ΚΕΦΑΛΑΙΟ 5

Μοιάζουν οι υπολογιστές με εμάς; Η άνοδος και η πτώση της τεχνητής νοημοσύνης

Τον Μάιο του 1997 ο υπολογιστής της IBM Deep Blue αντιμετώπισε σε μια παρτίδα σκάκι τον παγκόσμιο πρωταθλητή Γκάρι Κασπάροφ. Για μία εβδομάδα η πρόοδος του υπολογιστή εμφανιζόταν στα πρωτοσέλιδα των εφημερίδων – η παρτίδα αυτή είχε τη μεγαλύτερη δημοσιογραφική κάλυψη σε ολόκληρη την ιστορία αυτού του αρχαίου παιχνιδιού. Όταν ο Κασπάροφ έχασε, φάνηκε σαν να είχαμε μπει σε μια νέα εποχή. Ο υπολογιστής είχε τελικά θριαμβεύσει σε ένα διαγωνισμό που τον είχε φέρει αντιμέτωπο με τις ανθρώπινες ικανότητες σκέψης.

Πολλά θα μπορούσε να πει κανείς για να υπερασπιστεί τον άνθρωπο. Μία παρτίδα σκάκι έχει εκατομμύρια πιθανούς συνδυασμούς κινήσεων και αυτό το κάνει εξαιρετικά κατάλληλο για υπολογιστική ανάλυση. Επίσης ο υπολογιστής είχε σχεδιαστεί και προγραμματιστεί από ανθρώπους, επομένως ο αγώνας ήταν στην πραγματικότητα ένας θρίαμβος των ανθρώπων σχεδιαστών και των προγραμματιστών και όχι της μηχανής καθεαυτής. Παρ' όλα αυτά, ήταν έκδηλη η αίσθηση ότι η ψηφιακή εποχή είχε δημιουργήσει μια μηχανή που είχε μάθει να *σκέφτεται*. Σύμφωνα με τα λόγια του Grandmaster Yasser Seirawan, «Τι άλλο κάνουμε όταν παίζουμε ένα παιχνίδι σκακιού παρά σκεφτόμαστε;».

Αν μια μηχανή μπορεί να κερδίζει τον παγκόσμιο πρωταθλητή μιας από τις πιο απαιτητικές πνευματικά ασχολίες της ανθρωπότητας, σίγουρα θα μπορεί να αντικαταστήσει τους ανθρώπους και στις υπόλοιπες ασχολίες τους: την προώθηση προϊόντων, τα χρηματιστηριακά, την ιατρική και τη δικηγορία; Ή, αν όχι να τους αντικαταστήσει, να τους παρέχει προγράμματα τεχνητής νοημοσύνης ώστε ο καθένας να είναι πλέον ισάξιος με τα μεγαλύτερα μυαλά στο πεδίο του; Οι επιπτώσεις για την αξία της απλής ανθρώπινης σκέψης μοιάζουν τρομακτικές.

Η τεχνολογία είχε κατακτήσει το πρώτο επίπεδο της ιεραρχίας επικοινωνία-επεξεργασία-σκέψη όταν δημιουργήθηκε η υποδομή των τηλεπικοινωνιών με τα καλώδια, τις οπτικές ίνες και τα ραδιοκύματα. Είχε συνθλίψει το δεύτερο όταν εμφανίστηκαν οι ψηφιακοί υπολογιστές υψηλής ταχύτητας. Η μόνη πρόκληση που απέμενε ήταν να κατακτήσει και το τρίτο επίπεδο, να κατασκευάσει δηλαδή μια σκεπτόμενη μηχανή.

Στις αστυνομικές σειρές της τηλεόρασης επικρατεί το συναίσθημα. Οι κατώτεροι φωνάζουν στους διοικητές τους και αυτοί απαντούν. Βρίζονται μεταξύ τους και κλείνουν τις πόρτες με δύναμη. Ζουν στιγμές ενθουσιασμού ή απαισιοδοξίας, ακολουθούν το προαίσθημά τους, βρίσκουν στοιχεία σε ένα νεύμα ή μια ματιά και στο τέλος εξιχνιάζουν το έγκλημα. Η ζωή των πραγματικών αστυνομικών είναι περισσότερο βαρετή, έχει περισσότερη γραφειοκρατία και το ποσοστό εξιχνίασης εγκλημάτων είναι χαμηλότερο. Αυτό όμως που ισχύει τόσο στην πραγματικότητα όσο και στη μυθοπλασία είναι ότι η ατμόσφαιρα στο περιβάλλον εργασίας καθορίζεται κυρίως από την αλληλεπίδραση μεταξύ των ανθρώπων.

Το ίδιο ισχύει και στις άλλες επιχειρήσεις. Ανεξάρτητα

από τη δομή και τη φύση της εργασίας κάθε ανθρώπου σε ένα γραφείο, δισεκατομμύρια από τα εγκεφαλικά του κύτταρα μηρυκάζουν τα αιώνια ζητήματα διαπροσωπικών σχέσεων: Γιατί το αφεντικό με κοιτάζει έτσι; Γιατί εγώ εργάζομαι περισσότερο από τον Τζο; Τι πρέπει να κάνω για να πάρω προαγωγή ή αύξηση; Τι μπορώ να κάνω για να βρω μια καλύτερη δουλειά εδώ ή μια ακόμη καλύτερη στη νέα εταιρεία που άνοιξε απέναντι;

Τόσο οι υποστηρικτές όσο και οι πολέμιοι των υπολογιστών αναγνωρίζουν ότι αυτοί πρέπει να αναπτύξουν πιο ανθρώπινη προσωπικότητα, πριν αποκτήσουν σημαντικότερες θέσεις μέσα στις εταιρείες. Θα χρειαστεί να αποκτήσουν χιούμορ, διαίσθηση, ίσως ακόμη και περιστασιακούς φόβους, αλλιώς οι άνθρωποι δεν θα είναι διατεθειμένοι να συναναστρέφονται μαζί τους και οι υπολογιστές δεν θα μπορούν να τους αντιμετωπίσουν. Οι υπολογιστές θα πρέπει να επεξεργάζονται τις πληροφορίες, όχι με τη στενή έννοια του χειρισμού συμβόλων σύμφωνα με την αυστηρή λογική μιας μηχανής Τούρινγκ, αλλά με την ευρεία έννοια, κατά την οποία το ανθρώπινο μυαλό συσσωρεύει τη γνώση.

Εν ολίγοις, η νοημοσύνη των υπολογιστών πρέπει να γίνει πιο ανθρώπινη. Και υπάρχουν ερευνητές στα καλύτερα πανεπιστήμια και εργαστήρια ερευνών που πιστεύουν ότι μπορούν να πετύχουν κάτι τέτοιο.

Η ιδέα για τη δημιουργία τεχνητής νοημοσύνης –AI (Artificial Intelligence) όπως λέγεται–, μιας μηχανής δηλαδή που να μιμείται την ανθρώπινη ικανότητα σκέψης, υπάρχει από τότε που εμφανίστηκαν οι υπολογιστές. Το πρώτο λογισμικό τεχνητής νοημοσύνης δημιουργήθηκε ως ιδέα το 1955 από τον Herbert Simon, καθηγητή Ψυχολογίας και Επιστήμης των Υπολογιστών στο Πανεπιστήμιο Κάρνεγκι Μέλον, ο οποίος αργότερα κέρδισε το βραβείο Νόμπελ στα Οικονομικά. Μαζί με το συνάδελφό του Alan

Newell κατασκεύασαν ένα πρόγραμμα που μπορούσε να παράγει αποδείξεις για απλές μαθηματικές προτάσεις. Το ονόμασε Λογικό Θεωρητικό και ανακοίνωσε στους έκπληκτους φοιτητές του ότι είχε εφεύρει μια «σκεπτόμενη μηχανή». Το ότι η πρώτη εφαρμογή αυτής της μηχανής στην επίλυση προβλημάτων ήταν στον τομέα των μαθηματικών δεν προκαλεί έκπληξη, καθώς η σχέση μεταξύ των υπολογιστών και τις λογικής ανάλυσης είναι ένα διαρκές χαρακτηριστικό το οποίο θα αποτελούσε τόσο το μεγαλύτερο προτέρημά τους όσο και το μεγαλύτερο περιορισμό τους.

Ο Τούρινγκ ήταν αυτός που είχε θέσει το πλαίσιο γι' αυτή τη διένεξη, όπως και για πολλά άλλα θέματα στον τομέα των υπολογιστών. Το 1950, μιάμιση δεκαετία αφού είχε σχεδιάσει τη φανταστική μηχανή του, δημοσίευσε μια μελέτη με τον τίτλο «Μηχανήματα Υπολογισμών και Νοημοσύνης». Στη μελέτη αυτή υποστήριζε ότι το ερώτημα «Μπορεί μια μηχανή να σκεφτεί όπως ένας άνθρωπος;» θα μπορούσε να απαντηθεί αν παρατηρούσαμε τον τρόπο που *αλληλεπιδρά* η μηχανή με ανθρώπους. Με άλλα λόγια, το κριτήριο του Τούρινγκ, που αφορά τόσο τη φιλοσοφική όσο και την πρακτική πλευρά του θέματος, ήταν η ικανότητα της μηχανής να «μιμείται ανθρώπους». Αυτό που ονόμασε Παιχνίδι της Μίμησης –σήμερα λέγεται Τεστ Τούρινγκ– λειτουργεί ως εξής: εθελοντές κάθονται μπροστά από τερματικά που είναι συνδεδεμένα είτε με έναν άλλο άνθρωπο είτε με τη μηχανή που πρόκειται να δοκιμαστεί (αυτός που υποβάλλει τις ερωτήσεις δεν ξέρει ποιο από τα δύο συμβαίνει). Αν μπερδέψουν τις απαντήσεις του υπολογιστή με αυτές του ανθρώπου, το μηχάνημα έχει περάσει με επιτυχία τη δοκιμή.

Στα μέσα της δεκαετίας του 1960 ο Άρθουρ Κλαρκ μαζί με τον σκηνοθέτη Στάνλεϊ Κιούμπρικ έγραψαν το σενάριο για το έργο *2001: Η Οδύσσεια του Διαστήματος*. Σύμφωνα

με αυτό, ο υπολογιστής HAL έχει εμφανώς ανθρώπινες ιδιότητες – όχι μόνο νοημοσύνη αλλά και συναισθήματα και αυτοσυνειδησία. Η προηγούμενη πρόβλεψη του Κλαρκ για τον τομέα των υψηλών τεχνολογιών –η πρόβλεψη που είχε κάνει το 1945 ότι δορυφόροι θα χρησιμοποιούνταν στο διάστημα για την ανταλλαγή πληροφοριών– είχε ήδη επαληθευτεί. Η νέα πρόβλεψη για την κατασκευή υπολογιστών με ανθρώπινες ιδιότητες στην αρχή της νέας χιλιετίας φαινόταν όλο και περισσότερο πιθανή και μέχρι τη δεκαετία του 1970 εμφανίζονταν όλο και περισσότερα εργαστήρια τεχνητής νοημοσύνης σε ιδρύματα ερευνών και πανεπιστήμια. Η τεχνητή νοημοσύνη ήταν η νέα μόδα στη βιομηχανία των υπολογιστών.

Ήταν γνωστό ότι η αυστηρά σειριακή λογική, που υπήρχε στις ψηφιακές μηχανές, απείχε πολύ από τις διαδικασίες σκέψης του ανθρώπινου εγκεφάλου. Συνεπώς, η ηλεκτρονική τεχνολογία έπρεπε να ξεφύγει από τα δεσμά του μοντέλου γραμμικού υπολογισμού. Διάφορες ιδέες, βασισμένες στη δομή του εγκεφάλου, εμφανίστηκαν για να βοηθήσουν τους επιστήμονες στη νέα προσπάθειά τους.

Οι νευρολόγοι εξέτασαν προσεκτικά τα εγκεφαλικά κύτταρα και ανακάλυψαν ότι ο ανθρώπινος εγκέφαλος περνάει από δύο διακριτές διαδικασίες κατά τη μάθηση. Καθώς αρχίζουμε να σκεφτόμαστε, οι νευρώνες στον εγκέφαλό μας εκπέμπουν σήματα ο ένας προς τον άλλο κατά μήκος των δενδροειδών αξόνων και διακλαδώσεων που τους συνδέουν. Αλλά και οι ίδιες οι συνδέσεις εξελίσσονται με το χρόνο, γίνονται πιο δυνατές όσο επαναλαμβάνεται μια συγκεκριμένη δραστηριότητα, έτσι ώστε η δραστηριότητα αυτή «κωδικοποιείται» σε ένα βαθύτερο επίπεδο από το συνειδητό. Η ικανότητα του εγκεφάλου να μαθαίνει μέσω της ενίσχυσης αυτών των συνδέσεων –των συνάψεων– λέγεται *πλαστικότητα* και είναι περισσότερο

εμφανής στην παιδική ηλικία. Αντίθετα, η ικανότητα χρήσης του δικτύου των νευρώνων, μόλις δημιουργηθούν συνδέσεις, λέγεται *ισχύς* του εγκεφάλου και βρίσκεται στην ακμή της κατά την ενηλικίωση. Η διαφορά ανάμεσα στην πλαστικότητα και την ισχύ μπορεί να φανεί από το διαφορετικό τρόπο που μαθαίνουν μια γλώσσα ένα παιδί και ένας ενήλικας. Όταν ο Βιμ Ντούιζενμπεργκ, ο Ολλανδός ηγέτης της Ευρωπαϊκής Κεντρικής Τράπεζας, εμφανίζεται στην αμερικάνικη τηλεόραση για να εξηγήσει τη νομισματική πολιτική, συνθέτει περίπλοκες προτάσεις με μια γραμματική τέχνη που ένα παιδί από το Μπρονξ δεν μπορεί να συναγωνιστεί. Αλλά όταν ένα παιδί από το Μπρονξ μιλάει, το κάνει με απόλυτη φυσικότητα και χωρίς καμία προσπάθεια. Αντίθετα, η επιτηδευμένη ακρίβεια που χρησιμοποιεί ο Ολλανδός αποκαλύπτει ότι η γλώσσα που χειρίζεται είναι ξένη γι' αυτόν και χρειάζεται τις ικανότητες ανάλυσης που διαθέτει, για να συνθέσει αυτές τις προτάσεις. Ο Ντούιζενμπεργκ έχει μάθει να μιλάει Αγγλικά χρησιμοποιώντας την *ισχύ* του εγκεφάλου του, ενώ το παιδί από το Μπρονξ χρησιμοποιεί την *πλαστικότητά* του.

Το 1994 το Γερμανικό Βραβείο Επιχειρηματικής Καινοτομίας απονεμήθηκε στη Siemens Nixdorf για μια μηχανή που ονόμασαν Νευρωνικό Υπολογιστή Σύναψης-Ι (Synapse-I Neural Computer). Αυτή η ευρωπαϊκή επιχείρηση, που συνήθως δεν δίνει έμφαση σε λεκτικές υπερβολές, το διαφήμισε ως «σκεπτόμενο υπολογιστή», που μπορούσε να μάθει και να διδαχτεί. Δήλωσε ότι ήταν ικανός να πραγματοποιεί ιατρικές αναλύσεις, να ελέγχει περίπλοκες βιομηχανικές διαδικασίες και να «προβλέπει τη διακύμανση των επιτοκίων με μεγαλύτερη ακρίβεια από τους οικονομολόγους».

Οι μηχανές νευρωνικών δικτύων, των οποίων η Σύναψη-Ι είναι ένα παράδειγμα, είναι η τελευταία εκδήλωση της τάσης που ξεκίνησε στην αρχή της δεκαετίας του 1980,

ακολουθώντας την πρωτοποριακή εργασία του Ινστιτούτου Τεχνολογίας της Καλιφόρνια. Σε αυτή την εργασία, αντί να σχεδιαστεί ένας υπολογιστής πάνω σε έναν πολύ γρήγορο και πολύ προηγμένο κεντρικό επεξεργαστή, συνδέθηκαν παράλληλα πάρα πολλές μικροσκοπικές μονάδες επεξεργασίας με έναν τρόπο που θύμιζε τις συνάψεις των νευρώνων στον εγκέφαλο. Εταιρείες, με ονόματα όπως Εταιρεία Σκεπτόμενων Μηχανών, κατασκεύαζαν υλικό για την εκτέλεση των καινούργιων προγραμμάτων, υλικό με πολλούς επεξεργαστές που εργάζονταν παράλληλα και επέτρεπαν σε πολλές διαφορετικές «γραμμές λογικής» να εξετάζονται ταυτόχρονα πριν συγκλίνουν σε λύσεις.

Το όνομα που διάλεξε η Siemens Nixdorf για το μηχάνημά της είναι αποκαλυπτικό. Οι νευρώνες, οι συνάψεις, η ανατομία του εγκεφάλου γενικότερα χρησιμοποιήθηκε και σε άλλες περιπτώσεις ως έμπνευση για το σχεδιασμό υπολογιστών. Στα νευρωνικά δίκτυα υπολογιστών οι συνδέσεις μεταξύ των λογικών στοιχείων έχουν κατασκευαστεί ώστε να διαφέρουν, με τυχαίο τρόπο, σε ισχύ, ώστε να θυμίζουν τις διαφορές στα νευρικά σήματα που ταξιδεύουν κατά μήκος των συνάψεων στον εγκέφαλο. Ο υπολογιστής διδάσκεται μια συγκεκριμένη δραστηριότητα επαναληπτικά και καθώς μαθαίνει, η δύναμη των «συναπτικών» συνδέσεων προσαρμόζεται σταδιακά, μέχρι που η δραστηριότητα να γίνει περισσότερο ή λιγότερο γνωστή. Με άλλα λόγια, ο υπολογιστής δεν έχει προγραμματιστεί από την αρχή σύμφωνα με μια λογικά καθορισμένη μέθοδο, αλλά εξελίσσει σταδιακά ένα σύνολο συνδέσεων ανάμεσα σε πολλά ανεξάρτητα λογικά στοιχεία – μάλλον με τον τρόπο που ένα παιδί, μαθαίνοντας να κάνει ποδήλατο, προσαρμόζει ασυνείδητα τις συνδέσεις του εγκεφάλου έτσι ώστε να αντικατοπτρίζουν τις νέες γνώσεις που αποκτήθηκαν. Ο προγραμματιστής δεν ξέρει τον τρόπο με τον

οποίο μαθαίνει ο υπολογιστής, απλά βεβαιώνει ότι ο υπολογιστής συνεχίζει να προσπαθεί και μπορεί να ξεχωρίσει την επιτυχία από την αποτυχία.

Οι εταιρείες στο Χρηματιστήριο της Γουόλ Στριτ χρησιμοποιούν τέτοιες τεχνικές για να εντοπίζουν τις τάσεις στις τιμές των μετοχών μέσα στη χιονοστιβάδα των δεδομένων που προκύπτουν μετά από κάθε συνεδρίαση. Τα νευρωνικά δίκτυα, όπως λέγονται εν συντομία, μπορούν να ξεχωρίζουν πολύ εύκολα τα μοτίβα στα δεδομένα, τις τάσεις ή τους συσχετισμούς, που είναι πολύ δύσκολο να εντοπιστούν με τη χρήση παραδοσιακών μαθηματικών εργαλείων, γιατί ίσως να μη γνωρίζεις ποιο μοτίβο ψάχνεις. Η χρήση της τεχνικής των νευρωνικών δικτύων, από τη MasterCard International, για την αναγνώριση μοτίβων απάτης με πιστωτικές κάρτες, έχει αποτρέψει απώλειες που υπολογίζεται ότι φτάνουν τα 30 εκατομμύρια δολάρια ετησίως.

Μια περαιτέρω καμπή της βιολογικής αναλογίας γεννήθηκε από τη δαρβινική θεωρία της εξέλιξης των ειδών και ονομάστηκε γενετικός αλγόριθμος. Η ιδέα εδώ δεν είναι να κατασκευαστεί λογισμικό στην τελική του μορφή, αλλά να το αφήσουμε να δημιουργηθεί μέσω της δοκιμής και του λάθους. Μικρά κομματάκια κώδικα (ονομάστηκαν χρωμοσώματα, άλλος ένας όρος δανεισμένος από τη βιολογία) ενώνονται τυχαία σε πολλούς συνδυασμούς και τοποθετούνται σε ένα περιβάλλον όπου μόνο οι «δυνατότεροι» –δηλαδή αυτοί που μπορούν να αναπαραχθούν– επιβιώνουν. Το πρόγραμμα γράφεται με τρόπο που επιτρέπει στα χρωμοσώματα να αναπαράγουν τον εαυτό τους. Τυχαίες μεταλλάξεις τοποθετούνται για να βεβαιώσουν, όπως και στη βιολογική εξέλιξη, ότι υπάρχει το δυναμικό για πρόοδο. Οι χρηματιστηριακές εφαρμογές του γενετικού αλγορίθμου ήδη κερδίζουν την αποδοχή, καθώς το

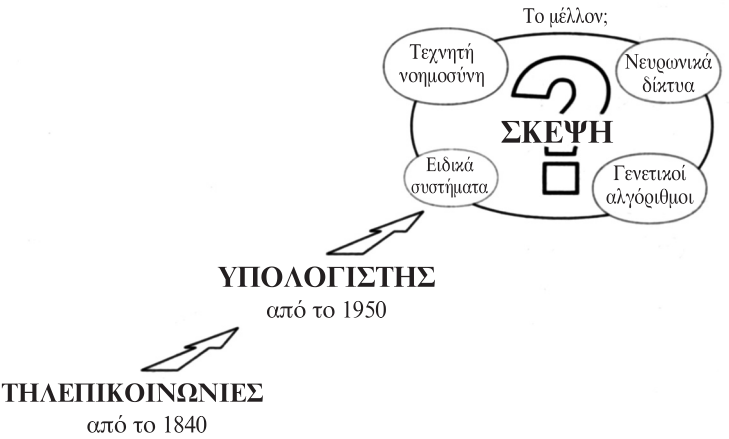
λογισμικό προσφέρει στοιχεία για τάσεις που ούτε η τεχνολογία των νευρωνικών δικτύων δεν μπορούσε να βρει.

Οι φανατικοί οπαδοί της τεχνητής νοημοσύνης βλέπουν τα έμπειρα συστήματα, τα νευρωνικά δίκτυα και τους γενετικούς αλγόριθμους ως σταδιακά βήματα –που το καθένα έχει τύχει αποδοχής στο σκληρό κόσμο του εμπορίου– προς τον τελικό στόχο δημιουργίας μιας μηχανής που να επιδεικνύει πνευματικές ιδιότητες ισάξιες με τις δικές μας.

Για τον Marvin Minsky του MIT, μία από τις εξέχουσες φυσιογνωμίες στην κοινότητα της τεχνητής νοημοσύνης, ο εγκέφαλος, παρόλο που πολλές από τις μικροδομές του παραμένουν μυστήριο, είναι απλά μια μηχανή στην οποία ηλεκτρικά και χημικά σήματα αλληλεπιδρούν και αποθηκεύονται. Εν ευθέτω χρόνω, πιστεύει, θα καταφέρουμε να καταλάβουμε τη δομή του και να την αντιγράψουμε ή τουλάχιστον να την αντιγράψουμε αρκετά πιστά, ώστε να δημιουργήσουμε μηχανές που θα έχουν σε γενικές γραμμές ανθρώπινη λογική.

Ωστόσο, φιλόσοφοι και επιστήμονες –και πολλοί από εμάς τους υπόλοιπους– έχουν την αίσθηση ότι η σκέψη είναι κάτι περισσότερο από υπολογισμούς. Η ιδέα ότι το μυαλό μας υπερβαίνει, κατά κάποιον τρόπο, τα όρια της απόλυτα λογικής σκέψης μάς ελκύει ενστικτωδώς, ωστόσο είναι δύσκολο να προσδιοριστεί. Η φιλοσοφία έχει αναγνωρίσει εδώ και πολύ καιρό ότι υπάρχει μια ριζική διαφορά μεταξύ της σύνταξης (του χειρισμού των συμβόλων) και της σημασιολογίας (της σημασίας). Αν μεταφέρουμε αυτόν το διαχωρισμό στη διαμάχη της πληροφορικής, θα λέγαμε ότι, παρόλο που οι υπολογιστές χειρίζονται *σύμβολα*, δεν είναι το ίδιο με το να τους προσδίδουν *σημασία*. Τα επιχειρήματα κατά της δυνατότητας της τεχνητής νοημοσύνης προβάλλουν το ερώτημα, με διαφορετικούς τρό-

που, αν η σκέψη είναι κάτι περισσότερο από το απλό ανακάτωμα συμβόλων σύμφωνα με μηχανικούς ή αλγοριθμικούς κανόνες.



Επικοινωνία, επεξεργασία και μετά... μια σκεπτόμενη μηχανή; Η επίτευξη αυτού του στόχου αντιμετωπίζει φιλοσοφικά και πρακτικά εμπόδια, που η πλειοψηφία θεωρεί αξεπέραστα.

Η διαμάχη αυτή έφτασε σε ένα σταυροδρόμι το 1931 με τη δημοσίευση του περίφημου Θεωρήματος της Μη Πληρότητας από τον Αυστριακό λογικολόγο Kurt Gödel. Ο σύγχρονός του Άλαν Τούρινγκ είχε δείξει με τη φανταστική μηχανή του ότι, αν δοθούν επαρκής χρόνος και μνήμη σε μια συσκευή με απλά κυκλώματα πυλών, μπορούσε να λύσει οποιοδήποτε πρόβλημα υπόκειται σε λογική επίλυση – οποιοδήποτε πρόβλημα μπορούσε να λυθεί αν «διαλυόταν» σε βασικά λογικά βήματα και μετά επανενωνόταν σ' ένα σύνολο. Ο Gödel, όπως και ο Τούρινγκ, ενδιαφερόταν να εξερευνήσει τα όρια της δύναμης της λογικής

σκέψης. Ανακάλυψε ότι, αν κάποιος ξεκινήσει με ένα σύνολο (λογικά συνεπών) προτάσεων που οδηγούν σε συγκεκριμένες συνέπειες ή συμπεράσματα, τότε η δύναμη της λογικής δεν θα μπορεί να οδηγήσει από αυτές τις θέσεις σε *όλες* τις συνέπειες ή τα συμπεράσματα.

Η ενστικτωδώς ελκυστική ιδέα ότι ο εγκέφαλός μας μπορεί κατά μία έννοια να υπερβεί τους περιορισμούς της αυστηρά λογικής σκέψης πήρε μια εντυπωσιακή ώθηση από το Θεώρημα της Μη Πληρότητας του Gödel. Με απλοποιημένους όρους έδειξε ότι δεν είναι δυνατό να επιλυθούν λογικά όλα τα λογικά προβλήματα (πόσο μάλλον αυτά που δεν εντάσσονται σε λογικούς όρους). Σύμφωνα με τους όρους του Τούρινγκ, δεν είναι όλα τα προβλήματα υπολογίσιμα.

Η ανακάλυψη έδωσε ένα καίριο χτύπημα στην ελπίδα που είχε αρχίσει να μεγαλώνει μεταξύ των φιλοσόφων ότι υπήρχε μια τελειότητα στη λογική, κάτι που πραγματικά θα την καθιστούσε το «χρυσό κανόνα» για την ανθρώπινη σκέψη. Ήταν μια ελπίδα που είχε ξεκινήσει από τον Πυθαγόρα και τον Πλάτωνα, αλλά εμφανιζόταν ιδιαίτερα έντονη στο ξεκίνημα του εικοστού αιώνα, πιο αξιοσημείωτα στις θεωρίες του Βρετανού φιλοσόφου Μπέρτραντ Ράσελ. Ο Ράσελ ενοχλήθηκε επειδή φαινόταν να υπάρχουν κενά στο λογικό πνεύμα, ιδιαίτερα από το γεγονός ότι μπορείς να σχηματίσεις προτάσεις που είναι απόλυτα σωστές στην κατασκευή τους, αλλά αποδεικνύονται λογικά παράδοξα. Για παράδειγμα:

> Υπάρχει ένας κουρέας στο χωριό που ξυρίζει όλους τους άντρες (και μόνο αυτούς) που δεν ξυρίζονται μόνοι τους.

Ακούγεται ως απόλυτα λογική πρόταση. Πραγματικά, αποτελεί μια λογική κατάσταση σχέσεων: κάθε άντρας ξυ-

ρίζεται μόνος του ή πάει στον κουρέα. Δυστυχώς, ένα παράδοξο προκύπτει στην ερώτηση αν ο *κουρέας* ξυρίζεται μόνος του ή όχι. Μια ολιγόλεπτη ανάλυση της πρότασης αποκαλύπτει ότι αν δεν ξυρίζεται μόνος του, τότε πρέπει να τον ξυρίζει ο κουρέας, δηλαδή ο εαυτός του, ενώ αν ο κουρέας ξυρίζεται μόνος του, δεν τον ξυρίζει ο κουρέας! Η πρόταση είναι, μιλώντας αυστηρά, χωρίς νόημα ή παράδοξη. Το παράδοξο του κουρέα, μεταφρασμένο στην επίσημη γλώσσα της λογικής, προβάλλει το ερώτημα:

> Μπορεί να υπάρχει ένα σύνολο όλων των συνόλων τα οποία δεν είναι μέλη του εαυτού τους;

Η απάντηση είναι ότι δεν υπάρχει απάντηση. Η πρόταση αυτή είναι εξίσου παράλογη σε αυτή την περισσότερο μαθηματική μορφή της.

Παγίδες όπως αυτή ανησυχούσαν πολύ τον Ράσελ. Αν θέλαμε να λειτουργήσει η λογική σκέψη, ήταν απαραίτητο να υπάρχει ένα σύνολο αξιωμάτων ή σημείων εκκίνησης που να αποτρέπουν παράλογες παγίδες όπως αυτή που αφορούσε τον κουρέα του χωριού.

Μαζί με ένα συνάδελφο, τον Alfred North Whitehead, ο Ράσελ ξεκίνησε να γράφει το *Principia Mathematica* από το 1910 ως το 1913, με στόχο να θέσουν ένα ολοκληρωμένο πλαίσιο λογικών θεμελίων που δεν έπεφταν σε αυτή την παγίδα. Ήταν ένα εντυπωσιακό εγχείρημα, αλλά απέτυχε. Όταν δημοσιεύτηκε το Θεώρημα του Gödel, ο κόσμος κατάλαβε (όπως και ο Ράσελ, που έζησε μέχρι το 1970) ότι το εγχείρημα αυτό *δεν θα μπορούσε ποτέ* να είχε πετύχει. Δεν υπάρχει πλήρες και συνεπές σύστημα λογικής: ένα λογικό σύστημα είναι αδύνατο να αποδείξει τις ίδιες του τις θέσεις χωρίς να χρησιμοποιήσει κάτι έξω από τον εαυτό του.

Το Θεώρημα του Gödel έγινε ένα σημείο καμπής στον

τρόπο που σκεφτόμαστε για τη λογική. Έκανε για τη μαθηματική λογική ό,τι έχει κάνει για τη φυσική η Κβαντική Θεωρία: έφερε ξανά ένα στοιχείο ασάφειας σε αυτό το πεδίο που έμοιαζε να γίνεται όλο και πιο ξεκάθαρο. Η διάκριση των Gödel-Τούρινγκ, μεταξύ προβλημάτων που είναι επιλύσιμα από αυτά που μπορούμε να τα περιγράψουμε ως προϊόντα μηχανικής και αυτών που δεν επιλύονται, είναι ένα από τα θεμέλια του επιχειρήματος ότι οι υπολογιστές δεν θα αντιγράψουν ποτέ την ανθρώπινη σκέψη. Με αυτή τη βάση ο φιλόσοφος John Lucas, του Πανεπιστημίου της Οξφόρδης, κατάφερε το 1960 να αναπτύξει τα σημαντικά επιχειρήματά του εναντίον ολόκληρης της ιδέας της τεχνητής νοημοσύνης, της ικανότητας των υπολογιστών να σκεφτούν. Υπάρχουν αριθμητικές αλήθειες που, όπως είπε ο μαθηματικός John Casti του Ινστιτούτου της Σάντα Φε, «εμείς οι άνθρωποι μπορούμε να δούμε ότι ισχύουν, αλλά μια μηχανή δεν μπορεί να το αποδείξει».

Υπάρχει ένα παλιό σχολικό αστείο:

Δάσκαλος: Γιάννη, πόσο μας κάνει δύο και δύο;

Γιάννης: Τέσσερα.

Δάσκαλος: Πολύ καλά.

Γιάννης: Τι εννοείτε *πολύ καλά*; Αυτό είναι τέλειο!

Ο δάσκαλος μένει εμβρόντητος, αλλά είναι γεγονός ότι ο Γιάννης έχει δίκιο. Το αποτέλεσμα του λογικού χειρισμού ψηφιακών οντοτήτων δεν είναι καλό ή κακό, καλύτερο ή χειρότερο: είναι τέλειο ή λάθος. (Ο δάσκαλος θα πρέπει να προσέχει αυτόν το μαθητή του, γιατί έχει κατανοήσει σε πολύ νεαρή ηλικία τη διαφορά μεταξύ αναλογικής και ψηφιακής σκέψης!)

Οι ήχοι, αντιθέτως, είναι αναλογικοί και έτσι ο δάσκαλος θα ήταν σε πιο σταθερό έδαφος στην ακόλουθη περίπτωση. Σε ένα μάθημα Γαλλικών:

Δάσκαλος: Γιάννη, πρόφερε τη λέξη «monsieur».

Γιάννης: Μεσιέ.

Δάσκαλος: Πολύ καλά.

Σε αυτή την περίπτωση ο Γιάννης δεν μπορεί να υποστηρίξει ότι «Αυτό είναι τέλειο!», γιατί δεν υπάρχει τέλειος τρόπος για να προφέρουμε μια λέξη. Μπορούμε να χαρακτηρίσουμε τον τρόπο με τον οποίο προφέρεται μια λέξη καθαρό ή θολό, μελωδικό ή επίπεδο, έντονο ή χαλαρό, εμφατικό ή αβέβαιο, ακόμη και όμορφο ή άσχημο, αλλά όχι *τέλειο.* (Οι ψηφιακοί υπολογιστές αντιμετωπίζουν μέχρι σήμερα σημαντικές δυσκολίες στην αναγνώριση της προφορικής, αναλογικής, ομιλίας.)

Μερικά χρόνια αργότερα σε ένα μάθημα ιστορίας:

Δάσκαλος: Ποιες ήταν οι αιτίες που προκάλεσαν τον Πρώτο Παγκόσμιο Πόλεμο;

Γιάννης: Οι αυξανόμενες στρατιωτικές εντάσεις μεταξύ των μεγαλύτερων ευρωπαϊκών δυνάμεων από το 1910 περίπου, που πυροδοτήθηκαν από τη δολοφονία του Αρχιδούκα Φραντς Φέρντιναντ στο Σαράγεβο τον Ιούνιο του 1914.

Δάσκαλος: Αυτή είναι σίγουρα μία από τις αναγνωρισμένες ερμηνείες. Μήπως μπορείς να επεκταθείς λίγο σε ό,τι αφορά τις αυξανόμενες εντάσεις;

Καθώς ο Γιάννης μεγαλώνει και διδάσκεται ιστορία, λογοτεχνία, οικονομικά και διοίκηση επιχειρήσεων, θα περιπλανηθεί σε περιοχές που απομακρύνονται όλο και περισσότερο από τη λογική. Θα ανακαλύψει ότι υπάρχουν δύο είδη προβλημάτων: αυτά στα οποία οι απαντήσεις είναι ξεκάθαρα σωστές ή λανθασμένες και αυτά που επιτρέπουν πολλές ερμηνείες.

Οι λογικές πράξεις —μπορούμε να τις πούμε ψηφιακές— διατυπώνουν ξανά γεγονότα που υπάρχουν ήδη πριν ξεκινήσει η πράξη. Δύο συν δύο ισοδυναμεί με τέσσερα, αφού το τέσσερα είναι ένας άλλος τρόπος να πούμε το δύο συν

δύο. Ένα από τα προβλήματα που τέθηκαν, το 1970, σε μαθητές που είχαν κάνει αίτηση για να σπουδάσουν μαθηματικά στο Πανεπιστήμιο του Κέμπριτζ ήταν να *αποδειχτεί* ότι δύο συν δύο κάνει τέσσερα. Όπως φάνηκε, ήταν μια εκπληκτικά δύσκολη άσκηση για τους μαθητές. Το κλειδί στην απάντηση είναι ο ορισμός των ακέραιων αριθμών, που υπάρχει σε κάθε λεξικό: κάθε ακέραιος έχει αξία ενός περισσότερου από τον ακέραιο που προηγείται. Συνεπώς, «δύο» σημαίνει «ένα περισσότερο από το ένα», «τρία» σημαίνει «ένα περισσότερο από το δύο» και «τέσσερα» σημαίνει «ένα περισσότερο από το τρία». Όταν το καταλάβουμε αυτό, η απόδειξη είναι εύκολη. «Τέσσερα» σημαίνει «τρία συν ένα», που σημαίνει «δύο συν ένα συν ένα», δηλαδή «ένα συν ένα συν ένα συν ένα». Συνεπώς, το «τέσσερα» και το «δύο συν δύο» είναι και οι δύο τρόποι για να πούμε «τέσσερις φορές το ένα».

Αυτό που θα μπορούσε να ονομαστεί νόμος της διατήρησης των πληροφοριών ισχύει για όλες τις λογικές πράξεις: η πληροφορία ούτε δημιουργείται ούτε καταστρέφεται. Αυτή η μοναδική συνειδητοποίηση διαχωρίζει τις λειτουργίες στις οποίες απευθύνεται η λογική σκέψη από αυτές στις οποίες δεν απευθύνεται. Η λογική σκέψη απευθύνεται στα προβλήματα εκείνα που περιέχουν ήδη την απάντηση στον ορισμό των λέξεων που συνιστούν την ερώτηση.

Ένα μικρό μέρος της διεργασίας που ακολουθούν οι άνθρωποι είναι καθαρά λογικό-ψηφιακό: χρησιμοποιεί ασπρόμαυρους κανόνες όπως αυτούς της αριθμητικής ή της τυπικής λογικής. Σε αυτούς τους τομείς οι μηχανές υπολογισμών, που μετά από πενήντα χρόνια εξακολουθούν να είναι κυρίως ψηφιακές, είναι ασύγκριτα ανώτερες σε ταχύτητα υπολογισμών σε σχέση με οποιοδήποτε ανθρώπινο μυαλό. Οι ψηφιακοί υπολογισμοί αφορούν τη χρήση τεχνητών μηχανημάτων στις πλευρές εκείνες της

ανθρώπινης σκέψης όπου ο χειρισμός των πνευματικών μεταβλητών είναι λογικός, διακριτός και ακριβής. Αλλά το πλήρες φάσμα της ανθρώπινης επεξεργασίας πληροφοριών είναι απείρως περίπλοκο και πολύπλευρο και κατά το μεγαλύτερό του μέρος αναλογικό, όπως στους πολύ σημαντικούς τομείς της αναγνώρισης ήχου και εικόνας. Στο βιβλίο του *What Will Be*, του 1997, ο Μιχάλης Δερτούζος του MIT αναφέρεται αναλυτικά στο μέλλον των νέων τεχνολογιών. Είναι γενικά αισιόδοξος για την τεχνολογία, αλλά προειδοποιεί:

> Δείξτε μου έναν υπολογιστή που να δει τηλεόραση και να αναγνωρίσει ότι ένα μικρό κοριτσάκι που κρατάει έναν άσπρο σκύλο από το λουρί και τρέχει να διασχίσει ένα δρόμο πρόκειται να χτυπηθεί από μια ανοιχτή BMW και θα σας δείξω έναν ψεύτικο υπολογιστή!

Στο χειρισμό αριθμών και αλφαβητικού κειμένου οι υπολογιστές είχαν πολύ μεγάλη επιτυχία πολύ σύντομα. Η τεχνολογία μπορούσε να κάνει υπολογισμούς γρηγορότερα από τον ανθρώπινο εγκέφαλο, ακόμη και στις ημέρες των πρώτων ψηφιακών μηχανών, και τα προγράμματα επεξεργασίας κειμένου είχαν εξελιχθεί αρκετά μέχρι τη δεκαετία του 1970. Θα ήταν λογικό να περιμέναμε παρόμοια πρόοδο και στην ερμηνεία των σημάτων εικόνας, ήχου και άλλων αισθήσεων, μόλις ανακαλύπτονταν οι μέθοδοι για την ψηφιοποίησή τους.

Κάτι τέτοιο θα σήμαινε να ξεχάσουμε ότι η ικανότητα των μηχανών να διαχειρίζονται ποσότητες και λέξεις ψηφιακά δεν οφειλόταν σε πενήντα χρόνια εξέλιξης στην τεχνολογία των υπολογιστών αλλά στην ανάπτυξη, κατά τα προηγούμενα πέντε χιλιάδες χρόνια, των ψηφιακών μορφών γραφής (10 αριθμοί και 24 γράμματα του αλφαβή-

του) και των λογικών κανόνων για το χειρισμό τους (οι κανόνες των μαθηματικών και της λογικής). Στην περίπτωση των μαγνητοσκοπημένων εικόνων, η κοινότητα των υπολογιστών έπρεπε να ξεκινήσει από την αρχή, σαν να προσπαθούσε να δουλέψει με τους αριθμούς πριν ανακαλυφθούν οι κανόνες της αριθμητικής ή με τις λέξεις πριν την εποχή του αλφάβητου και της γραμματικής.

Σήμερα τα προγράμματα στα καλύτερα εργαστήρια λογισμικού, σε όλο τον κόσμο, δεν μπορούν να συναγάγουν από τη φωτογραφία ενός δωματίου μια απλή περιγραφή του περιεχομένου: τις καρέκλες, τα τραπέζια, τις κουρτίνες, τα κατοικίδια και τους ανθρώπους. Ακόμη και όταν τα προγράμματα κατορθώσουν να αποκτήσουν αυτή τη σχετικά απλή ικανότητα –σε λίγο καιρό οι επιστήμονες θα το κατορθώσουν– θα εξακολουθήσουν να μην έχουν ιδέα τι *σημασία* έχει η παρουσία των αντικειμένων. Τι μπορεί να μας λέει η τοποθέτηση των καθισμάτων, το σχέδιο του χαλιού και η ράτσα του σκύλου για τον ιδιοκτήτη; Η εικόνα δεν μπορεί να γίνει *κατανοητή* από τη μηχανή. Αυτό που μπορεί να κάνει με μεγάλη ακρίβεια η μετατροπή από αναλογική σε ψηφιακή μορφή είναι να βοηθήσει στην αναγνώριση μοτίβων φωτός και ήχου. Ο υπολογιστής μπορεί να σαρώσει τις εικόνες και τους ήχους, ακόμη και να τα αναγνωρίσει, αν του έχουν δώσει ξεκάθαρες οδηγίες για το τι ψάχνει, αλλά αυτό που δεν μπορεί να συλλάβει είναι τι *σημαίνει* το μοτίβο, ακόμη και σε επίπεδα που θα θεωρούσαμε πολύ υποτυπώδη για έναν άνθρωπο.

Με την έλευση της ψηφιακής τεχνολογίας, οι μηχανές πέρασαν το όριο μεταξύ των δύο πρώτων επιπέδων της ιεραρχίας δεδομένα-πληροφορία-γνώση. Εκτός από το να μεταδίδουν δεδομένα, μπορούν πλέον να αναγνωρίζουν και να επεξεργάζονται το πληροφοριακό τους περιεχόμενο. Αλλά η επεξεργασία είναι περιορισμένη: μπορούν

να αναγνωρίζουν την πληροφορία, δεν μπορούν όμως να την κατανοήσουν.

Η διαφορά μεταξύ του υπολογίσιμου και του μη υπολογίσιμου είναι η ίδια με εκείνη μεταξύ του χειρισμού πληροφοριών με χρήση αλγορίθμων (κανόνων, διαδικασιών, οδηγιών) και μη αλγοριθμικής επεξεργασίας. Οι φιλόσοφοι και οι μαθηματικοί που είναι μυημένοι στα κείμενα του Gödel μιλούν για το υπολογίσιμο ενάντια στο μη υπολογίσιμο. Ακόμη κι ένας ανίδεος μπορεί να μιλήσει σωστά για τη διαφορά μεταξύ της λογικής και της διαισθητικής σκέψης. Είναι ένα πολύ παλιό θέμα: η λογική εναντίον της δημιουργικότητας. Ο χειρισμός που είναι λογικός έχει το πλεονέκτημα της ακρίβειας και της διαύγειας, αλλά δεν μπορεί να έχει ούτε ίχνος πρωτοτυπίας, λόγω της ίδιας της φύσης της αφαιρετικής λογικής. Είτε το ονομάσουμε δομημένο εναντίον αδόμητου είτε υπολογιστής εναντίον ανθρώπου, είναι το ίδιο.

Τα επιχειρήματα και των δύο πλευρών αυτής της διαμάχης φτάνουν ως τα όρια –ίσως και να τα ξεπερνούν– αυτού που μπορούν να μας πουν επί του θέματος η επιστήμη, τα μαθηματικά ή η φιλοσοφία. Ο σύγχρονος τρόπος σκέψης, στην επιστήμη και τα μαθηματικά, δεν έχει διευθετήσει τη διαφωνία που υπάρχει για την τεχνητή νοημοσύνη. Ίσως η επιστήμη και η λογική να αναπτυχθούν μέχρι το σημείο όπου το Μεγάλο Ερώτημα –αν η ανθρώπινη σκέψη υπερβαίνει τη λογική των μηχανών– θα λάβει την απάντησή του, με τον έναν ή τον άλλο τρόπο. Ή ίσως το ερώτημα να είναι ανάλογο με μία από τις αριθμητικές προτάσεις του Gödel που δεν μπορούν να αποδειχτούν: η διαίσθηση μας λέει την απάντηση, όμως η απάντηση θα είναι πάντα αναπόδεικτη. Η διαίσθησή μας κλίνει σαφώς προς την ιδέα ότι η νόηση δεν είναι καθαρά μηχανική και ότι το μυαλό είναι κάτι που δεν μπορεί να γίνει πλήρως κατανοη-

τό, πόσο μάλλον να αντιγραφεί πιστά εφαρμόζοντας τις δικές του πνευματικές ικανότητες (αντηχεί ξανά το Θεώρημα του Gödel). Υπάρχει ένα παλιό γνωμικό: «Αν το μυαλό ήταν αρκετά απλό για να γίνει κατανοητό, θα ήμασταν πολύ ηλίθιοι για να το κατανοήσουμε». Εκατό δισεκατομμύρια ανθρώπινοι νευρώνες θα μπορούσαν ίσως να κατανοήσουν το μυαλό της κατσαρίδας με τους 100.000 νευρώνες, αλλά μια κατσαρίδα δεν θα μπορούσε ποτέ να κατανοήσει το δικό της μυαλό. Με βάση το ίδιο επιχείρημα, που αφορά τους περιορισμούς μιας οντότητας να καταλάβει τον εαυτό της, ούτε εμείς θα μπορούσαμε να κατανοήσουμε το δικό μας μυαλό. (Στην πραγματικότητα βέβαια, δεν έχουμε την παραμικρή ιδέα ακόμη και για το πώς δουλεύει το μυαλό μιας κατσαρίδας.)

Κατά μια πολύ ευρεία έννοια, η παράλληλη επεξεργασία και τα νευρωνικά δίκτυα θυμίζουν χωρίς αμφιβολία μερικές από τις ιδιότητες του δικού μας εγκεφάλου. Αλλά τα βήματα που έχουν γίνει μέχρι τώρα προς αυτή την κατανόηση αποτελούν μόλις κάποιες δοκιμές. Το χαρακτηριστικό της δομής του εγκεφάλου, που φαίνεται να προσομοιώνουν τα νευρωνικά συστήματα, είναι η αρχή ότι υπάρχουν επεξεργαστικές οντότητες που ονομάζονται νευρώνες, η καθεμία εκ των οποίων συνδέεται με άλλους νευρώνες που πυροδοτούν σήματα κατά μήκος των συνδετικών οδών (η πυροδότηση γίνεται με τη μορφή παλμών που παραπέμπουν στην ψηφιοποίηση) και μετά από λίγο κάνουν τις συνδέσεις να ποικίλλουν σε δύναμη. Αν ψάξουμε λίγο βαθύτερα, η αναλογία γίνεται λίγο ή πολύ κατανοητή.

Το πρώτο πρόβλημα είναι ο βαθμός της πολυπλοκότητας. Μερικές χειρουργικές επεμβάσεις στον εγκέφαλο πραγματοποιούνται σε ασθενείς που δεν έχουν ναρκωθεί. Μέσα από αυτές η ιατρική επιστήμη έχει βρει ότι υπερβολικά λεπτομερείς αναμνήσεις συγκεκριμένων γεγονότων

μπορούν να δημιουργηθούν αν ερεθίσουμε μια ομάδα νευρικών κυττάρων με ένα ηλεκτρικό σήμα. Ένα περιστατικό, όπως η επίσκεψη σε έναν κινηματογράφο πολλά χρόνια πριν, μπορεί να επανέλθει στη μνήμη με τόσες λεπτομέρειες που υπό κανονικές συνθήκες θα είχαν ξεχαστεί ακόμη και με το τέλος εκείνου του απογεύματος: ο αριθμός της θέσης πάνω στο εισιτήριο, τα ρούχα που φορούσε καθένας από τους γύρω θεατές. Όπου υπήρχαν καταγεγραμμένα στοιχεία, αυτές οι μνήμες επαληθεύτηκαν. Σχεδόν κάθε γεγονός που έχουμε ζήσει φαίνεται να αφήνει κάποιο είδος ίχνους στον εγκέφαλό μας· η γνώση είναι έτοιμη να χρησιμοποιηθεί σε κάποια μελλοντική κατάσταση, ακόμη και αν δεν είναι εφικτό να την ανακαλέσουμε με τη θέλησή μας. Με εκατομμύρια οπτικών σημάτων να έρχονται μέσω του οπτικού νεύρου κάθε δευτερόλεπτο που είμαστε ξύπνιοι, για να μην αναφέρουμε και τα χιλιάδες σήματα που εισέρχονται από τα αυτιά, η δυνατότητα του μυαλού να αποθηκεύει αυτή τη συσσωρευμένη πληροφορία είναι ουσιαστικά αμέτρητη. Σύμφωνα με στοιχεία που προέκυψαν από μικροσκοπική εξέταση της νευρικής δομής, η χωρητικότητα αποθήκευσης είναι πραγματικά συγκλονιστική.

Απ' όσο μπορούμε να καταλάβουμε, τα βασικά σημεία αποθήκευσης είναι οι συνάψεις, που εμφανίζονται σε κάθε σύνδεση μεταξύ δύο νευρώνων. Καθένας από τους είκοσι δισεκατομμύρια νευρώνες συνδέεται με τους άλλους με τη βοήθεια ενός δέντρου με διακλαδιζόμενες συνδέσεις, τους δενδρίτες. Σε ακραίες περιπτώσεις μερικοί νευρώνες συνδέονται, μέσω συναπτικών συνδέσεων, με μέχρι 80.000 άλλους. Συνολικά, ίσως να υπάρχουν 100 *τρισεκατομμύρια* συνάψεις στον εγκεφαλικό φλοιό. Ο πιο περίπλοκος μικροεπεξεργαστής, αντιθέτως, έχει περίπου 10 *εκατομμύρια* συνδέσεις.

Πιο σημαντικό ωστόσο από την πρόοδό μας –ή την

έλλειψη αυτής– στο να φτάσουμε τον αριθμό και την έκταση των διασυνδέσεων των νευρώνων είναι το γεγονός ότι δεν έχουμε τον παραμικρό λόγο για να πιστεύουμε ότι γνωρίζουμε τι συμβαίνει *μέσα σε κάθε νευρώνα.* Η υπόθεση ότι ένα εγκεφαλικό κύτταρο μπορεί να συγκριθεί ως προς την υπολογιστική ισχύ με μία μόλις λογική πύλη ενός υπολογιστή δεν αντέχει σε λεπτομερή εξέταση. Κάθε νευρώνας είναι εξαιρετικά πολύπλοκος: περιέχει περίπου 100.000 διαφορετικές πρωτεΐνες, καθεμία από τις οποίες έχει πολύ πολύπλοκη μοριακή δομή. Ο Stuart Hameroff, του Πανεπιστημίου της Αριζόνα, εξετάζει μικρές δομές που ονομάζονται μικροσωληνίσκοι και αποτελούνται από την πρωτεΐνη τουμπουλίνη. Κάθε μόριο τουμπουλίνης φαίνεται ότι μπορεί να αλλάξει μορφή και συνεπώς είναι ικανό να αναπαραστήσει διαφορετικές καταστάσεις (όπως οι διακόπτες σε έναν υπολογιστή). Η ιδέα ότι τέτοιες πρωτεΐνες μπορούν να είναι εξαιρετικά μικροσκοπικές συσκευές επεξεργασίας πληροφοριών, με μόλις μερικά άτομα μήκος, είχε ήδη προταθεί με τη μορφή θεωρητικής εικασίας από τον Ρότζερ Πένροουζ. Ο Πένροουζ έδειξε ότι, σε αυτή την κλίμακα, τα φαινόμενα θα λειτουργούσαν σε κβαντικό επίπεδο, εννοώντας ότι θα λειτουργούσαν όχι με έναν αυστηρά αλγοριθμικό τρόπο μιας λογικής μηχανής, αλλά με τον πιο ασαφή τρόπο των πιθανοτήτων, που είναι τόσο γνωστός στη σωματιδιακή φυσική του εικοστού αιώνα. Οι ιδέες του έχουν υποστηριχθεί περισσότερο από τον Brian Josephson, του Πανεπιστημίου του Κέμπριτζ, που κέρδισε το βραβείο Νόμπελ στη Φυσική για την έρευνα σε ένα κβαντικό φαινόμενο σε ημιαγωγούς, που ονομάζεται Φαινόμενο Josephson[1]. Ο Josephson είχε από τότε στρέψει την

1. Ονομάζεται επίσης Μακρές Υπεραγώγιμες Επαφές Josephson. (Σ.τ.Ε.)

προσοχή του στα ψυχικά φαινόμενα, τα οποία θεωρεί ότι βασίζονται σε κβαντικά φαινόμενα παρόμοια με αυτά του εγκεφάλου.

Τα κύτταρα είναι θαύματα εξελικτικής προόδου. Στα πρώτα τρία από τα τέσσερα δισεκατομμύρια χρόνια που υπάρχει ζωή στη γη, οι μονοκύτταροι οργανισμοί ήταν τα μόνα ζωντανά πλάσματα. Πολλοί μονοκύτταροι οργανισμοί ήταν (και είναι) ικανοί να αισθανθούν το περιβάλλον, να πλησιάσουν πιο κοντά σε τροφή, να αποφύγουν πιθανούς κινδύνους και φυσικά να αναπαραχθούν. Με αυτά τα δεδομένα, μπορούμε μόνο εικασίες να κάνουμε σχετικά με την ικανότητα επεξεργασίας πληροφοριών που βρίσκεται στο εξαιρετικά πολύπλοκο εσωτερικό ενός νευρώνα. Φανταστείτε μια μηχανή που ενώνει είκοσι δισεκατομμύρια προηγμένους επεξεργαστές πληροφοριών χρησιμοποιώντας 100 τρισεκατομμύρια καλώδια και θα έχετε κατά προσέγγιση την πολυπλοκότητα στην οποία αναφερόμαστε. Κάτι τέτοιο μπορεί να σημαίνει ότι ένα ανθρώπινο ον –εσείς για παράδειγμα– βρίσκεται στην ίδια κατηγορία με τη συνολική υπολογιστική δύναμη όλου του κόσμου.

Είναι συνηθισμένο να διαβάζουμε σε δημοσιογραφικές αναφορές του ψηφιακού φαινομένου προτάσεις όπως: «-Μέχρι το έτος 2000 και κάτι οι υπολογιστές θα έχουν την ίδια χωρητικότητα μνήμης με τον ανθρώπινο εγκέφαλο». Προβλέψεις αυτού του είδους υποθέτουν ότι γνωρίζουμε πλήρως πού και πώς το μυαλό αποθηκεύει και επεξεργάζεται τις πληροφορίες.

Κάτι τέτοιο όμως είναι εντελώς αναληθές. Στην πραγματικότητα δεν έχουμε ιδέα. Ο Gerald Edelman γράφει στο *Brilliant Air, Brilliant Fire: On the Matter of the Mind* (1992): «Δεν έχουν βρεθεί πειστικά στοιχεία στο ανθρώπινο νευρικό σύστημα για το είδος των κωδικών που οι άνθρωποι χρη-

σιμοποιούν στην τηλεγραφία, στους υπολογιστές ή σε άλλες μορφές επικοινωνίας. Ο κόσμος είναι ένα κομμάτι ταινίας και το μυαλό δεν είναι ένας υπολογιστής». Ενώ έχουν γίνει τεράστια άλματα στη νευρολογική επιστήμη τα τελευταία χρόνια, κάθε ανακάλυψη φέρνει στην επιφάνεια καινούργια άγνωστα στοιχεία. Καθώς εστιάζουμε όλο και ισχυρότερα μικροσκόπια σε αυτά τα στοιχεία του εγκεφάλου, αντί να συγκλίνουμε σε μια απάντηση για τα μυστήρια της συμπεριφοράς των ανθρώπινων νευρώνων, αποκαλύπτουμε επιπλέον επίπεδα τεράστιας πολυπλοκότητας.

Ενώ οι φιλόσοφοι διαφωνούν για το Μεγάλο Ερώτημα –αν οι υπολογιστές θα μπορέσουν ποτέ να σκεφτούν όπως οι άνθρωποι– αυτό που θέλουν οι περισσότεροι από εμάς δεν είναι μια θεωρητική απάντηση σε ένα φιλοσοφικό ερώτημα αλλά μια ρεαλιστική απάντηση σε μια πρακτική ερώτηση: πού βρίσκεται στην πραγματικότητα το όριο μεταξύ, από τη μία πλευρά, εκείνων των ανθρώπινων πνευματικών δραστηριοτήτων τις οποίες οι υπολογιστές μπορούν να αντιγράψουν ή να ξεπεράσουν και, από την άλλη, αυτών που δεν μπορούν. Θα μπορούσε ένα πρόγραμμα στον υπολογιστή να αντικαταστήσει το δικηγόρο μου, το γιατρό μου, το αφεντικό μου, τον υφιστάμενό μου ή κάποιον άλλο από τους ανθρώπους με τους οποίους επικοινωνώ; Ή, εφόσον *είμαι* δικηγόρος, γιατρός, αφεντικό ή υφιστάμενος... θα μπορούσε ένα πρόγραμμα να αντικαταστήσει *εμένα*;

Η απάντηση στο Μεγάλο Ερώτημα δεν μας βοηθάει, όπως και να ’χει. Αν η απάντηση είναι ναι –θα μπορούσε καταρχήν να υπάρξει κάποιος συνδυασμός υλικού και λογισμικού που θα ήταν δυνατό να αντιγράψει την ανθρώπινη σκέψη– αυτός ο συνδυασμός ωστόσο θα είναι τόσο τρομακτικά μεγάλος και πολύπλοκος, ώστε να μη σχεδια-