

# 1

## ΕΙΣΑΓΩΓΗ

*Όπου θα προσπαθήσουμε να εξηγήσουμε γιατί θεωρούμε την τεχνητή νοημοσύνη θέμα άξιο σοβαρής μελέτης, και θα επιχειρήσουμε να προσδιορίσουμε τι ακριβώς είναι, το οποίο βέβαια είναι καλό να γίνει πριν καταπιαστούμε με το θέμα.*

Ονομάζουμε το είδος μας *homo sapiens* — άνθρωπος ο σοφός — επειδή οι νοητικές μας ικανότητες είναι πολύ σημαντικές για μας. Για χιλιάδες χρόνια προσπαθούμε να κατανοήσουμε το *πώς σκεπτόμαστε*: δηλαδή, πώς μια χούφτα ύλης μπορεί να αντιλαμβάνεται, να κατανοεί, να προβλέπει και να χειρίζεται έναν κόσμο πολύ μεγαλύτερο και πολύ πιο πολύπλοκο από τον εαυτό της. Το πεδίο της **τεχνητής νοημοσύνης** (artificial intelligence), ή για συντομία TN, πηγαίνει ακόμα πιο πέρα: Επιχειρεί όχι μόνο να κατανοήσει αλλά και να *κατασκευάσει* νοήμονες οντότητες.

Η TN είναι μία από τις νεότερες επιστήμες. Η σοβαρή δουλειά άρχισε να γίνεται λίγο μετά το Δεύτερο Παγκόσμιο Πόλεμο, και ο ίδιος ο όρος εμφανίστηκε το 1956. Μαζί με τη μοριακή βιολογία, η TN αναφέρεται συχνά από επιστήμονες άλλων κλάδων ως “το πεδίο όπου θα ήθελα περισσότερο να δουλέψω”. Ένας φοιτητής της Φυσικής θα ήταν κατανοητό να έχει την αίσθηση ότι όλες οι καλές ιδέες έχουν ήδη καλυφθεί από τους Γαλιλέο, Νεύτωνα, Αϊνστάιν, και άλλους. Η TN, από την άλλη, έχει ακόμα χώρο για πολλούς Αϊνστάιν πλήρους απασχόλησης.

Σήμερα, η TN συνδυάζει μια τεράστια ποικιλία επιμέρους πεδίων, τα οποία καλύπτουν ένα φάσμα που ξεκινά από γενικούς τομείς, όπως η μάθηση και η αντίληψη, και φτάνει σε συγκεκριμένες εργασίες όπως το σκάκι, η απόδειξη μαθηματικών θεωρημάτων, η συγγραφή ποίησης, και η διάγνωση ασθενειών. Η TN συστηματοποιεί και αυτοματοποιεί τις διανοητικές εργασίες, γι’ αυτό και μπορεί να έχει εφαρμογή σε οποιαδήποτε σφαίρα της ανθρώπινης διανοητικής δραστηριότητας. Με αυτή την έννοια, είναι πραγματικά ένα οικουμενικό πεδίο.

ΤΕΧΝΗΤΗ  
ΝΟΗΜΟΣΥΝΗ

### 1.1 ΤΙ ΕΙΝΑΙ Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

Υποστηρίξαμε ότι η TN είναι συναρπαστική, αλλά δεν είπαμε τι *είναι*. Στην Εικόνα 1.1, παρουσιάζονται ορισμοί της TN από οκτώ εγχειρίδια. Οι ορισμοί αυτοί έχουν διαφορές σε δύο κύριες διαστάσεις. Χονδρικά, οι επάνω ορισμοί ενδιαφέρονται περισσότερο για τις *διαδικασίες σκέψης* και τη *συλλογιστική*, ενώ οι κάτω ορισμοί ασχολούνται με τη *συμπεριφορά*. Οι ορισμοί στα αριστερά μετρούν την επιτυχία με βάση την εγγύτητα προς τις *ανθρώπινες επιδόσεις*, ενώ οι ορισμοί στα δεξιά τη μετρούν σε σχέση με μια *ιδανική* έννοια νοημοσύνης, την οποία θα ονομάσουμε **ορθολογικότητα** (rationality). Ένα σύστημα είναι ορθολογικό αν κάνει “το σωστό”, με δεδομένα όσα γνωρίζει.

ΟΡΘΟΛΟΓΙΚΟΤΗΤΑ

Συστήματα που σκέπτονται σαν τον άνθρωπο	Συστήματα που σκέπτονται ορθολογικά
<p>“Η συναρπαστική νέα προσπάθεια για να κάνουμε τους υπολογιστές να σκέπτονται ... <i>μηχανές με νόηση</i>, με την πλήρη και κυριολεκτική έννοια.” (Haugeland, 1985)</p> <p>“[Η αυτοματοποίηση των] δραστηριοτήτων που συσχετίζουμε με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση...” (Bellman, 1978)</p>	<p>“Η μελέτη των νοητικών ικανοτήτων με τη χρήση υπολογιστικών μοντέλων.” (Charniak και McDermott, 1985)</p> <p>“Η μελέτη των υπολογιστικών εργασιών που μας δίνουν τη δυνατότητα να αντιλαμβανόμαστε, να συλλογίζομαστε, και να ενεργούμε.” (Winston, 1992)</p>
Συστήματα που ενεργούν σαν τον άνθρωπο	Συστήματα που ενεργούν ορθολογικά
<p>“Η τέχνη της δημιουργίας μηχανών που πραγματοποιούν λειτουργίες οι οποίες απαιτούν νοημοσύνη όταν πραγματοποιούνται από ανθρώπους.” (Kurzweil, 1990)</p> <p>“Η μελέτη του πώς μπορούμε να κάνουμε τους υπολογιστές να κάνουν πράγματα στα οποία, προς το παρόν, οι άνθρωποι είναι καλύτεροι.” (Rich και Knight, 1991)</p>	<p>“Υπολογιστική Νοημοσύνη είναι η μελέτη της σχεδίασης ευφών πρακτόρων.” (Poole κ.α., 1998)</p> <p>“Η τεχνητή νοημοσύνη ασχολείται με τη ευφυή συμπεριφορά των τεχνουργημάτων.” (Nilsson, 1998)</p>
<p><b>Εικόνα 1.1</b> Μερικοί ορισμοί της τεχνητής νοημοσύνης, οργανωμένοι σε τέσσερις κατηγορίες.</p>	

Ιστορικά, έχουν ακολουθηθεί και οι τέσσερις προσεγγίσεις στην ΤΝ. Όπως θα περίμενε κανείς, υπάρχει κάποια διένεξη ανάμεσα στις προσεγγίσεις που εστιάζονται στον άνθρωπο και τις προσεγγίσεις που εστιάζονται στην ορθολογικότητα.<sup>1</sup> Μια ανθρωποκεντρική προσέγγιση θα πρέπει να είναι εμπειρική επιστήμη, με υποθέσεις και με πειραματική επιβεβαίωση. Μια ορθολογιστική (rationalist) προσέγγιση περιλαμβάνει ένα συνδυασμό μαθηματικών και τεχνολογίας. Η κάθε ομάδα έχει αμφισβητήσει αλλά και έχει βοηθήσει την άλλη. Ας εξετάσουμε πιο αναλυτικά τις τέσσερις προσεγγίσεις.

### Ανθρώπινη δράση: Η προσέγγιση με τη δοκιμασία Turing

#### ΔΟΚΙΜΑΣΙΑ TURING

Η δοκιμασία Turing (Turing Test), η οποία προτάθηκε από τον Alan Turing (1950), σχεδιάστηκε για να παρέχει έναν ικανοποιητικό λειτουργικό ορισμό της νοημοσύνης. Αντί να προτείνει μια εκτεταμένη και ενδεχομένως αντιφατική λίστα γνωρισμάτων που απαιτούνται για τη νοημοσύνη, ο Turing πρότεινε μια δοκιμασία που βασιζόταν στην αδυναμία να γίνει διάκριση από τις αναμφίβολα νοήμονες οντότητες — τους ανθρώπους. Ο υπολογιστής περνά τη δοκιμασία αν ένας άνθρωπος εξεταστής, αφού θέσει μερικές γραπτές ερωτήσεις, δεν μπορεί να συμπεράνει αν οι γραπτές απαντήσεις προέρχονται από άνθρωπο ή όχι. Στο Κεφάλαιο 26, εξετάζονται οι λεπτομέρειες της δοκιμασίας και το αν ένας υπολογιστής είναι πραγματικά νοήμων όταν μπορεί να την περάσει. Για την ώρα, θα σημειώσουμε ότι για να προγραμματιστεί ένας υπολογιστής για να πε-

<sup>1</sup> Θα πρέπει να επισημάνουμε ότι, όταν κάνουμε διάκριση μεταξύ *ανθρώπινης* και *ορθολογικής* συμπεριφοράς, δεν υπονοούμε αναγκαστικά ότι οι άνθρωποι είναι “παράλογοι”, με την έννοια της “συναισθηματικής αστάθειας” ή της “παραφροσύνης”. Αρκεί να παρατηρήσει κανείς ότι δεν είμαστε τέλειοι: Δεν είμαστε όλοι μεγάλοι μετρ στο σκάκι, ακόμα και αν γνωρίζουμε όλους τους κανόνες του παιχνιδιού· και δυστυχώς, δεν παίρνουν όλοι άριστα στις εξετάσεις. Μερικά συστηματικά σφάλματα της ανθρώπινης συλλογιστικής έχουν καταγραφεί από τους Kahneman κ.α. (1982).

ράσει τη δοκιμασία χρειάζεται να γίνουν πολλά. Ο υπολογιστής θα πρέπει να έχει τις εξής ικανότητες:

ΕΠΕΞΕΡΓΑΣΙΑ  
ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ  
ΑΝΑΠΑΡΑΣΤΑΣΗ  
ΓΝΩΣΗΣ  
ΑΥΤΟΜΑΤΟ-  
ΠΟΙΗΜΕΝΗ  
ΣΥΛΛΟΓΙΣΤΙΚΗ  
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

- ◇ **επεξεργασία φυσικής γλώσσας**, ώστε να μπορεί να επικοινωνεί ικανοποιητικά σε μια γλώσσα όπως η Αγγλική·
- ◇ **αναπαράσταση γνώσης**, ώστε να αποθηκεύει αυτά που γνωρίζει ή ακούει·
- ◇ **αυτοματοποιημένη συλλογιστική**, ώστε να χρησιμοποιεί τις αποθηκευμένες πληροφορίες για να απαντά ερωτήσεις και να παράγει νέα συμπεράσματα·
- ◇ **μηχανική μάθηση**, ώστε να προσαρμόζεται σε νέες περιστάσεις και να εντοπίζει ή να συμπεραίνει πρότυπα.

Η δοκιμασία Turing απέφυγε εσκεμμένα την άμεση φυσική αλληλεπίδραση μεταξύ του εξεταστή και του υπολογιστή, επειδή η *φυσική* προσομοίωση ενός ανθρώπου δεν είναι απαραίτητη για τη νοημοσύνη. Όμως, η λεγόμενη **πλήρης δοκιμασία Turing** (total Turing Test) περιλαμβάνει οπτικό σήμα, ώστε να μπορεί ο εξεταστής να εξετάζει τις αντιληπτικές ικανότητες του υποκειμένου και να έχει τη δυνατότητα να του δίνει φυσικά αντικείμενα “από το παραθυράκι”. Για να περάσει την πλήρη δοκιμασία Turing, ο υπολογιστής θα χρειαστεί να έχει:

ΠΛΗΡΗΣ ΔΟΚΙΜΑΣΙΑ  
TURING

- ◇ **μηχανική όραση**, ώστε να αντιλαμβάνεται αντικείμενα, και
- ◇ **ρομποτική**, ώστε να χειρίζεται αντικείμενα και να μετακινείται.

ΜΗΧΑΝΙΚΗ  
ΟΡΑΣΗ  
ΡΟΜΠΟΤΙΚΗ

Αυτά τα έξι πεδία αποτελούν το μεγαλύτερο μέρος της TN, και ο Turing αξίζει την αναγνώριση επειδή σχεδίασε μια δοκιμασία που διατηρεί την ισχύ της μετά από 50 χρόνια. Πάντως, οι ερευνητές της TN έχουν αφιερώσει ελάχιστη προσπάθεια στη δημιουργία μηχανών που μπορούν να περάσουν τη δοκιμασία Turing, καθώς πίστευαν ότι είναι πιο σημαντικό να μελετήσουν της αρχές στις οποίες βασίζεται η νοημοσύνη παρά να αντιγράψουν ένα πρωτότυπο. Η έρευνα για την “τεχνητή πτήση” έφτασε στην επιτυχία όταν οι αδελφοί Wright και άλλοι σταμάτησαν να προσπαθούν να μιμηθούν τα πουλιά και έμαθαν την αεροδυναμική. Τα χειρίδια αεροναυπηγικής δεν ορίζουν ως σκοπό του πεδίου τους την κατασκευή “μηχανών που πετούν τόσο όμοια με τα περιστέρια ώστε να μπορούν να ξεγελάσουν ακόμα και τα ίδια τα περιστέρια”.

### Ανθρώπινη σκέψη: Η προσέγγιση με γνωστικά μοντέλα

Για να πούμε ότι ένα δεδομένο πρόγραμμα σκέπτεται σαν άνθρωπος, θα πρέπει να έχουμε κάποιον τρόπο να προσδιορίσουμε το πώς σκέπτονται οι άνθρωποι. Θα χρειαστεί να μπούμε στο *εσωτερικό* της ίδιας της λειτουργίας του ανθρώπινου νου. Υπάρχουν δύο τρόποι να γίνει αυτό: με την ενδοσκόπηση — προσπαθώντας να συλλάβουμε τις ίδιες μας τις σκέψεις καθώς πραγματοποιούνται — και με ψυχολογικά πειράματα. Από τη στιγμή που θα έχουμε μια αρκετά ακριβή θεωρία της νόησης, θα είναι δυνατό να εκφράσουμε αυτή τη θεωρία ως πρόγραμμα υπολογιστή. Αν οι συμπεριφορές εισόδου-εξόδου και χρονισμού του προγράμματος συμφωνούν με τις αντίστοιχες ανθρώπινες συμπεριφορές, αυτό είναι τεκμήριο ότι μερικοί από τους μηχανισμούς του προγράμματος μπορεί να λειτουργούν όμοια και στον άνθρωπο. Για παράδειγμα, οι Allen Newell και Herbert Simon, οι οποίοι ανέπτυξαν το GPS (“General Problem Solver” — γενικός λύτης προβλημάτων — Newell και Simon, 1961) δεν απέστηκαν στο να λύνει το πρόγραμμά τους σωστά τα προβλήματα. Τους ενδιέφερε περισσότερο η σύγκριση της ακολουθίας συλλογιστικών βημάτων του με την αντίστοιχη ακολουθία συλλογιστικών βημάτων των ανθρώπων που έλυναν το ίδιο πρόβλημα. Το διεπιστημονικό πεδίο της **γνωστικής επιστήμης** (cognitive science) φέρνει στον ίδιο χώρο υπολογιστικά μοντέλα της TN και πειραματικές τεχνικές της ψυχολογίας, με σκοπό τη δημιουργία ακριβών και πειραματικά επαληθεύσιμων θεωριών για τον τρόπο λειτουργίας του ανθρώπινου νου.

ΓΝΩΣΤΙΚΗ ΕΠΙΣΤΗΜΗ

Η γνωστική επιστήμη είναι ένα συναρπαστικό πεδίο που αξίζει να έχει μια δική του εγκυκλοπαίδεια (Wilson και Keil, 1999). Σε αυτό το βιβλίο, δεν θα επιχειρήσουμε να περιγράψουμε όσα είναι γνωστά για την ανθρώπινη γνωστική λειτουργία. Θα σχολιάζουμε ευκαιριακά τις ομοιότητες ή τις διαφορές ανάμεσα στις τεχνικές της TN και την ανθρώπινη γνωστική λειτουργία. Η γνήσια γνωστική επιστήμη αναγκαστικά βασίζεται σε πειραματική έρευνα σε πραγματικούς ανθρώπους και ζώα, ενώ εμείς θα θεωρήσουμε ότι ο αναγνώστης έχει στη διάθεσή του μόνο έναν υπολογιστή για πειραματισμό.

Την πρόμη εποχή της TN, υπήρχε συχνά σύγχυση μεταξύ των δύο προσεγγίσεων· ένας συγγραφέας συχνά υποστήριζε ότι ένας αλγόριθμος αποδίδει καλά σε μια εργασία και *γι' αυτό το λόγο* είναι ένα καλό μοντέλο της ανθρώπινης λειτουργίας, ή το αντίστροφο. Οι σημερινοί συγγραφείς διαχωρίζουν τα δύο είδη ισχυρισμών, και η διάκριση αυτή έχει επιτρέψει τόσο στην TN όσο και στη γνωστική επιστήμη να αναπτύσσονται γρηγορότερα. Τα δύο πεδία συνεχίζουν να γονιμοποιούνται αμοιβαία, ιδίως στους τομείς της όρασης και της φυσικής γλώσσας. Ειδικότερα, ο τομέας της όρασης έχει να επιδείξει πρόσφατες προόδους, με μια ολοκληρωμένη προσέγγιση που παίρνει υπόψη ευρήματα της νευροφυσιολογίας και υπολογιστικά μοντέλα.

### Ορθολογική σκέψη: Η προσέγγιση με τους “νόμους της σκέψης”

Ο Έλληνας φιλόσοφος Αριστοτέλης ήταν από τους πρώτους που επιχειρήσαν να κωδικοποιήσουν τη “σωστή σκέψη”, δηλαδή, αδιάψευστες διαδικασίες συλλογιστικής. Οι αριστοτέλειοι **συλλογισμοί** (syllogisms) παρείχαν πρότυπα δομής επιχειρημάτων που έδιναν πάντα σωστά συμπεράσματα όταν ξεκινούσαν από σωστές υποθέσεις — για παράδειγμα, “Ο Σωκράτης είναι άνθρωπος· όλοι οι άνθρωποι είναι θνητοί· επομένως, ο Σωκράτης είναι θνητός”. Θεωρήθηκε ότι αυτοί οι νόμοι της σκέψης κυβερνούν τη λειτουργία της νόησης· από τη μελέτη τους ξεκίνησε το πεδίο που ονομάζεται **λογική** (logic).

Οι ερευνητές της λογικής του 19ου αιώνα ανέπτυξαν μια ακριβή σημειογραφία για τις προτάσεις που αναφέρονται σε όλα τα είδη πραγμάτων που υπάρχουν στον κόσμο και για τις σχέσεις μεταξύ τους. (Αντιπαραβάλετε με το συνηθισμένο συμβολισμό της αριθμητικής, ο οποίος παρέχει κυρίως σύμβολα για προτάσεις ισότητας και ανισότητας μεταξύ αριθμών.) Το 1965, υπήρχαν ήδη προγράμματα που μπορούσαν θεωρητικά να λύνουν *οποιοδήποτε* πρόβλημα ήταν δυνατό να λυθεί, αρκεί να ήταν διατυπωμένο με τη λογική σημειογραφία.<sup>2</sup> Η λεγόμενη **λογικιστική** (logicist) παράδοση στο χώρο της TN ελπίζει να οικοδομήσει με βάση τέτοια προγράμματα προκειμένου να δημιουργήσει ευφυή συστήματα.

Δύο είναι τα κύρια εμπόδια στην προσέγγιση αυτή. Πρώτον, δεν είναι εύκολο να παίρνει κανείς άτυπη γνώση και να τη διατυπώνει με τους τυπικούς όρους που απαιτούνται από τη σημειογραφία της λογικής, ιδιαίτερα όταν η γνώση είναι λιγότερο από 100% βέβαιη. Δεύτερον, υπάρχει μεγάλη διαφορά μεταξύ τού να μπορεί κανείς “θεωρητικά” να λύσει ένα πρόβλημα και του να μπορεί να το λύσει στην πράξη. Ακόμα και προβλήματα που περιλαμβάνουν μόλις μερικές δεκάδες γεγονότα μπορούν να εξαντλήσουν τους υπολογιστικούς πόρους οποιουδήποτε υπολογιστή, εκτός αν του δοθεί κάποια καθοδήγηση για το ποια συλλογιστικά βήματα θα δοκιμάσει πρώτα. Αν και τα δύο αυτά εμπόδια είναι παρόντα σε *οποιαδήποτε* προσπάθεια κατασκευής υπολογιστικών συστημάτων συλλογιστικής, εμφανίστηκαν πρώτα στη λογικιστική παράδοση.

### Ορθολογική δράση: Η προσέγγιση με ορθολογικούς πράκτορες

**Πράκτορας** (agent) είναι απλώς κάτι που “πράττει” (ο αγγλικός όρος agent επίσης προέρχεται από το λατινικό *agere*, που σημαίνει “πράττω”). Όμως, οι πράκτορες των υπολογιστών θα πρέπει να έχουν και κάποια άλλα γνωρίσματα που τους κάνουν να ξεχωρίζουν από τα απλά “προγράμ-

<sup>2</sup> Αν δεν υπάρχει λύση, το πρόγραμμα μπορεί να μη σταματήσει ποτέ να ψάχνει για λύση.

ΣΥΛΛΟΓΙΣΜΟΙ

ΛΟΓΙΚΗ

ΛΟΓΙΚΙΣΜΟΣ

ΠΡΑΚΤΟΡΑΣ

ματα”, όπως το να λειτουργούν κάτω από αυτόνομο έλεγχο, να αντιλαμβάνονται το περιβάλλον τους, να διατηρούνται για ένα παρατεταμένο χρονικό διάστημα, να προσαρμόζονται στις αλλαγές, και να έχουν την ικανότητα να αναλαμβάνουν τους στόχους κάποιου άλλου. **Ορθολογικός πράκτορας** (rational agent) είναι ένας πράκτορας που ενεργεί έτσι ώστε να επιτυγχάνει το καλύτερο αποτέλεσμα ή, όταν υπάρχει αβεβαιότητα, το καλύτερο αναμενόμενο αποτέλεσμα.

Στην προσέγγιση της τεχνητής νοημοσύνης με τους “νόμους της σκέψης”, η έμφαση δινόταν στη σωστή εξαγωγή συμπερασμάτων (συμπερασμό). Ο σωστός συμπερασμός μερικές φορές είναι μέρος τού να είναι ένας πράκτορας ορθολογικός, επειδή ένας τρόπος να ενεργεί κανείς ορθολογικά είναι να συλλογίζεται πρώτα λογικά για να καταλήξει στο συμπέρασμα ότι μια δεδομένη ενέργεια θα επιτύχει τους σκοπούς του, και έπειτα να ενεργεί σύμφωνα με αυτό το συμπέρασμα. Από την άλλη, ο σωστός συμπερασμός δεν είναι *το παν* στην ορθολογικότητα, επειδή υπάρχουν συχνά καταστάσεις όπου καμία ενέργεια δεν μπορεί να αποδειχτεί σωστή, και παρ’ όλα αυτά κάτι πρέπει να γίνει. Υπάρχουν επίσης τρόποι να ενεργεί κανείς ορθολογικά οι οποίοι δεν μπορούμε να πούμε ότι περιλαμβάνουν συμπερασμό. Για παράδειγμα, η αναπήδηση μακριά από μια καυτή θερμάστρα είναι μια αντανάκλαστική ενέργεια η οποία είναι συνήθως πιο επιτυχημένη από μια πιο αργή ενέργεια που γίνεται μετά από προσεκτική σκέψη.

Όλες οι δεξιότητες που χρειάζονται για τη δοκιμασία Turing υπάρχουν για να επιτρέπουν ορθολογικές ενέργειες. Χρειαζόμαστε λοιπόν την ικανότητα να αναπαριστούμε τη γνώση και να συλλογίζομαστε με αυτή επειδή αυτό μας επιτρέπει να καταλήγουμε σε καλές αποφάσεις σε ένα ευρύ φάσμα καταστάσεων. Χρειάζεται να μπορούμε να παράγουμε κατανοητές προτάσεις σε φυσική γλώσσα, επειδή η εκφώνηση αυτών των προτάσεων μας βοηθά να τα βγάζουμε πέρα σε μια πολύπλοκη κοινωνία. Χρειαζόμαστε τη μάθηση όχι μόνο για χάρη της πολυμάθειας, αλλά επειδή το να έχουμε μια καλύτερη ιδέα για το πώς λειτουργεί ο κόσμος μάς επιτρέπει να παράγουμε πιο αποτελεσματικές στρατηγικές για να τον αντιμετωπίσουμε. Χρειαζόμαστε οπτική αντίληψη όχι μόνο επειδή το να βλέπουμε είναι διασκεδαστικό, αλλά για να παίρνουμε μια καλύτερη ιδέα για το τι αποτέλεσμα μπορεί να έχει μια ενέργεια — για παράδειγμα, το να μπορούμε να δούμε μια νόστιμη μπουκιά μας βοηθά να κινηθούμε προς το μέρος της.

Γι’ αυτούς τους λόγους, η μελέτη της ΤΝ ως σχεδίασης ορθολογικών πρακτόρων έχει τουλάχιστον δύο πλεονεκτήματα. Πρώτον, είναι πιο γενική από την προσέγγιση με τους “νόμους της σκέψης”, επειδή ο σωστός συμπερασμός είναι μόνο ένας από τους πολλούς δυνατούς μηχανισμούς για την επίτευξη της ορθολογικότητας. Δεύτερον, προσφέρεται καλύτερα για επιστημονική ανάπτυξη από ό,τι οι προσεγγίσεις που βασίζονται στην ανθρώπινη συμπεριφορά ή την ανθρώπινη σκέψη, επειδή το πρότυπο της ορθολογικότητας είναι σαφώς ορισμένο και εντελώς γενικό. Η ανθρώπινη συμπεριφορά, από την άλλη, είναι καλά προσαρμοσμένη για ένα συγκεκριμένο περιβάλλον, και είναι κατά ένα μέρος προϊόν μιας περίπλοκης και σε μεγάλο βαθμό άγνωστης διαδικασίας εξέλιξης που απέχει πολύ ακόμα από την τελειότητα. Γι’ αυτό, το βιβλίο εστιάζεται στις γενικές αρχές των ορθολογικών πρακτόρων και στα στοιχεία με τα οποία κατασκευάζονται. Θα δούμε ότι, παρά τη φαινομενική απλότητα με την οποία μπορεί να διατυπωθεί το πρόβλημα, ανακύπτει μια τεράστια ποικιλία ζητημάτων όταν επιχειρήσουμε να το λύσουμε. Το Κεφάλαιο 2 σκιαγραφεί μερικά από αυτά τα ζητήματα πιο αναλυτικά.

Κάτι σημαντικό που πρέπει να έχουμε υπόψη μας: Δεν θα αργήσουμε να διαπιστώσουμε ότι η επίτευξη της τέλει ορθολογικότητας — το να κάνει κανείς πάντα το σωστό — δεν είναι εφικτή σε περίπλοκα περιβάλλοντα. Οι υπολογιστικές απαιτήσεις είναι υπερβολικά υψηλές. Για το μεγαλύτερο μέρος του βιβλίου, όμως, θα υιοθετήσουμε την υπόθεση εργασίας ότι η τέλεια ορθολογικότητα είναι ένα καλό σημείο εκκίνησης για την ανάλυσή μας. Απλουστεύει το πρόβλημα και μας παρέχει το κατάλληλο σκηνικό για την περισσότερη βασική ύλη του πεδίου. Τα Κεφάλαια 6 και 17 ασχολούνται ειδικά με το ζήτημα της **περιορισμένης ορθολογικότητας** (limited rationality) — την πραγματοποίηση της κατάλληλης ενέργειας όταν δεν υπάρχει αρκετός χρόνος για να γίνουν όλοι οι υπολογισμοί που θα θέλαμε.



## 1.2 ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Σε αυτή την ενότητα παρουσιάζουμε μια σύντομη ιστορία των κλάδων που συνεισέφεραν ιδέες, απόψεις, και τεχνικές στην ΤΝ. Όπως συμβαίνει πάντα με την ιστορία, θα αναγκαστούμε να εστιάσουμε την προσοχή μας σε ένα μικρό αριθμό ατόμων, περιστατικών και ιδεών, και να αγνοήσουμε άλλα που ήταν επίσης σημαντικά. Θα οργανώσουμε την ιστορία μας γύρω από μια σειρά ερωτήσεων. Δεν θα θέλαμε βέβαια να δημιουργηθεί η εντύπωση ότι αυτές οι ερωτήσεις είναι οι μόνες που απασχόλησαν τους σχετικούς κλάδους, ούτε ότι οι κλάδοι αυτοί δούλευαν όλοι έχοντας την ΤΝ ως τελικό στόχο τους.

### Φιλοσοφία (428 π.Χ. – σήμερα)

- Μπορούν να χρησιμοποιούνται τυπικοί κανόνες για την εξαγωγή έγκυρων συμπερασμάτων;
- Πώς προκύπτει η πνευματική νόηση από το φυσικό εγκέφαλο;
- Από πού προέρχεται η γνώση;
- Πώς η γνώση οδηγεί σε δράση;

Ο Αριστοτέλης (384 – 322 π.Χ.) ήταν ο πρώτος που διατύπωσε ένα ακριβές σύνολο νόμων που διέπουν το ορθολογικό μέρος της νόησης. Ανέπτυξε ένα άτυπο σύστημα συλλογισμών για τη σωστή συλλογιστική οι οποίοι θεωρητικά επέτρεπαν να παράγει κανείς συμπεράσματα μηχανικά με δεδομένες κάποιες αρχικές υποθέσεις. Πολύ αργότερα, ο Ramon Lull (; – 1315) είχε την ιδέα ότι μπορεί να πραγματοποιείται χρήσιμη συλλογιστική με ένα μηχανικό τεχνούργημα. Οι “τροχοί εννοιών” του (concept wheels) απεικονίζονται στο εξώφυλλο του βιβλίου. Ο Thomas Hobbes (1588 – 1679) πρότεινε ότι η συλλογιστική είναι σαν τους αριθμητικούς υπολογισμούς, και ότι “προσθέτουμε και αφαιρούμε στις σιωπηλές μας σκέψεις”. Η αυτοματοποίηση των υπολογισμών είχε ήδη διανύσει αρκετό δρόμο· γύρω στο 1500, ο Leonardo da Vinci (1452 – 1519) σχεδίασε μια μηχανική αριθμομηχανή, χωρίς όμως να την κατασκευάσει· πρόσφατες ανακατασκευές έδειξαν ότι το σχέδιο μπορούσε να λειτουργήσει. Η πρώτη γνωστή αριθμομηχανή κατασκευάστηκε γύρω στο 1623 από το Γερμανό επιστήμονα Wilhelm Schickard (1592 – 1635), αν και η αριθμομηχανή Pascaline, που κατασκευάστηκε το 1642 από τον Blaise Pascal (1623 – 1662), είναι πιο κοντινά στη σκέψη από ό,τι όλες οι πράξεις των ζώων”. Ο Gottfried Wilhelm Leibniz (1646 – 1716) κατασκεύασε μια μηχανική συσκευή που είχε σκοπό να πραγματοποιεί πράξεις σε έννοιες και όχι σε αριθμούς, αλλά το πεδίο της ήταν μάλλον περιορισμένο.

Έχοντας τώρα στη διάθεσή μας την ιδέα ενός συνόλου κανόνων που μπορούν να περιγράψουν το τυπικό, ορθολογικό μέρος της νόησης, το επόμενο βήμα ήταν να εξεταστεί η νόηση ως φυσικό σύστημα. Ο René Descartes (1596 – 1650) παρουσίασε την πρώτη σοβαρή μελέτη της διάκρισης μεταξύ νόησης και ύλης και των προβλημάτων που προκύπτουν. Ένα πρόβλημα που παρουσιάζει η καθαρά φυσική αντίληψη της νόησης είναι ότι φαίνεται να μην αφήνει πολύ χώρο για την ελεύθερη βούληση· αν η νόηση διέπεται αποκλειστικά από τους φυσικούς νόμους, τότε δεν έχει περισσότερη ελεύθερη βούληση από μια πέτρα που “αποφασίζει” να πέσει προς το κέντρο τη Γης. Αν και ένθερμος υποστηρικτής της δύναμης της συλλογιστικής, ο Descartes ήταν επίσης υποστηρικτής του **δυϊσμού** (dualism). Πίστευε ότι υπάρχει ένα μέρος της ανθρώπινης νόησης (ή της ψυχής, ή του πνεύματος) που βρίσκεται έξω από τη φύση, και δεν υπόκειται στους φυσικούς νόμους. Τα ζώα, από την άλλη, δεν είχαν αυτή τη διπλή ιδιότητα, και μπορούσαν να αντιμετωπίζονται σαν μηχανές. Μια εναλλακτική θεωρία, αντί του δυϊσμού, είναι ο **υλισμός** (materialism), ο οποίος υποστηρίζει ότι η λειτουργία του εγκεφάλου σύμφωνα με τους νόμους της Φυσικής *αποτελεί* τη νόηση. Η ελεύθερη βούληση είναι απλώς ο τρόπος με τον οποίο εμφανίζεται η αντίληψη των διαθέσιμων επιλογών σε μια διαδικασία επιλογής.

ΔΥΪΣΜΟΣ

ΥΛΙΣΜΟΣ

Με δεδομένη μια φυσική νόηση που χειρίζεται τη γνώση, το επόμενο πρόβλημα είναι να προσδιορίσουμε την πηγή της γνώσης. Το κίνημα του **εμπειρισμού** (empiricism), το οποίο ξεκίνησε με το *Novum Organum (Νέο Όργανο)*<sup>3</sup> του Francis Bacon (1561 – 1626), μπορεί να σκιαγραφηθεί με ένα απόφθεγμα του John Locke (1632 – 1704): “Δεν υπάρχει στην κατανόηση τίποτα που δε βρισκόταν πρώτα στις αισθήσεις”. Ο David Hume (1711 – 1776), στο σύγγραμμά του *A Treatise of Human Nature (Πραγματεία της ανθρώπινης φύσης — Hume, 1739)*, πρότεινε εκείνο που είναι σήμερα γνωστό ως **επαγωγή** (induction): Οι γενικοί κανόνες προκύπτουν με έκθεση σε επανειλημμένες συσχετίσεις μεταξύ των στοιχείων τους. Οικοδομώντας πάνω στη δουλειά του Ludwig Wittgenstein (1889 – 1951) και του Bertrand Russell (1872 – 1970), ο περίφημος Κύκλος της Βιέννης (Vienna Circle), με επικεφαλής τον Rudolf Carnap (1091 – 1970), ανέπτυξε το δόγμα του **λογικού θετικισμού** (logical positivism). Το δόγμα αυτό υποστηρίζει ότι όλη η γνώση μπορεί να χαρακτηριστεί ως λογικές θεωρίες οι οποίες σε τελική ανάλυση συνδέονται με **προτάσεις παρατήρησης** (observation sentences) που αντιστοιχούν σε αισθητηριακές εισόδους.<sup>4</sup> Η **θεωρία της επιβεβαίωσης** (confirmation theory), του Carnap και του Carl Hempel (1905 – 1997) επιχειρήσε να κάνει κατανοητό το πώς μπορεί να αποκτάται γνώση από την εμπειρία. Το βιβλίο του Carnap *The Logical Structure of the World (Η λογική δομή του κόσμου — 1928)* όριζε μια ρητή υπολογιστική διαδικασία για την εξαγωγή γνώσης από στοιχειώδεις εμπειρίες. Ήταν ίσως η πρώτη θεωρία που αντιμετώπιζε τη νόηση ως υπολογιστική διαδικασία.

ΕΜΠΕΙΡΙΚΙΣΜΟΣ  
ΕΠΑΓΩΓΗ  
ΛΟΓΙΚΟΣ ΘΕΤΙΚΙΣΜΟΣ  
ΠΡΟΤΑΣΕΙΣ  
ΠΑΡΑΤΗΡΗΣΗΣ  
ΘΕΩΡΙΑ  
ΕΠΙΒΕΒΑΙΩΣΗΣ

Το τελευταίο στοιχείο της φιλοσοφικής εικόνας της νόησης είναι η σχέση μεταξύ γνώσης και δράσης. Το ερώτημα αυτό είναι ζωτικό για την ΤΝ, επειδή η νοημοσύνη προϋποθέτει όχι μόνο συλλογιστική αλλά και δράση. Επίσης, μόνο κατανοώντας πώς αιτιολογούνται οι ενέργειες μπορούμε να κατανοήσουμε πώς μπορούμε να κατασκευάσουμε έναν πράκτορα που οι ενέργειές του να είναι αιτιολογημένες (ή ορθολογικές). Ο Αριστοτέλης υποστήριξε ότι οι ενέργειες αιτιολογούνται με λογική σύνδεση μεταξύ των στόχων και της γνώσης του αποτελέσματος της ενέργειας. (Το τελευταίο τμήμα του παρακάτω αποσπάσματος εμφανίζεται και στο εξώφυλλο του βιβλίου στα Αρχαία Ελληνικά.)

Πώς όμως συμβαίνει η σκέψη να συνοδεύεται μερικές φορές από δράση και άλλες φορές όχι, μερικές φορές από κίνηση και άλλες φορές όχι; Μοιάζει σαν να είναι σχεδόν το ίδιο όπως στην περίπτωση της συλλογιστικής και της εξαγωγής συμπερασμάτων για αναλλοίωτα αντικείμενα. Σε εκείνη την περίπτωση, όμως, ο σκοπός είναι μια θεωρητική πρόταση ... ενώ εδώ το συμπέρασμα που προκύπτει από τις δύο υποθέσεις είναι μια ενέργεια. ... Χρειάζομαι ένδυση· ένας χιτώνας είναι ένδυση. Χρειάζομαι ένα χιτώνα. Ό,τι χρειάζομαι, πρέπει να το φτιάξω· χρειάζομαι ένα χιτώνα. Πρέπει να φτιάξω ένα χιτώνα. Και το συμπέρασμα, “Πρέπει να φτιάξω ένα χιτώνα”, είναι μια ενέργεια. (Nussbaum, 1978, σελίδα 40.)

Στα *Ηθικά Νικομάχεια* (Βιβλίο III, 3, 1112β), ο Αριστοτέλης επεξηγεί περισσότερο αυτό το θέμα, προτείνοντας έναν αλγόριθμο:

Δεν διαλογιζόμαστε για τους σκοπούς αλλά για τα μέσα. Ένας γιατρός δεν διαλογίζεται για το αν θα θεραπεύσει, ούτε ένας ρήτορας για το αν θα πείσει, ... Δέχονται το σκοπό και εξετάζουν το πώς και με ποια μέσα μπορεί να επιτευχθεί και το αν φαίνεται να μπορεί να επιτευχθεί ευκολότερα και καλύτερα με αυτά τα μέσα· ενώ, αν μπορεί να επιτευχθεί μόνο με ένα μέσο, εξετάζουν το πώς θα επιτευχθεί με αυτό το μέσο, και με ποια μέσα θα επιτευχθεί αυτό το πώς, μέχρι να φτάσουν στον πρωταρχικό παράγοντα, ... και εκείνο που είναι τελευταίο στη σειρά της ανάλυσης εμφανίζεται ως πρώτο στη σειρά της πραγματοποίησης. Και αν φτάσουμε σε κάτι αδύνατο, εγκαταλείπουμε την αναζήτηση, για παράδειγμα, αν χρειαζόμα-

<sup>3</sup> Μια αναθεώρηση πάνω στο “Όργανον” του Αριστοτέλη.

<sup>4</sup> Στην εικόνα αυτή, όλες οι προτάσεις που έχουν νόημα μπορούν να επαληθεύονται ή να διαψεύδονται είτε με ανάλυση του νοήματος των λέξεων είτε με την πραγματοποίηση πειραμάτων. Επειδή αυτή η προσέγγιση αποκλείει το μεγαλύτερο μέρος της μεταφυσικής, το οποίο έγινε σκόπιμα, ο λογικός θετικισμός δεν ήταν δημοφιλής σε μερικούς κύκλους.

στε χρήματα και δεν μπορούμε να τα αποκτήσουμε· αν όμως κάτι φαίνεται δυνατό, επιχειρούμε να το κάνουμε.

Ο αλγόριθμος του Αριστοτέλη υλοποιήθηκε μετά από 2300 χρόνια από τους Newell και Simon στο πρόγραμμα GPS. Σήμερα, θα χαρακτηρίζαμε αυτό το πρόγραμμα ως σύστημα σχεδιασμού με οπισθοχώρηση (regression planning system — δείτε στο Κεφάλαιο 11).

Η ανάλυση με βάση το στόχο είναι χρήσιμη, αλλά δε μας λέει τι να κάνουμε όταν πολλές ενέργειες μπορούν να επιτύχουν το στόχο ή όταν καμία ενέργεια δεν μπορεί να τον επιτύχει πλήρως. Ο Antoine Arnauld (1612 – 1694) περιέγραψε σωστά έναν ποσοτικό μαθηματικό τύπο που προσδιορίζει ποια ενέργεια θα γίνεται σε τέτοιες περιπτώσεις (δείτε στο Κεφάλαιο 16). Ο John Stuart Mill (1806 – 1873), στο βιβλίο του *Utilitarianism* (*Χρησιμοθηρία* — Mill, 1863), υποστήριξε την ιδέα των ορθολογικών κριτηρίων απόφασης για όλες τις σφαίρες της ανθρώπινης δραστηριότητας. Η τυπική θεωρία αποφάσεων εξετάζεται στην επόμενη ενότητα.

### Μαθηματικά (περ. 800 – σήμερα)

- Ποιοι είναι οι τυπικοί κανόνες για την εξαγωγή ενός έγκυρου συμπεράσματος;
- Τι μπορεί να υπολογιστεί και τι όχι;
- Πώς συλλογίζομαστε με ασαφείς πληροφορίες;

Οι φιλόσοφοι επισήμαναν τις περισσότερες από τις σημαντικές ιδέες της ΤΝ, αλλά το άλμα για να γίνει μια τυπική επιστήμη απαιτούσε κάποιο επίπεδο μαθηματικής τυποποίησης σε τρεις θεμελιώδεις τομείς: στη λογική, στη θεωρία υπολογισμού και στις πιθανότητες.

Η ιδέα της τυπικής λογικής συναντάται από τον καιρό των αρχαίων Ελλήνων φιλοσόφων (δείτε στο Κεφάλαιο 7), αλλά η μαθηματική της ανάπτυξη ξεκίνησε πραγματικά με την εργασία του George Boole (1815 – 1864), ο οποίος επεξεργάστηκε τις λεπτομέρειες της προτασιακής λογικής ή λογικής Boole (Boole, 1847). Το 1879, ο Gottlob Frege (1848 – 1925) επέκτεινε τη λογική του Boole ώστε να συμπεριλάβει αντικείμενα και σχέσεις, δημιουργώντας τη λογική πρώτης τάξης (first-order logic) η οποία χρησιμοποιείται σήμερα ως το πιο βασικό σύστημα αναπαράστασης γνώσης.<sup>5</sup> Ο Alfred Tarski (1902 – 1983) παρουσίασε μια θεωρία της αναφοράς (theory of reference) η οποία δείχνει πώς μπορούν να συσχετίζονται τα αντικείμενα της λογικής με αντικείμενα του πραγματικού κόσμου. Το επόμενο βήμα ήταν να προσδιοριστούν τα όρια του τι μπορούσε να γίνει με τη λογική και τον υπολογισμό.

Ο πρώτος μη τετριμμένος **αλγόριθμος** θεωρείται ότι ήταν ο αλγόριθμος του Ευκλείδη για τον υπολογισμό του μέγιστου κοινού παρονομαστή. Η μελέτη των αλγορίθμων ως αντικειμένων αφ' εαυτών φτάνει μέχρι την εποχή του al-Khowarazmi, ενός Πέρση μαθηματικού του 9ου αιώνα, που τα συγγράμματά του εισήγαγαν επίσης τους αραβικούς αριθμούς και την άλγεβρα στην Ευρώπη. Ο Boole και άλλοι μελέτησαν αλγόριθμους λογικής παραγωγής συμπεράσματος (logical deduction) και, προς το τέλος του 19ου αιώνα, γίνονταν ήδη προσπάθειες να τυποποιηθεί η γενική μαθηματική συλλογιστική ως λογική παραγωγή συμπερασμάτων. Το 1900, ο David Hilbert (1862 – 1943) παρουσίασε έναν κατάλογο 23 προβλημάτων τα οποία, όπως σωστά πρόβλεψε, θα απασχολούσαν τους μαθηματικούς για το μεγαλύτερο μέρος του αιώνα. Το τελευταίο πρόβλημα ρωτούσε αν υπάρχει αλγόριθμος που μπορεί να αποφασίσει την αλήθεια οποιασδήποτε λογικής πρότασης που αφορούσε τους φυσικούς αριθμούς — το περίφημο *Entscheidungsproblem*, ή πρόβλημα της απόφασης. Ουσιαστικά, το ερώτημα που έθεσε ο Hilbert ήταν αν υπάρχουν θεμελιώδη όρια στην ισχύ των αποτελεσματικών διαδικασιών απόδειξης. Το 1930, ο Kurt Gödel (1906 – 1978) έδειξε ότι υπάρχει μια αποτελεσματική διαδικασία για την απόδειξη οποιασδήποτε αληθούς πρότασης της λογικής πρώτης τάξης του Frege και του Russell,

<sup>5</sup> Η σημειογραφία που πρότεινε ο Frege για τη λογική πρώτης τάξης δεν έγινε ποτέ δημοφιλής, για λόγους που γίνονται αμέσως φανεροί από το παράδειγμα στο εξώφυλλο του βιβλίου.



ΘΕΩΡΗΜΑ ΤΗΣ  
ΜΗ ΠΛΗΡΟΤΗΤΑΣ

όμως η λογική πρώτης τάξης δεν μπορούσε να συλλάβει την αρχή της μαθηματικής επαγωγής που χρειαζόταν για τον προσδιορισμό των φυσικών αριθμών. Το 1931, έδειξε ότι πραγματικά υπάρχουν όρια. Το **θεώρημα της μη πληρότητας** (incompleteness theorem) του Gödel έδειξε ότι, σε οποιαδήποτε γλώσσα που έχει αρκετή εκφραστική ισχύ για να περιγράψει τις ιδιότητες φυσικών αριθμών, υπάρχουν αληθείς προτάσεις οι οποίες είναι μη αποφασίσιμες (undecidable), με την έννοια ότι η αλήθεια τους δεν μπορεί να βεβαιωθεί με κανέναν αλγόριθμο.

Αυτό το θεμελιώδες συμπέρασμα μπορεί επίσης να ερμηνευτεί ότι δείχνει ότι υπάρχουν μερικές συναρτήσεις στους ακεραίους οι οποίες δεν μπορούν να αναπαρασταθούν από έναν αλγόριθμο — δηλαδή, δεν μπορούν να υπολογιστούν. Αυτό ώθησε τον Alan Turing (1912 – 1954) να επιχειρήσει να προσδιορίσει ακριβώς ποιες συναρτήσεις είναι δυνατό να υπολογιστούν. Η ιδέα αυτή είναι πραγματικά κάπως προβληματική, επειδή δεν μπορεί να δοθεί τυπικός ορισμός στην έννοια του υπολογισμού ή της αποτελεσματικής διαδικασίας. Όμως, η θέση των Church-Turing, η οποία υποστηρίζει ότι μια μηχανή Turing (Turing, 1936) είναι ικανή να υπολογίσει οποιαδήποτε υπολογίσιμη συνάρτηση, είναι γενικά παραδεκτό ότι παρέχει έναν επαρκή ορισμό. Ο Turing έδειξε επίσης ότι υπήρχαν μερικές συναρτήσεις τις οποίες καμία μηχανή Turing δεν μπορεί να υπολογίσει. Για παράδειγμα, καμία μηχανή δεν μπορεί γενικά να αποφανθεί αν ένα δεδομένο πρόγραμμα θα επιστρέψει απάντηση για μια δεδομένη είσοδο, ή αν θα εκτελείται για πάντα.

ΔΥΣΕΠΙΛΥΣΙΜΟΤΗΤΑ

Η μη αποφασισιμότητα (undecidability) και η μη υπολογισιμότητα (noncomputability) είναι βέβαια σημαντικές έννοιες για την κατανόηση του υπολογισμού, όμως η έννοια της **δυσεπιλυσιμότητας** (intractability) είχε πολύ μεγαλύτερες επιπτώσεις. Σε γενικές γραμμές, ένα πρόβλημα λέγεται δυσεπίλυτο αν ο χρόνος που χρειάζεται για την επίλυση στιγμιοτύπων του προβλήματος αυξάνεται εκθετικά με το μέγεθος των στιγμιοτύπων. Η διάκριση μεταξύ πολυωνυμικής και εκθετικής αύξησης της πολυπλοκότητας επισημάνθηκε για πρώτη φορά στα μέσα της δεκαετίας του 1960 (Cobham, 1964· Edmonds, 1965). Η διάκριση είναι σημαντική, επειδή η εκθετική αύξηση σημαίνει ότι ακόμα και όχι ιδιαίτερα μεγάλα στιγμιότυπα προβλημάτων δεν μπορούν να επιλυθούν μέσα σε κάποιο λογικό χρόνο. Γι' αυτό, θα πρέπει κανείς να επιδιώκει να διαιρεί το συνολικό πρόβλημα της παραγωγής ευφυούς συμπεριφοράς σε βατά (tractable) και όχι σε δυσεπίλυτα υποπροβλήματα.

NP-ΠΛΗΡΟΤΗΤΑ

Πώς μπορεί κανείς να αναγνωρίσει ένα δυσεπίλυτο πρόβλημα; Η θεωρία της **NP-πληρότητας** (NP-completeness), την οποία θεμελίωσαν οι Steven Cook (1971) και Richard Karp (1972), παρέχει μια μέθοδο. Ο Cook και ο Karp έδειξαν ότι υπάρχουν μεγάλες κλάσεις προβλημάτων κανονικής συνδυαστικής αναζήτησης και συλλογιστικής τα οποία είναι NP-πλήρη. Τα προβλήματα οποιαδήποτε κλάσης στην οποία μπορεί να αναχθεί η κλάση των NP-πλήρων προβλημάτων μάλλον είναι δυσεπίλυτα. (Αν και δεν έχει αποδειχθεί ότι τα NP-πλήρη προβλήματα είναι οπωσδήποτε δυσεπίλυτα, οι περισσότεροι θεωρητικοί πιστεύουν ότι είναι.) Τα συμπεράσματα αυτά έρχονται σε αντίθεση με την αισιοδοξία με την οποία ο λαϊκός τύπος χαιρέτισε τους πρώτους υπολογιστές — τους “ηλεκτρονικούς υπερεγκεφάλους” που ήταν “γρηγορότεροι από τον Αϊνστάιν!”. Παρά τη συνεχώς αυξανόμενη ταχύτητα των υπολογιστών, τα ευφυή συστήματα θα χαρακτηρίζονται από προσεκτική χρήση των πόρων. Για να το θέσουμε ωμά, ο κόσμος είναι ένα εξαιρετικά μεγάλο στιγμιότυπο προβλήματος! Τα τελευταία χρόνια, η TN έχει βοηθήσει να εξηγηθεί το γιατί μερικά στιγμιότυπα NP-πλήρων προβλημάτων είναι δύσκολα ενώ άλλα είναι εύκολα (Cheeseman κ.α., 1991).

ΠΙΘΑΝΟΤΗΤΑ

Εκτός από τη λογική και τον υπολογισμό, η τρίτη συνεισφορά των μαθηματικών στην TN είναι η θεωρία των **πιθανοτήτων**. Ο Ιταλός Gerolamo Cardano (1501 – 1576) διατύπωσε για πρώτη φορά την ιδέα της πιθανότητας, περιγράφοντάς τη με βάση τα δυνατά αποτελέσματα χαρτοπαικτικών συμβάντων. Η πιθανότητα έγινε σύντομα πολύτιμο μέρος όλων των ποσοτικών επισημών, βοηθώντας στην αντιμετώπιση αβέβαιων μετρήσεων και ατελών θεωριών. Ο Pierre Fermat (1601 – 1665), ο Blaise Pascal (1623 – 1662), ο James Bernoulli (1654 – 1705), ο Pierre

Laplace (1749 – 1827) και άλλοι προχώρησαν τη θεωρία και εισήγαγαν νέες στατιστικές μεθόδους. Ο Thomas Bayes (1702 – 1761) πρότεινε έναν κανόνα για την ενημέρωση των πιθανοτήτων υπό το φως νέων πειστηρίων. Ο κανόνας του Bayes και το πεδίο που προέκυψε, το οποίο ονομάστηκε ανάλυση Bayes, αποτελούν τη βάση των περισσότερων σημερινών προσεγγίσεων στην αβέβαιη συλλογιστική, στα συστήματα ΤΝ.

### Οικονομικά (1776 – σήμερα)

- Πώς πρέπει να παίρνουμε αποφάσεις ώστε να μεγιστοποιούμε την απολαβή;
- Πώς πρέπει να το κάνουμε όταν οι άλλοι ίσως να μη φέρονται ευνοϊκά;
- Πώς πρέπει να το κάνουμε όταν η απολαβή ίσως είναι στο απώτερο μέλλον;

Η οικονομική επιστήμη ξεκίνησε το 1776, όταν ο Σκωτσέζος φιλόσοφος Adam Smith (1723 – 1790) δημοσίευσε το σύγγραμμα *An Inquiry into the Nature and Causes of the Wealth of Nations* (*Έρευνα της φύσης και των αιτιών του πλούτου των εθνών*). Ενώ οι αρχαίοι Έλληνες και άλλοι είχαν επίσης κάνει συνεισφορές στην οικονομική σκέψη, ο Smith ήταν ο πρώτος που την αντιμετώπισε ως επιστήμη, βασιζόμενος στην ιδέα ότι οι οικονομίες μπορεί να θεωρηθεί ότι αποτελούνται από μεμονωμένους πράκτορες που μεγιστοποιούν τη δική τους οικονομική ευημερία. Οι περισσότεροι νομίζουν ότι η οικονομική επιστήμη ασχολείται με τα χρήματα, αλλά οι οικονομολόγοι θα απαντούσαν ότι στην πραγματικότητα μελετούν το πώς οι άνθρωποι κάνουν επιλογές που οδηγούν στα προτιμότερα αποτελέσματα. Η μαθηματική αντιμετώπιση των “προτιμότερων αποτελεσμάτων”, ή της **χρησιμότητας** (utility), διατυπώθηκε τυπικά για πρώτη φορά από τον Léon Walras (1834 – 1910), και βελτιώθηκε από τον Frank Ramsey (1931) και αργότερα από τον John von Neumann και τον Oskar Morgenstern (1944) στο βιβλίο τους *The Theory of Games and Economic Behavior* (*Θεωρία των παιγνίων και της οικονομικής συμπεριφοράς*).

ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

Η **θεωρία των αποφάσεων** (decision theory), η οποία συνδυάζει τη θεωρία των πιθανοτήτων με τη θεωρία της χρησιμότητας, παρέχει ένα τυπικό και πλήρες πλαίσιο για τη λήψη αποφάσεων (οικονομικών ή άλλων) που παίρνονται σε συνθήκες αβεβαιότητας — δηλαδή, σε περιπτώσεις όπου οι πιθανοτικές περιγραφές συλλαμβάνουν ικανοποιητικά το περιβάλλον εκείνου που παίρνει την απόφαση. Το πλαίσιο αυτό είναι κατάλληλο για “μεγάλες” οικονομίες, όπου ο κάθε πράκτορας δε χρειάζεται να δίνει σημασία στις ατομικές ενέργειες των άλλων πρακτόρων. Για “μικρές” οικονομίες, η κατάσταση μοιάζει πολύ περισσότερο με **παιχνίδι**: οι ενέργειες ενός παίκτη μπορούν να επηρεάσουν σημαντικά τη χρησιμότητα των ενεργειών ενός άλλου (είτε θετικά είτε αρνητικά). Η **θεωρία των παιγνίων** (game theory) του von Neumann και του Morgenstern (δείτε επίσης Luce και Raiffa, 1957) περιλάμβανε το απροσδόκητο συμπέρασμα ότι, σε μερικά παιχνίδια, ένας ορθολογικός πράκτορας θα πρέπει να ενεργεί με τυχαίο τρόπο, ή τουλάχιστον με έναν τρόπο που να φαίνεται τυχαίος στους αντιπάλους.

ΘΕΩΡΙΑ ΠΑΙΓΝΙΩΝ

Κατά το μεγαλύτερο μέρος, οι οικονομολόγοι δεν ασχολήθηκαν με την τρίτη ερώτηση, δηλαδή, το πώς μπορούμε να παίρνουμε ορθολογικές αποφάσεις όταν οι απολαβές των ενεργειών δεν είναι άμεσες αλλά προκύπτουν από μια *ακολουθία* πολλών ενεργειών. Το θέμα αυτό εξετάστηκε από το πεδίο της **επιχειρησιακής έρευνας** (operations research), το οποίο αναδύθηκε κατά το Δεύτερο Παγκόσμιο Πόλεμο από τις Βρετανικές προσπάθειες για βελτιστοποίηση των εγκαταστάσεων ραντάρ, και αργότερα βρήκε πολιτικές εφαρμογές στις πολύπλοκες διαχειριστικές αποφάσεις. Η δουλειά του Richard Bellman (1957) τυποποίησε μια κλάση προβλημάτων ακολουθιακών αποφάσεων που ονομάστηκαν **διαδικασίες αποφάσεων Markov**, τις οποίες θα μελετήσουμε στα Κεφάλαια 17 και 21.

ΕΠΙΧΕΙΡΗΣΙΑΚΗ  
ΕΡΕΥΝΑ

Η δουλειά που έγινε στα οικονομικά και στην επιχειρησιακή έρευνα συνέβαλε πολύ στην αντίληψή μας για τους ορθολογικούς πράκτορες, όμως για πολλά χρόνια η έρευνα στην ΤΝ ακολούθησε εντελώς διαφορετικά μονοπάτια. Ένας λόγος ήταν η προφανής **πολυπλοκότητα** της λήψης ορθολογικών αποφάσεων. Ο Herbert Simon (1916 – 2001), πρωτοποριακός ερευνητής της

ΙΚΑΝΟΠΟΙΗΤΙΚΟΤΗΤΑ

ΤΝ, κέρδισε το βραβείο Νόμπελ των οικονομικών το 1978 για την πρόμη εργασία του που έδειξε ότι τα μοντέλα που βασίζονται στην **ικανοποιητικότητα** (satisficing) — στη λήψη αποφάσεων που είναι “αρκετά καλές”, και όχι στον επίπονο υπολογισμό μιας βέλτιστης απόφασης — έδιναν μια καλύτερη περιγραφή της πραγματικής ανθρώπινης συμπεριφοράς (Simon, 1947). Στη δεκαετία του 1990, υπήρξε μια αναζωογόνηση του ενδιαφέροντος στις τεχνικές της θεωρίας αποφάσεων για τα συστήματα πρακτόρων (Wellman, 1995).

### Νευροεπιστήμες (1861 – σήμερα)

- Πώς επεξεργάζεται ο εγκέφαλος τις πληροφορίες;

ΝΕΥΡΟΕΠΙΣΤΗΜΕΣ

Οι **νευροεπιστήμες** (neuroscience) ασχολούνται με τη μελέτη του νευρικού συστήματος και ιδιαίτερα του εγκεφάλου. Ο ακριβής τρόπος με τον οποίο το μυαλό κάνει δυνατή τη σκέψη είναι από τα μεγαλύτερα μυστήρια της επιστήμης. Είχε εκτιμηθεί εδώ και χιλιάδες χρόνια ότι ο εγκέφαλος έχει κάποιου είδους ανάμειξη με τη σκέψη, επειδή υπήρχαν πειστήρια ότι δυνατά πλήγματα στο κεφάλι μπορούν να οδηγήσουν σε διανοητική αναπηρία. Ήταν επίσης γνωστό από παλιά ότι ο ανθρώπινος εγκέφαλος είναι κάπως διαφορετικός· γύρω στο 335 π.Χ., ο Αριστοτέλης έγραψε: “Από όλα τα ζώα, ο άνθρωπος έχει το μεγαλύτερο εγκέφαλο σε αναλογία με το μέγεθός του.”<sup>6</sup> Ωστόσο, μόλις στα μέσα του 18ου αιώνα αναγνωρίστηκε γενικά ότι ο εγκέφαλος είναι η έδρα της συναίσθησης. Προηγούμενες, υποψήφιες θέσεις ήταν η καρδιά, η σπλήνα και η επίφυση του εγκεφάλου.

ΝΕΥΡΩΝΕΣ

Η μελέτη του Paul Broca (1824 – 1889) για την αφασία (ανεπάρκεια της ομιλίας) σε ασθενείς με βλάβη του εγκεφάλου το 1861 αναζωογόνησε αυτό το πεδίο και έπεισε το ιατρικό κατεστημένο για την ύπαρξη εντοπισμένων περιοχών του εγκεφάλου που είναι υπεύθυνες για συγκεκριμένες γνωστικές λειτουργίες. Ειδικότερα, ο Broca έδειξε ότι η παραγωγή της ομιλίας είναι εντοπισμένη σε ένα τμήμα του αριστερού ημισφαιρίου που σήμερα ονομάζεται περιοχή Broca.<sup>7</sup> Την εποχή εκείνη, ήταν γνωστό ότι ο εγκέφαλος αποτελούνταν από νευρικά κύτταρα, ή **νευρώνες** (neurons), αλλά μόλις το 1873 ο Camillo Golgi (1843 – 1926) ανέπτυξε μια τεχνική χρωματισμού που έκανε δυνατή την παρατήρηση μεμονωμένων νευρώνων στον εγκέφαλο (δείτε στην Εικόνα 1.2). Η τεχνική αυτή χρησιμοποιήθηκε από τον Santiago Ramon y Cajal (1852 – 1934) στις πρωτοποριακές μελέτες του για τις νευρωνικές δομές του εγκεφάλου.<sup>8</sup>

Σήμερα, έχουμε κάποια δεδομένα για την αντιστοιχία μεταξύ περιοχών του εγκεφάλου και των μερών του σώματος τα οποία ελέγχουν ή από τα οποία δέχονται αισθητηριακά δεδομένα εισόδου. Αυτές οι αντιστοιχίες μπορούν να αλλάζουν ριζικά μέσα σε μερικές εβδομάδες, και μερικά ζώα φαίνεται να έχουν πολλαπλές αντιστοιχίες. Επίσης, δεν έχει γίνει εντελώς κατανοητό πώς άλλες περιοχές μπορούν να αναλαμβάνουν λειτουργίες όταν υπάρχει βλάβη σε μια περιοχή. Δεν υπάρχει καμία ικανοποιητική θεωρία για το πώς αποθηκεύεται μια μεμονωμένη ανάμνηση.

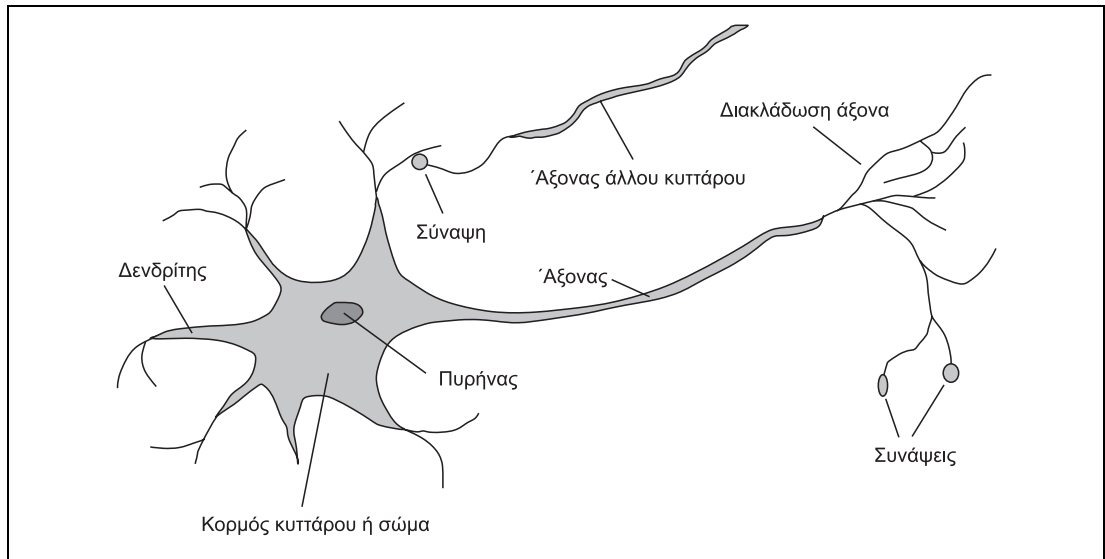
Η μέτρηση της ανέπαφης εγκεφαλικής δραστηριότητας ξεκίνησε το 1929 με την εφεύρεση του ηλεκτροεγκεφαλογράφου (electroencephalograph, EEG) από τον Hans Berger. Η πρόσφατη εξέλιξη της λειτουργικής απεικόνισης μέσω μαγνητικού συντονισμού (functional magnetic resonance imaging, fMRI — Ogawa κ.α., 1990) παρέχει στους νευροεπιστήμονες πρωτοφανείς λεπτομερείς εικόνες της εγκεφαλικής δραστηριότητας, οι οποίες επιτρέπουν μετρήσεις που παρουσιάζουν ενδιαφέρουσες αντιστοιχίες με τις γνωστικές διαδικασίες που βρίσκονται σε εξέλιξη. Οι

<sup>6</sup> Από τότε, έχει ανακαλυφθεί ότι μερικά είδη δελφινιών και φαλαινών έχουν αναλογικά μεγαλύτερο εγκέφαλο. Το αυξημένο μέγεθος του ανθρώπινου εγκεφάλου θεωρείται σήμερα ότι σε κάποιο βαθμό έγινε εφικτό λόγω πρόσφατων βελτιώσεων στο σύστημα ψύξης του.

<sup>7</sup> Πολλοί αναφέρουν τον Alexander Hood (1824) ως πιθανή προηγούμενη πηγή.

<sup>8</sup> Ο Golgi επέμενε στην πεποίθησή του ότι οι εγκεφαλικές λειτουργίες πραγματοποιούνται κυρίως σε ένα συνεχές μέσο στο οποίο ήταν ενσωματωμένοι οι νευρώνες, ενώ ο Cajal υποστήριζε το “νευρωνικό δόγμα”. Οι δύο τους μοιράστηκαν το βραβείο Νόμπελ το 1906, αλλά οι ομιλίες τους στην απονομή ήταν μάλλον ανταγωνιστικές.

παρατηρήσεις αυτές ενισχύονται από τις προόδους στην καταγραφή της νευρωνικής δραστηριότητας μεμονωμένων κυττάρων. Παρ' όλες αυτές τις προόδους, απέχουμε ακόμα πολύ από την κατανόηση του πώς λειτουργεί πραγματικά οποιαδήποτε από αυτές τις γνωστικές διαδικασίες.



**Εικόνα 1.2** Τα μέρη ενός νευρικού κυττάρου, ή νευρώνα (neuron). Κάθε νευρώνας αποτελείται από έναν κυτταρικό κορμό, ή σώμα (soma), ο οποίος περιέχει έναν κυτταρικό πυρήνα (nucleus). Από τον κυτταρικό κορμό διακλαδίζονται μερικές ίνες, που λέγονται δενδρίτες (dendrites), και μία μοναδική μεγάλη ίνα, που λέγεται άξονας (axon). Ο άξονας εκτείνεται σε μεγάλο μήκος, πολύ μεγαλύτερο από αυτό που φαίνεται στην κλίμακα του διαγράμματος. Συνήθως, οι άξονες έχουν μήκος 1 εκατοστό (100 φορές μεγαλύτερο από τη διάμετρο του κυτταρικού κορμού), αλλά μπορούν να φτάνουν και το 1 μέτρο. Ένας νευρώνας συνδέεται με 10 έως 100.000 άλλους νευρώνες, μέσω σημείων σύνδεσης που λέγονται συνάψεις (synapses). Τα σήματα διαδίδονται από νευρώνα σε νευρώνα με μια περίπλοκη ηλεκτροχημική αντίδραση. Τα σήματα αυτά ελέγχουν την εγκεφαλική δραστηριότητα βραχυπρόθεσμα, και επιτρέπουν επίσης μακρόχρονες αλλαγές στη θέση και τη συνδετικότητα των νευρώνων. Οι μηχανισμοί αυτοί θεωρείται ότι αποτελούν τη βάση της μάθησης στον εγκέφαλο. Το μεγαλύτερο μέρος της επεξεργασίας πληροφοριών γίνεται στον εγκεφαλικό φλοιό, το εξωτερικό στρώμα του εγκεφάλου. Η βασική μονάδα οργάνωσης φαίνεται να είναι μια στήλη ιστού με διάμετρο γύρω στο μισό εκατοστό που εκτείνεται σε όλο το βάθος του φλοιού, το οποίο είναι γύρω στα 4 χιλιοστά στον άνθρωπο. Μια τέτοια στήλη περιέχει περίπου 20.000 νευρώνες.

	Υπολογιστής	Ανθρώπινος εγκέφαλος
Υπολογιστικές μονάδες	1 CPU, $10^8$ πύλες	$10^{11}$ νευρώνες
Αποθηκευτικές μονάδες	$10^{10}$ bit RAM $10^{11}$ bit δίσκου	$10^{11}$ νευρώνες $10^{14}$ συνάψεις
Χρόνος κύκλου	$10^{-9}$ sec	$10^{-3}$ sec
Εύρος ζώνης	$10^{10}$ bit/sec	$10^{14}$ bit/sec
Ενημερώσεις μνήμης/sec	$10^9$	$10^{14}$

**Εικόνα 1.3** Πρόχειρη σύγκριση των καθαρών υπολογιστικών πόρων που διαθέτουν οι υπολογιστές (το 2003) και ο εγκέφαλος. Οι αριθμοί για τους υπολογιστές έχουν αυξηθεί όλοι τουλάχιστον κατά 10 φορές από την πρώτη έκδοση του βιβλίου, και αυτό αναμένεται να επαναληφθεί μέσα σε αυτή τη δεκαετία. Οι αριθμοί για τον εγκέφαλο δεν έχουν αλλάξει τα τελευταία 10.000 χρόνια.



Το πραγματικά εκπληκτικό συμπέρασμα είναι ότι *μια συλλογή μεμονωμένων κυττάρων μπορεί να οδηγήσει στη σκέψη, τη δράση, και τη συναίσθηση* ή, με άλλα λόγια, ότι *ο εγκέφαλος δημιουργεί νόηση* (Searle, 1992). Η μόνη πραγματικά εναλλακτική θεωρία είναι ο μυστικισμός: ότι υπάρχει κάποιο μυστηριώδες βασίλειο όπου λειτουργούν τα πνεύματα, το οποίο βρίσκεται πέρα από τις φυσικές επιστήμες.

Οι εγκέφαλοι και οι ψηφιακοί υπολογιστές πραγματοποιούν πολύ διαφορετικές εργασίες και έχουν διαφορετικές ιδιότητες. Η Εικόνα 1.3 δείχνει ότι οι νευρώνες που υπάρχουν στον συνήθη ανθρώπινο εγκέφαλο είναι 1000 φορές περισσότεροι από τις λογικές πύλες στη CPU ενός συνήθη τεχνολογικά προηγμένου υπολογιστή. Ο νόμος του Moore<sup>9</sup> προβλέπει ότι ο αριθμός των λογικών πυλών των CPU θα γίνει ίσος με τον αριθμό των νευρώνων του εγκεφάλου γύρω στο 2020. Φυσικά, δεν μπορούμε να συμπεράνουμε πολλά από τέτοιες προγνώσεις: επίσης, η διαφορά στην αποθηκευτική ικανότητα είναι δευτερεύουσα σε σύγκριση με τη διαφορά στην ταχύτητα μεταγωγής (switching) και την παραλληλία. Τα τσιπ των υπολογιστών μπορούν να εκτελούν μία εντολή σε ένα nanosecond ενώ οι νευρώνες είναι εκατομμύρια φορές πιο αργοί. Ο εγκέφαλος, όμως, αντισταθμίζει αυτό το μειονέκτημα με το παραπάνω, επειδή όλοι οι νευρώνες και οι συνάψεις είναι ταυτόχρονα ενεργοί, ενώ οι περισσότεροι σημερινοί υπολογιστές έχουν μόνο μία ή, στην καλύτερη περίπτωση, μερικές CPU. Επομένως, *αν και ένας υπολογιστής είναι ένα εκατομμύριο φορές γρηγορότερος σε καθαρή ταχύτητα μεταγωγής, ο εγκέφαλος είναι τελικά 100.000 φορές γρηγορότερος στη δουλειά που κάνει.*



## Ψυχολογία (1879 – σήμερα)

- Πώς σκέπτονται και πώς ενεργούν οι άνθρωποι και τα ζώα;

Οι ρίζες της επιστημονικής ψυχολογίας συνήθως εντοπίζονται στην εργασία του Γερμανού φυσικού Hermann von Helmholtz (1821 – 1894) και του μαθητή του Wilhelm Wundt (1832 – 1920). Ο Helmholtz εφάρμοσε την επιστημονική μέθοδο στη μελέτη της ανθρώπινης όρασης, και το σύγγραμμά του *Handbook of Physiological Optics* χαρακτηρίζεται ακόμα και σήμερα ως “η σημαντικότερη πραγματεία της φυσικής και της φυσιολογίας της ανθρώπινης όρασης” (Nalwa, 1993, σ.15). Το 1879, ο Wundt άνοιξε το πρώτο εργαστήριο πειραματικής ψυχολογίας στο Πανεπιστήμιο της Λειψίας. Ο Wundt επέμεινε σε προσεκτικά ελεγχόμενα πειράματα στα οποία οι συνεργάτες του πραγματοποιούσαν μια αντιληπτική ή συνειρμική εργασία ενώ ενδοσκοπούσαν τις διαδικασίες της σκέψης τους. Οι προσεκτικοί έλεγχοι συνέβαλαν σε μεγάλο βαθμό στο να γίνει η ψυχολογία επιστήμη, αλλά η υποκειμενική φύση των δεδομένων έκαναν απίθανο το να διαψευσθεί ποτέ η θεωρία κάποιου από ένα πείραμα. Οι βιολόγοι που μελετούσαν τη συμπεριφορά των ζώων, από την άλλη, δεν είχαν στη διάθεσή τους τέτοια ενδοσκοπικά δεδομένα και ανέπτυξαν μια αντικειμενική μεθοδολογία, όπως περιγράφεται από τον H. S. Jennings (1906) στη βαρυσήμαντη εργασία του *Behavior of the Lower Organisms* (*Συμπεριφορά των κατώτερων οργανισμών*). Με την εφαρμογή αυτής της μεθοδολογίας στους ανθρώπους, το κίνημα του **συμπεριφορισμού** (behaviorism), με ηγέτη τον John Watson (1878 – 1958), απέρριπτε οποιαδήποτε θεωρία χρησιμοποιούσε νοητικές διαδικασίες, με το επιχείρημα ότι η ενδοσκόπηση δεν μπορούσε να δώσει αξιόπιστα τεκμήρια. Οι συμπεριφοριστές επέμεναν στη μελέτη μόνο των αντικειμενικών μετρήσεων των αντιλήψεων (των *ερεθισμάτων*) που παρέχονταν σε ένα ζώο και των αντιδράσεών του (των *αποκρίσεων*). Διανοητικές κατασκευές όπως η γνώση, οι πεποιθήσεις, οι στόχοι και τα συλλογιστικά βήματα απορρίπτονταν ως αντιεπιστημονική “λαϊκή ψυχολογία” (folk psychology). Ο συμπεριφορισμός ανακάλυψε πολλά για τους αρουραίους και τα περιστέρια, αλλά είχε πολύ μικρότερη επιτυχία στην κατανόηση των ανθρώπων. Ωστόσο, άσκησε ισχυρή επιρροή στην ψυχολογία (ιδιαίτερα στις Ηνωμένες Πολιτείες) από το 1920 έως το 1960.

<sup>9</sup> Ο νόμος του Moore λέει ότι ο αριθμός των τρανζίστορ ανά τετραγωνική ίντσα διπλασιάζεται κάθε 1 έως 1,5 χρόνο. Η χωρητικότητα του ανθρώπινου εγκεφάλου διπλασιάζεται περίπου κάθε 2 έως 4 εκατομμύρια χρόνια.

Η άποψη του εγκεφάλου ως συσκευής επεξεργασίας πληροφοριών, η οποία είναι κύριο χαρακτηριστικό της **γνωστικής ψυχολογίας** (cognitive psychology), συναντάται ήδη στις εργασίες του William James<sup>10</sup> (1842 – 1910). Ο Helmholtz επίσης επέμενε ότι η αντίληψη περιλάμβανε μια μορφή ασυναίσθητου λογικού συμπερασμού. Η γνωστική άποψη επισκιάστηκε σε μεγάλο βαθμό από το συμπεριφορισμό στις Ηνωμένες Πολιτείες, αλλά στη “Μονάδα Εφαρμοσμένης Ψυχολογίας” του Cambridge, με την καθοδήγηση του Frederic Bartlett (1886 – 1969), η γνωστική μοντελοποίηση μπόρεσε να ευδοκιμήσει. Το βιβλίο *The Nature of Explanation (Η φύση της εξήγησης)* του Kenneth Craik (1943), μαθητή και διαδόχου του Bartlett, αποκατέστησε σθεναρά τη νομιμότητα “πνευματικών” όρων όπως οι πεποιθήσεις και οι στόχοι, υποστηρίζοντας ότι είναι εξίσου επιστημονικοί όπως, για παράδειγμα, η χρήση της πίεσης και της θερμοκρασίας όταν μιλάμε για τα αέρια, παρά το γεγονός ότι τα αέρια αποτελούνται από μόρια τα οποία δεν έχουν τίποτα από τα δύο. Ο Craik καθόρισε τα τρία βασικά βήματα ενός πράκτορα βασισμένου στη γνώση: (1) Το ερέθισμα πρέπει να μεταφράζεται σε μια εσωτερική αναπαράσταση, (2) η αναπαράσταση υπόκειται σε επεξεργασία από γνωστικές διαδικασίες για να προκύψουν νέες εσωτερικές αναπαραστάσεις, και (3) αυτές με τη σειρά τους μεταφράζονται ξανά σε δράση. Ο Craik εξήγησε με σαφήνεια γιατί η σχεδίαση αυτή είναι καλή για έναν πράκτορα:

Αν ο οργανισμός έχει μέσα στο κεφάλι του ένα “μοντέλο μικρής κλίμακας” της εξωτερικής πραγματικότητας και των δικών του δυνατών ενεργειών, μπορεί να δοκιμάζει διάφορες εναλλακτικές ενέργειες, να καταλήγει στην καλύτερη από αυτές, να αντιδρά σε μελλοντικές καταστάσεις πριν ανακύψουν, να αξιοποιεί τη γνώση παρελθόντων συμβάντων για να αντιμετωπίζει το παρόν και το μέλλον, και γενικά να αντιδρά με πολύ πιο ολοκληρωμένο, ασφαλές και σωστό τρόπο στα περιστατικά που αντιμετωπίζει. (Craik, 1943)

Μετά το θάνατο του Craik σε ποδηλατικό ατύχημα το 1945, τη δουλειά του συνέχισε ο Donald Broadbent, που το βιβλίο του *Perception and Communication (Αντίληψη και επικοινωνία — 1958)* περιλάμβανε μερικά από τα πρώτα μοντέλα επεξεργασίας πληροφοριών για ψυχολογικά φαινόμενα. Στο μεταξύ, στις Ηνωμένες Πολιτείες, η ανάπτυξη της υπολογιστικής μοντελοποίησης οδήγησε στη δημιουργία του πεδίου της **γνωστικής επιστήμης** (cognitive science). Το πεδίο αυτό μπορεί να θεωρηθεί ότι ξεκίνησε σε μια συνάντηση εργασίας (workshop) στο MIT το Σεπτέμβριο του 1956. (Όπως θα δούμε, αυτό συνέβη μόλις δύο μήνες μετά τη συνάντηση στην οποία “γεννήθηκε” η ίδια η TN.) Στη συνάντηση αυτή, ο George Miller παρουσίασε το “*Μαγικό αριθμό επτά*” (*The Magic Number Seven*), ο Noam Chomsky παρουσίασε τα “*Τρία μοντέλα της γλώσσας*” (*Three Models of Language*), και οι Allen Newell και Herbert Simon παρουσίασαν τη “*Μηχανή λογικής θεωρίας*” (*The Logic Theory Machine*). Αυτές οι τρεις βαρυσήμαντες εργασίες έδειξαν πώς μπορούσαν να χρησιμοποιηθούν υπολογιστικά μοντέλα στην ψυχολογία της μνήμης, της γλώσσας και της λογικής σκέψης, αντίστοιχα. Είναι σήμερα κοινά παραδεκτό μεταξύ των ψυχολόγων ότι “μια γνωστική θεωρία θα πρέπει να μοιάζει με πρόγραμμα υπολογιστή” (Anderson, 1980), δηλαδή, θα πρέπει να περιγράφει ένα λεπτομερές μηχανισμό επεξεργασίας πληροφοριών με τον οποίο θα μπορούσαν να υλοποιούνται μερικές γνωστικές λειτουργίες.

## Τεχνολογία υπολογιστών (1940 – σήμερα)

- Πώς μπορούμε να κατασκευάσουμε έναν αποδοτικό υπολογιστή;

Για την επιτυχία της TN, χρειάζονται δύο πράγματα: νοημοσύνη και ένα τεχνούργημα. Το τεχνούργημα της επιλογής μας είναι ο υπολογιστής. Ο σημερινός ψηφιακός ηλεκτρονικός υπολογιστής εφευρέθηκε ανεξάρτητα και σχεδόν ταυτόχρονα από επιστήμονες τριών εμπόλεμων χωρών στο Δεύτερο Παγκόσμιο Πόλεμο. Ο πρώτος *λειτουργικός* υπολογιστής ήταν ο ηλεκτρομηχανικός

<sup>10</sup> Ο William James ήταν αδελφός του λογοτέχνη Henry James. Λέγεται ότι ο Henry έγραφε λογοτεχνία σαν να ήταν ψυχολογία και ο William έγραφε ψυχολογία σαν να ήταν λογοτεχνία.

υπολογιστής Heath Robinson,<sup>11</sup> που κατασκευάστηκε το 1940 από την ομάδα του Alan Turing για ένα και μόνο σκοπό: την αποκρυπτογράφηση των γερμανικών μηνυμάτων. Το 1943, η ίδια ομάδα ανέπτυξε τον Colossus, μια ισχυρή μηχανή γενικής χρήσης που βασιζόταν σε λυχνίες κενού.<sup>12</sup> Ο πρώτος λειτουργικός *προγραμματίσιμος* υπολογιστής ήταν ο Z-3, που εφευρέθηκε από τον Konrad Zuse στη Γερμανία το 1941. Ο Zuse επινόησε επίσης τους αριθμούς κινητής υποδιαστολής και την πρώτη γλώσσα προγραμματισμού υψηλού επιπέδου, την Plankalkül. Ο πρώτος *ηλεκτρονικός* υπολογιστής, ο ABC, κατασκευάστηκε από τον John Atanasoff και το μαθητή του Clifford Berry μεταξύ 1940 και 1942 στο Πανεπιστήμιο της Iowa. Η έρευνα του Atanasoff είχε μικρή υποστήριξη ή αναγνώριση· ο υπολογιστής που αναδείχτηκε σε πιο σημαντικό πρόδρομο των σημερινών υπολογιστών ήταν ο ENIAC, ο οποίος κατασκευάστηκε στα πλαίσια ενός μυστικού στρατιωτικού έργου στο Πανεπιστήμιο της Pennsylvania από μια ομάδα όπου συμμετείχαν οι John Mauchly και John Eckert.

Στο μισό αιώνα που πέρασε από τότε, κάθε γενιά υλικού υπολογιστών έφερνε αύξηση στην ταχύτητα και τη χωρητικότητα και μείωση στις τιμές. Η απόδοση διπλασιάζεται κάθε 18 μήνες περίπου, και αυτός ο ρυθμός αύξησης προβλέπεται να διατηρηθεί για μία ή δύο δεκαετίες ακόμα. Μετά από αυτό, θα χρειαστούμε μοριακή τεχνολογία ή κάποια άλλη νέα τεχνολογία.

Φυσικά, συσκευές υπολογισμών υπήρχαν και πριν από τον ηλεκτρονικό υπολογιστή. Οι πρώτες αυτόματες μηχανές, που χρονολογούνται από το 17ο αιώνα, αναφέρθηκαν στην ενότητα για τη φιλοσοφία. Η πρώτη *προγραμματίσιμη* μηχανή ήταν ένας αργαλειός που επινοήθηκε το 1805 από τον Joseph Marie Jacquard (1752 – 1834), ο οποίος χρησιμοποιούσε διάτρητες κάρτες για την αποθήκευση οδηγιών για το μοτίβο που θα ύφαινε. Στα μέσα του 19ου αιώνα, ο Charles Babbage (1792 – 1871) σχεδίασε δύο μηχανές, χωρίς να ολοκληρώσει καμία από τις δύο. Η “Μηχανή Διαφορών” (Difference Engine), η οποία εμφανίζεται στο εξώφυλλο αυτού του βιβλίου, προοριζόταν για υπολογισμούς μαθηματικών πινάκων για τεχνολογικά και επιστημονικά έργα. Τελικά, κατασκευάστηκε και αποδείχτηκε ότι λειτουργεί το 1991 στο Μουσείο Επιστημών του Λονδίνου (Swade, 1993). Η “Αναλυτική Μηχανή” (Analytical Engine) του Babbage ήταν πολύ πιο φιλόδοξη· περιλάμβανε διευθυνσιοδοτήσιμη μνήμη, αποθηκευμένα προγράμματα και άλματα υπό συνθήκη, και ήταν το πρώτο τεχνούργημα που ήταν ικανό για καθολικό υπολογισμό. Η συνεργάτιδα του Babbage, Ada Lovelace, κόρη του ποιητή Λόρδου Byron, ήταν ίσως ο πρώτος προγραμματιστής που υπήρξε. (Η γλώσσα προγραμματισμού Ada ονομάστηκε προς τιμήν της.) Έγραψε προγράμματα για την ημιτελή Αναλυτική Μηχανή, και ακόμα υποστήριξε ότι η μηχανή μπορούσε να παίζει σκάκι ή να συνθέσει μουσική.

Η ΤΝ οφείλει επίσης πολλά στην πλευρά του λογισμικού της επιστήμης των υπολογιστών, η οποία της έδωσε τα λειτουργικά συστήματα, τις γλώσσες προγραμματισμού, και τα απαραίτητα εργαλεία για το γράψιμο σύγχρονων προγραμμάτων (και επιστημονικών δημοσιεύσεων γι’ αυτά). Αυτό όμως είναι ένα πεδίο όπου η οφειλή έχει ξεπληρωθεί· η εργασία στην ΤΝ έχει παραγάγει πολλές πρωτοποριακές ιδέες που βρήκαν το δρόμο τους πίσω στην κλασική επιστήμη των υπολογιστών, στις οποίες περιλαμβάνονται ο χρονομερισμός (time sharing), οι αλληλεπιδραστικοί ερμηνευτές (interactive interpreters), οι προσωπικοί υπολογιστές με παράθυρα και ποντίκι, τα περιβάλλοντα γρήγορης ανάπτυξης εφαρμογών, ο τύπος δεδομένων της συνδεδεμένης λίστας, η αυτόματη διαχείριση αποθήκευσης, και βασικές έννοιες του συμβολικού, συναρτησιακού, δυναμικού και αντικειμενοστρεφούς προγραμματισμού.

<sup>11</sup> Ο Heath Robinson ήταν ένας σκιτσογράφος διάσημος για τις απεικονίσεις εκκεντρικών και παράλογα περίπλοκων συσκευών για καθημερινές δουλειές όπως το βουτύρωμα των τoστ.

<sup>12</sup> Στη μεταπολεμική περίοδο, ο Turing ήθελε να χρησιμοποιήσει αυτούς τους υπολογιστές για έρευνα στην ΤΝ — για παράδειγμα, για ένα από τα πρώτα σκακιστικά προγράμματα (Turing κ.α., 1953). Οι προσπάθειές του απαγορεύτηκαν από τη βρετανική κυβέρνηση.

## Θεωρία ελέγχου και κυβερνητική (1948 – σήμερα)

- Πώς μπορούν τα τεχνουργήματα να λειτουργούν υπό το δικό τους έλεγχο;

Ο Κτησίβιος ο Αλεξανδρινός (περ. 250 π.Χ.) κατασκεύασε την πρώτη αυτοελεγχόμενη μηχανή: Ένα υδραυλικό ρολόι με ρυθμιστή που διατηρούσε τη ροή του νερού που περνούσε μέσα από αυτό σε ένα σταθερό, προβλέψιμο ρυθμό. Αυτή η εφεύρεση άλλαξε το τι μπορούσε να κάνει ένα τεχνουργήμα. Πριν από αυτό, μόνο έμβια όντα μπορούσαν να τροποποιούν τη συμπεριφορά τους αποκρινόμενα σε αλλαγές στο περιβάλλον. Άλλα παραδείγματα αυτορυθμιζόμενων συστημάτων ελέγχου με ανάδραση είναι ο ρεγουλαδόρος (governor) της ατμομηχανής, που εφευρέθηκε από τον James Watt (1736 – 1819), και ο θερμοστάτης, που εφευρέθηκε από τον Cornelis Drebbel (1572 – 1633), ο οποίος επινόησε επίσης το υποβρύχιο. Η μαθηματική θεωρία των σταθερών συστημάτων ανάδρασης αναπτύχθηκε το 19ο αιώνα.

Το κεντρικό πρόσωπο στη δημιουργία της λεγόμενης **θεωρίας ελέγχου** (control theory) ήταν ο Norbert Wiener (1894 – 1964). Ο Wiener ήταν ένας σπουδαίος μαθηματικός που συνεργάστηκε, μεταξύ άλλων, με τον Bertrand Russell, πριν εκδηλώσει ενδιαφέρον για τα βιολογικά και μηχανικά συστήματα ελέγχου και τη σχέση τους με τη γνωστική λειτουργία. Όπως και ο Craik (ο οποίος επίσης χρησιμοποίησε συστήματα ελέγχου ως ψυχολογικά μοντέλα), ο Wiener και οι συνεργάτες του Arturo Rosenblueth και Julian Bigelow αμφισβήτησαν το κατεστημένο του συμπεριφορισμού (Rosenblueth κ.α., 1943). Έβλεπαν τη σκόπιμη συμπεριφορά ως αποτέλεσμα ενός ρυθμιστικού μηχανισμού που προσπαθούσε να ελαχιστοποιεί το “σφάλμα” — τη διαφορά μεταξύ της παρούσας και της επιδιωκόμενης κατάστασης. Στα τέλη της δεκαετίας του 1940, ο Wiener, σε συνεργασία με τους Warren McCulloch, Walter Pitts και John von Neumann, οργάνωσε μια σειρά επιστημονικών διασκέψεων οι οποίες εξερεύνησαν τα νέα μαθηματικά και υπολογιστικά μοντέλα γνωστικής λειτουργίας και επηρέασαν πολλούς άλλους ερευνητές των επιστημών της συμπεριφοράς. Το βιβλίο του Wiener *Cybernetics* (*Κυβερνητική* — 1948) έγινε μπεστ σέλερ και αφύπνισε στο κοινό την ιδέα της δυνατότητας δημιουργίας μηχανών με τεχνητή νοημοσύνη.

Η σημερινή θεωρία ελέγχου, και ιδιαίτερα ο κλάδος που είναι γνωστός ως στοχαστικός βέλτιστος έλεγχος (stochastic optimal control), έχει στόχο τη σχεδίαση συστημάτων που μεγιστοποιούν μια **αντικειμενική συνάρτηση** (objective function) μέσα στο χρόνο. Αυτό αντιστοιχεί χονδρικά στην άποψή μας για την TN: σχεδίαση συστημάτων που συμπεριφέρονται βέλτιστα. Γιατί λοιπόν η TN και η θεωρία ελέγχου είναι δύο διαφορετικά πεδία, ιδιαίτερα αν πάρουμε υπόψη τη στενή σύνδεση μεταξύ των ατόμων που τα θεμελίωσαν; Η απάντηση βρίσκεται στη στενή σύζευξη ανάμεσα στις μαθηματικές τεχνικές που ήταν οικείες στους συμμετέχοντες και στα αντίστοιχα σύνολα προβλημάτων που είχε να αντιμετωπίσει η κάθε πλευρά. Ο μαθηματικός λογισμός και η άλγεβρα μητρώων, τα εργαλεία της θεωρίας ελέγχου, προσφέρονταν για συστήματα που περιγράφονται από σταθερά σύνολα συνεχών μεταβλητών· ακόμα, η ακριβής ανάλυση είναι γενικά εφικτή μόνο για τα *γραμμικά* συστήματα. Το πεδίο της TN, σε κάποιο βαθμό, θεμελιώθηκε ως ένας τρόπος να ξεπεραστούν οι μαθηματικοί περιορισμοί της θεωρίας ελέγχου τη δεκαετία του 1950. Τα εργαλεία του λογικού συμπερασμού και του υπολογισμού επέτρεψαν στους ερευνητές της TN να εξετάζουν προβλήματα όπως η γλώσσα, η όραση, και ο σχεδιασμός, τα οποία βρίσκονταν εντελώς έξω από το πεδίο των ερευνητών της θεωρίας ελέγχου.

## Γλωσσολογία (1957 – σήμερα)

- Πώς σχετίζεται η γλώσσα με τη σκέψη;

Το 1957, ο B. F. Skinner δημοσίευσε το *Verbal Behavior* (*Λεκτική συμπεριφορά*). Ήταν μια ολοκληρωμένη και λεπτομερής έκθεση της συμπεριφορικής προσέγγισης στη μάθηση της γλώσσας, γραμμένη από τον επιφανέστερο ειδικό του πεδίου. Κατά παράξενο τρόπο, όμως, μια κριτική του

ΘΕΩΡΙΑ  
ΕΛΕΓΧΟΥ

ΚΥΒΕΡΝΗΤΙΚΗ

ΑΝΤΙΚΕΙΜΕΝΙΚΗ  
ΣΥΝΑΡΤΗΣΗ



βιβλίου έγινε εξίσου διάσημη με το ίδιο το βιβλίο, και σχεδόν σκότωσε το ενδιαφέρον για το συμπεριφορισμό. Συγγραφέας αυτής της κριτικής ήταν ο Noam Chomsky, ο οποίος είχε ήδη δημοσιεύσει ένα βιβλίο για τη δική του θεωρία, το *Syntactic Structures* (*Συντακτικές δομές*). Ο Chomsky έδειξε πώς η συμπεριφορική θεωρία αδυνατούσε να αντιμετωπίσει την έννοια της δημιουργικότητας στη γλώσσα — δεν εξηγούσε πώς ένα παιδί μπορούσε να καταλαβαίνει και να επινοεί προτάσεις που δεν είχε ακούσει ποτέ πριν. Η θεωρία του Chomsky — η οποία βασιζόταν σε συντακτικά μοντέλα που έφταναν πίσω στο χρόνο μέχρι τον Ινδό γλωσσολόγο Panini (περ. 350 π.Χ.) — μπορούσε να εξηγήσει αυτό το φαινόμενο και, σε αντίθεση με τις προηγούμενες θεωρίες, είχε αρκετά τυπική μορφή ώστε να μπορεί θεωρητικά να προγραμματιστεί.

Η σύγχρονη γλωσσολογία και η TN, “γεννήθηκαν” λοιπόν περίπου την ίδια εποχή, και αναπτύχθηκαν μαζί, με κοινή περιοχή ένα υβριδικό πεδίο που ονομάζεται **υπολογιστική γλωσσολογία** (computational linguistics) ή **επεξεργασία φυσικής γλώσσας** (natural language processing). Το πρόβλημα της κατανόησης της γλώσσας σύντομα αποδείχτηκε ότι ήταν πολύ πιο σύνθετο από ό,τι φαινόταν το 1957. Η κατανόηση της γλώσσας απαιτεί κατανόηση του θέματος και των συμφραζομένων, και όχι μόνο κατανόηση της δομής των προτάσεων. Αυτό μπορεί να φαίνεται προφανές, αλλά η σημασία του δεν εκτιμήθηκε αρκετά μέχρι τη δεκαετία του 1960. Μεγάλο μέρος της πρώιμης δουλειάς στον τομέα της **αναπαράστασης γνώσης** (knowledge representation — η μελέτη του πώς η γνώση μπορεί να τεθεί σε μια μορφή με την οποία ένας υπολογιστής να μπορεί να συλλογίζεται) ήταν συνδεδεμένη με τη γλώσσα και έπαιρνε υπόψη τη γλωσσολογική έρευνα, η οποία με τη σειρά της ήταν συνδεδεμένη με δεκαετίες δουλειάς πάνω στη φιλοσοφική ανάλυση της γλώσσας.

ΥΠΟΛΟΓΙΣΤΙΚΗ  
ΓΛΩΣΣΟΛΟΓΙΑ

## 1.3 ΙΣΤΟΡΙΑ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Αφού τελειώσαμε με το υπόβαθρο, είμαστε τώρα έτοιμοι να ασχοληθούμε με την εξέλιξη της ίδια της TN.

### Η κυοφορία της τεχνητής νοημοσύνης (1943 – 1955)

Η πρώτη εργασία που αναγνωρίζεται σήμερα γενικά ως TN έγινε από τους Warren McCulloch και Walter Pitts (1943). Βασίστηκαν σε τρεις πηγές: στη γνώση της βασικής φυσιολογίας και λειτουργίας των νευρώνων του εγκεφάλου, σε μια τυπική ανάλυση της προτασιακής λογικής των Russell και Whitehead, και στη θεωρία υπολογισμού του Turing. Πρότειναν ένα μοντέλο τεχνητών νευρώνων, όπου κάθε νευρώνας χαρακτηρίζεται ως ενεργοποιημένος (“on”) ή απενεργοποιημένος (“off”), και μεταβαίνει στην ενεργοποιημένη κατάσταση αποκρινόμενος σε διέγερση από έναν επαρκή αριθμό γειτονικών νευρώνων. Η κατάσταση ενός νευρώνα θεωρούνταν ως “ουσιαστικά ισοδύναμη με μια πρόταση η οποία υποδεικνύει το επαρκές της ερέθισμα”. Έδειξαν, για παράδειγμα, ότι οποιαδήποτε υπολογίσιμη συνάρτηση μπορούσε να υπολογιστεί από κάποιο δίκτυο συνδεδεμένων νευρώνων και ότι όλοι οι λογικοί σύνδεσμοι (και, ή, όχι, κ.λπ.) μπορούσαν να υλοποιηθούν με απλές δικτυακές δομές. Οι McCulloch και Pitts υποστήριζαν επίσης ότι κατάλληλα ορισμένα δίκτυα μπορούσαν να μαθαίνουν. Ο Donald Hebb (1949) παρουσίασε έναν απλό κανόνα ενημέρωσης για την τροποποίηση των συνδετικών δυνάμεων μεταξύ νευρώνων. Ο κανόνας αυτός, που σήμερα ονομάζεται **μάθηση Hebb** (Hebbian learning), παραμένει μέχρι σήμερα ένα σημαντικό μοντέλο.

Δύο μεταπτυχιακοί φοιτητές του μαθηματικού τμήματος του Princeton, ο Marvin Minsky και ο Dean Edmonds, κατασκεύασαν τον πρώτο υπολογιστή νευρωνικού δικτύου το 1951. Ο SNARC, όπως ονομάστηκε, χρησιμοποιούσε 3000 λυχνίες κενού και ένα μηχανισμό αυτόματου πιλότου από βομβαρδιστικό B-24 για να προσομοιώνει ένα δίκτυο 40 νευρώνων. Η επιτροπή για το Ph.D. του Minsky έδειχνε σκεπτικισμό για το αν μια τέτοια εργασία μπορούσε να θεωρηθεί

μαθηματικά αλλά, όπως λέγεται, ο von Neumann είπε: “Αν δεν είναι σήμερα, θα είναι κάποια μέρα.” Αργότερα, ο Minsky απέδειξε σημαντικά θεωρήματα που έδειξαν τους περιορισμούς της έρευνας στα νευρωνικά δίκτυα.

Υπήρχαν μερικά πρώιμα παραδείγματα εργασιών που θα μπορούσαν να χαρακτηριστούν TN, αλλά ο Alan Turing ήταν εκείνος που διατύπωσε πρώτος ένα πλήρες όραμα για την TN το 1950, στο άρθρο του “Computing Machinery and Intelligence”. Στο άρθρο αυτό, παρουσίασε τη δοκιμασία Turing, τη μηχανική μάθηση, τους γενετικούς αλγόριθμους, και την ενισχυτική μάθηση.

## Η γέννηση της τεχνητής νοημοσύνης (1956)

Το Princeton φιλοξένησε επίσης μια άλλη σημαντική φυσιογνωμία της TN, τον John McCarthy. Μετά την αποφοίτησή του, ο McCarthy μετακόμισε στο Dartmouth College, το οποίο έμελλε να γίνει ο επίσημος τόπος γέννησης του πεδίου. Ο McCarthy έπεισε τους Minsky, Claude Shannon και Nathaniel Rochester να τον βοηθήσουν να φέρει σε επαφή Αμερικανούς ερευνητές που ενδιαφέρονταν για τη θεωρία των αυτομάτων, τα νευρωνικά δίκτυα και τη μελέτη της νοημοσύνης. Οργάνωσαν μια δίμηνη συνάντηση εργασίας στο Dartmouth το καλοκαίρι του 1956. Συμμετείχαν συνολικά 10 άτομα, μεταξύ των οποίων ο Trenchard More από το Princeton, ο Arthur Samuel από την IBM, και οι Ray Solomonoff και Oliver Selfridge από το MIT.

Δύο ερευνητές από το Carnegie Tech,<sup>13</sup> ο Allen Newell και ο Herbert Simon, μάλλον έκλεψαν την παράσταση. Αν και οι άλλοι είχαν ιδέες και σε μερικές περιπτώσεις προγράμματα για συγκεκριμένες εφαρμογές όπως η ντάμα, οι Newell και Simon είχαν ήδη ένα πρόγραμμα συλλογιστικής, το Logic Theorist (LT), για τον οποίο ο Simon είπε: “Εφεύραμε ένα πρόγραμμα που μπορεί να σκέπτεται μη αριθμητικά, και επομένως λύσαμε το κλασικό πρόβλημα του δυϊσμού νου-σώματος.”<sup>14</sup> Λίγο μετά τη συνάντηση, το πρόγραμμα μπορούσε να αποδείξει τα περισσότερα από τα θεωρήματα του Κεφάλαιου 2 του *Principia Mathematica* των Russell και Whitehead. Λέγεται ότι ο Russell ενθουσιάστηκε όταν ο Simon του έδειξε ότι το πρόγραμμα είχε βρει μια απόδειξη για ένα θεώρημα η οποία ήταν συντομότερη από εκείνη που υπήρχε στο *Principia*. Οι εκδότες του *Journal of Symbolic Logic*, όμως, δεν εντυπωσιάστηκαν το ίδιο· απέρριψαν μια δημοσίευση που υπογραφόταν από τους Newell, Simon, και το Logic Theorist.

Η συνάντηση του Dartmouth δεν οδήγησε σε νέες ριζοσπαστικές ανακαλύψεις, αλλά έφερε σε επαφή μεταξύ τους όλες τις μεγάλες φυσιογνωμίες του χώρου. Τα επόμενα 20 χρόνια, στο πεδίο θα κυριαρχούσαν αυτοί οι άνθρωποι, οι μαθητές τους, και οι συνεργάτες τους στο MIT, το CMU, το Stanford και την IBM. Το πιο μακρόχρονο επίτευγμα αυτής της συνάντησης ήταν μάλλον η συμφωνία να υιοθετηθεί η νέα ονομασία του McCarthy για το πεδίο: **τεχνητή νοημοσύνη** (artificial intelligence). Ίσως ο όρος “υπολογιστική ορθολογικότητα” θα ήταν καλύτερος, όμως ο όρος “τεχνητή νοημοσύνη” έμεινε.

Κοιτάζοντας την εισήγηση της συνάντησης εργασίας του Dartmouth (McCarthy κ.α., 1955), μπορούμε να δούμε γιατί ήταν ανάγκη να γίνει η TN ξεχωριστό πεδίο. Γιατί δεν μπορούσε όλη η δουλειά που γίνεται στην TN να γίνει κάτω από την ονομασία της θεωρίας ελέγχου ή της επιχειρησιακής έρευνας ή της θεωρίας αποφάσεων, τομείς που είχαν όλοι παρόμοιους στόχους με εκείνους της TN; Ή, γιατί δεν είναι η TN κλάδος των μαθηματικών; Η απάντηση στην πρώτη ερώτηση είναι ότι η TN ενστερνίστηκε από την αρχή την ιδέα της αντιγραφής ανθρώπινων λειτουργιών όπως η δημιουργικότητα, η αυτοβελτίωση και η χρήση γλώσσας. Κανένα από τα άλλα πεδία δεν ασχολείται με αυτά τα ζητήματα. Η απάντηση στη δεύτερη ερώτηση είναι η μεθοδολογία. Η TN είναι το μόνο από αυτά τα πεδία που είναι σαφώς ένας κλάδος της επιστήμης των υπολογιστών (αν και η

<sup>13</sup> Το σημερινό Πανεπιστήμιο Carnegie Mellon (CMU)

<sup>14</sup> Οι Newell και Simon επινόησαν επίσης μια γλώσσα επεξεργασίας λιστών, την IPL, προκειμένου να γράψουν το LT. Δεν είχαν μεταγλωττιστή (compiler), και το μετέφρασαν σε γλώσσα μηχανής οι ίδιοι. Για να αποφύγουν τα λάθη, δούλευαν παράλληλα, εκφωνώντας δυαδικούς αριθμούς ο ένας στον άλλο καθώς έγραφαν κάθε εντολή, για να βεβαιωθούν ότι συμφωνούσαν.

επιχειρησιακή έρευνα δίνει επίσης έμφαση στις προσομοιώσεις με υπολογιστή), και η TN είναι το μόνο πεδίο που προσπαθεί να δημιουργήσει μηχανές με σκοπό να λειτουργούν αυτόνομα μέσα σε πολύπλοκα, μεταβαλλόμενα περιβάλλοντα.

### Πρώιμος ενθουσιασμός, μεγάλες προσδοκίες (1952 – 1969)

Τα πρώτα χρόνια της TN ήταν γεμάτα επιτυχίες — σε περιορισμένο βαθμό, βέβαια. Με δεδομένους τους πρωτόγονους υπολογιστές και τα εργαλεία προγραμματισμού εκείνης της εποχής, και το γεγονός ότι μόλις λίγα χρόνια πριν οι υπολογιστές θεωρούνταν ως μηχανές που μπορούσαν να κάνουν αριθμητικές πράξεις και τίποτα περισσότερο, προκαλούσε έκπληξη όταν ένας υπολογιστής έκανε οτιδήποτε έστω και ελάχιστα έξυπνο. Το πνευματικό κατεστημένο, κατά κανόνα, προτιμούσε να πιστεύει ότι “μια μηχανή δεν θα μπορέσει ποτέ να κάνει το  $X$ ”. (Στο Κεφάλαιο 26 θα βρείτε έναν εκτεταμένο κατάλογο από  $X$  που συγκέντρωσε ο Turing.) Φυσικά, οι ερευνητές της TN απαντούσαν επιδεικνύοντας το ένα  $X$  μετά το άλλο. Ο John McCarthy αναφερόταν σε αυτή την περίοδο ως εποχή του “Κοίτα μαμά, χωρίς χέρια!”

Μετά την πρώτη επιτυχία των Newell και Simon, ακολούθησε ο “γενικός λύτης προβλημάτων” (General Problem Solver — GPS). Αντίθετα από το Logic Theorist, αυτό το πρόγραμμα ήταν σχεδιασμένο από την αρχή για να μιμείται τα ανθρώπινα πρωτόκολλα επίλυσης προβλημάτων. Μέσα στην περιορισμένη κλάση των γρίφων που μπορούσε να χειρίζεται, αποδείχτηκε ότι η σειρά με την οποία το πρόγραμμα εξέταζε τους υποστόχους και τις δυνατές ενέργειες ήταν παρόμοιος με τον τρόπο που προσέγγιζαν οι άνθρωποι τα ίδια προβλήματα. Έτσι, το GPS ήταν μάλλον το πρώτο πρόγραμμα με ενσωματωμένη την προσέγγιση της “ανθρώπινης σκέψης”. Η επιτυχία του GPS και των προγραμμάτων που ακολούθησαν ως μοντέλων της γνωστικής λειτουργίας οδήγησε τους Newell και Simon (1976) να διατυπώσουν την περίφημη υπόθεση του **φυσικού συστήματος συμβόλων** (physical symbol system), η οποία λέει ότι “ένα φυσικό σύστημα συμβόλων έχει τα αναγκαία και επαρκή μέσα για γενική ευφυή δράση”. Αυτό που εννοούσαν είναι ότι οποιοδήποτε σύστημα (άνθρωπος ή μηχανή) που εκδηλώνει νοημοσύνη πρέπει να λειτουργεί με το χειρισμό δομών δεδομένων που αποτελούνται από σύμβολα. Θα δούμε αργότερα ότι αυτή η υπόθεση έχει αμφισβητηθεί από πολλές πλευρές.

Στην IBM, ο Nathaniel Rochester και οι συνεργάτες του δημιούργησαν μερικά από τα πρώτα προγράμματα TN. Ο Herbert Gelernter (1959) δημιούργησε το Geometry Theorem Prover, το οποίο μπορούσε να αποδεικνύει θεωρήματα της γεωμετρίας τα οποία πολλοί φοιτητές των μαθηματικών εύρισκαν πολύ δύσκολα. Από το 1952, ο Arthur Samuel έγραψε μια σειρά προγραμμάτων που έπαιζαν ντάμα, τα οποία τελικά έμαθαν να παίζουν σε καλό ερασιτεχνικό επίπεδο. Στην πορεία, απέδειξε λανθασμένη την ιδέα ότι οι υπολογιστές μπορούν να κάνουν μόνο ότι τους λένε: το πρόγραμμά του έμαθε σύντομα να παίζει καλύτερα από το δημιουργό του. Το πρόγραμμα παρουσιάστηκε στην τηλεόραση το Φεβρουάριο του 1956, προκαλώντας πολύ μεγάλη εντύπωση. Όπως και ο Turing, ο Samuel αντιμετώπιζε δυσκολίες να βρει χρόνο πρόσβασης σε υπολογιστή. Εργαζόταν νύχτα, και χρησιμοποιούσε μηχανές που ήταν ακόμα υπό δοκιμή στο εργοστάσιο της IBM. Το Κεφάλαιο 6 καλύπτει το παίξιμο παιχνιδιών, και το Κεφάλαιο 21 περιγράφει και αναπτύσσει τις τεχνικές μάθησης που χρησιμοποιούσε ο Samuel.

Ο John McCarthy μετακόμισε από το Dartmouth στο MIT, και εκεί έκανε τρεις σημαντικές συνεισφορές μέσα σε μία ιστορική χρονιά, το 1958. Σε ένα υπόμνημα, το MIT AI Lab Memo No. 1, ο McCarthy όρισε τη γλώσσα υψηλού επιπέδου **Lisp**, η οποία έμελλε να γίνει η κυρίαρχη γλώσσα προγραμματισμού της τεχνητής νοημοσύνης. Η Lisp είναι χρονολογικά η δεύτερη σημαντική γλώσσα υψηλού επιπέδου που είναι ακόμα σε χρήση, ένα χρόνο νεότερη από τη FORTRAN. Με τη Lisp, ο McCarthy είχε το εργαλείο που χρειαζόταν, αλλά η πρόσβαση στους δυνάμεις και δαπανηρούς υπολογιστικούς πόρους ήταν επίσης σοβαρό πρόβλημα. Για την αντι-

μετώπιση αυτού του προβλήματος, ο ίδιος και άλλοι στο MIT επινόησαν το χρονομερισμό (time sharing). Πάλι το 1958, ο McCarthy έκανε μια δημοσίευση με τίτλο *Προγράμματα με κοινό νοῦ* (*Programs with Common Sense*), όπου περιέγραψε το Advice Taker, ένα υποθετικό πρόγραμμα που μπορεί να θεωρηθεί το πρώτο ολοκληρωμένο σύστημα TN. Όπως και το Logic Theorist και το Geometry Theorem Prover, το πρόγραμμα του McCarthy ήταν σχεδιασμένο να χρησιμοποιεί γνώση για να αναζητά λύσεις σε προβλήματα. Αντίθετα όμως από τα άλλα προγράμματα, το Advice Taker θα ενσωμάτωνε γενική γνώση του κόσμου. Για παράδειγμα, ο McCarthy έδειξε πώς μερικά απλά αξιώματα θα επέτρεπαν στο πρόγραμμα να παράγει ένα πλάνο μετάβασης στο αεροδρόμιο προκειμένου να προλάβει μια πτήση. Επίσης, το πρόγραμμα ήταν σχεδιασμένο έτσι ώστε να μπορεί να δέχεται νέα αξιώματα κατά την κανονική λειτουργία του, πράγμα που του επέτρεπε να αποκτά δεξιότητες σε νέους τομείς χωρίς να επαναπρογραμματίζεται. Το Advice Taker ενσωμάτωνε λοιπόν τις βασικές αρχές της αναπαράστασης γνώσης και της συλλογιστικής: ότι είναι χρήσιμο να έχουμε μια τυπικά διατυπωμένη και ρητή αναπαράσταση του κόσμου και του τρόπου που οι ενέργειες ενός πράκτορα επηρεάζουν τον κόσμο, και να μπορούμε να χειριζόμαστε αυτές τις αναπαραστάσεις με παραγωγικές διαδικασίες. Είναι αξιοσημείωτο πόσο μεγάλο μέρος αυτής της δημοσίευσης του 1958 παραμένει επίκαιρο ακόμα και σήμερα.

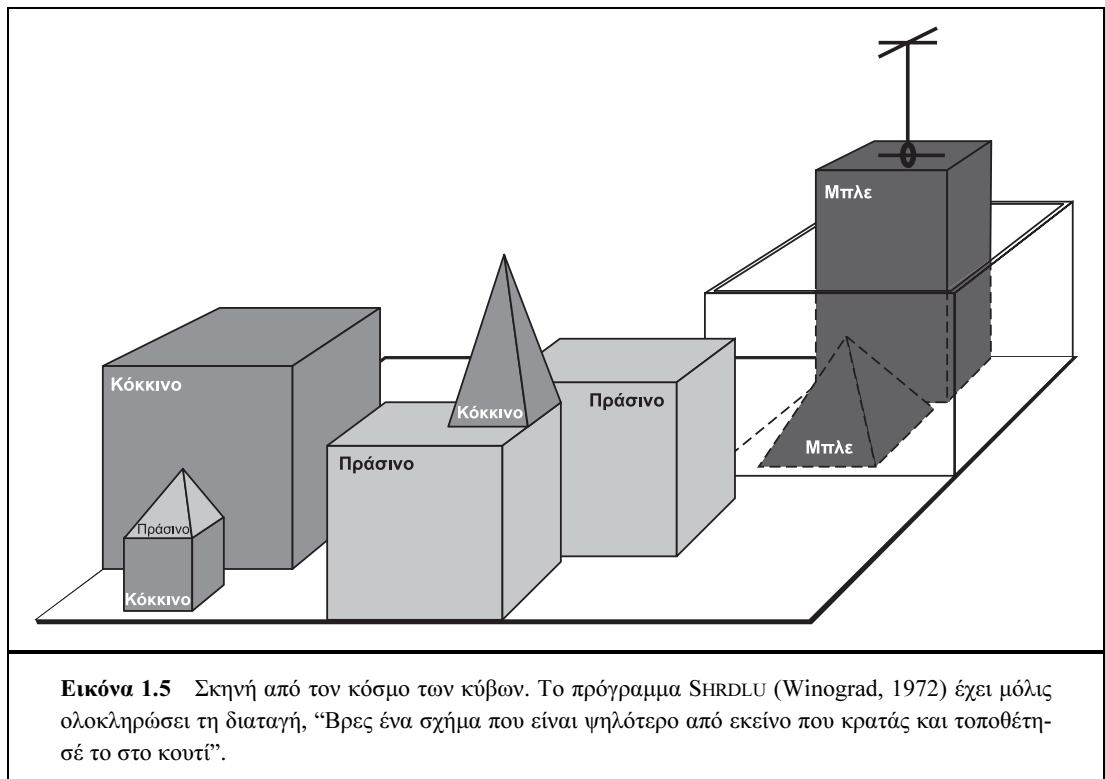
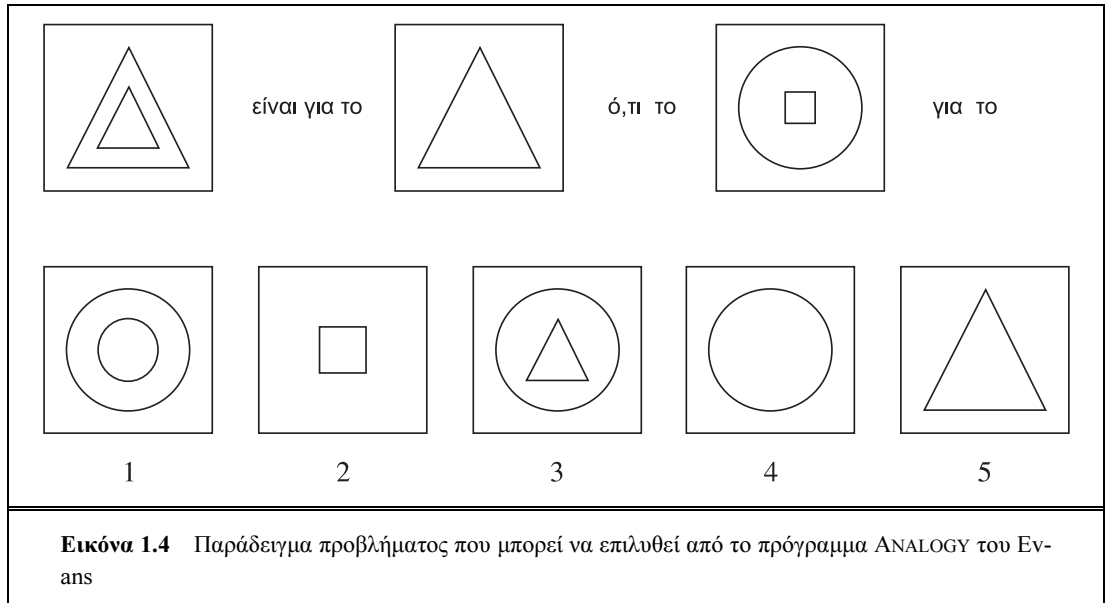
Το 1958 ήταν επίσης η χρονιά που ο Marvin Minsky μετακόμισε στο MIT. Όμως, η αρχική του συνεργασία με τον McCarthy δε διάρκεσε πολύ. Ο McCarthy έδινε έμφαση στην αναπαράσταση και τη συλλογιστική με την τυπική λογική, ενώ ο Minsky ενδιαφερόταν περισσότερο να δημιουργήσει προγράμματα που λειτουργούν και τελικά ανέπτυξε μια αντι-λογική άποψη. Το 1963, ο McCarthy ίδρυσε το εργαστήριο TN στο Stanford. Το σχέδιό του να χρησιμοποιήσει τη λογική για να κατασκευάσει το απόλυτο Advice Taker υποβοηθήθηκε από την ανακάλυψη της μεθόδου της ανάλυσης (resolution method — ένας πλήρης αλγόριθμος απόδειξης θεωρημάτων για τη λογική πρώτης τάξης· δείτε στο Κεφάλαιο 9) από τον J. A. Robinson. Στην εργασία στο Stanford, δινόταν έμφαση στις μεθόδους γενικής χρήσης για τη λογική συλλογιστική. Οι εφαρμογές λογικής περιλάμβαναν τα συστήματα απάντησης ερωτήσεων και σχεδιασμού του Cordell Green (Green, 1969) και το έργο του Shakey στη ρομποτική στο νέο Ινστιτούτο Έρευνας του Stanford (SRI). Αυτό το τελευταίο έργο, που εξετάζεται περισσότερο στο Κεφάλαιο 25, ήταν το πρώτο που επέδειξε την πλήρη ολοκλήρωση λογικής συλλογιστικής και φυσικής δράσης.

Ο Minsky επέβλεψε μια σειρά φοιτητών οι οποίοι επέλεξαν να ασχοληθούν με περιορισμένα προβλήματα που η επίλυσή τους φαινόταν να απαιτεί νοημοσύνη. Αυτά τα περιορισμένα πεδία ονομάστηκαν **μικρόκοσμοι** (microworlds). Τα πρόγραμμα SAINT του James Slagle (1963) μπορούσε να επιλύει προβλήματα του ολοκληρωτικού λογισμού κλειστής μορφής, συνήθη για τα πρωτοετή κολεγιακά μαθήματα. Το πρόγραμμα ANALOGY του Tom Evans (1968) επέλυε προβλήματα γεωμετρικών αναλογιών που εμφανίζονται στα τεστ ευφυΐας, όπως εκείνο της Εικόνας 1.4. Το πρόγραμμα STUDENT του Daniel Bobrow (1967) επέλυε αφηγηματικά αλγεβρικά προβλήματα όπως το παρακάτω:

Αν ο αριθμός των πελατών του Tom είναι διπλάσιος από το τετράγωνο του 20 τοις εκατό του αριθμού των διαφημίσεων που κάνει, και ο αριθμός των διαφημίσεων που κάνει είναι 45, πόσους πελάτες έχει ο Tom;

Ο πιο διάσημος μικρόκοσμος ήταν ο κόσμος των κύβων (blocks world), ο οποίος αποτελούνταν από ένα σύνολο στερεών γεωμετρικών σχημάτων τοποθετημένων πάνω σε ένα τραπέζι (ή συχνότερα, πάνω σε μια προσομοίωση τραπεζιού), όπως στην Εικόνα 1.5. Μια συνηθισμένη δουλειά σε αυτόν τον κόσμο είναι να αναδιαταχθούν τα σχήματα με έναν ορισμένο τρόπο, με τη χρήση ενός ρομποτικού βραχίονα που μπορεί να σηκώνει ένα κάθε φορά. Ο κόσμος των κύβων χρησιμοποιήθηκε στην εργασία για την όραση του David Huffman (1971), στην εργασία για την όραση και τη διάδοση περιορισμών του David Waltz (1975), στη θεωρία μάθησης του Patrick

Winston (1970), στο πρόγραμμα κατανόησης φυσικής γλώσσας του Terry Winograd (1972), και στο πρόγραμμα σχεδιασμού του Scott Fahlman (1974).



Οι πρώτες εργασίες που βασίστηκαν στα νευρωνικά δίκτυα των McCulloch και Pitts ευδοκίμησαν επίσης. Η εργασία των Winograd και Cowan (1963) έδειξε πώς ένας μεγάλος αριθμός στοιχείων μπορούσε να αναπαραστήσει συλλογικά μια μεμονωμένη έννοια, με αντίστοιχη ενίσχυση της ευρωστίας και την παραλληλία. Οι μέθοδοι μάθησης του Hebb ενισχύθηκαν από τον

Bernie Widrow (Widrow και Hoff, 1960· Widrow, 1962), ο οποίος ονόμασε τα δίκτυά του **adalines**, και από τον Frank Rosenblatt (1962) με τα **perceptron**. Ο Rosenblatt απέδειξε το **θεώρημα της σύγκλισης των perceptron**, δείχνοντας ότι οι αλγόριθμοι μάθησής του μπορούσαν να προσαρμόζουν την ισχύ των συνδέσεων ενός perceptron ώστε να συμφωνούν με οποιαδήποτε δεδομένα εισόδου, με την προϋπόθεση ότι μια τέτοια αντιστοιχία υπήρχε. Αυτά τα θέματα καλύπτονται στο Κεφάλαιο 20.

### Μια δόση ρεαλισμού (1966 – 1973)

Από την αρχή, οι ερευνητές της TN δε δίσταζαν να κάνουν προγνώσεις για την επερχόμενη επιτυχία τους. Οι παρακάτω δηλώσεις του Herbert Simon το 1957 αναφέρονται συχνά:

Ο σκοπός μου δεν είναι να σας καταπλήξω ή να σας σοκάρω — αλλά ο απλούστερο τρόπος να το θέσω είναι να πω ότι υπάρχουν σήμερα στον κόσμο μηχανές που σκέπτονται, μαθαίνουν, και δημιουργούν. Επίσης, η ικανότητά τους να κάνουν αυτά τα πράγματα πρόκειται να αυξηθεί γρήγορα μέχρι που — στο ορατό μέλλον — το φάσμα των προβλημάτων που θα μπορούν να χειρίζονται θα έχει την ίδια έκταση με το αντίστοιχο φάσμα της ανθρώπινης νόησης.

Όροι όπως το “ορατό μέλλον” μπορούν να ερμηνευτούν με διάφορους τρόπους, όμως ο Simon έκανε και μια πιο συγκεκριμένη πρόγνωση: ότι μέσα σε 10 χρόνια ένας υπολογιστής θα ήταν παγκόσμιος πρωταθλητής στο σκάκι και ότι ένα σημαντικό μαθηματικό θεώρημα θα αποδεικνυόταν από μια μηχανή. Οι προγνώσεις αυτές επαληθεύτηκαν (ή σχεδόν επαληθεύτηκαν) σε 40 χρόνια, και όχι σε 10. Η υπεραισιοδοξία του Simon οφειλόταν στις ελπιδοφόρες επιδόσεις των πρώιμων συστημάτων TN σε απλά παραδείγματα. Σχεδόν σε όλες τις περιπτώσεις, όμως, αυτά τα πρώιμα συστήματα κατέληξαν σε θλιβερή αποτυχία όταν δοκιμάστηκαν σε μια ευρύτερη επιλογή προβλημάτων ή σε δυσκολότερα προβλήματα.

Το πρώτο είδος δυσκολιών οφειλόταν στο γεγονός ότι τα περισσότερα από τα πρώιμα προγράμματα είχαν ελάχιστη ή καμία γνώση του θέματός τους· οι επιτυχίες τους οφείλονταν σε απλούς συντακτικούς χειρισμούς. Ένα αντιπροσωπευτικό περιστατικό συνέβη στις πρώτες προσπάθειες μηχανικής μετάφρασης, οι οποίες χρηματοδοτούνταν γενναία από το εθνικό συμβούλιο έρευνας των Η.Π.Α. με σκοπό να επιταχυνθεί η μετάφραση ρωσικών επιστημονικών δημοσιεύσεων ενόψει της εκτόξευσης του Sputnik το 1957. Αρχικά, θεωρήθηκε ότι ο απλός συντακτικός μετασχηματισμός με βάση την αγγλική και τη ρωσική γραμματική, και η αντικατάσταση των λέξεων με τη χρήση ενός ηλεκτρονικού λεξικού, θα αρκούσαν για να διατηρηθεί το ακριβές νόημα των προτάσεων. Το γεγονός είναι ότι η μετάφραση απαιτεί γενική γνώση του θέματος για να επιλυθούν οι αμφισημίες και να διευκρινιστεί το περιεχόμενο της πρότασης. Η διάσημη μετάφραση της φράσης “το πνεύμα είναι πρόθυμο αλλά η σάρκα ασθενής” σε “η βότκα είναι καλή αλλά το κρέας είναι σάπιο” δίνει μια εικόνα για τις δυσκολίες που αντιμετωπίζονται. Το 1966, μια έκθεση συμβουλευτικής επιτροπής έκανε τη διαπίστωση ότι “δεν υπάρχει μηχανική μετάφραση γενικού επιστημονικού κειμένου, ούτε προβλέπεται να υπάρξει άμεσα”. Όλη η αμερικανική κρατική χρηματοδότηση για ακαδημαϊκά έργα μετάφρασης ακυρώθηκε. Σήμερα, η μηχανική μετάφραση είναι ένα ατελές αλλά ευρύτατα χρησιμοποιούμενο εργαλείο για τεχνικά, εμπορικά, κρατικά και διαδικτυακά έγγραφα.

Το δεύτερο είδος δυσκολιών ήταν το δυσεπίλυτο πολλών από τα προβλήματα που επιχειρούσε να λύσει η TN. Τα περισσότερα από τα πρώιμα προγράμματα TN έλυναν προβλήματα δοκιμάζοντας διάφορους συνδυασμούς βημάτων μέχρι να βρεθεί η λύση. Η στρατηγική αυτή αρχικά είχε επιτυχία επειδή οι μικρόκοσμοι περιείχαν πολύ λίγα αντικείμενα και επομένως πολύ λίγες δυνατές ενέργειες και πολύ σύντομες ακολουθίες λύσης. Πριν αναπτυχθεί η θεωρία της υπολογιστικής πολυπλοκότητας, υπήρχε γενικά η αντίληψη η αλλαγή της κλίμακας μεγέθους σε μεγαλύτερα προβλήματα ήταν απλώς θέμα γρηγορότερου υλικού και περισσότερης μνήμης. Η αισιοδο-



ΜΗΧΑΝΙΚΗ ΕΞΕΛΙΞΗ

ξία που ακολούθησε την ανάπτυξη της απόδειξης θεωρημάτων με ανάλυση (resolution), για παράδειγμα, σύντομα μετριάστηκε όταν οι ερευνητές απέτυχαν να αποδείξουν θεωρήματα με περισσότερα από μερικές δεκάδες γεγονότα. *Το γεγονός ότι ένα πρόγραμμα μπορεί θεωρητικά να βρει μια λύση δε σημαίνει ότι το πρόγραμμα περιέχει κανέναν από τους μηχανισμούς που απαιτούνται για να βρει τη λύση πραγματικά.*

Η ψευδαίσθηση της απεριόριστης υπολογιστικής ισχύος δεν περιοριζόταν στα προγράμματα επίλυσης προβλημάτων. Τα πρώιμα πειράματα στην **μηχανική εξέλιξη** (machine evolution), ένα πεδίο που σήμερα ονομάζεται **γενετικοί αλγόριθμοι** (genetic algorithms — Friedberg, 1958· Friedberg κ.α., 1959) βασιζόταν στην αναμφισβήτητη σωστή πεποίθηση ότι, με την πραγματοποίηση μιας κατάλληλης σειράς μικρών μεταλλάξεων σε ένα πρόγραμμα κώδικα μηχανής, μπορούσε να δημιουργηθεί ένα πρόγραμμα με καλή απόδοση για οποιαδήποτε συγκεκριμένη απλή εργασία. Η ιδέα λοιπόν ήταν να δοκιμάζονται τυχαίες μεταλλάξεις με μια διαδικασία επιλογής που να διατηρεί εκείνες που φαίνονται χρήσιμες. Μετά από χιλιάδες ώρες χρόνου CPU δεν παρατηρήθηκε καμία πρόοδος. Οι σημερινοί γενετικοί αλγόριθμοι χρησιμοποιούν καλύτερες αναπαραστάσεις και έχουν δείξει μεγαλύτερη επιτυχία.

Η αποτυχία να τα βγάλει πέρα με το πρόβλημα της “συνδυαστικής έκρηξης” ήταν μία από τις κυριότερες κριτικές της TN που αναφέρονταν στην έκθεση του Lighthill (1973), η οποία αποτέλεσε τη βάση για την απόφαση της Βρετανικής κυβέρνησης να τερματίσει την υποστήριξη της έρευνας στην TN σε όλα τα πανεπιστήμια εκτός από δύο. (Η προφορική παράδοση μας δίνει μια κάπως διαφορετική και πιο ζωντανή εικόνα, με πολιτικές φιλοδοξίες και προσωπικές αντιπάθειες που η περιγραφή τους ξεφεύγει από το θέμα μας.)

Μια τρίτη δυσκολία οφειλόταν σε κάποιους θεμελιώδεις περιορισμούς των βασικών δομών που χρησιμοποιούνταν για την παραγωγή ευφυούς συμπεριφοράς. Για παράδειγμα, όπως αποδείχτηκε στο βιβλίο των Minsky και Papert *Perceptrons* (1969), αν και τα perceptron (μια απλή μορφή νευρωνικών δικτύων) μπορούσαν να μαθαίνουν οτιδήποτε μπορούσαν να αναπαραστήσουν, αυτά που μπορούσαν να αναπαραστήσουν ήταν ελάχιστα. Ειδικότερα, ένα perceptron δύο εισόδων δεν μπορούσε να εκπαιδευτεί να αναγνωρίζει πότε οι δύο είσοδοί του ήταν διαφορετικές. Αν και τα αποτελέσματά τους δεν είχαν εφαρμογή σε πιο πολύπλοκα, πολυεπίπεδα δίκτυα, η χρηματοδότηση της έρευνας στα νευρωνικά δίκτυα σύντομα μειώθηκε σχεδόν στο μηδέν. Η ειρωνεία είναι ότι οι νέοι αλγόριθμοι μάθησης με οπισθοδιάδοση (back-propagation learning) για τα πολυεπίπεδα δίκτυα, οι οποίοι θα προκαλούσαν τεράστια αναζωπύρωση της έρευνας στα νευρωνικά δίκτυα στα τέλη της δεκαετίας του 1980, στην πραγματικότητα είχαν επινοηθεί ήδη το 1969 (Bryson και Ho, 1969).

### Συστήματα βασισμένα στη γνώση: Το κλειδί της επιτυχίας; (1969 – 1979)

Η εικόνα της επίλυσης προβλημάτων που είχε προκύψει την πρώτη δεκαετία της έρευνας στην TN ήταν η εικόνα ενός μηχανισμού αναζήτησης γενικής χρήσης που προσπαθούσε να συνδυάσει στοιχειώδη συλλογιστικά βήματα για να βρει πλήρεις λύσεις. Οι προσεγγίσεις αυτές ονομάστηκαν **ασθενείς μέθοδοι** (weak methods) επειδή, αν και γενικές, δεν μπορούν να κλιμακωθούν σε μεγάλα ή δύσκολα στιγμιότυπα προβλημάτων. Η εναλλακτική λύση, αντί για τη χρήση ασθενών μεθόδων, είναι να χρησιμοποιηθεί η πιο ισχυρή μέθοδος της ειδικής για το συγκεκριμένο πεδίο γνώσης, η οποία επιτρέπει μεγαλύτερα συλλογιστικά βήματα και μπορεί να χειρίζεται ευκολότερα αντιπροσωπευτικές περιπτώσεις που παρουσιάζονται σε τομείς ειδικής εμπειρίας. Θα μπορούσε να πει κανείς ότι για να λυθεί ένα δύσκολο πρόβλημα θα πρέπει σχεδόν να γνωρίζουμε ήδη την απάντηση.

Το πρόγραμμα DENDRAL (Buchanan κ.α., 1969) ήταν ένα πρώιμο παράδειγμα αυτής της προσέγγισης. Αναπτύχθηκε στο Stanford, όπου ο Ed Feigenbaum (ένας πρώην φοιτητής του Herbert Simon), ο Bruce Buchanan (ένας φιλόσοφος που έγινε επιστήμονας των υπολογιστών) και ο

ΑΣΘΕΝΕΙΣ ΜΕΘΟΔΟΙ

Joshua Lederberg (ένας επιστήμονας της γενετικής βραβευμένος με Νόμπελ) συνεργάστηκαν για να λύσουν το πρόβλημα της συναγωγής της μοριακής δομής από τις πληροφορίες που παρέχονται από ένα φασματογράφο μάζας. Η είσοδος του προγράμματος αποτελείται από το βασικό χημικό τύπο του μορίου (π.χ.,  $C_6H_{13}NO_2$ ) και από το φάσμα μάζας το οποίο δίνει τις μάζες των διαφόρων θραυσμάτων του μορίου που παράγονται όταν βομβαρδίζεται με μια δέσμη ηλεκτρονίων. Για παράδειγμα, το φάσμα μάζας μπορεί να περιέχει μια κορυφή στο  $m=15$ , η οποία αντιστοιχεί στη μάζα ενός θραύσματος μεθυλίου ( $CH_3$ ).

Η απλοϊκή έκδοση του προγράμματος παρήγαγε όλες τις δυνατές δομές που ήταν συνεπείς με το χημικό τύπο, και έπειτα πρόβλεπε τι φάσμα μάζας θα παρατηρούνταν για την κάθε μία και το σύγκρινε με το πραγματικό φάσμα. Όπως ήταν αναμενόμενο, η εφαρμογή αυτής της μεθόδου ήταν δυσεπίλυτο πρόβλημα για τα αρκετά μεγάλα μόρια. Οι ερευνητές του DENDRAL συμβουλευτήκαν ειδικούς της αναλυτικής χημείας και διαπίστωσαν ότι αυτοί δούλευαν αναζητώντας στο φάσμα γνωστά μοτίβα κορυφών που ήταν ενδείξεις κοινών επιμέρους δομών στο μόριο. Για παράδειγμα, ο παρακάτω κανόνας χρησιμοποιείται για την αναγνώριση μιας υποομάδας κετόνης ( $C=O$ ), η οποία έχει μάζα 28:

εάν υπάρχουν δύο κορυφές στα  $x_1$  και  $x_2$  τέτοιες ώστε

(α)  $x_1 + x_2 = M + 28$  ( $M$  είναι η μάζα ολόκληρου του μορίου)

(β)  $x_1 - 28$  είναι άνω κορυφή

(γ)  $x_2 - 28$  είναι άνω κορυφή

(δ) Τουλάχιστον ένα από τα  $x_1$  και  $x_2$  δύο είναι άνω.

τότε υπάρχει υποομάδα κετόνης

Η αναγνώριση του ότι το μόριο περιέχει μια συγκεκριμένη επιμέρους δομή μειώνει πάρα πολύ τον αριθμό των δυνατών υποψηφίων. Το DENDRAL ήταν ισχυρό επειδή:

Όλη η σχετική θεωρητική γνώση για την επίλυση αυτών των προβλημάτων έχει αντιστοιχιστεί από τη γενική μορφή του [στοιχείου πρόγνωσης φάσματος] (“βασικές αρχές”) σε αποδοτικές ειδικές μορφές (“cookbook recipes” — συνταγές). (Feigenbaum κ.α., 1971)

Η σπουδαιότητα του DENDRAL ήταν ότι αποτέλεσε το πρώτο επιτυχημένο σύστημα με έμφαση στη γνώση: η πείρα του προερχόταν από πάρα πολλούς ειδικούς κανόνες. Τα επόμενα συστήματα περιλάμβαναν επίσης το κύριο θέμα της προσέγγισης του Advice Taker του McCarthy — το σαφή διαχωρισμό της γνώσης (σε μορφή κανόνων) από το συλλογιστικό στοιχείο.

Παίρνοντας υπόψη αυτό το μάθημα, ο Feigenbaum και άλλοι στο Stanford ξεκίνησαν το “έργο ευρητικού προγραμματισμού” (Heuristic Programming Project, HPP), για να ερευνήσουν την έκταση στην οποία μπορούσε να εφαρμοστεί η νέα μεθοδολογία των **έμπειρων συστημάτων** (expert systems) σε άλλα πεδία της ανθρώπινης εμπειρίας. Η επόμενη σημαντική προσπάθεια έγινε στο χώρο της ιατρικής διάγνωσης. Οι Feigenbaum, Buchanan και Dr. Edward Shortliffe ανέπτυξαν το MYCIN με σκοπό τη διάγνωση μολύνσεων του αίματος. Με γύρω στους 450 κανόνες, το MYCIN μπορούσε να τα καταφέρει εξίσου καλά με μερικούς ειδικούς, και πολύ καλύτερα από τους ασκούμενους γιατρούς. Είχε επίσης δύο μεγάλες διαφορές από το DENDRAL. Πρώτον, σε αντίθεση με τους κανόνες του DENDRAL, δεν υπήρχε κανένα γενικό θεωρητικό μοντέλο από το οποίο μπορούσαν να προκύψουν οι κανόνες του MYCIN. Οι κανόνες αναγκαστικά προέρχονταν από εκτεταμένες συνεντεύξεις με ειδικούς, οι οποίοι με τη σειρά τους είχαν αποκτήσει από βιβλία, από άλλους ειδικούς, και από την άμεση εμπειρία ιατρικών περιστατικών. Δεύτερον, οι κανόνες έπρεπε να αντανακλούν την αβεβαιότητα που χαρακτηρίζει την ιατρική γνώση. Το MYCIN περιλάμβανε ένα λογισμό της αβεβαιότητας που ονομαζόταν **συντελεστές βεβαιότητας** (certainty factors — δείτε στο Κεφάλαιο 13), ο οποίος φαινόταν (εκείνο τον καιρό) να συμφωνεί ικανοποιητικά με τον τρόπο που αξιολογούσαν οι γιατροί τη σημασία των τεκμηρίων στη διάγνωση.



Η σπουδαιότητα της γνώσης του πεδίου έγινε επίσης φανερή στο χώρο της κατανόησης της φυσικής γλώσσας. Αν και το σύστημα SHRDLU του Winograd για την κατανόηση φυσικής γλώσσας προξένησε πολλή έξαψη, η εξάρτησή του από τη συντακτική ανάλυση προκάλεσε μερικά από τα ίδια προβλήματα που παρατηρήθηκαν και στις προηγούμενες εργασίες πάνω στη μηχανική μετάφραση. Μπορούσε να ξεπερνά την αμφισημία και να καταλαβαίνει σε τι αναφέρονταν οι αντωνυμίες, αλλά αυτό συνέβαινε κυρίως επειδή ήταν σχεδιασμένο ειδικά για ένα χώρο — τον κόσμο των κύβων. Πολλοί ερευνητές, μεταξύ των οποίων ο Eugene Charniak, ένας μεταπτυχιακός φοιτητής του Winograd στο MIT, υποστήριξε ότι η σωστή κατανόηση της γλώσσας θα απαιτούσε γενική γνώση του κόσμου και μια γενική μέθοδο για τη χρήση αυτής της γνώσης.

Στο Yale, ο Roger Schank, ένας γλωσσολόγος που έγινε ερευνητής της TN, τόνισε αυτό το σημείο υποστηρίζοντας ότι “Η σύνταξη είναι κάτι το ανύπαρκτο”, το οποίο ενόχλησε πολλούς γλωσσολόγους αλλά συντέλεσε στην έναρξη μιας χρήσιμης συζήτησης. Ο Schank και οι φοιτητές του κατασκεύασαν μια σειρά προγραμμάτων (Schank και Abelson, 1977· Wilensky, 1978· Schank και Riesbeck, 1981· Dyer, 1983) που όλα είχαν σκοπό την κατανόηση φυσικής γλώσσας. Δινόταν όμως λιγότερη έμφαση στην ίδια τη γλώσσα, και περισσότερη στα προβλήματα της αναπαράστασης και της συλλογιστικής με την απαιτούμενη γνώση για την κατανόηση της γλώσσας. Τέτοια προβλήματα ήταν η αναπαράσταση στερεότυπων καταστάσεων (Cullingford, 1981), η περιγραφή της οργάνωσης της ανθρώπινης μνήμης (Rieger, 1976· Kolodner, 1983) και η κατανόηση των πλάνων και των στόχων (Wilensky, 1983).

Η ευρεία διάδοση της ανάπτυξης εφαρμογών για προβλήματα του πραγματικού κόσμου προκάλεσε ταυτόχρονη αύξηση των απαιτήσεων για πρακτικά σχήματα αναπαράστασης γνώσης. Αναπτύχθηκε μια πληθώρα διαφορετικών αναπαραστάσεων και γλωσσών συλλογιστικής. Μερικές βασίζονταν στη λογική — για παράδειγμα, στην Ευρώπη έγινε δημοφιλής η γλώσσα Prolog και στις Ηνωμένες Πολιτείες η οικογένεια PLANNER. Άλλοι, ακολουθώντας την ιδέα των **πλαίσια** (frames) του Minsky (1975), υιοθέτησαν μια πιο δομημένη προσέγγιση, συναρμολογώντας γεγονότα για συγκεκριμένους τύπους αντικειμένων και συμβάντων και διατάσσοντας αυτούς τους τύπους σε μια μεγάλη ταξονομική ιεραρχία ανάλογη με τη βιολογική ταξονομία.

## Η τεχνητή νοημοσύνη γίνεται βιομηχανία (1980 – σήμερα)

Το πρώτο επιτυχημένο έμπειρο σύστημα που εμφανίστηκε στην αγορά, το R1, άρχισε να λειτουργεί στην Digital Equipment Corporation (McDermott, 1982). Το πρόγραμμα βοηθούσε στη διαμόρφωση παραγγελιών για νέα υπολογιστικά συστήματα· εκτιμάται ότι το 1986 εξοικονομούσε για την εταιρία 40 εκατομμύρια δολάρια το χρόνο. Το 1988 το τμήμα TN της DEC είχε εγκαταστημένα 40 έμπειρα συστήματα και ακόμα περισσότερα στο στάδιο της ανάπτυξης. Η Du Pont είχε 100 σε χρήση και 500 υπό ανάπτυξη, εξοικονομώντας γύρω στα 10 εκατομμύρια δολάρια το χρόνο. Σχεδόν κάθε μεγάλη εταιρία στις Η.Π.Α. είχε δικό της τμήμα TN και είτε χρησιμοποιούσε έμπειρα συστήματα είτε έκανε σχετική έρευνα.

Το 1981 οι Ιάπωνες ανάγγειλαν το έργο “Fifth Generation” (Πέμπτη Γενιά), ένα 10-ετές σχέδιο για την κατασκευή ευφών υπολογιστών που θα χρησιμοποιούσαν τη γλώσσα Prolog. Η απάντηση των Η.Π.Α. ήταν να συστήσουν την εταιρία MCC (Microelectronics and Computer Technology Corporation) με τη μορφή ερευνητικής κοινοπραξίας που είχε σκοπό να εξασφαλίσει την εθνική ανταγωνιστικότητα. Και στις δύο περιπτώσεις, η TN ήταν μέρος μιας ευρύτερης προσπάθειας που περιλάμβανε τη σχεδίαση ολοκληρωμένων κυκλωμάτων και την έρευνα στις ανθρώπινες διασυνδέσεις. Όμως, τα τμήματα TN τόσο της MCC όσο και της Fifth Generation ποτέ δεν κατάφεραν να επιτύχουν τους φιλόδοξους στόχους τους. Στη Βρετανία, η έκθεση Alvey αποκατέστησε τη χρηματοδότηση που είχε διακοπεί από την έκθεση Lighthill.<sup>15</sup>

<sup>15</sup> Για να ξεπεράσουν την αμηχανία, επινόησαν ένα νέο πεδίο που ονομάστηκε IKBS (Intelligent Knowledge-Based Systems — ευφυή συστήματα βασισμένα στη γνώση), επειδή το πεδίο της Τεχνητής Νοημοσύνης είχε επίσημα καταργηθεί.

Συνολικά, η βιομηχανία της ΤΝ αναπτύχθηκε εκρηκτικά, και έφτασε από μερικά εκατομμύρια δολάρια το 1980 σε δισεκατομμύρια δολάρια το 1988. Σύντομα μετά από αυτό, ακολούθησε μια περίοδος που ονομάστηκε “χειμώνας της ΤΝ”, κατά την οποία πολλές εταιρίες υπέφεραν, καθώς δεν κατάφεραν να ανταποκριθούν στις υπερβολικές υποσχέσεις τους.

### Η επιστροφή των νευρωνικών δικτύων (1986 – σήμερα)

Αν και η επιστήμη των υπολογιστών είχε σε μεγάλο βαθμό εγκαταλείψει το πεδίο των νευρωνικών δικτύων από τα τέλη της δεκαετίας του 1970, η δουλειά συνεχίστηκε σε άλλα πεδία. Φυσικοί όπως ο John Hopfield (1982) χρησιμοποίησαν τεχνικές της στατιστικής μηχανικής για να αναλύσουν τις ιδιότητες αποθήκευσης και βελτιστοποίησης των δικτύων, αντιμετωπίζοντας τις συλλογές κόμβων ως συλλογές ατόμων. Ψυχολόγοι, όπως ο David Rumelhart και ο Geoff Hinton, συνέχισαν τη μελέτη των μοντέλων νευρωνικών δικτύων για τη μνήμη. Όπως θα δούμε στο Κεφάλαιο 20, η πραγματική ώθηση ήρθε στα μέσα της δεκαετίας του 1980, όταν τουλάχιστον τέσσερις διαφορετικές ομάδες επανεφεύραν τον αλγόριθμο μάθησης με οπισθοδιάδοση που είχε επινοηθεί για πρώτη φορά το 1969 από τους Bryson και Ho. Ο αλγόριθμος εφαρμόστηκε σε πολλά προβλήματα μάθησης της επιστήμης των υπολογιστών και της ψυχολογίας, και η ευρεία διάδοση των αποτελεσμάτων στη συλλογή *Parallel Distributed Processing* (Rumelhart και McClelland, 1986) προκάλεσε πολύ μεγάλο ενδιαφέρον.

Τα λεγόμενα **συνδεδεστικά** (connectionist) μοντέλα ευφών συστημάτων αντιμετωπίζονταν από μερικούς ως άμεσοι ανταγωνιστές τόσο των συμβολικών μοντέλων που προωθούνταν από τους Newell και Simon όσο και της λογικιστικής προσέγγισης του McCarthy και άλλων (Smolensky, 1988). Φαίνεται ίσως προφανές ότι σε κάποιο επίπεδο οι άνθρωποι χειρίζονται σύμβολα — το βιβλίο του Terrence Deacon *The Symbolic Species* (1997) υποστηρίζει μάλιστα ότι αυτό είναι το *καθοριστικό χαρακτηριστικό* των ανθρώπων, όμως οι πιο ένθερμοι συνδεδεμιστές αμφισβήτησαν το αν ο χειρισμός συμβόλων είχε πραγματικά οποιονδήποτε ρόλο στην εξήγηση των λεπτομερών μοντέλων της γνωστικής λειτουργίας. Η ερώτηση αυτή παραμένει αναπάντητη, αλλά η τρέχουσα άποψη είναι ότι η συνδεδεστική και η συμβολική προσέγγιση είναι συμπληρωματικές και όχι ανταγωνιστικές.

### Η Τεχνητή Νοημοσύνη γίνεται επιστήμη (1987 – σήμερα)

Τα τελευταία χρόνια γνωρίσαμε μια επανάσταση τόσο στο περιεχόμενο όσο και στη μεθοδολογία της δουλειάς στην τεχνητή νοημοσύνη.<sup>16</sup> Σήμερα είναι πιο συνηθισμένο οι ερευνητές να οικοδομούν πάνω σε υπάρχουσες θεωρίες και όχι να προτείνουν νέες, να θεμελιώνουν τους ισχυρισμούς τους σε αυστηρά θεωρήματα ή σε απτά πειραματικά πειστήρια και όχι στη διαίσθηση, και να δίνουν βαρύτητα σε πραγματικές εφαρμογές και όχι σε παραδείγματα-παιχνίδια.

Το πεδίο της ΤΝ γεννήθηκε εν μέρει ως αντίδραση στους περιορισμούς των υπαρχόντων πεδίων, όπως η θεωρία ελέγχου και η στατιστική, αλλά σήμερα ενστερνίζεται αυτά τα πεδία. Όπως το έθεσε ο David McAllester (1998),

<sup>16</sup> Μερικοί χαρακτήρισαν αυτή την εξέλιξη ως νίκη των **μεθοδικών** (neats) — εκείνων που πιστεύουν ότι οι θεωρίες για την ΤΝ θα πρέπει να βασίζονται στη μαθηματική αυστηρότητα — έναντι των **τσαπατσούληδων** (scruffies) — εκείνων που προτιμούσαν να δοκιμάζουν πολλές ιδέες, να γράφουν διάφορα προγράμματα, και και μετά να αποτιμούν τι φαίνεται να λειτουργεί. Και οι δύο προσεγγίσεις είναι σημαντικές. Μια μετατόπιση προς την “τακτικότητα” υποδηλώνει ότι το πεδίο έχει φτάσει σε ένα επίπεδο σταθερότητας και ωριμότητας. Το αν αυτή η σταθερότητα θα διαταραχθεί από μια νέα “άτακτη” ιδέα είναι άλλο ζήτημα.

Την πρόιμη εποχή της TN φαινόταν λογικό ότι νέες μορφές συμβολικού υπολογισμού, π.χ., τα πλαίσια και τα σημασιολογικά δίκτυα, θα έκαναν μεγάλο μέρος της κλασικής θεωρίας ξεπερασμένο. Αυτό οδήγησε σε μια μορφή απομονωτισμού κατά την οποία η TN διαχωρίστηκε σε μεγάλο βαθμό από την υπόλοιπη επιστήμη των υπολογιστών. Σήμερα αυτό ο απομονωτισμός εγκαταλείπεται. Αναγνωρίζεται ότι η μηχανική μάθηση δε θα πρέπει να είναι απομονωμένη από τη θεωρία πληροφοριών, ότι η αβέβαιη συλλογιστική δε θα πρέπει να είναι απομονωμένη από τη στοχαστική μοντελοποίηση, ότι η αναζήτηση δε θα πρέπει να είναι απομονωμένη από την κλασική βελτιστοποίηση και έλεγχο, και ότι η αυτοματοποιημένη συλλογιστική δε θα πρέπει να είναι απομονωμένη από τις τυπικές μεθόδους και τη στατική ανάλυση.

Από την άποψη της μεθοδολογίας, η TN έχει τελικά υπαχθεί σαφώς στην επιστημονική μέθοδο. Για να γίνονται δεκτές οι υποθέσεις, πρέπει να υποβάλλονται σε αυστηρά εμπειρικά πειράματα και τα αποτελέσματα να αναλύονται στατιστικά για το βαθμό σημαντικότητάς τους (Cohen, 1995). Με τη χρήση του Διαδικτύου και των μεριζόμενων χώρων αποθήκευσης πειραματικών δεδομένων και κώδικα, είναι σήμερα δυνατό τα πειράματα να αναπαράγονται.

Ένα τέτοιο παράδειγμα είναι το πεδίο της αναγνώρισης ομιλίας (speech recognition). Τη δεκαετία του 1970 δοκιμάστηκε ένα ευρύ φάσμα διαφορετικών αρχιτεκτονικών και προσεγγίσεων. Πολλές από αυτές ήταν μάλλον περιστασιακές και εύθραυστες και εφαρμόστηκαν μόνο σε μερικά ειδικά επιλεγμένα παραδείγματα. Τα τελευταία χρόνια έχουν κυριαρχήσει στο χώρο προσεγγίσεις που βασίζονται σε **κρυφά μοντέλα Markov** (hidden Markov models, HMM). Δύο απόψεις των κρυφών μοντέλων Markov είναι εκείνες που μας ενδιαφέρουν. Πρώτον, βασίζονται μια αυστηρή μαθηματική θεωρία. Αυτό έδωσε στους ερευνητές της ομιλίας τη δυνατότητα να οικοδομούν πάνω σε πολλές δεκαετίες μαθηματικών αποτελεσμάτων που προέρχονται από άλλα πεδία. Δεύτερον, τα μοντέλα αυτά παράγονται με μια διαδικασία εκπαίδευσης που βασίζεται σε μια μεγάλη συλλογή πραγματικών δεδομένων ομιλίας. Αυτό εξασφαλίζει ότι η απόδοση είναι ανθεκτική και σε αυστηρές τυφλές δοκιμές τα κρυφά μοντέλα Markov βελτιώνουν σταθερά τις επιδόσεις τους. Η τεχνολογία ομιλίας και το σχετικό πεδίο της αναγνώρισης χειρόγραφων χαρακτήρων ήδη περνούν σε διαδεδομένες βιομηχανικές και καταναλωτικές εφαρμογές.

Τα νευρωνικά δίκτυα επίσης υπόκεινται σε αυτή την τάση. Μεγάλο μέρος της δουλειάς στα νευρωνικά δίκτυα τη δεκαετία του 1980 έγινε στα πλαίσια μιας προσπάθειας να ξεκαθαριστεί τι ήταν δυνατό να γίνει και να μελετηθεί σε τι διαφέρουν τα νευρωνικά δίκτυα από τις “παραδοσιακές” τεχνικές. Με τη χρήση της βελτιωμένης μεθοδολογίας και των θεωρητικών πλαισίων, το πεδίο έφτασε σε μια αντίληψη όπου τα νευρωνικά δίκτυα μπορούν να συγκρίνονται με αντίστοιχες τεχνικές από τη στατιστική, την αναγνώριση προτύπων και τη μηχανική μάθηση και να χρησιμοποιείται η τεχνική που υπόσχεται περισσότερο στην κάθε εφαρμογή. Ως αποτέλεσμα αυτών των εξελίξεων, η λεγόμενη τεχνολογία **εξόρυξης δεδομένων** (data mining) οδήγησε στη γέννηση μιας νέας ανθηρής βιομηχανίας.

Το βιβλίο του Judea Pearl *Probabilistic Reasoning in Intelligent Systems (Πιθανοτική συλλογιστική στα ευφυή συστήματα* — 1988) οδήγησε σε αναβίωση της αποδοχής της θεωρίας πιθανοτήτων και της θεωρίας αποφάσεων στην TN, μετά από μια αναζωπύρωση του ενδιαφέροντος που συνοψιζόταν στο άρθρο του Peter Cheeseman “In Defense of Probability” (1985). Ο φορμαλισμός του **δικτύου Bayes** (Bayesian network) επινοήθηκε με σκοπό την αποδοτική αναπαράσταση της αβέβαιης γνώσης και την αυστηρή συλλογιστική με αυτή. Αυτή η προσέγγιση ξεπερνά σε μεγάλο βαθμό πολλά προβλήματα των συστημάτων πιθανοτικής συλλογιστικής των δεκαετιών του 1960 και 1970· σήμερα, κυριαρχεί στην έρευνα της TN πάνω στην αβέβαιη συλλογιστική και τα έμπειρα συστήματα. Η προσέγγιση αυτή μπορεί να καλύψει τη μάθηση από την εμπειρία και συνδυάζει τα καλύτερα στοιχεία της κλασικής TN και των νευρωνικών δικτύων. Η δουλειά του Judea Pearl (1982) και των Eric Horvitz και David Heckerman (Horvitz και Heckerman, 1986· Horvitz κ.α., 1986) προώθησε την ιδέα των *κανονιστικών* (normative) έμπειρων συστημάτων.

των: εκείνων που ενεργούν ορθολογικά σύμφωνα με τους νόμους της θεωρίας αποφάσεων και δεν προσπαθούν να μιμηθούν τα βήματα της σκέψης των ανθρώπων-ειδικών. Το λειτουργικό σύστημα Windows<sup>TM</sup> περιλαμβάνει πολλά κανονιστικά διαγνωστικά έμπειρα συστήματα για την επανόρθωση προβλημάτων. Αυτό το θέμα καλύπτεται στα Κεφάλαια 13 έως 16.

Παρόμοιες ειρηνικές επαναστάσεις έγιναν στη ρομποτική, την υπολογιστική όραση και την αναπαράσταση γνώσης. Η καλύτερη κατανόηση των προβλημάτων και των ιδιοτήτων της πολυπλοκότητάς τους, σε συνδυασμό με την εξέλιξη των μαθηματικών μέσων, οδήγησε σε εφαρμόσιμα ερευνητικά προγράμματα και εύρωστες μεθόδους. Σε πολλές περιπτώσεις, η τυποποίηση και η εξειδίκευση οδήγησαν επίσης σε κατακερματισμό· θέματα όπως η όραση και η ρομποτική απομονώνονται όλο και περισσότερο από τον κύριο κορμό της δουλειάς στην ΤΝ. Η ενοποιητική άποψη που βλέπει την ΤΝ ως σχεδίαση ορθολογικών πρακτόρων μπορεί να επαναφέρει την ενότητα σε αυτά τα ανόμοια πεδία.

### Η εμφάνιση των ευφύων πρακτόρων (1995 – σήμερα)

Ενθαρρυμένοι ίσως από την πρόοδο στην επίλυση των επιμέρους προβλημάτων της ΤΝ, οι ερευνητές άρχισαν επίσης να επανεξετάζουν το πρόβλημα του “ολοκληρωμένου πράκτορα” (whole agent). Η εργασία των Allen Newell, John Laird και Paul Rosenbloom στο SOAR (Newell, 1990· Laird κ.α., 1987) είναι το πιο γνωστό παράδειγμα αρχιτεκτονικής ολοκληρωμένων πρακτόρων. Το λεγόμενο κίνημα των “εγκαταστημένων” (situated) έχει σκοπό την κατανόηση της λειτουργίας πρακτόρων που είναι ενσωματωμένοι σε πραγματικά περιβάλλοντα με συνεχείς αισθητηριακές εισόδους. Ένα από τα σημαντικότερα περιβάλλοντα για ευφυείς πράκτορες είναι το Διαδίκτυο. Τα συστήματα ΤΝ έχουν γίνει τόσο κοινά στις εφαρμογές του Ιστού ώστε η κατάληξη “-bot” έχει περάσει στην καθημερινή γλώσσα. Τεχνολογίες ΤΝ βρίσκονται επίσης κάτω από πολλά εργαλεία του Διαδικτύου, όπως οι μηχανές αναζήτησης, τα συστήματα παροχής συμβουλών και τα συστήματα κατασκευής ιστοσελίδων.

Εκτός από την πρώτη έκδοση αυτού του βιβλίου (Russell και Norvig, 1995), άλλα πρόσφατα βιβλία έχουν επίσης υιοθετήσει την άποψη των πρακτόρων (Poole κ.α., 1998· Nilsson, 1998). Μία συνέπεια της προσπάθειας κατασκευής ολοκληρωμένων πρακτόρων είναι η διαπίστωση ότι τα επιμέρους πεδία της ΤΝ που ήταν προηγουμένως απομονωμένα μπορεί να χρειαστεί τώρα να επανοργανωθούν κάπως, προκειμένου τα αποτελέσματά τους να μπορούν να συνδυαστούν όλα μαζί. Ειδικότερα, είναι σήμερα γενικά παραδεκτό ότι τα αισθητηριακά συστήματα (όρασης, σόναρ, αναγνώρισης ομιλίας, κ.λπ.) δεν μπορούν να δίνουν εντελώς αξιόπιστες πληροφορίες για το περιβάλλον. Γι' αυτό, τα συστήματα συλλογιστικής και σχεδιασμού πρέπει να είναι σε θέση να αντιμετωπίζουν την αβεβαιότητα. Μια δεύτερη σημαντική επίπτωση της άποψης των πρακτόρων είναι ότι η ΤΝ έχει έρθει σε πολύ στενότερη επαφή με άλλα πεδία, όπως η θεωρία ελέγχου και τα οικονομικά, που ασχολούνται επίσης με πράκτορες.

## 1.4 ΣΥΓΧΡΟΝΗ ΤΕΧΝΟΛΟΓΙΑ

Τι μπορεί να κάνει η ΤΝ σήμερα; Είναι δύσκολο να δοθεί μια σύντομη απάντηση, επειδή υπάρχουν τόσο πολλές δραστηριότητες σε τόσο πολλά επιμέρους πεδία. Θα παραθέσουμε μερικά παραδείγματα εφαρμογών· άλλες εφαρμογές παρουσιάζονται σε ολόκληρο το βιβλίο.

**Αυτόνομος σχεδιασμός και χρονοπρογραμματισμός:** Εκατό εκατομμύρια μίλια από τη Γη, το πρόγραμμα Remote Agent της NASA έγινε το πρώτο πρόγραμμα αυτόνομου σχεδιασμού σε τροχιά, με σκοπό να ελέγχει το χρονοπρογραμματισμό των λειτουργιών ενός διαστημικού σκάφους (Jonsson κ.α., 2000). Το Remote Agent δημιουργούσε πλάνα για την επίτευξη στόχων υψηλού επιπέδου καθορισμένων από το έδαφος και παρακολουθούσε τη λειτουργία του σκάφους

καθώς τα πλάνα εκτελούνταν — ανιχνεύοντας προβλήματα καθώς παρουσιάζονταν, κάνοντας διαγνώσεις και αποκαθιστώντας τη σωστή λειτουργία.

**Παιχνίδια:** Το Deep Blue της IBM έγινε το πρώτο πρόγραμμα υπολογιστή που νίκησε τον παγκόσμιο πρωταθλητή σε σκακιστικό αγώνα, όταν επικράτησε του Garry Kasparov με σκορ 3,5–2,5 σε μια εξαντλητική μονομαχία (Goodman και Keene, 1997). Ο Kasparov δήλωσε ότι ένοιωσε ένα “νέο είδος ευφυΐας” στην απέναντι πλευρά της σκακιάρας. Το περιοδικό *Newsweek* χαρακτήρισε τον αγώνα ως “το τελευταίο οχυρό του εγκεφάλου”. Η αξία των μετοχών της IBM αυξήθηκε κατά 18 δισεκατομμύρια δολάρια.

**Αυτόνομος έλεγχος:** Το σύστημα υπολογιστικής όρασης ALVINN εκπαιδεύτηκε να κατευθύνει ένα αυτοκίνητο ώστε να παραμένει σε μια λωρίδα. Τοποθετήθηκε στο ελεγχόμενο από υπολογιστή όχημα NAVLAB του πανεπιστημίου Carnegie Mellon και το οδήγησε να διασχίσει τις Ηνωμένες Πολιτείες — για 2850 μίλια, είχε τον έλεγχο του τιμονιού του οχήματος στο 98% του χρόνου. Ένας άνθρωπος αναλάμβανε το τιμόνι το υπόλοιπο 2%, κυρίως στις ράμπες εξόδου. Το όχημα NAVLAB έχει βιντεοκάμερες που μεταδίδουν εικόνες του δρόμου στο ALVINN, το οποίο στη συνέχεια υπολογίζει την καλύτερη κατεύθυνση για το τιμόνι με βάση την πείρα από προηγούμενες δοκιμές.

**Διάγνωση:** Τα προγράμματα ιατρικής διάγνωσης που βασίζονται σε ανάλυση πιθανοτήτων έχουν κατορθώσει να λειτουργούν σε επίπεδο έμπειρου γιατρού σε πολλούς τομείς της ιατρικής. Ο Heckerman (1991) περιγράφει μια περίπτωση όπου ένας διάσημος ειδικός στην παθολογία των λεμφαδένων χλευάζει μια διάγνωση του προγράμματος για μια ιδιαίτερα δύσκολη περίπτωση. Οι κατασκευαστές του προγράμματος του προτείνουν να ζητήσει από τον υπολογιστή εξήγηση για τη διάγνωση. Η μηχανή επισημαίνει τους κύριους παράγοντες που επηρέασαν την απόφασή της και εξηγεί τη λεπτή αλληλεπίδραση μεταξύ πολλών από τα συμπτώματα της συγκεκριμένης περίπτωσης. Τελικά, ο ειδικός συμφωνεί με το πρόγραμμα.

**Σχεδιασμός εφοδιασμού:** Κατά την κρίση στον Περσικό Κόλπο, το 1991, οι αμερικανικές δυνάμεις χρησιμοποίησαν ένα εργαλείο δυναμικής ανάλυσης και επανασχεδιασμού (Dynamic Analysis and Replanning Tool, DART — Cross και Walker, 1994) για να κάνει αυτοματοποιημένο σχεδιασμό του εφοδιασμού και χρονοπρογραμματισμό των μεταφορών. Αυτό αφορούσε μέχρι 50.000 οχήματα, φορτία και ανθρώπους ταυτόχρονα, και έπρεπε να παίρνει υπόψη σημεία εκκίνησης, προορισμούς, δρομολόγια και να επιλύει τις συγκρούσεις μεταξύ όλων των παραμέτρων. Οι τεχνικές σχεδιασμού της TN έκαναν δυνατό να παράγεται μέσα σε ώρες ένα πλάνο το οποίο θα χρειαζόταν εβδομάδες με τις παλαιότερες μεθόδους. Η αμερικανική στρατιωτική υπηρεσία έρευνας (Defense Advanced Research Project Agency, DARPA) δήλωσε ότι αυτή η εφαρμογή και μόνο ξεπλήρωσε με το παραπάνω τα 30 χρόνια που επένδυσε η DARPA στην TN.

**Ρομποτική:** Πολλοί χειρουργοί σήμερα χρησιμοποιούν ρομποτικούς βοηθούς στη μικροχειρουργική. Το HipNav (DiGioia κ.α., 1996) είναι ένα σύστημα που χρησιμοποιεί τεχνικές υπολογιστικής όρασης για να δημιουργήσει ένα τρισδιάστατο μοντέλο της εσωτερικής ανατομίας του ασθενή και μετά χρησιμοποιεί ρομποτικό έλεγχο για να καθοδηγήσει την εισαγωγή ενός προσθετικού γοφού.

**Κατανόηση γλώσσας και επίλυση προβλημάτων:** Το PROVERB (Littman κ.α., 1999) είναι έναν πρόγραμμα υπολογιστή που λύνει σταυρόλεξα καλύτερα από τους περισσότερους ανθρώπους, χρησιμοποιώντας περιορισμούς για τις δυνατές συμπληρώσεις λέξεων, μια μεγάλη βάση δεδομένων προηγούμενων σταυρολέξων και μια ποικιλία πηγών πληροφοριών, στις οποίες περιλαμβάνονται online λεξικά και βάσεις δεδομένων, όπως ένας κατάλογος κινηματογραφικών ταινιών και των ηθοποιών που εμφανίζονται σ’ αυτές. Για παράδειγμα, το πρόγραμμα βρίσκει ότι στην περιγραφή “Nice Story” (που μπορεί να σημαίνει είτε “Ωραία ιστορία” είτε “Ιστορία στη Nice”) η απάντηση μπορεί να είναι “ETAGE”, επειδή η βάση δεδομένων του περιέχει το ζεύγος περιγραφής-απάντησης “Story in France/ETAGE” και επειδή αναγνωρίζει ότι τα πρότυπα “Nice

Χ” και “X in France” έχουν συχνά την ίδια απάντηση. Το πρόγραμμα δε γνωρίζει ότι η Nice είναι πόλη της Γαλλίας, αλλά μπορεί να λύσει το σταυρόλεξο.

Αυτά είναι μόνο μερικά από τα συστήματα τεχνητής νοημοσύνης που υπάρχουν σήμερα. Δεν είναι ούτε μαγεία ούτε επιστημονική φαντασία — είναι επιστήμη, τεχνολογία και μαθηματικά, για τα οποία αυτό το βιβλίο παρέχει μια εισαγωγή.

## 1.5 ΣΥΝΟΨΗ

Το κεφάλαιο αυτό ορίζει την ΤΝ και παρουσιάζει το πολιτιστικό υπόβαθρο στο οποίο αναπτύχθηκε. Μερικά σημαντικά σημεία:

- Οι διάφοροι ερευνητές αντιμετωπίζουν την ΤΝ διαφορετικά. Δύο σημαντικά ερωτήματα είναι: Μας ενδιαφέρει η σκέψη ή η συμπεριφορά; Θέλουμε ένα ανθρώπινο μοντέλο, ή θέλουμε να δουλέψουμε με βάση ένα ιδανικό πρότυπο;
- Σε αυτό το βιβλίο, υιοθετούμε την άποψη ότι η νοημοσύνη ασχολείται κυρίως με την **ορθολογική δράση**. Ιδανικά, ένας **ευφυής πράκτορας** (intelligent agent) πραγματοποιεί την καλύτερη δυνατή ενέργεια σε μια κατάσταση. Θα μελετήσουμε το πρόβλημα της δημιουργίας πρακτόρων που είναι ευφυείς με αυτή την έννοια.
- Οι φιλόσοφοι (ήδη από το 400 π.Χ.) έκαναν την ΤΝ αντιληπτή εξετάζοντας τις ιδέες ότι η νόηση μοιάζει από κάποιες απόψεις με μηχανή, ότι λειτουργεί βασιζόμενη σε γνώση κωδικοποιημένη σε κάποια εσωτερική γλώσσα, και ότι η σκέψη μπορεί να χρησιμοποιείται για να επιλέγεται η ενέργεια που θα πραγματοποιηθεί.
- Οι μαθηματικοί παρείχαν τα εργαλεία για το χειρισμό προτάσεων λογικής βεβαιότητας καθώς και αβέβαιων πιθανοτικών προτάσεων. Έθεσαν επίσης τις βάσεις για την κατανόηση του υπολογισμού και για τη συλλογιστική σχετικά με τους αλγόριθμους.
- Οι οικονομολόγοι τυποποίησαν το πρόβλημα της λήψης αποφάσεων που μεγιστοποιούν το αναμενόμενο αποτέλεσμα για εκείνον που παίρνει την απόφαση.
- Οι ψυχολόγοι υιοθέτησαν την ιδέα ότι οι άνθρωποι και τα ζώα μπορούν να θεωρηθούν μηχανές επεξεργασίας πληροφοριών. Οι γλωσσολόγοι έδειξαν ότι η χρήση της γλώσσας ταιριάζει σε αυτό το μοντέλο.
- Οι επιστήμονες των υπολογιστών παρείχαν τα τεχνουργήματα που έκαναν δυνατή την εφαρμογή της ΤΝ. Τα προγράμματα ΤΝ τείνουν να είναι μεγάλα, και δε θα μπορούσαν να λειτουργήσουν χωρίς τις τεράστιες προόδους που έχει κάνει η βιομηχανία των υπολογιστών στα ζητήματα της ταχύτητας και της μνήμης.
- Η θεωρία ελέγχου ασχολείται με τη σχεδίαση συσκευών που ενεργούν με το βέλτιστο τρόπο με βάση την ανάδραση από το περιβάλλον. Αρχικά, τα μαθηματικά εργαλεία της θεωρίας ελέγχου ήταν πολύ διαφορετικά από της ΤΝ, αλλά τα δύο πεδία έρχονται πιο κοντά.
- Η ιστορία της ΤΝ είχε κύκλους επιτυχίας, άτοπης αισιοδοξίας και επακόλουθης μείωσης του ενθουσιασμού και της χρηματοδότησης. Υπήρξαν επίσης κύκλοι εμφάνισης νέων δημιουργικών προσεγγίσεων και συστηματικής βελτίωσης των καλύτερων από αυτές.
- Η ΤΝ έκανε γρηγορότερες προόδους την τελευταία δεκαετία, λόγω της μεγαλύτερης χρήσης της επιστημονικής μεθόδου στον πειραματισμό και τη σύγκριση των διαφόρων προσεγγίσεων.
- Οι πρόσφατες πρόοδοι στην κατανόηση της θεωρητικής βάσης της νοημοσύνης συνοδεύτηκαν από παράλληλες βελτιώσεις των δυνατοτήτων των πραγματικών συστημάτων. Τα επιμέρους πεδία της ΤΝ ενοποιήθηκαν περισσότερο και η ΤΝ βρήκε κοινό έδαφος με άλλους κλάδους.

---

**ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΚΑΙ ΙΣΤΟΡΙΚΕΣ ΣΗΜΕΙΩΣΕΙΣ**

Η μεθοδολογική θέση της τεχνητής νοημοσύνης διερευνάται στο *The Sciences of the Artificial* του Herb Simon (1981), όπου εξετάζονται τομείς της έρευνας που ασχολούνται με πολύπλοκα τεχνουργήματα. Το βιβλίο αυτό εξηγεί πώς η ΤΝ μπορεί να θεωρηθεί και ως εμπειρική επιστήμη και ως μαθηματικά. Ο Cohen (1995) κάνει μια επισκόπηση της πειραματικής μεθοδολογίας στο χώρο της ΤΝ. Οι Ford και Hayes (1995) δίνουν μια μάλλον προκατειλημμένη άποψη για τη χρησιμότητα της δοκιμασίας Turing.

Το βιβλίο *Artificial Intelligence: The Very Idea* του John Haugeland (1985) κάνει έναν καλογραμμένο απολογισμό των φιλοσοφικών και πρακτικών προβλημάτων της ΤΝ. Η γνωστική επιστήμη (cognitive science) περιγράφεται από πολλά πρόσφατα βιβλία (Johnson-Laird, 1988· Stillings κ.α., 1995· Thagard, 1996) και από την *Encyclopedia of the Cognitive Sciences* (Wilson και Keil, 1999). Ο Baker (1989) καλύπτει το συντακτικό μέρος της σύγχρονης γλωσσολογίας και οι Chierchia και McConnell-Ginet (1990) καλύπτουν τη σημασιολογία. Οι Jurafsky και Martin (2000) καλύπτουν την υπολογιστική γλωσσολογία.

Οι πρώιμες προσπάθειες για την ΤΝ περιγράφονται στο *Computers and Thought* των Feigenbaum και Feldman (1963), στο *Semantic Information Processing* του Minsky (1968) και στη σειρά *Machine Intelligence* με επιμέλεια του Donald Michie. Ένας μεγάλος αριθμός σημαντικών επιστημονικών δημοσιεύσεων έχουν ανθολογηθεί από τους Webber και Nilsson (1981) και από τον Luger (1995). Πρώιμες δημοσιεύσεις για τα νευρωνικά δίκτυα υπάρχουν συγκεντρωμένες στη συλλογή *Neurocomputing* (Anderson και Rosenfeld, 1988). Η *Encyclopedia of AI* (Shapiro, 1992) περιέχει άρθρα ανασκόπησης για σχεδόν όλα τα θέματα της ΤΝ. Τα άρθρα αυτά συνήθως παρέχουν μια καλή εισαγωγή στη βιβλιογραφία της έρευνας πάνω στο κάθε θέμα.

Η πιο πρόσφατη δουλειά εμφανίζεται στα πρακτικά των μεγάλων συνεδρίων της ΤΝ: του διετούς International Joint Conference on AI (IJCAI), του επίσης διετούς European Conference on AI (ECAI), και του National Conference on AI, το οποίο συνήθως αναφέρεται ως AAAI από την επωνυμία του οργανισμού που το έχει υπό την αιγίδα του. Τα κύρια περιοδικά για τη γενική ΤΝ είναι το *Artificial Intelligence*, το *Computational Intelligence*, το *IEEE Transactions on Pattern Analysis and Machine Intelligence*, το *IEEE Intelligent Systems* και το ηλεκτρονικό *Journal of Artificial Intelligence Research*. Υπάρχουν επίσης πολλά συνέδρια και περιοδικά αφιερωμένα σε ειδικούς τομείς, τα οποία θα καλύψουμε στα σχετικά κεφάλαια. Οι κύριες επαγγελματικές ενώσεις για την ΤΝ είναι η American Association for Artificial Intelligence (AAAI), το Special Interest Group in Artificial Intelligence (SIGART) της ACM, και η Society for Artificial Intelligence and Simulation of Behaviour (AISB). Το περιοδικό *AI Magazine* της AAAI περιέχει πολλά θεματικά και εκπαιδευτικά άρθρα και η ιστοσελίδα του, [aaai.org](http://aaai.org), περιέχει νέα και πληροφορίες υποβάθρου.

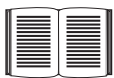
---

**ΑΣΚΗΣΕΙΣ**

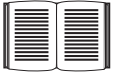
Σκοπός αυτών των ασκήσεων είναι να παρακινήσουν τη συζήτηση και μερικές μπορεί να ανατεθούν ως εργασίες εξαμήνου. Εναλλακτικά, μπορεί να γίνουν κάποιες προκαταρκτικές προσπάθειες τώρα, και να αναθεωρηθούν μετά την ολοκλήρωση του βιβλίου.

**1.1** Ορίστε με δικά σας λόγια τους όρους: (α) νοημοσύνη, (β) τεχνητή νοημοσύνη, (γ) πράκτορας.

**1.2** Διαβάστε την πρωτότυπη δημοσίευση του Turing για την ΤΝ (Turing, 1950). Στο κείμενο αυτό, ο Turing εξετάζει πολλές ενδεχόμενες αντιρρήσεις στο προτεινόμενο εγχείρημά του και στη δοκιμασία του για τη νοημοσύνη. Ποιες αντιρρήσεις εξακολουθούν να έχουν κάποια βαρύτητα; Είναι οι ανασκευές του σωστές; Μπορείτε να σκεφτείτε νέες αντιρρήσεις που προκύπτουν



από τις εξελίξεις αφότου έγραψε τη δημοσίευση; Στο κείμενό του, ο Turing έκανε την πρόγνωση ότι μέχρι το έτος 2000 ένας υπολογιστής θα έχει πιθανότητα 30% να περάσει μια πεντάλεπτη δοκιμασία Turing με ένα μη έμπειρο εξεταστή. Τι πιθανότητα νομίζετε ότι θα είχε ένας υπολογιστής σήμερα; Σε άλλα 50 χρόνια;

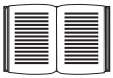


**1.3** Το βραβείο Loebner απονέμεται κάθε χρόνο στο πρόγραμμα που πλησιάζει περισσότερο στο να περάσει μια έκδοση της δοκιμασίας Turing. Ερευνήστε ποιο πρόγραμμα ήταν το τελευταίο που κέρδισε το βραβείο Loebner. Ποιες τεχνικές χρησιμοποιεί; Πώς συμβάλλει στην πρόοδο της σύγχρονης τεχνολογίας της TN;

**1.4** Υπάρχουν καλά γνωστές κλάσεις προβλημάτων τα οποία παρουσιάζουν ανεξέλεγκτη δυσκολία για τους υπολογιστές (δυσεπίλυτα — intractable), και άλλες κλάσεις προβλημάτων που είναι αποδειγμένα μη αποφασίσιμα (undecidable). Μήπως αυτό σημαίνει ότι η TN είναι αδύνατη;

**1.5** Ας υποθέσουμε ότι επεκτείνουμε το πρόγραμμα ANALOGY του Evans έτσι ώστε να μπορεί να επιτύχει βαθμολογία 200 σε ένα συνηθισμένο τεστ ευφυΐας. Θα είχαμε τότε ένα πρόγραμμα πιο ευφυές από τον άνθρωπο; Εξηγήστε.

**1.6** Πώς μπορεί η ενδοσκόπηση — η αναφορά των εσωτερικών σκέψεων κάποιου — να είναι ανακριβής; Θα μπορούσα εγώ να κάνω λάθος για το τι σκέπτομαι; Συζητήστε το.



**1.7** Ψάξτε στη βιβλιογραφία της TN για να βρείτε αν οι υπολογιστές μπορούν σήμερα να κάνουν τις παρακάτω εργασίες:

- α. Να παίξουν καλό πινγκ-πονγκ.
- β. Να οδηγήσουν στο κέντρο του Καΐρου.
- γ. Να ψωνίσουν οπωρικά για μια εβδομάδα από την αγορά.
- δ. Να ψωνίσουν οπωρικά για μια εβδομάδα από τον Ιστό.
- ε. Να παίξουν καλό μπριτζ αγωνιστικού επιπέδου.
- στ. Να ανακαλύψουν και να αποδείξουν νέα μαθηματικά θεωρήματα.
- ζ. Να γράψουν μια σκόπιμα αστεία ιστορία.
- η. Να δώσουν έγκυρες νομικές συμβουλές σε έναν εξειδικευμένο νομικό τομέα.
- θ. Να μεταφράσουν προφορικά Αγγλικά σε προφορικά Σουηδικά σε πραγματικό χρόνο.
- ι. Να πραγματοποιήσουν μια πολύπλοκη χειρουργική επέμβαση.

Για τις εργασίες που είναι σήμερα ανέφικτες, προσπαθήστε να βρείτε ποιες είναι οι δυσκολίες και να κάνετε μια πρόγνωση για το πότε, αν ποτέ, θα ξεπεραστούν.

**1.8** Μερικοί συγγραφείς έχουν υποστηρίξει ότι η αντίληψη και οι κινητικές ικανότητες είναι το σημαντικότερο μέρος της νοημοσύνης, και ότι οι ικανότητες “ανώτερου επιπέδου” είναι κατ’ ανάγκη παρασιτικές — απλές προσθήκες σε αυτές τις υποκείμενες λειτουργίες. Είναι βέβαια αλήθεια ότι το μεγαλύτερο μέρος της βιολογικής εξέλιξης και ένα μεγάλο μέρος του εγκεφάλου αφορούν την αντίληψη και τις κινητικές ικανότητες, ενώ το πεδίο της TN βρίσκει τις εργασίες όπως τα παιχνίδια και η λογική εξαγωγή συμπερασμάτων από πολλές απόψεις ευκολότερες από την αντίληψη και τη δράση στον πραγματικό κόσμο. Νομίζετε ότι η παραδοσιακή εστίαση της TN στις γνωστικές ικανότητες ανώτερου επιπέδου είναι άστοχη;

**1.9** Γιατί η βιολογική εξέλιξη τείνει να οδηγεί σε συστήματα που ενεργούν ορθολογικά; Ποιους στόχους είναι σχεδιασμένα να επιτυγχάνουν τα συστήματα αυτού του είδους;

**1.10** Είναι οι αντανάκλαστικές ενέργειες (όπως η απομάκρυνση του χεριού σας από ένα φούρνο που καίει) ορθολογικές; Είναι ευφυείς;

**1.11** “Και βέβαια οι υπολογιστές δεν μπορεί να είναι ευφυείς — μπορούν να κάνουν μόνο ό,τι τους λένε οι προγραμματιστές τους.” Είναι η δεύτερη πρόταση αληθής; Συνεπάγεται την πρώτη;



**1.12** “Και βέβαια τα ζώα δεν μπορεί να είναι ευφυή — μπορούν να κάνουν μόνο ό,τι τους λένε τα γονίδια τους.” Είναι η δεύτερη πρόταση αληθής; Συνεπάγεται την πρώτη;

**1.13** “Και βέβαια τα ζώα, οι άνθρωποι και οι υπολογιστές δεν μπορεί να είναι ευφυείς — μπορούν να κάνουν μόνο ό,τι υπαγορεύουν οι νόμοι της Φυσικής στα άτομα από τα οποία αποτελούνται.” Είναι η δεύτερη πρόταση αληθής; Συνεπάγεται την πρώτη;