

## Εγώ, το ρομπότ: Η επιστροφή των ρομποτόφιλων

Θανάσης Πρωτόπαπας  
Ξυλόκαστρο, 9 Ιουλίου 2007

Έχω την εντύπωση πως ξεκινήσαμε με ένα στημένο ερώτημα, καταδικασμένο να οδηγεί σε ψεύτικα παράδοξα και να μείνει αναπάντητο, και πως δυσκολευόμαστε να το ελευθερώσουμε από τις προκαταλήψεις της αρχικής του διατύπωσης.

Ελπίζω στη σημερινή μου ομιλία να κάνω σαφές γιατί έχω αυτήν την εντύπωση και να βοηθήσω ώστε να ξεπεράσουμε το σκόπελο της επιφανειακής ασυνεννοησίας. Ειπώθηκε πέρσι, προς το τέλος της συνάντησής μας, πως δεν κάνουμε διάλογο αλλά λέει ο καθένας λίγο-πολύ τα δικά του και δεν πάμε πουθενά. Για το λόγο αυτό αποφάσισα να μην προσθέσω κάτι νέο με τη σημερινή μου παρέμβαση, αλλά να απαντήσω στην περσινή ομιλία του Ανδρέα Παπανικολάου, εντοπίζοντας σημεία συναίνεσης και σημεία απόκλισης. Διάλεξα την ομιλία του Ανδρέα διότι φαίνεται να τοποθετείται σε μια διαμετρικά αντίθετη θέση από τη δική μου, δηλαδή διαφωνούμε σε βασικά ζητήματα, άρα υπάρχουν διαφορές να εξεταστούν. Όμως παρόλα αυτά οι περισσότερες θέσεις του Ανδρέα είναι ξεκάθαρα διατυπωμένες, σαφείς, και μέσα σ' ένα πλαίσιο τέτοιο που νομίζω πως μπορώ να τις καταλάβω και να πω κάτι ως απάντηση, κάτι που δεν θα μπορούσα να κάνω σ' ένα ψυχαναλυτικό, θεολογικό ή λογοτεχνικό κείμενο.

Το σχέδιο της παρέμβασής μου είναι να σκιαγραφήσω την πορεία της ανάπτυξης των θέσεων του Ανδρέα στο συγκεκριμένο κείμενο, να ασκήσω κριτική σε επιμέρους σημεία με τα οποία διαφωνώ, και στο τέλος να αναφερθώ στον πυρήνα των θέσεων, αναπτύσσοντας σε αντιπαράθεση τις δικές μου, «ρομποτικές», θέσεις. Θα σταθώ περισσότερο στα σημεία που διαφωνούμε, με κίνδυνο να δοθεί η εντύπωση πως διαφωνώ με τα πάντα. Παρότι υπάρχει πολύ κοινό έδαφος, νομίζω πως είναι χρησιμότερο να ασχοληθούμε με τις διαφορές παρά με τις ομοιότητες.

Θα αρχίσω με δύο αφορισμούς του Ανδρέα, οι οποίοι δείχνουν τη διαφορετικότητα του σημείου αφετηρίας του από το δικό μου. Ο πρώτος είναι ότι θεωρεί πρόβλημα για οποιαδήποτε θεωρία «ότι ήταν γραφτό να πω αυτήν εδώ τη στιγμή ακριβώς αυτό που λέω και όχι κάτι έστω και ελάχιστα αλλιώτικο». Ο δεύτερος είναι ότι «οι αξίες και πεποιθήσεις περί πραγμάτων δεν αποδίδονται με αλγόριθμους». Λέω ότι πρόκειται για αφορισμούς διότι δεν βλέπω καμία απόδειξη, ούτε καν επιχειρήματα, που να υποστηρίζει τις θέσεις αυτές. Μου φαίνονται σαν a priori κρίσεις, λίγο-πολύ ταμπού, που χρησιμοποιούνται για να περιορίσουν το είδος των απαντήσεων που επιτρέπεται να δώσουμε στις σημαντικές ερωτήσεις. Όμως οι κρίσεις αυτές δεν ανήκουν εδώ. Για παράδειγμα, το αν «ήταν γραφτό», δηλαδή αν το σύμπαν είναι ντετερμινιστικό, δεν είναι θέμα της προσωπικής μας γνώμης και προτίμησης, ούτε της βιολογίας ούτε της συνείδησης. Είναι κάτι που πρέπει να απαντήσει η φυσική. Αν η τυχαιότητα που βλέπουμε σήμερα σε κβαντικό επίπεδο μπορεί να αποδοθεί σε ντετερμινιστικούς νόμους, που δεν τους γνωρίζουμε ακόμα, τότε καθετί ήταν γραφτό από τη στιγμή της γέννησης του σύμπαντος είτε μας αρέσει είτε όχι. Αν η τυχαιότητα είναι θεμελιακή και έχει μακροσκοπικές επιπτώσεις τότε τίποτα δεν ήταν γραφτό, πάλι είτε μας αρέσει είτε όχι. Ο λόγος που δεν υποστηρίζω το ένα ή το άλλο είναι αφενός γιατί δεν γνωρίζω τη σωστή απάντηση, και αφετέρου γιατί πιστεύω ότι η σωστή απάντηση δεν έχει σημαντικές συνέπειες για το αντικείμενο της συζήτησής μας, όσο κι αν μας φαίνεται προτιμότερη η μια ή η άλλη εκδοχή.

Ο δεύτερος αφορισμός αναφέρεται σε αξίες και πεποιθήσεις, πράγματα που δεν ξέρουμε τι είναι. Τουλάχιστον δεν το ξέρουν οι ψυχολόγοι και γνωσιακοί επιστήμονες που προσπαθούν να τα μελετήσουν. Το να λέμε σήμερα αν οι αξίες αποδίδονται ή όχι με αλγόριθμους το βρίσκω πρόωρο. Το αν οποιαδήποτε νοητική διεργασία αποδίδεται με αλγόριθμους, και τι είδους αλγόριθμοι θα 'ταν αυτοί, είναι ένα εξαιρετικά ενδιαφέρον και σημαντικό ερώτημα, το οποίο ελπίζω κάποτε ν' απαντηθεί. Προς το παρόν πιστεύω πως δεν πρέπει να αποφασίζουμε χωρίς στοιχεία ποια είναι η απάντηση και να τη χρησιμοποιούμε μετά για να υποστηρίξουμε άλλες θέσεις, διότι το αποτέλεσμα είναι να μένουν και οι άλλες θέσεις αστήρικτες.

Βάζοντας στην άκρη αυτές τις επιμέρους τοποθετήσεις, το διάγραμμα της ανάπτυξης των θέσεων που εκφράζονται στην περυσινή ομιλία του Ανδρέα είναι το εξής: Καταρχήν αναφέρεται η ύπαρξη υποκειμενικών εμπειριών στον άνθρωπο, ως φαινόμενο προς εξήγηση, στο πλαίσιο της εξέλιξης του είδους. Τίθεται το ερώτημα του σκοπού που εξυπηρετεί η υποκειμενικότητα, για να έχει εξελιχθεί με φυσική επιλογή ως αρμοστικό γνώρισμα. Αναγνωρίζεται ότι η επίγνωση δεν συνάδει πάντα με την πράξη, και εξετάζονται δύο ενδεχόμενα για τις αποκλίσεις: το ένα είναι ο δυϊσμός, ο οποίος χαρακτηρίζεται προβληματικός χωρίς να απορρίπτεται, και το άλλο χαρακτηρίζεται απαξιωτικά ως «ρομποτάκια» που ακολουθούν μόνο τους νόμους της φυσικής. Παρατίθενται συγκινησιακές επικλήσεις στα πολιτικά ή σεξουαλικά δικαιώματα που θα ήμασταν διατεθειμένοι να παραχωρήσουμε σε ένα υποθετικό μελλοντικό κατασκευάσμα της ρομποτικής, το Βασιλάκη απ' τ' Άνω Λιόσια. Και ενώ ρητά αναφέρεται πως δεν μπορούμε να είμαστε λιγότερο ή περισσότερο σίγουροι για το αν ο Βασιλάκης έχει συνείδηση ή όχι, απ' όσο είμαστε σίγουροι ο ένας για τον άλλο, η συζήτηση ξαφνικά μεταφέρεται στο σκοπό της συνείδησης. Υποστηρίζεται πως η εξελικτική ανάδυση της συνείδησης αποτελεί παράδοξο. Εν συνεχεία ακολουθεί ένα τμήμα του κειμένου όπου παρατίθενται ορισμένα προβλήματα για το Βασιλάκη: αν διαθέτει τάση αυτοπροστασίας, αν είναι η τάση αυτή συνειδητή, αν είναι γενετική, και πώς προέκυψε γενικά η δομή και η τάξη στην έμβια ύλη κατά την εξέλιξη. Παρατίθεται ο αφορισμός περί ντετερμινισμού που προανέφερα και το λανθασμένο επιχείρημα των δημιουργιστών για την αύξηση της εντροπίας. Και το κείμενο τελειώνει με αναφορές στις ενδεχόμενες δυνατότητες μιας τεχνητής νοημοσύνης να προσεγγίσει την ανθρώπινη νοημοσύνη, στις οποίες θα αναφερθώ παρακάτω.

Αυτή είναι με δυο λόγια η πορεία του κειμένου: χτίζει καταρχήν τη θέση ότι αν δεν ισχύει ο δυϊσμός τότε πρέπει να είμαστε αυτόματα, και εν συνεχεία πυροβολεί εναντίον του προφανούς συμπεράσματος (ότι άρα είμαστε αυτόματα). Η αυτοπαγίδευση στη γωνία με το παράδοξο ακολουθεί τη συνταγή της αρχικής ερώτησης που τέθηκε για το συμπόσιο αυτό, στην οποία θεωρείται εκ των προτέρων δεδομένο πως η χαρά μας θα εξανεμιστεί αν μάθουμε ότι η χαρά είναι εγκεφαλική λειτουργία και όχι κάτι άλλο, άφατο. Η δικιά μου θέση είναι η αντίθετη: αποδεχόμαστε το συμπέρασμα του «αυτόματου» και αξιοποιούμε τις διαθέσιμες ενδείξεις για να διερευνήσουμε πώς με ποιους εγκεφαλικούς μηχανισμούς, μια βιολογική μηχανή έχει τη νόηση που έχουμε και τη συνείδηση για την οποία συζητάμε, και πώς παράγει τη χαρά που νιώθουμε όποτε βρίσκουμε την απάντηση σε ένα δύσκολο ερώτημα, σαν το ερώτημα του πώς παράγει τη χαρά ο εγκέφαλος.

Νομίζω συμφωνούμε απολύτως με τον Ανδρέα ότι αν η υποκειμενικότητα καθεαυτή είναι αιτιακά αδρανής τότε δεν μπορεί να έχει αποτελέσει αντικείμενο φυσικής επιλογής. Αυτό άλλωστε ισχύει και για οτιδήποτε άλλο, δεδομένου ότι η φυσική επιλογή μπορεί να δρά μόνο σε ιδιότητες με αρμοστικές επιπτώσεις. Αν η συνείδηση είναι είναι όχι απλώς ομοούσιος αλλά αδιαχώριστα ταυτισμένη με κάποια όψη ή ιδιότητα της εγκεφαλικής λειτουργίας, τότε δεν έχουμε κανένα πρόβλημα, διότι αποδεχόμαστε όλοι ότι η εγκεφαλική λειτουργία μπορεί κάλλιστα να έχει εξελιχθεί, με φυσική επιλογή, ώστε να έχει τα χαρακτηριστικά εκείνα που, μεταξύ άλλων, της προσδίδουν και υποκειμενικότητα, δηλαδή συνείδηση. Αν όμως η συνείδηση είναι οτιδήποτε πάνω ή πέρα από τη λειτουργία του εγκεφάλου, τότε, όπως λέει ο Ανδρέας έχουμε δύο επιλογές: Ή πρόκειται για άλλου είδους ουσία, δηλαδή μιλάμε για κλασικό δυϊσμό, με όλα τα γνωστά προβλήματα, ή μιλάμε για αιτιακά αδρανές επιφαινόμενο της εγκεφαλικής λειτουργίας, τέτοιο ώστε να παράγεται μεν από το εγκέφαλο αλλά να διαφέρει από τη λειτουργία του. Το ενδεχόμενο του δυϊσμού θα μου επιτρέψετε να μην το πάρω σοβαρά υπόψη μου, διότι δε βλέπω να μας οδηγεί σε κανένα χρήσιμο συμπέρασμα. Το ενδεχόμενο του επιφαινόμενου είναι εκείνο που, ως αιτιακά αδρανές, δεν μπορεί να έχει επιλεγεί εξελικτικά.

Γιατί απορρίπτεται άνευ εξέτασης και μετά βδελυγμίας το ενδεχόμενο να ταυτίζεται η νόηση, συνειδητή και ασυνειδητή, με την εγκεφαλική λειτουργία; Ίσως μπορεί κανείς να βασίσει μια απάντηση στις εισαγωγικές θέσεις του κειμένου, όπου αναφέρεται, χωρίς καμία τεκμηρίωση, ότι η συνείδηση αποτελεί προνόμιο του ανθρώπου, επί λέξει «είναι εμφανές ότι αυτό που βάζει τον άνθρωπο πάνω απ' όλα τα έμβια όντα είναι η ύπαρξη των υποκειμενικών εμπειριών». Για μένα αυτό δεν είναι καθόλου εμφανές. Θα έλεγα πως σήμερα δεν γνωρίζουμε αν και σε ποιο βαθμό άλλα είδη

έχουν επίγνωση του περιβάλλοντος ή του εαυτού τους. Μπορούμε να υποθέσουμε ότι η ομοιότητα των εγκεφαλικών μηχανισμών αντιστοιχεί σε ομοιότητα νοητικών καταστάσεων, και στο βαθμό που ισχύει κάτι τέτοιο δεν αποκλείεται καθόλου να διαθέτουν σημαντικό βαθμό επίγνωσης άλλα πρωτεύοντα, άλλα θηλαστικά, και ποιος ξέρει πόσοι οργανισμοί ακόμα, χωρίς φυσικά να είναι σε θέση να δηλώσουν κάτι τέτοιο, μια και τους λείπει η γλώσσα. Αυτό δε σημαίνει πως μπορούμε χωρίς στήριξη να απαγορεύουμε σε πιθήκους και σκύλους να πονούν, να πεινούν, ή να χαίρονται.

Στη συνέχεια το κείμενο διακρίνει την επίγνωση από τη μηχανιστική λειτουργία, θέτοντας με καθαρά δυϊστικούς όρους το ερώτημα της σκοπιμότητας της συνείδησης και εξασφαλίζοντας έτσι ότι θα οδηγηθεί σε παράδοξο. Όταν προϋποθέτει κανείς ότι άλλο πράγμα η συνείδηση και άλλο η εγκεφαλική λειτουργία, προφανώς θα πρέπει να αναρωτηθεί μετά ποιος ο ρόλος της συνείδησης και ποια η σχέση μεταξύ των δύο – και αναγκαστικά να καταλήξει στο δυισμό ή στην αιτιακή αδράνεια του επιφαινόμενου τίποτα. Δεν υπάρχει λύση σε αυτό το κατασκευασμένο δίλημμα, διότι προκύπτει εξαιτίας της αρχικής συνθήκης. Η απάντηση που προτείνουν οι συνεπείς μονιστές είναι ότι συνείδηση είναι κάποια αναπαραστασιακή κατάσταση της εγκεφαλικής λειτουργίας, όχι ότι παράγεται από αυτήν αλλά ότι ταυτίζεται με αυτήν. Αυτό σημαίνει πως οποιοσδήποτε οργανισμός, ή οποιαδήποτε κατασκευή επιτυγχάνει την κατάλληλη κατάσταση και σχέση με το περιβάλλον, θα επιτυγχάνει αυτομάτως και την αντίστοιχη υποκειμενικότητα. Το πλεονέκτημα της θέσης αυτής είναι ότι δεν οδηγεί σε παράδοξο αλλά σε ερωτήματα που μπορούν να εξεταστούν εμπειρικά, στο βαθμό που κανείς αποδέχεται τις αυτοαναφορές ως προς την υποκειμενικότητα (όποιος δεν τις αποδέχεται είναι σολιμιστής και εκτός συζήτησης ούτως ή άλλως).

Σε κάποιο σημείο του κειμένου βρίσκουμε τη φράση «θα μπορούσαμε να ζούμε ερήμην ...του εαυτού μας», με την οποία ο Ανδρέας υποθέτω ότι θέλει να καταδείξει το ανόητο του μονισμού. Αυτό που νομίζω εγώ πως δείχνει είναι ότι αντιφάσκει, παίζοντας συγχρόνως σε δύο ταμπλό. Ξεχνά να προσδιορίσει ποιοι είμαστε εμείς που ζούμε, και ποιος είναι ο εαυτός μας που λείπει. Πού ορίζονται, τι είναι, και πώς μπορεί να είναι δύο διαφορετικά πράγματα; Ξαναλέει παρακάτω, το κείμενο, για τη «λανθασμένη πεποίθηση ότι ενώ ο εγκεφαλικός μηχανισμός μας λύνει και δένει εμείς νομίζουμε ότι κάποιος άλλος ευθύνεται» – μα τι είμαστε εμείς αν όχι εγκεφαλικός μηχανισμός; Και τι μπορεί να σημαίνει «νομίζουμε» παρά μια κατάσταση του εγκεφαλικού μηχανισμού μας;

Αναφορικά με τα αμφιλεγόμενα πειράματα του Libet, αναφέρει ο Ανδρέας πως κάποιες εγκεφαλικές περιοχές ενεργοποιούνται πριν την αίσθηση της απόφασης. Συμφωνώ απολύτως πως οι ενδείξεις αυτές δεν είναι πειστικές, παρότι αντίθετα με τον Ανδρέα θα μου άρεσε να τις πιστέψω. Όμως δε χρειαζόμαστε συγκεκριμένες ενδείξεις για να παραδεχτούμε πως κάποια εγκεφαλική λειτουργία, κάπου και κάπως, θα πρέπει να προηγείται της συνειδητής απόφασης, εφόσον η απόφαση είναι νοητική λειτουργία, που επιτελείται στον εγκέφαλο, και θα πρέπει να προκαλείται από άλλες, προηγούμενες, λειτουργίες του εγκεφάλου, διότι δεν υπάρχει τίποτα άλλο από το οποίο να μπορεί να προκληθεί. Ανεξάρτητα από το τι ακριβώς βρήκε ο Libet, είναι αυτονόητο πως δεν μπορεί μια εγκεφαλική λειτουργία να είναι αναίτια. Και όσο δε γνωρίζουμε ουσίες και δυνάμεις έξω από τον εγκέφαλο που να μπορούν να αλλοιώσουν τη λειτουργία του, είμαστε αναγκασμένοι να αποδεχτούμε πως κάποια εγκεφαλική λειτουργία θα προηγείται της απόφασης και θα την προκαλεί αιτιακά.

Παρακάτω αναφέρεται επί λέξει το εξής: «πολλές φορές συμπεριφερόμεθα με τρόπο που υποδηλώνει πως αντιληφθήκαμε το ποιον μιας αισθητικής εμπειρίας καίτοι δεν έχουμε επίγνωση πως το αντιληφθήκαμε». Εδώ υπάρχει εννοιολογική σύγχυση, αλλά δεν την καταλογίζω στον Ανδρέα. Το πρόβλημα είναι γενικότερο στη φιλοσοφία του νου. Τι θα πει αντιληφθήκαμε, όταν δεν έχουμε επίγνωση; Και, πιο σημαντικό, πώς είναι δυνατό να μην έχουμε επίγνωση ενός ποιόντος που αντιληφθήκαμε, όταν «ποιον» ορίζουμε πως είναι η υποκειμενική όψη του αντιλήμματος; Υποκειμενικότητα χωρίς επίγνωση; Αυτό δεν σημαίνει τίποτα. Απλώς καταδεικνύει κατηγορηματικά πόσο αποτυχημένη είναι η έννοια των ποιόντων, που δημιουργεί περισσότερες αντιφάσεις παρά λύσεις.

Προχωρώ παρακάτω, παρακάμπτοντας κάμποση ρομποτολογία, στην οποία θα επανέλθω αργότερα, για να σχολιάσω μερικά ακόμα επιμέρους σημεία. Ο Ανδρέας δικαιολογεί την τελεολογική

διερεύνηση με αναφορά στην κοινή πρακτική, και νομίζω πως μέχρι ενός σημείου έχει απόλυτο δίκιο. Πράγματι, απ' όσο καταλαβαίνω, όλοι οι βιολόγοι προσπαθούν να κατανοήσουν μηχανισμούς υποθέτοντας πως πρόκειται για μηχανισμούς που εξυπηρετούν κάποιο σκοπό. Και οι εξελικτικοί βιολόγοι προσπαθούν να κατανοήσουν τα χαρακτηριστικά των οργανισμών υποθέτοντας πως εξελίχθηκαν με φυσική επιλογή επειδή καθιστούν τους οργανισμούς αρμοστικότερους, στο πλαίσιο των υπολοίπων εσωτερικών και εξωτερικών χαρακτηριστικών. Νομίζω πως η ουσία της παρατήρησης είναι σωστή, με μια διαφορά: η ύπαρξη συγκεκριμένου σκοπού δεν αποτελεί παρά υπόθεση εργασίας για την έρευνα, και μπορεί να αποδειχτεί λανθασμένη υπόθεση για κάποιες περιπτώσεις. Δεν θεωρείται αδιάσειστα δεδομένο πως καθετί έχει το σκοπό του και τον επιτελεί καλά, διότι οι αλληλεξαρτήσεις των επιμέρους γνωρισμάτων μεταξύ τους και με το περιβάλλον, καθώς και η εξελικτική ιστορία του κάθε οργανισμού, θέτουν τεράστιους περιορισμούς στη διαμόρφωση των οργανισμών. Αυτό σημαίνει, για μένα, πως καλά κάνουμε και ρωτάμε «προς τι η συνείδηση» (ή οποιοδήποτε άλλο γνώρισμα οποιοδήποτε οργανισμού). Θα πρέπει να είμαστε όμως διατεθειμένοι να αποδεχτούμε, αν τα στοιχεία το υποδείξουν, πως οποιοδήποτε γνώρισμα μπορεί να είναι τυχαίο, άχρηστο, ή ακόμα και βλαπτικό, για το συγκεκριμένο οργανισμό στο συγκεκριμένο σημερινό περιβάλλον, εξαιτίας της εξελικτικής συγκυρίας. Πάντως δεν υποστηρίζω πως η συνείδηση αποτελεί τέτοιο γνώρισμα, αφού η θέση μου είναι πως η συνείδηση δεν αποτελεί καν διακριτό γνώρισμα προς επιλογή, αλλά ιδιότητα της εξελιγμένης εγκεφαλικής λειτουργίας, άρα δεν έχει και δεν μπορεί να έχει διακριτό σκοπό, άλλον από εκείνο της υπόλοιπης εγκεφαλικής λειτουργίας.

Στο σημείο αυτό θα ήθελα να προτείνω κάποιες διορθώσεις. Πρόκειται για σημεία όπου δεν βλέπω ζήτημα διαφοράς προσωπικών απόψεων. Το πρώτο έχει να κάνει με το αν είναι θερμοδυναμικά εφικτή η ευταξία, και η απάντηση είναι απολύτως ναι. Χωρίς να παριστάνω τον ειδικό φυσικό, παρότι πτυχιούχος φυσικής, πρέπει να επισημάνω πως η αύξηση της εντροπίας αφορά σε κλειστά συστήματα, ή σε ολόκληρο το σύμπαν, και όχι σε τοπικά ανοιχτά συστήματα όπως είναι ο πλανήτης μας. Οι φυσικοί εδώ μπορούν να βεβαιώσουν ότι δεν υπάρχει κανένα πρόβλημα στην αυτο-οργάνωση και τη διατήρηση της τάξης συστημάτων που μπορούν να ανταλλάσσουν ενέργεια με το περιβάλλον τους. Το θέμα της θερμοδυναμικής δεν έχει καμία σχέση με τη συζήτησή μας ούτε και μπορεί να προσφέρει επιχειρήματα υπέρ ή κατά οποιασδήποτε πλευράς.

Το δεύτερο σημείο έχει και αυτό να κάνει με την οργάνωση, την τάξη και την πολυπλοκότητα, και το ρόλο της τυχαιότητας σε αυτή. Πρόκειται για μια εκδοχή του κλασικού πλέον επιχειρήματος υπέρ του σχεδιαστή, και ατυχώς το επιστρατεύει ο Ανδρέας, ο οποίος δεν δηλώνει υπέρμαχος δημιουργών και δημιουργημάτων. Λέει, για την ακρίβεια: «το μυστήριο είναι πώς προκύψαμε εμείς, αφού 'τυχαίως' δεν είναι εύλογο και αφού 'μας σχεδίασε ο ύψιστος' δεν αποτελεί νόμιμη επιστημονική αιτιολογία». Αυτό ακριβώς το ψευτοδίλημμα είναι που έλυσε ο Δαρβίνος, όταν έδειξε πως η αυξανόμενη αρμοστικότητα, στοιχείο εμφανώς αντίθετο με την «τυχαιότητα», μπορεί να προκύψει, για την ακρίβεια προκύπτει υποχρεωτικά, όποτε υπάρχουν κληρονομήσιμες διαφορές με αρμοστικές συνέπειες. Η εναλλακτική υπόθεση στον ύψιστο δεν είναι η τυχαιότητα, αλλά η φυσική επιλογή. Η φυσική επιλογή είναι μηχανισμός μη τυχαίας μείωσης της ποικιλίας, ενώ η αναπαραγωγή με ανασυνδυασμό και οι μεταλλάξεις είναι μηχανισμοί αύξησης της ποικιλίας, λίγο ή πολύ τυχαίοι. Η εξέλιξη μπορεί να οδηγήσει σε πολύπλοκους οργανισμούς, καλά ταιριασμένους στο οικολογικό τους πλαίσιο, ακριβώς επειδή η διαφορική αναπαραγωγή δεν είναι τυχαία αλλά εξαρτάται από κληρονομήσιμες διαφορές.

Η τρίτη διόρθωση έχει να κάνει με τη συνεργατικότητα. Μιλά ο Ανδρέας για χημικές ενώσεις και άλλες οργανικές οντότητες που δομούν πολύπλοκες κοινωνίες όπου, επί λέξει, «τα μέλη, αντί να αποζητούν... όπως πρώτα, εγωιστικά τη δική τους ξέχωρη επιβίωση αποζητούν πρωτίστως να εξυπηρετήσουν το πέραν αυτών κοινό καλό, το καλό της οντότητας που συναπαρτίζουν». Πρόκειται για παρανόηση, και στα δύο επίπεδα όπου μπορεί να προσδιοριστεί. Στο επίπεδο των ίδιων των γονιδίων, όταν τυχαίνει η επίδραση κάποιου γονιδίου X να οδηγήσει σε αρμοστικότερο φαινότυπο παρουσία κάποιου άλλου γονιδίου Ψ, παρά κάποιου Ω, τότε είναι πολύ πιθανή συνέπεια ότι ο συνδυασμός ΧΨ θα αποβεί συχνότερος από κάποιον ανταγωνιστικό συνδυασμό ΧΩ. Κανείς δεν αποζητά τίποτα ούτε εξυπηρετεί τίποτα, απλώς κάποιες οντότητες επιβιώνουν περισσότερο από άλλες, με αποτέλεσμα εμείς να βλέπουμε περισσότερα ΧΨ από ΧΩ. Στο επίπεδο των φαινοτύπων τώρα, αν

ένα κύτταρο ενωθεί με κάποιο άλλο κύτταρο και αποτελέσει μακροπρόθεσμα «οργανίδιο» του, δε σημαίνει ότι κανείς απεμπόλησε την εθνική του κυριαρχία. Τα κύτταρα, και γενικότερα οι όποιες οργανικές οντότητες σε επίπεδο πάνω από το γενετικό, δεν αποτελούν αυτοαναπαραγόμενες μονάδες, αλλά αποτέλεσμα της έκφρασης των αυτοαναπαραγόμενων μονάδων που είναι το γενετικό υλικό. Αν λοιπόν η έκφραση των γονιδίων που οδηγεί σε συνεργατικές οντότητες, είναι πιο αποτελεσματική από εκείνη ανταγωνιστικών γονιδίων που παράγουν μεμονωμένες οντότητες, τότε είναι πολύ πιθανό ότι θα βρούμε τις ολόενα και πιο πολύπλοκες δομές και κοινωνίες που αναφέρονται. Απλώς τα γονίδια επιβιώνουν, ως αναπαραγόμενες πληροφοριακές μονάδες, μέσα από συνεργατική φαινοτυπική στρατηγική. Ούτε εγωϊστικές πρωτεΐνες ούτε υπηρέτες της κοινωνίας.

Κανένα λοιπόν πρόβλημα δεν αποτελεί η καταγωγή της ευταξίας, και τίποτα δεν προσφέρει στο αντιρομποτικό μανιφέστο. Ένα σχόλιο ακόμα στην επίκληση της τυχαιότητας. Η τυχαιότητα υποτίθεται ότι καλείται για να αντιπαρατεθεί στον ανεπιθύμητο ντετερμινισμό, όμως δεν λύνει κανένα πρόβλημα. Αν το κακό με το ντετερμινισμό είναι ότι δεν αφήνει χώρο στην ατομική υπευθυνότητα, η τυχαιότητα σίγουρα δεν βοηθάει. Δε νομίζω πως θα νιώθει κανείς πιο υπεύθυνος για τις πράξεις του αν αυτές καθορίζονται από τυχαία γεγονότα παρά αν έχουν προδιαγραφεί από τις απαρχές του σύμπαντος. Η ατομική υπευθυνότητα ορίζεται μέσα στο πλαίσιο ενός ενιαίου εγκεφαλικά κατασκευασμένου εαυτού. Ακόμα κι αν οι αποφάσεις του εαυτού αυτού είναι εξαρχής προκαθορισμένες, παραμένουν αποφάσεις του εαυτού, του συγκεκριμένου υπολογιστικού βιολογικού συστήματος. Όσο τα νευρικά κυκλώματα συγκροτούν ατομικούς εαυτούς, φορείς και υποκείμενα αισθημάτων, επιθυμιών και δράσεων, η έννοια της ατομικής υπευθυνότητας θα είναι απολύτως ασφαλής.

Θα ήθελα εδώ να διαμαρτυρηθώ για μια σύγκριση που κάνει ο Ανδρέας μεταξύ της «υπεράσπισης της ρομποτικής θεωρίας» και «την ευλαβή ανάγνωση της όποιας θεογονίας και όποιας βίβλου». Τέτοιες θέσεις παραβλέπουν τη θεμελιώδη διαφορά μεταξύ της επιστήμης και της θρησκείας, η οποία είναι η επίγνωση της ενδεχόμενης διάψευσης. Το αν κάποιος υποστηρίζει επίμονα μια θέση, ακόμα και χωρίς επαρκή στοιχεία, δε σημαίνει πως δεν είναι έτοιμος να την εγκαταλείψει αν παρασχεθούν στοιχεία για το αντίθετο. Η θρησκεία όχι μόνο δεν είναι ανοικτή σε στοιχεία και διαψεύσεις, αλλά ρητά αποκηρύσσει τον ορθό λόγο και τη διερεύνηση, εμμένοντας αποκλειστικά στην πίστη. Μάλιστα όσο πιο τρελά και αυταποδείκτως αδύνατα είναι αυτά που πιστεύει κανείς τόσο μεγαλύτερες τιμές του επιφυλάσονται. Συνεπώς η διαφορά δεν είναι πόσα στοιχεία έχουμε σήμερα υπέρ μιας θέσης την οποία υποστηρίζουμε επειδή μας αρέσει, αλλά πόσο ασχολούμαστε με τη συλλογή σχετικών στοιχείων και τι είμαστε διατεθειμένοι να κάνουμε στην περίπτωση που τα στοιχεία είναι ασύμβατα με τη θέση μας.

Το κείμενο του Ανδρέα καταλήγει με μια σειρά από προβληματικές προτάσεις. Λέει πως «κανένα ρομπότ...υπαρκτό ή δυνάμει υπαρκτό...δεν μπορεί να είναι όμοιο του ανθρώπου»· πως «τα όρια της νοημοσύνης του ανθρώπου μετατοπίζονται και διευρύνονται...και κανείς δεν ξέρει πόσο θα διευρυνθούν στο μέλλον»· πως «κανένα ρομπότ δεν μπορεί να ανακαλύψει καινούριες αλήθειες»· και πως «οι αξίες και πεποιθήσεις περί πραγμάτων δεν αποδίδονται με αλγορίθμους». Σε τίποτα από αυτά δεν σκοπεύω να απαντήσω μια και δεν στηρίζονται πουθενά και ούτε βλέπω να μπορούν να συνδεθούν με κάποια υποθετική εμπειρική βάση. Νομίζω όμως πως μεγάλο μέρος του προβλήματος βρίσκεται στην έννοια του ρομπότ.

Είναι λοιπόν ώρα να έρθουμε στο κεντρικό αυτό ζήτημα και βασικό σημείο διαφωνίας, το ρομπότ. Για να μπορέσουμε να συνεννοηθούμε, καλό είναι να συμφωνήσουμε τι είναι και τι δεν είναι ρομπότ, και για τι είδους ρομπότ μιλάμε. Με τη λέξη «ρομπότ» γενικά μπορεί να εννοούμε οποιαδήποτε προγραμματιζόμενη ηλεκτρομηχανική κατασκευή, ή ακόμα και οποιαδήποτε μη βιολογική κατασκευή, αρκεί να είναι προγραμματισμένη για να ανταποκρίνεται σε κάποιο βαθμό σε ερεθίσματα με λιγότερο ή περισσότερο ευέλικτη και προσαρμοστική συμπεριφορά. Μπορεί να είναι οτιδήποτε έχουμε κατασκευάσει που δεν χρειάζεται χειριστή ή τηλεκοντρόλ.

Καταρχήν ρομπότ είναι οι μηχανικοί ελεγχόμενοι βραχίονες που συναρμολογούν αυτοκίνητα, οι ηλεκτρικές σκούπες που γυρνάνε το σπίτι και σκουπίζουν μόνες τους, το αυτοκινούμενο όχημα

εξερεύνησης της επιφάνειας του Άρη και κάτι χαζοχαρούμενα πλαστικά σκυλάκια που περπατούν ατσούμπαλα και κάνουν ηχογραφημένα «γαβ». Όλα αυτά είναι για μένα το ίδιο είδος ρομπότ. Πρόκειται για εξαιρετικά απλές κατασκευές, προγραμματισμένες να κάνουν μία ή δύο δουλειές το πολύ, τις οποίες κάνουν καλούτσικα μέχρι πολύ καλά, ανάλογα με το πόσο επαναληπτικές είναι και πόσο σταθερό το πλαίσιο μέσα στο οποίο τις εκτελούν. Οι συσκευές αυτές καταρρέουν εντελώς μόλις τοποθετηθούν σε λίγο διαφορετικό πλαίσιο από αυτό για το οποίο κατασκευάστηκαν. Αυτές οι κατασκευές είναι σήμερα το όριο της ρομποτικής τεχνολογίας. Νομίζω μπορούμε όλοι να συμφωνήσουμε πως δεν μιλάμε γι αυτό το πράγμα. Δεν πιστεύω κάποιος να έχει κατά νου πως θεωρούμε το νου, ή οποιοδήποτε ζωντανό οργανισμό, συγκρίσιμο με τέτοιου είδους ντενεκέδια. Δε φαντάζομαι πως θα 'κανε κανείς τον κόπο να επιχειρηματολογήσει γιατί δεν είμαστε τέτοιου είδους ρομπότ.

Τι άλλο μπορεί να είναι ρομπότ; Αυτό για το οποίο θεωρώ πως μιλάμε, για το Βασιλάκη απ' τ' Άνω Λιόσια δηλαδή, πρέπει να είναι μια κατασκευή μεγάλης πολυπλοκότητας, συγκρίσιμης με αυτήν του ανθρώπου. Φυσικά πρέπει να είναι κάτι κατασκευασμένο από τον άνθρωπο. Για να έχει όμως δικαίωμα στο όνομα Βασιλάκης και στη συζήτηση αυτή, θα πρέπει να είναι επιφανειακά μη διακρίσιμη από οποιονδήποτε άλλο Βασιλάκη. Κάτι τέτοιο είναι σήμερα έξω όχι μόνο από την ρομποτική τεχνολογία αλλά και από την επιστημονική φαντασία. Αυτό δεν είναι καθόλου νεφελώδες ή ασαφές. Όμως δεν είναι και αυτό που αναφέρει ο Ανδρέας, «το είδος του πράγματος που φτιάχνουν ή που θα μπορούσαν να φτιάξουν οι τεχνικοί που ασχολούνται με το θέμα». Οι τεχνικοί που ασχολούνται σήμερα με το θέμα δεν ξέρουν καλά-καλά πώς να φτιάξουν κάτι που να μη κουτουλάει στους τοίχους αν τους βάσουμε άλλο χρώμα ή αν αλλάξουμε τη θέση του φωτιστικού. Λοιπόν πρόκειται για κάτι εντελώς αδύνατο με την τεχνολογία που έχουμε και με όποια τεχνολογία μπορούμε να φανταστούμε. Ίσως να παραμείνει για πάντα τεχνολογικά αδύνατο. Δεν έχω ιδέα. Αν ποτέ γίνει δυνατό τότε η γνωσιακή επιστήμη θα έχει θριαμβεύσει και θα μπορούν όλοι οι επιστήμονες του κλάδου να πάνε ήσυχοι στα σπίτια τους, ή να ασχοληθούν με το ψάρεμα.

Έστω όμως ότι μπορεί κάποτε να γίνει. Μπορούμε να φανταστούμε, με λίγη ή περισσότερη δυσκολία, ένα μηχάνημα το οποίο είναι κατασκευασμένο, κύτταρο προς κύτταρο, κατ' εικόνα και ομοίωση ενός συγκεκριμένου Βασιλάκη. Κάθε νευρώνας ένας επεξεργαστής, κάθε νευράξονας και καλώδιο. Το ίδιο και για τα υπόλοιπα κύτταρα των άλλων ιστών του σώματος, σε ό,τι χρειάζεται για να είναι στην όψη και στη συμπεριφορά σαν άνθρωπος – ακόμα και να τρώει παϊδάκια και να πίνει τσίπουρα.

Νομίζω ότι για ένα τέτοιο ρομπότ μιλάμε, ή τουλάχιστον, κατά τη γνώμη μου, γι αυτό έχει νόημα να μιλήσουμε και σε αυτό αναφέρονται όσα ακολουθούν.

Μια μη βιολογική μηχανή λοιπόν, κατασκευασμένη από τον άνθρωπο κατά απόλυτη λειτουργική ομοίωση του ανθρώπου. Από μέταλλα, καλώδια ή κεραμικά δεν έχει σημασία. Εκείνο που έχει σημασία για τη συζήτησή μας είναι ο Βασιλάκης να είναι κατασκευασμένος έτσι ώστε να περνάει το τεστ του Turing, δηλαδή να μην μπορεί να διακριθεί από άνθρωπο συζητώντας μαζί του. Προσοχή, αυτό είναι πολύ σημαντικό, και είναι όρος απαράβατος της συζήτησης: Ο Βασιλάκης πρέπει να είναι φτιαγμένος έτσι ώστε να μην ξεχωρίζει από τους άλλους ανθρώπους στη συμπεριφορά, με κανέναν τρόπο. Είναι κατασκευασμένος ώστε να λέει ό,τι θα έλεγε κάθε άλλος άνθρωπος και να κάνει ό,τι θα έκανε καθένας από μας. Μόνο που δεν αποτελείται από κύτταρα αλλά από ηλεκτρονικούς μικροεπεξεργαστές.

Συνεπώς, στο λογισμικό του Βασιλάκη δεν μας ενδιαφέρει να κωδικοποιήσουμε καμία Principia Mathematica και κανένα άλλο λογικό σύστημα, παρά μόνο με ακρίβεια το νοητικό σύστημα του ανθρώπου. Διαφορετικά δεν είναι Βασιλάκης, είναι μια περίπλοκη ηλεκτρική σκούπα, και δε νομίζω κανέναν από τους υποστηρικτές της ρομποτικής θεωρίας, που λέει ο Ανδρέας, να υποστηρίζουν ότι ο νους είναι τέτοιου είδους αυτόματο.

Ο Βασιλάκης αυτός ίσως θυμίζει σε κάποιους το κατασκεύασμα στη φιλοσοφία του νου με το όνομα «ζόμπι». Ζόμπι είναι μια φανταστική ύπαρξη, που υπάρχει μόνο στη φιλοσοφική μυθολογία· μια ύπαρξη που υποτίθεται πως έχει όλα τα νοητικά χαρακτηριστικά του ανθρώπου εκτός από συνείδηση.

Είναι η ίδια ανοησία που βλέπουμε στη διαφήμιση του Johnie Walker, που μιλάει και φέρεται σαν άνθρωπος, αλλά «χωρίς συναίσθημα». Είναι ο Dr. Spock και ο Data, ο C3PO και ο κάθε κινηματογραφικός ζωντανός νεκρός με απλανές βλέμα. Νομίζω πως κάτι τέτοιο φαντάζεται ο Ανδρέας όταν λέει ότι ζούμε ερήμην του εαυτού μας και ότι αν τυχόν ο Βασιλάκης έχει υποκειμενικές εμπειρίες, που δεν το πιστεύει, αυτές θα 'ναι «ψευδείς». Τι θα πεί δηλαδή «ψευδείς» υποκειμενικές εμπειρίες; Αν αυτό που έχει είναι και υποκειμενικές και εμπειρίες, τι άλλο χρειαζόμαστε;

Η ιδέα των ζόμπι, αν και ελκυστική για πολλούς δεισιδής, είναι τουλάχιστον προβληματική έως και αδύνατη για κάθε συνεπή μονιστή. Αυτό είναι προφανές μόλις αναλογιστούμε τις πλήρεις συνέπειες των προϋποθέσεων με βάση τις οποίες προσδιορίσαμε το Βασιλάκη. Καταρχήν είπαμε πως ο Βασιλάκης περνά το τεστ Turing. Αυτό σημαίνει πως δεν θα έλεγε ποτέ «εγώ δεν μπορώ να αισθανθώ όπως εσύ», γιατί αυτό θα τον πρόδιδε αμέσως. Αντίθετα, θα ήταν κατασκευασμένος ώστε να κάνει «αχ» στο τσίμπημα, να παίρνει περίλυπη έκφραση όταν αντιλαμβάνεται υπαινιγμό για την κατασκευή του, να αναστενάζει βαριά όταν βλέπει να περνά μια αιθέρια ύπαρξη, να συζητά για τις λεπτές αποχρώσεις του σωματικού και ψυχικού πόνου, και να φέρνει αντιρρήσεις για τη συναισθηματική βαρύτητα του Μιρό ή του Κλέε. Αυτό θα πει «μη διακρίσιμος». Το σωστό ρομπότ λοιπόν, αντίθετα από τον κ. Data, δεν μπορεί να δηλώνει επίγνωση της μη επίγνώσής του, αλλά αντίθετα θα πρέπει να είναι κατασκευασμένο έτσι ώστε να συμπεριφέρεται όπως ακριβώς και όλοι οι υπόλοιποι άνθρωποι. Πρόκειται για μη διαπραγματεύσιμη, εξ' ὄρισμού ιδιότητα του Βασιλάκη, να δηλώνει ότι έχει επίγνωση και οι δηλώσεις αυτές να είναι απόλυτα ανθρώπινες.

Αναφορικά με το Βασιλάκη αυτόν, κανείς μας δεν μπορεί να υποστηρίξει ότι έχει ή δεν έχει υποκειμενικότητα και αυτοσυνείδηση με περισσότερη εμπειρική βεβαιότητα απ' ότι για οποιονδήποτε άλλον από μας. Αυτό είναι κάτι που ρητά δηλώνει κι ο Ανδρέας, άρα δεν αποτελεί κριτική. Η κριτική είναι η αποφυγή του αναγκαίου για μένα συμπεράσματος και η προσφυγή στο μύθευμα του ζόμπι.

Ποιος, και με ποια κριτήρια, μπορεί να αποφασίσει ότι ο Βασιλάκης δεν έχει συνείδηση; Αν μας φταίει το υλικό κατασκευής, δεν θα πρέπει πρώτα να έχουμε μια ιδέα για ποιο λόγο να είναι η συνείδηση κατασκευαστικό ζήτημα; Δηλαδή όποιος ακρωτηριασμένος έχει συναίσθηση του προσθετικού χεριού του ψεύδεται; Ο τυφλός που νιώθει την άκρη του μπαστουνιού ως σημείο έκτασης του σώματος, τι ακριβώς νιώθει; Αν αντικαταστήσουμε ένα νευρώνα στον εγκέφαλο με μικροεπεξεργαστή, έχουμε ακόμα αυτοσυνείδηση; Αν αντικαταστήσουμε δύο, ή τρεις νευρώνες; Στους πόσους νευρώνες πάνω-κάτω υπερσκελίζεται η εμφάνιση και η συμπεριφορά από το κριτήριο του υλικού, ώστε ο άνθρωπος να γίνει ρομπότ; Το ξέρω ότι αυτό είναι φιλοσοφικούρα, και δεν θα σταθώ πολύ, όμως για κείνους που παίρνουν σοβαρά επιχειρήματα τύπου ζόμπι νομίζω θα 'ναι απαραίτητο να δώσουν μια σαφή και τεκμηριωμένη απάντηση και στο πρόβλημα αυτό που αναγκαστικά προκύπτει.

Για μένα ο Βασιλάκης που αποτελείται από επεξεργαστές και όχι κύτταρα, είναι τόσο άνθρωπος όσο ανθρώπινα φέρεται. Αν μου μιλήσει για τη νοσταλγία της πατρίδας, τον πόνο της απώλειας, τον ανοιξιιάτικο έρωτα ή τη μεθυστική μυρωδιά του αγρού, δεν έχω λόγο να μην του αποδώσω τα ίδια υποκειμενικά αντιλήμματα που θεωρώ πως έχει ο καθένας από μας. Προφανώς και θα έχει πολιτικά δικαιώματα καθώς και οποιοδήποτε άλλο δικαίωμα απορρέει από την ανθρώπινη υπόστασή του. Και μια και δεν έχω κόρη για να φέρω αντίρρηση στις σχέσεις της μαζί του (αν υποθέσουμε ότι θα μου έπεφτε λόγος) δηλώνω ότι δεν έχω αντίρρηση να έχουν σχέσεις μαζί του οι κόρες της Ρηνιώς.

Με την έννοια αυτή θεωρώ λοιπόν πως είμαστε όλοι ρομπότ. Είμαστε σαν το Βασιλάκη, φτιαγμένοι από κομματάκια που υπόκεινται στους φυσικούς νόμους και σε τίποτα άλλο. Τα κομματάκια αυτά δεν είναι ηλεκτρονικοί μικροεπεξεργαστές, όπως του Βασιλάκη, αλλά βιολογικά κύτταρα, όμως αυτό δεν έχει καμία σημασία για τα λειτουργικά χαρακτηριστικά τους και για το τι συγκροτούν όταν συνεργάζονται πολλά μαζί σε συγκεκριμένες διατάξεις, όπως είναι ένα ανθρώπινο νευρικό σύστημα. Δε νομίζω πως είναι κανείς σε θέση να καταρρίψει τη θέση για την υποκειμενικότητα του Βασιλάκη χωρίς αναφορά σε ουσιοκρατικά κριτήρια, δηλαδή χωρίς δεισμό. Θα σεβαστώ την επιλογή του οποιουδήποτε να είναι ένας συνεπής δειστής, αν οι συνέπειες των βασικών του θέσεων τον οδηγούν

εκεί. Δεν βλέπω κανέναν άλλον τρόπο απόδοσης συνείδησης με τον οποίο ο Ανδρέας να κρατήσει εμένα και να πετάξει το Βασιλάκη, αν υποθέσουμε ότι επιθυμεί κάτι τέτοιο. Εγώ, ως συνεπής μονιστής, αναλογίζομαι και αποδέχομαι τις συνέπειες των θεμελιωδών επιλογών μου. Οι συνέπειες αυτές περιλαμβάνουν αναγκαστικά ότι είμαστε βιολογικές μηχανές, «ρομπότ», εξελιγμένα όχι κατασκευασμένα, μια και η φυσική επιλογή μπορεί, όπως έδειξε ο Δαρβίνος, να οδηγήσει σε πολύπλοκους οργανισμούς χωρίς σχέδιο ούτε σχεδιαστή.

Αν αφήσουμε στην άκρη τις ατομικές επιλογές για την ουσία και το ρόλο της ψυχής, που περιορίζουν αναγκαστικά την οπτική γωνία από την οποία βλέπει και ερμηνεύει κανείς τα πράγματα, υπάρχει νομίζω και μια επιπλέον δυσχέρεια στη σύγκλιση. Αυτή αφορά στον προσδιορισμό του προς εξήγηση φαινομένου. Το προς εξήγηση φαινόμενο, δηλαδή η συνείδηση, είναι η υποκειμενική όψη της εγκεφαλικής λειτουργίας, αυτό που καθένας αισθάνεται ότι έχει επίγνωση και αυτεπίγνωση. Τι σημαίνει να έχει κανείς συνείδηση; Πώς, τότε και γιατί γίνεται αυτό το πράγμα από τον εγκέφαλο; Ότι και αν είναι αυτό, απαιτεί τη συγκρότηση ενός ενιαίου υποκειμένου ως φορέα της υποκειμενικότητας. Το υποκείμενο αυτό είναι ένα εγκεφαλικό κατασκεύασμα που «αντιλαμβάνεται», «γνωρίζει», «αποφασίζει» και «δρα» – και που αναπαριστά άλλα όμοια εγκεφαλικά κατασκευάσματα που «αντιλαμβάνονται», «αποφασίζουν» κλπ. Το υποκείμενο είναι το προς εξήγηση φαινόμενο, μαζί με ό,τι αναπαράσταση αποτελεί το «περιεχόμενο» του, δηλαδή το περιεχόμενο της συνείδησης.

Θεμελιωδώς, το υποκείμενο και οι αναπαραστάσεις του δεν παρέχει εξηγήσεις ή θεωρίες, παρά μόνο στοιχεία προς εξήγηση. Η οποιαδήποτε υποκειμενική αίσθηση πρέπει να εξηγηθεί, όχι να χρησιμοποιηθεί ως στοιχείο για να εξηγηθεί κάτι άλλο. Για παράδειγμα, η αίσθηση του σώματος, η ενιαία και συνεχής αίσθηση του νοήμονος ιστορικού εαυτού, η αίσθηση της εκούσιας πράξης, η αίσθηση της αντίληψης του οπτικού πεδίου, και όλα τα υπόλοιπα, όλα αυτά είναι το προς εξήγηση φαινόμενο. Δεν είναι λογικό, και δεν στέκει εμπειρικά, να θεωρούμε εξήγηση για μια πράξη ότι «σκέφτηκα και αποφάσισα να δράσω» – αυτό είναι μέρος της ερώτησης, όχι της απάντησης. Δηλαδή, η πλήρης ερώτηση δεν είναι μόνο «πώς ή γιατί κουνήθηκε το χέρι» αλλά «πώς ή γιατί συγκροτήθηκε η υποκειμενική αίσθηση της επιθυμίας, της πρόθεσης, της απόφασης και της εντέλει εκούσιας πράξης και πώς αυτό σχετίζεται, αν σχετίζεται, με τη φυσική κίνηση του χεριού;» Είναι προφανές από τη διατύπωση αυτή ότι οι «εκ των έσω» απόπειρες απάντησης είναι καταδικασμένες να αστοχούν διότι δεν εφάπτονται καν με το ερώτημα, αλλά αποτελούν περαιτέρω στοιχεία προς εξήγηση και περιγραφή. Η μόνη αποδεκτή περιγραφή πρέπει να γίνει με όρους εξωτερικούς του εαυτού, στο επίπεδο της εγκεφαλικής λειτουργίας.

Η πρόταση που βρίσκω πιο ελκυστική, και που προσπάθησα να σκιαγραφήσω πέρυσι, είναι ότι το βασικό στοιχείο της συνείδησης είναι μια όψη της εγκεφαλικής λειτουργίας που συγκροτεί το υποκείμενο, και ότι η όψη αυτή μάλλον έχει να κάνει με μια αναπαράσταση της σχέσης του οργανισμού με το τρέχον εσωτερικό και εξωτερικό πλαίσιο λειτουργίας, ή περιβάλλον. Η ιδέα είναι ότι ένα σύνθετο νευρικό σύστημα δεν έχει ως ρόλο να ανταποκρίνεται απλώς σε περιβαλλοντικά ερεθίσματα, ούτε και να αποθηκεύει ίχνη της ιστορίας του. Αυτά τα απλά πράγματα μπορεί σήμερα να τα κάνει ένα στοιχειώδες συνδεδεστικό δίκτυο, το οποίο σε τίποτα δεν μοιάζει να προσεγγίζει ούτε την πολυπλοκότητα, ούτε την ευελιξία, αλλά το κυριότερο, ούτε τη διαρκή αυτορύθμιση των βιολογικών νευρικών συστημάτων. Το νευρικό σύστημα μοιάζει να προσαρμόζεται διαρκώς σε μια σχέση με το περιβάλλον, αναπαριστώντας και προσδοκώντας τις αλλαγές που προκαλεί η δράση του. Δεν γνωρίζουμε ακόμα πώς συγκροτείται ο ενιαίος φορέας που επιτρέπει στους οργανισμούς να δρουν συγκροτημένα, να περπατούν με όλα τα πόδια προς την ίδια κατεύθυνση, να πιάνουν αυτό που κοιτάζουν, ή να απαντούν σε μια επίθεση με φυγή ή άμυνα αλλά όχι και τα δύο μαζί. Προφανώς είναι σημαντικό να συγκροτείται αυτός ο ενιαίος φορέας. Από τη στιγμή λοιπόν που ένα νευρικό σύστημα εξασφαλίζει μια διαρκή και ενιαία σχέση του οργανισμού με το εσωτερικό και εξωτερικό περιβάλλον, τίθεται το βασικό θεμέλιο για το περιεχόμενο της συνείδησης. Ίσως χρειάζεται κάτι παραπάνω για μια «πλήρη» υποκειμενική επίγνωση, ίσως παίζει ρόλο η γλώσσα, ίσως μια αναπαράσταση του ίδιου του οργανισμού ως μοντέλου πρόγνωσης, ποιος ξέρει.



Αν κάτι τέτοιο πλησιάζει την πραγματικότητα, τότε:

- Πρώτον. Κάθε οργανισμός που διαθέτει στοιχειωδώς σύνθετο νευρικό σύστημα, το οποίο λειτουργεί με αυτόν τον τρόπο, έχει κάποιο βαθμό συνείδησης και υποκειμενικότητας (σίγουρα οι χιμπατζήδες, μάλλον οι σκύλοι, ίσως και τα χταπόδια και οι παπαγάλοι, γιατί όχι και οι κότες).
- Δεύτερον. Από τη στιγμή που οποιοδήποτε μη βιολογικό υπολογιστικό σύστημα μπαίνει σε μια σχέση φυσικής αλληλεπίδρασης με το περιβάλλον και υποστηρίζει τις σχετικές αναπαραστάσεις των σχέσεων, των προσδοκιών κλπ., αυτομάτως θα αναδύεται και η αντίστοιχη υποκειμενικότητα, με τον ίδιο τρόπο και για τον ίδιο λόγο που αναδύεται και στα βιολογικά συστήματα.

Ας ξαναθυμηθούμε το Βασιλάκη. Είναι κατασκευασμένος να αναφέρεται στις πιο λεπτές αντιληπτικές και συναισθηματικές αποχρώσεις, όπως και μεις. Το κάνει με ημιαγώγιμα ηλεκτρονικά ενώ εμείς με βιοχημικά κύτταρα. Αυτό που θέλω να πω είναι πως το σύστημα που «νιώθει» και «μιλάει» είναι το ίδιο σύστημα: μια λειτουργική διάταξη αναπαραστάσης η οποία έχει περιεχόμενο (θέλετε συγκινησιακό, θέλετε αποβλεπτικό, δεν έχει σημασία για τη συζήτηση αυτή). Στην περίπτωση μας, δηλαδή των ανθρώπων, η λειτουργική αυτή διάταξη μπορεί να επικοινωνήσει αναφορικά με το περιεχόμενό της επειδή έχει και γλώσσα: και μπορεί να πείσει τον άλλον ότι όντως έχει το περιεχόμενο αυτό.

Η θέση μου δεν είναι ότι δεν υπάρχει υποκειμενικότητα : αντίθετα, πρόκειται για τη θέση ότι η υποκειμενικότητα ταυτίζεται με μια αναπαραστασιακή κατάσταση της εγκεφαλικής λειτουργίας, καθώς και οποιοδήποτε άλλου υπολογιστικά ισοδύναμου μη βιολογικού συστήματος. Αναλογιστείτε ένα κατασκεύασμα, με προγραμματισμένες νοητικές λειτουργίες που συμπεριλαμβάνουν όλο το φάσμα των αισθημάτων και συναισθημάτων. Ένα κατασκεύασμα που αναπαριστά μια ενιαία σχέση εαυτού με το περιβάλλον και μπορεί ν' αναφερθεί στη σχέση αυτή. Μιλά και φέρεται σαν εμάς. Πώς είναι δυνατό να εξαρτάται, το τι πιστεύουμε για την υποκειμενική του επίγνωση, από το τι πιστεύουμε για τη διαδικασία κατασκευής του: αν δηλαδή τα κύτταρα αναπτύχθηκαν σε κύηση ή τα ένωσε ένας μελλοντικός βιολόγος-μηχανικός, ή αν αντί για κύτταρα έχει κεραμικά, κυκλώματα άνθρακα ή πυριτίου. Η απόδοση υποκειμενικότητας είναι ζήτημα λειτουργικό, όχι κατασκευαστικό. Και με την έννοια αυτή είμαστε όλοι βιολογικά ρομπότ, των οποίων οι γενετικά προκαθορισμένες νοητικές ικανότητες περιλαμβάνουν δυνητικά και την ικανότητα να οργανώνουμε συζητήσεις για τη συνείδηση. Ίσως και την ικανότητα να καταλάβουμε πώς γίνεται αυτό.

Τίποτα από τα παραπάνω δεν αποτελεί απόδειξη, ή ένδειξη, πως «η ρομποτική θεωρία», όπως την αποκαλεί ο Ανδρέας, είναι σωστή. Ελπίζω όμως πως έδειξα ότι η περυσινή επίθεση του Ανδρέα αποτελείται από μια σειρά ασύνδετων παρανοήσεων και δεν υπονομεύει στο παραμικρό τη θέση ότι είμαστε βιολογικές μηχανές, περίπλοκα οργανικά αυτόματα, και ότι η ανθρώπινη συνείδηση είναι μια ιδιότητα της εγκεφαλικής λειτουργίας, όψη της λειτουργίας κυττάρων σύμφωνα με τους φυσικούς νόμους και τίποτα παραπάνω. Μένει στους υποστηρικτές της ρομποτικής θεωρίας να δείξουν τι ακριβώς σημαίνει αυτό και πώς γίνεται και στους πολέμιούς της να συγκροτήσουν κάποιο ουσιαστικό αντίθετο επιχείρημα παρά να δηλώνουν απλώς τη δυσαρέσκειά τους για τις αναπόφευκτες συνέπειες που έχει η αποδοχή της «ρομποτικής θεωρίας».