

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΦΙΛΟΣΟΦΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΣΠΟΥΔΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΤΟΜΕΑ ΦΙΛΟΣΟΦΙΑΣ**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΕΙΝΑΙ ΔΥΝΑΤΗ ΜΙΑ ΛΕΙΤΟΥΡΓΙΣΤΙΚΗ ΕΞΗΓΗΣΗ
ΤΟΥ ΝΟΥ ΚΑΙ ΤΗΣ ΖΩΗΣ;
ΜΙΑ ΚΡΙΤΙΚΗ ΣΤΑ ΕΠΙΣΤΗΜΟΛΟΓΙΚΑ ΠΡΟΓΡΑΜΜΑΤΑ ΤΗΣ
ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΗΣ ‘ΤΕΧΝΗΤΗΣ ΖΩΗΣ’.**

ΚΑΣΣΑΝΔΡΑ ΜΠΕΪΚΑΚΗ

**ΕΠΙΒΛΕΠΟΥΣΑ:
ΣΤΑΥΡΟΥΛΑ ΤΣΙΝΟΡΕΜΑ
ΑΝΑΠΛΗΡΩΤΡΙΑ ΚΑΘΗΓΗΤΡΙΑ
ΕΠΟΠΤΕΣ: ΒΕΝΙΕΡΗ ΜΑΡΙΑ, ΡΟΥΣΣΟΠΟΥΛΟΣ ΓΙΩΡΓΟΣ**

ΡΕΘΥΜΝΟ 2002

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΙΣΑΓΩΓΗ	σελ. 1
ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ: Η ΓΝΩΣΙΑΚΗ ΕΠΙΣΤΗΜΗ	
1. Εισαγωγή	σελ. 10
2. Δύο μεταφυσικές προσεγγίσεις του νου:	
α. Υλισμός	σελ. 10
β. Δυισμός	σελ. 13
3. Μέθοδοι εξήγησης του νου:	
α. Φυσικαλισμός	σελ. 14
β. Λειτουργισμός	σελ. 16
γ. Η μη-αναγωγιστική, προθεσιακή προσέγγιση	σελ. 18
4. Συγκεκριμένες επιστημονικές προκείμενες της Γνωσιακής Επιστήμης: Γνωσιακή Ψυχολογία και Τεχνητή νοημοσύνη	σελ. 19
5. Τα εννοιολογικά θεμέλια του προγράμματος της Γνωσιακής Επιστήμης (Jerry Fodor)	σελ. 25
6. Φιλοσοφικές παραδοχές και μεθοδολογία της Γνωσιακής Επιστήμης	σελ. 30
ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ: ΑΝΑΣΥΓΚΡΟΤΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΤΗΣ ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ	
1. Το ‘κινέζικο δωμάτιο’	
2. α. Εισαγωγή	σελ. 36
β. Το πείραμα	σελ. 37
γ. Το επιχείρημα	σελ. 39
2. Κριτική	σελ. 45
ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ: ‘ΤΕΧΝΗΤΗ ΖΩΗ’	
1. Το εννοιολογικό πλαίσιο της ‘Τεχνητής Ζωής’	

και η μεθοδολογία της	σελ. 59
2. Κριτική στο πρόγραμμα της ‘Τεχνητής Ζωής’	σελ. 66
ΣΥΜΠΕΡΑΣΜΑΤΑ/ ΕΠΙΛΕΓΟΜΕΝΑ	σελ. 77
ΒΙΒΛΙΟΓΡΑΦΙΑ	σελ. 81

ΕΙΣΑΓΩΓΗ

Ποιες μπορεί να είναι οι δυνατότητες ενός ηλεκτρονικού υπολογιστή; Υπάρχουν τόσο εξελιγμένα υπολογιστικά μοντέλα και προγράμματα, ώστε να δικαιούμαστε να πούμε ότι ένας υπολογιστής σκέφτεται, αποφασίζει, αισθάνεται πόνο, είναι έξυπνος ή ανόητος; Μπορεί ο κατάλληλα προγραμματισμένος υπολογιστής να επιδείξει νοήμονα συμπεριφορά ή ακόμη και να διαθέτει νόηση; Ένα ρομπότ μπορεί να χαρεί, να αναπαραχθεί, να αναπτυχθεί; Με άλλα λόγια, μπορούμε να υποστηρίξουμε ότι ένα μηχανικό κατασκευάσμα, βασισμένο στην λειτουργία των ηλεκτρονικών υπολογιστών, διαθέτει νόηση ή είναι όντως ζωντανό; Τέτοιου είδους ερωτήματα είναι τόσο επιστημονικά όσο και φιλοσοφικά και οι απαντήσεις τους μπορούν να έχουν ευρύτερες συνέπειες ως προς την κοινωνική ηθική, στην αυτοκατανόησή μας, τον καταλογισμό ευθυνών κτλ.

Ειδικά στις μέρες μας, με την αλματώδη πρόοδο της τεχνολογίας, τα ερωτήματα που αναφέρθηκαν αρχικά φαίνονται να είναι περισσότερο καίρια και οι απαντήσεις σ' αυτά περισσότερο σημαντικές από ποτέ. Αυτό συμβαίνει, διότι, αν αληθεύει το ότι οι υπολογιστές έχουν νόηση, ή ότι τα ρομπότ είναι τόσο ζωντανά όσο και ένα φυτό ή ζώο, τότε το ήδη διαμορφωμένο κοσμοείδωλό μας για τον νου, την ζωή, και ιδιαιτέρως η ιδέα μας για την ανθρώπινη φύση θα αλλάξει ριζικά και θα πρέπει να ανασυγκροτηθεί εκ νέου.

Όλη αυτή η συζήτηση έχει ως αφετηρία το εγχείρημα που καταβάλλει ένας κλάδος της σύγχρονης επιστήμης, η Γνωσιακή Επιστήμη (Cognitive Science), προκειμένου να προσομοιώσει τα ανθρώπινα νοητικά φαινόμενα στο μοντέλο των υπολογιστών. Η Γνωσιακή Επιστήμη αποτελεί ένα επιστημονικό πρόγραμμα, το οποίο, παράλληλα με τις εξελίξεις και την αρωγή και άλλων επιστημών, πρεσβεύει ότι είναι δυνατόν να αποκρυπτογραφήσει και να εξηγήσει επιστημονικά την φύση των νοητικών φαινομένων, στη βάση του

μοντέλου της λειτουργίας των ηλεκτρονικών υπολογιστών. Παράλληλα με την Γνωσιακή Επιστήμη, και πολύ προσφάτως, (τις τελευταίες δύο δεκαετίες) έχει εμφανιστεί και ένα νέο επιστημονικό πεδίο, το οποίο τελεί υπό διαμόρφωση, και του οποίου το πρόγραμμα στοχεύει στη ‘δημιουργία’ έμβιων οργανισμών, μέσω της προσομοίωσής τους ως λειτουργίες του ηλεκτρονικού υπολογιστή. Πρόκειται για την ‘Τεχνητή Ζωή’ (Artificial Life), η οποία πρεσβεύει πως είναι σε θέση να κατανοήσει το φαινόμενο της ζωής γενικότερα, μέσω της υπολογιστικής προσομοίωσης, και στη συνέχεια να κατασκευάσει ζωντανούς οργανισμούς.

Τα παραπάνω προγράμματα έχουν τις ρίζες τους στην περίοδο της ανακάλυψης των ηλεκτρονικών υπολογιστών (τη δεκαετία του ’50) και πήραν ώθηση από τις απόψεις του Alan Turing¹ και του John von Neumann². Ο Alan Turing (1913-1954), Βρετανός μαθηματικός και θεωρητικός της λογικής, το 1950 υποστήριξε ότι μια μηχανή μπορεί να σκέπτεται, αν είναι σε θέση να ‘περάσει’ τη δοκιμασία του. Η δοκιμασία αυτή βασίζεται στην κεντρική ιδέα του ‘παιχνιδιού της μίμησης’. Την φράση αυτή την εισηγήθηκε ο ίδιος ο Turing. Το παιχνίδι αυτό παίζεται με τρεις παίκτες: έναν άνδρα (Α), μια γυναίκα (Β) και ένα τρίτο πρόσωπο (Γ), που θα παίζει το ρόλο του ανακριτή, ανεξαρτήτως φύλου, το οποίο βρίσκεται απομονωμένο από τα πρόσωπα Α και Β και τους υποβάλλει ερωτήσεις. Ο σκοπός του παιχνιδιού είναι ο ανακριτής να βρει ποιο από τα δύο πρόσωπα είναι ο άνδρας και ποιο η γυναίκα. Από τα πρόσωπα αυτά το ένα (Α) ως σκοπό έχει να εξαπατήσει τον ανακριτή, ενώ το άλλο (Β) τον βοηθάει. Αυτό που απασχολεί τον Turing είναι τι θα συνέβαινε αν τη θέση του Α την έπαιρνε μια μηχανή.

Με άλλα λόγια, το ζήτημα που έθεσε ο Turing είναι το κατά πόσο μπορεί μία μηχανή να σκέφτεται και να επικοινωνεί με άλλα

¹ Turing, A.M., [1950], “Computing Machinery and Intelligence.”, *Mind* 59, 433-460

σκεπτόμενα όντα και, αν ναι, τι είδους μηχανή θα ήταν αυτή. Καταλήγει λοιπόν ότι η μηχανή αυτή θα μπορούσε να είναι ένας ηλεκτρονικός ψηφιακός υπολογιστής. Σκοπός του ‘παιχνιδιού’ αυτού τώρα θα είναι να καταλάβει ο πρώτος παίκτης ποιος από τους δύο είναι ο άνθρωπος και ποιος ο υπολογιστής. Το ‘παιχνίδι της μίμησης’ τώρα πλέον διευρύνεται και ο υπολογιστής προσπαθεί να αποδείξει ότι είναι άνθρωπος, ενώ ο άνθρωπος ότι δεν είναι ο υπολογιστής. Ο ίδιος ο Turing πίστευε πως ένας ψηφιακός υπολογιστής, προγραμματισμένος με κανόνες και γεγονότα, θα μπορούσε να επιδείξει νοήμονα συμπεριφορά.³ Κι όχι μόνο αυτό, αλλά θεώρησε ότι, εάν μια μηχανή μπορέσει να ‘περάσει’ με επιτυχία το τεστ του, θα διαθέτει νόηση. Ο Turing θεώρησε ότι μια τουλάχιστον κατηγορία μηχανών είναι σε θέση όχι απλά να προσομοιώσει την ανθρώπινη σκέψη, αλλά ακόμη και να την υπερβεί. Και ακόμη περισσότερο, οι μηχανές αυτές δικαιούμαστε να υποστηρίξουμε ότι διαθέτουν νόηση. Αυτό μια υπολογιστική μηχανή μπορεί να το κάνει, εάν οι συνάψεις των καλωδίων της αναπαριστούν αυτές του ανθρώπινου εγκεφάλου και εάν χειρίζεται τα σύμβολα, με τα οποία τροφοδοτείται, με τον ίδιο τυπικό τρόπο που τα χειρίζονται οι άνθρωποι.

Λίγα χρόνια αργότερα, το 1956, εμφανίστηκε ο John von Neumann (1903-1957), σπουδαίος Ούγγρος μαθηματικός, ο οποίος έδωσε νέα πνοή στο αίτημα του Turing σχετικά με την προσομοίωση της ανθρώπινης νόησης σε ηλεκτρονικό υπολογιστή και την εξήγησή της με λειτουργικούς όρους. Η προσφορά του von Neumann ήταν ότι κατάφερε να λειτουργεί η μνήμη του υπολογιστή με δύο διαφορετικούς τρόπους: Αφενός η κύρια μνήμη της ‘μηχανής von Neumann’ περιέχει τα σύμβολα πάνω στα οποία θα γίνονται οι πράξεις, αφετέρου η ίδια μνήμη είναι σε θέση να περιέχει και τα προγράμματα που εκτελεί η μηχανή. Κατάφερε, με άλλα λόγια, να

² von Neumann, J., [1956], “The General and Logical Theory of Automata”, ανατύπωση στο *The World of Mathematics (Νέα Υόρκη: Simon and Schuster)*, σ. 2077

τοποθετηθούν μαζί στην κύρια μνήμη τόσο ο κώδικας του προγράμματος όσο και τα δεδομένα του. Η μεγαλύτερη επιρροή, ωστόσο, που άσκησε ο von Neumann, στο ζήτημα που μας απασχολεί, είναι ότι υποστήριξε πως ο εγκέφαλος λειτουργεί αποκλειστικά ως αναλογικός υπολογιστής.⁴

Από μεθοδολογική άποψη, τόσο η Γνωσιακή Επιστήμη όσο και η Τεχνητή Ζωή βασίζονται στο φορμαλιστικό μοντέλο και στην λειτουργιστική μέθοδο, σε συνδυασμό με την σύγχρονη τεχνολογία που χρησιμοποιείται στους ηλεκτρονικούς υπολογιστές. Η μέθοδος αυτή εφαρμόζεται από την πρώτη με σκοπό να εξηγηθούν τα ανώτερα νοητικά φαινόμενα του ανθρώπου, μέσω της προσομοίωσής τους σε έναν ηλεκτρονικό υπολογιστή. Αυτή η προσπάθεια γίνεται μέσα από μια διαδικασία που θεωρεί ως δεδομένο ότι η νόηση δεν είναι τίποτε άλλο από υπολογισμός, ο οποίος επιτυγχάνεται εξαιτίας των τυπικών/ συντακτικών σχέσεων που υπάρχουν στο υλισμικό του εγκεφάλου – το οποίο όμως θα μπορούσε να το αντικαταστήσει οποιοδήποτε άλλο υλισμικό υπόστρωμα.

Παράλληλα, η ‘Τεχνητή Ζωή’, μια σύγχρονη, πρωτοποριακή τάση, υπόσχεται όχι μόνο πως θα εξηγήσει τις ανώτερες νοητικές λειτουργίες, αλλά και όλες τις υπόλοιπες ζωικές και βιολογικές λειτουργίες. Οι παραπάνω αναδυόμενες επιστήμες/ προγράμματα, κατά πρώτο λόγο, ‘υπόσχονται’ πως θα κατορθώσουν τελικά να αποκωδικοποιήσουν τις λειτουργίες του νου και να δώσουν επιστημονικές εξηγήσεις για τη φύση των νοητικών φαινομένων. Κατά δεύτερο λόγο, πιστεύουν ότι οι ηλεκτρονικοί υπολογιστές μπορούν να προσομοιώσουν όχι μόνο τη νόηση, αλλά και το φαινόμενο της ζωής.

³ Dreyfus, H. L., [1998], σ.1

⁴ Την ίδια περίοδο, ανάμεσα σε άλλους, οι Newell και Simon εκδίδουν μελέτες προσομοίωσης των νοητικών διεργασιών με βάση το μοντέλο των ηλεκτρονικών υπολογιστών: Newell, A., και Simon, H.A., [1961], *Computer Simulation of Human Thinking*, The RAND Corporation, P-2276 (20 Απριλίου)

Η ζωή, σε αυτήν την περίπτωση βεβαίως, και υπό αυτήν την προοπτική, κατανοείται, όπως και η νόηση, ως μιας μορφής λειτουργία ή ως ένα σύνολο λειτουργιών.

Οι όροι συγκρότησης αυτών των δύο προγραμμάτων αποτελούν το αντικείμενο της πραγμάτευσης που ακολουθεί. Τι σημαίνει όμως ότι τα παραπάνω εγχειρήματα επιδιώκουν να εξηγήσουν *επιστημονικά* τη νόηση;

Το ανώτατο ορθολογικό επίτευγμα της κοινωνίας μας και του πολιτισμού θεωρείται πως είναι η επιστημονική γνώση. Καθεμιά από τις επιμέρους επιστήμες (φυσική, ψυχολογία, βιολογία, κ.ά.) έχει ως αντικείμενό της μια ορισμένη περιοχή της εμπειρικής πραγματικότητας, την οποία διερευνά με μια ορισμένη μέθοδο. Τα βασικά γνωρίσματα των επιμέρους επιστημών μπορούν να συνοψισθούν ως εξής: Αρχικά, ως αντικείμενα των επιστημών αυτών θεωρούνται και λαμβάνονται ορισμένες κάθε φορά επιμέρους περιοχές του κόσμου της εμπειρίας. Ως χώρος της εμπειρίας μπορεί να θεωρηθεί οποιαδήποτε περιοχή της φύσης, δηλαδή τα επιμέρους φυσικά φαινόμενα, τα καλλιτεχνικά δημιουργήματα, οι γλώσσες, τα ιστορικά γεγονότα κ.ά. Το δεύτερο χαρακτηριστικό των επιστημών είναι ότι εξηγούν τα εμπειρικά δεδομένα, βασισμένες σε άλλα εμπειρικά δεδομένα και πάντοτε στη βάση αιτιακών σχέσεων μεταξύ των δεδομένων αυτών. Τέλος, καθεμιά από αυτές τις επιστήμες έχει την δική της μέθοδο, και εξετάζει το αντικείμενό της από μια ορισμένη σκοπιά, χωρίς να λαμβάνει υπόψη της άλλες ενδεχομένως πλευρές του. Στο πλαίσιο αυτό κινείται και η Γνωσιακή Επιστήμη (και θέλει να κινηθεί ως επιστήμη και το πρόγραμμα της 'Τεχνητής Ζωής').

Το ερώτημα που θα μπορούσε να τεθεί από ορισμένους είναι ποια είναι η θέση της φιλοσοφίας και ποια σκοπιμότητα εξυπηρετεί η ενασχόλησή της με τα παραπάνω θέματα. Ο ρόλος της φιλοσοφίας, σε σχέση με τις επιστήμες, είναι πολυδιάστατος. Κατά πρώτο λόγο, η φιλοσοφία έρχεται στο προσκήνιο ως κριτική της μεταφυσικής των

επιστημών, δηλαδή του δογματισμού, στον οποίο μπορεί να οδηγήσουν τη σκέψη. Τα επιστημονικά πορίσματα ορισμένες φορές απολυτοποιούνται και δημιουργούν τη βεβαιότητα και την αισιοδοξία ότι μπορούν να εξηγήσουν τα πάντα, όλες τις πτυχές της ζωής με συγκεκριμένες επιστημονικές μεθόδους. Αυτή η απολυτοποίηση είναι που δίνει αισιοδοξία σχετικά με το ότι η επιστήμη είναι σε θέση να δώσει απαντήσεις για όλους τους τομείς του επιστητού. Ακριβώς για αυτό ο ρόλος της φιλοσοφίας επί της σκέψης/ επιστήμης είναι διαφωτιστικός: Η φιλοσοφία θέτει όρια στην απολυτοκρατία που μπορεί να απορρέει από την απόλυτη πίστη στην επιστήμη, δηλαδή θέτει όρια στον επιστημονισμό. Κατά δεύτερο λόγο, η φιλοσοφία εμφανίζεται στο προσκήνιο και ως διαφωτισμός της σκέψης, άρα ως άνοιγμα του δρόμου για την επιστήμη (κριτική της μεταφυσικής). Επιπλέον, ο φιλοσοφικός στοχασμός συνθέτει το νέο κάθε φορά κοσμοείδωλο, όπως αυτό διαμορφώνεται από τα πορίσματα που προσφέρουν οι επιμέρους επιστήμες, και καθορίζει κατ' αυτόν τον τρόπο τη θέση του ανθρώπου εντός αυτού. Η φιλοσοφική σκέψη, οδηγεί σε μια σύνθεση αυτών των επιστημονικών επιτευγμάτων, ώστε να εξηγηθεί ο κόσμος και να δοθεί μια εικόνα του ανθρώπου και της σχέσης του με αυτόν. Τέλος, η φιλοσοφία μπορεί να αντιμετωπισθεί ως διάχυση των έλλογων επιστημονικών περιεχομένων στην κοινωνία και, κατ' επέκταση, ως διαφωτισμός του κοινού νου και ως άρση της άγνοιας της καθημερινής εμπειρίας. Σε τελική ανάλυση, η φιλοσοφία επιδιώκει να προσδιορίσει τους όρους του 'ευ ζην', θέτοντας την επιστήμη και τα πορίσματά της στην υπηρεσία του κοινωνικού συνόλου και στην καλυτέρευση της ανθρώπινης ζωής.

Η φιλοσοφία δεν αμφισβητεί τα αποτελέσματα των εμπειρικών ερευνών τη Γνωσιακής Επιστήμης και της 'Τεχνητής Ζωής', αλλά αναστοχάζεται γύρω από τις έννοιες, τις αρχές και τις μεθόδους τους, των οποίων η εγκυρότητα προϋποτίθεται σε όλα τα στάδια της επιστημονικής έρευνας. Η φιλοσοφία καλείται να επεξεργαστεί τα

επιχειρήματα των επιστημών αυτών και να τα αναδιατυπώσει με τέτοιο τρόπο, ώστε να τα ανασυγκροτήσει, να εντοπίσει, εάν υπάρχουν, κάποια θεωρητικά ή μεθοδολογικά προβλήματα και να οδηγήσει σε αποσαφήνιση και διαφωτισμό της σκέψης γύρω από αυτά.

Στα δύο πρώτα κεφάλαια της εργασίας αυτής θα μας απασχολήσουν οι αρχές και η μεθοδολογία της Γνωσιακής Επιστήμης: Αρχικά, θα εκθέσουμε θέματα εννοιολογικής θεμελίωσης και μεθοδολογίας της επιστήμης αυτής και στη συνέχεια θα επιχειρήσουμε να ελέγξουμε το εννοιολογικό τους σθένος. Στην προσπάθειά μας αυτή σημαντική βοήθεια θα μας προσφέρει ο John Searle, ο οποίος ασκεί ριζική κριτική τόσο στις μεθόδους όσο και στα πορίσματα των γνωσιακών επιστημόνων, όσον αφορά την εξήγηση των νοητικών φαινομένων. Προκειμένου να ασκήσει κριτική στο οικοδόμημα της γνωσιακής επιστήμης, ο Searle προβαίνει σε ένα νοητικό πείραμα και αποδεικνύει, σε ένα πρώτο επίπεδο πώς οι μέθοδοι και η εννοιολογική συγκρότηση του οικοδομήματος αυτού στηρίζονται σε αμφιλεγόμενα εννοιολογικά και μεθοδολογικά θεμέλια. Επομένως, οι θέσεις της Γνωσιακής Επιστήμης είναι, σύμφωνα με τον Searle, άνευ νοήματος και εννοιολογικά προβληματικές και η άποψή της ότι τα νοητικά φαινόμενα είναι καθαρά υπολογιστικά, σε ένα πρώτο επίπεδο, είναι εσφαλμένη τουλάχιστον όσον αφορά το πεδίο του νοήματος.

Στο επόμενο κεφάλαιο, θα εκθέσουμε το πρόγραμμα της ‘Τεχνητής Ζωής’ και θα ανασυγκροτήσουμε την επιχειρηματολογία και τις μεθόδους της.

Τέλος, θα επιχειρήσουμε να θέσουμε τη μεθοδολογία και τα πορίσματα της Γνωσιακής Επιστήμης και της ‘Τεχνητής Ζωής’ στο γενικότερο πλαίσιο συζήτησης στο οποίο υπάγονται. Το πλαίσιο αυτό δεν είναι κανένα άλλο από το φιλοσοφικό εγχείρημα, που οδηγεί στα βάθη της νεοτερικής φιλοσοφικής σκέψης, να εξηγηθεί τόσο ο νους όσο και η ζωή, με όρους φορμαλιστικούς και λειτουργικούς. Η Γνωσιακή Επιστήμη θεμελιώνεται στην άποψη πως τα φαινόμενα του

νου μπορούν να αναχθούν σε ένα μηχανικό-τυπικό σύστημα, το οποίο κατανοείται μέσω των λειτουργιών του, οι οποίες με τη σειρά τους ανάγονται σε αιτιακές σχέσεις. Με άλλα λόγια, τα νοητικά φαινόμενα ανάγονται σε λειτουργίες και θεωρούνται ως ένα τυπικό/ συντακτικό σύστημα, το οποίο εξηγείται με βάση αποκλειστικά τυπικές/ αιτιακές σχέσεις – όπως ακριβώς συμβαίνει με έναν ηλεκτρονικό υπολογιστή και τα μέρη του.

Τα συμπεράσματα, στα οποία θα καταλήξουμε, έχουν στόχο να διευρύνουν την κατανόησή μας, ώστε να διαμορφωθεί μια λίγο ως πολύ ολοκληρωμένη και συνεκτική εικόνα σχετικά με τη νόηση και τη ζωή, κι έτσι να συνεισφέρουν στο εγχείρημα που κάνει ο άνθρωπος να αυτο-κατανοηθεί. Επίσης, με τα πορίσματα της παρούσας εργασίας στοχεύουμε να σκιαγραφήσουμε τα όρια της σχέσης που μπορεί να υπάρξει μεταξύ των επιστημονικών επιτευγμάτων της εποχής μας με την προσπάθεια να κατανοηθούν η νόηση, η συνείδηση και η ζωή, που αποτελούν κλασικά φιλοσοφικά ερωτήματα.

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ : Η ΓΝΩΣΙΑΚΗ ΕΠΙΣΤΗΜΗ

1. ΕΙΣΑΓΩΓΗ

Οι επιστήμες της γνωσιακής και της εξελικτικής ψυχολογίας, της επιστήμης των υπολογιστών, της νευροεπιστήμης και της γλωσσολογίας απαρτίζουν, κατά κύριο λόγο, την Γνωσιακή Επιστήμη (Cognitive Science), η οποία συγκεφαλαιώνει μια ορισμένου τύπου προσπάθεια να εξηγηθούν επιστημονικά οι νοητικές λειτουργίες. Οι βασικές παραδοχές της Γνωσιακής Επιστήμης είναι οι εξής: α. Η νόηση εξισούται με τη λειτουργία της, και β. η λειτουργία εξισούται με υπολογισμό. Επομένως, η νόηση κατανοείται ως ένα υπολογιστικό φαινόμενο και το τελευταίο γίνεται κατανοητό σε επίπεδο λειτουργικό και τυπικό. Συγκεκριμένα, η Γνωσιακή Επιστήμη μελετά 1. αυτό που εισέρχεται/ ερεθίζει (input) το 'σύστημα' το οποίο κατανοεί, δηλαδή τον νου, 2. τις λειτουργίες που λαμβάνουν χώρα στο εσωτερικό του συστήματος αυτού και 3. Την εξωτερική αντίδραση (output), η οποία προκύπτει ως αποτέλεσμα αυτής της αιτιακής αλυσίδας σχέσεων.

2. ΔΥΟ ΜΕΤΑΦΥΣΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΤΟΥ ΝΟΥ

α. Υλισμός (Materialism): Μια από τις πλέον θεμελιώδεις οντολογικές φιλοσοφικές που έχουν υπάρξει ιστορικά (πάντα με τις διάφορες εκδοχές της), η οποία ως σκοπό έχει την σπουδή και την ερμηνεία του πώς υπάρχει ο κόσμος, όλα δηλαδή τα φαινόμενα, άρα και τα νοητικά φαινόμενα εντός αυτού, είναι ο Υλισμός. Πρόκειται για μια μεταφυσική θεώρηση σχετικά με το τι και το πώς υπάρχει. Ο υλισμός, ήδη από την προσωκρατική εποχή και την δημοκρίτεια ατομική θεωρία, υποστηρίζει πως οτιδήποτε υπάρχει είναι ύλη ή μπορεί να αναχθεί στην ύλη. Επομένως, όλα τα νοητικά φαινόμενα ανάγονται, με τη σειρά τους, σε αλληλεπιδράσεις της ύλης.

Οι φιλοσοφικές παραδοχές και η μεθοδολογία της Γνωσιακής Επιστήμης δεν βασίζονται σε αυτήν την μεταφυσική, ωστόσο, στο

πλαίσιο μιας συστηματοποίησης των δυνατών εναλλακτικών προσεγγίσεων του νου, θεωρούμε σκόπιμο να την αναφέρουμε, διότι, συνιστά μια σημαντική πρόταση για την εξήγηση των νοητικών φαινομένων, η οποία έχει εκφραστεί σε διάφορες εκδοχές στην ιστορία των ιδεών. Με τον όρο ‘υλισμό’, λοιπόν, εννοούμε μια μονιστική θεωρία, η οποία, αφενός, δέχεται πως ό,τι υπάρχει μπορεί να αναχθεί στο υλικό πεδίο. Ο υλισμός μελετά τις αιτιακές σχέσεις ανάμεσα στα υλικά σώματα και θεωρεί ότι όλα τα φαινόμενα είναι υλικά ή αναγώγιμα σε αυτά, και εξηγήσιμα με αιτιακές, φυσικές κατηγορίες – άρα αντικείμενα εξήγησης της φυσικής επιστήμης (physical). Για έναν υλιστή, ένας άνθρωπος είναι απλώς ένα φυσικό αντικείμενο που διαφέρει από τα άλλα φυσικά αντικείμενα μόνο ως προς την ιδιαίτερη συνθετότητα του φυσικού οργανισμού του και την ιδιαίτερη συνθετότητα των φυσικών του ικανοτήτων. Ο άνθρωπος δεν έχει καμία μη-φυσική ιδιότητα.⁵

Για τους υλιστές, η δυστική θεωρία αδυνατεί να εξηγήσει επαρκώς την ενότητα του νου με το σώμα. Είναι προβληματική, κατ’ αυτούς, διότι δεν μπορεί να αποδώσει, με συνεκτικό τρόπο, τα νοητικά και τα φυσικά χαρακτηριστικά ενός υποκειμένου, καθώς ‘σπάει’ την ενότητά του. Εξάλλου, σύμφωνα με τους υλιστές, ο δυισμός δεν συμφωνεί με τη σύγχρονη επιστημονική εικόνα του κόσμου, αφού δεν λαμβάνει υπόψη του την σύγχρονη επιστημονική γνώση για τον εγκέφαλο και τη λειτουργία του.⁶

Η σύγχρονη εκδοχή του υλισμού αποκαλείται “Central-State Materialism”. Κύριος εκφραστής της είναι ο David Armstrong. Σύμφωνα με αυτήν την θεωρία “όλες οι νοητικές εκδηλώσεις είναι ενδεχομένως ταυτόσημες με τις εκδηλώσεις του εγκεφάλου ή του

⁵ Armstrong, D.M.,(1968)

⁶ Armstrong D.M., [1968], σσ. 29-30

κεντρικού νευρικού συστήματος”.⁷ Οι βασικές αρχές του υλισμού, όπως διατυπώνονται από τον Armstrong, είναι οι εξής:

A) Η συνείδηση δεν έχει καμία σημασία (ή έχει δευτερεύουσα σημασία) για τα νοητικά φαινόμενα. “Η συνείδηση είναι απλά μια επιπλέον νοητική κατάσταση, μια κατάσταση κατευθυνόμενη (directed), η οποία υποβοηθά τις γνήσιες εσωτερικές καταστάσεις”⁸.

B) Η επιστήμη είναι αντικειμενική και αποτελεί την πιο ασφαλή μέθοδο για την αντικειμενική σπουδή της νόησης. Όπως χαρακτηριστικά δηλώνει ο Armstrong, “ιστορικά, η επιστημονική έρευνα έχει αποδειχθεί ότι είναι ο μόνος τρόπος με τον οποίο έχει επιτευχθεί ομοφωνία γνώμων σχετικά με θεωρητικά ζητήματα κι έχει κατορθώσει να δώσει στα ζητήματα αυτά σοβαρές και ευφυείς εξηγήσεις...Επομένως η επιστημονική μελέτη μπορεί να μας δώσει ένα εντελώς ιδιαίτερο κύρος σε δεδομένα ζητήματα της θεωρίας που συζητάμε τώρα, όπως το ποια είναι η φύση του νου”.⁹

Γ) Σε αντίθεση με τον συμπεριφορισμό, η θεωρία του υλισμού δέχεται ότι οι νοητικές καταστάσεις είναι υπαρκτές και δεν είναι τίποτε άλλο από φυσικές καταστάσεις του εγκεφάλου.¹⁰ Επομένως, θεωρεί ότι υπάρχει μια αιτιακή σχέση μεταξύ της νόησης και του σώματος, χωρίς να δέχεται το γεγονός ότι το προθετικό πεδίο είναι μη αναγώγιμο και ότι μπορεί να εξηγηθεί αποκλειστικά με νοητικούς όρους.

Δ) Ο υλισμός είναι σε θέση να εξηγήσει την αλληλεπίδραση του νου με το σώμα, την οποία αποδίδει στην αλληλεπίδραση του εγκεφάλου με το σώμα. Ουσιαστικά, κατ’ αυτήν την θεωρία ο νους ταυτίζεται με τον εγκέφαλο, στηρίζεται δηλαδή σε μια θεμελιώδη προϋπόθεση αρχών: Ο νους είναι υλικό αντικείμενο. Μ’ αυτόν τον τρόπο ο υλισμός θεωρεί ότι έλυσε το πρόβλημα που προξενεί ο δυισμός.

⁷ Audi, R., [1995], σ. 600

⁸ Armstrong D.M., [1968], σ. 94

⁹ Armstrong D.M., [1968], σ. 52

¹⁰ Armstrong D.M., [1968], σ. 75

Γενικότερα οι υλιστές δέχονται ότι υπάρχουν νόες, αλλά θεωρούν ότι είναι ‘ουσίες’ διαφορετικές από αυτές του Descartes. Δεν αρνούνται την ύπαρξη ‘εσωτερικών’ καταστάσεων¹¹, αλλά θεωρούν τον νου όχι χωριστά από τον εγκέφαλο, αλλά ταυτισμένον με αυτόν.¹²

β. Δυισμός (Dualism): Το δεύτερο κυρίαρχο φιλοσοφικό ρεύμα για το τι και πώς υπάρχει (επομένως αφορά το τι είναι και το πώς υπάρχουν και τα νοητικά φαινόμενα) στηρίζεται στις ακριβώς αντίθετες αρχές από αυτές του υλισμού, και είναι ο δυισμός. Στην κλασική εκδοχή του, ο δυισμός εμφανίζεται στη νεότερη φιλοσοφία με τον René Descartes, ο οποίος προσπάθησε να καταδείξει ότι η φύση περιέχει τους νόμους της λειτουργίας της μέσα της, οι οποίοι είναι αιτιακοί. Η φύση, κατ’ αυτόν τον τρόπο, μπορεί να ελεγχθεί πειραματικά. Από το άλλο μέρος, “η μεταφυσική έγινε, έτσι, γενικευμένη ή φιλοσοφική φυσική κι η πρόθεση ‘μετά-’ δεν υποδήλωνε τώρα πια τη μετάβαση από τον εμπειρικό κόσμο στην σφαίρα του υπερβατικού πνεύματος, παρά την εύρεση των έσχατων δομών και νομοτελειών του ίδιου του εμπειρικού κόσμου, οι οποίες, μολονότι δεν υπάρχουν έξω απ’ αυτόν τον ίδιο, όμως δεν είναι δυνατό να διατυπωθούν θεωρητικά, αν δεν υπερβαθεί το επίπεδο της άμεσης αισθητής εμπειρίας”.¹³

Στο παραπάνω πλαίσιο, όμως, δεν ‘ταίριαζε’ το νοητικό / ψυχικό πεδίο, το οποίο δεν μπορούσε να ενταχθεί στο πλαίσιο της λειτουργίας του φυσικού πεδίου και να εξηγηθεί με τους μαθηματικά διατυπώσιμους νόμους του. Έτσι, ο Descartes εισηγήθηκε την άποψη ότι η πραγματικότητα συντίθεται από δύο χωριστές ουσίες, το σώμα και τον νου/ την ψυχή. Ο άνθρωπος είναι μια σύνθετη οντότητα

¹¹ Armstrong D.M., [1968], σ. 72

¹² Χαρακτηριστική είναι η δήλωση του Armstrong [1968]: “The mind is simply the brain ”, σ. 73

¹³ Κονδύλης, Π., [1983], σ. 225

αποτελούμενη από τον νου, ο οποίος είναι ειδικά συνδεδεμένος με ένα σώμα (στο σώμα περιλαμβάνεται και ο εγκέφαλος). Ο νους και το σώμα είναι τελείως διαφορετικές φύσεις το ένα από το άλλο. Ο Descartes δεχόταν πως μεταξύ του νου και των υπόλοιπων φυσικών πραγμάτων πρέπει να υπάρχει αιτιακή αλληλεπίδραση.

Για πρώτη φορά, ο Descartes κάνει λόγο για τη διάκριση μεταξύ της ψυχής και του σώματος στο *Λόγος περί της μεθόδου*: “Απ’ αυτό κατάλαβα πως είμουν μια υπόσταση που ολόκληρη η ουσία ή η φύση της δεν είναι παρά το να σκέπτεται και δεν έχει ανάγκη, για να υπάρχει, από κανένα τόπο κι ούτε εξαρτιέται από τίποτα το υλικό. Έτσι που, αυτό το εγώ, δηλαδή η ψυχή, χάρη στην οποία είμαι ό,τι είμαι, είναι εντελώς ξέχωρη από το σώμα...Κι αν ακόμα το σώμα δεν υπήρχε διόλου, πάλι η ψυχή δεν θα έπαυε να είναι ό,τι είναι.”¹⁴ Η ψυχή, επίσης, είναι αθάνατη σε σχέση με το θνητό ανθρώπινο σώμα.¹⁵ Ανακεφαλαιώνοντας: Ο Descartes θεωρούσε πως υπάρχουν δύο είδη οντοτήτων ή ουσιών: 1) οι φυσικές οντότητες (*res extensa*), οι οποίες ως κύριο γνώρισμα έχουν την έκταση και 2) ο νους (*res cogitans*), ο οποίος δεν έχει κανένα φυσικό χαρακτηριστικό, αφού δεν μερίζεται και δεν καταλαμβάνει χώρο.

Ο δυισμός δεν κατάφερε να εξηγήσει ικανοποιητικά τον τρόπο με τον οποίο σχετίζονται ο νους με το σώμα. Έτσι, στην σύγχρονη εκδοχή του, ο δυισμός δεν επιμένει στο δυισμό της ουσίας, δηλαδή δέχεται ότι είναι δυνατόν ο νους και σώμα να έχουν την ίδια οντολογία ως προς τη φύση τους, αλλά υπερασπίζεται την άποψη ότι οι ιδιότητες των φαινομένων είναι οντολογικά διαχωρισμένες από τις φυσικές τους ιδιότητες. Κύριοι εκπρόσωποι της άποψης αυτής (που είναι γνωστή και ως επιφαινομεναλισμός) είναι ο K. K. Campbell¹⁶, ο T. Huxley¹⁷, ο F. Jackson¹⁸ και ο W.S. Robinson¹⁹.

¹⁴ Descartes, R., [1976], σ. 32

¹⁵ Descartes, R., [1976], σ. 54

¹⁶ Campbell, K.K., [1970], “*Body and Mind*”. Doubleday

3. ΜΕΘΟΔΟΙ ΕΞΗΓΗΣΗΣ ΤΟΥ ΝΟΥ

Το δεύτερο ερώτημα (πέρα από το τι και πώς υπάρχει κάτι) είναι πώς μπορούμε να γνωρίσουμε αυτό που υπάρχει, απάντηση η οποία μπορεί να επηρεάσει και την αντίληψή μας για το τι υπάρχει.

α. Φυσικαλισμός (physicalism): Ο φυσικαλισμός, ως μέθοδος εξήγησης του πώς θα μάθουμε γι' αυτό που υπάρχει, άρα και για τα νοητικά φαινόμενα, θεωρεί πως θα μάθουμε γι' αυτό που υπάρχει μέσω της καλύτερης κάθε φορά φυσικής μας θεωρίας. Ο φυσικαλισμός – μολονότι στο πλαίσιο της παρούσας εργασίας εκλαμβάνεται ως μεθοδολογική πρόταση για το πώς γνωρίζω κάτι (δηλαδή μέσω της φυσικής επιστήμης)– έχει επιπροσθέτως οντολογικές μεταφυσικές συνέπειες, κατά το ότι υποστηρίζει πως αυτό που υπάρχει είναι αυτό που εξηγείται ή είναι δυνατόν να εξηγηθεί από την καλύτερη κάθε φορά φυσική μας θεωρία. (μεθοδισμός).

Οι φυσικαλιστές, λοιπόν, επηρεασμένοι εμφανώς από το μεταφυσικό ρεύμα του υλισμού, επιχειρούν να εξηγήσουν τα νοητικά φαινόμενα ανάγοντάς τα σε ένα καθαρά υλισμικό/ φυσικαλιστικό επίπεδο. Θεωρούν πως όλα τα φαινόμενα (χημικά, βιολογικά, κοινωνιολογικά, ψυχολογικά κ.ά.) είναι κατά βάση φυσικά. Η φράση 'φυσικά φαινόμενα' μπορεί να εφαρμοστεί, κατά πρώτο λόγο, σε επίπεδο *ταυτότητας τύπων*, όπου όλες οι ειδικές ιδιότητες (π.χ. οι χημικές ή βιολογικές ιδιότητες) των αντικειμένων της εμπειρίας μας μπορούν να θεωρηθούν ότι είναι ή ότι ταυτίζονται με φυσικές

¹⁷ Huxley, T., [1893], "On the hypothesis that animals are automata". *Fortnightly Review* 95: 555-80. Reprinted in *Collected Essays*. London, 1893

¹⁸ Jackson, F., [1982], Epiphenomenal qualia. *Philosophical Quarterly* 32:127-136

¹⁹ Robinson, W.S., [1969], "Epiphenomenalism, Laws and Properties", *Philosophical Studies* 69: 1-34

ιδιότητες.²⁰ Κατά δεύτερο λόγο, η φράση ‘φυσικά φαινόμενα’ είναι δυνατόν να εξηγηθεί με όρους *επιγένεσης* (supervenience) των ειδικών κατηγοριών στις φυσικές. Η επιγένεση δεν σημαίνει τίποτε άλλο παρά ότι “δύο συστήματα δεν μπορούν να διαφέρουν χημικά ή βιολογικά... ή κατά οποιονδήποτε άλλο τρόπο, χωρίς να διαφέρουν φυσικά.”²¹ Αυτό ισχύει και αντίστροφα, δηλαδή δύο φυσικώς ταυτόσημα συστήματα είναι χημικώς, βιολογικώς, κτλ., ταυτόσημα.²²

Οι θιασώτες του φυσικαλισμού θεωρούν ότι, προκειμένου να εξηγηθεί η νόηση (‘ειδική κατηγορία’), θα πρέπει να ενταχθεί σε μια φυσική κατηγορία ή, με άλλα λόγια, να αναχθεί σε ένα φυσιολογικό επίπεδο. Η κατηγορία αυτή δεν είναι άλλη από τον εγκέφαλο, τους νευρώνες και τις συνάψεις τους.²³ Υπό την προοπτική αυτή, αντικείμενο μελέτης γίνεται η φυσική αλληλεπίδραση των παραπάνω μεταξύ τους.

β. Λειτουργισμός: Υπάρχει ένα άλλο μεθοδολογικό εγχείρημα εξήγησης των νοητικών φαινομένων που είναι ο λειτουργισμός. Πρόκειται ακριβώς για το εγχείρημα, στο οποίο βασίζεται το οικοδόμημα της Γνωσιακής Επιστήμης. Η τελευταία δεν ανάγει τα νοητικά φαινόμενα στο φυσικαλιστικό, αλλά σε ένα λειτουργικό επίπεδο, δηλαδή θεωρεί πως μπορεί να τα εξηγήσει με βάση ένα ενδιάμεσο πεδίο μεταξύ του νοητικού και του υλικού, εκείνο της λειτουργίας τους²⁴, το οποίο αποδίδει η νέα εν προκειμένω επιστήμη (η Γνωσιακή Επιστήμη) χωρίς καταφυγή στη φυσική επιστήμη. Η λειτουργιστική μέθοδος θεωρεί ως δεδομένο ότι οποιαδήποτε νοητική

²⁰ Ένα χαρακτηριστικό παράδειγμα διατύπωσης του φυσικαλισμού ταυτότητας τύπων είναι το εξής: η ιδιότητα να είναι κάτι υδρογόνο μπορεί να ταυτιστεί με την φυσική ιδιότητα να έχει άτομα με ένα πρωτόνιο και ένα ηλεκτρόνιο. (Papineau, D., [2002], σ. 12)

²¹ Papineau, D., [2002], σ. 13

²² Με όρους επιγένεσης λ.χ. το γεγονός ότι ένα άτομο-σύστημα χαίρεται σε σχέση με ένα άλλο άτομο-σύστημα που λυπάται, δηλαδή το γεγονός ότι δύο συστήματα διαφέρουν ψυχολογικά, σημαίνει ότι διαφέρουν και φυσικά, δε σχετίζονται δηλαδή με την ίδια φυσική μοριακή διάταξη. Αντιστρόφως, το γεγονός ότι δύο άτομα χαίρονται, για έναν φυσικαλιστή, δεν σημαίνει τίποτε άλλο παρά ότι τα άτομα αυτά είναι ψυχολογικά ταυτόσημα.

²³ Τη θεωρία αυτή υποστηρίζουν κυρίως οι Smart J.J.C. [1959/1990], Feigl, [1958] και Place, U.T., [1956/1980]

κατάσταση (επιθυμία, πεποιθήσεις κτλ.) είναι μια εσωτερική αιτιακή/ τυπική διαμεσολάβηση μεταξύ της αντίληψης (ερεθίσματος) και της εκδηλώσιμης συμπεριφοράς που ακολουθεί.

Για τους γνωσιακούς επιστήμονες, και υπό την επίδραση του μοντέλου των υπολογιστών, η νόηση, μπορεί να εξηγείται μάλλον σε σχήματα ροής πληροφοριών παρά με αναφορά στις υλικές δομές, όπου πραγματώνονται τα υπό διερεύνηση γνωσιακά συμβάντα. Ο H. Putnam, άλλωστε, παρέβαλλε τις νοητικές καταστάσεις με τις λειτουργικές ή 'λογικές' καταστάσεις ενός υπολογιστή. Όπως ένα υπολογιστικό πρόγραμμα μπορεί να πραγματοποιηθεί σε έναν μεγάλο αριθμό διαφορετικών υλικών διατάξεων, έτσι και τα νοητικά φαινόμενα (ως ενός είδους πρόγραμμα) είναι δυνατόν να πραγματοποιούν σε διαφορετικά υλικά υποστρώματα.²⁵ Ο νους λειτουργεί ανεξάρτητα από το υλικό πεδίο, αφού διαφορετικές υλικές πραγματώσεις μπορούν να στηρίζουν το ίδιο πρόγραμμα (π.χ. το πρόγραμμα BASIC μπορεί να εκτελεστεί από δύο υπολογιστές με διαφορετική γλώσσα και κατασκευασμένους από διαφορετικό υλικό). Το υλισμικό επίπεδο, όπου λαμβάνουν χώρα τα νοητικά φαινόμενα, είναι εντελώς αδιάφορο. Ο νους, ακριβώς όπως ο υπολογιστής, χρησιμοποιεί αλγόριθμους, οι οποίοι, ανεξάρτητα από το υλισμικό πεδίο (δηλαδή το επίπεδο πραγμάτωσης του προγράμματος), στηρίζουν το ίδιο πρόγραμμα.

Ο νους, τελικά, εξηγείται ως λειτουργία, η οποία γίνεται με απολύτως τυπικό τρόπο και προκαλεί συγκεκριμένα συμπεριφορικά αποτελέσματα. Τα νοητικά φαινόμενα εξηγούνται ως λειτουργίες – ακριβώς όπως εξηγείται λ.χ. ο ηλεκτρισμός – και δεν διαθέτουν κανενός είδους περιεχόμενο.

²⁴ Blackwell, B., [1994], σ. 318

²⁵ Putnam, H., [1960]

γ. Η μη αναγωγιστική προθεσιακή προσέγγιση²⁶: Σύμφωνα με μια τρίτη προσέγγιση, η οποία αποκτά ολοκληρωμένη και συστηματική μορφή στην καντιανή κριτική του νατουραλισμού, αλλά και στην κριτική του Frege κατά του ψυχολογισμού, τα νοητικά φαινόμενα είναι δυνατόν να εξηγηθούν μόνο με όρους προθεσιακών-νοητικών, δηλαδή αποκλειστικά σε επίπεδο νοήματος. Σύμφωνα με αυτήν την οπτική, για τη νόηση δεν μπορεί να δοθεί παρά μία μη αναγωγική πρωτογενής εξήγηση. Η εξήγηση αυτή βασίζεται στο προθεσιακό λεξιλόγιο και δεν θεωρεί ότι τα νοητικά φαινόμενα μπορούν να 'επιστημονικοποιηθούν' και να εξηγηθούν με κατηγορίες που δεν διαθέτουν την έννοια 'περιεχόμενο'. Αυτό σημαίνει ότι το περιεχόμενο της νόησης δεν επιδέχεται αιτιακή αναγωγή γιατί είναι αυτοεξηγούμενο. Η κατανόηση της νόησης λαμβάνει χώρα μόνο στο πεδίο του νοήματος, ως ανερμηνεύσιμο.

Τα νοητικά φαινόμενα, λοιπόν, έχει επιχειρηθεί να εξηγηθούν εντός των πλαισίων της εμφάνισής τους και να ερμηνευτούν με νοητικούς όρους χωρίς να αναχθούν σε άλλο επίπεδο. Υπάρχουν και δύο εγχειρήματα, τα οποία ανάγουν τη νόηση: το πρώτο σε επίπεδο λειτουργιών (όπου τα νοητικά φαινόμενα εξηγούνται ως μηχανικά/τυπικά φαινόμενα) και το δεύτερο σε υλισμικό επίπεδο (όπου εξηγούνται ως αμιγώς φυσικά φαινόμενα σε επίπεδο φυσικής αλληλεπίδρασης του εγκεφάλου με τους νευρώνες κτλ.).

4. ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΕΠΙΣΤΗΜΟΝΙΚΕΣ ΠΡΟΚΕΙΜΕΝΕΣ ΤΗΣ ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ

²⁶ Τη θεωρία αυτή υποστηρίζει ο John McDowell [1994], ο οποίος έλκει τις απόψεις του από τον G. Frege, ο οποίος θεωρεί ότι δεν χρειάζεται κανενός είδους γεγονός για να εξηγηθεί η λογική, γιατί αυτή είναι που εξηγεί τα γεγονότα [1976], τον L. Wittgenstein [1953], § 217, και τον Imm. Kant [1929]

Πριν προχωρήσουμε στην εξέταση των βασικών αρχών της Γνωσιακής Επιστήμης, θα αναφερθούμε στην *Γνωσιακή Ψυχολογία* και την *Τεχνητή Νοημοσύνη*, διότι αυτές, κατά κύριο λόγο, ‘υποστηρίζουν’ εννοιολογικά το πρόγραμμα της Γνωσιακής Επιστήμης. Η *Γνωσιακή Ψυχολογία (Cognitive Psychology)* χρησιμοποίησε τις δεκαετίες του ’30 και του ’40 ως βασικό τύπο θεωρίας για την εξήγηση της αντίληψης και των νοητικών διεργασιών τη θεωρία του **συμπεριφορισμού (behaviorism)**. Από την πλευρά της ψυχολογίας, ο συμπεριφορισμός εμφανίστηκε ως θεωρία, η οποία θα εξηγούσε την αντίληψη, την σκέψη και την συμπεριφορά. Κύριοι εκπρόσωποι της θεωρίας αυτής είναι ο J. Watson²⁷ και ο F.B. Skinner²⁸.

Σύμφωνα με το συμπεριφοριστικό μοντέλο, προκειμένου να προσεγγιστούν τα παραπάνω φαινόμενα, πρέπει να οριστούν από το εξωτερικό ερέθισμα (input) που επηρεάζει το άτομο, το οποίο με τη σειρά του εκδηλώνει μια ορισμένη/ συγκεκριμένη συμπεριφορά (output)²⁹ σε συγκεκριμένο περιβάλλον. Η συμπεριφορά ‘ενισχύεται’ θετικά ή αρνητικά, και αυξάνει ή μειώνει, αντιστοίχως, την πιθανότητα να εκδηλωθεί η ίδια συμπεριφορά σε παρόμοιες συνθήκες. Η ενίσχυση αυτή προέρχεται από περιβαλλοντικές τυχαίες επιδράσεις σχεδόν ταυτόχρονα με την αρχική συμπεριφορά.

Σε καμία περίπτωση ο συμπεριφορισμός δεν δέχεται ότι υπάρχουν εσωτερικές νοητικές διεργασίες, μη αναγώγιμες σε εξωτερικές συμπεριφορικές τάσεις ή διαθέσεις για συμπεριφορά. Επιπλέον, οι θιασώτες του συμπεριφορισμού, όταν εξηγούν την συμπεριφορά, δεν χρησιμοποιούν νοητικούς, προθεσιακούς όρους όπως ‘θέλω’, ‘επιθυμώ’, ‘προτίθεμαι’ κ.ά. Η χρήση τέτοιου είδους όρων, σύμφωνα με τους συμπεριφοριστές, δεν έχει καμία εξηγητική ισχύ. “Η γλώσσα των νοητικών καταστάσεων αναφέρεται σε κάτι ,το οποίο δεν είναι

²⁷ Watson, J., *Psychology from the standpoint of a behaviorist* : Philadelphia: Lippincott , 1919

²⁸ Skinner, B.F., [1953], *Science and Human Behavior*. New York: Macmillan

²⁹ Machamer, P., [1999], σ. 491

παρατηρήσιμο για λόγους αρχής, κι έτσι δεν έχει καμία θέση στην εμπειρική επιστήμη”.³⁰

Στον ψυχολογικό συμπεριφορισμό ασκήθηκε έντονη κριτική. Από τους σημαντικότερους επικριτές ήταν ο γλωσσολόγος Noam Chomsky³¹, ο οποίος επισήμανε πως ο συμπεριφορισμός αδυνατούσε να περιλάβει στην εξήγησή του την ικανότητα του ανθρώπου να παραγάγει έναν αμέτρητο αριθμό προτάσεων σε μια φυσική γλώσσα, αρκετές από τις οποίες μπορεί να μην είχαν εκφωνηθεί ποτέ στο παρελθόν. Η εκφώνηση μιας τέτοιας πρότασης είναι μια συμπεριφορά που δεν έχει ξανασυμβεί, και, στην περίπτωση αυτή, η ενίσχυση αδυνατεί να εξηγήσει τον τρόπο με τον οποίο οι άνθρωποι μαθαίνουν να παράγουν για πρώτη φορά προτάσεις.

Το ρεύμα του συμπεριφορισμού επηρέασε και τη φιλοσοφία. Ο φιλοσοφικός/ αναλυτικός συμπεριφορισμός είναι μια θεωρία, η οποία γεννήθηκε ως αντίδραση στον καρτεσιανής προέλευσης δυισμό. Εκπρόσωπος της θεωρίας αυτής είναι ο Hempel.³² Ο αναλυτικός συμπεριφορισμός είναι μια θεωρία, η οποία υποστηρίζει ότι “οι ψυχολογικές προτάσεις είναι μεταφράσιμες, χωρίς απώλεια περιεχομένου, σε προτάσεις που αφορούν συμπεριφορές ή διαθέσεις για συμπεριφορά.”³³ Πρόκειται για την άποψη ότι η συμπεριφορά είναι θεμελιώδης για την κατανόηση των νοητικών φαινομένων. Ο συμπεριφορισμός εφαρμόζει την επαληθευσιοκρατική θεωρία του νοήματος, σύμφωνα με την οποία “το νόημα των προτάσεων έγκειται σε όσα παρατηρήσιμα συμβάντα ή περιστατικά τις επαληθεύουν ή επιβεβαιώνουν την αλήθεια τους”.³⁴ Με άλλα λόγια, για το συμπεριφορισμό, οποιαδήποτε [δια]νοητική³⁵ ιδιότητα ή το νόημα

³⁰ Machamer, P., [1999], σ. 492

³¹ Chomsky, N., [1959], *Review of Verbal Behavior*, by B. F. Skinner. *Language* 35: 26-58

³² Hempel, [1935]

³³ Παγωνδιώτης, Κ., [2001], σ. 38

³⁴ Audi, R., [1995], σ. 67

³⁵ Ο αγγλικός όρος ‘mind’ αποδίδεται από τους Έλληνες μεταφραστές είτε ως ‘νους’ είτε ως ‘διάνοια’. Εμείς υιοθετούμε την πρώτη εκδοχή και στο εξής σε όποια βιβλιογραφική αναφορά χρησιμοποιούμε θα τον αποδίδουμε ως ‘νου’.

οποιασδήποτε πρότασης που αποδίδει μια συγκεκριμένη νοητική ιδιότητα, συνίσταται στην συμπεριφορά ή στην συμπεριφορική τάση που συνεπάγεται η πρόταση.

Με άλλα λόγια, λοιπόν, για τον συμπεριφορισμό το μόνο που είναι δυνατόν να υπάρχει και μπορεί να γίνει αντικείμενο παρατήρησης, άρα επιστημονικής εξήγησης, είναι οι φανερές και φυσικά εξηγήσιμες συμπεριφορές. Ακόμα παραπέρα, ο συμπεριφορισμός διατυπώνει μια οντολογική θέση: οι συμπεριφορές αυτές είναι όλο κι όλο αυτό που υπάρχει. Επομένως, όσον αφορά το νοητικό ή προθεσιακό πεδίο, τότε μόνο υφίσταται, όταν μπορεί να γίνει αντικείμενο παρατήρησης της έκδηλης συμπεριφοράς.

Ωστόσο, και ο φιλοσοφικός συμπεριφορισμός αντιμετώπισε προβλήματα και του ασκήθηκε έντονη κριτική. Η κριτική αυτή αφορούσε κυρίως στο ότι μόνη η μελέτη της συμπεριφοράς απορρίπτει τους εσωτερικούς μηχανισμούς ή τις διεργασίες, που διαμεσολαβούν τη σχέση 'ερεθίσματος' και 'απόκρισης'. Πρώτα πρώτα, εάν τα νοητικά φαινόμενα εξηγηθούν μόνο με συμπεριφορικούς όρους, τότε δεν συνυπολογίζεται στην εξήγησή τους η 'εσωτερικότητά' τους. Επίσης συχνά ένα ορισμένο ερέθισμα δεν ακολουθείται πάντοτε από την συμπεριφορά που συνήθως έχουν οι άνθρωποι (για παράδειγμα κάποιος μπορεί να πονάει, αλλά να μην το δείχνει). Τέλος, για την εξήγηση/ ορισμό νοητικών όρων, εκτός από τους συμπεριφορικούς όρους απαιτείται η εισαγωγή και νοητικών όρων³⁶. Για παράδειγμα, το νοητικό κατηγορημα 'πιστεύει ότι πεινάει' με συμπεριφορικό ορισμό θα γινόταν 'αν ο X πεινούσε, θα άνοιγε το ψυγείο για να φάει'. Ωστόσο, υπάρχει η περίπτωση κάποιος να κάνει δίαιτα. Τότε, ο ορισμός θα γινόταν 'αν ο X πεινούσε και ήθελε να κάνει δίαιτα, ο X δεν θα έτρωγε'. Στο δεύτερο ορισμό προστέθηκε ο όρος 'ήθελε', ο οποίος είναι προθετικός. Δεν μπορούμε, επομένως, να έχουμε αμιγώς συμπεριφορικό λεξιλόγιο όταν αναφερόμαστε σε πεποιθήσεις, διότι

³⁶ Goldman, A.I., [1998], σ. 94

είμαστε αναγκασμένοι να εισάγουμε προηγουμένως το προθετικό λεξιλόγιο (π.χ. επιθυμώ).

Η θεωρία, την οποία υιοθέτησε αργότερα η Γνωσιακή Ψυχολογία, και η οποία προήλθε ως απάντηση για τη λύση των προβλημάτων του συμπεριφορισμού, είναι ο **λειτουργισμός (functionalism)**. Οι πιο σημαντικοί εκπρόσωποι του λειτουργισμού είναι ο πρώιμος Hilary Putnam³⁷ και ο David Lewis³⁸. Λειτουργισμός, σε αδρές γραμμές, είναι η μέθοδος που, ως ‘επίγονος’ του συμπεριφορισμού, θεωρεί ότι οι νοητικές καταστάσεις μπορούν να εξηγηθούν με αναφορά την εκδηλούμενη συμπεριφορά του συστήματος. Σε αντίθεση όμως με αυτόν, αναγνωρίζει ότι υπάρχουν εσωτερικές νοητικές καταστάσεις, οι οποίες, προσδιορίζονται από τα ερεθίσματα/ αιτίες που τις προκαλούν και από τα αποτελέσματα που παρουσιάζουν. Τα νοητικά φαινόμενα ταυτίζονται με λειτουργίες και καθίστανται ανεξάρτητα από το μέσο υλοποίησής τους. Σύμφωνα με τον λειτουργισμό, η νόηση αποτελεί μια διεργασία, η οποία “χαρακτηρίζεται επαρκώς στο επίπεδο της αφηρημένης οργάνωσης πράξεων επί συμβόλων³⁹”.

Για το λειτουργισμό, τα νοητικά φαινόμενα δεν είναι τίποτε άλλο παρά λειτουργίες και αποδίδονται με όρους της πληροφορικής και όχι με προθετικούς όρους. Σύμφωνα με τη λειτουργιστική υπόθεση, ο ανθρώπινος νους είναι ακριβώς σαν ένα πρόγραμμα που μπορεί να πραγματοποιηθεί με πολλές διαφορετικές μηχανικές και υλικές μορφές. Η νόηση, επομένως, είναι μία λειτουργία που ορίζεται από συντακτικές σχέσεις και μηχανικό χειρισμό συμβόλων.

Οι σχέσεις μεταξύ 1. του εξωτερικού ερεθίσματος με κάποια νοητική κατάσταση, 2. των νοητικών καταστάσεων μεταξύ τους και 3. των νοητικών καταστάσεων με έκδηλη συμπεριφορά διέπονται από

³⁷ Putnam, H., [1963], “Brains and Behavior,” in R. Butler, ed., *Analytical Philosophy, Second Series*. Oxford: Basil Blackwell. και Putnam, H., [1975], “Philosophy and Our Mental Life”, in *Mind, Language and Reality: Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.

³⁸ Lewis, D.K., [1972] και [1983]

³⁹ Μαραγκός, [1996], σ. 114

αιτιακούς νόμους, δηλαδή από αμιγώς τυπικές/ μηχανικές διαδικασίες. Το νόημα των νοητικών καταστάσεων δεν εξηγείται παρά μόνο εάν μάθουμε τους αιτιακούς νόμους που συνδέουν το ερέθισμα με τις νοητικές καταστάσεις και τη συμπεριφορά. Τέλος, σύμφωνα με την βασική υπόθεση του λειτουργισμού, “ο ανθρώπινος νους είναι ακριβώς σαν ένα πρόγραμμα⁴⁰ που μπορεί να πραγματοποιηθεί με πολλές διαφορετικές μηχανικές και υλικές μορφές, μια από τις οποίες είναι και το νευρωνικό υλισμικό.”⁴¹

Όπως φάνηκε από τον παραπάνω ορισμό, ο λειτουργισμός αποφεύγει τις ‘παγίδες’ του συμπεριφορισμού (αφού τουλάχιστον αναγνωρίζει την ύπαρξη νοητικών καταστάσεων), ωστόσο αποδίδει τον ιδιαίτερο χαρακτήρα των νοητικών καταστάσεων όχι σε εγγενείς ιδιότητες του νου, αλλά στο ότι προσδιορίζονται από αιτιακές σχέσεις. Πρώτα πρώτα σχέσεις με εξωτερικά συμβάντα, όπως εισροές (input) ερεθισμάτων και συμπεριφορικές εκροές (output), και δεύτερον σχέσεις με άλλες εσωτερικές καταστάσεις.

Τόσο ο συμπεριφορισμός όσο και ο λειτουργισμός, τέλος, συγγενεύουν με τον υλισμό, γιατί εν τέλει ορίζουν τα νοητικά κατηγορήματα μόνο με αναφορά στις σχέσεις τους προς φυσικές εισροές και εκροές.

Ο δεύτερος επιστημονικός κλάδος, που ‘στηρίζει’ το πρόγραμμα της Γνωσιακής Επιστήμης, είναι η *Τεχνητή Νοημοσύνη (Artificial Intelligence)*, η οποία οφείλει την εμφάνιση και την ανάπτυξή της στη θεωρία του υπολογισμού, που γεννιέται από την επιστήμη των μαθηματικών. Η θεωρία του υπολογισμού, βρήκε, μετά από τον δεύτερο παγκόσμιο πόλεμο, μια πρακτική εφαρμογή με την ανάπτυξη των ηλεκτρονικών ψηφιακών υπολογιστών αποθηκευμένου προγράμματος. “Ήταν φανερό στους σχεδιαστές των πρώτων

⁴⁰ Με τον όρο ‘πρόγραμμα’ εννοούμε ένα “σύνολο εντολών –ενός αλγόριθμου–, και ενός συνόλου πληροφοριών –μιας δομής δεδομένων... Αλγόριθμοι και δομές δεδομένων αλληλοπροσδιορίζονται: ποιες πράξεις είναι εφικτές και λογισμικά βέλτιστες εξαρτάται από το είδος των προς επεξεργασία πληροφοριών.” (Μαραγκός, [1996], σ. 120)

ψηφιακών υπολογιστών ότι οι μηχανές αυτές θα μπορούσαν να χρησιμοποιηθούν για να προσομοιώσουν, και από μερικές απόψεις να ξεπεράσουν, τις ανθρώπινες νοητικές ικανότητες.”⁴²

Η Τεχνητή Νοημοσύνη δεν είναι τίποτε άλλο από το εγχείρημα που γίνεται σήμερα, από τους σχεδιαστές προγραμμάτων για ηλεκτρονικούς υπολογιστές, να δημιουργήσουν και να εφαρμόσουν κατάλληλα προγράμματα, τα οποία να προσεγγίζουν και να αντιμετωπίζουν προβλήματα με νοήμονα τρόπο. Τέτοιου είδους προβλήματα μπορεί να είναι η νίκη σε ένα παιχνίδι, η επιτέλεση υπολογισμών, η προσομοίωση ενός τυφώνα ή οποιουδήποτε άλλου φυσικού φαινομένου, η ανακάλυψη φυσικών νόμων κ.ά.

Ωστόσο, το πρόγραμμα της Τεχνητής Νοημοσύνης έχει περιορισμένη εξηγητική εμβέλεια και δεν φιλοδοξεί να συνδέσει τις υπολογιστικές διαδικασίες που λαμβάνουν χώρα σε έναν υπολογιστή με τις νοητικές διεργασίες των ανθρώπων. Επίσης, ο ισχυρισμός ότι ένα υπολογιστικό πρόγραμμα περιγράφει τον τρόπο, με τον οποίο πράγματι λειτουργούν οι άνθρωποι, για την Τεχνητή Νοημοσύνη είναι αδιάφορος και δεν λαμβάνεται υπόψη στο πρόγραμμά της.⁴³ Με άλλα λόγια, η Τεχνητή Νοημοσύνη δεν επιδιώκει να απαντήσει στο εάν και οι άνθρωποι σκέφτονται μέσω τέτοιων διαδικασιών.

Αν θελήσουμε να επισημάνουμε τη σχέση μεταξύ της Τεχνητής Νοημοσύνης και της Γνωσιακής Ψυχολογίας, θα μπορούσαμε να πούμε ότι τα αντικείμενά τους σχετίζονται στενά, αφού και οι δυο ασχολούνται με το ίδιο ζήτημα, πώς δηλαδή θα μπορούσε να επέλθει μια νοήμων πράξη μέσω υπολογισμού. Από την άλλη, διαφέρουν ως προς το στόχο τους: Η Γνωσιακή Ψυχολογία ως στόχο έχει να διατυπώσει ακριβείς περιγραφές των υπολογιστικών διεργασιών, οι οποίες κάνουν τους ανθρώπους να σκέφτονται, να συλλογίζονται, να θυμούνται, να αναγνωρίζουν, να επιλύουν προβλήματα κ.ά. Με άλλα

⁴¹ Machamer, P., [1999], σσ. 486-487

⁴² Glymour, C., [1999], σ. 541

λόγια, η Γνωσιακή Ψυχολογία ενδιαφέρεται να εξηγήσει τον τρόπο με τον οποίο οι άνθρωποι μαθαίνουν ή εκτελούν οποιοδήποτε γνωσιακό έργο. Αντίθετα, το πεδίο της Τεχνητής Νοημοσύνης θεωρείται πιο αφηρημένο: “Δεν προτείνονται τόσο περιγραφικά όσο εξηγητικά υποδείγματα, με την έννοια ότι ενδιαφέρει το πώς θα ήταν δυνατόν να προκύψει το υπό μελέτη φαινόμενο ως αποτέλεσμα υπολογιστικής δραστηριότητας, ανεξαρτήτως του υλικού υποστρώματος που την διεκπεραιώνει.”⁴⁴

4. ΤΑ ΕΝΝΟΙΟΛΟΓΙΚΑ ΘΕΜΕΛΙΑ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

ΤΗΣ ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ (Jerry Fodor)

Θεωρητικός εκπρόσωπος της Γνωσιακής Επιστήμης είναι ο Jerry Fodor⁴⁵, του οποίου τις απόψεις επιλέξαμε να εκθέσουμε, διότι είναι ένας από τους βασικούς θεμελιωτές του θεωρητικού πλαισίου της. Το συνολικό έργο του, και ειδικότερα το *Language of Thought* (1975), έπαιξε σπουδαίο ρόλο στη διαμόρφωση της επιστήμης αυτής. Ο Fodor προσφέρει στο επιστημολογικό πρόγραμμα της Γνωσιακής Επιστήμης μια *νατουραλιστική εξήγηση* για τον νου μέσω της *Αναπαραστασιακής-Υπολογιστικής Θεωρίας του Νου*.

Ο Fodor δεν υιοθετεί τη μέθοδο του φυσικαλισμού, σχετικά με το ζήτημα της φύσης του νου, και δεν δέχεται την υλιστική θέση πως καθετί που υπάρχει και που μπορεί να γνωσθεί είναι (ή μπορεί να γίνει) αντικείμενο της φυσικής επιστήμης.⁴⁶ Αντίθετα, προκρίνει τον *νατουραλισμό*: Πρόκειται για μία συγκεκριμένη αντίληψη για τη φύση εν γένει, και για τη φύση του νου ειδικότερα. Ο νατουραλισμός αντιμετωπίζει τα φαινόμενα, για παράδειγμα τα βιολογικά, τα χημικά κ.ά., ως ‘φυσικά’, με την ευρεία έννοια του ‘natural’, δηλαδή ως τμήματα της φύσης, (αλλά όχι κατ’ ανάγκην ως ‘φυσικά’ με την έννοια

⁴³ Glymour, C., [1999], σ. 510

⁴⁴ Μαραγκός, [1996], σ. 112

⁴⁵ Fodor, J.A., [1975], *The Language of Thought*. J.J.Katz, D.T. Langendoen, G.A. Miller ed., Harvard University Press

του 'physical', δηλαδή δεν εμπίπτουν απαραίτητως στο εξηγητικό πεδίο της φυσικής επιστήμης⁴⁷). Με άλλα λόγια, η φύση δεν είναι τίποτε άλλο από "αντικείμενο της αιτιακής-νομολογικής περιγραφής, του οποίου τα θεμελιώδη χαρακτηριστικά εξηγούνται από τις κατηγορίες της μοντέρνας φυσικής επιστήμης, σε αντιδιαστολή προς το λεξιλόγιο που εννοιολογικοποιεί τα υποκείμενα και τις δραστηριότητές τους [δηλαδή το προθετικό λεξιλόγιο]"⁴⁸. Ο νατουραλισμός διευρύνεται και μπορεί να αφορά και τα νοητικά φαινόμενα, τα οποία, με όποιον τρόπο κι αν εξηγηθούν, η εξήγησή τους θα παραμείνει στο φυσιοκρατικό πλαίσιο.⁴⁹

Πριν αναφερθούμε στην *αναπαραστασιακή θεωρία του νου*, την οποία πρότεινε ο Fodor ως επιστημονική εξήγηση για τον νου, σκόπιμο θα ήταν να ορίσουμε τον όρο 'αναπαράσταση' και να την διαχωρίσουμε από τον όρο 'παράσταση'. Ως 'παράσταση', στη νεότερη φιλοσοφία, εκλαμβάνεται εκείνη η νοητική οντότητα που μεσολαβεί μεταξύ του υποκειμένου και του κόσμου και δίνει την δυνατότητα στο πρώτο να έχει έμμεση πρόσβαση στον δεύτερο. Παράσταση μπορεί να είναι μια ιδέα (impression), μια εποπτεία (intuition), μια έννοια (concept) ή γενικότερα ένα αισθητηριακό δεδομένο (sense data), ανάλογα με την συγκεκριμένη θεωρία που υφίσταται. Όλα τα παραπάνω είναι προσβάσιμα στην πρωτοπρόσωπη, υποκειμενική παρατήρηση. Για την εξήγηση των παραστάσεων δεν υιοθετείται το μοντέλο της γλώσσας.

Αντίθετα, ως 'αναπαράσταση' μπορεί να θεωρηθεί, γενικώς, οτιδήποτε μπορεί να σταθεί στη θέση, να πάρει τη θέση κάποιου άλλου

⁴⁶ Fodor, J.A., [1975], σ. 24

⁴⁷ Η θεωρία του φυσικαλισμού, όπως είδαμε, ανάγει τα νοητικά φαινόμενα σε φυσικά και πρεσβεύει ότι μπορεί να 'επιστημονικοποιήσει' τη μελέτη της νοητικής ζωής με την αρωγή της φυσικής επιστήμης, της βιολογίας κ.ά. Ο φυσικαλισμός, όπως αναφέραμε και προηγουμένως (βλ. σσ. 15-16) είναι όρος σχεδόν ταυτόσημος με τον υλισμό, ενώ ο νατουραλισμός όχι.

⁴⁸ Τσινόρεμα, Σ., [1999], σσ. 2-3

⁴⁹ Παγωνδιώτης, Κ., [2001], σ. 46

πράγματος.⁵⁰ Αναπαράσταση λ.χ. μπορεί να θεωρηθεί ένας χάρτης, μια πρόταση, ένα μοντέλο αεροπλάνου, μια εικόνα κτλ. Ως *περιεχόμενο* της αναπαράστασης θεωρείται το πράγμα, του οποίου τη θέση παίρνει η αναπαράσταση. (Για παράδειγμα, περιεχόμενο ενός χάρτη συνιστά ο τόπος, τον οποίο αυτός απεικονίζει).

Η μορφή της αναπαράστασης έχει, βεβαίως, σχέση με την φυσική εμφάνισή της. Ωστόσο, η υλική υπόσταση της αναπαράστασης δεν είναι στο σύνολό της σημαντική για τον προσδιορισμό της μορφής. Σε έναν χάρτη, για παράδειγμα, δεν μας ενδιαφέρει τόσο το υλικό της κατασκευής του, όσο τα σχήματα, τα χρώματα κτλ., με τα οποία εικονίζεται το αντικείμενό του. Κατ' αυτόν τον τρόπο, σε κάθε αναπαράσταση υπάρχει η *σημαίνουσα μορφή* και ο *φορέας της αναπαράστασης*, ο οποίος διακρίνεται από την συνολική υλική υπόστασή της.⁵¹ Μ' αυτήν ακριβώς τη διάκριση μπορούμε να επιτύχουμε την απομόνωση των χαρακτηριστικών ενός σώματος, τα οποία μπορούν να αναπαρασταθούν, από τα υπόλοιπα χαρακτηριστικά.

Ένα ακόμη στοιχείο, το οποίο μας βοηθάει να οριοθετήσουμε την έννοια της αναπαράστασης είναι το να δούμε ποια λειτουργία αυτή επιτελεί. Η λειτουργία της, λοιπόν, συνίσταται στο ότι *αναφέρεται* σε αυτό του οποίου τη θέση παίρνει (αρκεί, βεβαίως, να μπορεί δυνάμει να χρησιμοποιηθεί). Ο Fodor εισάγει την έννοια της αναπαράστασης, προκειμένου να δώσει ένα επιστημονικό εργαλείο για την εξήγηση των νοητικών φαινομένων. Ως νοητικά φαινόμενα εκλαμβάνει εκείνα που η *'λαϊκή/ δημώδης ψυχολογία'* ή η *'ψυχολογία του κοινού νου'* (folk psychology)⁵² θεωρεί ως τέτοια: πρόκειται για τις πεποιθήσεις, τις επιθυμίες και τις προθέσεις. Τα φαινόμενα αυτά έχουν σημασιολογικές και αιτιακές ιδιότητες (μια πεποίθηση, δηλαδή μπορεί να είναι αληθής

⁵⁰ Παγωνδιώτης Κ., [2001], σ. 77

⁵¹ Παγωνδιώτης Κ., [2001], σσ. 78-79

⁵² Η δημώδης ψυχολογία είναι μια θεωρία, η οποία ξεκινά από τη θέση ότι "η ταξινόμηση των νοητικών φαινομένων, όπως αυτή εκφράζεται μέσα στην ψυχολογία του κοινού νου, είναι κατά βάση ορθή και γι' αυτό πρέπει να αποτελέσει το θεμέλιο για μια επιστημονική ψυχολογία.

ή ψευδής και να γίνει η αιτία κάποιας πράξης του υποκειμένου αντίστοιχα). Τα νοητικά φαινόμενα, που αναφέρθηκαν παραπάνω, είναι γνωστά ως ‘προτασιακές στάσεις’ (propositional attitudes).

Το πρόβλημα με τις προτασιακές στάσεις είναι πώς το υποκείμενο, που τις έχει, συνδέεται με αυτές (γιατί, εάν δεν υπάρχει σχέση μια φυσική σχέση, υπάρχει ο ‘κίνδυνος’ να υποπέσουμε σε δυισμό). Ο Fodor εισάγει, λοιπόν, την έννοια μιας *εσωτερικής αναπαράστασης*, η οποία παίζει διαμεσολαβητικό ρόλο μεταξύ του υποκειμένου και των προτασιακών στάσεων. Αυτή η αναπαράσταση έχει ως περιεχόμενο κάθε προτασιακή στάση και είναι γλωσσική, ‘γραμμένη’ δηλαδή στη ‘γλώσσα της σκέψης’. Η αναπαράσταση είναι ασυνείδητη για το υποκείμενο, το οποίο δεν μπορεί ποτέ να έχει πρόσβαση σ’ αυτήν. Επίσης, οι αναπαραστάσεις δεν είναι προσβάσιμες ούτε στους εξωτερικούς παρατηρητές. (Πρόσβαση σε αυτές έχουν μόνο τα αυτόνομα γνωσιακά ‘υποσυστήματά’ του υποκειμένου⁵³).

Η παραπάνω γλωσσική αναπαράσταση είναι τυπική, δηλαδή έχει έναν ανεμήνευτο συμβολισμό (σύμβολα), ο οποίος λειτουργεί βασισμένος σε συντακτικούς κανόνες, που καθορίζουν τους δυνατούς τρόπους συνδυασμού των συμβόλων. Ο συνδυασμός αυτός γίνεται με βάση τη μορφή των συμβόλων και όχι το περιεχόμενό τους. Η γλώσσα της σκέψης έχει, βεβαίως, κάποιον υλικό φορέα, επομένως εκλαμβάνεται από τον Fodor ως ένα είδος φυσικού φαινομένου⁵⁴, πράγμα που θεωρείται πως εξασφαλίζει τον νατουραλισμό των νοητικών φαινομένων, άρα την επιστημονική εξήγησή τους. Η ΑΥΘΝ διεκδικεί πως είναι, επομένως, μια θεωρία, η οποία, προσφέρει στην Γνωσιακή Επιστήμη μια επιστημονική, νατουραλιστική εξήγηση για τον νου.

Η σχέση μεταξύ του οργανισμού και της νοητικής αναπαράστασης, σύμφωνα με τον Fodor, μπορεί να οριστεί ως μια

⁵³ Παγωνδιώτης Κ., [2001], σσ. 75-76

⁵⁴ Παγωνδιώτης, Κ., [1999], σ. 65

λειτουργία, η οποία υλοποιείται υπολογιστικά σε ένα αναπαραστασιακό σύστημα.⁵⁵ Η θεωρία αυτή υιοθετεί, κατά πρώτο λόγο, το μοντέλο του υπολογισμού⁵⁶, το οποίο, σύμφωνα με τον Fodor, είναι το καταλληλότερο για να παραστήσει τη λειτουργική σχέση μεταξύ του ανθρώπινου οργανισμού και κάθε νοητικής αναπαράστασης. Η αναπαράσταση αυτή μπορεί, λοιπόν, να συνδέσει το υποκείμενο με τον κόσμο με έναν τρόπο τυπικό στο πλαίσιο ενός υπολογιστικού συστήματος. Η σύνδεση αυτή συνιστά μια λειτουργία, η οποία επιτυγχάνεται μέσω του μοντέλου της γλώσσας.

5. ΦΙΛΟΣΟΦΙΚΕΣ ΠΑΡΑΔΟΧΕΣ ΚΑΙ ΜΕΘΟΛΟΓΙΑ ΤΗΣ ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ

Η Γνωσιακή Επιστήμη, με βάση τα παραπάνω, έλκεται από την παραδοχή ότι οι νοητικές λειτουργίες είναι υπολογιστικές: Η βασική ιδέα του υπολογιστικού μοντέλου για τη νόηση είναι ότι η νόηση αποτελεί το πρόγραμμα (software⁵⁷) και ο ανθρώπινος εγκέφαλος αποτελεί τη ‘συσκευή’, το υλισμικό (hardware) ενός υπολογιστικού συστήματος.⁵⁸ Η νόηση λοιπόν είναι για τον εγκέφαλο ό,τι είναι το πρόγραμμα για τον ηλεκτρονικό υπολογιστή. Σύμφωνα με τη Γνωσιακή Επιστήμη, μέσω της μελέτης ενός υπολογιστικού συστήματος μπορούμε να εξηγήσουμε την ανθρώπινη νόηση.

Για τη Γνωσιακή Επιστήμη δεν παρουσιάζει ιδιαίτερο ενδιαφέρον ούτε η μελέτη του εγκεφάλου αυτού καθαυτόν, ούτε η σπουδή της συνείδησης αυτής καθαυτήν. Βέβαια, οι μηχανισμοί της γνώσης που

⁵⁵ Fodor, J., [1975], σσ. 27-29

⁵⁶ Βλ. Fodor, J., [1975], σ. 99: “Τα διαθέσιμα μοντέλα της γνωστικής διαδικασίας την χαρακτηρίζουν ως θεμελιωδώς υπολογιστική. Γι’ αυτό το λόγο προϋποτίθεται ένα αναπαραστασιακό σύστημα όπου λαμβάνει χώρα ο υπολογισμός.”

⁵⁷ Το λογισμικό (software), είναι “η επινοήση γλωσσών προγραμματισμού και η κατάστρωση προγραμμάτων, με στόχο τη λειτουργική ρύθμιση της μηχανής προς επιτέλεση καλά ορισμένου έργου”. (Μαραγκός, [1996], σ. 111)

⁵⁸ Η άποψη αυτή αναπτύχθηκε και υποστηρίχθηκε από τον A. Turing [1950], Feigenbaum και Feldman [1963], από τον Block [1990] κ.ά.

ενδιαφέρουν την Γνωσιακή Επιστήμη εφαρμόζονται στον εγκέφαλο και σίγουρα βρίσκουν ένα πεδίο έκφρασης στη συνείδηση, αλλά εκείνο που ενδιαφέρει τους εκπροσώπους της είναι το ενδιάμεσο επίπεδο, όπου οι πραγματικές γνωστικές διεργασίες είναι απρόσιτες στη συνείδηση. Επιπλέον, όλες αυτές οι διαδικασίες είναι υπολογιστικές. Η βασική υπόθεση πίσω από τη γνωστική επιστήμη είναι το ότι ο εγκέφαλος αποτελεί έναν υπολογιστή και ότι οι νοητικές διεργασίες είναι υπολογιστικές. Οι γνωστικές νοητικές διεργασίες είναι ασυνείδητες· ως επί το πλείστον είναι ασυνείδητες κατ' αρχήν· και είναι υπολογιστικές.⁵⁹

Αυτό που πρέπει να επισημάνουμε, είναι το γεγονός ότι το υλικό πεδίο εφαρμογής του υπολογισμού (δηλαδή το υλισμικό) δεν έχει σημασία για τη Γνωσιακή Επιστήμη. Μπορεί το πεδίο αυτό να συγκροτείται από νευρώνες, σιλικόνη, μηχανικά καλώδια ή οτιδήποτε άλλο. Μπορεί αυτές οι διεργασίες να εφαρμόζονται στην πραγματικότητα στον εγκέφαλο, θα μπορούσαν, όμως, να βρουν εφαρμογή και σε αναρίθμητα άλλα συστήματα. Το ότι βρίσκουν εφαρμογή στον εγκέφαλό μας είναι χωρίς ιδιαίτερη σημασία.

Το εγχείρημα της Γνωσιακής Επιστήμης, όπως είδαμε παραπάνω, διαφέρει από το εγχείρημα άλλων επιστημών ως προς τη μέθοδο προσέγγισης των νοητικών φαινομένων, τα οποία θεωρεί ότι μπορούν να εξηγηθούν ως υπολογιστικά προγράμματα (software) σε λειτουργικό επίπεδο. Αυτό σημαίνει ότι οι νοητικές λειτουργίες και όλα εν γένει τα νοητικά φαινόμενα δεν μπορούν να εξηγηθούν με όρους προθεσιακούς ή νοητικούς, δηλαδή σε επίπεδο νοήματος, αλλά με όρους λειτουργικούς/ τυπικούς. Από το άλλο μέρος, η Γνωσιακή Επιστήμη δεν εξηγεί κατ' ανάγκη τα νοητικά φαινόμενα ανάγοντάς τα σε υλισμικό (δηλαδή φυσικαλιστικό) επίπεδο.⁶⁰

⁵⁹ Η εξήγηση αυτή προκύπτει από το έργο του Turing [1950], του Putnam [1967] και του Fodor [1975]

⁶⁰ Μπορεί κάποιος γνωσιακός επιστήμονας να θεωρεί ότι το λειτουργικό επίπεδο μπορεί, εν τέλει, να αναχθεί περαιτέρω στο υλικό-νευροφυσιολογικό (του εγκεφάλου) επίπεδο.

Αποφεύγοντας γενικά αυτήν την αναγωγή στο υλισμικό πεδίο, οι οπαδοί της Γνωσιακής Επιστήμης αρνούνται ότι το πρόγραμμά τους είναι αναγωγιστικό, αφού ο νους δεν ανάγεται σε φυσικαλιστικό επίπεδο⁶¹. Ωστόσο, το πρόγραμμα αυτό, αφού εξηγεί τα νοητικά φαινόμενα ανάγοντάς τα σε ένα ενδιάμεσο του νοητικού και του φυσικαλιστικού επίπεδο, δηλαδή στο λειτουργικό, σημαίνει ότι τελικά προβαίνει σε κάποιου είδους αναγωγή. Αυτό συμβαίνει γιατί θεωρεί πως τα νοητικά φαινόμενα δεν είναι τίποτε άλλο από υπολογιστικά φαινόμενα, τα οποία λειτουργούν καθαρά μηχανικά, και εξισώνονται με τη λειτουργία τους. Έτσι, αυτό που απαιτείται, ώστε να εξηγηθούν τα φαινόμενα αυτά, είναι να αναχθούν σε τυπικά, άρα να εξηγηθούν οι λειτουργίες τους.

Κατόπιν τούτου, μπορεί εύλογα να υποστηριχθεί ότι η Γνωσιακή Επιστήμη στηρίζει τις παραδοχές της, σχετικά με την φύση και την ερμηνεία της νόησης, στο μεθοδολογικό πρόγραμμα του *αναγωγισμού* (*reductionism*), το οποίο, σε γενικές γραμμές, αρνείται την ύπαρξη μη αναγώγιμων, εγγενών νοητικών φαινομένων. Ο αναγωγισμός είναι μια μέθοδος, η οποία στηρίζεται στην ιδέα ότι “ορισμένα πράγματα είναι δυνατό ν’ αποδειχτεί πως δεν είναι τίποτε άλλο παρά κάποια άλλα πράγματα.”⁶² Όπως φαίνεται, λοιπόν, ο αναγωγισμός παραπέμπει σε μια σχέση ταυτότητας, όπου τα πράγματα Α μιας ομάδας είναι δυνατόν να αναχθούν σε πράγματα μιας άλλης ομάδας Β, αν και μόνο αν τα Α δεν είναι τίποτε άλλο από Β. Για την φυσική επιστήμη, για παράδειγμα, τα χρώματα (Α) δεν είναι τίποτε άλλο από ορισμένες συχνότητες (Β), με αποτέλεσμα τα χρώματα να ανάγονται στις συχνότητες αυτές. Σύμφωνα με τα παραπάνω, γίνεται φανερό πως η συνθετότητα των αναγόμενων φαινομένων (π.χ. χρώμα) βρίσκεται σε υψηλότερο επίπεδο από το επίπεδο συνθετότητας των αναγόντων (π.χ.

⁶¹ Fodor, J., [1975], σσ. 9-10

⁶² Searle, J., [1997] σ. 136

συχνότητα). Ουσιαστικά, στην αναγωγή το όλο εξηγείται από τα μέρη και το σύνθετο από το απλό.

Το συχνότερα απαντώμενο είδος αναγωγής είναι η οντολογική αναγωγή: “Είναι η αναγωγή όπου τα αντικείμενα ορισμένων τύπων μπορεί να αποδειχτεί ότι δεν είναι παρά αντικείμενα άλλων τύπων.”⁶³ Για παράδειγμα, τα στερεά ή υγρά σώματα μπορεί να αποδειχθεί ότι δεν είναι τίποτε άλλο παρά αποκλειστικά και μόνο συγκεντρώσεις μορίων ύλης, ο ήχος μπορεί να αποδειχθεί ότι δεν είναι τίποτε άλλο από κύματα με ορισμένη συχνότητα, το φως μπορεί να αποδειχθεί ότι δεν είναι τίποτε άλλο από κύματα, τα γονίδια μπορεί να αποδειχθεί ότι δεν είναι τίποτε άλλο από μόρια DNA κ.ά.

Η Γνωσιακή Επιστήμη χρησιμοποιεί την οντολογική, σε συνδυασμό με την αιτιώδη αναγωγή, η οποία αφορά σχέσεις δύο οποιουδήποτε τύπου πραγμάτων “που μπορεί να έχουν αιτιώδεις δυνατότητες, όπου η ύπαρξη και a fortiori οι αιτιώδεις δυνατότητες της αναγόμενης οντότητας αποδεικνύονται πλήρως εξηγήσιμες, με βάση τις αιτιώδεις δυνατότητες των αναγόντων φαινομένων.”⁶⁴ Έτσι, για παράδειγμα, ορισμένα σώματα είναι υγρά και αυτό το γεγονός έχει αιτιώδεις συνέπειες: Τα υγρά σώματα δεν έχουν συγκεκριμένο σχήμα και μπορούν να πάρουν το σχήμα του δοχείου στο οποίο βρίσκονται, δεν είναι αδιαπέραστα από άλλα αντικείμενα κτλ. Αυτές οι δυνατότητες μπορούν να εξηγηθούν αιτιωδώς από τις αιτιώδεις δυνατότητες των παλμικών κινήσεων των μορίων των υγρών σωμάτων.

Η Γνωσιακή Επιστήμη, λοιπόν, για να εξηγήσει τα νοητικά φαινόμενα, τα ανάγει σε λειτουργικά. Καταρχήν, η Γνωσιακή Επιστήμη θεωρεί τη νόηση, τη συνείδηση, την προθετικότητα και όλα τα νοητικά φαινόμενα ως σύνθετα φαινόμενα (ή ως το όλο), τα οποία μπορούν να αναχθούν οντολογικά σε απλούστερα φαινόμενα (ή σε

⁶³ Searle, J., [1997] σ. 136

⁶⁴ Searle, J., [1997], σσ. 137-138

μέρη). Το απλό φαινόμενο, στο οποίο θα αναχθεί η νόηση, δεν είναι κανένα άλλο από την λειτουργία. Η βασική ιδέα, που η Γνωστική Επιστήμη θέτει ως προϋπόθεση, για να κάνει αυτήν την αναγωγή, είναι ότι τα νοητικά φαινόμενα μπορούν να εξηγηθούν κατά έναν ικανοποιητικότερο τρόπο, εάν καταστεί δυνατό να αναχθούν σε μια βασική εμπειρική θεωρία. Η φύση της θεωρίας αυτής εξασφαλίζεται από το γεγονός ότι χαρακτηρίζει πραγματικούς αιτιακούς μηχανισμούς.

Η Γνωστική Επιστήμη μεθοδολογικά χρησιμοποιεί (μέσω του συμπεριφορισμού) τον αιτιώδη αναγωγισμό για να οδηγηθεί τελικά στον οντολογικό αναγωγισμό. Η γενική αρχή στις περιπτώσεις αυτές φαίνεται να είναι η εξής: από τη στιγμή που μια ιδιότητα κάποιων αντικειμένων (λ.χ. η αδιαπερατότητα των στερεών) αποτελεί συνέπεια κάποιων ιδιοτήτων των απλούστερων δομικών μερών τους (λ.χ. των μορίων τους), έχουμε αυτόματα μιαν αιτιώδη αναγωγή ('αιτία της αδιαπερατότητας των στερεών είναι η αντίστοιχη ιδιότητα των μορίων τους') κι αυτό οδηγεί σε οντολογική αναγωγή (λ.χ. το στερεό είναι το σύνολο των μορίων από τα οποία αποτελείται). Κατ' αναλογία, αφού οι τυπικές / λειτουργικές σχέσεις μεταξύ των νευρώνων, των κυττάρων και του εγκεφάλου είναι η αιτία των νοητικών φαινομένων (αιτιώδης αναγωγή), τότε οι λειτουργικές αυτές σχέσεις θα πρέπει να είναι τα ίδια τα νοητικά φαινόμενα.

- Για το οικοδόμημα αυτό της Γνωστικής Επιστήμης πολλοί φιλόσοφοι διατύπωσαν τις αντιρρήσεις τους και εξέθεσαν επιχειρηματολογία, η οποία ασκεί κριτική στη δυνατότητά της να εξηγήσει επιστημονικά τη νόηση. Από αυτούς, ο J. Searle πρόβαλλε ένα νοητικό πείραμα, το οποίο είναι γνωστό ως το 'πείραμα του κινέζικου δωματίου'. Με βάση το πείραμα αυτό, θα προσπαθήσουμε να ανασυγκροτήσουμε το μοντέλο της Γνωστικής Επιστήμης και να αποτιμήσουμε την εξηγητική του εμβέλεια.

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ: ΑΝΑΣΥΓΚΡΟΤΗΣΗ ΚΑΙ ΚΡΙΤΙΚΗ ΤΟΥ
ΜΟΝΤΕΛΟΥ ΤΗΣ ΓΝΩΣΙΑΚΗΣ ΕΠΙΣΤΗΜΗΣ

1. “ΤΟ ΚΙΝΕΖΙΚΟ ΔΩΜΑΤΙΟ”

α. Εισαγωγή: Η “Τεχνητή Νοημοσύνη”, όπως αναφέρθηκε προηγουμένως, είναι μια προσπάθεια να προσομοιωθούν κάποια φαινόμενα (βιολογικά, φυσικά κ.ά.) σε προγράμματα ηλεκτρονικών υπολογιστών. Τέτοιου είδους προσομοίωση μπορεί να γίνει, για παράδειγμα, για τα μετεωρολογικά φαινόμενα, το χρηματιστήριο κτλ. Την παραπάνω εκδοχή της τεχνητής νοημοσύνης ο Searle την αποκαλεί ‘Ασθενή Τεχνητή Νοημοσύνη’ (Weak AI)⁶⁵ και την αποδέχεται. Αντιτίθεται, όμως, στην “Ισχυρή Τεχνητή Νοημοσύνη (ITN)” (Strong AI), η οποία πρεσβεύει ότι “ένας υπολογιστής θα μπορούσε να έχει – και πρέπει να θεωρηθεί πως έχει στην πραγματικότητα – σκέψεις, αισθήματα και νοητικές ικανότητες, απλώς και μόνο επειδή εκτελεί ένα κατάλληλο πρόγραμμα με τα κατάλληλα δεδομένα και αποτελέσματα”.⁶⁶ Ουσιαστικά, λοιπόν, η ITN είναι ένα ερευνητικό πεδίο που έχει ως στόχο την προσομοίωση της νόησης σε ψηφιακούς υπολογιστές. Το πεδίο αυτό αποδέχεται πως η σκέψη μπορεί να ταυτιστεί με υπολογισμό, μέσω των λειτουργιών της, και ότι η σκέψη είναι υπολογισμός. Εκπρόσωποι της ITN είναι οι H.A. Simon και A. Newell⁶⁷, ο R.C. Schank⁶⁸ κ.ά.

Ο Searle, με το νοητικό πείραμα του κινέζικου δωματίου, επιχειρεί να ανασκευάσει τις απόψεις όσων θεωρούν ότι οι γνωσιακές ικανότητες του ανθρώπου είναι δυνατό να ταυτιστούν με τις λειτουργίες ενός ηλεκτρονικού υπολογιστή. Η επιχειρηματολογία του Searle έρχεται να κρίνει το πρόγραμμα του Schank (1977), το οποίο επιδιώκει να προσομοιώσει την ικανότητα των ανθρώπων να κατανοούν ιστορίες και να δίνουν απαντήσεις με βάση το περιεχόμενο

⁶⁵ Searle, J., [1997], σ. 28

⁶⁶ Searle, J., [1997], σ. 23

⁶⁷ Newell, A.J.C. Shaw και Simon, H.A., [1963], “Chess-Playing Programs and the Problem of Complexity”, στο *Computers and Thought*, Edward A. Feigenbaum Julian Feldman (επιμ.). (Νέα Υόρκη: McGraw-Hill.) και Allen Newell, J.C. Shaw και H.A. Simon, [1964], “Elements of a Theory of Human Problem Solving”, στο Harper, R.J.C. et al.: *The Cognitive Process. Readings, Englewood Cliffs*, N.J.

⁶⁸ Οι απόψεις αυτές προκύπτουν από το έργο του Schank, R.C., et al., [1977], Panel on Natural Language Processing (IJCAI-77, *Proceedings*)

των ιστοριών αυτών. Δηλώνει ότι θα προσπαθήσει να αναιρέσει τις δύο βασικές θέσεις της ITN, οι οποίες εκτίθενται στο έργο του Schank και οι οποίες είναι πως: α) ο υπολογιστής, εάν είναι κατάλληλα προγραμματισμένος, διαθέτει *κυριολεκτικά* γνωσιακές καταστάσεις και β) τα προγράμματα αυτά εξηγούν τη γνωσιακή⁶⁹ ικανότητα του ανθρώπου.

Προκειμένου ο Searle να ελέγξει και να ανασκευάσει τις παραπάνω θέσεις, προβαίνει σε ένα νοητικό πείραμα. Πρόκειται για στρατηγική απόδειξης που μεθοδολογικά βρίσκεται στον αντίποδα των εμπειρικών επιστημών, διότι δεν γίνεται με εμπειρικά δεδομένα, αλλά με έννοιες. Εφαρμόζοντας αυτό το πείραμα δεχόμαστε τις θέσεις τις οποίες πρόκειται να ελέγξουμε και αναζητούμε τις συνέπειές τους. Εάν καταλήξουμε σε συμπέρασμα ασυνάρτητο και άτοπο, θα έχουμε τελικά αναιρέσει την αρχική θέση.

Το ερώτημα λοιπόν του Searle είναι “πώς θα είχαν τα πράγματα αν όντως η νόησή μας λειτουργεί βάσει των αρχών που η Γνωσιακή Επιστήμη θέτει ως θεμέλιο για τη λειτουργία όλων των νοήσεων;”⁷⁰

β. Το πείραμα: Έστω ότι κάποιο άτομο βρίσκεται κλεισμένο σε ένα δωμάτιο με ένα φύλλο χαρτί γεμάτο κινέζικα ιδεογράμματα. Το άτομο αυτό δεν γνωρίζει καθόλου κινέζικα, ούτε να γράφει, ούτε να ομιλεί και ούτε καν μπορεί να αναγνωρίσει την κινέζικη γραφή από οποιαδήποτε άλλη. Για το άτομο αυτό, τα κινέζικα ιδεογράμματα είναι ‘ορνιθοσκαλίσματα’ χωρίς νόημα. Έστω ότι, κατόπιν, στο άτομο αυτό δίνεται ένα δεύτερο φύλλο χαρτί στα κινέζικα, καθώς και ένα σύνολο κανόνων γραμμένων στην μητρική γλώσσα του ατόμου, προκειμένου να συσχετίσει το δεύτερο χαρτί με το πρώτο. Οι κανόνες αυτοί είναι

⁶⁹ Ο μεταφραστής του κειμένου του Searle αποδίδει τον όρο “cognitive” με τον όρο “γνωστική”. Ωστόσο, θεωρούμε πως η καλύτερη απόδοσή του είναι “γνωσιακή” και αυτήν την μετάφραση ακολουθούμε εδώ.

⁷⁰ Searle, J., [1993], σ. 409

κατανοητοί από το άτομο και του επιτρέπουν να συσχετίσει το σύνολο του πρώτου χαρτιού (ανερμήνευτα σύμβολα) με τα κινέζικα του δεύτερου (ανερμήνευτα σύμβολα). Η συσχέτιση αυτή θα γίνει καθαρά τυπικά, δηλαδή θα στηρίζεται στο ότι τα σύμβολα είναι αναγνωρίσιμα μόνο από τη μορφή τους. Στη συνέχεια, δίνεται στο άτομο ένα τρίτο χαρτί γραμμένο στα κινέζικα μαζί με κάποιες οδηγίες, πάλι στη μητρική γλώσσα του ατόμου, οι οποίες του επιτρέπουν να συσχετίσει τα σύμβολα του τρίτου χαρτιού με τα σύμβολα των δύο πρώτων. Οι κανόνες αυτοί θα βοηθήσουν το άτομο να σχεδιάσει το ίδιο κάποια κινέζικα σύμβολα με ορισμένο σχήμα ως 'απάντηση' σε ορισμένους τύπους σχημάτων που του δόθηκαν στο τρίτο φύλλο χαρτιού (ένα σχήμα 'κάπως έτσι', ακολουθείται από ένα σχήμα 'κάπως αλλιώς').

Αυτοί που προμηθεύουν στο άτομο τα χαρτιά με τα σύμβολα και τις οδηγίες ονομάζουν το περιεχόμενο του πρώτου χαρτιού 'γραφή', του δεύτερου 'ιστορία', του τρίτου 'ερωτήσεις', του χαρτιού που δίνει το άτομο με το συσχετισμό συμβόλων 'απαντήσεις στις ερωτήσεις' και τις οδηγίες στη μητρική γλώσσα 'πρόγραμμα'. Όλα αυτά, βεβαίως, το έγκλειστο άτομο τα αγνοεί.

Ας υποθεθεί, ακόμη, ότι το άτομο που βρίσκεται κλεισμένο στο δωμάτιο γίνεται πολύ επιδέξιο στο χειρισμό των κινέζικων συμβόλων και οι προγραμματιστές πολύ επιδέξιοι στο να γράφουν προγράμματα, ώστε, από μια εξωτερική οπτική γωνία, οι απαντήσεις, οι οποίες δίνονται στις ερωτήσεις, να φαίνονται εντελώς όμοιες με αυτές που θα έδινε κάποιο άτομο, το οποίο είχε ως μητρική γλώσσα τα κινέζικα.

γ. Το επιχείρημα: Πριν να προχωρήσουμε στο επιχείρημα του Searle, θα πρέπει να αποσαφηνίσουμε τα εξής: Οι νοητικές καταστάσεις, σύμφωνα με τη φιλοσοφία του νου χωρίζονται σε δύο κατηγορίες, δηλαδή στα *αισθήματα* (π.χ. πόνος, πείνα κτλ.) και στις *προτασιακές στάσεις*, δηλαδή τις προθέσεις, τις πεποιθήσεις και τις

επιθυμίες⁷¹. Οι προτασιακές στάσεις έχουν με τη σειρά τους τρία εγγενή χαρακτηριστικά: Τη *σύνταξη*, την *πραγματολογία* και τη *σημασιολογία*. Ο νους διέπεται από σύνταξη, δηλαδή έχει μια τυπική/συντακτική δομή και τα σύμβολά του ακολουθούν ορισμένους λογικούς κανόνες. Η πραγματολογία, από την άλλη μεριά, αφορά το πώς χρησιμοποιεί κανείς μια προτασιακή στάση, τι κάνει με αυτήν. Όλες οι προτασιακές στάσεις έχουν χρήστες. Τέλος, οι νοητικές καταστάσεις διαθέτουν σημασιολογία, δηλαδή ανήκουν σε αντικείμενα και καταστάσεις του κόσμου διαφορετικές από αυτές τις ίδιες και αφορούν τον εμπειρικό κόσμο. Εδώ, επίσης, εντάσσονται οι όροι ‘*προθετικότητα*’ και ‘*αναφορικότητα*’. Με τον όρο σημασιολογία, εννοούμε ότι οι εσωτερικές νοητικές λειτουργίες έχουν ένα νοηματικό περιεχόμενο, δηλαδή αναφέρονται σε κάτι. Όταν, για παράδειγμα, σκεφτόμαστε τη θάλασσα ή ένα ποτήρι κρασί, οι σκέψεις μας στρέφονται γύρω από κάτι, διαθέτουν ένα νόημα. Αυτό το κάτι, που βρίσκεται πέρα από την τυπική δομή του εγκεφάλου είναι η σημασία, δηλαδή το περιεχόμενο του νου. Με άλλα λόγια, ακόμη κι αν ο νους διαχειρίζεται σύμβολα, τα σύμβολα αυτά έχουν πάντοτε περιεχόμενο.

Οι κανόνες που έχουν δοθεί στο άτομο καθορίζουν τον χειρισμό των συμβόλων με τυπικό τρόπο. Το άτομο που βρίσκεται στο δωμάτιο, βεβαίως, δεν κατανοεί τα κινέζικα με τον ίδιο τρόπο που τα κατανοούν οι Κινέζοι και για το ίδιο τα κινέζικα σύμβολα δεν έχουν κανένα νόημα, είναι δηλαδή ανερμήνευτα. Ακόμη κι αν το άτομο γίνει πολύ καλό στο χειρισμό των συμβόλων και κατορθώσει να δίνει ‘απαντήσεις’ τόσο καλά όσο και ένας Κινέζος, πάλι τα σύμβολα που χειρίζεται δεν θα έχουν αναφορά ούτε ερμηνεία, επομένως ούτε και περιεχόμενο. Με άλλα λόγια, δεν θα καταστεί δυνατό για το άτομο να κατανοήσει τα σύμβολα ή τη σημασία τους. Είναι, λοιπόν, φανερό πως οι απαντήσεις οι οποίες δίνονται είναι ανερμήνευτες και, φυσικά,

⁷¹ Goldam, A.I., [1997], σσ. 99-100

δίχως νόημα και σημασιολογία, αφού το άτομο δε γνωρίζει την αναφορά τους. Αυτό συμβαίνει γιατί, τα νοητικά φαινόμενα εξηγούνται σε τυπικό/ συντακτικό επίπεδο, με αποτέλεσμα το νοητικό περιεχόμενό τους (ειδητικά γνωρίσματα του οποίου είναι κατεξοχήν η αναφορικότητα και η σημασιολογία τους) να μην εμπλέκεται καθόλου στην εξήγησή τους.

Όσον αφορά τα κινέζικα, το άτομο συμπεριφέρεται ακριβώς όπως ένας ηλεκτρονικός υπολογιστής: Εκτελεί διεργασίες υπολογισμού πάνω σε στοιχεία που έχουν προσδιοριστεί με τυπικό τρόπο, άρα το άτομο αυτό αποτελεί απλώς μια εξατομίκευση του προγράμματος του υπολογιστή. Όπως φάνηκε από τα παραπάνω, οι είσοδοι και οι έξοδοι του ατόμου είναι όμοιες με αυτές ενός άλλου ατόμου, το οποίο γνωρίζει τα κινέζικα. Κι ενώ το πρώτο άτομο θα μπορούσε να έχει το πιο εξελιγμένο πρόγραμμα, θα εξακολουθούσε να μην καταλαβαίνει τίποτε, ακόμη κι αν οι παρατηρητές είχαν την αντίθετη άποψη. Με άλλα λόγια, δεν υπάρχει κάποιο *κριτήριο* που θα διαχωρίζει το πραγματικό 'συμβάν' της νόησης, ανεξάρτητα απ' την οπτική γωνία του παρατηρητή.

Οι παρατηρητές κάθε άλλου φαινομένου, όπως λόγου χάρι οι φυσικοί ή οι γιατροί, εκ των πραγμάτων, είναι σε θέση να παρατηρήσουν και να μελετήσουν μόνο την εξωτερική ή φυσική 'συμπεριφορά' του συστήματος, το οποίο μελετούν. Τα νοητικά φαινόμενα, όμως, παρουσιάζουν μια ιδιαιτερότητα, η οποία δεν μας επιτρέπει να αρκεστούμε μόνο στην παρατήρηση της συμπεριφοράς: Κάθε νοητικό φαινόμενο, κάθε πεποίθηση και κάθε εσωτερική κατάσταση είναι πάντοτε κάποιου ατόμου και αναφέρεται σε κάτι. Από την οπτική γωνία ενός τρίτου προσώπου, δεν είναι δυνατόν να προσεγγιστούν ή να μελετηθούν αυτές οι υποκειμενικές νοητικές διεργασίες, ούτε και η υποκειμενικότητά τους. Η οπτική γωνία του τρίτου προσώπου αδυνατεί να προσεγγίσει αυτή την 'εσωτερικότητα' των νοητικών καταστάσεων, με αποτέλεσμα να αποκλείει το

υποκειμενικό στοιχείο τους και να παραγνωρίζει, τελικά, ένα από τα κύρια ζητήματα της έρευνας, δηλαδή το ερώτημα υπό ποια έννοια και πώς κατανοεί κάθε άτομο.

Ας δούμε τι συμβαίνει εάν στη θέση του ατόμου τοποθετήσουμε έναν υπολογιστή: Το ίδιο τυπικό πρόγραμμα χρήσης κινέζικων συμβόλων χωρίς περιεχόμενο και πραγματολογία, το οποίο διαθέτει το άτομο στο δωμάτιο, διαθέτει και οποιοσδήποτε υπολογιστής. Η λειτουργία του ηλεκτρονικού υπολογιστή εντοπίζεται στο να υλοποιεί προγράμματα, τα οποία ορίζονται με τυπικό τρόπο. Επομένως, οποιοσδήποτε υπολογιστής διαθέτει μόνο τυπικά σύμβολα. Και τα τυπικά σύμβολα, σε όποια 'γλώσσα' κι αν είναι 'γραμμένα' ως υπολογιστικό πρόγραμμα, δεν έχουν περιεχόμενο, είναι δηλαδή χωρίς αναφορά και χωρίς σημασιολογία.

Η υλοποίηση ενός τυπικού υπολογιστικού προγράμματος μπορεί, από τη σκοπιά ενός εξωτερικού παρατηρητή, να σημαίνει ότι ο υπολογιστής διαθέτει νοητικές καταστάσεις, αλλά στην πραγματικότητα στερείται νοήματος και αναφορικότητας. Αυτό συμβαίνει διότι, στην περίπτωση των κινέζικων, ο υπολογιστής είναι το ίδιο το άτομο, ενώ στην περίπτωση όπου ο υπολογιστής δεν είναι το άτομο, δε θα έχει τίποτε περισσότερο από ό,τι έχει αυτό, όταν βρίσκεται στην κατάσταση όπου δεν καταλαβαίνει τίποτε.⁷² Ο Searle, λέγοντας ότι ο υπολογιστής δεν έχει αυτό που έχουμε εμείς όταν κατανοούμε, αναφέρεται, όπως φάνηκε παραπάνω, στο ότι από τον υπολογιστή και γενικότερα από το λειτουργικό μοντέλο εξήγησης των νοητικών φαινομένων, λείπει η αναφορικότητα και η σημασιολογία. Όταν χρησιμοποιούμε τον όρο 'κατανοώ', εννοούμε ότι κατανοούμε το νόημα. Προκειμένου, όμως, να νοηματοδοτηθούν τα σύμβολα του υπολογιστικού προγράμματος, πρέπει να κατανοηθούν, πράγμα που δεν συμβαίνει στην περίπτωση του υπολογιστή, καθώς η κατανόηση προϋποθέτει την κατανόηση της σημασιολογικής αναφοράς των

⁷² Searle, J., [1993], σ. 410

συμβόλων. Από αυτήν την οπτική, η θέση της ITN βρίσκεται σε γνωσιολογικό αδιέξοδο: ο ηλεκτρονικός υπολογιστής δεν είναι δυνατόν να διαθέτει γνωσιακές λειτουργίες.

Ωστόσο, ακόμη κι αν σ' έναν παρατηρητή φαίνεται ότι ο υπολογιστής κατανοεί, αυτό δεν σημαίνει ότι το σύστημα πράγματι το κάνει. Αυτό το οποίο τονίζει με το επιχείρημά του ο Searle είναι ότι ο υπολογιστής μπορεί να διαθέτει σύνταξη, δηλαδή να είναι σε θέση να ταξινομεί και να χειρίζεται τυπικά τα σύμβολα με τα οποία τον προμηθεύουν οι προγραμματιστές του (π.χ. ένα σύμβολο 'κάπως έτσι' ακολουθείται από ένα σύμβολο 'κάπως αλλιώς'). Ωστόσο δεν είναι σε θέση να κατανοεί τη σημασία αυτών των συμβόλων. Με άλλα λόγια, τα υπολογιστικά/ λειτουργικά προγράμματα στερούνται σημασιολογίας – και κατανόηση χωρίς σημασιολογία, και άρα χωρίς νόημα δεν μπορεί να υπάρξει. Από αυτήν την οπτική, η θέση της ITN έχει ανααιρεθεί: ο ηλεκτρονικός υπολογιστής δεν είναι δυνατόν να διαθέτει νοητικές/ γνωσιακές λειτουργίες.

Όσον αφορά, επίσης, στη δεύτερη θέση της ITN, ότι δηλαδή ο υπολογιστής διαθέτει κυριολεκτικά νόηση, βασισμένοι στο πείραμα του Searle, μπορούμε να υποστηρίξουμε ότι ο υπολογιστής και το πρόγραμμα δεν παρέχουν επαρκείς συνθήκες για να υπάρξει διαδικασία κατανόησης. Μπορεί βεβαίως, να υπάρχει λειτουργία που γίνεται με τυπικό τρόπο, άλλα δεν υπάρχει κατανόηση. Ακόμη κι αν για την κατανόηση των 'κινέζικων' δε χρειάζεται τίποτε παραπάνω από μια πιο εντατική μελέτη που θα οδηγήσει στον απρόσκοπτο χειρισμό των τυπικών συμβόλων, τίποτε δεν αποδεικνύει πως ισχύει το ίδιο για όλες τις γλώσσες ή για τους υπολογιστές. Ο Searle δέχεται ότι σε ένα υπολογιστικό πρόγραμμα μπορεί να πληρούνται οι όροι της ορθής σύνταξης⁷³, δηλαδή τα σύμβολα του προγράμματος να

⁷³ Αργότερα ο Searle (1990) και (1992/1997, σσ. 234-239) θεωρεί ότι τα προγράμματα των ηλεκτρονικών υπολογιστών δεν πληρούν ούτε την προϋπόθεση της ύπαρξης τυπικής/ συντακτικής σχέσης μεταξύ συμβόλων. Το λάθος που γίνεται από τους γνωσιακούς επιστήμονες εντοπίζεται στο ότι οι ίδιοι ερμηνεύουν ως συντακτικά τα φυσικά

σχετίζονται μεταξύ τους τυπικά, όμως, αυτό δεν μας επιτρέπει να ταυτίζουμε τη σύνταξη με το νόημα.

Εφόσον τα σύμβολα των υπολογιστών σχετίζονται μεταξύ τους τυπικά, αλλά δεν διαθέτουν νόημα, δεν μπορούμε να θεωρήσουμε επαρκές το υπολογιστικό μοντέλο για την εξήγηση της νόησης. Ας πάρουμε ως παράδειγμα, την περίπτωση όπου στο δωμάτιο οι ερωτήσεις δίνονταν στη μητρική γλώσσα του ατόμου και το άτομο απαντούσε σ' αυτήν την γλώσσα. Στην εκδοχή αυτή, η διαφορά με την περίπτωση των κινέζικων θα ήταν προφανής: Τα σύμβολα που θα χρησιμοποιούνταν στην δεύτερη περίπτωση, θα είχαν νόημα και περιεχόμενο για το άτομο, και το ίδιο θα παρήγαγε σύμβολα επίσης με νόημα και περιεχόμενο για το ίδιο. Η διάκριση, λοιπόν, μεταξύ νοητικού και μη νοητικού πρέπει να είναι εγγενής στο σύστημα, διότι μόνο σ' αυτήν την περίπτωση τα νοητικά φαινόμενα, εκτός από σύνταξη, έχουν και περιεχόμενο και πραγματολογία.

Ουσιαστικά, το 'πείραμα του κινέζικου δωματίου' αφήνει να διαφανεί πως η κατανόηση και η γνωστική ικανότητα του ανθρώπου δε σχετίζονται και ούτε μπορούν να εξηγηθούν επαρκώς και αναγκαίως μόνο με όρους συντακτικών/ υπολογιστικών διεργασιών. Η νόηση δεν έγκειται απλώς σε τυπικές λειτουργικές σχέσεις – τις οποίες, όπως έδειξε το πείραμα, μπορεί ένα νοήμον άτομο να τις ακολουθεί, χωρίς ωστόσο να τις καταλαβαίνει. Η θεωρία αυτή, εν τέλει, δεν καταφέρνει να εξηγήσει αυτό το οποίο επαγγέλλεται, δηλαδή το περιεχόμενο της νόησης και το νόημα, που την καθιστά τέτοια. Δεν μπορεί να εξηγήσει και να διακρίνει "τις αρχές στις οποίες βασίζεται η λειτουργία της νόησης από εκείνες στις οποίες βασίζεται η λειτουργία των μη νοητικών συστημάτων."⁷⁴ Τα χαρακτηριστικά του νου ο υπολογιστής μπορεί να τα προσομοιώσει ('μας φαίνεται ως εάν να

χαρακτηριστικά του υπολογιστή (και η σύνταξη δεν είναι κάτι εγγενές στη φυσική). Το πραγματικό πρόβλημα, σύμφωνα με τον Searle, είναι ότι οι σχέσεις μεταξύ των συμβόλων θεωρούνται συντακτικές, επειδή αυτόν τον ρόλο τους αποδίδουν οι παρατηρητές.

⁷⁴ Searle, J., [1993], σ. 416

κατανοεί’), αλλά δεν μπορεί να τα αναπαράγει με νοηματικά γονιμοποιημένο τρόπο. Και η προσομοίωση, βεβαίως, δεν είναι αναπαραγωγή.

Στο σημείο αυτό, σκόπιμο θα ήταν να διασαφηνιστεί και να οριοθετηθεί ο όρος ‘κατανόηση’, ο οποίος αποτελεί βασική έννοια της επιχειρηματολογίας του Searle. Ο Searle, στο πλαίσιο του πειράματός του, την ορίζει ως την “κατοχή νοητικών (προθετικών) καταστάσεων και την αλήθεια (εγκυρότητα, επιτυχία) τους”.⁷⁵ Συχνά, οι συνομιλητές του υποστηρίζουν ότι ο όρος αυτός δεν είναι μονοσήμαντος, αλλά μπορεί να δεχτεί διαβαθμίσεις και να διακριθεί σε διάφορα επίπεδα, ανάλογα με τη σημασία του. Έτσι, όταν αυτοί δηλώνουν πως ο υπολογιστής κατανοεί, εννοούν κάτι αόριστο και μη προσδιορισμένο εννοιολογικά. Θεωρούν, λοιπόν, ότι λύνουν το εννοιολογικό πρόβλημα που αντιμετωπίζουν με τη σχετικότητα που αποδίδουν στην έννοιά τους. Ωστόσο, είτε η λέξη κατανόηση μπορεί να έχει διαφορετικές νοηματικές αποχρώσεις είτε όχι, οι εκπρόσωποί της δεν μπορούν να την ορίσουν και κυριολεκτικά διατυπώνουν άποψη για κάτι το οποίο δε γνωρίζουν, γεγονός που υποσκάπτει την υπόθεση εργασίας τους.

2. ΚΡΙΤΙΚΗ

Το πρόγραμμα της ITN προτείνεται και εφαρμόζεται από τους γνωσιακούς επιστήμονες, οι οποίοι θεωρούν δεδομένα και προβλέψιμα τόσο τα αποτελέσματά του, όσο και την προσφορά του στην εξήγηση της νόησης. Ωστόσο, όπως ήδη διαφάνηκε, το πρόγραμμα αυτό χωλαίνει στην εννοιολογική του συγκρότηση.

Πρώτα πρώτα, θα πρέπει να παραθέσουμε τον διαχωρισμό που υπάρχει ανάμεσα στα πεδία της ‘οντολογίας’, της ‘επιστημολογίας’ και της ‘αιτιότητας’. Υπάρχει διαφορά ανάμεσα στις απαντήσεις που μπορούν να δοθούν στις ερωτήσεις: ‘Τι είναι αυτό;’ (οντολογία), ‘Πώς

⁷⁵ Searle, J., [1993], σ. 412

θα μάθουμε γι' αυτό;' (επιστημολογία) και 'Τι κάνει αυτό;' (αιτιότητα). Ας δούμε λοιπόν, υπό το φως της παραπάνω επισήμανσης, τι συμβαίνει με την Γνωσιακή Επιστήμη .

Από την επιστημολογική σκοπιά, ο νατουραλισμός της Γνωσιακής Επιστήμης θέτει έμμεσα **το πρόβλημα του ' τρίτου προσώπου'**, δηλαδή τις δυσκολίες που δημιουργεί η παρατήρηση των νοητικών φαινομένων από μια 'εξωτερική' σε σχέση με το άτομο οπτική γωνία, εκείνη δηλαδή του επιστήμονα-παρατηρητή: Τα ερωτήματα μετατοπίζονται από την υποκειμενικότητα των νοητικών καταστάσεων στην αντικειμενικότητα της εξωτερικής συμπεριφοράς και της παρατήρησής της, από την οπτική γωνία του τρίτου προσώπου. Η συνείδηση, και όλα εν γένει τα νοητικά φαινόμενα, ως εσώτερα φαινόμενα, περισσότερο αγνοούνται εντελώς παρά εξηγούνται υπό το φως του συμπεριφορικού και του λειτουργιστικού μοντέλου. Τα φαινόμενα αυτά είτε επαναπροσδιορίζονται είτε παραβλέπονται πλήρως και απογυμνώνονται από τα κύρια και εγγενή γνώρισμά τους: την υποκειμενικότητα, την εσωτερικότητα και την πρωτοπρόσωπη ποιοτική τους διαφοροποίηση .⁷⁶

Αυτό συμβαίνει γιατί οι γνωσιακοί επιστήμονες, φοβούμενοι μήπως υποπέσουν στο 'σφάλμα' του καρτεσιανού δυισμού και στον 'αντιεπιστημονικό' χαρακτήρα του, δηλαδή την απόδοση 'μυστηριωδών' ιδιοτήτων στο νοητικό, εξετάζουν τα νοητικά φαινόμενα πάντα υπό την οπτική του τρίτου προσώπου, με αποτέλεσμα να τα αντικειμενικοποιούν πλήρως, δηλαδή να τα αντιμετωπίζουν ως φυσικά φαινόμενα που παρατηρούνται από κάθε παρατηρητή. Ωστόσο, η οντολογία της νόησης εγγενώς είναι μια οντολογία πρώτου προσώπου, επομένως μια οντολογία πρωτίστως υποκειμενική. Τα νοητικά φαινόμενα συνδέονται κατά τρόπο ουσιαστικό με τη συνείδηση. Η συνείδηση, όμως, είναι στην ουσία της υποκειμενική, επομένως και η οντολογία του νοητικού είναι στην

⁷⁶ Searle, J., [1997], σ. 24

ουσία της υποκειμενική οντολογία. Οι νοητικές καταστάσεις είναι πάντοτε νοητικές καταστάσεις κάποιου συγκεκριμένου ατόμου.

Εξάλλου, και από την επιστημολογική σκοπιά, η άποψη του πρώτου προσώπου διαφέρει από την άποψη του τρίτου προσώπου. Στο σημείο αυτό, πρέπει να υπογραμμίσουμε ότι δεν θα πρέπει να συγχέεται η επιστημολογική με την οντολογική έννοια της υποκειμενικότητας: Για την εύρεση της αλήθειας είναι υποχρεωμένος κανείς να αποκλείει και να αποβάλει τις υποκειμενικές προκαταλήψεις του, ώστε να οδηγηθεί σε αντικειμενικά και έγκυρα συμπεράσματα. Διαφορετικό όμως (και εντελώς θεμιτό) είναι να διαπιστώσει κανείς ότι η αντικειμενική αλήθεια που θα βρει περιέχει μη αναγώγιμα υποκειμενικά στοιχεία. Η συνείδηση, λόγου χάρη, επιστημολογικά είναι μια αντικειμενική έννοια, η οποία οντολογικά είναι μη αναγώγιμη σε αντικειμενική παρατήρηση, από την οπτική του τρίτου προσώπου.

Τόσο ο συμπεριφορισμός όσο και ο λειτουργισμός συγχέουν την τριτοπρόσωπη παρατήρηση (την επιστημολογία) με τις υποκειμενικές διεργασίες που λαμβάνουν χώρα εντός του υποκειμένου (την οντολογία). Η ITN ταυτίζει οντολογικά αυτές τις δύο εξηγήσεις, με αποτέλεσμα να συγχέει την προσομοίωση του νου, δηλαδή το υπολογιστικό μοντέλο, με την ίδια τη νόηση. Το μοντέλο της ITN, και κατ' επέκταση της Γνωσιακής Επιστήμης, εξαιτίας της παραπάνω σύγχυσης, δεν επιχειρεί να εξηγήσει ποιο είναι το νόημα και το περιεχόμενο των νοητικών φαινομένων, πράγματα που είναι εγγενή στοιχεία της νόησης. Έτσι, βλέπουμε, σε τελική ανάλυση, να εξαφανίζεται αυθαιρέτως η νοητική ζωή, την οποία προγραμματικά το γνωσιακό μοντέλο επαγγέλλεται πως θα εξηγήσει.

Επιπλέον, η απλή παρατήρηση συμπεριφοράς-εκροών αδυνατεί να χρησιμεύσει στη Γνωσιακή Επιστήμη ως βάση για μια εμπειρική εξηγητική θεωρία του νου. Τα νοητικά φαινόμενα, από οντολογική άποψη, δεν μπορούν να σχετιστούν μόνο με τη συμπεριφορά. Θα ήταν

λάθος να ακολουθήσουμε τη λογική που λέει ‘ίδια συμπεριφορά, άρα, ίδια νοητικά φαινόμενα’. Αυτό συμβαίνει γιατί, όπως ήδη έδειξε το νοητικό πείραμα του Searle, ένα σύστημα μπορεί να παρουσιάζει μια ορισμένη συμπεριφορά ως εάν να κατανοούσε (πάντα σε σχέση με την οπτική γωνία του τρίτου προσώπου), χωρίς ωστόσο να κατανοεί τίποτε. Εξάλλου, η συμπεριφορά φαίνεται να μην είναι ταυτόσημη με τη συνείδηση, αφού ένα άτομο μπορεί να διαθέτει συνείδηση, αλλά τα νοητικά του φαινόμενα να μην εκδηλώνονται με τη μορφή συμπεριφοράς (στην ιατρική το σύνδρομο Guillain – Barré).⁷⁷ Από την άλλη μεριά, δυο διαφορετικά άτομα, εκ των οποίων το ένα δρα ενσυνείδητα, ενώ το άλλο αντανακλαστικά, θα μπορούσαν να παρουσιάζουν την ίδια συμπεριφορά.

Η συμπεριφορά, εξάλλου, δεν αποτελεί παρά την εκδήλωση μιας εσωτερικής νοητικής κατάστασης/ πραγματικότητας ‘τινός’ προσώπου. Όταν πρόκειται να εξηγήσουμε τη νόηση άλλων ατόμων, η συμπεριφορά αυτή καθαυτή δε μας αρκεί· τη βάση της εξήγησής μας, την αποτελεί ο συνδυασμός της συμπεριφοράς με τη γνώση των αιτιωδών λόγων και των νοητικών διεργασιών όπου στηρίζεται η συγκεκριμένη συμπεριφορά.

Γίνεται, λοιπόν, φανερό πως οι εσωτερικές νοητικές καταστάσεις δεν μπορούν να γνωσθούν μόνο με τη μελέτη της συμπεριφοράς, αφού αυτή δεν αποτελεί επαρκή συνθήκη για την εξήγηση της ύπαρξης της νόησης. Επομένως, όταν εξετάζουμε και θέλουμε να πιστοποιήσουμε την ύπαρξη και να εξηγήσουμε τη φύση των νοητικών φαινομένων και των νοητικών καταστάσεων ως τέτοιων, δεν είναι δεσμευτικό ούτε επαρκές να στηριζόμαστε στην εξωτερική συμπεριφορά.

Τα ίδια περίπου προβλήματα αντιμετωπίζει και ο **λειτουργισμός**, διότι, μπορεί να μην αγνοεί εντελώς την ύπαρξη και την ‘εσωτερικότητα’ των νοητικών φαινομένων, όμως ενδιαφέρεται μόνο για τις *σχέσεις* των ‘εισροών’, των ‘εσωτερικών καταστάσεων’ μεταξύ

⁷⁷ Searle, J., [1997], σσ. 89-90

τους και των ‘εκροών’ σε τυπικό/ συντακτικό και όχι σε νοηματικό επίπεδο. Για παράδειγμα, η έννοια ‘νοιώθω δίψα’, για τους λειτουργιστές, θα ερμηνευόταν με τη βοήθεια του εξής σχήματος: Αν ο Χ έχει μείνει πολλές ώρες χωρίς να πάρει υγρά (εισροή), θα έχει την τάση να διψά και να θέλει να πει (εσωτερική κατάσταση) και τελικά θα ανοίξει τη βρύση (εκροή) για να ικανοποιήσει τη δίψα του.⁷⁸

Ο λειτουργισμός, επομένως, αναγνωρίζει, όπως προαναφέραμε, την ύπαρξη εσωτερικών νοητικών καταστάσεων, όμως αδιαφορεί για τις ‘εγγενείς’ ποιότητές τους: Παρατηρεί και δέχεται ότι κάποιο άτομο έχει μια τάση/ πρόθεση/ πεποίθηση για ένα ζήτημα, αλλά δεν ενδιαφέρεται για το πώς είναι/ πώς αισθάνεται αυτό το άτομο, όταν διαθέτει αυτές τις νοητικές καταστάσεις. Ουσιαστικά, το λειτουργιστικό μοντέλο, καθώς επιχειρεί να εξηγήσει – με βάση πάντα την τριτοπρόσωπη παρατήρηση, της οποίας τις αδυναμίες έχουμε ήδη αναφέρει – όλες τις νοητικές καταστάσεις, τις παρακάμπτει με έντεχνο τρόπο, τις αρνείται σχεδόν προγραμματικά, και αφαιρεί από αυτές το ειδητικό γνώρισμά τους, που είναι η εσωτερικότητα και η υποκειμενικότητά τους.

Το οικοδόμημα της Γνωσιακής Επιστήμης αντιμετωπίζει, εκτός των άλλων, σοβαρά προβλήματα και στο επίπεδο της μεθοδολογίας του. Ο **αναγωγισμός**, που υιοθετείται ως μέθοδος της επιστημονικής εξήγησης των νοητικών φαινομένων, αντιμετωπίζει τις νοητικές διεργασίες (ανάμεσά τους και τη συνείδηση) ως φυσικές εμπειρικές καταστάσεις και τις μελετά όπως μελετά οποιοδήποτε άλλο φυσικό φαινόμενο. Κατά πόσο, όμως, η Γνωσιακή Επιστήμη μπορεί να αποδείξει την εγκυρότητα της μεθόδου αυτής για τη μελέτη του νου;

Οι αναγωγιστικές μεθοδολογίες λειτουργούν μόνο με την αποσιώπηση ή την παραγνώριση εγγενών χαρακτηριστικών του νου, όπως είναι η συνείδηση. Οι μεθοδολογίες αυτές, δηλαδή, παραλείπουν

⁷⁸ Goldman, A.I., [1998], σσ. 96-97

ενδιάμεσα συλλογιστικά βήματα: Προκειμένου να υλοποιηθεί οποιοδήποτε αναγωγιστικό πρόγραμμα, πρέπει να στηριχτεί στην εξήγηση/ ανάλυση αυτού, το οποίο θα αναχθεί. Στην αντίθετη περίπτωση, εάν, δηλαδή, κάποιο βήμα παραλειφθεί, τότε το πρόβλημα έχει τεθεί εσφαλμένα. (Εάν λ.χ. οι ιδιότητες των στερεών αντικειμένων πρόκειται να αναχθούν στις ιδιότητες των ατόμων από τα οποία αποτελούνται, τότε απαραίτητως πρέπει να εξηγηθεί εκ των προτέρων η ατομική φύση των στερεών.) Κατά τον ίδιο τρόπο, για την αναγωγή της νόησης σε τυπικό υπολογισμό, πρέπει εκ των προτέρων να εξηγηθούν η συνείδηση, η προθετικότητα, το περιεχόμενο και όλα εν γένει τα νοητικά φαινόμενα. Αντίθετα, η ITN ενώ δεν καταβάλλει καμία προσπάθεια να εξηγήσει τα χαρακτηριστικά της νόησης, διευρύνει την αναγωγή που κάνει και φτάνει τελικά να τα συμπεριλάβει στην τελική εξήγηση του νου ως λειτουργία.

Αντιθέτως, με τη μέθοδο της ITN, δεν υπάρχει τρόπος να μελετήσουμε τα φαινόμενα του νου, χωρίς, άμεσα ή έμμεσα, να μελετήσουμε τη **συνείδηση**. “Ο βασικός λόγος είναι ότι, στην πραγματικότητα, δεν έχουμε ιδέα γι’ αυτό που ονομάζουμε νοητικό, πνευματικό, παρά μόνο μέσω της έννοιας της συνείδησης που έχουμε διαμορφώσει”.⁷⁹ Ας δούμε κάποια από τα βασικά χαρακτηριστικά της συνείδησης με κριτήριο την απουσία τους, σε περιπτώσεις παθολογικών καταστάσεων:

Αρχικά, η συνείδηση εκδηλώνεται με *λογικούς τύπους αυστηρά καθορισμένους*: 1. Υπάρχουν οι πέντε αισθήσεις μας (όραση, αφή, όσφρηση, ακοή και γεύση), οι οποίες μπορούν να λειτουργούν ευχάριστα ή δυσάρεστα αναφορικά με κάτι. 2. Υπάρχει, επιπρόσθετα, η ‘αίσθηση της ισορροπίας’, καθώς και τα σωματικά αισθήματα: με τον όρο ‘σωματικά αισθήματα’, εννοούμε τα αισθήματα που συνδέονται φανερά με τη φυσιολογία του σώματος (λ.χ. οι σωματικοί πόνοι), αλλά και με την αισθητήρια αντίληψή των ανθρώπων για τη

⁷⁹ Searle, J., [1997], σσ. 35-36

θέση, π.χ., που έχουν τα χέρια τους και το κεφάλι τους για το τι αισθάνονται στο αριστερό τους γόνατο.⁸⁰ 3. Τέλος, λογικός τύπος της συνείδησης είναι και η ροή της σκέψης: Η ροή της σκέψης περιλαμβάνει λέξεις, παραστάσεις (οπτικές ή οτιδήποτε άλλο), αλλά και άλλα στοιχεία, μη λεκτικά ή μη εικονικά.

Η συνείδηση, επίσης, διακρίνεται για την *ενότητα* – *αλληλουχία* της. Η ενότητα αυτή εντοπίζεται σε δύο τουλάχιστον διαστάσεις, την ‘οριζόντια’ και την ‘κατακόρυφη’. “Η οριζόντια ενότητα είναι η οργάνωση των ενσυνείδητων εμπειριών μέσα από σύντομα χρονικά διαστήματα. Για παράδειγμα, όταν εκφέρω ή σκέφτομαι μια πρόταση, ακόμα κι αν είναι πολύ μακριά, η επίγνωση εκείνου που είπα ή σκέφτηκα στην αρχή συνεχίζεται, ακόμα και όταν έχω πάψει να εκφέρω ή να σκέφτομαι αυτό το πρώτο μέρος της πρότασης.”⁸¹ Από την άλλη μεριά, η κατακόρυφη ενότητα “είναι ζήτημα ταυτόχρονης επίγνωσης όλων των ποικίλων χαρακτηριστικών οποιασδήποτε ενσυνείδητης κατάστασης.”⁸²

Η συνείδηση, ακόμη, χαρακτηρίζεται (εκτός από τις περιπτώσεις που βρισκόμαστε σε μια κατάσταση, π.χ. λύπη, χωρίς να υπάρχει κάτι, το οποίο μας προκαλεί την συγκεκριμένη κατάσταση) από *αναφορικότητα/ προθετικότητα*. Η συνείδηση, με άλλα λόγια, αναφέρεται πάντα σε κάτι, ακόμη κι αν αυτό το κάτι δεν υπάρχει ως φυσικό αντικείμενο. Πάντα έχουμε συνείδηση του ενός ή του άλλου πράγματος, και αυτό το ‘του’, στην έκφραση ‘συνείδηση του’, είναι το ‘του’ της αναφορικότητας, εκείνου, δηλαδή, προς το οποίο κατευθύνεται η συνείδηση. Κάθε αναφορική κατάσταση εμπεριέχει μια εποπτική μορφή (*aspectual shape*), δηλαδή μια αναφορά ‘ως προς’. “Το να έχουμε κατά νου τον προοπτικό χαρακτήρα της ενσυνείδητης εμπειρίας είναι χρήσιμο για να μας θυμίζει ότι κάθε αναφορικότητα

⁸⁰ Searle, J., [1997], σ. 152

⁸¹ Searle, J., [1997], σ. 154

⁸² Searle, J., [1997], σ. 154

είναι εποπτική ως προς ορισμένες μόνο πλευρές (απόψεις ή επόψεις).”

83

Επίσης, κάθε ενσυνείδητη εμπειρία είναι πάντα εμπειρία από μια προοπτική, από μια οπτική γωνία, δηλαδή είναι πάντοτε μια εμπειρία *τινός*. Βεβαίως, όταν αναφερόμαστε σε μια εμπειρική κατάσταση (π.χ. στον πόνο), δεν αναφερόμαστε στον πόνο εν γένει, αλλά σε ‘αυτόν’ τον πόνο, δηλαδή στην υποκειμενική, εξατομικευμένη εμπειρία. Κύριο γνώρισμα της συνείδησης είναι η *υποκειμενικότητα*. Οι ενσυνείδητες εμπειρίες των ατόμων είναι εμπειρίες από μια συγκεκριμένη οπτική γωνία (την οπτική γωνία του ατόμου που έχει μια εμπειρία – ‘κανείς δεν μπορεί να αισθανθεί τον πόνο μου’). Η υποκειμενικότητα περιέχει απαραίτητα την πλευρά του ‘πώς το αισθάνεται κανείς’ των ενσυνείδητων καταστάσεων. Έτσι, είναι λογικό, η ποιότητα και η υποκειμενικότητα των ενσυνείδητων καταστάσεων να μην είναι προσιτή ως τέτοια στην παρατήρηση του τρίτου προσώπου.

Πολλοί μελετητές επιχειρηματολόγησαν σχετικά με το μη αναγώγιμο της συνείδησης. Ένας από τους πλέον διακεκριμένους είναι ο Thomas Nagel, ο οποίος διαφωνεί με την αναγωγή των νοητικών φαινομένων σε φυσικά και επισημαίνει ότι αυτό που καθιστά καταλυτικό το επιχειρήμά του είναι η αδυναμία αντικειμενικής παρατήρησης των υποκειμενικών εμπειριών μας. Ασκεί κριτική στο πρόγραμμα της Γνωσιακής Επιστήμης, η οποία αγνοεί τη συνείδηση και δεν λαμβάνει υπόψη της τον *υποκειμενικό* χαρακτήρα της εμπειρίας και, από λογική άποψη, εν τέλει αποκλείει την ύπαρξή του, με τρόπο που παραποιεί τον ίδιο τον χαρακτήρα της εμπειρίας.⁸⁴

Εάν τώρα προσπαθήσουμε να αναγάγουμε οντολογικά οποιοδήποτε υποκειμενικό μας αίσθημα (π.χ. πόνος) απλά και μόνο στις αντικειμενικές μορφές των νευρικών ερεθισμάτων, τότε θα παραλείπαμε τα υποκειμενικά χαρακτηριστικά του. Αυτό θα

⁸³ Searle, J., [1997], σ. 155

⁸⁴ Nagel, T., [1993], σ. 454

συνέβαινε, γιατί τα χαρακτηριστικά του πρώτου προσώπου δεν είναι ίδια με αυτά του τρίτου προσώπου. Εξάλλου, ακόμη κι αν υπήρχε η δυνατότητα να εξηγηθούν οι σωματικοί πόνοι αποκλειστικά με νευροφυσιολογικούς όρους, ακόμη και τότε δε θα μπορούσε κανείς να κατανοήσει τι πραγματικά σημαίνει πόνος, εάν δε γνώριζε πώς αισθάνεται κανείς όταν πονά, δηλαδή το περιεχόμενο της εμπειρίας.

Ο Nagel στηρίζει το αντι-αναγωγιστικό του επιχείρημα στο γεγονός ότι κάθε υποκειμενικό φαινόμενο συνδέεται ουσιαστικά με μια μοναδική οπτική γωνία, και μοιάζει αναπόφευκτο ότι αυτή η οπτική γωνία πρέπει να εγκαταλείπεται από κάθε αντικειμενική και φυσική θεώρηση. Με άλλα λόγια, εστιάζει το πρόβλημα του αναγωγισμού στην τριτοπρόσωπη παρατήρηση και διαχωρίζει το 'πώς είναι για το ίδιο το υποκείμενο να έχει μια συγκεκριμένη εμπειρία' και το 'με τι μοιάζει (σε εμάς) όταν έχει αυτή την εμπειρία'.

Ο Nagel, όπως είδαμε, θεωρεί πως είναι αδύνατον να αναχθούν οι εμπειρίες και ορισμένα ενσυνείδητα συναισθήματα (λ.χ. πόνος), των οποίων την υποκειμενική εμπειρία έχουμε, στην τριτοπρόσωπη αντικειμενική παρατήρηση, χωρίς ωστόσο να κάνει λόγο για τις προθεσιακές στάσεις, δηλαδή τις πεποιθήσεις, τις επιθυμίες και τις προθέσεις. Το γεγονός αυτό, επιβάλλει την αναδιατύπωση και διεύρυνση του επιχειρήματος κατά του ψυχοσωματικού αναγωγισμού ώστε να περιλαμβάνει και τα υψηλότερα επίπεδα της νόησης.

Την προσπάθεια αυτή αναλαμβάνει ο John R. Searle, ο οποίος με ένα επιχείρημα καθαρά οντολογικό, αποδεικνύει πως οι προθεσιακές στάσεις δεν μπορούν να αναχθούν στη φυσική πραγματικότητα: Ας δούμε ένα παράδειγμα οντολογικής αναγωγής: Όταν θέλουμε λ.χ. να αναγάγουμε οντολογικά τις ιδιότητες ενός τραπέζιού στις αιτίες που δημιουργούν αυτές τις ιδιότητες, σε πρώτη φάση ορίζουμε τα υποκειμενικά επιφανειακά χαρακτηριστικά (λ.χ. στερεότητα, σκληρότητα), καθώς και τα αντικειμενικά και ανεξάρτητα από τον παρατηρητή χαρακτηριστικά (λ.χ. αδιαπερατότητα) του. Κατόπιν,

αναζητούμε τις αιτίες του φαινομένου (παλμικές κινήσεις των μορίων του τραπέζιου), οι οποίες προσδίδουν στο τραπέζι τις ιδιότητές του (υποκειμενικές ή αντικειμενικές). Το τραπέζι πια δεν είναι τίποτε άλλο από μια συγκεκριμένη οργάνωση μορίων.

Από το παραπάνω παράδειγμα, φάνηκε ότι το ζητούμενο της αναγωγής είναι να αποκλειστούν τα επιφανειακά τυχαία χαρακτηριστικά και η αρχική έννοια να επαναπροσδιοριστεί σε σχέση με τα αίτια που παράγουν αυτά τα επιφανειακά χαρακτηριστικά. Αυτό, όμως, δε σημαίνει ότι η έννοια/ πράγμα ‘τραπέζι’ έχει εξαφανιστεί, απλά ότι έχει οριστεί με νέο τρόπο χωρίς να λαμβάνεται υπόψη “η υποκειμενική έκφανση, αλλά οι υποκείμενες φυσικές αιτίες”.⁸⁵

Το ερώτημα, το οποίο τίθεται τώρα, είναι ποιος λόγος μας εμποδίζει να επαναπροσδιορίσουμε τη συνείδηση, και εν γένει τις νοητικές στάσεις, από την άποψη των νευροβιολογικών διεργασιών με τον ίδιο τρόπο που επαναπροσδιορίσαμε το τραπέζι ως συγκεκριμένες κινήσεις μορίων. Τα αναγωγιστικά προγράμματα, εκτός των άλλων, έχουν ως σκοπό να αποκλείσουν τις υποκειμενικές εμπειρίες και να τις αφήσουν έξω από τον επαναπροσδιορισμό των πραγματικών φαινομένων. Τα φαινόμενα αυτά ορίζονται εκ νέου με βάση εκείνα τα χαρακτηριστικά που μπορούν να παρατηρηθούν από εξωτερικούς παρατηρητές. Στην περίπτωση της συνείδησης, όμως, αυτό που μας ενδιαφέρει είναι καθαυτές οι υποκειμενικές εμπειρίες, τις οποίες δεν υπάρχει τρόπος να αποκλείσουμε.

Με άλλα λόγια, γενικά στις αναγωγές προσπαθούμε να διακρίνουμε την πραγματικότητα από τις φαινομενικές εκφάνσεις της. Αντίθετα στην περίπτωση της συνείδησης, πραγματικότητα είναι οι ίδιες οι εκφάνσεις της. Επομένως, δεν μπορούμε σε αυτήν να εφαρμόσουμε το σχήμα που εφαρμόζουμε, όταν ανάγουμε άλλα φαινόμενα: Στην περίπτωση των άλλων φαινομένων διακρίνουμε την ‘αντικειμενική φυσική τους πραγματικότητα’ από την ‘υποκειμενική

⁸⁵ Searle, J., [1997], σ. 144

έκφρασή τους' πριν την αναγωγή (λ.χ. το χρώμα είναι φωτεινά κύματα ορισμένης συχνότητας, ανεξάρτητα από το αν αυτός που το βλέπει έχει αχρωματοψία). Στην περίπτωση, όμως, της συνείδησης, η πραγματικότητά της είναι ίδια η έκφρασή της' κατά συνέπεια, "ο σκοπός της αναγωγής θα χανόταν, αν επιχειρούσαμε ν' αποκλείσουμε την έκφραση και ορίζαμε απλώς τη συνείδηση με βάση την υποκειμενική φυσική πραγματικότητα".⁸⁶

Επιπλέον, το πρόγραμμα της ITN, εξ ορισμού, δε θεωρεί ως δεδομένο ότι γνωρίζει την ίδια την λειτουργία του εγκεφάλου και ακριβώς για το λόγο αυτό χρησιμοποιεί τους προσομοιωτές, προκειμένου δηλαδή να εξηγήσει αυτή τη λειτουργία. Επιπλέον – και σε αντίθεση με τα παραπάνω – η ITN θέτει ως δεδομένο ότι ο υπολογιστής -εξομοιωτής κατανοεί και υποστηρίζει ότι οι νοητικές διεργασίες δεν είναι τίποτε άλλο από υπολογιστικές διαδικασίες πάνω σε τυπικά στοιχεία και ότι μπορούν να πραγματοποιηθούν μέσα από τις πιο διαφορετικές εγκεφαλικές διαδικασίες, όπως ακριβώς ένα πρόγραμμα υπολογιστή μπορεί να εκτελεστεί μέσα από διαφορετικά υλικά μέρη. Ωστόσο, υπάρχει, στο σημείο αυτό ένα προφανές λογικό άλμα: ο συλλογισμός αποδεικνύεται μη έγκυρος, αφού δεν κάνει τίποτε άλλο από **λήψη του ζητούμενου**.

Εάν ισχύουν τα παραπάνω, η αναγωγή της συνείδησης σε νατουραλιστικό επίπεδο ισοδυναμεί με την εξουδετέρωσή της από το εξηγητικό σχήμα του νοητικού. Επομένως, εάν αγνοείται μια θεμελιώδης διάστασή του, γίνεται μια παραποίηση του εκ μέρους της Γνωσιακής Επιστήμης, η οποία εννοιολογικά, μεθοδολογικά και λογικά – τυπικά στηρίζεται κατ' αυτόν τον τρόπο σε επιστημολογικά προβληματικά θεμέλια.

Ακόμη κι αν η Γνωσιακή Επιστήμη ήταν επαρκώς συγκροτημένη εννοιολογικά και μεθοδολογικά, θα αντιμετώπιζε πρόβλημα στο επίπεδο της **επιστημονικής εξήγησης και θεμελίωσής** της. Πριν να

⁸⁶ Searle, J., [1997], σσ. 145-146

αναφερθούμε διεξοδικά στη θέση αυτή, σκόπιμο θα ήταν να ορίσουμε και να αποσαφηνίσουμε τον όρο ‘επιστημονική εξήγηση’: “Το πρώτο βήμα στην πορεία της διασάφησης της έννοιας ‘επιστημονική εξήγηση’, είναι να χαράξουμε μια σαφή διαχωριστική γραμμή μεταξύ της εξήγησης του *γιατί* ένα συγκεκριμένο φαινόμενο συμβαίνει, και της δικαιολόγησης της πεποίθησης *ότι* αυτό συμβαίνει.”⁸⁷

Έπειτα, πρέπει να σημειώσουμε ότι κάθε επιστημονική εξήγηση “αποτελείται από δύο μέρη, το *εξηγητέο* (explanandum) και το *εξηγούν* (explanans). Το εξηγητέο είναι το γεγονός που οφείλει να εξηγηθεί. Αυτό το γεγονός μπορεί να είναι ένα *επιμέρους* γεγονός, όπως η έκρηξη του διαστημικού λεωφορείου Τσάλεντζερ, ή ένα *γενικό* γεγονός, όπως ο νόμος της διατήρησης της ορμής. Μια πρόταση λέγεται *εξηγητέα πρόταση* (explanandum-statement) όταν ισχυρίζεται ότι το εξηγητέο συνέβη ή ισχύει.”⁸⁸ Το εξηγούν, αφετέρου, είναι “αυτό που πραγματοποιεί την εξήγηση. Αποτελείται από οποιαδήποτε γεγονότα, επιμέρους ή γενικά, τα οποία επικαλούμαστε για να εξηγήσουμε το εξηγητέο. Τις προτάσεις που αφορούν αυτά τα γεγονότα μπορούμε να τις ονομάσουμε *εξηγούσες προτάσεις* (explanans-statements).”⁸⁹

Με βάση τις παραπάνω διευκρινίσεις, μπορούμε να εντοπίσουμε στο πεδίο της Γνωσιακής Επιστήμης το εξής σφάλμα: Τα επιχειρήματά της συγχέουν το εξηγητέο (νόηση) με το εξηγούν (προγράμματα). Αυτό γίνεται αντιληπτό από την θεμελιώδη κιόλας θέση της Γνωσιακής Επιστήμης: Υποστηρίζει ότι ο υπολογιστής που είναι εφοδιασμένος με κατάλληλα προγράμματα φέρεται σαν να διαθέτει νόηση και υπόσχεται ότι με αυτή την υπόθεση εργασίας θα εξηγήσει τα ανθρώπινα νοητικά φαινόμενα (‘τι είναι η νόηση’ – εξηγητέα πρόταση). Στην πορεία, όμως, η Γνωσιακή Επιστήμη καταλήγει, χωρίς να στηρίζει εννοιολογικά και μεθοδολογικά αυτή την άποψη, στο ότι ο

⁸⁷ Salmon, W. C., [1999], σ. 4

⁸⁸ Salmon, W. C., [1999], σ. 7

παραπάνω υπολογιστής διαθέτει όντως νόηση, είναι ένα σκεπτόμενο ον (‘νόηση είναι αυτό που κάνει ο υπολογιστής’ – εξηγούσα πρόταση). Τελικά, λοιπόν, με βάση τη νόηση του υπολογιστή, εξηγεί την ανθρώπινη νόηση, αφού συγγέει το ‘ότι’ συμβαίνει κάτι με το ‘γιατί’ συμβαίνει. Πράγματι, ο ηλεκτρονικός υπολογιστής φαίνεται σε εμάς να κατανοεί, και έχουμε το δικαίωμα να υποθέσουμε τον τρόπο με τον οποίο αυτό συμβαίνει: Με τις ‘γλώσσες προγραμματισμού’ καθίσταται εφικτή η “επικοινωνία χειριστή και ηλεκτρονικού υπολογιστή⁹⁰” μέσα από ‘εντολές’, δηλαδή από δηλώσεις για το τι πρέπει να επιτελέσει με τυπικό τρόπο το λογισμικό (software) του υπολογιστή. Στη συνέχεια, έχουμε την ‘απάντηση’, η οποία προήλθε έπειτα από τον τυπικό χειρισμό των συμβόλων που ενεργοποιήθηκαν στο μηχάνημα/συσσκευή (hardware). Η διαδικασία αυτή όμως δεν μας νομιμοποιεί, ώστε να χρησιμοποιήσουμε τις διαπιστώσεις που προκύπτουν για να απαντήσουμε στο ερώτημα: ‘γιατί συμβαίνει αυτό’;

Τέλος, όπως όλοι γνωρίζουμε, “μεγάλο μέρος της επιστημονικής γνώσης μας εξαρτάται τόσο από την παρατήρηση όσο και από τη **μαθηματική (αυστηρή) συναγωγή συμπερασμάτων.**”⁹¹ Στην περίπτωση της Γνωσιακής Επιστήμης έχουμε τη σύγχυση των δύο στοιχείων που προαναφέραμε: Σύμφωνα με τους γνωσιακούς επιστήμονες ο ηλεκτρονικός υπολογιστής φέρεται ως εάν να διαθέτει νόηση (παρατήρηση), πράγμα το οποίο θεωρείται ότι αποδεικνύει πως πράγματι διαθέτει (μαθηματική(;) συναγωγή συμπεράσματος). Η συλλογιστική πορεία που ακολουθείται, είναι εσφαλμένη, αφού γίνεται λήψη του ζητουμένου, επομένως η σύγχυση παρατήρησης και συναγωγής δεν μπορεί να μας παράσχει συνεκτική και συνεπή γνώση της ανθρώπινης νόησης. Η Γνωσιακή Επιστήμη με αυτόν την

⁸⁹ Salmon, W. C., [1999], σ. 7

⁹⁰ Μαραγκός, Γ., [1996], σ. 118

⁹¹ Earman, J., και Salmon, W., [1999], σ. 50

ασυνέπεια υποσκάπτει συγκροτησιακά τα συλλογιστικά της θεμέλια,
άρα και τα συμπεράσματα στα οποία καταλήγει.

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ: 'ΤΕΧΝΗΤΗ ΖΩΗ'

1.ΤΟ ΕΝΝΟΙΟΛΟΓΙΚΟ ΠΛΑΙΣΙΟ ΤΗΣ ‘ΤΕΧΝΗΤΗΣ ΖΩΗΣ’ ΚΑΙ Η ΜΕΘΟΔΟΛΟΓΙΑ ΤΗΣ

Η ‘Τεχνητή Ζωή’ (Artificial Life) είναι ένα νέο επιστημονικό πρόγραμμα, με ζωή τριών περίπου δεκαετιών. Τα σπέρματα και τις βασικές αρχές του σχεδίου της ‘Τεχνητής Ζωής’ τα βρίσκουμε στον J. Von Neumann⁹², τον J. H. Holland⁹³ και τον C. G. Langton⁹⁴. Η ALife, πιο συγκεκριμένα, είναι το όνομα μιας ομάδας εφαρμογών, οι οποίες προέρχονται από πολλά επιστημονικά πεδία, μεταξύ άλλων, και κυρίως, την Πληροφορική, τις θεωρίες Αυτοοργάνωσης και Πολυπλοκότητας, και την Τεχνητή Νοημοσύνη. Η ‘Τεχνητή Ζωή’ διεκδικεί το χαρακτήρα μιας υπό διαμόρφωση επιστήμης, η οποία επιχειρεί να μελετήσει τον γενικότερο χαρακτήρα των θεμελιωδών διαδικασιών της ζωής, με όρους υπολογιστικούς-λειτουργικούς. Σύμφωνα με την ‘Τεχνητή Ζωή’, οι ιδιότητες της ζωής μπορούν να αποσπαστούν από τις φυσικές/ υλικές τους λεπτομέρειες (physical details), οι οποίες αφορούν τον τρόπο με τον οποίο αυτές οι ιδιότητες πραγματώνονται.⁹⁵ Οι διαδικασίες που φιλοδοξεί να προσομοιώσει η ‘Τεχνητή Ζωή’ περιλαμβάνουν την αυτό-οργάνωση, την εξέλιξη, την αναπαραγωγή, τη μάθηση, και την προσαρμοστικότητα, βασικές δηλαδή διαδικασίες του φαινομένου της ζωής. Ο κύριος προσανατολισμός της Τεχνητής Ζωής, σύμφωνα με τον Langton, συνίσταται στο να “επιχειρήσουμε να δημιουργήσουμε ζωή σε τεχνητό

⁹² von Neumann, J., [1966], *Theory of Self-Reproducing Automata*, (ed.) And completed by A. W. Burks (Urbana, Ill, University of Illinois Press).

⁹³ Holland, J. H., [1975], *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (Ann Arbor: University of Michigan Press).

⁹⁴ Langton, C.G., [1984], ‘Self-Reproduction in Cellular Automata’, *Physica D*, 10/1-2: 135-44.

Langton, C.G., [1986], ‘Studying Artificial Life with Cellular Automata’, *Physica D*, 22: 120-49. και Langton, C.G., [1989], *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 6; Redwood City, Calif.: Addison-Wesley).

⁹⁵ Sober, E., [1996], σ. 371

περιβάλλον, χρησιμοποιώντας τα ίδια είδη χημικών οργανιδίων σε περιβάλλον διαφορετικό από αυτό, από το οποίο αποτελούμαστε.”⁹⁶

Πρακτικός στόχος της ALife είναι να κατασκευαστεί μια νέα γενιά ρομπότ, τα οποία θα μιμούνται τη ζωή, προκειμένου, σε αρχικό τουλάχιστον επίπεδο, να βοηθήσουν τη βιολογία να κατανοήσει πώς κινούνται τα ζώα. Ουσιαστικά, στα ρομπότ αυτά η ‘Τεχνητή Ζωή’ επαγγέλλεται πως θα αναπαράγει τις θεμελιώδεις λειτουργίες των έμβιων όντων, με βάση το μοντέλο του υπολογιστή. Ακόμη, φιλοδοξία όλων αυτών των επιστημόνων είναι να κατασκευάσουν έναν κόσμο από νέα βιο-εμψυχωμένα ρομπότ, τα οποία θα βοηθούν τους ανθρώπους. Τελικός σκοπός είναι να δημιουργούνται στο εξής αυτόνομα ρομπότ-ζώα, τα οποία δεν θα κατευθύνονται από κάποιο συγκεκριμένο πρόγραμμα. Αυτό, ωστόσο, που δεν μπορεί να επιτευχθεί ακόμη είναι η αυτοκίνηση⁹⁷, πόσο μάλλον η αυτονομία.

Η επιστήμη, η οποία σχετίζεται άμεσα με την ‘Τεχνητή Ζωή’, είναι η βιολογία, η οποία θεωρεί κάθε ζωντανό οργανισμό ως μια σύνθετη βιοχημική μηχανή. Η βιολογία αρχίζει από τη συνολική θεώρηση του οργανισμού και, στη συνέχεια, δουλεύει αναλυτικά και τον εξετάζει ξεκινώντας από τα σύνθετα και καταλήγοντας στα απλούστερα δομικά μέρη του (όργανα, ιστοί, κύτταρα, οργανίδια, μόρια). Η ‘Τεχνητή Ζωή’, αποτελεί, από το άλλο μέρος, σε σχέση με τη βιολογία, μια συνθετική προσέγγιση στο φαινόμενο της ζωής: Επιχειρεί να συνθέσει, δηλαδή, έναν ζωντανό οργανισμό ξεκινώντας από την κατασκευή των πιο απλών δομικών μερών του, για να καταλήξει στο σύνολο.⁹⁸ “Η “Τεχνητή Ζωή” περιλαμβάνει διάφορα εγχειρήματα, τέτοια ώστε (1) να συνθέσει την εξέλιξη, (2) στους ηλεκτρονικούς υπολογιστές, και (3) να ενδιαφερθεί για οτιδήποτε προκύψει από τη διαδικασία”⁹⁹, ακόμη κι αν τα αποτελέσματα δεν

⁹⁶ Langton, C.G., [1996] σ. 50

⁹⁷ Taubes, G., [2000], σ. 80

⁹⁸ Langton, C.G., [1996], σ. 40

⁹⁹ Langton, C.G., [1996], σ. 40

έχουν αναλογίες με τον φυσικό κόσμο. Έχει, λοιπόν, για τους υποστηρικτές της, επιστημονικό ενδιαφέρον να γνωρίζουμε ποια είδη πραγμάτων μπορούν κατά κανόνα να συμβούν – ακόμη κι αν δεν λαμβάνουν χώρα εδώ στη γη.

Η ALife είναι για την θεωρητική βιολογία, λίγο ως πολύ, ό,τι είναι η Τεχνητή Νοημοσύνη (TN) για την θεωρητική ψυχολογία. Στην πραγματικότητα, αφότου ο νους επιδεικνύει πολλές θεμελιώδεις ιδιότητες της ζωής¹⁰⁰, και ουσιαστικά αποτελεί μια προέκτασή της, η TN, σύμφωνα με τους οπαδούς της συγκεκριμένης προσέγγισης, μπορεί να μελετά μια υπο-περιοχή της ALife. Εξάλλου, έναυσμα για την ανάπτυξη και των δύο έδωσε το έργο των ίδιων ανθρώπων, του A. Turing και του J. von Neumann.

Οι παραπάνω επιστήμες αποδέχονται ως κοινό πρότυπο, για τις περιοχές τις οποίες μελετούν, την λειτουργία του ηλεκτρονικού υπολογιστή, και θεωρούν ότι μπορούν να ‘αναπαραστήσουν’ σε οποιοδήποτε υλισμικό τόσο τις ‘συνάψεις’ του ανθρώπινου εγκεφάλου, όσο και αυτές κάθε έμβιου οργανισμού. Με τον όρο ‘συνάψεις’, εννοούμε τα κομβικά σημεία του εγκεφάλου, τα οποία η νευροεπιστήμη θεωρεί πως ερεθίζονται και στη συνέχεια επιτελούν την συμπεριφορική έκφραση των νοητικών φαινομένων. “Η Τεχνητή Νοημοσύνη χρησιμοποιεί την τεχνολογία του υπολογισμού, ως ένα μοντέλο για τη νοημοσύνη. Η ‘Τεχνητή Ζωή’, αφετέρου, επιχειρεί να αναπτύξει ένα νέο υπολογιστικό υπόδειγμα, το οποίο θα βασίζεται στις φυσικές διαδικασίες, οι οποίες υποστηρίζουν τους ζωντανούς οργανισμούς.”¹⁰¹

Η ‘Τεχνητή Ζωή’ εμφανίζει δύο μορφές, την ‘Ασθενή’ (*Weak Artificial Life*) και την ‘Ισχυρή’ (*Strong Artificial Life*). Στην ‘ασθενή’ μορφή της, η ‘Τεχνητή Ζωή’ χρησιμοποιεί τους ηλεκτρονικούς υπολογιστές ως εργαλεία, προκειμένου να μελετήσει τις διαδικασίες

¹⁰⁰ Audi, R., [1995], σ. 50

¹⁰¹ Langton, C.G., [1996], σ.50

της ζωής. Την ‘Ασθενή Τεχνητή Ζωή’ την ενδιαφέρει να μιμηθεί και να εξομοιώσει πλήρως τα φαινόμενα της ζωής. “Οι ηλεκτρονικοί υπολογιστές μπορούν να χρησιμοποιηθούν για να μας βοηθήσουν να δούμε το τι θα συμβεί σε μια γέφυρα εάν υπερφορτωθεί, μα ο υπολογιστής δεν είναι ο ίδιος γέφυρα.”¹⁰² Από την άλλη μεριά, η ‘Ισχυρή Τεχνητή Ζωή’ θεωρεί τους υπολογιστές ως αντικείμενό της. Με άλλα λόγια, στο ερευνητικό της πρόγραμμα αποδέχεται την ιδέα ότι οι ηλεκτρονικοί υπολογιστές αποτελούν υποδείγματα των βιολογικών διαδικασιών. Λίγο ως πολύ, δηλαδή, οι υπολογιστές, από την οπτική γωνία της ‘Ισχυρής Τεχνητής Ζωής’, παρουσιάζουν χαρακτηριστικά της ίδιας της ζωής και είναι *και οι ίδιοι* ζωντανοί.

Η ‘Τεχνητή Ζωή’, λοιπόν, ως υπόθεση εργασίας αποδέχεται ότι η ουσιώδης φύση των θεμελιωδών διαδικασιών της ζωής μπορεί να εκπληρωθεί αν σχετιστεί με απλά υπολογιστικά μοντέλα¹⁰³, ανεξάρτητα από το υλισμικό πραγμάτωσής τους. Αυτό σημαίνει πως βασική της ιδέα είναι ότι, όπως η νόηση, εν τέλει και τα φαινόμενα της ζωής, γενικότερα, μπορούν να πραγματώνονται σε διαφορετικά υλικά συστήματα. Δεν έχει, επομένως, καμιά σημασία το συγκεκριμένο πλαίσιο-οργανισμός εντός του οποίου υπάρχει, και εκδηλώνεται, κάθε φαινόμενο της ζωής. Οι βιολογικές δομές, τα κύτταρα ή τα όργανα κάθε έμβιου όντος μπορούν να αντικατασταθούν από οποιοδήποτε άλλο υλικό, είτε πρόκειται για καλώδιο, ή για σιλικόνη, ή για μέταλλο, ή για οτιδήποτε άλλο.

Η ALife μελετά τους μηχανισμούς της ζωής μέσα από τη ‘δημιουργία’ τεχνητών βιολογικών φαινομένων στο σύμπαν των υπολογιστών, στηριζόμενη τόσο στις υπολογιστικές όσο και στις βιολογικές επιστήμες. Τα παραπάνω βιολογικά φαινόμενα θα προσομοιώνονται με τη λειτουργία των ηλεκτρονικών υπολογιστών. “Η ‘Τεχνητή Ζωή’, λοιπόν, χρησιμοποιεί προτάσεις της πληροφορικής

¹⁰² Sober, E., [1996], σ. 361

¹⁰³ Bedau, M. A., [1996], σ. 343

και το πρότυπο των ηλεκτρονικών υπολογιστών για να μελετήσει τη ζωή γενικά και τη ζωή στη γη ειδικότερα”¹⁰⁴ Ο πυρήνας του όλου εγχειρήματος έγκειται στη δημιουργία περίπλοκων οργανισμών μέσα στον ψηφιακό μικρόκοσμο, των οποίων η πολυπλοκότητα είναι πολλές φορές συγκρίσιμη με εκείνη των οργανισμών που συναντούμε στη φύση. Οι έμβιοι οργανισμοί αποτελούν δυναμικά συστήματα, με την έννοια ότι είναι “συστήματα για καθένα απ’ τα οποία έχουμε μια αυστηρή ανάλυση του τρόπου με τον οποίο εξελίσσονται μέσα στο χρόνο. Τέτοια συστήματα μπορούν να θεωρηθούν όσα διασφαλίζουν έναν πεπερασμένο αριθμό μεταβλητών καταστάσεων, οι οποίες δοθέντων των ενδιαφερόντων του παρατηρητή, συλλαμβάνουν επαρκώς την κατάσταση του συστήματος μια δεδομένη χρονική στιγμή από την εξίσωση, η οποία ελέγχει κάθε στιγμή τον τρόπο με τον οποίο το σύστημα αλλάζει.”¹⁰⁵

Η ALife, ως μεθοδολογία, χρησιμοποιεί την προσομοίωση στο μοντέλο του ηλεκτρονικού υπολογιστή. “Η ‘Τεχνητή Ζωή’ χρησιμοποιεί ιδέες της πληροφορικής και το πρότυπο των υπολογιστών για να μελετήσει τη ζωή γενικά και τη ζωή στη γη ειδικότερα.”¹⁰⁶ Η ‘Τεχνητή Ζωή’ χρησιμοποιεί, κατά βάση, ‘προεκτάσεις’ και εφαρμογές του *συμπεριφορισμού* και του *λειτουργισμού*. Όπως αναφέραμε στο πρώτο κεφάλαιο, με τη μέθοδο του λειτουργισμού, τα νοητικά φαινόμενα ορίζονται με αναφορά προς τις αιτιακές – τυπικές σχέσεις τους με άλλα νοητικά φαινόμενα. Ο πόνος, η ζήλια, οι πεποιθήσεις κ.ά., εάν υιοθετήσουμε αυτή τη μέθοδο, δεν είναι ούτε βιώματα, ούτε διαθέτουν εν γένει νοητικό περιεχόμενο, το οποίο εξηγείται με προθεσιακό λεξιλόγιο, ούτε φυσικές καταστάσεις του νου. Τα νοητικά φαινόμενα, δηλαδή, δεν εξηγούνται φυσικαλιστικά ή νευροβιολογικά, αλλά είναι λειτουργίες, των οποίων οι τυπικές σχέσεις μπορούν να πραγματοποιηθούν σε οποιοδήποτε

¹⁰⁴ Boden, M.A., [1996], σ. 1

¹⁰⁵ Wheeler, M., [1996], σ. 222

υλικό πεδίο (σιλικόνη, καλώδια κ.ά.). Με άλλα λόγια, τα νοητικά φαινόμενα είναι το λογισμικό που μπορεί να πραγματώνεται σε οποιοδήποτε υλισμικό. Το ίδιο συμβαίνει και με την ‘Τεχνητή Ζωή’, η οποία επεκτείνει το παραπάνω πρόγραμμα από τα νοητικά φαινόμενα στα φαινόμενα της ζωής εν γένει.

Θεμελιωτής της ‘Τεχνητής Ζωής’ είναι ο C. G. Langton¹⁰⁷, ο οποίος, βλέπει τη ζωή ως ένα σύστημα ζωτικών λειτουργιών. Για τον Langton, η πληροφορία (information), η επικοινωνία (communication) και η ερμηνεία (interpretation) αποτελούν πραγματικά υπολογιστικά συστατικά τυπικών περιγραφικών συστημάτων. Θεωρεί, ακόμη, ότι η πληροφορία έχει σχέση, έτσι κι αλλιώς, με τη βιολογία και τους γενετικούς κώδικες των οργανισμών. Τελικά, η ζωή αποτελεί ένα εμπειρικό φαινόμενο, το οποίο μπορεί να εξηγηθεί με βάση την βιολογική λειτουργία του¹⁰⁸. Η τελευταία κατανοείται τυπικά/υπολογιστικά.

Από την άποψη αυτή, όπως φαίνεται, το πρόγραμμα της ‘Τεχνητής Ζωής’ αντιμετωπίζει τους ζωντανούς οργανισμούς και την ζωή ως λειτουργίες, τις οποίες θα παρουσίαζε ένα οποιοδήποτε σύστημα, του οποίου τα χαρακτηριστικά θα καθορίζονταν από τις τυπικές και αιτιακές σχέσεις των μερών του συστήματος αυτού.

Όσα αναφέρθηκαν παραπάνω, αποτυπώνουν τόσο το γενικό πλαίσιο της ‘Τεχνητής Ζωής’ όσο και το θεωρητικό υπόβαθρο, στο οποίο στηρίζει το πρόγραμμά της. Το υπόβαθρο αυτό είναι καταρχήν ο *συμπεριφορισμός*, τον οποίο η ‘Τεχνητή Ζωή’ ‘προσαρμόζει’ στο πρόγραμμά της: Εάν τα τεχνήματά της παρουσιάζουν όμοια συμπεριφορά με αυτή των έμβιων όντων, τότε τα τεχνήματα αυτά είναι έμβια. Αυτό θεωρεί πως μπορεί να το επιτύχει κατασκευάζοντας μηχανές, οι οποίες δομικά να είναι απολύτως όμοιες και να

¹⁰⁶ Boden, M.A., [1996], σ.1

¹⁰⁷ Το βασικό του κείμενο είναι το: “Self-Reproduction in Cellular Automata”, *Physica D*, 1984, 10: 135-44

¹⁰⁸ Boden, M.A., [1996], σσ. 8-9

αναπαριστούν βήμα-προς-βήμα τη δομή των ζωντανών οργανισμών. Η ‘Τεχνητή Ζωή’ έχει, σύμφωνα με τον Langton, την ιδιαίτερα μεγάλη τύχη ότι πολλοί από τους μηχανισμούς, οι οποίοι γίνονται αιτία, ώστε να εκδηλωθεί κάποια συγκεκριμένη συμπεριφορά από έναν ζωντανό οργανισμό, είναι ήδη γνωστοί. Σε αντίθεση με την Τεχνητή Νοημοσύνη, η ‘Τεχνητή Ζωή’ διαθέτει ήδη το πρότυπο, στο οποίο θα ‘προσαρμόσει’ τα ‘ευρήματά’ της. Αυτό είναι σε θέση να το επιτύχει, διότι το αντικείμενό της, η ζωή, θα αποτελεί ‘αντανάκλαση’ της ‘αρχιτεκτονικής’ δομής των φυσικών ζωντανών οργανισμών, οι οποίοι συνίστανται από πολλά εκατομμύρια επιμέρους μερών, καθένα απ’ τα οποία εκδηλώνει μια ορισμένη συμπεριφορά.

Επιπλέον, η ‘Τεχνητή Ζωή’ ακολουθεί, ως μεθοδολογική στρατηγική, τον *αναγωγισμό*. Ως υπόθεση εργασίας, δέχεται ότι ένας ζωντανός οργανισμός δεν είναι τίποτε άλλο από ένα περίπλοκο σύστημα, στο οποίο υπάρχουν τυπικές σχέσεις μεταξύ των δομικών μερών του. Οι λειτουργίες αυτές εξαρτώνται από τις τυπικές σχέσεις που υπάρχουν ανάμεσα στα μέρη του συγκεκριμένου συστήματος. Για παράδειγμα, η συμπεριφορά/ λειτουργία των ιστών μπορεί να αναπαραχθεί από τυπικές σχέσεις που υπάρχουν μεταξύ των κυττάρων, τα οποία συγκροτούν κάθε ιστό, ενώ η λειτουργία των οργάνων μπορεί να αναπαραχθεί ως λειτουργική/ τυπική σχέση που υπάρχει μεταξύ των ιστών. Τα όργανα, με τη σειρά τους, λειτουργούν τυπικά με τέτοιο τρόπο, ώστε να επηρεάζουν αιτιακά την κίνηση των άκρων.

Με τον τρόπο αυτό, όλες οι εκδηλώσεις των ζωντανών οργανισμών (κίνηση, αναπαραγωγή, θρέψη κ.ά.) αντιμετωπίζονται και ανάγονται σε επίπεδο τυπικών λειτουργιών. Αυτές οι λειτουργίες, πάντα σύμφωνα με το πρόγραμμα της ‘Τεχνητής Ζωής’, συνιστούν την έννοια της ζωής. Η ζωή, επομένως, δεν θεωρείται τίποτε άλλο παρά ένα σύνολο φαινομένων, όπου “τα φαινόμενα των υψηλότερων επιπέδων εξαρτώνται από αλληλεπιδράσεις διαδικασιών, οι οποίες

λαμβάνουν χώρα σε απλούστερα επίπεδα.”¹⁰⁹ Αυτού του είδους η αναγωγή αποτελεί την αιτιώδη αναγωγή, σύμφωνα με την οποία οι αιτιώδεις δυνατότητες της αναγόμενης οντότητας αποδεικνύονται πλήρως εξηγήσιμες, με βάση τις αιτιώδεις δυνατότητες των αναγόντων φαινομένων. Η ‘Τεχνητή Ζωή’ ανάγει, λοιπόν, το φαινόμενο της ζωής σε φαινόμενο λειτουργικό-τυπικό, και υπό αυτήν την προϋπόθεση το εξηγεί. Θεωρεί, επίσης, ως δεδομένο ότι θα δημιουργήσει και ζωή, αναπαριστώντάς το.

2. ΚΡΙΤΙΚΗ ΣΤΟ ΠΡΟΓΡΑΜΜΑ ΤΗΣ ‘ΤΕΧΝΗΤΗΣ ΖΩΗΣ’

Πριν προχωρήσουμε στην κριτική των επιχειρημάτων και της μεθοδολογίας που υιοθετεί το πρόγραμμα της ‘Τεχνητής Ζωής’, σκόπιμο θα ήταν να κάνουμε έναν διαχωρισμό μεταξύ δύο όρων. Οι όροι αυτοί έχουν μια λεπτή νοηματική διαφορά, που καταλήγει να είναι καθοριστική για προγράμματα όπως αυτό της ‘Τεχνητής Ζωής’. Πρόκειται για τους όρους ‘μίμηση’, ‘μηχανική προσομοίωση’ και ‘ταύτιση’. Από τη μια μεριά, με τον όρο ‘μίμηση’, αναφερόμαστε στο ότι είναι δυνατόν ορισμένα φαινόμενα να αποδοθούν μέσα από μια ορισμένη άλλη διαδικασία. Η συμπεριφορά του χρηματιστηρίου ή η πτητική συμπεριφορά ενός αεροπλάνου, για παράδειγμα, μπορεί να περιγραφεί/ αναπαρασταθεί στο πρόγραμμα ενός ηλεκτρονικού υπολογιστή. Ο όρος ‘μηχανική προσομοίωση’ είναι συνώνυμος με τον όρο ‘μίμηση’. Όταν, όμως, ταυτίζουμε ένα φαινόμενο με ένα άλλο, εννοούμε ότι το πρώτο φαινόμενο είναι το δεύτερο.

Προσαρμόζοντας τους παραπάνω ορισμούς στις κατασκευές της ‘Τεχνητής Ζωής’ μπορούμε να τους ‘παραφράσουμε’: ‘μιμούμαι’ ή ‘προσομοιώνω τη ζωή’ σημαίνει ότι επιτυγχάνω να κατασκευάσω

¹⁰⁹ Boden, M.A., [1996], σ. 9

μηχανές, οι οποίες με την τυπική-λειτουργική σχέση που υπάρχει μεταξύ των μερών τους, είναι δυνατόν να ενεργούν όπως ακριβώς ενεργεί το ζωντανό πρότυπο, με βάση το οποίο κατασκευάστηκαν. Όταν, από την άλλη μεριά, χρησιμοποιούμε τον όρο ‘ταύτιση’, εννοούμε ότι κάτι είναι *ταυτόσημο* κάτι άλλο. Οι θιασώτες της ‘Τεχνητής Ζωής’ χρησιμοποιούν τον όρο ‘προσομοίωση’ με την σημασία του όρου ‘ταύτιση’¹¹⁰. Με άλλα λόγια, λέγοντας ότι ‘προσομοιώνουν’ τη ζωή με υπολογιστές/ ρομπότ, εννοούν πως το δημιουργημά τους είναι το πρότυπό τους, δηλαδή είναι και το ίδιο ζωντανό. Αν, για παράδειγμα, κατασκευαστεί ένα μηχανικό ‘έντομο’ που θα μιμείται το ζωντανό του υπόδειγμα, τότε το έντομο αυτό δεν θα κάνει τίποτε άλλο από το να ενεργεί όπως το ‘πρότυπό’ του (προσομοίωση). Στην περίπτωση, αντίθετα, που το παραπάνω ‘έντομο’ θα θεωρηθεί ως ζωντανό, τότε αναφερόμαστε σε ταύτιση. (Στο εξής θα χρησιμοποιούμε τον όρο ‘προσομοίωση’ με την σημασία που του αποδίδεται από την ‘Τεχνητή Ζωή’).

Η ‘Τεχνητή Ζωή’ επαγγέλλεται πως θα εξηγήσει και θα δημιουργήσει οργανισμούς, οι οποίοι θα ‘μιμούνται’, μέσω προσομοίωσης, τη ζωή και τα γνωρίσματά της. Ορισμένοι θεωρούν ότι μια από τις εφαρμογές της ‘Τεχνητής Ζωής’ θα είναι εξαιρετικά χρήσιμη για τους ανθρώπους: Θα καταστεί δυνατό να αντικαθίστανται άρρωστα μέρη του ανθρώπινου οργανισμού με άλλα τεχνητά, νέα και υγιή. Είναι απόλυτα θεμιτό και χρήσιμο οι ανακαλύψεις και τα πορίσματα της αναδυόμενης αυτής επιστήμης να χρησιμοποιηθούν προς όφελος της ανθρωπότητας – θεωρητικά, η προοπτική αυτή εντάσσεται στο πλαίσιο προγράμματος της ‘Τεχνητής Ζωής’.¹¹¹ Ωστόσο, πριν απασχολήσει τους επιστήμονες το θέμα των εφαρμογών

¹¹⁰ Για παράδειγμα, ο Taubes ([2000], σ. 83) χρησιμοποιεί τους όρους προσομοίωση και δημιουργία, άρα ταύτιση, για το ίδιο τέχνημα: “Η ‘Τεχνητή Ζωή’ θα προσομοιώσει τις θεμελιώδεις λειτουργίες των έμβιων όντων... Ακόμη, φιλοδοξία όλων αυτών των επιστημόνων είναι να κατασκευάσουν έναν κόσμο από νέα *βιο-εμψυχωμένα* ρομπότ...”

¹¹¹ Minsky, M., [1994], σ. 88

της, θα πρέπει να αποσαφηνιστούν τα εννοιολογικά ζητήματα που ανακύπτουν από το παραπάνω πρόγραμμα.

Αυτό, που θέλουμε να θίξουμε με τα παραπάνω είναι το γεγονός ότι η ‘Τεχνητή Ζωή’ αντιμετωπίζει πρόβλημα σχετικά με το πώς κατανοεί τις υπολογιστικές διαδικασίες και με το πώς καταλήγει στο συμπέρασμα ότι αυτές οι υπολογιστικές διαδικασίες *είναι* η ζωή. Ο βασικός στόχος της ‘Τεχνητής Ζωής’ είναι, όπως είδαμε, να προσομοιώσει τις λειτουργίες των ζωντανών οργανισμών, ώστε να βοηθήσει τη βιολογία να μελετήσει καλύτερα αυτές τις λειτουργίες. Όμως, προβαίνει σε μια φιλοσοφικό-λογική αυθαιρεσία και διαπράττει ένα φιλοσοφικό-λογικό άλμα όταν υποστηρίζει ότι ένα τέχνημα διαθέτει *όντως* ζωή, εάν παρουσιάζει κάποιο από τα γνωρίσματά της. Με άλλα λόγια, η ‘Τεχνητή Ζωή’ δεν κάνει τίποτε άλλο από **λήψη του ζητουμένου**, αφού προϋποθέτει αυτό το οποίο επιχειρεί να αποδείξει, μιας και για να αποδείξει πως τα τεχνήματά της είναι ζωντανά, προϋποθέτει πως είναι ζωντανά και τα μελετά ως τέτοια. Είναι, όμως, εντελώς διαφορετικό το να ‘αναπαριστά’ κανείς ένα φαινόμενο από το να αποφανθεί ότι το φαινόμενο αυτό *είναι* η αναπαράστασή του. Με άλλα λόγια, η ‘Τεχνητή Ζωή’ όταν αποφαινεται πως μια κατασκευή είναι ζωντανή, προϋποθέτει ότι διαθέτει ζωή αντί να το αποδεικνύει.

Επιπλέον, και η μεθοδολογία της ‘Τεχνητής Ζωής’ αντιμετωπίζει προβλήματα. Η βιολογία είναι η επιστήμη που μελετά τη ζωή στη γη, η οποία είναι βασισμένη στον κύκλο του άνθρακα. Οι γνώσεις μας για τη ζωή αφορούν μόνο τη ζωή με την μορφή που την συναντούμε στη γη. Δεν είναι, όμως, απίθανο να υπάρχουν και άλλα είδη ζωής (σε κάποιον άλλον πλανήτη), που να παρουσιάζουν διαφορετικά χαρακτηριστικά από αυτά της ζωής στον πλανήτη μας. Το γεγονός αυτό αφαιρεί από τη θεωρία της Τεχνητής Ζωής το αντιπροσωπευτικό δείγμα στις προκειμένες των επιχειρημάτων της. Προκειμένου, όμως, να καταλήξουμε σε ασφαλή συμπεράσματα σε κάθε επαγωγικό συλλογισμό, πρέπει να αποφεύγουμε την γενίκευση και να

στηριζόμαστε σε επαρκή στοιχεία στις προκείμενές μας. Ο συλλογισμός της ‘Τεχνητής Ζωής’ – απ’ τη στιγμή που η έλλειψη άλλων παραδειγμάτων περιορίζει σοβαρά την προσπάθειά της για την κατανόηση του τι είναι ζωή και του τι θα μπορούσε να αποτελεί έναν ζωντανό οργανισμό – συνιστά μια **ατελή επαγωγή**.

Εξάλλου, η **επιστημονική εξήγηση**, την οποία προσφέρει η ‘Τεχνητή Ζωή’, σχετικά με τα χαρακτηριστικά της ζωής, δεν είναι έγκυρη. Αυτό συμβαίνει, γιατί η επιστημονική της εξήγηση *συγχέει το εξηγητέο* (ζωή) *με το εξηγούν* (προσομοίωση σε υλισμικό επίπεδο). Με άλλα λόγια, η ‘Τεχνητή Ζωή’ προσομοιώνοντας τα γνωρίσματα της ζωής στοχεύει να τα εξηγήσει (‘γιατί συμβαίνουν’ / ‘τι είναι’), ενώ την ίδια στιγμή εξισώνει την προσομοίωση αυτή (‘το γεγονός ότι μοιάζουν’) με την πραγματική ζωή. Με τον τρόπο αυτό, δεν διαχωρίζει το *ότι* συμβαίνει κάτι από το *γιατί* συμβαίνει, αλλά τα ταυτίζει, με αποτέλεσμα να μην προσφέρει την επιστημονική εξήγηση, την οποία επιχειρεί να δώσει.

Η ‘Τεχνητή Ζωή’ καταλήγει στα συμπεράσματά της, υιοθετώντας μεθοδολογικά μια **λειτουργιστική εξήγηση** των φαινομένων της ζωής. Η βιολογία θεωρεί πως κάθε κύτταρο επιτελεί και εξισούται με τη λειτουργία του. Με τα μεγάλα βήματα που έγιναν τις τελευταίες δεκαετίες στο ζήτημα της αποκρυπτογράφησης του γονιδιώματος, το οποίο θεωρείται πως μεταφέρει τις γενετικές πληροφορίες, το κύτταρο ερμηνεύεται ως ‘υποδοχέας’ πληροφοριών. Η λειτουργία του κυττάρου θεωρείται πως είναι η επεξεργασία πληροφοριών. Με άλλα λόγια, η λειτουργία των κυττάρων έγκειται στο φορμαλισμό τους, και τα ίδια αποτελούν το ‘λογισμικό’ του ζωντανού οργανισμού.

Όπως φάνηκε από τα παραπάνω, το πρόγραμμα της ‘Τεχνητής Ζωής’ θεωρεί πως οι ζωντανοί οργανισμοί συνίστανται από τυπικούς υπομηχανισμούς (modulus), οι οποίοι συνδέονται μεταξύ τους με τυπικές σχέσεις. Οι υπομηχανισμοί αυτοί αλληλεπιδρούν και, σύμφωνα με το πρόγραμμα της ‘Τεχνητής Ζωής’, το γεγονός αυτό

είναι που προσδίδει ζωή σε έναν έμβιο οργανισμό. Έτσι, λοιπόν, σύμφωνα με την άποψη της ‘Τεχνητής Ζωής’, η κίνηση, η αυτό-οργάνωση, η θρέψη, η αναπαραγωγή κ.ά. είναι αποτελέσματα τυπικών και μόνο σχέσεων μεταξύ των οργάνων κάθε έμβιου όντος. Για τον λόγο αυτό και υπό την προϋπόθεση αυτή, εξάλλου, η ‘Τεχνητή Ζωή’ δέχεται ότι η ζωή μπορεί να υλοποιηθεί σε οποιοδήποτε υλισμικό υπόστρωμα.

Ο E. Sober θεωρεί πως ο λειτουργισμός είναι “μια απελευθερωτική δύναμη. Μας οδηγεί στο να δούμε την ανθρώπινη γνώση (cognition) και τους οργανισμούς που υπάρχουν στη γη ως παραδείγματα του νου και της ζωής. Για να κατανοήσουμε τον νου και τη ζωή, οφείλουμε να τα διαχωρίσουμε από τις φυσικές λεπτομέρειες.”¹¹² Ο ίδιος μελετητής θεωρεί ότι ο λειτουργισμός εμπεριέχει τον κίνδυνο να μπερδευτεί η μαθηματική δομή ενός προγράμματος με το εμπειρικό περιεχόμενό του.¹¹³ Με άλλα λόγια, αναγνωρίζει ότι το λειτουργικό πρόγραμμα είναι ένα πρόγραμμα καθαρά τυπικό-φορμαλιστικό, άρα προσιδιάζει στη μέθοδο που ακολουθείται για την επίλυση μαθηματικών προβλημάτων. Αντίθετα όμως, η ζωή σε κάθε εκδήλωσή της δεν μπορεί να αποκοπεί, ούτε να μην προσδιοριστεί σε συνάρτηση με την έννοια της *εμπειρίας*.

Ο T. Nagel πραγματεύεται το ζήτημα των **υποκειμενικών εμπειριών** των έμβιων όντων, καθώς και την ποιοτική τους διαφοροποίηση από οργανισμό σε οργανισμό, χρησιμοποιώντας ως παράδειγμα τις νυχτερίδες. Οι νυχτερίδες, όπως όλοι γνωρίζουμε, διαθέτουν ορισμένα αισθητήρια όργανα (λ.χ. το σόναρ), τα οποία τις βοηθούν να έχουν μια μορφή αντίληψης. Αυτά τα όργανα δεν παρουσιάζουν ομοιότητες με κανένα όργανο του ανθρώπινου οργανισμού. Από την άλλη μεριά, οι άνθρωποι ‘δεσμεύονται’ από τη δική τους σωματική διάπλαση και τις δικές τους διανοητικές

¹¹² Sober, E., [1996], σ.376

¹¹³ Sober, E., [1996], σ.371

ικανότητες, οι οποίες τους βάζουν όρια στην αντίληψη και την κατανόηση όλης γενικά της πραγματικότητας. Αυτό έχει ως αποτέλεσμα οι άνθρωποι να μην μπορούν να κατανοήσουν την ‘εσωτερικότητα’ και την ποιοτική διαφοροποίηση των εμπειριών των νυχτερίδων και, γενικότερα, οποιουδήποτε άλλου ζωντανού οργανισμού.

Θα επισημάνουμε, στο σημείο αυτό ότι οι ζωντανοί οργανισμοί πέρα από τις λειτουργίες που επιτελούν, κάποιες από τις οποίες θα μπορούσαν ίσως, σε ένα αρχικό επίπεδο, να προσομοιωθούν σε έναν υπολογιστή ή σε ένα ρομπότ, διαθέτουν κάποια βασικά γνωρίσματα, τα οποία είναι αδύνατο να απομονωθούν από έναν συγκεκριμένο οργανισμό: είναι **γνωρίσματα εσωτερικά/ εγγενή** και σχετίζονται με τις εμπειρίες των έμβιων όντων. Τέτοιου είδους εμπειρίες μπορεί να είναι ο φόβος, ο πόνος, ή ο τρόπος με τον οποίο οι ζωντανοί οργανισμοί αντιλαμβάνονται τα χρώματα ή τον ήχο κτλ. Αυτές οι υποκειμενικές εμπειρίες είναι πάντοτε εμπειρίες *τινός*, και είναι αδύνατον να εξηγηθούν από την οπτική του παρατηρητή, δηλαδή του τρίτου προσώπου.

Οι εμπειρίες έχουν πάντοτε έναν συγκεκριμένο φορέα, ή αλλιώς είναι πάντοτε εμπειρίες που ανήκουν σε συγκεκριμένο υποκείμενο (**οντολογία του πρώτου προσώπου**). Πιθανόν κάποιος να φέρει ως αντίρρηση στα παραπάνω ότι θα μπορούσε, με κάποιο τρόπο, ένας άνθρωπος να βιώσει την εμπειρία του να είναι νυχτερίδα – ή οτιδήποτε άλλο – προσαρμόζοντας στο σώμα του τα όργανα (ή προσομοιώσεις οργάνων) του θηλαστικού αυτού. Στην περίπτωση αυτή, θα μπορούσαμε να υποστηρίξουμε ότι, ακόμη κι αν προσπαθήσουμε να φανταστούμε τον εαυτό μας με τις ιδιότητες κάποιου άλλου οργανισμού, ακόμη κι αν φανταζόμασταν κάποιον συνδυασμό από προσθήκες, αφαιρέσεις και τροποποιήσεις, ακόμη κι αν είχαμε την εμφάνιση μιας νυχτερίδας ή ενός μυρμηγκιού, θα ήταν αδύνατο να είχαμε την εμπειρία των όντων αυτών με τον τρόπο που τα ίδια την

βιώνουν – ή κι αν είχαμε κάποιου είδους εμπειρία, αυτή δεν θα έμοιαζε με την εμπειρία των συγκεκριμένων όντων. Με άλλα λόγια, η εμπειρία κάθε ζωντανού οργανισμού όχι μόνο είναι ποιοτικά διαφοροποιημένη από την εμπειρία κάποιου άλλου οργανισμού – άρα μη προσβάσιμη πλήρως στον τριτοπρόσωπο παρατηρητή, αλλά και οι ίδιοι οι άνθρωποι-παρατηρητές δεν μπορούν να ‘ξεφύγουν’, και να αποδεσμευτούν από τον δικό τους ποιοτικό, υποκειμενικό προσδιορισμό.

Όταν οι υποκειμενικές εμπειρίες δεν μπορούν να απομονωθούν από τον φορέα τους, πόσο μάλλον τα ανώτερα νοητικά φαινόμενα, τα οποία είναι επίσης φαινόμενα τινός και, όπως αναπτύχθηκε στο δεύτερο κεφάλαιο, είναι κατεξοχήν φαινόμενα μη πλήρως προσπελάσιμα από την οπτική γωνία του τρίτου προσώπου. Όσον αφορά την πιθανότητα δημιουργίας ρομπότ, το οποίο θα αποτελεί προσομοίωση ενός ανθρώπου, τότε, εκτός από τις ασυνείδητες ‘εσωτερικές’ καταστάσεις που διαθέτουν τα ‘άλογα’ έμβια όντα, η ‘Τεχνητή Ζωή’ θα πρέπει να προσομοιώσει και τα νοητικά φαινόμενα. Όμως, το ότι η Τεχνητή Νοημοσύνη χωλαίνει συγκροτησιακά (επιστημολογικά, μεθοδολογικά και λογικά) επηρεάζει ανάλογα το οικοδόμημα της ‘Τεχνητής Ζωής’. (Δεν θα πρέπει να ξεχνάμε ότι στο πλαίσιο της ‘Τεχνητής Ζωής’ συγκαταλέγεται η Τεχνητή Νοημοσύνη).

Όπως είδαμε παραπάνω, είναι αδύνατη η πρόσβαση του τριτοπρόσωπου αντικειμενικού παρατηρητή στις εμπειρίες των ζωντανών οργανισμών. Επιπλέον, τα έμβια όντα έχουν ορισμένα χαρακτηριστικά γνωρίσματα, όπως την αυτό-οργάνωση, την αυτοκίνηση, την αυτενέργεια, τον αυτοέλεγχο και το ένστικτο της αυτοσυντήρησης (στο ένστικτο αυτό περιλαμβάνεται και η ικανότητα της προσποίησης). Το πρώτο συνθετικό ‘αυτό-’ είναι εκείνο που προσδίδει στο φαινόμενο της ζωής την μοναδικότητα και το ειδητικό του γνώρισμα, δηλαδή το υποκειμενικό του περιεχόμενο. Είναι φανερό το γεγονός ότι όλες οι αντιδράσεις και η συμπεριφορά των ζωντανών

οργανισμών είναι απόρροια εξωτερικών ερεθισμάτων και φυσικά ακολουθούν ορισμένους τυπικούς/ συντακτικούς κανόνες. Ωστόσο, τίποτε δεν μας επιτρέπει να παραγνωρίσουμε το γεγονός ότι υπάρχει επιπλέον στους ζωντανούς οργανισμούς το ποιοτικό στοιχείο, ένα πρωτογενές επίπεδο **‘υποκειμενικότητας’**. Με άλλα λόγια, υπάρχουν κάποιες εσωτερικές διεργασίες με συγκεκριμένο περιεχόμενο, δηλαδή το ‘τι’ τους, οι οποίες λαμβάνουν χώρα σε κάθε ζωντανό ον. Από τη στιγμή που ο λειτουργισμός μελετά μόνο τις μηχανικές σχέσεις μεταξύ των ερεθισμάτων και των αντιδράσεων των οργανισμών και παραγνωρίζει αυτό το ‘περιεχόμενο’, δεν είναι σε θέση να αποτελέσει ικανοποιητική μέθοδο για την εξήγηση της ζωής.

Επιπλέον, κρίσιμο ζήτημα για την ‘Τεχνητή Ζωή’ αναδεικνύεται η επιτυχημένη προσομοίωση, ή δημιουργία, ενός χαρακτηριστικού γνωρίσματος των ζωντανών οργανισμών, της **αναπαραγωγής**. “Χωρίς αυτό οι μηχανισμοί (φυσικής) επιλογής θα πρέπει επίσης να *προκαθοριστούν* από τον επιστήμονα που κάνει την εξομοίωση. Έτσι, η τεχνητή επιλογή ποτέ δε θα μπορέσει να λειτουργήσει ως δημιουργός της φυσικής επιλογής.”¹¹⁴ Αυτό το γεγονός έχει καταλυτικές συνέπειες στο πρόγραμμα της ‘Τεχνητής Ζωής’: Οι οργανισμοί-τεχνήματα, που θα κατασκευαστούν, σύμφωνα με τον τρόπο αυτό, δεν θα είναι ελεύθεροι να αναπτύξουν από μόνοι τους τις υγιείς και φυσιολογικές λειτουργίες τους, ούτε θα μπορούν να εξελιχθούν με τρόπο φυσιολογικό, αβίαστο και ανεπηρέαστο από παράγοντες, οι οποίοι θα είναι εξωγενείς στο σύστημα. Ουσιαστικά, οι προσομοιώσεις, οι οποίες γίνονται με προκαθορισμένα γονίδια και προκαθορισμένες ‘φυσιολογικές’ λειτουργίες δεν συνιστούν ζωντανούς οργανισμούς.

Το ζήτημα του **προκαθορισμού** εγείρει επίσης πολλά προβλήματα, σχετικά με το κατά πόσο οι κατασκευές της ‘Τεχνητής Ζωής’ θα διαθέτουν πράγματι τα ιδιαίτερα χαρακτηριστικά της πραγματικής ζωής, καθώς και την ανεξάρτητη πορεία που αυτή

¹¹⁴ Ray, T.S., [1996] σ. 112

ακολουθεί. Κάθε ζωντανός οργανισμός είναι κατασκευασμένος με τέτοιο τρόπο, ώστε να παρουσιάζει τη δική του αυτόνομη πορεία σε όλους τους τομείς της εξέλιξης και ύπαρξής του, μέχρι το θάνατό του. Ακόμη και ζώα ή φυτά του ίδιου είδους, τα οποία έχουν την ίδια βιολογική δομή, τις ίδιες γενετικές πληροφορίες και τα οποία αποτελούνται από τον ίδιο ‘συνδυασμό’ κυττάρων, ιστών, οργάνων κ.ά. παρουσιάζουν διαφορετική συμπεριφορά και έχουν διαφορετική ανάπτυξη και εξέλιξη. Η προκαθορισμένη πορεία των τεχνημάτων της ‘Τεχνητής Ζωής’ αφαιρεί από αυτά τουλάχιστον ένα χαρακτηριστικό της ζωής, άρα δεν μπορούν να είναι ζωντανά.

Η πορεία, την οποία ακολουθεί ξεχωριστά κάθε έμβιο ον, είναι προκαθορισμένη, βέβαια, σε ένα βαθμό (σύμφωνα με τη βιολογία, αυτό οφείλεται κυρίως στον γενετικό του κώδικα), αλλά η περαιτέρω ανάπτυξη του οργανισμού αυτού γίνεται αυτόνομα και εξατομικευμένα. Η ‘Τεχνητή Ζωή’ δεν είναι σε θέση να υποστηρίζει ότι τα προγράμματά της είναι – ή τουλάχιστον στο μέλλον θα είναι – ζωντανά, από τη στιγμή που δεν μπορεί να αποφύγει τον προκαθορισμό της συμπεριφοράς, της κίνησης, της αυτοσυντήρησης κ.ά. Και δεν μπορεί να τον αποφύγει, επειδή τα δημιουργήματά της αναγκαστικά θα βασίζονται μόνο σε μια τυπική αλληλουχία και αλληλεπίδραση υλικών, τα οποία θα συνδέονται και θα αλληλεπιδρούν με κοινό παρονομαστή ένα πρόγραμμα υλοποίησης μηχανικών εντολών. Το γεγονός αυτό στερεί τη δυνατότητα από τα τεχνήματα της ‘Τεχνητής Ζωής’ από **την αυτονομία**.

Ένα επιπλέον πρόβλημα που θα πρέπει να αντιμετωπίσει και να επιλύσει η ‘Τεχνητή Ζωή’ είναι και **το ζήτημα του θανάτου**. Όλοι οι ζωντανοί οργανισμοί έχουν μια αναπόφευκτη κοινή κατάληξη, κατά την οποία η ζωή τους τερματίζεται εξαιτίας φυσικών, ή μη, παραγόντων. Τότε, οι λειτουργίες της ζωής παύουν και κάθε οργανισμός, αφού έχει τερματίσει την πορεία του, αποσυντίθεται. Αφού ο θάνατος είναι η μοιραία κατάληξη της ζωής, θα πρέπει και τα

τεχνήματα της ‘Τεχνητής Ζωής’, εάν είναι πράγματι προσομοιώσεις της ζωής, να πεθαίνουν. Αυτό, βεβαίως, θα πρέπει να συμβαίνει στα ‘ρομπότ’ ανεξάρτητα από το πρόγραμμα με το οποίο έχουν τροφοδοτηθεί από τους προγραμματιστές τους. Προς το παρόν, ωστόσο, η ‘Τεχνητή Ζωή’ δεν μπορεί να δώσει μια ικανοποιητική απάντηση στο κατά πόσο τα τεχνήματα που θα δημιουργεί θα είναι σε θέση να λειτουργούν ανεξάρτητα και αυτόνομα.

Σύμφωνα με τους επιστήμονες, η ζωή στη γη εμφανίστηκε έπειτα από την έκρηξη που σημειώθηκε στην περίοδο του Καμβρίου πριν από περίπου 600 εκατομμύρια χρόνια. Την περίοδο εκείνη, αλλά και μέχρι σήμερα, υπήρχε μεγάλη ποικιλία μορφών ζωής, η οποία μεταβλήθηκε πολλές φορές, εξαιτίας διαφόρων αιτιών (παγετώνες, γεωλογικές αλλαγές κ.ά.). Παράλληλα, οι οργανισμοί μεταβλήθηκαν και προσαρμόστηκαν στις εκάστοτε συνθήκες που επικρατούσαν κατά περιόδους στον πλανήτη μας. Με άλλα λόγια, όλοι οι ζωντανοί οργανισμοί, εκτός από όσα προαναφέραμε, παρουσιάζουν και τα χαρακτηριστικά της **ποικιλίας**, της **προσαρμοστικότητας** και της **εξέλιξης**.¹¹⁵ Η ‘Τεχνητή Ζωή’ δεν κάνει τίποτε άλλο από το να μιμείται τους ζωντανούς οργανισμούς και, εκ των πραγμάτων, δεν είναι σε θέση να αναπαράγει, να δημιουργήσει ή έστω να προκαθορίσει για το μέλλον τα παραπάνω γνωρίσματα.

Τέλος, η ‘Τεχνητή Ζωή’ στηρίζεται σε μια αυθαίρετη μεθοδολογική προϋπόθεση. Προϋποθέτει, δηλαδή, ότι τα φαινόμενα της ζωής δεν είναι τίποτε άλλο από τυπικά φαινόμενα και λειτουργίες. Όπως βλέπουμε, γίνεται **αναγωγή** του φαινομένου της ζωής σε σχέσεις τυπικών στοιχείων σε ένα σύστημα, πράγμα που αποδεικνύεται προβληματικό για το πρόγραμμά της. Η ζωή συνίσταται, εν μέρει, σε λειτουργίες, αλλά σε λειτουργίες ενός φορέα, τινός συστήματος. Η αναγωγιστική μέθοδος μπορεί να εξυπηρετήσει, έως έναν βαθμό, την ‘Τεχνητή Ζωή’, στο επίπεδο της κατανόησης των λειτουργιών που

¹¹⁵ Ray, T.S., [1996], σ. 113

λαμβάνουν χώρα σε έναν ζωντανό οργανισμό. Υπάρχουν, όμως, ορισμένα χαρακτηριστικά της ζωής, τα οποία δεν είναι δυνατό να αναχθούν σε τυπικό επίπεδο, όπως λ.χ. αυτά που αφορούν την αυτό-οργάνωση των έμβιων όντων.

ΣΥΜΠΕΡΑΣΜΑΤΑ/ ΕΠΙΛΕΓΟΜΕΝΑ

Στην εργασία αυτή εξετάσαμε τις προσπάθειες που γίνονται από την Γνωσιακή Επιστήμη, προκειμένου να εξηγηθούν τα νοητικά φαινόμενα του ανθρώπου. Η Γνωσιακή Επιστήμη ως βασική παραδοχή έχει το ότι η νόηση εξισούται με υπολογισμό και μπορεί να εξηγηθεί ως λειτουργία. Επίσης, δεν έχει νόημα το να αναγνωρίζεται οποιοδήποτε περιεχόμενο, ποιότητα ή ιδιάζουσα υποκειμενικότητα/εσωτερικότητα στα νοητικά φαινόμενα. Το μόνο που έχει σημασία για την Γνωσιακή Επιστήμη είναι οι λειτουργίες των φαινομένων αυτών. Από την άλλη πλευρά, αναφέραμε το πρόγραμμα μιας σύγχρονης

πρωτοποριακής τάσης, της ‘Τεχνητής Ζωής’ – που δεν έχει, ωστόσο, καθιερωθεί ακόμη ως επιστήμη – η οποία αντιμετωπίζει τα βιολογικά φαινόμενα ως λειτουργίες και στοχεύει στην προσομοίωση όλων ανεξαιρέτως των ζωικών φαινομένων με υπολογιστικά μοντέλα. Τόσο η πρώτη όσο και η δεύτερη, ειδικότερα, θεωρούν πως τα ανώτερα νοητικά και βιολογικά φαινόμενα είναι δυνατόν να εξηγηθούν, αν προσομοιωθούν με βάση το μοντέλο των ηλεκτρονικών υπολογιστών.

Αφού εκθέσαμε το εννοιολογικό πλαίσιο και τις μεθόδους των παραπάνω προγραμμάτων και ανασυγκροτήσαμε την επιχειρηματολογία τους, προχωρήσαμε στον έλεγχο των εννοιολογικών τους αποθεμάτων και ελέγξαμε τα λογικά και νοηματικά όρια των θέσεών τους. Με άλλα λόγια, ενδιαφερθήκαμε να δούμε το εννοιολογικό σθένος αυτών των προγραμμάτων. Μας απασχόλησε να βρούμε αν, τελικά, είναι δυνατόν η νόηση να εξισούται με υπολογισμό και αν, επίσης, το φαινόμενο της ζωής πράγματι δεν είναι τίποτε άλλο από ένα φορμαλιστικό ‘συμβάν’.

Καταλήξαμε στο ότι τα προγράμματα της Γνωσιακής Επιστήμης και της ‘Τεχνητής Ζωής’ θα μπορούσαν να είναι γόνιμα και να ανοίγουν προοπτικές σχετικά με τα αντικείμενά τους, εάν το εξηγητικό εύρος που διεκδικούν ήταν πιο περιορισμένο και δεν πίστευαν ότι αυτό που εξηγούν (δηλ. η υπολογιστική διαδικασία) είναι ο νους ή η ίδια η ζωή (λήψη του ζητουμένου). Επομένως, υπάρχουν ορισμένα εννοιολογικά όρια στο εγχείρημά τους. Η υπεραισιοδοξία που διακατέχει τα παραπάνω προγράμματα έγκειται στο ότι θεωρούν πως είναι δυνατόν να επιστημονικοποιηθούν ο νους και η ζωή. Ωστόσο, για το εγχείρημα αυτό είναι απαραίτητο να λάβουν σοβαρά υπόψη τους τις προϋποθέσεις που συγκροτούν τα προς εξήγηση αντικείμενά τους.

Σε αυτό το σημείο, ο ρόλος της φιλοσοφίας, ως κριτική της μεταφυσικής της επιστήμης, είναι καθοριστικός. Η φιλοσοφία είναι σε θέση να ανακαλύψει τα όρια κάθε είδους επιστημονικού εγχειρήματος και να ‘αποκαθαίρει’ το έδαφος, ώστε η εμπειρική έρευνα να

κατορθώσει να καταλήξει σε γόνιμα πορίσματα. Η σχέση μεταξύ φιλοσοφίας και επιστήμης, εξάλλου, είναι αμφίδρομη: αρκετοί επιστημονικοί κλάδοι αναδύθηκαν από τη φιλοσοφική έρευνα κάθε εποχής και, αντίστροφα, η φιλοσοφία πάντοτε χρησιμοποίησε τα κυρίαρχα μοντέλα από κάθε σύγχρονή της επιστήμη, προκειμένου να εξηγήσει και να λύσει τα προβλήματα που η ίδια αντιμετωπίζει.

Η εμφάνιση των ερωτημάτων που αφορούν τη νόηση και η ιδέα της προσέγγισης των νοητικών φαινομένων ως λειτουργιών ξεκίνησε από τον 17^ο αιώνα (T. Hobbes¹¹⁶, D. Hume¹¹⁷) – την ίδια περίοδο που η διαμόρφωση της μαθηματικής φυσικής αποφέρει τους πρώτους καρπούς της με τα επιστημονικά επιτεύγματα του Κοπέρνικου και του Γαλιλαίου.

Η προσέγγιση της σκέψης/ Λόγου μέσω της λειτουργίας του προτείνεται στην ολοκληρωμένη μορφή της από τον Im. Kant. Ο Kant ανέλυσε το σύνολο της εμπειρικής γνώσης, συμπεριλαμβανομένης της αντίληψης, βάσει κανόνων, πράγμα που κάνει κατανοητό πως τα νοητικά φαινόμενα αντιμετωπίζονται και εξηγούνται με βάση μια κανονιστική μέθοδο. Η νόηση και ο Λόγος μπορούν να εξεταστούν “όπως λειτουργούν μέσα σ’ ένα σύστημα εννοιών και αρχών αναφερόμενων σε αντικείμενα εν γένει, ανεξάρτητα δηλ. από την ιδιαίτερη υφή τους.”¹¹⁸ Με άλλα λόγια, ο Λόγος δεν είναι τίποτε άλλο από την λειτουργία κανόνων σε κάθε περίπτωση, αν η συγκεκριμένη περίπτωση έχει ήδη περιγραφεί χωρίς αμφισημίες βάσει γενικών γνωρισμάτων που αναφέρονται στον κανόνα. Ο κανόνας αυτός τίθεται και οργανώνει την μη δομημένη εμπειρία και καθιστά τη νόηση τυπική δομή. Η ύπαρξη κανόνων για τη φιλοσοφία του Kant είναι καθοριστική και αφορά τόσο την a priori γνώση όσο και τις καθαρές έννοιες της νόησης (τις κατηγορίες). Η δυνατότητα να γνωρίσουμε ένα αντικείμενο μας παρέχεται μέσω των παραπάνω ‘εργαλείων’, τα οποία

¹¹⁶ Hobbes T., [1656]

¹¹⁷ Hume D., 1739-40]

ορίζονται από κανόνες. Αφού ο νους, λοιπόν, διέπεται από συγκεκριμένους κανόνες, τότε μπορεί να εξηγηθεί με βάση τις λειτουργίες του.

Η φιλοσοφία, λοιπόν, από τον 17ο αιώνα, προκειμένου να εξηγήσει τα νοητικά φαινόμενα, παύει να τα θεωρεί ως ποιότητες και αμφισβητεί την γνώση που προσφέρει η οντολογική προσέγγισή τους. Μέσα σε μια εξορθολογικοποιημένη φύση και σε έναν κόσμο που διέπεται από συγκεκριμένους και αυστηρούς κανόνες λειτουργίας, τα νοητικά φαινόμενα θα αναχθούν και θα εξηγηθούν, σύμφωνα με πολλούς στοχαστές, με τη βοήθεια μιας μεθόδου φορμαλιστικής και θα αντιμετωπιστούν ως λειτουργίες. Αυτό που είναι σημαντικό πλέον δεν είναι το ‘τι είναι’ τα νοητικά φαινόμενα, αλλά το ‘πώς λειτουργούν’. Με άλλα λόγια, μια θεμελιώδης κατεύθυνση της φιλοσοφίας της νεοτερικότητας επιχείρησε να εξηγήσει το νου με όρους φορμαλιστικούς και λειτουργικούς μετά από τον διαχωρισμό του νου από τον κόσμο της έκτασης – όμως, δεν ανήγαγε στο νατουραλιστικό επίπεδο την λειτουργία των νοητικών φαινομένων.

Το νέο στοιχείο που προστίθεται με τα επιστημολογικά προγράμματα της Γνωσιακής Επιστήμης και της ‘Τεχνητής Ζωής’, είναι ότι ο λειτουργισμός, σε συνδυασμό με την επιστημονική και τεχνολογική πρόοδο, επιχειρείται να μετατραπεί σε πρόγραμμα μιας εμπειρικής, μηχανιστικής επιστήμης (νατουραλισμός). Στην παρούσα εργασία επιχειρήσαμε να εκθέσουμε και να ασκήσουμε κριτική στα παραπάνω εγχειρήματα, δηλαδή στα επιστημολογικά προγράμματα της Γνωσιακής Επιστήμης και της ‘Τεχνητής Ζωής’ και, ειδικότερα, να ελέγξουμε κατά πόσο ο λειτουργισμός, ο οποίος χρησιμοποιείται από αυτά ως εξηγητικό μοντέλο για τη νόηση και τη ζωή, είναι σε θέση να προσφέρει πράγματι αυτήν την εξήγηση. Μια τέτοιου είδους εξήγηση δεν λαμβάνει υπόψη της, αποσιωπά και αφαιρεί από τα νοητικά φαινόμενα και τα φαινόμενα της ζωής τα ουσιώδη, συστατικά και

¹¹⁸ Κονδύλης, Π., [1983], σ. 354

ποιοτικά χαρακτηριστικά τους. Αυτό στο οποίο καταλήξαμε είναι ότι απλώς η περιγραφή των τυπικών λειτουργιών ενός συστήματος, και η συντακτική θέση των συμβόλων που αυτό περιέχει, δεν συνιστούν ικανοποιητική εξήγηση του νου και της ζωής.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Armstrong, D.M., [1968], *A Materialistic Theory of Mind*, London: Routledge and Kegan Paul
2. Audi R., [1995], *The Cambridge Dictionary of Philosophy*, Cambridge University Press
3. Bedau M.A., [1996], “The Nature of Life”, στο *The Philosophy of Artificial Life*, Boden M., Oxford University Press, New York, σσ. 332-357
4. Blackwell B., [1994], *A Companion to the Philosophy of Mind*, Guttenplan S. (ed.), Cambridge, Massachusetts.
5. Block N., [1990], “The Computer Model of the Mind”, in D. Osherson and E. E. Smith (eds.), *An Invitation to Cognitive Science 3*, Cambridge, MA: MIT Press, σσ. 247-289.
6. Boden M., [1996], *The Philosophy of Artificial Life*, Oxford University Press, New York

7. Campell K.K., [1970], *Body and Mind*, Doubledau, 1970
8. Chomsky N., [1959], “Review of Verbal Behavior”, by B. F. Skinner, *Language* 35: 26-58
9. Descartes R. [1976], *Λόγος περί της μεθόδου*, εισαγ., μτφ., σημειώσεις Χρ. Χρηστίδης, Β έκδοση, Παπαζήσης, Αθήνα
10. Dreyfus H.L., [1992/2001], *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές*, μτφ. Π. Καρλέτσα, Πανεπιστημιακές Εκδόσεις Κρήτης (ΠΕΚ), Ηράκλειο
11. Earman J. και Salmon W., [1999], “Επικύρωση επιστημονικών υποθέσεων”, στον συλλογικό τόμο *Εισαγωγή στη φιλοσοφία της επιστήμης*, Μ. Salmon, J. Earman κ.ά. (επιμ.), Μτφ. Π. Θεοδώρου, Κ. Παγωνδιώτης, Γ. Φουρτούνης, ΠΕΚ, Ηράκλειο, σσ. 51-137
12. Feigenbaum E.A. and Feldman J., (eds.), [1963], *Computers and Thought*. New York: McGray-Hill Company
13. Feigl H., [1958], “The ‘Mental’ and the ‘Physical’”, στο Feigl H., Scriven M. & Maxwell G. (eds.),: *Concepts, Theories, and the Mind Body Problem* (Minnesota Studies in the Philosophy of Science 2), University of Minnesota Press
14. Fodor J.A., [1975], *The Language of Thought*, J.J. Katz, D.T. Langendoen, G.A. Miller (eds.), Harvard University Press
15. Frege G., [1976], “The Thought: a Logical Inquiry”, (Translation A. and M. Quinton), στο P.F. Strawson (ed.),: *Philosophical Logic*, Oxford University Press
16. Glymour C., [1999], “Γνωσιολογία των ανδροειδών: Υπολογισμός, Τεχνητή Νοημοσύνη και φιλοσοφία της επιστήμης”, στον συλλογικό τόμο *Εισαγωγή στη φιλοσοφία της επιστήμης*. Μ. Salmon, J. Earman κ.ά. (επιμ.), Μτφ. Π. Θεοδώρου, Κ. Παγωνδιώτης, Γ. Φουρτούνης, ΠΕΚ, Ηράκλειο, σσ. 509-564
17. Goldman A.I., [1993/1998], *Γνωσιοεπιστήμη: Φιλοσοφικές εφαρμογές*, μτφ. Γ. Μαραγκός, Οδυσσέας, Αθήνα

18. Haugeland J., [1992], *Τεχνητή Νοημοσύνη*, Μτφ. Σ. Ζαχαρίου, Κάτοπτρο, Αθήνα
19. Hempel C.G., [1935/1980], “The Logical Analysis in Psychology”, στο Block 1980
20. Hobbes T., [1656], *Elements of Philosophy*, Αποσπάσματα του βιβλίου στο Molesworth, 1839-45, τόμ. Ι
21. Holland J. H., [1975], *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, (Ann Arbor: University of Michigan Press)
22. Hume D., [1739-40], *A Treatise of Human Nature*, αποσπάσματα του βιβλίου στο Selby-Bigge, 1888
23. Huxley T., [1995], “On the Hypothesis that Animals are Automata”, *Fortnightly Review* 95: 555-80. Reprinted in *Collected Essays*, London
24. Jackson F., [1982], Epiphenomenal Qualia, *Philosophical Quarterly* 32:127-136
25. Kant Im., [1929], *Critique of Pure Reason*, Macmillan Education
26. Κονδύλης Π.[1983], *Η κριτική της μεταφυσικής στη νεότερη σκέψη (Από τον όψιμο μεσαίωνα ως το τέλος του Διαφωτισμού)*, Γνώση, Αθήνα
27. Langton C.G., [1984], ‘Self-Reproduction in Cellular Automata’, *Physica D*, 10/1-2: 135-44
28. Langton C.G., [1986], ‘Studying Artificial Life with Cellular Automata’, *Physica D*, 22: 120-49
29. Langton C.G., [1989], *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 6: Redwood City, Calif.: Addison-Wesley)
30. Langton C.G., [1996], “Artificial Life”, στο Boden M. (ed.), *The Philosophy of Artificial Life*, Oxford University Press, New York, σσ. 39-94
31. Lewis D., [1972], “Psychological and Theoretical Identifications”, *Australasian Journal of Philosophy* 50, επανεκδόθηκε στο Block 1980

32. Lewis D., [1983], “Extrinsic Properties”, *Philosophical Studies* 44, σσ. 197-200
33. Machamer P., [1999], “Φιλοσοφία της ψυχολογίας”, στον συλλογικό τόμο *Εισαγωγή στη φιλοσοφία της επιστήμης*, Μ. Salmon, J. Earman κ.ά. (επιμ.), Μτφ. Π. Θεοδώρου, Κ. Παγωνδιώτης, Γ. Φουρτούνης, ΠΕΚ, Ηράκλειο, σσ. 483-508
34. Μαραγκός Γ., [1996], *Γνωσιολογισμοί*. Οδυσσέας, Αθήνα
35. McDowell J., [1994], *Mind and World*, Harvard University Press
36. Minsky M., [1994], “Will Robots Inherit the Earth?”, *Scientific American*, σσ. 87-91
37. Nagel T., [1974/1993], “What it is like to be a Bat?”, *Philosophical Review* 83, Επανεκδόθηκε στον συλλογικό τόμο Hofstadet D.R. και Dennet D.C., (1981), *The Mind's I, / Το εγώ της νόησης*, (1993): “Τι νοιώθει κανείς όντας νυχτερίδα;”, Μτφ. Μ. Αντωνοπούλου, Κάτοπτρο, Αθήνα, σσ. 453-468
38. Neumann J. von, [1956], “The General and Logical Theory of Automata”, ανατύπωση στο *The World of Mathematics* (Νέα Υόρκη: Simon and Schuster), σ. 2077
39. Neumann von, J., [1966], *Theory of Self-Reproducing Automata*, (ed.) and completed by A. W. Burks (Urbana, III, University of Illinois Press)
40. Newell A.J.C., και Simon H.A., [1961], *Computer Simulation of Human Thinking*, The RAND Corporation, P-2276 (20 Απριλίου)
41. Newell A.J.C., Shaw και Simon H.A., [1963], “Chess-Playing Programs and the Problem of Complexity”, στο *Computers and Thought*, Edward A., Feigenbaum Julian Feldman (επιμ.). (Νέα Υόρκη: McGraw-Hill)
42. Newell A.J.C., Shaw και Simon H.A., [1964], “Elements of a Theory of Human Problem Solving”, στο Harper R.J.C. et al.: *The Cognitive Process, Readings, Englewood Cliffs*, N.J.
43. Παγωνδιώτης Κ., [1999], “Υπάρχει γλώσσα της σκέψης;”, στο *Γλώσσα και Νόηση*, Κατή Δ., Κονδύλη Μ. & Νικηφορίδου Κ. (επιμ.), εκδ. Αλεξάνδρεια, Αθήνα

44. Παγωνδιώτης Κ., [2001], *Το πρόβλημα των νοητικών αναπαραστάσεων στη γνωσιακή επιστήμη: προς μια μη-αναπαραστασιακή περιγραφή των νοητικών φαινομένων*, Διδακτορική διατριβή, Αθήνα
45. Παπαδημητρίου Ε. Γ., [1988], *Θεωρία της επιστήμης και ιστορία της φιλοσοφίας*, Gutenberg, Αθήνα
46. Parineau D., [1993/2002], *Philosophical Naturalism*, Leader Books 1993/ *Φιλοσοφικός Νατουραλισμός*, Μτφ. Τ. Μπούκη, Leader Press, Αθήνα
47. Place U.T, [1956], “Is Consciousness a Brain Process?”, *British Journal of Psychology*, 47
48. Putnam H., [1960], “Minds and Machines”, in *Dimensions of Mind*, S. Hook (ed.), New York: Collier Books
49. Putnam H., [1963], “Brains and Behavior,” in R. Butler, (ed.), *Analytical Philosophy*, Second Series, Oxford: Basil Blackwell
50. Putnam H., [1967], “The Mental Life of Some Machines”, H.N. Castaneda (ed.), *Intentionality, Minds and Perception*, Wayne State University Press
51. Putnam H., [1975], *Language, Mind and Knowledge*, Minneapolis: Minnesota Press
52. Ray, T.S., [1996], “An Approach to the Synthesis of Life”, στο Boden M., (ed.), *The Philosophy of Artificial Life*, Oxford University Press, New York, σσ. 111-145
53. Robinson W.S., [1969], “Epiphenomenalism, Laws and Properties”, *Philosophical Studies* 69: 1-34
54. Salmon W.C., [1999], “Επιστημονική εξήγηση”, στον συλλογικό τόμο *Εισαγωγή στη φιλοσοφία της επιστήμης*, Μ. Salmon, J. Earman κ.ά. (επιμ.), Μτφ. Π. Θεοδώρου, Κ. Παγωνδιώτης, Γ. Φουρτούνης, ΠΕΚ, Ηράκλειο, σσ. 3-50
55. Schank R.C. et al., [1977], “Panel on Natural Language Processing” (IJCAI-77, *Proceedings*)

56. Searle J., [1980/1993], “Minds, Brains and Programs”, *The Behavioral and Brain Science* 3: 417-424, Επανεκδόθηκε στον συλλογικό τόμο Hofstadet D.R., και Dennet D.C., (1981), *The Mind's I/ Το εγώ της νόησης*, (1993): “Νοήσεις, εγκεφαλοι και προγράμματα”, Μτφ. Μ. Αντωνοπούλου, Κάτοπτρο, Αθήνα, σσ. 407-430
57. Searle J., [1984/1994], *Νους, Εγκέφαλος και Επιστήμη*, Μτφ. Κ. Χατζηκυριάκου, ΠΕΚ, Ηράκλειο
58. Searle J., [1992/1997], *Ανακαλύπτοντας ξανά το νου*, μτφ. Μ. Κορνήλιος, Γκοβόστης, Αθήνα
59. Searle J., [1990], “Is the Brain a Digital Computer?”, Address APA, *Proceedings of American Assosiation*
60. Skinner B.F., [1953], *Science and Human Behavior*, New York: Macmillan
61. Sloman A., [1978], “The Computer Revolution in Philosophy”, (*Philosophy, Science and Models of Mind*), Harvester Press and Humanities Press
62. Smart J.J.C., [1959/1990], “Sensations and Brain Processes”, *Philosophical Review* 68, (Επανεκδόθηκε στο Rosenthal D., (ed.), *The Nature of Mind*, Oxford University Press)
63. Sober E., [1996], “Learning from Functionalism – Prospects for Strong Artificial Life”, στο Boden M., (ed.), *The Philosophy of Artificial Life*, Oxford University Press, New York, σσ. 361-378
64. Taubes G., [2000], “Biologists and Create a New Generation of Robots That Imitate Life”, *Science*, Vol. 288, σσ. 80-83
65. Τσινόρεμα Σ., [1999], *Πανεπιστημιακές παραδόσεις*
66. Turing A.M., [1950], “Computing Machinery and Intelligence”, *Mind* 59: 433-460
67. Watson J., *Psychology from the Standpoint of a Behaviorist*, Philadelphia: Lippincott, 1919

68. Wheeler M., [1996], "From Robots to Rothko: The Bringing forth of Worlds", στο Boden M., (ed.), *The Philosophy of Artificial Life*, Oxford University Press, New York, σσ. 209-236
69. Wittgenstein L., [1953], *Philosophical Investigations*, (translator G.E.M. Anscombe), Blackwell