Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2006). The role of physical embodiment in human-robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 117-122). Hatfield, United Kingdom: IEEE.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., &Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 872-877). Jeju, South Korea: IEEE.

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J. H., & Cacioppo, J. T. (2010). Making sense by making sentient: effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, 99(3), 410-435.doi: 10.1037/a0020240

Weiss, A., Wurhofer, D., &Tscheligi, M. (2009). "I love this dog"—children's emotional attachment to the robotic dog AIBO. *International Journal of Social Robotics,* 1(3), 243-248.

Winnicott, D. (1971/ 2005). *Playing and Reality.* London: Routledge.

Wrobel, T. A., & Dye, A. L. (2003). Grieving pet death: Normative, gender, and attachment issues. Omega: *Journal of Death and Dying*, 47, 385-393.

Χουντουμάδη, Α. & Πατεράκη, Ε. (2008). *Λεξικό Ψυχολογίας*. Αθήνα: Τόπος.

ΑΝΘΡΩΠΟΙ ΚΑΙ ΡΟΜΠΟΤ ΣΤΟΝ 21ο ΑΙΩΝΑ

ΠΡΟΟΠΤΙΚΕΣ ΣΥΝΥΠΑΡΞΗΣ ΚΑΙ ΣΥΝΑΙΣΘΗΜΑΤΙΚΩΝ ΣΥΝΔΕΣΕΩΝ

Παναγιώτης Χαλάτσης

Στις αρχές του 21ου αιώνα, οι τεχνολογικές εξελίξεις που χαρακτηρίζουν τη λεγόμενη *τέταρτη βιομηχανική επανάσταση,* περιλαμβάνουν την *τεχνητή νοημοσύνη,* τα *μεγάλα δεδομένα,* τη *ρομποτική,* το *διαδίκτυο των πραγμάτων*, και την *επαυξημένη πραγματικότητα.* Ειδικά η Τεχνητή Νοημοσύνη (ΤΝ) έχει εισβάλει στην καθημερινότητά μας, έστω και αν συνήθως δεν συνειδητοποιούμε την ύπαρξή της. Τα σύγχρονα ρομπότ της αυτοκινητοβιομηχανίας, οι αυτόματοι μεταφραστές της Google, η ρομποτική χειρουργική, το χρηματιστήριο, είναι μερικές μόνο από τις ολοένα αυξανόμενες εφαρμογές της. Τα υπάρχοντα συστήματα ΤΝ υπερβαίνουν ήδη τις ανθρώπινες δυνατότητες σε διάφορους τομείς, όπως στους μαθηματικούς υπολογισμούς, στα παιχνίδια στρατηγικής, στην αυτόματη μετάφραση γλωσσών, στη διάγνωση διαφόρων ειδών καρκίνου, στην αναγνώριση οπτικών δεδομένων κ.ά.

Ωστόσο, η νοημοσύνη αυτών των συστημάτων είναι περιορισμένη σε ένα μόνο γνωστικό έργο και καθόλου δεν μοιάζει

με την ανθρώπινη. Για παράδειγμα, το AlphaGo, ένα από τα πιο εξελιγμένα συστήματα ΤΝ, μπορεί να κατατρόπωσε όλους τους πρωταθλητές παγκοσμίως στο πολύπλοκο παιχνίδι στρατηγικής Go, αλλά δεν μπορεί να παίξει ντάμα ή σκάκι, να κάνει αναγνώριση οπτικών δεδομένων ή αυτόματη μετάφραση ανάμεσα σε γλώσσες κλπ. Εντούτοις, μια μεγάλη αλλαγή έχει συμβεί στη σύγχρονη ΤΝ· χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα, μεθόδους ενισχυτικής μάθησης αλλά και μάθησης χωρίς επίβλεψη, είναι σε θέση να μαθαίνει, να κάνει προβλέψεις και να παίρνει αποφάσεις χωρίς να έχει προγραμματιστεί διεξοδικά για το συγκεκριμένο έργο που επιτελεί.

Χαρακτηριστικότερο παράδειγμα της σύγχρονης ΤΝ είναι το προαναφερθέν λογισμικό AlphaGo της Google, που ενώ διέθετε ελάχιστες σε σχέση με τις $2{,}08168199382\times10^{170}$ πιθανές κινήσεις του παιχνιδιού, κατάφερε να μάθει παίζοντας κυρίως εναντίον του εαυτού του (Silver et al., 2017). Αυτό κάνει πολλούς ερευνητές να υποστηρίζουν ότι βρισκόμαστε στην αρχή μιας εποχής *τεχνολογικής μοναδικότητας*, όπου η ΤΝ θα αποκτήσει απρόβλεπτες δυνατότητες, οπότε και όλα θα είναι εφικτά από τεχνικής άποψης, συμπεριλαμβανομένων και ευφυών, ενσυνείδητων και ενσυναίσθητων μηχανών (Kurzweil, 2005). Όμως πόσο είναι εφικτό αυτό πραγματικά; Προς το παρόν η ρομποτική έχει πετύχει σε μεγάλο βαθμό να κατασκευάσει ρομπότ που κινητοποιούν το ανθρώπινο συναίσθημα, αλλά απέχει πολύ από την κατασκευή ρομπότ που σκέφτονται και αντιλαμβάνονται τον κόσμο όπως οι άνθρωποι. Έτσι, η δημιουργία μηχανών με ανθρώπινη νοημοσύνη και συνειδητότητα αποτελεί το Ιερό Δισκοπότηρο των ερευνητών της ΤΝ, όμως διχάζονται ως προς το αν και πότε θα επιτευχθεί.

Σύμφωνα με τον Tegmark (2017), υπάρχουν αυτοί που βασιζόμενοι στην εκθετική αύξηση των υπολογιστικών δυ-

νατοτήτων που ορίζει ο Νόμος του Moore (Moore, 1997), πιστεύουν ότι αυτό θα επιτευχθεί στον αιώνα μας, ενώ άλλοι θεωρούν ότι η ανάπτυξη Τεχνητής Γενικής Νοημοσύνης (που αντιστοιχεί στην ανθρώπινη), συναισθημάτων και συνειδητότητας δεν μπορεί να επιτευχθεί στα επόμενα 200 τουλάχιστον χρόνια, οπότε δεν αξίζει να ασχοληθούμε επί του παρόντος. Τέλος, μια τρίτη ομάδα πρεσβεύει ότι αν και οι νοήμονες μηχανές θα αργήσουν να δημιουργηθούν, δεν πρέπει να αναβάλλουμε τις σχετικές ερευνητικές προσπάθειες. Έτσι, αφενός έχουμε ερευνητές που ασχολούνται με τη δημιουργία μη συνειδητών μηχανών, που αν μη τι άλλο θα μας δίνουν την εντύπωση ότι μας νιώθουν και μας καταλαβαίνουν, κι αφετέρου έχουμε αυτούς που αποσκοπούν στη δημιουργία υπερευφυών μηχανών, που όντως θα νοούν και θα αισθάνονται τον κόσμο όπως οι άνθρωποι.

Οι πρώτοι θεωρούν ότι έχει επέλθει η εποχή όπου τα ρομπότ θα μας είναι απαραίτητα ως κοινωνικοί σύντροφοι, βοηθοί και συνεργάτες, οπότε προσανατολίζονται στη δημιουργία όλο και πιο εξελιγμένων *κοινωνικών* ή με μια άλλη ορολογία *συναισθηματικών* ρομπότ ή *Τεχνητών Συντρόφων*, τα οποία θα «εκμεταλλεύονται» τις ανθρώπινες συναισθηματικές ιδιαιτερότητες, όπως, για παράδειγμα, το να αναγνωρίζουν σε κάποιον βαθμό τα συναισθήματά μας, να μας κοιτούν στα μάτια, να ακολουθούν την κίνησή μας, να προσομοιάζουν ενδιαφέρον για εμάς κ.ο.κ, ώστε να αποδεχτούμε τη συνύπαρξη μαζί τους (Breazeal, 2002· Picard, 2000). Οι δεύτεροι πιστεύουν ότι ποτέ δεν θα υπάρξει ουσιαστική επικοινωνία ή και σχέση ανθρώπων με ρομπότ, αν αυτά δεν αντιλαμβάνονται και δεν συνειδητοποιούν τον κόσμο όπως οι άνθρωποι. Ωστόσο, υπάρχει και μια τρίτη κατηγορία ερευνητών, οι οποίοι θεωρούν ότι η σχέση των ανθρώπων με μηχανές είναι πολύ πιθανόν να αποκτήσει μια συμβιωτική μορφή, μέσα από την ενσωμάτωση ρομποτικών συστημάτων στον ίδιο τον άνθρωπο (Harari, 2016· Kurzweil,

2005· Tegmark, 2017). Πρόκειται ουσιαστικά για συνένωση ανθρώπου-μηχανής, δηλαδή για έναν υβριδικό οργανισμό που υπόσχεται να μας αναβαθμίσει σε υπεράνθρωπους. Αναφερόμαστε σε εμφυτευμένες μνήμες, τεχνητή όραση και ακοή, γενετικά τροποποιημένα γονίδια, τεχνητά μέλη, νανορομπότ που ανιχνεύουν ασθένειες στο ανθρώπινο σώμα κ.ο.κ.

Στις επόμενες ενότητες θα εστιαστούμε στους λόγους για τους οποίους πολλοί άνθρωποι ενδέχεται να δεθούν συναισθηματικά με τα μη συνειδητά, πλην τεχνολογικά προηγμένα κοινωνικά ρομπότ, τα οποία στις επόμενες δεκαετίες θα διατίθενται ευρέως στην αγορά για οικιακή χρήση και στη συνέχεια θα εξετάσουμε συνοπτικά θεωρητικές θέσεις και ερευνητικές προσεγγίσεις αναφορικά με την πιθανότητα δημιουργίας ενσυνείδητων ρομπότ. Τα θέματα αυτά θα διατρέχονται από ευρύτερους προβληματισμούς σχετικά με τις πιθανές επιπτώσεις που η συνύπαρξη και συναισθηματική σύνδεση με ανθρωπόμορφα ρομπότ θα έχει για τη φύση του ανθρώπου και των διαπροσωπικών του σχέσεων.

Θα δεθούν συναισθηματικά οι άνθρωποι με τα κοινωνικά ρομπότ;

Δύο βασικές παράμετροι ενδέχεται να οδηγήσουν σημαντικό μέρος των ανθρώπων στη δόμηση συναισθηματικών συνδέσεων με τα σύγχρονα και διαρκώς εξελισσόμενα κοινωνικά ρομπότ, μέσα στον 21ο αιώνα: αφενός οι ψυχολογικές ευαισθησίες που χαρακτηρίζουν το ανθρώπινο είδος θα επιτρέψουν στα ρομπότ να δίνουν την εντύπωση ότι μας καταλαβαίνουν και νοιάζονται για εμάς, αφετέρου οι επιστημονικές προσεγγίσεις σχετικά με το πώς λειτουργεί ο νους, θα επηρεάσουν τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται τον εαυτό τους, δίνοντας έτσι χώρο σε νέου τύπου σχέσεις, οι

οποίες ενδέχεται να συμπεριλαμβάνουν και συναισθηματικές συνδέσεις με ρομπότ.

Ανθρώπινες εξελικτικές ευαισθησίες

Τα Tamagotchi ήταν μικρά φορητά ηλεκτρονικά «ζωάκια», σχήματος αυγού, της ιαπωνικής εταιρείας Bandai, τα οποία μπήκαν για μια δεκαετία στη ζωή των παιδιών στα τέλη της δεκαετίας του '90. Δεν έκλειναν με On/Off, ήθελαν πάντα να «νοιάζεσαι» γι' αυτά. Έπρεπε να τα «ταΐζεις», να τα «κοιμίζεις», να παίζεις μαζί τους, ειδάλλως «στενοχωριούνταν», «αρρώσταιναν» ή και «πέθαιναν». Πάνω από 70 εκατομμύρια παιδιά «ανέθρεψαν» Tamagotchi. Αν «πέθαιναν», μπορούσες να τους κάνεις επανεκκίνηση, όμως για τα περισσότερα παιδιά δεν θα ήταν πια το ίδιο, το δικό τους Tamagotchi, οπότε τα θρηνούσαν στα αλήθεια και τα έθαβαν είτε κανονικά στη γη είτε σε ψηφιακά «νεκροταφεία». Τα Tamagotchi μας έμαθαν ότι ο άνθρωπος δεν φροντίζει μόνο ό,τι αγαπάει, αλλά αγαπάει και ό,τι φροντίζει (Turkle, 2011).

Πρόσφατα, η Julie Carpenter (2016) μελέτησε το πώς οι πυροτεχνουργοί του στρατού των ΗΠΑ νιώθουν απέναντι στα ρομπότ εξουδετέρωσης εκρηκτικών που χειρίζονται. Περιγράφει την αντίφαση που εμφανίζεται στις αφηγήσεις τους, καθώς ενώ αναγνωρίζουν ότι οι «βοηθοί» τους είναι μηχανές, συγχρόνως δηλώνουν ότι νιώθουν ένα είδος συναισθηματικής απώλειας, όταν ένα ρομπότ με το οποίο συνεργάζονταν για καιρό εκραγεί και καταστραφεί. Τα ρομπότ είναι μηχανές, αλλά στον νου των πυροτεχνουργών αποκτούν κάποιες ανθρώπινες ιδιότητες, με τον ίδιο ασυνείδητο τρόπο που τα παιδιά θρηνούσαν τα Tamagotchi. Αυτό λέγεται ανθρωπομορφισμός, η τάση δηλαδή του ανθρώπου να αποδίδει ανθρώπινες ιδιότητες σε μη ανθρώπινα όντα.

Στον τομέα της ΤΝ αρχικά παρατηρήθηκε από τον ερευνητή του MIT Weizenbaum (1966) και ονομάστηκε *φαινόμενο ELIZA,* παίρνοντας το όνομα του αντίστοιχου λογισμικού επικοινωνίας με ανθρώπους (chatbot). Ο Weizenbaum παρατήρησε ότι οι αλληλεπιδράσεις με το λογισμικό προκαλούσαν συναισθηματικές αντιδράσεις στους ανθρώπους. Για παράδειγμα, στο «ευχαριστώ» του *ELIZA* απέδιδαν ένα είδος ευγνωμοσύνης και στο «θα ήθελες να σε βοηθήσω σε κάτι;» μια ιδέα συμπάθειας ή φροντίδας. Λίγα χρόνια αργότερα, οι ερευνητές Reeves και Nass (1996), βασισμένοι σε πειραματικές έρευνες στο πανεπιστήμιο Stanford, έδειξαν πως οι άνθρωποι τείνουν να συμπεριφέρονται σε μηχανές όπως θα φέρονταν σε ανθρώπους. Είναι ιδιαίτερα ενδιαφέρον που ένας σημαντικός στοχαστής του 20ού αιώνα, ο John Dewey (1934), είχε επισημάνει ότι τα εργαλεία που χρησιμοποιεί ο άνθρωπος, συχνά μπορούν ως διαμεσολαβητές μιας αισθητικής εμπειρίας να του ασκήσουν τέτοια γοητεία, ώστε για αυτόν να αποκτήσουν ένα είδος «προσωπικότητας».

Αυτή η τάση ανθρωπομορφισμού που μας χαρακτηρίζει, έχει προεκτάσεις σε διάφορες ψυχολογικές αποκρίσεις μας απέναντι σε μηχανές, αν οι δημιουργοί τους γνωρίζουν πώς να την «εκμεταλλευτούν». Σε μια παρουσίαση συνεδρίου το 1999, η ερευνήτρια Freedom Baird (όπως αναφέρει η Turkle, 2011) ζήτησε από κάποιους από το κοινό να κρατήσουν ανάποδα για όση ώρα μπορούσαν μια κούκλα Barbie (χωρίς δυνατότητα φωνής), ένα Furby (ένα μαλλιαρό ηλεκτρονικό «ζωάκι» που κυκλοφόρησε το 1998 και όταν το αναποδογύριζες κλαψούριζε και έλεγε «εγώ φοβάμαι») και ένα αληθινό χάμστερ. Εν αντιθέσει με το χάμστερ, οι άνθρωποι δεν είχαν πρόβλημα να κρατήσουν τη Barbie και το Furby ανάποδα. Όμως όταν έπειτα από λίγο το Furby άρχιζε να κλαψουρίζει λέγοντας «εγώ φοβάμαι», τότε στον νου αρκετών μεταμορφώθηκε σε «ζωντανό», οπότε το γύριζαν πάλι κανονικά.

Φαίνεται ότι οι άνθρωποι, καθώς επικοινωνούν με μηχανές, βιώνουν μια σύγκρουση. Ενώ αναγνωρίζουν ότι επικοινωνούν με κάτι τεχνητό, καθώς αυτό εμφανίζει κάποιες ανθρώπινες ιδιότητες και δεξιότητες, παγιδεύονται ανάμεσα σε αντιφατικά μεταξύ τους γνωστικά δεδομένα και, προκειμένου να λύσουν αυτήν τη γνωστική ασυμφωνία (Festinger, 1957), συχνά του αποδίδουν μια, έστω αχνή, ανθρώπινη διάσταση.

Τα σύγχρονα κοινωνικά ρομπότ μπορεί να μην έχουν συνειδητότητα ή συναισθήματα, αλλά κάνουν πολύ περισσότερα από την εποχή του *ELIZA*. Για παράδειγμα, η Octavia, το ρομπότ-πυροσβέστης του αμερικανικού Ναυτικού, διακρίνει άτομα από τη χροιά της φωνής, αναγνωρίζει με την αφή διάφορα αντικείμενα και είναι εξοπλισμένο με μια *θεωρία του νου*, ώστε να διακρίνει κάποιες ψυχολογικές καταστάσεις των ανθρώπων. Το εμπορικό ρομπότ Pepper της Soft Bank Robotics, που βγήκε στην ιαπωνική αγορά το 2015, κάνει οπτική επαφή, αναγνωρίζει τις κύριες ανθρώπινες συγκινήσεις, λέει αστεία και ξεναγεί ανθρώπους σε διάφορους χώρους. Συνοπτικά, ο προσανατολισμός των ερευνητών είναι να δημιουργήσουν μηχανές που θα είναι σε θέση να αναγνωρίζουν τη συναισθηματική διάθεση του ανθρώπου και στη συνέχεια να ανταποκρίνονται, κάνοντας γνωστική εκτίμηση της κατάστασης.

Οι σχεδιαστές κοινωνικών ρομπότ προβληματίζονται ιδιαίτερα σε σχέση με το ποια θα πρέπει να είναι η εμφάνισή τους. Τα περισσότερα από τα κοινωνικά ρομπότ που σχεδιάζονται έχουν εμφανισιακή αντιστοιχία με τον άνθρωπο (π.χ. κεφάλι, μάτια, χέρια), αλλά σίγουρα δεν μοιάζουν με εμάς. Υπάρχει η άποψη ότι όσο πιο πολύ μας μοιάζουν τόσο ευκολότερα θα τα αποδεχτούμε (Levy, 2007), γι' αυτό και απώτερος στόχος είναι η δημιουργία ανδροειδών, δηλαδή ρομπότ με ανθρώπινη εμφάνιση, δέρμα, ομιλία, κίνηση κ.ο.κ. Πολλοί υποστηρίζουν ότι αν δεν είμαστε έτοιμοι να δημιουργήσουμε ανδροειδή με πλήρη ανθρώπινα χαρακτηριστικά, καλύτερα να μην το επι-

χειρήσουμε, καθώς το «περίπου σαν άνθρωποι» θα μας οδηγήσει σε κάποιο είδος αποστροφής ή και φόβου προς αυτά. Είναι το φαινόμενο της «ανοίκειας κοιλάδας», που πρώτος επεσήμανε ο Masahiro Mori (1970). Αν κάτι μοιάζει πολύ με άνθρωπο, περιμένουμε να φερθεί όπως ένας άνθρωπος. Αν αυτό δεν συμβεί, τότε εισερχόμαστε σε έναν *ανοίκειο χώρο,* με τη φροϋδική έννοια του ανοίκειου, όπου κάτι που μέχρι τώρα μας φαινόταν γνωστό και συνηθισμένο, τώρα μας φαίνεται περίεργο, αλλόκοτο και απόκοσμο (Freud, 1919).

Όπως και να έχει, οι ανθρώπινες ψυχολογικές μεταβιβάσεις σε άψυχες μηχανές θα είναι ο κύριος παράγοντας που σταδιακά θα επιτρέψει συναισθηματικές συνδέσεις αρκετών ανθρώπων μαζί τους. Ο τρόπος που ανατραφήκαμε και μάθαμε να αντιλαμβανόμαστε τον εαυτό μας και να συνδεόμαστε με τους άλλους (κάτι που η ψυχανάλυση ονομάζει *σχέση αντικειμένου*) θα καθορίσει το είδος των μεταβιβάσεών μας προς τις μηχανές. Για άλλους το ρομπότ τους θα είναι ο φίλος, για άλλους ο θαυμαστής, ο πατέρας ή το θύμα τους, για αρκετούς απλώς ένα άψυχο αντικείμενο που τους παρέχει κάποιες υπηρεσίες και πιθανόν για μερικούς πολλά από τα παραπάνω κατά περίσταση. Η πρακτική υποστήριξη που τα ρομπότ θα παρέχουν στους χρήστες τους θα οδηγήσει στη σταδιακή εξοικείωση μαζί τους, στοιχείο που θα ενισχύσει περαιτέρω τέτοιου είδους μεταβιβάσεις.

Πρακτικές αναγκαιότητες και εξοικείωση με τα ρομπότ

Σημειώσαμε πως η τάση του ανθρώπου για ανθρωπομορφισμό, οι ψυχολογικές προβολές και οι ασυνείδητες μεταβιβάσεις θα οδηγήσουν σε κάποια είδη συναισθηματικών συνδέσεων των ανθρώπων με μηχανές. Ο Levy (2007) υποστηρίζει ότι τέτοιες συναισθηματικές συνδέσεις είναι αναπόφευκτες,

κυρίως λόγω της σταδιακής εξοικείωσής μας, καθώς, όπως έδειξε ο Zajonc (1980), όσο πιο πολύ εξοικειώνεσαι με κάτι τόσο αυξάνονται οι πιθανότητες να το προτιμήσεις ή και να το συμπαθήσεις. Στη διαδικασία αυτή, κεντρικό ρόλο θα διαδραματίσει η εξέλιξη των κυρίαρχων κοινωνικών αναπαραστάσεων. Για παράδειγμα, στην Ιαπωνία όπου η ανιμιστική πεποίθηση ότι ακόμα και τα άψυχα αντικείμενα έχουν ένα είδος ψυχικής οντότητας, η αποδοχή των κοινωνικών ρομπότ έγινε με μεγάλη ευκολία. Κατά τον Levy (ό.π.), η αρχική αποδοχή τους από ένα μέρος των ανθρώπων σταδιακά θα μειώσει τις προκαταλήψεις των υπόλοιπων, οπότε και οι συνδέσεις με μηχανές θα γίνουν κοινωνικά αποδεκτές. Είναι αναμενόμενο ότι η αποδοχή αυτή θα είναι ευκολότερη για τους νεότερους, τους λεγόμενους «ψηφιακούς ιθαγενείς» (Prensky, 2001:1), για τους οποίους τα όρια ανάμεσα στο εικονικό και το πραγματικό είναι ήδη πιο δυσδιάκριτα (Floridi, 2010).

Για να εξοικειωθούμε όμως με τα ρομπότ, θα πρέπει πρώτα να έχουν γίνει απαραίτητο μέρος της ζωής μας. Ήδη οι ψηφιακοί ευφυείς προσωπικοί βοηθοί, όπως η *Alexa* της Amazon, η *Siri* της Apple, ή το *Assistant* της Google, έχουν μπει στην καθημερινότητα πολλών εκατομμυρίων ανθρώπων. Στις αρχές του 2017, ο εκτιμώμενος αριθμός όσων χρησιμοποιούσαν τέτοιους βοηθούς συστηματικά ήταν πάνω από ένα δισεκατομμύριο, ενώ ο αριθμός αυτός αυξάνεται.[1] Αν εξετάσουμε τώρα τα κοινωνικά ρομπότ, βλέπουμε ότι μέχρι το 2018 πουλήθηκαν 2.680.000 μονάδες.[2] Στόχος των εταιρειών που τα παράγουν είναι, όπως υποστηρίζουν, η εξυπηρέτηση των ανθρώπινων αναγκών – από τις πρακτικές υποχρεώσεις της καθημερινότητας ως το να καταλαβαίνουν σε κάποιον

[1] https://www.statista.com/chart/5621/users-of-virtual-digital-assistants/
[2] https://www.statista.com/statistics/755677/social-and-entertainment-robot-sales-worldwide/

βαθμό τη διάθεσή μας και να τη βελτιώνουν αν χρειαστεί, μαθαίνοντας τις ιδιαιτερότητες και τις προτιμήσεις μας. Οι προσωπικοί αυτοί άυλοι ή υλικοί βοηθοί μπορούν να βρίσκουν τη μουσική ή την ταινία που τους ζητάς, να παραγγέλνουν και να πληρώνουν με την κάρτα σου τα προϊόντα που θέλεις, να σου βρίσκουν φτηνά εισιτήρια, να πιάνουν κουβέντα ή να παίζουν επιτραπέζια μαζί σου, να σε βοηθούν να επικοινωνείς με τους δικούς σου, να λύνουν απορίες σου, να σου θυμίζουν τα φάρμακα που παίρνεις κ.ο.κ.

Η βοήθεια που αυτές οι μηχανές δίνουν και υπόσχονται να δίνουν όλο και πιο επιτυχημένα στο επίπεδο διαχείρισης πληροφοριών στους ιδιοκτήτες τους, αποκτά ένα επιπλέον ενδιαφέρον: Με δεδομένο ότι το ανθρώπινο γνωστικό σύστημα, λόγω της περιορισμένης χωρητικότητάς του, δυσκολεύεται να επεξεργαστεί ταυτοχρόνως πολλές πληροφορίες –τουλάχιστον συνειδητά– και ότι στην εποχή μας οι πληροφορίες που το άτομο καλείται να διαχειριστεί, διακινούνται σε τεράστιες ποσότητες και με τεράστιες ταχύτητες, η γνωστική απόδοση των ανθρώπων μειώνεται και ο χρόνος ανταπόκρισής τους στα εξωτερικά γεγονότα επιβραδύνεται (Ντάβου, 2000). Οι εταιρείες ΤΝ το αντιλαμβάνονται αυτό ως έναν κερδοφόρο ορίζοντα· οι ρομποτικοί βοηθοί θα αναπληρώνουν αυτή την εγγενή «αδυναμία» μας, καθώς θα αποθηκεύουν και θα διαχειρίζονται αναμνήσεις των ιδιοκτητών τους που αλλιώς θα ξεχνιόνταν (O' Hara et al., 2006), ώστε, όποτε χρειάζεται, να παρουσιάζουν όλα τα διαθέσιμα στοιχεία γύρω από ένα θέμα, βοηθώντας μας να παίρνουμε αποφάσεις. Βέβαια, ήδη οι άνθρωποι χρησιμοποιούν τους υπολογιστές και τα κινητά τους τηλέφωνα ως εξωτερικά συστήματα μνήμης και φαίνεται να δημιουργείται μια συμβιωτική σχέση ανθρώπου-μηχανών, όσον αφορά τη διαχείριση πληροφοριών (Σιδηροπούλου, 2019, Sparrow, Liu & Wegner, 2011).

Τα ρομπότ θα ενισχύσουν αυτήν τη σχέση, καθώς θα εί-

ναι σε θέση να καταγράφουν όχι μόνο αυτό που επιλέγει ο ιδιοκτήτης τους, αλλά και όσα τα ίδια θα αποθηκεύουν αναφορικά με τις συναισθηματικές του αντιδράσεις, τις συζητήσεις με άλλους ανθρώπους, τους μονολόγους του, τις συνήθειές του και στο εγγύς μέλλον τα βιομετρικά του στοιχεία (π.χ. παλμοί, πίεση), που θα συνδυάζονται με τις παραπάνω καταστάσεις της καθημερινότητας. Αναμνήσεις που φυσιολογικά θολώνουν και χάνονται στον ασυνείδητο ψυχικό μας χώρο θα είναι πλέον δυνάμει πάντα διαθέσιμες. Ωστόσο, η δυνατότητα αποκάλυψης στοιχείων του ασυνειδήτου μας που θα παρέχουν τα ρομπότ, ίσως αποδειχθεί «δίκοπο μαχαίρι», εκτός κι αν αυτά μπορούν να διαχειριστούν θεραπευτικά τις πιθανές αναδύσεις τραυματικών αναμνήσεων. Ήδη προωθούνται στην αγορά υπολογιστικά προγράμματα που προσφέρουν υπηρεσίες γνωστικο-συμπεριφοριστικής θεραπείας, με κάποιους να ισχυρίζονται την ύπαρξη θετικών αποτελεσμάτων (Fitzpatrick, Darcy & Vierhile, 2017). Είναι αναπόφευκτο να προβληματιστούμε σε σχέση με την αποτελεσματικότητα θεραπευτικών προσεγγίσεων από μηχανές που δεν διαθέτουν ενσυναίσθηση, ώστε να διαχειριστούν το αναδυόμενο συναίσθημα των «πελατών» τους και τις ποιοτικές παραμέτρους των αφηγήσεών τους.

Μία ακόμα ανθρώπινη ανάγκη την οποία τα κοινωνικά ρομπότ υπόσχονται να καλύψουν, αφορά την όλο και αυξανόμενη αίσθηση μοναξιάς που βιώνουν οι άνθρωποι. Στο Ηνωμένο Βασίλειο, το 2017 περισσότεροι από 9.000.000 άνθρωποι δήλωσαν ότι βιώνουν μοναξιά,[3] ενώ σύμφωνα με πρόσφατη έρευνα της Ψυχολογικής Εταιρείας της Αυστραλίας, το 25% των Αυστραλών νιώθει συνέχεια μοναξιά και πάνω από τους

[3] https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/active-communities/rb_dec17_jocox_commission_finalreport.pdf

μισούς αρκετά συχνά.[4] Επιπλέον, στις ΗΠΑ, η γενιά της ψηφιακής εποχής, ακριβώς αυτή που με την αέναη ψηφιακή συνδεσιμότητά της αποφεύγει να ζήσει έστω και λίγες στιγμές μοναχικότητας, δηλώνει πιο μόνη από κάθε προηγούμενη (Konrath, O'Brien & Hsing, 2011· Twenge, 2017), ενώ ήδη εδώ και δύο δεκαετίες κοινωνικά ρομπότ ή ρομποτικά ζωάκια χρησιμοποιούνται στην Ιαπωνία για να κρατούν συντροφιά σε ανθρώπους σε οίκους ευγηρίας.

Με βάση ότι η επιθυμία μας για κοινωνικές συνδέσεις και διαπροσωπική κατανόηση είναι το πιο χαρακτηριστικό στοιχείο του είδους Sapiens (Christakis, 2019), η αυξανόμενη αίσθηση μοναξιάς ανοίγει περισσότερο τον δρόμο για σχέσεις με μηχανές που δίνουν την εντύπωση ότι νοιάζονται για εμάς. Αν επιπλέον σκεφτούμε ότι μια φιλία βοηθά τα άτομα να κατανοήσουν τον εαυτό τους, κυρίως μέσα από την αυτο-αποκάλυψη και την αφήγηση του εαυτού (Χρηστάκης & Χαλάτσης, 2010), μια μηχανή που «ακούει» τον άνθρωπο, του δίνει στοιχειώδη ανατροφοδότηση και του επιτρέπει ασυνείδητες μεταβιβάσεις, μπορεί αν μη τι άλλο να εκπληρώσει κάποιες από τις λειτουργίες των φιλικών σχέσεων. Άλλωστε, ένα ρομπότ με δυνατότητες μηχανικής μάθησης θα προσαρμόζεται στην προσωπικότητα των ιδιοκτητών του, με αποτέλεσμα αυτό να αποκτά ένα είδος «προσωπικότητας» διαφορετικής από άλλων ρομπότ με διαφορετικούς ιδιοκτήτες, εφόσον θα μπορούσε να είναι λιγότερο ή περισσότερο φλύαρο, να δίνει πολλή ή λίγη φροντίδα, να είναι υπάκουο ή αντιδραστικό κ.ο.κ.

Στη διαδικασία μιας τέτοιας προσαρμογής, μεγάλο πλεονέκτημα των συστημάτων ΤΝ είναι η ικανότητά τους για συνεχή συνδεσιμότητα και επικαιροποίηση (Harari, 2018). Για παράδειγμα, τα αυτο-οδηγούμενα αυτοκίνητα, που ίσως κάνουν επίσημα το ντεμπούτο τους στην Ολυμπιάδα του 2021 στο Τόκυο, θα είναι συνδεδεμένα μεταξύ τους, επικοινωνώντας ασταμάτητα τις απαραίτητες πληροφορίες, ώστε η οδήγηση να είναι ασφαλής. Κάτι ανάλογο μπορούμε να υποθέσουμε ότι θα ισχύσει και για τα ρομπότ, τα οποία θα αναζητούν μεταξύ τους πληροφορίες σχετικά με τρόπους αντίδρασης απέναντι στον ιδιοκτήτη τους, ανάλογα με την ψυχολογική του κατάσταση. Αν λοιπόν ορίσουμε ως πολιτισμό (culture) το σύνολο των ιδεών (και τεχνουργημάτων) που παράγονται από μια ομάδα και μεταδίδονται κοινωνικά, επηρεάζοντας την ατομική συμπεριφορά (Christakis, 2019), τότε ίσως θα αρχίσουμε σταδιακά να αναφερόμαστε ακόμα και σε έναν αντίστοιχο πολιτισμό (ή κουλτούρες) των ρομπότ.

[4] https://psychweek.org.au/wp/wp-content/uploads/2018/11/Psychology-Week-2018-Australian-Loneliness-Report.pdf

Με βάση τα παραπάνω, αντιλαμβανόμαστε πως οι προαναφερθείσες ανθρωπομορφικές μας τάσεις απέναντι στα ρομπότ θα βρουν εύφορο έδαφος και η ιδέα των συναισθηματικών ρομπότ θα αποκτήσει δυναμική.

Αλλαγές στην αντίληψη του τι σημαίνει «άνθρωπος»

Αρχικά, μια βασική αλλαγή στο πλαίσιο του μετανεωτερικού κόσμου ήταν ότι ο εαυτός έπαψε να νοείται ως μια σταθερή οντότητα και μετατράπηκε σε ένα αναστοχαστικό σχέδιο που υποστηρίχτηκε από μια πληθώρα πόρων (όπως η ψυχοθεραπεία και τα εγχειρίδια αυτοβοήθειας), που ωθούν σε έναν συνεχή προβληματισμό σχετικά με το «τι μπορεί να είμαι» αλλά και «τι μπορώ να γίνω» (Giddens, 1992). Καθώς μάλιστα μπήκαμε στην ψηφιακή επανάσταση, οι δυνατότητες για πειραματισμό με διαδικτυακές εικονικές ταυτότητες και με υβριδικούς εαυτούς (όπως τα άβαταρ), πολλαπλασιάστηκαν, οπότε τρόπον τινά ο «εαυτός» μπορεί όλο και περισσότερο να «συναρμολογηθεί» (Turkle, 2011). Τα Μέσα Κοινωνικής

Δικτύωσης ωθούν τους ανθρώπους να αυτο-προσδιορίζονται δημόσια, κατακερματίζοντας τον εαυτό τους σε ήδη προκαθορισμένες κατηγορίες (από το τι μουσική ακούν και τι βιβλία διαβάζουν, έως το ποια είναι τα ψυχολογικά και συναισθηματικά τους χαρακτηριστικά ή τα ιδανικά τους), με αποτέλεσμα ο εαυτός να «αντικειμενικοποιείται» ώστε να γίνεται κατανοητός αλλά και συμβατός με τους άλλους (Illouz, 2007).

Όπως τόνισε ο Baudrillard (1994), ανέκαθεν οι μορφές επικοινωνίας που χρησιμοποιούνταν σε μια κοινωνία, καθόριζαν τις ίδιες τις κοινωνικές σχέσεις των μελών της. Είτε πρόκειται για Μέσα Κοινωνικής Δικτύωσης (Davou & Sidiropoulou, 2017) είτε για κοινωνικά ρομπότ (Peltu & Wilks, 2010), αυτά τα νέα τεχνολογικά εργαλεία επηρεάζουν έντονα τους τρόπους με τους οποίους οι άνθρωποι αντιλαμβάνονται και βιώνουν τον εαυτό τους, τους άλλους και τις σχέσεις. Η ασώματη και πανταχού παρούσα ηλεκτρονική επικοινωνία δημιουργεί έναν ψηφιακό μεταβατικό χώρο (αντιστοίχως με την έννοια του μεταβατικού χώρου που εισήγαγε ο Winnicott, 1971), όπου φαντασία και πραγματικότητα περιπλέκονται και όπου κάθε είδους ατομικές προσδοκίες, επιθυμίες και φαντασιώσεις μπορούν να προβληθούν (Suler, 2016), δημιουργώντας έτσι ένα είδος «πολιτισμού της προσομοίωσης» (Turkle, 2010:30). Σε αυτόν τον μεταβατικό χώρο ακόμα και η επικοινωνία μας με μια μηχανή, όπως ένα κοινωνικό ρομπότ, μπορεί να μετατραπεί σε επικοινωνία με κάτι που, όπως σημειώνει η Turkle, (2011:18) είναι «αρκετά ζωντανό για σχέση».

Όμως δεν είναι μόνο οι επιδράσεις των νέων τεχνολογιών που δημιουργούν έναν νέο τρόπο αντίληψης του εαυτού, που θα ευνοήσει την αποδοχή των ρομπότ και τις συναισθηματικές συνδέσεις μαζί τους. Η ίδια η επιστήμη και ο τρόπος που προσεγγίζει τη διερεύνηση του ανθρώπινου νου, οδηγεί σε αντιστοιχίες με τον τρόπο που λειτουργούν τα υπολογιστικά συστήματα, δημιουργώντας έτσι σταδιακά την αναπαράσταση μιας ομοιότητας ανάμεσα σε ανθρώπους και μηχανές. Η επιστήμη πάντα συνήθιζε να μελετά τον ανθρώπινο νου ορμώμενη από τις κυρίαρχες τεχνολογικές καινοτομίες κάθε εποχής, για παράδειγμα άλλοτε σαν μηχανικό ρολόι, άλλοτε σαν ατμομηχανή, σαν τηλεφωνικό δίκτυο κ.ο.κ. (Kaku, 2014). Μετά το τέλος του Β΄ Παγκοσμίου Πολέμου, η επιστήμη στράφηκε στην αντιμετώπιση του ανθρώπινου εγκεφάλου και νου σαν κάποιου είδους ηλεκτρονικό υπολογιστή. Μια τέτοια προσέγγιση του ανθρώπινου νου αναπόφευκτα παραπέμπει στο πώς η ΤΝ αντιλαμβάνεται τον νου των «κοινωνικών ρομπότ».

Σχετικά, η Turkle (2011) παρατηρεί πως από τη στιγμή που η έννοια της νοημοσύνης άρχισε να χρησιμοποιείται στη ρομποτική, απέκτησε μια πιο μονοδιάστατη μορφή, που αναφέρεται μόνο στη γνωστική πλευρά του ανθρώπινου νου, ενώ ο Harari (2016:97) σημειώνει ότι η κυρίαρχη ιδέα της σύγχρονης βιολογίας συνοψίζεται στο ότι ο νους και το συναίσθημά μας λειτουργούν υπακούοντας σε «βιοχημικούς αλγόριθμους». Ο Harari προφανώς αναφέρεται στα οικουμενικά χαρακτηριστικά του ανθρώπινου συναισθήματος και νου, που κωδικοποιήθηκαν στα γονίδιά μας καθώς ευνοούσαν την επιβίωση και την αναπαραγωγή του είδους. Ωστόσο, η χρήση της έννοιας των βιοχημικών αλγορίθμων φέρνει τη βιολογία εγγύτερα στην επιστήμη των υπολογιστών, και προωθεί μια κοινωνική αναπαράσταση που πρεσβεύει ότι όπως ακριβώς η συμπεριφορά μιας μηχανής είναι προβλέψιμη, έτσι και η κατανόηση της λειτουργίας αυτών των βιοχημικών αλγορίθμων θα κάνει διαφανή και προβλέψιμη την ανθρώπινη συμπεριφορά.

Πράγματι, όπως κάποιοι νευροεπιστήμονες οραματίζονται, ίσως αρκεί να αποκρυπτογραφήσουμε τους εγκεφαλικούς μηχανισμούς που δημιουργούν, για παράδειγμα, τα συναισθήματα και στη συνέχεια να τους αναπαραγάγουμε αλγοριθμικά, δημιουργώντας έτσι συναισθήματα στην επερχόμενη γενιά ρομπότ (Pfaff, 2015). Στην ίδια λογική, το «Blue

Brain Project», ένα πρόγραμμα που ξεκίνησε το 2005 στο Πολυτεχνείο της Λωζάννης, με χρηματοδότηση από την ελβετική κυβέρνηση και την Ευρωπαϊκή Επιτροπή, έχει ως στόχο να δομήσει λεπτομερείς ψηφιακές προσομοιώσεις του εγκεφάλου των ποντικιών και εντέλει του ανθρώπου. Δικαίως, πολλοί ερευνητές και στοχαστές διαφωνούν με αυτήν την άποψη περί μελέτης του εγκεφάλου *εν κενώ*, εφόσον δεν λαμβάνει υπόψη τον ρόλο του φυσικού και κοινωνικο-πολιτισμικού περιβάλλοντος (Christakis, 2019, Sigman, 2017). Όπως σημειώνει ο Damasio[5], η έννοια του αλγόριθμου αγνοεί παντελώς την επίδραση του φυσικού υποστρώματος.

Ας διασαφηνίσουμε με ένα παράδειγμα: όταν ανακαλύφθηκε ο θετικός ρόλος της ορμόνης ωκυτοκίνης στη δόμηση συναισθηματικών δεσμών και στην ανάπτυξη της ενσυναίσθησης, η μηχανιστική προσέγγιση βιάστηκε να πιστέψει ότι η χορήγησή της στους ανθρώπους θα οδηγούσε σε έναν καλύτερο κόσμο. Όμως μετά την αρχική αισιοδοξία, στη συνέχεια φάνηκε ότι ατομικοί και διατομικοί παράγοντες έπαιζαν καθοριστικό ρόλο στον τρόπο απόκρισης στη λήψη ωκυτοκίνης. Σε κάποιες δε περιπτώσεις, η χορήγησή της οδηγούσε ακόμα και σε αρνητικές κοινωνικές συμπεριφορές, όπως φθόνο και κακόβουλη χαρά (Shamay-Tsoory, Fischer, Dvash, Harari, Perach-Bloom & Levkovitz, 2009). Διαπιστώθηκε λοιπόν ότι ο τρόπος ερμηνείας μιας κοινωνικής περίστασης ως ασφαλούς ή ανασφαλούς από τα άτομα στα οποία χορηγούνταν (Bartz, Zaki, Bolger N & Ochsner, 2011) ή οι προσδοκίες των οικείων τους σχετικά με τη δράση της (Guastella & Hickie, 2016) έπαιζαν καθοριστικό ρόλο στο αν η ωκυτοκίνη θα οδηγούσε ή όχι σε βελτίωση της κοινωνικής συμπεριφοράς.

Έγινε λοιπόν σαφές πως τα επίπεδα ωκυτοκίνης επηρεάζονται από τη συμπεριφορά των άλλων με τους οποίους αλληλεπιδρούμε (Christakis, 2019), είτε πρόκειται για ανθρώπους είτε για ζώα, όπως στην περίπτωση της αύξησης της ωκυτοκίνης που παρατηρείται όταν ένας άνθρωπος και ο σκύλος του κοιτάζονται στα μάτια (Nagasawaetal., 2015). Έτσι, οι μηχανιστικές προσεγγίσεις του ανθρώπου δείχνουν να είναι ένας αποπροσανατολιστικός αναγωγισμός· ο άνθρωπος δεν δρα, δεν σκέπτεται, δεν αισθάνεται εν κενώ, άρα δεν μπορεί και να μελετηθεί εν κενώ. Όπως σημειώνει η Ντάβου (2000), αντί οι γνωστικές διεργασίες να προσεγγιστούν μελετώντας συνδυαστικά τη βιολογική, τη συγκινησιακή και την κοινωνική φύση του ανθρώπου, αυτές αντιμετωπίζονται μηχανιστικά, δίνοντας βαρύτητα στη μελέτη των μηχανισμών επεξεργασίας της πληροφορίας εκτός πλαισίου και αγνοώντας το προσωπικό νόημα και τη σημασία που κάθε άτομο αποδίδει στις εμπειρίες του.

Παράλληλα με την κυρίαρχη αντίληψη της βιολογίας ότι ο άνθρωπος είναι αποτέλεσμα της δράσης βιοχημικών αλγορίθμων, η επιστήμη της πληροφορικής μετά την εποχή του Alan Turing προώθησε την ιδέα ότι τα πάντα, από ένα μουσικό κομμάτι έως έναν ιό, είναι πληροφοριακά δεδομένα. Αυτή η προσέγγιση, σύμφωνα με τον Harari (2016), σταδιακά δημιουργεί ένα είδος νέας «θρησκείας», την οποία αποκαλεί νταταϊσμό (dataism). Ο νταταϊσμός, συνδυάζοντας τη βιολογία και την πληροφορική, αντιλαμβάνεται τον άνθρωπο σαν ένα σύνολο πληροφοριακών δεδομένων, οπότε επαγωγικά απορρίπτει την οντολογική απόσταση ανάμεσα στα ζώα και τις μηχανές. Αυτή η θεώρηση ίσως οδηγεί την ανθρωπότητα σε μια νέα περίοδο αυτο-κατανόησης. Μετά τη συνειδητοποίηση ότι δεν είμαστε το κέντρο του σύμπαντος (Κοπέρνικος), ότι δεν διαφέρουμε από τα άλλα ζώα (Δαρβίνος), κι ότι δεν είμαστε διάφανοι στον εαυτό μας (Φρόιντ), στο παρόν είναι πιθανόν να αρχίσουμε ασυνείδητα να δεχόμαστε την ιδέα ότι είμαστε ένα οργανωμένο σύνολο βιοχημικών δεδομένων (Floridi, 2010). Όμως είμαστε πράγματι; Ή μήπως πρόκειται για μια μονόπλευρη

[5] https://www.huffpost.com/entry/algorithmic-human-life_b_10699712

αντίληψη της ανθρώπινης πολυπλοκότητας, η οποία ευνοεί ιδιαίτερες οικονομικές και πολιτικές σκοπιμότητες;

Ας δούμε σχετικά την όλο και αυξανόμενη αγορά γκάτζετ «αυτογνωσίας μέσω των αριθμών» που προωθούν την ιδέα του «ποσοτικοποιημένου εαυτού».[6] Πρόκειται για εργαλεία που επιτρέπουν μετρήσεις της φυσικής δραστηριότητας, της ποιότητας του ύπνου, της διατροφικής ποιότητας, της στάσης του σώματος, της δραστηριότητας των εγκεφαλικών κυμάτων κ.ο.κ., με στόχο τη βελτιστοποίηση της σωματικής, νοητικής και συναισθηματικής μας κατάστασης. Πέρα όμως από την αμφιλεγόμενη επιστημονική εγκυρότητα μιας ιδέας που υπόσχεται καλύτερη υγεία, αγνοώντας παντελώς κάθε ψυχολογικό παράγοντα, είναι εύκολο να διαπιστώσουμε τη διαπλοκή συμφερόντων. Αφενός τα δεδομένα των γκάτζετ αυτών είναι ευνοήτως στη διάθεση κάθε ενδιαφερόμενου χωρίς τη συγκατάθεση των χρηστών, αφετέρου ήδη μία από τις μεγαλύτερες ασφαλιστικές εταιρείες των ΗΠΑ ανακοίνωσε την παροχή οικονομικών κινήτρων σε πελάτες που αποδέχονται την επίσημη πρόσβαση της εταιρείας στα προσωπικά δεδομένα των σχετικών ψηφιακών συσκευών της.[7] Ίσως δεν είμαστε μακριά από τη δημιουργία προδιαγραφών υγιεινών πρακτικών, στις οποίες οι άνθρωποι θα υποχρεώνονται να ανταποκρίνονται, αν θέλουν να καλύπτονται από το ασφαλιστικό σύστημα.

Ρομπότ με συνειδητότητα

Σημειώσαμε στην εισαγωγή πως το Ιερό Δισκοπότηρο της ΤΝ είναι η κατασκευή ρομπότ με συνειδητότητα, κάτι που μεταξύ άλλων θα προσέδιδε το στοιχείο της αμοιβαιότητας στις σχέσεις των ανθρώπων μαζί τους. Για να μπορέσουμε

[6] https://quantifiedself.com/about/what-is-quantified-self/
[7] https://www.bbc.com/news/technology-45590293

να δημιουργήσουμε μηχανές με συνειδητότητα, θα πρέπει πρώτα να συμφωνήσουμε στο τι είναι η συνειδητότητα, και ακολούθως στο πώς προκύπτει, τουλάχιστον στον άνθρωπο.

Ως προς το πρώτο εξακολουθεί να μην υπάρχει ένας κοινά αποδεκτός ορισμός (Dietrich, 2007). Οι ορισμοί της συνειδητότητας περιλαμβάνουν έννοιες όπως αυτοεπίγνωση, ενσυναίσθηση, υποκειμενικότητα, ικανότητα σύνδεσης των πληροφοριών σε ένα αφήγημα, ικανότητα αντίληψης της συνέχειας του χρόνου, ικανότητα αντίληψης αισθητηριακών ερεθισμάτων, δυνατότητα επιλογής, ικανότητα εικασιών σχετικά με τη νοητική και συναισθηματική κατάσταση των άλλων (θεωρία του νου) κ.ο.κ. Είναι εμφανές ότι όλες οι παραπάνω έννοιες προϋποθέτουν την ικανότητα αντίληψης μιας ατομικότητας που αντιλαμβάνεται τη διαφορά της από τους άλλους και το περιβάλλον της (Dennett, 1991).

Ως προς το δεύτερο, δηλαδή το πώς προκύπτει η συνειδητότητα, μέχρι στιγμής δεν υπάρχει καμιά επιστημονική απάντηση. Ενώ υπάρχουν δεδομένα που δείχνουν μια στενή σχέση ανάμεσα στον συνειδητό νου και τον εγκέφαλο, η φύση αυτής της σχέσης παραμένει άγνωστη (Tononi & Koch, 2015). Συγχρόνως, πρέπει να απαντήσουμε σε πολλά ερωτήματα, όπως το αν τα βρέφη ή κάποια ζώα έχουν συνειδητότητα, τι συμβαίνει με τη συνειδητότητα όταν είμαστε σε βαθύ ύπνο, αν υπάρχει συνειδητότητα όταν κάποιος ονειρεύεται ή βρίσκεται σε κώμα κ.ο.κ.

Αναγωγικές προσεγγίσεις της Τεχνητής Νοημοσύνης

Ένας κλάδος της ΤΝ με βάση τον καρτεσιανό άυλο νου –αλλά και διεθνή επιστημονικά προγράμματα όπως το προαναφερθέν «Blue Brain Project»– υπαινίσσεται ότι αν αποκωδικοποιήσουμε το πώς λειτουργεί ο ανθρώπινος νους και το συναίσθημα, θα μπορέσουμε να δημιουργήσουμε άυλα λογισμικά

που θα σκέφτονται και θα νιώθουν (Haugeland, 1985). Όμως η σύγχρονη νευροεπιστήμη έχει καταδείξει πως το σώμα είναι όργανο σκέψης και συναισθήματος και ότι εγκέφαλος και σώμα βρίσκονται σε συνεχή αλληλεπίδραση, μέσα από την οποία παράγεται η πραγματικότητα που βιώνουμε (Damasio, 1994) – άρα, όπως εύστοχα σημειώνει η Turkle (2011), αν θέλουμε να φτιάξουμε ενσυνείδητες και ενσυναίσθητες μηχανές, θα πρέπει πρώτα να τους δώσουμε ένα σώμα. Τα παραπάνω σημαίνουν ότι ο νους και η συνειδητότητα προκύπτουν από την ύλη; Όχι απαραίτητα, αλλά καθώς μόνο η ύλη είναι επιστημονικά διαθέσιμη για διερεύνηση, οι νευροεπιστήμες επικεντρώθηκαν στη μελέτη του εγκεφάλου.

Οι κυρίαρχες επιστημονικές προσεγγίσεις σήμερα προσπαθούν να λύσουν το αίνιγμα της συνειδητότητας διερευνώντας τις νευρωνικές της παραμέτρους και διαδικασίες, όπως αυτές αναπαριστώνται στον εγκέφαλο (Chalmers, 2000) – δηλαδή ποια δίκτυα νευρώνων κινητοποιούνται και συνεργάζονται όταν βρισκόμαστε εν συνειδήσει, όταν βλέπουμε, σκεφτόμαστε, νιώθουμε ή επιλέγουμε κάτι, ή και αντίστροφα, ποια συνειδητή εμπειρία θα έχει ένα άτομο όταν διεγερθούν συγκεκριμένα δίκτυα νευρώνων. Η προσέγγιση αυτή έχει κάνει θαυμαστές προόδους, όμως ένα πρόβλημα που ανακύπτει είναι ότι οι ερευνητές συνήθως ταυτίζουν τα τεκταινόμενα των νευρωνικών διαδικασιών με την ίδια την υποκειμενική εμπειρία (Crick & Koch, 1998). Για παράδειγμα, η ενεργοποίηση νευρωνικών δικτύων στον αμφιβληστροειδή χιτώνα, όταν κάποιος συνειδητοποιεί ότι βλέπει κάτι, δεν σημαίνει απαραίτητα ότι αυτά τα δίκτυα προκαλούν αυτή τη συνειδητοποίηση· μπορεί να έχεις μια εικόνα στον αμφιβληστροειδή σου χωρίς να τη συνειδητοποιείς, αλλά και μπορεί να βλέπεις μια εικόνα ενώ ονειρεύεσαι, χωρίς αυτή να υπάρχει στον αμφιβληστροειδή σου (Koch, 2004).

Άρα, μήπως δεν αρκεί η κατανόηση των νευρωνικών διαδικασιών για να προσεγγίσουμε το θέμα της συνειδητότητας; Πολλοί υποστηρίζουν ότι εφόσον ψυχεδελικές ουσίες (π.χ. LSD) και βιοχημικές ή δομικές δυσλειτουργίες του εγκεφάλου (π.χ. άτομα με διηρημένο εγκέφαλο) μπορούν να παραγάγουν μια άλλη κατάσταση συνειδητότητας, αυτό είναι μια απόδειξη ότι η συνειδητότητα είναι δημιούργημα νευρωνικών διαδικασιών. Ήδη, 70 χρόνια νωρίτερα, ο φιλόσοφος Gilbert Ryle (1949) σημείωνε ότι «νους» είναι η ίδια η εγκεφαλική λειτουργία κι όχι κάποιου είδους «φάντασμα μέσα στη μηχανή». Ωστόσο, μια τέτοια προσέγγιση είναι σαν να δεχόμαστε ότι, για παράδειγμα, το ραδιόφωνο παράγει τη μουσική που ακούμε από αυτό, οπότε κατ' αντιστοιχία μπορούμε να αναρωτηθούμε μήπως ο εγκέφαλος δεν είναι ο δημιουργός αλλά το εργαλείο έκφρασης της συνειδητότητας. Επ' αυτού, ο Chalmers (1996) πρότεινε ότι ο τρόπος προσέγγισης της συνειδητότητας από τις νευροεπιστήμες μπορεί να λύσει το *εύκολο πρόβλημα* (πώς ο εγκέφαλος επεξεργάζεται τις πληροφορίες), αλλά όχι το *δύσκολο πρόβλημα* της συνειδητότητας (πώς δηλαδή προκύπτει η υποκειμενική εμπειρία).

Όπως εύστοχα έθεσε τον προβληματισμό και ο φιλόσοφος Thomas Nagel (1974), όσο και να κατανοήσουμε τις νευροβιολογικές εγκεφαλικές διαδικασίες μιας νυχτερίδας, ποτέ δεν θα μάθουμε τη φαινομενική της συνείδηση –πώς δηλαδή η ίδια βιώνει τον κόσμο–, εφόσον ο «κόσμος» της είναι εντελώς διαφορετικός από τον δικό μας. Κάτι αντίστοιχο συμβαίνει και μεταξύ των ανθρώπων· όταν, δηλαδή, πίνουμε ένα ποτήρι κρασί, ακούμε ένα τραγούδι, νιώθουμε έναν πόνο κ.ο.κ., δεν έχουμε όλοι την ίδια υποκειμενική εμπειρία και δεν υπάρχει, προς το παρόν, τρόπος αυτή η εμπειρία να περιγραφεί αντικειμενικά ή να μετρηθεί.

Μια επιστημονικά αποδεκτή ερμηνεία είναι πως το διαφορετικό βιολογικό, ψυχολογικό, κοινωνικό και πολιτισμικό

μας υπόβαθρο, οδηγεί σε διαφορετικές υποκειμενικές εμπειρίες. Όμως αυτό μας επαναφέρει στο *δύσκολο πρόβλημα* της συνειδητότητας του Chalmers (1996) – δηλαδή, με ποιον ακριβώς τρόπο ένα παρόμοιο ερέθισμα μετουσιώνεται σε μια διαφορετική υποκειμενική εμπειρία. Η ΤΝ μπορεί να αναγνωρίσει ένα δέντρο, αλλά δεν έχει υποκειμενική εμπειρία του δέντρου· το δέντρο για την ΤΝ είναι μια ομάδα pixel που με βάση τα δεδομένα που διαθέτει και τα εξελιγμένα νευρωνικά της δίκτυα, μπορεί να τα ερμηνεύσει ως «δέντρο». Στην ταινία «Blade Runner», όταν ο κεντρικός ήρωας Rick Deckard ανακαλύπτει ότι η Rachel δεν είναι άνθρωπος αλλά ανδροειδές, αναρωτιέται «πώς μπορεί να ξέρει τι είναι;», ακριβώς όπως ο Heidegger (1962) αναρωτήθηκε τι είναι αυτό που μας κάνει να σκεφτόμαστε κάτι συγκεκριμένο εδώ και τώρα, δηλαδή πώς γίνεται ένα νοήμον ον να νοεί κάτι ως τέτοιο και όχι ως κάτι άλλο.

Σχετικά με το θέμα της δυνατότητας δημιουργίας ενσυνείδητων μηχανών, τα πράγματα περιπλέκονται περαιτέρω, καθώς τα παραπάνω μας οδηγούν αναπόφευκτα στον ρόλο του ασυνείδητου· είναι πια ευρέως αποδεκτό ότι μεγάλο μέρος της διαδικασίας μέσω της οποίας αντιλαμβανόμαστε τον κόσμο, πραγματοποιείται σε ασυνείδητο επίπεδο. Τα δεδομένα δείχνουν ότι από τα 10^7bits πληροφορίας που εισέρχονται στον εγκέφαλό μας ανά δευτερόλεπτο, συνειδητοποιούμε το πολύ 50 (Tegmark, 2017). Η συνειδητή επεξεργασία πληροφοριών είναι κάτι σαν τον Γενικό Διευθυντή του νου, σε αντιστοιχία με αυτόν μιας εταιρείας· αντιλαμβάνεται τη γενική κατάσταση και παίρνει αποφάσεις χωρίς να είναι ενήμερος για τις λεπτομέρειες των δραστηριοτήτων στα δεκάδες τμήματα της επιχείρησης (Kaku, 2014). Με απλά λόγια, η υποκειμενική εμπειρία ενός τριαντάφυλλου δημιουργείται κυρίως μέσα από ασυνείδητες διεργασίες, διαφορετικές για τον καθέναν (αν και σε μεγάλο βαθμό κοινωνικο-πολιτισμικά καθορισμένες, καθώς προκύπτουν από τον συλλογικό διαμοιρασμό ατομικών εμπειριών).

Η συνειδητότητα ως πληροφορία και η Τεχνητή Νοημοσύνη

Σχετικά πρόσφατα, σημαντικοί ερευνητές της συνειδητότητας (Tononi, 2004· Tononi & Koch, 2015) αντέστρεψαν τον τρόπο προσέγγισης του θέματος. Προτείνουν ότι το να προσπαθούμε να βρούμε πώς προκύπτει η συνειδητότητα από τις διεργασίες του εγκεφάλου, είναι σαν να επιχειρούμε να αποστάξουμε τον νου από την ύλη, κάτι που ίσως αποδειχτεί αδύνατον – οπότε γιατί να μην πάμε αντίστροφα, επιχειρώντας να ανακαλύψουμε ποιες είναι οι απαραίτητες προϋποθέσεις ώστε ένα υλικό σύστημα να έχει έναν ελάχιστο ή μεγαλύτερο βαθμό συνειδητότητας; Έτσι, η θεωρία της Ενοποίησης των Πληροφοριών (Tononi, 2004), στηριζόμενη εν μέρει σε ιδέες πανψυχισμού, υποστηρίζει ότι η συνειδητότητα είναι μια θεμελιώδης ιδιότητα *φυσικών συστημάτων* των οποίων τα μέρη είναι σε θέση να αλληλεπιδρούν, να ανταλλάσσουν πληροφορίες και να βρίσκονται σε σχέσεις αιτίου-αιτιατού μεταξύ τους. Τα ευρήματα του Tononi με ασθενείς σε κώμα, σε αναισθησία, σε βαθύ ύπνο ή υπό την επήρεια ουσιών, υποστηρίζουν αρκετά την πολλά υποσχόμενη θεωρία του. Επιπρόσθετα, η θεωρία δείχνει ότι η συνειδητότητα είναι διαβαθμισμένη ιδιότητα και ότι κάποιος βαθμός συνειδητότητας υπάρχει και σε μονοκύτταρους οργανισμούς. Ωστόσο, προβλέπει ότι μια μηχανή δεν θα αποκτήσει ποτέ συνειδητότητα, αν δεν υπάρχει δυνατότητα αμφίδρομης ανταλλαγής πληροφοριών αλλά και σχέσης αιτίου-αιτιατού μεταξύ των μερών της (Tononi & Koch, ό.π.).

Σε σχετικά ίδιο μήκος κύματος, επιστήμονες της ΤΝ από τον χώρο της Φυσικής, πιστεύουν ότι ακόμα κι αν ανακα-

λύψουμε τις νευρωνικές παραμέτρους της συνειδητότητας, ποτέ δεν θα μπορέσουμε να τις μεταφέρουμε σε μηχανές ΤΝ, εφόσον το «σώμα» και ο «εγκέφαλος» που θα διαθέτουν θα απέχουν πολύ από την πολυπλοκότητα των αντίστοιχων ανθρώπινων. Έτσι, προσανατολίζονται και αυτοί στη διερεύνηση των προϋποθέσεων που καθιστούν ένα φυσικό σύστημα στοιχειωδώς συνειδητό. Θεωρούν λοιπόν ότι η ανάδυση της συνειδητότητας από την ύλη σχετίζεται με κάποια συγκεκριμένα μοτίβα διάταξης και κίνησης στον χώρο των υποατομικών σωματιδίων αυτής της ύλης (Tegmark, 2017). Σε αυτό το πλαίσιο, ο Tegmark προτείνει την απλοποίηση του ορισμού της συνειδητότητας, ώστε να μπορεί να περιλάβει και μηχανικά συστήματα ΤΝ· την ορίζει ως τη δυνατότητα Υποκειμενικής Εμπειρίας και υποστηρίζει ότι αν δημιουργήσουμε ένα τεχνητό σύστημα με μεγάλες δυνατότητες αποθήκευσης και επεξεργασίας πληροφοριών, και με τα διάφορα υποσυστήματά του να έχουν την ικανότητα πολύπλευρης ανταλλαγής πληροφοριών λειτουργώντας ως ένα αυτόνομο σύνολο, τότε υπάρχουν πολλές πιθανότητες αυτό το σύστημα να έχει ένα είδος συνειδητότητας.

Ωστόσο, όπως σημειώνει ο Tegmark (2017), πολλοί θεωρούν την ενασχόληση με το θέμα της συνειδητότητας των ρομπότ σαν μια απέλπιδα προσπάθεια να λύσουμε ένα μη επιστημονικό πρόβλημα, οπότε δεν αξίζει να ασχολούμαστε. Ο Harari (2016, σ. 453) συνοψίζει αυτή τη θεώρηση λέγοντας «... αν καταφέρουμε να αντικαταστήσουμε όχι μόνο τους οδηγούς ταξί και τους γιατρούς αλλά και τους δικηγόρους, τους ποιητές και τους μουσικούς με ανώτερα υπολογιστικά προγράμματα, γιατί θα έπρεπε να ανησυχούμε εάν αυτά τα προγράμματα δεν έχουν συνειδητότητα και υποκειμενική εμπειρία;»

Συζήτηση

Η ανθρώπινη φαντασία είχε προ πολλού συλλάβει την ιδέα ανθρώπων-μηχανών· ο χάλκινος μυθικός φύλακας της Μινωικής Κρήτης Τάλως, το εβραϊκό golem, οι μικροσκοπικοί «υπηρέτες» karakuri της περιόδου Edo στην Ιαπωνία, είναι μερικά παραδείγματα που δείχνουν την επιθυμία των ανθρώπων για δημιουργία όντων «κατ' εικόνα και καθ' ομοίωσιν» (Mayor, 2018). Τον 19ο αιώνα η Mary Shelley επανέφερε την ιδέα της τεχνητής ζωής με τον Frankenstein, ενώ στον 20ό αιώνα η σχετική με τα ρομπότ μυθιστοριογραφία και φιλμογραφία έθεσε προβληματισμούς αναφορικά με το κατά πόσο αυτές οι μηχανές θα μπορούσαν να αποκτήσουν ανθρώπινες ιδιότητες και χαρακτηριστικά. Στις αρχές του 21ου αιώνα, η ΤΝ είναι πλέον μέρος της ζωής μας, ενώ η ρομποτική διδάσκεται ήδη σε πολλά δημοτικά σχολεία.

Στο μεταξύ οι τεχνολογικές εξελίξεις οδήγησαν στη δημιουργία των πρώτων *κοινωνικών ρομπότ*, τα οποία καταφέρνουν να προσομοιώσουν ενδιαφέρον και νοιάξιμο προς τους ιδιοκτήτες τους, με αποτέλεσμα πολλοί από αυτούς να τα προικίζουν με ένα ιδιαίτερο νόημα και να δένονται συναισθηματικά μαζί τους. Οι δημιουργοί τους και οι εταιρείες που τα πουλούν ισχυρίζονται ότι θα μας συμπληρώνουν και θα κάνουν τη ζωή μας ευκολότερη και λιγότερο μοναχική. Φαίνεται πως τα ρομπότ στο εγγύς μέλλον θα αποτελέσουν μέρος της ζωής εκατομμυρίων ανθρώπων. Εντούτοις, είναι μάλλον ασαφές το πώς θα διαχειριστούμε αυτές τις τεχνολογικές καινοτομίες και πώς αυτές θα επηρεάσουν τον τρόπο που αντιλαμβανόμαστε τον εαυτό μας και τις σχέσεις.

Με βάση τα δεδομένα που παρατέθηκαν, ανακύπτουν σοβαροί προβληματισμοί. Ένα θετικό είναι ότι η προσπάθεια για δημιουργία εξελιγμένης τεχνητής νοημοσύνης και ευφυ-

ών ρομπότ εντείνει τις ερευνητικές απόπειρες κατανόησης του ανθρώπινου νου, όμως οι μέχρι τώρα πολύ αξιόλογες προσπάθειες δεν έχουν καταφέρει να λύσουν το αίνιγμα της λειτουργίας του, περιοριζόμενες στη διερεύνηση των διαδικασιών και των νευρωνικών παραμέτρων μέσω των οποίων αντιλαμβανόμαστε τον κόσμο και επεξεργαζόμαστε πληροφορίες. Φυσικά δεν αποκλείεται μελλοντικά η όλο και πλουσιότερη γνώση που αποκομίζουν οι νευροεπιστήμονες, να επιτρέψει την αποκωδικοποίηση της λειτουργίας του νου και της συνειδητότητας, όμως προς το παρόν φαίνεται να απέχουμε πολύ από αυτήν την προοπτική. Η αρνητική συνέπεια αυτής της αδυναμίας κατανόησης είναι ότι οδηγεί το κυρίαρχο ρεύμα των νευροεπιστημών να ανάγει την πολυπλοκότητα του ανθρώπινου νου σε μετρήσιμες μόνο ιδιότητες, που θα μπορούσαν να «μεταφερθούν» σε μηχανές. Ο νους γίνεται αντιληπτός με μηχανιστικούς όρους, οπότε ο υπαινιγμός που επαγωγικά προκύπτει είναι ότι τελικά οι άνθρωποι σκεφτόμαστε σαν μηχανές, οπότε αρκεί να αφήσουμε τον έλεγχο των αποφάσεών μας και γενικά της ύπαρξής μας στη βιοτεχνολογία και την πληροφορική, μια και αυτές μας κατανοούν καλύτερα από όσο εμείς τον εαυτό μας (Harari, 2018). Δεν είναι δύσκολο λοιπόν να φανταστούμε πώς η αποδοχή μιας τέτοιας ιδέας ανοίγει περαιτέρω τον δρόμο στον έλεγχο της σκέψης και της συμπεριφοράς μας, εντείνοντας τους ήδη υπάρχοντες τρόπους και μηχανισμούς ελέγχου (π.χ. θρησκεία, ΜΜΕ, κοινωνική επιρροή και κάθε είδους προπαγάνδα).

Επιπλέον, η αναγωγιστική αυτή αναπαράσταση ενός νου μαθηματικά μετρήσιμου και άρα προβλέψιμου οδηγεί στην παραγνώριση ανθρώπινων ιδιοτήτων που μπορεί να μην είναι ακόμα επιστημονικά κατανοητές, αλλά εντούτοις είναι χρήσιμες και λειτουργικές. Για παράδειγμα, έχει δειχθεί ο ιδιαίτερος ρόλος της διαίσθησης στη λήψη αποφάσεων σχετικά με πολύπλοκα θέματα (Dijksterhuis, Bos, Nordgren, &

van Baaren 2006· Sigman, 2017). Μια διαισθητική απόφαση είναι μια απόφαση στηριγμένη στη γνώση του ασυνείδητου. Το γεγονός ότι δεν κατανοούμε επιστημονικά πώς λειτουργεί η διαίσθηση, δεν θα έπρεπε να αποθαρρύνει την αποδοχή της ούτε να μας απομακρύνει από την ερευνητική της προσέγγιση. Αν αφεθούμε στον μετρήσιμο ορθολογισμό των μηχανών, μήπως κινδυνεύουμε να χάσουμε τελικά αυτή την ικανότητα; Μήπως τελικά η εναρμόνιση με τον τρόπο λειτουργίας τους περιορίσει σταδιακά τον καθοριστικό ρόλο των συναισθημάτων, της φαντασίας, της δημιουργικής σκέψης, του χιούμορ ή της διορατικότητας στην ύπαρξή μας;

Ένα ακόμα θέμα προβληματισμού είναι το πώς θα αντιδράσουν συναισθηματικά οι άνθρωποι στην προσομοίωση συναισθηματικής ανταπόκρισης που τα καλοσχεδιασμένα κοινωνικά ρομπότ μπορούν ήδη να προσφέρουν. Θα παραδοθούν στην ανθρωπομορφική τάση που μας χαρακτηρίζει και θα «βολευτούν» με μονόπλευρες φαντασιωτικές σχέσεις; Αβίαστα εδώ προκύπτει το «γιατί να γίνουμε φίλοι με ρομπότ, όταν γύρω μας υπάρχουν άνθρωποι». Είναι ευνόητο να υποθέσουμε ότι για να φερθεί ένα ρομπότ υποστηρικτικά ή φιλικά, θα πρέπει να είναι «φτιαγμένο στα μέτρα του πελάτη», δηλαδή να μαθαίνει συνεχώς την προσωπικότητα του ιδιοκτήτη του (κάτι που υπόσχεται η ΤΝ της αυτο-ενισχυτικής αυτόνομης μάθησης). Ποιος θα ορίζει όμως τις βασικές παραμέτρους ώστε να ταιριάζει με τον κάτοχό του; Θα αποφασίζουμε εμείς οι ίδιοι πώς θέλουμε να είναι οι τεχνητοί μας «φίλοι»; Τι άλλο δείχνει αυτό παρά μια επιθυμία για σχέσεις που θα ελέγχουμε απόλυτα – και πώς κάτι τέτοιο δεν θα οδηγήσει σταδιακά στην απαλοιφή των τόσο δημιουργικών ανθρώπινων συναισθηματικών συγκρούσεων και τελικά στην αναπόδραστη πλήξη μιας προβλέψιμης οιονεί σχέσης;

Όμως, ό,τι κι αν συμβεί και όπως και να εξελιχθούν οι δυνατότητες των τεχνητών συντρόφων, το «συναισθηματικό»

τους εύρος θα απέχει πολύ από το ανθρώπινο, οπότε αναπόφευκτα η εξοικείωση σε συναισθηματικές ανταλλαγές και σχέσεις με ρομπότ θα υποβιβάσει σταδιακά τις προσδοκίες μας για την ποιότητα των σχέσεών μας με ανθρώπους (Turkle, 2011) και επαγωγικά τις προσπάθειές μας για κατανόηση των άλλων. Ωστόσο, αυτή ακριβώς η επένδυση στη σχέση και στις δυσκολίες που συνεπάγεται, είναι που σηματοδοτεί τη διαδικασία κατανόησης του ίδιου του εαυτού μας, εφόσον υπάρχουμε και αποκτάμε αυτογνωσία μόνο μέσω των άλλων. Τι θα συμβεί άραγε όταν αυτοί οι «τεχνητοί άλλοι» δεν θα είναι παρά αντανακλάσεις των αναγκών και των εξιδανικεύσεών μας; Δεν είναι ορατός ο κίνδυνος οπισθοχώρησης κάθε έννοιας συλλογικότητας; Είναι προφανής εδώ η αντιφατικότητα της σύγχρονης τεχνολογίας· αφενός υποστηρίζει ότι η συνεχής συνδεσιμότητα και η σχεδιαζόμενη διασύνδεση ανθρώπου - μηχανής θα μας παράσχει πρόσβαση σε έναν ωκεανό παγκόσμιας σκέψης (άρα στην απόλυτη συλλογικότητα), αφετέρου προωθεί την ιδέα των πλεονεκτημάτων της επικοινωνίας με τα ρομπότ, ωθώντας μας σε μια «αυτιστικού» τύπου αυτο-επικοινωνία. Ήδη αρκετά δεδομένα (βλ. ενότητα 2.2) δείχνουν ότι η συλλογικότητα του διαδικτυακού «παγκόσμιου χωριού» τείνει να οδηγεί τη νέα γενιά σε μια αυξανόμενη ψυχολογική απομόνωση.

Σημειώσαμε επίσης ότι πολλοί ερευνητές θεωρούν πως ποτέ δεν θα υπάρξουν αμοιβαίες σχέσεις με ρομπότ, αν αυτά δεν αποκτήσουν συνειδητότητα. Για να επιτευχθεί αυτό, η επιστήμη πρέπει πρώτα να απαντήσει στο αν και πώς η ανθρώπινη συνειδητότητα προκύπτει από βιοχημικούς μηχανισμούς του εγκεφάλου, αν αποτελεί επιφαινόμενο της ανθρώπινης νοημοσύνης, μια αναδυόμενη δηλαδή ιδιότητά της, αν αποτελεί ένα υποπροϊόν της ροπής μας για την αφήγηση και της παρεπόμενης κοινωνικής κατασκευής της πραγματικότητας είτε εν τέλει αν προκύπτει από κάτι άλλο που προς το παρόν

δεν γνωρίζουμε[8]. Σε κάθε περίπτωση, η προσπάθεια δημιουργίας τεχνητών ενσυνείδητων ανθρώπων, πέρα από την επιστημονική πρόκληση που ανέκαθεν γοήτευε την ανθρώπινη ψυχοσύνθεση, φαίνεται να κρύβει και μια βαθιά επιθυμία υπέρβασης του θανάτου. Δεν μπορούμε να ξέρουμε αν θα επιτευχθεί, αλλά από μια άποψη ίσως αξίζει να δοκιμάσουμε, για το ίδιο το νόημα της ζωής. Αν πράγματι το σύμπαν είναι άσκοπο, χωρίς νόημα, όπως επεσήμανε ο νομπελίστας Steven Weinberg (1992), ίσως είμαστε εμείς, τα ενσυνείδητα όντα, που το νοηματοδοτούμε· σε έναν γήινο κόσμο όπου η βιολογική ζωή είναι πιθανό να εξαφανιστεί σε μερικές χιλιάδες ή εκατομμύρια χρόνια, η διατήρηση της συνειδητότητας μέσα από μηχανές, ίσως είναι ένας τρόπος να συνεχίσουμε να το κάνουμε - κι επιπλέον, τι άλλο είναι αυτό παρά μια εκδοχή αθανασίας;

Στον 21ο αιώνα, η τεχνολογία και η επιστήμη επιτρέπουν να δημιουργούμε μηχανές (και να οραματιζόμαστε όλο και πιο εξελιγμένες), που αφενός μπορούν να συνεισφέρουν στην ευημερία μας, αφετέρου μπορούν να οδηγήσουν στην απώλεια των ιδιαίτερων χαρακτηριστικών της ανθρώπινης φύσης. Η επιστημονική εξέλιξη ποτέ δεν εμποδίστηκε από Λουδιτικού τύπου αντιδράσεις. Ο άνθρωπος, αντίθετα με κάθε άλλο έμβιο ον, τείνει προς την αέναη ανακάλυψη και καινοτομία, διαμορφώνοντας το φυσικό και κοινωνικοπολιτισμικό περιβάλλον του και σε βάθος χρόνου την ίδια τη βιολογική λειτουργία και τη γενετική του κληρονομιά. Το πώς θα διαχειριστούμε την πρόκληση της ΤΝ θα καθορίσει την πορεία του ανθρώπινου

[8] Για παράδειγμα, σύμφωνα με κάποιες θεωρήσεις της κβαντικής φυσικής, είναι η συνειδητότητα του υποκειμένου αυτή η οποία κάνει ένα κύμα πιθανοτήτων να καταρρέει –δηλαδή τις πολλαπλές εκδοχές της πραγματικότητας να πραγματώνονται σε μία, αυτή που βιώνουμε– ενώ σύμφωνα με άλλες είναι ακριβώς από την κατάρρευση του κύματος πιθανοτήτων που αναδύεται η συνειδητότητα (Gao, 2014).

πολιτισμού. Η ΤΝ μπορεί να αξιοποιηθεί για την ορθολογική διαχείριση των παγκόσμιων πόρων και τη βελτίωση της ζωής όλων των ανθρώπων, αλλά συγχρόνως και για τον απόλυτο έλεγχό τους και την καταπίεση κάθε μορφής δημοκρατίας.

Είναι θετικό ότι υπάρχει μια τεράστια αύξηση των ερευνών σχετικά με την ασφάλεια της ΤΝ, εφόσον έχει γίνει πλέον σαφές ότι οι τεράστιες δυνατότητές της μπορούν να αλλάξουν τον κόσμο μας προς το καλύτερο, αλλά ενέχουν και κινδύνους μεγαλύτερους κι από την πυρηνική απειλή (Tegmark, 2017). Αν η ΤΝ αποκτήσει ανθρώπινες δυνατότητες και τα ρομπότ σκέφτονται και συνειδητοποιούν τον κόσμο όπως οι άνθρωποι, δεν θα αποκτήσουν συνεπώς και τη δυνατότητα ελεύθερης βούλησης, διαμορφώνοντας πλέον τη δική τους ηθική και θέτοντας δικούς τους στόχους; Τι θα σημαίνει αυτό για τους ανθρώπους; Πέρα όμως από την προστασία των ανθρώπων από τον οποιοδήποτε κίνδυνο των εξελιγμένων ρομπότ, είναι σημαντικό να συζητηθεί σοβαρά το πώς ενδέχεται να διαμορφωθεί η ίδια η ανθρώπινη φύση, μέσα από τη συνύπαρξή μας με αυτά τα τεχνητά όντα. Διότι αν και ένα μεγάλο μέρος των ανθρώπων του σήμερα πιστεύει ότι τα ρομπότ θα είναι απλώς εργαλεία που θα ικανοποιούν πρακτικές ανάγκες μας, η ιστορία του πολιτισμού έχει δείξει ότι τα εργαλεία που δημιουργούμε, στη συνέχεια διαμορφώνουν την αντίληψη της ύπαρξής μας και τη μορφή των κοινωνικών και διαπροσωπικών μας σχέσεων.

Βιβλιογραφικές αναφορές

Bartz, J. A., Zaki, J., Bolger, N., Ochsner, K. N. (2011). Social effects of oxytocin in humans context and person matter. *Trends in Cognitive Sciences*, 15(7): 301-9. doi: 10.1016/j.tics.2011.05.002.

Baudrillard, J. (1994). *Simulacra and simulation*. Ann Arbor: University of Michigan Press.

Breazeal, C. (2002). *Designing Sociable Robots.* Cambridge, MA: MIT Press.

Carpenter, J. (2016). *Culture and Human-Robot Interaction in Militarized Spaces: A War Story*. London: Routledge.

Chalmers, D. (1996). *The Conscious Mind: In search of a fundamental theory.* Oxford: Oxford University Press.

Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (ed.), *Neural correlates of consciousness: empirical and conceptual questions.* Cambridge, MA: MIT Press.

Christakis, N. (2019). *Blueprint: The evolutionary origins of a good society*. New York: Little, Brown Spark.

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex,* 8(2), 97-107.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Davou, B., & Sidiropoulou, A. (2017). Family life around screens: Some thoughts on the impact of ICTs on psychological development and the development of relationships. *Contemporary Family Therapy: An International Journal,* 39(4), 261-270.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts.* New York: Viking.

Dennett, D. C. (1991). *Consciousness explained*. New York: Little, Brown and Co.

Dewey, J. (1934). *Art as experience*. Oxford, England: Minton, Balch.

Dietrich, A. (2007). *Introduction to Consciousness*. New York and Basingstoke United Kingdom: Palgrave Macmillan.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science,* 311(5763), 1005-1007.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, Ill: Row, Peterson.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR mental health,* 4(2), e19. doi:10.2196/mental.7785.

Floridi, L. (2010). Artificial Companions and their philosophical

challenges. In Y. Wilks (ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues.* Amsterdam: John Benjamins Pub. Co.

Freud, S. (1919). *The 'Uncanny'*. The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XVII (1917-1919): An Infantile Neurosis and Other Works, 217-256.

Gao, S. (2014). *Consciousness Encounters Quantum Physics: A Scientific Argument for Panpsychism*. Amazon.co.uk: Kindle Store.

Giddens, A. (1992). *The transformation of intimacy: Sexuality, love and eroticism in modern societies.* Cambridge, UK: Polity Press.

Guastella, A. J., Hickie, I. B.(2016). Oxytocin Treatment, Circuitry, and Autism: A Critical Review of the Literature Placing Oxytocin Into the Autism Context. *Biological Psychiatry*, 79(3), 234-42.

Harari, Y. N. (2016). *Homo deus: A brief history of tomorrow*. New York: Hachette.

Harari, Y. N. (2018). *21 Lessons for the 21st Century*. New York: Spiegel & Grau.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Heidegger, M. (1962). *Being and time.* New York: Harper and Row.

Illouz, E. (2007). *Cold intimacies: The making of emotional capitalism*. Cambridge, UK: Polity Press.

Kaku, M. (2014). *The future of the mind: the scientific quest to understand, enhance, and empower the mind.* New York: Doubleday.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Denver, Colo: Roberts and Co.

Konrath, S. H., O'Brien, E. H., & Hsing, C. (2011). Changes in Dispositional Empathy in American College Students Over Time: A Meta-Analysis. *Personality and Social Psychology Review*, 15(2), 180-198.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking.

Levy, D. (2007). *Love and sex with robots: The evolution of human-robot relationships.* London: Harper-Collins.

Mayor, A. (2018). *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. New Jersey: Princeton University Press.

Moore, G. E. (1998). Cramming more components onto integrated circuits. *Proc. IEEE, 86*, 82–85. doi: 10.1109/JPROC.1998.658762.

Mori, M. (1970). The Uncanny Valley Phenomenon. *Energy*, 7(4), 33-35.

Nagasawa, M., Mitsui, S., En, S., Ohtani, N., Ohta, M., Sakuma, Y., Onaka, T., Mogi, K. Kikusui, T. (2015). Oxytocin-gaze positive loop and the coevolution of human-dog bonds. *Science,* 348, 333-336.

Ντάβου, Μπ. (2000). *Οι διεργασίες της σκέψης στην εποχή της πληροφορίας: Θέματα γνωστικής ψυχολογίας και επικοινωνίας.* Αθήνα: Παπαζήσης.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4) 435-450.

O'Hara, K., Morris, K., Shadbolt, N., Hitch, G. J., Hall, W., & Beargrie, N. (2006). Memories for life: a review of the science and technology. *Journal of the Royal Society Interface,* 3, 351-365. doi: 10.1098/rsif.2006.0125.

Peltu, M. & Wilks, Y. (2010). Summary and discussion of the issues. In Y. Wilks (ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues*. Amsterdam: John Benjamins Pub. Co.

Picard, R. W. (2000). *Affective computing*. Cambridge, Mass: MIT Press.

Pfaff, D. W. (2015). *A neuroscientist looks at robots.* New Jersey: World Scientific.

Prensky, M. (2001). Digital natives, digital immigrants Part 1. *On the horizon*, 9(5), 1-6. doi: 10.1108/10748120110424816.

Reeves, B., and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford, CA: CSLI.

Ryle, G. (1949). *The concept of the mind.* New York: Hutchinson's University Library.

Σιδηροπούλου, Α. (2019). *Ψυχολογικές διαδρομές στην ψηφιακή εποχή: Από το multitasking στην πολυδιεργασία της ύπαρξης.* Αθήνα: Παπαζήσης.

Shamay-Tsoory, S.G., Fischer, M, Dvash, J, Harari, H, Perach-Bloom, N, Levkovitz, Y. (2009). Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biological Psychiatry*, 66(9):864-870. doi:10.1016/j.biopsych.2009.06.009.

Sigman, M. (2017). *The Secret Life of the Mind: How Your Brain*

Thinks, Feels, and Decides. New York: Little, Brown and Company.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550, 354-359.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science,* 333(6043), 776-778.

Suler, J. R. (2016). *Psychology of the digital age: Humans become electric.* New York: Cambridge University Press.

Tegmark, M. (2017). *Life 3.0: being human in the age of artificial intelligence.* New York: Alfred Knopf.

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370 (1668). doi: 10.1098/rstb.2014.0167.

Tononi, G. (2004). An Information Integration Theory of Consciousness. *BMC Neuroscience, 5*, 42-42. doi:10.1186/1471-2202-5-42.

Turkle, S. (2010). In Good Company? On the threshold of robotic Companions. In Y. Wilks (ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues.* Amsterdam: John Benjamins Pub. Co.

Turkle, S. (2011). *Alone together : why we expect more from technology and less from each other.* New York :Basic Books.

Twenge, J. M. (2017). *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy and completely unprepared for adulthood (and what this means for the rest of us).* New York: Atria Books.

Weinberg, S. (1992). *Dreams of a Final Theory: The Search for the Fundamental Laws of Nature.* New York: Pantheon Books.

Weizenbaum, J. (1966). ELIZA-A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM*, 9, 36-45. doi: 10.1145/365153.365168.

Winnicott, D. (1971). *Playing and reality.* London: Routledge.

Χρηστάκης, Ν., & Χαλάτσης, Π. (2010). *Φιλικές σχέσεις: Νοήματα και πρακτικές στις ομόφυλες και στις ετερόφυλες φιλίες*. Αθήνα: Ελληνικά Γράμματα.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist,* 35(2), 151-175.