ενώ αυξάνεται κατά τα πρώτα στάδια αύξησης της ενεργητικότητας [5].

φλοιού απαλείφει για πάντα την ικανότητα ενός ανθρώπου να δέχεται την εμπειρία των χρωμάτων (είτε αυτή έχει να κάνει με την όραση, τη φαντασία, τη μνήμη ή ένα όνειρο), ενώ η καταστροφή άλλων μερών επιλεκτικά απαλείφει τη δυνατότητα να αντιληφθεί τα σχήματα. Προφανώς υπάρχει κάτι σε αυτά τα διαφορετικά μέρη του φλοιού που είναι υπεύθυνο για τη διαφορετική τους συνεισφορά στην ποιότητα της εμπειρίας. Αλλά τι είναι αυτό το κάτι;

Η IIT υποστηρίζει ότι, όπως η *ποσότητα* της συνείδησης που παράγεται από ένα σύμπλεγμα στοιχείων προκύπτει από το μέγεθος της ενσωματωμένης πληροφορίας που δημιουργείται μεταξύ των μερών της, η *ποιότητα* της συνείδησης προκύπτει από το σύνολο των πληροφοριακών σχέσεων που παράγει ο μηχανισμός της. Ας σκεφτούμε πάλι το παράδειγμα με τη φωτοδίοδο. Η φωτοδίοδος, λοιπόν, αναγνωρίζει τη φωτεινότητα ενώ ο άνθρωπος ξεχωρίζει το «φως» μέσα από μια πληθώρα πιθανών καταστάσεων, παράγοντας έτσι και πολύ περισσότερη πληροφορία. Αυτό είναι αποτέλεσμα του ότι το φως δεν είναι απλώς το αντίθετο του σκοταδιού, αλλά ουσιαστικά διαφορετικό και από οποιοδήποτε άλλο χρώμα, σχήμα, μυρωδιά, ήχο και ούτω καθεξής [4].

Σε αυτό το σημείο πρέπει να δοθεί έμφαση στο ότι πρακτικά κάποιος εκείνη τη στιγμή δεν διαχωρίζει την εμπειρία «φως» έναντι όλων των υπόλοιπων εναλλακτικών, παρά την ξεχωρίζει ατομικά, δηλαδή καταλήγει σε αυτή με έναν ιδιαίτερο τρόπο. Ας σκεφτούμε το παράδειγμα ενός δυαδικού μετρητή, ικανού να διακρίνει μεταξύ των τεσσάρων αριθμών: 00, 01, 10, 11. Όταν ο μετρητής δίνει τη δυαδική τιμή «3», δεν διαλέγει απλώς την τιμή 11 από το σύνολο, αφού έτσι θα ήταν ανιχνευτής. Ως μετρητής, το σύστημα του πρέπει να μπορεί να ξεχωρίζει καθεμία από τις τιμές με διαφορετικό τρόπο την καθεμία. Φυσικά αυτό το κάνει μέσο των μηχανισμών του που είναι σε θέση να κάνουν το σύστημα να καταλάβει αν ασχολείται με το πρώτο ή το δεύτερο ψηφίο και αν αυτό είναι 0 ή 1. Ο κάθε μηχανισμός δηλαδή έχει τη δική του συνεισφορά, ενώ ενεργούν ταυτόχρονα. Αντίστοιχα, όταν κάποιος βλέπει φως, μηχανισμοί στον εγκέφαλο του του επιτρέπουν αυτόματα να το ξεχωρίσει από οποιαδήποτε άλλη πιθανή μορφή εμπειρίας. Έτσι, σύμφωνα με την IIT, αυτοί οι μηχανισμοί συνεργάζονται και παράγουν ενσωματωμένη πληροφορία καθορίζοντας ένα σετ πληροφοριακών σχέσεων που αποφασίζει απόλυτα και μονοσήμαντα την ποιότητα της εμπειρίας.

Για να εξεταστεί πώς αυτό το κομμάτι της θεωρίας μπορεί να πάρει έναν μαθηματικό-πρακτικό σχηματισμό, ας θεωρήσουμε πάλι το σύμπλεγμα *n* δυαδικών στοιχείων X(mech, x_1) που έχει έναν συγκεκριμένο μηχανισμό και βρίσκεται σε μια συγκεκριμένη κατάσταση. Ο μηχανισμός προκύπτει από ένα σύνολο συνδέσεων X^{conn} μεταξύ των στοιχείων. Έστω τώρα ότι κάθε πιθανή κατάσταση του συστήματος απαρτίζει έναν άξονα ή διάσταση στον χώρο των qualia (Q), που διαθέτει 2^n διαστάσεις. Κάθε άξονας χαρακτηρίζεται από την πιθανότητα p της κατάστασης, με τιμές από 0 έως 1, ώστε να προκύπτει ένα εύρος (κατανομή πιθανοτήτων για τις καταστάσεις του συμπλέγματος) που να αντιστοιχεί σε ένα σημείο στο Q (**Σχήμα 5**).

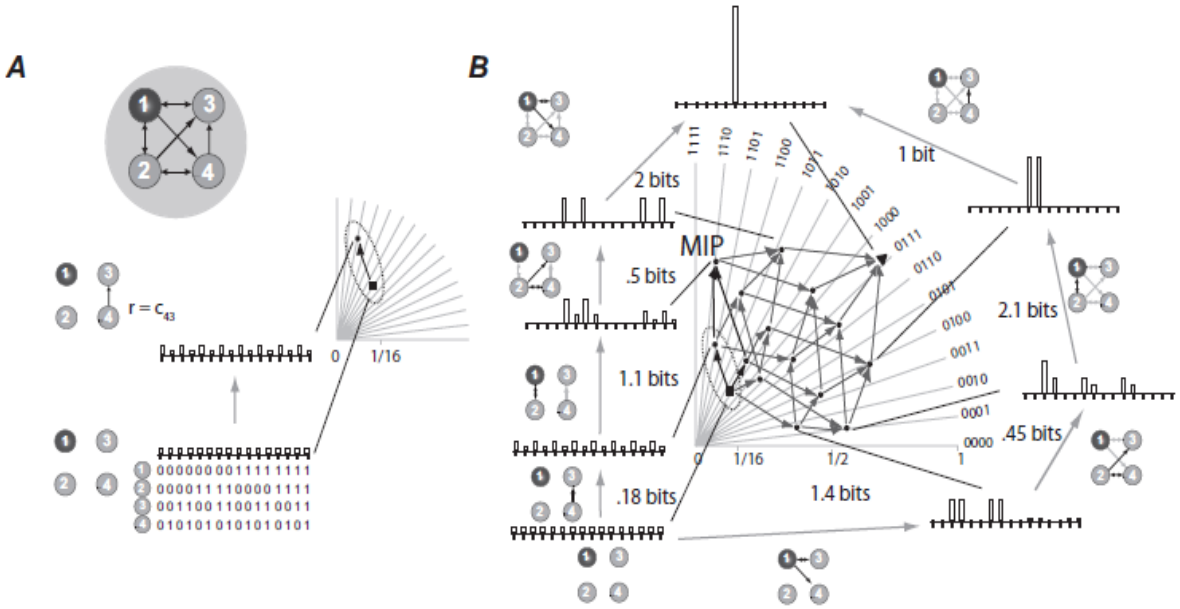
Σε αυτό το σημείο χρειάζεται να εξεταστεί πώς οι συνδέσεις μεταξύ των στοιχείων του συμπλέγματος καθορίζουν τις κατανομές πιθανότητας, δηλαδή πώς μια ομάδα μηχανισμών καθορίζει ένα σετ πληροφοριακών σχέσεων. Αρχικά ας θεωρήσουμε ένα σύμπλεγμα με όλες τις συνδέσεις μεταξύ των στοιχείων του απενεργοποιημένες, αποκλείοντας έτσι οποιαδήποτε αλληλεπίδραση (**Σχήμα 5Α**). Χωρίς μηχανισμό, η κατάσταση x_1 δεν παρέχει κάποια πληροφορία σχετικά με την προηγούμενη κατάσταση του συστήματος, αφού χωρίς αλληλεπίδραση όλες οι

καταστάσεις είναι ισοπίθανες. Στον χώρο Q, αυτή η κατανομή πιθανότητας είναι ένα σημείο που προβάλλει σε όλους τους άξονες με p = $1/2^n$. Στη συνέχεια έστω ότι μία σύνδεση ανήκει στο μηχανισμό και όλες οι υπόλοιπες αποτελούν εξωτερικό θόρυβο (**Σχήμα 5Α**). Όπως και με τη φωτοδίοδο, ο μηχανισμός που προκύπτει μαζί με την κατάσταση, στην οποία βρίσκεται το σύστημα, αυξάνουν την πιθανότητα ορισμένων προηγούμενων καταστάσεων από τις οποίες μπορεί να προήλθε η x_1, δημιουργώντας ένα πραγματικό εύρος. Στον χώρο Q, το πραγματικό εύρος που προκύπτει από αυτή τη σύνδεση είναι ένα σημείο που προβάλλει σε άλλους άξονες με διαφορετικές τιμές του p. Έτσι η σύνδεση δημιουργεί μια πιο συγκεκριμένη κατανομή, παράγοντας ταυτόχρονα πληροφορία (αφού μειώνει την αβεβαιότητα). Πιο γενικά, μπορούμε να πούμε ότι η σύνδεση ειδικεύει μια πληροφοριακή σχέση, μια σχέση δηλαδή μεταξύ δύο κατανομών πιθανότητας. Αυτή η πληροφοριακή σχέση αντιστοιχεί σε ένα διάνυσμα στον Q (q-διάνυσμα) που ξεκινάει από το σημείο που αντιστοιχεί σε κατανομή με μέγιστη εντροπία (p=$1/2^n$) και καταλήγει στο σημείο που αντιστοιχεί στο πραγματικό εύρος που προκύπτει λόγω της σύνδεσης. Το μήκος (απόκλιση) του q-διανύσματος εκφράζει το *πόσο* η σύνδεση επηρέασε την κατανομή (πόση ενεργό πληροφορία παράγει, τη σχετική εντροπία μεταξύ των δύο κατανομών), ενώ η κατεύθυνση εκφράζει τον τρόπο με τον οποίο η σύνδεση επηρέασε την κατανομή. Παρόμοια, αν εξεταστεί κάθε σύνδεση ξεχωριστά θα προκύψουν πολλά και διαφορετικά q-διανύσματα.

Ας θεωρήσουμε τώρα όλους τους πιθανούς συνδυασμούς συνδέσεων (**Σχήμα 5Β**). Προς στιγμήν έστω ότι υπάρχει μια επιπλέον σύνδεση μαζί με την αρχική. Μαζί αυτές ειδικεύουν ένα καινούριο πραγματικό εύρος (ένα νέο σημείο στον Q) και παράγουν περισσότερη πληροφορία από ότι η καθεμία ξεχωριστά, αφού δίνουν μια πιο συγκεκριμένη κατανομή. Στην άκρη λοιπόν του q-διανύσματος που είχαμε από την πρώτη σύνδεση, μπορούμε να προσθέσουμε το δεύτερο που προκύπτει από τη δεύτερη σύνδεση, σχηματίζοντας μια «γωνία» στον Q (η πρόσθεση των διανυσμάτων θα δώσει το ίδιο αποτέλεσμα όποιο και να τοποθετηθεί πρώτο). Ο κάθε συνδυασμός συνδέσεων, λοιπόν, μας δίνει τη δική του q-γωνία από κολλημένα διανύσματα. Γενικά, όσο περισσότερες συνδέσεις έχουμε, τόσο περισσότερο θα σχηματίζεται το πραγματικό εύρος και θα διαφέρει από το αρχικό.

Τελικά, ας θεωρήσουμε την ταυτόχρονη συνεισφορά όλων των συνδέσεων του συμπλέγματος (**Σχήμα 5Β**). Προκύπτει λοιπόν το τελικό πραγματικό ρεπερτόριο, το σημείο στο οποίο όλες οι q-γωνίες συγκλίνουν. Μαζί, όλες οι q-γωνίες στον Q οριοθετούν ένα quale, ένα συμπαγές με 2^n διαστάσεις σχήμα. Στο κάτω μέρος βρίσκεται η κατανομή μέγιστης εντροπίας και μέσω q-διανυσμάτων καταλήγουμε στην κορυφή και το πραγματικό εύρος του συμπλέγματος ως σύνολο. Το σχήμα προκύπτει από όλες τις πληροφοριακές σχέσεις που δημιουργούνται από τις αλληλεπιδράσεις μεταξύ των στοιχείων του συμπλέγματος. Να σημειωθεί ότι το ίδιο ακριβώς σύμπλεγμα με τον ίδιο μηχανισμό θα μας δώσει διαφορετικό σχήμα στον Q αν βρίσκεται σε διαφορετική κατάσταση από την x_1.

Αξίζει σε αυτό το σημείο να αναφερθούν ορισμένες ιδιότητες των πληροφοριακών σχέσεων ή q-διανυσμάτων. Αρχικά, οι πληροφοριακές σχέσεις εξαρτώνται από την πορεία. Ως πορεία εννοούμε οποιοδήποτε σημείο στον Q αντιστοιχεί σε ένα εύρος και παράχθηκε από ένα

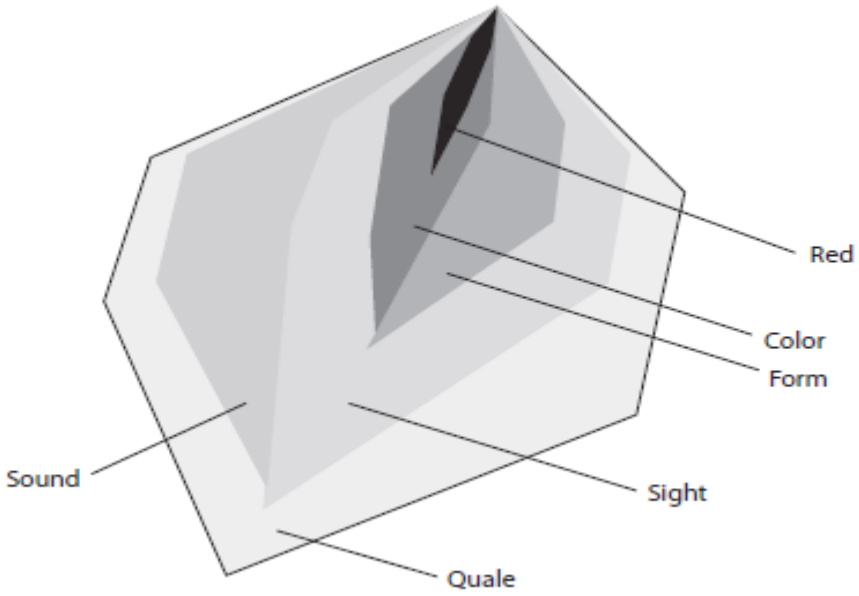


Σχήμα 5: Qualia. (A): Το σύστημα στο εσωτερικό του είναι ίδιο με αυτό στο **Σχήμα 2A**. Ο χώρος Q για ένα σύστημα με 4 στοιχεία έχει 16 διαστάσεις. Στην κατάσταση x_1 = 1000 το σύμπλεγμα παράγει ένα quale ή σχήμα στον Q, ως εξής. Η κατανομή μέγιστης εντροπίας για τις 16 καταστάσεις είναι ένα σημείο (μαύρο τετράγωνο) που αποθέτει ίδια πιθανότητα (p=1/16) σε όλες τις καταστάσεις (άξονες) κοντά στην αρχή αξόνων του Q. Ενεργοποιώντας μια σύνδεση r μεταξύ των στοιχείων 3 και 4 παίρνουμε ότι εφόσον το στοιχείο 3 έχει την τιμή 0, τότε η πιθανότητα στην προηγούμενη κατάσταση το στοιχείο 4 να ήταν 1 γίνεται p=0.25 (αντί για 0.5 που ήταν), ενώ προφανώς το 4 να ήταν πριν 0 γίνεται p=0.75. Υπάρχει λοιπόν αλλαγή στην πραγματική κατανομή πιθανότητας του συστήματος. Έτσι από τη σύνδεση r προκύπτει ένα νέο σημείο στον Q, που ενώνεται με το αρχικό με ένα q-διάνυσμα, το οποίο εκφράζει γεωμετρικά την πληροφοριακή σχέση που προκύπτει από τη σύνδεση. Το μήκος του χαρακτηρίζει το πόσο επηρεάστηκε η κατανομή και η κατεύθυνση τον τρόπο που έγινε αυτό. (B): Ενεργοποιώντας περισσότερες συνδέσεις αλλάζει συνεχώς το πραγματικό εύρος και δημιουργούνται νέα σημεία και q-διανύσματα στον Q. Η εικόνα δείχνει 16 από τα 399 στον Q, που μπορούν να προκύψουν από τους διάφορους συνδυασμούς ενεργοποίησης των συνδέσεων των τεσσάρων στοιχείων. Προφανώς όποια πορεία κι αν ακολουθηθεί (με όποια σειρά και να ενεργοποιήσουμε τις συνδέσεις αυτές θα καταλήξουν στο ίδιο σημείο. Στο σχήμα φαίνονται επίσης δύο διαδρομές και το ρεπερτόριο που αντιστοιχεί σε κάθε σημείο πριν γίνει η συνέχεια για το επόμενο, καθώς και η ενεργός πληροφορία που αντιστοιχεί σε κάθε q-διάνυσμα [5].

συγκεκριμένο υποσύνολο συνδέσεων. Μπορεί να δειχθεί ότι ένα q-διάνυσμα που παράχθηκε από την προσθήκη μιας επιπλέον σύνδεσης μπορεί να αλλάξει μέγεθος και κατεύθυνση ανάλογα με την, μέχρι εκείνη τη στιγμή, πορεία. Στο **Σχήμα 5**, όταν είναι απομονωμένα τα στοιχεία 4 και 3, η σύνδεση τους r παράγει ένα μικρό q-διάνυσμα (0.18 bits) με μια συγκεκριμένη κατεύθυνση. Όταν όμως έχουν προηγηθεί όλες οι συνδέσεις εκτός της r τότε αυτή δημιουργεί ένα μεγαλύτερο q-διάνυσμα (1 bit) που δείχνει σε διαφορετική κατεύθυνση.

Ένα άλλο σημαντικό χαρακτηριστικό των q-διανυσμάτων είναι η περιπλοκή ($\gamma > 0$). Ένα q-διάνυσμα είναι μπλεγμένο αν οι βαθύτερες συνδέσεις από αυτό παράγουν όλες μαζί περισσότερη πληροφορία από ότι το άθροισμα αυτής που παράγει ξεχωριστά η καθεμία. Έτσι, η περιπλοκή χαρακτηρίζει τις πληροφοριακές σχέσεις που ως σύνολο είναι «δυνατότερες» από το απλό άθροισμα των μερών τους (**Σχήμα 5B**). Γεωμετρικά, η περιπλοκή είναι αυτή που «στρεβλώνει» το σχήμα του quale και δεν έχουμε απλώς κάθετα διανύσματα μεταξύ των αξόνων. Έχει επίσης και ορισμένες συνέπειες. Για παράδειγμα ένα μπλεγμένο q-διάνυσμα μπορεί να υποστηριχθεί ότι ειδικεύει μια *έννοια*, τα βαθύτερα στοιχεία της οποίας δεν μπορούν να

αποσυντεθούν και να την επανασχηματίσουν ως άθροισμα ομάδων αυτών των στοιχείων. Επιπλέον, όπως το Φ χαρακτηρίζει τα συμπλέγματα, έτσι και η περιπλοκή *γ* χαρακτηρίζει φόρμες. Κατά αναλογία με τα συμπλέγματα, οι φόρμες είναι ομάδες q-διανυσμάτων που είναι πιο πυκνά μπλεγμένα από τα κοντινά q-διανύσματα. Μπορούν να θεωρηθούν ως συγκρότημα πληροφοριακών σχέσεων που απαρτίζουν διακριτά «υποσχήματα» στον Q (**Σχήμα 7**). Πιο συγκεκριμένα οι φόρμες θα αναλυθούν στη συνέχεια, καθώς παίζουν σημαντικό ρόλο στην κατανόηση της δομής της εμπειρίας.



Σχήμα 6: Σχηματική απεικόνιση για τις φόρμες και υπο-φόρμες. Μια φόρμα αντικατοπτρίζεται ως ένα πολύγωνο quale και είναι ένα σύνολο q-διανυσμάτων που είναι πιο πυκνά μπλεγμένα από τα γειτονικά και μπορεί να θεωρηθεί ως ένα συγκρότημα πληροφοριακών σχέσεων που συγκροτούν ένα διακριτό υποσχήμα στον Q. Δύο διαφορετικές φόρμες, για παράδειγμα, μπορούν να αντιστοιχούν στις λεπτομέρειες του ήχου και της όρασης. Μια υπο-φόρμα είναι ένα σύνολο ακόμα πιο μπλεγμένων q-διανυσμάτων. Το χρώμα και το σχήμα μπορούν να αποτελούν δύο υπο-φόρμες της οπτικής φόρμας. Στοιχειώδεις φόρμες είναι αυτές οι οποίες δεν μπορούν να κάνουν πιο συγκεκριμένη μια ποσότητα, όπως το χρώμα κόκκινο [5].

Πώς γίνεται, λοιπόν, μέσα από αυτές τις έννοιες να γίνει καλύτερα κατανοητή η ποιότητα της συνείδησης; Δεν είναι εύκολη η εξοικείωση με έναν πολυδιάστατο χώρο που δύσκολα σχεδιάζεται. Μέχρι τώρα η IIT υποστηρίζει ότι οι πληροφοριακές σχέσεις που προκύπτουν από τον μηχανισμό και την κατάσταση ενός συμπλέγματος μεταφράζεται ως ένα σχήμα (quale) στον χώρο Q. Μάλιστα υποστηρίζεται ότι αυτό το σχήμα καθορίζει ολοκληρωτικά την ποιότητα της συνείδησης.

Για κάθε ξεχωριστή εμπειρία προκύπτουν και διαφορετικά σχήματα στον Q, ακόμα και αν προέρχονται από το ίδιο σύμπλεγμα και παράγεται η ίδια ποσότητα Φ. Όταν ένα στοιχείο του συμπλέγματος ενεργοποιείται, παράγει πληροφορία, ακόμα και αν αυτή από μόνη της είναι κάτι ασήμαντο, και αλλάζει το σχήμα του quale. Επιπλέον, αν δύο εμπειρίες είναι παρόμοιες τότε και τα σχήματα τους είναι παρόμοια, και τόσο διαφορετικές στο βαθμό που διαφέρουν τα σχήματα τους. Έτσι το σύνολο των σχημάτων που παράγει ένα σύστημα για όλες τις διαφορετικές

καταστάσεις αποτελεί και το σύνολο των εμπειριών του συστήματος. Σημαντικό είναι επίσης το ότι, αφού στο σχηματισμό του quale συμβάλλουν το ίδιο κατάσταση και μηχανισμός, τότε δύο διαφορετικά συστήματα στην ίδια κατάσταση θα παράγουν δύο διαφορετικές εμπειρίες. Από την άλλη πλευρά υπάρχει η πιθανότητα δύο διαφορετικά συστήματα να παράγουν ακριβώς την ίδια εμπειρία. Για παράδειγμα ας σκεφτούμε πάλι τη φωτοδίοδο. Αναφέρθηκε πως οι καταστάσεις (00,01,10,11), με την κατάσταση x_1=11, έχουν κατανομή μέγιστης εντροπίας (1/4,1/4,1/4,1/4) και πραγματικό εύρος (0,0,1/2,1/2). Αυτό στον Q αντιστοιχεί σε ένα q-διάνυσμα, μήκους 1bit. Έστω ότι γίνεται αντικατάσταση του αισθητήρα φωτός με έναν αντίστοιχο θερμοκρασίας, αποκτώντας ουσιαστικά ένα θερμίστορ. Παρά το ότι η φυσική ιδιότητα του αισθητήρα έχει αλλάξει, σύμφωνα με την IIT, η μικρή αυτή εμπειρία πρέπει να είναι η ίδια, καθώς η πληροφοριακή σχέση μεταξύ των δύο συσκευών δεν έχει αλλάξει (αισθητήρας, ανιχνευτής). Αντίστοιχα όταν μια λογική πύλη OR είναι ενεργή (σε κατάσταση λογικού 1) και μια λογική πύλη AND μη ενεργή παράγουν την ίδια ελάχιστη εμπειρία. Φυσικά αν μιλήσουμε για πιο πολύπλοκα συστήματα που παράγουν υψηλό Φ είναι δύσκολο να παράγουν ακριβώς την ίδια εμπειρία.

Ένα άλλο χαρακτηριστικό είναι ότι, εφόσον η εμπειρία είναι ενσωματωμένη πληροφορία, οι πληροφοριακές σχέσεις ενός συμπλέγματος συνεισφέρουν στην εμπειρία, ενώ αυτές εκτός του κύριου συμπλέγματος, όπως κύκλοι στο φλοιό ή συμπλέγματα υπεύθυνα για τις αισθήσεις μας, δεν συνεισφέρουν στην ποσότητα ή την ποιότητα της εμπειρίας.

Είναι σε αυτό το σημείο σημαντικό να σημειωθεί πού συνεισφέρει το Φ σε αυτό το σημείο (**Σχήμα 7Α**). Η ελάχιστη διαμέριση πληροφορίας (minimum information partition-MIP) είναι ένα απλό σημείο στον Q. Είναι αυτό που προκύπτει αν υποτεθεί ότι υπάρχουν συνδέσεις μόνο στο εσωτερικό των ελάχιστων μερών ενός συστήματος και όχι συνδέσεις μεταξύ τους. Το Φ λοιπόν είναι το διάνυσμα που ξεκινάει από αυτό το σημείο και καταλήγει στο σημείο του πραγματικού εύρους ολόκληρου το συστήματος με όλες τις συνδέσεις. Έτσι το Φ θα είναι για παράδειγμα 0 σε ένα σύστημα που αποτελείται από δύο ανεξάρτητα μεταξύ τους συμπλέγματα (**Σχήμα 7Β**).

Οι παραπάνω έννοιες έχουν ως στόχο να παρέχουν ένα πλαίσιο για μια σωστή και κατανοητή μετάφραση ποιοτικών χαρακτηριστικών που προκύπτουν από τη φαινομενολογία στη γλώσσα των μαθηματικών, κυρίως με τις πληροφοριακές σχέσεις στον Q. Ιδανικά, στόχος είναι η ανάπτυξη των εννοιών σε σημείο που θα είναι εφικτή η δημιουργία μιας γεωμετρικής αναπαράστασης μετά από φαινομενολογική μελέτη του ανθρώπινου εγκεφάλου. Μάλιστα σε αυτή την περίπτωση, θεωρητικά, θα υπάρχει και η δυνατότητα απεικόνισης και άλλων συστημάτων, όπως το σύστημα που χρησιμοποιούν για τη μετακίνηση τους οι νυχτερίδες, και η σύγκριση με τη δικιά μας.

Προς το παρόν, λόγω συνδυαστικών προβλημάτων που προκύπτουν από το πολυδιάστατο σχήμα των quale, η παραγωγή τους υφίσταται μόνο για συστήματα με λίγα στοιχεία και απλώς υπάρχει η ελπίδα ότι μέσω της «γλώσσας» του Q θα γίνει εφικτή η αντιστοίχιση χαρακτηριστικών τους με δικά μας φαινομενολογικά χαρακτηριστικά και νευροψυχολογικές παρατηρήσεις. Μία τέτοια λίστα είναι η παρακάτω [6].

(i) Η εμπειρία διαιρείται σε βασικές φόρμες, όπως είναι οι αισθήσεις της όρασης, ακοής, αφής, όσφρησης, γεύσης (και άλλες), και υπο-φόρμες, όπως είναι η όψη του χρώματος και του σχήματος. Σύμφωνα με την IIT, στον χώρο Q οι φόρμες είναι πυκνά μπλεγμένα q-διανύσματα που

διαμορφώνουν υποσχήματα μέσα στο quale. Ακόμη πιο πυκνά μπλεγμένα q-διανύσματα είναι οι υπο-φόρμες και λειτουργούν αντίστοιχα και όπως έχει εξηγηθεί (**Σχήμα 6**).

(ii) Μερικές εμπειρίες φαίνεται να είναι «στοιχειώδεις», δηλαδή δεν μπορούν να αποσυντεθούν περισσότερο. Τέτοια παραδείγματα είναι το χρώμα κόκκινο, ένας πόνος, μια φαγούρα, εμπειρίες για τις οποίες είναι δύσκολο, αν όχι απίθανο, να εντοπίσουμε μια περαιτέρω πιο αναλυτική φαινομενολογική δομή. Σύμφωνα με την IIT, αυτές οι στοιχειώδεις εμπειρίες αντιστοιχούν στις υπο-φόρμες που δεν περιέχουν ακόμα πιο πυκνά μπλεγμένες υπο-υπο-φόρμες.

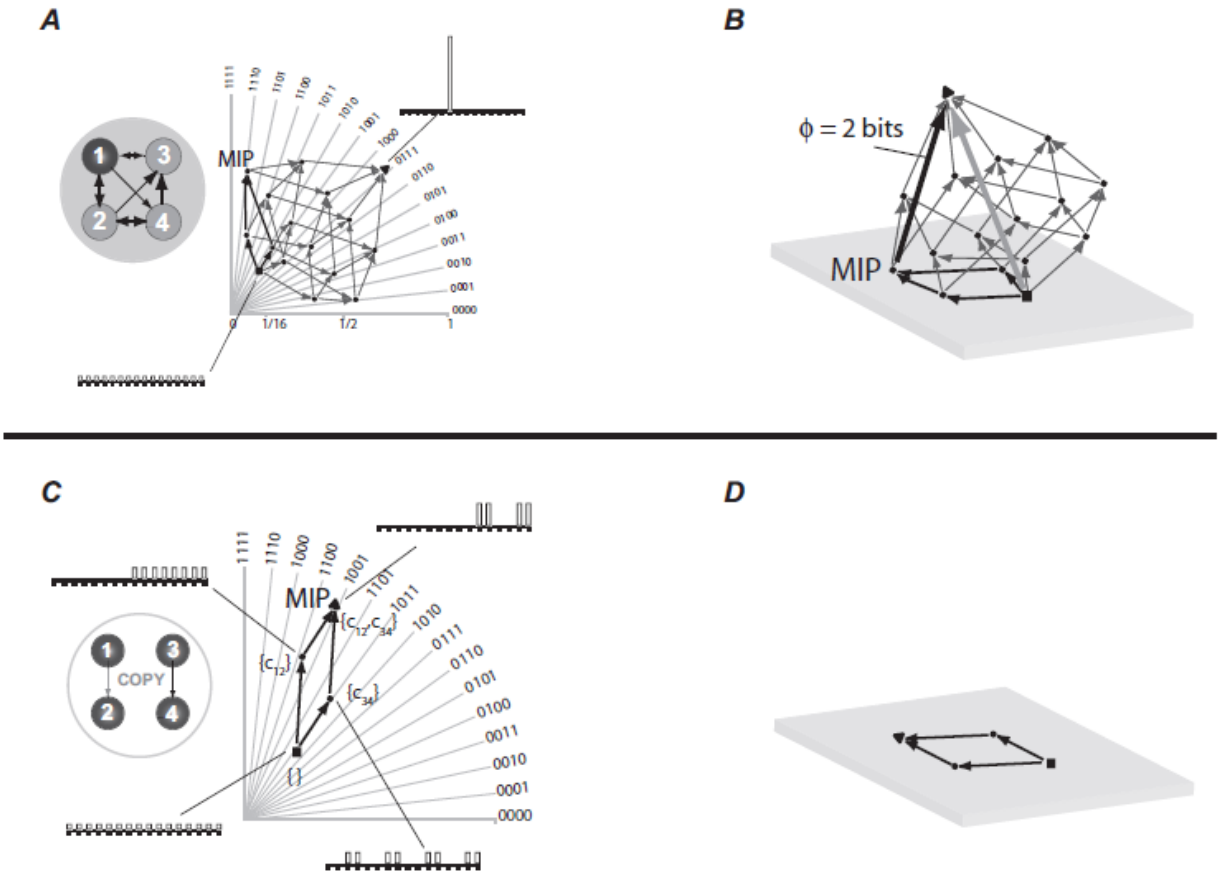
(iii) Κάποιες εμπειρίες είναι ομογενείς, ενώ άλλες είναι σύνθετες. Παραδείγματα είναι η καθολική εμπειρία του χρώματος μπλε όταν κοιτάμε έναν καθαρό ουρανό και η εμπειρία ενός εμπορικού κέντρου σε ώρα αιχμής, αντίστοιχα. Στον Q οι ομογενείς εμπειρίες αντιστοιχούν σε απλά σχήματα, ενώ οι σύνθετες σε ιδιαίτερα πολύπλοκα σχήματα με πολλές διακριτές φόρμες και υπο-φόρμες.

(iv) Ορισμένες εμπειρίες είναι ιεραρχικά οργανωμένες. Αν κάποιος σκεφτεί ένα πρόσωπο, αμέσως το βλέπει ως σύνολο ότι είναι το πρόσωπο κάποιου, αλλά επίσης αναγνωρίζει και τα μέρη-χαρακτηριστικά του όπως μαλλιά, μάτια, στόμα, μύτη και με τη σειρά τους αυτά ότι έχουν συγκεκριμένα χαρακτηριστικά. Η υποκειμενική εμπειρία προκύπτει από πληροφοριακές σχέσεις που είναι μπλεγμένες (δεν μπορούν να αποτελέσουν προϊόν από ανεξάρτητα στοιχεία) μεταξύ ιεραρχικών επιπέδων. Για παράδειγμα, οι πληροφοριακές σχέσεις που σχηματίζουν την εμπειρία ενός προσώπου είναι πιο πυκνά μπλεγμένες από αυτές που δημιουργούνται αν δούμε έναν παράξενο συνδυασμό χαρακτηριστικών προσώπου σε έναν Κυβιστικό πίνακα.

(v) Μερικές εμπειρίες μοιάζουν μεταξύ τους. Το χρώμα μπλε είναι διαφορετικό από το κόκκινο, αλλά σίγουρα είναι πολύ πιο διαφορετικό από ένα τετράεδρο σχήμα. Στο πλαίσιο της IIT, στον Q τα χρώματα αντιστοιχούν σε διαφορετικά υποσχήματα ενός όμοιου σχήματος, όπως και τα σχήματα.

(vi) Οι εμπειρίες μπορούν να «εκλεπτυστούν» μέσω της μάθησης και αλλαγών στις συνδέσεις. Έστω ότι κάποιος μαθαίνει αρχικά να ξεχωρίζει το νερό από το κρασί, στη συνέχεια τα κόκκινα κρασιά από τα λευκά και εντέλει αναγνωρίζει και συγκεκριμένες ποικιλίες κρασιού. Πιθανώς, πίσω από αυτή τη φαινομενολογική εξέλιξη υπάρχει και μια νευροβιολογική εξέλιξη. Νευρώνες που αρχικά συνδέονταν αδιακρίτως στις ίδιες εισόδους γίνονται πιο ειδικοί και χωρίζονται σε υποομάδες με μερικώς ξεχωριστές εισόδους. Με αυτή τη διαδικασία αυξάνεται ο αριθμός και αλλάζουν οι κατευθύνσεις των q-διανυσμάτων μεταβάλλοντας και το σχήμα του quale.

(vii) Οι στοιχειώδεις φόρμες βρίσκονται στην «κορυφή» της εμπειρίας. Έστω ότι κάποιος μπορεί να δει το χρώμα κόκκινο κανονικά. Έρευνες έχουν αποδείξει ότι υπεύθυνη για την κατανόηση των χρωμάτων είναι μια ομάδα νευρώνων στην περιοχή V8 του εγκεφάλου. Δυσλειτουργία στη συγκεκριμένη περιοχή μπορεί να αποτρέψει το άτομο να έχει αντίληψη του κόκκινου. Αυτό σημαίνει ότι πλέον δεν μπορεί να το δει, να το φανταστεί, να το θυμηθεί ή να το ονειρευτεί, μπορεί απλώς να μιλάει για αυτό, όπως εμείς μιλάμε για τον τρόπο με τον οποίο προσανατολίζεται μια νυχτερίδα. Κατά τα υπόλοιπα βέβαια το άτομο έχει πλήρως τη συνείδησή του. Αντίθετα μοιάζει απίθανο για έναν ασθενή που βρίσκεται σε κατάσταση φυτού, ακόμα και αν υπάρχει ενεργητικότητα στην περιοχή V8 του εγκεφάλου του ενώ κατά τα άλλα είναι αναίσθητος, να μπορεί να αντιληφθεί τα χρώματα όπως εμείς.



Σχήμα 7: (A): Το σύστημα του Σχήματος 5. (B) Οι q-γωνίες που καταλήγουν στο σημείο ελάχιστης διαμέρισης πληροφορίας (MIP) σχηματίζουν μια βάση πάνω στην οποία ξεδιπλώνεται το υπόλοιπο σύμπλεγμα. Τα q-διανύσματα εκτός της βάσης είναι αυτά που οφείλονται στις πληροφοριακές σχέσεις μεταξύ των ελάχιστων μερών του συστήματος, ενώ η βάση προκύπτει από τις πληροφοριακές σχέσεις καθαρά στο εσωτερικό αυτών. Το μεγάλο γκρι βέλος μας δίνει τη συνολική πληροφορία που δημιουργεί όλο το σύστημα. (C): Το σύστημα του Σχήματος 2A. Το quale που προκύπτει από δύο φωτοδιόδους που τις θεωρούμε ένα σύστημα. Αφού όμως το σύστημα αποτελείται από δύο ανεξάρτητα μεταξύ τους μέρη δεν θεωρείται οντότητα. (D): Σε αυτή την περίπτωση η κατάληξη του quale είναι το MIP: Δεν ξεδιπλώνεται αλλά μένει στη βάση του, αφού δεν έχουμε πληροφοριακές σχέσεις μεταξύ των μερών του συστήματος [5].

2.6 Τα συμπεράσματα και τα προβλήματα της IIT

Ανακεφαλαιώνοντας, η IIT ισχυρίζεται ότι η ποσότητα της συνείδησης δίνεται από την ενσωματωμένη πληροφορία (Φ) που παράγεται από ένα σύμπλεγμα αλληλοεπιδρώντων στοιχείων, ενώ η ποιότητα της από το σχήμα στον χώρο Q που προκύπτει από τις πληροφοριακές σχέσεις μεταξύ αυτών. Μάλιστα αυτό το μέχρι τώρα θεωρητικό πλαίσιο είναι σύμφωνο με τις μέχρι τώρα παρατηρήσεις της νευροβιολογίας και της νευροφυσιολογίας και ουσιαστικά επεκτείνει τη γλώσσα της φαινομενολογίας σε γλώσσα μαθηματικών.

Το γεγονός αυτό έδωσε τη δυνατότητα στη θεωρία να ελεγχθεί και σε ένα περιβάλλον σχετικά με την εξέλιξη μικρών, προσαρμοστικών κυκλωμάτων, αποτελούμενα από λογικές πύλες και καλούμενα να αντιμετωπίσουν μπλοκ διαφορετικού μεγέθους, είτε να τα πιάσουν είτε να τα

αποφύγουν, σε ένα παιχνίδι τύπου Τέτρις. Για τη λύση αυτού του προβλήματος απαιτείται ενσωμάτωση των πληροφοριών εισόδου και μνήμη. Τα ανεπτυγμένα συστήματα αξιολογήθηκαν με βάση το σύνολο της ενσωματωμένης πληροφορίας που επεξεργάστηκαν-παρήγαν καθώς και τον αριθμό των συμπλεγμάτων που αξιοποιούσαν. Τελικά το αποτέλεσμα ήταν πως κατά τη διάρκεια του παιχνιδιού η ενσωματωμένη πληροφορία αυξανόταν, τα συμπλέγματα αυξάνονταν και όλα αυτά είχαν εξάρτηση από την πολυπλοκότητα του παιχνιδιού και κυρίως τις απαιτήσεις του σε μνήμη. Αυτό που ουσιαστικά έδειξαν τα αποτελέσματα ήταν ότι για ένα περιβάλλον με συγκεκριμένες εισόδους και εσωτερικό μηχανισμό γίνεται να αναπτυχθούν υψηλά ενσωματωμένα δίκτυα («εγκεφάλους») με πολλά συμπλέγματα, οδηγώντας ταυτόχρονα στην αύξηση της εσωτερικής τους πολυπλοκότητας [7].

Αυτό εντέλει που κάνει η IIT είναι να παίρνει τα αυταπόδεικτα χαρακτηριστικά της εμπειρίας ως αξιώματα και να τα μεταφράζει σε μια μορφή υπολογιστική και εν συνεχεία αυτή να την αντιστοιχεί στη δομή του εγκεφάλου. Το πρώτο αξίωμα είναι πως η εμπειρία υπάρχει εσωτερικά, δηλαδή για κάθε σύστημα-εγκέφαλο η κάθε εμπειρία αντιμετωπίζεται εσωτερικά με έναν τρόπο που ισχύει μόνο για εκείνο. Το δεύτερο είναι το αξίωμα της συγκρότησης που λέει πως η κάθε εμπειρία είναι δομημένη από διακριτά μέρη που την αποτελούν. Με αυτά αναπτύσσει ο εγκέφαλος μια σχέση αιτίου-αποτελέσματος, που μεταφράζεται στα διάφορα ρεπερτόρια και quale που παράγει. Το τρίτο αξίωμα είναι αυτό που λέει ότι η εμπειρία αντιμετωπίζεται ως σύνολο και όχι ως κομμάτια εμπειριών, ενώ το τέταρτο είναι πως είναι απεριόριστη στο χωροχρόνο. Αυτό σημαίνει πως κοιτάζοντας διαρκώς ένα σταθερό τοπίο συνεχώς δεχόμαστε πληροφορία, απλά είναι σταθερά η ίδια.

Φυσικά όπως σημειώθηκε στην αρχή η IIT ακόμα είναι μια θεωρία και σίγουρα δε γίνεται αποδεκτή από όλους. Υπάρχουν αρκετοί που την αμφισβητούν θεωρώντας πως πολλοί από τους ισχυρισμούς της δεν έχουν βάση. Για παράδειγμα δε γίνεται δεκτό πως η συνείδηση προκύπτει από την απόρριψη εναλλακτικών καταστάσεων. Χρειάζονται αποδείξεις για κάτι το οποίο η IIT χρησιμοποιεί ως βάση και μάλιστα αυταπόδεικτη. Το δεύτερο μεγάλο κενό βρίσκεται στον χώρο των qualia, καθώς εφόσον η δουλειά των νευρώνων μπορεί να αναπαρασταθεί απλώς από διανύσματα (όσο πολύπλοκα κι αν είναι αυτά) με βάση το αν δίνουν σήμα ή όχι, τότε γιατί να μη μπορούμε να αντικαταστήσουμε έναν νευρώνα με ένα νανορομπότ; Προφανώς κάτι τέτοιο όχι μόνο δεν είναι απλό, αλλά κατά πάσα πιθανότητα αδύνατο να συμβεί.

Τελικά αυτή είναι κατά βάση η Θεωρία Ενσωματωμένων Πληροφοριών, την οποία και δεν μπορούμε να κρίνουμε επιστημονικά, αλλά σίγουρα μπορούμε να αναγνωρίσουμε ότι η δουλειά που έχει γίνει είναι σοβαρή. Σίγουρα χρειάζεται πολύ περισσότερη, όμως μας φέρνει ένα βήμα πιο κοντά στον κόσμο της συνείδησης.

3. Anil Seth και Αιτιώδης Πυκνότητα

3.1 Εισαγωγή

Το σκεπτικό του Anil Seth ξεκινάει από το ότι υπάρχουν ορισμένοι βασικοί τρόποι να μελετήσουμε τα συνειδητά συμβάντα στον εγκέφαλο. Ένας είναι να συγκρίνουμε συνειδητές και μη συνειδητές καταστάσεις, όπως για παράδειγμα ενδείξεις κατά τη διάρκεια ενός εγκεφαλικού επεισοδίου. Υπάρχει επίσης ευρύ επιστημονικό υλικό για τη διαμόρφωση των νευρώνων και τη δραστηριοποίηση στον εγκεφαλικό φλοιό κατά τη διάρκεια του περπατήματος, του ύπνου με κίνηση ή όχι των ματιών, της κωματώδους κατάστασης και της γενικής αναισθησίας. Τα δεδομένα σε συνειδητή κατάσταση συνήθως προκύπτουν εθελοντικά από ανθρώπους που καλούνται να απαντήσουν σε απλά ερωτήματα. Τέτοια είναι συνήθως η περιγραφή ενός απλού γεγονότος, όπως μια κούπα με καφέ ή μια περιστρεφόμενη καρέκλα. Τέτοια συμβάντα είναι εύκολο να αναφερθούν με ακρίβεια υπό τις κατάλληλες συνθήκες, δηλαδή άμεση περιγραφή χωρίς περισπασμούς.

Μάλιστα αυτού του είδους τα πειράματα είναι που έφεραν ένα κύμα αντίστοιχων ερευνών καθώς η σύγκριση τους έδειξε να υπάρχουν ορισμένες φυσιολογικές ομοιότητες σε συνειδητή και ασυνείδητη κατάσταση. Έτσι με τον καιρό άρχισε να εξασθενεί η ιδέα ότι δεν μπορούμε να αποκτήσουμε επιστημονικές γνώσεις σχετικά με τη συνείδηση και προέκυψε η ανάγκη για θεωρητικά πλαίσια και νευροεπιστημονικά μοντέλα ικανά να στηρίξουν τα πειραματικά δεδομένα, καθώς και να οδηγήσουν σε επιπλέον μελλοντικά πειράματα.

Στόχος λοιπόν του Anil Seth είναι για την επιστήμη της συνείδησης να αναπτύξει και να εξετάσει το τι μπορεί να ονομαστεί «επεξηγηματικές συσχετίσεις»: ποιες νευρολογικές διαδικασίες δηλαδή όχι μόνο σχετίζονται, αλλά ουσιαστικά αποτελούν θεμελιώδη χαρακτηριστικά της συνειδητής εμπειρίας. Ένα τέτοιο χαρακτηριστικό είναι ότι οι συνειδητές σκηνές είναι ταυτόχρονα ολοκληρωμένες (γίνονται κατανοητές ως ένα σύνολο) και διακρινόμενες (αποτελούνται από πολλά διαφορετικά είδη εμπειριών) και μαζί αποτελούν τη δυναμική πολυπλοκότητα της συνείδησης. Το να βρεθεί ένα μέτρο που να μπορεί να μετρήσει και τα δύο αυτά χαρακτηριστικά μέσω της δυναμικής και της λειτουργίας των νευρώνων θα δώσει και τη βάση για να μετρήσουμε τη συνείδηση. Έχουν ήδη αναπτυχθεί τέτοιου είδους μέτρα κυρίως όμως που κατηγοριοποιούν το επίπεδο συνείδησης σε μια κλίμακα (με τα άκρα να είναι κατάσταση θανάτου ή κώματος και η κατάσταση πλήρους συνείδησης [8].

Ένα, λοιπόν, από τα μέτρα για τη νευρική πολυπλοκότητα είναι αυτό που χρησιμοποιεί την αιτιώδη πυκνότητα (causal density-CD). Για τις διάφορες μαθηματικές σχέσεις που θα χρησιμοποιήσουμε παρακάτω σημειώνουμε ότι γράμματα σε **bold** γενικά αναπαριστούν διανύσματα ενώ τα κεφαλαία πίνακες ή τυχαίες μεταβλητές. Όλα τα διανύσματα θεωρούμε πως είναι κάθετα. Με το συμβολισμό $\oplus$ θα συμβολίζουμε την κάθετη σύζευξη δύο διανυσμάτων. Έτσι για $\mathbf{x}=(x_1,\ldots,x_n)^T$ και $\mathbf{y}=(y_1,\ldots,y_m)^T$ τότε $\mathbf{x}\oplus\mathbf{y}$ δίνει το διάνυσμα $(x_1,\ldots,x_n,y_1,\ldots,y_m)^T$. Επιπλέον για τυχαία διανύσματα $\mathbf{X}$ και $\mathbf{Y}$ θα συμβολίζουμε $\Sigma(\mathbf{X})$ τον $n \times n$ πίνακα $\mathrm{cov}(X_i,X_j)$ και με $\Sigma(\mathbf{X}, \mathbf{Y})$ τον $n \times m$ πίνακα $\mathrm{cov}(X_i,Y_\alpha)$. Ακόμα θα κάνουμε χρήση της ποσότητας:

$$\Sigma(\mathbf{X} \mid \mathbf{Y}) =: \Sigma(\mathbf{X}) - \Sigma(\mathbf{X}, \mathbf{Y})\Sigma(\mathbf{Y})^{-1}\Sigma(\mathbf{X}, \mathbf{Y})^{T}.$$

Αυτή την ποσότητα θα την καλούμε συνδιακύμανση του **X** δεδομένου **Y**. Αν το $\boldsymbol{X}_{\mathrm{t}}$ είναι ένα τυχαίο διάνυσμα σε διακριτό χρόνο χρησιμοποιούμε το ανάπτυγμα $\boldsymbol{X}_{\mathrm{t}}^{(p)}$=$\boldsymbol{X}_t \oplus \boldsymbol{X}_{t-1} \oplus \ldots \oplus \boldsymbol{X}_{t-p+1}$ για να δηλώσουμε το ίδιο το $\boldsymbol{X}_t$, μαζί με p-1 καθυστερήσεις. Δεδομένης μιας καθυστέρησης p, συνήθως χρησιμοποιούμε το συμβολισμό $\boldsymbol{X}_{\mathrm{t}}^{-} \equiv \boldsymbol{X}_{\mathrm{t}-1}^{(p)}$ για την καθυστερημένη μεταβλητή, εφόσον δε δημιουργείται κάποια σύγχυση.

3.2 Αιτιώδης πυκνότητα

Η αιτιώδης πυκνότητα είναι ένα μέτρο για τη συνολική αιτιατή διαδραστικότητα που υφίσταται σε ένα σύστημα. Είναι ουσιαστικά το όργανο της αιτιότητας του Γκρέιντζερ (G-αιτιότητα), που αποτελεί το μέτρο της αιτιατής επιρροής βασισμένη σε χρονολογικά συμπεράσματα. Έτσι, δεδομένων δύο μεταβλητών X και Y, το X είναι G-αιτιατό ως προς το Y, αν από στατιστικής άποψης, το Y βοηθάει στην πρόβλεψη της μελλοντικής κατάστασης του X σε μεγαλύτερο βαθμό από ότι το X μπορεί να προβλέψει το δικό του μέλλον. Πιο γενικά, το X είναι G-αιτιατό ως προς το Y, δεδομένης κατάστασης Z, όπου το X εξαρτάται από το Z, αν το Y βοηθάει στην πρόβλεψη της μελλοντικής κατάστασης του X περισσότερο από ότι το X και το Z μαζί προβλέπουν τη μελλοντική κατάσταση του X [8].

Δεδομένων χρονολογικών γεγονότων, η G-αιτιότητα τυπικά εφαρμόζεται χρησιμοποιώντας το πλαίσιο της γραμμικής αυτοοπισθοδρόμησης. Για να μετρήσουμε τη G-αιτιότητα του Y (προβλέπουσα μεταβλητή) στο X (προβλεπόμενη μεταβλητή) δεδομένου Z (μεταβλητή εξάρτησης), συγκρίνουμε τις παρακάτω αυτοοπισθοδρομικές σχέσεις:

$$X_t = \mathrm{A}(X_{\mathrm{t}-1}^{(p)} \oplus Z_{\mathrm{t}-1}^{(r)}) + \boldsymbol{\varepsilon}_t$$

και $$X_t = A'(X_{\mathrm{t}-1}^{(p)} \oplus Y_{t-1}^{(q)} \oplus Z_{\mathrm{t}-1}^{(r)}) + \boldsymbol{\varepsilon}'_t \qquad (3.1)$$

Έτσι, η προβλεπόμενη μεταβλητή X οπισθοδρομείται αρχικά στις προηγούμενες p καταστάσεις της συν r καθυστερήσεις της μεταβλητής εξάρτησης Z, και στη συνέχεια δέχεται επιπλέον q καθυστερήσεις από την προβλέπουσα μεταβλητή Y (τα p, q και r μπορούν να επιλεχθούν με βάση το Μπεϋζιανό κριτήριο πληροφορίας). Το μέγεθος της G-αιτιότητας (αιτιατή αλληλεπίδραση) δίνεται από το λογάριθμο του λόγου των υπολειπόμενων διακυμάνσεων:

$$F_{Y\to X|Z} =: \ln\left(\frac{var(\varepsilon_t)}{var(\varepsilon'_t)}\right) = \ln\left(\frac{\Sigma(\varepsilon_t)}{\Sigma(\varepsilon'_t)}\right) = \ln\left(\frac{\Sigma(X|\boldsymbol{X}^{-}\oplus\boldsymbol{Z}^{-})}{\Sigma(X|\boldsymbol{X}^{-}\oplus\boldsymbol{Y}^{-}\oplus\boldsymbol{Z}^{-})}\right), \qquad (3.2)$$

όπου ο τελευταίος όρος εκφράζει την G-αιτιότητα σε μεγέθη μερικών συνδιακυμάνσεων. Δεδομένου ενός σετ τιμών G-αιτιότητας μεταξύ των στοιχείων ενός συστήματος **X**, μια απλή μορφή της αιτιώδης πυκνότητας (CD) μπορεί να οριστεί ως ο μέσος όρος όλων των μερικών G-αιτιοτήτων των στοιχείων:

$$\mathrm{CD}(\mathbf{X}) =: \frac{1}{n(n-1)} \sum_{i\neq j} F_{X_i\to X_j|\boldsymbol{X}_{[ij]}}, \qquad (3.3)$$

όπου το $\boldsymbol{X}_{[ij]}$ δηλώνει το υποσύστημα του **X** με τις μεταβλητές X_i και X_j να παραλείπονται, ενώ το n είναι ο συνολικός αριθμός των μεταβλητών. Η αιτιώδης πυκνότητα παρέχει ένα ιεραρχημένο

μέτρο για τη δυναμική πολυπλοκότητα του συστήματος, κατά το οποίο αν τα στοιχεία του είναι εντελώς ανεξάρτητα θα δώσει αποτέλεσμα μηδέν, όπως και αν οι αλληλεπιδράσεις μεταξύ τους δεν επηρεάζουν την κατάσταση τους. Υψηλές τιμές θα επιτευχθούν μόνο όταν τα στοιχεία συμπεριφέρονται διαφορετικά το ένα με το άλλο, με τρόπο ώστε να προσφέρουν κάποια πιθανή πληροφορία πρόβλεψης και η αλληλεπίδραση τους είναι τέτοια, ώστε αυτή η πληροφορία να καθίσταται πρακτικά χρήσιμη.

Όπως και τα περισσότερα μέτρα που βασίζονται σε χρονολογικά γεγονότα, η G-αιτιότητα κανονικά εκτιμάται μεταξύ μονοδιάστατων μεταβλητών, που πιθανόν εξαρτόνται από μια ομάδα άλλων μεταβλητών. Ωστόσο, οι σχετικές αιτιατές αλληλεπιδράσεις σε ένα σύστημα ίσως λαμβάνουν χώρα μεταξύ των γκρουπ των μεταβλητών. Για παράδειγμα, στα νευρολογικά συστήματα, κάποιος μπορεί να επιθυμεί να εξετάσει τις αιτιατές αλληλεπιδράσεις μεταξύ ενός συνόλου νευρώνων, ή, σε μακροσκοπικό επίπεδο, τις αλληλεπιδράσεις μεταξύ δικτύων νευρώνων σε περιοχές ενδιαφέροντος που συνεισφέρουν στον εγκέφαλο. Πιο γενικά, οι μετρούμενες μεταβλητές περιορίζονται από μεθόδους απόκτησης δεδομένων και δε χρειάζεται να υπάρχει ξεκάθαρος σχεδιασμός της αποσύνθεσης του υπό εξέταση συστήματος.

Ευτυχώς είναι εύκολο να επεκτείνουμε τη μέτρηση της G-αιτιότητας από απλώς μονοδιάστατες μεταβλητές (X, Y, Z) σε σετ-πίνακες μεταβλητών (**X**, **Y**, **Z**). Ορίζουμε λοιπόν τη νέα G-αιτιότητα ως εξής:

$$F_{\boldsymbol{X}\to\boldsymbol{X}|\boldsymbol{Z}} =: \ln\left(\frac{|\Sigma(\varepsilon_t)|}{|\Sigma(\varepsilon'_t)|}\right) = \ln\left(\frac{|\Sigma(\boldsymbol{X}|\boldsymbol{X}^- \oplus \boldsymbol{Z}^-)|}{|\Sigma(\boldsymbol{X}|\boldsymbol{X}^- \oplus \boldsymbol{Y}^- \oplus \boldsymbol{Z}^-)|}\right), \quad (3.4)$$

όπου το | * | αντιπροσωπεύει την ορίζουσα του πίνακα και το |Σ(ε)| είναι η γενική διακύμανση της υπολειπόμενης διακύμανσης του πίνακα Σ(ε), η οποία ποσοτικοποιεί τον όγκο των υπολειμμάτων. Με τον τρόπο αυτό αποκτούμε ορισμένα πλεονεκτήματα. Εν συντομία, ο σχηματισμός της ορίζουσας αντιστοιχεί στην εντροπία μεταφοράς σύμφωνα με Γκαουσιανές υποθέσεις, και μπορεί να χρησιμοποιηθεί ως άθροισμα κλασικών μονοδιάστατων G-αιτιοτήτων.

Είναι ουσιαστικά μια επέκταση για τη CD στην οποία οι G-αιτιώδεις αλληλεπιδράσεις αξιολογούνται μεταξύ διαμερίσεων του συστήματος. Για ένα σύστημα **X**, ορίζουμε το $CD_{k\to r}(\mathbf{X})$, ως το μέσο όρο της παραπάνω ποσότητας για ένα υποσύστημα μεγέθους k έως ένα υποσύστημα μεγέθους r, δεδομένου του υπόλοιπου συστήματος.

$$CD_{k\to r}(\mathbf{X}) =: \frac{1}{n_{k,r}} \sum_{i=1}^{n_{k,r}} F_{\boldsymbol{V}_i^k \to \boldsymbol{U}_i^r | \boldsymbol{W}_i^{n-k-r}}, \quad (3.5)$$

Όπου το $\mathbf{X} = \boldsymbol{V}_i^k \cup \boldsymbol{U}_i^r \cup \boldsymbol{W}_i^{n-k-r}$ δηλώνει την i-οστή από τις $n_{k,r} = \binom{n}{k}\binom{n-k}{r}$ διακριτές τριμερείς διαμερίσεις του **X** σε ξεχωριστά υποσυστήματα μεγέθους k, r και (n – k - r) αντίστοιχα. Έτσι προκύπτει το BCD ως ο μέσος όρος του $CD_{k\to(n-k)}(\mathbf{X})$ με προβλέπτη μεγέθους k:

$$\mathrm{BCD}(\mathbf{X}) =: \frac{1}{n-1} \sum_{k=1}^{n-1} CD_{k\to(n-k)}(\boldsymbol{X}). \quad (3.6)$$

Αυτή η ποσότητα μπορεί να παρέχει ένα περισσότερο ιεραρχημένο μέτρο δυναμικής πολυπλοκότητας από το CD ως προς την ανάλυση ενός συστήματος σε πολλαπλές κλίμακες. Όπως θα εξηγηθεί στη συνέχεια, είναι στενά συνδεδεμένη με τη γνωστή «νευρική πολυπλοκότητα», τις κοινές δηλαδή πληροφορίες μεταξύ διαμερίσεων σε συστήματα νευρώνων.

Σε αυτό το σημείο είναι σημαντικό να προσθέσουμε δύο επιπλέον παρατηρήσεις σχετικά με την αιτιώδη πυκνότητα. Αρχικά, τα CD και BCD μπορούν να εκτιμηθούν εντός μιας συγκεκριμένης ζώνης συχνοτήτων, κάτι που μπορεί να αποδειχθεί χρήσιμο σε περιπτώσεις που αυτά τα όρια έχουν και συγκεκριμένες νευροφυσιολογικές ερμηνείες. Δεύτερον, σε κάθε πολύπλοκο σύστημα, είναι συνήθως εφικτό να μελετήσουμε μόνο ένα υποσύνολο των σχετικών μεταβλητών, κάτι που μπορεί να οδηγήσει σε στρεβλωμένα αιτιατά συμπεράσματα, ενώ προκύπτουν από συνήθεις αιτίες. Μία προσέγγιση σε αυτό το πρόβλημα είναι να «αγνοήσουμε» τις κρυμμένες επιρροές με το να προσθέσουμε έναν επιπλέον όρο στις εξισώσεις του Granger που είναι ευαίσθητος στις συσχετίσεις μεταξύ των υπολειμμάτων.
Μια συνήθης κριτική της G-αιτιότητας είναι ότι η τυπική της εφαρμογή σε γραμμικά μοντέλα αυτοοπισθοδρόμησης προφανώς εξαιρεί την ισχύ της για μη γραμμικές αλληλεπιδράσεις. Ωστόσο η μη γραμμική επέκταση της G-αιτιότητας υπάρχει, όμως είναι αρκετά πολύπλοκη και δύσκολο να εφαρμοστεί στην πράξη. Ένα εναλλακτικό πλαίσιο παρέχεται μέσω της εντροπίας μεταφοράς, που αποτελεί ένα μέτρο της διακριτής πληροφορίας που μεταφέρεται μεταξύ των διαμερίσεων και των συστημάτων. Η εντροπία μεταφοράς *T* ορίζεται από τη διαφορά στις εντροπίες:

$$T_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}} =: \mathrm{H}(\mathbf{X}|\, \boldsymbol{X}^{-} \oplus \boldsymbol{Z}^{-}) - \mathrm{H}(\mathbf{X}|\, \boldsymbol{X}^{-} \oplus \boldsymbol{Y}^{-} \oplus \boldsymbol{Z}^{-}), \qquad (3.7)$$

και ποσοτικοποιεί, με έναν φυσικό μη γραμμικό τρόπο, το βαθμό στον οποίο η γνώση των προηγούμενων καταστάσεων του **Y** μειώνει την αβεβαιότητα για τη μελλοντική κατάσταση του **X**, δεδομένου του **Z**.
Αν και εδώ και καιρό έχει αναγνωριστεί η συσχέτιση της G-αιτιότητας και της εντροπίας μεταφοράς, μόλις πρόσφατα επισημοποιήθηκε μια μεταξύ τους μαθηματική ισότητα. Έτσι, για Γκαουσιανές, προκύπτει ότι η μεταξύ τους σχέση είναι απλή, και συγκεκριμένα έχει τη μορφή:

$$F_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}} = 2T_{\boldsymbol{Y}\to\boldsymbol{X}|\boldsymbol{Z}}. \qquad (3.8)$$

Η ισότητα (3.8) βασίζεται στο ότι οι σχέσεις μεταξύ σχετικής εντροπίας, μερικής συνδιακύμανσης και πρόβλεψης γραμμικής αυτοοπισθοδρόμησης οδηγούν πάντα σε σφάλμα. Οι ουσιαστικές σχέσεις είναι οι ακόλουθες. Αρχικά, για Γκαουσιανές μεταβλητές, η σχετική εντροπία είναι μια συνάρτηση της ορίζουσας του αντίστοιχου πίνακα μερικής συνδιακύμανσης:

$$\mathrm{H}(\mathbf{X}|\, \mathbf{Y}) = \frac{1}{2}\ln\left(|\Sigma(\mathbf{X}|\, \mathbf{Y})|\right) + \frac{1}{2}n\ln(2\pi\, \mathrm{e}), \qquad (3.9)$$

όπου n είναι η διάσταση του **X**. Δεύτερον, η σχετική συνδιακύμανση του **X** δεδομένου **Y** είναι ακριβώς ο πίνακας συνδιακύμανσης των υπολειμμάτων μιας γραμμικής οπισθοδρόμησης του **X** στο **Y**:

$$\Sigma(\boldsymbol{\varepsilon}) \equiv \Sigma(\mathbf{X}|\, \mathbf{Y}). \qquad (3.10)$$

Αξίζει να σημειωθεί ότι η ισότητα (3.10) ισχύει για οποιαδήποτε (στατικά) **X** και **Y**, είτε Γκαουσιανά είτε όχι. Μαζί, αυτές οι εκφράσεις επιτρέπουν στο *T* να γραφτεί με όρους γραμμικής οπισθοδρόμησης και έτσι να συσχετιστεί άμεσα με το *F*.

Η ισότητα μεταξύ *F* και *T* είναι σημαντική γιατί υποδεικνύει ότι, για Γκαουσιανές μεταβλητές, η γραμμική οπισθοδρόμηση αντιπροσωπεύει όλες τις συσχετίσεις μεταξύ των μεταβλητών, δικαιολογώντας ταυτόχρονα το CD ως μέτρο δυναμικής πολυπλοκότητας.
Εν συνεχεία, η νευρική πολυπλοκότητα *C* ενός συστήματος **X** με n στοιχεία δίνεται από τον τύπο:

$$C(\mathbf{X}) =: \sum_{k=1}^{n-1}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} H\left(\boldsymbol{U}_k^j\right) - \frac{k}{n}H(\boldsymbol{X})\right), \qquad (3.11)$$

όπου το $\boldsymbol{U}_k^j$ είναι η κατάσταση του j-οστού υποσυστήματος με k στοιχεία και $n_k=\binom{n}{k}$. Αυτό το μέτρο ποσοτικοποιεί το εύρος στο οποίο η εντροπία των υποσυστημάτων είναι μεγαλύτερη από την κανονικοποιημένη εντροπία του συνόλου, όπου κανονικοποίηση είναι ο λόγος του μεγέθους του υποσυστήματος προς το μέγεθος του συνόλου. Οι προσδοκώμενες διαφορές για κάθε μέγεθος υποσυστήματος προστίθενται.
Μια σημαντική διαφορά μεταξύ του *C* και της αιτιώδους πυκνότητας είναι ότι το *C* το αφορούν μόνο οι στατικές συνεισφορές των καταστάσεων του συστήματος και των μερών του, ενώ η αιτιώδης πυκνότητα ενδιαφέρεται στο να προβλέψει το παρόν και το μέλλον ενός συστήματος βασιζόμενη στο παρελθόν του. Ωστόσο, για Γκαουσιανές μεταβλητές, είναι δυνατό να δειχθεί ότι το BCD είναι ισοδύναμο με μια τροποποιημένη μορφή των *C* και *C'*, κατά την οποία οι εντροπίες αντικαθιστούνται από τις σχετικές εντροπίες των καταστάσεων στο παρών δοθέντων των προηγούμενων καταστάσεων. Έτσι έχουμε:

$$C'(\mathbf{X}) =: \sum_{k=1}^{n-1}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} H\left(\boldsymbol{U}_k^j \middle| \left(\boldsymbol{U}_k^j\right)^-\right) - \frac{k}{n}H(\boldsymbol{X}|\,\boldsymbol{X}^-)\right). \qquad (3.12)$$

Για να δούμε την ισότητα, επαναπροσδιορίζουμε το *C'* με όρους διαμέρισης του **X**. Σημειώνουμε τις διαμερίσεις του **X** ως $(\boldsymbol{U}_k^j, \boldsymbol{V}_k^j)$ και αποτελούν την j-οστή διαμέριση με το μικρότερο στοιχείο $\boldsymbol{U}_k^j$ να αποτελείται από k στοιχεία. Έτσι έχουμε:

$$C'(\mathbf{X}) =: \sum_{k=1}^{n/2}\left(\frac{1}{n_k}\sum_{j=1}^{n_k}\left(H\left(\boldsymbol{U}_k^j \middle| \left(\boldsymbol{U}_k^j\right)^-\right) + H\left(\boldsymbol{V}_k^j \middle| \left(\boldsymbol{V}_k^j\right)^-\right) - H(\boldsymbol{X}|\,\boldsymbol{X}^-)\right)\right) \qquad (3.13)$$

Υποθέτουμε ότι δεν υπάρχουν κρυμμένα ή εξωγενή στοιχεία που να επηρεάζουν το **X**, ώστε δεδομένου $\boldsymbol{X}^-$ η τωρινή κατάσταση του **X** να είναι ανεξάρτητη και να προκύπτει:

$$\mathrm{H}(\mathbf{X}|\,\boldsymbol{X}^-) = \mathrm{H}(\boldsymbol{U}_k^j|\,\boldsymbol{X}^-) + \mathrm{H}(\boldsymbol{V}_k^j|\,\boldsymbol{X}^-). \qquad (3.14)$$

Πλέον μπορούμε να εκφράσουμε το *C'* συναρτήσει του *T*, και ως εκ τούτου του *F*:

$$C'(\mathbf{X}) =: \sum_{k=1}^{n-1}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} T_{\boldsymbol{V}_k^j \to \boldsymbol{U}_k^j}\right) = \sum_{k=1}^{n-1}\left(\frac{1}{2n_k}\sum_{j=1}^{n_k} F_{\boldsymbol{V}_k^j \to \boldsymbol{U}_k^j}\right), \qquad (3.15)$$

Όπου τώρα $(\boldsymbol{U}_k^j, \boldsymbol{V}_k^j)$ είναι η j-οστή διαμέριση με το $\boldsymbol{U}_k^j$ να αποτελείται από k στοιχεία. Η άμεση ισότητα μεταξύ αιτιώδους πυκνότητας και νευρικής πολυπλοκότητας δίνεται από τον τύπο:

$$\mathrm{BCD}(\mathbf{X}) = \frac{1}{2(n-1)}\,C'(\mathbf{X})$$

3.3 Προσομοιώσεις

Παρουσιάζουμε τα αποτελέσματα από τον υπολογισμό των CD και BCD για κάποια παραδειγματικά Μαρκοβιανά Γκαουσιανά συστήματα. Αυτά τα συστήματα είναι της μορφής:

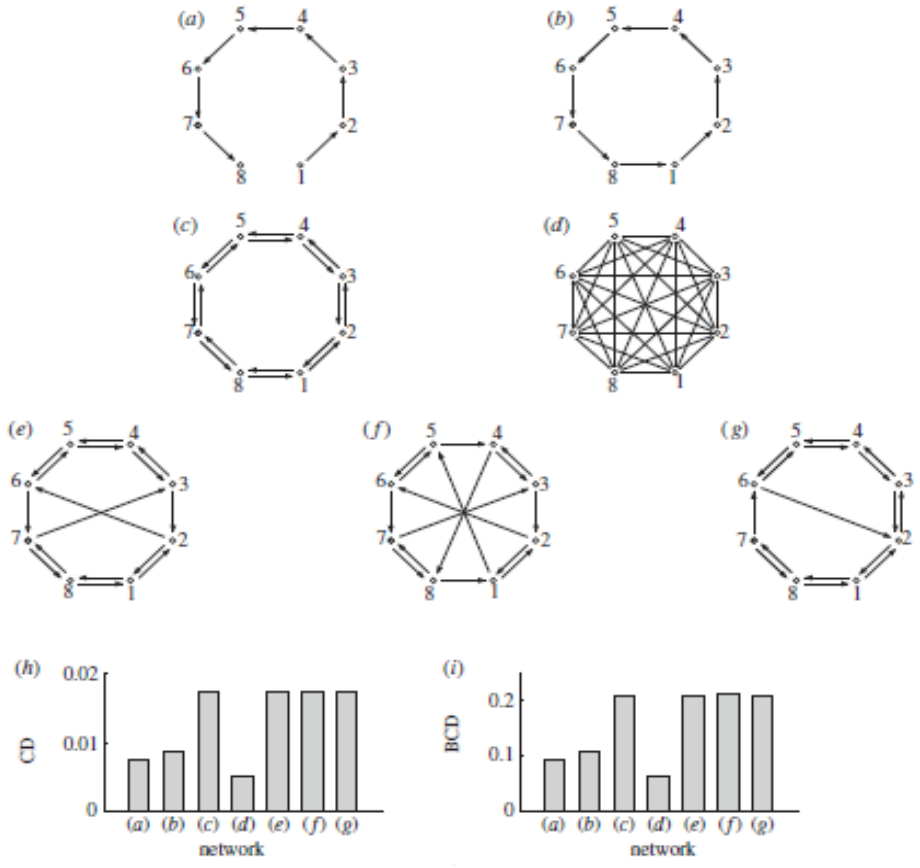
$$\boldsymbol{X}_t = \mathrm{A} \cdot \boldsymbol{X}_{t-1} + \boldsymbol{E}_t$$

Όπου το $\boldsymbol{X}_t$ περιέχει n μεταβλητές, το A είναι πίνακας συνεκτικότητας και κάθε στοιχείο του $\boldsymbol{E}_t$ είναι ανεξάρτητη Γκαουσιανή τυχαία μεταβλητή με μέσο 0 και διακύμανση 1. Θεωρήσαμε εφτά συστήματα με n=8 και συνδεσιμότητα όπως φαίνεται στο **Σχήμα 8** (a-g). Αναφερόμαστε στα συστήματα ως '8a', '8b' και ούτω καθεξής. Οι προκύπτουσες τιμές των CD και BCD δίνονται στο **Σχήμα 8** (h,i). Όλες οι τιμές έχουν υπολογιστεί αναλυτικά.

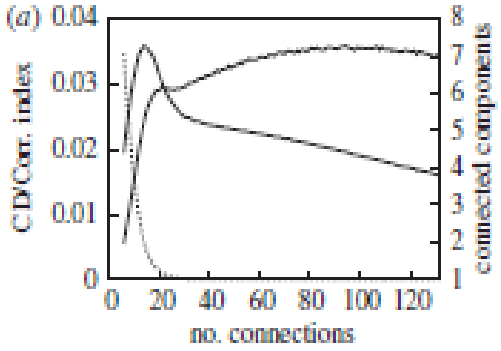
Συνολικά όλες οι τιμές που έχουν υπολογιστεί ανταποκρίνονται στις προσδοκίες μας για αυτά τα παραδείγματα. Ένα σύστημα με αμφίδρομες συνδέσεις (9c) παράγει μεγαλύτερες τιμές από ένα με μονόπλευρες συνδέσεις (9b), το οποίο με τη σειρά του παράγει μεγαλύτερες τιμές από ένα σύστημα με λιγότερες από αυτό μονόπλευρες συνδέσεις (9a). Επιπλέον, όπως περιμέναμε, ένα ομογενές σύστημα (9d) παρέχει χαμηλές τιμές για όλες τις μετρήσεις. Ίσως σε αντίθεση με τις προσδοκίες μας, η δημιουργία σποραδικών «συντομότερων μονοπατιών» σε ένα σύστημα με αμφίδρομες συνδέσεις (9e-g) δεν αυξάνει τις τιμές των μετρήσεων σε σχέση με το αρχικό (9c).

Για να εξετασθεί η τάση της συμπεριφοράς των μετρήσεων ως συνάρτηση της πυκνότητας συνδέσεων, η παρακάτω διαδικασία επαναλήφθηκε 100 φορές. Ξεκινώντας από ένα σύστημα n=8 στοιχείων χωρίς καμία σύνδεση, όπου το η κατάσταση του καθενός επηρεαζόταν μόνο από την εσωτερική του δομή (ισχύς 0.5), συμπληρώθηκε σταδιακά ο πίνακας συνεκτικότητας με τυχαία σειρά. Κάθε φορά που γινόταν προσθήκη μιας σύνδεσης, η ισχύς των συνδέσεων κανονικοποιούταν με τέτοιο τρόπο ώστε (i) η συνολική ισχύς συνδέσεων (μαζί με την εσωτερική δομή) που αντιστοιχεί σε κάθε στοιχείο να είναι σταθερά 0.5 και (ii) όλες οι συνδέσεις ενός στοιχείου να είναι ισοδύναμες.

Για επιβεβαίωση της γενικότητας των παρατηρήσεων από την παραπάνω διαδικασία, αυτή επαναλήφτηκε σε δίκτυο με n=12 στοιχεία και k τυχαία επιλεγμένες συνδέσεις μεταξύ των στοιχείων. Πρέπει να τονίσουμε βέβαια ότι για μικρό k ($k < n/2$), οι συνδέσεις που πραγματοποιήθηκαν έβγαζαν αποτελέσματα πολύ κοντά στο 0 οπότε και απορρίφθηκαν. Έγιναν 20.000 δοκιμές και υπολογισμοί του CD για κάθε k αριθμό συνδέσεων και τα αποτελέσματα φαίνονται στο **Σχήμα 9**, όπου φαίνεται πώς CD και δείκτης συσχέτισης στρέφονται σε σχέση με την πυκνότητα σύνδεσης. Συγκεκριμένα το CD παίρνει τη μεγαλύτερη τιμή του λίγο πριν αρχίσει να πυκνώνει η συνεκτικότητα του δικτύου. Εντέλει συμπεραίνεται ότι όντως το CD επηρεάζεται από τα σημαντικά χαρακτηριστικά της τοπολογίας και της δυναμικής του δικτύου, όπως ακριβώς προτείνεται από το θεωρητικό του πλαίσιο. Συνεπές στη διαίσθηση πάνω στην οποία στηρίχτηκε η πρόταση του, το CD είναι ευαίσθητο στο δυναμικό καθεστώς της μεταξύ των στοιχείων ανεξαρτησίας και της δυνατής συσχέτισης τους.



Σχήμα 8: Μέτρηση CD, BCD για Μαρκοβιανά Γκαουσιανά συστήματα. (a-g): Συνεκτικά διαγράμματα για εφτά συστήματα που ορίζουν και τον αντίστοιχο πίνακα συνεκτικότητας A. Η ισχύς κάθε εισερχόμενης σύνδεσης για τα συστήματα (a-c) και (e-g) είναι 0.25. Για το σύστημα (d) η ισχύς κάθε σύνδεσης είναι 1/14 και συνολικά το άθροισμα το άθροισμα των εισερχόμενων συνδέσεων για κάθε στοιχείο είναι 0.5. (h) CD. (i) BCD [8].



Σχήμα 9: Μέτρηση δυναμικής πολυπλοκότητας ως συνάρτηση της πυκνότητας σύνδεσης σε δίκτυα 12 στοιχείων, με κανονικοποιημένη ισχύ σύνδεσης. Το CD είναι η συνεχόμενη γραμμή που ξεκινάει κοντά στο 0 και η άλλη είναι ο δείκτης συσχέτισης [8].

Σε αυτό το σημείο είναι καλό να ξεκαθαρίσουμε εν συντομία δύο όρους που έχουν χρησιμοποιηθεί μέχρι τώρα. Αρχικά, το Μπεϋζιανό κριτήριο πληροφορίας, αποτελεί ένα στατιστικό μοντέλο επιλογής μοντέλων/παραμέτρων ανάμεσα σε ένα πεπερασμένο σετ μοντέλων/παραμέτρων.

Αναπτύχθηκε από τον Gideon E. Schwarz το 1978 και η λειτουργία του είναι να αυξάνει την πιθανότητα για μια διαδικασία να επιστρέψει σωστό αποτέλεσμα αυξάνοντας τις παραμέτρους της. Ωστόσο όσο περισσότερες παράμετροι προστίθενται τόσο μεγαλύτερη είναι η πιθανή απόκλιση του τελικού αποτελέσματος από το πραγματικό.

Επιπλέον γίνεται συνεχώς λόγος για Μαρκοβιανά Γκαουσιανά συστήματα, τα οποία και εξετάστηκαν. Ένα σύστημα είναι Γκαουσιανό αν και μόνο αν η ποσότητα $\boldsymbol{X}_{t1,\ldots,tk}=(\boldsymbol{X}_{t1},\ldots,\boldsymbol{X}_{tk})$ είναι Γκαουσιανή τυχαία μεταβλητή. Αυτό σημαίνει ότι ο οποιοσδήποτε γραμμικός συνδυασμός των $(\boldsymbol{X}_{t1},\ldots,\boldsymbol{X}_{tk})$ έχει μια ενιαία κανονική συνεισφορά στο σύστημα. Μαρκοβιανό είναι ένα σύστημα για το οποίο υπάρχει η δυνατότητα να γίνει πρόβλεψη για τη μελλοντική του κατάσταση με βάση μόνο την τωρινή του κατάσταση και η πιθανή γνώση του ιστορικού του να είναι τελείως ανεξάρτητη από αυτή τη μελλοντική κατάσταση.

3.4 Γενικά συμπεράσματα

Περιεγράφηκαν τα μέτρα της δυναμικής πολυπλοκότητας με βάση την αιτιώδη πυκνότητα, που είναι εύκολα εφαρμόσιμα σε δεδομένα με χρονολογική σειρά και έχουν θεωρητικά τη δυνατότητα να αντιμετωπίζουν τις συνεκτικές δράσεις διαφοροποιημένων λειτουργικών συστημάτων. Αυτή η βάση ενθαρρύνει τη χρήση τους για τη μέτρηση των επιπέδων συνείδησης, δοθέντων νευρολογικών δεδομένων συγκρίνοντας τις περιπτώσεις κανονικής συνειδητής κατάστασης και κατάστασης χωρίς (ή με μειωμένη) συνείδηση όπως γενική αναισθησία, ύπνος χωρίς όνειρα, κώμα, κατάσταση φυτού, επιληπτικό επεισόδιο ή ακόμα και ψυχιατρική ανωμαλία [9].

Η μέτρηση της αιτιώδους πυκνότητας βασίζεται στην ικανότητα πρόβλεψης μεταξύ γεγονότων σε χρονολογική σειρά, λειτουργώντας με τη χρήση της G-αιτιότητας, και στόχος αυτής της προσπάθειας είναι η σύλληψη της συνολικής διαδραστικότητας που παρατηρείται σε ένα σύστημα για την τέλεση κάποιας ενέργειας. Τα CD και BCD χαρακτηρίζουν αντίστοιχα, (i) τη συνολική αιτιώδη διαδραστικότητα μεταξύ ομάδων από στοιχεία του συστήματος και (ii) τη συνολική αιτιώδη διαδραστικότητα μεταξύ διαμερίσεων όλου του συστήματος.

Παρά την πρόοδο σε θεωρητικό επίπεδο που έχει παρουσιαστεί μέχρι τώρα, η εφαρμογή της αιτιώδους πυκνότητας σε νευρολογικά δεδομένα αποτελεί ακόμα μια μεγάλη πρόκληση για αρκετούς λόγους. Για αριθμητική διευκόλυνση (π.χ. υπολογισμούς άμεσα από εμπειρικές μήτρες συνδιακυμάνσεων) τα δεδομένα πρέπει να έχουν το χαρακτηριστικό της συνεστιακής σταθεροποίησης, δηλαδή οι μέσοι όροι και οι υπόλοιπες διαφορές να μένουν σταθερά μέσα στο χρόνο. Ωστόσο τα νευρολογικά δεδομένα συνοδεύονται συχνά από θορύβους, έχοντας μεγάλη μεταβλητότητα. Μια λύση είναι να γίνεται συνεχώς αφαίρεση των θορύβων εφόσον αυτό είναι εφικτό. Εναλλακτικά, μπορεί κάποιος να μελετήσει δεδομένα για μικρά χρονικά διαστήματα τα οποία τοπικά μπορεί να είναι σταθερά. Για ορισμένους τύπους δεδομένων (μαγνητο/ηλεκτροεγκεφαλογράφημα-M/EEG) χρησιμοποιούνται φίλτρα για την καλύτερη δυνατή αφαίρεση των θορύβων, τα οποία όμως δεν συμβάλουν θετικά στην ακρίβεια των αποτελεσμάτων μας.

Τα νευρικά σήματα είναι συνήθως μη γραμμικά όπως και μη στατικά. Αν και η μέτρηση που έχει προταθεί καλύπτει και περίπτωση μη γραμμικότητας, τα αυτοοπισθοδρομικά συνήθως

υπολογίζονται από γραμμικά μοντέλα. Ωστόσο, όπως δείξαμε η κατανομή μιας Γκαουσιανής κατάστασης ισορροπίας αποδίδει ένα γραμμικό αυτοοπισθοδρομικό μοντέλο, ενώ οι ποσότητες πληροφορίας τυπικά εξαρτώνται από Γκαουσιανές κατανομές ώστε να μπορούν να είναι και πρακτικά υπολογίσιμες. Αυτά τα δύο σημεία μαζί υπογραμμίζουν τη σημασία της Γκαουσιανής υπόθεσης. Ευτυχώς, οι Γκαουσιανές προσεγγίσεις φαίνονται να έχουν επαρκή ισχύ στη νευροεπιστήμη.

Τα νευρολογικά δεδομένα σχετικά με τη συνείδηση τυπικά αποκτούνται από νευρολογικές μεθόδους απεικόνισης όπως τα M/EEG και τη Λειτουργική Απεικόνιση Μαγνητικού Συντονισμού (fMRI). Κάθε μέθοδος θέτει και τις δικές της προκλήσεις. Τα σήματα από M/EEG προσφέρουν υψηλή χρονική ανάλυση που ταιριάζει για την ανάλυση μας σε γεγονότα χρονολογικής σειράς, απαιτούν όμως μια μη μοναδική μοντελοποίηση αντιστροφής για τη μετακίνηση από το χώρο του αισθητήρα (στο κρανίο) σε έναν βαθύτερο χώρο-πηγή του σήματος, προσδίδοντας ασάφεια στη μεταγενέστερη ερμηνεία των αποτελεσμάτων. Η χωρική ανάλυση του EEG είναι επίσης χαμηλή σε σύγκριση με το fMRI και προκύπτουν επιπλέον προκλήσεις σχετικά με τα φίλτρα που πρέπει να χρησιμοποιηθούν. Το fMRI έρχεται σε αντίθεση με τα M/EEG παρέχοντας υψηλή χωρική ανάλυση σε βάρος όμως της χρονικής ανάλυσης. Το σήμα για το επίπεδο του αίματος σε οξυγόνο μετριέται από το fMRI και αντιστοιχεί σε ορισμένες αργές διαδικασίες του μεταβολισμού που σχετίζονται με νευρική δράση (αν και αυτές οι συσχετίσεις δεν έχουν ακόμα κατανοηθεί πλήρως), και το δείγμα είναι τυπικά της τάξης του 0.3-1 Hz. Αυτά λοιπόν τα δείγματα θέτουν προκλήσεις για την ακριβή εκτίμηση των αυτοοπισθοδρομικών μοντέλων ακόμα και για τις μήτρες συνδιακύμανσης. Επιπλέον, η μεταβλητότητα των αιμοδυναμικών αποκρίσεων στις διαφορετικές περιοχές του εγκεφάλου μπορούν να μπερδέψουν τα αιτιώδη αποτελέσματα, υπονομεύοντας μεταγενέστερα συμπληρωματικές μετρήσεις της πολυπλοκότητας. Παρά όλα αυτά, τα αιτιώδη αποτελέσματα βασισμένα σε μετρήσεις με χρονολογική σειρά μέσω fMRI είναι μια περιοχή ιδιαίτερης δραστηριοποίησης της έρευνας και πολλά υποσχόμενη για νέες προσεγγίσεις, όπως για παράδειγμα η ενσωμάτωση των αυτοοπισθοδρομικών μοντέλων σε χωροχρονικά μοντέλα που περιλαμβάνουν αιμοδυναμικές παραμέτρους.

Ο William James περιέγραψε τη συνείδηση ως ένα «ρεύμα» ή μια «διαδικασία». Η αιτιώδης πυκνότητα ως μέτρο δυναμικής πολυπλοκότητας έρχεται σε συμφωνία με αυτή την άποψη. Αυτό συμβαίνει διότι εξαρτάται από στατικά στατιστικά στοιχεία των υποκείμενων νευρικών δυναμικών θεωρώντας ότι η μέτρηση του επιπέδου συνείδησης χαρακτηρίζεται από το ότι (i) το συνειδητό επίπεδο είναι σταθερό κατά τη διάρκεια στατικών στιγμών στη δράση του εγκεφάλου και (ii) το συνειδητό επίπεδο αλλάζει όταν αλλάζουν και οι λειτουργικές συνδέσεις, τροποποιώντας τα στατικά στατιστικά.

Φυσικά ακόμα είναι μεγάλη η σημασία περαιτέρω πειραματισμού αυτού του μοντέλου πάνω στα ενστικτώδη κυρίως, μέχρι τώρα, χαρακτηριστικά του φαινόμενου-στόχου (συνείδηση). Αυτός ο πειραματισμός μπορεί να γίνει με ποικίλους τρόπους για να φέρει αποτελέσματα ή και να οδηγήσει σε ακόμα περισσότερες επιπλοκές ως προς την πρακτική εφαρμογή του ή αμφισβήτηση στις αρχικές θεωρητικές αρχές. Άλλωστε αυτή είναι η μεγαλύτερη πρόκληση μέχρι τώρα, η μείωση του χάσματος ανάμεσα σε θεωρία και πράξη και η εξέταση αν αυτό το θεωρητικό μοντέλο (ή οποιοδήποτε άλλο) μπορεί να ανταποκριθεί και στην πραγματικότητα και όχι μόνο στις δικές του θεωρητικές υποθέσεις και αξίες.

4. Karl Friston και Αρχή Ελεύθερης Ενέργειας

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα γίνει μια ανάλυση της Αρχής Ελεύθερης Ενέργειας (ΑΕΕ – Free Energy Principle) βασικός εκφραστής της οποίας είναι ο Βρετανός νευροεπιστήμονας Karl Friston. Η ΑΕΕ έχει προταθεί με σκοπό να παρέχει μια συνολική θεωρία σχετικά με τον εγκέφαλο, τα ενσωματωμένα δεδομένα και μια θεωρία σχετική με τη δράση, την αντίληψη και τη διαδικασία εκμάθησης του εγκεφάλου. Η θεωρία και οι εφαρμογές της ΑΕΕ συνδυάζουν πληροφορίες από το έργο «Αντίληψη ως τεκμήριο» του Hermann von Helmholtz, τη θεωρία της μηχανικής μάθησης καθώς και τη στατιστική θερμοδυναμική. Έχει γίνει λοιπόν προσπάθεια να εντοπιστούν αρχές σχετικά με τη λειτουργία του εγκεφάλου βασισμένες σε νόμους συντήρησης και στη νευρολογική ενέργεια.

Η Αρχή Ελεύθερης Ενέργειας είναι, λοιπόν, ένα απλό αξίωμα με πολύπλοκα συμπεράσματα. Ισχυρίζεται ότι οποιαδήποτε προσαρμοστική αλλαγή στον εγκέφαλο θα ελαχιστοποιήσει την ελεύθερη ενέργεια. Αυτή η ελαχιστοποίηση μπορεί να γίνει είτε μέσα σε λίγα milliseconds είτε ύστερα από μεγαλύτερο διάστημα. Ουσιαστικά, η αρχή αυτή είναι εφαρμόσιμη σε οποιοδήποτε βιολογικό σύστημα αντιστέκεται στην τάση για αποσύνθεση, δηλαδή από μονοκύτταρους οργανισμούς μέχρι ένα κοινωνικό δίκτυο.

Η ΑΕΕ είναι μια προσπάθεια να εξηγηθεί η δομή και η λειτουργία του εγκεφάλου, ξεκινώντας από το αδιάσειστο δεδομένο ότι υπάρχουμε· ένα δεδομένο που περιορίζει τις αλληλεπιδράσεις μας με τον κόσμο, κάτι που μελετάται για πολλά χρόνια στην εξελικτική βιολογία και στη θεωρία συστημάτων. Ωστόσο, πρόσφατες πρόοδοι στη στατιστική φυσική και στη μηχανική μάθηση δείχνουν ένα απλό σχέδιο που επιτρέπει στα βιολογικά συστήματα να συμμορφώνονται με αυτούς τους περιορισμούς. Αν κάποιος κοιτάξει τον εγκέφαλο ως μία εφαρμογή αυτού του σχεδίου (ελαχιστοποιώντας τους περιορισμούς λόγω διαταραχών), σχεδόν όλες οι πτυχές της ανατομίας και της φυσιολογίας του εγκεφάλου αρχίζουν να βγάζουν νόημα [10].

4.2 Ελεύθερη ενέργεια και αυτο-οργάνωση

Τι είναι, λοιπόν, η ελεύθερη ενέργεια; Ελεύθερη ενέργεια είναι μια ποσότητα θεωρητικής πληροφορίας που οδηγεί στην απόδειξη για ένα μοντέλο δεδομένων. Εδώ, τα δεδομένα είναι αισθητήρες εισόδου και το μοντέλο είναι σχεδιασμένο από τον εγκέφαλο. Πιο συγκεκριμένα, η ελεύθερη ενέργεια είναι μεγαλύτερη από την αρνητική έκβαση ή «έκπληξη» για τα δεδομένα των αισθητήρων, δεδομένου του μοντέλου του πώς αυτά παράχθηκαν. Κρίσιμο είναι, σε αντίθεση με την έκπληξη, ότι η ελεύθερη ενέργεια μπορεί να εκτιμηθεί γιατί αποτελεί μια συνάρτηση των δεδομένων εισόδου των αισθητήρων και των καταστάσεων του εγκεφάλου. Στην ουσία, υπό απλοποιημένες υποθέσεις, είναι απλώς η ποσότητα του σφάλματος πρόβλεψης.

Η βάση της ΑΕΕ είναι απλή αλλά θεμελιώδης. Ξεκινάει από το γεγονός ότι οι αυτο-οργάνωτοι βιολογικοί παράγοντες αντιστέκονται σε οποιαδήποτε τάση προς αταξία και έτσι ελαχιστοποιούν

την εντροπία των καταστάσεων των αισθητήρων τους. Έπειτα από εργοδικές υποθέσεις, η εντροπία προκύπτει ότι είναι:

$$H(\mathrm{y}) = -\int p(y|m) \ln p(y|m)\, dy = \lim_{T\to\infty} \frac{1}{T} \int_0^T -\ln p(y|m) dt \qquad (4.1)$$

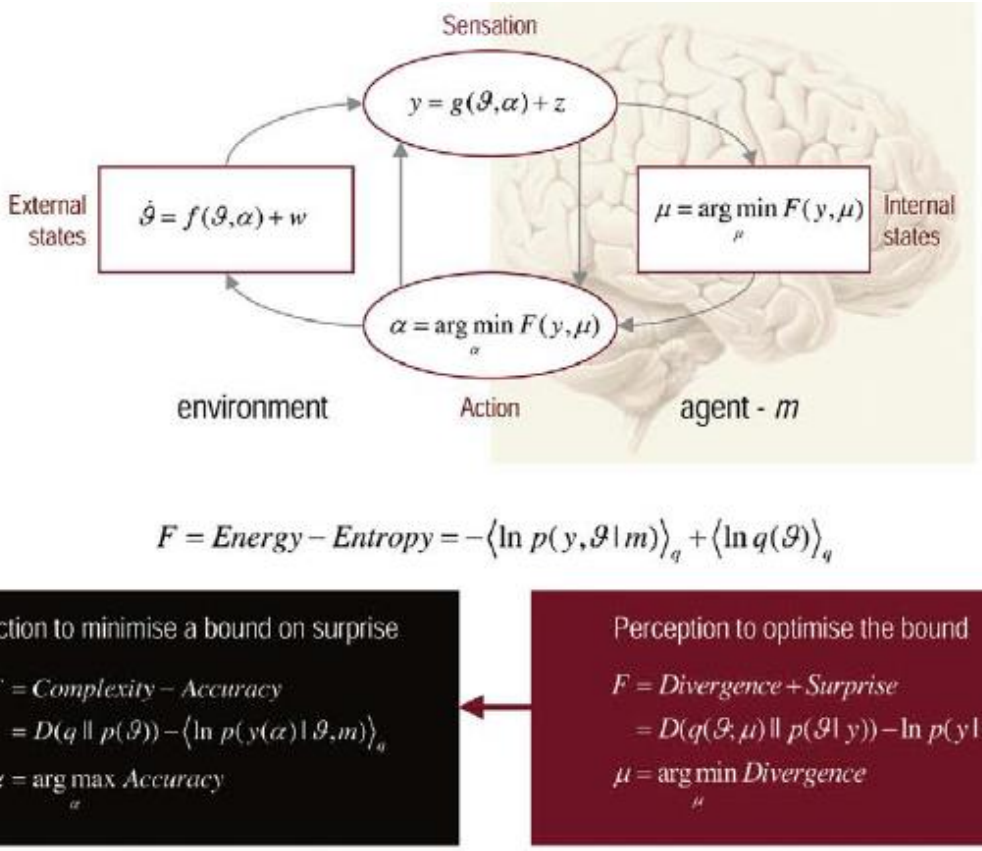
Η ελεύθερη ενέργεια είναι επομένως μία συνάρτηση της πυκνότητας αναγνώρισης και της εισόδου του αισθητήρα. Συμπεριλαμβάνει δύο όρους, την προβλεπόμενη ενέργεια υπό συγκεκριμένη πυκνότητα και την εντροπία αυτής. Η ενέργεια είναι απλώς η έκπληξη σχετικά με τη συλλογική δράση της εισόδου του αισθητήρα *y* και των επιδράσεων της *θ*. Η ελεύθερη ενέργεια εξαρτάται από δύο πυκνότητες· μία που παράγει τα δείγματα του αισθητήρα και τις αιτίες τους *p(y, θ)* και μία πυκνότητα αναγνώρισης μόνο για τις αιτίες *q(θ, μ)*. Η τελευταία καθορίζεται από τα χαρακτηριστικά της, *μ*, τα οποία θεωρούμε ότι κωδικοποιούνται από τον εγκέφαλο. Αυτό σημαίνει ότι η ελεύθερη ενέργεια προκαλεί ένα μοντέλο *m* για κάθε σύστημα και μια πυκνότητα αναγνώρισης με βάσεις τις αιτίες ή τις παραμέτρους αυτού του μοντέλου. Δεδομένης της λειτουργικής μορφής αυτών των πυκνοτήτων η ελεύθερη ενέργεια μπορεί να εκτιμηθεί, αφού είναι μια συνάρτηση της εισόδου του αισθητήρα και επαρκούς στατιστικής. Η ΑΕΕ δηλώνει ότι όλες οι ποσότητες που μπορούν να αλλάξουν ελαχιστοποιούν την ελεύθερη ενέργεια (**Σχήμα 10**).

Είναι εύκολο να δείξουμε πώς η οπτικοποίηση της πυκνότητας αναγνώρισης αντιστοιχεί την υποθετική πυκνότητα σε περιβαλλοντικές αιτίες, δοθέντων των δεδομένων του αισθητήρα. Αυτό μπορεί να φανεί εκφράζοντας την ελεύθερη ενέργεια ως έκπληξη *-ln(p(y|m))* συν την απόκλιση της πυκνότητας αναγνώρισης και της υποθετικής πυκνότητας. Επειδή αυτή η διαφορά είναι πάντα θετική, η ελαχιστοποίηση της ελεύθερης ενέργειας κάνει την πυκνότητα αναγνώρισης μια προσέγγιση της πραγματικής μεταγενέστερης πιθανότητας. Αυτό σημαίνει ότι το σύστημα σιωπηρά αναφέρεται ή αντιπροσωπεύει τις αιτίες των δειγμάτων του αισθητήρα του με έναν Μπεϋζιανό τρόπο. Ταυτόχρονα, η ελεύθερη ενέργεια γίνεται ένα στενό όριο της έκπληξης, η οποία ελαχιστοποιείται μέσω της δράσης [10].

Η τροποποίηση του περιβάλλοντος για ελαχιστοποίηση της ελεύθερης ενέργειας μέσω της δράσης επιβάλει ένα δείγμα δεδομένων του αισθητήρα που είναι συνεπές με τη συγκεκριμένη αναπαράσταση. Αυτό ουσιαστικά μπορεί να είναι μια εκ νέου διευθέτηση της ελεύθερης ενέργειας ως ένα μείγμα ακρίβειας και πολυπλοκότητας. Είναι όμως σημαντικό να τονιστεί ότι η δράση επηρεάζει μόνο την ακρίβεια. Αυτό σημαίνει ότι ο εγκέφαλος θα επαναρυθμίσει τους επιθηλιακούς του αισθητήρες για τα δείγματα εισόδου που προβλέπονται από την αναπαράστασή του, δηλαδή θα ελαχιστοποιήσει το σφάλμα πρόβλεψης.

Επιστρέφοντας στην εξίσωση (4.1) προκύπτει, λοιπόν, ότι η ελαχιστοποίηση της εντροπίας αντιστοιχεί στην καταστολή της έκπληξης με την πάροδο του χρόνου. Εν συντομία, για να υπάρχει ένας καλά ορισμένος παράγοντας πρέπει να κατέχει ένα ορισμένο εύρος καταστάσεων, όπως για παράδειγμα ένα ψάρι στο νερό. Αυτό σημαίνει ότι το ισοζύγιο πυκνότητας των παραγόντων, που περιγράφουν την πιθανότητα, ένας από αυτούς να βρίσκεται σε μια συγκεκριμένη κατάσταση, πρέπει να έχει χαμηλή εντροπία. Μια κατανομή με χαμηλή εντροπία απλώς σημαίνει ότι μόνο ένας μικρός αριθμός καταστάσεων είναι κατειλημμένος τον περισσότερο χρόνο. Επειδή η εντροπία είναι ο μακροχρόνιος μέσος όρος της έκπληξης, οι παράγοντες πρέπει να αποφεύγουν καταστάσεις έκπληξης (ψάρι έξω από το νερό). Ωστόσο, υπάρχει το πρόβλημα ότι οι παράγοντες

δεν μπορούν να εκτιμήσουν αμέσως την έκπληξη, κάτι που θα σήμαινε πως θα γνωρίζαμε όλες τις πιθανές καταστάσεις του κόσμου που θα μπορούσαμε να δεχτούμε ως είσοδο με τις αισθήσεις μας. Όμως, ένας παράγοντας μπορεί να αποφύγει την ανταλλαγή έκπληξης με το περιβάλλον ελαχιστοποιώντας την ελεύθερη ενέργεια του, αφού η ελεύθερη ενέργεια είναι πάντα μεγαλύτερη της έκπληξης.



Σχήμα 10: Πάνω μέρος: Σχηματική ανάλυση των ποσοτήτων που ορίζουν την ελεύθερη ενέργεια. Αυτές περιλαμβάνουν καταστάσεις του εγκεφάλου μ και ποσότητες που περιγράφουν ανταλλαγές με το περιβάλλον· η είσοδος του αισθητήρα y=g(θ,α)+z και η δράση α που αλλάζει τον τρόπο με τον οποίο παίρνεται το δείγμα από το περιβάλλον. Το περιβάλλον περιγράφεται από εξισώσεις κίνησης, $\dot{\theta}$=f(θ, α)+w, που προσδιορίζουν τη δυναμική των περιβαλλοντικών αιτιών θ. Οι καταστάσεις και η δράση του εγκεφάλου αλλάζουν για να ελαχιστοποιήσουν την ελεύθερη ενέργεια, η οποία είναι μια συνάρτηση της εισόδου των αισθητήρων και της πιθανολογικής αναπαράστασης (πυκνότητα αναγνώρισης) $q(\theta, \mu)$ κωδικοποιημένης από το μ. Κάτω μέρος: Εναλλακτικές εκφράσεις για την ελεύθερη ενέργεια που δείχνουν τι συνεπάγεται η ελαχιστοποίηση της. Από άποψη δράσης, η ελεύθερη ενέργεια μπορεί να κατασταλεί μόνο με την αύξηση της ακρίβειας των δεδομένων του αισθητήρα. Αντίθετα, η βελτιστοποίηση των καταστάσεων του εγκεφάλου κάνει την αναπαράσταση μια κατά προσέγγιση υποθετική πυκνότητα των αιτιών της εισόδου του αισθητήρα. Αυτή η βελτιστοποίηση κάνει την ελεύθερη ενέργεια να περιορίζει περισσότερο την έκπληξη και αποφεύγονται μη προσδοκόμενα συμβάντα στον αισθητήρα [10].

Μαθηματικά, η διαφορά μεταξύ ελεύθερης ενέργειας και έκπληξης είναι η απόκλιση μεταξύ της πιθανολογικής αναπαράστασης (πυκνότητα αναγνώρισης) κωδικοποιημένης από τον παράγοντα και της πραγματικής κατανομής των αιτιών της εισόδου του αισθητήρα. Αυτή η αναπαράσταση

επιτρέπει στον εγκέφαλο να ελαττώσει την ελεύθερη ενέργεια αλλάζοντας τη δική του αναπαράσταση, κάνοντας έτσι την πυκνότητα αναγνώρισης κατά προσέγγιση ίση με την υποθετική πυκνότητα. Αυτό αντιστοιχεί στο Μπεϋζιανό συμπέρασμα σχετικά με τις άγνωστες καταστάσεις του κόσμου που προκαλούν τα δεδομένα του αισθητήρα. Εν συντομία, η ΑΕΕ υπάγεται στη Μπεϋζιανή υπόθεση σχετικά με τον εγκέφαλο. Αξίζει να σημειωθεί ότι έχει δειχθεί αποτελεσματικά πως βιολογικοί παράγοντες πρέπει να δρούνε με έναν Μπεϋζιανό τρόπο για να αποφύγουν την ανταλλαγή κάποιας έκπληξης με το περιβάλλον.

Ωστόσο η αντίληψη είναι μόνο η μισή ιστορία, καθώς η ελεύθερη ενέργεια είναι ένας καλός αντιπρόσωπος της έκπληξης, ωστόσο αυτό δεν αλλάζει τις ίδιες τις αισθήσεις και τις εκπλήξεις τους. Για να μειώσουμε την έκπληξη πρέπει να αλλάξουμε την είσοδο του αισθητήρα. Εδώ έρχεται η ΑΕΕ και λέει ότι και η δράση μας βοηθάει στην ελαχιστοποίηση της ελεύθερης ενέργειας. Είμαστε ανοιχτά συστήματα σε συνεχή ανταλλαγή με το περιβάλλον, δηλαδή το περιβάλλον μας παρέχει ως εισόδους τις καταστάσεις του, εμείς τις επεξεργαζόμαστε, δρούμε πάνω του και τις αλλάζουμε. Αυτή η ανταλλαγή γίνεται με τις αισθήσεις μας και τα τελεστικά όργανα (όπως οι φωτοϋποδοχείς και οι μυς οφθαλμοκίνησης). Αν αλλάξουμε το περιβάλλον ή τη σχέση μας μαζί του, αλλάζουν και οι είσοδοι. Επομένως, η δράση μπορεί να μειώσει την ελεύθερη ενέργεια αλλάζοντας είσοδο, ενώ η αντίληψη μειώνει την ελεύθερη ενέργεια αλλάζοντας την πρόβλεψη.

Μαθηματικά, η ΑΕΕ απαιτεί ένα δείγμα των πληροφοριών του αισθητήρα που να προσαρμόζεται στις απαιτήσεις μας. Αυτό βέβαια δε σημαίνει ότι μπορούμε απλώς να κλείσουμε κάποιο κανάλι αισθητήρα για να αποφύγουμε κάποια έκπληξη, όμως μπορούμε να αλλάξουμε την είσοδο μέσω της δράσης. Για παράδειγμα, δεν μπορούμε να αποφύγουμε τον πόνο, αν δεν απομακρύνουμε το επιβλαβές ερέθισμα που τον προκαλεί. Προσπαθούμε λοιπόν μέσα από τις εισόδους του περιβάλλοντος και τη δράση μας να εξασφαλίσουμε ότι οι προβλέψεις μας θα πραγματοποιηθούν και δε θα υπάρχει καμία έκπληξη. Η δυνατότητα δράσης «επιβάλει» στην αντίληψη να κάνει το δυνατόν περισσότερο φιλαλήθεις προβλέψεις, ελαχιστοποιώντας έτσι την ελεύθερη ενέργεια.

Η ΑΕΕ απαιτεί επίσης από τον εγκέφαλο να αναπαριστά τις αιτίες των εισόδων στον αισθητήρα. Η φύση αυτής της αναπαράστασης υπαγορεύεται από φυσιολογικούς και ανατομικούς περιορισμούς. Ανεξάρτητα από τη μορφή του, ο εγκέφαλος πρέπει να κωδικοποιήσει μια πυκνότητα αναγνώρισης με τα φυσικά χαρακτηριστικά της. Αυτά έχουν τον ρόλο των επαρκών στατιστικών, δηλαδή αριθμοί που χαρακτηρίζουν μια κατανομή, όπως ο μέσος όρος και η διασπορά.

Προφανώς οι αιτίες για τα δεδομένα του αισθητήρα μπορούν να αλλάξουν με διαφορετικά χρονοδιαγράμματα. Για παράδειγμα, οι περιβαλλοντικές καταστάσεις μπορεί να κωδικοποιούνται από νευρολογικά δυναμικά σε κλίμακα millisecond, ενώ αιτιατές σταθερές ή παράμετροι που αλλάζουν αργά μπορεί να κωδικοποιούνται από την ισχύ των συνδέσεων. Αυτές οι ποσότητες (καταστάσεις και παράμετροι) ανήκουν στα ντετερμινιστικά δυναμικά του κόσμου. Ωστόσο είναι επίσης απαραίτητο να παρουσιαστούν και οι τυχαίες επιδράσεις, όπως για παράδειγμα το εύρος τυχαίων διακυμάνσεων σε καταστάσεις. Αυτό επιφέρει μια τρίτη κλάση ποσοτήτων (ακρίβεια) που δημιουργούν αβεβαιότητα για τις καταστάσεις. Η ακρίβεια είναι σημαντικό κομμάτι της αναπαράστασης που προκύπτει από την τυχαιότητα στον κόσμο και αναλύεται αργότερα.

Σύμφωνα με την ΑΕΕ, τα επαρκή στατιστικά που αντιπροσωπεύουν και τις τριών ειδών ποσότητες θα αλλάξουν με στόχο την ελαχιστοποίηση της ελεύθερης ενέργειας. Αυτό παρέχει μια δομημένη εξήγηση για την αντίληψη, τη μνήμη και την προσοχή, δηλαδή και για τον αντιληπτικό συμπερασμό (οπτικοποίηση της συναπτικής δραστηριότητας για την κωδικοποίηση των καταστάσεων του περιβάλλοντος), και για την αντιληπτική μάθηση και μνήμη (οπτικοποίηση των συναπτικών δεσμών που κωδικοποιούν κάτι απρόβλεπτο ή αιτιατές σταθερές), και για την προσοχή (νευρορυθμιστική οπτικοποίηση της συναπτικής διαδικασίας για την κωδικοποίηση της ακρίβειας των καταστάσεων).

Αυτή η οπτικοποίηση μπορεί να διατυπωθεί ως μια κλίση της ελεύθερης ενέργειας για να αποδώσει διαφορικές εξισώσεις, που να ορίζουν δυναμική αναγνώρισης για συναπτική δραστηριότητα και αποτελεσματικότητα. Αυτή η δυναμική εξαρτάται από τη μορφή του γεννητικού μοντέλου του εγκεφάλου και των επαρκών στατιστικών που κωδικοποιεί. Αν θεωρήσουμε ότι η πυκνότητα αναγνώρισης είναι μια πολυδιάστατη Γκαουσιανή πυκνότητα, τότε η δυναμική αναγνώρισης υιοθετεί εύλογα νευρολογικές μορφές: η οπτικοποίηση των επαρκών στατιστικών των καταστάσεων μοιάζει ακριβώς με προβλεπτικό κώδικα, που περιλαμβάνει επαναλαμβανόμενη παράδοση μηνυμάτων μεταξύ πληθυσμών που κωδικοποιούν προβλέψεις και σφάλματα προβλέψεων. Η οπτικοποίηση των επαρκών στατιστικών των παραμέτρων είναι τυπικά ίδια με τη συσχετισμένη πλαστικότητα, ενώ η οπτικοποίηση των επαρκών στατιστικών της ακρίβειας είναι όμοια με την αφομοίωση του σφάλματος πρόβλεψης από τα ενισχυμένα συστήματα μάθησης.

Υπό την υπόθεση Laplace, η δυναμική αναγνώρισης γίνεται σύστημα αποδεικτικής συσσώρευσης, στο οποίο αλλάζει μέσω της νευρολογικής δραστηριότητας το σφάλμα πρόβλεψης. Επιπλέον, κάποιος μπορεί να καταλάβει την ιεραρχική παράταξη των περιοχών του φλοιού του εγκεφάλου και τη φύση των μηνυμάτων που περνάνε μέσα από τα επίπεδα του φλοιού ώστε να φτάσουμε στην ελαχιστοποίηση του σφάλματος πρόβλεψης υπό τα ιεραρχημένα δυναμικά μοντέλα του περιβάλλοντος. Τα ιεραρχικά μοντέλα είναι σημαντικά επειδή τυπικά είναι ισοδύναμα με τα εμπειρικά Μπεϋζιανά μοντέλα, στα οποία τα υψηλότερα επίπεδα παρέχουν εκ των προτέρων περιορισμούς στα χαμηλότερα επίπεδα. Αυτό επιτρέπει σε κάποιον να ερμηνεύσει από πάνω προς τα κάτω τις επιδράσεις στον εγκέφαλο εκδηλώνοντας κάποιο ερέθισμα. Υπό αυτή την οπτική, η καταπίεση της ελεύθερης ενέργειας σημαίνει ότι κάθε επίπεδο προσπαθεί να εξαλείψει τα σφάλματα πρόβλεψης για το ίδιο και για κάθε επίπεδο κάτω από αυτό, κάτι που οδηγεί σε μια επαναλαμβανόμενη αυτοοργάνωτη δυναμική που συγκλίνει σε μια συνεχή αναπαράσταση των αιτιών της εισόδου του αισθητήρα, σε πολλαπλά επίπεδα.

4.3 Νέες προοπτικές

Μέχρι τώρα έγινε προσπάθεια να εξηγήσουμε την προαναφερθείσα διατύπωση αναλύοντας πολλές εμπειρικές πλευρές της ανατομίας και της φυσιολογίας καταλήγοντας στην οπτικοποίηση της ελεύθερης ενέργειας. Κάποιος μπορεί να εξηγήσει ένα αξιοσημείωτο εύρος δεδομένων, όπως για παράδειγμα την ιεραρχημένη κατανομή των περιοχών του φλοιού, λειτουργικές ασυμμετρίες ανάμεσα σε συνδέσεις και άλλα γνωστικά φαινόμενα. Ωστόσο, η προσπάθεια που γίνεται πλέον

είναι επικεντρωμένη σε υποψήφια θέματα που θα μπορούσαν να προσφέρουν νέες οπτικές για τη δομή της νευροεπιστήμης ακόμα κι αν είναι αμφισβητήσιμες εξ αρχής. Αυτά τα παραδείγματα ενισχύουν τη σημασία της ακρίβειας (ή της αβεβαιότητας) μέσω της νευρολογικής διαμόρφωσης.

Ένας άξονας κλειδί είναι πώς ο εγκέφαλος κωδικοποιεί την πυκνότητα αναγνώρισης. Η ΑΕΕ δέχεται αυτή την πυκνότητα, η οποία όμως πρέπει να αναπαρασταθεί από τα επαρκή στατιστικά της. Είναι επομένως δεδομένο ότι ο εγκέφαλος αναπαριστά κατανομές πιθανότητας μέσω των ενεργειών των αισθητήρων. Αλλά ποια είναι η μορφή της κατανομής και ποια αυτή των επαρκών στατιστικών που απαρτίζουν τον κώδικα πιθανοτήτων του εγκεφάλου; Υπάρχουν δύο ειδών υποθετικές μορφές, η ελεύθερη και η φιξαρισμένη. Οι προτάσεις για την ελεύθερη μορφή περιλαμβάνουν κώδικες φιλτραρίσματος σωματιδίων και πιθανοτικού πληθυσμού. Σχετικά με το φιλτράρισμα σωματιδίων, η πυκνότητα αναγνώρισης αντιπροσωπεύεται από ένα δείγμα πυκνότητας του συνόλου των νευρώνων, η δραστηριότητα των οποίων κωδικοποιεί την τοποθεσία των σωματιδίων στο χωροχρόνο. Οι προτάσεις για τη φιξαρισμένη μορφή είναι συνήθως είτε «πολυεθνικές» είτε Γκαουσιανές. Οι πολυεθνική μορφή θεωρεί ότι ο κόσμος είναι μία από αρκετές διακριτές καταστάσεις και συνήθως σχετίζονται με κρυφά Μαρκοβιανά μοντέλα. Αντίθετα, η Γκαουσιανή μορφή επιτρέπει συνεχής και συσχετιζόμενες καταστάσεις.

Κάθε σχέδιο που οπτικοποιεί τα επαρκή στατιστικά αυτών των μορφών πρέπει να συμμορφώνεται με την ΑΕΕ. Οπότε γιατί έχουμε επικεντρωθεί στην προσέγγιση του Laplace; Αρχικά, οι προτάσεις ελεύθερης μορφής δεν κλιμακώνονται. Για παράδειγμα, για να αναπαρασταθεί ένα πρόσωπο με περίπου 30 χαρακτηριστικά, θα χρειαζόμασταν έναν αντιληπτικό χωροχρόνο 30 διαστάσεων με περισσότερους νευρώνες από αυτούς που έχει ο εγκέφαλος. Το επιχείρημα βέβαια υπέρ της ελεύθερης μορφής είναι ότι έχει τη δυνατότητα να κωδικοποιήσει περίπλοκες πυκνότητες αναγνώρισης. Ωστόσο, είναι σχεδόν ασήμαντο να αναπαριστά κάποιος μη Γκαουσιανές μορφές υπό την προσέγγιση Laplace χρησιμοποιώντας μη γραμμικούς μετασχηματισμούς των μεταβλητών. Επιπλέον δεν υπάρχει καμία ηλεκτροφυσιολογική ή ψυχοφυσική απόδειξη που να υποδεικνύει ότι ο εγκέφαλος μπορεί να κωδικοποιήσει πολυτροπικές προσεγγίσεις, ενώ όντως υπάρχουν δείγματα που δείχνουν ότι η πυκνότητα αναγνώρισης είναι μονοτροπική. Από την άλλη οι πολυτροπικές προσεγγίσεις και τα κρυφά Μαρκοβιανά μοντέλα έχουν μεγάλη απλότητα και δεν μπορούν να αναπαραστήσουν εξαρτήσεις ανάμεσα σε καταστάσεις. Αντίθετα, η προσέγγιση Laplace μπορεί να χειριστεί συνεχείς και συσχετιζόμενες καταστάσεις επαρκώς, ενώ αυτό οφείλεται στο ότι η πυκνότητα αναγνώρισης προσδιορίζεται απολύτως από το μέσο όρο της.

Υπό τα ιεραρχικά μοντέλα της αντίληψης, είναι υποχρεωτικό να οπτικοποιήσουμε τη σχετική ακρίβεια των αποδείξεων των αισθητήρων. Νευροβιολογικά, αυτό αντιστοιχεί στη διαμόρφωση του βάρους των μονάδων σφάλματος. Αυτή η οπτικοποίηση είναι κρίσιμη για να φτάσουμε σε συμπέρασμα και είναι σαν να υπολογίζουμε το σίγουρο σφάλμα σε ένα τεστ, ενώ η σημασία του να αναπαριστούμε την ακρίβεια είναι ακόμα μεγαλύτερη στην ιεραρχική απόδειξη καθώς ελέγχει τη σχετική επιρροή των προσδοκιών μας σε διαφορετικά επίπεδα.

4.4 Η Αρχή Ελεύθερης Ενέργειας

Παρά το πλήθος εμπειρικών δεδομένων στη νευροεπιστήμη, υπάρχουν σχετικά λίγες διεθνείς θεωρίες σχετικά με το πώς λειτουργεί ο εγκέφαλος. Η Αρχή Ελεύθερης Ενέργειας (ΑΕΕ) για προσαρμοστικά συστήματα προσπαθεί να παρέχει μια συνολική πρόταση για τη δράση, την αντίληψη και τη μάθηση. Παρά το ότι αυτή η αρχή έχει χαρακτηριστεί ως μια ενοποιημένη θεωρία για τον εγκέφαλο, η ικανότητα της να συνδυάζει διαφορετικές οπτικές της λειτουργίας του εγκεφάλου δεν έχει εδραιωθεί ακόμα.

Η ΑΕΕ λέει ότι κάθε αυτό-οργανωτικό σύστημα που βρίσκεται σε ισορροπία με το περιβάλλον του πρέπει να ελαχιστοποιεί την ελεύθερη ενέργεια. Η Αρχή είναι βασικά ένας μαθηματικός σχηματισμός για το πώς τα προσαρμοστικά συστήματα (δηλαδή βιολογικοί παράγοντες, όπως τα ζώα ή ο εγκέφαλος) αντιστέκονται στη φυσική τάση για αταξία. Αν και πρόκειται για μια αρκετά απλή βασική ιδέα, τα συμπεράσματα της είναι ποικίλα και πολύπλοκα. Η ποικιλία επιτρέπει στην Αρχή να ασχολείται με πολλές πτυχές της δομής και της λειτουργίας του εγκεφάλου και να τείνει στην πιθανότητα να ενοποιηθούν οι διαφορετικές οπτικές για το πώς λειτουργεί ο εγκέφαλος.

Το καθοριστικό χαρακτηριστικό των βιολογικών συστημάτων είναι ότι διατηρούν τις καταστάσεις και τη μορφή τους μπροστά σε ένα συνεχώς εναλλασσόμενο περιβάλλον. Από την πλευρά του εγκεφάλου, ως περιβάλλον θεωρείται και το εσωτερικό και το εξωτερικό περιβάλλον. Αυτή η διατήρηση της τάξης παρατηρείται σε πολλά επίπεδα και διακρίνεται βιολογικά από άλλα αυτο-οργανωτικά συστήματα, καθώς όντως η φυσιολογία των βιολογικών συστημάτων μπορεί να αναλυθεί σχεδόν πλήρως από την ομοιόσταση τους. Πιο συγκεκριμένα, το εύρος των φυσιολογικών καταστάσεων και των καταστάσεων των αισθητήρων στις οποίες περιορίζεται ένας οργανισμός, καθορίζει και το φαινότυπο του οργανισμού. Μαθηματικά, αυτό σημαίνει ότι η πιθανότητα αυτών των καταστάσεων των αισθητήρων πρέπει να έχει χαμηλή εντροπία, δηλαδή να υπάρχει μεγάλη πιθανότητα το σύστημα να βρεθεί σε κάποια από έναν μικρό αριθμό πιθανών καταστάσεων και μικρή πιθανότητα να βρεθεί σε κάποια από όλες τις υπόλοιπες. Η εντροπία ουσιαστικά χαρακτηρίζει την «έκπληξη». Συγκεκριμένα, ένα ψάρι έξω από το νερό θα ήταν μια κατάσταση έκπληξης, τόσο συναισθηματικά όσο και μαθηματικά, άρα θα είχε και υψηλή εντροπία. Ωστόσο, ας σημειωθεί ότι έκπληξη και εντροπία εξαρτώνται και από τον παράγοντα, αφού κάτι που μπορεί να αποτελεί έκπληξη για κάποιον μπορεί να μην αποτελεί για κάποιον άλλο. Οι βιολογικοί παράγοντες πρέπει επομένως να ελαχιστοποιήσουν τον μακροχρόνιο μέσο όρο έκπληξης για να εξασφαλίσουν ότι η εντροπία των αισθητήρων τους παραμένει χαμηλά [11].

Συντομότερα, η μακροπρόθεσμη επιτακτική διατήρηση των καταστάσεων σε φυσιολογικά όρια μεταφράζεται σε βραχυπρόθεσμη αποφυγή της έκπληξης. Η έκπληξη δεν σχετίζεται απλώς με την τωρινή κατάσταση, η οποία δεν μπορεί και να αλλάξει, αλλά και με τη μετακίνηση από μία κατάσταση σε μια άλλη, η οποία μπορεί να αλλάξει. Αυτή η κίνηση μπορεί να είναι περίπλοκη καθώς επισκέπτεται ένα μικρό σετ καταστάσεων, οι οποίες όμως είναι συμβατές με την επιβίωση (όπως για παράδειγμα η ασφαλής οδήγηση ενός αυτοκινήτου). Αυτή την κίνηση οπτικοποιεί η ΑΕΕ.

Μέχρι τώρα, αυτό που είπαμε είναι ότι οι βιολογικοί παράγοντες πρέπει να αποφεύγουν τις εκπλήξεις για να εξασφαλίζουν ότι οι καταστάσεις τους θα παραμένουν σε φυσιολογικά πλαίσια.

Αλλά πώς το κάνουν αυτό; Ένα σύστημα δεν μπορεί να γνωρίζει αν οι αισθήσεις του θα επιφέρουν έκπληξη και δε θα μπορούσε να τις αποφύγει ακόμα και αν το γνώριζε. Εδώ είναι που έρχεται η ελεύθερη ενέργεια, λειτουργώντας ως ένα άνω όριο για την έκπληξη, που σημαίνει ότι αν οι παράγοντες ελαχιστοποιούν την ελεύθερη ενέργεια, ουσιαστικά ελαχιστοποιούν και την έκπληξη. Είναι σημαντικό ότι η ελεύθερη ενέργεια μπορεί να αξιολογηθεί επειδή αποτελεί συνάρτηση δύο πραγμάτων στα οποία ο παράγοντας έχει πρόσβαση: τις καταστάσεις των αισθητήρων του και την πυκνότητα αναγνώρισης που κωδικοποιείται από τις εσωτερικές του καταστάσεις (όπως για παράδειγμα τη νευρολογική δραστηριότητα και την ισχύ των συνδέσεων). Η πυκνότητα αναγνώρισης είναι μια πιθανολογική αναπαράσταση του τι προξένησε μια συγκεκριμένη αίσθηση.

Αυτή η μεταβλητή δομή της ελεύθερης ενέργειας εισάχθηκε στη στατιστική φυσική για να μετατρέψει δύσκολα πιθανολογικά ενσωματωμένα προβλήματα σε εύκολα οπτικοποιημένα προβλήματα. Είναι μια θεωρητική πληροφοριακή ποσότητα (όπως και η έκπληξη), σε αντίθεση με τη θερμοδυναμική ποσότητα. Η μεταβλητή ελεύθερη ενέργεια έγινε αντικείμενο εκμετάλλευσης στη μηχανική μάθηση και τη στατιστική για να λύσει πολλά συμπερασματικά και μαθησιακά προβλήματα. Στη δική μας περίπτωση, η έκπληξη καλείται αρνητικό αποδεικτικό μοντέλο. Αυτό σημαίνει ότι η ελαχιστοποίηση της έκπληξης είναι το ίδιο με τη μεγιστοποίηση της απόδειξης των αισθητήρων για την ύπαρξη ενός παράγοντα, αν θεωρήσουμε τον παράγοντα ως μοντέλο του κόσμου του. Στο παρόν περιεχόμενο η ελεύθερη ενέργεια παρέχει την απάντηση σε μια θεμελιώδη ερώτηση: Πώς τα αυτο-οργανωτικά προσαρμοστικά συστήματα αποφεύγουν τις καταστάσεις έκπληξης; Με την ελαχιστοποίηση της ελεύθερης ενέργειάς τους. Τι περιλαμβάνει αυτό όμως;

4.5 Συμπεράσματα: δράση και αντίληψη

Οι παράγοντες μπορούν να καταπιέσουν την ελεύθερη ενέργεια αλλάζοντας τα δύο πράγματα από τα οποία εξαρτάται, δηλαδή την είσοδο των αισθητήρων δρώντας πάνω στο περιβάλλον και την πυκνότητα αναγνώρισης αλλάζοντας τις εσωτερικές καταστάσεις. Αυτή η διάκριση καθορίζει τη δράση και την αντίληψη. Κάποιος μπορεί να καταλάβει καλύτερα τη σημασία αυτής της διάκρισης θεωρώντας τρεις μαθηματικά ισοδύναμες σχέσεις για την ελεύθερη ενέργεια.

Η πρώτη σχέση εκφράζει την ελεύθερη ενέργεια ως ενέργεια μείον εντροπία. Αυτή η σχέση είναι σημαντική για τρεις λόγους. Πρώτον, συνδέει την έννοια της ελεύθερης ενέργειας όπως χρησιμοποιείται στη θεωρία πληροφοριών με έννοιες που χρησιμοποιούνται στη στατιστική θερμοδυναμική. Δεύτερον, δείχνει ότι η ελεύθερη ενέργεια μπορεί να αξιολογηθεί από έναν παράγοντα επειδή η ενέργεια είναι η έκπληξη σχετικά με ένα κοινό περιστατικό των αισθήσεων και τις αντιληπτές αιτίες του, ενώ η εντροπία προκύπτει απλώς από την πυκνότητα αναγνώρισης του ίδιου του οργανισμού. Τρίτον, δείχνει ότι η ελεύθερη ενέργεια βασίζεται σε ένα γεννητικό μοντέλο του κόσμου, που εκφράζεται με όρους πιθανοτήτων των αισθήσεων και των αιτιών τους που δρουν μαζί. Αυτό σημαίνει ότι κάθε παράγοντας πρέπει να έχει ένα εσωτερικό γεννητικό μοντέλο για το πώς οι αιτίες αντιδρούν για να παράξουν τα δεδομένα του αισθητήρα. Αυτό το μοντέλο καθορίζει τόσο τη φύση του παράγοντα όσο και την ποιότητα των ορίων της ελεύθερης ενέργειας σχετικά με την έκπληξη [12].

Η δεύτερη σχέση εκφράζει την ελεύθερη ενέργεια ως έκπληξη συν έναν όρο απόκλισης. Η (αντιληπτική) απόκλιση είναι απλώς η διαφορά μεταξύ της πυκνότητας αναγνώρισης και της σχετικής πυκνότητας (ή μεταγενέστερης πυκνότητας) των αιτιών μιας αίσθησης, δεδομένων των σημάτων του αισθητήρα. Αυτή η σχετική πυκνότητα αντιπροσωπεύει τη βέλτιστη πιθανή υπόθεση σχετικά με τις πραγματικές αιτίες. Η διαφορά μεταξύ των δύο πυκνοτήτων είναι πάντα μη αρνητική και η ελεύθερη ενέργεια είναι επομένως το άνω όριο της έκπληξης. Έτσι, ελαχιστοποιώντας την ελεύθερη ενέργεια με την αλλαγή της πυκνότητας αναγνώρισης (χωρίς να αλλάξουν τα δεδομένα των αισθητήρων) μειώνεται και η αντιληπτική απόκλιση, και εντέλει η πυκνότητα αναγνώρισης γίνεται σχετική πυκνότητα και η ελεύθερη ενέργεια γίνεται έκπληξη.

Η Τρίτη σχέση εκφράζει την ελεύθερη ενέργεια ως πολυπλοκότητα μείον ακρίβεια, χρησιμοποιώντας όρους από το μοντέλο της βιβλιογραφίας συγκρίσεων. Πολυπλοκότητα είναι η διαφορά μεταξύ της πυκνότητας αναγνώρισης και της προγενέστερης πυκνότητας των αιτιών. Η τελευταία είναι γνωστή και ως Μπεϋζιανή έκπληξη και είναι η διαφορά της προγενέστερης πυκνότητας, που κωδικοποιεί τις πεποιθήσεις για την κατάσταση του περιβάλλοντος πριν αφομοιωθούν τα δεδομένα των αισθητήρων, και των μεταγενέστερων πεποιθήσεων, που κωδικοποιούνται από την πυκνότητα αναγνώρισης. Η ακρίβεια είναι απλώς η έκπληξη σχετικά με τις αισθήσεις που προσδοκούνται υπό μια συγκεκριμένη πυκνότητα αναγνώρισης. Αυτή η σχέση δείχνει ότι η ελαχιστοποίηση της ελεύθερης ενέργειας με την αλλαγή των δεδομένων του αισθητήρα (χωρίς να αλλάξει η πυκνότητα αναγνώρισης) πρέπει να αυξάνει την ακρίβεια των προβλέψεων ενός παράγοντα. Εν συντομία, ο παράγοντας θα λάβει επιλεκτικά δείγματα των εισόδων του αισθητήρα που προσδοκεί. Αυτό είναι γνωστό και ως ενεργό συμπέρασμα. Ένα ενστικτώδες παράδειγμα αυτής της διαδικασίας (αναφερόμενοι στη συνείδηση) είναι η αίσθηση που έχουμε κινούμενοι στο σκοτάδι σε έναν γνωστό χώρο: προσδοκούμε ότι θα έρθουμε σε επαφή με κάποιο αντικείμενο σύντομα και προσπαθούμε να επιβεβαιώσουμε αυτή την προσδοκία.

Συνολικά, η ελεύθερη ενέργεια βασίζεται σε ένα μοντέλο για το πώς παράγονται τα δεδομένα των αισθητήρων και η πυκνότητα αναγνώρισης με βάση τις παραμέτρους του μοντέλου. Η ελεύθερη ενέργεια μπορεί να μειωθεί μόνο αλλάζοντας την πυκνότητα αναγνώρισης ώστε να αλλάξουν και οι σχετικές προσδοκίες ή αλλάζοντας το δείγμα του αισθητήρα ώστε να συμμορφώνεται με τις προσδοκίες.

4.6 Η Μπεϋζιανή υπόθεση για τον εγκέφαλο

Η Μπεϋζιανή υπόθεση για τον εγκέφαλο χρησιμοποιεί τη Μπεϋζιανή θεωρία πιθανότητας για να σχηματίσει την αντίληψη ως μια κατασκευαστική διαδικασία βασισμένη σε εσωτερικά ή γεννητικά μοντέλα. Η υποβόσκουσα ιδέα είναι ότι ο εγκέφαλος έχει ένα μοντέλο του περιβάλλοντος, το οποίο προσπαθεί να οπτικοποιήσει χρησιμοποιώντας τις εισόδους των αισθητήρων. Αυτή η ιδέα συσχετίζεται με την ανάλυση μέσω σύνθεσης και τα επιστημολογικά αυτόματα. Από αυτή την άποψη, ο εγκέφαλος είναι μια συμπερασματική μηχανή που ενεργά προβλέπει και εξηγεί τις αισθήσεις του. Βάση αυτής της υπόθεσης είναι το πιθανολογικό μοντέλο που μπορεί να παράγει προβλέψεις, απέναντι στα δείγματα των αισθητήρων που εξετάζονται ώστε να ανανεώσει τις πεποιθήσεις σχετικά με τις αιτίες τους. Έτσι η αντίληψη γίνεται η διαδικασία

αντιστροφής του πιθανολογικού μοντέλου και οδηγεί σε πρόσβαση στη μεταγενέστερη πιθανότητα των αιτιών, δεδομένων των δεδομένων των αισθητήρων. Αυτή η αντιστροφή είναι το ίδιο με την ελαχιστοποίηση της διαφοράς μεταξύ της πυκνότητας αναγνώρισης και της μεταγενέστερης πυκνότητας για την καταπίεση της ελεύθερης ενέργειας. Όντως, ο σχηματισμός της ελεύθερης ενέργειας αναπτύχθηκε για την εξομάλυνση των δύσκολων συμπερασματικών προβλημάτων μετατρέποντας τα σε ευκολότερα προβλήματα οπτικοποίησης. Αυτή η διαδικασία θωράκισε ορισμένες προσεγγιστικές τεχνικές μοντέλων αναγνώρισης και σύγκρισης. Προκύπτουν έτσι αρκετά ενδιαφέροντα θέματα με το συνδυασμό της Μπεϋζιανής υπόθεσης και της ΑΕΕ, ενώ θα επικεντρωθούμε στα δύο βασικά.

Το πρώτο είναι η μορφή του γεννητικού μοντέλου και πώς αυτό εκδηλώνεται στον εγκέφαλο. Μια κριτική της Μπεϋζιανής προσέγγισης είναι ότι αγνοεί το πώς σχηματίζονται οι προγενέστερες πεποιθήσεις, που είναι απαραίτητες για το συμπερασμό. Ωστόσο, η κριτική απορρίπτεται λόγω των ιεραρχικών γεννητικών μοντέλων, στα οποία γίνεται οπτικοποίηση όλων των προγενέστερων. Στα ιεραρχικά μοντέλα, αιτίες σε ένα επίπεδο δημιουργούν υφιστάμενες αιτίες σε ένα χαμηλότερο επίπεδο, ενώ τα δεδομένα των αισθητήρων δημιουργούνται στο χαμηλότερο επίπεδο. Αυτό είναι σημαντικό, καθώς όλα τα προγενέστερα είναι συνδεδεμένα ιεραρχικά, παίρνουν πληροφορίες από τα δεδομένα των αισθητήρων και δίνουν τη δυνατότητα στον εγκέφαλο να οπτικοποιήσει τις προγενέστερες προσδοκίες του. Αυτή η οπτικοποίηση κάνει κάθε ιεραρχικό επίπεδο ορατό στα υπόλοιπα, εξασφαλίζοντας μια εσωτερική συνεχή αναπαράσταση των αιτιών των αισθητήρων σε πολλαπλά επίπεδα περιγραφής. Όχι μόνο τα ιεραρχικά μοντέλα παίζουν σημαντικό ρόλο στη στατιστική, αλλά χρησιμοποιούνται και από τον εγκέφαλο, δεδομένου και του ιεραρχικού κανονισμού τους αισθητήρες του φλοιού.

Το δεύτερο θέμα είναι η μορφή της πυκνότητας αναγνώρισης που κωδικοποιείται από τα φυσικά χαρακτηριστικά του εγκεφάλου, όπως η συναπτική δραστηριότητα, αποτελεσματικότητα και κέρδος. Γενικά κάθε πυκνότητα κωδικοποιείται από τα επαρκή στατιστικά της (για παράδειγμα το μέσο όρο και τη διακύμανση μιας Γκαουσιανής κατανομής). Ο τρόπος με τον οποίο ο εγκέφαλος κωδικοποιεί αυτά τα στατιστικά τοποθετεί σημαντικούς περιορισμούς στη μορφή των σχημάτων που προκύπτουν από την αναγνώριση. Το εύρος τους είναι από σχήματα ελεύθερης μορφής, που χρησιμοποιούν ένα μεγάλο αριθμό επαρκών στατιστικών, μέχρι σχήματα απλούστερης μορφής, που έχουν όμως ισχυρότερες υποθέσεις σχετικά με τη μορφή της πυκνότητας αναγνώρισης, ώστε όλα να μπορούν να κωδικοποιηθούν από έναν μικρό αριθμό επαρκών στατιστικών. Η απλούστερη υποτιθέμενη μορφή είναι η Γκαουσιανή, που απαιτεί μόνο το σχετικό μέσο όρο ή προσδοκία, που είναι γνωστός ως υπόθεση Laplace και υπό την οποία η ελεύθερη ενέργεια είναι απλώς η διαφορά μεταξύ της πρόβλεψης του μοντέλου και των αναπαραστάσεων που προβλέπονται. Τότε η ελαχιστοποίηση της ελεύθερης ενέργειας αντιστοιχεί στην εξήγηση των σφαλμάτων πρόβλεψης. Αυτή η διαδικασία είναι γνωστή ως προβλεπτικός κώδικας και έχει γίνει ένα διάσημο πλαίσιο για την κατανόηση των νευρικών μηνυμάτων που μεταδίδονται μεταξύ των ιεραρχικών επιπέδων του φλοιού. Σε αυτό το σχήμα, οι μονάδες του σφάλματος πρόβλεψης συγκρίνουν τις σχετικές προσδοκίες με τις από πάνω προς τα κάτω προβλέψεις ώστε να αποδώσουν ένα σφάλμα πρόβλεψης. Αυτό το σφάλμα πρόβλεψης προωθείται για να οδηγήσει τις μονάδες σε ένα ανώτερο επίπεδο που κωδικοποιεί τις σχετικές προσδοκίες, οι οποίες οπτικοποιούν τις από πάνω προς τα κάτω προσδοκίες, ώστε να μειωθεί το σφάλμα πρόβλεψης στο παρακάτω επίπεδο. Η αμοιβαία

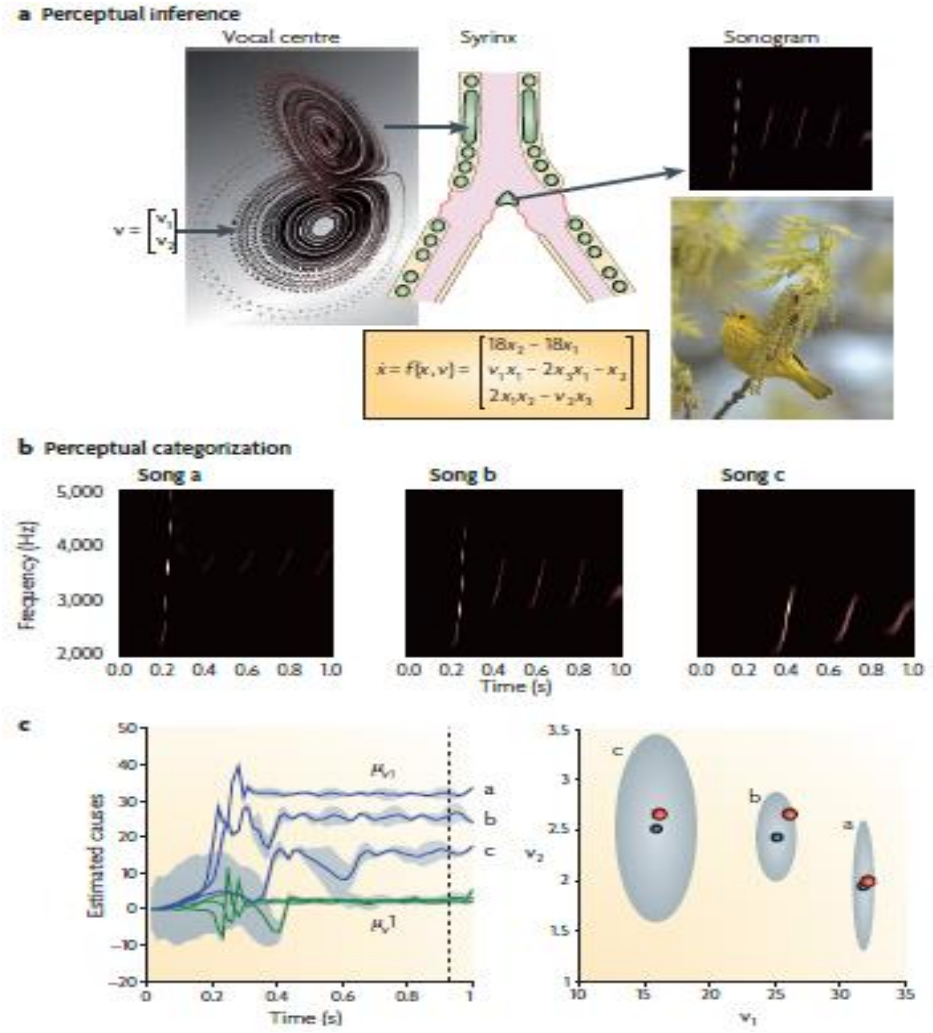
ανταλλαγή προβλέψεων και σφαλμάτων πρόβλεψης συνεχίζεται έως ότου το σφάλμα πρόβλεψης ελαχιστοποιηθεί σε όλα τα επίπεδα και οι σχετικές προσδοκίες οπτικοποιούνται. Αυτό το σχήμα έχει επικαλεστεί για να εξηγηθούν πολλά χαρακτηριστικά των πρώιμων οπτικών αντιδράσεων. Το **Σχήμα 11** παρέχει ένα παράδειγμα αντιληπτικής κατηγοριοποίησης χρησιμοποιώντας αυτό το σχήμα [12].

Η μετάδοση μηνυμάτων με αυτόν τον τρόπο είναι συνεπής με τις λειτουργικές ασυμμετρίες στις πραγματικές ιεραρχίες του φλοιού, όπου οι συνδέσεις που μεταφέρουν το σφάλμα πρόβλεψης είναι κατευθυντήριες, ενώ αυτές που παράγουν τη μη γραμμική είσοδο των αισθητήρων είναι κατευθυντήριες με ρυθμιστικά χαρακτηριστικά.

4.7 Η αρχή του επαρκούς κώδικα

Η αρχή του επαρκούς κώδικα ισχυρίζεται ότι ο εγκέφαλος οπτικοποιεί την κοινή πληροφορία μεταξύ των αισθήσεων και της εσωτερικής αναπαράστασης, υπό τον περιορισμό της αποδοτικότητας αυτών των αναπαραστάσεων. Αυτός ο τρόπος σκέψης διατυπώθηκε αρχικά από τον Barlow με βάση την αρχή συνεχούς ελάττωσης και πήρε μορφή αργότερα με βάση την αρχή μέγιστης πληροφορίας. Έχει βρει εφαρμογή στη μηχανική μάθηση, οδηγώντας σε μεθόδους όπως η ανεξάρτητη ανάλυση περιεχομένων, και στη νευροβιολογία, συμβάλλοντας στην κατανόηση της φύσης των νευρικών αντιδράσεων. Αυτή η αρχή είναι ιδιαίτερα αποτελεσματική στην πρόβλεψη των εμπειρικών χαρακτηριστικών των κλασικών αντιληπτικών πεδίων και παρέχει τις αρχές για την εξήγηση του διαχωρισμού των επεξεργαζόμενων ροών στις οπτικές ιεραρχίες. Έχει επεκταθεί ώστε να καλύπτει δυναμικές και κινητικές τροχιές, ενώ ακόμα χρησιμοποιείται για το συμπερασμό των μεταβολικών περιορισμών στις νευρολογικές διαδικασίες.

Πολύ απλά, η αρχή μέγιστης πληροφορίας λέει ότι η νευρική δραστηριότητα πρέπει να κωδικοποιεί την πληροφορία των αισθητήρων με έναν επαρκή και φειδωλό τρόπο. Εξετάζει την αντιστοίχιση μεταξύ δύο ομάδων μεταβλητών (των καταστάσεων των αισθητήρων και των μεταβλητών που αντιπροσωπεύουν αυτές τις καταστάσεις). Σε πρώτη ματιά, φαίνεται να αποκλείει μια πιθανολογική αναπαράσταση, επειδή αυτό θα συμπεριλάμβανε την αντιστοίχιση μεταξύ των καταστάσεων των αισθητήρων και της πυκνότητας αναγνώρισης. Ωστόσο, η αρχή



Σχήμα 11: Το τραγούδι των πουλιών και αντιληπτική κατηγοριοποίηση: (a): Το γεννητικό μοντέλο του τραγουδιού πουλιών που χρησιμοποιείται σε αυτή την προσομοίωση περιλαμβάνει δύο ελεγχόμενες παραμέτρους (ή αιτιατές καταστάσεις) v_1 και v_2, οι οποίες δίνουν δύο επιπλέον παραμέτρους σε μια συνθετική σήραγγα ώστε να παράγουν «τιτιβίσματα» που ρυθμίστηκαν σε πλάτος και συχνότητα (ένα παράδειγμα δίνεται στο υπερηχογράφημα). Τα τιτιβίσματα παρουσιάστηκαν ως ερεθίσματα σε ένα συνθετικό πουλί για να δούμε αν μπορεί να συμπεράνει τις υποβόσκουσες αιτιατές καταστάσεις και επομένως να κατηγοριοποιήσει το τραγούδι. Αυτό εμπεριέχει την ελαχιστοποίηση της ελεύθερη ενέργειας αλλάζοντας την εσωτερική αναπαράσταση (μ_{v1}, μ_{v2}) των ελεγχόμενων παραμέτρων. (b): Τρία προσομοιωμένα τραγούδια δίνονται σε μορφή υπερηχογραφήματος. Το καθένα εμπεριέχει μια σειρά από τιτιβίσματα, η συχνότητα και ο αριθμός των οποίων μειώνεται προοδευτικά από το τραγούδι a ως το c, ενώ μειώνεται και η αιτιατή κατάσταση. (c): Η γραφική στα αριστερά απεικονίζει τις σχετικές προσδοκίες (μ_{v1}, μ_{v2}) των αιτιατών καταστάσεων, δοθέντων ως συνάρτηση του χρόνου για τα τρία τραγούδια. Δείχνει ότι οι αιτίες αναγνωρίζονται έπειτα από περίπου 600 ms με υψηλή σχετική ακρίβεια της τάξης του 90%. Η γραφική στα δεξιά δείχνει τη σχετική πυκνότητα για τις αιτίες λίγο πριν το τέλος του χρόνου. Οι μπλε τελείες αντιστοιχούν στις σχετικές προσδοκίες και οι γκρι περιοχές στο 90% των σχετικών προσδοκώμενων περιοχών. Αξίζει να σημειωθεί ότι αυτές περικλείουν τις πραγματικές τιμές (κόκκινες τελείες) των v_1, v_2 που χρησιμοποιήθηκαν για να παραχθούν τα τραγούδια. Αυτά τα αποτελέσματα διευκρινίζουν τη φύση της αντιληπτικής κατηγοριοποίησης υπό το συμπερασματικό σχήμα που χρησιμοποιήθηκε. Συγκεκριμένα η αναγνώριση σηματοδοτεί την αντιστοίχιση από μια συνεχώς εναλλασσόμενη και χαοτική είσοδο του αισθητήρα με ένα σταθερό σημείο στον αντιληπτικό χώρο [12].

μέγιστης πληροφορίας μπορεί να εφαρμοστεί στα επαρκή στατιστικά μιας πυκνότητας αναγνώρισης. Σε αυτό το πλαίσιο, η αρχή μέγιστης πληροφορίας γίνεται μια ειδική περίπτωση της ΑΕΕ, η οποία προκύπτει όταν αγνοούμε την αβεβαιότητα σε πιθανολογικές αναπαραστάσεις. Αυτό φαίνεται εύκολα με την παρατήρηση ότι τα σήματα των αισθητήρων παράγονται από αιτίες. Αυτό σημαίνει ότι είναι επαρκές να αναπαραστήσουμε τις αιτίες για να προβλέψουμε αυτά τα σήματα. Πιο τυπικά, η αρχή μέγιστης πληροφορίας μπορεί να γίνει κατανοητή ως αποσύνθεση της ελεύθερης ενέργειας σε πολυπλοκότητα και ακρίβεια, καθώς η κοινή πληροφορία οπτικοποιείται όταν οι σχετικές προσδοκίες μεγιστοποιούν την ακρίβεια (ή ελαχιστοποιούν το σφάλμα πρόβλεψης) και η αποδοτικότητα εξασφαλίζεται από την ελαχιστοποίηση της πολυπλοκότητας. Αυτό εγγυάται ότι δεν εφαρμόζονται υπερβολικές παράμετροι στο γεννητικό μοντέλο και οδηγεί σε μια φειδωλή αναπαράσταση των δεδομένων των αισθητήρων που συμμορφώνονται στους προγενέστερους περιορισμούς των αιτιών τους. Ενδιαφέρον είναι ότι εξελιγμένες τεχνικές οπτικοποίησης χρησιμοποιούν την οπτικοποίηση της ελεύθερης ενέργειας για να εξαλείψουν άφθονες παραμέτρους, δηλώνοντας ότι η οπτικοποίηση ελεύθερης ενέργειας ίσως παρέχει μια καλή εξήγηση για το συναπτικό κλάδεμα και την ομοιόσταση που λαμβάνουν χώρα στον εγκέφαλο κατά τη διάρκεια της ανάπτυξης των νευρώνων και του ύπνου.

4.8 Προδιάθεση και προσοχή

Αιτιατές τακτικότητες κωδικοποιημένες με συναπτική αποτελεσματικότητα ελέγχουν την ντετερμινιστική εξέλιξη των καταστάσεων του περιβάλλοντος. Ωστόσο, στοχαστικές διακυμάνσεις σε αυτές τις καταστάσεις παίζουν σημαντικό ρόλο στην παραγωγή των δεδομένων του αισθητήρα. Το εύρος τους παρουσιάζεται συνήθως ως ακρίβεια, που κωδικοποιεί την αξιοπιστία των σφαλμάτων πρόβλεψης. Η ακρίβεια είναι σημαντική, ειδικά σε ιεραρχικά σχήματα, επειδή ελέγχει τη σχετική επιρροή των από κάτω προς τα πάνω σφαλμάτων πρόβλεψης και των από πάνω προς τα κάτω προβλέψεων. Οπότε πώς κωδικοποιείται η ακρίβεια στον εγκέφαλο; Στον προβλεπτικό κώδικα, η ακρίβεια ρυθμίζει το εύρος των σφαλμάτων πρόβλεψης, ώστε τα σφάλματα πρόβλεψης με υψηλή ακρίβεια να έχουν μεγαλύτερη επίπτωση στις μονάδες που κωδικοποιούν τις σχετικές προσδοκίες. Αυτό σημαίνει ότι η ακρίβεια αντιστοιχεί στο συναπτικό κέρδος των μονάδων πρόβλεψης σφάλματος. Οι πιο προφανείς υποψήφιοι για τον έλεγχο του κέρδους είναι κλασικοί νευρικοί ρυθμιστές όπως η ντοπαμίνη και η ακετυλοχολίνη, οι οποίες παρέχουν και μια σύνδεση με τις θεωρίες προσοχής και αβεβαιότητας. Ένας άλλος υποψήφιος είναι η γρήγορη συγχρονισμένη προσυναπτική είσοδος που μειώνει τις δραστικές μετασυναπτικές σταθερές και αυξάνει το σύγχρονο κέρδος.

4.9 Νευρικός Δαρβινισμός και μάθηση αξιών

Στη θεωρία επιλογής νευρικών ομάδων, η εμφάνιση νευρολογικών συγκροτημάτων γίνεται υπό το φως επιλεκτικής πίεσης. Η θεωρία έχει τέσσερα στοιχεία: επιγενετικοί μηχανισμοί δημιουργούν ένα πρωτεύον εύρος νευρικών συνδέσεων, που εκκαθαρίζονται με βάση την εξαρτώμενη από την εμπειρία πλαστικότητα για να παράγουν ένα δευτερεύον ρεπερτόριο νευρικών ομάδων. Αυτά επιλέγονται και διατηρούνται μέσω επαναλαμβανόμενων σημάτων μεταξύ των νευρικών ομάδων.

Η πλαστικότητα προκύπτει από συσχετισμένη προ- και μετασυνατική δραστηριότητα και ρυθμίζεται από την αξία. Η αξία λαμβάνει σήμα από αυξανόμενα νευρορυθμιστικά μεταβιβαστικά συστήματα και ελέγχει το ποιες νευρικές ομάδες θα επιλέγονται και ποιες όχι. Η ομορφιά του νευρικού Δαρβινισμού είναι ότι ομαδοποιεί διακριτές επιλεκτικές διαδικασίες. Με άλλα λόγια, αποφεύγει την επιλογή μιας ατομικής μονάδας και εκμεταλλεύεται την ιδέα της επιλογής επιλεκτικών μηχανισμών. Σε αυτή την περίπτωση, η αξία παρέχει εξελικτική αξία με την επιλογή νευρικών ομάδων που μελετούν προσαρμοστικές από ερέθισμα σε ερέθισμα ενώσεις και από ερέθισμα σε ανταπόκριση συνδέσεις. Αυτή η ικανότητα της αξίας εξασφαλίζεται από τη φυσική επιλογή, με την έννοια ότι τα νευρικά συστήματα αξίας είναι τα ίδια υποκείμενα σε επιλεκτική πίεση.

Αυτή η θεωρία, συγκεκριμένα η εξαρτώμενη από την αξία μάθηση, έχει βαθιές συνδέσεις με την ενισχυμένη μάθηση και σχετικές προσεγγίσεις στη μηχανική, όπως ο δυναμικός προγραμματισμός και οι ενότητες χρονικής διαφοράς. Αυτό προκύπτει επειδή τα νευρικά συστήματα αξίας ενισχύουν τις μεταξύ τους συνδέσεις, επιτρέποντας επομένως στον εγκέφαλο να σηματοδοτήσει μια κατάσταση αισθητήρα ως αξιόλογη αν και μόνο αν αυτή οδηγεί σε μια άλλη αξιόλογη κατάσταση. Αυτό εξασφαλίζει ότι οι παράγοντες περνάνε από μια διαδοχή καταστάσεων, οι οποίες έχουν τη δυνατότητα πρόσβασης σε καταστάσεις με γενετικά προσδιορισμένη έμφυτη αξία. Πώς όμως αυτό συσχετίζεται με την οπτικοποίηση της ελεύθερης ενέργειας;

Η απάντηση είναι απλή, η αξία είναι αντιστρόφως ανάλογη της έκπληξης, με την έννοια ότι η πιθανότητα ενός φαινοτύπου να βρίσκεται σε μια συγκεκριμένη κατάσταση αυξάνεται ανάλογα με την αξία της κατάστασης. Επιπλέον, η εξελικτική αξία του φαινοτύπου είναι η μέση αρνητική έκπληξη όλων των καταστάσεων που βιώνει, δηλαδή είναι απλώς η αρνητική εντροπία. Όντως, το όλο νόημα της ελαχιστοποίησης της ελεύθερης ενέργειας είναι να εξασφαλίσει ότι οι παράγοντες ξοδεύουν τον περισσότερο χρόνο τους σε έναν μικρό αριθμό αξιόλογων καταστάσεων. Αυτό σημαίνει ότι η ελεύθερη ενέργεια είναι το συμπλήρωμα της αξίας, ενώ ο μακροχρόνιος μέσος όρος της είναι το συμπλήρωμα της προσαρμοστικής καταλληλότητας. Αλλά πώς γνωρίζουν οι παράγοντες τι είναι αξιόλογο; Με άλλα λόγια, πώς μια γενιά λέει στην επόμενη ποιες καταστάσεις έχουν αξία; Αξία ή έκπληξη καθορίζονται από τη μορφή του γεννητικού μοντέλου του παράγοντα, που καθορίζει την αξία των καταστάσεων των αισθητήρων που κληρονομούνται μέσω γενετικών και επιγενετικών μηχανισμών. Αυτό σημαίνει ότι οι προγενέστερες προσδοκίες μπορούν να ορίσουν ένα μικρό αριθμό καταστάσεων με έμφυτη αξία. Αυτό επιτρέπει στη φυσική επιλογή να οπτικοποιήσει προγενέστερες προσδοκίες και να εξασφαλίσει ότι αυτές είναι συνεπείς με το φαινότυπο του παράγοντα. Πιο απλά, αξιόλογες καταστάσεις είναι οι καταστάσεις που ο παράγοντας προσδοκεί να είναι συχνότερες. Αυτές οι προσδοκίες περιορίζονται από τη μορφή του γεννητικού μοντέλου, το οποίο καθορίζεται γενετικά και συμπληρώνεται συμπεριφορικά, υπό ενεργό συμπερασμό.

Είναι σημαντικό να εκτιμηθεί ότι οι προγενέστερες προσδοκίες περιλαμβάνουν όχι μόνο το τι θα ληφθεί σα δείγμα από το περιβάλλον αλλά και το πώς θα ληφθεί αυτό το δείγμα. Αυτό σημαίνει ότι η φυσική επιλογή μπορεί να εξοπλίσει τους παράγοντες με προγενέστερες προσδοκίες ώστε να εξερευνήσουν το περιβάλλον μέχρι να συναντηθούν καταστάσεις με την έμφυτη αξία.

Τόσο ο νευρικός Δαρβινισμός όσο και η ΑΕΕ προσπαθούν να κατανοήσουν τις σωματικές αλλαγές ενός ατόμου στην περίπτωση της εξέλιξης. Ο νευρικός Δαρβινισμός επικαλείται τις επιλεκτικές διαδικασίες, ενώ η ΑΕΕ χρησιμοποιεί την οπτικοποίηση του συνόλου των δυναμικών με μορφή εντροπίας και έκπληξης. Το κύριο θέμα που προκύπτει εδώ είναι ότι οι κληροδοτούμενες προγενέστερες προσδοκίες μπορούν να σηματοδοτήσουν πράγματα ως εμφύτως αξιόλογα, αλλά πώς μπορούν οι σηματοδοτημένες καταστάσεις να προκαλέσουν προσαρμοστική συμπεριφορά; Η απάντηση σε αυτό το ερώτημα θα αναλυθεί αργότερα.

4.10 Θεωρία βέλτιστου ελέγχου και θεωρία παιγνίων

Η αξία είναι το κέντρο για θεωρίες σχετικές με τη λειτουργία του εγκεφάλου που βασίζονται στην ενισχυμένη μάθηση και το βέλτιστο έλεγχο. Η βασική ιδέα που υποστηρίζει αυτές τις προσεγγίσεις είναι ότι ο εγκέφαλος οπτικοποιεί την αξία, η οποία αναμένεται ως ανταμοιβή ή ωφέλεια (ενώ το συμπλήρωμα της αναμένεται ως απώλεια ή κόστος). Αυτό φαίνεται στη συμπεριφορική ψυχολογία ως ενισχυμένη μάθηση, στην πληροφορική νευροεπιστήμη και στη μηχανική μάθηση ως παραλλαγές του δυναμικού προγραμματισμού και στα οικονομικά ως θεωρία προσδοκώμενου κέρδους. Η ιδέα του προσδοκώμενου κέρδους ή κόστους είναι σημαντική εδώ, αφού μιλάμε για το προσδοκώμενο κόστος για μελλοντικές καταστάσεις, δεδομένης μιας συγκεκριμένης πολιτικής που ορίζει τη δράση ή τις επιλογές. Μια πολιτική καθορίζει τις καταστάσεις στις οποίες θα μεταπηδήσει ένας παράγοντας από κάποια δεδομένη κατάσταση. Αυτή η πολιτική πρέπει να έχει πρόσβαση σε σποραδικές ωφέλιμες καταστάσεις χρησιμοποιώντας μια ζημιογόνα συνάρτηση, η οποία σηματοδοτεί τις καταστάσεις ως ωφέλιμες ή όχι. Το πρόβλημα για το πώς οπτικοποιείται αυτή η πολιτική αντιμετωπίζεται από τη θεωρία βέλτιστου ελέγχου όπως η εξίσωση Bellman και οι παραλλαγές της, η οποία εκφράζει την αξία ως συνάρτηση της βέλτιστης πολιτικής και μιας ζημιογόνας συνάρτησης. Αν κάποιος μπορεί να λύσει την εξίσωση Bellman, μπορεί και να συσχετίσει κάθε κατάσταση των αισθητήρων με μια αξία και να οπτικοποιήσει την πολιτική εξασφαλίζοντας ότι η επόμενη κατάσταση είναι η πιο αξιόλογη από τις διαθέσιμες καταστάσεις. Γενικά, είναι αδύνατο να βρεθεί ακριβής λύση για την εξίσωση Bellman, αλλά υπάρχουν αρκετές προσεγγίσεις, με εύρος από απλά μοντέλα Rescorla-Wagner μέχρι πιο περιεκτικούς σχηματισμούς όπως η Q-μάθηση. Το κόστος παίζει επίσης σημαντικό ρόλο στη θεωρία Μπεϋζιανής απόφασης, στην οποία οι βέλτιστες αποφάσεις ελαχιστοποιούν το προσδοκώμενο κόστος στην περίπτωση αβεβαιότητας σχετικά με το αποτέλεσμα [12].

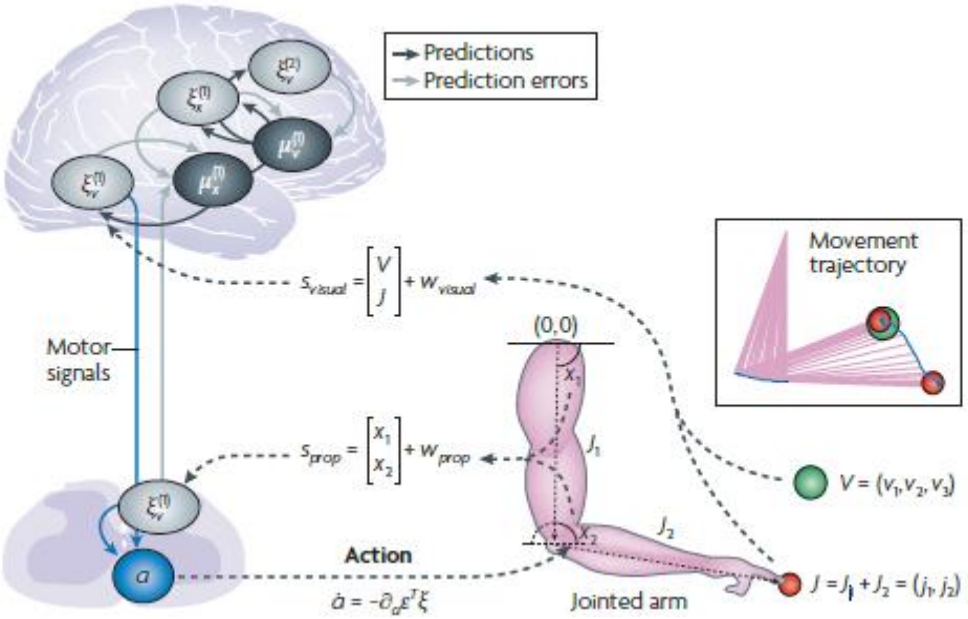
Ποια είναι λοιπόν η συμβολή της ελεύθερης ενέργειας; Αν κάποιος θεωρήσει ότι η βέλτιστη πολιτική εκτελεί μια βαθμιαία αύξηση της αξίας, τότε είναι εύκολο να δείξουμε ότι η αξία είναι αντιστρόφως ανάλογη της έκπληξης. Αυτό σημαίνει ότι η ελεύθερη ενέργεια είναι ένα άνω όριο προσδοκώμενου κόστους, το οποίο έχει λογική αφού η θεωρία βέλτιστου ελέγχου θεωρεί πως η δράση ελαχιστοποιεί το προσδοκώμενο κόστος, ενώ η ΑΕΕ δηλώνει ότι αυτή ελαχιστοποιεί την ελεύθερη ενέργεια. Αυτό είναι σημαντικό καθώς εξηγεί γιατί οι παράγοντες πρέπει να ελαχιστοποιούν το προσδοκώμενο κόστος. Επιπρόσθετα, η ελεύθερη ενέργεια παρέχει μια ποσοτικοποιημένη σύνδεση μεταξύ των συναρτήσεων κόστους της ενισχυμένης μάθησης και της αξίας στην εξελικτική βιολογία. Τελικά, η δυναμική προοπτική παρέχει μια μηχανιστική επίγνωση στο πώς προσδιορίζονται οι πολιτικές στον εγκέφαλο. Σύμφωνα με την αρχή βελτιστότητας το

κόστος είναι ο βαθμός της αλλαγής της αξίας, που εξαρτάται από τις αλλαγές στις καταστάσεις των αισθητήρων. Αυτό δηλώνει ότι οι βέλτιστες πολιτικές μπορούν να οριστούν από τις προγενέστερες προσδοκίες σχετικά με την κίνηση των καταστάσεων των αισθητήρων. Πιο απλά, τα προγενέστερα καθορίζουν ένα συγκεκριμένο σημείο στο οποίο, όταν φτάσουν οι καταστάσεις, η αξία θα σταματήσει να αλλάζει και το κόστος θα έχει ελαχιστοποιηθεί. Ένα απλό παράδειγμα φαίνεται στο **Σχήμα 12**, στο οποίο η κίνηση ενός χεριού σε ανάκληση προσομοιώνεται χρησιμοποιώντας μόνο τις προγενέστερες προσδοκίες ότι το χέρι θα φτάσει σε ένα συγκεκριμένο σημείο-στόχο. Αυτή η αναπαράσταση επεξηγεί το πώς ο υπολογιστικός έλεγχος κίνησης μπορεί να διατυπωθεί με βάση τα προγενέστερα και την καταπίεση των σφαλμάτων πρόβλεψης των αισθητήρων. Πιο γενικά, δείχνει το πώς οι ανταμοιβές και οι στόχοι μπορούν να θεωρηθούν ως προγενέστερες προσδοκίες, τις οποίες η δράση είναι υποχρεωμένη να εκπληρώσει. Επίσης δηλώνει το πώς η φυσική επιλογή μπορεί να οπτικοποιήσει τη συμπεριφορά μέσω του γενετικού προσδιορισμού των κληροδοτούμενων ή έμφυτων προγενέστερων που εξαναγκάζουν τη μάθηση των εμπειρικών προγενέστερων και της ακόλουθης με βάση το στόχο δράσης.

Πρέπει να σημειωθεί ότι απλά η προσδοκία ότι θα φτάσουμε σε κάποιες καταστάσεις μπορεί να μην επαρκεί ώστε όντως να παρευρεθούμε σε αυτές. Αυτό ισχύει διότι κάποιος μπορεί να πρέπει να τις προσεγγίσει μέσω πολλών ενδιάμεσων καταστάσεων (για παράδειγμα να προηγείται η ανάγκη αποφυγής ενός εμποδίου) ή να πρέπει να συμμορφωθεί με τους φυσικούς περιορισμούς της δράσης. Αυτά είναι δυσκολότερα προβλήματα για την κατάληξη σε μια απομακρυσμένη ανταμοιβή τα οποία αντιμετωπίζουν η ενισχυμένη μάθηση και ο βέλτιστος έλεγχος. Σε αυτές τις περιπτώσεις, η εξέταση των δυναμικών πυκνότητας, πάνω στις οποίες βασίζεται η ΑΕΕ, προτείνει ότι επαρκεί να συνεχίζεται η κίνηση έως ότου φτάσουμε σε μια εκ των προτέρων γνωστή κατάσταση. Αυτό συμπεριλαμβάνει την καταστροφή απροσδόκητων καθορισμένων σημείων του περιβάλλοντος κάνοντας τα ασταθή (όπως για παράδειγμα η αλλαγή θέσης όταν καθόμαστε κάπου άβολα). Μαθηματικά, αυτό σημαίνει την υιοθέτηση μιας πολιτικής που εξασφαλίζει μια θετική απόκλιση στις ζημιογόνες καταστάσεις. Στο **Σχήμα 13** φαίνεται η λύση ενός τέτοιου προβλήματος όπου ένα απλό προγενέστερο επιφέρει μια τέτοιου είδους πολιτική. Προγενέστερα τέτοιου είδους μπορούν να παρέχουν έναν δομημένο τρόπο για την κατανόηση της αντιστάθμισης εξερεύνησης και εκμετάλλευσης και άλλων σχετικών θεμάτων στην εξελικτική βιολογία. Η σιωπηρή χρήση των προγενέστερων για την προτροπή δυναμικής αστάθειας επίσης παρέχει μια βασική σύνδεση για τις προσεγγίσεις της θεωρίας δυναμικών συστημάτων για τον εγκέφαλο που δίνουν έμφαση στη σημασία των περιοδευόντων δυναμικών, της μεταστατότητας και της αυτό-οργανωτικής κρισιμότητας. Αυτά τα δυναμικά φαινόμενα παίζουν σημαντικό ρόλο στις συνεργατικές και αυτοπαθείς πτυχές της προσαρμοστικής συμπεριφοράς [12].

Συμπερασματικά, ο βέλτιστος έλεγχος και η θεωρία παιγνίων ξεκινούν με την ιδέα του κόστους και προσπαθούν να κατασκευάσουν συναρτήσεις αξίας των καταστάσεων, οι οποίες μεταγενέστερα οδηγούν τη δράση. Η διατύπωση της ελεύθερης ενέργειας με το όριο της ελεύθερης ενέργειας για την αξία των καταστάσεων, οι οποίες προσδιορίζονται από τα προγενέστερα με την κίνηση για τις κρυφές καταστάσεις του περιβάλλοντος. Αυτά τα προγενέστερα μπορούν να ενσωματώσουν κάθε συνάρτηση κόστους ώστε να εξασφαλίσουν ότι θα αποφευχθούν οι ζημιογόνες καταστάσεις. Ουσιαστικά, το πρόβλημα του να βρίσκουμε

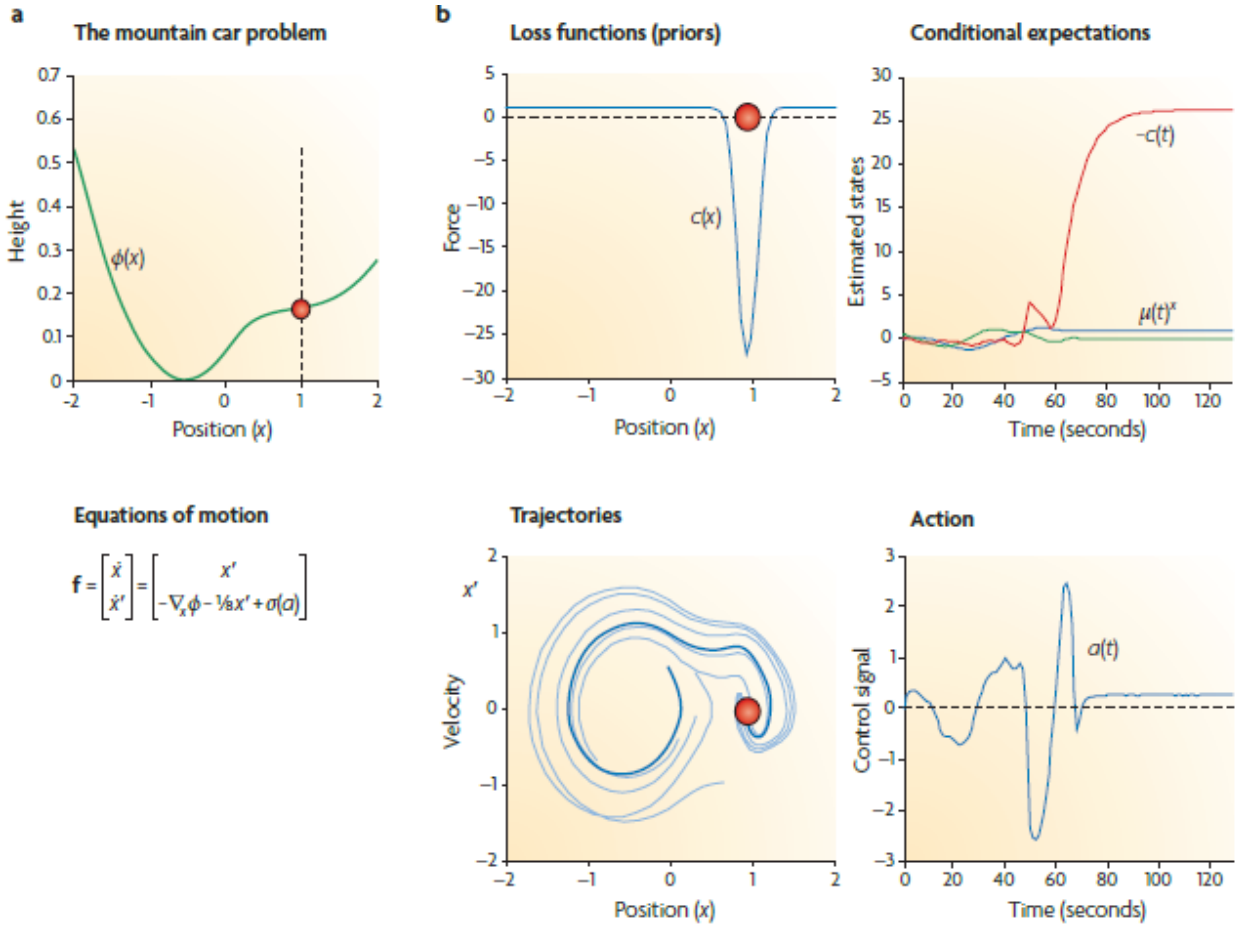
ανταμοιβή στο περιβάλλον είναι η φυσική λύση του προβλήματος ελαχιστοποίησης της εντροπίας για τις καταστάσεις ενός παράγοντα.



Σχήμα 12: Μια επίδειξη ενός χεριού σε κίνηση. Το κάτω δεξιά κομμάτι του σχήματος δείχνει έναν κινητήρα, που περιλαμβάνει ένα χέρι σε ανάκληση με δύο αρθρώσεις και δύο κρυφές καταστάσεις, καθεμία από τις οποίες αντιστοιχεί σε μια συγκεκριμένη θέση των δύο αρθρώσεων: η τωρινή θέση του δάχτυλου (κόκκινος κύκλος) είναι το άθροισμα των διανυσμάτων που περιγράφουν την τοποθεσία της κάθε άρθρωσης. Εδώ, οι αιτιώδεις καταστάσεις του περιβάλλοντος είναι η θέση και η λάμψη του στόχου (πράσινος κύκλος). Το χέρι υπακούει στους μηχανισμούς του Νεύτωνα, ορισμένους ως προς τη γωνιώδη αδράνεια και τριβή. Το αριστερό κομμάτι του σχήματος παρουσιάζει ότι ο εγκέφαλος αισθάνεται τις κρυφές καταστάσεις άμεσα ως προς την ιδιοδεκτική είσοδο (S_{prop}) που δίνει σήμα για τις γωνιώδεις θέσεις (x_1, x_2) των αρθρώσεων και έμμεσα μέσω της όψης της τοποθεσίας του δάχτυλου στο χώρο (J_1, J_2). Επιπλέον, μέσω της οπτικής εισόδου (S_{visual}) ο παράγοντας αισθάνεται την τοποθεσία στόχο (v_1, v_2) και τη λάμψη (v_3). Τα σφάλματα πρόβλεψης του αισθητήρα περνάνε σε υψηλότερα επίπεδα του εγκεφάλου για να οπτικοποιήσουν τις σχετικές προσδοκίες των κρυφών καταστάσεων (των θέσεων των αρθρώσεων) και των αιτιωδών καταστάσεων (του στόχου). Οι επακόλουθες προβλέψεις επιστρέφονται πίσω ώστε να καταπιεστούν τα σφάλματα προβλέψεων του αισθητήρα. Ταυτόχρονα, τα σφάλματα πρόβλεψης προσπαθούν να καταπιέσουν τον εαυτό τους αλλάζοντας την είσοδο των αισθητήρων μέσω της δράσης. Οι γκρι και μπλε γραμμές δηλώνουν αμφίδρομο πέρασμα μηνυμάτων μεταξύ των νευρικών πληθυσμών που κωδικοποιούν το σφάλμα πρόβλεψης και των σχετικών προσδοκιών. Οι μπλε γραμμές αντιπροσωπεύουν φθίνοντα σήματα ελέγχου κίνησης από τις μονάδες πρόβλεψης σφάλματος του αισθητήρα. Το γεννητικό μοντέλο του παράγοντα σε συνδυασμό με τα προγενέστερα των κρυφών καταστάσεων δημιουργούν έναν ελαστικό δεσμό μεταξύ του δάχτυλου και του στόχου. Αυτό προκαλεί μια προγενέστερη προσδοκία ότι το δάχτυλο θα φτάσει στο στόχο, εφόσον έχει και τη σωστή κλίση. Στο κουτί φαίνεται η κίνηση που θα προκληθεί από τη δράση ώστε να επιτευχθεί ο στόχος. Οι κόκκινοι κύκλοι δείχνουν την αρχική και τελική θέση του δάχτυλου [12].

4.11 Συμπεράσματα και μελλοντικές κατευθύνσεις

Είδαμε λοιπόν ότι αρκετές θεωρίες σχετικά με τη λειτουργία του εγκεφάλου μπορούν να βρουν μια κοινή στέγη υπό την αντίληψη του Helmholtz που θεωρεί τον εγκέφαλο ως ένα γεννητικό μοντέλο του κόσμου στον οποίο ζει, με τα πιο αξιοσημείωτα παραδείγματα να περιλαμβάνουν την ενσωμάτωση του Μπεϋζιανού εγκέφαλου και την υπολογιστική θεωρία ελέγχου κίνησης, τις αντικειμενικές συναρτήσεις που προκύπτουν από προβλεπτικό κώδικα και την αρχή μέγιστης



Σχήμα 13: Η επίλυση του προβλήματος του τζιπ αυτοκινήτου με προγενέστερες προσδοκίες. (a): Πώς η παράδοξη αλλά προσαρμοστική συμπεριφορά (για παράδειγμα, η απομάκρυνση από έναν στόχο για να εξασφαλίσουμε ότι θα τον ασφαλίσουμε αργότερα) προκύπτει από απλά προγενέστερα με την κίνηση κρυφών καταστάσεων του περιβάλλοντος. Δεδομένο είναι το τοπίο ή η συνάρτηση πιθανής ενέργειας (με ελάχιστο x=-0.5) η οποία ασκεί δυνάμεις στο τζιπ. Το αυτοκίνητο δίνεται στη θέση-στόχο στο λόφο x=1 και συμβολίζεται με τον κόκκινο κύκλο. Οι εξισώσεις κίνησης του τζιπ δίνονται κάτω από τη γραφική. Για x=0 η δύναμη στο τζιπ δεν μπορεί να ξεπεραστεί από τον παράγοντα, γιατί μια συνάρτηση πίεσης (-1<σ<1) εφαρμόζεται ώστε να την εμποδίσει να γίνει μεγαλύτερη του 1. Αυτό σημαίνει ότι ο παράγοντας μπορεί να φτάσει στο στόχο μόνο ξεκινώντας από τα μισά του αριστερού λόφου ώστε να έχει αρκετή ορμή και να φτάσει στην άλλη πλευρά. (b): Τα αποτελέσματα του ενεργούς συμπερασμού υπό τα προγενέστερα που αποσταθεροποιούν σταθερά σημεία έξω από το πεδίο ορισμού του στόχου. Τα προγενέστερα κωδικοποιούνται από μια συνάρτηση κόστους c(x) (πάνω αριστερά), που δρα σαν αρνητική τριβή. Όταν η τριβή είναι αρνητική το αμάξι προσδοκά να πάει γρηγορότερα. Οι συμπερασμένες κρυφές καταστάσεις (πάνω δεξιά: θέση με μπλε, ταχύτητα με πράσινο και η αρνητική διάλυση με κόκκινο) δείχνουν ότι το αυτοκίνητο εξερευνεί το τοπίο του μέχρι να βρεθεί στο στόχο και ότι τότε η τριβή αυξάνεται (οπότε και το κόστος μειώνεται) για να εμποδίσουν το αμάξι να φύγει από το στόχο (με πιθανή πτώση από το λόφο). Η επακόλουθη τροχιά δίνεται με μπλε (κάτω αριστερά). Οι πιο αχνές γραμμές παρέχουν ένα υπόδειγμα τροχιών από άλλες δοκιμές, με διαφορετικά σημεία εκκίνησης. Στον πραγματικό κόσμο, η τριβή είναι σταθερά. Ωστόσο, το αυτοκίνητο «προσδοκεί» η τριβή να αλλάξει καθώς αυτό αλλάζει θέση, επιβάλλοντας έτσι την εξερεύνηση ή την εκμετάλλευση. Αυτές οι προσδοκίες εκπληρώνονται μέσω της δράσης (κάτω δεξιά) [12].

πληροφορίας, τον ιεραρχικό συμπερασμό και τις θεωρίες προσοχής, την ενσωμάτωση της αντίληψης στη φυσική επιλογή και τη σύνδεση μεταξύ οπτικού ελέγχου και πιο εξωτικών φαινομένων στη θεωρία δυναμικών συστημάτων. Η σταθερή βάση σε όλες αυτές τις θεωρίες είναι ότι ο εγκέφαλος οπτικοποιεί ένα όριο (ελεύθερης ενέργειας) σχετικά με την έκπληξη ή το

συμπλήρωμά της, την αξία. Αυτό δηλώνεται ως αντίληψη (δηλαδή την αλλαγή των προβλέψεων) ή ως δράση (δηλαδή την αλλαγή των αισθήσεων που προβλέπονται). Σημαντικό είναι ότι αυτές οι προβλέψεις εξαρτώνται από τις προγενέστερες προσδοκίες, που οπτικοποιούνται σε διαφορετικά χρονοδιαγράμματα και καθορίζουν τι θεωρείται αξιόλογο.

Τι προμηνύει όμως η ΑΕΕ για το μέλλον; Αν η βασική της συνεισφορά είναι να ενοποιεί καθιερωμένες θεωρίες, τότε η απάντηση είναι «όχι πολλά». Αντίθετα, μπορεί να παρέχει ένα πλαίσιο στο οποίο οι πρόσφατες σχετικές διαμάχες μπορούν να λυθούν, αν για παράδειγμα μέσω της ντοπαμίνης κωδικοποιείται η αξία, το σφάλμα πρόβλεψης ή η έκπληξη, κάτι ιδιαίτερα σημαντικό για την κατανόηση συνθηκών όπως ο εθισμός, η ασθένεια Πάρκινσον και η σχιζοφρένεια. Όντως, η διατύπωση της ελεύθερης ενέργειας έχει ήδη χρησιμοποιηθεί για να εξηγηθούν τα θετικά συμπτώματα της σχιζοφρένειας από άποψη λάθους συμπερασμού.

5. Άλλες σχετικές θεωρίες

5.1 Καθολικός χώρος εργασίας

Ξεκινάμε τις αναφορές μας σε άλλες σχετικές με τη συνείδηση θεωρίες, χωρίς βέβαια να γίνει κάποια ιδιαίτερα εκτεταμένη ανάλυση, με τη θεωρία καθολικού χώρου εργασίας, η οποία ως κύριο εκφραστή της έχει τον καθηγητή Θεωρητικής Νευροβιολογίας Bernard Baars. Για τον Baars η συνείδηση είναι μια λειτουργική βιολογική προσαρμογή , ένα είδος πύλης, μια λειτουργία «πρόσβασης, μετάδοσης και ανταλλαγής πληροφοριών» καθώς και άσκησης καθολικού συντονισμού και ελέγχου». Έτσι, κατά τον Baars η συνείδηση είναι απαραίτητη για την ενοποίηση της αντίληψης, της σκέψης και της δράσης και για την προσαρμογή σε νέες καταστάσεις και στην παροχή πληροφοριών στο σύστημα του εαυτού. Συνεπώς, στα πλαίσια της θεωρίας αυτής, οποιοδήποτε κομμάτι πληροφορίας είναι συνειδητό εφόσον αναμεταδίδεται σε πολλές περιοχές του μη-συνειδητού εγκεφάλου.

Προκειμένου να προσεγγίσει το θέμα της συνείδησης χρησιμοποιεί ως στοιχεία τα αποτελέσματα της «συγκριτικής ανάλυσης», που προκύπτουν μέσω μιας σύγκρισης ανάμεσα σε όμοιες συνειδητές και μη-συνειδητές διεργασίες, θεωρώντας πως με αυτό το τρόπο αντιμετωπίζει άμεσα το πρόβλημα της υποκειμενικής εμπειρίας.

Έτσι για παράδειγμα, μια σύγκριση είναι μεταξύ του τεράστιου αριθμού μη συνειδητών διεργασιών και της πολύ περιορισμένης χωρητικότητας της συνείδησης, που φαίνεται να δρα ως πύλη, δημιουργώντας πρόσβαση σε διάφορα μέρη του νευρικού συστήματος. Με τον τρόπο αυτό η συνείδηση δημιουργεί καθολική πρόσβαση στις διεργασίες και τα αποτελέσματα αυτών που λαμβάνουν χώρα στα διάφορα νευρωνικά δίκτυα του εγκεφάλου. Ο Baars στοχεύει στην κατανόηση της εστιακής συνείδησης εύκολα περιγραφόμενων γεγονότων όπως πχ. βλέπω μια τυπωμένη σελίδα, τα οποία μπορούν να αναφερθούν με ακρίβεια, όταν αυτό γίνεται άμεσα, χωρίς διασπαστές και με παρουσία εξωτερικού παρατηρητή που μπορεί να επαληθεύσει, δηλαδή με επαληθεύσιμη δημόσια αναφορά.

Ο εγκέφαλος είναι ένα ογκώδες σύνολο νευρωνικών δικτύων, στρωμάτων και συνδέσεων εξειδικευμένων σε συγκεκριμένα έργα, το μεγαλύτερο μέρος των οποίων δρα παράλληλα, χωρίς να γίνεται συνειδητό και με μεγάλη αποτελεσματικότητα στην εκτέλεση έργων ρουτίνας. Το ερώτημα, λέει ο Baars βρίσκεται στο γιατί οι συνειδητές πλευρές του συστήματος είναι τόσο περιορισμένες ενώ οι ασυνείδητες τόσο μεγάλες; Για να απαντήσει σε αυτό το ερώτημα επιστρατεύει την περιορισμένη- σύμφωνα με το νόμο του Miller - ενεργό μνήμη, την επιλεκτικότητα της συνειδητής εμπλοκής μας με ένα ρεύμα πληροφοριών και την ύπαρξη αυτού του τεράστιου συνόλου μη συνειδητών διεργασιών που δρουν παράλληλα και κατανεμημένα, ιδιότητες που επιλέχθηκαν μέσα από εκατομμύρια χρόνια βιολογικής εξέλιξης και προσαρμογής. Τα παραπάνω, υποστηρίζει ο Baars, είναι στοιχεία που ενισχύουν την ιδέα του πως «μπορούμε να δημιουργήσουμε πρόσβαση σε οποιοδήποτε μέρος του εγκεφάλου μας χρησιμοποιώντας τη συνείδηση» [13].

Για παράδειγμα, ασυνείδητη, λέει ο Baars είναι και η τεράστια αυτοβιογραφική μας μνήμη, που βρίσκεται με κάποιο τρόπο «αποθηκευμένη» σε νευρωνικά δίκτυα του εγκεφάλου, στης οποίας όμως τις λεπτομέρειες αποκτούμε πρόσβαση μέσω της συνείδησης. Έτσι, απλώς η συνείδηση των αποτελεσμάτων είναι αρκετή για να δημιουργήσει πρόσβαση σε σύνθετα ασυνείδητα συστήματα. Αυτή η ικανότητα να δημιουργεί πρόσβαση σε δισεκατομμύρια νευρώνες είναι που καθιστά τη συνείδηση εξαιρετικά χρήσιμη, είναι το όργανο δημοσιοποίησης του εγκεφάλου, μια λειτουργία πρόσβασης, μετάδοσης και ανταλλαγής πληροφοριών, όπως και άσκησης ολικού συντονισμού και ελέγχου.

Η θεωρία του Baars, προτείνει πως το νοητικό μας σύστημα δομείται πάνω σε έναν καθολικό χώρο εργασίας τον οποίο παραλληλίζει με μια σκηνή στο θέατρο του νου. Σ' αυτό το «θέατρο», η σκηνή αντιπροσωπεύεται από την ενεργό μνήμη, όπου μη-συνειδητές επεξεργασίες ανταγωνίζονται προκειμένου να αποκτήσουν πρόσβαση στον φωτεινό προβολέα της προσοχής, και τα αποτελέσματά τους να γίνουν συνειδητά και να «αναμεταδοθούν» καθολικά στο μη συνειδητό κοινό (πχ. τη γλώσσα και τη μνήμη). Για τον Baars, αυτή η καθολική αναμετάδοση συνιστά τη συνείδηση.

Η θεωρία του Baars συνιστά μια απόπειρα προσέγγισης της συνείδησης, από την οποία όμως δεν λείπουν τα αναπάντητα ερωτήματα., όπως αυτά που αφορούν τη σχέση της συνείδησης με έννοιες όπως η προσοχή και η ενεργός μνήμη, ο εαυτός καθώς και το πρόβλημα νου-σώματος.

Ωστόσο αυτή η θεωρία έχει επηρεάσει αρκετούς μελετητές προσφέροντας ένα γόνιμο θεωρητικό πλαίσιο για τη συνείδηση, απαγκιστρωμένο από «δύσκολα προβλήματα», «υποκειμενικότητες» και «κουάλια» μέσα στο οποίο μπορούν να ενταχθούν και να συμφιλιωθούν δεδομένα από τη σύγχρονη νευροβιολογική μελέτη του εγκεφάλου

5.2 Νευρωνικός καθολικός χώρος εργασίας

Ένας από τους πρώτους ερευνητές που χρησιμοποίησε το πλαίσιο του Καθολικού Χώρου Εργασίας είναι ο Stanislas Dehaene (2001), ο οποίος υποστηρίζει πως το πρόβλημα της συνείδησης πρέπει να απλοποιηθεί έτσι ώστε να μπορεί να ελεγχθεί στο εργαστήριο. Για το λόγο αυτό επιλέγει από τις διάφορες έννοιες της συνείδησης τη μόνη που μπορεί να ελεγχθεί πειραματικά, αυτό δηλαδή που ονομάζουμε «πρόσβαση στη συνείδηση», τη μεταβατική έννοια της συνείδησης κατά την οποία είμαστε σε θέση να αντιληφθούμε ένα αισθητηριακό ερέθισμα.

Στο «νευρωνικό καθολικό χώρο εργασίας» του Dehaene η «πρόσβαση στη συνείδηση» είναι διαφορετική από την «εγρήγορση», η οποία αποτελεί προϋπόθεση της πρώτης. Για παράδειγμα, η «εγρήγορση» ή αλλιώς η μη μεταβατική έννοια της συνείδησης, σύμφωνα με το μοντέλο νευρωνικού χώρου εργασίας είναι μια διαβαθμισμένη μεταβλητή, ένα ελάχιστο επίπεδο της οποίας είναι απαραίτητο για τη τοποθέτηση των θαλαμοφλοιικών συστημάτων σε προσληπτική κατάσταση, χαμηλώνοντας το κατώφλι για αισθητηριακά εισιόντα. Η νευρωνική κατάσταση που αντιστοιχεί στην εγρήγορση περιλαμβάνει δραστηριότητα πυρήνων του στελέχους με μεγάλες προβολές στο θάλαμο και το φλοιό. Η εγρήγορση αποτελεί αναγκαία, αλλά όχι επαρκή συνθήκη για τη μεταβατική έννοια της συνείδησης, δηλαδή την «πρόσβαση στη συνειδητή αναφορά», η

οποία έχει συσχετιστεί με δραστηριότητα σε περιοχές του ταινιωτού, εξωταινιωτού και βρεγματομετωπιαίου φλοιού. Με βάση δεδομένα από παρόμοιες μελέτες, έχει προταθεί πως η συνειδητή αντίληψη ενός συγκεκριμένου οπτικού χαρακτηριστικού (πχ. χρώμα) βρίσκεται στην εξωταινιωτή περιοχή που εξειδικεύεται στο χαρακτηριστικό αυτό.

Ωστόσο, ο Dehaene και οι συνεργάτες του υποστηρίζουν πως η πρώιμη αισθητηριακή ενεργοποίηση είναι αναγκαία αλλά όχι επαρκής για τη συνειδητή πρόσβαση καθώς σε πολλές μελέτες παρατηρείται δραστηριότητα σε εξωταινιωτές περιοχές όταν οι συμμετέχοντες αρνούνται πως έχουν δει κάποιο ερέθισμα [14].

Στην άποψη αυτή συνηγορούν δεδομένα από μελέτες fMRI όπως αυτή των Beck, G Rees, C. D. Frith, & Lavie, οι οποίοι προσπάθησαν να διαφοροποιήσουν τα νευρωνικά αντίστοιχα ανάμεσα στη συνθήκη κατά την οποία οι συμμετέχοντες μπορούσαν να εντοπίσουν την αλλαγή στα ερεθίσματα που προβάλλονταν σε μια οθόνη, και τη συνθήκη κατά την οποία δεν εντόπιζαν την αλλαγή. Τα αποτελέσματα της μελέτης έδειξαν πως όταν οι συμμετέχοντες μπορούσαν να εντοπίσουν την αλλαγή, υπήρχε αυξημένη δραστηριότητα στο βρεγματικό και το δεξί πλαγιοραχιαίο προμετωπιαίο φλοιό, καθώς και σε περιοχές του εξωταινιωτού οπτικού φλοιού. Αντιθέτως στις περιπτώσεις μη εντοπισμού της αλλαγής κάποια αυξημένη ενεργοποίηση παρουσιαζόταν σε περιοχές του εξωταινιωτού, αλλά όχι στις πρόσθιες μετωπιαίες, ή στις βρεγματικές περιοχές, υποδεικνύοντας συσχέτιση των τελευταίων με την παρουσία της συνείδησης. Παρόμοια αποτελέσματα είχαν ο Dehaene και οι συνεργάτες του, όταν προσπάθησαν να διαφοροποιήσουν τα νευρωνικά αντίστοιχα κατά την παρουσίαση κρυμμένων λέξεων σε σχέση με μη κρυμμένες. Στην πρώτη περίπτωση η ενεργοποίηση περιοριζόταν σε αριστερές εξωταινιωτές περιοχές, στην αριστερή ατρακτοειδή έλικα και σε αριστερές προκεντρικές περιοχές, σε αντίθεση με τη συνθήκη παρουσίασης των μη κρυμμένων λέξεων που προκαλούσε και ενεργοποίηση επιπλέον προμετωπιαίων περιοχών. Ακόμη, οι Gross και συνεργάτες υποστήριξαν πως αλλαγές στις απαιτήσεις της οπτικής προσοχής που αφορά μια συνθήκη σχετίζεται με αλλαγές στο συγχρονισμό φάσης νευρωνικής πυροδότησης μεταξύ του δικτύου προσοχής το οποίο περιλαμβάνει, μετωπιαίες, βρεγματικές και κροταφικές περιοχές. Επιπλέον, παρά την άποψη που υποστηρίζει πως η ορατότητα συγκεκριμένων οπτικών χαρακτηριστικών όπως πχ. η φωτεινότητα, σχετίζεται μόνο με την ενεργοποίηση συγκεκριμένων περιοχών του πρώιμου οπτικού φλοιού, οι Haynes, Driver, & Rees έδειξαν πως διακυμάνσεις στην ορατότητα τέτοιων χαρακτηριστικών σχετίζονταν με την ενεργοποίηση μετέπειτα οπτικών περιοχών καθώς και βρεγματο-κροταφικών. Οι Kranczioch, Debener, Schwarzbach, Goebel, & Engel, σε μια μελέτη attentional blink έδειξαν πως, ενώ η ενεργοποίηση ινιακοκροταφικών περιοχών μπορεί να αντικατοπτρίζει κυρίως τη διάρκεια της προσοχής, οι μετωποβρεγματικές περιοχές φαίνεται να εμπλέκονται σε ένα ευρέως κατανεμημένο δίκτυο που ελέγχει την οπτική συνείδηση. Έτσι, βασιζόμενοι σε μελέτες όπως οι παραπάνω, ο Dehaene προτείνει πως η γρήγορη βρεγματομετωπιαία ενεργοποίηση και η από πάνω προς τα κάτω ενίσχυση των οπίσθιων -σχετικών με τα αισθητηριακά ερεθίσματα- περιοχών μετά την ενεργοποίηση των πρόσθιων «συνειρμικών», φαίνεται πως είναι τα πρότυπα ενεργοποίησης που δραστηριοποιούνται συστηματικά κατά τη συνθήκη πρόσβασης στη συνειδητή αναφορά, δηλαδή κατά τη συνθήκη όπου το υποκείμενο μπορεί να αναφέρει με κάποιο τρόπο, λεκτικό ή μη πως έχει αντιληφθεί ένα ερέθισμα.

Στην παρούσα προσέγγιση σημαντικό θεωρείται το θέμα της προσοχής σε σχέση με την ένταση του ερεθίσματος και τη συνειδητή αντίληψη. Έτσι, οι συγγραφείς υποστηρίζουν πως προκειμένου να υπάρξει πρόσβαση στη συνειδητή αντίληψη χρειάζεται να υπάρχει και από κάτω προς τα πάνω αισθητηριακή ένταση (του ερεθίσματος) και από πάνω προς τα κάτω ενίσχυση μέσω εστίασης της προσοχής για να προκύψει συνειδητή αντίληψη. Για αυτό και πρέπει πάντα να ελέγχεται με υποκειμενική αναφορά για κάθε δοκιμή. Έτσι, με βάση τα παραπάνω, ο Dehaene σε αντίθεση με το κλασσικό δυαδικό διαχωρισμό ανάμεσα σε συνειδητή και μη συνειδητή επεξεργασία, προτείνει δύο βασικά είδη μη συνειδητής επεξεργασίας. Συγκεκριμένα:

Α. Η υπο-οριακή επεξεργασία ορίζεται ως η κατάσταση όπου η από κάτω προς τα πάνω ενεργοποίηση δεν είναι επαρκής ώστε να διεγείρει μια μεγάλης κλίμακας ανατροφοδοτική κατάσταση σε ένα μεγάλο δίκτυο νευρώνων με μεγάλους άξονες. Ουσιαστικά πρόκειται για τη περίπτωση όπου ένα μικρής έντασης ερέθισμα, προκαλεί μια μικρή ενεργοποίηση που σύντομα όμως σβήνει χωρίς να προκαλέσει συνείδηση του ερεθίσματος.

Β. Η προσυνειδητή επεξεργασία αφορά μια διεργασία όπου ένα ερέθισμα είναι ικανό να γίνει συνειδητό, κουβαλά δηλαδή δυνάμει αρκετή ενεργοποίηση για πρόσβαση στη συνείδηση αλλά προσωρινά μένει σε μια μη συνειδητή αποθήκευση λόγω έλλειψης από πάνω προς τα κάτω ενίσχυσης μέσω προσοχής.

Η προσέγγιση του Dehaene καθίσταται ιδιαίτερα γοητευτική καθώς εκτός από τη σχετική απόσταση που λαμβάνει από τις φιλοσοφικές τοποθετήσεις, παρέχει ένα πλαίσιο που στηρίζεται σε εμπειρικά δεδομένα και μπορεί να διαψευστεί πειραματικά από αυτά, ενώ όπως ο ίδιος σημειώνει το παραπάνω θεωρητικό πλαίσιο θα μπορούσε να συμφιλιώσει αλλά και να συμπληρώσει άλλες σημαντικές θεωρίες. Ο Dehaene και οι συνεργάτες του πιστεύουν πως η φαινομενική συνείδηση είναι μια ψευδαίσθηση βασισμένη στην υποκειμενική μας διαίσθηση πως έχουμε πλούσια εμπειρία μιας οπτικής σκηνής ακόμα και αν δεν μπορούμε να την αναφέρουμε, και δεν ταυτίζεται με την προσυνειδητή επεξεργασία κατά την οποία η ένταση ενός ερεθίσματος μπορεί να είναι αρκετά δυνατή για να προκαλέσει μια από κάτω προς τα πάνω ενεργοποίηση, ωστόσο λείπει η προσοχή για να περάσει η πληροφορία σε συνειδητό επίπεδο.

5.3 Εστιακή ανατροφοδότηση

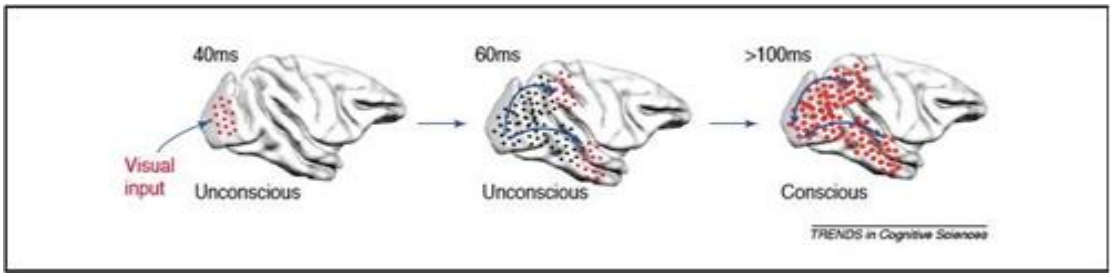
Μια επιπλέον ενδιαφέρουσα θεωρία είναι αυτή της εστιακής ανατροφοδότησης που προτάθηκε από τον καθηγητή γνωστικής νευροεπιστήμης Viktor Lamme. Ο Lamme, έχοντας εντοπίσει τα προβλήματα που προκύπτουν από την προσπάθεια ορισμού της συνείδησης με βάση μόνο ενδοσκοπικές ή συμπεριφορικές παρατηρήσεις, προτείνει έναν επιστημονικό ορισμό που να βασίζεται στη σύγκλιση νευρωνικών και συμπεριφορικών μετρήσεων. Συγκεκριμένα επιχειρεί να προσδιορίσει τη συνείδηση με ένα νευρωνικό τρόπο, ταυτίζοντάς την με την ανατροφοδοτική επεξεργασία. Για να υποστηρίξει την παραπάνω άποψη, επικαλείται μελέτες που υποδεικνύουν πως αυτό που είναι απαραίτητο για τη συνειδητή επεξεργασία είναι η εμπλοκή νευρώνων στις οπτικές περιοχές σε ανατροφοδοτική, συντονισμένη επεξεργασία μέσω αλληλεπίδρασης με άλλες υψηλότερου και χαμηλότερου επιπέδου περιοχές. Παραδείγματα συνείδησης με αδυναμία αναφοράς περιλαμβάνουν περιπτώσεις με ασθενείς με διατομή του μεσολόβιου, που αδυνατούν

να ονομάσουν ερεθίσματα που παρουσιάζονται στο αριστερό οπτικό τους πεδίο, αλλά μπορούν να τα ζωγραφίσουν, αφαιρώντας με τον τρόπο αυτό τη σύγχυση της συνείδησης με τη γλωσσική ικανότητα. Παρά όλα αυτά, το πρόβλημα της επαλήθευσης της συνειδητής εμπειρίας χωρίς την παρεμβολή καμίας άλλης νοητικής λειτουργίας παραμένει, και κατά τον Lamme, αυτό σχετίζεται άμεσα με τη διαμάχη περί της ύπαρξης μιας «μη-προσβάσιμης» ή φαινομενικής συνείδησης. Κατά τον ίδιο, η θεωρία του καθολικού χώρου εργασίας δε λύνει το πρόβλημα της φαινομενικής συνείδησης, αν και επιχειρεί να το ενσωματώσει με τη μορφή μιας «προσυνειδητής επεξεργασίας», ενώ η τοποθέτηση του Dehaene πως ένα τέτοιο ερώτημα δε φαίνεται να είναι επιστημονικά διαχειρίσιμο, μάλλον αποφεύγει μια απάντηση, παρά την παρέχει. Αντιθέτως, ο Lamme επιχειρεί να εξηγήσει τη φαινομενική συνείδηση διαχωρίζοντας την οπτική συνείδηση από την οπτική προσοχή, και ταυτίζοντάς τη φαινομενική συνείδηση με τη δεύτερη.

Στη θεωρία της εστιακής ανατροφοδότησης γίνεται διαχωρισμός ανάμεσα σε τρεις ιεραρχικούς τύπους νευρωνικής επεξεργασίας που σχετίζεται με τη συνείδηση ενός οπτικού ερεθίσματος **(Σχήμα 14)**. Το πρώτο στάδιο περιλαμβάνει μια εμπροσθοδρομική ενεργοποίηση κατά την οποία η πληροφορία μεταδίδεται από των ταινιωτό οπτικό φλοιό προς εξωταινιωτές περιοχές καθώς και προς βρεγματικές και κροταφικές περιοχές χωρίς να συνοδεύεται από συνειδητή εμπειρία του οπτικού ερεθίσματος.

Κατά το δεύτερο στάδιο η πληροφορία μεταδίδεται από τις περιοχές αυτές ξανά πίσω στις πρώιμες οπτικές περιοχές. Αυτό το στάδιο περιλαμβάνει την «εστιασμένη ανατροφοδοτική επεξεργασία», και είναι αυτές οι ανατροφοδοτικές αλληλεπιδράσεις ανάμεσα στις πρώιμες και τις ανώτερες περιοχές που οδηγεί τη συνειδητή εμπειρία, η οποία στο στάδιο αυτό έχει την έννοια της φαινομενικής εμπειρίας [15].

Το τελευταίο στάδιο χαρακτηρίζεται από ευρεία ανατροφοδοτική επεξεργασία, η οποία περιλαμβάνει καθολικές/ευρείες νευρωνικές αλληλεπιδράσεις -παρόμοιες με αυτές που προτείνει το μοντέλο του καθολικού χώρου εργασίας- και επεκτείνεται πως τα μετωποβρεγματικά δίκτυα. Το στάδιο αυτό περιλαμβάνει επιτελικές και λειτουργικές διεργασίες που είναι απαραίτητες για τη συνειδητή πρόσβαση και δυνατότητα αναφοράς του ερεθίσματος.



Σχήμα 14: Τα τρία στάδια της νευρωνικής επεξεργασίας της συνείδησης: Η συνειδητή οπτική εμπειρία απαιτεί ανατροφοδοτική επεξεργασία. Τα οπτικά εισιόντα φτάνουν στις πρώιμες οπτικές περιοχές (V1) 40ms μετά την εμφάνιση του ερεθίσματος. Τότε η οπτική πληροφορία άμεσα μεταδίδεται εμπροσθοδρομικά στις εξωταινιωτές περιοχές και τον βρεγματικό και κροταφικό φλοιό (60ms). Αυτή η εμπροσθοδρομική μετάδοση της πληροφορίας είναι μη συνειδητή. Στα περίπου 100ms, οι πρώϊμες οπτικές περιοχές και υψηλότερες ιεραρχικά περιοχές αποκτούν ανατροφοδοτική αλληλεπίδραση, που είναι απαραίτητη για τη συνειδητή οπτική εμπειρία [15].

Τέλος τρεις είναι οι βασικές παρατηρήσεις του Lamme, οι οποίες δίνονται με τη μορφή απαντήσεων σε τρία σημαντικά ερωτήματα:

1. Είναι η εμπλοκή του μετωποβρεγματικού δικτύου το νευρωνικό αντίστοιχο της συνειδητής επεξεργασίας, όπως υποστηρίζει η θεωρία του καθολικού χώρου εργασίας, ή μήπως είναι η ανατροφοδοτική επεξεργασία; Η απάντηση του Lamme υποστηρίζει το δεύτερο ενδεχόμενο, επικαλούμενος μελέτες masking που υποστηρίζουν πως η αυτο-ενεργοποίηση φλοιϊκών νευρώνων, ακόμη και μετωποβρεγματικών είναι ανεπαρκής για τη συνειδητή εμπειρία.

2. Είναι η ανατροφοδοτική επεξεργασία (ΑΕ) πραγματικά διαφορετική από την εμπροσθοδρομική; Ο Lamme, απαντά πως η ΑΕ ικανοποιεί τον κανόνα του Hebb όπου οι προ και μετασυναπτικοί νευρώνες είναι ταυτόχρονα ενεργοί, γεγονός που προκαλεί ενεργοποίηση της διαδικασίας της συναπτικής πλαστικότητας, που είναι η νευρωνική βάση της μάθησης και της μνήμης. Αντιθέτως μια εμπροσθοδρομική ενεργοποίηση δεν έχει αποτέλεσμα διαρκείας. Μια τέτοια θεώρηση δε μπορεί να προσδώσει ιδιαίτερο ρόλο στα μετωποβρεγματικά κυκλώματα.

3. Σύμφωνα με τη θεωρία του Tononi, η συνείδηση συνίσταται στη δυνατότητα ενός συστήματος να ολοκληρώσει, να ενοποιήσει τις πληροφορίες. Αυτή η ικανότητα εξαρτάται από τη δυνατότητα του συστήματος να διαμορφώσει ένα δυναμικό πυρήνα, χαρακτηριζόμενο από δυνατές αμοιβαίες αλληλεπιδράσεις ανάμεσα στα στοιχεία του, και ικανό να αλλάξει σε πολλές διαφορετικές καταστάσεις. Τέτοιοι δυναμικοί πυρήνες, η παραπάνω θεωρία υποστηρίζει πως διαμορφώνονται εντός του θαλαμοφλοιϊκού δικτύου. Στην ίδια θεωρία οι ανατροφοδοτικές αλληλεπιδράσεις ανάμεσα σε φλοιϊκούς νευρώνες είναι το κρίσιμο χαρακτηριστικό της συνείδησης, και όχι το ποιες περιοχές συμμετέχουν σε αυτές. Τέτοια συμπλέγματα αλληλεπιδράσεων μπορούν να λαμβάνουν χώρα ταυτόχρονα και κάθε ένα να συνιστά μια διαφορετική συνειδητή αναπαράσταση.

5.4 Μικροσυνειδήσεις

Τελευταία θεωρία στην οποία θα γίνει αναφορά είναι αυτή των μικροσυνειδήσεων, η οποία βρίσκει ως βασικό εκφραστή τον βρετανό νευροβιολόγο Semir Zeki. Το θεωρητικό πλαίσιο που υποστηρίζει ο Zeki και οι συνεργάτες του συνοψίζεται στην υπόθεση πως η συνείδηση αποτελείται από πολλές μικρο-συνειδήσεις που κατανέμονται στον χώρο και το χρόνο και πως η ενοποιημένη συνείδηση για την οποία συνήθως μιλάμε είναι δυνατή μόνο με τη χρήση της γλώσσας και της επικοινωνίας. Οι πολλαπλές αυτές συνειδήσεις συνιστούν μια ιεραρχία, με τη "συνθετική, υπερβατική" ενοποιημένη συνείδηση (αυτή του εαυτού μου ως ένα πρόσωπο που αντιλαμβάνεται) να βρίσκεται στη κορυφή της. Την υπόθεση αυτή στηρίζουν συμπεράσματα από μελέτες που αφορούν τη λειτουργία του οπτικού φλοιού και οι οποίες μπορούν να συνοψιστούν στις εξής κατηγορίες:

1. Ο οπτικός φλοιός αποτελείται από παράλληλα, λειτουργικά εξειδικευμένα συστήματα επεξεργασίας, με χρονική ιεραρχία. Ο οπτικός φλοιός αποτελείται από παράλληλα λειτουργικά εξειδικευμένα συστήματα επεξεργασίας, όπου το καθένα περιλαμβάνει διάφορα στάδια ή κόμβους, που ολοκληρώνουν το έργο τους σε διαφορετικούς χρόνους. Για παράδειγμα, το χρώμα γίνεται αντιληπτό πριν τον προσανατολισμό, ο οποίος γίνεται αντιληπτός πριν την κίνηση με

περίπου 30 και 40ms καθυστέρηση αντίστοιχα. Προκειμένου να μελετήσουν αυτό το ενδεχόμενο, οι Moutoussis & Zeki άλλαζαν το χρώμα και την κατεύθυνση της κίνησης ενός αφηρημένου προτύπου τετραγώνων σε μια οθόνη και ζητούσαν από τους συμμετέχοντες να αντιστοιχίσουν το χρώμα του προτύπου των τετραγώνων με την κατεύθυνση της κίνησης. Τα αποτελέσματα της μελέτης αυτής έδειξαν διαφορετικούς χρόνους αντίληψης της κίνησης και του χρώματος, υποστηρίζοντας πως ο εγκέφαλος «δένει» οπτικά χαρακτηριστικά που γίνονται αντιληπτά ταυτόχρονα, παρά χαρακτηριστικά που παρουσιάζονται ή συμβαίνουν μαζί σε πραγματικό χρόνο. Κάθε περιοχή εντός του οπτικού συστήματος έχει συγκεκριμένες συνδέσεις αφενός με τις περιοχές V1 και V2 και αφετέρου με επιπλέον περιοχές στον κροταφικό, βρεγματικό και μετωπιαίο λοβό. 'Ετσι οι συγγραφείς ορίζουν ένα σύστημα επεξεργασίας ως ένα σύστημα που περιλαμβάνει τα εξειδικευμένα κύτταρα της V1 και V2 καθώς και τις εξειδικευμένες περιοχές στις οποίες προβάλουν, καθώς και τις επιπλέον προβολές μιας εξειδικευμένης περιοχής.

2. Κάθε σύστημα επεξεργασίας αποτελείται από πολλά στάδια ή κόμβους. Για παράδειγμα, το κινητικό σύστημα αποτελείται από τη στιβάδα 4B της V1, τις παχιές λωρίδες της V2, την περιοχή V5 και άλλες περιοχές σχετιζόμενες με τη κίνηση που την περιλαμβάνουν. Κάθε μια από αυτές τις περιοχές συνιστά έναν κόμβο του συστήματος επεξεργασίας της κίνησης ενώ οι εμπροσθόδρομες συνδέσεις εντός του συστήματος είναι τύπου “like-with-like’’, δηλαδή “όμοιο με όμοιο”. Με αυτό οι συγγραφείς εννοούν πως, για παράδειγμα τα κύτταρα της στιβάδας 4B που είναι επιλεκτικά της κατεύθυνσης συνδέονται είτε άμεσα, είτε μέσω των παχιών λωρίδων της V2 με τη περιοχή V5, που είναι επίσης πλούσια σε επιλεκτικά της κατεύθυνσης κύτταρα. Τέλος, οι εμπροσθόδρομες συνδέσεις εντός ενός συστήματος επεξεργασίας οδηγούν σε κύτταρα με αυξανόμενο μέγεθος υποδεκτικού πεδίου και συνθετότητα, με ιεραρχικό τρόπο.

3. Τα συστήματα επεξεργασίας μπορούν να λειτουργούν αυτόνομα. Κλινικά δεδομένα, όπως αυτά που προέκυψαν από τη μελέτη μιας περίπτωσης τυφλής όρασης, αλλά και από μελέτες περιπτώσεων αχρωματοψίας ή ακινητοψίας υποστηρίζουν πως αυτά τα συστήματα επεξεργασίας μπορούν να λειτουργούν αυτόνομα, καθώς βλάβη που εστιάζεται σε ένα σύστημα (όπως στην περιοχή που σχετίζεται με την αντίληψη του χρώματος, όπως η V4) περιορίζει συγκεκριμένα την αντίληψη του χαρακτηριστικού στο οποίο το σύστημα αυτό εξειδικεύεται, δηλαδή στο χρώμα. Ακόμη βλάβη σε ένα συγκεκριμένο κόμβο ενός συστήματος επεξεργασίας που αφήνει προηγούμενους κόμβους άθικτους έχει ως αποτέλεσμα μια υποβαθμισμένη αντιληπτική ικανότητα για το συγκεκριμένο χαρακτηριστικό, που συνδέεται άμεσα με τις φυσιολογικές δυνατότητες των κυττάρων που δεν έχουν επηρεαστεί από τη βλάβη. Για παράδειγμα, ασθενείς με βλάβη στην περιοχή V5 δεν είναι ικανοί να αντιληφθούν αντικείμενα που βρίσκονται σε γρήγορη κίνηση, όμως μπορούν να τα αντιληφθούν όταν βρίσκονται σε πιο αργή, προφανώς αντικατοπτρίζοντας τις δυνατότητες των κόμβων που έχουν μείνει άθικτοι από τη βλάβη. Αντιθέτως, το σύστημα που μένει άθικτο όταν όλα τα άλλα έχουν υποστεί βλάβη, μπορεί να λειτουργήσει σχετικά φυσιολογικά.

4. Το “δέσιμο” (binding) των μικροσυνειδήσεων. Ανατομικά δεδομένα υποδεικνύουν πως δεν υπάρχει κάποιος τελικός ολοκληρωτικός σταθμός στον εγκέφαλο, ο οποίος να λαμβάνει εισιόντα από όλες τις οπτικές περιοχές. Αντιθέτως κάθε κόμβος έχει πολλαπλές εξόδους, και κανένας κόμβος δεν είναι μόνο προσληπτικός. Με βάση τα παραπάνω δεδομένα, οι Zeki & Bartels

προτείνουν πως κάθε κόμβος ενός συστήματος επεξεργασίας της αντίληψης δημιουργεί τη δική του «μικροσυνείδηση». Αν όντως κάποιο δέσιμο το οποίο προσδίδει την ολοκληρωμένη εικόνα που έχουμε για τον οπτικό κόσμο λαμβάνει χώρα, αυτό πρέπει να είναι το δέσιμο ανάμεσα σε μικροσυνειδήσεις παραγόμενες σε διαφορετικούς κόμβους. Καθώς οποιεσδήποτε δύο μικροσυνειδήσεις που παράγονται σε δύο κόμβους μπορούν να «δεθούν», η αντιληπτική ολοκλήρωση (δηλαδή η ενοποίηση διαφορετικών αισθητηριακών πληροφοριών) δεν είναι ιεραρχική αλλά παράλληλη και μετα-συνειδητή επίγνωση της, όπως μέσω συγχρονισμένης πυροδότησης ή οποιαδήποτε άλλης μορφής επικοινωνία ανάμεσα σε σύνολα κυττάρων. Κατά τους συγγραφείς, το «δέσιμο» πρέπει να έχει ως αποτέλεσμα μια αλλαγή στη δραστηριότητα των κόμβων που εμπλέκονται, ώστε παραλλαγμένες μικροσυνειδήσεις να παράγονται στον καθένα από αυτούς. Αντιθέτως, ο νευρωνικός μηχανισμός που προσδίδει ιδιότητες σε αυτά τα κύτταρα, των οποίων η δραστηριότητα έχει κάποιο νευρωνικό αντίστοιχο, είναι ιεραρχικός και προ-συνειδητός, και οι συγγραφείς αναφέρονται σε αυτόν ως παραγωγικό δέσιμο, για να το διαχωρίσουν από το δέσιμο που μπορεί να συμβαίνει ανάμεσα στις μικροσυνειδήσεις.

5. Τα ιεραρχικά επίπεδα της Συνείδησης. Ο Zeki διακρίνει τρία ιεραρχικά επίπεδα συνείδησης που περιλαμβάνουν τα επίπεδα της μικρο-συνείδησης, της μακρο-συνείδησης και της ενοποιημένης συνείδησης. Οι μικροσυνειδήσεις, που αντιστοιχούν σε κόμβους των διαφόρων συστημάτων επεξεργασίας, όπως αναφέρεται και προηγουμένως μπορεί να αφορούν χαρακτηριστικά όπως το χρώμα ή η κίνηση. Η μακροσυνείδηση αναφέρεται στο αποτέλεσμα του δεσίματος μεταξύ δύο ή περισσοτέρων μικροσυνειδήσεων και άρα σε μια πιο ολοκληρωμένη μορφή αντίληψης, ενώ η ενοποιημένη συνείδηση, το υψηλότερο ιεραρχικά επίπεδο συνείδησης, αναφέρεται στη συνείδηση του εαυτού μου ως ατόμου που αντιλαμβάνεται. Το κάθε επίπεδο εξαρτάται από την παρουσία του προηγούμενου ενώ εντός του κάθε επιπέδου μπορεί να υποτεθεί η ύπαρξη μιας χρονικής ιεραρχίας. Τα δύο πρώτα επίπεδα συνείδησης με τις δικές τους χρονικές ιεραρχίες οδηγούν στην τελική ενοποιημένη συνείδηση, αυτή του εαυτού μου ως άτομο που αντιλαμβάνεται [16].

Εν κατακλείδι η θεωρία των μικροσυνειδήσεων του Zeki, υποστηρίζει την ύπαρξη τριών σταδίων συνειδητής επεξεργασίας που ξεκινά από τις μικροσυνειδήσεις, η οποία συνίσταται στη επεξεργασία και ταυτόχρονα αντίληψη οπτικών χαρακτηριστικών όπως το χρώμα ή η κίνηση, προχωρά στη μακροσυνείδηση η οποία προκύπτει από τον συνδυασμό αλληλοεπιδρώντων κόμβων και μικροσυνειδήσεων και αφορά πιο ολοκληρωμένη αντίληψη (βλέπω έναν κόκκινο κύκλο που κινείται), καταλήγοντας στην ανώτερη ιεραρχικά ενοποιημένη συνείδηση, αυτή του εαυτού μου ως ατόμου που αντιλαμβάνεται.

6. Συνείδηση και Τεχνητή Νοημοσύνη

6.1 Εισαγωγή

Υπολογιστικός λογισμός είναι η θεωρία ότι ο ανθρώπινος εγκέφαλος είναι ουσιαστικά ένας υπολογιστής, αν και πρακτικά δεν είναι ένας προγραμματισμένος ηλεκτρονικός υπολογιστής. Η Τεχνητή Νοημοσύνη (ΤΝ) είναι το πεδίο της επιστήμης των υπολογιστών που εξερευνά υπολογιστικά μοντέλα για τη λύση προβλημάτων, των οποίων η πολυπλοκότητα είναι σε επίπεδο που μπορεί να τα επιλύσει ένας ανθρώπινος εγκέφαλος. Ένας ερευνητής ΤΝ δεν χρειάζεται για τη δουλεία του να συμφωνεί στις απόψεις με έναν ερευνητή υπολογιστικού λογισμού, ωστόσο οι περισσότεροι ερευνητές ΤΝ είναι σε κάποιο βαθμό και ερευνητές υπολογιστικού λογισμού. Όταν βέβαια μιλάμε για προβλήματα φαινομενικής συνείδησης μόνο ένα πολύ μικρό ποσοστό των ερευνητών ΤΝ θεωρούν ότι η λύση μπορεί να βρεθεί μέσω της τεχνητής νοημοσύνης.

Ίσως, μάλιστα, η αναφορά στον υπολογιστικό λογισμό ως μια θεωρία να μην είναι απόλυτα ορθή. Κάποιος θα αναφερόταν σε αυτόν ως «υπόθεση υπό ανάπτυξη» ή «δόγμα». Για άλλους όμως ο όρος αυτός είναι πρακτικά υπαρκτός και υπόκειται σε μοντέρνα πεδία έρευνας πάνω στη γνωστική ψυχολογία, στη γλωσσολογία και σε ορισμένη έκταση στη νευροεπιστήμη. Από την άλλη πλευρά, όλη αυτή η συζήτηση και αναζήτηση τυπικών ή υπολογιστικών μοντέλων για τη λειτουργία του εγκεφάλου μπορεί να κριθεί έως και μάταια αν αποδειχθεί ότι δεν υπάρχει κάποια υπολογιστική βάση στον τρόπο λειτουργίας του εγκεφάλου, παρά ένας ενεργειακός-μηχανικός χειρισμός. Ο υπολογιστικός λογισμός έχει εξελιχθεί σε μια γόνιμη υπόθεση υπό ανάπτυξη, αν και αυτοί που δεν τον αποδέχονται θεωρούν τη συνεισφορά του σχεδόν μηδαμινή.

Ορισμένοι ερευνητές υπολογιστικού λογισμού θεωρούν ότι ο εγκέφαλος δεν είναι τίποτα περισσότερο από έναν υπολογιστή. Πολλοί άλλοι είναι περισσότερο προσεκτικοί και διακρίνουν ενότητες που είναι αρκετά πιθανό να έχουν καθαρά υπολογιστικό χαρακτήρα (όπως το σύστημα όρασης) και αυτές που είναι λιγότερο πιθανό να ισχύει κάτι τέτοιο (όπως η δημιουργικότητα ή ο ρομαντισμός). Άλλωστε δεν υπάρχει ανάγκη να εξηγηθούν όλα με μια υπολογιστική βάση, αλλά να υπάρξει μεγαλύτερη επίγνωση στα κομμάτια τα οποία μπορεί αυτή να καλύψει [17].

Ίσως η ενότητα του εγκεφάλου που είναι το πιο πιθανό να εξαιρεθεί από αυτές που έχουν υπολογιστικό χαρακτήρα είναι αυτή της συνείδησης, αναφερόμενοι σε αυτή από την άποψη του «δύσκολου προβλήματος» της συνείδησης, δηλαδή την ικανότητα ενός φυσιολογικού συστήματος να αποκτά ζωντανές εμπειρίες και αυτές να έχουν εσωτερικές ιδιότητες για το σύστημα όπως είναι το κόκκινο χρώμα της ντομάτας ή το πόσο καυτερή είναι μια πιπεριά. Αυτό βέβαια δεν έχει εμποδίσει την προσπάθεια των ερευνών, οι οποίες για τα μέχρι τώρα δεδομένα έχουν κάνει πολύ σημαντική πρόοδο, χωρίς να έχουν καταφέρει να παρουσιάσουν ουσιαστικά πειστήρια πως η συνείδηση έχει υπολογιστική βάση. Κάτι τέτοιο θα άνοιγε ταυτόχρονα τη δυνατότητα να δημιουργηθεί μια μηχανή με συνείδηση φαινομενικά. Ο πιο σημαντικός λόγος που οδηγεί σε αυτό το κενό είναι η μέχρι τώρα αδυναμία μας να εξηγήσουμε το αν υπάρχει η δυνατότητα εφαρμογής υψηλού επιπέδου γνωστικοί αλγόριθμοι προς όφελος της ΤΝ με βάση νευροϋπολογιστικών διαδικασιών. Η γεφύρωση αυτού του κενού θα μπορούσε να συνεισφέρει σε επιπλέον πρόοδο

σχετικά με τη συνείδηση των μηχανών, με το σχηματισμό γενικής τεχνητής νοημοσύνης και με την κατανόηση της θεμελιώδους φύσης της συνείδησης.

6.2 Συνείδηση σε μηχανές

Υπάρχει λοιπόν η πιθανότητα κέρδους από τη μελέτη της συνείδησης μέσα στα πλαίσια της ΤΝ; Με μια γρήγορη ματιά στη σχετική βιβλιογραφία θα παρατηρούσε κανείς ότι οι ερευνητές της ΤΝ, με ελάχιστες εξαιρέσεις, αγνοούν τις αυξανόμενες προσπάθειες των τελευταίων δεκαετιών για εξερεύνηση της φύσης της συνείδησης με υπολογιστικά μέσα. Ως αποτέλεσμα, ένα ξεχωριστό πεδίο, γνωστό ως τεχνητή συνείδηση ή συνείδηση μηχανών, έχει αναδυθεί και αναπτύσσεται σχετικά αυτόνομα από το πεδίο της ΤΝ.

Η δουλειά στην τεχνητή συνείδηση επικεντρώνεται στη δημιουργία υπολογιστικών μοντέλων ποικίλων πτυχών του συνειδητού εγκεφάλου, είτε σε μορφή λογισμικού σε κάποιον υπολογιστή είτε ενσωματωμένο σε φυσικά ρομποτικά συστήματα. Αυτή η δουλειά κινητοποιείται κυρίως από την επιθυμία να αναπτυχθεί η κατανόηση μας για την ανθρώπινη συνείδηση και τη σχέση της με τη γνωστική διαδικασία, αλλά και περιστασιακά από τις απόψεις ότι η συνείδηση των μηχανών (ή η παρουσία αυτής) θα μπορούσε να αυξήσει τη λειτουργικότητα των μελλοντικών συστημάτων ΤΝ, ή εντέλει να οδηγήσει και σε φαινομενικά συνειδητές μηχανές.

Η πρόοδος που έχει σημειωθεί στο συγκεκριμένο τομέα είναι ουσιώδης και η φύση της μπορεί να γίνει κατανοητή με το διαχωρισμό μεταξύ δύο στόχων στην έρευνα για τη συνείδηση των μηχανών: προσομοίωση εναντίον εκδήλωσης συνείδησης από μια μηχανή. Η δουλειά πάνω στην προσομοιωμένη συνείδηση έχει φτάσει σε σημείο όπου η χρήση προσομοιώσεων σε υπολογιστή έχει γίνει μια συνεχώς και περισσότερο αποδεκτή προσέγγιση της επιστημονικής μελέτης της συνείδησης. Ωστόσο, οι προσπάθειες να παραχθεί μια φαινομενικά συνειδητή μηχανή μέχρι τώρα έχουν αποδειχθεί λιγότερο επιτυχείς.

Για την προσομοιωμένη συνείδηση στόχος είναι να αποτυπωθούν ορισμένες από τις νευρικές, συμπεριφορικές και γνωστικές συσχετίσεις της συνείδησης σε ένα υπολογιστικό μοντέλο, με παρόμοια λειτουργία όπως τώρα χρησιμοποιείται ο υπολογιστής για άλλες φυσικές διαδικασίες (για παράδειγμα μοντέλα για τον καιρό). Δεν υπάρχει κάτι ιδιαίτερα μυστηριώδες σχετικά με αυτή τη δουλειά. Και δεν υπάρχει κανένας ισχυρισμός ότι η φαινομενική συνείδηση είναι όντως παρούσα σε αυτή την κατάσταση. Αντίθετα, με την εφαρμοσμένη συνείδηση το θέμα είναι το εύρος στο οποίο ένα σύστημα ΤΝ έχει όντως φαινομενική συνείδηση. Αποκτά ή έχει υποκειμενικές εμπειρίες; Αυτή είναι μια αρκετά πιο δύσκολη και αμφιλεγόμενη ερώτηση.

Από την οπτική της προσομοιωμένης συνείδησης είναι εύκολο κάποιος να δει ότι έχει επιτευχθεί ουσιώδης πρόοδος. Ορισμένα παραδείγματα είναι η δημιουργία νευυροϋπολογιστικών μοντέλων που αυξάνουν τη δραστηριοποίηση ενός καθολικού χώρου εργασίας όταν επωμίζονται δύσκολα ή πολύπλοκα καθήκοντα που συνδέονται με συνειδητή δραστηριότητα από ανθρώπους, το απροσδόκητο πόρισμα της Θεωρίας Ενσωματωμένων Πληροφοριών ότι οι μονάδες πύλης είναι τα πιο συνειδητά συστατικά ενός νευροελεγκτή, η δημιουργία ενδιαφέρων συμπεριφορών στη ρομποτική, η επίδειξη ότι ορισμένα ειδικά ρομπότ μπορούν να αναγνωρίσουν τον εαυτό τους στον

καθρέφτη σε πειράματα που έγιναν με βάση την ιδέα του αν τα ζώα έχουν αυτοαναγνώριση. Προφανώς, αυτά και άλλα υπολογιστικά μοντέλα των συσχετίσεων της συνείδησης έχουν παρέχει χρήσιμες πληροφορίες για εξειδικευμένες μελέτες σχετικά με τη συνείδηση. Ως αποτέλεσμα, τέτοια μοντέλα θεωρούνται όλο και περισσότερο ως μια αποδεκτή προσέγγιση στη διεύρυνση της επιστημονικής έρευνας της συνείδησης και ίσως βοηθήσουν και στον περιορισμό του πολύ μεγάλου αριθμού διαφορετικών θεωριών σχετικά με την ύπαρξη της [18].

Η κατάσταση είναι αρκετά διαφορετική από την πλευρά της εφαρμοσμένης συνείδησης. Αρκετοί ερευνητές έχουν ισχυριστεί ότι γνωρίζουν πώς να δημιουργήσουν φαινομενικά συνειδητά τεχνουργήματα. Και πάλι μερικά παραδείγματα είναι ότι οποιοδήποτε σύστημα διατηρεί μια αντιστοιχία μεταξύ υψηλού επιπέδου, συμβολικά αναπαριστώμενων εννοιών και χαμηλού επιπέδου οντοτήτων-δεδομένων και έχει και ένα συλλογιστικό σύστημα που κάνει χρήση αυτών των συμβόλων τότε έχει και υποκειμενικές εμπειρίες που μπορούν να αντιστοιχηθούν με σημάδια αυτογνωσίας, ότι σε ένα πλαίσιο καθολικού χώρου εργασίας είναι πιθανό να αναπτυχθούν μοντέλα που δείχνουν με ολοφάνερο τρόπο εύλογες νευροϋπολογιστικές βάσεις για εμφάνιση και πρόσβαση στη συνείδηση, ότι ένα σύστημα έχει υποκειμενική εμπειρία στο βαθμό που έχει την ικανότητα να ενσωματώνει πληροφορίες. Παρά αυτούς τους ισχυρισμούς, καμία υπάρχουσα υπολογιστική προσέγγιση στην τεχνητή συνείδηση δεν έχει παρουσιάσει ακόμα ένα ολοκληρωμένο σχέδιο ή επίδειξη εφαρμοσμένης συνείδησης σε μηχανή, ή έστω ξεκάθαρες αποδείξεις ότι η εφαρμοσμένη μηχανική συνείδηση είναι εφικτή. Φυσικά, ούτε η αναίρεση της παραπάνω πιθανότητας είναι καθολικά αποδεκτή. Έτσι η τρέχουσα κατάσταση θέτει το ζήτημα του τι μπορεί να γίνει για να εξακριβωθεί το αν υπάρχουν πιθανότητες για εφαρμοσμένη συνείδηση στις μηχανές. Η διαλεύκανση αυτού του θέματος εξαρτάται από τη σαφή εξακρίβωση των βασικών ορίων της επιπλέον προόδου και έρευνας που είναι ευάγωγες και η υπερπήδηση τους.

6.3 Το υπολογιστικό επεξηγηματικό κενό

Ποιο είναι το κυρίως πρακτικό πρόβλημα, όμως, που προς στιγμήν για τη δημιουργία της εκδήλωσης συνείδησης από κάποια μηχανή; Ξεκάθαρα, υπάρχει έλλειψη γνωστών αντίστοιχων υποψηφιοτήτων. Αυτό περιλαμβάνει την απουσία ενός γενικώς αποδεκτού ορισμού της συνείδησης, την περιορισμένη κατανόηση μας για τις νευροβιολογικές συσχετίσεις της, όπως και άλλα «εγκεφαλικά προβλήματα» σχετικά με τις μηχανές (πώς θα γνωρίζουμε σίγουρα ότι μια μηχανή διαθέτει συνείδηση;). Αν και όλες αυτές οι δυσκολίες είναι ουσιώδεις, από την πλευρά του μηχανικού υπολογιστών καμία από αυτές δε φαίνεται να επαρκούν για την έλλειψη της προόδου πάνω στο ζήτημα. Άλλωστε, αξιοσημείωτη πρόοδος έχει επιτευχθεί στην τεχνητή νοημοσύνη και τεχνητή ζωή χωρίς να υπάρχουν γενικώς αποδεκτοί ορισμοί ούτε της νοημοσύνης ούτε της ζωής. Η τρέχουσα επαρκής κατανόηση της νευροβιολογικής βάσης της συνείδησης δεν εμποδίζουν κάποιον να διεξάγει μια υπολογιστική μοντελοποίηση των υπαρχουσών θεωριών, ενώ τα υπόλοιπα εγκεφαλικά προβλήματα δεν εμποδίζουν τις μελέτες σε άλλους τομείς όπως η φαινομενική συνείδηση και η αυτογνωσία σε ανθρώπους και ζώα.

Προσφάτως προτάθηκε ένα επιπλέον λιγότερο αναγνωρίσιμο εμπόδιο, το υπολογιστικό επεξηγηματικό κενό, που είναι επίσης μεγάλης σημασίας. Ο ορισμός του είναι σαφής: είναι η

τρέχουσα έλλειψη κατανόησης του πώς μπορεί η επεξεργασία υψηλού επιπέδου γνωστικών πληροφοριών να αντιστοιχηθεί με χαμηλού επιπέδου νευρικούς υπολογισμούς. Εδώ η επεξεργασία υψηλού επιπέδου γνωστικών πληροφοριών περιλαμβάνει την επίλυση γνωστικών προβλημάτων στόχου, τη λήψη και εκτέλεση αποφάσεων, το σχεδιασμό, τη γλώσσα και γενικά γνωστικές διαδικασίες που είναι ευρέως αποδεκτό πως έστω εν μέρει είναι συνειδητά προσπελάσιμες. Ο όρος χαμηλού επιπέδου νευρικοί υπολογισμοί δηλώνει τα είδη των υπολογισμών που μπορούν να επιτευχθούν από δίκτυα ή τεχνητούς νευρώνες όπως αυτοί που μελετώνται πλέον από τη σύγχρονη επιστήμη των υπολογιστών, τη μηχανική, την ψυχολογία και τη νευροεπιστήμη.

Ενώ το υπολογιστικό επεξηγηματικό κενό είναι εντέλει σχετικό με τα εγκεφαλικά προβλήματα, δεν αποτελεί καθ' εαυτόν ένα εγκεφαλικό πρόβλημα. Αντίθετα, είναι ένα κενό στη γενικότερη κατανόηση μας του πώς οι υπολογισμοί (αλγόριθμοι και δυναμικές καταστάσεις) που υποστηρίζουν τον συλλογισμό και την επίλυση προβλημάτων στόχου σε υψηλό επίπεδο επεξεργασίας γνωστικών πληροφοριών μπορούν να μπορούν να αντιστοιχηθούν σε είδη υπολογισμών/ αλγορίθμων/ καταστάσεων που μπορούν να υποστηριχθούν από χαμηλού επιπέδου νευρικά δίκτυα. Με άλλα λόγια, αποτελεί ένα καθαρά υπολογιστικό θέμα, χωρίς να υπάγεται απόλυτα στην τεχνητή νοημοσύνη, τους υπολογιστές ή ακόμα την επιστήμη των υπολογιστών. Είναι μάλιστα ανεξάρτητο από τα σχετικά απαραίτητα εξαρτήματα, όπως τα ηλεκτρονικά κυκλώματα του υπολογιστή ή τα νευρικά κυκλώματα του εγκεφάλου. Το υπολογιστικό επεξηγηματικό κενό μπορεί να συσχετιστεί με την πρόσφατη φιλοσοφία σχετικά με το μυαλό, ότι οι υποκειμενικές πρώτου-προσώπου εμπειρίες ενός ατόμου δεν περιορίζονται στα παραδοσιακά qualia περιλαμβάνοντας την αντίληψη και την αντίδραση, αλλά επίσης περικλείουν τη συμβουλευτική σκέψη και υψηλού επιπέδου γνώση.

Αν κάποιος αποδεχτεί ότι το υπολογιστικό επεξηγηματικό κενό είναι σημαντικό πρόβλημα, τότε επακολουθεί ότι το θέμα της συνείδησης των μηχανών είναι μεγαλύτερης θεμελιώδους σημασίας για την ΤΝ από ότι γενικά αναγνωρίζεται. Για παράδειγμα, στην ΤΝ το υπολογιστικό επεξηγηματικό κενό σχετίζεται με μακροχρόνια διαφωνία που αφορά τη σχετική αξία των από πάνω προς τα κάτω και από κάτω προς τα πάνω προσεγγίσεων για τη δημιουργία μηχανικής νοημοσύνης. Αυτή η συνεχιζόμενη συζήτηση έχει χάσει κατά πολύ το σημείο ότι αυτές οι δύο προσεγγίσεις δεν είναι τόσο πολύ αντικρουόμενες εναλλακτικές παρά συμπληρώματα σε σχέση με αυτό που επιδιώκουν να επιτύχουν στη νοημοσύνη [18].

Οι από πάνω προς τα κάτω μέθοδοι της ΤΝ έχουν διακριθεί στην υψηλού επιπέδου γνωστική μοντελοποίηση, όπως η επίλυση προβλημάτων στόχου, η λήψη αποφάσεων, η κατανόηση φυσικών γλωσσών και ο προγραμματισμός σχεδίων, αλλά έχουν υπάρξει λιγότερο επιτυχημένα στην αναγνώριση προτύπων και στον έλεγχο ακολουθητικών συμπεριφορών που σχετίζονται με τη συνείδηση και σε αναφερόμενα χαρακτηριστικά της γνωστικής διαδικασίας. Οι από πάνω προς τα κάτω μέθοδοι έχουν επίσης βρεθεί να είναι εύθραυστες, αποτυγχάνοντας, για παράδειγμα, δραματικά σε πλαίσια θορύβου, απροσδόκητων γεγονότων ή μικρών αλλαγών στο περιεχόμενο της μνήμης. Αντίθετα, οι από κάτω προς τα πάνω νευροϋπολογιστικές μέθοδοι έχουν κατά βάση τα αντίθετα πλεονεκτήματα και μειονεκτήματα. Είναι ιδιαίτερα αποτελεσματικές και εύρωστες στη μάθηση χαμηλού επιπέδου προτύπων ταξινόμησης («είσοδος») και χαμηλού επιπέδου

συμβάντων ελέγχου («έξοδος»), αλλά σε καμία περίπτωση δεν είναι τόσο αποτελεσματικές για υψηλού επιπέδου γνωστικά προβλήματα. Για παράδειγμα, τα νευροϋπολογιστικά έχει αποδειχθεί πως είναι πολύ αποτελεσματικά στη μάθηση διαδικαστικής γνώσης, που κατά πολύ πραγματοποιείται με έναν αυτόματο μη-συνειδητό τρόπο. Τέτοια παραδείγματα είναι οι νευροελεγκτές για την κίνηση ενός ρομποτικού χεριού και οι ελεγκτές που καθορίζουν την κατεύθυνση σε ένα αυτόνομο αυτοκινητάκι. Σε σύγκριση με τις από πάνω προς τα κάτω μεθόδους, οι νευροϋπολογιστικές μέθοδοι είναι λιγότερο εύθραυστες στα πλαίσια εξωτερικού θορύβου ή ουσιωδών τυχαίων αλλαγών στην αποθηκευμένη πληροφορία (όπως τυχαία απώλεια προσομοιωμένων νευρώνων).

6.4 Μια πρώτη διάκριση για τη συνείδηση

Η λέξη συνείδηση, όπως και πολλοί επιστημονικοί όροι, χρησιμοποιείται ευρέως με διαφορετικές έννοιες. Σε ιατρικά πλαίσια, συνήθως χρησιμοποιείται με μια αμετάβατη σημασία σχετικά με την αξιολόγηση της επαγρύπνησης ενός ατόμου. Η διευκρίνηση των μηχανισμών επαγρύπνησης του εγκεφάλου είναι ένας ουσιώδης επιστημονικός στόχος, με σοβαρά επακόλουθα στην κατανόηση του ύπνου, της αναισθησίας, των καταστάσεων «φυτό» και «κόμμα». Φυσικά όσον αφορά τις μηχανές αυτή η ερμηνεία έχει ελάχιστη σημασία. Ωστόσο, είναι σημαντικό να διακριθούν δύο άλλες διαστάσεις της συνείδησης σχετικά με τις μηχανές. Θα διαχωρίζονται μέσω της χρήσης των όρων καθολική διαθεσιμότητα (C1) και αυτοέλεγχος (C2).

Η καθολική διαθεσιμότητα αντιστοιχεί στη μεταβατική σημασία της συνείδησης, για παράδειγμα όταν ένας οδηγός έχει συνείδηση του φαναριού. Αναφέρεται στη σχέση ανάμεσα σε ένα γνωστικό σύστημα και ένα συγκεκριμένο αντικείμενο σκέψης, όπως μια νοητική αναπαράσταση για το λαμπάκι του ντεπόζιτου. Αυτό το αντικείμενο φαίνεται να επιλέγεται για επιπλέον επεξεργασία, συμπεριλαμβανομένης και μιας λεκτικής ή μη-λεκτικής αναφοράς. Η πληροφορία από αυτή την πλευρά της συνείδησης γίνεται καθολικά διαθέσιμη στον οργανισμό, καθώς μπορούμε να την επαναφέρουμε στη σκέψη μας, να δράσουμε με βάση αυτή και να μιλήσουμε για αυτή. Αυτή η έννοια είναι συνώνυμη με το να έχει κάποιος μια πληροφορία κατά νου, δηλαδή μέσα στο αχανές εύρος σκέψεων που μπορούν να γίνουν συνειδητές κάθε στιγμή, μόνο αυτή που είναι καθολικά διαθέσιμη αποτελεί το περιεχόμενο της C1.

Ο αυτοέλεγχος εκφράζει μια αυτοπαθή σημασία της συνείδησης. Αναφέρεται σε μια αυτο-αναφορική σχέση, στην οποία το γνωστικό σύστημα είναι σε θέση να ελέγχει τις δικές του επεξεργασίες και να αποκτά πληροφορίες σχετικά με τον εαυτό του. Οι άνθρωποι γνωρίζουν πολλά για τους εαυτούς τους, συμπεριλαμβανομένης μιας μεγάλης ποικιλίας πληροφοριών όπως η μορφή και η θέση του σώματος τους, αν γνωρίζουν ή αντιλαμβάνονται κάτι ή αν έχουν κάνει κάτι λάθος. Αυτή η έννοια της συνείδησης αντιστοιχεί σε αυτό που κοινώς καλείται ενδοσκόπηση, ή όπως αλλιώς το καλούνε οι ψυχολόγοι «μετα-γνωστικότητα», δηλαδή η ικανότητα κάποιου να συλλαμβάνει και να χρησιμοποιεί τις εσωτερικές αναπαραστάσεις των δικών του γνώσεων και δυνατοτήτων.

Έχει προταθεί ότι τα C1 και C2 απαρτίζουν ορθογώνιες διαστάσεις των υπολογισμών συνείδησης. Αυτό δε σημαίνει ότι C1 και C2 δεν περιλαμβάνουν επικαλυπτόμενα φυσικά υποστρώματα, όπως

για παράδειγμα έχει αποδειχθεί ότι και οι δύο εξαρτώνται από τον προμετωπιαίο φλοιό. Ωστόσο, εμπειρικά και εννοιολογικά, έχει εκφραστεί η άποψη πως αυτές οι δύο διαχωρίζονται, καθώς μπορεί να υπάρξει η C1 χωρίς τη C2, σε περίπτωση που κάποια αναφερόμενη διαδικασία μπορεί να μη συνοδεύεται και από μετα-γνωστικότητα, όπως και η C2 χωρίς τη C1, όταν μια αυτοελεγκτική διαδικασία εξελιχθεί χωρίς να είναι συνειδητά αναφερόμενη. Φυσικά υπάρχουν και διαδικασίες οι οποίες δεν περιλαμβάνουν ούτε τη C1 ούτε τη C2 και, επομένως, καλούνται «ασυνείδητες». Ήταν ένα από τους απώτερους στόχους-οράματα του Turing, ότι η επεξεργασία ακόμα και εκλεπτυσμένων πληροφοριών μπορεί να γίνει αντιληπτή από κάποιο αυτόματο χωρίς μυαλό. Η γνωστική νευροεπιστήμη επιβεβαιώνει πως πολύπλοκοι υπολογισμοί όπως η αναγνώριση προσώπου ή ομιλίας, η εκτίμηση για ένα παιχνίδι σκάκι και η ανάλυση προτάσεων και νοημάτων, συμβαίνουν ασυνείδητα στον ανθρώπινο εγκέφαλο, υπό συνθήκες που δεν παρουσιάζεται ούτε καθολική διαθεσιμότητα, ούτε αυτοέλεγχος. Ο εγκέφαλος φαίνεται να δρα, εν μέρει, ως μια αντιπαράθεση ειδικευμένων επεξεργαστών που λειτουργούν ασυνείδητα και αντιστοιχούν στη λειτουργία ανεστραμμένων δικτύων βαθιάς μάθησης [19].

6.5 Η μοναδικότητα

Πέραν βέβαια όλων των πρακτικών προβλημάτων που έχουν αναφερθεί η έλλειψη της γνώσης μας πάνω στο θέμα έχει επηρεάσει και σε ιδεολογικό επίπεδο την κοινή γνώμη. Η διαμάχη του αν τα ρομπότ θα ξεπεράσουν σε νοημοσύνη τον άνθρωπο και οι συνέπειες αυτού του γεγονότος υφίστανται ακόμα και, ενώ ίσως προκύπτει από την έλλειψη πρακτικής εμπειρίας πάνω στο θέμα που φέρνει μέχρι και φόβο, δεν υπάρχουν και ακράδαντα αντικρουόμενα επιχειρήματα. Έτσι εκφράζεται και η άποψη ότι ένα «έξυπνο» ρομπότ με συνείδηση μπορεί να φτάσει σε θέση να σχεδιάσει ένα ακόμα πιο έξυπνο ρομπότ χωρίς την ανθρώπινη παρεμβολή. Αυτό θα οδηγήσει σε μια τεχνολογική έκρηξη με τις πιο έξυπνες μηχανές να επικρατούν και να επιβάλλονται των ανθρώπων. Αυτή είναι η ουσία του επιχειρήματος της μοναδικότητας, το οποίο αν και ανάμεσα στα όρια επιστημονικής φαντασίας και θεωρίας, έχει αναδυθεί πιο έντονα τον τελευταίο καιρό λόγω της εκθετικής επιτάχυνσης στην ανάπτυξη της τεχνολογίας, που οδήγησε τις μηχανές να αντικαθιστούν τη δουλειά των ανθρώπων για αρκετές εργασίες.

Η επικρατέστερη, ωστόσο, άποψη είναι ότι αυτές οι συζητήσεις περί μοναδικότητας μπερδεύουν τη δύναμη ενός ολοκληρωμένου κυκλώματος επεξεργασίας με την κατανόηση του πώς η βιολογική μηχανική οδηγεί στη νοημοσύνη των ανθρώπων. Η δημιουργία απλώς ταχύτερων μηχανών δεν βοηθάει στην ψηφιακή αρχιτεκτονική για την ανάπτυξη κάποιας θεωρίας που προσεγγίζει τη βιολογική αρχιτεκτονική. Για παράδειγμα, η ανάπτυξη της ταχύτητας για τις μηχανές που παίζουν σκάκι δεν έχει οδηγήσει και σε εξέλιξη των αλγορίθμων για την επιλογή των κινήσεων. Φυσικά η έρευνα σε μεγαλύτερο βάθος αυτών των μηχανών απαιτεί και την ταχύτητα εξαντλητικών εκτιμήσεων για τις πιθανές κινήσεις να είναι υψηλότερη, ωστόσο το νόημα παραμένει το ίδιο. Αντίστοιχα μόνο η σοβαρή ενασχόληση με τον σχεδιασμό της νοημοσύνης των μηχανών είναι αυτή που θεμελιώνει την ανάπτυξη αυτής. Για να φτάσουμε σε ένα σημείο μοναδικότητας ένας άνθρωπος πρέπει να είναι σε θέση να μοντελοποιήσει το πώς αυτός χρησιμοποιεί το μυαλό του για να σχεδιάσει ένα ρομπότ πιο έξυπνο από τον ίδιο. Αυτό δεν εξαρτάται μόνο από τον τεράστιο όγκο γνώσεων που απαιτούνται σε πλαίσια μηχανικής,

ψυχολογίας και νευροεπιστήμης, αλλά επίσης έχει και λογικά κενά. Ο σχεδιασμός ενός καλύτερου ποδηλάτου είναι τελείως διαφορετικός από τον σχεδιασμό του σχεδιαστή ενός καλύτερου ποδηλάτου. Έτσι είναι άλλο να είσαι σχεδιαστής της πιο εξελιγμένης ΤΝ και άλλο σχεδιαστής ενός σχεδιαστή. Επομένως δεν υπάρχει καμία εγγύηση ότι η διαρκής εξέλιξη των αλγορίθμων ΤΝ θα οδηγήσει και στο επίπεδο της θεωρίας της μοναδικότητας [3].

Υπάρχει, φυσικά, και ο αντίλογος ότι ο τρόπος για να σχεδιάσει κάποιος ένα σύστημα που να σχεδιάζει καλύτερα από τον ίδιο είναι να χρησιμοποιήσει «απλώς» μια επαναστατική μέθοδο προσομοίωσης. Τέτοια είναι να χτίσει ένα σύστημα που θα παράγει αρκετά σχέδια, θα τα τεστάρει, θα κρατάει τα πιο αποδοτικά, με βάση αυτά θα δημιουργεί επιπλέον σχέδια και θα επαναλαμβάνει συνεχώς αυτή τη διαδικασία. Αλλά επειδή ο χώρος σχεδίων είναι τόσο αχανής, αυτές οι επαναστατικές μέθοδοι πρέπει να εφαρμόζονται από την αρχή πάρα πολλές φορές για να σιγουρευτεί ότι δεν καταλήγουν μόνο τοπικά σε επιθυμητό αποτέλεσμα. Σε αυτό πρέπει να προστεθεί ότι ακόμα δεν είναι γνωστές γενικές μέθοδοι για τη σύγκριση της ικανότητας του σχεδίου δύο συστημάτων. Έτσι, ακόμα κι αν παραβλεφθούν οι απαιτήσεις για την εύρεση ενός μονοπατιού βελτιστοποίησης σε ένα τεράστιο εύρος σχεδίων, είναι πολύ ισχνό το στήριγμα της άποψης ότι επαναστατικά συστήματα μπορούν να οδηγήσουν σε μορφές ΤΝ με ανώτερες σχεδιαστικές ικανότητες από τον άνθρωπο.

6.6 Γενική τεχνητή νοημοσύνη

Η Γενική Τεχνητή Νοημοσύνη (ΓΤΝ) δεν είναι μια αυστηρά ορισμένη δραστηριότητα, αλλά θέτει μια πρόκληση σε όσους ασχολούνται με την ΤΝ με την υποστήριξη ότι υπάρχει νοημοσύνη στις καθημερινές δραστηριότητες των ανθρώπων, των οποίων αξίζει να γίνει μοντελοποίηση ώστε να κατανοήσουμε καλύτερα το τι είναι νοημοσύνη. Οι ερευνητές ΓΤΝ ισχυρίζονται πως η νοημοσύνη που χρειάζεται για να μπει κάποιος σε μια άγνωστη κουζίνα και να φτιάξει μια κούπα με καφέ είναι, μέχρι τώρα, μια ανεκπλήρωτη πρόκληση. Αυτό, θεωρούν, πως απαιτεί μια πιο ολιστική και συστηματική προσέγγιση από ότι οι προσπάθειες στη συμβατή ΤΝ που στοχεύουν στο σχεδιασμό καλύτερης όρασης, αντίληψης της γλώσσας, ικανότητας προγραμματισμού και αλγορίθμων μετακίνησης.

Ωστόσο, παρά τις αρκετές προσπάθειες να οριστεί σαφώς η ΓΤΝ, η δουλειά που έχει γίνει πρακτικά δεν είναι και απόλυτα ξεχωριστή από αυτή που θα γινόταν σε ένα εργαστήριο ΤΝ για αντίστοιχες περιπτώσεις. Παρά όλα αυτά, ο Ben Goertzel, ένας από τους κύριους συνηγόρους της ΓΤΝ έχει δημοσιεύσει μια θεωρητική, φουτουριστική ματιά με τίτλο «10 χρόνια από τη Μοναδικότητα, αν προσπαθήσουμε πραγματικά». Η θέση του είναι πως η σταθερή πρόοδος στη ΓΤΝ θα οδηγήσει σε μια «μοναδικότητα», η οποία περιγράφεται από τον ίδιο ως: «σε κάποιο συγκεκριμένο σημείο, δε θα είμαστε τα πιο έξυπνα γενικώς πλάσματα στον κόσμο». Φέρνει ισχυρό αντίλογο στην ιδέα ότι αυτή η κατάσταση προμηνύει το τέλος του ανθρώπινου είδους και το βλέπει περισσότερο ως μια επανισορρόπηση της δουλειάς και των ευθυνών μεταξύ ανθρώπων και μηχανών. Σε κάθε περίπτωση, το μεγαλύτερο μέρος της μέχρι τώρα εργασίας πάνω στη νοημοσύνη των ρομπότ φαίνεται να είναι στα αρχικά στάδια μιας ασύμπτωτης σε σχέση με αυτή των ανθρώπων, οπότε η πρόβλεψη πως η δεύτερη θα ξεπεράσει την πρώτη μάλλον είναι πρώιμη.

Ίσως το πιο διαλλακτικό υλικό σχετικά με τη μοναδικότητα και τη ΓΤΝ έχει γραφτεί από τον Murray Shanahan. Ο ίδιος υποστηρίζει (όπως και ο Goertzel) πως η τροχιά για να ξεπεραστεί η νοημοσύνη των περισσότερων ανθρώπων από κάποιο ρομπότ ίσως να υπάρχει και οι αναγνώστες του βιβλίου του πρέπει να είναι προετοιμασμένοι για αυτό. Χωρίς έλλειψη ενδιαφέροντος παραθέτει τις τεχνολογίες που μπορεί να οδηγήσουν σε τέτοιου είδους σενάρια που είναι τόσο ωφέλιμα όσο και επιζήμια για το ανθρώπινο είδος και θεωρεί πως επικρατεί το στοιχείο της επιλογής για την τελική εξέλιξη της κατάστασης [3].

6.7 Η είσοδος των C1 και C2 στις μηχανές

Πώς θα μπορούσαν, λοιπόν, οι υπολογιστές να εμπλουτιστούν με υπολογισμούς των C1 κι C2; Θα γίνει πάλι αναφορά στο παράδειγμα του αυτοκινήτου με το λαμπάκι του καυσίμου. Στις μέχρι τώρα μηχανές, αυτό το λαμπάκι είναι ένα πρότυπο παράδειγμα ενός ασυνείδητου δυαδικού σήματος. Όταν η ένδειξη είναι ανοιχτή, όλοι οι υπόλοιποι επεξεργαστές της μηχανής παραμένουν ανενημέρωτοι και απαράλλαχτοι, καθώς καύσιμο συνεχίζει κανονικά να παρέχεται στο καρμπυρατέρ και το αυτοκίνητο προσπερνά τα βενζινάδικα χωρίς να σταματήσει (παρά το ότι μπορεί να εμφανίζονται στο σύστημα GPS του). Τα σύγχρονα αυτοκίνητα ή κινητά τηλέφωνα είναι απλές συλλογές ειδικευμένων ενοτήτων που ουσιαστικά έχουν άγνοια της μεταξύ τους ύπαρξης. Το προίκισμα αυτών των μηχανών με καθολική διαθεσιμότητα πληροφορίας (C1) θα επέτρεπε σε αυτές τις ενότητες να μοιράζονται πληροφορίες και να συνεισφέρουν στην αντιμετώπιση του επικείμενου προβλήματος (όπως ακριβώς κάνουν οι άνθρωποι όταν βλέπουν την ένδειξη έλλειψης καυσίμου).

Αν και η ΤΝ έχει συναντήσει αξιοσημείωτη επιτυχία στην επίλυση συγκεκριμένων προβλημάτων, η υλοποίηση πολλαπλών διαδικασιών σε ένα μόνο σύστημα και ο ευέλικτος συντονισμός τους παραμένουν δύσκολα προβλήματα. Τη δεκαετία του ’60, υπολογιστικές αρχιτεκτονικές με το όνομα «Συστήματα Μαυροπίνακα» σχεδιάστηκαν ειδικευμένα για να ποστάρουν πληροφορίες και να τις κάνουν διαθέσιμες και για άλλες ενότητες με έναν ευέλικτο και ερμηνεύσιμο τρόπο, παρόμοιο με τον καθολικό χώρο εργασίας. Μια πρόσφατη αρχιτεκτονική με το όνομα Pathnet χρησιμοποιεί έναν γενετικό αλγόριθμο για να μαθαίνει ποιο μονοπάτι των πολλών ειδικευμένων νευρικών δικτύων του είναι καταλληλότερο για μια συγκεκριμένη εργασία. Αυτή η αρχιτεκτονική επιδεικνύει ευρωστία, ευέλικτη επίδοση και γενίκευση μεταξύ των εργασιών και ίσως να αποτελεί ένα πρώτο βήμα προς μία πρώτης μορφής συνειδητή ευελιξία.

Για να γίνει βέλτιστη χρήση της πληροφορίας που παρέχεται από το λαμπάκι καυσίμου, θα ήταν επίσης χρήσιμο για το αυτοκίνητο να κατέχει μία βάση δεδομένων των δικών του δυνατοτήτων και ορίων. Αυτός ο τύπος αυτοελέγχου (C2) θα περιλάμβανε μια ενσωματωμένη εικόνα του εαυτού του, όπως για παράδειγμα γνώση της τοποθεσίας του και της κατανάλωσης καυσίμου, καθώς και των εσωτερικών δεδομένων του, όπως γνώση ότι διαθέτει ένα σύστημα GPS που μπορεί να εντοπίσει κοντινά σημεία ανατροφοδότησης. Μία μηχανή με τη δυνατότητα αυτοελέγχου θα κρατούσε μία λίστα των υποπρογραμμάτων της, θα υπολόγιζε προσεγγιστικά τις πιθανότητες επιτυχίας για μια ποικιλία εργασιών και συνεχώς θα τις ανανέωνε. Τα περισσότερα σύγχρονα συστήματα μηχανικής μάθησης στερούνται οποιουδήποτε αυτοελέγχου, αφού υπολογίζουν χωρίς

να παρουσιάζουν το εύρος και τα όρια της γνώσης τους ή το γεγονός ότι άλλοι μπορεί να έχουν μια διαφορετική οπτική γωνία από τη δική τους. Υπάρχουν βέβαια ορισμένες εξαιρέσεις: Τα Μπεϋζιανά δίκτυα ή προγράμματα υπολογίζουν με βάση κατανομές πιθανοτήτων και επομένως παρακολουθούν το πόσο πιθανό είναι οι υπολογισμοί τους να είναι σωστοί. Ακόμα και όταν ο βασικός υπολογισμός πραγματοποιείται από έναν κλασικό περιελιγμό νευρικού δικτύου (Convolutional Neural Network – CNN), και επομένως δεν είναι διαθέσιμος για ενδοσκόπηση, είναι εφικτό να εκπαιδευτεί ένα δεύτερο, ιεραρχικά ανώτερο νευρικό δίκτυο που θα προβλέπει την εκτέλεση εργασιών του πρώτου. Αυτή η προσέγγιση, κατά την οποία ένα σύστημα επαναχαρακτηρίζει τον εαυτό του, έχει χαρακτηριστεί ότι οδηγεί στην εμφάνιση εσωτερικών μοντέλων που είναι φύσει μεταγνωστικά και επιτρέπουν σε έναν παράγοντα να αναπτύξει μια κατανόηση του εαυτού του. Το Pathnet χρησιμοποιεί μια σχετική αρχιτεκτονική για να εντοπίζει ποιοι εσωτερικοί σχηματισμοί είναι περισσότερο επιτυχείς για μια συγκεκριμένη εργασία και χρησιμοποιεί αυτή τη γνώση για να καθοδηγήσει μεταγενέστερες επεξεργασίες. Τα ρομπότ έχουν, επίσης, προγραμματιστεί για να ελέγχουν τη γνωστική τους πρόοδο και τη χρησιμοποιούν για να προσανατολίζουν τους πόρους που μεγιστοποιούν την απόκτηση πληροφοριών σχετικά με ένα πρόβλημα, υποδεικνύοντας έτσι μια μορφή περιέργειας [19].

Ένα σημαντικό στοιχείο της C2 που λαμβάνει σχετικά λίγη προσοχή είναι ο έλεγχος της πραγματικότητας. Οι Μπεϋζιανές προσεγγίσεις στην TN έχουν αναγνωρίσει τη χρησιμότητα των γενετικών μοντέλων μάθησης που μπορούν από κοινού να χρησιμοποιηθούν για πραγματική αντίληψη (παρόν), για προσδοκώμενο σχεδιασμό (μέλλον) και για ενδοσκοπική ανάλυση (παρελθόν). Στους ανθρώπους, οι ίδιες αισθητήριες περιοχές λαμβάνουν μέρος τόσο στην αντίληψη όσο και στη φαντασία. Ως τέτοιοι, κάποιοι μηχανισμοί χρειάζεται να ξεχωρίζουν τις αυτοδημιουργούμενες δραστηριότητες από αυτές που τις προκάλεσε κάποιος εξωτερικός παράγοντας. Μια ισχυρή μέθοδος για την εκπαίδευση των γενετικών μοντέλων, με το όνομα αντιφατική μάθηση, περιλαμβάνει την ύπαρξη ενός δεύτερου δικτύου «ανταγωνισμού» απέναντι σε ένα γενετικό δίκτυο ώστε να εκτιμά την αυθεντικότητα των αυτοδημιουργούμενων αναπαραστάσεων. Όταν ένας τέτοιος έλεγχος της πραγματικότητας (C2) συνδυαστεί με C1 μηχανισμούς, μπορεί να προκύψει μια μηχανή που να παρουσιάζει πολύ καλή μίμηση της ανθρώπινης συνείδησης από την άποψη της υποστήριξης καθολικής πρόσβασης σε αντιληπτικές αναπαραστάσεις, έχοντας ταυτόχρονα μια άμεση αίσθηση ότι το περιεχόμενο της αποτελεί μια γνήσια αποτύπωση της κατάστασης του περιβάλλοντος εκείνη τη στιγμή.

6.8 Περαιτέρω πιθανότητες και επιπτώσεις

Λέξη κλειδί, βέβαια, στο παραπάνω συμπέρασμα είναι η μίμηση σχετικά με την ανθρώπινη συνείδηση. Για την ανάπτυξη, όμως, μηχανών με συνείδηση λογίζεται η γεφύρωση του υπολογιστικού επεξηγηματικού κενού ως κρίσιμης σημασίας, καθώς κάτι τέτοιο θα επέτρεπε την άμεση και ξεκάθαρη σύγκριση μεταξύ νευροϋπολογιστικών μηχανισμών που σχετίζονται με συνειδητές/αναφερόμενες υψηλού επιπέδου γνωστικές διαδικασίες και νευροϋπολογιστικών μηχανισμών που σχετίζονται με την επεξεργασία χαμηλού επιπέδου ασυνείδητων πληροφοριών.

Ουσιαστικά, θα επέτρεπε να αποφασιστεί αν υπάρχουν υπολογιστικές συσχετίσεις της συνείδησης με αντίστοιχη έννοια, όπως υπάρχουν νευροβιολογικές συσχετίσεις της συνείδησης. Ο όρος «υπολογιστικές συσχετίσεις της συνείδησης» προτίθεται να σημαίνει ελάχιστοι υπολογιστικοί μηχανισμοί επεξεργασίας που σχετίζονται ειδικά με συνειδητές πτυχές της γνωστικής λειτουργίας, αλλά όχι με ασυνείδητες πτυχές. Όσον αφορά το υπολογιστικό επεξηγηματικό κενό, ιδιαίτερη σημασία έχουν οι νευροϋπολογιστικές συσχετίσεις της συνείδησης, οι οποίες σχετίζονται για παράδειγμα με την αναπαράσταση, την αποθήκευση, την επεξεργασία και την τροποποίηση της πληροφορίας που λαμβάνει χώρα στα νευρικά δίκτυα.

Αν κάποιος αποδεχτεί ότι το υπολογιστικό επεξηγηματικό κενό είναι σημαντικό και ένα ουσιώδες εμπόδιο στην εκδήλωση μηχανικής συνείδησης, τότε το άμεσο ερευνητικό πρόγραμμα γίνεται αποφασιστικό για το πώς μπορεί να γεφυρωθεί αυτό το κενό. Ενθαρρυντικά, έχει γίνει μια τεράστια προσπάθεια τα τελευταία χρόνια από ερευνητές, συμπεριλαμβανομένων αρκετών της ΤΝ, που εξετάζουν θέματα σχετικά με ένα τέτοιο πρόγραμμα, ακόμα κι αν αυτή η συσχέτιση δε γίνεται σαφής. Ως παράδειγμα, μπορεί να αναφερθεί ότι ένας αριθμός προηγούμενων μοντέλων που είχαν ρητό στόχο την εξήγηση της συνείδησης βασίζονται στη διάκριση μεταξύ τοπικών συμβολικών αναπαραστάσεων, που συχνά χρησιμοποιούνται στην από πάνω προς τα κάτω ΤΝ έναντι κατανεμημένων νευρικών αναπαραστάσεων. Αυτά τα υβριδικά μοντέλα αρχίζουν με τον ισχυρισμό ότι η επεξεργασία συμβολικής πληροφορίας είναι καθ' εαυτή η βάση για την επεξεργασία συνειδητής πληροφορίας, και έτσι ουσιαστικά ενισχύουν το υπολογιστικό επεξηγηματικό κενό παρά παρέχουν κάποιου είδους λύση σε αυτό. Αντίθετα, άλλα μοντέλα σχετικά με τη μηχανική συνείδηση, μπορούν να θεωρηθούν ως προτάσεις συγκεκριμένων υπολογιστικών συσχετίσεων της συνείδησης, όπως η καθολική επεξεργασία πληροφορίας, χρησιμοποιώντας αναπαραστάσεις οι οποίες αναφέρονται από άλλες αναπαραστάσεις, αυτοέλεγχο, ποικίλες πτυχές των μηχανισμών προσοχής και τη γείωση των συμβόλων στα δεδομένα των αισθητήρων. Επιπρόσθετες υπολογιστικές συσχετίσεις της συνείδησης προτείνονται από μελέτες πάνω στις νευρογνωστικές αρχιτεκτονικές που αναμφισβήτητα σχετίζονται με την ΤΝ και στοχεύουν στη χαρτογράφηση υψηλότερων γνωστικών λειτουργιών στους νευροϋπολογιστικούς μηχανισμούς. Αντίστοιχα παραδείγματα περιλαμβάνουν το στοχευμένο γνωστικό έλεγχο της μνήμης και την εκμάθηση μιας γλώσσας.

Αν μπορέσει να εξακριβωθεί ένας ικανοποιητικός αριθμός νευροϋπολογιστικών συσχετίσεων, τότε ίσως αναπτυχθεί μια άμεση κατεύθυνση για τη διερεύνηση της πιθανότητας εκδήλωσης συνείδησης στις μηχανές, για την εξακρίβωση πιθανών χαρακτηριστικών που θα δρουν ως αντικειμενικά κριτήρια για την παρουσία/απουσία φαινόμενης συνείδησης σε μηχανές και ανθρώπους, και ίσως ακόμα για την καλύτερη κατανόηση της θεμελιώδους φύσης της συνείδησης. Η αναζήτηση των νευροϋπολογιστικών συσχετίσεων της συνείδησης είναι μια απαραίτητη προσπάθεια ανεξάρτητα από το τελικό αποτέλεσμα όλης αυτής της εργασίας. Ακόμα και αν δε βρεθεί καμία διαφορά κατά τη νευροϋπολογιστική εκτέλεση συνειδητών και ασυνείδητων γνωστικών λειτουργιών, αυτό θα είχε τεράστιες επιπτώσεις στη μέχρι τώρα οπτική μας σχετικά με τα προβλήματα μυαλού-εγκεφάλου [18].

Με άλλα λόγια, ένας πλήρης χαρακτηρισμός της υψηλού επιπέδου γνωστικής λειτουργίας με νευροϋπολογιστικούς όρους μπορεί να δείξει πώς η υποκειμενική εμπειρία εγείρεται μηχανιστικά.

Επιπλέον, ακόμα και αν η γεφύρωση του υπολογιστικού επεξηγηματικού κενού δεν παρέχει έναν πλήρη μηχανιστικό λογαριασμό για τις υποκειμενικές εμπειρίες, θα παρέχει και πάλι σημαντικές συνθήκες για τη διερεύνηση φαινόμενης συνείδησης σε κάποιο τεχνούργημα. Η επίτευξη προόδου σε αυτόν τον τομέα μπορεί ακόμα να απομυθοποιήσει το «δύσκολο πρόβλημα». Τουλάχιστον, το υπολογιστικό επεξηγηματικό κενό υποδεικνύει ότι το ζήτημα της μηχανικής συνείδησης αξιώνει πολύ μεγαλύτερη προσοχή από την ΤΝ, από ότι ιστορικά έχει λάβει.

6.9 Σύγχρονη πρακτική δράση

Πολύ μεγάλης σημασίας παραμένει το γεγονός ότι ο αυτοματισμός συνεχώς εξελίσσεται και η ρομποτική θα συνεχίσει να αλλάζει τη φύση του χώρου εργασίας. Αυτό βέβαια δημιουργεί φόβους πως οι μηχανές μπορούν να απειλήσουν των βίο ορισμένων εργαζομένων. Άλλωστε, αρκετοί ερευνητές της ΤΝ συνηγορούν πως ο αληθινός δρόμος προς την τεχνητή νοημοσύνη ανθρώπινου επιπέδου είναι η κατασκευή ενός εκπαιδευόμενου ρομπότ που να μαθαίνει πώς να πραγματοποιεί κάθε εργασία για την οποία πληρώνεται ο άνθρωπος. Όλη αυτή η κατάσταση αποτελεί και εύφορο έδαφος για την εμφάνιση τρομακτικών προφητικών άρθρων στο γενικό τύπο.

Αυτό το μοτίβο είναι το ίδιο όπως και στην περίπτωση της μοναδικότητας: κάποιος κάνει μια αναπόδεικτη, δυσοίωνη πρόβλεψη χωρίς κανένα έμπρακτο κύρος και καταφέρνει να εγείρει τον πεσιμισμό και το φόβο. Ευτυχώς, υπάρχουν καλύτερα βασισμένες απόψεις οι οποίες, ίσως επειδή είναι πιο εκλογικευμένες, δεν φτάνουν στην πρώτη γραμμή των ειδήσεων. Μία από τις επικρατούσες είναι πως, σε αντίθεση με τους σημερινούς χειρότερους φόβους των ανθρώπων, η ρομποτική μπορεί να βοηθήσει στην ανάπτυξη και όχι στο θάνατο της γνώσης του εργάτη, αρκεί οι υπεύθυνοι να είναι σε θέση να προετοιμάσουν το προσωπικό για αναπόφευκτες αλλαγές στη μέχρι τώρα εργασία τους, εξελίσσοντας τα χαρακτηριστικά τους ακόμα και με επανεκπαίδευση αν χρειαστεί.

Η ισχύουσα φυσικά συνθήκη στον σύγχρονο κόσμο είναι ότι οι δραστηριότητες νοημοσύνης χωρίζονται σε αυτές που είναι αλγοριθμικά προβλεπόμενες από κανόνες, συγκεκριμένα μοτίβα και στρατηγικές και αυτές που δεν μπορούν να αναπαρασταθούν αλγοριθμικά, όπως οι διαπροσωπικές σχέσεις με το προσωπικό ενός ξενοδοχείου, οι δημόσιες σχέσεις στη γενική βιομηχανική πρακτική ή οι εφευρετικοί τρόποι εκμάθησης στην εκπαίδευση. Αυτή η διχοτόμηση των ανθρώπινων εργασιακών ικανοτήτων και των αλγορίθμων απλώς σημαίνει ότι τα ρομπότ κατά βάση είναι ικανά να αντικαταστήσουν τον άνθρωπο σε δουλειές που ορίζονται ξεκάθαρα από κανόνες. Αλλά και για να είναι η δουλειά των ρομπότ αποτελεσματική πρέπει να βοηθάει και την άλλη πλευρά των ανθρώπινων ικανοτήτων: την καινοτομία και το ένστικτο. Η συνύπαρξη αυτών των δύο στα ρομπότ μπορεί να ανακατευθύνει ανθρώπους που κάνουν επαναληπτικές και νοητικά μη αποδοτικές δουλειές στην καινοτομία μέσω της εκπαίδευσης.

Η ουσία είναι, λοιπόν, ότι ένα αλγοριθμικό ρομπότ χρειάζεται τη φαντασία του ανθρώπινου εγκεφάλου προκειμένου να εξασφαλίσει ένα αποτελεσματικό σύστημα και κάποιος πρέπει να εξετάσει τα κύρια πρότζεκτ σχετικά με τον εγκέφαλο που μπορούν να οδηγήσουν σε μηχανές που όχι μόνο θα τρέχουν έναν αλγόριθμο αλλά θα λειτουργούν ενστικτωδώς και με φαντασία. Μαζικές είναι οι πρωτοβουλίες μοντελοποίησης του εγκεφάλου σε Ευρώπη, Αμερική και Κίνα με την

ελπίδα ότι η ενστικτώδης, εφευρετική και δημιουργική λειτουργία του εγκεφάλου μπορεί πρώτον να κατανοηθεί και έπειτα πιθανώς να αναπαραχθεί σε ένα ρομπότ [3].

Το 2013 η Ευρωπαϊκή Κομισιόν αποφάσισε να χρηματοδοτήσει ένα πρόγραμμα «Ανθρώπινος Εγκέφαλος» ύψους 1,19 δισεκατομμυρίων ευρώ σε βάθος 10 χρόνων. Ευρέως πολυπειθαρχικό, το πρόγραμμα έχει αρκετές κατευθυντήριες γραμμές:

1. Νευροπληροφορική: επίσημες μελέτες του τρόπου που τα τυπικά νευρικά μοντέλα επεξεργάζονται την πληροφορία.
2. Προσομοίωση του εγκεφάλου: αντί για τυποποίηση της νευρικής δομής, γίνεται προσπάθεια να μιμηθεί η νευρική αρχιτεκτονική του εγκεφάλου.
3. Προηγμένες αρχιτεκτονικές: υπερ-προγραμματισμός, παράλληλες προσεγγίσεις για τη δημιουργία νευρικών συστημάτων
4. Ιατρική πληροφορική: Κατανόηση και διαχείριση ασθενειών μέσω της τεχνολογίας της πληροφορικής
5. Νευρομορφικός προγραμματισμός: σχεδιασμός και εκτέλεση ηλεκτρονικών εξαρτημάτων νευρολογικής μορφής
6. Νευρορομποτική: μελέτη των νευροεπιστημονικών μοντέλων ενσωματωμένα και επηρεασμένα από τα πλαίσια της ρομποτικής

Το πρόγραμμα διευθύνεται από τον Henry Markram της Πολυτεχνικής Σχολής της Λωζάνης και είναι ισχυρά βασισμένο στη φιλοσοφία ότι ο καλύτερος τρόπος για να προσεγγιστούν οι ικανότητες του εγκεφάλου είναι να εισαχθεί όσο περισσότερος βιολογικός ρεαλισμός γίνεται για τις πιο απαιτητικές προσομοιώσεις του εγκεφάλου που πραγματοποιούνται σε υπερ-υπολογιστές. Τέτοιος ρεαλισμός πρέπει να σχεδιάζεται από τα αποτελέσματα της βιολογικής νευροεπιστήμης που συνεχώς ανανεώνονται. Το πρόγραμμα ένωσε 100 από τα μεγαλύτερα εργαστήρια κυρίως της Ευρώπης για τη συνεισφορά της γνώσης και των αποτελεσμάτων τους. Ωστόσο, 3 χρόνια μετά την έναρξη του, το πρότζεκτ άρχισε να αντιμετωπίζει προβλήματα. Η αντίρρηση βασιζόταν στη συνειδητοποίηση ότι όσο περίπλοκα και ακριβά και να ήταν τα υπολογιστικά μοντέλα του εγκεφάλου, δεν θα μπορέσουν να αποδώσουν κάποια θεωρία που να σχετίζεται με το συνειδητό κομμάτι του μυαλού. Η θεωρία πρέπει να προέλθει από βαθιά γνώση της νευροεπιστήμης. Είναι περίπου το ίδιο με το να πει κάποιος, πως αν επενδυθούν πολλά χρήματα σε ένα τηλεσκόπιο θα αποκαλυφθούν όλα τα μυστικά του σύμπαντος, χωρίς να επενδυθεί κάτι στην αστρονομία ως επιστήμη. Το πρόγραμμα λοιπόν συνεχίζει στη βάση ότι υπολογιστικές επενδύσεις είναι πολύ πιθανό να εγείρει πρόοδο στον έξυπνο προγραμματισμό και όχι πληροφορίες για τη φύση της συνείδησης. Για τη δημιουργία ενός μοντέλου με τις ενστικτώδεις και εφευρετικές λειτουργίες του εγκεφάλου, κάποιος πρέπει να βασιστεί σε υποθέσεις για το πώς αυτές προκύπτουν.

7. Η εφαρμογή

7.1 Εισαγωγή

Έχοντας κάνει μια πρώτη περιγραφή για τις υπάρχουσες θεωρίες σχετικά με τη μέτρηση της συνείδησης, στο σημείο αυτό θα γίνει μια πιο πρακτική προσέγγιση σε αυτή των Ενσωματωμένων Πληροφοριών (Integrated Information Theory – IIT). Η κεντρική υπόθεση είναι ότι ένα φυσικό σύστημα πρέπει να πληροί πέντε απαιτήσεις («αξιώματα»), ώστε να αποτελεί ένα φυσικό υπόστρωμα μιας υποκειμενικής εμπειρίας: (1) εσωτερική ύπαρξη (το σύστημα πρέπει να είναι σε θέση να ξεχωρίζει τον εαυτό του), (2) σύνθεση (πρέπει να αποτελείται από μέρη που έχουν αιτιατή ισχύ πάνω στο σύνολο), (3) πληροφορία (η αιτιατή ισχύς του πρέπει να είναι συγκεκριμένη), (4) ενσωμάτωση (η αιτιατή ισχύς του δεν πρέπει να υπόκειται σε αυτή των μερών του), (5) εξαίρεση (πρέπει να είναι μέγιστα μη αναστρέψιμο).

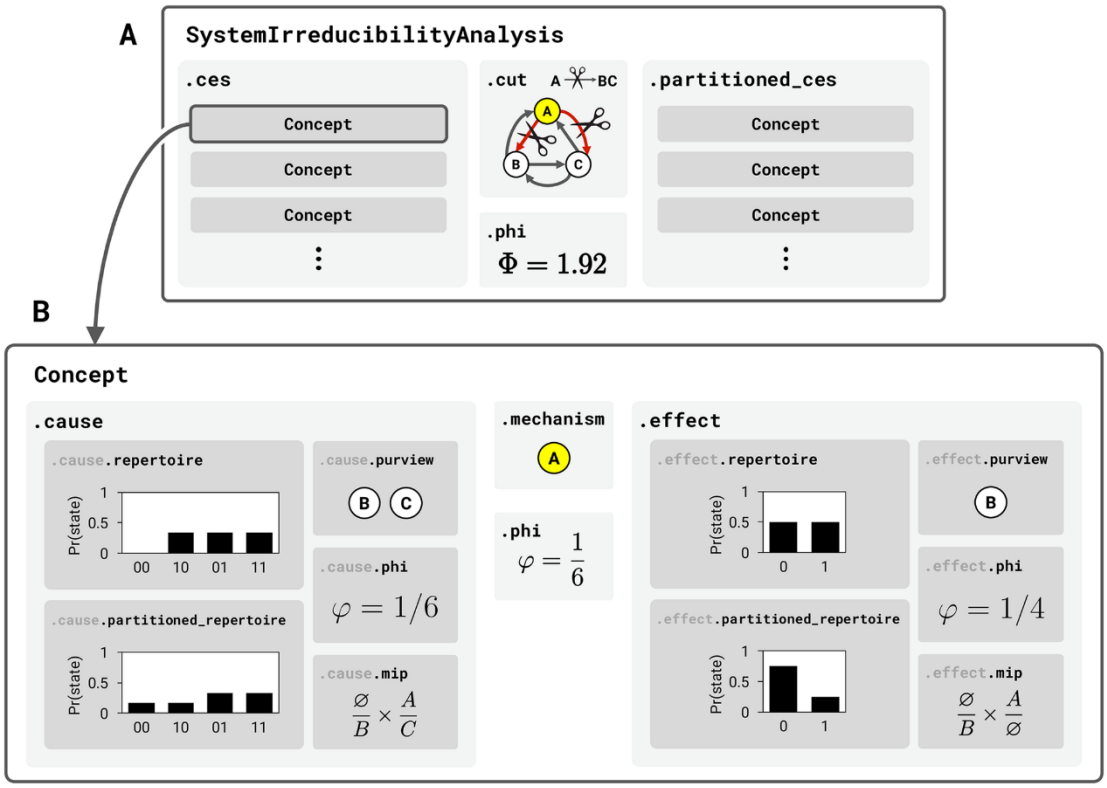
Από αυτά τα αξιώματα, η IIT αναπτύσσει ένα μαθηματικό πλαίσιο για την εκτίμηση της δομής αιτίου-αποτελέσματος (cause-effect structure – CES) ενός φυσικού συστήματος, που είναι εφαρμόσιμο σε διακριτά δυναμικά συστήματα. Αυτό το πλαίσιο έχει αποδειχτεί όχι μόνο χρήσιμο σε μελέτες για τη συνείδηση, αλλά και εφαρμόσιμο σε έρευνες πάνω σε συγκεκριμένες βιολογικές ερωτήσεις.

Το βασικό μέτρο της ισχύος αιτίου-αποτελέσματος, δηλαδή η ενσωματωμένη πληροφορία (υποδηλωμένη ως Φ), ποσοτικοποιεί πόσο μη αναστρέψιμη είναι η CES ενός συστήματος ως προς τα μέρη του. Το Φ επίσης αποτελεί ένα γενικό μέτρο πολυπλοκότητας, που συλλαμβάνει το εύρος στο οποίο ένα σύστημα είναι τόσο ενσωματωμένο όσο και (πληροφοριακά) διαφοροποιημένο.

Εδώ, λοιπόν, θα περιγραφεί το *PyPhi*, ένα λογισμικό πακέτο γραμμένο σε Python που υποστηρίζει αυτό το πλαίσιο της IIT για αιτιατή ανάλυση και ξετυλίγει τη συνολική CES ενός διακριτού Μαρκοβιανού δυναμικού συστήματος δυαδικών στοιχείων. Το λογισμικό επιτρέπει στο χρήστη μια εύκολη μελέτη αυτών των δομών και προσφέρει μια ανανεωμένη αναφορά τη εφαρμογής των φορμαλισμών της IIT.

7.2 Λειτουργία Εφαρμογής

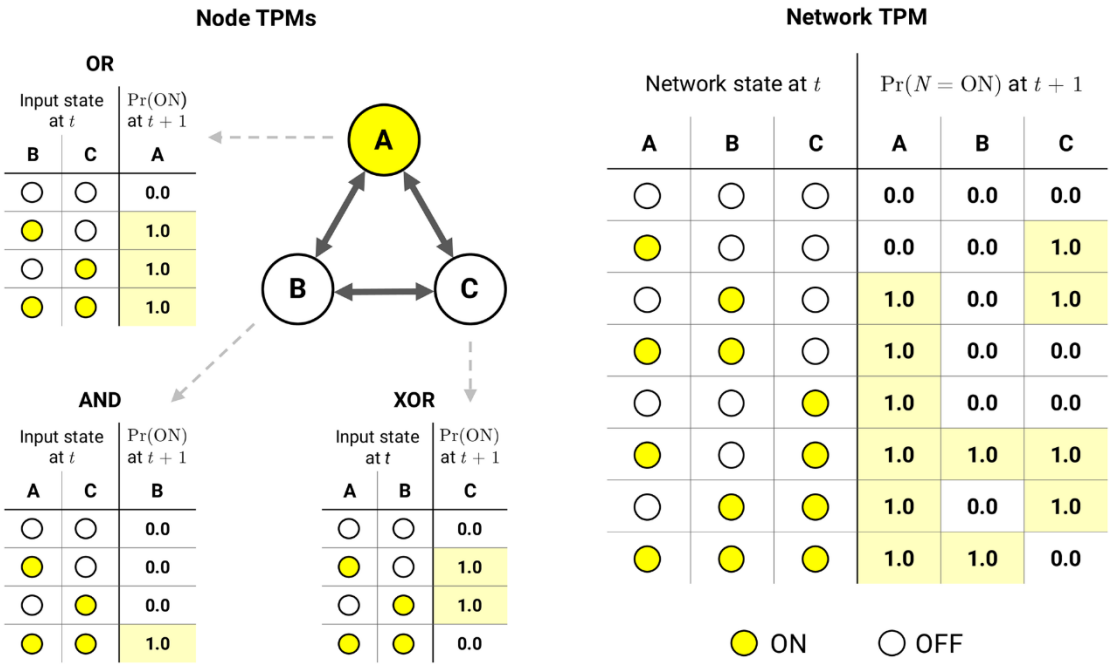
Το λογισμικό έχει δύο κύριες λειτουργίες: (1) να ξετυλίξει τη συνολική CES ενός διακριτού δυναμικού συστήματος αλληλεπιδρώντων στοιχείων και να υπολογίσει το αντίστοιχο Φ, και (2) να υπολογίσει το μέγιστα μη αναστρέψιμο εύρος αιτίου-αποτελέσματος ενός συγκεκριμένου σετ στοιχείων μέσα στο σύστημα [20]. Η πρώτη λειτουργία πραγματοποιείται με την εντολή *pyphi.compute.major_complex()*, η οποία επιστρέφει το αντικείμενο *SystemIrreducibilityAnalysis* (**Σχήμα 15Α**). Η CES του συστήματος περιέχεται στο χαρακτηριστικό *ces*, ενώ το Φ στο *phi*.



Σχήμα 15: (A): Το αντικείμενο *SystemIrreducibilityAnalysis* είναι η βασική έξοδος του λογισμικού. Αντιπροσωπεύει τα αποτελέσματα της ανάλυσης του συστήματος. Περιλαμβάνει αρκετά χαρακτηριστικά: Το αντικείμενο *CauseEffectStructure* που περιέχει όλα τα *Concepts* του συστήματος, το αντικείμενο *Cut* που αντιπροσωπεύει τη διαμέριση ελάχιστης πληροφορίας (Minimum Information Partition - MIP) του συστήματος, το *partitioned_ces* είναι το *CuaseEffectStructure* των *Concepts* που καθορίζεται από το σύστημα αφού εφαρμοστεί το MIP, και το *phi*, δηλαδή την τιμή Φ που μετράει τη διαφορά των μη διαμερισμένων και των διαμερισμένων CES. (B): Ένα *Concept* αντιπροσωπεύει τη μέγιστα μη αναστρέψιμη αιτία (MIC) και το μέγιστα μη αναστρέψιμο αποτέλεσμα (MIE) ενός μηχανισμού σε μία κατάσταση. Το χαρακτηριστικό *mechanism* περιέχει τους δείκτες των στοιχείων του μηχανισμού. Τα χαρακτηριστικά *cause* και *effect* περιέχουν τα αντικείμενα *MaximallyIrreducibleCause* και *MaximallyIrreducibleEffect* που περιγράφουν τα MIC και MIE του μηχανισμού αντίστοιχα. Το καθένα περιλαμβάνει από ένα όριο (purview), εύρος (repertoire), MIP, διαμερισμένο εύρος (partitioned_repertoire), και τιμή φ (phi). Το χαρακτηριστικό *phi* περιέχει την τιμή φ, που είναι η ελάχιστη των τιμών φ από τα MIC και MIE [20].

Η CES προκύπτει από τα αντικείμενα *Concept*, τα οποία αποτελούν την έξοδο της δεύτερης κύριας λειτουργίας, που πραγματοποιείται με την εντολή *Subsystem.concept()* (**Σχήμα 15B**). Κάθε *Concept* καθορίζεται από ένα σετ στοιχείων μέσα στο σύστημα. Ένα *Concept* περιέχει ένα μέγιστα μη αναστρέψιμο αίτιο και αποτέλεσμα (*cause_repertoire* και *effect_repertoire*), τα οποία είναι οι κατανομές πιθανότητας που συλλαμβάνουν το πώς τα στοιχεία του μηχανισμού στη συγκεκριμένη κατάσταση περιορίζουν την προηγούμενη και την επόμενη κατάσταση του συστήματος αντίστοιχα.

Το σημείο εκκίνησης για την ανάλυση με βάση την IIT είναι ένα διακριτό Μαρκοβιανό δυναμικό σύστημα *S*, αποτελούμενο από *n* αλληλεπιδρώντα στοιχεία. Ένα τέτοιο σύστημα μπορεί να αναπαρασταθεί από έναν γράφο διασυνδεδεμένων κόμβων, καθένας από τους οποίους φέρει μια λειτουργία που δίνει ως έξοδο την κατάσταση του κόμβου για την επόμενη χρονική στιγμή $t+1$ δεδομένης της κατάστασης των γονέων του και της προηγούμενης χρονικής στιγμής t (**Σχήμα 16**).



Σχήμα 16: Ένα δίκτυο κόμβων και το TPM του. Κάθε κόμβος έχει το δικό του TPM- σε αυτή την περίπτωση, τον πίνακα αληθείας μιας ντετερμινιστικής λογικής πύλης. Το κίτρινο σηματοδοτεί την κατάσταση '1', ενώ το άσπρο την ΄0΄. Το TPM του συστήματος (δεξιά) συντίθεται από τα TPMs των κόμβων του (αριστερά), ενώ εδώ δείχνεται σε μορφή κατάσταση-κατά-κόμβο. Πρέπει να σημειωθεί ότι για την αναπαράσταση του TPM στο PyPhi, η κατάσταση του πρώτου κόμβου ποικίλει, σύμφωνα με τη σύμβαση μικρότερης σημαντικότητας [20].

Μέχρι στιγμής, το PyPhi μπορεί να αναλύσει τόσο ντετερμινιστικά όσο και στοχαστικά διακριτά Μαρκοβιανά δυναμικά συστήματα αποτελούμενα από στοιχεία δύο καταστάσεων.

Κάθε τέτοιο διακριτό δυναμικό σύστημα ορίζεται πλήρως από τον πίνακα μετάβασης πιθανότητας (Transition Probability Matrix – TPM), που περιλαμβάνει όλες της πιθανότητες για την κατάσταση μετάβασης από το t στο t+1. Μπορεί να προκύψει από τη γραφική αναπαράσταση του συστήματος, διαταράσσοντας το σύστημα σε κάθε πιθανή του κατάσταση και παρατηρώντας την κατάσταση που ακολουθεί την επόμενη χρονική στιγμή (για στοχαστικά συστήματα χρειάζονται επαναλαμβανόμενες δοκιμές της παραπάνω διαδικασίας, για να προκύψουν ασφαλώς οι πιθανότητες μετάβασης των καταστάσεων). Στο PyPhi, το TPM είναι η θεμελιώδης αναπαράσταση του συστήματος.

Τυπικά, αν S_t είναι η τυχαία μεταβλητή της κατάστασης του συστήματος τη χρονική στιγμή t, το TPM καθορίζει τη σχετική κατανομή πιθανότητας για την επόμενη κατάσταση S_{t+1}, δεδομένης της υπάρχουσας κατάστασης s_t:

$$\Pr(S_{t+1} \mid S_t = s_t), \text{ για κάθε } s_t \in \Omega_s,$$

όπου το Ω_s δηλώνει το σύνολο των πιθανών καταστάσεων. Επιπλέον, δεδομένης μιας οριακής κατανομής για τις προηγούμενες καταστάσεις του συστήματος, το TPM καθορίζει πλήρως την ενοποιημένη κατανομή για την κατάσταση μετάβασης. Εδώ η IIT επιβάλλει ομοιομορφία στην οριακή κατανομή των προηγούμενων καταστάσεων, επειδή ο σκοπός της ανάλυσης είναι να

συλληφθούν οι άμεσες αιτιατές σχέσεις κατά τη διάρκεια μίας χρονικής στιγμής χωρίς την παρουσία παραγόντων σύγχυσης, όπως πιθανές επιρροές από τις καταστάσεις του συστήματος τη στιγμή *t*-1 και νωρίτερα. Έτσι, η οριακή κατανομή αντιστοιχεί σε μια μεσολαβητική (αιτιατή), και όχι μέσω παρατήρησης, κατανομή της κατάστασης.

Επιπρόσθετα, η IIT θεωρεί ότι δεν υπάρχουν στιγμιαία αίτια, δηλαδή θεωρείται ότι τα στοιχεία του δυναμικού συστήματος επηρεάζουν το ένα το άλλο μόνο από μία χρονική στιγμή προς την επόμενη. Επομένως το σύστημα απαιτείται να ικανοποιεί την ακόλουθη Μαρκοβιανή συνθήκη, η οποία καλείται ιδιότητα σχετικής ανεξαρτησίας: η κατάσταση κάθε στοιχείου στο *t*+1 πρέπει να είναι ανεξάρτητη από την κατάσταση των υπόλοιπων, δεδομένης μιας κατάστασης του συστήματος στο *t*.

$$\Pr(S_{t+1} \mid S_t = s_t) = \prod_{N \in S} \Pr(N_{t+1} \mid S_t = s_t), \text{ για κάθε } s_t \in S.$$

Για συστήματα δυαδικών στοιχείων, το TPM που ικανοποιεί τη σχέση Χ μπορεί να αναπαρασταθεί σε κατάσταση-κατά-κόμβο μορφή (σχέση Ψ, δεξιά), καθώς είναι αναγκαίο να αποθηκευτεί μόνο η οριακή κατανομή κάθε στοιχείου, αντί για τη συνολική ενωμένη κατανομή.

Στο PyPhi, το σύστημα υπό ανάλυση αναπαρίσταται από το αντικείμενο *Network*. Ένα *Network* δημιουργείται περνώντας το TPM του ως πρώτη μεταβλητή στη διαδικασία: *network = pyphi.Network(tpm)*. Επίσης, υπάρχει και η επιλογή ενός συνεκτικού πίνακα (CM), όπου

$$[CM]_{i\,j} = \begin{cases} 1 & \text{αν υπάρχει ένωση μεταξύ των στοιχείων } i \text{ και } j \\ 0 & \text{αλλιώς} \end{cases}$$

Αυτό γίνεται περνώντας ως μεταβλητή τη λέξη κλειδί *cm*: *network=pyphi.Network(tpm, cm=cm)*. Επειδή το TPM καθορίζει πλήρως το σύστημα, η παροχή ενός CM δεν είναι απαραίτητη. Ωστόσο, μια σαφής συνεκτική πληροφορία μπορεί να χρησιμοποιηθεί για να κάνει πιο αποδοτικούς τους υπολογισμούς, ειδικά για αραιά δίκτυα, αφού το PyPhi μπορεί να αποκλείσει ορισμένες αιτιατές επιρροές εξ' αρχής αν υπολείπονται συνδέσεων. Ας σημειωθεί ότι η παροχή ενός λανθασμένου CM μπορεί να οδηγήσει σε μη ακριβές αποτέλεσμα. Αν δεν δίνεται ο CM, το PyPhi υποθέτει πλήρη συνδεσιμότητα, δηλαδή ότι το κάθε στοιχείο μπορεί να επηρεάζει όλα τα υπόλοιπα, κάτι που οδηγεί με βεβαιότητα σε σωστά αποτελέσματα [20].

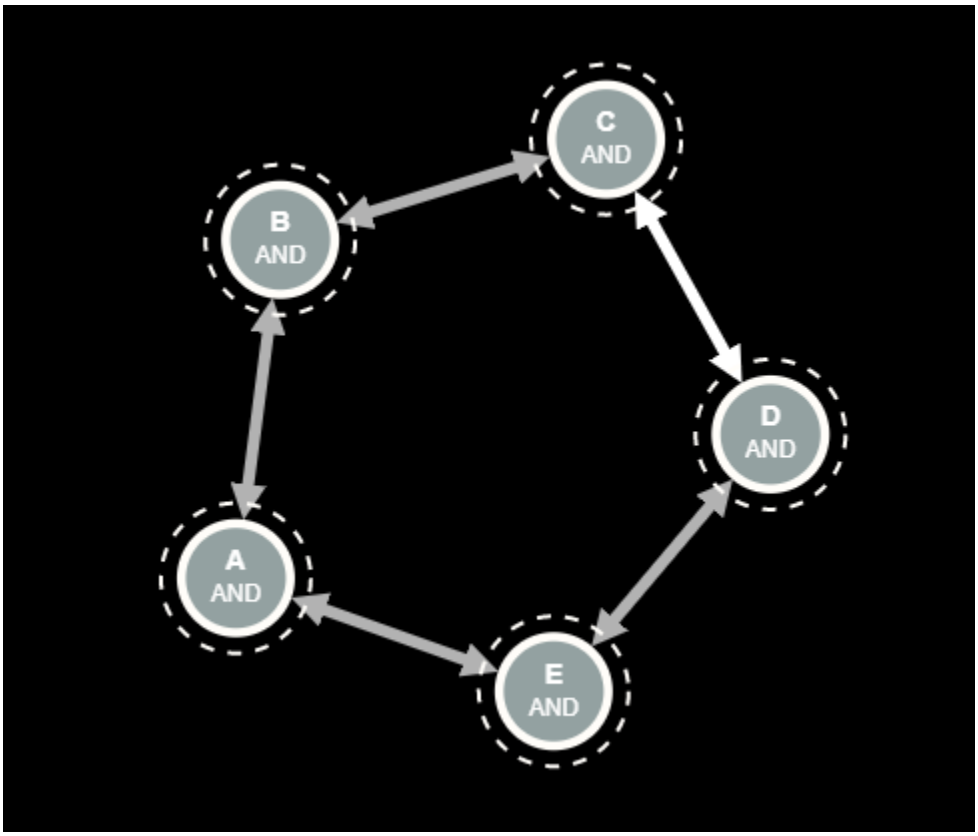
Τέλος, από τη στιγμή που δημιουργηθεί το *Network*, ένα υποσύνολο στοιχείων του συστήματος (που καλείται υποψήφιο σύστημα), μαζί με μια συγκεκριμένη κατάσταση του συστήματος, μπορεί να επιλεχθεί για ανάλυση, με τη δημιουργία του αντικειμένου *Subsystem*.

7.3 Παραδείγματα

Στη συνέχεια θα παρατεθούν παραδείγματα 3 κυκλωμάτων για καλύτερη κατανόηση του τι προσφέρει η εφαρμογή. Φυσικά, λόγω του αρχικού σταδίου στο οποίο βρίσκεται η γενικότερη έρευνα, τα κυκλώματα των παραδειγμάτων είναι απλά και δεν γίνεται να αντιστοιχηθούν με κάποιο τμήμα μίας διαδικασίας του εγκεφάλου. Ωστόσο, έχουν συγκριτική αξία και προσφέρουν οικειοποίηση με την έννοια του Φ.

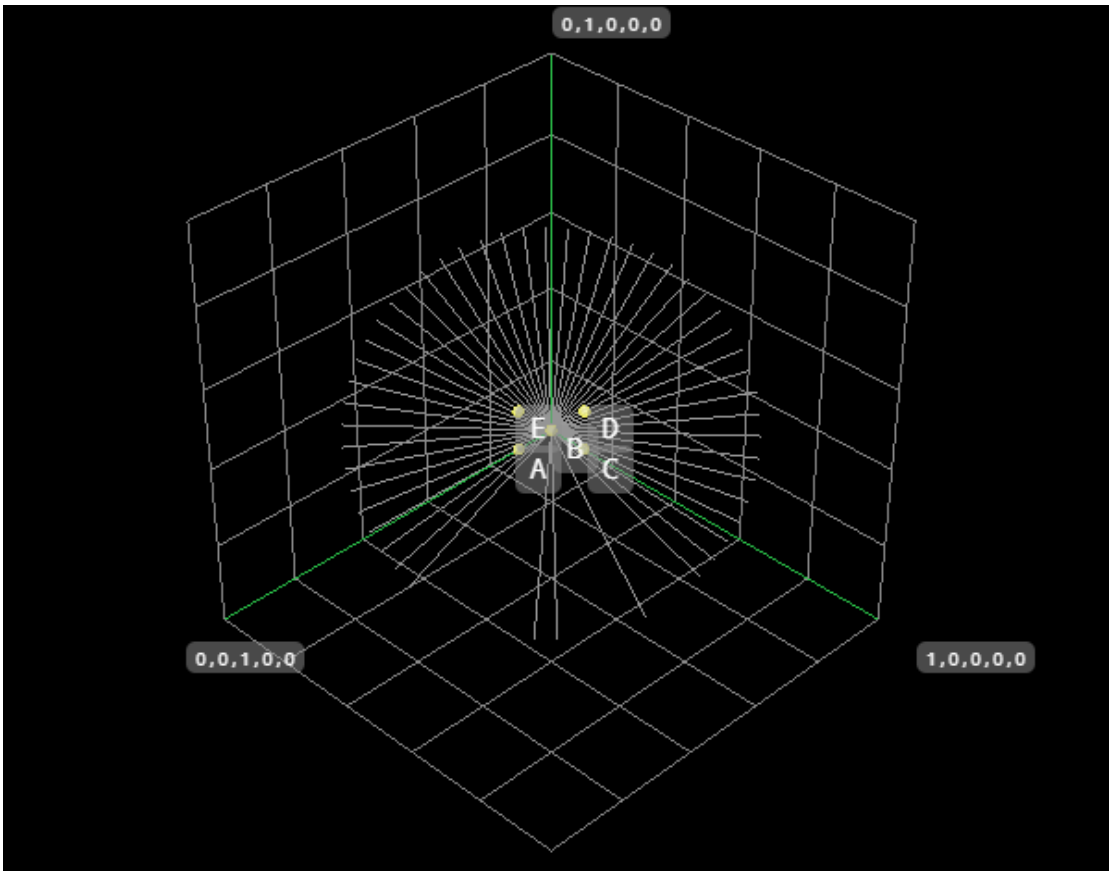
Παράδειγμα 1:

Για το πρώτο παράδειγμα, χρησιμοποιείται ένα σύστημα 5 λογικών πυλών AND (A, B, C, D, E) που συνδέονται κυκλικά με τις γειτονικές τους με κανάλι αμφίδρομης διάδοσης. Το σύστημα, λοιπόν, έχει τη μορφή του **Σχήματος 17**.



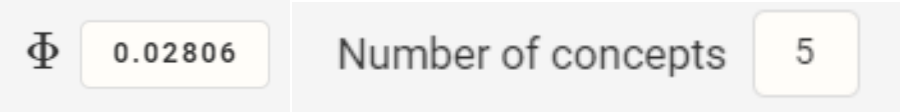
Σχήμα 17: Μορφή συστήματος 5 πυλών AND αμφίδρομης διάδοσης.

Αν θεωρηθεί ότι οι τιμές των πυλών στην τωρινή κατάσταση είναι (A,B,C,D,E)=(0,0,0,0,0), με την εντολή *pyphi.compute.major_complex()* παρέχονται οι πληροφορίες για το Φ του συστήματος, ποιο είναι το κύριο σύμπλεγμα, μια μορφή της CES του συστήματος, καθώς και τον αριθμό των *Concepts* που περιλαμβάνονται. Στα **Σχήματα 18-20** βρίσκονται αυτά τα στοιχεία για την παραπάνω περίπτωση περίπτωση.



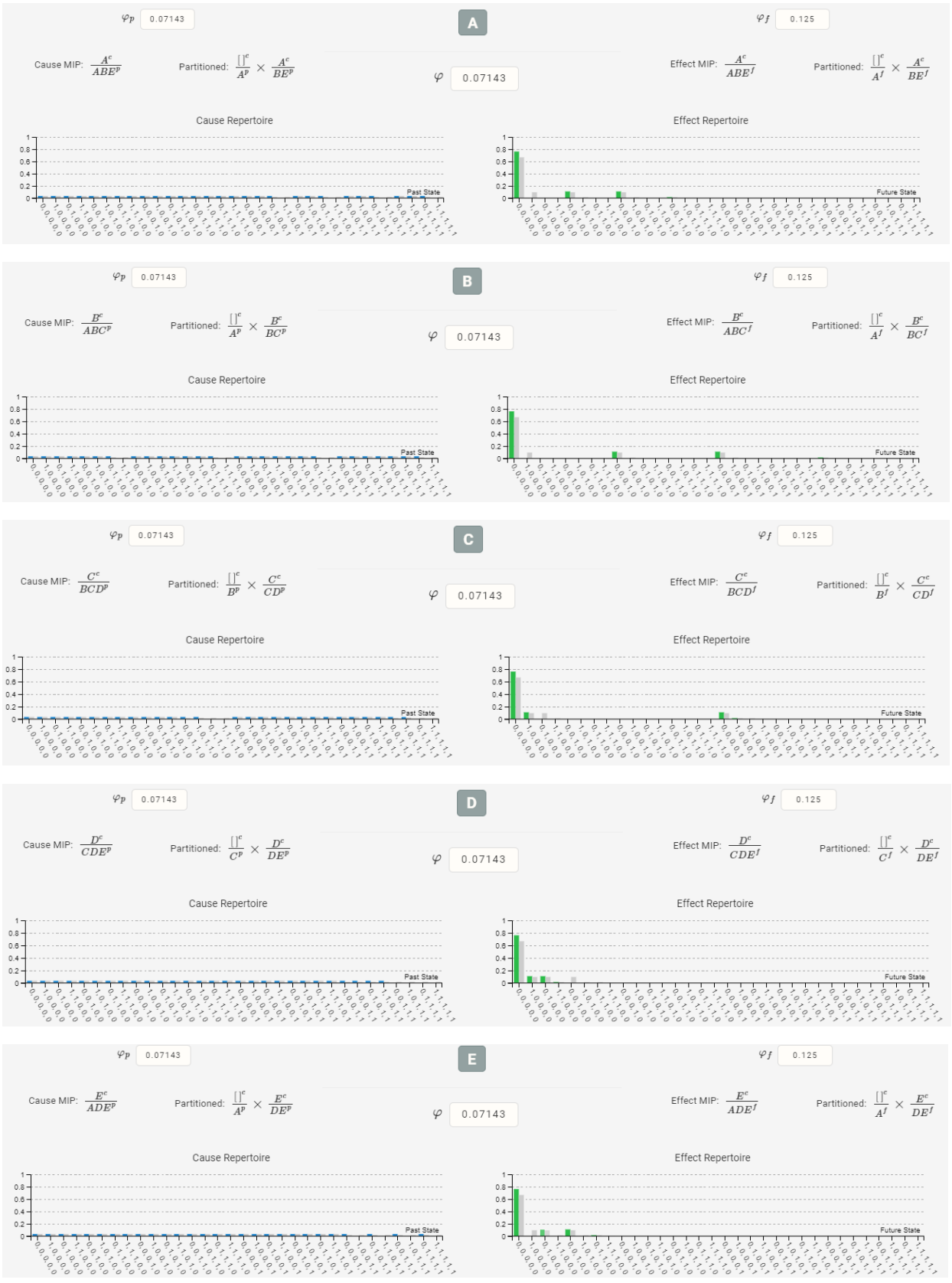
Σχήμα 18: Μορφή CES συστήματος κόμβων AND (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Σχήμα 19: Κύριο σύμπλεγμα συστήματος κόμβων AND (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).



Σχήμα 20: Τιμή Φ και αριθμός *Concepts* συστήματος κόμβων AND (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Από το αντικείμενο *SystemIrreducibilityAnalysis* προκύπτουν και οι ειδικότερες πληροφορίες για κάθε *Concept*, οι οποίες παρουσιάζονται στο **Σχήμα 21**.



Σχήμα 21: Το κάθε *Concept* συστήματος κόμβων AND (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Για την καλύτερη εύρεση των αποτελεσμάτων, παρατίθεται πίνακας με συγκεντρωμένες τις βασικότερες μετρήσεις.

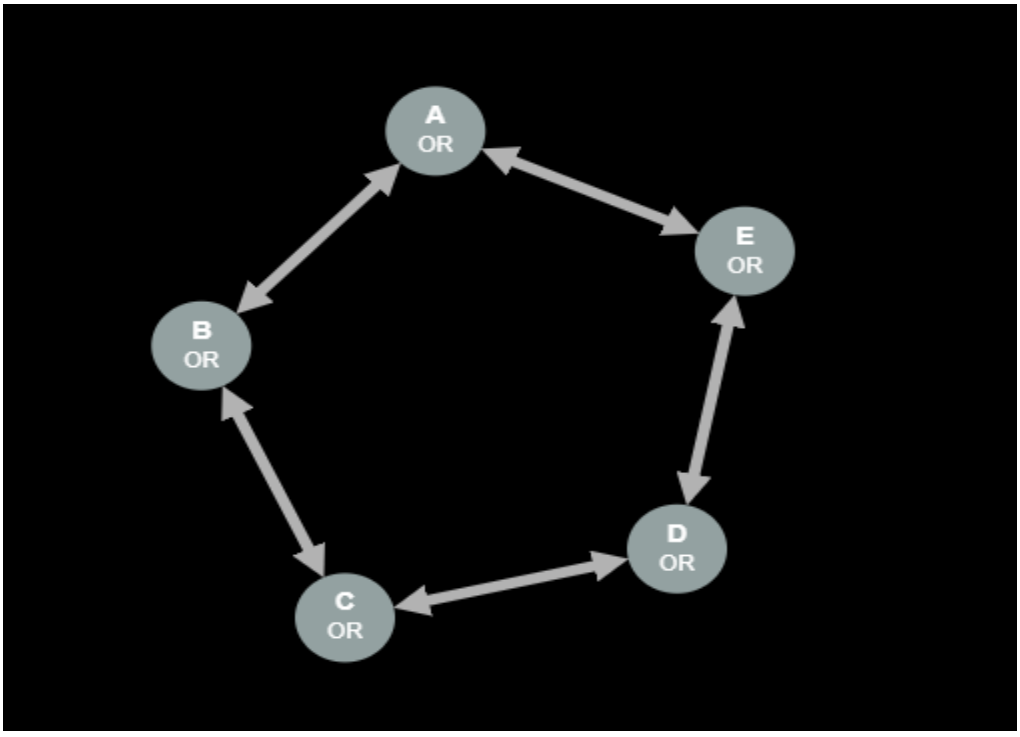
Πίνακας 1: Πίνακας μετρήσεων συστήματος 5 πυλών AND αμφίδρομης διάδοσης.

(A,B,C,D,E)	**Main Complex**	**Φ Value**	**N.o.C**	**φ of Con.A**	**φ of Con.B**	**φ of Con.C**	**φ of Con.D**	**φ of Con.E**
(0,0,0,0,0)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.07143	0.07143	0.07143
(0,0,0,0,1)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.07143	0.07143	0.125
(0,0,0,1,0)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.07143	0.125	0.07143
(0,0,0,1,1)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.07143	0.125	0.125
(0,0,1,0,0)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.125	0.07143	0.07143
(0,0,1,0,1)	A,B	0.06945	2	0.16667	0,16667	-	-	-
(0,0,1,1,0)	A,B,C,D,E	0.02806	5	0.07143	0.07143	0.125	0.125	0.07143
(0,0,1,1,1)	A,B	0.06945	2	0.16667	0.16667	-	-	-
(0,1,0,0,0)	A,B,C,D,E	0.02806	5	0.07143	0.125	0.07143	0.07143	0.07143
(0,1,0,0,1)	C,D	0.06945	2	-	-	0.16667	0.16667	-
(0,1,0,1,0)	A,E	0.06945	2	0.16667	-	-	-	0.16667
(0,1,0,1,1)	-	0	0	-	-	-	-	-
(0,1,1,0,0)	A,B,C,D,E	0.02806	5	0.07143	0.125	0.125	0.07143	0.07143
(0,1,1,0,1)	-	0	0	-	-	-	-	-
(0,1,1,1,0)	A,E	0.06945	2	0.16667	-	-	-	0.16667
(0,1,1,1,1)	C,D	0.1875	2	-	-	0.25	0.25	-
(1,0,0,0,0)	A,B,C,D,E	0.02806	5	0.125	0.07143	0.07143	0.07143	0.07143
(1,0,0,0,1)	A,B,C,D,E	0.02806	5	0.125	0.07143	0.07143	0.07143	0.125
(1,0,0,1,0)	B,C	0.06945	2	-	0.16667	0.16667	-	-
(1,0,0,1,1)	B,C	0.06945	2	-	0.16667	0.16667	-	-
(1,0,1,0,0)	D,E	0.06945	2	-	-	-	0.16667	0.16667
(1,0,1,0,1)	-	0	0	-	-	-	-	-
(1,0,1,1,0)	-	0	0	-	-	-	-	-

(1,0,1,1,1)	D,E	0.1875	2	-	-	-	0.25	0.25
(1,1,0,0,0)	A,B,C,D,E	0.02806	5	0.125	0.125	0.07143	0.07143	0.07143
(1,1,0,0,1)	C,D	0.06945	2	-	-	0.16667	0.16667	-
(1,1,0,1,0)	-	0	0	-	-	-	-	-
(1,1,0,1,1)	A,E	0.1875	2	0.25	-	-	-	0.25
(1,1,1,0,0)	D,E	0.06975	2	-	-	-	0.16667	0.16667
(1,1,1,0,1)	A,B	0.1875	2	0.25	0.25	-	-	-
(1,1,1,1,0)	B,C	0.1875	2	-	0.25	0.25	-	-
(1,1,1,1,1)	A,B	0.1875	2	0.25	0.25	-	-	-

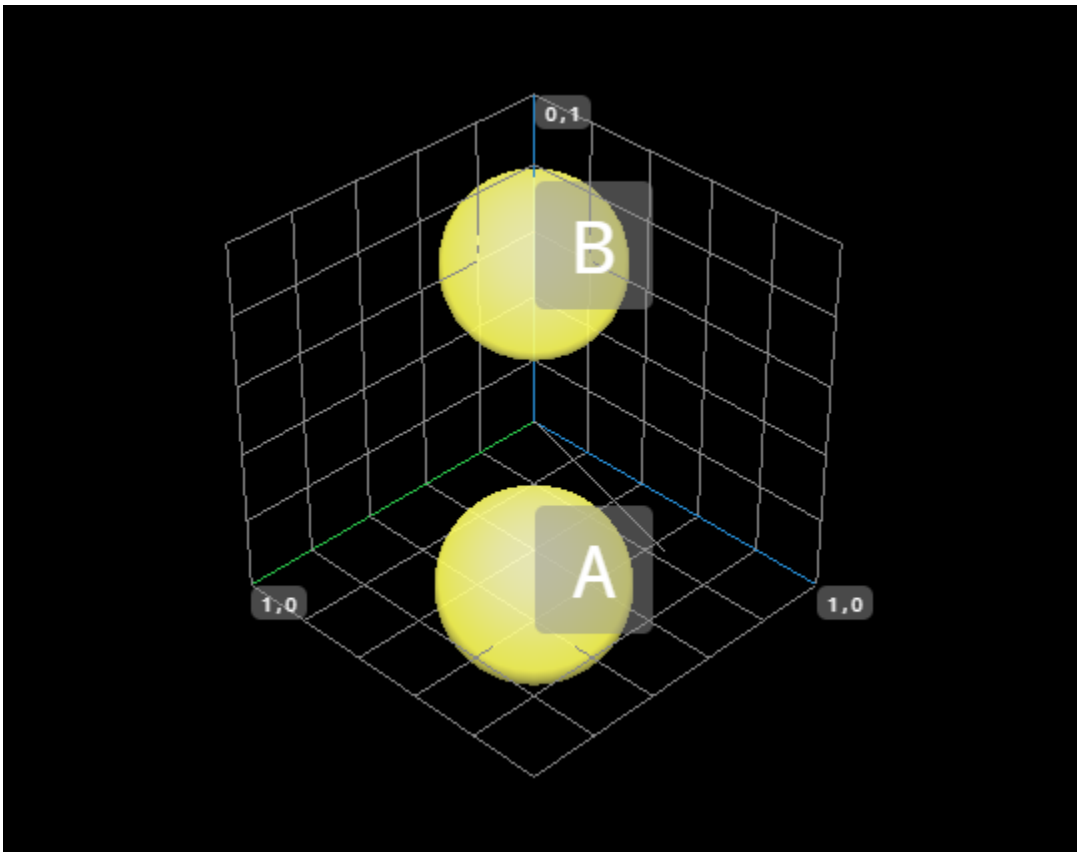
Παράδειγμα 2:

Για το δεύτερο παράδειγμα, χρησιμοποιείται ένα σύστημα 5 λογικών πυλών OR (A, B, C, D, E) που συνδέονται κυκλικά με τις γειτονικές τους με κανάλι αμφίδρομης διάδοσης. Το σύστημα, λοιπόν, έχει τη μορφή του **Σχήματος 23**.

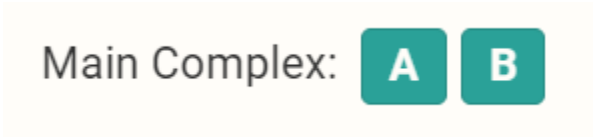


Σχήμα 22: Μορφή συστήματος 5 πυλών OR αμφίδρομης διάδοσης.

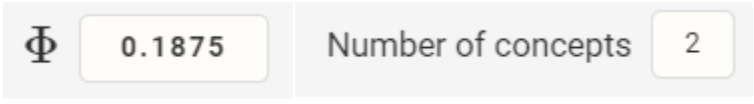
Ακολουθείται η ίδια διαδικασία με το *Παράδειγμα 1* χρησιμοποιώντας αρχικά τις τιμές (A,B,C,D,E)=(0,0,0,0,0). Στα **Σχήματα 23-25** βρίσκονται τα πρώτα αποτελέσματα που προκύπτουν.



Σχήμα 23: Μορφή CES συστήματος κόμβων OR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

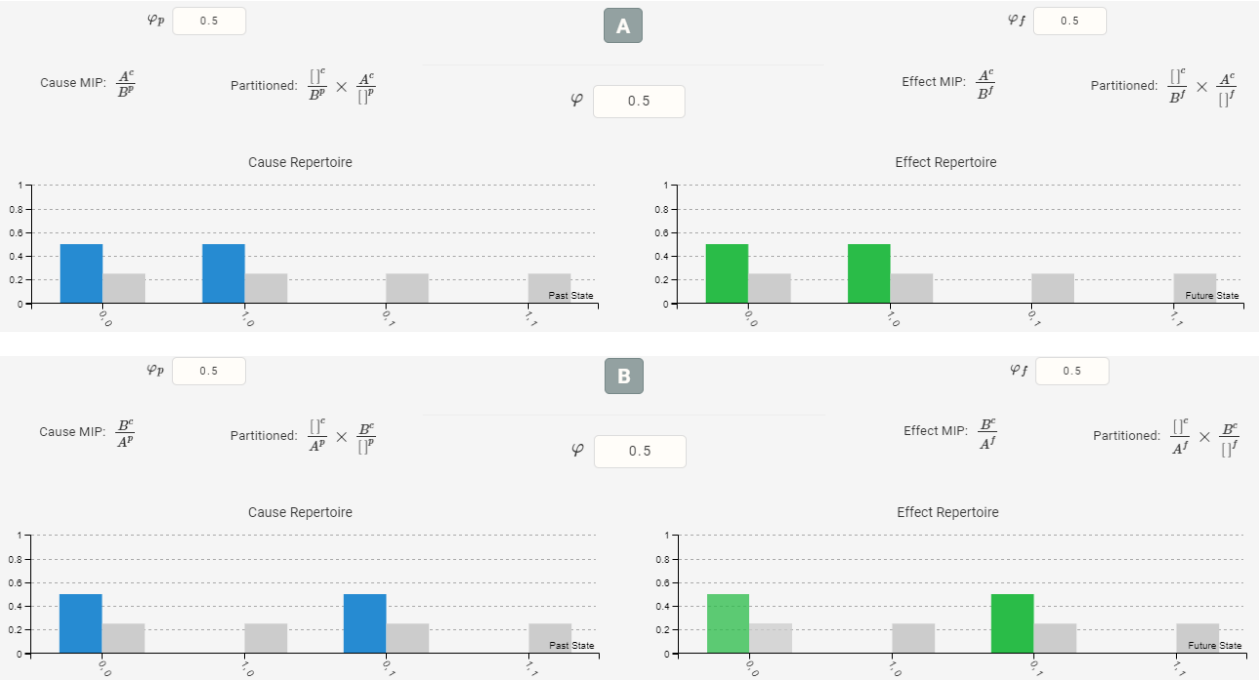


Σχήμα 24: Κύριο σύμπλεγμα συστήματος κόμβων OR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).



Σχήμα 25: Τιμή Φ και αριθμός *Concepts* συστήματος κόμβων OR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Από το αντικείμενο *SystemIrreducibilityAnalysis* προκύπτουν και οι ειδικότερες πληροφορίες για κάθε *Concept*, οι οποίες παρουσιάζονται στο **Σχήμα 26**.



Σχήμα 26: Το κάθε *Concept* συστήματος κόμβων OR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Για την καλύτερη εύρεση των αποτελεσμάτων, παρατίθεται πίνακας με συγκεντρωμένες τις βασικότερες μετρήσεις.

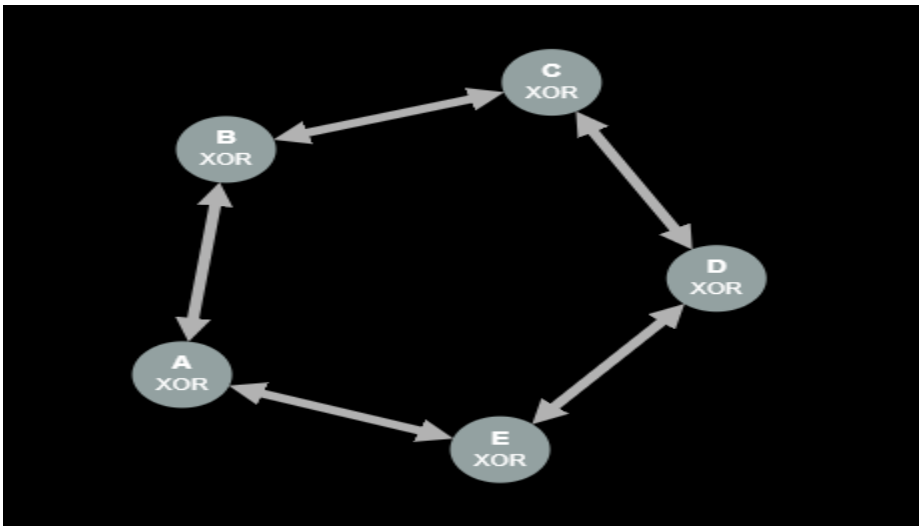
Πίνακας 2: Πίνακας μετρήσεων συστήματος 5 πυλών OR αμφίδρομης διάδοσης.

(A,B,C,D,E)	Main Complex	Φ Value	N.o.C	φ of Con.A	φ of Con.B	φ of Con.C	φ of Con.D	φ of Con.E
(0,0,0,0,0)	A,B	1	2	0.5	0.5	-	-	-
(0,0,0,0,1)	A,E	1	2	0.5	-	-	-	0.5
(0,0,0,1,0)	A,B	1	2	0.5	0.5	-	-	-
(0,0,0,1,1)	D,E	1	2	-	-	-	0.5	0.5
(0,0,1,0,0)	A,E	1	2	0.5	-	-	-	0.5

(0,0,1,0,1)	A,E	1	2	0.5	-	-	-	0.5
(0,0,1,1,0)	C,D	1	2	-	-	0.5	0.5	-
(0,0,1,1,1)	C,D,E	0.21528	3	-	-	0.25	0.16667	0.25
(0,1,0,0,0)	A,B	1	2	0.5	0.5	-	-	-
(0,1,0,0,1)	B,C	1	2	-	0.5	0.5	-	-
(0,1,0,1,0)	A,B	1	2	0.5	0.5	-	-	-
(0,1,0,1,1)	D,E	1	2	-	-	-	0.5	0.5
(0,1,1,0,0)	B,C	1	2	-	0.5	0.5	-	-
(0,1,1,0,1)	B,C	1	2	-	0.5	0.5	-	-
(0,1,1,1,0)	B,C,D	0.21528	3	-	0.25	0.16667	0.25	-
(0,1,1,1,1)	A,B,C,D,E	0.13889	5	0.25	0.16667	0.16667	0.16667	0.16667
(1,0,0,0,0)	A,B	1	2	0.5	0.5	-	-	-
(1,0,0,0,1)	A,E	1	2	0.5	-	-	-	0.5
(1,0,0,1,0)	A,B	1	2	0.5	0.5	-	-	-
(1,0,0,1,1)	A,D,E	0.21528	3	0.25	-	-	0.25	0.16667
(1,0,1,0,0)	A,E	1	2	0.5	-	-	-	0.5
(1,0,1,0,1)	A,E	1	2	0.5	-	-	-	0.5
(1,0,1,1,0)	C,D	1	2	-	-	0.5	0.5	-
(1,0,1,1,1)	A,B,C,D,E	0.13889	5	0.16667	0.25	0.16667	0.16667	0.16667
(1,1,0,0,0)	A,B	1	2	0.5	0.5	-	-	-
(1,1,0,0,1)	A,B,E	0.21528	3	0.16667	0.25	-	-	0.25
(1,1,0,1,0)	A,B	1	2	0.5	0.5	-	-	-
(1,1,0,1,1)	A,B,C,D,E	0.13889	5	0.16667	0.16667	0.25	0.16667	0.16667
(1,1,1,0,0)	A,B,C	0.21528	3	0.25	0.16667	0.25	-	-
(1,1,1,0,1)	A,B,C,D,E	0.13889	5	0.16667	0.16667	0.16667	0.25	0.16667
(1,1,1,1,0)	A,B,C,D,E	0.13889	5	0.16667	0.16667	0.16667	0.16667	0.25
(1,1,1,1,1)	A,B,C,D,E	0.13889	5	0.16667	0.16667	0.16667	0.16667	0.16667

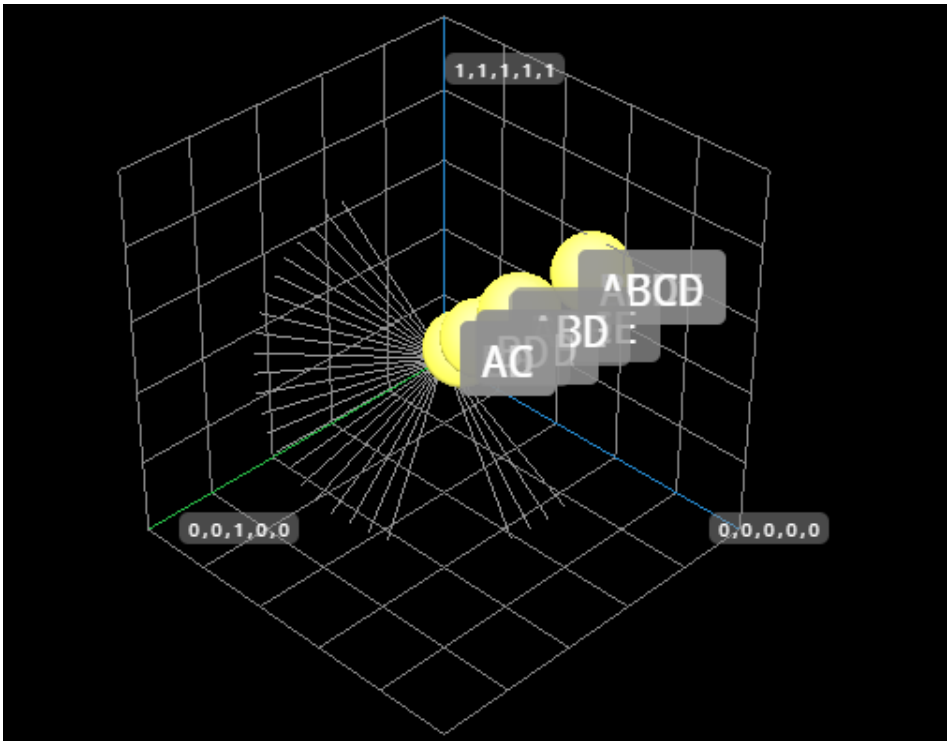
Παράδειγμα 3:

Για το τρίτο και τελευταίο παράδειγμα, χρησιμοποιείται ένα σύστημα 5 λογικών πυλών XOR (A, B, C, D, E) που συνδέονται κυκλικά με τις γειτονικές τους με κανάλι αμφίδρομης διάδοσης. Το σύστημα, λοιπόν, έχει τη μορφή του **Σχήματος 27**.



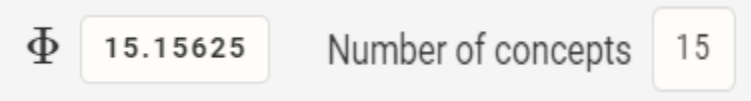
Σχήμα 27: Μορφή συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

Ακολουθείται η ίδια διαδικασία με τα παραπάνω παραδείγματα χρησιμοποιώντας αρχικά τις τιμές (A,B,C,D,E)=(0,0,0,0,0). Στα **Σχήματα 28-30** βρίσκονται τα πρώτα αποτελέσματα που προκύπτουν.



Σχήμα 28: Μορφή CES συστήματος κόμβων XOR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Σχήμα 29: Κύριο σύμπλεγμα συστήματος κόμβων OR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).



Σχήμα 30: Τιμή Φ και αριθμός *Concepts* συστήματος κόμβων XOR (A,B,C,D,E) με τωρινή κατάσταση (0,0,0,0,0).

Από το αντικείμενο *SystemIrreducibilityAnalysis* προκύπτουν και οι ειδικότερες πληροφορίες για κάθε *Concept*, οι οποίες παραλείπονται λόγω πλήθους στο συγκεκριμένο παράδειγμα. Παρατίθενται, ωστόσο κανονικά τα αποτελέσματα της προσομοίωσης για κάθε πιθανή τωρινή κατάσταση στους πίνακες που ακολουθούν.

Πίνακας 3: Πρώτος πίνακας μετρήσεων συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

(A,B,C,D,E)	**Main Complex**	**Φ Value**	**N.o.C**	**φ of Con.A**	**φ of Con.B**	**φ of Con.C**	**φ of Con.D**	**φ of Con.E**
(0,0,0,0,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,0,0,0,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,0,0,1,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,0,0,1,1)	A,B,C,D,E	15.15625	15	-	-	-	-	0
(0,0,1,0,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,0,1,0,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,0,1,1,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,0,1,1,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,1,0,0,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,1,0,0,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,1,0,1,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,1,0,1,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,1,1,0,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(0,1,1,0,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-

(0,1,1,1,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(0,1,1,1,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,0,0,0,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,0,0,0,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,0,0,1,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,0,0,1,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,0,1,0,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,0,1,0,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,0,1,1,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,0,1,1,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,1,0,0,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,1,0,0,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,1,0,1,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,1,0,1,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,1,1,0,0)	A,B,C,D	1.75	4	-	0.5	0.5	-	-
(1,1,1,0,1)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,1,1,1,0)	A,B,C,D,E	15.15625	15	-	-	-	-	-
(1,1,1,1,1)	A,B,C,D	1.75	4	-	0.5	0.5	-	-

Πίνακας 4: Δεύτερος πίνακας μετρήσεων συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

(A,B,C,D,E)	**φ of Con. AB**	**φ of Con. AC**	**φ of Con. AD**	**φ of Con. AE**	**φ of Con. BC**	**φ of Con. BD**	**φ of Con. BE**	**φ of Con. CD**
(0,0,0,0,0)	-	0.5	0.5	-	-	0.5	0.5	-
(0,0,0,0,1)	-	0.5	-	-	-	0.5	-	-
(0,0,0,1,0)	-	0.5	-	-	-	0.5	-	-
(0,0,0,1,1)	-	0.5	0.5	-	-	0.5	0.5	-
(0,0,1,0,0)	-	0.5	-	-	-	0.5	-	-

(0,0,1,0,1)	-	0.5	0.5	-	-	0.5	0.5	-
(0,0,1,1,0)	-	0.5	0.5	-	-	0.5	0.5	-
(0,0,1,1,1)	-	0.5	-	-	-	0.5	-	-
(0,1,0,0,0)	-	0.5	-	-	-	0.5	-	-
(0,1,0,0,1)	-	0.5	0.5	-	-	0.5	0.5	-
(0,1,0,1,0)	-	0.5	0.5	-	-	0.5	0.5	-
(0,1,0,1,1)	-	0.5	-	-	-	0.5	-	-
(0,1,1,0,0)	-	0.5	0.5	-	-	0.5	0.5	-
(0,1,1,0,1)	-	0.5	-	-	-	0.5	-	-
(0,1,1,1,0)	-	0.5	-	-	-	0.5	-	-
(0,1,1,1,1)	-	0.5	0.5	-	-	0.5	0.5	-
(1,0,0,0,0)	-	0.5	0.5	-	-	0.5	0.5	-
(1,0,0,0,1)	-	0.5	0.5	-	-	0.5	0.5	-
(1,0,0,1,0)	-	0.5	0.5	-	-	0.5	0.5	-
(1,0,0,1,1)	-	0.5	-	-	-	0.5	-	-
(1,0,1,0,0)	-	0.5	0.5	-	-	0.5	0.5	-
(1,0,1,0,1)	-	0.5	-	-	-	0.5	-	-
(1,0,1,1,0)	-	0.5	-	-	-	0.5	-	-
(1,0,1,1,1)	-	0.5	0.5	-	-	0.5	0.5	-
(1,1,0,0,0)	-	0.5	0.5	-	-	0.5	0.5	-
(1,1,0,0,1)	-	0.5	-	-	-	0.5	-	-
(1,1,0,1,0)	-	0.5	-	-	-	0.5	-	-
(1,1,0,1,1)	-	0.5	0.5	-	-	0.5	0.5	-
(1,1,1,0,0)	-	0.5	-	-	-	0.5	-	-
(1,1,1,0,1)	-	0.5	0.5	-	-	0.5	0.5	-
(1,1,1,1,0)	-	0.5	0.5	-	-	0.5	0.5	-
(1,1,1,1,1)	-	0.5	-	-	-	0.5	-	-

Πίνακας 5: Τρίτος πίνακας μετρήσεων συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

(A,B,C,D,E)	**φ of Con. CE**	**φ of Con. DE**	**φ of Con. ABC**	**φ of Con. ABD**	**φ of Con. ABE**	**φ of Con. ACD**	**φ of Con. ACE**	**φ of Con. ADE**
(0,0,0,0,0)	0.5	-	-	0.5	-	0.5	0.5	-
(0,0,0,0,1)	-	-	-	-	-	-	-	-
(0,0,0,1,0)	-	-	-	-	-	-	-	-
(0,0,0,1,1)	0.5	-	-	0.5	-	0.5	0.5	-
(0,0,1,0,0)	-	-	-	-	-	-	-	-
(0,0,1,0,1)	0.5	-	-	0.5	-	0.5	0.5	-
(0,0,1,1,0)	0.5	-	-	0.5	-	0.5	0.5	-
(0,0,1,1,1)	-	-	-	-	-	-	-	-
(0,1,0,0,0)	-	-	-	-	-	-	-	-
(0,1,0,0,1)	0.5	-	-	0.5	-	0.5	0.5	-
(0,1,0,1,0)	0.5	-	-	0.5	-	0.5	0.5	-
(0,1,0,1,1)	-	-	-	-	-	-	-	-
(0,1,1,0,0)	0.5	-	-	0.5	-	0.5	0.5	-
(0,1,1,0,1)	-	-	-	-	-	-	-	-
(0,1,1,1,0)	-	-	-	-	-	-	-	-
(0,1,1,1,1)	0.5	-	-	0.5	-	0.5	0.5	-
(1,0,0,0,0)	0.5	-	-	0.5	-	0.5	0.5	-
(1,0,0,0,1)	0.5	-	-	0.5	-	0.5	0.5	-
(1,0,0,1,0)	0.5	-	-	0.5	-	0.5	0.5	-
(1,0,0,1,1)	-	-	-	-	-	-	-	-
(1,0,1,0,0)	0.5	-	-	0.5	-	0.5	0.5	-
(1,0,1,0,1)	-	-	-	-	-	-	-	-
(1,0,1,1,0)	-	-	-	-	-	-	-	-
(1,0,1,1,1)	0.5	-	-	0.5	-	0.5	0.5	-
(1,1,0,0,0)	0.5	-	-	0.5	-	0.5	0.5	-
(1,1,0,0,1)	-	-	-	-	-	-	-	-

(1,1,0,1,0)	-	-	-	-	-	-	-	-
(1,1,0,1,1)	0.5	-	-	0.5	-	0.5	0.5	-
(1,1,1,0,0)	-	-	-	-	-	-	-	-
(1,1,1,0,1)	0.5	-	-	0.5	-	0.5	0.5	-
(1,1,1,1,0)	0.5	-	-	0.5	-	0.5	0.5	-
(1,1,1,1,1)	-	-	-	-	-	-	-	-

Πίνακας 6: Τέταρτος πίνακας μετρήσεων συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

(A,B,C,D,E)	**φ of Con. BCD**	**φ of Con. BCE**	**φ of Con. BDE**	**φ of Con. CDE**	**φ of Con. ABCD**	**φ of Con. ABCE**	**φ of Con. ABDE**	**φ of Con. ACDE**
(0,0,0,0,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,0,0,0,1)	-	-	-	-	-	-	-	-
(0,0,0,1,0)	-	-	-	-	-	-	-	-
(0,0,0,1,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,0,1,0,0)	-	-	-	-	-	-	-	-
(0,0,1,0,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,0,1,1,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,0,1,1,1)	-	-	-	-	-	-	-	-
(0,1,0,0,0)	-	-	-	-	-	-	-	-
(0,1,0,0,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,1,0,1,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,1,0,1,1)	-	-	-	-	-	-	-	-
(0,1,1,0,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(0,1,1,0,1)	-	-	-	-	-	-	-	-
(0,1,1,1,0)	-	-	-	-	-	-	-	-
(0,1,1,1,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,0,0,0,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5

(1,0,0,0,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,0,0,1,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,0,0,1,1)	-	-	-	-	-	-	-	-
(1,0,1,0,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,0,1,0,1)	-	-	-	-	-	-	-	-
(1,0,1,1,0)	-	-	-	-	-	-	-	-
(1,0,1,1,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,1,0,0,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,1,0,0,1)	-	-	-	-	-	-	-	-
(1,1,0,1,0)	-	-	-	-	-	-	-	-
(1,1,0,1,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,1,1,0,0)	-	-	-	-	-	-	-	-
(1,1,1,0,1)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,1,1,1,0)	-	0.5	0.5	-	0.5	0.5	0.5	0.5
(1,1,1,1,1)	-	-	-	-	-	-	-	-

Πίνακας 7: Πέμπτος πίνακας μετρήσεων συστήματος 5 πυλών XOR αμφίδρομης διάδοσης.

(A,B,C,D,E)	**φ of Con. BCDE**	**φ of Con. ABCDE**
(0,0,0,0,0)	0.5	-
(0,0,0,0,1)	-	-
(0,0,0,1,0)	-	-
(0,0,0,1,1)	0.5	-
(0,0,1,0,0)	-	-
(0,0,1,0,1)	0.5	-
(0,0,1,1,0)	0.5	-
(0,0,1,1,1)	-	-
(0,1,0,0,0)	-	-

(0,1,0,0,1)	0.5	-
(0,1,0,1,0)	0.5	-
(0,1,0,1,1)	-	-
(0,1,1,0,0)	0.5	-
(0,1,1,0,1)	-	-
(0,1,1,1,0)	-	-
(0,1,1,1,1)	0.5	-
(1,0,0,0,0)	0.5	-
(1,0,0,0,1)	0.5	-
(1,0,0,1,0)	0.5	-
(1,0,0,1,1)	-	-
(1,0,1,0,0)	0.5	-
(1,0,1,0,1)	-	-
(1,0,1,1,0)	-	-
(1,0,1,1,1)	0.5	-
(1,1,0,0,0)	0.5	-
(1,1,0,0,1)	-	-
(1,1,0,1,0)	-	-
(1,1,0,1,1)	0.5	-
(1,1,1,0,0)	-	-
(1,1,1,0,1)	0.5	-
(1,1,1,1,0)	0.5	-
(1,1,1,1,1)	-	-

7.4 Συμπεράσματα

Τα παραδείγματα, όπως έχει προαναφερθεί, δεν είναι ικανά για να κατανοήσει κάποιος το πόσο «συνείδηση» διαθέτει ένα σύστημα. Γίνεται, ωστόσο, εύκολα αντιληπτό ότι ένα σύστημα με πύλες AND, παρόμοιο με αυτό των παραδειγμάτων, δίνει λιγότερη πληροφορία (maxΦ = 0.1875), σε σχέση με το αντίστοιχο πυλών OR (maxΦ = 1) και πυλών XOR (maxΦ = 15.15625). Μάλιστα δεν

είναι απίθανο ένα σύστημα πυλών AND να μη μας δίνει καθόλου πληροφορία. Αξίζει να σημειωθεί ότι η απλότητα των συστημάτων (κυρίως αυτών με πύλες AND και OR) προκύπτει και από το ότι δεν είναι λίγες οι φορές που το κύριο σύμπλεγμα του συστήματος δεν απαιτεί τη συμμετοχή όλων των κόμβων, αφού η πληροφορία μπορεί να αποδοθεί και από λιγότερους.

Ένα επιπλέον χαρακτηριστικό, για κάθε σύστημα ξεχωριστά, είναι πως όσο λιγότερα *Concepts* συμμετέχουν στο κύριο σύμπλεγμα του συστήματος τόσο μικρότερο είναι και το Φ που αυτό αποδίδει. Από την άλλη, το Φ που αποδίδει το κάθε *Concept* σε ένα σύστημα με κύριο σύμπλεγμα λίγων στοιχείων είναι μεγαλύτερο ή ίσο από αυτό που αποδίδεται από το κάθε *Concept* σε σύστημα μεγαλύτερου κύριου συμπλέγματος. Το πρώτο μέρος αυτής της παρατήρησης είναι απόλυτα φυσιολογικό, με βάση τις θέσεις της IIT, αφού εξ' ορισμού το σύμπλεγμα προκύπτει από την ομάδα στοιχείων που παράγουν ενσωματωμένη πληροφορία, η οποία δεν περιέχεται σε κάποια μεγαλύτερη ομάδα με υψηλότερο δείκτη Φ. Όταν, λοιπόν, το σύστημα είναι συγκεκριμένο, είναι λογικό να προκύπτει περισσότερη πληροφορία στις περιπτώσεις που αξιοποιούνται οι αιτιατές αλληλεπιδράσεις μεταξύ περισσότερων στοιχείων. Ωστόσο, το δεύτερο μέρος αυτής της παρατήρησης είναι συγκυριακό και αποδίδεται κυρίως στην ομοιότητα των στοιχείων που αποτελούν το κάθε σύστημα, καθώς και στην απλότητα των συστημάτων.

Ενδιαφέρον επίσης παρουσιάζει η τεράστια αβεβαιότητα που υπάρχει για την προηγούμενη χρονικά κατάσταση των *Concepts* στο σύστημα πυλών AND, σε αντίθεση βέβαια με την επόμενη χρονικά κατάσταση. Αυτό το φαινόμενο είναι που περιορίζει σε μεγάλο βαθμό και την τιμή του Φ που προκύπτει από το συγκεκριμένο σύστημα, αλλά είναι αναγκαίο να αναφερθεί, ώστε να γίνει καλύτερα αντιληπτό πόσο δύσκολη είναι γενικότερα αυτή η προσπάθεια μέτρησης της συνείδησης (είτε στα πλαίσια της IIT είτε όχι), αφού η αναφορά γίνεται για ένα σύστημα πέντε στοιχείων δυαδικής εξόδου.

Τέλος, για να μην υπάρχει κάποια παρεξήγηση, αξίζει να αναφερθεί ότι το CES δεν έχει κάποια σχέση με τον χώρο των Qualia, παρά το ότι και αυτός είναι πολυδιάστατος. Ουσιαστικά αποτελεί την οπτικοποίηση των πληροφοριών που παίρνουμε από τα *Concepts*, στην παρούσα όμως χρονική στιγμή είναι κατανοητό πως δεν είναι και ιδιαίτερα εύχρηστος, κυρίως λόγω της δυσκολίας αποτύπωσης σε μία οθόνη περισσότερων από τρεις διαστάσεις.

7.5 Μελλοντικές Κατευθύνσεις

Είναι, λοιπόν, φανερό πως γίνεται αναφορά σε μια εφαρμογή που είναι ακόμα υπό ανάπτυξη, βρισκόμενη στο δρόμο που επιθυμεί η IIT. Αρκετές επιπρόσθετες δυνατότητες βρίσκονται υπό ανάπτυξη. Αυτές περιλαμβάνουν τη δυνατότητα υπολογισμού του Φ σε περισσότερες χωροχρονικές βαθμίδες, όπως και τον υπολογισμό της «πραγματικής αιτιότητας». Φυσικά το λογισμικό ενημερώνεται κατάλληλα, ώστε να ακολουθεί και οποιαδήποτε πιθανή εξέλιξη ή αλλαγή πάνω στους ισχυρισμούς της IIT.

Από εκεί και πέρα, η παρούσα διπλωματική εργασία θα μπορούσε να αποτελέσει τη βάση για περαιτέρω έρευνα στο μέλλον. Η αξιοποίησή της μπορεί να γίνει με διάφορους τρόπους. Πρώτον η πρόοδος της βιοϊατρικής έρευνας και τεχνολογίας στον τομέα της συνείδησης συνεχώς κάνει

βήματα μπροστά. Οι θεωρίες για τη μέτρηση της συνείδησης εξελίσσονται και γίνονται, αργά αλλά σταθερά, όλο και περισσότερο αποδεκτές στο αντίστοιχο επιστημονικό κοινό. Στο παρόν κείμενο γίνεται μια πρώτη ανάλυση των επικρατέστερων μέχρι στιγμής θεωριών, αυτό όμως είναι απλά ένα βήμα για κάποια ανάλυση σε μεγαλύτερο βάθος, ειδικά σε περίπτωση που κάποια από αυτές αρχίσει να ξεχωρίζει. Εν συνεχεία, εξίσου ενδιαφέρον παρουσιάζει και ο τομέας της τεχνητής συνείδησης και της προσπάθειας που γίνεται να αποκτήσουν συνείδηση οι μηχανές. Εδώ υπάρχει μια πρώτη ανάλυση των βασικών προβλημάτων που αντιμετωπίζονται, αλλά και των προσπαθειών που εφαρμόζονται ώστε αυτά να ξεπεραστούν. Προκύπτει, έτσι, η επιλογή της εμβάθυνσης της κατανόησης των προβλημάτων αυτών, όπως και της έρευνας στο κομμάτι της φαινόμενης συνείδησης, πάνω στο οποίο πλέον επικεντρώνονται οι ερευνητές μηχανικής συνείδησης. Τέλος, όπως αναφέρθηκε, η βελτίωση της εφαρμογής PyPhi στο άμεσο μέλλον είναι βέβαιη και μπορεί να αποτελέσει ένα χρησιμότατο εργαλείο για οποιονδήποτε έχει τη θέληση να μελετήσει καλύτερα τη Θεωρία Ενσωματωμένων Πληροφοριών.

Αυτό που μπορεί να ειπωθεί με σιγουριά είναι ότι έγινε αναφορά σε ένα ζήτημα, η έρευνα γύρω από το οποίο βρίσκεται ακόμα στις απαρχές της, αφού ελάχιστες γενιές ερευνητών έχουν ασχοληθεί με την πρακτική πλευρά του θέματος. Γίνεται προσπάθεια μέτρησης της συνείδησης και απόδοσης της συνείδησης σε κάποια μηχανή, χωρίς να υπάρχει καν ένας σαφής καθολικά αποδεκτός ορισμός αυτής. Το γεγονός αυτό κάνει αξιοσημείωτα δύσκολο, λεπτό σε χειρισμό και ενδιαφέρον το εν λόγω θέμα. Ταυτόχρονα, όμως, προκύπτει ότι τόσο τα δεδομένα για την παρουσίαση των θεωριών όσο και αυτά των παραδειγμάτων της εφαρμογής δεν είναι απόλυτα, παρά ισχύουν τη δεδομένη χρονική στιγμή και αποτελούν μία βάση για περαιτέρω έρευνα στο μέλλον.

Βιβλιογραφία

[1] L. BonJour, “What is it like to be a human (instead of a bat)?,” *Am. Philos. Q.*, vol. 50, no. 4, pp. 373–385, 2013.

[2] Κ. Παρασκευή, “Σύγχρονες προσεγγίσεις στην Συνείδηση Από τη φιλοσοφία του νου στα νευρωνικά μοντέλα της συνείδησης Αθήνα Εισηγητές Μουτούσης Κωνσταντίνος Ειρήνη Σκαλιώρα,” 2011.

[3] I. Aleksander, “Partners of humans: A realistic assessment of the role of robots in the foreseeable future,” *J. Inf. Technol.*, vol. 32, no. 1, pp. 1–9, 2017.

[4] M. Oizumi, L. Albantakis, and G. Tononi, “From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0,” *PLoS Comput. Biol.*, vol. 10, no. 5, 2014.

[5] W. T. Dixon, “Consciousness as Integrated Information: a Provisional Manifesto,” *Biol. Bull.*, vol. 215, no. 3, pp. 216–242, 2008.

[6] F. Fallon, “Integrated information theory,” *Routledge Handb. Conscious.*, vol. 10, no. 2015, pp. 137–148, 2018.

[7] M. A. Cerullo, “The Problem with Phi: A Critique of Integrated Information Theory,” *PLOS Comput. Biol.*, vol. 11, no. 9, p. e1004286, 2015.

[8] A. K. Seth, A. B. Barrett, and L. Barnett, “Causal density and integrated information as measures of conscious level,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 369, no. 1952, pp. 3748–3767, 2011.

[9] A. K. Seth, “Consciousness : Theories and Models,” pp. 131–136, 2009.

[10] K. Friston, “The free-energy principle: a rough guide to the brain?,” *Trends Cogn. Sci.*, vol. 13, no. 7, pp. 293–301, 2009.

[11] C. L. Buckley, C. S. Kim, S. McGregor, and A. K. Seth, “The free energy principle for action and perception: A mathematical review,” *J. Math. Psychol.*, vol. 81, pp. 55–79, 2017.

[12] K. Friston, “The free-energy principle: a unified brain theory?,” *Nat. Rev. Neurosci.*, vol. 11, no. 2, pp. 127–38, 2010.

[13] B. J. Baars, “IN THE THEATRE OF CONSCIOUSNESS Global Workspace

Theory, A Rigorous Scientific Theory of Consciousness."

[14] S. Dehaene, J.-P. Changeux, L. Naccache, J. Sackur, and C. Sergent, "Conscious, preconscious, and subliminal processing: a testable taxonomy," *Trends Cogn. Sci.*, vol. 10, no. 5, pp. 204–211, May 2006.

[15] V. A. F. Lamme, "Towards a true neural stance on consciousness."

[16] S. Zeki, "The disunity of consciousness," *Trends Cogn. Sci.*, vol. 7, no. 5, pp. 214–218, May 2003.

[17] P. Kavitha, B. Krishna Moorthy, P. S. Sudharshan, and T. Aarthi, "Mapping artificial intelligence and education," *Proc. 2018 Int. Conf. Commun. Comput. Internet Things, IC3IoT 2018*, vol. 6, pp. 165–168, 2019.

[18] J. A. Reggia, "Conscious machines: The AI perspective," *Nat. Humans Mach. — A Multidiscip. Discourse Pap. from 2014 AAAI Fall Symp. Conscious*, pp. 34–37, 2014.

[19] S. Dehaene, H. Lau, and S. Kouider, "What is consciousness, and could machines have it?," *Science (80-.).*, vol. 358, no. 6362, pp. 486–492, 2017.

[20] W. G. P. Mayner, W. Marshall, L. Albantakis, G. Findlay, R. Marchman, and G. Tononi, "PyPhi: A toolbox for integrated information theory," *PLOS Comput. Biol.*, vol. 14, no. 7, p. e1006343, Jul. 2018.