

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ



**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΤΟ ΘΕΩΡΗΜΑ ΤΟΥ GÖDEL.
ΑΠΟΔΕΙΞΗ, ΣΥΝΕΠΕΙΕΣ ΚΑΙ
ΠΡΟΕΚΤΑΣΕΙΣ ΣΤΗ ΦΙΛΟΣΟΦΙΑ ΤΟΥ ΝΟΥ**

ΤΖΑΝΗΣ ΔΗΜΗΤΡΙΟΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΚΟΛΕΤΣΟΣ Γ.

**ΑΘΗΝΑ
ΟΚΤΩΒΡΙΟΣ 2005**

*Μπορείς να βρεις την αλήθεια με τη λογική
μόνο αν την έχεις ήδη βρει χωρίς αυτή.*

G.K. Chesterton

ΠΕΡΙΕΧΟΜΕΝΑ

ΜΕΡΟΣ Ι

A. Εισαγωγή	5
B. Ιστορική Αναδρομή	7

ΜΕΡΟΣ ΙΙ

1. Γενικά στοιχεία	13
2. Βασικά εργαλεία της απόδειξης	15
2.1 Στοιχεία Μαθηματικής Λογικής	15
2.1.1 Προτασιακή Λογική	15
2.1.2 Προτασιακός Λογισμός	17
2.1.3 Κατηγορηματική Λογική	19
2.1.4 Κατηγορηματικός Λογισμός	23
2.1.5 Ανακεφαλαίωση και μία σημαντική εφαρμογή	26
2.2 Αναδρομικές συναρτήσεις και κατηγορήματα	32
2.2.1 Εισαγωγικοί ορισμοί	32
2.2.2 Ορισμός της κλάσης των πρωταρχικά αναδρομικών και της κλάσης των αναδρομικών συναρτήσεων	33
2.2.3 Επιπλέον ορισμοί αναδρομικών συναρτήσεων και ένας γενικός κανόνας	35
2.3 Αναπαραστάσιμες συναρτήσεις και κατηγορήματα	38
3. Κύριο μέρος της απόδειξης	41
3.1 Αριθμητικοποίηση της Λογικής	41
3.2 Αναδιατύπωση και απόδειξη του 1 ^{ου} θεωρήματος μη πληρότητας	45
3.3 Το 2 ^ο θεώρημα μη-πληρότητας	52
4. Συνέπειες, εφαρμογές των θεωρημάτων μη-πληρότητας και άλλα σχετικά αποτελέσματα	55
4.1 Συνέπειες μαθηματικές και μη	55
4.2 Ανάλογα αποτελέσματα του θεωρήματος του Gödel	59
4.3 Αρχική μορφή του θεωρήματος του Gödel	64

ΜΕΡΟΣ ΙΙΙ

1. Εισαγωγή	67
2. Περίοδος προ Lucas	69
2.1 A.M. Turing ‘Computing Machinery and Intelligence (1950)	69
2.2 Hilary Putnam ‘Minds and Machines’ (1960)	76
3. Το βασικό επιχείρημα του J.R. Lucas	81
3.0 Ένα σχόλιο πάνω στη χρήση του θεωρήματος μη-πληρότητας και ένα συναφές παράδειγμα	81
3.1 J.R. Lucas ‘Minds and Machines’ (1961)	83
3.2 Η απάντηση στον H. Putnam και μία διαφορετική διατύπωση του επιχειρήματος	90
4. Πρώτη περίοδος αντιπαράθεσης (1961 – 1988)	95
4.1 Αντιρρήσεις πάνω στη βάση των υποθέσεων του J.R. Lucas	95
4.1.1 Περί της φύσης των μηχανών και της αξιοπιστίας της συλλογιστικής του Lucas	96
4.1.2 Περί των ορίων της διαλεκτικής προσέγγισης	97
4.1.3 Σύνδεση με το κατασκευαστικό πρόβλημα στα θεμέλια των μαθηματικών	98
4.2 Μία διαφορετική προσέγγιση των συνεπειών του θεωρήματος του Gödel στη μηχανιστική θέση	101
4.2.1 Paul Benacerraf ‘God, the Devil and Gödel’ (1967)	101
4.2.2 Η απάντηση του J.R. Lucas	108
4.2.3 Μία συνολική κριτική στο νέο πρόβλημα και ένα νέο συμπέρασμα	110
4.3 Ανακεφαλαίωση και μία σύντομη παρουσίαση κάποιων επιπλέον αντιρρήσεων	113
5. Οι απόψεις του Gödel για το θέμα	117
6. Δεύτερη περίοδος αντιπαράθεσης (1989 – σήμερα)	123
6.1 Το πρώτο επιχείρημα του R. Penrose και οι αδυναμίες του	127
6.2 Το δεύτερο επιχείρημα του ιδίου και οι σχετικές αντιρρήσεις	131
6.3 Πιο σύγχρονες και πιο απόμακρες προεκτάσεις του θ. του G.	140
7. Τελικά συμπεράσματα	143
<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>	151
<u>ΠΑΡΑΡΤΗΜΑ</u>	155

ΜΕΡΟΣ Ι

Α. Εισαγωγή

Η παρούσα διπλωματική εργασία έχει σαν βασικό θέμα επεξεργασίας το ‘Θεώρημα του Gödel’ και αποτελείται από τρία μέρη. Στο πρώτο αυτό μέρος της εργασίας συμπεριλαμβάνονται κάποιες βασικές εισαγωγικές έννοιες μαζί με μία σύντομη ιστορική αναδρομή, στο δεύτερο μέρος θα παρουσιαστεί αναλυτικότερα το θεώρημα και η απόδειξή του και τέλος, στο τρίτο μέρος θα παρουσιαστεί μια σειρά επιχειρημάτων που συνδέουν το θεώρημα με τη φιλοσοφία, και πιο συγκεκριμένα τη φιλοσοφία του νου.

Ένα από τα πρώτα πράγματα που πρέπει να εξηγηθεί σε αυτή την εισαγωγή είναι σε πιο αποτέλεσμα αναφερόμαστε με τον όρο ‘Θεώρημα του Gödel’. Συνήθως πρόκειται για το **1^ο Θεώρημα μη-πληρότητας** στο οποίο αποδεικνύει ο Gödel την ύπαρξη μη-αποκρίσιμων [undecidable - που δεν μπορούμε να αποφανθούμε αν είναι αληθής ή ψευδής] προτάσεων σε κάθε τυπικό σύστημα που πληροί κάποιες προϋποθέσεις, καθιστώντας το σύστημα αυτό μη-πλήρες. Ένα πόρισμα αυτού του θεωρήματος που καλείται **2^ο Θεώρημα μη-πληρότητας** χρησιμοποιεί το παραπάνω αποτέλεσμα για να αποφανθεί περί της δυνατότητας απόδειξης της συνέπειας [consistency - η ιδιότητα να μη μπορεί να αποδειχθεί μία αντίφαση δηλ. ότι ταυτόχρονα ισχύει μία πρόταση και η άρνησή της] τέτοιων τυπικών συστημάτων. Για τους σκοπούς της εργασίας αυτής, όταν αναφερόμαστε στο ‘Θεώρημα του Gödel’ θα εννοούμε κατά βάση το 1^ο θεώρημα μη-πληρότητας, ενώ για το πόρισμα περί της συνέπειας θα αναφέρεται ξεχωριστά όποτε χρειάζεται.

Για να δοθεί τώρα μία συνολική εικόνα του πλαισίου στο οποίο κινείται η παρούσα διπλωματική εργασία, αμέσως παρακάτω αναλύονται σε πολύ γενικές γραμμές τα τρία μέρη της και αναφέρεται συνοπτικά ο σκοπός που εξυπηρετεί το καθένα.

Σε αυτό το 1^ο μέρος, πέραν των εισαγωγικών στοιχείων, θα παρατεθεί παρακάτω και μία σύντομη ιστορική αναδρομή η οποία έχει σα στόχο να μπορέσουμε να δούμε από ποια λογική παράδοση ξεκίνησε η πορεία που οδήγησε στα θεωρήματα

του Gödel, ποιο ήταν το αντίκτυπό τους σε αυτή την παράδοση, και πού οδηγήθηκαν τα πράγματα από εκεί και πέρα. Η αναδρομή χαρακτηρίζεται σύντομη διότι όλα αυτά εκτυλίχθηκαν σχετικά πρόσφατα και κυρίως τις δεκαετίες του 1920 και '30, οπότε η μαθηματική λογική πήρε τη σύγχρονη μορφή της και τα θέματα που προέκυψαν μία πιο ξεκάθαρη διάσταση.

Όσον αφορά το 2^ο μέρος τα πράγματα είναι αρκετά ξεκάθαρα. Πρόκειται για μία παρουσίαση των εμπλεκόμενων μαθηματικών εννοιών και της τυπικής απόδειξης καθενός από τα αποτελέσματα που μας ενδιαφέρουν. Σαν στόχος τίθεται η απόλυτη κατανόηση, όχι μόνο των θεωρημάτων, αλλά και του τρόπου με τον οποίο αυτά αποδεικνύονται. Διότι, όντας η απόδειξή τους κατασκευαστική, μας δίνει, πέρα από ένα μαθηματικό εργαλείο και ένα εργαλείο γενικότερης σημασίας που μπορεί να χρησιμοποιηθεί για την εξαγωγή και άλλων συμπερασμάτων, πέρα των μαθηματικών.

Αυτό το γεγονός είναι που συνδέει όλα αυτά που φαίνονται να απασχολούν κυρίως το μαθηματικό κόσμο με τη φιλοσοφία του νου, όπως θα δούμε στο 3^ο και τελευταίο μέρος της εργασίας. Συγκεκριμένα, θα ακολουθήσουμε την πορεία ενός επιχειρήματος που εκφράστηκε για πρώτη φορά το 1961 από τον Άγγλο φιλόσοφο J.R. Lucas και που ήταν η αφορμή να αρχίσει μία εκτεταμένη αρθρογραφία που συνεχίζεται μέχρι τις μέρες μας και σχετίζεται με τις πιο φιλοσοφικές προεκτάσεις του θεωρήματος του Gödel. Τελικός σκοπός είναι να διαπιστώσουμε μέχρι που μπορεί να προεκταθεί αυτό το μαθηματικό αποτέλεσμα και τι υλικό έχει να προσφέρει για σκέψη πάνω σε φιλοσοφικά αλλά και μαθηματικά ζητήματα.

Σημειώσεις: Όσον αφορά το τυπικό μέρος της παρουσίασης που θα ακολουθήσει, σημειώνονται τα παρακάτω:

Για την **ορολογία** επιλέγονται οι ελληνικοί όροι αλλά αναφέρονται μέσα σε [αγκύλες] οι αντίστοιχοι αγγλικοί, οι οποίοι είναι πιο διαδεδομένοι και γενικά λιγότερο διφορούμενοι από μερικούς αντίστοιχους ελληνικούς. Όταν συναντώνται έννοιες που έχουν διττή ερμηνεία στα ελληνικά, θα αναφέρονται και οι προτεινόμενες από την βιβλιογραφία εναλλακτικές διατυπώσεις.

Η **βιβλιογραφία** που βρίσκεται στο τέλος της εργασίας είναι ταξινομημένη χρονολογικά και οι αναφορές σε αυτή θα γίνονται με τη χρήση του *αύξοντος αριθμού* που αντιστοιχεί σε κάθε αναφορά, εγκλεισμένου σε άγκιστρα, π.χ. {#04}.

B. Ιστορική Αναδρομή

Η αποσαφήνιση της δομής της ανθρώπινης σκέψης είναι ένα καθαρά φιλοσοφικό πρόβλημα το οποίο είναι δύσκολο να απαντηθεί συλλήβδην. Το ερώτημα όμως μπορεί να αναδιατυπωθεί κατά περίπτωση των διαφόρων επιστημών, ούτως ώστε να μπορεί να πάρει μια πιο σαφή και κατανοητή μορφή. Μία από τις καλύτερες υποψήφιος, και αυτή για την οποία η απάντηση θα ήταν καθοριστικής σημασίας, είναι αναμφίβολα τα μαθηματικά. Πράγματι, η μαθηματική σκέψη μπορεί πιο φυσικά από οποιαδήποτε ίσως άλλη, να θεωρηθεί δομημένη πάνω σε αυστηρά καθορισμένες αρχές από τις οποίες θα μπορούμε, μέσω μαθηματικών συλλογισμών, να παράγουμε ασφαλή συμπεράσματα. Μπορούμε όμως να παράγουμε με κάποιον τέτοιο τρόπο *όλα* τα ορθά μαθηματικά συμπεράσματα, ήτοι όλες τις μαθηματικές αλήθειες;

Το πρόγραμμα που είχε προτείνει ο Leibnitz για τα μαθηματικά ήταν η αρχή της αναζήτησης ενός τέτοιου καλού τυπικού συστήματος που θα μας επιτρέψει δίχως παρέκκλιση να αποφαινόμαστε για όλες τις μαθηματικές προτάσεις την αλήθεια ή το ψεύδος τους. Συγκεκριμένα, ο Leibnitz οραματίστηκε τη δημιουργία μίας τυπικής γλώσσας που θα περιέγραφε όλες τις μαθηματικές έννοιες και προτάσεις, μαζί με τη δημιουργία μιας *συνεπούς* μαθηματικής θεωρίας που θα μπορούσε να αποδείξει την αλήθεια όλων των μαθηματικών προτάσεων. Η προφανής αντιστοίχιση της σκέψης με τις αναγωγές σε ένα τυπικό σύστημα που υπονοείται από το πρόγραμμα, μαζί με την άποψη ότι η λογική είναι η μελέτη των νόμων που διέπουν τον νου, (εκφρασμένη ποικιλοτρόπως από τους Boole, Frege και Gentzen, μεταξύ άλλων), οδηγούν τον Γερμανό μαθηματικό David Hilbert να ξεκινήσει ο ίδιος ένα αντίστοιχο πρόγραμμα για να φτάσει τελικά στην απόλυτη και αδιαπραγμάτευτη Θεμελίωση των Μαθηματικών.

Το φιλόδοξο αυτό σχέδιο ξεκινάει από τις αρχές του προηγούμενου αιώνα και έχει μείνει στην ιστορία σαν 'Πρόγραμμα του Hilbert'. Οι ακριβείς απαιτήσεις αυτού του εγχειρήματος είναι μεν περίπλοκες, αλλά φωτίζουν καλύτερα τους στόχους του Hilbert, οπότε θα γίνει μία προσπάθεια αποσαφήνισης των εμπλεκόμενων ιδεών και εννοιών αμέσως παρακάτω. Να σημειωθεί επίσης ότι το 'Πρόγραμμα' δεν ήταν ένα μόνο πράγμα ή μία μόνο ιδέα αλλά ένα σύμπλεγμα ιδεών και φιλοσοφικών τάσεων το οποίο όμως είχε έναν πολύ σαφή κεντρικό προσανατολισμό, τον οποίο και ενδιαφερόμαστε να προσεγγίσουμε μέσω της ανάλυσης που ακολουθεί.

Το Πρόγραμμα του Hilbert

Μετά τις αποτυχημένες προσπάθειες θεμελίωσης των πρωταρχικών αντικείμενων και των συλλογιστικών μεθόδων μέσω της πρώιμης Θεωρίας Συνόλων, στα τέλη του 19^{ου} αι., προέκυψε εύλογα ο προβληματισμός που οδηγεί στην κεντρική ιδέα πίσω από το Πρόγραμμα. Το ερώτημα ήταν πως θα απαλλαχθεί μία θεωρία που θεμελιώνει τα μαθηματικά από τις αντιφάσεις και τα παράδοξα, όχι μόνο επειδή δεν θα τα ανακάλυπτε κανείς τη δεδομένη ιστορική στιγμή αλλά διότι θα είχε αποδειχθεί ότι αυτά δε μπορούν να προκύψουν.

Προέκυψε έτσι η κεντρική αναζήτηση για μία *απόδειξη* της συνέπειας των Μαθηματικών, τα οποία χωρίζονται πλέον σε ‘πραγματικά’ [real] και ιδεατά [ideal] ανάλογα με τις μεθόδους που χρησιμοποιούν για να εξάγουν συμπεράσματα. Για να εξηγήσουμε όμως τα κριτήρια τα οποία διαχωρίζουν τις ιδεατές από τις πραγματικές μεθόδους και να γίνουν πιο σαφείς οι στόχοι του Προγράμματος, θα πρέπει να κάνουμε μία νύξη στην πρακτική υλοποίησή των βασικών του ιδεών, η οποία ξεκίνησε τις αρχές του 20^{ου} αι. και το 1921 περιλαμβάνει δύο βασικές απαιτήσεις:

1. Την τυποποίηση όλων των μαθηματικών σε *αξιοματική μορφή*.
2. Την απόδειξη της συνέπειας αυτής της αξιωματικοποίησης με χρήση μόνο *περατοκρατικών* [finitary] μεθόδων.

Αυτές οι δύο απαιτήσεις βέβαια προϋποθέτουν κάποια *πρωταρχικά αντικείμενα* μέσω των οποίων θα ορίζονται τα αξιώματα και τα οποία τελικά θα επεξεργαζόμαστε μέσω των όποιων μεθόδων. Ο Hilbert επέλεξε γι’ αυτό το σκοπό τα *σημάδια* [signs], αντικείμενα τα οποία δεν υπεισέρχονται σε λογικές σχέσεις, είναι αναγνωρίσιμα ανεξαρτήτως χώρου, χρόνου ή μεθόδων παραγωγής τους και έχουν αντικειμενικές ιδιότητες. «Διαισθητικά παρόντα σαν άμεση εμπειρία προηγούμενη κάθε σκέψης», όπως έγραφε ο ίδιος το 1926.

Το επόμενο βασικό θέμα, αυτό της τυποποίησης, καλύφθηκε ικανοποιητικά από την έκδοση του συγγράμματος Principia Mathematica [PM, από εδώ και στο εξής] των Whitehead και Russel. Αυτή η μεγαλειώδης συγγραφική προσπάθεια πέτυχε όντως να ανταποκριθεί κατά μεγάλο βαθμό στις ανάγκες του Hilbert για μία αυστηρή τυποποίηση της λογικής, τυποποίηση η οποία δεν θα περιελάμβανε καμία εγγενή αντίφαση και θα επέτρεπε έτσι την απρόσκοπτη επεξεργασία θεωριών.

Έχοντας λοιπόν στα χέρια μας τα βασικά εργαλεία που θα χρειαστούμε, προχωρούμε στην εξήγηση των καθ' εαυτών απαιτήσεων. Όσον αφορά την πρώτη, αποζητούμε ουσιαστικά την αναγωγή μιας θεωρίας σε ένα σύστημα *συνεπών και ανεξάρτητων μεταξύ τους αξιωμάτων*, σύστημα από το οποίο θα παράγουμε όλες τις αποδείξιμες προτάσεις (και άρα όλες τις *αληθείς* προτάσεις) μιας θεωρίας, μέσω κατάλληλα ορισμένων κανόνων συνεπαγωγής. Ο Hilbert πίστευε πως ο πρέπων τρόπος για να εξελίξουμε κάθε επιστημονικό πεδίο απαιτεί μία τέτοια αξιωματική θεώρηση που σαν απώτερο στόχο θα είχε την εξάλειψη κάθε ανάγκης για *διαίσθηση* [intuition] και, κατά συνέπεια, τη διευκόλυνση της ανάλυσης των βασικών εννοιών και αξιωμάτων. Η διαίσθηση αποτελεί και ένα παράδειγμα ιδεατής συλλογιστικής μεθόδου, άρα, με βάση και τα παραπάνω, η εξάλειψή της συμπίπτει με την αναγωγή της σε μεθόδους των πραγματικών μαθηματικών.

Αυτές οι μέθοδοι των πραγματικών μαθηματικών είναι ακριβώς αυτές οι οποίες χαρακτηρίζονται ως περατοκρατικές και χρησιμοποιούνται *αποκλειστικά* στα πραγματικά μαθηματικά. Σαφής ορισμός για την έννοια της περατοκρατικότητας δυστυχώς *δεν υπάρχει*, η μόνη ξεκάθαρη απαίτηση που προβάλλει ο Hilbert είναι η *απαγόρευση εισαγωγής στις περατοκρατικές μεθόδους συλλογισμού αφηρημένων* [abstract] *εννοιών*. Σύμφωνα με τα σημερινά δεδομένα, θα μπορούσαμε να πούμε ότι, η έννοια που πλησιάζει περισσότερο στην περιγραφή της περατοκρατικότητας είναι η έννοια της πρωταρχικής αναδρομής [primitive recursion - βλ. Μέρος II σελ. 33 για περισσότερες λεπτομέρειες πάνω στην έννοια της (πρωταρχικής) αναδρομής].

Συνδυάζοντας όλα τα παραπάνω, και χωρίς περισσότερες λεπτομέρειες, καταλήγουμε στο ότι οι προτάσεις των πραγματικών μαθηματικών, οι οποίες προκύπτουν από την περατοκρατική επεξεργασία ενός αξιωματικού συστήματος, είναι όλες οι προτάσεις της μορφής $\forall x S(x)$, όπου S μία αναδρομική σχέση. Η κλάση προτάσεων που δημιουργείται αντιστοιχεί στην κλάση Π_1^0 της αριθμητικής ιεραρχίας που χρησιμοποιούμε σήμερα.

Επανερχόμαστε τώρα στο βασικό μας θέμα για να διαπιστώσουμε ότι ο βασικός στόχος του Hilbert ήταν τελικά η *πλήρης αναγωγή των ιδεατών μαθηματικών στα πραγματικά*. Προϋπόθεση βέβαια αυτής της αναγωγής ήταν η αντίστοιχη αναγωγή αρχικά των ιδεατών μεθόδων στις πραγματικές, αναγωγή η οποία συνίσταται στην απόδειξη του ότι: *Οποιαδήποτε πραγματική πρόταση που μπορεί να αποδειχθεί με τη βοήθεια ιδεατών μεθόδων, μπορεί να αποδειχθεί και στα πραγματικά μαθηματικά*.

Η αναγωγή που αναφέρθηκε πιο πάνω έμεινε στην ιστορία σαν η αναζήτηση για την ‘Καθαρότητα των Μεθόδων’ [purity of methods] και αναπαρίσταται σχηματικά ως εξής:

$$\boxed{\text{Ιδεατά μαθηματικά} \mapsto \phi} \quad \Rightarrow \quad \boxed{\text{Πραγματικά μαθηματικά} \mapsto \phi}$$

όπου ϕ θα είναι μία οποιαδήποτε πραγματική (Π_1^0) πρόταση. Σημειώστε επίσης ότι η απαίτηση για την Καθαρότητα των μεθόδων είναι απολύτως ισοδύναμη⁽¹⁾ με την *απαίτηση μίας απόδειξης της συνέπειας των (πραγματικών ή ιδεατών) μαθηματικών με τη χρήση μόνο περατοκρατικών μεθόδων.*

Τώρα που είναι πιο ξεκάθαρο το πλαίσιο στο οποίο κινείται το Πρόγραμμα, γίνεται, πιστεύω, επίσης πιο ξεκάθαρο ότι, μέσω *αυτού* του πλαισίου, το οποίο δεν θα περιελάμβανε τίποτα που να μπορεί να μας οδηγήσει στην ανάγκη χρήσης αφηρημένων (όπως η διαίσθηση) εννοιών, η απόλυτη και αδιαπραγμάτευτη Θεμελίωση των Μαθηματικών φαντάζει ένας άμεσα εφικτός στόχος, από όποια μαθηματικά (ιδεατά ή πραγματικά) κι αν προέρχεται.

Ας προσπαθήσουμε όμως τώρα να δώσουμε ένα σύντομο παράδειγμα εφαρμογής των παραπάνω σε μία θεωρία. Διαλέγουμε να δομήσουμε την περατοκρατική θεωρία αριθμών για την οποία ο ίδιος ο Hilbert πίστευε ότι προϋποτίθεται από *κάθε* ανθρώπινο νου και για έναν ακόμη λόγο που θα γίνει εμφανής αργότερα. Καταρχήν, για πρωταρχικά της αντικείμενα διαλέγουμε τα αριθμητικά [numerals], τα οποία αποτυπώνονται ως |, ||, ||| κτλ. (είναι προφανής η αντιστοιχία με τους φυσικούς αριθμούς). Για αξιωματικό σύστημα επιλέγεται η αριθμητική του Robinson⁽²⁾ (βλ. Μέρος II, σελ. 29) και σαν λογικό σύστημα ο Κατηγορηματικός Λογισμός, ένα πλήρες σύστημα λογικής κατά τα πρότυπα του Hilbert. Είναι, πιστεύω, φανερό και από την πολύ στοιχειώδη δομή των αξιωμάτων του Robinson, ότι μία τέτοια θεωρία είναι πράγματι ‘καθαρή’ κατά τα πρότυπα του Προγράμματος. Παρατηρείστε όμως ότι η απαίτηση για την αυτή την καθαρότητα είναι ισοδύναμη με την απαίτηση για την απόδειξη της συνέπειας ακόμα και μίας *ιδεατής* θεωρίας αριθμών, όπως είναι η αριθμητική του Peano...

(1): βλ. σελ. 125 για μία πιο εμπεριστατωμένη αιτιολόγηση αυτής της ισοδυναμίας.

(2): Η συγκεκριμένη επιλογή δεν είναι γενικά αποδεκτό ότι πληροί τις προϋποθέσεις του Hilbert αλλά τις πλησιάζει αρκετά.

Η παρέμβαση του Gödel

Εδώ είναι που μπαίνει στην εικόνα ο Gödel, ο οποίος τον Σεπτέμβριο του 1930 ανακοινώνει σε ένα συνέδριο το 1^ο θεώρημα μη-πληρότητας, το οποίο καθιστά το αξιωματικό σύστημα της αριθμητικής του Peano (PA, από εδώ και στο εξής) εγγενώς μη-πλήρες. Ο J. von Neumann, που ήταν στο ακροατήριο, καταλαβαίνει τη σημασία του αποτελέσματος για το πρόγραμμα του Hilbert και βρίσκει σαν πόρισμα, ανεξάρτητα από τον Gödel, το 2^ο θεώρημα μη-πληρότητας, το οποίο δείχνει *ότι το σύστημα PA δεν αποδεικνύει την τυποποίηση του ισχυρισμού ότι το PA είναι συνεπές*. Ο Bernays, συνεργάτης του Hilbert από το 1917, παραδέχεται και αυτός ότι, εφόσον η περατοκρατική συλλογιστική είναι τυποποιήσιμη στο PM (γεγονός που προσκρούει στην ασάφεια της έννοιας της περατοκρατικότητας, άρα δεν είναι αποδεκτό χωρίς προβλήματα), τότε *μια απόδειξη της συνέπειας του PA με περατοκρατικές μεθόδους είναι αδύνατη*. Η Καθαρότητα των Μεθόδων φαίνεται να είναι αδύνατο να επιτευχθεί.

Παρ' όλα αυτά, ο ίδιος ο Gödel τελειώνει το άρθρο του 1931 στο οποίο αποδεικνύει τα δύο θεωρήματα μη-πληρότητας {#01} ως εξής: «Θέλω να σημειώσω ιδιαίτερα ότι το Θεώρημα XI (Σημ: εννοεί το 2^ο θεώρημα μη-πληρότητας) δεν αντιβαίνει στην τυποποιημένη (φορμαλιστική) αντίληψη του Hilbert. Διότι αυτή η αντίληψη προϋποθέτει μόνο την ύπαρξη μιας απόδειξης για τη συνέπεια στην οποία μόνο περατοκρατικές μέθοδοι απόδειξης χρησιμοποιούνται, και είναι πιθανό να υπάρχουν περατοκρατικές αποδείξεις που *δεν μπορούν* να εκφραστούν στο φορμαλισμό του PA». Ίσως με αυτόν τον τρόπο ο Gödel να ήθελε να υπερασπίσει το Πρόγραμμα του Hilbert, το οποίο στη βάση του άγγιζε και τις δικές του φιλοσοφικές ανησυχίες όσον αφορά τα Θεμέλια των Μαθηματικών, αλλά κατά γενική ομολογία, το Πρόγραμμα με την κλασική του μορφή είχε καταρρεύσει.

Από εκεί και πέρα, στην 'μετά Gödel εποχή', έγιναν και συνεχίζουν να γίνονται προσπάθειες αναβίωσης των ιδεωδών του Hilbert σε ένα διαφορετικό βέβαια πλαίσιο που καλείται 'Σχετικοποιημένο Πρόγραμμα του Hilbert', προσπάθειες που βασίζονται κατά ένα μεγάλο μέρος και στην ασάφεια της έννοιας της περατοκρατικότητας. Επίσης συνεχίζονται και οι σχετικές *προσπάθειες έρευνας νέων αξιωμάτων* τα οποία θα δομούν συστήματα που θα αποφεύγουν τη μη-πληρότητα. Οι ανησυχίες του Gödel κινήθηκαν και προς αυτή την κατεύθυνση, όπως εκφράζονται

στο άρθρο του *‘Τι είναι η υπόθεση του συνεχούς του Cantor;’*, όπου αναλύει την ανάγκη ύπαρξης νέων αξιωμάτων για τη Θεωρία Συνόλων και καταδεικνύει με ποιον τρόπο, εποπτικά, μπορούμε να διαπιστώσουμε την εγκυρότητά τους.

Να αναφέρουμε σε αυτό τη σημείο ότι υπήρχε και διαδόθηκε ευρέως εκείνη την περίοδο και μία φιλοσοφική τάση η οποία *a priori* δεν δεχόταν τίποτα από τα παραπάνω, και συγκεκριμένα δεν απέδιδε νόημα στο να ψάχνει κανείς τα Θεμέλια των Μαθηματικών. Αυτή η τάση με πιο γνωστό θιασώτη της τον Brouwer, είναι ο ιντουισιονισμός [intuitionism - συναντάται και σαν ‘εποπτεία’] ο οποίος θεωρεί τις μαθηματικές οντότητες κατασκευάσματα των ανθρώπων-μαθηματικών και όχι αντικείμενα υπαρκτά ανεξάρτητα από τον ανθρώπινο νου. Όπως θα διαπιστώσουμε αργότερα (κεφ.5 του Μέρους III), αυτή η τάση βρίσκεται στην ακριβώς αντίπερα όχθη των πεποιθήσεων του Gödel.

Για να ολοκληρώσουμε όμως την ιστορία και να προχωρήσουμε παρακάτω, να αναφέρουμε ότι το 1936 ο Gentzen *απέδειξε* την συνέπεια της πρωτοβάθμιας αριθμητικής του Peano, *αλλά χρησιμοποίησε μεθόδους μη-τυποποιήσιμες σ’ αυτή*, συγκεκριμένα, υπερπεπερασμένη επαγωγή [transfinite induction] και τον διατακτικό αριθμό ϵ_0 . Τέτοιου είδους μέθοδοι όμως επ’ ουδενί δεν μπορούν να χαρακτηριστούν ως περατοκρατικές και άρα δεν αλλάζει το αντίκτυπο των θεωρημάτων μη-πληρότητας που αναφέρθηκε νωρίτερα.

Κατά γενική ομολογία λοιπόν το όραμα του Hilbert είχε καταρριφθεί και ίσως τελικά να μην αποτελεί μεγάλη έκπληξη το γεγονός ότι μια θεωρία που ήθελε να αναγάγει όλους τους μαθηματικούς συλλογισμούς στη δική της (περατοκρατική) συλλογιστική αποδείχθηκε ανεπαρκής. Διότι, μπορεί τα μαθηματικά είναι μια αυστηρά δομημένη επιστήμη, όμως με την αναζήτηση ολόκληρης της αλήθειας που *δυναμικά* περιέχουν να γίνονται τόσο πολύπλοκα ώστε τελικά η πολυπλοκότητά τους να ανάγεται στην πολυπλοκότητα της ανθρώπινης σκέψης. Το να ισχυριζόμαστε όμως ότι έχουμε βρει μοντέλο/αλγόριθμο ή οτιδήποτε άλλο που να περιγράφει ακριβώς *ολόκληρη* την ανθρώπινη σκέψη είναι κάτι που, μέχρι σήμερα τουλάχιστον, ακούγεται υπερφιλόδοξο, για να μην πούμε τελείως αδύνατο. Αυτή η άποψη είναι που άπτεται ακριβώς της συζήτησης που θα παρουσιαστεί στο τελευταίο μέρος αυτής της εργασίας, υπό το πρίσμα των θεωρημάτων μη-πληρότητας και της ‘μηχανιστικής θέσης’ [mechanist thesis], κάτι ανάλογο του προγράμματος του Hilbert αλλά εφαρμοσμένο στην Τεχνητή Νοημοσύνη [Artificial Intelligence – AI].

ΜΕΡΟΣ II

1. Γενικά στοιχεία

Το έτος 1931 ο αυστριακός μαθηματικός Kurt Gödel δημοσίευσε το άρθρο του *‘Για τις τυπικές μη-αποκρίσιμες προτάσεις του Principia Mathematica και συγγενικών συστημάτων I’* {#01}, άρθρο που περιέχει τα περίφημα θεωρήματα μη-πληρότητας. Το άρθρο αυτό περιέχει, εκτός από τις αποδείξεις των δύο θεωρημάτων, και άλλα συγγενή αποτελέσματα από τα οποία τα περισσότερα μπορούν να θεωρηθούν σαν λήμματα που βοηθάνε στην απόδειξη του κύριου αποτελέσματος (1^ο θεώρημα μη-πληρότητας), καθώς και διάφορες συνέπειες του αποτελέσματος αυτού. Όσον αφορά τον αριθμό ‘I’ ο οποίος εμφανίζεται στον τίτλο του άρθρου, ο Gödel αρχικά σχεδίαζε να επεκταθεί σε εφαρμογές πάνω και σε άλλα συστήματα για να ισχυροποιήσει το αποτέλεσμα του αλλά η άμεση αποδοχή της επιστημονικής αρτιότητας που διέπει το άρθρο από τη μαθηματική κοινότητα τον έκανε να αλλάξει γνώμη.

Στις μέρες μας βέβαια σπάνια χρησιμοποιείται η αρχική μορφή της απόδειξης των θεωρημάτων καθότι με το πέρασμα των ετών τα αρχικά αποτελέσματα ισχυροποιήθηκαν, επεκτάθηκαν και εφαρμόζονται σε πιο ‘ασθενή’ συστήματα και η απόδειξη απέκτησε συνολικά μια πιο σύγχρονη μορφή. Μία από αυτές τις μορφές που χρησιμοποιούν την τρέχουσα σημειογραφία της μαθηματικής λογικής και αποσαφηνίζουν τις χρησιμοποιούμενες έννοιες από μηδενική βάση θα χρησιμοποιήσουμε, κυρίως, στην ανάλυση της απόδειξης που ακολουθεί. Σε αυτή την απόδειξη (επιμελημένη από τον κ. Γιώργο Κολέτσο) η σειρά των αποτελέσματα δεν είναι η ίδια με αυτή του αρχικού άρθρου ενώ εμφανίζονται και επιπλέον αποτελέσματα, αλλά διατηρείται η ουσία της αποδεικτικής μεθόδου, η οποία είναι τελικά και αυτή που μας ενδιαφέρει. Φυσικά, θα επισημανθούν όσα σημεία του αρχικού άρθρου είναι διαφωτιστικά για τη μέθοδο ενώ θα χρησιμοποιηθεί επίσης, σαν ένα τρίτο σημείο αναφοράς, η απόδειξη που βρίσκεται στις *‘Σημειώσεις Μαθηματικής Λογικής’* του K.I. Δημητρακοπούλου, για την όσο το δυνατόν καλύτερη

παρουσίαση της σειράς των αποτελεσμάτων. Να σημειωθεί, τέλος, ότι, μιας και το ενδιαφέρον μας επικεντρώνεται στο βασικό κατασκευαστικό επιχείρημα που καταδεικνύει τη μη-πληρότητα, οι αποδείξεις των ενδιάμεσων αποτελεσμάτων τις περισσότερες φορές θα παραλείπονται, αν και όλα τα εργαλεία που χρησιμοποιούνται θα αναλύονται με πάσα λεπτομέρεια.

Σημ.: Η ολοκληρωμένη μέθοδος με τις αποδείξεις όλων των θεωρημάτων και λημμάτων, βρίσκεται στο Παράρτημα, το οποίο επίσης περιλαμβάνει ενδιαφέροντα πράγματα εκτός απόδειξης, όπως εφαρμογές του θεωρήματος του Gödel, ανάλογα αποτελέσματα όπως το θεώρημα απροσδιοριστίας του Tarski, σχόλια πάνω στη σχέση θεωρίας αριθμών και θεωρίας συνόλων κ.ά.

Μετά από την αναλυτική παρουσίαση της απόδειξης, στο τέλος αυτού του μέρους θα σκιαγραφηθεί ένα μέρος των πολλών μαθηματικών και φιλοσοφικών συνεπειών που έχουν σχέση με το θεώρημα του Gödel και θα δοθούν κάποια συγγενικά του αποτελέσματα, αλλά και εκλαϊκευμένες αποδόσεις του, οι οποίες χωρίς να είναι απαραίτητα τυπικά ορθές, είναι αυτές που δίνουν αυτό το χαρακτηριστικό στίγμα του ενδιαφέροντος που εξάπτει το θεώρημα σε μαθηματικούς και μη.

2. Βασικά εργαλεία της απόδειξης

2.1 Στοιχεία μαθηματικής λογικής

Είναι εύλογο να υποτεθεί ότι ο αναγνώστης γνωρίζει κάποια βασικά στοιχεία μαθηματικής λογικής αλλά θα ήταν χρήσιμο να γίνει μία συνολική παρουσίασή της, ώστε, όχι μόνο για να ξεκαθαριστούν θέματα συμβολισμού και ορολογίας, αλλά και να αναλυθεί πλήρως το υπόβαθρο που οδηγεί στα θεωρήματα μη-πληρότητας. Ένας ακόμη λόγος είναι ότι πολλά επιχειρήματα από το 3^ο μέρος της εργασίας χρησιμοποιούν στοιχεία μαθηματικής λογικής και θα πρέπει να είναι σαφές με ποιον τρόπο δομούνται, για να μπορέσουμε να διαπιστώσουμε (και) την μαθηματική εγκυρότητά τους. Άλλωστε, με το θεώρημα του Gödel δεν έχουν ασχοληθεί μόνο μαθηματικοί αλλά και φιλόσοφοι και πολλές φορές απαντώνται επιχειρήματα αμιγώς μαθηματικά, τα οποία προσπαθούν να δείξουν ότι σε κάποια φιλοσοφική ανάλυση δε χρησιμοποιούνται σωστά τα αποτελέσματα του Gödel, ή ότι εμπλέκονται τυπικά μαθηματικά σφάλματα στην πορεία μιας απόδειξης κτλ. Έτσι, αρχίζουμε από τα πιο βασικά κομμάτια της μαθηματικής λογικής για να περάσουμε σε όλο και πιο σύνθετα θέματα, παραλείποντας όσα εισάγουν πολλές τεχνικές λεπτομέρειες.

Υπενθυμίζεται ότι το τυπικό σύστημα λογικής που αναλύεται εδώ, εκτελεί ανάλογη λειτουργία με αυτό που περιείχε το PM, και το οποίο χρησιμοποίησε ο Gödel για να αποδείξει τα δύο θεωρήματα μη-πληρότητας.

2.1.1 Προτασιακή Λογική

Ας αρχίσουμε από την Προτασιακή Λογική και τη γλώσσα που την περιγράφει. Αυτή η γλώσσα μοιάζει με τη φυσική μας γλώσσα και δομείται έτσι ώστε να προσφέρει έναν έγκυρο τρόπο διατύπωσης γενικών επιχειρημάτων. Για τα μαθηματικά επιχειρήματα βασιζόμαστε σε μία πιο σύνθετη γλώσσα που εισάγει μαθηματικά σύμβολα και επιτρέπει την διατύπωση προτάσεων μαθηματικού περιεχομένου (βλ. § 2.1.3).

Το αλφάβητο της γλώσσας της προτασιακής λογικής δεν αποτελείται από λέξεις, αλλά από σύμβολα-προτάσεις που με τη χρήση λογικών συμβόλων οδηγούν σε νέες προτάσεις. Συγκεκριμένα, το αλφάβητο αυτό περιλαμβάνει:

α) *Προτασιακές μεταβλητές* (αριθμήσιμες το πλήθος και συμβολιζόμενες με p, q, r κτλ.) και

β) *Λογικά σύμβολα* που αποτελούνται από την αριστερή και δεξιά παρένθεση (συμβολίζονται με $($ και $)$ αντίστοιχα) και τους λογικούς συνδέσμους, οι οποίοι είναι: η άρνηση (συμβ. $\neg$), η σύζευξη (συμβ. $\wedge$), η διάζευξη (συμβ. $\vee$) και η συνεπαγωγή (συμβ. $\rightarrow$).

Με τη χρήση αυτών των συμβόλων, κατασκευάζουμε τους *προτασιακούς τύπους* οι οποίοι ορίζονται αναδρομικά και αναπαριστούν τις *καλά ορισμένες εκφράσεις* που μπορούν να προκύψουν από αυτή τη χρήση. Για παράδειγμα, ο τύπος $p \rightarrow q$ είναι μια καλά δομημένη έκφραση ενώ ο τύπος $\neg p(\rightarrow q$ δεν είναι, όπως αντίστοιχα στη φυσική γλώσσα η πρόταση «Αν 'κάτι', τότε 'κάτι άλλο'» έχει νόημα, ενώ η πρόταση «'Όχι 'κάτι', αν τότε 'κάτι άλλο'» δεν έχει.

Τώρα προχωρούμε για να απονείμουμε (ή αποδώσουμε) στις μεταβλητές μας (και κατ' επέκταση σε κάθε προτασιακό τύπο μέσω γνωστών κανόνων) μία *τιμή αλήθειας* (ή *αληθοτιμή*), η οποία τιμή έχει δύο μόνο επιλογές: ή θα είναι αληθής (True, με συμβολισμό T) είτε θα είναι ψευδής (False, με συμβολισμό F). Ορίζουμε επίσης μία απονομή αλήθειας, έστω V , να ικανοποιεί έναν προτασιακό τύπο ϕ , αν $V(\phi) = T$, δηλαδή αν μέσω αυτής της απονομής ο τύπος ϕ βρίσκεται να είναι αληθής. Στη συνέχεια ορίζουμε για ένα σύνολο προτασιακών τύπων, έστω Σ , ότι το Σ 'ταυτολογικά συνεπάγεται' έναν άλλο τύπο ϕ , (συμβ. $\Sigma \models \phi$) αν *κάθε απονομή αλήθειας που ικανοποιεί όλους τους τύπους του Σ , ικανοποιεί και τον ϕ* . Για παράδειγμα, αν το σύνολο Σ αποτελείται από μόνο έναν τύπο ψ και ισχύει ότι $\psi \models \phi$, τότε όποια αληθοτιμή απονεμηθεί στον τύπο ψ αναγκαστικά η ίδια θα απονεμηθεί και στον τύπο ϕ . Έτσι φτάνουμε στον ορισμό της έννοιας της **ταυτολογίας** σαν ένα τύπο ϕ για τον οποίο ισχύει ότι $\emptyset \models \phi$, δηλαδή έναν τύπο τον οποίο ικανοποιεί *κάθε* απονομή αλήθειας. Αντίθετα, **αντίφαση** καλούμε έναν τύπο ϕ αν και μόνον αν η άρνησή του, $\neg\phi$, είναι ταυτολογία.

Ένα ακόμη ενδιαφέρον στοιχείο που προκύπτει για την προτασιακή λογική, είναι το γεγονός ότι *υπάρχει αλγόριθμος μέσω του οποίου μπορούμε να αποφανθούμε εάν ένα πεπερασμένο σύνολο προτασιακών τύπων ταυτολογικά συνεπάγεται έναν προτασιακό τύπο ή όχι*. Η ύπαρξη ενός τέτοιου αλγορίθμου δεν είναι καθόλου τετριμμένη, όπως θα διαπιστώσουμε αργότερα στην πορεία δόμησης της λογικής.

2.1.2 Προτασιακός Λογισμός

Έχοντας αυτούς τους ορισμούς στη διάθεσή μας, προχωρούμε στον ορισμό ενός τυπικού συστήματος που θα αποδεικνύει με αυστηρό τρόπο όλες τις αληθείς προτάσεις του προτασιακού λογισμού και θα αποτελείται από κάποια αξιώματα και κανόνες απαγωγής (Σημ: επίσης χρησιμοποιείται ο όρος ‘κανόνες συναγωγής’). Χρησιμοποιούμε το παρακάτω αξιωματικό σύστημα, το οποίο συμβολίζουμε σαν $\mathcal{A}_0$.

Αξιωματικό σύστημα τύπου Hilbert για τον προτασιακό λογισμό:

Λογικά αξιώματα που ισχύουν για οποιουδήποτε προτασιακούς τύπους ϕ, ψ, χ :

A1. $\phi \rightarrow (\psi \rightarrow \phi)$

A2. $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))$

A3. $(\neg \psi \rightarrow \neg \phi) \rightarrow ((\neg \psi \rightarrow \phi) \rightarrow \psi)$

Κανόνας απαγωγής:

Modus Ponens (MP, από εδώ και στο εξής), σύμφωνα με τον οποίο,

Από τους τύπους $\phi \rightarrow \psi$ και ϕ απάγουμε τον τύπο ψ .

Σημ.: Παρατηρούμε ότι στον ορισμό των λογικών αξιωμάτων χρησιμοποιούμε αποκλειστικά το σύνολο συνδέσμων $\{\neg, \rightarrow\}$, κάτι απόλυτα θεμιτό αφού το σύνολο αυτό είναι *επαρκές* για την περιγραφή όλων των λογικών συνδέσμων.

Στα λογικά αξιώματα που αναφέραμε, μπορούμε να προσθέσουμε σαν δεδομένα για την παραγωγή αποδείξεων και επιπλέον παραδοχές - θεωρήματα, υπό τη μορφή προτασιακών τύπων, φτιάχνοντας έτσι μία θεωρία, T . Αν μέσω των θεωρημάτων της T , των λογικών αξιωμάτων και του MP αποδεικνύεται ένας τύπος ϕ , λέμε ότι η T αποδεικνύει τον ϕ και γράφουμε $T \vdash \phi$. Ορίζουμε τώρα τυπικά την έννοια της συνέπειας που αναφέραμε στη εισαγωγή του 1^{ου} μέρους.

Ορισμός: Μία θεωρία T είναι **συνεπής** αν και μόνον αν δεν μπορεί να αποδεικνύει αντιφάσεις, δηλαδή δεν υπάρχει τύπος ϕ τέτοιος ώστε η T να αποδεικνύει τον ϕ και ταυτόχρονα την άρνησή του, $\neg \phi$. Αλλιώς, αν υπάρχει ϕ τέτοιος ώστε $T \vdash \phi \wedge \neg \phi$, η T είναι **αντιφατική**.

Αν βέβαια ο σκοπός της δημιουργίας του προτασιακού λογισμού είναι η απόδειξη αληθειών, θα ήταν κρίσιμης σημασίας η γνώση του γεγονότος ότι οτιδήποτε μπορεί να αποδειχθεί στα πλαίσιά του είναι αληθές, αλλά και αντίστροφα, αν κάτι ισχύει (είναι αληθές) να ξέρουμε ότι υπάρχει μία τυπική απόδειξη γι' αυτό. Αυτή τη γνώση μας την προσφέρουν δύο πολύ βασικά αποτελέσματα, το οποία θα μας απασχολήσουν και αργότερα. Συγκεκριμένα, αυτά είναι:

Το Θεώρημα της Ορθότητας του προτασιακού λογισμού, το οποίο λέει ότι:

Αν $T \vdash \phi$, τότε $T \models \phi$, δηλαδή ότι αν οποιοσδήποτε τύπος ϕ είναι αποδείξιμος σε μια συνεπή θεωρία T , τότε ο ϕ είναι αληθής σε αυτή τη θεωρία, και το

Θεώρημα της Πληρότητας του προτασιακού λογισμού, το οποίο λέει ότι:

Αν $T \models \phi$, τότε $T \vdash \phi$, δηλαδή ότι αν μια συνεπής θεωρία T ταυτολογικά συνεπάγεται ένα τύπο ϕ , τότε υπάρχει απόδειξη για τον ϕ στη θεωρία αυτή.

Τα παραπάνω θεωρήματα ισχύουν για οποιαδήποτε συνεπή θεωρία (οι αντιφατικές θεωρίες δε μας ενδιαφέρουν, αφού μέσω μιας αντίφασης μπορούμε να αποδείξουμε όλες τις προτάσεις, είτε αληθείς, είτε ψευδείς). Αν στη θέση της T μπει η κενή θεωρία ($\emptyset$), έχουμε τα αντίστοιχα αποτελέσματα για τον καθαρό προτασιακό λογισμό, δηλαδή το σύστημα $\mathcal{A}_0$ χωρίς άλλα δεδομένα, οπότε προκύπτει ότι τα θεωρήματα του $\mathcal{A}_0$ είναι οι ταυτολογίες (από ορθότητα), και μόνον αυτές (από πληρότητα).

Σε σχέση όσα ακολουθήσουν, το θεώρημα της πληρότητας που αναφέρεται εδώ δεν έχει σχέση με το θεώρημα μη-πληρότητας του Gödel. Αυτό συμβαίνει διότι είναι διαφορετικές οι δύο έννοιες της πληρότητας που αναφέρονται στους τίτλους των θεωρημάτων. Το θεώρημα πληρότητας που αναφέρεται εδώ έχει να κάνει μόνο με την ιδιότητα που περιγράφει (γι' αυτό δεν αναφέρεται πουθενά μέχρι τώρα η έννοια της πληρότητας από μόνη της), ενώ το θεώρημα μη-πληρότητας αναφέρεται στην μη-πληρότητα ενός αξιωματικού συστήματος το οποίο περιγράφει μία μαθηματική θεωρία. Το τι σημαίνει όμως πλήρες αξιωματικό σύστημα είναι κάτι δεν έχει οριστεί ακόμα και σίγουρα δεν σχετίζεται με το $\mathcal{A}_0$ διότι όπως επισημάνθηκε, η προτασιακή λογική δεν μπορεί να περιγράψει μαθηματικές θεωρίες. Αυτή η δυνατότητα θα δοθεί με τον ορισμό της κατηγορηματικής λογικής που ακολουθεί, οπότε στη συνέχεια θα οριστεί και η έννοια της πληρότητας ενός αξιωματικού συστήματος.

2.1.3 Κατηγορηματική Λογική

Αρχίζουμε όπως και στην Προτασιακή Λογική με την παρουσίαση της γλώσσας της Κατηγορηματικής Λογικής (ή Λογικής των Κατηγορημάτων), η οποία θα μας δώσει τη δυνατότητα έκφρασης μαθηματικών προτάσεων. Για να υπάρξει αυτή η δυνατότητα, πρέπει να εμπλουτιστεί το αλφάβητο της προτασιακής λογικής με σύμβολα που θα επιδέχονται μαθηματικής ερμηνείας. Αυτά τα νέα σύμβολα καλούνται *μη-λογικά σύμβολα* και είναι τα παρακάτω:

α) *Σύμβολα σταθερών*, τα οποία συμβολίζουν σταθερά αντικείμενα και μπορεί να είναι αριθμήσιμα ή πεπερασμένα το πλήθος ή να απουσιάζουν και τελείως.

β) *Σύμβολα κατηγορημάτων*, τα οποία συμβολίζουν τις ιδιότητες των μαθηματικών αντικειμένων της θεωρίας με την οποία ασχολούμαστε και είναι κάποιου βαθμού n , ανάλογα με τις n -μελείς σχέσεις που συμβολίζουν. Για παράδειγμα, αν τα x και y είναι αριθμοί, ένα δυαδικό κατηγορημα $Q(x, y)$ μπορεί να συμβολίζει μία σχέση ισότητας μεταξύ τους.

γ) *Σύμβολα συναρτήσεων*, τα οποία κατέχουν τον ρόλο των συναρτήσεων και μπορεί να είναι n θέσεων με ανάλογο τρόπο σαν τα κατηγορήματα. Για παράδειγμα, η πράξη της πρόσθεσης δύο φυσικών είναι μία δυαδική συνάρτηση που έχει σαν πεδίο ορισμού το $\mathbb{N} \times \mathbb{N}$ και πεδίο τιμών το $\mathbb{N}$. Αυτά, όπως και τα σύμβολα των κατηγορημάτων μπορούν να απουσιάζουν από μία θεωρία ή να είναι πεπερασμένα ή αριθμήσιμα το πλήθος.

Τα σύμβολα συναρτήσεων τα συμβολίζουμε με F, G, H κ.ο.κ. ενώ αυτά των κατηγορημάτων με P, Q, R κ.ο.κ.

Διατηρούμε επίσης τα λογικά σύμβολα της προτασιακής λογικής, καθότι τελικός σκοπός μας είναι η διατύπωση μαθηματικών επιχειρημάτων και αποδείξεων και η χρήση τους είναι απαραίτητη για να μεταχειριζόμαστε τα μαθηματικά αντικείμενα. Φυσικά, γίνονται οι απαραίτητες προσθήκες και τροποποιήσεις ώστε αυτά να ανταποκρίνονται στις νέες απαιτήσεις μας και έτσι ολοκληρώνεται το αλφάβητο με τα παρακάτω σύμβολα:

α) *Μεταβλητές*, ένα αριθμήσιμο σύνολο συμβόλων $x_1, x_2, \dots, x_n, \dots, y$ κτλ.,

β) *Λογικούς συνδέσμους*, τα γνωστά $\neg, \wedge, \vee, \rightarrow$, καθώς και τη διπλή συνεπαγωγή, $\leftrightarrow$,

γ) *Κόμμα και παρενθέσεις*, με το γνωστό συμβολισμό και

δ) *Ποσοδείκτες*, τον καθολικό (συμβ. $\forall$) και τον υπαρξιακό (συμβ. $\exists$) ποσοδείκτη.

Ο τρόπος με τον οποίο δομήθηκε το αλφάβητο της κατηγορηματικής λογικής μας επιτρέπει, με την κατάλληλη επιλογή μη-λογικών συμβόλων (αφού τα λογικά σύμβολα είναι τα ίδια σε κάθε γλώσσα), να παράγουμε πολλές διαφορετικές γλώσσες. Ανάλογα με αυτή την επιλογή παίρνουμε γλώσσες οι οποίες περιγράφουν διαφορετικές μαθηματικές θεωρίες. Για παράδειγμα, για μία γλώσσα για την αριθμητική μπορούμε να επιλέξουμε για μη-λογικά σύμβολα, τα σύμβολα σταθερών 0 και 1, τα σύμβολα κατηγορημάτων δευτέρου βαθμού = και $\leq$ και τα σύμβολα συναρτήσεων δύο θέσεων + και $\cdot$. Όλα αυτά βέβαια θα συμβολίζουν τις γνωστές πράξεις, σχέσεις και σταθερές που συναντούμε στην απλή αριθμητική, αλλά η ερμηνεία είναι ένα ζήτημα που θα αναλυθεί παρακάτω.

Εντελώς ανάλογα με την προτασιακή λογική, ορίζουμε τώρα τις καλοσχηματισμένες εκφράσεις που προκύπτουν από τους συνδυασμούς των παραπάνω συμβόλων, τις εκφράσεις δηλαδή που έχουν κάποιο νόημα. Επαγωγικά ορίζουμε τους *όρους* οι οποίοι είναι οι μεταβλητές, τα σύμβολα σταθερών καθώς και οι εκφράσεις του τύπου $F(a_1, \dots, a_n)$, όπου τα $a_1, \dots, a_n$ είναι όροι και το F σύμβολο συνάρτησης. Επίσης ορίζουμε τους *ατομικούς τύπους* ως κάθε έκφραση της μορφής $R(a_1, \dots, a_n)$, όπου τα $a_1, \dots, a_n$ είναι όροι και R σύμβολο κατηγορήματος n-οστού βαθμού. Τέλος, όπως και στην προτασιακή λογική, με τη χρήση των λογικών συνδέσμων και των ποσοδεικτών ορίζουμε, επίσης επαγωγικά, τους *τύπους*, ξεκινώντας την επαγωγή από τους ατομικούς τύπους. Έτσι δομούμε εκφράσεις που έχουν κάποιο νόημα, όπως για παράδειγμα το γεγονός ότι «για κάθε μεταβλητή x , αν 'κάτι' ισχύει για τη x , τότε ισχύει 'κάτι άλλο' για τη x », με τον τύπο $\forall x(Q(x) \rightarrow P(x))$.

Ένα βασικό θέμα που προκύπτει από αυτούς τους ορισμούς είναι το κατά πόσο θα πρέπει να ορίσουμε συγκεκριμένη τιμή για κάθε μεταβλητή που εμφανίζεται σε κάθε τύπο ή μας αρκεί να είναι αόριστη, όπως στο προηγούμενο παράδειγμα. Χωρίς περισσότερες λεπτομέρειες, αρκεί να πούμε ότι στην περίπτωση του παραδείγματος, η μεταβλητή καλείται *δεσμευμένη*, διότι βρίσκεται στην εμβέλεια του καθολικού (ή και του υπαρξιακού, σε άλλες περιπτώσεις) ποσοδείκτη. Σε περίπτωση που μια μεταβλητή δεν δεσμεύεται από κάποιον ποσοδείκτη και χρειάζεται να της αποδώσουμε συγκεκριμένη τιμή για να έχει νόημα η πρότασή μας, καλείται *ελεύθερη*. Ακολουθώντας, ορίζουμε μία μεταβλητή x να είναι *αντικαταστάσιμη* από ένα όρο t σε ένα τύπο $\phi(x)$ αν με αυτή την ενέργεια, στον νέο τύπο $\phi(t)$ που προκύπτει, όλες οι

μεταβλητές του t παραμένουν ελεύθερες, δηλαδή πρακτικά, αν με την αντικατάσταση αφήνουμε την εμβέλεια των ποσοδεικτών ανέπαφο.

Τελειώνοντας με τους ορισμούς των αντικειμένων και καλά ορισμένων εκφράσεων, ορίζουμε σαν *πρόταση*, έναν τύπο που δεν έχει ελεύθερες μεταβλητές. Οι προτάσεις αυτές είναι που θα ερμηνεύονται τελικά σαν αληθείς ή ψευδείς στα πλαίσια της ερμηνείας μίας μαθηματικής θεωρίας, όπως αντίστοιχα οι τύποι της προτασιακής λογικής στα πλαίσια μίας απονομής αλήθειας.

Μέχρι τώρα στην περιγραφή της γλώσσας της κατηγορηματικής λογικής, δεν έχουμε δώσει μαθηματικό νόημα στα αντικείμενα που την απαρτίζουν. Αυτό γίνεται με την εισαγωγή των Δομών (ή Ερμηνειών), η οποία θα μας οδηγήσει τελικά στη διαμόρφωση μαθηματικών θεωριών.

Μία Δομή (ή Ερμηνεία) είναι ένα σύστημα $\mathcal{A}$ που αποτελείται από:

1. Ένα μη κενό σύνολο A που καλείται *πεδίο της δομής* και είναι το σύμπαν των αντικειμένων στα οποία αναφερόμαστε μέσω της γλώσσας που ορίσαμε (π.χ. στο σύστημα της αριθμητικής τα αντικείμενα είναι οι αριθμοί και το σύμπαν τους το $\mathbb{N}$).
2. Μία αντιστοίχιση ενός στοιχείου του A σε κάθε σύμβολο σταθεράς της γλώσσας (π.χ. του αριθμού 0 σε ένα σύμβολο σταθεράς '0'),
3. Μία αντιστοίχιση συνάρτησης n μεταβλητών από το A^n στο A σε κάθε σύμβολο συνάρτησης (π.χ. της πρόσθεσης στο σύμβολο '+') και ανάλογα,
4. Μία αντιστοίχιση μίας n -μελούς σχέσης $R \subseteq A^n$ σε κάθε σύμβολο κατηγορήματος της γλώσσας (π.χ. για τη γλώσσα της θεωρίας συνόλων, της σχέσης του 'ανήκειν', στο σύμβολο ' $\in$ ').

Η παραπάνω διαδικασία συνίσταται στην απόδοση μίας πραγματικότητας στα μη-λογικά σύμβολα της γλώσσας. Ανάλογα βέβαια με τη θεωρία που θέλουμε να αναπτύξουμε, διαλέγουμε το συγκεκριμένο αριθμό και τη συγκεκριμένη ερμηνεία των συμβόλων αυτών, ώστε η επιλογή μας να αντανakλά τις απαιτήσεις της εκάστοτε θεωρίας. Ουσιαστικά λοιπόν, αυτό που κάναμε είναι να δώσουμε νόημα στα 'α-νόητα' μέχρι τώρα μη-λογικά στοιχεία της προτασιακής λογικής για να δομήσουμε συγκεκριμένες μαθηματικές θεωρίες. Αν υποθέσουμε μία αναλογία με τη ζωγραφική, θα μπορούσαμε να πούμε ότι διαλέγουμε από την παλέτα των άπειρων συνδυασμών χρωμάτων αυτά ακριβώς που θέλουμε για να ζωγραφίσουμε κάτι συγκεκριμένο, όπως για παράδειγμα το μαύρο και το άσπρο για να ζωγραφίσουμε μία σκακιέρα.

Αυτό που μένει τώρα που έχει νόημα μια γλώσσα που περιγράφει μία θεωρία, είναι να αποδώσουμε την αλήθεια ή το ψεύδος των προτάσεων της στην εκάστοτε ερμηνεία της. Η απονομή αλήθειας ήταν σχετικά απλή στα πλαίσια της προτασιακής λογικής, αλλά τώρα που στους τύπους εμπλέκονται και ελεύθερες μεταβλητές, τα πράγματα είναι πιο σύνθετα. Ο Tarski έδωσε τον ορισμό της αλήθειας των προτάσεων επαγωγικά, απονέμοντας αρχικά τιμές στις μεταβλητές μέσω μίας συνάρτησης $s:V \rightarrow A$, όπου V το σύνολο των μεταβλητών της γλώσσας και A το πεδίο της εκάστοτε δομής. Στη συνέχεια περνάμε στην απονομή τιμής στις σταθερές της δομής, καθώς και στους όρους της, μέσω μιας αυστηρώς δομημένης διαδικασίας που δίνει προφανή αποτελέσματα. Για παράδειγμα, στην αριθμητική από τη στιγμή που δώσουμε τιμή για δύο μεταβλητές, έστω $x_1=3$ και $x_2=2$, τότε η έκφραση $x_1 \cdot x_2 + 1$ θα πάρει αυτόματα την τιμή 7. Το τι ακριβώς σημαίνει για ένα τύπο ϕ να αληθεύει σε μία δομή $\mathcal{A}$ για μία δοθείσα αποτίμηση s (συμβ. $\mathcal{A} \models \phi[s]$), βγαίνει επίσης από τους ορισμούς που έδωσε ο Tarski (σημ. μία άλλη διατύπωση του ίδιου πράγματος είναι: η $\mathcal{A}$ ικανοποιεί τον τύπο ϕ με την αποτίμηση s). Χωρίς να μπορούμε σε τεχνικές λεπτομέρειες, αναφέρουμε απλώς ότι οι ορισμοί αυτοί επιβεβαιώνουν τη διαισθητική προσέγγιση της μαθηματικής αληθείας. Για παράδειγμα, ένας από αυτούς τους ορισμούς λέει ότι ‘η δομή $\mathcal{A}$ ικανοποιεί τον τύπο $\psi \wedge \chi[s]$ αν και μόνον αν ικανοποιεί τον $\psi[s]$ και τον τύπο $\chi[s]$, κάτι που το περιμέναμε από τον ορισμού του λογικού συμβόλου ‘ $\wedge$ ’. Γενικότερα, λέμε ότι:

- Ένας τύπος ϕ είναι αληθής σε μία ερμηνεία $\mathcal{A}$, όταν για κάθε απονομή αλήθειας s , έχουμε ότι $\mathcal{A} \models \phi[s]$ και αντίστοιχα,
- Ένας τύπος ϕ είναι ψευδής σε μία ερμηνεία $\mathcal{A}$, όταν για κάθε απονομή αλήθειας s έχουμε ότι $\mathcal{A} \not\models \phi[s]$ (με ‘ $\not\models$ ’ συμβολίζουμε ότι κάτι δεν συνεπάγεται ταυτολογικά κάτι άλλο, δηλ. ότι το δεξί μέλος δε βγαίνει αληθές σύμφωνα με τις υποθέσεις του αριστερού).

Οι ορισμοί αυτοί κινούνται στα πλαίσια μιας συγκεκριμένης ερμηνείας και άρα αναφέρονται σε συγκεκριμένες θεωρίες. Υπάρχουν όμως και προτάσεις που ισχύουν ανεξαρτήτως συγκεκριμένων ερμηνειών, που αποτελούν δηλαδή γενικές αλήθειες της κατηγορηματικής λογικής, όπως ακριβώς οι ταυτολογίες της προτασιακής λογικής.

Ορίζουμε λοιπόν για έναν τύπο ϕ να είναι *έγκυρος ή λογικά αληθής* αν και μόνον αν ο ϕ αληθεύει για κάθε αποτίμηση s στην οποιαδήποτε τυχαία δομή $\mathcal{A}$. Επίσης, ανάλογα με την προτασιακή λογική, για ένα σύνολο μη-λογικών προτάσεων Σ , μπορούμε να πούμε ότι το Σ ‘*συνεπάγεται λογικά*’ έναν τύπο ϕ (συμβ. $\Sigma \models \phi$) αν κάθε ερμηνεία που ικανοποιεί το σύνολο τύπων Σ για κάθε αποτίμηση s , ικανοποιεί και τον τύπο ϕ . Τέλος, ορίζουμε μία δομή $\mathcal{A}$ να είναι *μοντέλο* ενός συνόλου προτάσεων, Σ , μίας γλωσσάς, αν για κάθε $t \in \Sigma$, έχουμε ότι: $\mathcal{A} \models t$.

2.1.4 Κατηγορηματικός Λογισμός

Για να ολοκληρωθεί αυτή η προσπάθεια παρουσίασης της μαθηματικής λογικής, μένει μόνο να ορίσουμε ένα σύστημα αξιωμάτων ανάλογο με αυτό του προτασιακού λογισμού, σύστημα το οποίο θα αποδεικνύει με αυστηρό τρόπο όλες τις αλήθειες του κατηγορηματικού λογισμού. Αυτή η διαδικασία είναι απόλυτα απαραίτητη διότι, σε αντίθεση με τον προτασιακό λογισμό, εδώ *δεν υπάρχει αλγόριθμος με τη βοήθεια του οποίου να αποφαινόμαστε για την λογική αλήθεια μίας πρότασης* (με άλλα λόγια ο κατηγορηματικός λογισμός είναι μη-αποκρίσιμος όσον αφορά τις μαθηματικές αλήθειες). Ένας από τους λόγους είναι ότι στον ορισμό του έγκυρου τύπου που δόθηκε πιο πάνω, αναφέραμε ότι πρέπει να ελεγχθεί η αλήθεια του για *κάθε* δομή $\mathcal{A}$ και ο αριθμός των δυνατών δομών είναι άπειρος! Άρα μας είναι απαραίτητος ο ορισμός ενός τυπικού αξιωματικού συστήματος με το οποίο θα αποδεικνύουμε όλες τις λογικές και μαθηματικές αλήθειες. Από τη στιγμή που θα έχουμε μία τέτοια απόδειξη για έναν τύπο, θα μπορούμε να αποφανθούμε αν αυτός είναι έγκυρος ή όχι, εφόσον φυσικά ισχύει το θεώρημα της ορθότητας για τον κατηγορηματικό λογισμό.

Το σύνολο συνδέσμων που απαιτούνται για τη δόμηση του αξιωματικού συστήματος διαλέγεται να είναι και πάλι το ελάχιστο δυνατό επαρκές. Χρησιμοποιούνται λοιπόν μόνο οι σύνδεσμοι $\{\neg, \rightarrow, \forall\}$, μέσω των οποίων όλοι οι υπόλοιποι μπορούν να προκύψουν.

Συμβολίζουμε το σύστημα αυτό με $\mathcal{A}_1$, εύλογα, διότι αποτελεί μία επέκταση του αντίστοιχου του προτασιακού λογισμού με δύο παραπάνω αξιώματα και έναν κανόνα απαγωγής.

Αξιοματικό σύστημα τύπου Hilbert για τον κατηγορηματικό λογισμό:

Λογικά αξιώματα που ισχύουν για οποιουσδήποτε τύπους ϕ, ψ, χ :

A1. $\phi \rightarrow (\psi \rightarrow \phi)$

A2. $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))$

A3. $(\neg \psi \rightarrow \neg \phi) \rightarrow ((\neg \psi \rightarrow \phi) \rightarrow \psi)$

A4. $\forall x \phi(x) \rightarrow \phi(t)$, με την προϋπόθεση ότι η x είναι αντικαταστάσιμη από τον όρο t στον τύπο $\phi(x)$.

A5. $\forall x(\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \forall x\psi)$, με την προϋπόθεση ότι ο τύπος ϕ δεν περιέχει καμία ελεύθερη εμφάνιση της μεταβλητής x .

Κανόνες απαγωγής:

1. Ο γνωστός **Modus Ponens** και ο

2. **Κανόνας της Γενίκευσης** (Gen, από εδώ και στο εξής), σύμφωνα με τον οποίο:

Από τον τύπο ϕ απάγουμε τον $\forall x\phi$.

Σημ.: Η έννοια του τελευταίου κανόνα είναι η γνωστή μέθοδος απόδειξης στα μαθηματικά που λέει ότι 'αν θέλω να αποδείξω έναν ισχυρισμό $\phi(x)$ για κάθε x , αρκεί να αποδείξω τον ισχυρισμό για ένα τυχαίο x '.

Μία θεωρία T δομείται με όμοιο τρόπο με τον προτασιακό λογισμό, δηλαδή με την εισαγωγή ενός συνόλου προτάσεων της γλώσσας του κατηγορηματικού λογισμού σαν επιπλέον υποθέσεις που καλούνται μη-λογικά αξιώματα. Δεδομένων μη-λογικών και λογικών αξιωμάτων και με τη χρήση των κανόνων απαγωγής ορίζεται η έννοια της απόδειξης σαν μία πεπερασμένη ακολουθία τύπων $\phi_1, \phi_2, \dots, \phi_n$, όπου οι τύποι εκτός του ϕ_n αντιπροσωπεύουν τα βήματα που έγιναν με βάση των δεδομένων και των κανόνων απαγωγής. Εφόσον τα βήματα αυτά είναι έγκυρα, από μία απόδειξη προκύπτει ο τύπος ϕ_n σαν *θεώρημα*. Συμβολικά, αν ϕ είναι ένα θεώρημα μιας θεωρίας T , γράφουμε $T \vdash \mathcal{A}_1 \phi$, που σημαίνει ότι *υπάρχει μία τυπική απόδειξη στο σύστημα $\mathcal{A}_1$ μαζί με το σύνολο υποθέσεων T , για τον τύπο ϕ* .

Το ερώτημα τώρα που έχουν οριστεί όλες οι απαραίτητες έννοιες, είναι αν ισχύουν τα θεωρήματα ορθότητας και πληρότητας που ίσχυαν αντίστοιχα στον προτασιακό λογισμό. Η απάντηση είναι καταφατική, και συγκεκριμένα έχουμε:

Το Θεώρημα της Ορθότητας του κατηγορηματικού λογισμού, το οποίο λέει ότι:

Αν $T \vdash \phi$, τότε $T \models \phi$, δηλαδή ότι αν οποιοσδήποτε τύπος ϕ είναι τυπικά αποδείξιμος σε μια συνεπή θεωρία T , τότε ο ϕ είναι αληθής στα πλαίσιά της, και το Θεώρημα της Πληρότητας του κατηγορηματικού λογισμού (Gödel, 1930), το οποίο λέει ότι:

Αν $T \models \phi$, τότε $T \vdash \phi$, δηλαδή ότι αν μια συνεπής θεωρία T λογικά συνεπάγεται ένα τύπο ϕ , τότε υπάρχει τυπική απόδειξη για τον ϕ στη θεωρία αυτή.

Με αυτά τα δύο αποτελέσματα εξασφαλίζουμε ότι αυτά που μπορούμε να αποδείξουμε σε μία θεωρία είναι ακριβώς αυτά που αληθεύουν σε αυτή. Να επισημάνουμε επίσης ότι το παραπάνω θεώρημα πληρότητας, ένα από τα πιο σημαντικά αποτελέσματα της μαθηματικής λογικής, και πάλι δεν έχει σχέση με αυτό της μη-πληρότητας, αν και τα δύο τα έχει αποδείξει ο Gödel, για τους ίδιους λόγους που έχουν αναφερθεί στο τέλος της παραγράφου 2.1.2.

Έχουμε φτάσει πλέον κοντά στην ολοκλήρωση της παρουσίασης της λογικής, αλλά πριν ανακεφαλαιώσουμε και πάμε παρακάτω, για να δούμε τελικά που αναφέρεται η μη-πληρότητα που μας απασχολεί, ας δούμε ένα ισοδύναμο αποτέλεσμα με τα θεωρήματα ορθότητας και πληρότητας, στο οποίο μάλιστα βασίζεται η απόδειξη του θεωρήματος της πληρότητας:

Γενικευμένο θεώρημα πληρότητας: Αν Σ είναι ένα σύνολο προτάσεων, τότε το Σ είναι ικανοποιήσιμο αν και μόνον αν είναι συνεπές.

Σημ.: Ένα σύνολο προτάσεων Σ είναι ικανοποιήσιμο αν υπάρχει έστω και ένα μοντέλο του. Για τη συνέπεια ισχύει ο γνωστός ορισμός, αναφερόμενος σε ένα σύνολο προτάσεων αντί μία θεωρία, πράγματα ισοδύναμα, αφού μια θεωρία λέγεται συνεπής αν το σύνολο των μη-λογικών της αξιωμάτων είναι συνεπές.

2.1.5 Ανακεφαλαίωση και μία σημαντική εφαρμογή

Ανακεφαλαιώνοντας, καλό θα ήταν αρχικά να αναφερθούν και άλλες διαδεδομένες ορολογίες για όλα αυτά που συναντήσαμε μέχρι τώρα και να δοθεί η ερμηνεία τους, διότι ενδέχεται να απαντηθούν αργότερα ή σε άλλα συγγράμματα λογικής.

Πρώτα απ' όλα, οι όροι 'Πρωτοβάθμια Λογική' ή 'Πρωτοβάθμιος Κατηγορηματικός λογισμός' περιγράφουν το ίδιο πράγμα, δηλαδή τον κατηγορηματικό λογισμό που περιγράφηκε πιο πάνω, με αυτά τα αξιώματα και κανόνες συνεπαγωγής (αν και μερικοί συγγραφείς προτιμούν τον κανόνα της γενίκευσης να τον παρουσιάζουν σαν θεώρημα). Η λέξη 'πρωτοβάθμιος/α' χρησιμοποιείται για να χαρακτηρίσει τη γλώσσα που χρησιμοποιούμε, και κατ' επέκταση φυσικά το αξιωματικό μας σύστημα, και σημαίνει ότι οι μεταβλητές μας παίρνουν τιμές από τα στοιχεία ενός συνόλου. Αυτή η επιλογή δεν είναι η μόνη δυνατή αφού υπάρχουν δευτεροβάθμιες γλώσσες που οι τιμές των μεταβλητών τους είναι *υποσύνολα* κάποιου συνόλου κ.ο.κ. Η 'Δευτεροβάθμια Λογική' που δομείται με βάση τέτοιες γλώσσες χρησιμοποιείται μεν, αλλά χωρίς ευρεία αποδοχή, αφού υπάρχει η πεποίθηση ότι δεν είναι 'σωστή λογική' ή ότι είναι κατά βάση η θεωρία συνόλων με άλλη ορολογία. Ίσως μια πιο αποδεκτή επέκταση της πρωτοβάθμιας λογικής να έδινε ενδιαφέροντα αποτελέσματα πληρότητας ή μη-πληρότητας αλλά ελλείψει της, συνεχίζουμε να ασχολούμαστε κυρίως με την πρωτοβάθμια.

Όπως έχει ήδη αναφερθεί, σε αυτή τη λογική το βασικό που μπορούμε να κάνουμε είναι να φτιάξουμε και να επεξεργαστούμε μαθηματικές θεωρίες, όπως η Θεωρία Αριθμών, η Θεωρία Συνόλων, η Θεωρία Ομάδων κτλ. Οι θεωρίες αυτές στηρίζονται φυσικά σε κάποια αξιώματα, οπότε καλούνται επίσης και 'αξιωματικά συστήματα'. Η επιλογή της αξιωματικής θεμελίωσης επίσης δεν είναι η μόνη δυνατή επιλογή, αλλά φαίνεται να είναι η πιο έγκυρη επιστημονικά και είναι αυτή που χρησιμοποιείται σε σχεδόν καθολικό βαθμό.

Ένα από αυτά τα αξιωματικά συστήματα, για τη θεωρία της αριθμητικής, έχουμε ήδη συναντήσει στην ιστορική αναδρομή του 1^{ου} μέρους, την αριθμητική του Peano (PA). Αυτό που με λεπτομέρεια θα γίνει τώρα, είναι η βήμα προς βήμα δόμηση της γενικής πρωτοβάθμιας αριθμητικής, η οποία βέβαια περιλαμβάνει το σύστημα της PA. Η πορεία της δόμησης θα είναι αυτή που ακολουθήθηκε και στη γενική περίπτωση της πρωτοβάθμιας λογικής.

Πρωτοβάθμια Αριθμητική

Ξεκινάμε εύλογα από την επιλογή του αλφαβήτου της γλώσσας η οποία θα περιγράψει την πρωτοβάθμια αριθμητική. Η γλώσσα της αριθμητικής συμβολίζεται με $\mathcal{L}_A$ και περιλαμβάνει, πέρα από τα λογικά, τα παρακάτω μη-λογικά σύμβολα:

1. Ένα σύμβολο σταθεράς, το $\{\bar{0}\}$.
2. Δύο σύμβολα κατηγορημάτων, τα $\{<\}$ και $\{=\}$, δευτέρου βαθμού και τα δύο.
3. Τρία σύμβολα συναρτήσεων, τα $\{+\}$ και $\{\cdot\}$, δύο θέσεων και το $\{S\}$, μίας θέσης.

Σημειώστε ότι αυτή η επιλογή δεν είναι η μόνη δυνατή, άλλωστε δεν είναι ίδια με αυτή που είχε αναφερθεί σαν παράδειγμα για την αριθμητική στην αρχή της παραγράφου 2.1.3. Η συγκεκριμένη επιλογή έγινε με βάση την απόδειξη του θεωρήματος μη-πληρότητας που βρίσκεται στο Παράρτημα, και ειδικότερα το κομμάτι που αναφέρεται στην πρωτοβάθμια αριθμητική.

Η γλώσσα $\mathcal{L}_A$ περιλαμβάνει όπως είδαμε και το σύμβολο κατηγορήματος $\{=\}$. Αυτό το σύμβολο που προφανώς θα πάρει την ερμηνεία της ισότητας, μερικοί συγγραφείς το θεωρούν λογικό σύμβολο, οπότε τα αξιώματα που περιγράφουν τις βασικές ιδιότητες της ισότητας, τα περιλαμβάνουν ανάμεσα στα λογικά αξιώματα A1- A5. Εδώ δεν ακολουθείται αυτή η σύμβαση, θεωρούμε δηλαδή το $\{=\}$ σαν μη-λογικό σύμβολο, οπότε δίνουμε τώρα τα αξιώματα που θα πρέπει να ενσωματώνει μία ‘θεωρία με ισότητα’, όπως είναι η πρωτοβάθμια αριθμητική.

Αξιώματα της ισότητας:

ΙΣ1. $\forall x(x = x)$

ΙΣ2. $x = y \rightarrow (\phi(x, x) \rightarrow \phi(x, y))$, όπου x, y οποιεσδήποτε μεταβλητές, $\phi(x, x)$ τυχαίος τύπος και $\phi(x, y)$ ο τύπος που προκύπτει από την αντικατάσταση (μερικών ή όλων των ελεύθερων εμφανίσεων) της x από την y , εφόσον οι ελεύθερες εμφανίσεις της x που αντικαθίστανται είναι αντικαταστάσιμες από την y .

Είναι επίσης προφανές ότι τα δύο αυτά αξιώματα ισχύουν σε κάθε θεωρία που βασίζεται σε μία γλώσσα που περιλαμβάνει το σύμβολο της ισότητας, αν και τα ίδια δεν συμπεριλαμβάνονται ούτε στα λογικά, ούτε στα μη-λογικά αξιώματα της εκάστοτε θεωρίας.

Τώρα που έχουμε το αλφάβητο της γλώσσας που διαλέξαμε για να περιγράψει τη θεωρία μας, πρέπει να αποδώσουμε νόημα στα μη-λογικά σύμβολά της, νόημα που φυσικά θα συνάδει με τις αρχές της αριθμητικής. Όπως είδαμε, αυτό γίνεται με τον ορισμό μίας δομής (ή ερμηνείας). Όπως και στην περίπτωση επιλογής της γλώσσας $\mathcal{L}_A$, από τις διάφορες δομές που υπάρχουν για την περιγραφή της αριθμητικής, διαλέγουμε το ‘στάνταρ μοντέλο’ που υπάρχει στην απόδειξη του παραρτήματος (βλ. σελ. xiii).

Αυτή η **Κανονικοποιημένη Δομή της αριθμητικής** συμβολίζεται με $\mathbb{N}$ και αποτελείται από:

1. Το μη-κένό σύνολο $\{\omega\}$, σύμπαν της δομής, με αντικείμενα τα αριθμητικά.
2. Το μηδενικό στοιχείο $\{0\}$, σαν ερμηνεία του συμβόλου $\bar{0}$.
3. Τη διμελή σχέση διάταξης $\{<\}$, σαν ερμηνεία του συμβόλου κατηγορήματος δευτέρου βαθμού $<$.
4. Τη συνάρτηση μίας μεταβλητής $\{s\}$ που δηλώνει τον επόμενο, σαν ερμηνεία του συναρτησιακού συμβόλου S .
5. Τις συναρτήσεις δύο μεταβλητών $\{+\}$ (πρόσθεση) και $\{\cdot\}$ (πολλαπλασιασμός), σαν ερμηνεία των αντίστοιχων συναρτησιακών συμβόλων.

Σημ.1: Τα άγκιστρα γύρω από τα διάφορα σύμβολα δεν είναι μέρος του εκάστοτε συμβόλου, μπαίνουν για αποφυγή συγχύσεων με τη ροή του κειμένου.

Σημ.2: Τα αριθμητικά, τα οποία όπως είδαμε στο 1^ο μέρος είχε διαλέξει ο Hilbert για πρωταρχικά αντικείμενα της αριθμητικής, στην ουσία καλύπτουν όλους τους γνωστούς μας φυσικούς αριθμούς, αφού αυτοί μπορούν να οριστούν ως $S(\bar{0}), SS(\bar{0}), SSS(\bar{0})$ κ.ο.κ., δηλαδή σαν οι επόμενοι του μηδενικού στοιχείου. Για ευκολία στη γραφή, τα συμβολίζουμε από εδώ και στο εξής σαν $\bar{1}, \bar{2}, \bar{3}$ κ.ο.κ.

Σημ.3: Ο συμβολισμός της δομής με το σύμβολο των φυσικών αριθμών αντικατοπτρίζει το γεγονός ότι δουλεύουμε σε ένα πιο ‘πρωτόγονο’ σύμπαν από αυτό των φυσικών αριθμών. Φυσικά με τον ορισμό των αριθμητικών και της χρήσης της συνάρτησης του επομένου που έχει οριστεί στη δομή αυτή, μπορούμε να αναπαράγουμε ακριβώς το σύμπαν των φυσικών, οπότε προκύπτει και ο συμβολισμός της κανονικοποιημένης δομής.

Έχοντας ορίσει τη γλώσσα $\mathcal{L}_A$ και τη σημασιολογία του αλφάβητου που την περιγράφει, το μόνο που μένει είναι να ορίσουμε, για να ολοκληρωθεί η πρωτοβάθμια αριθμητική, είναι ένα κατάλληλο σύνολο μη-λογικών αξιωμάτων. Το σύνολο αυτό θα αποτελείται φυσικά από προτάσεις που θα εμπεριέχουν τις στοιχειώδεις ιδιότητες που ισχύουν στην απλή αριθμητική και μέσω αυτών των προτάσεων - αξιωμάτων και του συστήματος $\mathcal{A}_1$, θα παράγουμε τις αληθείς προτάσεις της αριθμητικής μέσω τυπικών αποδείξεων. Εννοείται βέβαια ότι μπορούν να οριστούν πολλά διαφορετικά σύνολα μη-λογικών αξιωμάτων, ανάλογα με το ποιες στοιχειώδεις ιδιότητες της αριθμητικής θέλουμε να περιγράψουμε μέσω αυτών. Τα αξιωματικά συστήματα που θα προκύψουν είναι αυτά που θα μελετήσουμε ως προς την πληρότητά τους, αφού ορίσουμε την έννοια. Εννοείται επίσης ότι βασική επιθυμητή ιδιότητα είναι η συνέπεια του συνόλου των μη-λογικών αξιωμάτων, ώστε να προκύψουν από αυτά συνεπείς θεωρίες.

Αξιωματικό σύστημα Robinson για την αριθμητική:

Με βάση τα παρακάτω μη-λογικά αξιώματα (S1 - S9), φτιάχνουμε την θεωρία 'αριθμητική του Robinson'. Συμβολίζουμε αυτό το αξιωματικό σύστημα ως S_0 .

$$\mathbf{S1.} \quad S(x) \neq \bar{0}$$

$$\mathbf{S2.} \quad S(x) = S(y) \rightarrow x = y$$

$$\mathbf{S3.} \quad x + \bar{0} = x$$

$$\mathbf{S4.} \quad x + S(y) = S(x + y)$$

$$\mathbf{S5.} \quad x \cdot \bar{0} = \bar{0}$$

$$\mathbf{S6.} \quad x \cdot S(y) = (x \cdot y) + x$$

$$\mathbf{S7.} \quad \neg(x < \bar{0})$$

$$\mathbf{S8.} \quad x < S(y) \leftrightarrow x < y \vee x = y$$

$$\mathbf{S9.} \quad x < y \vee x = y \vee y < x$$

Σημ.1: Προφανώς η δομή $\mathbb{N}$ αποτελεί μοντέλο για το σύνολο των μη-λογικών αυτών αξιωμάτων, αφού οι προτάσεις S1-S9 αποτελούν αληθείς προτάσεις της αριθμητικής.

Σημ.2: Το S_0 αποτελεί ένα πεπερασμένο αξιωματικό σύστημα.

Αν τώρα προσθέσουμε στο σύστημα S_0 ένα άπειρο⁽¹⁾ σχήμα αξιώματος το οποίο θα περιγράφει μία γνωστή μέθοδο απαγωγής στην αριθμητική, τη ‘μαθηματική επαγωγή’, παίρνουμε μια πιο πλούσια θεωρία, την ‘αριθμητική του Peano’. Συμβολίζουμε το αξιωματικό σύστημα $S_1 - S_{10}$, ως S . Επίσης, η δομή $\mathbb{N}$ αποτελεί μοντέλο και του S , εφόσον η μαθηματική επαγωγή ισχύει στην αριθμητική.

S10. (Αξίωμα της μαθηματικής επαγωγής) $(\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x))) \rightarrow \forall x\phi(x)$, όπου $\phi(x)$ είναι ένας τυχαίος τύπος της γλώσσας $\mathcal{L}_A$.

Στην ιστορική αναδρομή του 1^{ου} μέρους έχουμε αναφέρει ότι αυτό ακριβώς το σύστημα, το οποίο συμβολίσαμε ως ‘PA’, είναι που υπόκειται στο θεώρημα μη-πληρότητας του Gödel. Θα δούμε αργότερα το ίδιο ισχύει επίσης για το S_0 , όπως και για οποιαδήποτε άλλη επέκταση του, καθώς και για άλλα αξιωματικά συστήματα, όπως η θεωρία συνόλων των Zermelo-Fraenkel (ZF, από εδώ και στο εξής).

Τώρα που έχουμε πλήρη εικόνα μιας ολοκληρωμένης μαθηματικής θεωρίας, μπορεί να δοθεί ο ορισμός της πληρότητας ενός αξιωματικού συστήματος ώστε να ξεκαθαριστεί το τι ακριβώς δείχνει το θεώρημα μη-πληρότητας και ποια η σχέση του με το αποτέλεσμα της πληρότητας που έχουμε δει να ισχύει στον κατηγορηματικό λογισμό.

Ορισμός: Ένα αξιωματικό σύστημα καλείται **πλήρες**, αν για κάθε πρόταση που πηγάζει από τη θεωρία που περιγράφει, μπορούμε να αποδείξουμε αυτήν ή την άρνησή της. Με άλλη διατύπωση, αν δεν υπάρχουν μη-αποκρίσιμες προτάσεις στα πλαίσια αυτής της θεωρίας.

Έτσι ολοκληρώνουμε τους βασικούς ορισμούς που θα μας απασχολήσουν αργότερα, και μπορούμε, έχοντας μια συνολική εικόνα, να επισημάνουμε και να ξεκαθαρίσουμε κάποια πράγματα σε σχέση με τον παραπάνω ορισμό και τα θεωρήματα ορθότητας και πληρότητας του κατηγορηματικού λογισμού.

(1): Ο χαρακτηρισμός ‘άπειρο’ έχει να κάνει με την ισχύ του αξιώματος για κάθε τύπο ϕ της γλώσσας της αριθμητικής. Από την άλλη, τα αξιώματα S_1-S_9 αναπαριστούν συγκεκριμένες αληθείς προτάσεις, οπότε χαρακτηρίζονται πεπερασμένα.

Στα πλαίσια μίας οποιασδήποτε συνεπούς θεωρίας, το θεώρημα της πληρότητας μας δίνει μία απαραίτητη ιδιότητα, χωρίς την οποία δε θα μπορούσαμε να προχωρήσουμε, ήτοι την ιδιότητα ότι αν ένας τύπος είναι αληθής στην θεωρία αυτή, τότε υπάρχει γι' αυτόν μία τυπική απόδειξη που πηγάζει από τα λογικά, τα μη-λογικά αξιώματά της και τους κανόνες απαγωγής του συστήματος $\mathcal{A}_1$. Παρατηρούμε όμως ότι αυτή η πρόταση δεν μας εξασφαλίζει ότι μπορούμε να αποδείξουμε *κάθε* αληθή πρόταση της θεωρίας μας! *Αυτό θα συνέβαινε εάν δε μπορούσαμε να βγούμε έξω από τα πλαίσια μιας θεωρίας χρησιμοποιώντας μόνο το σύστημα $\mathcal{A}_1$ και τα αξιώματά της θεωρίας.* Τότε πράγματι, οτιδήποτε αποδεικνύαμε μέσα σε μία θεωρία θα ήταν αληθές (από θεώρημα ορθότητας) και όλες οι αληθείς προτάσεις της θεωρίας μας θα είχαν μία τυπική απόδειξη, από το θεώρημα της πληρότητας. Τότε θα λέγαμε ότι έχουμε ένα πλήρες αξιωματικό σύστημα που περιγράφει μια θεωρία, η οποία φυσικά θα περιείχε μόνο αποκρίσιμες προτάσεις, για τις οποίες θα ήμασταν σίγουροι αν υπάρχει απόδειξη για την κάθε μία ή όχι. Στη συνέχεια, από το θεώρημα της ορθότητας, θα διαπιστώναμε επίσης αν ήταν αληθείς ή όχι. Όλα αυτά βέβαια ισχύουν κατά μείζονα λόγο και για την πρωτοβάθμια αριθμητική και τα συστήματα S_0 και S .

Έτσι περιγράφεται η πληρότητα στην οποία αναφέρεται το θεώρημα μη-πληρότητας του Gödel και καταλαβαίνουμε, δεδομένης της παραπάνω ανάλυσης, τη σπουδαιότητα αλλά και τη δυσκολία του εγχειρήματος της απόδειξης του. Για να μπούμε λίγο στο κλίμα, να πούμε ότι, δεδομένου του θεωρήματος πληρότητας και των συνεπειών που είδαμε ότι αυτό θα είχε σε ένα πλήρες αξιωματικό σύστημα, η μεγαλοφυΐα του Gödel έγκειται στο ότι κατάφερε και έφτιαξε *με τα εργαλεία του συστήματος και μόνον*, μία μετα-πρόταση που έβγαине έξω από το σύστημα. Επιπλέον, αυτή η μετα-πρόταση μπορούσε να περιγράψει μία ιδιότητα του εαυτού της, η οποία όμως ιδιότητα έλεγε ακριβώς ότι η πρόταση δεν αποδεικνύεται και άρα είναι μη-αποκρίσιμη!

Για να επιτευχθεί βέβαια κάτι τέτοιο με επιστημονικά άρτιο και αυστηρό τρόπο, χρειάζονται πολλά νέα δεδομένα, ορισμοί και ενδιάμεσα αποτελέσματα (άλλωστε, και το αρχικό άρθρο του Gödel ήταν ιδιαίτερα μεγάλο σε έκταση για τα δεδομένα μιας απόδειξης). Η σειρά που θα παρουσιαστούν όλα αυτά από εδώ και πέρα, είναι κατά βάση η σειρά που ακολουθείται στην απόδειξη του παραρτήματος, χωρίς βέβαια τα κομμάτια που ήδη έχουμε αναλύσει, όπως την πρωτοβάθμια αριθμητική.

2.2 Αναδρομικές συναρτήσεις και κατηγορήματα

2.2.1 Εισαγωγικοί ορισμοί

Σημειογραφία της δομής της αριθμητικής ($\mathbb{N}$):

ω = Σύνολο φυσικών αριθμών.

Μεταβλητές (συμβ. $x, y, z, \dots$) και σταθερές (συμβ. $a, b, c, \dots$) έχουν πεδίο ορισμού το ω

Συναρτήσεις n-θέσεων: $F, G, H, \dots : \omega^n \rightarrow \omega$.

Κατηγορήματα n-αδικά: $P, Q, R, \dots$, υποσύνολα του ω^n .

Μία ακολουθία μεταβλητών $x_1, \dots, x_n$ θα συμβολίζεται ως $\vec{x}$ και όταν γράφουμε $\forall \vec{x}$ εννοείται $\forall x_1, \dots, \forall x_n$.

Δίνουμε τώρα έναν ορισμό που συνδέει τα κατηγορήματα με τις συναρτήσεις, ή μάλλον αντιστοιχεί σε κάθε κατηγορημα μία συγκεκριμένη συνάρτηση:

Ορισμός χαρακτηριστικής συνάρτησης: Σε κάθε κατηγορημα P , αντιστοιχεί μία

χαρακτηριστική συνάρτηση C_P , τέτοια ώστε: $C_P(\vec{x}) = \begin{cases} 0, & \text{αν } P(\vec{x}) \\ 1, & \text{αν } \neg P(\vec{x}) \end{cases}$.

Ορίζουμε επίσης την έννοια της υπολογιστής συνάρτησης σαν μία συνάρτηση F για την οποία υπάρχει αλγόριθμος $\mathcal{P}_F$ τέτοιος ώστε με την εισαγωγή δεδομένων [input] $a_1, \dots, a_n$ να εξάγεται [παράγεται output] το $F(a_1, \dots, a_n)$ μέσω του $\mathcal{P}_F$, σε πεπερασμένο χρόνο.

Για ένα κατηγορημα, λέμε ότι είναι υπολογιστό αν και μόνον αν η χαρακτηριστική του συνάρτηση είναι υπολογιστή.

Πρόβλημα απόφασης [decision problem] ονομάζουμε το πρόβλημα που απαντά ως προς την ύπαρξη διαδικασίας απόφασης για ένα σύνολο B σε ένα υπερσύνολό του, A . Αυτή η διαδικασία είναι ένας αλγόριθμος που υπολογίζει, δεδομένου οποιουδήποτε $x \in A$, αν $x \in B$ ή $x \notin B$ μέσα σε πεπερασμένο χρόνο.

Για παράδειγμα, αν $A = \omega$ και ένα κατηγορημα P τέτοιο ώστε $P(x) \leftrightarrow x \in B$, τότε υπάρχει διαδικασία απόφασης για το $B \subseteq A$ αν και μόνον αν το P είναι υπολογιστό, δηλαδή το συγκεκριμένο πρόβλημα απόφασης λύνεται (καλύτερα, απαντά) αν το P είναι υπολογιστό, ενώ δεν απαντά αν το P δεν είναι υπολογιστό.

2.2.2 Ορισμός της κλάσης των πρωταρχικά αναδρομικών και της κλάσης των αναδρομικών συναρτήσεων

Ορίζουμε αρχικά μία κλάση Σ συναρτήσεων με τα παρακάτω αξιώματα:

R1. (Βασικές συναρτήσεις) Οι παρακάτω συναρτήσεις ανήκουν εξ' ορισμού στην κλάση Σ :

Η μηδενική συνάρτηση	$Z: \omega \rightarrow \omega$ τ.ω. $Z(x) = 0 \quad \forall x \in \omega$
Η συνάρτηση του επόμενου	$\sigma: \omega \rightarrow \omega$ τ.ω. $\sigma(x) = x+1 \quad \forall x \in \omega$
Η συνάρτηση της πρόσθεσης	$+: \omega^2 \rightarrow \omega$ τ.ω. $+(x_1, x_2) = x_1 + x_2$
Η συνάρτηση του πολλαπλασιασμού	$\cdot: \omega^2 \rightarrow \omega$ τ.ω. $\cdot(x_1, x_2) = x_1 \cdot x_2$
Η συνάρτηση του μικροτέρου	$C<: \omega^2 \rightarrow \omega$ τ.ω. $C<(x_1, x_2) = x_1 < x_2$
Η συνάρτηση προβολής	$I: \omega^n \rightarrow \omega$ τ.ω. $I_i^n(x_1, \dots, x_n) = x_i$

R2. (Σχήμα της αντικατάστασης) Αν οι συναρτήσεις $G, H_1, \dots, H_r \in \Sigma$ και η συνάρτηση F δίνεται από τον τύπο $F(\vec{x}) = G(H_1(\vec{x}), H_2(\vec{x}), \dots, H_r(\vec{x}))$, τότε $F \in \Sigma$.

R3. (Σχήμα αναδρομής) Αν $G, H \in \Sigma$ και η F δίνεται από τον επαγωγικό ορισμό:

$$\begin{cases} F(0, \vec{x}) = G(\vec{x}) \\ F(y+1, \vec{x}) = H(F(y, \vec{x}), y, \vec{x}) \end{cases}$$

τότε $F \in \Sigma$.

Ορισμός: Η κλάση των **πρωταρχικά (ή πρωτογονικά) αναδρομικών συναρτήσεων** είναι η *μικρότερη* κλάση συναρτήσεων που ικανοποιεί τα αξιώματα R1, R2 και R3. Τη συμβολίζουμε σαν 'Πρωτ.Αν.' για συντομία.

Με αυτόν τον τρόπο ορίζουμε και μπορούμε να αναγνωρίσουμε τις πρωταρχικά αναδρομικές συναρτήσεις του τύπου $F(x_1, \dots, x_n)$, ενώ αποδεικνύεται ότι μπορούμε να αναγνωρίσουμε, αντιμεταθέσουμε ή προσθέσουμε μεταβλητές στις $x_1, \dots, x_n$ χωρίς να βγούμε έξω από την κλάση Πρωτ.Αν.

Σημ.: Όλες οι 'συνηθισμένες' συναρτήσεις, όπως του τετραγώνου, η εκθετική κτλ., βγαίνουν τελικά ότι είναι πρωταρχικά αναδρομικές.

Για να ορίσουμε τώρα την κλάση των αναδρομικών συναρτήσεων, χρειαζόμαστε έναν τελεστή που ονομάζεται **μ-τελεστής** και ορίζεται ως εξής:

Αν $P(\vec{x}, y)$ ένα κατηγορημα τέτοιο ώστε $\forall \vec{y} \exists x P(\vec{x}, y)$, τότε ορίζουμε τον μ-τελεστή σαν το ελάχιστο κ τ.ω. $P(\vec{x}, \kappa)$. Συμβολίζουμε αυτή την ελάχιστη τιμή για την οποία το κατηγορημα ισχύει σαν $\mu y P(\vec{x}, y)$.

Με βάση τον μ-τελεστή παίρνουμε ένα ακόμα αξίωμα που αφορά συναρτήσεις:

R4. (μ-σχήμα) Αν $G \in \Sigma$ και $\forall \vec{y} \exists x (G(\vec{y}, x) = 0)$, τότε $\mu x (G(\vec{y}, x) = 0) \in \Sigma$

Ορισμός: Η κλάση των **αναδρομικών συναρτήσεων** είναι η *μικρότερη* κλάση συναρτήσεων που ικανοποιεί τα αξιώματα R1, R2 και R4. Για συντομία τη συμβολίζουμε 'Αναδ.'.

Σημ.: Η κλάση των αναδρομικών συναρτήσεων είναι υπερ-κλάση αυτής των πρωταρχικά αναδρομικών, αφού ισχύει για αυτήν και το R3 αξίωμα, από ένα αποτέλεσμα που θα δούμε αργότερα.

Η επέκταση των παραπάνω ορισμών στα κατηγορήματα γίνεται εύλογα μέσω των χαρακτηριστικών του συναρτήσεων. Συγκεκριμένα:

Ορίζουμε ένα κατηγορημα να είναι αναδρομικό (αντ. πρωταρχικά αναδρομικό) αν η χαρακτηριστική του συνάρτηση είναι αναδρομική (αντ. πρωταρχικά αναδρομική).

Έχοντας ορίσει τις κλάσεις που περιλαμβάνουν τις αναδρομικές συναρτήσεις (και κατηγορήματα φυσικά), είναι φυσικό να ψάξουμε τα 'όριά' τους, δηλαδή να διαπιστώσουμε ποιες πράξεις πάνω σε αναδρομικές συναρτήσεις και κατηγορήματα μας δίνουν επίσης αναδρομικές συναρτήσεις ή κατηγορήματα, καθώς και να ορίσουμε συγκεκριμένες αναδρομικές συναρτήσεις. Ο απώτερος σκοπός είναι να βγάλουμε ένα γενικό κανόνα για τις αναδρομικές συναρτήσεις, ώστε να διαπιστώνουμε με αυτόν την αναδρομικότητα και όχι με τον (δύσχρηστο) ορισμό.

Χωρίς να επεκταθούμε σε τεχνικές λεπτομέρειες, μέσω των προτάσεων-κανόνων **C1-C10** (βλ. Παράρτημα, σελ. iii-vi) διαπιστώνουμε κάτι από τα παραπάνω, όπως το ότι χρησιμοποιώντας λογικούς ποσοδείκτες, σταθερές και τα κατηγορήματα $<$, $\leq$, και $=$, παραμένουμε στην κλάση των αναδρομικών συναρτήσεων. Επίσης σε

αυτούς τους κανόνες ορίζουμε την ‘πλον’ συνάρτηση (συμβ. ‘ $\div$ ’) που είναι η γνωστή μας αφαίρεση με τη διαφορά ότι δεν ορίζεται αποτέλεσμα έξω από το πεδίο των φυσικών αριθμών, ήτοι το πεδίο τιμών της φράσσεται από κάτω στο μηδέν. Ορίζεται ακόμη ο *φραγμένος* μ-τελεστής ως εξής:

Ορισμός: Για κάθε κατηγορημα $P(\vec{x}, y)$ ορίζουμε τον φραγμένο μ-τελεστή μέσω του τύπου $\mu_{x < z} P(\vec{y}, x) = \mu x (P(\vec{y}, x) \vee x = z)$.

Ο τελεστής αυτός δίνει το ελάχιστο $x < z$, αν αυτό υπάρχει, ή το z αλλιώς, τέτοιο ώστε να ισχύει η ιδιότητα του P . Το άνω φράγμα για την τιμή του x είναι αυτό που τον διαφοροποιεί από τον μ-τελεστή και του δίνει την ονομασία ‘φραγμένος’. Αποδεικνύεται (κανόνας C7) ότι ο φραγμένος μ-τελεστής δε μας οδηγεί έξω από την κλάση των αναδρομικών συναρτήσεων.

Ένας πολύ χρήσιμος ορισμός που προκύπτει επίσης από τους προαναφερθέντες κανόνες, είναι αυτός που περιγράφει τον **σαφή ορισμό** μιας συνάρτησης σε σχέση με άλλες συναρτήσεις και κατηγορήματα. Ο σαφής ορισμός προκύπτει όταν ορίζω μία συνάρτηση χρησιμοποιώντας *μόνο* τον μ-τελεστή και την αντικατάσταση πάνω σε άλλες συναρτήσεις ή/και κατηγορήματα. Προκύπτει ότι:

Αν μία συνάρτηση έχει σαφή ορισμό σε σχέση με αναδρομικές συναρτήσεις και κατηγορήματα, τότε είναι επίσης αναδρομική.

2.2.3 Επιπλέον ορισμοί αναδρομικών συναρτήσεων και ένας γενικός κανόνας

Σε αυτό το υποκεφάλαιο θα ορίσουμε μερικές ακόμα αναδρομικές συναρτήσεις που είναι απαραίτητες για την πορεία της απόδειξης, ενώ στο τέλος του θα δώσουμε ένα γενικό κανόνα που μας βοηθά στην αναγνώριση της αναδρομικότητας μίας συνάρτησης.

Παρατήρηση: Ασχολούμαστε μόνο με τις αναδρομικές συναρτήσεις διότι αποτελούν υπερ-κλάση των πρωταρχικά αναδρομικών και, όπως θα δούμε παρακάτω, είναι αυτές που χρειαζόμαστε για να βγάλουμε το τελικό συμπέρασμα. Ένας ακόμη λόγος είναι η αδυναμία των πρωταρχικά αναδρομικών συναρτήσεων να περιγράψουν την έννοια της υπολογιστότητας, κάτι που σκιαγραφήθηκε σαν συμπέρασμα στην ιστορική αναδρομή του 1^{ου} μέρους και θα αναλυθεί περισσότερο παρακάτω.

Αρχίζουμε με τον ορισμό της **β-συνάρτησης του Gödel**, μίας βοηθητικής, στην ουσία, συνάρτησης που τη χρησιμοποιούμε για να ορίσουμε άλλες χρήσιμες συναρτήσεις.

Ορισμός: Η β-συνάρτηση του Gödel είναι μία *δυαδική αναδρομική συνάρτηση* τέτοια ώστε: (i) $\beta(x, y) \leq x \div 1$

(ii) Για κάθε n και για κάθε ακολουθία φυσικών αριθμών $a_1, \dots, a_{n-1}$, υπάρχει αριθμός α τέτοιος ώστε $\beta(\alpha, i) = a_i \quad \forall i < n$.

Την ύπαρξη της β-συνάρτησης καθώς και την αναδρομικότητά της, τις αποδεικνύουμε με τη βοήθεια μίας άλλης παρεμφερούς συνάρτησης, της δ-συνάρτησης και της συνάρτησης ζεύγους, Pair. Για περισσότερες λεπτομέρειες ο αναγνώστης παραπέμπεται στο Παράρτημα (σελ. ix-x).

Η β-συνάρτηση επεκτείνεται σε μία 1-1 συνάρτηση, την $\prec \dots \succ: \omega^n \rightarrow \omega$, τέτοια ώστε, $\prec y_0, \dots, y_{n-1} \succ = \mu x [\beta(x, 0) = n \wedge \beta(x, 1) = y_0 \wedge \dots \wedge \beta(x, y) = y_n]$. Ο αριθμός που προκύπτει από τη χρήση του μ-τελεστή, καλείται **‘αριθμός ακολουθίας’** του $\prec y_0, \dots, y_{n-1} \succ$ και για διαφορετικές ακολουθίες παίρνει διαφορετικές τιμές, αφού η συνάρτηση είναι 1-1. Έτσι, έχοντας έναν αριθμό ακολουθίας μπορούμε να είμαστε σίγουροι ότι αυτός αντιστοιχεί σε μία συγκεκριμένη ακολουθία και σε καμία άλλη, μία ιδιότητα πολύ σημαντική για την ανάλυσή μας. Εξ’ ορισμού, ο αριθμός ακολουθίας της κενής ακολουθίας είναι το μηδέν ($\prec \emptyset \succ = 0$). Η συνάρτηση αυτή αποδεικνύεται ότι είναι επίσης αναδρομική.

Με τη βοήθεια της β-συνάρτησης, ορίζουμε επίσης δύο ακόμα χρήσιμες συναρτήσεις που μας δίνουν εκφράσεις για το μήκος μιας ακολουθίας και για την επιλογή ενός στοιχείου της. Συγκεκριμένα έχουμε:

Ορισμός: Ορίζουμε τις αναδρομικές συναρτήσεις ‘μήκος του x ’ και ‘ i -οστό στοιχείο του x ’ ως: (i) $lh(x) = \beta(x, 0)$ και (ii) $(x)_i = \beta(x, i+1)$, αντίστοιχα.

Με αυτό τον επιπλέον ορισμό είμαστε τώρα σε θέση να ορίσουμε ένα κατηγορημα που θα απαντά ως προς το αν ένας αριθμός είναι αριθμός ακολουθίας, καθώς και μία συνάρτηση που μας δίνει τη δυνατότητα να ‘μειώνουμε’ ακολουθίες.

Ορισμός: Το κατηγορημα $Seq(x)$ που σημαίνει ‘το x είναι αριθμός ακολουθίας’, ορίζεται από τον τύπο: $Seq(x) = \forall y, y < x [lh(y) = lh(x) \rightarrow \exists i, i < lh(x) ((y)_i \neq (x)_i)]$ και είναι αναδρομικό.

Σημ.: Αν $x = \prec (x)_0, \dots, (x)_{n-1} \succ$, τότε ο τύπος $Seq(x) \wedge lh(x) = n$ μας δίνει το x , όπως προκύπτει από τους παραπάνω ορισμούς.

Ορισμός: Η **συνάρτηση αναγωγής [reduct function]** είναι μία δυαδική αναδρομική συνάρτηση που ορίζεται από τον τύπο: $\text{Red}(x, i) = \mu y [\text{lh}(y) = i \wedge \forall j < i ((y)_j = (x)_j)]$ και έχει σαν χαρακτηριστική ιδιότητα ότι μειώνει το μήκος ακολουθιών, για παράδειγμα, $\text{Red}\{(y_0, \dots, y_{n-1}), i\} = (y_0, \dots, y_{i-1}) \quad (i \leq n)$.

Σε αυτό το σημείο, κλείνοντας το κεφάλαιο των αναδρομικών συναρτήσεων, θα δώσουμε έναν ορισμό γενικής αναδρομής και θα καταλήξουμε μέσω αυτού σε έναν κανόνα αναγνώρισης της αναδρομικότητας μίας συνάρτησης.

Γενική αναδρομή: Αν G μία αναδρομική συνάρτηση και F μία συνάρτηση που ορίζεται από τον τύπο $F(y, \vec{x}) = G(\bar{F}(y, \vec{x}), y, \vec{x})$, τότε η F είναι αναδρομική.

Η συνάρτηση $\bar{F}$ που εμφανίζεται στο ορισμό της γενικής αναδρομής, εκπορεύεται από την $F(y, \vec{x})$ και ορίζεται αναδρομικά ως εξής:

$$\bar{F}(y, \vec{x}) = \begin{cases} \bar{F}(0, \vec{x}) = \langle \emptyset \rangle = 0 \\ \bar{F}(y, \vec{x}) = \langle F(0, \vec{x}), \dots, F(y-1, \vec{x}) \rangle \end{cases}$$

Σαν **πόρισμα** της γενικής αναδρομής, το αξίωμα R_3 ισχύει και για την κλάση των αναδρομικών συναρτήσεων αφού προκύπτει ότι οποιαδήποτε αναδρομική συνάρτηση F ικανοποιεί το R_3 , εφόσον μπορεί να προκύψει επαγωγικά ως εξής:

$$\begin{cases} F(0, \vec{x}) = G(\vec{x}) \\ F(y+1, \vec{x}) = H(F(y, \vec{x}), y+1, \vec{x}) \end{cases}$$

οπότε αν οι G, H αναδρομικές, η F θα είναι επίσης αναδρομική.

Παρατήρηση: Για τα κατηγορήματα ισχύει επίσης η γενική αναδρομή.

Γενικός Κανόνας: Ένας ορισμός της μορφής « $F(y, \vec{x}) =$ κάποιος αναδρομικός ορισμός που περιλαμβάνει την F », δίνει αναδρομική συνάρτηση F , υπό την προϋπόθεση ότι όποτε απαντάται το $F(w, \vec{x})$, να μπορούμε να αποδείξουμε ότι $w < y$.

2.3 Αναπαραστάσιμες συναρτήσεις & κατηγορήματα

Στο κεφάλαιο αυτό θα ορίσουμε την έννοια της αναπαραστάσιμης συνάρτησης και κατηγορήματος σε μία θεωρία T . Σε σχέση με την ολοκληρωμένη απόδειξη του παραρτήματος, ο ορισμός υπάρχει στο κεφάλαιο ‘Πρωτοβάθμια Αριθμητική’ (σελ. xiii) στο οποίο καλύπτονται τα αξιωματικά συστήματα των Robinson και Peano (βλ. § 2.1.5) και αποδεικνύονται κάποια θεωρήματα της θεωρίας S_0 . Μιας και έχει γίνει εκτενής αναφορά στην πρωτοβάθμια αριθμητική δεν θεωρείται σκόπιμο να επαναληφθεί η ανάλυση των θεωριών, οπότε περνάμε κατευθείαν στο θέμα του τίτλου του κεφαλαίου. [Σημ: Για τη γενική γλώσσα του κατηγορηματικού λογισμού στο κεφάλαιο αυτό του παραρτήματος, χρησιμοποιείται το σύμβολο $\mathcal{L}$, ενώ παρακάτω χρησιμοποιείται το *ίδιο* σύμβολο για να περιγράψει τη γλώσσα της αριθμητικής. Στα πλαίσια της εργασίας θα συνεχιστεί η χρήση του συμβόλου $\mathcal{L}_A$ για την περιγραφή της τελευταίας.]

Ορισμός: Έστω μία θεωρία T στην $\mathcal{L}_A$. Ένα **κατηγόρημα** $R \subseteq \omega^n$ είναι **αναπαραστάσιμο** στην T αν υπάρχει κάποιος τύπος $\phi(x_1, \dots, x_n)$ τέτοιος ώστε:

- $R(a_1, \dots, a_n) \Rightarrow \vdash_T \phi(\bar{a}_1, \dots, \bar{a}_n)$, δηλαδή αν το R ισχύει, τότε υπάρχει απόδειξη για τον ϕ στην T , και
- $\neg R(a_1, \dots, a_n) \Rightarrow \vdash_T \neg \phi(\bar{a}_1, \dots, \bar{a}_n)$, δηλαδή αν το R δεν ισχύει, υπάρχει απόδειξη της άρνησης του ϕ στην T .

Η απλή συνεπαγωγή μπορεί να αντικατασταθεί από τη διπλή σε μία *συνεπή* θεωρία, αφού μόνο με την υπόθεση της συνέπειας θα έχουμε εξασφαλισμένη την μη-ύπαρξη απόδειξης για τον τύπο $\neg \phi$, δεδομένης της ύπαρξης απόδειξης για τον τύπο ϕ .

Ορισμός: Μία **συνάρτηση** $F: \omega^n \rightarrow \omega$ είναι **αναπαραστάσιμη** σε μία θεωρία T αν υπάρχει τύπος $\phi(x_1, \dots, x_n, y)$ τέτοιος ώστε:

$F(a_1, \dots, a_n) = b \Rightarrow \vdash_T \phi(\bar{a}_1, \dots, \bar{a}_n, y) \leftrightarrow y = \bar{b}$, δηλαδή αν η συνάρτηση πάρει μία τιμή b , τότε υπάρχει απόδειξη στη T για το ότι ο τύπος $\phi(x_1, \dots, x_n, y)$ είναι έγκυρος αν και μόνον αν $y = b$.

Οι δύο παραπάνω ορισμοί συνδέονται μέσω του επόμενου αποτελέσματος.

Πρόταση: Αν T οποιαδήποτε επέκταση της θεωρίας S_0 και κατηγορημα $R \subseteq \omega^n$, τότε το R είναι αναπαραστάσιμο στην T αν και μόνον αν η χαρακτηριστική του συνάρτηση είναι αναπαραστάσιμη στην T .

Η έννοια την αναπαραστασιμότητας θα μας χρησιμεύσει αργότερα που θα αναπαραστήσουμε με αυστηρά καθορισμένο τρόπο τα αντικείμενα της λογικής και της αριθμητικής. Για την ώρα θα δούμε δύο πολύ σημαντικά αποτελέσματα που προκύπτουν άμεσα :

Θεώρημα 2.3.1: Αν υποθέσουμε ότι μία συνάρτηση F είναι αναπαραστάσιμη στην S_0 από τον τύπο $\phi(x_1, \dots, x_n, y)$, τότε η F είναι υπολογιστή. Αυτό συμβαίνει διότι υπάρχει μια διαδικασία (αλγόριθμος) υπολογισμού του $F(a_1, \dots, a_n)$ η οποία είναι η εξής: Απαριθμούμε τα θεωρήματα του S_0 μέχρι να φτάσουμε στο $\vdash_{S_0} \phi(\bar{a}_1, \dots, \bar{a}_n, \bar{b})$, για κάποιο $b \in \omega$. Τότε σταματάμε και ξέρουμε ότι $F(a_1, \dots, a_n) = b$.

Θεώρημα 2.3.2: Κάθε αναδρομική συνάρτηση είναι αναπαραστάσιμη στο S_0 .

Το θεώρημα 2.3.2, το οποίο προφανώς ισχύει και για κάθε αναδρομικό κατηγορημα, είναι αυτό το αποτέλεσμα στο οποίο θέλαμε να καταλήξουμε για να περάσουμε στο κύριο μέρος της απόδειξης. Αυτό που μας δίνει και είναι το σημαντικό, είναι ότι το βασικό μας εργαλείο, οι αναδρομικές συναρτήσεις (και τα κατηγορήματα), πάνω στο οποίο θα βασιστεί η περαιτέρω ανάλυση, μπορεί να πάρει αριθμητική μορφή, δηλαδή *μπορεί να αναπαρασταθεί στη γλώσσα της αριθμητικής*. Αυτή η δυνατότητα είναι που επέτρεψε στον Gödel να εκφράσει τη μη-αποκρίσιμη πρόταση από την οποία προκύπτει η μη-πληρότητα της αριθμητικής. Διότι, όπως έχουμε δει, δεδομένου του θεωρήματος πληρότητας για τον κατηγορηματικό λογισμό, η μόνη δυνατότητα απόδειξης της μη-πληρότητας, έπρεπε να περιλαμβάνει μετα-προτάσεις, οι οποίες όμως θα πρέπει να είναι επίσης προτάσεις τις αριθμητικής, εφόσον θέλουμε να μην ξεφύγουμε από τα πλαίσιά της, δηλαδή αναπαραστάσιμες προτάσεις στη γλώσσα $\mathcal{L}_A$. Το σκεπτικό γύρω από την επιλογή της κατάλληλης μετα-πρότασης η οποία προκύπτει τελικά μη-αποκρίσιμη, αλλά και την πλήρη διαδικασία δόμησης μίας τέτοιας πρότασης θα δούμε αναλυτικά στο επόμενο κεφάλαιο, οπότε και θα ολοκληρωθεί η απόδειξη του 1^{ου} θεωρήματος μη-πληρότητας.

3. Κύριο μέρος της απόδειξης

3.1 Αριθμητικοποίηση της Λογικής

Η διαδικασία αναπαράστασης των απαιτούμενων στοιχείων που θα μας οδηγήσουν στην αναπαράσταση στη γλώσσα της αριθμητικής, προτάσεων που αφορούν την ίδια τη θεωρία της αριθμητικής (μετα-προτάσεων), είναι το βασικό αντικείμενο επεξεργασίας αυτού του κεφαλαίου. Καταλαβαίνουμε βέβαια αμέσως ότι αυτή η αναπαράσταση θα πρέπει να συνίσταται στην κωδικοποίηση όλων των στοιχείων της θεωρίας της αριθμητικής (αξιωμάτων, κανόνων απαγωγής, συμβόλων κτλ.) με *αριθμητικά*, καθότι αυτά είναι τα αντικείμενα του σύμπαντος της θεωρίας της αριθμητικής επεξεργάζεται. Αυτή ακριβώς την κωδικοποίηση αναφέρουμε σαν ‘Αριθμητικοποίηση της Λογικής’ και είναι η κεντρική ιδέα που οδήγησε στην απόδειξη των θεωρημάτων μη-πληρότητας. Φυσικά, η υλοποίηση μιας τέτοιας ιδέας θέλει προσεκτική επεξεργασία για να μην υποπέσουμε σε αντιφάσεις, ή διαφορούμενα νοήματα. Η ολοκληρωμένη μορφή της υπάρχει στην απόδειξη του παραρτήματος (βλ. κεφ. 3, σελ. xxι-xxvi). Στο παρόν κεφάλαιο θα σκιαγραφήσουμε τα κυριότερα μόνο σημεία της μεθόδου.

Εύλογα, αρχίζουμε την κωδικοποίηση που σκιαγραφήθηκε από το αλφάβητο της γλώσσας $\mathcal{L}_A$, ήτοι από την κωδικοποίηση των λογικών και μη-λογικών συμβόλων της, εκτός των μεταβλητών. Αυτά τα σύμβολα κωδικοποιούνται με τους *περιττούς αριθμούς* από το 1 έως το 17, ενώ οι μεταβλητές με *άρτιους αριθμούς* εν γένει. Ακολουθούμε τώρα την επαγωγική πορεία που περιγράφηκε στην παράγραφο 2.1.3 για να κωδικοποιήσουμε τους όρους, τους ατομικούς τύπους και τους τύπους της γλώσσας της αριθμητικής.

Με την ακριβή διαδικασία που περιγράφεται στο παράρτημα, μέχρι τώρα, για παράδειγμα, ο τύπος $S(x_1 + x_2) = x_3 \cdot x_4$, θα μπορούσε να κωδικοποιηθεί σαν 17-1-3-2-4-5-6-8 (χρησιμοποιήθηκε η τυπικά σωστή μορφή του: ‘ $= (S(+ (x_1, x_2)), \cdot (x_3, x_4))$ ’). Καταλαβαίνει κανείς πώς με αυτόν τον τρόπο, θα καταλήξουμε σε τεράστιους αριθμούς όταν προσπαθήσουμε να κωδικοποιήσουμε προτάσεις ολόκληρες (και πράγματι, στο αρχικό άρθρο κάτι τέτοιο συνέβαινε). Για να απλουστευτεί λίγο η κωδικοποίηση, χρησιμοποιούμε τον ‘αριθμό ακολουθίας’ (βλ. § 2.2.3) και

καταλήγουμε σε ένα συγκεκριμένο αριθμό για κάθε όρο, τύπο ή ατομικό τύπο που κωδικοποιούμε. Για παράδειγμα, για έναν όρο της μορφής $\Phi = t_1 + t_2$, όπου τα t_1, t_2 μεταβλητές ή σταθερές, βγάζουμε τον αριθμό ακολουθίας $\prec \oplus, \lceil t_1 \rceil, \lceil t_2 \rceil \succ$ και τον συμβολίζουμε $\lceil \Phi \rceil$. Όλοι αυτοί οι αριθμοί που χρησιμοποιούμε στην παραπάνω κωδικοποίηση, ονομάζονται Gödel-αριθμοί, προς τιμή του εμπνευστή αυτής της μεθόδου.

Προσοχή: Ο συμβολισμός διαφέρει από αυτόν του παραρτήματος όπου οι Gödel-αριθμοί σηματοδοτούνται με δύο μικρότερα ‘γάμμα’ πάνω δεξιά και πάνω αριστερά του κωδικοποιημένου συμβόλου.

Με δύο απλές παρατηρήσεις διαπιστώνουμε πως έχουμε στα χέρια μας έναν πολύ καλό μηχανισμό κωδικοποίησης, μια υλοποίηση δηλαδή της αρχικής μας ιδέας που μας δίνει τις σωστές βάσεις για να συνεχίσουμε.

Παρατήρηση 1^η: Όπως έχουμε ορίσει τον αριθμό ακολουθίας (‘1-1’ συνάρτηση) και την κωδικοποίηση των όρων, τύπων και ατομικών τύπων (επαγωγικά), εξασφαλίζουμε ότι η κωδικοποίησή μας δεν οδηγεί σε διφορούμενα αποτελέσματα, δηλαδή διαφορετικές εκφράσεις έχουν διαφορετικούς Gödel-αριθμούς.

Παρατήρηση 2^η: Από τον ορισμό του κατηγορήματος ‘Seq’ (βλ. § 2.2.3) προκύπτει μία αλγοριθμική διαδικασία που καθορίζει αν ένας τυχαίος αριθμός a είναι Gödel-αριθμός ή όχι και ανακατασκευάζει την έκφραση που αναπαριστά ο a αν αυτός όντως είναι Gödel-αριθμός. Αυτό γίνεται υποθέτοντας για τον a ότι ισχύει το $\text{Seq}(a)$, οπότε αυτός είναι της μορφής $a = \prec (a)_0, \dots, (a)_{n-1} \succ$ με $n = \text{lh}(a)$. Αν $n = 2$ ή 3 , τότε ο a δεν είναι Gödel-αριθμός όπως προκύπτει από τους επαγωγικούς ορισμούς των όρων, τύπων κτλ. Αλλιώς, ελέγχω αν το $(a)_0$ είναι μονός αριθμός μικρότερος του 17 και αν δεν ισχύει τότε πάλι δεν έχω Gödel-αριθμό, ενώ αν ισχύει, ψάχνω τα $(a)_1, \dots, (a)_{n-1}$ κ.ο.κ., οπότε βήμα βήμα μπορώ να ανακατασκευάσω την αρχική έκφραση. Το ότι δεν είναι κάθε τυχαίος αριθμός Gödel-αριθμός προκύπτει από τον ορισμό του αριθμού ακολουθίας σαν μία συνάρτηση ‘1-1’ αλλά όχι ‘επί’, δηλαδή μία συνάρτηση της οποίας το πεδίο τιμών δεν καλύπτει όλους τους φυσικούς αριθμούς.

Αυτό που μένει τώρα είναι να αντιστοιχίσουμε *προτάσεις* και φυσικά *αποδείξεις* σε Gödel-αριθμούς ώστε να ολοκληρωθεί η αριθμητικοποίηση. Φυσικά, για να γίνει κάτι τέτοιο πρέπει να κωδικοποιηθεί πλήρως το αξιωματικό σύστημα $\mathcal{A}_1$ του κατηγορηματικού λογισμού, εφόσον σύμφωνα με αυτό δομούμε αποδείξεις για να βγάλουμε τις αληθείας προτάσεις μίας θεωρίας.

Αυτή η διαδικασία περιγράφεται αναλυτικά στο κομμάτι της απόδειξης του παραρτήματος με τίτλο ‘Ορισμοί για την αριθμητικοποίηση της λογικής’. Η διαδικασία είναι η εξής: Ορίζουμε αναδρομικές (και άρα αναπαραστάσιμες στην αριθμητική) συναρτήσεις και κατηγορήματα που παίρνουν σαν τιμές Gödel-αριθμούς, αρχίζοντας από τις μεταβλητές, τις εκφράσεις τους τύπους και τους όρους. Προχωρούμε με τους ορισμούς για τους κανόνες αντικατάστασης, των ελεύθερων μεταβλητών, των δύο κανόνων συνεπαγωγής (Modus Ponens και Gen) και των λογικών και μη-λογικών αξιωμάτων. Η όλη διαδικασία είναι σχετικά απλή και βασίζεται στην αρχική κωδικοποίηση που περιγράφηκε παραπάνω. Για παράδειγμα, το κατηγορήμα $Lax(a)$ δηλώνει ότι ο αριθμός a είναι Gödel-αριθμός ενός λογικού αξιώματος (A1-A5, βλ. §2.1.4) και δίνεται από έναν τύπο που περιγράφει την κωδικοποίηση που μπορεί να γίνει με τα υπάρχοντα εργαλεία σε καθένα από αυτά τα αξιώματα.

Έτσι, καταλήγουμε σε έναν ορισμό που αναθέτει σε μία *ακολουθία τύπων* έναν Gödel-αριθμό, ώστε να καταλήξουμε στην ανάθεση Gödel-αριθμών σε αποδείξεις.

Ορισμός: Σε κάθε πεπερασμένη ακολουθία από τύπους $\phi_1, \dots, \phi_n$ αναθέτουμε με μία ‘1-1’ αντιστοιχία έναν αριθμό $a = \langle \lceil \phi_1 \rceil, \dots, \lceil \phi_n \rceil \rangle$. Με αυτόν τον τρόπο, κάθε ακολουθία τύπων έχει μοναδικό Gödel-αριθμό.

Με βάση αυτόν τον ορισμό και την μέχρι τώρα διαδικασία, είμαστε τελικά σε θέση να ορίσουμε ένα κατηγορήμα που θα απαντά αν ένας Gödel-αριθμός αντιστοιχεί σε μία ακολουθία τύπων που αποτελούν απόδειξη για κάποιον άλλο τύπο.

Ορισμός: $\text{Prf}_T(a) \Leftrightarrow$ Υπάρχει μία ακολουθία τύπων $\phi_1, \dots, \phi_n$ η οποία αποτελεί μία απόδειξη για τον τύπο ϕ_n τέτοια ώστε $a = \langle \lceil \phi_1 \rceil, \dots, \lceil \phi_n \rceil \rangle$.

Σημ.: Για την απόδειξη της αναδρομικότητας του $\text{Prf}_T(a)$ χρειαζόμαστε μία συνθήκη που δεν έχουμε ορίσει ακόμα, ότι η θεωρία T είναι αξιωματοποιημένη (βλ. επόμενη παράγραφο, 3.2). Θα δούμε αργότερα πως ουσιαστικά μόνο με τέτοιου είδους θεωρίες ασχολούμαστε, οπότε το κατηγορήμα θεωρείται εν γένει αναδρομικό.

Για να αποδώσουμε, τέλος, έναν συγκεκριμένο Gödel-αριθμό σε μία απόδειξη, χρησιμοποιούμε το δυαδικό κατηγορήμα $\text{Prf}_T(a)$ που ορίζεται ως εξής:

$\text{Prf}_T(a, b) \Leftrightarrow \text{Prf}_T(b) \wedge a = (b)_{\text{lh}(b) - 1}$, δηλαδή ο b είναι ο **Gödel-αριθμός απόδειξης** της απόδειξης που αποδεικνύει τον τύπο με Gödel-αριθμό a .

Το κατηγορήμα αυτό αποδεικνύεται επίσης αναδρομικό, με την προϋπόθεση που αναφέρεται στην παραπάνω σημείωση.

Σαν αποτέλεσμα της όλης διαδικασίας που περιγράφηκε μέχρι τώρα, δίνεται ακόμη η δυνατότητα απόδειξης του αντίστροφου του θεωρήματος 2.3.2, δηλαδή του εξής αποτελέσματος:

Θεώρημα 3.1.1: Κάθε αναπαραστάσιμη συνάρτηση στο S_0 είναι αναδρομική.

Πόρισμα 3.1.2: Αν T είναι μία αξιωματοποιημένη επέκταση του S_0 (π.χ. η PA με το αξιωματικό σύστημα S), τότε οι αναπαραστάσιμες σε αυτήν συναρτήσεις και κατηγορήματα είναι *μόνον* οι αναδρομικές συναρτήσεις και κατηγορήματα.

Με αυτόν τον τρόπο φτάνουμε στον στόχο της κωδικοποίησης ολόκληρης της θεωρίας της αριθμητικής με φυσικούς (Gödel-)αριθμούς και ολοκληρώνεται έτσι η δυνατότητα έκφρασης προτάσεων της θεωρίας μέσω αριθμών. Πετύχαμε δηλαδή την *πλήρη* αριθμητικοποίηση της λογικής.

Αυτό που μένει τώρα να γίνει, είναι να βρεθεί η ακριβής μετα-πρόταση που θα επεξεργαστούμε με την αριθμητική ώστε να προκύψει η μη-πληρότητά της, αφού όμως πρώτα αναδιατυπώσουμε το αποτέλεσμα αυτό, ώστε να ταιριάζει με την μέχρι τώρα θεώρησή μας. Γι' αυτό το λόγο, στο κεφάλαιο που ακολουθεί, ορίζονται όλες οι απαραίτητες νέες έννοιες πριν από την κύρια απόδειξη, έννοιες όπως αυτή της 'αξιωματοποιημένης θεωρίας', που αναφέρθηκε παραπάνω.

3.2 Αναδιατύπωση και απόδειξη του 1^{ου} θεωρήματος μη-πληρότητας

Για τη διευκόλυνση της σημειογραφίας, θέτουμε καταρχήν δύο σύνολα για μία θεωρία T της γλώσσας της αριθμητικής, τα εξής:

- Το σύνολο των Gödel-αριθμών που αντιστοιχούν στα μη-λογικά αξιώματά της: $Ax_T = \{ \ulcorner \phi \urcorner \mid \phi = \text{μη λογικό αξίωμα της } T \}$
- Το σύνολο των Gödel-αριθμών που αντιστοιχούν στα θεωρήματά της: $Thm_T = \{ \ulcorner \phi \urcorner \mid \vdash_T \phi \}$

Επιπλέον, ορίζουμε τα παρακάτω:

- **Αναδρομικό σύνολο** καλείται ένα σύνολο $A \subseteq \omega$ αν και μόνον αν το κατηγορημα $x \in A$ είναι αναδρομικό.
- **Αναδρομικά αριθμήσιμο [recursively enumerable – r.e.] κατηγορημα** καλείται ένα κατηγορημα $P(\vec{x})$ αν υπάρχει ένα άλλο αναδρομικό κατηγορημα $R(\vec{x}, y)$, τέτοιο ώστε να ισχύει $P(\vec{x}) \leftrightarrow \exists y R(\vec{x}, y)$.
- **Αναδρομικά αριθμήσιμο σύνολο** καλείται ένα σύνολο $A \subseteq \omega$ αν και μόνον αν το κατηγορημα $x \in A$ είναι αναδρομικά αριθμήσιμο.

Όσον αφορά μια θεωρία T , έχουμε τους παρακάτω ορισμούς:

- Μία θεωρία T καλείται **(αναδρομικά) αξιωματοποιημένη** αν το σύνολο Ax_T είναι αναδρομικό.
- Μία θεωρία T καλείται **(αναδρομικά) αξιωματοποιήσιμη** αν υπάρχει μία αξιωματοποιημένη θεωρία T' τέτοια ώστε για κάθε τύπο ϕ να ισχύει ότι: $\vdash_T \phi \Leftrightarrow \vdash_{T'} \phi$, δηλαδή οτιδήποτε αποδεικνύεται στην T αν και μόνον εάν αποδεικνύεται στην T' .

Σημ.: Οι παραπάνω ορισμοί είναι που μας παρέχουν τη δυνατότητα να εκφράσουμε τα αποτελέσματα μη-πληρότητας και άλλα συναφή τους μέσω μίας νέας, αλλά ισοδύναμης με τις πρωτότυπες διατυπώσεις, οπτικής γωνίας. Αυτή η θεώρηση θα αναλυθεί περισσότερο μετά το πέρας της αποδεικτικής διαδικασίας, οπότε θα έχουμε όλα τα αποτελέσματα στη διάθεση μας.

Θεώρημα 3.2.1: Αν η T είναι μία αξιωματοποιημένη θεωρία, τότε το σύνολο Thm_T είναι αναδρομικά αριθμήσιμο.

Απόδειξη: Ισχύει ότι $\text{Thm}_T(a) \leftrightarrow \exists b \text{Pf}_T(a, b)$ και ότι αν η T αξιωματοποιημένη, τότε το Pf_T είναι αναδρομικό. Από τον ορισμό του αναδρομικά αριθμήσιμου συνόλου έπεται αμέσως το αποτέλεσμα.

□

Ορίζουμε επίσης μία έννοια συναφή αυτής της πληρότητας, την έννοια της αποκρισιμότητας μιας θεωρίας.

Ορισμός: Μία θεωρία είναι **αποκρίσιμη** αν το σύνολο Thm_T είναι αναδρομικό.

Για να διαπιστώσουμε ότι ο παραπάνω ορισμός εκφράζει καλά την έννοια της αποκρισιμότητας, παρατηρούμε ότι: Υιοθετώντας την θέση του Church [Church Thesis – βλ. επίσης § 4.1 του παρόντος μέρους για περισσότερες λεπτομέρειες], της οποίας μία διατύπωση λέει ότι «ένα κατηγορημα είναι υπολογιστό αν και μόνον αν είναι αναδρομικό» έχουμε ότι: Μία αποκρίσιμη θεωρία έχει αναδρομικό σύνολο θεωρημάτων, άρα το κατηγορημα $P(x) \leftrightarrow x \in \text{Thm}_T$ είναι αναδρομικό άρα και υπολογιστό. Αυτό σημαίνει ότι *μία θεωρία είναι αποκρίσιμη αν και μόνον αν υπάρχει αλγόριθμος που απαντά σε πεπερασμένο χρόνο αν μία πρόταση της θεωρίας είναι θεώρημά της*.

Ο Gödel έδειξε ότι για κάθε συνεπή επέκταση της αριθμητικής του Robinson, δεν υπάρχει τέτοιος αλγόριθμος. Αυτό το αποτέλεσμα αναφέρεται σαν ‘Θεώρημα του Church’ στην απόδειξη του παραρτήματος και η απόδειξή του εμπεριέχει μία συλλογιστική ανάλογη αυτής του θεωρήματος μη-πληρότητας, οπότε θα την παρουσιάσουμε αναλυτικά.

Θεώρημα Church: Κάθε συνεπής επέκταση της θεωρίας S_0 είναι μη-αποκρίσιμη.

Απόδειξη: Από τον ορισμό της αποκρίσιμης θεωρίας, αρκεί να δειχτεί ότι για μία τυχαία συνεπή θεωρία $T \supseteq S_0$, το Thm_T δεν είναι αναδρομικό.

Ορίζουμε ένα κατηγορημα P ως εξής: $P(a) \leftrightarrow \neg \text{Thm}_T(\text{Sub}(a, \lceil x_1 \rceil, \text{Num}(a)))$, όπου x_1 ελεύθερη μεταβλητή της γλώσσας $\mathcal{L}_A$, σταθεροποιημένη.

[Σημ: Η συνάρτηση $\text{Num}:\mathbb{N}\rightarrow\omega$, είναι μία αναδρομική συνάρτηση η οποία, δοθέντος ενός αριθμού a , επιστρέφει τον Gödel-αριθμό του αριθμητικού $\bar{a}$. Δηλαδή, ισχύει ότι $\text{Num}(a)=\lceil \bar{a} \rceil$ για κάθε αριθμό a .]

Συνέχεια απόδειξης: Από τον ορισμό του κατηγορήματος P , προκύπτει ότι αυτό ισχύει αν και μόνον αν δεν ανήκει στο Thm_T , ο Gödel-αριθμός ενός τύπου ο οποίος προκύπτει από την αντικατάσταση της x_1 από το αριθμητικό $\bar{a}$, στο τύπο με Gödel-αριθμό a .

Προφανώς, το κατηγορήμα P είναι αναδρομικό αν το Thm_T είναι αναδρομικό, δηλαδή αν η θεωρία T είναι αποκρίσιμη.

Προς εις άτοπο, έστω ότι είναι έτσι, οπότε το P είναι αναδρομικό άρα και αναπαραστάσιμο στην T από έναν τύπο με μία ελεύθερη μεταβλητή, έστω $\phi(x_1)$ (η χρήση της x_1 είναι απόλυτα θεμιτή).

Έστω επίσης k να είναι ο Gödel-αριθμός του τύπου $\phi(x_1)$, ήτοι $k=\lceil \phi(x_1) \rceil$.

Αν τώρα πάρουμε σαν όρισμα του P το k , δηλαδή τον ίδιο τον Gödel-αριθμό που περιγράφει τον τύπο που το αναπαριστά, έχουμε:

$P(k)=P(\lceil \phi(x_1) \rceil)\Leftrightarrow \neg\text{Thm}_T(\lceil \phi(\bar{k}) \rceil)\Leftrightarrow \not\vdash_T \phi(\bar{k})$, όπως προκύπτει αμέσως από τα σχόλια πάνω στον ορισμό του P που έγιναν πιο πάνω. Εν προκειμένω, $a\equiv k$ και ο τύπος με Gödel-αριθμό a είναι ο $\phi(x_1)$.

Όμως, από τον ορισμό του αναπαραστάσιμου κατηγορήματος (βλ. § 2.3), για το P θα πρέπει να ισχύουν τα παρακάτω:

$$P(k)\Rightarrow \vdash_T \phi(\bar{k}) \text{ και } \neg P(k)\Rightarrow \vdash_T \neg \phi(\bar{k}).$$

Έτσι, αν ισχύει το $P(k)$ έχουμε μια αντίφαση, αφού προηγουμένως δείξαμε ότι $P(k)\Leftrightarrow \not\vdash_T \phi(\bar{k})$, ενώ αν ισχύει το $\neg P(k)$, τότε η T μπορεί να αποδείξει τον τύπο $\neg \phi(\bar{k})$, αλλά και τον $\phi(\bar{k})$, εφόσον

$$P(k)\Leftrightarrow \not\vdash_T \phi(\bar{k})\Rightarrow \neg P(k)\Leftrightarrow \neg \not\vdash_T \phi(\bar{k})\Rightarrow \neg P(k)\Leftrightarrow \vdash_T \phi(\bar{k}),$$

άρα είναι μη-συνεπής, επίσης άτοπο.

Έτσι καταλήγουμε ότι το Thm_T δε μπορεί να είναι αναδρομικό, άρα κάθε συνεπής επέκταση της θεωρίας S_0 είναι μη-αποκρίσιμη.

□

Σχόλια: Η βασική ιδέα της παραπάνω απόδειξης και αυτή που θα μας ακολουθήσει στην απόδειξη των θεωρημάτων μη πληρότητας είναι η αυτο-αναφορά που χρησιμοποιήθηκε για την εύρεση της αντίφασης. Το γεγονός ότι χρησιμοποιήσαμε σαν όρισμα του κατηγορήματος P το ίδιο (σχεδόν) το κατηγορήμα δεν περιλαμβάνει καμία λανθασμένη φαυλοκυκλική συλλογιστική. Όπως επισημαίνει και ο ίδιος ο Gödel σε μία υποσημείωση του πρωτότυπου άρθρου του (όχι ακριβώς για το ίδιο αποτέλεσμα βέβαια), το ότι χρησιμοποιούμε τον Gödel-αριθμό του P σαν όρισμα του ίδιου του P για να βγει η αντίφαση, γίνεται σχεδόν κατά τύχη. Θα μπορούσαμε δηλαδή τυχαία να διαπιστώσουμε ότι η χρησιμοποίηση αυτού του Gödel-αριθμού οδηγεί σε αντίφαση και εκ των υστέρων να διαπιστώναμε ότι αυτός αναφέρεται όντως στον τύπο του κατηγορήματος P . Αυτή η ιδέα της αυτο-αναφοράς εξηγείται και αποδεικνύεται παρακάτω ως ένα λήμμα το οποίο δεν ήταν σαφώς γραμμένο στην αρχική απόδειξη αλλά σαν αποτέλεσμα δικαιολογεί απολύτως τη χρήση της η οποία αποδείχτηκε τελικά η ‘αχίλλειος πτέρνα’ της θεωρίας της αριθμητικής.

Θα δούμε τώρα κάποια επιπλέον ενδιάμεσα αποτελέσματα και ανάμεσά τους την αναλυτική διατύπωση του λήμματος της αυτο-αναφοράς.

Λήμμα 3.2.2: Ένα κατηγορήμα P είναι αναδρομικό αν και μόνον αν τα P και $\neg P$ είναι αναδρομικά αριθμήσιμα.

Θεώρημα 3.2.3: Αν μία θεωρία T είναι αξιωματοποιημένη και πλήρης τότε είναι αποκρίσιμη.

Απόδειξη: Με χρήση του λήμματος 3.2.2 και του δεδομένου της πληρότητας.

Παρατήρηση: Από το παραπάνω θεώρημα φαίνεται ότι το θεώρημα του Church είναι άμεση συνέπεια του αποτελέσματος του Gödel για τη μη πληρότητα της αριθμητικής. Αυτό συμβαίνει διότι ισχύει το αντιθετο-αντίστροφο του θεωρήματος 3.2.3, ήτοι ότι αν μία αξιωματοποιημένη θεωρία δεν είναι πλήρης, τότε δεν είναι ούτε αποκρίσιμη, είναι δηλαδή μη-αποκρίσιμη. Ο Gödel όμως, όπως θα διαπιστώσουμε και παρακάτω, έδειξε με το θεώρημα του τη μη-πληρότητα κάθε συνεπούς επέκτασης της S_0 , οπότε άμεσα έπεται το αποτέλεσμα του Church.

Θεώρημα 3.2.4: (Λήμμα της αυτο-αναφοράς) Έστω $\phi(x)$ οποιοσδήποτε τύπος με μία ελεύθερη μεταβλητή. Τότε μπορούμε να κατασκευάσουμε μία πρόταση Ψ τέτοια ώστε $\vdash_{S_0} \Psi \leftrightarrow \phi(\ulcorner \Psi \urcorner)$.

Απόδειξη: Η απόδειξη χρησιμοποιεί την ίδια τεχνική με αυτή του θεωρήματος του Church, ήτοι ορίζεται και πάλι ένα αναδρομικό κατηγορημα, αναπαρίσταται από έναν τύπο και δημιουργείται μία αυτο-αναφορική πρόταση που μας δίνει το ζητούμενο. Για περισσότερες λεπτομέρειες, ο αναγνώστης παραπέμπεται στις σελίδες xxvii και xxviii του Παραρτήματος.

□

Η σπουδαιότητα του παραπάνω λήμματος έγκειται στο ότι ισχύει για *οποιοδήποτε* τύπο και δίνει τη δυνατότητα κατασκευής μίας πρότασης με μία πολύ συγκεκριμένη ιδιότητα η οποία ιδιότητα είναι *αποδείξιμη* από τη θεωρία S_0 . Στην ουσία, η κατασκευή αυτή μας δίνει τη δυνατότητα να εναλλάξουμε την πρόταση Ψ με τον τύπο που προκύπτει αν η ελεύθερη μεταβλητή του $\phi(x)$ πάρει την τιμή του Gödel-αριθμού της πρότασης Ψ . Έτσι, όπως θα δούμε αμέσως παρακάτω, με την κατάλληλη επιλογή του τύπου $\phi(x)$, παίρνουμε τη μη-αποκρίσιμη πρόταση που ζητάμε, την Ψ , για την ύπαρξη της οποίας καλυπτόμαστε από το λήμμα της αυτο-αναφοράς.

1^ο Θεώρημα μη-πληρότητας του Gödel: Κάθε συνεπής αξιωματοποιημένη επέκταση του αξιωματικού συστήματος S_0 είναι μη-πλήρης.

Απόδειξη: Η απόδειξη συνίσταται στην κατασκευή μίας (συγκεκριμένης) μη-αποκρίσιμης πρότασης για κάθε τέτοια επέκταση T .

Εφόσον η T είναι αξιωματοποιημένη, το κατηγορημα $\text{Pft}(x, y)$ είναι αναδρομικό και άρα αναπαραστάσιμο στην T από έναν τύπο, έστω $\Theta(x, y)$, ο οποίος εκφράζει το γεγονός ότι υπάρχει μία απόδειξη για τον τύπο με Gödel-αριθμό y , απόδειξη της οποίας ο 'Gödel-αριθμός απόδειξης' είναι ο x .

Έστω τώρα ο τύπος $\phi(x)$ να είναι ο $\forall y \neg \Theta(x, y)$ που σημαίνει ότι το x δεν είναι ο Gödel-αριθμός απόδειξης κανενός αποδείξιμου τύπου (θεωρήματος) της θεωρίας T . Εφαρμόζοντας το λήμμα αυτό-αναφοράς για τον τύπο $\Theta(x, y)$, μπορώ να κατασκευάσω μία πρόταση Ψ , τέτοια ώστε:

$$\vdash_T \Psi \leftrightarrow \forall y \neg \Theta(\ulcorner \Psi \urcorner, y) \quad (*)$$

Αυτή η μετα-πρόταση (που αποδεικνύεται από την T , άρα είναι αληθής) λέει για την πρόταση Ψ ότι αυτή είναι αληθής αν και μόνον αν δεν είναι αποδείξιμη (αφού το 2^ο μέλος της λέει ότι ο Gödel-αριθμός της Ψ δεν είναι Gödel-αριθμός απόδειξης).

Ισχυρισμός: Η πρόταση Ψ είναι μη-αποκρίσιμη πρόταση της T .

Απόδειξη ισχυρισμού:

- Έστω ότι η T αποδεικνύει την Ψ . Τότε έχω:

$$\begin{aligned} \vdash_T \Psi &\Rightarrow \vdash_T \forall y \neg \Theta(\ulcorner \Psi \urcorner, y) && \text{απο } (*) \\ &\Rightarrow \vdash_T \neg \Theta(\ulcorner \Psi \urcorner, \bar{a}) && \forall a \in \omega \\ &\Rightarrow \neg \text{Pft}(\ulcorner \Psi \urcorner, a) && \forall a \in \omega \\ &\Rightarrow \not\vdash_T \Psi && \text{'Ατοπο.} \end{aligned}$$

- Έστω ότι η T αποδεικνύει την άρνηση της Ψ . Τότε:

$$\begin{aligned} \vdash_T \neg \Psi &\Rightarrow \vdash_T \exists y \Theta(\ulcorner \Psi \urcorner, y) && \text{απο } (*) \\ &\Rightarrow \mathbb{N} \models \exists y \Theta(\ulcorner \Psi \urcorner, y) && \text{απο θ.Ορθότητας} \\ &\Rightarrow \mathbb{N} \models \Theta(\ulcorner \Psi \urcorner, \bar{a}) && \text{για καποιο } a \in \omega \\ &\Rightarrow \not\vdash_T \neg \Theta(\ulcorner \Psi \urcorner, \bar{a}) && \text{απο θ.Πληροτητας} \\ &\Rightarrow \text{Pft}(\ulcorner \Psi \urcorner, a) && \text{αφου ο } \Theta \text{ αναπαριστα το Pft} \\ &\Rightarrow \vdash_T \Psi && \text{'Ατοπο} \end{aligned}$$

Έτσι καταλήξαμε στο ότι η T δε μπορεί να αποδείξει ούτε την Ψ αλλά ούτε και την άρνησή της, οπότε αυτή η πρόταση είναι όντως μη-αποκρίσιμη στην T .

Από τον ορισμό του πλήρους αξιωματικού συστήματος που έχει δοθεί, βγαίνει απευθείας το συμπέρασμα ότι όντως το αξιωματικό σύστημα της T είναι μη-πλήρες, ή ότι η θεωρία T είναι μη-πλήρης (προφανώς οι δύο διατυπώσεις είναι ισοδύναμες).

□

Σχόλια:

- Αυτή φυσικά, δεν είναι η πρωτότυπη διατύπωση και απόδειξη του θεωρήματος, αλλά κρατάει ανέπαφη την κεντρική ιδέα, η οποία είναι ακριβώς η κατασκευή της μη-αποκρίσιμης πρότασης η οποία λέει για τον εαυτό της ότι είναι μη-αποδείξιμη.
- Ο Gödel απέδειξε το θεώρημα για το αξιωματικό σύστημα του Peano (S), το οποίο όμως όντας μία επέκταση αυτού του Robinson (S_0), συγκαταλέγεται στο παραπάνω αποτέλεσμα. Το γεγονός ότι το θεώρημα μη-πληρότητας ισχύει όσο και αν επεκταθεί αξιωματικά το S_0 είχε επίσης επισημανθεί στο αρχικό άρθρο.
- Τέλος, να σημειωθεί ότι αρχικά, η απόδειξη του ότι το S αποδεικνύει την άρνηση της Ψ , είχε δοθεί με δεδομένη την ω -συνέπεια⁽¹⁾ του S, μία συνθήκη ισχυρότερη από τη συνέπεια. Ο J.B. Rosser ήταν αυτός ο οποίος, το 1936, απέδειξε ότι η ω -συνέπεια μπορεί να αντικατασταθεί από τη συνέπεια, οπότε και το θεώρημα πήρε τη σύγχρονη μορφή του.

Μία πολύ σημαντική παρατήρηση: Όπως είχε επισημάνει ο Gödel στο άρθρο του, αν και η πρόταση Ψ (από εδώ και στο εξής θα την αποκαλούμε Gödel-πρόταση) είναι μη-αποκρίσιμη στα πλαίσια της αριθμητικής, παραμένει αληθής με βάση μεταμαθηματικούς συλλογισμούς. Αυτό συμβαίνει διότι η μετα-πρόταση (*) της απόδειξης, λέει για την Gödel-πρόταση ότι αυτή είναι αληθής αν και μόνον αν είναι μη-αποδείξιμη. Όντας όμως η Gödel-πρόταση μη-αποκρίσιμη, είναι όντως και μη-αποδείξιμη, άρα η Gödel-πρόταση είναι αληθής, απλά *δε μπορούμε να διαπιστώσουμε την αλήθεια της στα πλαίσια της θεωρίας από την οποία εκπορεύτηκε*. Ένας άλλος τρόπος να διαπιστώσουμε την αλήθεια της Gödel-πρότασης είναι με την εις άτοπο απαγωγή. Έτσι, αν υποθεθεί ψευδής, από τη σχέση (*) θα πηγάζει ότι αυτή είναι επίσης και αποδείξιμη, αλλά το να αποδεικνύεται σε μία θεωρία μία ψευδής πρόταση, είναι φυσικά αδιανόητο.

Αυτή η πολύ σημαντική διαπίστωση αποτελεί αναπόσπαστο κομμάτι πολλών επιχειρημάτων από αυτά που θα αναλυθούν στο 3^ο και τελευταίο μέρος της εργασίας και θα την αναφέρουμε ξανά τότε, διότι δεν είναι καθόλου τετριμμένο αυτό που διαπιστώθηκε και πολλές φορές δημιουργείται σύγχυση ως προς το περιεχόμενό του.

(1): Για τον ορισμό της ω -συνέπειας, βλ. § 4.3 του ίδιου μέρους.

3.3 Το 2^ο θεώρημα μη-πληρότητας

Έχει ήδη αναφερθεί ότι το 2^ο θεώρημα μη-πληρότητας αποτελεί ένα πόρισμα του 1^{ου} και είχε κεφαλαιώδη αντίκτυπο στο Πρόγραμμα του Hilbert. Πράγματι, η ουσία του θεωρήματος έγκειται στην εφαρμογή του 1^{ου} θεωρήματος μη-πληρότητας σε μία πρόταση που εμπεριέχει έναν ορισμό ισοδύναμο της συνέπειας μιας θεωρίας (συγκεκριμένα εδώ, της αριθμητικής του Peano που βασίζεται στο αξιωματικό σύστημα S). Αυτή η πρόταση βγαίνει μη-αποδείξιμη στη θεωρία αυτή, και κατά συνέπεια η συνέπεια της θεωρίας προκύπτει επίσης μη-αποδείξιμη στα πλαίσια της θεωρίας αυτής. Συγκεκριμένα για την αριθμητική, έχουμε το:

2^ο Θεώρημα μη-πληρότητας του Gödel: Αν η S είναι συνεπής θεωρία, και Cons μία πρόταση που ισοδυναμεί με τη συνέπειά της, τότε $\not\vdash_s \text{Cons}$, δηλαδή η συνέπεια της S δεν αποδεικνύεται από την θεωρία S.

Απόδειξη: Θα εκφράσουμε καταρχήν την έννοια της συνέπειας σαν την ιδιότητα της S να *μη μπορεί να αποδείξει μία αντίφαση*, και συγκεκριμένα την αντίφαση $0=1$. [Σημ: Προφανώς κάθε άλλη αντίφαση της αριθμητικής μπορεί να προκύψει από αυτή]. Χρησιμοποιούμε για αυτό το σκοπό το αναδρομικό κατηγορήμα που εκφράζει την αποδειξιμότητα στην S, δηλαδή το $\text{Pfs}(x, y)$, και την αναπαράστασή του στην θεωρία μας μέσω του τύπου $\Theta_s(y, x)$.

Τότε η πρόταση που εκφράζει τη συνέπεια θα είναι η ακόλουθη:

$$\text{Cons} = \forall y \neg \Theta_s(\overline{0=1}, y).$$

Έστω τώρα Ψ_s η μη-αποκρίσιμη πρόταση του 1^{ου} θεωρήματος μη πληρότητας. Δείξαμε ότι, δεδομένης της συνέπειας της S, αυτή η πρόταση δεν αποδεικνύεται από την S, δηλαδή $\text{Cons} \not\vdash_s \Psi_s$, αλλά και ότι η S αποδεικνύει, δεδομένης της συνέπειάς της, ότι η πρόταση Ψ_s δεν έχει απόδειξη στην S, δηλαδή $\text{Cons} \vdash_s \forall y \neg \Theta_s(\overline{\Psi_s}, y)$. Αντικαθιστώντας σε αυτή τη σχέση την έκφραση της συνέπειας, παίρνουμε:

$$\vdash_s \forall y \neg \Theta_s(\overline{0=1}, y) \rightarrow \forall y \neg \Theta_s(\overline{\Psi_s}, y) \quad (**)$$

Αν υποθέσουμε το ζητούμενο, δηλαδή ότι $\vdash_s \text{Cons}$, τότε από τη σχέση (**) και τον κανόνα MP, προκύπτει ότι $\vdash_s \forall y \neg \Theta(\overline{\lceil \Psi_s \rceil}, y)$.

Όμως από κατασκευή της Ψ_s έχω ότι $\vdash_s \Psi_s \leftrightarrow \forall y \neg \Theta(\overline{\lceil \Psi_s \rceil}, y)$, άρα και πάλι από MP, βγαίνει ότι $\vdash_s \Psi_s$, πράγμα που είναι άτοπο λόγω του 1^{ου} θεωρήματος μη-πληρότητας.

Άρα η υπόθεσή μας είναι λάθος, άρα $\not\vdash_s \text{Cons}$, δηλαδή η αριθμητική του Peano δεν αποδεικνύει τη συνέπειά της.

□

Σχόλιο: Παρατηρούμε ότι η ευρέως διαδεδομένη ερμηνεία της απόδειξης του 2^{ου} θεωρήματος μη πληρότητας, ότι δηλαδή χρησιμοποιούμε στη θέση της Gödel-πρότασης την πρόταση για τη συνέπεια και έτσι βγαίνει το αποτέλεσμα, δεν είναι απολύτως αληθές. Το αποτέλεσμα του 1^{ου} θεωρήματος μη-πληρότητας χρησιμοποιείται όντως για την απόδειξη, αλλά η συνεπαγωγή:

Συνέπεια θεωρίας $\Rightarrow$ Ύπαρξη μη-αποκρίσιμης πρότασης στη θεωρία αυτή είναι που χρησιμοποιείται τελικά, και όχι κάποια πρόταση που αντικαθιστά την Gödel-πρόταση, η οποία παραμένει η ίδια.

Παρά το αποτέλεσμα που αποδείξαμε παραπάνω, ο Gentzen το 1936 κατάφερε να αποδείξει τη συνέπεια της αριθμητικής του Peano χρησιμοποιώντας όμως μία μέθοδο που δεν καλύπτεται από την πρωτοβάθμια αριθμητική του Peano. [Σημ: έχουμε αναφερθεί σε αυτό το αποτέλεσμα ξανά στην ιστορική αναδρομή του 1^{ου} μέρους]. Συγκεκριμένα, χρησιμοποίησε αποδείξεις των οποίων τα μήκη είναι *άπειροι* αριθμοί και όχι πεπερασμένοι. Η μέθοδος αυτή καταδεικνύει αυτό που γράφτηκε στο τέλος της προηγούμενης παραγράφου, ήτοι το γεγονός ότι η αλήθεια των μετα-προτάσεων που δεν αποδεικνύονται από μία θεωρία είναι δυνατόν να φανεί, αλλά μόνο μέσω μιας μεθόδου που ξεπερνά τα όρια της συγκεκριμένης θεωρίας. Βέβαια έτσι ίσως προκύπτουν προβλήματα αξιοπιστίας αυτών των μεθόδων, ειδικά αν μιλούμε για τη ‘διαίσθηση’ (τους μετα-μαθηματικούς συλλογισμούς που αναφέραμε μετά το 1^ο θεώρημα μη-πληρότητας) ή για ‘άπειρες’ αποδείξεις (όπως εδώ), διότι αυτές οι μέθοδοι μοιάζουν να βρίσκονται πιο κοντά στη φιλοσοφία παρά στα μαθηματικά.

4. Συνέπειες, εφαρμογές των θεωρημάτων μη-πληρότητας και άλλα σχετικά αποτελέσματα

4.1 Συνέπειες μαθηματικές και μη

Όπως έχει ήδη αναφερθεί στην ιστορική αναδρομή του 1^{ου} μέρους, η παρέμβαση του Gödel επέφερε μία σημαντική αλλαγή πλεύσης και έναν επαναπροσδιορισμό των στόχων που είχαν τεθεί για τα κλασικά μαθηματικά. Μεγάλο ρόλο σε αυτή την αλλαγή έπαιξε η δυνατότητα επέκτασης του αποτελέσματος της μη-πληρότητας, όχι μόνο σε κάθε επέκταση της αριθμητικής του Peano, αλλά και στην θεωρία συνόλων των Zermelo-Fraenkel, όπως και σε κάθε άλλη θεωρία που περιλαμβάνει τουλάχιστον την θεωρία αριθμών. Έτσι, το όραμα για καθαρότητα των μεθόδων που θα οδηγούσαν αποδεδειγμένα σε πλήρεις και συνεπείς θεωρίες που απαρτίζουν τα θεμέλια των μαθηματικών φαίνεται ανεκπλήρωτο. Όμως, αν και αυτή η εξέλιξη ήταν δυσάρεστη για μεγάλο κομμάτι του μαθηματικού κόσμου, δεν είχε μόνο αρνητικά αποτελέσματα.

Στην προσπάθεια να γίνουν κατανοητά και να ενσωματωθούν με κάποιο τρόπο αυτά τα αποτελέσματα στα ήδη υπάρχοντα μαθηματικά, δημιουργήθηκαν νέες θεωρίες και προέκυψαν εντελώς νέα ερωτήματα για τις μαθηματικές θεωρίες. Έτσι, ερωτήματα όπως το που μας οδηγεί η επιλογή αξιωματικής επέκτασης πλήρων, αλλά αδύναμων, αξιωματικών συστημάτων, όπως η αριθμητική που περιλαμβάνει μόνο μία εκ των δύο θεμελιωδών πράξεων (πρόσθεση – πολλαπλασιασμός), επέκτειναν την κατανόηση και εμβάθυνση πάνω σε αυτές τις θεωρίες. Επίσης, δημιουργήθηκε η ανάγκη αναζήτησης νέων αξιωμάτων, ιδιαίτερα για τη θεωρία συνόλων η οποία παίζει μεγάλο ρόλο στα σύγχρονα μαθηματικά, μία αναζήτηση που άσχετα αν αποδειχθεί μάταιη ή όχι, προσφέρει έδαφος για νέους γόνιμους προβληματισμούς.

Πολύ σημαντική προσθήκη στη μαθηματική συλλογιστική ήταν και η προσθήκη της έννοιας της αποκρισιμότητας. Διότι, η πληρότητα φαίνεται να ισχύει σε συστήματα που περιγράφουν ένα μέρος μαθηματικών θεωριών, όπως η αριθμητική του Robinson (RA), η οποία, καθώς στερείται του αξιώματος της μαθηματικής επαγωγής, περιγράφει μόνο πολύ τετριμμένες αλήθειες της θεωρίας αριθμών. Από την άλλη όμως, η αποκρισιμότητα φαίνεται ότι είναι πιο δύσκολο να ισχύει, ακόμα και για πολύ βασικές θεωρίες. Στο κεφάλαιο του παραρτήματος με τίτλο ‘Εφαρμογές’

(σελ. xxxi), αποδεικνύονται μία σειρά τέτοιων αποτελεσμάτων που αφορούν μεταξύ άλλων τη δομή των φυσικών αριθμών, την $RA^{(1)}$, αλλά και τον (γενικό) κατηγορηματικό λογισμό⁽²⁾. Με το αποτέλεσμα της μη-αποκρισιμότητας των παραπάνω, ένα βασικό ερώτημα είναι πλέον αν μία πρόταση (ή μία θεωρία) είναι αποκρίσιμη ή όχι. Με την προσπάθεια απάντησης αυτού του ερωτήματος και άλλων συναφών, δημιουργήθηκε η Θεωρία Υπολογιστότητας⁽³⁾ που σύντομα περιγράφεται παρακάτω, για λόγους που θα γίνουν αμέσως προφανείς.

Θεωρία Υπολογιστότητας [computability theory]

Η θεωρία αυτή γεννήθηκε από την *ανάγκη μαθηματικής περιγραφής της έννοιας του αλγορίθμου*, έννοια που συνεπάγεται την έννοια της υπολογιστότητας, αφού υπολογιστό πρόβλημα καλούμε ένα πρόβλημα το οποίο επιλύεται μέσω μίας μηχανιστικής διαδικασίας (αλγορίθμου). Έχουμε ήδη ορίσει την έννοια της ‘υπολογιστής συνάρτησης’ (βλ. § 2.2.1) και αντίστοιχα, ‘υπολογιστό κατηγορημα’, θα είναι ένα κατηγορημα για το οποίο υπάρχει αλγόριθμος που να αποκρίνεται [decides] ως προς το αν αυτό ισχύει ή όχι. Υπήρξαν διάφοροι επιστήμονες (Turing, Church, Kleene,...) που ασχολήθηκαν με το πρόβλημα αποσαφήνισης των εννοιών ‘αλγόριθμος’, ‘υπολογιστός’, ‘αποκρίσιμο πρόβλημα απόφασης’ κτλ., και κατέληξαν σε διαφορετικά μοντέλα περιγραφής τους. Έτσι προέκυψαν οι ‘μηχανές Turing [Turing Machines]’, ο λ-λογισμός’ και άλλα μοντέλα, που όμως προέκυψαν όλα ισοδύναμα μεταξύ τους. Αυτό είναι το αποτέλεσμα της θέσης του Church.

Θέση του Church [Church Thesis]: Όλα τα σύγχρονα ή και μελλοντικά μοντέλα περιγραφής της έννοιας ‘υπολογιστός’ είναι μεταξύ τους μηχανιστικά ισοδύναμα, δηλαδή η ύπαρξη αλγορίθμου που επιλύει ένα πρόβλημα σε ένα μοντέλο, συνεπάγεται την ύπαρξη δυνατότητας μηχανιστικής κατασκευής αλγορίθμου που επιλύει το ίδιο πρόβλημα σε ένα άλλο μοντέλο.

(1): Η αποκρισιμότητα της RA δεν αποδεικνύεται αυτή καθ’ εαυτή αλλά προκύπτει με προφανή τρόπο από τα υπόλοιπα αποτελέσματα.

(2): Το γεγονός αυτό το είχαμε επισημάνει χωρίς απόδειξη στην αρχή της παραγράφου 2.1.4.

(3): Διαδεδομένη είναι γενικά και η χρήση του όρου ‘υπολογίσιμος’ αντί για υπολογιστός.

Σε όλη την ανάλυση του μέρους στο οποίο βρισκόμαστε συναντήσαμε αυτές τις έννοιες πολλές φορές, αρχίζοντας από το γεγονός ότι η προτασιακή λογική είναι μία αποκρίσιμη θεωρία και καταλήγοντας μέσω των θεωρημάτων 2.3.1 και 2.3.2 και την θέσης του Church⁽⁴⁾, μέχρι να διαπιστώσουμε την ισοδυναμία των εννοιών ‘υπολογιστή’, ‘αναδρομική’ και ‘αναπαραστάσιμη’ συνάρτηση. Στα πλαίσια βέβαια της θεωρίας υπολογιστότητας, υπάρχουν και μη-αναδρομικές συναρτήσεις, αλλά και έννοιες που αδυνατούν να περιγραφούν από την έννοια της υπολογιστότητας, όπως η έννοια της πρωταρχικής αναδρομής. Υπό αυτό το πρίσμα, η προσπάθεια αναγωγής όλων των συναρτήσεων στην κλάση Πρωτ.Αν., κάτι που μοιάζει με έναν από τους στόχους του προγράμματος του Hilbert, είναι καταδικασμένη εκ των προτέρων, εφόσον η κλάση αυτή αδυνατεί να περιγράψει την έννοια της υπολογιστότητας.

Υπάρχουν, στα πλαίσια αυτής της θεωρίας, πολλά ακόμη ενδιαφέροντα αποτελέσματα καθώς και μερικά ανάλογα με το θεώρημα μη-πληρότητας, όπως η ύπαρξη μη-υπολογιστών συναρτήσεων και αντίστοιχα η ύπαρξη μη-αποκρίσιμων υπολογιστικών προβλημάτων (όπως το Halting Problem⁽⁵⁾) που φανερά ανάλογη με την ύπαρξη μη-αποκρίσιμων προτάσεων στην αριθμητική. Ορίζεται μάλιστα μία ολόκληρη κλάση από τέτοια μη επιλύσιμα προβλήματα, ενώ ισχύει και το παρακάτω πολύ ενδιαφέρον αποτέλεσμα:

Θεώρημα Rice: Κάθε μη τετριμμένη ιδιότητα υπολογιστών συναρτήσεων αντιπροσωπεύει ένα μη-επιλύσιμο πρόβλημα.

[Σημ: Η έννοια ‘μη τετριμμένη ιδιότητα’ σημαίνει ουσιαστικά μετα-ιδιότητα, δηλαδή μία ιδιότητα των υπολογιστών συναρτήσεων που αναφέρεται στις υπολογιστές συναρτήσεις.]

Κλείνοντας αυτή την πολύ μικρή εισαγωγή στη θεωρία υπολογιστότητας, είναι απαραίτητο να αναφερθούμε σε ένα ιδιαίτερο υπολογιστικό μοντέλο, τη μηχανή Turing [TM, από εδώ και στο εξής]. Αυτό το μοντέλο χρησιμοποιείται ευρέως στην επιχειρηματολογία περί των δυνατοτήτων της τεχνητής νοημοσύνης, διότι αποτελεί τον πρόγονο του σύγχρονου ηλεκτρονικού υπολογιστή, πρέπει όμως να έχουμε στο μυαλό μας ότι, λόγω της CT, έχει τις ίδιες υπολογιστικές δυνατότητες με αυτόν.

(4): ή μάλλον μέσω μίας ισοδύναμης διατύπωσής του, που λέει ότι: ‘Μία συνάρτηση είναι υπολογιστή αν και μόνον αν είναι αναδρομική’. Αυτή απαντάται και ως ‘θέση των Church-Turing’, ειδικότερα αν το μοντέλο υπολογιστότητας είναι μία TM.

(5): Το HP απαντά στο ερώτημα αν ένα συντακτικά ορθό πρόγραμμα θα σταματήσει κάποτε ή θα τρέχει για πάντα.

Συνεχίζεται τώρα η σκιαγράφηση των άμεσων ή έμμεσων συνεπειών των θεωρημάτων του Gödel, με κάποιες συνέπειες στη φιλοσοφία των μαθηματικών.

Όπως έχει γίνει φανερό μέχρι τώρα, στην εποχή πριν τα αποτελέσματα της μη-πληρότητας, ήταν διαδεδομένη η άποψη ότι όλες οι διαισθητικές μέθοδοι που οδηγούν σε μαθηματικά αποτελέσματα μπορούν να αντικατασταθούν από απαγωγές στα πλαίσια ενός αξιωματικού συστήματος. Όπως αναφέρει ο Alan Turing⁽⁶⁾, «Η μαθηματική συλλογιστική ανάγεται στο συνδυασμό δύο παραγόντων: της διαίσθησης [intuition] και της ιδιοφυΐας [ingenuity]. Η διαίσθηση συνίσταται στην αυτόματη κρίση που δεν είναι αποτέλεσμα μίας συνειδητής πορείας σκέψεων». Αυτή η διαίσθηση είναι που έπρεπε να αντικατασταθεί, αλλά τελικά φαίνεται ότι δεν είναι δυνατόν να υπάρξει ένα τυπικό σύστημα λογικής που να το κάνει αυτό πλήρως. Έτσι, η αναγκαιότητα της ύπαρξης της έγινε αποδεκτή από το σύνολο των λογικών, αν και παραμένει η προσπάθεια περιορισμού της σε όσο γίνεται στενότερα πλαίσια.

Από την άλλη όμως, αυτή η εξέλιξη έδωσε ένα ακόμη επιχείρημα σε μία μειοψηφία που διακηρύττει τη γενικότερη ανικανότητα της αφαιρετικής (ή παραγωγικής) μεθόδου [deduction] να περιγράψει τη μαθηματική σκέψη. Κατά τη γνώμη μου, αυτή είναι μία εύλογη ένσταση απέναντι σε μία αντίληψη για τα μαθηματικά που περιγράφει όλη την ουσία τους στο να ψάχνει κανείς να βρει την αλήθεια τους στα πλαίσια ενός αξιωματικού συστήματος. Διότι με αυτόν τον τρόπο, αντικαθιστούμε τη διαίσθηση από την υπομονή που χρειάζεται να έχει κάποιος για να βρει την απόδειξη που χρειάζεται για ένα θεώρημα στο (πλήρες) σύνολο των μηχανιστικά απαριθμημένων δυνατών μαθηματικών αποδείξεων. Προσωπικά, δε βρίσκω τίποτα το ενδιαφέρον σε μία τέτοια προοπτική.

Αν όμως όντως το θεώρημα του Gödel καταδεικνύει σαν ανεπαρκή την αφαιρετική μέθοδο, πρέπει να βρεθεί ο αντικαταστάτης της. Εδώ θα παραθέσουμε μία την πρόταση του Carlo Cellucci, όπως αυτή περιγράφεται στο άρθρο ‘The Scope of Logic: deduction, abduction, analogy’, πρόταση ενδιαφέρουσα και αντιπροσωπευτική πιστεύω άλλων ανάλογων προτάσεων. Ο Cellucci αντιπροτείνει την αναλυτική μέθοδο [analytic method], η οποία λειτουργεί με συνεχείς αναγωγές άλυτων προβλημάτων σε άλλα (πιθανώς άλυτα) προβλήματα. Αυτά τα νέα προβλήματα μπορεί να λύνονται με τη αφαιρετική μέθοδο, μπορεί όμως να απαιτούν με τη σειρά τους επίσης χρήση της αναλυτικής μεθόδου.

(6): βλ. A.M. Turing, ‘Systems of logic based on ordinals’ σελ. 208-9.

Αυτή η διαδικασία, είναι φανερό ότι μας αποδεσμεύει από τη χρήση κλειστών τυπικών συστημάτων, και απαιτεί για να λειτουργήσει ανοιχτά συστήματα [open systems]. Αυτά είναι συστήματα που μπορούν να αντλήσουν πληροφορίες από άλλα συστήματα και το γενικότερο μαθηματικό περιβάλλον. Έτσι, κάθε ανοιχτό σύστημα προσφέρει μία μερική αναπαράσταση μίας συγκεκριμένης περιοχής μαθηματικής γνώσης, και αντλεί τις πληροφορίες στις οποίες υπολείπεται από άλλα συστήματα. Καθότι τα συστήματα αυτά βασίζονται στην αναλυτική μέθοδο, ο Cellucci υποστηρίζει ότι τελικά το θεώρημα του Gödel αναιρεί την θεώρηση των μαθηματικών που στηρίζεται στα κλειστά συστήματα.

Ο ίδιος ο Gödel βέβαια, πιθανότατα θα διαφωνούσε κάθετα με αυτές τις σκέψεις, άλλωστε (βλ. Μέρος I, σελ. 11) ο ίδιος τελειώνει το άρθρο που περιέχει τα θεωρήματα μη-πληρότητας με τρόπο που αφήνει να εννοηθεί ότι συμφωνούσε με τη βασική ιδέα πίσω από το πρόγραμμα του Hilbert. Παρ' όλα αυτά, πιστεύω πως τέτοιες τοποθετήσεις μόνο κακό δεν κάνουν στα μαθηματικά, ίσως μάλιστα να οδηγούν και τους υπέρμαχους των κλειστών τυπικών συστημάτων σε μία προσπάθεια να εξελίξουν με γονιμότερο τρόπο τα αξιωματικά συστήματα. Όπως και να έχει πάντως, οι συνέπειες των θεωρημάτων μη-πληρότητας στη φιλοσοφία των μαθηματικών και στα ίδια τα μαθηματικά είναι πολλές και πολύπλευρες, σκιαγραφήθηκαν εδώ μόνο μερικές που είναι ιδιαίτερα χαρακτηριστικές.

4.2 Ανάλογα αποτελέσματα του θεωρήματος του Gödel

Η ιδέα που οδήγησε στο 1^ο θεώρημα μη-πληρότητας⁽⁷⁾ δεν είναι μία ιδέα χωρίς ιστορία αλλά ούτε χωρίς συνέχεια. Σε αυτή την παράγραφο θα παραθέσουμε μία σειρά αποτελεσμάτων, ακολουθώντας την ιστορική τους σειρά, τα οποία εμπεριέχουν τη συλλογιστική που οδήγησε στο θεώρημα μη-πληρότητας.

- **Το παράδοξο του ψεύτη.** Ο Επιμενίδης από την Κρήτη λέει το εξής: «Οι Κρήτες λένε πάντα ψέματα». Ο ίδιος λέει αλήθεια ή ψέματα;

Αυτό είναι το αρχαιότερο αυτοαναφορικό παράδοξο στην ιστορία. Ο λόγος για τον

(7): Το 2^ο θεώρημα έχει σαν αποτέλεσμα κάτι πολύ συγκεκριμένο και δε θα μας απασχολήσει σε αυτή την παράγραφο.

οποίο αποτελεί παράδοξο είναι λίγο πολύ προφανές. Ο Επιμενίδης, αφού κατάγεται από την Κρήτη, υποκείται ο ίδιος στην αλήθεια των λόγων του. Άρα λέει πάντα ψέματα. Άρα και αυτό που λέει είναι ψέματα, δηλαδή ότι οι Κρήτες λένε πάντα ψέματα. Άρα οι Κρήτες λένε πάντα αλήθεια, άρα και ο ίδιος λέει πάντα αλήθεια. Άρα αληθεύει ότι οι Κρήτες λένε πάντα ψέματα κ.ο.κ. μέχρι το άπειρο.

Μία σύγχρονη μορφή του παραπάνω παράδοξου που μοιάζει περισσότερο με το θεώρημα του Gödel είναι η παρακάτω:

- **Το παράδοξο της ψεύδους πρότασης.** Η ‘αλήθεια’ για τις προτάσεις της ελληνικής γλώσσας δεν είναι ορίσιμη στην ελληνική γλώσσα.

Απόδειξη: Έστω ότι είναι. Τότε είναι ορίσιμη και η αντίθετη έννοια του ‘ψεύδους’. Έστω Ψ να είναι η πρόταση: «Αυτή η πρόταση είναι ψευδής». Αφού η φράση ‘αυτή είναι πρόταση’ αναφέρεται στην Ψ , έχουμε ότι:

$$\Psi \Leftrightarrow \text{«Αυτή η πρόταση είναι ψευδής»} \Leftrightarrow \text{«η } \Psi \text{ είναι ψευδής»} \Leftrightarrow \text{όχι } \Psi.$$

Αντίφαση.

- **Το διαγώνιο επιχείρημα του Cantor.** [Πρόκειται για την μαθηματική εκδοχή του παράδοξου του ψεύτη και είναι το επιχείρημα στο οποίο μπορούμε να ανάγουμε όλα τα άλλα μαθηματικά επιχειρήματα που θα αναφερθούν!]

Το εφαρμόζουμε για να δείξουμε ότι:

Το σύνολο των απείρων δυαδικών ακολουθιών δεν είναι αριθμήσιμο⁽⁸⁾

Απόδειξη: Έστω ότι είναι αριθμήσιμο. Τότε μπορεί να δημιουργηθεί μία (άπειρη) λίστα ακολουθιών, έστω $f_0, f_1, f_2, \dots$, που θα περιέχει όλες τις ακολουθίες που αποτελούνται από 0 και 1.

Έστω τώρα η ακολουθία f που είναι η συμπληρωματική ακολουθία της διαγώνιας ακολουθίας $f_i(i)$ που δημιουργείται από την παραπάνω απαρίθμηση, ήτοι, $f(i) = 1 - f_i(i) \quad \forall i$.

Αυτή η ακολουθία όμως, *διαφέρει από κάθε ακολουθία f_i της λίστας που απαριθμεί τις δυαδικές ακολουθίες σε τουλάχιστον μία θέση* (και συγκεκριμένα στη θέση i). Αλλά η f αποτελεί και αυτή με τη σειρά της μία δυαδική ακολουθία, οπότε βρήκαμε μία δυαδική ακολουθία εκτός της παραπάνω λίστας, άτοπο.

(8): Αριθμήσιμο είναι ένα σύνολο που είναι είτε πεπερασμένο, ή ισοπληθικό με το σύνολο των φυσικών αριθμών.

- **Το παράδοξο του Russell.** [Αναφέρεται στην συνολοθεωρία του Cantor και δείχνει τα όρια των ιδιοτήτων που μπορούν να οριστούν σαν αντικείμενα ενός συνόλου]. Δεν μπορεί για κάθε συνθήκη $P(x)$ να υπάρχει ένα σύνολο $\{x : P(x)\}$ με στοιχεία που την ικανοποιούν.

Απόδειξη: Αν ισχύει κάτι τέτοιο, τότε μπορεί να οριστεί και το παρακάτω σύνολο:

$$R = \{x : x \text{ είναι σύνολο και } x \notin x\}$$

Αμέσως από τον ορισμό αυτού του συνόλου έπεται η αντίφαση, διότι μπορεί να γίνει αυτοαναφορά στο R , οπότε έχουμε: $R \in R \Leftrightarrow R \notin R$. Άτοπο.

- **1^ο Θεώρημα μη-πληρότητας του Gödel (εκλαϊκευμένη εκδοχή).**
Οποιαδήποτε αξιωματοποιημένη ‘επαρκής’⁽⁹⁾ θεωρία είναι μη πλήρης. Συγκεκριμένα, η πρόταση «Αυτή η πρόταση δεν είναι αποδείξιμη» είναι αληθής αλλά όχι αποδείξιμη στη θεωρία αυτή.

Απόδειξη: [Εδώ, αξιωματοποιημένη θεωρία σημαίνει ότι αυτή έχει σύνολο αξιωμάτων που παράγεται μηχανιστικά, ορισμός ισοδύναμος με αυτόν της § 3.2]. Δεδομένης της μηχανιστικής παραγωγής των αξιωμάτων, έστω **αποδείξιμο** (με αυτή τη γραμματοσειρά!) να είναι το σύνολο των αριθμών που κωδικοποιούν προτάσεις αποδείξιμες μέσω αυτών των αξιωμάτων.

Τότε, για κάθε πρόταση Ψ , ισχύει ότι:

$$\lceil \Psi \rceil \in \text{αποδείξιμο} \Leftrightarrow \eta \Psi \text{ είναι αποδείξιμη} \quad (1),$$

όπου $\lceil \Psi \rceil$ δηλώνει τον αριθμό της πρότασης (αντίστοιχο του Gödel-αριθμού).

Εφόσον όμως το σύνολο των αξιωμάτων είναι μηχανιστικά παραγόμενο, τέτοιο θα είναι και το σύνολο των αποδείξεων και άρα και αυτό των θεωρημάτων. Τότε όμως, τέτοιο είναι και το σύνολο **αποδείξιμο**. Αλλά η έννοια ‘μηχανιστικά παραγόμενο’ είναι ανάλογη του ‘υπολογιστού’ και άρα το σύνολο **αποδείξιμο** είναι ορίσιμο σε επαρκείς θεωρίες (αντ. αναπαραστάσιμο στην αρχική απόδειξη).

Έτσι, έστω η πρόταση Ψ να είναι η πρόταση: «Αυτή η πρόταση είναι μη-αποδείξιμη». Τότε από λήμμα αυτό-αναφοράς, η Ψ υπάρχει και είναι η λύση της:

$\Psi \Leftrightarrow \lceil \Psi \rceil \notin \text{αποδείξιμο}$. Και από την (1), προκύπτει ότι:

$$\Psi \Leftrightarrow \lceil \Psi \rceil \notin \text{αποδείξιμο} \Leftrightarrow \eta \Psi \text{ δεν είναι αποδείξιμη} \quad (2)$$

(9): Σαν επαρκής θεωρείται μία θεωρία που είναι ‘αρκετά ισχυρή’ (βλ. Παράρτημα σελ. xxxv) και που υπόκειται φυσικά σε αριθμητικοποίηση κατά Gödel.

Η πρόταση Ψ όμως, θα είναι είτε αληθής ή ψευδής.

Αν η Ψ είναι ψευδής, τότε από (2) η Ψ είναι αποδείξιμη.

Άτοπο, αφού δε νοείται απόδειξη μίας ψευδούς πρότασης σε μία συνεπή θεωρία.

Άρα η Ψ είναι αληθής. Αλλά από (2) είναι και μη-αποδείξιμη.

Άρα όντως η Ψ είναι αληθής και μη-αποδείξιμη!

Σημ: Όπως γίνεται αμέσως αντιληπτό, αυτή δεν είναι μία τυπική απόδειξη του θεωρήματος του Gödel. Τέτοιες εκλαϊκευμένες αποδόσεις κυκλοφορούν γενικά σε πολλές και διάφορες μορφές που είναι σαφώς πιο γοητευτικές στα μάτια του μέσου αναγνώστη από την τυπικά σωστή απόδειξη. Αυτή η συγκεκριμένη είναι αρκετά αντιπροσωπευτική, αν και μάλλον υπερβολικά αναλυτική σε σχέση με άλλες αντίστοιχες.

Πριν προχωρήσουμε για να δούμε ένα αποτέλεσμα του Tarski πιο ισχυρό από αυτό του Gödel, σύμφωνα με το οποίο η έννοια της αλήθειας δεν είναι πλήρως εκφράσιμη στην αριθμητική, θα αναφέρουμε κάποια ενδιάμεσα απαραίτητα αποτελέσματα, στο αυστηρό κλίμα και συμβολισμό του κεφαλαίου 3.

Ορισμός: Ένα κατηγορημα $P \subseteq \omega^n$ καλείται **αριθμητικό** αν υπάρχει ένας τύπος $\phi(x_1, \dots, x_n)$, τέτοιος ώστε $\forall a_1, \dots, a_n \in \omega, P(a_1, \dots, a_n) \Leftrightarrow \mathbb{N} \models \phi(\bar{a}_1, \dots, \bar{a}_n)$, δηλαδή το P ισχύει αν και μόνον αν ο τύπος ϕ αληθεύει στη δομή της αριθμητικής.

Ένα σύνολο A λέγεται αριθμητικό αν και μόνον αν το κατηγορημα $x \in A$ είναι αριθμητικό.

Θεώρημα⁽¹⁰⁾: Κάθε αναδρομικά αριθμήσιμο (οπότε κατά μείζονα λόγο αναδρομικό) κατηγορημα P είναι αριθμητικό.

➤ **Θεώρημα μη ορισιμότητας του Tarski.** Το σύνολο Tr των Gödel-αριθμών των αληθών τύπων της δομής $\mathbb{N}$, δεν είναι αριθμητικό.

Απόδειξη: Το σύνολο Tr , αν το εκφράσουμε με τον μαθηματικό συμβολισμό που χρησιμοποιούμε, είναι στην ουσία το σύνολο $\{\ulcorner \phi \urcorner \mid \mathbb{N} \models \phi\}$, όπου ϕ τύπος.

(10): Αυτό το θεώρημα επίσης έχει αποδειχθεί από τον Gödel στο διάσημο άρθρο του {#01 – Θεώρημα VII}.

Έστω ότι αυτό το σύνολο αυτό είναι όντως αριθμητικό. Τότε το κατηγορημα $P(a) \leftrightarrow a \in Tr$ είναι επίσης αριθμητικό, άρα υπάρχει ένας τύπος $\phi(x)$ τέτοιος ώστε $\forall a \in \omega, P(a) \leftrightarrow \mathbb{N} \models \phi(\bar{a})$, δηλαδή,

$$a \in Tr \leftrightarrow \mathbb{N} \models \phi(\bar{a}) \quad (1)$$

Έτσι, το σύνολο Tr γράφεται σαν: $Tr = \{a \mid a \in \omega \wedge \mathbb{N} \models \phi(\bar{a})\}$.

Εφαρμόζοντας τώρα το λήμμα της αυτό-αναφοράς για τον τύπο $\neg\phi(x)$ μαζί με το θεώρημα της ορθότητας, έπεται η ύπαρξη μίας πρότασης Ψ τέτοιας ώστε:

$$\mathbb{N} \models \Psi \leftrightarrow \neg\phi(\overline{\lceil \Psi \rceil})$$

Τότε όμως έχουμε ότι,

$$\begin{aligned} \mathbb{N} \models \Psi &\leftrightarrow \mathbb{N} \models \neg\phi(\overline{\lceil \Psi \rceil}) \\ &\leftrightarrow \lceil \Psi \rceil \notin Tr && \text{απο (1)} \\ &\leftrightarrow \mathbb{N} \not\models \Psi && \text{απο ορισμο του συνολου } Tr \end{aligned}$$

Αντίφαση.

□

Παρατήρηση: Αυτό το τελευταίο αποτέλεσμα αποτελεί επέκταση αυτού του Gödel διότι μας δείχνει ακριβώς ότι το σύνολο των αληθών προτάσεων της αριθμητικής δεν μπορεί να οριστεί πλήρως στα πλαίσια της αριθμητικής. Αυτό το βλέπουμε καλύτερα αν ανέβουμε σημασιολογικά ένα επίπεδο πιο πάνω από αυτό της παραπάνω διατύπωσης του θεωρήματος, ώστε το ‘σύνολο των Gödel-αριθμών αληθών τύπων’ γίνεται το ‘σύνολο των αληθών τύπων’. Αυτό το σύνολο όμως είναι επίσης μη αριθμητικό, ή μάλλον για την ακρίβεια είναι μεγαλύτερο από αριθμητικό, άρα έχουμε το ζητούμενο.

Τα παραπάνω αποτελέσματα αποτελούν ένα μικρό κομμάτι σε σχέση με τα όσα σχετίζονται γενικότερα με το διαγώνιο επιχείρημα του Cantor. Για παράδειγμα, σε ένα συγκεκριμένο μη-αποκρίσιμο πρόβλημα της θεωρίας υπολογιστότητας, το Halting Problem, το οποίο βγαίνει μη-αποκρίσιμο με αυτή τη μέθοδο, μπορούμε να ανάγουμε *κάθε* άλλο μη-αποκρίσιμο πρόβλημα! Μνημονεύεται και πάλι το διαγώνιο επιχείρημα, διότι είναι σαφής η αναλογία της βασικής του ιδέας με το θεώρημα του Gödel. Οπως υπάρχει το διαγώνιο στοιχείο μίας (πιθανά) πλήρους απαρίθμησης και

είναι στην ουσία ένα νέο στοιχείο της απαρίθμησης, έτσι και από την προσπάθεια απόδειξης όλων των αληθών προτάσεων, προκύπτουν προτάσεις που οι ίδιες δεν είναι αποδείξιμες, αν και παραμένουν αληθείς.

Από την παραπάνω χρονική πορεία αυτών των αποτελεσμάτων, προκύπτει και ένα ακόμη ενδιαφέρον στοιχείο, που περισσότερο βέβαια μοιάζει με κάποιο μαθηματικό ανέκδοτο. Ο Cantor, εμπνευστής της θεωρίας συνόλων, διατύπωσε για πρώτη φορά το διαγώνιο επιχείρημα. Όμως ο Russell, με βάση αυτό το επιχείρημα έδειξε τις ατέλειες της συνολοθεωρίας του Cantor και ουσιαστικά επέφερε την εγκατάλειψή της, χάριν την ZF θεωρίας συνόλων που αποφεύγει με έντεχνο τρόπο το παράδοξό του. Ο Russell στη συνέχεια (μαζί με τον Whitehead βέβαια), θέλοντας να απαλείψει μια για πάντα τις αντιφάσεις από τα μαθηματικά δημιούργησε τη θεωρία των τύπων που περιγράφεται στο Principia Mathematica, και λογικά έδωσε μεγάλη βαρύτητα στο να μην πέσει θύμα του δικού του παράδοξου. Όμως τελικά ο Gödel, με μία ακόμα πιο πολύπλοκη εφαρμογή του διαγώνιου (και πάλι!) επιχειρήματος, έδειξε τις ατέλειες του PM να περιγράψει πλήρως την έννοια της μαθηματικής συνέπειας και κατέρριψε με τη σειρά του τους ευσεβείς πόθους του Russell. Και ευτυχώς ίσως που το αποτέλεσμα του είναι καθολικό και αμετάκλητο, ώστε να μην υπάρξουν άλλες μαθηματικές απογοητεύσεις, πάνω σε αυτό τουλάχιστον το θέμα.

4.3 Αρχική μορφή του θεωρήματος του Gödel

Πριν περάσουμε στο 3^ο και τελευταίο μέρος της εργασίας, είναι απαραίτητο να δούμε την αρχική διατύπωση του 1^{ου} θεωρήματος μη-πληρότητας⁽¹¹⁾. Η ανάγκη αυτή πηγάζει από το γεγονός ότι μερικοί χρησιμοποιούν το αρχικό αποτέλεσμα στη διατύπωση των (κυρίως, μαθηματικών) επιχειρημάτων τους, όπως θα διαπιστώσουμε σε όσα θα ακολουθήσουν στο επόμενο μέρος. Ουσιαστικά, η διαφορά της σύγχρονης και της αρχικής διατύπωσης, έγκειται στο ότι -όπως έχει ήδη αναφερθεί μετά το τέλος της απόδειξης του 1^{ου} θεωρήματος μη-πληρότητας- η άρνηση της Gödel-πρότασης είχε αρχικά αποδειχθεί με υπόθεση την ω -συνέπεια και όχι τη (συνήθη) συνέπεια της αριθμητικής του Peano. Αυτή την έννοια της ω -συνέπειας είναι που πρέπει να αποσαφηνίσουμε για να μην υπάρχουν προβλήματα κατανόησης στη συνέχεια.

(11): Μόνο του 1^{ου}, διότι η αρχική διατύπωση του 2^{ου} θεωρήματος μη-πληρότητας δε διαφέρει από αυτή της παραγράφου 3.3.

Ορισμός: Έστω ένα σύνολο προτάσεων T , στη γλώσσα της αριθμητικής $\mathcal{L}_A$. Λέμε ότι το T είναι **ω-συνεπές** αν και μόνον αν για κάθε τύπο $\phi(x)$, να ισχύει ότι:

$$\text{Αν } T \vdash \neg\phi(\bar{n}) \quad \forall \bar{n} \in \mathbb{N}, \text{ τότε } T \not\vdash \exists x\phi(x).$$

Δηλαδή, αν από το σύνολο T αποδεικνύεται η άρνηση του τύπου ϕ για κάθε πιθανό όρισμά του, τότε επίσης από το T δεν αποδεικνύεται ότι υπάρχει κάποιο x για το οποίο να ισχύει ο $\phi(x)$.

Σημ.: Αμέσως έπεται από τον ορισμό, ότι αν $\mathbb{N} \models T$, τότε το T είναι ω-συνεπές. Η αντίστοιχη θεωρία που πιθανόν προκύψει από τη χρήση του συνόλου T ως το σύνολο των μη-λογικών αξιωμάτων της, καλείται (εύλογα) ω-συνεπής.

Πρόταση: Αν το T είναι ω-συνεπές, τότε το T είναι συνεπές. Το αντίστροφο δεν ισχύει, δηλαδή γενικά, η συνέπεια δεν συνεπάγεται την ω-συνέπεια.

Η παραπάνω πρόταση, συνδέει τις έννοιες της ω-συνέπειας και της συνέπειας και δείχνει επίσης ότι η ω-συνέπεια που εμπλέκεται στην αρχική διατύπωση του 1^{ου} θεωρήματος μη-πληρότητας (βλ. αμέσως παρακάτω), είναι όντως *ισχυρότερη συνθήκη* από αυτή της συνέπειας.

1^ο Θεώρημα μη-πληρότητας (αρχική μορφή): Υπάρχει μία πρόταση Ψ της $\mathcal{L}_A$, τέτοια ώστε:

1. Αν το S είναι συνεπές, τότε $S \not\vdash \Psi$ και
2. Αν το S είναι ω-συνεπές, τότε $S \not\vdash \neg\Psi$.

Μία σημαντική σημείωση: Κάτι που παραλείφθηκε να αναφερθεί σε σχέση με το 1^ο θεώρημα μη-πληρότητας, είναι το ότι υπάρχει μία (τουλάχιστον) μη-αποκρίσιμη πρόταση που έχει *μαθηματικό περιεχόμενο*, δεν είναι δηλαδή αυτοαναφορικού τύπου όπως η Gödel-πρόταση. Αυτή η πρόταση εκφράζει μία παραλλαγή ενός θεωρήματος του Ramsey και ανακαλύφθηκε το 1977 από τους J. Paris και L. Harrington.

Με την παρουσίαση αυτού του τελευταίου αποτελέσματος, ολοκληρώνεται το 2^ο μέρος της εργασίας και υπάρχει πλέον το υπόβαθρο που χρειάζεται για να έχουμε μία σωστή θεώρηση των αποτελεσμάτων του Gödel, της αποδεικτικής μεθόδου που οδήγησε σε αυτά και του γενικότερου μαθηματικού πλαισίου στο οποίο κινούνται. Το υπόβαθρο αυτό είναι απαραίτητο, διότι, στο μέρος της εργασίας που ακολουθεί, θα δούμε χρήσεις και διατυπώσεις του θεωρήματος του Gödel που καμία σχέση δεν έχουν με όσα έχουμε ασχοληθεί μέχρι τώρα. Αυτές οι ‘αναπροσαρμογές’ του αποτελέσματος της μη-πληρότητας μπορεί βέβαια να οδηγούν συνήθως σε πολύ ενδιαφέρουσες φιλοσοφικές σκέψεις, αλλά καλό θα ήταν να έχουμε στο μυαλό μας πάντα, σαν σημείο αναφοράς, τη μαθηματική διάσταση των πραγμάτων. Άλλωστε, μέσα από τη χρήση της μαθηματικής λογικής διατυπώνονται και τα περισσότερα επιχειρήματα που θέλουν να καταδείξουν λογικά λάθη που προκύπτουν από τη σχεδόν αυθαίρετη χρήση του θεωρήματος χάριν φιλοσοφικών σκέψεων. Επιπλέον, αυτές οι φιλοσοφικές σκέψεις έχουν σαν αρχικό τους ερέθισμα το θεώρημα του Gödel όπως αυτό ισχύει τυπικά, ενώ οι εμπνευστές τους επίσης χρησιμοποιούν μαθηματικά λογικά επιχειρήματα για να αντεπεξέλθουν στην αντιπαράθεση.

Καταλαβαίνουμε λοιπόν ότι το 2^ο μέρος που τελειώνει εδώ, μας παρέχει, εκτός των άλλων, και όλα τα εφόδια που χρειαζόμαστε για να αντεπεξέλθουμε σε μία ολοκληρωμένη ανάλυση των απόψεων, επιχειρημάτων και ιδεών που θα προκύψουν στο 3^ο και τελευταίο μέρος της διπλωματικής αυτής εργασίας.

ΜΕΡΟΣ III

1. Εισαγωγή

Το αντικείμενο του μέρους στο οποίο εισερχόμαστε, αποτελεί ένα κομμάτι των ευρύτερων φιλοσοφικών προεκτάσεων που προέκυψαν από το Θεώρημα του Gödel. Έχοντας επισημάνει επανειλημμένα τις σημαντικές αλλαγές στην μαθηματική και λογική σκέψη και τις ποικίλες συνέπειες στη φιλοσοφία των μαθηματικών που προέκυψαν από αυτό, καταλαβαίνει κανείς ότι μία συνολική ανάλυση του αντίκτυπου των αποτελεσμάτων του Gödel καθίσταται αδύνατη στα πλαίσια μίας και μόνο εργασίας. Έτσι, επιλέχθηκε να αναλυθεί το κομμάτι των προεκτάσεων που έχει να κάνει με τη φιλοσοφία του νού, και πιο συγκεκριμένα μελετάται το αντίκτυπο των θεωρημάτων του Gödel στην φιλοσοφική θεώρηση η οποία εξισώνει τις ανθρώπινες νοητικές λειτουργίες με τις μηχανιστικές διεργασίες.

Ο προβληματισμός για το αν ο άνθρωπος είναι στην ουσία μία μηχανή ή κάτι παραπάνω από ένα σύνολο μηχανιστικών διαδικασιών ξεκινάει από πολύ παλιά, αλλά αποκτά μεγαλύτερη δημοτικότητα τον 17^ο αι., όταν οι εξελίξεις στον τομέα των φυσικών επιστημών ενίσχυσαν τις σκέψεις που κινούνται προς αυτή την κατεύθυνση. Ο Καρτέσιος, για παράδειγμα, υποστήριζε ότι τα ζώα θα μπορούσαν να περιγραφούν σαν απλές μηχανές και ίσως κάτι ανάλογο να μπορούσε να ισχύει και για το ανθρώπινο *σώμα*. Υπήρχαν όμως και πιο τολμηροί διανοητές, όπως ο De la Mettrie, οι οποίοι ήταν διατεθειμένοι να επεκτείνουν αυτή την αναλογία και στις λειτουργίες του ανθρώπινου *νου*, χωρίς όμως ιδιαίτερη σαφήνεια, εφόσον δεν υπήρχε ακόμα μία ξεκάθαρη εικόνα για τη φύση της μηχανής και των διαδικασιών που θα μπορούσε αυτή να αναπαράγει. Στις μέρες μας βέβαια, η εικόνα αυτή έχει σαφώς περισσότερο ξεκαθαριστεί και η συζήτηση έχει μία πολύ καλύτερη βάση πάνω στην οποία μπορεί να κινηθεί. Οι δύο κύριες σύγχρονες αντιμαχόμενες θεωρίες είναι ο **Μηχανισμός [Mechanism]** που ισχυρίζεται ότι οι διεργασίες του νου είναι, στην ουσία μηχανιστικές, και ο **Νοητισμός [Mentalism]** που ισχυρίζεται ότι ο νους λειτουργεί εγγενώς διαφορετικά από οποιοδήποτε μηχανιστικό κατασκεύασμα.

Ο κύριος λοιπόν σκοπός αυτού του μέρους της εργασίας είναι να διαπιστωθούν οι πιθανές επιπτώσεις του θεωρήματος του Gödel στον Μηχανισμό. Τον βασικό άξονα γύρω από τον οποίο θα δομηθεί το 3^ο αυτό μέρος αποτελεί ένα άρθρο του John Randolph Lucas {‘Minds, machines and Gödel’ - #05} που δημοσιεύθηκε για πρώτη φορά το 1961 στο περιοδικό Philosophy και στο οποίο διατυπώνεται για πρώτη φορά με σαφήνεια ένα ολοκληρωμένο επιχείρημα ενάντια στο Μηχανισμό, επιχείρημα που έχει σαν βάση του το θεώρημα του Gödel. Ο αντίκτυπος της βασικής θέσης του Lucas ήταν σημαντικός και η αφορμή να ξεκινήσει μία διαδικασία ανταπαντήσεων μέσω άρθρων σε διάφορα περιοδικά φιλοσοφίας, διαδικασία η οποία έχει να προσφέρει άφθονο υλικό για φιλοσοφικές αλλά και μαθηματικές σκέψεις, καθότι οι περισσότεροι από τους αντιφρονούντες των απόψεων του Lucas, είτε προσπαθούν με χρήση άλλων φιλοσοφικών σκέψεων ή αποτελεσμάτων της λογικής να τις απορρίψουν, ή ψάχνουν άλλες, πιο συμβατές με τις ιδέες τους συνέπειες του θεωρήματος του Gödel στο Μηχανισμό. Πρόκειται δηλαδή για μία γόνιμη αντιπαράθεση από την οποία προέκυψαν θέματα που συνεχίζουν να απασχολούν την επιστημονική κοινότητα μέχρι σήμερα. Έτσι, μέσω της ανάλυσης που ακολουθεί, εκτός της σύγκρισης και αξιολόγησης των αντιπαρατιθέμενων επιχειρημάτων που αφορούν το αρχικό επιχείρημα του Lucas, θα έχουμε την ευκαιρία να διαπιστώσουμε και την ιστορική εξέλιξη των ιδεών, ώστε στο τέλος της εργασίας να υπάρξει μία συνολική εικόνα για τις φιλοσοφικές προεκτάσεις του θεωρήματος του Gödel σε αυτό το συγκεκριμένο και πολύ ενδιαφέρον κομμάτι της φιλοσοφίας του νου.

Σημ: Πριν περάσουμε στο κύριο μέρος της ανάλυσης, κρίνεται αναγκαίο να γίνει μία σύντομη αναφορά στην πορεία παρουσίασης των άρθρων τα οποία αποτελούν την κύρια βιβλιογραφική πηγή των θεμάτων που θα μας απασχολήσουν. Η παρουσίαση χωρίζεται ουσιαστικά σε τρεις χρονικές περιόδους, αυτή που προηγείται των θέσεων του Lucas, την περίοδο όπου συζητούνται αυτές οι θέσεις, και την πιο πρόσφατη περίοδο όπου προκύπτουν νέα θέματα αντιπαράθεσης. Τα θέματα της κάθε περιόδου ομαδοποιούνται με βάση τη συνάφεια των ιδεών που εκφράζουν, ενώ ενδιάμεσα παρατίθενται και όποια άλλα ενδιαφέροντα θέματα προκύπτουν, όπως οι απόψεις του ίδιου του Gödel πάνω στην αναλογία ανθρώπου-μηχανής όπως αυτή εκφράζεται παραπάνω, ή ευρύτερες προεκτάσεις του θεωρήματός του οι οποίες προέκυψαν με αφορμή τη συζήτηση που θα παρακολουθήσουμε.

2. Περίοδος προ Lucas

Όπως φαίνεται από τον τίτλο του κεφαλαίου στο οποίο εισερχόμαστε, το σημείο από το οποίο ξεκινάμε την ανάλυσή μας, δεν είναι το άρθρο του J.R. Lucas που περιέχει το επιχείρημα ενάντια στο Μηχανισμό. Ξεκινάμε από δύο άρθρα που προηγήθηκαν χρονικά αυτού του Lucas, και στην ουσία έδωσαν το αρχικό υλικό πάνω στο οποίο βασίστηκαν όλα τα υπόλοιπα (ειδικά αυτό του A.M. Turing που ακολουθεί αμέσως μετά). Περιέχουν δηλαδή βασικές θέσεις και προβληματισμούς που θα μας απασχολήσουν εκτενέστερα παρακάτω, οπότε κρίνεται απαραίτητη μία, όσο το δυνατόν πληρέστερη, παρουσίασή τους.

2.1 A.M. Turing. ‘Computing Machinery and Intelligence’ (1950)

Αυτό το άρθρο {#02} του Alan Turing είναι ίσως το πιο διάσημο από όλα όσα θα παρατεθούν στη συνέχεια και περιλαμβάνει πολλά βασικά στοιχεία για τις μηχανές και τις υπολογιστικές δυνατότητες που μπορούν να έχουν. Θα μπορούσαμε να πούμε πως πρόκειται ουσιαστικά για μία άρτια καταγραφή της βασικής θέσης του Μηχανισμού, η οποία, με βάση τη θεώρηση του Turing μπορεί να εκφραστεί ως εξής:

Μηχανιστική Θέση (εκδοχή Turing): Κάθε ανθρώπινη νοητική διαδικασία μπορεί να αναπαρασταθεί πειστικά από μία αντίστοιχη μηχανιστική (δηλ. παραγόμενη από μηχανή), σε τέτοιο βαθμό ώστε να μη μπορεί να γίνει σαφής διαχωρισμός μεταξύ των δύο από κάποιον τρίτο (άνθρωπο ή μηχανή).

Η παραπάνω διατύπωση αποτελεί μία ιδιαίτερη εκδοχή της γενικότερης μηχανιστικής θέσης, η οποία πολύ απλά ισχυρίζεται ότι *οι μηχανές μπορούν να σκέπτονται*, πράγμα που φυσικά αντικατοπτρίζει το δόγμα του Μηχανισμού, όπως αυτό αναφέρθηκε στην εισαγωγή. Καθότι όμως το γενικότερο ερώτημα πίσω από τη μηχανιστική θέση (για την δυνατότητα ‘σκέψης’ των μηχανών), φαινόταν στον Turing ασαφές και υπαίτιο δημιουργίας εντυπώσεων μη-επιθυμητών, πρότεινε ο ίδιος σε αυτό το άρθρο την *αντικατάστασή* του με ένα τεστ που έμεινε στην ιστορία σαν το ‘τεστ του Turing’ [Turing’s Test]. Με βάση την ορολογία αυτού του τεστ, που αναλύεται παρακάτω, προέκυψε η παραπάνω (πρωτότυπη) διατύπωση της μηχανιστικής θέσης.

Το τεστ του Turing

Σε τρία μη επικοινωνούντα δωμάτια, τοποθετούμε μια μηχανή (M), έναν άνθρωπο (A) και έναν ανακριτή (K). Οι τρεις τους επικοινωνούν μέσω τυπογραφημένων μηνυμάτων και δεν υπάρχει καμία οπτική ή άλλου είδους επαφή μεταξύ τους. Ο σκοπός του K είναι να καταλάβει ποιος είναι ο άνθρωπος και ποια η μηχανή μόνο μέσω ερωτήσεων που έχει δικαίωμα να υποβάλλει στον καθένα. Υποθέτουμε επίσης πως ο σκοπός της M είναι να μπερδέψει τον K, ενώ σκοπός του A είναι να τον βοηθήσει να βρει τη σωστή απάντηση.

Υπό αυτές τις υποθέσεις, πετυχημένο θεωρείται το τεστ για μία μηχανή που θα καταφέρει να κάνει τον K να απαντήσει λάθος σε κάποιο ποσοστό ανάλογο του ποσοστού που θα απαντούσε λάθος αν στη θέση της M ήταν μία γυναίκα και ο K προσπαθούσε να μαντέψει το φύλο των ανακρινόμενων. Η πιο προφανής στρατηγική που θα βοηθήσει τη μηχανή να περάσει το τεστ, φαίνεται να είναι αυτή του να προσπαθεί να δώσει απαντήσεις που αυθόρμητα θα δίνονταν από έναν άνθρωπο, να αναπαράγει δηλαδή την ανθρώπινη συλλογιστική διαδικασία πειστικά. Αν βέβαια υπάρχει μία τέτοια μηχανή, μπορούμε άμεσα να αποδεχθούμε την εκδοχή της μηχανιστικής θέσης του Turing, άρα και την ευρύτερη μηχανιστική θέση.

Όσον αφορά την αξιοπιστία του τεστ, δεδομένου ότι οι ερωτήσεις του K δεν υπόκεινται σε περιορισμούς, ο Turing πιστεύει ότι με αυτό *μπορούμε να ελέγξουμε οτιδήποτε θεωρεί καθένας ότι έχει να κάνει με την ανθρώπινη έκφραση*, η εγκυρότητά του δηλαδή γίνεται δεκτή χωρίς περαιτέρω ανάλυση του θεωρητικού του υποβάθρου.

Μετά την παρουσίαση του τεστ, ο Turing προχωρά στον ορισμό της έννοιας της ‘μηχανής’ και πιο συγκεκριμένα της μηχανής που λαμβάνει μέρος στο τεστ. Τα κριτήριά που βάζει για να βγάλει αυτόν τον ορισμό είναι:

- ✓ Να μπορούμε να εισάγουμε στη μηχανή οποιαδήποτε τεχνική μηχανολογίας έχουμε στη διάθεσή μας,
- ✓ Να επιτρέψουμε την πιθανότητα κατασκευής μίας μηχανής της οποίας δεν θα μπορούμε να περιγράψουμε απολύτως τη λειτουργία της, ας πούμε διότι η μέθοδος κατασκευής της είναι ακόμη σε πειραματικό στάδιο και
- ✓ Να μην επιτρέπεται η εισαγωγή βιολογικών στοιχείων σε κανένα κομμάτι της μηχανής.

Με βάση αυτά τα τρία κριτήρια, ο μόνος τύπος μηχανής που μπορεί να πάρει μέρος στο τεστ είναι ένας **ψηφιακός υπολογιστής [digital computer]**, ένα θεωρητικό μοντέλο που σαν βασική αρχή του έχει ότι μπορεί φέρει εις πέρας οποιαδήποτε διαδικασία που μπορεί να εκτελέσει και ένας ανθρώπινος υπολογιστής. Ένας ψηφιακός υπολογιστής αποτελείται από τρία μέρη: (i) τη μνήμη, (ii) την εκτελεστική μονάδα και (iii) τη μονάδα ελέγχου, με βάση τα οποία εκτελεί τις λειτουργίες του. Περαιτέρω τεχνικές λεπτομέρειες που αφορούν τη φύση του ψηφιακού υπολογιστή δεν θα μας απασχολήσουν, για την μετέπειτα ανάλυσή μας να αναφέρουμε μόνο ότι οι κανόνες που ακολουθεί ένας τέτοιος υπολογιστής είναι σταθεροί και *δεν υπάρχει δυνατότητα αυτόκλητης παρέκκλισης* από αυτούς, ενώ η μνήμη τους μπορεί να επεκταθεί θεωρητικά απεριόριστα. Ένα ακόμη ενδιαφέρον στοιχείο είναι το ‘τυχαίο στοιχείο’ [random element] που έχουμε τη δυνατότητα να εισάγουμε, προγραμματίζοντας τον υπολογιστή να ‘ρίχνει ένα ζάρι’ και να αποθηκεύει το αποτέλεσμα σε ένα κομμάτι της μνήμης του.

Ένα θέμα βέβαια που αμέσως προκύπτει από τον ορισμό του ψηφιακού υπολογιστή, είναι το κατά πόσο μπορεί ένας τέτοιος υπολογιστής να προσομοιώσει κάθε υπολογιστική δυνατότητα. Αυτή η δυνατότητα φαίνεται να δίνεται από έναν συνδυασμό της θέσης του Church και των σχολίων πάνω στην ‘παγκοσμιότητα των ψηφιακών υπολογιστών’ του Turing που περιέχονται στο εν λόγω άρθρο, όπου στην ουσία *δείχνονται ισοδύναμοι οι ψηφιακοί υπολογιστές με τις μηχανές Turing*.

Με βάση λοιπόν το μοντέλο που περιγράφει με κάθε λεπτομέρεια ο Turing στο άρθρο του, μπορούμε τελικά να αντικαταστήσουμε το ερώτημα: ‘Μπορούν οι μηχανές να σκέπτονται;’ με το: ‘Υπάρχουν ιδεατοί ψηφιακοί υπολογιστές που μπορούν να τα πάνε καλά στο τεστ του Turing;’

Έχουμε αφήσει να εννοηθεί ότι ο ίδιος ο Turing ρέπει προς μία καταφατική απάντηση των παραπάνω ερωτημάτων, ή μάλλον του τελευταίου, και κατ’ επέκταση προς μία αποδοχή της μηχανιστικής θέσης. Πράγματι, όπως θα δούμε παρακάτω, ο Turing επιβεβαιώνει αυτή του τη θέση αλλά όχι με απόλυτο τρόπο. Περισσότερο φαίνεται να εστιάζει στην αντίκρουση των αντίπαλων θέσεων της θεώρησής του, για να ισχυροποιήσει τη δική του από τη μία, και να ορίσει πλήρως το πρόβλημα από την άλλη. Αυτές οι απόψεις ενάντια στη μηχανιστική θέση, αναφέρονται επιγραμματικά παρακάτω, για λόγους που θα γίνουν αμέσως προφανείς.

1. **Η θεολογική αντίρρηση.** Η σκέψη είναι λειτουργία της ανθρώπινης ψυχής και ο Θεός έδωσε ψυχή μόνο στον άνθρωπο και όχι στα ζώα ή τα υλικά αντικείμενα και τις μηχανές.
2. **Η ‘κρύβομαι πίσω από το δάχτυλό μου’ αντίρρηση.** Οι συνέπειες της ύπαρξης σκεπτόμενων μηχανών θα είναι καταστροφικές. Καλύτερα να μην σκεφτόμαστε καν το ενδεχόμενο.
3. **Η μαθηματική αντίρρηση.** Διάφορα αποτελέσματα της μαθηματικής λογικής (το θεώρημα του Gödel και άλλα σχετικά αποτελέσματα των Church, Kleene, Turing...) μπορούν να χρησιμοποιηθούν για να καταδειχθούν τα όρια των δυνατοτήτων των μηχανιστικών διαδικασιών.
4. **Το επιχείρημα της συνείδησης.** Μόνο αν μία μηχανή μπορεί να γράψει μουσική ή ένα ποίημα λόγω συναισθημάτων και σκέψεων δικών της και όχι τυχαίων θέσεων συμβόλων, μπορούμε να θεωρήσουμε ότι φτάνει στα επίπεδα του ανθρώπινου νου.
5. **Επιχειρήματα διαφόρων αδυναμιών.** Αυτά είναι της μορφής ‘Μπορεί να μπορούν οι μηχανές να κάνουν όλα τα πράγματα που ισχυρίζεσαι, αλλά δεν μπορείς να δημιουργήσεις μία μηχανή που να κάνει το X πράγμα’. Όπου X μπορεί να είναι: να έχει χιούμορ, να μαθαίνει εμπειρικά, να καινοτομεί κτλ.
6. **Η αντίρρηση της Lady Lovelace.** Μία μηχανή δε μπορεί να δημιουργήσει τίποτα. Μπορεί μόνο να κάνει αυτό για το οποίο φτιάχτηκε να κάνει.
7. **Το βιολογικό επιχείρημα.** Μία μηχανή με διακριτές καταστάσεις όπως η TM δεν μπορεί να αναπαράγει τη συνέχεια του ανθρώπινου νευρικού συστήματος και άρα τη λειτουργία του εγκεφάλου.
8. **Το επιχείρημα της απρόβλεπτης συμπεριφοράς.** Δεν είναι δυνατόν να παραχθεί ένα σετ κανόνων που να αναπαράγει την ανθρώπινη συμπεριφορά διότι οι άνθρωποι μπορούν να παραβλέπουν ακόμα και αυστηρά δομημένους κανόνες αν αυτό κριθεί σκόπιμο σε σχέση με τις περιστάσεις.
9. **Το μεταφυσικό επιχείρημα.** Υπάρχουν επιστημονικώς αποδεκτά μεταφυσικά φαινόμενα, όπως για παράδειγμα η τηλεπάθεια, με τη χρήση των οποίων καθίσταται αδύνατη οποιοδήποτε επικράτηση της μηχανής έναντι του ανθρώπου στο τεστ του Turing.

Όλα αυτά τα επιχειρήματα παρουσιάζουν ευρύτερο ενδιαφέρον όσον αφορά το γενικότερο θέμα που μας απασχολεί, αλλά και ειδικότερα όσον αφορά το βασικό επιχείρημα του Lucas, το οποίο προκύπτει τελικά να είναι μία υποπερίπτωση κάποιου συνδυασμού τους. Συγκεκριμένα, θα δούμε παρακάτω πως είναι στην ουσία ένας συνδυασμός κυρίως των αντιρρήσεων (3) και (5) από τις παραπάνω, ενώ διάφορα άλλα επιχειρήματα πέραν του βασικού, περιέχουν στοιχεία και από τις αντιρρήσεις (4) και (6). Εύλογα λοιπόν κρίνεται λοιπόν χρήσιμο να αναφερθούν οι απαντήσεις που δίνει ο Turing σε όλες τις παραπάνω αντιρρήσεις, με ιδιαίτερη έμφαση φυσικά σε αυτές που μας ενδιαφέρουν περισσότερο με βάση τα όσα θα ακολουθήσουν.

Για τις δύο πρώτες αντιρρήσεις δεν έχουμε να πούμε πολλά πράγματα, εύκολα δείχνεται ότι δεν αποτελούν ισχυρά 'χαρτιά' ενάντια στη μηχανιστική θέση, περισσότερο μοιάζουν με προκαταλήψεις ενάντια σε μία καινοτόμα ιδέα. Η επόμενη μαθηματική αντίρρηση όμως είναι που περιέχει το Θεώρημα του Gödel το οποίο αποτελεί το κύριο θέμα επεξεργασίας μας, οπότε αξίζει να επεκταθούμε.

Καταρχήν αναφέρουμε ότι ο Turing έχει αποδείξει ένα αποτέλεσμα που ταιριάζει περισσότερο στην ανάλυσή του, το παρακάτω:

Δεν υπάρχει TM που να μπορεί να καθορίσει σε πεπερασμένο αριθμό βημάτων αν οποιοσδήποτε τύπος του κατηγορηματικού λογισμού είναι θεώρημά του ή όχι.

Γενικεύοντας, ένα επιχείρημα που πηγάζει από τη μαθηματική λογική είναι πιθανό να μπορεί να καταδείξει την ανικανότητα μιας μηχανής να δώσει σωστή απάντηση σε κάποιο ερώτημα ή, πιο εγκεκριμένα, την αδυναμία της να δώσει απάντηση σε κάποιο *αυτοαναφορικό* ερώτημα (οπότε έχουμε να κάνουμε και με την αντίρρηση (5)). Η απάντηση του Turing σε αυτό είναι ότι μπορεί να υπάρχουν όρια στις δυνατότητες μίας συγκεκριμένης μηχανής, αλλά δεν έχει αποδειχθεί ότι το ίδιο δεν ισχύει και για τις δυνατότητες της ανθρώπινης συλλογιστικής. Έτσι η αίσθηση ανωτερότητας που μπορεί να έχει κάποιος απαντώντας σε ένα ερώτημα που αδυνατεί να απαντήσει μία μηχανή είναι πλασματική διότι μπορεί αντίστοιχα η μηχανή να μπορεί να απαντήσει σωστά εκεί που ο άνθρωπος κάνει λάθος. Συν τοις άλλοις, ο Turing ισχυρίζεται ότι δεν μπορούμε με αυτό τον τρόπο να καταδείξουμε την αδυναμία *όλων των δυνατών* μηχανών.

Η απάντηση αυτή έχει κάποια ενδιαφέροντα στοιχεία αλλά κατά τη γνώμη μου δεν είναι ολοκληρωμένη. Ένα πιο πειστικό επιχείρημα που δεν αναφέρεται ευθέως θα ήταν το εξής: Εφόσον δεν ξέρουμε ότι η ανθρώπινη συλλογιστική δεν έχει περιορισμούς, τότε, υποθέτοντας ότι έχει, μπορούμε να πούμε ότι αυτή την εγγενή ιδιότητα φυσικό είναι να την προσομοιώνει μία μηχανή με ανάλογες δυνατότητες, άρα δεν καταρρίπτεται η μηχανιστική θέση. Από την άλλη, ο Turing προσπαθεί να μειώσει την αξία του επιχειρήματος, με το σκεπτικό ότι τελικά είναι ένα μικρόψυχο επιχείρημα που θέλει τους ανθρώπους ανώτερους από τις μηχανές. Αυτή η τακτική δεν είναι ιδιαίτερα πειστική, ούτε το γεγονός ότι δεν μπορεί κανείς να αντιπαρατεθεί με όλες τις μηχανές. Διότι στα πλαίσια μίας φιλοσοφικής συζήτησης, οι έννοιες έχουν μία παγκοσμιότητα και δεν μπορούμε να εξειδικεύουμε τους ρόλους σε έναν συγκεκριμένο άνθρωπο και μία συγκεκριμένη μηχανή. Θα δούμε παρακάτω πως ο Lucas αποφεύγει έντεχνα αυτή τη δυσκολία για να διατυπώσει το επιχείρημά του, και γιατί δεν καλύπτεται ουσιαστικά από αυτή την απάντηση.

Η επόμενη αντίρρηση (4) εκλαμβάνεται από τον Turing σαν μία αμφισβήτηση της εγκυρότητας του τεστ του, οπότε απαντά στη βάση αυτή ικανοποιητικά. Βέβαια το θέμα συνείδηση έχει πολύ πιο βαθιές προεκτάσεις, και μάλιστα, θα διαπιστώσουμε ότι ο Lucas συνδέει τη συνείδηση με το 2^ο θεώρημα μη-πληρότητας (!), μέσω μίας συλλογιστικής που θα συναντήσει, εύλογα, σφοδρές αντιδράσεις.

Όσον αφορά την 5^η ομάδα επιχειρημάτων, εκτός από τη σύνδεσή τους με το μαθηματικό επιχείρημα που αναφέραμε, προκύπτουν και άλλα ενδιαφέροντα ζητήματα. Ένα από αυτά είναι αυτό της ενδοσκόπησης, δηλαδή το ερώτημα *κατά πόσο μία μηχανή μπορεί να γίνει το αντικείμενο της ίδιας της διεργασίας της*. Υπάρχει το αποτέλεσμα που λέει ότι ένα πρόγραμμα μπορεί να έχει σαν έξοδο τον ίδιο του τον κώδικα από τη θεωρία υπολογιστότητας, ενώ μέσω του ελέγχου μία μηχανή μπορεί να ελέγχει τις δομικές αλλαγές της αλλά και να κατασκευάζει νέα προγράμματα με βάση κάποιους κανόνες. Έτσι το ερώτημα φαίνεται να απαντάται υπό μία έννοια καταφατικά, αλλά θα το ξανασυναντήσουμε αργότερα σε άλλη μορφή, οπότε θα δούμε ότι μία άμεση απάντηση δεν είναι τόσο απλή υπόθεση.

Η αντίρρηση της Lady Lovelace είναι μία αντίρρηση από το παρελθόν, όπου η έννοια της μηχανής είχε τελείως διαφορετικό περιεχόμενο από αυτό που έχει σήμερα. Σαν απάντηση, ο Turing αντιτάσσει τις μηχανές με δυνατότητα μάθησης [learning machines] που βέβαια την εποχή που έγραφε το άρθρο του, ήταν ακόμη στα πολύ πρώτα τους βήματα. Αυτού του είδους οι μηχανές θα πρέπει ίσως να μας

προβληματίσουν αργότερα, όταν ο Lucas θα παραθέσει το επιχείρημά του για όλες τις μηχανές, διότι μπορεί πίσω από την ιδέα αυτή να κρύβεται μία πειστική απάντηση στους αντι-μηχανιστικούς προβληματισμούς.

Τα τρία τελευταία επιχειρήματα δεν θα μας απασχολήσουν ιδιαίτερα, θα ασχοληθούμε μόνο με αυτό που αναφέρει ο Turing για να διαλύσει τυχόν παρερμηνείες της λέξης 'μηχανικός' που χρησιμοποιείται, σε σχέση με την ηλεκτρική φύση του ψηφιακού υπολογιστή. Το απόσπασμα που μεταφέρουμε αυτούσιο, αναφέρεται στην πρώτη εφαρμογή της ιδέας πίσω από τον ψηφιακό υπολογιστή που (σχεδόν) πραγματοποιήθηκε από τον Charles Babbage το 1839 χρησιμοποιώντας μόνο μηχανικά μέσα. «Το γεγονός ότι η μηχανή του Babbage ήταν τελείως μηχανική μας βοηθά να ξεφύγουμε από μία προκατάληψη. Συχνά θεωρείται ότι έχει σημασία το γεγονός ότι οι σύγχρονοι ψηφιακοί υπολογιστές είναι ηλεκτρικοί και το νευρικό μας σύστημα είναι επίσης ηλεκτρικό. Εφόσον η μηχανή του Babbage δεν ήταν ηλεκτρική, και αφού όλου οι ψηφιακοί υπολογιστές είναι κατά μία έννοια ισοδύναμοι με αυτήν, βλέπουμε ότι η χρήση του ηλεκτρισμού δεν μπορεί να έχει καμία θεωρητική σημασία.»

Η πεποίθηση του ίδιου του Turing ήταν ότι σε πενήντα χρόνια από τη συγγραφή του άρθρου, θα υπάρχουν ψηφιακοί υπολογιστές που θα μπορούν να τα πάνε τόσο καλά στο τεστ του, ώστε ένας μέσος ανακριτής να μην έχει περισσότερο από 70% πιθανότητα για να κάνει τη σωστή επιλογή μετά από 5 λεπτά ανάκρισης. Το έτος 2000 βέβαια ήλθε και παρήλθε, και παρ' ότι οι δυνατότητες των σύγχρονων υπολογιστών, όσον αφορά τη μνήμη και την ταχύτητα, σίγουρα ξεπέρασαν ακόμα και τις πιο αισιόδοξες προβλέψεις του Turing, ακόμη εκκρεμεί η εκπλήρωση της εικασίας του. Υπάρχει όμως, πλέον, τουλάχιστον ένα πνευματικό πεδίο στο οποίο ο υπολογιστής ανταγωνίζεται με επιτυχία τον άνθρωπο, το σκάκι⁽¹⁾, οπότε κάποιος αισιόδοξος υποστηρικτής του, μπορεί να πει ότι έγινε η αρχή για να εκπληρωθεί κάποτε η ελπίδα που εκφράζει στο τέλος του άρθρου του, ότι «μπορούμε να ελπίζουμε ότι οι μηχανές τελικά θα μπορούν να συναγωνιστούν τον άνθρωπο σε όλα τα καθαρά πνευματικά πεδία».

(1): Είναι γνωστό ότι το 1997 ο 'υπερυπολογιστής Deep Blue' της IBM κατάφερε να κερδίσει σε μία σκακιστική αναμέτρηση τον παγκόσμιο πρωταθλητή, Garry Kasparov. Σημειώστε ότι το σκάκι αναφέρεται από τον Turing σαν ένα πεδίο «αφηρημένης ανθρώπινης δραστηριότητας» το οποίο θα μπορούσε να αποτελεί ένα καλό υποψήφιο για το ξεκίνημα της αναζήτησης πεδίων συναγωνισμού ανθρώπου-μηχανής!

2.2 Hilary Putnam. ‘Minds and Machines’ (1960)

Το επόμενο άρθρο που θα μας απασχολήσει πριν από αυτό του Lucas, είναι αυτό του τίτλου του κεφαλαίου {#04}, στο οποίο ο H. Putnam παραθέτει μία σειρά επιχειρημάτων που προσπαθούν να δώσουν μία απάντηση στο δυϊστικό πρόβλημα μυαλού και σώματος, μέσω της εισαγωγής μίας αναλογίας ανθρώπινων και μηχανιστικών λειτουργιών. Αυτή η αναλογία συνίσταται στο συσχετισμό του **μυαλού** με τις **λογικές καταστάσεις** ενός υπολογιστή και του **σώματος** με τις **δομικές καταστάσεις** ενός υπολογιστή (συγκεκριμένα το άρθρο ασχολείται με τις μηχανές Turing). Μετά την εισαγωγή αυτής της αναλογίας, μέσω διαφόρων γλωσσολογικών επιχειρημάτων και την κατάδειξη ‘λογικά περιέργων’ [deviant] προτάσεων, ο Putnam φτάνει στο συμπέρασμα ότι το δυϊστικό πρόβλημα δεν αποτελεί ένα αυθεντικό θεωρητικό πρόβλημα ή και αν ακόμα αποτελεί, η ‘λύση’ του δεν θα συνεισφέρει τελικά σε τίποτα απολύτως! Τελικά δηλαδή το όλο πρόβλημα φαίνεται να είναι για τον Putnam μόνο θέμα *λεκτικό* και τίποτα παραπάνω.

Για την ανάλυσή μας βέβαια μας ενδιαφέρει μόνο το κομμάτι της αναλογίας ανθρώπου-μηχανής, αναλογία που για να είναι έγκυρη για την όλη ανάλυση, προϋποθέτει την αποδοχή μίας μορφής της μηχανιστικής θέσης. Διότι, μπορεί ο ίδιος ο Putnam, να δηλώνει ότι δεν υιοθετεί μία μηχανιστική κοσμοθεώρηση, με την έννοια ότι δεν υποστηρίζει ότι οι άνθρωποι είναι μηχανές, ούτε ότι αυτές σκέπτονται, αλλά βιάζεται να διαλύσει γρήγορα τις αμφιβολίες για την αναλογία μυαλού-λογικών καταστάσεων μηχανής για να προχωρήσει με την ανάλυσή του.

Συγκεκριμένα, σε κάποιο σημείο, χωρίς ιδιαίτερη σύνδεση με τα προηγούμενα ή επόμενα επιχειρήματα που αναφέρει, προσπαθεί να καταρρίψει μία θέση των Nagel και Newman που υπάρχει στο βιβλίο τους ‘Gödel’s Proof’ και είναι η εξής: «Το θεώρημα του Gödel υποδεικνύει ότι η δομή και οι δυνατότητες του ανθρώπινου εγκεφάλου είναι πολύ πιο σύνθετες και λεπτοφυείς από οτιδήποτε ανάλογο υπάρχει μέχρι τώρα σαν μηχανή.» Αυτή είναι μία πρωτόλεια διατύπωση του επιχειρήματος του Lucas, και προσπαθεί να καταδείξει την αδυναμία των TM (και κάθε μηχανής) να μπορούν να λειτουργήσουν σαν μοντέλα του ανθρώπινου νου. Φυσικά δεν δίνεται κάποιος συγκεκριμένος τρόπος σύμφωνα με τον οποίο μπορεί να αποδειχθεί κάτι τέτοιο, οπότε πρόκειται στην ουσία περισσότερο για μία *διαισθητική* προσέγγιση του προβλήματος.

Ο Putnam θεωρεί το διαισθητικό αυτό αποτέλεσμα είναι λάθος, και ότι το λάθος οφείλεται σε παρερμηνεία του θεωρήματος του Gödel. Χωρίς να τον απασχολεί ποια μέσα χρησιμοποιούνται για την περιγραφή λογικών συστημάτων με την ορολογία μηχανής και αντίστροφα, (κάτι που επισημαίνει εύστοχα ο Turing ότι θα πρέπει να μας απασχολεί, στο προηγούμενο άρθρο που παρατέθηκε) υποθέτει ότι:

Υπάρχει μία μηχανή Turing, έστω T η οποία με ‘αναπαριστά’, με την έννοια ότι οτιδήποτε μπορεί να αποδείξω εγώ, μπορεί να το αποδείξει και η μηχανή.

Στη συνέχεια, εφόσον οι Nagel και Newman δεν διατυπώνουν ένα σαφές επιχείρημα, ο Putnam θεωρεί πως ισχυρίζονται ότι, με βάση του θεωρήμα του Gödel, μπορώ ‘εγώ’ να βρω μία πρόταση που η T δεν μπορεί να αποδείξει, αλλά εγώ μπορώ, οπότε η T δεν με αναπαριστά, οπότε δεν είμαι μία TM.

Αυτή η θέση απορρίπτεται από τον Putnam με βάση την παρακάτω συλλογιστική που παρατίθεται αυτούσια: «...το μόνο που μπορεί κάποιος να κάνει είναι βρει μία μη-αποκρίσιμη πρόταση, έστω U, και να αποδείξει το ακόλουθο για αυτήν:

Αν η T είναι συνεπής, τότε η U αληθεύει. (1)

Αλλά η μηχανή μας μπορεί επίσης να αποδείξει την (1)! Και δεν μπορεί να αποδείξει την πρόταση U (υποθέτοντας τη συνέπεια), την οποία όμως ούτε εγώ μπορώ να αποδείξω, εκτός εάν μπορώ να αποδείξω την συνέπεια της T, πράγμα απίθανο αν η T είναι πολύπλοκη».

Σημ: Η πρόταση (1) εκπορεύεται από μία συλλογιστική που περιέχεται στο 2^ο θεώρημα μη-πληρότητας και όχι το πρώτο. Συγκεκριμένα, (βλ. Μέρος II, σελ. 53) με την υπόθεση της συνέπειας, που είναι το ζητούμενο προς απόδειξη στο 2^ο θεώρημα μη-πληρότητας, καταλήγουμε στο ότι η T αποδεικνύει την Gödel-πρόταση U, άρα προκύπτει η (1). Στη συνέχεια βέβαια, από αυτό ακριβώς το συμπέρασμα οδηγούμαστε σε άτοπο, αλλά αυτό δεν φαίνεται να απασχολεί τον Putnam.

Η παραπάνω απάντηση θεωρείται η πλέον ‘κλασική’ για να καταρρίψει κανείς επιχειρήματα σαν αυτό του Lucas. Ο Lucas όμως δεν ισχυρίζεται ακριβώς αυτό που υποθέτει ο Putnam, αλλά θα δώσει ένα διαφορετικό επιχείρημα του οποίου βέβαια η εγκυρότητα είναι άκρως συζητήσιμη σε πολλές ακόμα βάσεις, όπως θα δούμε

παρακάτω. Όσον αφορά το ίδιο το επιχείρημα, ο Putnam φαίνεται ότι *καταρρίπτει τη δική του εκδοχή του επιχειρήματος*, το οποίο όμως είχε εξ' αρχής στο μυαλό του να απορρίψει, οπότε προκύπτει ένα θέμα αξιοπιστίας. Αν δεχθούμε πάντως την ανάλυσή του, αυτή καθ' εαυτή η αξιοπιστία του επιχειρήματος έγκειται στο να ορίσει κανείς τις αποδεικτικές δυνατότητες του ανθρώπου. Διότι αν αυτές δεν μπορούν να ξεπεράσουν αυτές ενός τυπικού συστήματος (άρα και του T), τότε φυσικά ισχύει a priori ότι δεν μπορούμε να αποδείξουμε τίποτα παραπάνω από το T! Για μερικούς βέβαια, το ότι μπορούμε να αποδείξουμε γενικά περισσότερα πράγματα για ένα τυπικό σύστημα από όσα μπορούν να αποδειχθούν μέσα σε αυτό το σύστημα είναι άμεση συνέπεια των θεωρημάτων του Gödel. Για τη συγκεκριμένη λοιπόν περίπτωση, το θέμα είναι αν μπορεί ο άνθρωπος να διαπιστώσει την αλήθεια της U ή όχι και αν μπορεί να το κάνει, σε ποια πλαίσια αποδεικτικής δυνατότητας μπορεί αυτό να γίνει. Το όλο θέμα είναι κεφαλαιώδους σημασίας για όλη την ανάλυση που ακολουθεί, και παρακάτω, έχοντας περισσότερα δεδομένα θα μπορέσουμε να πούμε περισσότερα γι' αυτό. Όσον αφορά την απάντηση του Lucas στην αντίρρηση του Putnam ή μάλλον σε μία αναδιατύπωσή της που αφορά άμεσα το βασικό του επιχείρημα, αυτή θα τη δούμε στην παράγραφο 3.3, αμέσως μετά την ανάλυση του άρθρου του Lucas.

Επιστρέφοντας στο άρθρο του Putnam, δεδομένου του αντεπιχειρήματός του, θεωρεί τελικά ότι οι Nagel και Newman έχουν λάθος, οπότε μπορούμε με ασφάλεια να υποθέσουμε ότι ισχυρίζεται ότι όντως μία TM μπορεί να λειτουργήσει σαν μοντέλο του νου.

Η ευρύτερη αντιμετώπισή του όμως σε σχέση με το Μηχανισμό περιπλέκεται ακόμη παραπάνω όταν φτάνει στο τέλος των συμπερασμάτων του, οπότε αναφέρει τα εξής: «...αν το πρόβλημα μυαλού-σώματος αναγνωρίζεται σαν ένα πρόβλημα που είναι παραπάνω από θεμελιώδες, τότε μόνο τρεις εναλλακτικές υπάρχουν: είτε (α) *κανένα* επιχείρημα που χρησιμοποιήθηκε ποτέ από φιλόσοφους δεν μπορεί να ρίξει το *παραμικρό* φως σε αυτή την υπόθεση, είτε (β) κάποιο φιλοσοφικό επιχείρημα για τη μηχανιστική θέση είναι έγκυρο, ή (γ) κάποιο δυϊστικό επιχείρημα δείχνει ότι *και* οι άνθρωποι *αλλά και* οι μηχανές Turing έχουν ψυχή!».

Καθότι όμως, όπως διαπιστώσαμε στην αρχή, ο ίδιος πιστεύει ότι το εν λόγω πρόβλημα δεν είναι τίποτα παραπάνω από θέμα θεμελιωδών συμβάσεων της γλώσσας, είναι σα να μας δείχνει ότι καμία από τις τρεις εναλλακτικές δεν μπορεί να θεωρηθεί αξιόπιστη. Η επιχειρηματολογία του όμως βασίζεται στην αναλογία

ανθρώπου-μηχανής, για την οποία έχει υποθέσει ότι οι TM μπορούν να λειτουργήσουν σαν μοντέλα του νου, όπως είδαμε παραπάνω. Αλλά αυτό είναι μια μορφή έγκυρου μηχανιστικού επιχειρήματος, άρα η εναλλακτική (β) είναι εφικτή, άρα η αρχική του θέση τίθεται υπό αμφισβήτηση! Από την άλλη, αν δεχθούμε την αντίθετή της, δηλαδή ότι όντως το πρόβλημα μυαλού-σώματος είναι παραπάνω από θεμελιώδες και έχει φιλοσοφική αξία, τότε αμέσως απορρίπτεται η εναλλακτική (α) διότι αποτελεί μέρος της επιχειρηματολογίας του Putnam πάνω σε αυτό ακριβώς το θέμα. Εφόσον όμως η (α) λέει ότι *κανένα* επιχείρημα δεν μπορεί να ρίξει το παραμικρό φως στην υπόθεση μυαλού-σώματος, και αυτή είναι μία διαπίστωση για αυτή την υπόθεση, πρέπει να την απορρίψουμε διότι *αυτή η διαπίστωση ρίχνει όντως κάποιο φως* στο δυϊστικό πρόβλημα! Άρα μένουμε με τις εναλλακτικές (β) και (γ), οι οποίες όμως αποτελούν επίσης θέσεις της ευρύτερης μηχανιστικής θεώρησης.

Με βάση τα παραπάνω, φαίνεται ότι ο Putnam, παρ' ότι ισχυρίζεται ότι δεν ασπάζεται τη μηχανιστική κοσμοθεωρία, δεν μπορεί να αποφεύγει την αποδοχή τουλάχιστον μίας μορφής της μηχανιστικής θέσης. Επιπλέον, οι αντιφάσεις που διαφαίνονται στα τελικά συμπεράσματα φαίνεται να δείχνουν πως η αναλογία ανθρώπου-μηχανής που επιχειρείται και φαίνεται να συνάδει με τον Μηχανισμό, δεν υποστηρίζεται με αξιόπιστο τρόπο. Περισσότερες σκέψεις πάνω στο θέμα της θέσης του Putnam σίγουρα μπορούν να γίνουν αλλά δεν είναι αυτό το αντικείμενο της εργασίας αυτής. Κρατάμε λοιπόν από το άρθρο κυρίως την επιχειρηματολογία που έχει να κάνει με το θεώρημα του Gödel, και προχωρούμε να διαπιστώσουμε ποια ακριβώς είναι η θέση του Lucas πάνω στο θέμα, η οποία είναι και το κύριο σημείο που μας ενδιαφέρει.

3. Το βασικό επιχείρημα του J.R. Lucas

3.0 Ένα σχόλιο πάνω στη χρήση του θεωρήματος μη-πληρότητας και ένα συναφές παράδειγμα

Πριν αναλυθεί το άρθρο του Lucas που περιέχει το βασικό του επιχείρημα ενάντια στη μηχανιστική θέση, πρέπει να ξεκαθαριστεί η χρήση που γίνεται στο αποτέλεσμα του Gödel, από τον Lucas και άλλους (κυρίως) φιλόσοφους. Φαίνεται ότι συστηματικά χρησιμοποιείται σε εκλαϊκευμένες αποδόσεις του θεωρήματος, η έκφραση:

Η Gödel-πρόταση λέει για τον εαυτό της ότι είναι μη αποδείξιμη. (*)

Αυτό, όπως έχουμε δει στην τυπική απόδειξη δεν είναι απολύτως ακριβές, διότι στην πραγματικότητα, η μετα-πρόταση που προκύπτει από το λήμμα της αυτό-αναφοράς εκφράζει το γεγονός ότι η Gödel-πρόταση είναι αληθής αν και μόνον αν δεν είναι αποδείξιμη. Παρ' όλα αυτά, η πιο απλή διατύπωση (*) χρησιμοποιείται και από τον ίδιο τον Gödel στο άρθρο που αποδεικνύει τα θεωρήματα μη-πληρότητας, οπότε δεχόμαστε ότι δεν ξεφεύγει από την κεντρική ιδέα του θεωρήματος. Την ίδια παραδοχή κάνουμε και για τις παραλλαγές της απόδειξης που βασίζονται στο διαγώνιο επιχείρημα, που επίσης δεν είναι τυπικά απόλυτα σωστές, αλλά βασίζονται στην ιδέα από την οποία έτσι κι αλλιώς προέκυψε το θεώρημα μη-πληρότητας.

Να σημειώσουμε ακόμη, ότι, θα αναφερόμαστε εφεξής στους όρους 'μηχανιστική θέση' και 'Μηχανισμός' αποκλειστικά με βάση τους αρχικούς ορισμούς που δόθηκαν στην εισαγωγή και όχι με βάση την εκδοχή του Turing. Άλλωστε, όπως έγινε φανερό και από την ανάλυση του άρθρου του Putnam, τα επιχειρήματα που παρατίθενται σαν κύριο στόχο έχουν την απάντηση του αρχικού ερωτήματος δηλαδή το αν οι μηχανές σκέπτονται. Ισοδύναμο είναι επίσης το ερώτημα *αν μπορεί να υπάρξει μία μηχανή που να αποτελεί μοντέλο του ανθρώπινου νου*, με την έννοια ότι μπορεί να προσομοιώσει όλες τις λειτουργίες του.

Μετά από αυτά τα πρώτα σχόλια, παραθέτουμε τώρα μία διαγώνια διαδικασία απόδειξης του θεωρήματος του Gödel, η οποία βρίσκεται στο βιβλίο του Rudy Rucker 'Infinity and the mind'. Η συγκεκριμένη επιλογή έγινε κυρίως λόγω της διαλεκτικής φύσεως της διαδικασίας, για να γίνει μία πρώτη επαφή με τα διαλεκτικά επιχειρήματα που θα μας απασχολήσουν εκτενέστερα παρακάτω.

**Μία διαλεκτική διαδικασία απόδειξης του 1^{ου} θεωρήματος μη-
πληρότητας του Gödel, από τον R. Rucker**

1. Κάποιος συστήνει στον Gödel μία ΠΑΜΑ, μία μηχανή που θεωρείται ότι είναι ΠΑγκόσμια Μηχανή Αλήθειας, ικανή να απαντήσει σωστά *οποιαδήποτε* ερώτηση.
2. Ο Gödel ζητά το πρόγραμμα και το σχεδιασμό των κυκλωμάτων της ΠΑΜΑ. Το πρόγραμμά της μπορεί να είναι πολύπλοκο αλλά δεν μπορεί να είναι, στην πράξη, παρά μόνο πεπερασμένο. Ας συμβολίσουμε το πρόγραμμα αυτό P(ΠΑΜΑ).
3. Χαμογελώντας, ο Gödel γράφει την εξής πρόταση: «Η μηχανή που κατασκευάστηκε με βάση το P(ΠΑΜΑ) ποτέ δεν θα πει ότι αυτή η πρόταση είναι αληθής». Ας καλέσουμε αυτή την πρόταση G. *Σημειώστε επίσης ότι η G είναι ισοδύναμη με την πρόταση: «η ΠΑΜΑ ποτέ δεν θα πει ότι η G είναι αληθής».*
4. Τώρα ο Gödel γελάει θριαμβευτικά και ρωτάει την ΠΑΜΑ αν η G είναι αληθής ή όχι.
5. Αν η ΠΑΜΑ πει ότι η G είναι αληθής, τότε η πρόταση «η ΠΑΜΑ ποτέ δεν θα πει ότι η G είναι αληθής», είναι ψευδής. Αν όμως ισχύει αυτό, τότε και η G θα είναι ψευδής (αφού $G = \text{«η ΠΑΜΑ ποτέ δεν θα πει ότι η G είναι αληθής»}$). Άρα, αν η ΠΑΜΑ πει ότι η G είναι αληθής, τότε προκύπτει ότι η G είναι ψευδής, και η ΠΑΜΑ έχει διαπιστώσει το ψεύδος της. Άρα η ΠΑΜΑ ποτέ δεν θα πει ότι η G είναι αληθής, αφού μπορεί να επιβεβαιώσει μόνο αληθείς προτάσεις.
6. Έχουμε σιγουρευτεί ότι η ΠΑΜΑ δεν θα πει ποτέ ότι η G είναι αληθής. Άρα η πρόταση που λέει «η ΠΑΜΑ ποτέ δεν θα πει ότι η G είναι αληθής» είναι αληθής πρόταση. Άρα η G είναι αληθής, αφού η G είναι ακριβώς η παραπάνω πρόταση.
7. ‘Ξέρω μία αλήθεια που η ΠΑΜΑ δεν μπορεί να την εκφράσει’ λέει ο Gödel. ‘Ξέρω ότι η G είναι αληθής, άρα η ΠΑΜΑ δεν είναι πραγματικά παγκόσμια’.

Μπορεί να διακρίνει κανένας σε αυτή την αλληλουχία προτάσεων, τόσο το παράδοξο του ψεύτη, όσο και την απόδειξη του Gödel για το θεώρημα μη-πληρότητας. Αυτό που κάνει τον Rucker, και όσους χρησιμοποιούν ανάλογα επιχειρήματα, να αισθάνονται ασφαλείς ότι αυτά που λένε δεν είναι απλά γλωσσικές παραδοξολογίες, είναι το γεγονός ότι πίσω από τέτοιου τύπου επιχειρηματολογίες υπάρχει ένα αποδεκτό επιστημονικό υπόβαθρο, η τυπική απόδειξη του θεωρήματος του Gödel. Το πόσο μακριά μπορεί βέβαια κανείς να το τραβήξει με άλλοθι τον Gödel είναι ένα εύλογο και συνάμα δύσκολο ερώτημα.

Παρ' όλα αυτά, τελικά για τον Rucker το θεώρημα μας δείχνει ότι η λογική σκέψη ποτέ δεν θα φτάσει την απόλυτη αλήθεια, μία ιδέα που θα την κρατήσουμε.

3.1 J.R. Lucas. 'Minds, machines and Gödel' (1961)

Ο τίτλος του κεφαλαίου είναι και ο τίτλος του άρθρου στο οποίο για πρώτη φορά δημοσιεύτηκε με σαφήνεια το μαθηματικό επιχείρημα που έχει σα βάση το θεώρημα του Gödel και σκοπό να καταρρίψει τη βασική θέση του Μηχανισμού. Με τα λόγια του ίδιου του Lucas: «Το θεώρημα του Gödel φαίνεται σε μένα ότι αποδεικνύει ότι ο Μηχανισμός έχει λάθος, δηλαδή ότι το ανθρώπινο μυαλό δεν μπορεί να εξηγηθεί σαν μία μηχανή». Ο τελικός σκοπός του είναι να διατυπώσει ένα ολοκληρωμένο επιχείρημα, απαντώντας στις κυριότερες αντιρρήσεις που μπορεί να προκύψουν, ώστε να διαμορφώσει ένα πλαίσιο συζήτησης που να εκφράζει με σαφήνεια αυτό που ο ίδιος αλλά και άλλοι πιστεύουν ότι δείχνει το θεώρημα όσον αφορά τα όρια 'σκέψης' των μηχανών σε σχέση με τη σκέψη του ανθρώπου.

Ο Lucas χρησιμοποιεί την έκφραση (*) του θεωρήματος που περιγράφηκε στο προηγούμενο κεφάλαιο, ήτοι ότι σε κάθε συνεπές τυπικό σύστημα δεν μπορεί να αποδειχθεί ο τύπος που λέει ότι 'αυτός ο τύπος είναι μη-αποδείξιμος στο σύστημα αυτό', και άρα αυτός δεν είναι αληθής για το σύστημα. Αυτόν όμως τον τύπο μπορούμε εμείς, που είμαστε έξω από το σύστημα, να δούμε σαν αληθή (βλ. Μέρος II, σελ. 51) και σε αυτή ακριβώς τη δυνατότητα θα βασιστεί όλη η ανάλυση του άρθρου. Είναι χαρακτηριστικό ότι χρησιμοποιείται η έκφραση 'βλέπω να είναι αληθές' αντί του 'αποδεικνύω', για να αποτυπωθεί ακριβώς ότι δεν υπάρχει απόδειξη τυπική για την αλήθεια της Gödel-πρότασης. Άλλωστε, αν υπήρχε, αυτή δεν θα ήταν μη-αποκρίσιμη, αλλά αληθής στα πλαίσια του συστήματος.

Μετά από αυτή τη σύντομη παρουσίαση, ο Lucas προσπαθεί να δείξει γιατί ακριβώς θα πρέπει το θεώρημα του Gödel να ισχύει για τις μηχανές, και συγκεκριμένα για τα ‘αυτόματα’ [automata], ο ευρύτερος ορισμός των οποίων περιλαμβάνει και τις μηχανές Turing. Συγκεκριμένα, ισχυρίζεται ότι, επί της ουσίας, το να θεωρείται κάτι σαν μηχανή είναι να αποτελεί ένα ‘συμπαγές στιγμιότυπο’ [concrete instantiation] ενός τυπικού συστήματος. Αυτή η θέση φαίνεται ότι είναι γενικά αποδεκτή και αποτελεί έναν τρόπο να ξεπεραστεί η δυσκολία σύνδεσης τυπικών συστημάτων και μηχανών που προβάλλει ο Turing σαν ένα θέμα που θα πρέπει να λυθεί πριν αποκτήσει βάση οποιοδήποτε επιχείρημα που πηγάζει από τη μαθηματική λογική. Έχοντας λοιπόν ο Lucas τη βάση του καλυμμένη, μπορούμε να εκφράσουμε το βασικό επιχείρημά του ως εξής:

Επιχείρημα του Lucas: Για κάθε μηχανή που περιλαμβάνει την απλή αριθμητική και είναι συνεπής, θα υπάρχει ένας τύπος που δεν θα μπορεί να παραχθεί αληθής από αυτή (συγκεκριμένα η μη-αποκρίσιμη πρόταση του Gödel). Κάθε άνθρωπος όμως, ακολουθώντας τα συλλογιστική πίσω από το θεώρημα του Gödel, θα μπορεί να διαπιστώσει την αλήθεια της. Άρα καμία μηχανή δεν μπορεί να αποτελεί ένα πλήρες μοντέλο του νου και άρα, *ο νους είναι στην ουσία του διαφορετικός από μία μηχανή.*

Η παραπάνω διατύπωση αποτελεί το βασικό επιχείρημα του άρθρου. Βέβαια για να γίνει πλήρως αποδεκτή μία τέτοια θέση χρειάζεται μία πλήρης και εμπεριστατωμένη επιπλέον ανάλυση, πράγμα που επιχειρείται στο υπόλοιπο του άρθρου. Από αυτά τα επιπλέον σημεία που επισημαίνονται για να ισχυροποιηθεί το βασικό επιχείρημα, θα δούμε τα σημαντικότερα αμέσως παρακάτω.

Καταρχήν επιχειρείται η αποσαφήνιση της έννοιας της μηχανής σαν μία συσκευή που αποτελείται από ένα σύνολο κανόνων που καθοδηγούν απόλυτα τη λειτουργία της. Ανάλογα, ένα μηχανοποιημένο μοντέλο για το νου, τον θεωρεί απόλυτα καθορισμένης λειτουργίας, με την έννοια ότι εισάγοντάς του κάποιο δεδομένο, θα μπορούμε να ξέρουμε το ακριβές εξαγόμενο αποτέλεσμα, αν μπορούμε να διαθέσουμε για το σκοπό αυτό αρκετό χρόνο. Αυτή η δυνατότητα απόλυτης περιγραφής της λειτουργίας, θεωρείται από τον Lucas σχεδόν συνώνυμο των μηχανών και δεν χάνεται ούτε με την εισαγωγή τυχαίων στοιχείων σε αυτές. «Οι μηχανές είναι καθορισμένες [definite]: Οτιδήποτε είναι ασαφές [indefinite] ή άπειρο [infinite], δεν θα πρέπει να θεωρείται μηχανή», όπως ακριβώς γράφει σε ένα σημείο.

Σημειώστε σε αυτό το σημείο ότι με αυτή τη θεώρηση, ο Lucas ουσιαστικά απορρίπτει το δεύτερο από τα τρία κριτήρια που είχε προβάλλει ο Turing για μία μηχανή (βλ. παραπάνω, σελ. 70). Αυτή η απόρριψη φαίνεται ότι σιωπηλά γίνεται αποδεκτή από όλους, δεν δεχόμαστε δηλαδή, με βάση τις αρχές που έχουμε στο μυαλό μας ότι διέπουν τη λειτουργία μίας μηχανής, ότι *πραγματικά* δεν μπορούμε να ξέρουμε κάθε λειτουργία της. Γι' αυτό το λόγο και δε συναντήθηκε σε καμία ανάλυση από όσες ακολουθήσουν, καμία αντίδραση που να αφορά αυτό το κομμάτι ορισμού των μηχανών που μόλις περιγράφηκε.

Με αυτά λοιπόν τα αρχικά δεδομένα, μπορούμε να φτιάξουμε μία μηχανή που θα παράγει όλες τις αληθείς προτάσεις της αριθμητικής μέσω τυπικών αποδείξεων και για κάθε τέτοια απόδειξη να μπορεί καθένας να ακολουθήσει και καταγράψει ολοκληρωμένη την πορεία των απαγωγών μέσω των οποίων η μηχανή καταλήγει στο αποτέλεσμα. Επιπλέον, κάθε υποψήφιο μοντέλο για το νου θα μπορεί να κάνει κάτι ανάλογο, εφόσον η παραγωγή μαθηματικών αποδείξεων είναι μία δεδομένη ανθρώπινη ιδιότητα. Άρα το βασικό επιχείρημα μπορεί να διατυπωθεί για αυτό το μοντέλο, οπότε να έχουμε στα χέρια μας μία αλήθεια που δεν μπορεί να την παράγει η μηχανή και ναδειχθεί ότι τελικά ο νους είναι διαφορετικός από τη μηχανή. Όπως γράφει και ο Lucas, «δεν είναι ότι δε μπορούμε να κατασκευάσουμε μία μηχανή που να προσομοιώνει *οποιαδήποτε* επιθυμητή ιδιότητα του νου, απλά, δεν μπορούμε να κατασκευάσουμε μία μηχανή που να προσομοιώνει *κάθε* ιδιότητα του νου».

Ο Lucas βέβαια αναγνωρίζει πως αυτό το συμπέρασμα μπορεί να ξενίζει αρκετά ή και να δημιουργεί υποψίες όσον αφορά την αξιοπιστία του, οπότε σπεύδει να απαντήσει σε πιθανές αντιρρήσεις. Η πρώτη αντίρρηση που μπορεί να απαντηθεί άμεσα είναι αυτή που αναφέρεται στην τελευταία διατύπωση του συμπεράσματος. Όπως εύστοχα επισημαίνεται, εφόσον μέσω μίας έγκυρης μαθηματικής συλλογιστικής μπορούμε να διαπιστώσουμε ότι για οποιοδήποτε φυσικό αριθμό μπορεί να βρεθεί ένας μεγαλύτερός του, αλλά δεν υπάρχει ένας φυσικός που να είναι μεγαλύτερος από όλους τους άλλους, ανάλογα έγκυρη είναι και η διαπίστωση της προηγούμενης παραγράφου.

Στη συνέχεια απαντάται συνολικά το ερώτημα αν μπορεί να φτιαχτεί μία μηχανή που μπορεί να καλύπτει το ελάττωμα που βρέθηκε στην πρώτη για την οποία διατυπώθηκε το μαθηματικό επιχείρημα. Η απάντηση βασίζεται στο γεγονός ότι το θεώρημα του Gödel ισχύει *για κάθε* αξιωματοποιημένη επέκταση ενός τυπικού συστήματος, οπότε ακόμα και αν εντάξουμε την Gödel-πρόταση στα αξιώματα του

συστήματος, το σύστημα που θα προκύψει θα συνεχίζει να είναι μη-πλήρες. Έτσι, αναλόγως, για μία μηχανή που θα ενσωματώνει την μη-αποκρίσιμη πρόταση στις αληθείς προτάσεις που μπορεί να παράγει, θα υπάρχει μία νέα Gödel-πρόταση για αυτή τη μηχανή και άρα θα ισχύει και πάλι το μαθηματικό επιχείρημα. Αυτή η διαδικασία μπορεί να συνεχιστεί μέχρι το 'άπειρο', ή μάλλον ακόμα και αν προσθέσουμε άπειρα νέα αξιώματα, και πάλι δεν λύνεται το πρόβλημα⁽¹⁾.

Με αυτή την τελευταία παρατήρηση ο Lucas απαντά και στην αντίρρηση που λέει ότι εφόσον η εύρεση μη-αποκρίσιμων προτάσεων είναι μία σταθερή διαδικασία, θα μπορούσαμε να την προγραμματίσουμε μέσα στη μηχανή-μοντέλο του νου. Εδώ προκύπτει ένα ενδιαφέρον θέμα. Αν υπάρχει αυτή η δυνατότητα, τότε φαίνεται ότι μία τέτοια μηχανή δεν θα μπορούσε καν να υπάρξει. Διότι η πρόσθεση αξιωμάτων δεν θα τελείωνε ποτέ, άρα πρακτικά, δεν θα ολοκληρωνόταν ποτέ μία τέτοια μηχανή! Ο Lucas υποθέτει ότι θα μπορούν να προστεθούν σε αυτή άπειρα αξιώματα, αλλά η αντίρρηση δεν λέει αυτό. Λέει ότι θα υπάρχει μία διαδικασία που θα προσθέτει αξιώματα, και αυτή είναι που θα είναι άπειρη, οπότε η απάντηση δεν είναι απόλυτα αξιόπιστη. Κατά τη γνώμη του γράφοντος, πιο πειστική θα ήταν μία απάντηση που θα επικεντρωνόταν στη δυνατότητα κατασκευής ενός τέτοιου προγράμματος. Πώς γίνεται μπορούμε να κατασκευάσουμε μία μηχανή που θα παίρνει σαν δεδομένο το ίδιο της το πρόγραμμα και θα το επεξεργάζεται; Και το κομμάτι αυτό που θα επεξεργάζεται τα δεδομένα, δεν θα αποτελεί το ίδιο μέρος της επεξεργασίας;

Με βάση όλα τα παραπάνω, φαίνεται να καταρρίπτονται τα επιχειρήματα του Turing για την αδυναμία του ανθρώπου να υπερσχύσει όλων των μηχανών ταυτόχρονα και ότι απλά η κάθε υπερίσχυση είναι δείγμα μικροψυχίας. Διότι η συλλογιστική του Lucas συνίσταται κυρίως στην κατάδειξη της εγγενούς *διαφορετικότητας* του ανθρώπινου νου από τις μηχανές και όχι της ανωτερότητάς του. Εκτός αυτού, το θέμα δεν είναι η αντιμετώπιση όλων των μηχανών ταυτόχρονα, αλλά της ιδέας που κρύβεται πίσω από τη μηχανιστική θέση. Το επιχείρημα δηλαδή είναι στην ουσία του **διαλεκτικό** και στην αρχική του μορφή κινείται στο πλαίσιο που περιγράφεται στην επόμενη σελίδα:

(1): Για μία αντίρρηση πάνω στα όρια αυτής της διαδικασίας, βλ. § 4.1.2 του ίδιου μέρους.

Τον πρώτο λόγο έχει ο Μηχανιστής που παράγει μία μηχανή και ισχυρίζεται ότι αυτή αποτελεί ένα μοντέλο του νου. Τότε ο Lucas (και ο οποιοσδήποτε 'Lucas'), δείχνει ότι μπορεί να κάνει κάτι που η μηχανή αυτή δεν μπορεί να το κάνει, άρα δεν αποτελεί μοντέλο του δικού του, και άρα συνολικά κανενός νου. Ο Μηχανιστής μπορεί στη συνέχεια να παράγει μία νέα μηχανή ή να αλλάξει την ήδη υπάρχουσα, αλλά ο Lucas έχει το δικαίωμα να ελέγξει τη νέα μηχανή για αδυναμίες. Αν δεν μπορεί να βρεθεί καμία τέτοια αδυναμία, τότε η μηχανιστική θέση ισχύει. Αλλά αν για κάθε νέα μηχανή μπορεί να βρεθεί και μία νέα αδυναμία, όπως και φαίνεται να ισχύει με βάση το μαθηματικό επιχείρημα, τότε η μηχανιστική θέση δεν ισχύει και άρα συνολικά ο Μηχανισμός έχει λάθος.

Αυτή η μορφή του επιχειρήματος είναι που φαίνεται ότι αντικατοπτρίζει καλύτερα τις απόψεις του Lucas, διότι είναι αυτή που χρησιμοποιεί κατά κόρον σε επόμενα άρθρα του, με τα οποία απαντά στις επιπλέον αντιρρήσεις που εκφράζονται για την εγκυρότητα του επιχειρήματός του.

Στη συνέχεια, και για μεγάλο μέρος του άρθρου, ο Lucas ασχολείται με βαθύτερες αντιρρήσεις, που περισσότερο στοχεύουν στην ιδέα πίσω από κάθε χρήση μαθηματικών επιχειρημάτων ενάντια στο μηχανισμό. Για παράδειγμα, υπάρχει η ένσταση που λέει ότι εφόσον οι άνθρωποι δεν βασίζονται μόνο στην αφαιρετική μέθοδο για την εξαγωγή των συμπερασμάτων τους, μία μηχανή-μοντέλο του νου θα πρέπει να περιλαμβάνει και μη αφαιρετική σκέψη. Έτσι όμως, αφού το θεώρημα του Gödel αναφέρεται σε αφαιρετικά συστήματα, δεν θα υπήρχε πρόβλημα με τις μη-αποκρίσιμες προτάσεις. Αυτές θα αντιμετωπίζονταν από το μη αφαιρετικό κομμάτι της μηχανής. Μία τέτοια κατασκευή βέβαια μπορεί να οδηγήσει σε απαράδεκτα αποτελέσματα, όπως το να δεχθεί η μηχανή σαν αληθή όχι την Gödel-πρόταση, αλλά την άρνησή της, κάτι που αντιβαίνει με τον ισχυρισμό ότι αυτή η μηχανή αποτελεί μοντέλο για το νου.

Από τις παραπάνω απόψεις, γίνεται εμφανές ότι ο Lucas έχει μία τάση να μειώνει την ιδέα της εισαγωγής άλλων μεθόδων πλην της αφαιρετικής στις μηχανές, προσπαθώντας να δείξει ότι έτσι θα καταλήγαμε σε γελοία αποτελέσματα. Υπάρχει όμως μία εμπεριστατωμένη θεωρία πίσω από αυτή την ιδέα, σαν αυτή που αναλύσαμε στο 2^ο μέρος της εργασίας, στις φιλοσοφικές συνέπειες του θεωρήματος (βλ. Μέρος

II, § 4.1, σελ. 58-59), ενώ υπάρχει και μία αντίρρηση που δείχνει ότι, στην πράξη, όντως χρησιμοποιούνται άλλες μέθοδοι συλλογισμού στον τομέα της έρευνας της Τεχνητής Νοημοσύνης (βλ. παρακάτω, σελ. 96). Σίγουρα βέβαια η ορθή ενσωμάτωση τέτοιων στοιχείων μπορεί να αποδειχθεί εξαιρετικά δύσκολη, αλλά ίσως να οδηγούσε σε ενδιαφέροντα αποτελέσματα, που δεν θα ήταν απαραίτητα παράλογα.

Μία ακόμη βασική αντίρρηση αφορά την έννοια της συνέπειας. Ως γνωστόν, το θεώρημα του Gödel ισχύει μόνο για συνεπή συστήματα. Πως είμαστε όμως τόσο σίγουροι ότι η μηχανή-μοντέλο του νου είναι συνεπής, ώστε να εφαρμόζεται σε αυτή το μαθηματικό επιχείρημα;⁽²⁾ Και κατά συνέπεια, πως μπορούμε να ξέρουμε ότι οι άνθρωποι είναι συνεπείς; Υπάρχει περίπτωση να είμαστε μηχανές, απλά να είμαστε μη-συνεπείς μηχανές, όπως είχε προτείνει ο Putnam σε μία συζήτηση με τον Lucas;

Στην προσπάθεια απάντησης αυτών των ερωτημάτων, αναπόφευκτα εμπλέκεται και το αποτέλεσμα του 2^{ου} θεωρήματος μη-πληρότητας. Σύμφωνα με αυτό, μία μηχανή-συμπαγές στιγμιότυπο ενός τυπικού συστήματος, δεν μπορεί να αποδείξει η ίδια τη συνέπειά της. Μπορεί ίσως να το κάνει ο άνθρωπος, αλλά σε περίπτωση μεγάλης πολυπλοκότητας ίσως αυτό είναι αδύνατο να γίνει. Εδώ εγείρεται ένα νέο ερώτημα, το κατά πόσο ο άνθρωπος έχει δικαίωμα να ζητάει τη συνέπεια από μία μηχανή, εφόσον ο ίδιος ίσως να μην μπορεί να αποδείξει τη δικιά του! Ο Lucas αντιτάσσει ότι ο άνθρωπος μπορεί να υπόκειται συχνά σε λάθη, όμως δεν ανέχεται την ασυνέπεια και προσπαθεί να την απαλείψει όπου αυτή εμφανίζεται. Πάνω σε αυτή τη βάση γίνεται μία εκτενής ανάλυση, η οποία όμως φαίνεται να καταλήγει τελικά ότι η συνέπεια είναι θέμα ‘τετριμμένο’ για τον άνθρωπο, κάτι σαν το γεγονός ότι ξέρει ότι υπάρχει, απλά δεν μπορεί να το αποδείξει με τυπικό τρόπο. Με τα λόγια του ίδιου του Lucas: «Τα παράδοξα της συνείδησης προκύπτουν από το γεγονός ότι ένα ενσυνείδητο ον μπορεί να έχει αντίληψη του εαυτού του, όπως και άλλων πραγμάτων, αν και δεν μπορεί πραγματικά να ερμηνευτεί σαν χωρισμένο σε πολλά μέρη. Αυτό σημαίνει ότι ένα ενσυνείδητο ον μπορεί να πραγματεύεται αυτοαναφορικές ερωτήσεις με έναν τρόπο που καμία μηχανή δεν μπορεί, διότι ένα ενσυνείδητο ον μπορεί να σκέφτεται για τον εαυτό του και τις λειτουργίες του και ταυτόχρονα να μην είναι άλλο από αυτό που έκανε αυτή τη σκέψη...δεν χρειάζεται κάτι παραπάνω για να το κάνει αυτό: είναι ήδη πλήρες και δεν έχει ‘Αχίλλειο πτέρνα’».

(2): Μία απάντηση σε αυτό δίνεται σε ένα μεταγενέστερο άρθρο του Lucas (βλ. σελ.109) αλλά δεν είναι απόλυτα ικανοποιητική.

Αυτή η κατάληξη της ανάλυσης του Lucas μπορεί να αποτελεί ένα πολύ καλό γενικό επιχείρημα για τη διαφορά ανθρώπου και μηχανής, με βάση την ύπαρξη συνείδησης στον άνθρωπο, αλλά αν σκεφτούμε από πού ξεκίνησε, φαίνεται να ξεφεύγει αρκετά από το αρχικό πλαίσιο. Αυτό συμβαίνει διότι εξ' αρχής, η συνέπεια είναι μία έννοια δομική και το 2^ο θεώρημα μη-πληρότητας δε μπορεί να μας διαφωτίσει για τη φύση της. Απλά επισημαίνει την αδυναμία απόδειξής της στα πλαίσια ενός τυπικού συστήματος, με την προϋπόθεση ότι αυτό είναι όντως συνεπές. Δομούμε δηλαδή ένα συνεπές σύστημα και μετά προσπαθούμε να δούμε τι δυνατότητες έχει. Αυτό δεν σημαίνει ότι έχουμε αμφιβολίες για τη συνέπειά του, αυτή θεωρείται δεδομένη αφού στη βάση της έχουμε δομήσει όλο το σύστημα. Παράδοξα όπως αυτό της συνολοθεωρίας του Cantor, μπορεί να προκύψουν, αλλά και πάλι δεν απορρίπτουμε συνολικά μία θεωρία σαν μη-συνεπή, αλλά προσπαθούμε να την αναδομήσουμε ώστε να εξαλειφθούν οι αντιφάσεις. Και φυσικά έχουμε κάθε δικαίωμα να κάνουμε κάτι τέτοιο, χωρίς να χρειάζεται να αναφερθούμε στη συνέπεια της ανθρώπινης συλλογιστικής. Τελικά, κατά τη γνώμη μου, το 'λάθος' στην ανάλυση του Lucas περί της συνέπειας, είναι ότι προσπαθεί να αναλύσει τη (δεδομένη) φύση της χρησιμοποιώντας όμως ένα αποτέλεσμα που αναφέρεται σε ένα επιφανόμενο της. Δεν χρειάζεται όμως το αποτέλεσμα του Gödel για να μιλήσει κανείς γενικά για τη συνέπεια και αυτό φαίνεται από την αμιγώς φιλοσοφική φύση του τελικού συμπεράσματός του.

Αυτά είναι τα κυριότερα σημεία του άρθρου του J.R. Lucas μαζί με κάποια πρώτα σχόλια που αφορούν το θέμα που μας απασχολεί. Όπως θα διαπιστώσουμε στη συνέχεια, για κάθε ένα από τα επιχειρήματα που αναλύθηκαν υπάρχουν ποικίλες διαφωνίες, άλλες εύστοχες και άλλες λιγότερο πειστικές. Μέσα από μία σειρά άρθρων θα δούμε τις περισσότερες από αυτές, όπως και τις ανταπαντήσεις του Lucas ανά περίπτωση. Θα διαπιστώσουμε έτσι πως το βασικό επιχείρημα του Lucas είναι αρκετά προσεκτικά διατυπωμένο, ώστε να διατηρεί το ενδιαφέρον του ακόμα και μετά τη διατύπωση όλων των πιθανών αντιρρήσεων.

Όσον αφορά τις έμμεσες αντιρρήσεις των Turing και Putnam, που εκφράστηκαν πριν από τη δημοσίευση του άρθρου του Lucas, σκιαγραφήθηκαν ήδη οι απαντήσεις στους προβληματισμούς του Turing και στο επόμενο κεφάλαιο θα δούμε τον τρόπο με τον οποίο απαντάται η 'κλασική' αντίρρηση του H. Putnam.

3.2 Μία απάντηση στον H. Putnam και μία διαφορετική διατύπωση του επιχειρήματος

Η απάντηση αυτή έρχεται μέσα από ένα μεταγενέστερο άρθρο του Lucas {#11} στο οποίο περιέχεται επίσης και μία μονολογική εκδοχή του διαλεκτικού επιχειρήματος, η οποία βασίζεται σε μία ιδέα και πάλι του Putnam και παρουσιάζει ενδιαφέρον.

Έχουμε δει ότι ο Putnam βάζει την αλήθεια της Gödel-πρότασης σε ένα πλαίσιο όπου αυτή μπορεί να αναχθεί από μία υπόθεση και συγκεκριμένα την υπόθεση της συνέπειας. Έχει επισημανθεί ήδη ότι η T δεν μπορεί να αποδείξει τη συνέπειά, της, αλλά δεν υπάρχει λόγος να πιστεύουμε ότι ο άνθρωπος μπορεί να το κάνει, άρα και οι δύο μένουν με την απόδειξη του ότι:

Αν T συνεπής, τότε U

οπότε η διαφορά ανθρώπου-μηχανής φαίνεται να έχει εξαφανιστεί. Το λεπτό σημείο που πρέπει να προσεχθεί σε αυτό το σημείο, είναι ότι, στην ανάλυση του Lucas, αλλά και στα σχόλια του ίδιου του Gödel, δεν παράγεται η αλήθεια της U από την υπόθεση της συνέπειας, αλλά από την ίδια της την αυτοαναφορική φύση.

Ο Lucas, αν και σε ένα μεταγενέστερο άρθρο του {#17} υπεραμύνεται της αυτοαναφορικότητας της Gödel-πρότασης, διαλέγει εδώ μία άλλη οδό αντίκρουσης και ισχυρίζεται ότι μπορεί να διαπιστώσει τη συνέπεια της T, άρα μέσω του κανόνα MP να δει και ‘επίσημα’ την αλήθεια της U. Ο τρόπος με τον οποίο ισχυρίζεται ότι μπορεί να διαπιστώσει κανείς τη συνέπεια της μηχανής περιγράφεται στις σελίδες 109 και 110 αλλά κρίνεται αμφίβολης αξιοπιστίας, για λόγους που αναλύονται σε εκείνο το σημείο της εργασίας. Πιο σαφή είναι τα αντεπιχειρήματα του Lucas, απέναντι σε μία άλλη εκδοχή του επιχειρήματός του που είναι μονολογική, έχει προταθεί από τον Putnam και αποτελείται από 5 βήματα:

- (1). Αν T είναι μη-συνεπής, τότε όχι-(T = Εγώ)
- (2). Βλέπει_{Εγώ} (Αν T συνεπής, τότε U)
- (3). Αν T συνεπής, τότε Βλέπει_{Εγώ}(U)
- (4). Αν T συνεπής, τότε όχι- Βλέπει_T(U)
- (5). Αν T συνεπής, τότε όχι-(T = Εγώ),

όπου Βλέπει_{Εγώ}(...) σημαίνει «Μπορώ να δω ότι το ... είναι αληθές» και αντίστοιχα, Βλέπει_T(...) σημαίνει «Η T μπορεί να παράγει σαν αληθές το...».

Η παραπάνω μορφή του διαλεκτικού επιχειρήματος μπορεί να μην είναι απολύτως ισοδύναμη με την αρχική, αλλά φαίνεται να είναι αυτή που διατηρεί καλύτερα το πνεύμα της, στα πλαίσια ενός ‘μονολόγου’ που καταλήγει σε άτοπο και αποδεικνύει την επιθυμητή θέση. Άλλωστε, ο αναγνώστης μπορεί να διαπιστώσει εύκολα ότι τα 5 βήματα ‘μαζεύουν’ το διαλεκτικό επιχείρημα, μαζί με τα σχόλια που έγιναν για τις μη-συνεπείς μηχανές στο κεφάλαιο 3.1 (βλ. σελ. 83-89).

Κατά τον Putnam, το πρόβλημα στην παραπάνω συλλογιστική πορεία φαίνεται να υπάρχει στη σύνδεση των βημάτων (2) και (3), οπότε όλη η αποδεικτική πορεία καθίσταται μη έγκυρη. Φυσικά, ο Putnam βασίζεται και στην ανυπαρξία τυπικών κανόνων που να ορίζουν την έννοια ‘βλέπω κάτι σαν αληθές’. Πράγματι, στα πλαίσια της ανάλυσης του Lucas δεν μπορεί εξ’ ορισμού να τυποποιηθεί, μπορεί όμως να χρησιμοποιηθεί με τέτοιο τρόπο ώστε να απαντηθεί η αντίρρηση του Putnam. Συγκεκριμένα ο Lucas αντιτάσσει ότι αν **εφαρμοστεί το Βλέπει_{Εγώ} σε όλους τους τύπους των βημάτων (1) έως (5)**, τότε δεν υπάρχει κανένα πρόβλημα στη σύνδεση των βημάτων (2*) και (3*), όπως αυτά προκύπτουν ως εξής:

- (1*). Βλέπει_{Εγώ}(Αν T είναι μη-συνεπής, τότε όχι-(T = Εγώ))
- (2*). Βλέπει_{Εγώ}(Βλέπει_{Εγώ} (Αν T συνεπής, τότε U))
- (3*). Βλέπει_{Εγώ}(Αν T συνεπής, τότε Βλέπει_{Εγώ}(U))
- (4*). Βλέπει_{Εγώ}(Αν T συνεπής, τότε όχι- Βλέπει_T(U))
- (5*). Βλέπει_{Εγώ}(Αν T συνεπής, τότε όχι-(T = Εγώ)),

όπου, όσον αφορά το (2*), μπορεί να γραφεί ισοδύναμα σαν το (2), για προφανείς λόγους.

Τώρα όμως, αν μπορώ να δω ότι ‘Αν η T συνεπής, τότε U’, τότε θα μπορώ να δω επίσης ότι ‘Αν η T είναι συνεπής, μπορώ να δω την U’. Διότι, ακολουθώντας τη συλλογιστική του θεωρήματος του Gödel, μπορώ να κάνω την απαραίτητη μετασκέψη του βήματος (3*), οπότε η πορεία (1*) έως (5*) είναι έγκυρη, άρα, καταλήγοντας και πάλι σε άτοπο, η μηχανιστική θέση καταρρίπτεται, ήτοι ‘Εγώ δεν είμαι η T’.

Για να ισχυροποιήσει τη θέση του, ο Lucas επίσης διαπιστώνει ότι η παραδοχή της (2) και της *άρνησης* της (3) θα τον οδηγούσε να δεχθεί ότι, μπορεί να βλέπει ότι ‘Αν η T είναι συνεπής, τότε U’, αλλά από την άλλη, ‘Αν και η T είναι συνεπής, δεν μπορεί ο Lucas να δει την U’. Αυτό που φαίνεται όμως πιθανό να ισχύει

από τη μεριά ενός τρίτου προσώπου, ο Lucas θεωρεί ότι θα είναι παράλογο να ισχύει για τον ίδιο ο οποίος διαπιστώνει τις παραπάνω προτάσεις. Διότι από τη στιγμή που κάποιος *θα δεχθεί σαν υπόθεση* ότι η T είναι συνεπής, και θα μπορεί να δει ότι αν ισχύει αυτό, τότε η U αληθεύει (μέσω της (2)), τότε θα δεχθεί την αλήθεια της U. Αλλά τότε η άρνηση της (3) θα τον οδηγήσει σε αντίφαση.

Διαπιστώνουμε με τα παραπάνω ότι ακόμη και μετά από αυτή την αναδιατύπωση των επιχειρημάτων, ουσιαστικά ο καθένας από τους δύο εμμένει στη θέση του χωρίς να μπορεί να δεχθεί βασικές αρχές του άλλου. Ας θυμηθούμε ότι ο Putnam θεωρεί ότι η πρόταση ‘εγώ είμαι μία μηχανή’ έχει να κάνει με την *ισοδυναμία της αποδεικτικής δύναμης* του ενός και του άλλου. Ο Lucas από την άλλη υποστηρίζει ότι αυτό που ξέρει σαν αληθές δεν μπορεί να τυποποιηθεί έτσι ώστε να εντάσσεται στην αυστηρή αποδεικτική του δύναμη, άλλωστε αν συνέβαινε κάτι τέτοιο θα ήταν πράγματι ισοδύναμος με μία μηχανή. Ο όλη οπτική του φαίνεται και από την αναδιατύπωση των 5 προτάσεων, η οποία συνίσταται ακριβώς στο ότι ο άνθρωπος μπορεί να δει από ένα επίπεδο πιο πάνω την αλήθεια των πραγμάτων. Αυτή η δυνατότητα είναι που αμφισβητείται από τον Putnam, γι’ αυτό και, δεχόμενοι της προτάσεις (1) έως (5), καταλήγουμε ότι δεν μπορούμε να δείξουμε τίποτα παραπάνω από μία μηχανή, ενώ με τις (1*) έως (5*) καταλήγουμε στο αντίθετο αποτέλεσμα.

Πριν περάσουμε στο επόμενο κεφάλαιο και καταπιαστούμε πλέον πιο αυστηρά με το θέμα που μας απασχολεί, έχει ενδιαφέρον να διαπιστώσουμε ότι οι προθέσεις του Lucas δεν είναι τόσο κοντόφθαλμες όσο μπορεί να φαίνεται αρχικά. Απεναντίας, ο τελικός σκοπός του είναι να συνεισφέρει στην προσπάθεια να απαλλαχθεί η επιστημονική και φιλοσοφική σκέψη από τον ‘μπαμπούλα’ του μηχανιστικού ντετερμινισμού που θέλει τους ανθρώπους που μιλάνε επιστημονικά να είναι κάποιου είδους αυτόματα, χωρίς ελευθερίες και παρεκκλίσεις από το αυστηρό πλαίσιο στο οποίο αναφέρονται. Στη συνέχεια, καθησυχάζοντας τους όποιους προβληματισμούς για τις συνέπειες μίας τέτοιας θέσης, αναφέρει ότι μία πιθανή κατάρριψη αυτής της θεώρησης, φυσικά δεν θα σημαίνει ότι η επιστημονική πρόοδος θα συνεχίσει να υφίσταται και να μας διαφωτίζει. Απλά, πάντα θα υπάρχει χώρος να ειπωθεί και κάτι παραπάνω.

Στο ίδιο μήκος κύματος με τα παραπάνω αναφέρθηκε στο τέλος της ιστορικής αναδρομής του πρώτου μέρους ότι ‘ο Μηχανισμός αποτελεί το ανάλογο του προγράμματος του Hilbert’. Διότι θα μπορούσε να πει κάποιος πως, όπως το

Πρόγραμμα είχε σα σκοπό την πλήρη αναγωγή της μαθηματικής σκέψης σε ένα κλειστό σύστημα συγκεκριμένων και αυστηρά δομημένων μαθηματικών, έτσι και ο Μηχανισμός φαίνεται να στοχεύσει στην αναγωγή όλων των λειτουργιών του ανθρώπινου εγκεφάλου σε μηχανιστικές διεργασίες των οποίων η πολυπλοκότητα μπορεί να μελετηθεί στα πλαίσια ενός συστήματος με απόλυτη ακρίβεια. Το αν βέβαια το θεώρημα του Gödel μπορεί να δια φωτίσει τα όρια των μηχανιστικών λειτουργιών με την ίδια σαφήνεια που έδειξε τα όρια των περατοκρατικών μαθηματικών, είναι το θέμα που προσπαθούμε να δια φωτίσουμε σε αυτή τη διπλωματική εργασία. Και ακριβώς επειδή δεν έχουμε να κάνουμε πλέον με αυστηρά μαθηματικά αλλά με φιλοσοφικές σκέψεις, πρέπει να δούμε το θέμα από όλες τις δυνατές οπτικές γωνίες, ώστε να σκιαγραφηθούν όλες οι απόψεις και να αξιολογηθεί κάθε επιχείρημα που υποστηρίζει την μία ή την άλλη πλευρά.

4. Πρώτη περίοδος αντιπαράθεσης (1961-1988)

Με την δεδομένη την ιστορική πορεία παρουσίασης που έχει επιλεχθεί, στο παρόν και το επόμενο κεφάλαιο αναλύονται οι απόψεις που παρουσιάζονται σε κάθε χρονική περίοδο, χωρίς βέβαια αυτές να ακολουθούν την αυστηρή χρονική σειρά με την οποία εκφράστηκαν. Το όριο που διαχωρίζει τις δύο περιόδους επιλέχθηκε να είναι το έτος 1989 οπότε εκδόθηκε το βιβλίο του Roger Penrose ‘The emperor’s new mind’. Ο λόγος για τον οποίο έγινε αυτή η επιλογή είναι ότι από τότε και μετά, ο Penrose φαίνεται να ‘παίρνει τη σκυτάλη’ από τον Lucas για να υπερασπιστεί εκ νέου τις δυνατότητες του ανθρώπινου νου και φυσικά να γίνει η αφορμή να αρχίσει μία νέα αντιπαράθεση πάνω στο ίδιο θέμα, αλλά με άλλους πρωταγωνιστές αυτή τη φορά.

Στην πρώτη περίοδο που πραγματεύεται το παρόν κεφάλαιο, αναλύονται τα άρθρα τα οποία περιέχουν τις πρώτες αντιρρήσεις πάνω στη χρήση του μαθηματικού επιχειρήματος, όπως αυτό παρουσιάστηκε στο προηγούμενο κεφάλαιο και φυσικά οι απαντήσεις του Lucas σε αυτές. Μία ανάλογη πορεία ανάλυσης θα ακολουθήσουμε και για τη δεύτερη περίοδο αντιπαράθεσης.

4.1 Αντιρρήσεις πάνω στη βάση των υποθέσεων του J.R. Lucas

Πριν περάσουμε στην ανάλυση επιχειρημάτων που στοχεύουν στην κατάρριψη των αρχικών υποθέσεων του Lucas, κρίνεται σκόπιμο να αναφερθούν και μερικές απόψεις που αποσυνδέουν το θέμα τελείως από τη βάση του. Για παράδειγμα, μπορεί να απορρίπτεται εξ’ αρχής η θεώρηση των μηχανών ως μόνο μηχανές Turing, οπότε η όλη ανάλυση που παρουσιάσαμε παραπάνω χάνει το νόημά της. Το κύριο πράγμα όμως που θα διαπιστώσουμε, είναι ότι υπάρχει μία έντονη κριτική που επικεντρώνει τα βέλη της στην αυτή καθ’ εαυτή χρήση του θεωρήματος του Gödel για την εξαγωγή συμπερασμάτων σε σχέση με το Μηχανισμό. Θεωρείται δηλαδή αυτή η χρήση, είτε ότι είναι εντελώς ανάρμοστη, είτε ότι είναι άσκοπη, ή ακόμη και ότι δεν αποτελεί το κύριο θέμα και κακώς επικεντρώνεται η προσπάθεια κατάρριψης της μηχανιστικής θέσης σε αυτή. Αυτή η θεώρηση των πραγμάτων έχει πράγματι πολλούς υποστηρικτές (και κυρίως μαθηματικούς-λογικούς), οπότε κρίνεται σκόπιμο να αναλυθεί περαιτέρω.

4.1.1 Περί της φύσης μηχανών και της αξιοπιστίας της συλλογιστικής του Lucas

Ένα χρόνο μετά τη δημοσίευση του άρθρου του J.R. Lucas, δημοσιεύτηκαν στο ίδιο περιοδικό (Philosophy) δύο σύντομα άρθρα που εκφράζουν τις πρώτες αντιρρήσεις πάνω στην αξιοπιστία της επιχειρηματολογίας του.

Συγκεκριμένα, ο C.H. Whiteley {#06} ισχυρίζεται ότι η ιδέα του επιχειρήματος του Lucas δεν είναι παρά ένα ‘λογικό μαγικό κόλπο’ με το οποίο παγιδεύεται η μηχανή στην αυτό-αναφορά. Αντίστοιχα, αντιπαραθέτει μία ανάλογη πρόταση η οποία παγιδεύει τον Lucas, την πρόταση «Αυτός ο τύπος δεν μπορεί να βεβαιωθεί με συνεπή τρόπο από τον Lucas», η οποία επιτελεί κατά τη γνώμη του την ίδια λειτουργεί με την μη-αποδείξιμη πρόταση που δεν μπορεί να αποδείξει η μηχανή. Αυτό το επιχείρημα μοιάζει στην ουσία του με το παράδοξο του ψεύτη. Το θεώρημα όμως του Gödel δεν είναι μία παραδοξολογία, αλλά έχει στέρεη επιστημονική βάση, όπως έχουμε επανειλημμένα επισημάνει. Αυτό το γεγονός βέβαια δεν γίνεται φανερό αν διαβάσει κανείς μόνο το άρθρο του Lucas, και πράγματι ο Whiteley ισχυρίζεται ότι δεν έχει άλλη πηγή γνώσης για το θεώρημα, οπότε είναι λογικό να του φαίνεται η χρήση του από τον Lucas σαν κάποιο ‘μαγικό κόλπο’ της λογικής.

Μία βαθύτερη αντίρρηση που έχει να κάνει με τη φύση των μηχανών εκφράζει ο F.H. George {#07}. Σύμφωνα με αυτόν, οι μηχανές που προορίζονται για να προσομοιώνουν την ανθρώπινη νοημοσύνη δεν βασίζονται στην αφαιρετική μέθοδο και μόνο, αλλά χρησιμοποιούν και επαγωγικές μεθόδους συλλογισμού⁽¹⁾, οπότε ξεφεύγουν από την αναλογία με τυπικά συστήματα. Το θέμα που τίγεται εδώ έχει αναλυθεί και από τον Lucas, ο οποίος υποστηρίζει ότι μηχανές που ξεφεύγουν από τα όρια της αφαιρετικής μεθόδου δεν είναι αυτές πάνω στις οποίες στηρίζεται η μηχανιστική θέση, μάλιστα δεν θα έπρεπε να καλούνται καν ‘μηχανές’.

Φαίνεται ότι το ζήτημα που εγείρει ο George, αποτελεί τελικά μία αμφισβήτηση της εγκυρότητας της θέσης του Church. Διότι, αν η υπολογιστότητα συνδέεται με την αναδρομικότητα και μόνον, τότε οι μηχανές στις οποίες αναφέρεται, θα πρέπει να ξεφεύγουν από τα όρια της αναδρομικότητας και να έχουν περαιτέρω υπολογιστικές δυνατότητες. Με την CT όμως σε ισχύ και αποτελέσματα όπως αυτό

(1): Η επαγωγική συλλογιστική [inductive reasoning] αφορά την εξαγωγή γενικών συμπερασμάτων από ένα σύνολο παραδειγμάτων. Για παράδειγμα, δεδομένου του γεγονότος ότι: ‘Αυτά τα ρούχα είναι του Γιάννη’, και του αποτελέσματος: ‘Αυτά τα ρούχα είναι μαύρα’, μπορούμε να εξαγάγουμε το συμπέρασμα: ‘Όλα τα ρούχα του Γιάννη είναι μαύρα’.

του θεωρήματος του Rice (βλ. Μέρος II, σελ. 57) κάτι τέτοιο δεν φαίνεται να είναι ακόμα εφικτό. [Σημ: Για τη σύνδεση της θέσης του Church με το θέμα της συζήτησης, βλ. επίσης παρακάτω, § 4.1.3] Έτσι, για την ώρα τουλάχιστον, δεν μπορούν να αλλάξουν ριζικά τη βάση της συζήτησης αντιρρήσεις που έχουν να κάνουν με τις υπολογιστικές δυνατότητες των μηχανών, εκτός εάν πρακτικά μπορεί να παραχθεί μία μηχανή που θα μας κάνει να αναθεωρήσουμε το θεωρητικό πλαίσιο πάνω στο οποίο μέχρι τώρα τις έχουμε δομήσει.

4.1.2 Περί των ορίων της διαλεκτικής προσέγγισης

Το 1967, ο I.J. Good {#08} προσπαθεί να θέσει το θέμα σε μία τελείως άλλη βάση, ψάχνοντας τα όρια της διαλεκτικής προσέγγισης του Lucas. Με βάση τα γραφόμενά του, για οποιαδήποτε μηχανή, ας πούμε F_1 , η οποία υπόκειται στο αποτέλεσμα του Gödel, μπορούμε να βρούμε μία βελτιωμένη μηχανή F_2 , για την οποία όμως επίσης ισχύει το αποτέλεσμα του Gödel, οπότε φτιάχνουμε μία νέα μηχανή F_3 κ.ο.κ. Το όριο αυτής της διαδικασίας είναι που απασχολεί τον Good. Διότι αν μπορούμε να βελτιώσουμε ‘άπειρα’ τη μηχανή, θα καταλήξουμε στην σε μία μηχανή F_ω , όπου ω είναι ο πρώτος υπερπεπερασμένος διατακτικός αριθμός, αλλά η πορεία βελτίωσης μπορεί να συνεχιστεί με μία μηχανή $F_{\omega+1}$ και συνεχίζοντας, θα φτάσουμε τελικά στον πρώτο μη-κατασκευάσιμο υπερπεπερασμένο διατακτικό, τον τ , του οποίου η ύπαρξη συνδέεται με το αξίωμα της επιλογής. Εφόσον όμως μέχρι εκείνο το σημείο μπορούμε να δώσουμε οδηγίες για την κατασκευή και αποδοχή της Gödel-πρότασης από τη μηχανή, η θέση του Lucas απορρίπτεται.

Η απάντηση του Lucas {#09} στην αντίρρηση αυτή, είναι ότι δεν αποτυπώνει η παραπάνω πορεία σκέψης την ουσία του επιχειρήματός του. Η ουσία είναι ότι για κάθε μοντέλο του ανθρώπινου νου που μπορεί να παρουσιάσει ο Μηχανιστής, ο άνθρωπος θα μπορεί να πάει ένα βήμα πιο πάνω, όχι με την έννοια της ανωτερότητας, απλά για να δείξει ότι είναι διαφορετικός. Το ότι μπορεί να ενσωματωθεί σε μία νέα μηχανή το δεδομένο που έδειξε τη διαφορετικότητα *δεν αναιρεί το γεγονός ότι αυτή δείχτηκε για τη συγκεκριμένη μηχανή-μοντέλο του νου*. Αυτό το ‘σχήμα αντίκρουσης’ είναι που μπορεί να ισχύει επ’ άπειρον, καταρρίπτοντας τη μηχανιστική θέση.

Η μίνι αντιπαράθεση των δύο ολοκληρώνεται με ένα άρθρο του Good {#13 - 1969} που έχει τον, μάλλον προκλητικό, τίτλο «Το θεώρημα του Gödel είναι άσχετο θέμα». Ο Good και πάλι ισχυρίζεται ότι εφόσον το διαλεκτικό επιχείρημα μπορεί να

συνεχιστεί και πέρα από το ω , και με κάποιον τρόπο ο άνθρωπος να αποδείξει ότι μπορεί να βρίσκει πάντα μία 'βελτίωση' στη μηχανή. Αλλά αν αυτό είναι ισοδύναμο με το να ισχυριζόμαστε ότι μπορεί να κατασκευαστεί ο πρώτος μη-κατασκευάσιμος υπερπεπερασμένος διατακτικός, πράγμα εξ' ορισμού αδύνατο. Έτσι η κατάδειξη της ισχύος του Νοητισμού ανάγεται τελικά στην κατασκευαστική διαδικασία της υπερπεπερασμένης ιεραρχίας αρίθμησης, και δεν έχει να κάνει σε τίποτα με το θεώρημα του Gödel.

Το τελευταίο συμπέρασμα δεν φαίνεται πολύ ελκυστικό, παρ' όλα αυτά την ευρύτερη άποψη του Good, ότι δηλαδή το θεώρημα του Gödel δεν έχει να δώσει κάτι σημαντικό στο θέμα της ισοδυναμίας ανθρώπινου νου και μηχανής, τη συμμερίζονται και πολλοί άλλοι λογικοί και φιλόσοφοι. Πάνω σε αυτό ο Lucas έχει μία απάντηση, την οποία όμως θα δούμε παρακάτω (βλ. σελ. 108), χάριν της πορείας που έχουμε επιλέξει για να παρουσιάσουμε τις διάφορες απόψεις. Όσον αφορά τα παραπάνω, ο Good φαίνεται ότι παρουσιάζει, αν και όχι πολύ συνειδητά, μία εύλογη αντίρρηση για τα όρια μέχρι τα οποία θα μπορεί κανείς να παρουσιάζει το συγκεκριμένο μαθηματικό επιχείρημα για να δείχνει τη διαφορετικότητά του από μία μηχανή. Το ερώτημα δηλαδή, με την ευρύτερη έννοια, είναι σε ποιο σημείο εξαντλείται η δημιουργικότητα του ανθρώπου; Ο Lucas υποστηρίζει ότι αυτή είναι άπειρη, ενώ ο Good υποστηρίζει ότι κάποτε εξαντλούνται τα όριά της, άρα θα πρέπει αλλού να αναζητήσουμε επιχειρήματα για την κατάρριψη του μηχανισμού.

4.1.3 Σύνδεση με το κατασκευαστικό πρόβλημα στα θεμέλια των μαθηματικών

Ο J. Webb στο άρθρο του 'Metamathematics and the philosophy of mind' {#12}, θέτει ένα παρόμοιο θέμα με τον Good, ήτοι προσπαθεί να δείξει ότι τελικά, «κάθε απόπειρα φιλοσοφικής σκέψης που προέρχεται από τα μετα-μαθηματικά, είναι απλώς μία δραματική μεταφορά του προβλήματος της κατασκευής στα θεμέλια των μαθηματικών». Έτσι, εφόσον ακόμα όχι μόνο δεν έχουν λυθεί βασικά προβλήματα θεμελίωσης των μαθηματικών, αλλά ακόμα και το αντίκτυπο των θεωρημάτων του Gödel είναι αμφιλεγόμενο, δεν μπορούμε να το επεκτείνουμε χωρίς πρόβλημα σε άλλους τομείς, ακόμα και της φιλοσοφίας. Αυτή η γενικότερη θέση δεν σταματά βέβαια τον Webb να κάνει τα σχόλιά του πάνω στις φιλοσοφικές ανησυχίες του

Lucas, πάντα βέβαια σε αναλογία με τα ανοιχτά θέματα του κατασκευαστικού προβλήματος [constructivity problem] των θεμελίων των μαθηματικών.

Όσον αφορά το κύριο επιχείρημα του Lucas, ο Webb θεωρεί πως όσο η θέση του Church παραμένει απλή θέση και δεν έχει λυθεί το πρόβλημα, το επιχείρημα στερείται αξιόπιστης βάσης. Πέρα από αυτή την αρχική παρατήρηση, το θέμα της αλήθειας της Gödel-πρότασης, για τον Webb αποτελεί μόνο θέμα επιλογής της επέκτασης του τυπικού συστήματος στο οποίο αναφερόμαστε. Πράγματι, επιλέγοντας σαν νέο αξίωμα ενός μη-πλήρους συστήματος την Gödel-πρόταση ή την άρνησή της, καταλήγουμε σε *δύο διαφορετικές* επεκτάσεις του εν λόγω συστήματος. Έτσι, το επιχείρημα του Lucas φαίνεται να βασίζεται μόνο στην επιλογή της επέκτασης που έχει σαν αξίωμα την Gödel-πρόταση, και όχι στην αυτοαναφορική φύση της τελευταίας. Επιπλέον, η ίδια η αυτοαναφορά αποτελεί, κατά τον Webb, μόνο θέμα επιλογής αριθμητικοποίησης κατά Gödel⁽²⁾, οπότε αποτελεί στην ουσία ένα *επιφανόμενο* και όχι την ουσία της Gödel-πρότασης. Με αυτόν τον τρόπο, και με τη χρήση διαφόρων αποτελεσμάτων που αφορούν τις επεκτάσεις τυπικών συστημάτων, ο Webb δεν αφήνει περιθώρια στον Lucas. Διότι, αν στηριχθεί στην αυτοαναφορά, τα αποτελέσματά του θα είναι λάθος, αφού αυτή αποτελεί επιφανόμενο, ενώ αν στηριχτεί στην αλήθεια της Gödel-πρότασης, το πρόβλημα ανάγεται σε ανοιχτά προβλήματα των θεμελίων των μαθηματικών και δεν έχει να κάνει σε τίποτα με τη φύση των μηχανών. Με τα δικά του λόγια: «...το νους-μηχανή-Gödel πρόβλημα δεν μπορεί να γίνει αντικείμενο κατανοητής μεταχείρισης μέχρι τουλάχιστον να ξεκαθαριστούν τα κατασκευαστικά προβλήματα στα θεμέλια των μαθηματικών».

Υπό αυτό το πρίσμα, τα υπόλοιπα επιχειρήματα του Lucas αντιμετωπίζονται αναλόγως. Ενδιαφέρον έχει μία παρατήρηση που έχει να κάνει με την θέση του Church και πάλι και σύμφωνα με την οποία: «Όταν το θέμα διατυπώνεται χωρίς φανφάρες, καταλήγει να είναι απλά ένα επιχείρημα για την CT. Ο Lucas την αποδέχεται ενώ ο Μηχανιστής όχι». Βέβαια αυτό όπως έχουμε δει δεν είναι απόλυτα ακριβές, διότι σκοπός του Lucas είναι να απορρίψει τον Μηχανισμό που *στηρίζεται στην CT*. Άλλωστε μηχανές οι οποίες ξεπερνούν υπολογιστικά τις TM δεν θεωρούνται από τον Lucas ούτε καν μηχανές.

(2): Πράγματι, μπορούμε να φανταστούμε ότι, στην αριθμητικοποίηση που επιλέχθηκε στην σελ. 41 του 2^{ου} μέρους, εναλλάσσοντας για παράδειγμα τους αριθμούς κωδικοποίησης του $\neg$ και του $\wedge$, η Gödel-πρότασή που θα προκύψει, θα είναι, αν τη δούμε κωδικοποιημένη, διαφορετική από αυτή που προέκυψε με την αρχική κωδικοποίηση.

Όπως θα περίμενε κανείς, το άρθρο {#17} με το οποίο ο Lucas συνολικά απαντά στον Webb αποτελεί ουσιαστικά μία αναδιατύπωση των φιλοσοφικών του θέσεων του, με συγκεκριμένες επιπλέον επισημάνσεις. Για την Gödel-πρόταση συγκεκριμένα, ο Lucas αντιτάσσει ότι δεν είναι η αυτοαναφορά που κάνει την πρόταση αυτή που είναι, αλλά αυτή η ιδιότητα που μας επιτρέπει με μεταμαθηματικούς συλλογισμούς να τη δούμε αληθή. Και αυτό, όσο επιφανόμενο και να είναι, είναι κάτι που η μηχανή δεν μπορεί να κάνει. Επίσης, σε ένα σχόλιο του Webb για την απόδειξη της συνέπειας της αριθμοθεωρίας από τον Gentzen, απαντά ότι ο Μηχανιστής που θα συμπεριλάβει μία τέτοια απόδειξη στη μηχανή του, παύει να είναι Μηχανιστής με την αυστηρή έννοια του όρου, όπως ακριβώς μία μηχανή που δεν υπόκειται στην CT, παύει να είναι μηχανή. Βέβαια αυτή η άποψη δεν στερείται κάποιας ουσίας, διότι αν για την απόδειξη της συνέπειας, χρησιμοποιούνται έννοιες όπως ‘άπειρες’ αποδείξεις κτλ., αυτό αμφισβητεί την ουσία της μηχανής όπως τη θεωρεί ο Lucas, να είναι δηλαδή απόλυτα καθορισμένη και πεπερασμένη.

Συνολικά για το θέμα, όπως έχουμε ήδη αναφέρει, ελλείπει εναλλακτικών, και με την CT ακόμα σε ισχύ, υποχρεούμαστε να θέτουμε το πρόβλημα στη βάση πάνω στην οποία το θέτει ο Lucas. Οι αντιρρήσεις βέβαια που εκφράζονται σε σχέση με τη θέση του Church, εκτός από την όποια αξία έχουν τη δεδομένη χρονική στιγμή, καταλαβαίνει κανείς πως θα αποκτήσουν πολύ μεγαλύτερη σε περίπτωση μίας ενδεχόμενης απόρριψης της θέσης, οπότε θα βοηθήσουν στο να γίνουν οι απαραίτητες αναθεωρήσεις πάνω στα θέματα τα οποία συζητούμε.

4.2 Μία διαφορετική προσέγγιση των συνεπειών του θεωρήματος του Gödel στη μηχανιστική θέση

4.2.1 Paul Benacerraf. 'God, the Devil and Gödel' (1967)

Το κύριο άρθρο του J.R. Lucas που αναλύθηκε στην παράγραφο 3.1, δεν προκάλεσε μόνο στείρες αρνητικές αντιδράσεις, όπως μπορεί να φαίνεται μέχρι τώρα, αλλά είχε σαν αποτέλεσμα να δημιουργηθεί και η ανάγκη εκ νέου αποσαφήνισης των συνεπειών των αποτελεσμάτων του Gödel στη διαμάχη Μηχανισμού–Νοητισμού. Στο άρθρο που θα αναλύσουμε σε αυτή την παράγραφο {#10}, θα δούμε πως ο P. Benacerraf προσπαθεί να καταδείξει τα λογικά λάθη που περιέχει η ανάλυση του Lucas και επίσης, πώς, μέσω της προσπάθειας αυτής, βγάζει ένα νέο συμπέρασμα για τις επιπτώσεις του θεωρήματος του Gödel στη μηχανιστική θέση.

Καταρχήν να αναφέρουμε ότι ο τίτλος του άρθρου, 'Ο Θεός, ο Διάβολος και ο Gödel', προέρχεται από μία ρήση του A. Weil: «Ο Θεός υπάρχει, εφόσον τα μαθηματικά είναι συνεπή, και ο Διάβολος υπάρχει επειδή δεν μπορούμε να το αποδείξουμε». Ο Gödel, κατά τον Benacerraf, φαίνεται να είναι ο συνδετικός κρίκος, εφόσον (πολύ χοντρικά) απέδειξε ότι αν τα μαθηματικά είναι συνεπή, τότε δεν μπορούμε να αποδείξουμε τη συνέπειά τους. Από αυτή την ανάλαφρη αρχή, ο Benacerraf παίρνει την αφορμή να μιλήσει για το μηχανισμό που διέπει την εξαγωγή φιλοσοφικών συμπερασμάτων από μαθηματικά αποτελέσματα. Ο μηχανισμός αυτός στην περίπτωση του θεωρήματος του Gödel, συνίσταται στη σύνδεση των μαθηματικών αποτελεσμάτων με τη φιλοσοφική άποψη για τη φύση των μαθηματικών, και το τι αποτελεί απόδειξη. Για παράδειγμα, όπως έχουμε ήδη διαπιστώσει, μία θεώρηση των μαθηματικών σε ορισμένα αυστηρά περατοκρατικά πλαίσια, μαζί με το 2^ο θεώρημα μη-πληρότητας, οδήγησε στην απόρριψη του προγράμματος του Hilbert.

Όσον αφορά το τι αποτελεί απόδειξη και τι όχι, αυτό εξαρτάται και πάλι από την ευρύτερη φιλοσοφική θεώρηση, και είναι τελικά αυτό που καθορίζει το πώς εκλαμβάνει κανείς τις συνέπειες του θεωρήματος του Gödel στο συγκεκριμένο κομμάτι της φιλοσοφίας του νου το οποίο μελετάμε. Για τον Benacerraf (και άλλους) το 'βλέπω κάτι σαν αληθές' δεν αποτελεί επαρκή απόδειξη για να βασιστεί επάνω της μία, ακόμα και φιλοσοφική, ανάλυση, οπότε είτε καταρρίπτεται όλη η συλλογιστική

του Lucas, ή τα αποτελέσματά της αποδυναμώνονται αισθητά. Από την άλλη, υπάρχει και η άποψη, που εκφράζεται μεταξύ άλλων από τους A. Anderson και D. Hofstadter {#21}, ότι, εκτός των άλλων, το θεώρημα του Gödel μας δείχνει επίσης ότι η ‘αποδειξιμότητα’ είναι μία πιο αδύναμη έννοια από την ‘αλήθεια’, με την έννοια ότι δεν μπορεί να την καλύψει πλήρως. Παρατηρούμε εδώ ότι, υπό αυτό το πρίσμα, το θεώρημα του Gödel φαίνεται να ‘προτρέπει’ τον Lucas (και τον κάθε Lucas) να θεωρήσει τα πράγματα όπως τα θεωρεί. Δίνει δηλαδή, εκτός από την δυνατότητα να μπορεί κανείς να αμφισβητήσει τη μηχανιστική θέση μέσω αυτού, και τη ‘μετα-δυνατότητα’ να μπορεί κανείς να υπερασπιστεί τη χρήση του, με βάση το ίδιο το θεώρημα! Αυτό είναι ένα ενδιαφέρον θέμα στο οποίο θα επανέλθουμε όταν ολοκληρωθεί η εικόνα του θέματος, στο τέλος του 3^{ου} μέρους.

Επιστρέφουμε τώρα στο κύριο μέρος του άρθρου. Ο Benacerraf βασικά στοχεύει στο να απορρίψει τη θέση του Lucas ξεκινώντας όμως από τις υποθέσεις του, και όχι αμφισβητώντας τις. Έτσι, δέχεται την αναλογία των TM με τα τυπικά συστήματα, με την γνωστή όμως επιφύλαξη που αφορά την ισχύ της θέσης του Church. Επίσης δέχεται την σύντομη παρουσίαση της απόδειξης του θεωρήματος του Gödel, αν και παραθέτει μία πιο ολοκληρωμένη, η οποία στηρίζεται στην αρχική διατύπωση του θεωρήματος (βλ. Μέρος II, § 4.3).

Στη συνέχεια ο Benacerraf διαφωνεί ανοιχτά με το βασικό επιχείρημα του Lucas, με το σκεπτικό ότι δεν μπορεί να αποδειχθεί *τυπικά* η αλήθεια της Gödel-πρότασης, και η μη τυπική απόδειξη μπορεί να αναπαραχθεί από τη μηχανή. Διότι αφού η μηχανή (που ονομάζεται στο άρθρο ‘Maud’) μπορεί, μέσω του 2^{ου} θεωρήματος μη-πληρότητας να αποδείξει ότι: $\text{Con}(\text{Maud}) \rightarrow H$, όπου H η Gödel-πρόταση, τότε «δεν μπορεί να ‘πειστεί’ η μηχανή ότι η H αποτελεί όντως αληθή αλλά μη-αποδείξιμη πρόταση;». Μία θετική απάντηση μπορεί να μην είναι άμεσα αποκλειόμενη από τα θεωρήματα του Gödel, όπως ισχυρίζεται ο Benacerraf, ο οποίος εδώ ακολουθεί τα χνάρια του Putnam (βλ. § 2.2), χρησιμοποιώντας και το 2^ο θεώρημα μη-πληρότητας για να διαπιστώσει εάν η Gödel-πρόταση είναι αληθής ή όχι.

Η διαπίστωση της αλήθειας της Gödel-πρότασης μέσω της συλλογιστικής του Lucas, είναι το κύριο σημείο που απασχολεί τον Benacerraf, ο οποίος κάνει σαφές ότι *δεν θεωρεί αξιόπιστη τη μη τυπική απόδειξη της αλήθειας της Gödel-πρότασης*. Με αυτό το δεδομένο, η ανακατασκευή του επιχειρήματος του Lucas κρίνεται

απαραίτητη και την ακολουθούμε βήμα προς βήμα παρακάτω, για να δούμε το τελικό συμπέρασμα που βγάζει ο Benacerraf για τις πιθανές προεκτάσεις του θεωρήματος του Gödel στην θεωρία του Μηχανισμού. Οι αρχικές υποθέσεις του είναι ότι το ανθρώπινο μυαλό είναι στην καλύτερη περίπτωση μία TM, χρησιμοποιεί και τα δύο θεωρήματα του Gödel και καταλήγει σε μία αντίφαση, μέσω της οποίας βγαίνει το τελικό συμπέρασμα.

Μία ανακατασκευή του επιχειρήματος του Lucas από τον P. Benacerraf

Βήμα 1^ο: Έστω ένα σύνολο $S = \{x \mid \text{Μπορώ να αποδείξω } x\}$ που αναπαριστά το σύνολο των πραγμάτων που ‘εγώ’ μπορώ να αποδείξω με την αφαιρετική μέθοδο και είναι σωστά αλλά με την ευρύτερη έννοια της ορθότητας που υποθέτει ο Lucas. Παίρνοντας την κλειστότητα αυτού του συνόλου υπό τους κανόνες απαγωγής της πρωτοβάθμιας λογικής με την ισότητα, έχω το σύνολο $S^* = \{x \mid S \vdash x\}$ το οποίο δεν είναι απαραίτητα κλειστό.

Βήμα 2^ο: Εφόσον κάθε στοιχείο του S είναι αληθές και η πρωτοβάθμια λογική διατηρεί την αλήθεια, κάθε στοιχείο του S^* είναι επίσης αληθές και άρα το S^* είναι ένα συνεπές σύνολο προτάσεων. Άρα απέδειξα (έστω και μη τυπικά) ότι το S^* είναι συνεπές, άρα εξ’ ορισμού " $Con(S^*)$ " $\in S$. Αλλά ισχύει ότι $S \subseteq S^*$, άρα " $Con(S^*)$ " $\in S^*$. Αυτή η τελευταία διαπίστωση αναλογεί (κατά τον Benacerraf) στην υπόθεση του Lucas ότι μπορεί να αποδείξει τη συνέπειά του, ή μάλλον στο ότι ξέρει ότι είναι συνεπής.

Βήμα 3^ο: Αφού το S^* είναι συνεπές, συνεπής θα είναι και όλα τα αναδρομικά αριθμήσιμα [i.e.] υποσύνολά του, έστω W_x . Δηλαδή $\forall x (W_x \subseteq S^* \rightarrow Con(W_x))$. Ανάλογα με το βήμα 2, αυτό το γεγονός το έχω αποδείξει, άρα: " $\forall x (W_x \subseteq S^* \rightarrow Con(W_x))$ " $\in S$ και κατά συνέπεια, " $\forall x (W_x \subseteq S^* \rightarrow Con(W_x))$ " $\in S^*$

Σημειογραφία: Το W_x του προηγούμενου βήματος δηλώνει το εξαγόμενο σύνολο αληθών προτάσεων μίας μηχανής W (με 'αριθμό προγράμματος' x) που αποδεικνύει θεωρήματα. Έτσι, για κάθε j , W_j είναι ο Gödel-αριθμός ενός r.e. συνόλου θεωρημάτων, αυτά δηλαδή που αποδεικνύει η μηχανή με πρόγραμμα j .

Βήμα 4^ο – Επιπλέον υποθέσεις: Οι υποθέσεις του Lucas, μπορούν να συμβολιστούν τώρα ως εξής: (1): " $Q \subseteq W_j$ " $\in S^*$, όπου σαν Q δηλώνεται η αριθμητική. Αυτή η

συνθήκη δηλώνει ότι *μπορώ να αποδείξω ότι η W_j είναι επαρκής για την αριθμητική.*

(2): " $W_j \subseteq S^*$ " $\in S^*$, δηλαδή *μπορώ να αποδείξω ότι το σύνολο των θεωρημάτων που αποδεικνύει η W_j είναι υποσύνολο του δικού μου συνόλου εξαγομένων θεωρημάτων.*

(3): $S^* \subseteq W_j$. Αυτή η υπόθεση αποτελεί την *άρνηση* της υπόθεσης του Lucas, ήτοι ότι *μία TM μπορεί να αποδείξει περισσότερα πράγματα από εμένα.*

Οι παραπάνω υποθέσεις θα χρησιμοποιηθούν με σκοπό να διαπιστωθεί αν υπάρχει μία TM που να ικανοποιεί και τις τρεις, ή θα οδηγηθούμε σε αντίφαση, οπότε θα δούμε τι συμπεράσματα μπορεί να προκύψουν.

Σημ: Από την υπόθεση (1), αμέσως προκύπτει ότι $Q \subseteq W_j$

Βήμα 5^ο: Εφόσον η W_j είναι επαρκής για την αριθμητική και ανάλογη ενός τυπικού συστήματος, από το 1^ο θεώρημα μη-πληρότητας έπεται η ύπαρξη μίας μη-αποκρίσιμης πρότασης H τέτοιας ώστε $H \notin W_j$ και $\neg H \notin W_j$. Επίσης, από το 2^ο θεώρημα μη-πληρότητας έχω:

$$\text{"Con}(W_j) \rightarrow H" \in W_j \quad (4)$$

Σημειώστε ότι το Con εδώ *δεν* γράφεται με *πλάγια γράμματα*, διότι αντιπροσωπεύει την συνέπεια στα πλαίσια της W_j . Πρόκειται δηλαδή στην ουσία για την κωδικοποίηση κατά Gödel της έννοιας της συνέπειας.

Βήμα 6^ο: Παίρνω την τελευταία πρόταση του 3^{ου} βήματος και αντικαθιστώ την καθολικότητα με ένα συγκεκριμένο στιγμιότυπο j . Αυτό μπορώ να το κάνω εφόσον το S^* είναι κλειστό ως προς την πρωτοβάθμια λογική. Άρα έχω:

$$"W_j \subseteq S^* \rightarrow \text{Con}(W_j)" \in S^* \quad (5)$$

Από (2) και (5), μέσω του κανόνα MP, παίρνω ότι: $"\text{Con}(W_j)" \in S^*$. Αυτό το γεγονός σημαίνει ότι για να μπορώ να εξάγω σαν συμπέρασμα τη συνέπειά της W_j , δεν αρκεί να μπορώ να αποδείξω οτιδήποτε μπορεί να αποδείξει η W_j , αλλά πρέπει επίσης να μπορώ να αποδείξω ότι μπορώ να αποδείξω οτιδήποτε αποδεικνύει η W_j ! Αυτό ακριβώς είναι το νόημα της υπόθεσης (2).

Βήμα 7^ο: Σε αυτό το βήμα επιχειρείται η σύνδεση των δύο συνεπειών, Con και Con , μέσω του τύπου:

$$"\forall x(Q \subseteq W_x \rightarrow (\text{Con}(W_x) \equiv \text{Con}(W_x)))" \in S^* \quad (6)$$

Για να ισχύει το μέρος της πρότασης που είναι μέσα στα εισαγωγικά, θα πρέπει να μπορούμε να αποδείξουμε ότι για κάθε τυπικό σύστημα που περιλαμβάνει την αριθμητική, υπάρχει ένας τύπος ο οποίος εκφράζει μέσα σε αυτό το σύστημα, τη συνέπειά του. Αλλά για να μπορώ να κάνω κάτι τέτοιο, μία ‘κρυφή’ επιπλέον υπόθεση είναι απαραίτητη. Αυτή είναι ότι πρέπει να μου δοθεί το πρόγραμμα x της μηχανής W_x με τέτοιο τρόπο ώστε να υπάρχει η δυνατότητα κατασκευής της έκφρασης που δηλώνει τη συνέπεια. Ο Benacerraf ισχυρίζεται πως αυτή η δυνατότητα δεν είναι αυτονόητη, διότι ο Lucas δεν πείθει ότι αν του δοθεί μία μηχανή-‘μαύρο κουτί’ μπορεί να καθορίζει το πρόγραμμά της μόνο από τα εξαγόμενα, χωρίς να μπορεί να διαπιστώσει τη λειτουργία της. Παρ’ όλα αυτά, μιας και οι υποθέσεις του Lucas γίνονται δεκτές για να καταλήξει το επιχείρημα, η υπόθεση (1) γίνεται επίσης δεκτή, αφού αντιπροσωπεύει ακριβώς αυτή την ‘ενόραση’ στην εσωτερική λειτουργία μιας μηχανής, που μας δίνει το συμπέρασμα ότι μπορεί να εκφράσει τη συνέπειά της.

Με τα παραπάνω δεδομένα και με το δεδομένο της πληρότητας της μηχανής, ανάλογα με το βήμα 6, παίρνω το j στιγμιότυπο του τύπου (6) και έχω:

$$"Q \subseteq W_j \rightarrow (\text{Con}(W_j) \equiv \text{Con}(W_j))" \in S^* \quad (7)$$

Βήμα 8^ο: Από τους την υπόθεση (1) και τον τύπο (7), μέσω του κανόνα MP προκύπτει τελικά ότι: $"\text{Con}(W_j)" \in S^*$. Και με βάση την υπόθεση (3), άμεσα έχουμε:

$$"\text{Con}(W_j)" \in W_j \quad (8)$$

Βήμα 9^ο: Από (4) και (8) μέσω του κανόνα MP, προκύπτει ότι: $H \in W_j$, άτοπο διότι έχει υποτεθεί ότι η W_j είναι συνεπής (βλ. βήμα 5^ο). Αν βέβαια η W_j προκύπτει μη-συνεπής, από υπόθεση (2) το S^* είναι επίσης μη-συνεπές, καταλήγοντας και πάλι σε αντίφαση με βάση το βήμα 2^ο.

Οι συνέπειες του παραπάνω αποτελέσματος κατά τον Benacerraf

Η αντίφαση που προκύπτει μετά από την επισκόπηση της όλης διαδικασίας, οφείλεται, κατά τον Benacerraf, στις τρεις υποθέσεις του 4^{ου} βήματος, εφόσον δεχθούμε τις αρχικές υποθέσεις του 1^{ου} βήματος. Ο ίδιος υποστηρίζει πως οι τρεις υποθέσεις *δεν μπορούν να ιδωθούν ξεχωριστά*, οπότε πρέπει να αρνηθούμε *και τις τρεις μαζί*. Έτσι, «...είτε δεν μπορώ να δείξω ότι η W_j είναι επαρκής για την αριθμητική, ή, αν είμαι ένα υποσύνολο του W_j , τότε δεν μπορώ να αποδείξω ότι μπορώ να αποδείξω οτιδήποτε μπορεί να αποδείξει η W_j . Φαίνεται επίσης ότι είναι συνεπές με όλα αυτά, ότι είμαι όντως μία TM, αλλά μία με τόσο πολύπλοκο πρόγραμμα που δεν μπορώ να διακρίνω ποια ακριβώς. Ανάλογα, αν είμαι μία TM, τότε δεν μπορώ να ξέρω ποια είμαι». Στο τέλος του άρθρου, διατυπώνεται η τελική θέση του Benacerraf ως εξής: «Αν αγνοήσουμε την πιθανότητα να είναι λάθος η υπόθεση (1), τότε μπορούμε να αναδιατυπώσουμε τον αντίκτυπο του θεωρήματος του Gödel ως εξής. Αν είμαι μία TM, τότε από τη φύση μου δεν μπορώ να υπακούσω στη σωκρατική νουθεσία: ‘Γνώθι σαυτόν’», που σημαίνει ότι θα υπάρχουν πάντα πράγματα που δεν μπορώ να ξέρω για τον εαυτό μου.

Να σημειώσουμε, τέλος, ότι ο Benacerraf αναφέρει ότι η παρουσίαση αυτού του αποτελέσματος σε έναν φίλο του, επέφερε το συμπέρασμα ότι «η Ψυχολογία όπως την ξέρουμε, με αυτό τον τρόπο είναι αδύνατο να υπάρχει»! Αυτό το προκλητικό σχόλιο είναι όντως συμβατό με το τελικό συμπέρασμα και δείχνει με τον πιο χαρακτηριστικό τρόπο το πόσο μακριά μπορούμε να φτάσουμε στην εξαγωγή συμπερασμάτων με βάση το θεώρημα του Gödel. Βέβαια η αξιοπιστία του, συνδέεται άμεσα με την αξιοπιστία της ανάλυσης του Benacerraf, η οποία όμως είναι τουλάχιστον αμφιλεγόμενη, όπως διαφαίνεται από τα παρακάτω σχόλια, αλλά και την εμπεριστατωμένη κριτική του Charles Chihara πάνω στη βάση των υποθέσεών της (βλ. παρακάτω, § 4.3).

Σχόλια: Η παραπάνω ανάλυση του Benacerraf, αν και λογικοφανής, φαίνεται να στηρίζεται σε αμφιλεγόμενες και ίσως ακόμη και διφορούμενες υποθέσεις. Συγκεκριμένα, οι 3 βασικές υποθέσεις του βήματος 4, τις οποίες θέλει τελικά να απορρίψει ο Benacerraf, φαίνεται ότι είναι *ήδη αντιφατικές μεταξύ τους!* Διότι, η (3) αποτελεί σχεδόν προφανώς την άρνηση της (2), ή μάλλον και οι δύο μαζί μπορούν να αντικατασταθούν από την υπόθεση $W_j = S^*$ που μεταφράζεται ως: 'Είμαι ισοδύναμος με μία TM'. Το περίεργο είναι ότι αυτό το διαπιστώνει και ο ίδιος, αν και προσπαθεί να το αποφύγει λέγοντας ότι η συνθήκη που δείχνει την ισοδυναμία θα ήταν η $W_j = S$. Και πάλι όμως, αν, σύμφωνα πάντα με τον Benacerraf, οτιδήποτε αποδεικνύει είναι αληθές, πως μπορούν να συνυπάρχουν οι (2) και (3), αφού από τη (2) προκύπτει ότι $W_j \subseteq S^*$, άρα και πάλι βγαίνει η ισότητα; Τίποτα στην παραπάνω ανάλυση δεν ξεκαθαρίζει με ποια κριτήρια χρησιμοποιείται τότε η (2) και τότε η (3) και γιατί δεν χρησιμοποιείται η ισότητα. Και όμως, με δεδομένο την υπόθεση (1) και ότι $W_j = S^*$, θα καταλήγαμε και πάλι σε άτοπο μέσω του τύπου (8). Τότε όμως θα απορρίπταμε είτε την υπόθεση (1) – που ο ίδιος παραδέχεται ότι μάλλον ισχύει, πέρα από την αντίρρηση του 'μαύρου κουτιού' – ή την υπόθεση $W_j = S^*$, άρα την υπόθεση 'είμαι μία TM'. Μα το ίδιο απορρίπτει και ο Lucas με το επιχείρημά του!

Εκτός των παραπάνω όμως, είδαμε επίσης ότι ο Benacerraf διαπιστώνει ότι πρέπει τελικά να απορριφθούν και οι τρεις υποθέσεις του 4^{ου} βήματος (συμβ. $Y(1), Y(2), Y(3)$). Σχηματικά, ισχυρίζεται ότι:

$$\text{Θεώρημα Gödel} \rightarrow \neg(Y(1) \wedge Y(2) \wedge Y(3)) \Leftrightarrow$$

$$\text{Θεώρημα Gödel} \rightarrow \neg Y(1) \vee \neg Y(2) \vee \neg Y(3)$$

Ενώ όμως ο ίδιος αναλύει τις επιπτώσεις την απόρριψης των υποθέσεων (1) και (2), οι επιπτώσεις μιας πιθανής απόρριψης της (3) δεν συζητούνται καν, διότι η (3) έχει γίνει δεκτή *a priori*. Μία απόρριψή της όμως, θα οδηγούσε στο συμπέρασμα ότι μπορώ όντως να αποδείξω περισσότερα πράγματα από μία TM, δηλαδή και πάλι στην θέση του Lucas. Ενώ λοιπόν αυτή η θέση έχει ίσες πιθανότητες να ισχύει, μαζί με τις αρνήσεις των άλλων δύο υποθέσεων, ο Benacerraf δεν ασχολείται καν μαζί της, ενώ προφανώς αποτελεί ένα μέρος των πιθανών συμπερασμάτων του.

Όπως θα διαπιστωθεί παρακάτω, υπάρχουν και άλλα προβλήματα με τις υποθέσεις του Benacerraf, οπότε συνολικά τα συμπεράσματά του, λογικό είναι να επιδέχονται μίας έντονης αμφισβήτησης.

4.2.2 Η απάντηση του J.R. Lucas

Ο Lucas απάντησε στο παραπάνω άρθρο αρκετά εμπεριστατωμένα, σε ένα άρθρο {#11} που θα μπορούσε να σταθεί και από μόνο του σαν ολοκληρωμένη παρουσίαση των βασικών θέσεών του. Θα δούμε τώρα τα κυριότερα σημεία του και θα διαπιστώσουμε πως με το χρόνο το επιχείρημά του γίνεται ακόμα πιο λεπτό και εύκαμπτο, ώστε να αντιστέκεται σε ακόμη περισσότερες αντιρρήσεις.

Το άρθρο αρχίζει όπως αυτό του Benacerraf, φιλοσοφώντας πάνω στη χρήση μαθηματικών αποτελεσμάτων σε τομείς που δεν άπτονται των άμεσων συνεπειών τους. Συγκεκριμένα για το θεώρημα του Gödel, ο Lucas γράφει: «Το θεώρημα του Gödel, όπως πολλές άλλες αλήθειες, μπορεί να χρησιμοποιηθεί με δύο τρόπους: μπορεί να χρησιμοποιηθεί σαν μία τυπική απόδειξη που δίνει συγκεκριμένα συντακτικά αποτελέσματα για μία συγκεκριμένη κλάση τυπικών συστημάτων, μπορεί όμως να χρησιμοποιηθεί και σαν ένα συγκεκριμένου τύπου ή στιλ επιχείρημα, το οποίο μπορούμε να κατανοήσουμε, και, όταν το έχουμε πλήρως αποσαφηνίσει, να μπορούμε να το προσαρμόσουμε και εφαρμόσουμε σε αναρίθμητες διαφορετικές καταστάσεις». Αυτή η θέση συνοψίζει, πιστεύω, όλη την ουσία πίσω από κάθε πιθανή ή απίθανη χρήση του θεωρήματος του Gödel για την εξαγωγή οποιαδήποτε μη μαθηματικών συμπερασμάτων. Η αξία βέβαια του κάθε συμπεράσματος δε μπορεί να είναι η ίδια και θα κριθεί αναλόγως, αλλά η ώθηση και η τροφή για σκέψεις που δίνεται από το συγκεκριμένο αποτέλεσμα δε μπορεί να γίνει η ίδια αντικείμενο καμίας κακόβουλης κριτικής.

Στη συνέχεια ο Lucas παρουσιάζει και πάλι το διαλεκτικό του επιχείρημα, προσθέτοντας ότι πάντα αναφέρεται, *όχι σε έναν συγκεκριμένο άνθρωπο* που μπορεί να ακολουθήσει την αποδεικτική πορεία του θεωρήματος του Gödel, αλλά στο «ιδεατό και αθάνατο Εγώ», το οποίο μπορεί κάθε φορά να δειχθεί διαφορετικό από μία μηχανή-μοντέλο του νου. Αυτή η επισήμανση απαντά και διαφωνίες που διατυπώνονται ως προς το ότι το μαθηματικό επιχείρημα μπορεί να χρησιμοποιηθεί μόνο από ανθρώπους-μαθηματικούς⁽¹⁾. Φυσικά, κινούμενοι στο χώρο των ιδεών, το ζητούμενο είναι να διαπιστώσουμε αν η ουσία *ολόκληρης της ανθρώπινης σκέψης* μπορεί να είναι ανάλογη με την ουσία της *μηχανιστικής ή αλγοριθμικής διαδικασίας*, και όχι αν ένας συγκεκριμένος άνθρωπος μπορεί να αποδειχθεί ανάλογος με μία

(1): Μία τέτοια διατυπώνεται στο άρθρο του D. Coder 'Gödel's Theorem and Mechanism' {#15}

συγκεκριμένη μηχανή. Η γενικότερη αυτή διαπίστωση αναφέρεται στο συγκεκριμένο άρθρο από τον Lucas σε σχέση με την κριτική του Benacerraf στο διαλεκτικό επιχείρημα, κριτική που δεν αναλύσαμε πιο πριν, διότι είναι μάλλον πρόχειρη και βασίζεται στη λάθος υπόθεση ότι ο Lucas τα βάζει με τον άνθρωπο που φτιάχνει τη μηχανή και όχι με την ίδια τη μηχανή. Πάντως ο Benacerraf βρίσκει ενδιαφέρουσα τη διαλεκτική προσέγγιση και το αντίστοιχο επιχείρημα πειστικότερο.

Γυρνώντας και πάλι στο άρθρο του Lucas, διαπιστώνουμε ότι οι απόψεις του για τη σχέση αποδειξιμότητας και αλήθειας, ουσιαστικά ταυτίζονται με αυτές που έχουμε παραθέσει στην αρχή του κεφαλαίου. Σαν παραπάνω πειστήριο, παρατίθεται το αποτέλεσμα του Tarski για την ανυπαρξία δυνατότητας τυπικής αναπαράστασης της έννοιας της αλήθειας, κάτι που επίσης υπονομεύει την προσπάθεια του Benacerraf να 'χαλιναγωγήσει' ακόμα και την έννοια της μη τυπικής απόδειξης σε τυπικά πλαίσια.

Όσον αφορά τη βασική επιχειρηματολογία του Benacerraf, ο Lucas ισχυρίζεται ότι αυτή καταλήγει σε ένα συμπέρασμα που είναι κενό νοήματος. Διότι, αν η αυτογνωσία είναι αδύνατη, τότε αν ρωτήσει κανείς τον Μηχανιστή του Benacerraf τι τύπου μηχανή είναι, η απάντηση θα είναι 'ο εαυτός σου'! Αλλά τότε η μηχανιστική θέση στερείται της θεωρητικής της βάσης, η οποία είναι ότι όλες οι ανθρώπινες λειτουργίες μπορούν να αναπαρασταθούν με τέτοιο (μηχανιστικό) τρόπο ώστε να είναι απόλυτα εξηγήσιμες. Υπάρχουν βέβαια και αντιρρήσεις που πηγάζουν από τις τρεις υποθέσεις του 4^{ου} βήματος, δεν προσθέτουν όμως κάτι ουσιαστικό στην ανάλυση που έγινε παραπάνω.

Στο ζήτημα που αφορά την αναγνώριση της συνέπειας μίας μηχανής, αναγνώριση την οποία προβάλλει σαν απαίτηση ο Benacerraf αλλά και άλλοι, ο Lucas παραθέτει έναν παράδοξο αλλά ενδιαφέρον τρόπο ελεγχθεί αν μία μηχανή είναι συνεπής: Προτείνει να ρωτήσουμε τον κατασκευαστή της μηχανής αν η μηχανή του μπορεί να αποδείξει την Gödel-πρόταση. Δεδομένης της γνώσης που έχει πάνω στη λειτουργία της μηχανής ο κατασκευαστής της, θεωρείται αυτονόητο ότι ξέρει την απάντηση. Έτσι, αν απαντήσει ότι η μηχανή μπορεί να αποδείξει την Gödel-πρόταση, τότε θα ξέραμε ότι η μηχανή είναι μη-συνεπής, ενώ στην αντίθετη περίπτωση θα ξέραμε ότι είναι συνεπής. Βέβαια και στις δύο περιπτώσεις επαληθεύεται η άρνηση της μηχανιστικής θέσης, εφόσον μία συνεπής μηχανή δε μπορεί να είναι μοντέλο για

το νου, ενώ σε περίπτωση συνέπειας, το μαθηματικό επιχείρημα μπορεί να χρησιμοποιηθεί για να καταλήξουμε στο ίδιο ακριβώς συμπέρασμα.

Βέβαια η παραπάνω μέθοδος προσθέτει μία δεδομένη ασάφεια στα πράγματα, διότι μέχρι τώρα ο Μηχανιστής δεν έπαιζε κάποιο σημαντικό ρόλο, εκτός από το να είναι υπέρμαχος των θέσεών του. Αν όμως τον χρησιμοποιήσουμε για την εξαγωγή συμπερασμάτων, τότε θα πρέπει να τον συμπεριλάβουμε και στην κριτική μας, οπότε η όλη φιλοσοφική ‘διαμάχη’, δεν θα είναι πλέον μία διαμάχη ιδεών αλλά μία διαμάχη μεταξύ ανθρώπων, από την οποία δε νομίζω να εκπορευόταν τίποτα ιδιαιτέρως ουσιαστικό. Προσωπικά βρίσκω πιο δελεαστικό το επιχείρημα που αρχικά είχε διατυπώσει ο Lucas και επαναλαμβάνει στο τέλος αυτού του άρθρου, ότι δηλαδή η συνέπεια τελικά είναι θέμα θεμελιώδους επιλογής του ανθρώπου. Και μπορεί να μην είναι δυνατόν να αποδειχθεί τυπικά, όμως χωρίς αυτή δεν δεχόμαστε να προχωρήσουμε βήμα παραπέρα, επιστημονικά μιλώντας.

4.2.3 Μία συνολική κριτική στο νέο πρόβλημα και ένα νέο συμπέρασμα

Ο Charles Chihara στο άρθρο του ‘On alleged refutations of mechanism’ {#18} επιχειρεί μία συνολική θεώρηση των επιχειρημάτων που χρησιμοποιούν το θεώρημα του Gödel ενάντια στον Μηχανισμό και δίνει κυρίως βάση στην ανακατασκευή του επιχειρήματος του Lucas από τον Benacerraf. Ο απώτερος σκοπός του είναι να δείξει ότι τελικά τα αποτελέσματα και των δύο δεν περιέχουν τα στοιχεία αυτά που είναι σε θέση να καταρρίψουν το Μηχανισμό, ούτε καν να βγάλουν άλλα συμπεράσματα, διότι βασίζονται σε λάθος αρχές. Ο τρόπος με τον οποίο επιλέγει να επιτύχει το στόχο του ο Chihara είναι πραγματικά πρωτότυπος και εντυπωσιακός. Αντί να απορρίψει εξ’ αρχής τις διφορούμενες (όπως προκύπτουν) ιδέες, προσπαθεί να *ισχυροποιήσει* τα αποτελέσματα, και κυρίως αυτό του Benacerraf! Έτσι, καταλήγοντας στο πιο ‘ισχυρό’ αποτέλεσμα και δείχνοντας την ασάφεια που το διέπει, μπορεί να διαπιστώσει τελικά ότι η μηχανιστική θέση δεν μπορεί να καταρριφθεί από επιχειρήματα που πηγάζουν από το θεώρημα του Gödel. Η ανάλυση του Benacerraf προτιμάται για την μεγαλύτερη σαφήνιά της, σε αντίθεση με τις απόψεις του Lucas οι οποίες εκτιμώνται ασαφείς και διφορούμενες, κυρίως αυτές περί συνέπειας του ανθρώπου, οι οποίες βέβαια, όπως έχουμε ήδη επισημάνει, αποτελούν έτσι κι αλλιώς το αδύναμο σημείο της συλλογιστικής του.

Αυτό που φαίνεται κυρίως να απασχολεί τον Chihara είναι ο ορισμός του συνόλου $S = \{x \mid \text{Μπορώ να αποδείξω } x\}$, στο οποίο εμπλέκεται η έννοια της (απόλυτης) απόδειξης, έννοια η οποία δεν αποσαφηνίζεται πλήρως από τον Benacerraf. Για παράδειγμα, υπό μία άλλη ερμηνεία της ‘απόλυτης απόδειξης’, όπου οι κανόνες και τα αξιώματα που χρησιμοποιούνται για μία ‘απόλυτη απόδειξη’ είναι αυταπόδεικτα, θα μπορούσε να πει κάποιος ότι μπορεί να ανακαλύψει το πρόγραμμα του εμπειρικά, άρα να ξέρει ποια ΤΜ είναι, απλά να μη μπορεί να το αποδείξει.

Περαιτέρω σκεπτικισμό όσον αφορά τα αποτελέσματα του Benacerraf προκαλεί μία ανάλυση του παραρτήματος του αρχικού του άρθρου, στο οποίο δεν είχαμε αναφερθεί μέχρι τώρα. Στο παράρτημα αυτό, θέλοντας ο Benacerraf να διασφαλίσει την αλήθεια των όσων περιέχονται στο σύνολο S , χρησιμοποιεί το σύνολο Q (βλ. σελ. 104) και τις επεκτάσεις του Q' και Q'' , που ενσωματώνουν αντίστοιχα τους κανόνες απαγωγής: ‘οτιδήποτε μπορώ να αποδείξω είναι αληθές’ και ‘μπορώ να αποδείξω ότι οτιδήποτε μπορώ να αποδείξω στο Q' είναι αληθές’. Δεδομένης της καθολικότητας που χαρακτηρίζει αυτά τα σύνολα, δεν θα έπρεπε ίσως να ξαφνιαζόμασταν που ο Chihara δείχνει ότι το Q'' είναι μη-συνεπές σύστημα! Φυσικά αυτό προκύπτει με μία ακόμα παραλλαγή του διαγώνιου επιχειρήματος, και βέβαια τίθεται πλέον σοβαρό θέμα αξιοπιστίας των συμπερασμάτων του Benacerraf.

Παρά την διαπίστωση αυτή, ο Chihara συνεχίζει και προσπαθεί να ανασκευάσει το επιχείρημα ώστε να αποφεύγει τις αντιφάσεις. Συγκεκριμένα προτείνει την αλλαγή του συνόλου S στο παρακάτω:

$$S = \{x \mid x \text{ είναι ο Gödel-αριθμός μιας πρότασης του } Q \text{ που μπορώ να αποδείξω}\}.$$

Με αυτόν τον νέο ορισμό αποφεύγουμε το παράδοξο που είδαμε παραπάνω και καταλήγουμε πάλι στο συμπέρασμα του Benacerraf: Είτε το S δεν είναι αναδρομικά αριθμήσιμο, άρα δεν μπορεί να αναπαραχθεί από μία μηχανή, ή, αν είναι, δεν μπορώ να ανακαλύψω ποτέ το ‘πρόγραμμά του’. Έτσι ο Μηχανισμός φαίνεται και πάλι να καταρρέει⁽²⁾ διότι, αν το S είναι όντως αναδρομικά αριθμήσιμο (με βάση τις αρχές του Μηχανισμού), τότε «θα ήταν κατανοητό ότι η πρόοδος των εμπειρικών επιστημών θα μας επέτρεπε κάποια μέρα να βρούμε το πρόγραμμά του, κάτι που από το παραπάνω επιχείρημα δείχνεται ότι είναι αδύνατο».

(2): Παρατηρούμε εδώ ότι ουσιαστικά είναι σαν να διαπιστώνει ο Chihara, με βάση την ανάλυση του Benacerraf, αυτό ακριβώς το οποίο ο τελευταίος θέλει να αποφύγει, ήτοι ότι από το θεώρημα του Gödel καταρρίπτεται ο Μηχανισμός! Κάτι ανάλογο διαπιστώσαμε και με τα σχόλια τα οποία ακολούθησαν των συμπερασμάτων του Benacerraf στο προηγούμενο κεφάλαιο.

Παρ' όλα αυτά, ο Chihara παραμένει ανικανοποίητος όσον αφορά την ιδέα της (απόλυτης) απόδειξης που χρησιμοποιείται για να περιγράψει τις ευρύτερες αποδεικτικές δυνατότητες του ανθρώπου και δεν μπορεί να δεχθεί εύκολα ένα τέτοιο αποτέλεσμα. Βασιζόμενος στη σύνδεση της παραπάνω έννοιας με αυτή της γνώσης και εφόσον η γνώση είναι κάτι συνεχώς μεταβαλλόμενο και πιο ρευστό από την αποδεικτική ικανότητα, προτείνει την παρακάτω θεώρηση:

Δεχόμενοι τη μηχανιστική θέση, περιμένουμε από μία ΤΜ να αναπαράγει όλα τα πιθανά εξαγόμενα σύνολα αριθμητικής από τον άνθρωπο, τα οποία όμως εξαρτώνται κάθε φορά από την δεδομένη κατάσταση του ανθρώπου (γνώσεις, ηλικία κτλ.). Εφόσον πρέπει σε αυτή τη μηχανή να εισάγουμε κάποια δεδομένα για να μας παράγει το οτιδήποτε, έστω T_a να είναι το σύνολο των αριθμών που μπορεί να παράγει μία μηχανή με την εισαγωγή δεδομένων a . Με την ίδια λογική, έστω S_a να είναι το σύνολο των Gödel-αριθμών προτάσεων της Q που μπορώ να αποδείξω τη δεδομένη χρονική στιγμή a .

Με βάση τα παραπάνω, το επιχείρημα του Benacerraf αποκτά μία καλύτερη βάση, αλλά οι υποθέσεις του χάνουν ακόμα περισσότερο σε αξιοπιστία, διότι πλέον δεν είναι σταθερό το σύνολο των πραγμάτων που μπορώ να αποδείξω. Έτσι, πριν μάθω το πρόγραμμα μίας ΤΜ που υπολογίζει το S , με περιγράφει ένα σύνολο S_a , αλλά μετά από αυτή τη γνώση με περιγράφει ένα άλλο σύνολο, ας πούμε S_{a+1} . Βάσει αυτής της ιδέας, ο Chihara τελικά θέλει να προβάλλει ακόμα πιο επιτακτικά την ανάγκη περαιτέρω διερεύνησης και αποσαφήνισης των εμπλεκόμενων εννοιών στη συζήτηση για τις επιπτώσεις του θεωρήματος του Gödel στο Μηχανισμό. [Σημ: Αυτή η ιδέα υλοποιείται με πιο σαφή τρόπο από τον D. Lewis στο κεφάλαιο που ακολουθεί, βλ. σελ. 115].

Μετά από όλη αυτή την ανάλυση ο Chihara καταλήγει πως φαίνεται τελικά το καλύτερο δυνατό αποτέλεσμα που μπορεί να προκύψει σαν φιλοσοφική προέκταση του θεωρήματος του Gödel είναι ότι 'δε θα μπορέσω ποτέ να ανακαλύψω το πρόγραμμά μου'. Εφόσον όμως αυτό είναι στην ουσία το ίδιο συμπέρασμα με αυτό του Benacerraf, του οποίου η ανάλυσή βρέθηκε τουλάχιστον ελλιπής, φαίνεται ότι όχι μόνο δεν προκύπτουν σοβαροί λόγοι για την κατάρριψη του Μηχανισμού μέσω του θεωρήματος του Gödel, αλλά ακόμα και όποια άλλα αποτελέσματα σχετικά μπορεί να προκύψουν δεν έχουν ιδιαίτερη ισχύ.

4.3 Ανακεφαλαίωση και μία σύντομη παρουσίαση κάποιων επιπλέον αντιρρήσεων

Τα άρθρα και οι απόψεις που παρουσιάστηκαν σε αυτό το κεφάλαιο, μας παρέχουν μία ολοκληρωμένη εικόνα όσον αφορά τη χρήση του θεωρήματος του Gödel για την κατάρριψη του Μηχανισμού από τον J.R. Lucas. Είδαμε ότι η κριτική στη βασική θέση του Lucas ήταν σφόδρα και πολυεπίπεδη, με την έννοια ότι τα επιχειρήματά του έγιναν αντικείμενο επίθεσης από πολλές διαφορετικές οπτικές γωνίες, σα να μην μπορεί να βρεθεί ένας κοινός άξονας πάνω στον οποίο να επικεντρώνεται αυτή η κριτική. Το γεγονός αυτό δείχνει ότι σίγουρα οι προεκτάσεις του θεωρήματος του Gödel στη φιλοσοφία είναι ένα ζήτημα που έχει προξενεί μεγάλο ενδιαφέρον άρα δε θα μπορούσε να είναι στην ουσία του ανεδαφικό και κενό νοήματος, όπως θέλουν να το παρουσιάζουν μερικοί από τους κριτικούς (για παράδειγμα ο J. Webb). Άλλωστε, θα διαπιστώσουμε παρακάτω ότι και ο ίδιος ο Gödel υπερασπίστηκε τη χρήση των αποτελεσμάτων του για φιλοσοφικούς σκοπούς σε μία ομιλία του {#03a} η οποία αν και είναι παλαιότερη των περισσότερων άρθρων που παρουσιάστηκαν, παρέμεινε ανέκδοτη μέχρι πρότινος και γι' αυτό το λόγο δεν παρουσιάστηκε μέχρι τώρα. Θα δούμε όμως στο επόμενο κεφάλαιο, πως με αυτή την ομιλία ο Gödel αναφέρεται και στο ζήτημα που κυρίως μας ενδιαφέρει, αφού τελειώσουμε με την ανακεφαλαίωση και θυμηθούμε τα κύρια σημεία του.

Με βάση λοιπόν τις απόψεις και τα επιχειρήματα που έχουν παρουσιαστεί μέχρι τώρα, μπορούμε να διαπιστώσουμε ότι ο τίτλος του κάθε υποκεφαλαίου (4.1 και 4.2) του κεφαλαίου στο οποίο βρισκόμαστε, περιγράφει ουσιαστικά τις δύο μεγάλες κατηγορίες κριτικής πάνω στο επιχείρημα του Lucas. Η πρώτη κατηγορία παρουσιάζει και τη μεγαλύτερη ποικιλία σε απόψεις, διότι, μη δεχόμενη τη χρήση των θεωρημάτων του Gödel για φιλοσοφικούς σκοπούς ή απορρίπτοντας μερικές εκ των βασικών υποθέσεων, είναι ανοιχτή σε όλα τα πεδία αντιπαράθεσης. Η δεύτερη κατηγορία είναι αυτή που απορρίπτει τη συλλογιστική του Lucas, όχι λόγω της χρήσης, αλλά λόγω της *λάθος* χρήσης του θεωρήματος του Gödel. Σε αυτή την κατηγορία διαπιστώνουμε ότι γίνονται και οι περισσότερες αναδιατυπώσεις των αρχικών επιχειρημάτων με τέτοιο τρόπο ώστε να εκφράζονται καλύτερα οι απόψεις του κάθε κριτικού. Παρατηρούμε επίσης την τάση αυτών των ανακατασκευών προς μία μεγαλύτερη μαθηματική (και λογική) ακρίβεια που θα 'πρέπει' να διέπει το φιλοσοφικό λόγο που σχετίζεται με το θεώρημα του Gödel.

Μία τέτοια (νέα) ανακατασκευή θα δούμε εν συντομία αμέσως τώρα, μέσω μίας σειρά σύντομων άρθρων των D. Lewis και J.R. Lucas, με τη βοήθεια επίσης των οποίων θα θυμηθούμε και τα κύρια σημεία της κριτικής πάνω στο βασικό επιχείρημα του Lucas.

Στο πρώτο άρθρο του Lewis {#14} γίνεται η προσπάθεια ‘καλούπωσης’ της διαδικασίας που ακολουθεί ο Lucas για την κατάρριψη του Μηχανισμού, μέσω κάποιων βασικών αξιωμάτων και ενός «άπειρου κανόνα απαγωγής» ο οποίος από ένα σύνολο προτάσεων, βγάζει σαν αληθή την πρόταση της συνέπειας που αφορά αυτό το σύνολο. Το σύστημα που προκύπτει (‘Αριθμητική του Lucas’), αποτελεί ένα καλά δομημένο σύστημα μέσω του οποίου ο Lucas μπορεί να παράγει θεωρήματα κινούμενος εντός των ορίων του. Το πρόβλημα όμως κατά τον Lewis είναι ότι ο Lucas δεν μπορεί να ελέγξει την αλήθεια όλων των εξαγομένων προτάσεων-θεωρημάτων της αριθμητικής του, οπότε τελικά, μπορεί να είναι όντως μία μηχανή της οποίας απλά τα εξαγόμενα είναι μπορεί να είναι πολύ περισσότερα από αυτά μιας μηχανής που παράγει θεωρήματα της αριθμητικής του Peano.

Βλέπουμε εδώ την προσπάθεια να αποκτήσει μία καλύτερη (πιο μαθηματική) βάση το επιχείρημα του Lucas, προσπάθεια που έγινε επίσης από τον Putnam αλλά και τον Benacerraf. Το τελικό συμπέρασμα μοιάζει επίσης με αυτό του Benacerraf, οπότε δεν εκπλήσσει το γεγονός ότι η απάντηση του Lucas {#16} είναι στην ουσία μία αναδιατύπωση των θέσεών του με έμφαση στο διαλεκτικό επιχείρημα και άρα στη δυνατότητά του να παράγει αληθείς προτάσεις όποτε χρειάζεται, χωρίς απαραίτητα να χρησιμοποιεί ολόκληρη την αριθμητική του. Φαίνεται λοιπόν ότι γενικότερα, ο τρόπος με τον οποίο αναδιατυπώνονται οι αρχικές θέσεις του Lucas, στην προσπάθειά του να είναι πιο αυστηρός, πάντα χάνει κάτι από την ουσία της διαλεκτικής προσέγγισης, άρα η αντιπαράθεση εκφυλίζεται σε υπεράσπιση των βασικών θέσεων του καθένα από τους αντιμαχόμενους.

Επανερχόμαστε στην αντιπαράθεση των δύο, η οποία συνεχίζεται με τον Lewis {#22} να προσπαθεί να πάρει πιο σοβαρά (και να απορρίψει τελικά) αυτή τη φορά το *διαλεκτικό* επιχείρημα μέσω μίας προσέγγισης που ουσιαστικά υλοποιεί μία ιδέα του C. Chihara, όπως αυτή παρουσιάστηκε στη σελ. 112 (βλ. παραπάνω). Η ιδέα αυτή είναι βασικά ότι το πιθανό εξαγόμενο (αριθμητικό) σύνολο του Lucas είναι *μεταβλητό*, αναλόγως εάν ο Μηχανιστής τον προκαλέσει ή όχι.

Έστω λοιπόν O_L το εξαγόμενο σύνολο σε κατάσταση αδράνειας και O_L^M το εξαγόμενο σύνολο του Lucas όταν αυτός προκαλείται από τον Μηχανιστή. Προφανώς το O_L^M ισούται με το O_L + την Gödel-πρόταση ϕ_M που αντιστοιχεί στη μηχανή M η οποία κατά τον Μηχανιστή αποτελεί το μοντέλο του Lucas. Πρέπει σε αυτό το σημείο να επισημανθεί ότι η ϕ_M θεωρείται ακριβέστερα η πρόταση που εκφράζει τη συνέπεια της μηχανής M , και όχι η αυθεντική Gödel-πρόταση. Ο Lewis υποθέτει επίσης την ύπαρξη μίας μηχανής M , της οποίας το εξαγόμενο σύνολο O_M αλλάζει και γίνεται O_M^N όταν αυτή ‘κατηγορείται’ ότι αναπαρίσταται από μία άλλη μηχανή N . Με τις επιπλέον υποθέσεις ότι κάθε πρόταση του O_L είναι αληθής και ότι το O_L περιλαμβάνει την αριθμητική του Peano, ο Lucas ισχυρίζεται ότι, υποθέτοντας ότι $O_L = O_M$ και $O_L^M = O_M^M$, ήτοι ότι αυτός είναι η M , καταλήγουμε σε άτοπο.

Ο Lewis όμως αμφισβητεί την εγκυρότητα αυτού του αποτελέσματος εφόσον ακόμα δεν έχει ξεκαθαριστεί αν η ϕ_M ανήκει στο O_M ή στο O_M^M . Στην πρώτη όμως περίπτωση, προκύπτει ότι η ϕ_M είναι αληθής πρόταση που περιλαμβάνεται στο O_L^M αλλά και στο O_M^M , κάτι το οποίο είναι απόλυτα θεμιτό, ενώ στη δεύτερη, η ϕ_M είναι όντως μη-αποδείξιμη στα πλαίσια του O_M^M αλλά δεν υπάρχει εγγύηση ότι είναι αληθής στο O_L^M , παρά μόνο στο υποσύνολό του, O_L .

Ο Lucas απαντά {#24} ότι και οι δύο περιπτώσεις είναι αδύνατες. Στην πρώτη περίπτωση, εφόσον η ϕ_M ανήκει στο O_M , δεν είναι πλέον το O_M που συγκρίνεται με το O_L^M , αλλά το O_M^M , οπότε το επιχείρημα πρέπει να ξαναρχίσει από την αρχή για τη νέα μηχανή O_M^M , η οποία όμως θα έχει μία άλλη Gödel-πρόταση, ας πούμε ϕ_M^M κ.ο.κ. Η δεύτερη περίπτωση είναι με τη σειρά της αδύνατη διότι αν υποτεθεί ότι όντως δεν μπορούμε να δούμε την αλήθεια της ϕ_M , η οποία τώρα είναι η πρόταση για τη συνέπεια του O_M^M , το οποίο ισούται με το O_L^M από υπόθεση, τότε δεν μπορούμε επίσης να διαπιστώσουμε τη συνέπεια του O_L^M . Αλλά αυτό μπορούμε όντως να το κάνουμε, διότι αν υποτεθεί το O_L^M μη-συνεπές, τότε αυτό θα σήμαινε ότι υπάρχει στην O_L μία απόδειξη για την άρνηση της (αρχικής) ϕ_M , άρα η M θα ήταν μη-συνεπής και άρα όχι ένα μοντέλο του νου για τους λόγους που έχουν ήδη αναφερθεί.

Η παραπάνω ανάλυση έγινε διότι, αν μεταφραστεί σε απλά ελληνικά, μας δίνει όλη την επιχειρηματολογία των κεφαλαίων 4.1 και 4.2 σε μία σελίδα! Αν δούμε, για παράδειγμα πιο προσεκτικά την πρώτη περίπτωση που αναλύει ο Lewis, παρατηρούμε ότι πρόκειται ουσιαστικά για το επιχείρημα που λέει ότι αν μπορεί να βρεθεί η Gödel-πρόταση μίας μηχανής τότε μπορούμε να τη συμπεριλάβουμε σε μία νέα μηχανή (Good κ.ά.). Στη δεύτερη περίπτωση, έχουμε ουσιαστικά το επιχείρημα του Putnam, ήτοι ότι ναι μεν δεν μπορεί η μηχανή να δει την αλήθεια της Gödel-πρότασης, αλλά ούτε και ο άνθρωπος μπορεί. Ο Lucas απαντά βέβαια σε κάθε περίπτωση αναλόγως με όσα έχουμε επισημάνει μέχρι τώρα.

Ένα ακόμη ενδιαφέρον στοιχείο που προκύπτει από τα παραπάνω και μία τάση που διαγράφει όλη η πορεία της αντιπαράθεσης είναι η στροφή προς το 2^ο θεώρημα της μη-πληρότητας για να καταρριφθεί το αντι-μηχανιστικό συμπέρασμα που προκύπτει από το 1^ο. Αυτή η στροφή του ενδιαφέροντος γίνεται άμεσα αντιληπτή επισκοπώντας ολόκληρο το κεφάλαιο 4 και παρατηρώντας την στροφή των επιχειρημάτων προς την κατεύθυνση της συνέπειας. Πράγματι το θέμα της συνέπειας του ανθρώπου είναι ένα λεπτό θέμα, το οποίο τελικά ανάγεται σε θέμα πίστης. Απέναντι στην άποψη ότι οι άνθρωποι μπορεί τελικά να είναι μη-συνεπείς μηχανές (Putnam, G. Lee Bowie - #23) το μόνο που μπορεί να αντιταχθεί είναι η πίστη στη συνέπεια του ανθρώπου όπως αυτή έχει εκφραστεί από τον Lucas, με βάση το γεγονός ότι όπου εμφανίζεται η μη-συνέπεια ο άνθρωπος δεν την ανέχεται⁽³⁾.

Εδώ ολοκληρώνεται η παρουσίαση της πρώτης περιόδου αντιπαράθεσης για τις πιθανές συνέπειες του θεωρήματος του Gödel πάνω στο Μηχανισμό, αντιπαράθεση που αποδείχθηκε εξαιρετικά γόνιμη σε (κυρίως κριτικές) ιδέες. Η γονιμότητα αυτή μπορεί να μη σημαίνει βέβαια από μόνη της τίποτα, ή μπορεί τελικά να καταδεικνύει την αδυναμία {Bowie - #23} ή την ορθότητα {Lucas - #27} του επιχειρήματος έναντι στον Μηχανισμό. Και μπορεί ο καθένας να έχει τη δική του άποψη πάνω σε αυτό, δε μπορεί να αρνηθεί όμως κανείς το ενδιαφέρον που παρουσιάζει η συζήτηση και να μην είναι πεπεισμένος ότι το θεώρημα του Gödel αποτελεί, αν μη τι άλλο, μία ξεχωριστή περίπτωση μαθηματικού αποτελέσματος.

(3): Ακόμη όμως και αυτό αμφισβητείται από τον A. Hutton {#19}, ο οποίος καταλήγει στο ενδιαφέρον συμπέρασμα ότι, με την υπόθεση της συνέπειας, δε μπορεί να ξέρει κάποιος αν είναι συνεπής ή όχι. Δυστυχώς, η μέθοδος που ακολουθεί (χρησιμοποιώντας πιθανοθεωρητικές έννοιες και στατιστική!) είναι τόσο αμφιλεγόμενη που το αποτέλεσμά του χάνει κάθε ισχύ.

5. Οι απόψεις του Gödel για το θέμα

Διακόπτεται σε αυτό το σημείο η ιστορική πορεία παρουσίασης που έχει επιλεχθεί για να δούμε τις απόψεις του εμπνευστή των θεωρημάτων μη-πληρότητας πάνω στο θέμα που μας απασχολεί. Η κύρια πηγή από όπου αντλούνται αυτές οι απόψεις είναι μία ομιλία του Gödel στο πανεπιστήμιο Brown όπου συνεδρίασε στις 26 Δεκεμβρίου του 1951 η Αμερικάνικη μαθηματική κοινότητα, ομιλία που είναι πιο γνωστή σαν ‘η ομιλία του Gibbs’ [Gibbs’ Lecture]. Οι λόγοι για τους οποίους δεν παρουσιάστηκε πιο πριν αυτή η πηγή είναι ότι, αφ’ ενός, μέχρι ενός χρονικού σημείου δεν είχε εκδοθεί η ομιλία, οπότε δεν συμπεριλαμβανόταν σε καμία επιχειρηματολογία, και αφ’ ετέρου διότι για να παρακολουθήσουμε τις απόψεις του Gödel πρέπει να έχουμε ήδη μία καλή εικόνα και κατανόηση των θεμάτων και εννοιών που εμπλέκονται. Θα πρέπει βέβαια να σημειώσουμε πως ο Gödel δεν πήρε μέρος σε καμία διαμάχη όπως αυτή που παρουσιάστηκε παραπάνω και οι απόψεις που εκφράζει κινούνται σε άλλα πλαίσια από αυτά που είδαμε μέχρι τώρα. Παρ’ όλα αυτά είναι σημαντική η οπτική γωνία που μας προσφέρεται μέσω αυτών των απόψεων, για το απλό λόγο ότι προέρχονται από τον άνθρωπο ο οποίος ίσως καλύτερα από τον καθένα είχε εντρυφήσει στα θεωρήματα μη-πληρότητας και τη συλλογιστική από την οποία αυτά προέκυψαν.

Το πρώτο πράγμα που γίνεται αμέσως φανερό από την ομιλία του Gibbs είναι ότι ο Gödel πίστευε ότι τα αποτελέσματα της μη-πληρότητας της αριθμητικής έχουν σοβαρές μαθηματικές συνέπειες στα θεμέλια των μαθηματικών αλλά και σοβαρότατο φιλοσοφικό αντίκτυπο πάνω στο ίδιο ζήτημα. Το κύριο μέλημά του σε αυτή την ομιλία είναι να καταδείξει τις σημαντικότερες από αυτές τις συνέπειες, με απώτερο στόχο να δείξει την αδυναμία της θεωρίας του ιντουσιονισμού να περιγράψει ικανοποιητικά τις θεμελιώδεις μαθηματικές έννοιες. Πρέπει δηλαδή να έχουμε στο μυαλό μας σε σχέση με όσα αναλυθούν παρακάτω, ότι ο Gödel προσπαθεί να υπερασπιστεί τελικά την ρεαλιστική/πλατωνική μαθηματική κοσμοθεωρία, και ότι οι απόψεις που μας ενδιαφέρουν αποτελούν ένα βήμα προς αυτή την κατεύθυνση. Αυτό βέβαια δε σημαίνει ότι από μόνες τους δεν έχουν αξία, απλά προσπαθούμε να δώσουμε το γενικότερο πλαίσιο στο οποίο κινούμαστε και να τονίσουμε τη διαφορά του από αυτό που συναντήσαμε μέχρι τώρα.

Η κεντρική θέση του Gödel, γύρω από την οποία κινούνται και όλα όσα αναλύει στην ομιλία του, είναι ότι φαίνεται να προκύπτει εύλογα από διάφορα μετα-μαθηματικά αποτελέσματα το *μη-πλήρες* και το *ανεξάντλητο* των μαθηματικών. Αυτή η θέση στηρίζεται τόσο στα δικά του αποτελέσματα για τη μη-πληρότητα, όσο και από τις προσπάθειες αξιωματικής θεμελίωσης της θεωρίας συνόλων, προσπάθειες που φαίνονται να οδηγούν όντως σε ανεξάντλητες (παρ' όλα αυτά δημιουργικές – μη-αλγοριθμικές) διαδικασίες. Τα θέματα της θεωρίας συνόλων βέβαια δεν είναι το κύριο κομμάτι του ενδιαφέροντός μας, οπότε περνάμε να δούμε πως ο Gödel χρησιμοποιεί τα δύο θεωρήματά του προς την θεμελίωση της φιλοσοφικής του θέσης.

Με τη χρήση της έννοιας της μηχανής Turing και ενός είδους αναλογίας διοφαντικών εξισώσεων και κατηγορημάτων, όπως περιγράφεται σύντομα στο Παράρτημα (σελ. xxv), το 1^ο θεώρημα μη πληρότητας αναδιατυπώνεται ως εξής: «Οποιοδήποτε καλά ορισμένο σύστημα αξιωμάτων και κανόνων απαγωγής επιλεγεί, πάντα θα υπάρχουν διοφαντικά προβλήματα του τύπου που περιγράφηκε που είναι μη-αποκρίσιμα σε αυτό το σύστημα, δεδομένου μόνο ότι δεν προκύπτει μέσω του συστήματος καμία ψευδής πρόταση αυτού του τύπου»⁽¹⁾. Το 'καλά ορισμένο σύστημα' που περιγράφει ο Gödel στη συνέχεια, δείχνεται ότι είναι ισοδύναμο με μία μηχανή Turing η οποία θα μπορεί να καταγράψει όλες τις προτάσεις που μπορούν να προκύψουν μέσα από τα αξιώματα και τους κανόνες του συστήματος. Μέσω βέβαια μίας τέτοιας αναλογίας, το 1^ο θεώρημα μη-πληρότητας είναι ισοδύναμο με το γεγονός ότι «δεν υπάρχει πεπερασμένη διαδικασία για τη λύση του προβλήματος απόφασης που έχει να κάνει με τις διοφαντικές εξισώσεις του τύπου που περιγράφηκε».

Αν δεχθούμε τώρα ότι ο ανθρώπινος νους είναι ικανός να αντιληφθεί και να φέρει εις πέρας συλλογιστικές διαδικασίες που είναι εν δυνάμει άπειρες, αυτό το αποτέλεσμα μοιάζει να *συνηγορεί* με το συμπέρασμα του J.R. Lucas για τις πιθανές συνέπειες του θεωρήματος. Πάνω σε αυτή την ιδέα όμως θα επεκταθούμε αργότερα καθότι ο Gödel, σε αντίθεση με τον Lucas, δεν ασχολείται παραπάνω με τις συνέπειες του 1^{ου} θεωρήματος μη-πληρότητας, αλλά στρέφει κυρίως την προσοχή του στο 2^ο. Άλλωστε, περισσότερο ταιριάζει αυτή η διατύπωσή στο βασικό επιχείρημα του R. Penrose, το οποίο όμως θα δούμε και θα αναλύσουμε στο επόμενο κεφάλαιο, όταν οι δυνατότητες της πεπερασμένης/αλγοριθμικής διαδικασίας, θα αποκτήσουν έναν πιο ουσιαστικό ρόλο στην αντι-μηχανιστική επιχειρηματολογία.

(1): Προφανώς αυτή η διατύπωση είναι απόλυτα ισοδύναμη με αυτή που διατυπώθηκε στο 2^ο μέρος της εργασίας.

Κατά τον Gödel λοιπόν, το 2^ο θεώρημα μη-πληρότητας, σύμφωνα με το οποίο «...για οποιοδήποτε καλά ορισμένο σύστημα αξιωμάτων και κανόνων η πρόταση που καταδεικνύει την συνέπειά του είναι μη-αποδείξιμη⁽²⁾...αρκεί να είναι συνεπές και να επαρκεί για παράγει ένα μέρος της περατοκρατικής αριθμητικής των ακεραίων», είναι που δείχνει με τον πιο προφανή τρόπο την μη-πληρότητα των μαθηματικών. Διότι, «καθίσταται αδύνατο για κάποιον να φτιάξει ένα τέτοιο σύστημα και να κάνει την παρακάτω διαπίστωση με συνεπή τρόπο: ‘Αντιλαμβάνομαι (με μαθηματική βεβαιότητα) ότι όλα αυτά τα αξιώματα και οι κανόνες που έχω ορίσει είναι σωστά, και περαιτέρω πιστεύω ότι περιλαμβάνουν όλα τα μαθηματικά.’ Αν κάποιος διατυπώσει αυτή την πρόταση, αντιφάσκει. Διότι αν αντιλαμβάνεται τα αξιώματα για τα οποία συζητούμε σαν σωστά, αντιλαμβάνεται επίσης (με την ίδια βεβαιότητα) ότι είναι συνεπή. Άρα έχει μία μαθηματική επίγνωση που δεν παράγεται από τα αξιώματά του.»

Την παραπάνω συλλογιστική πορεία την έχουμε επανειλημμένα συναντήσει, όμως ο Gödel δεν αρκείται σε αυτό το συμπέρασμα, ή μάλλον σπεύδει να ξεκαθαρίσει ότι αυτή η εγγενής μη-πληρότητα αναφέρεται στα μαθηματικά με την *αντικειμενική έννοια*, ήτοι στα μαθηματικά που περιλαμβάνουν όλες τις *αληθείς* μαθηματικές προτάσεις. Αν όμως, από την άλλη, θεωρήσει κάποιος τα μαθηματικά με την *υποκειμενική έννοια*, δηλαδή αυτά που περιέχουν μόνο τις *αποδείξιμες* μαθηματικές προτάσεις, τότε «δεν αποκλείεται ότι μπορεί να υπάρχει ένας πεπερασμένος κανόνας ο οποίος θα παράγει όλα τα προφανή του [Σημ: εννοεί ‘του συστήματος’ – το οποίο περιλαμβάνει όλα τα υποκειμενικά μαθηματικά] αξιώματα». Σε αυτή όμως την τελευταία περίπτωση, επισημαίνει ο Gödel, δεν θα μπορούμε να ξέρουμε με την ανθρώπινη κατανόησή μας ότι αυτός ο κανόνας υπάρχει και ότι όντως μπορεί να παράγει νέα αξιώματα που να είναι σωστά! Θα μπορούμε μόνο *εκ των υστέρων* να διαπιστώνουμε την ορθότητα αυτών των αξιωμάτων, και μόνον εμπειρικά ίσως θα μπορούσαμε να παράγουμε έναν επαγωγικό κανόνα για την ορθότητά τους. Αυτό όμως το γεγονός είναι συμβατό με το ότι «ο ανθρώπινος νους (στο πεδίο των *καθαρών μαθηματικών*) είναι ισοδύναμος με μία πεπερασμένη μηχανή η οποία όμως δεν μπορεί να αντιληφθεί πλήρως τη λειτουργία της». [Σημ: Παρατηρούμε εδώ την προφανή αντιστοιχία του αποτελέσματος με αυτό του P. Benacerraf, βλ. § 4.2.1].

(2): Συγκεκριμένα εδώ ο Gödel χρησιμοποιεί τον όρο ‘μη-αποδείκτικός’ [undemonstrable], ο οποίος όμως μπορεί να θεωρηθεί ισοδύναμος με τον όρο ‘μη-αποδείξιμος’.

Έχουμε λοιπόν τελικά δύο δυνατότητες θεώρησης των μαθηματικών και από τη μεν αντικειμενική συμπεραίνουμε ότι τα μαθηματικά είναι μη-πλήρη, ενώ από την υποκειμενική προκύπτει η ισοδυναμία νου και μηχανής Turing η οποία με τη σειρά της οδηγεί στο συμπέρασμα ότι, «όχι μόνο τα αντικειμενικά μαθηματικά δεν μπορούν να συμπεριληφθούν στην ολότητά τους σε ένα αξιωματικό σύστημα, αλλά επιπλέον ότι υπάρχουν *απολύτως* μη-επιλύσιμα⁽³⁾ διοφαντικά προβλήματα του τύπου που περιγράφηκε παραπάνω». Έτσι, το τελικό συμπέρασμα του Gödel είναι *διαζευκτικό* και έχει την παρακάτω μορφή:

Είτε ο ανθρώπινος νους μπορεί άπειρα να ξεπερνά τις δυνατότητες οποιασδήποτε πεπερασμένης μηχανής, ή υπάρχουν απολύτως μη-επιλύσιμα διοφαντικά προβλήματα του τύπου που έχει αναφερθεί.

Το συμπέρασμα αυτό φαίνεται στον Gödel ότι έχει μεγάλο φιλοσοφικό ενδιαφέρον και πράγματι μπορεί να οδηγήσει σε διάφορες σκέψεις υπέρ ή κατά της μηχανιστικής θεώρησης. Όμως ο ίδιος δε δίνει σαφή απάντηση για το ποια από τις δύο εναλλακτικές πιστεύει ότι είναι περισσότερο εφικτή. Μάλλον χρησιμοποιεί και τις δύο για να δείξει με πειστικό τρόπο, ότι, υιοθετώντας είτε τη μία, είτε την άλλη άποψη, οδηγούμαστε τελικά σε μία πλατωνική θεώρηση των μαθηματικών πραγμάτων. Διότι όπως διαπιστώνει, από τη μία δεν μπορεί να αναχθεί η λειτουργία του νου στις διεργασίες του μυαλού και από την άλλη, φαίνεται ότι τα μαθηματικά δεν μπορεί να είναι δημιούργημα του ανθρώπου, καθότι ο δημιουργός οφείλει να γνωρίζει όλες τις ιδιότητες των δημιουργημάτων του, ήτοι δεν μπορούν αυτά να έχουν ιδιότητες πέρα από αυτές που τις έχει προσδώσει.

Με αυτή την τακτική ο Gödel αφήνει ουσιαστικά το θέμα τελείως ανοιχτό. Μάλιστα, από το διαζευκτικό συμπέρασμα και μέσα από τα σχόλια που κάνει στο υπόλοιπο μέρος του άρθρου, έχουν βγει όλων των ειδών τα συμπεράσματα. Ο Lucas {#35} υποστηρίζει ότι ο Gödel τελικά απορρίπτει τη δεύτερη εναλλακτική, άρα η θέση του είναι κατά της μηχανιστικής θεώρησης, ο G. Boolos πιστεύει ότι η δεύτερη εναλλακτική είναι η πιθανότερη να ισχύει και ότι το διαζευκτικό συμπέρασμα στέκεται σε πλήρη αντίθεση με τα όποια συμπεράσματα ενάντια στη μηχανιστική

(3): Σαν 'απολύτως μη-επιλύσιμα' προβλήματα θεωρούνται αυτά τα προβλήματα που είναι μη-αποκρίσιμα, όχι μόνο στα πλαίσια ενός συστήματος, αλλά μέσω οποιασδήποτε μαθηματικής απόδειξης που μπορεί να σκεφτεί ανθρώπινος νους.

θέση κατά καιρούς έχουν βγει (όπως προκύπτει από το εισαγωγικό σημείωμα που προηγείται της έκδοσης της ομιλίας Gibbs), ενώ ο M. Davis {#29} πιάνεται από το σχόλιο που είναι ανάλογο με το συμπέρασμα του Benacerraf (βλ. προ-προηγούμενη σελίδα) για να διαπιστώσει ότι καταρρίπτεται το αντι-μηχανιστικό επιχείρημα του Penrose που έχει σαν βάση και πάλι το θεώρημα του Gödel.

Για να διαπιστώσουμε τι από όλα αυτά βρίσκεται πιο κοντά στις πεποιθήσεις του Gödel αναγκαστικά θα πρέπει να καταφύγουμε σε άλλες πηγές που είναι μεν λίγες, μπορούν όμως να ρίξουν λίγο φως στην υπόθεση. Καταρχήν αναφέρουμε ένα σχόλιό του που βρίσκεται στις *‘Παρατηρήσεις πάνω στα προβλήματα των μαθηματικών στο συνέδριο για τα 200 χρόνια του Princeton’* {#03b}, σύμφωνα με το οποίο οι (μηχανιστικές) απόψεις του Turing δεν είναι αρκετά πειστικές. Ο λόγος, κατά τον Gödel, είναι ότι ο νους δεν είναι στατικός και σε αντίθεση με μία μηχανή πεπερασμένων καταστάσεων αναπτύσσεται διαρκώς μέσω ιδεών που προέρχονται από τη διαίσθηση και όχι κάποια αυστηρώς δομημένη διαδικασία. Σαν παράδειγμα τέτοιας χρήσης του νου φέρνεται η εισαγωγή νέων αξιωμάτων στη θεωρία συνόλων, μία διαδικασία που είναι μεν επαναληπτική, οπότε δεν υπάρχει λόγος να υποθέσουμε ότι δεν μπορεί να επεκταθεί μέχρι το άπειρο, αλλά δεν είναι κάθε φορά αυστηρά καθορισμένη, δηλαδή δεν παράγεται μέσω κάποιας μηχανιστικής διαδικασίας. Παρ’ όλα αυτά, ο ίδιος ο Gödel, σε συζητήσεις με τον H. Wang⁽⁴⁾ παραδέχεται ότι το επιχείρημα του Turing γίνεται έγκυρο με δύο ακόμη παραδοχές που είναι γενικά αποδεκτές με βάση τη σύγχρονη θεώρηση των πραγμάτων, και οι οποίες είναι οι εξής: 1) ότι δεν υπάρχει νους χωριστά από τη ύλη και 2) ότι οι φυσικοί νόμοι έχουν ένα πεπερασμένο όριο ακρίβειας. Από την τελευταία παραδοχή φαίνεται να προκύπτει ότι ο εγκέφαλος λειτουργεί όντως σαν ψηφιακός υπολογιστής, οπότε έχουμε την επιβεβαίωση των ‘πιστεύω’ του Turing. Η αναλογία των δύο προτάσεων γίνεται φανερή αν (μαζί με τον Gödel) αποδεχθούμε τη διαφοροποίηση που υφίσταται η έννοια του εγκεφάλου από αυτή του νου, με την έννοια του δυϊσμού ανάμεσα στις φυσικές διεργασίες (κινήσεις σωματιδίων) και τις νοητικές διεργασίες (σκέψη) του μυαλού. Αυτή η ιδέα είναι ενδιαφέρουσα και φαίνεται ότι ο Gödel της δίνει αρκετή βαρύτητα, γεγονός που βέβαια πηγάζει από τις ευρύτερες φιλοσοφικές του αντιλήψεις. Με αυτή αποδοχή αυτού του δυϊσμού πάντως, αποδεχόμενοι την παραδοχή (1) μπορούμε να καταλάβουμε καλύτερα πως μαζί με τη (2) προκύπτει

(4): βλ. Hao Wang *From Mathematics to Philosophy*. London Routledge & Kegan Paul, σελ. 324-326

η αναλογία που αναφέραμε πιο πριν. Σε ένα τέτοιο πλαίσιο βέβαια έχει κάποια βάση το μηχανιστικό επιχείρημα, αν και πάλι ο Gödel 'θολώνει το τοπίο' θεωρώντας ότι η παραδοχή (1) αποτελεί προκατάληψη της εποχής μας η οποία θα απορριφθεί, άρα και πάλι θεωρεί ότι δε μπορούμε να δεχθούμε αβίαστα τις θέσεις του Turing.

Τελικά, το μόνο ασφαλές συμπέρασμα που μπορεί κανείς να βγάλει όσον αφορά τη θέση του Gödel για τις δυνατότητες του νου σε σχέση με τις δυνατότητες των μηχανιστικών διαδικασιών είναι ότι, εφόσον και οι δύο εναλλακτικές της βασικής του θέσης εξυπηρετούν το σκοπό για τον οποίο τις διατύπωσε, δεν υπήρχε η ανάγκη να πάρει μία σαφή θέση. Και πράγματι φαίνεται ότι ανάλογα με τη χρονική περίοδο ή την ιδέα που ήθελε να αναπτύξει, είτε δεχόταν τη μία και απέρριπτε την άλλη, είτε δεχόταν και τις δύο για διαφορετικούς λόγους την καθεμία, ή (κυρίως) χρησιμοποιούσε αποκλειστικά τη διαζευκτική μορφή για να επεξεργαστεί μετά τα συμπεράσματα που προέκυπταν από την απόρριψη ή την αποδοχή της διάζευξης. Επίσης ενδεικτικό της έλλειψης ενδιαφέροντος του Gödel για το καθ' εαυτό θέμα που μας απασχολεί σε αυτό το 3^ο μέρος της εργασίας είναι το γεγονός ότι, αν και οι απόψεις του Lucas και οι πρώτες αντιρρήσεις εκφράστηκαν όσο ακόμα ζούσε, δεν έχει καταγραφεί καμία συγκεκριμένη αντίδραση από τη μεριά του.

Ολοκληρώνοντας το παρόν κεφάλαιο αξίζει να παρατηρήσουμε τη σύνδεση του διαζευκτικού συμπεράσματος με τα τελευταία σχόλια πάνω στη σύνδεση της θεωρία αριθμών και της θεωρίας συνόλων που βρίσκονται στη σελ. xxxvii του Παραρτήματος. Με βάση σύγχρονα του Gödel αποτελέσματα, βλέπουμε ότι, με την υπόθεση της ύπαρξης – σαν αξίωμα – ενός μη προσβάσιμου διατακτικού αριθμού μεγαλύτερου του $\aleph_0$, ύπαρξη που είναι πολύ πιθανό να υφίσταται, καταλήγουμε όντως στην αποδοχή της ύπαρξης ενός μη-επιλύσιμου διοφαντικού προβλήματος του τύπου που περιγράφει ο Gödel. Έτσι φαίνεται πολύ πιθανό να ισχύει η δεύτερη εναλλακτική της διάζευξης, κάτι που επισημαίνει και ο Boolos, χωρίς βέβαια αυτό να σημαίνει ότι δεν είναι πιθανό να ισχύει *μαζί* με την πρώτη εναλλακτική. Τώρα βέβαια, το αν από το διαζευκτικό επιχείρημα μαζί με την παραπάνω διαπίστωση μπορεί να προκύψει ένα επιχείρημα υπέρ του Μηχανισμού, εξαρτάται άμεσα και από την εγκυρότητα της συλλογιστικής του Gödel μέσω της οποίας φτάνει σε αυτή τη διάζευξη. Το θέμα όμως αυτό είναι ανοιχτό σε συζήτηση, όπως είναι και το ευρύτερο θέμα των απόψεων του Gödel, άρα κανείς δεν μπορεί με ασφάλεια να χρησιμοποιήσει την ομιλία του Gibbs για να εκμαιεύσει απόλυτα επιχειρήματα υπέρ ή κατά κάποιας θέσης που αφορά τη σύνδεση νοητικών και μηχανιστικών λειτουργιών.

6. Δεύτερη περίοδος αντιπαράθεσης (1989 - σήμερα)

Επανερχόμαστε τώρα στην παρουσίαση απόψεων που σχετίζονται άμεσα με το θέμα της δυνατότητας μηχανιστικής αναπαράστασης του νου και περνάμε στην επόμενη ιστορική περίοδο που ασχολείται με αυτό και επεκτείνεται μέχρι τις μέρες μας. Σε αυτή την περίοδο είναι προφανές ότι η ορολογία που χρησιμοποιούνταν μέχρι τώρα έχει αλλάξει, με βάση τις εξελίξεις τόσο στα θεωρητικά θέματα, όπως στη θεωρία υπολογιστότητας (βλ. Μέρος I, σελ. 56), όσο και στην πρακτική τους εφαρμογή, με τις έρευνες πάνω στον τομέα της Τεχνητής Νοημοσύνης. Θα διαπιστώσουμε ωστόσο, ότι με την κατάλληλη αναλογία των σημερινών με των τότε εννοιών, τα επιχειρήματα τελικά δεν διαφέρουν και τόσο στην ουσία τους, μάλιστα θα μπορούσε να πει κανείς ότι μερικά από αυτά είναι απόλυτα ίδια με τα αντίστοιχά τους παλαιότερα και η ίδια διαπίστωση μπορεί φυσικά να επεκταθεί και στον τομέα των συμπερασμάτων.

Όπως και στην πρώτη περίοδο αντιπαράθεσης, ο βασικός μας άξονας θα είναι και πάλι ένα επιχείρημα ενάντια στη μηχανιστική θέση, όπως αυτό εκφράζεται πλέον από τον Roger Penrose στα δύο βιβλία του *'The Emperor's New Mind'* (1989) και (κυρίως στο) *'Shadows of the Mind'* (1994). Αντίστοιχα με το επιχείρημα του Lucas, αυτό του Penrose δεν συνάντησε μικρότερη κριτική, όπως φαίνεται από τις αντιρρήσεις που θα αναλύσουμε παρακάτω, αντιρρήσεις που είναι λιγότερες για το μόνο λόγο ότι το επιχείρημα είναι πιο πρόσφατο. Βέβαια, δεν θα πρέπει να δημιουργεί η εντύπωση ότι εκτός των Lucas και Penrose δεν υπάρχουν άλλοι υποστηρικτές των αντι-μηχανιστικών συνεπειών του θεωρήματος του Gödel, ή ότι ξαφνικά όλοι ασχολούνται με το νέο επιχείρημα και αυτό του Lucas ξεχάστηκε. Απλά, από τη μία, οι απόψεις των δύο και οι απαντήσεις τους στις διάφορες κριτικές καλύπτουν σε πολύ μεγάλο ποσοστό την επιχειρηματολογία υπέρ των συνεπειών του θεωρήματος του Gödel, ενώ από την άλλη, οι νεότερες κριτικές του επιχειρήματος του Lucas δεν έχουν να προσθέσουν σημαντικά πράγματα σε όσα έχουν ήδη διατυπωθεί και αναλυθεί. Αν πάντως μπορεί να βγει κάποια κρίση από τη μακροβιότητα της συζήτησης, αυτή θα πρέπει να είναι ότι το όλο θέμα φαίνεται να απασχολεί διαχρονικά την επιστημονική κοινότητα σε ένα πλαίσιο που ακροβατεί ανάμεσα στη φιλοσοφία και τις πιο απόμακρες περιοχές της μαθηματικής σκέψης.

Πριν προχωρήσουμε για να δούμε ποια ακριβώς είναι τα νέα επιχειρήματα του Penrose και προς ποια κατεύθυνση κινούνται οι νέες αντιρρήσεις είναι απόλυτα απαραίτητο να αναφέρουμε τη νέα ορολογία που χρησιμοποιείται και κυρίως την αναλογία της με την παλαιότερη. Εκτός από ότι αυτή η παρουσίαση θα μας βοηθήσει προφανώς να καταλάβουμε τα νέα επιχειρήματα καλύτερα, θα απλουστεύσει επίσης κατά πολύ την ευρύτερη ανάλυση, διότι οι ομοιότητες με τα παλιότερα επιχειρήματα θα γίνουν πιο φανερές, άρα θα αποφευχθεί η επανάληψη των ίδιων πραγμάτων με άλλα λόγια.

Αρχίζουμε με τη διαπίστωση ότι ο όρος ‘Μηχανισμός’ που κατά κόρον συναντήσαμε στην πρώτη περίοδο αντιπαράθεσης, απουσιάζει σχεδόν ολοκληρωτικά από την δεύτερη. Η αναφορά πλέον στη θεωρία που υποστηρίζει τη δυνατότητα μηχανιστικής αναπαράστασης του νου γίνεται μέσω του όρου ‘Υπολογιστικισμός’ [**Computationalism**] που είναι η θεωρία που υποστηρίζει την δυνατότητα υπολογισμού μέσω αλγορίθμου των διεργασιών του νου. Ο όρος προφανώς προέρχεται σαφώς από την ευρύτερη ορολογία της θεωρίας Υπολογιστότητας και περιγράφει ουσιαστικά το ίδιο πράγμα με το Μηχανισμό. Απλά, περιγράφει με μεγαλύτερη ακρίβεια το τι εννοείται με το ερώτημα αν ‘ο νους είναι ισοδύναμος με μία μηχανή’, αντικαθιστώντας με το ερώτημα ‘αν οι διεργασίες της σκέψης μπορούν να μετουσιωθούν σε έναν αλγόριθμο που θα μπορούμε κάποτε να υπολογίσουμε, και άρα να τον ενσωματώσουμε σε μία μηχανή’.

Με βάση τα παραπάνω επήλθε και η φυσική εξέλιξη της επιχειρηματολογίας που βασίζεται στο θεώρημα του Gödel. Ο Penrose δηλαδή προσπαθεί να δείξει ότι *δεν υπάρχει αλγόριθμος* μέσω του οποίου να μπορεί να δει κανείς την αλήθεια της Gödel-πρότασης, θέλοντας να καταρρίψει τη θεωρία του Υπολογιστικισμού. Φαίνεται ότι αυτή είναι μία πιο ξεκάθαρη απαίτηση από αυτή του Lucas, και μπορούμε να πούμε ότι γενικότερα οι εξελίξεις στις διάφορες θεωρίες που έχουν να κάνουν με αυτά τα θέματα, είναι ικανή να μας δώσει μία παραπάνω σαφήνεια. Για παράδειγμα, η σχέση μηχανών Turing με τυπικά συστήματα, σχέση που μέχρι τώρα ήταν γενικά αποδεκτή, αλλά όχι τελείως σαφής, εγκαθιδρύεται με το παρακάτω αποτέλεσμα:

Λήμμα του Craig: Για κάθε TM που αποδέχεται μία συγκεκριμένη κλάση προτάσεων, υπάρχει ένα αντίστοιχο αξιωματικό σύστημα που αποδέχεται την ίδια κλάση προτάσεων.

Από τη μία λοιπόν τα νέα δεδομένα εδραιώνουν καλύτερα και με πιο σαφή τρόπο τη βάση της συζήτησης, από την άλλη όμως χρησιμοποιούνται σε αυτή τη δεύτερη περίοδο και όροι οι οποίοι επιτελούν την αντίθετη λειτουργία. Συγκεκριμένα, ο Penrose χρησιμοποιεί την έννοια της *ορθότητας* στη θέση αυτής της συνέπειας, σαν την προϋπόθεση για να ισχύει το αποτέλεσμα του Gödel, κάτι δεν είναι τυπικά αθέμιτο (βλ. παρακάτω), αλλά θα δούμε ότι δυσχεραίνει σημαντικά την προσπάθεια αποσαφήνισης των νέων επιχειρημάτων του. [Σημ: Υπενθυμίζεται ότι η ορθότητα είναι η ιδιότητα μίας θεωρίας να παράγει σαν αληθή μία πρόταση την οποία έχει αποδείξει (για περισσότερες λεπτομέρειες βλ. Μέρος II, σελ. 25)]. Ίσως βέβαια η επιλογή αυτή να στόχευε στην αποφυγή των αντιρρήσεων που έχουν να κάνουν με το ζήτημα της συνέπειας του ανθρώπου ή της μηχανής που θα τον αναπαραστή (οι οποίες ήταν πάμπολλες όπως έχει ήδη διαπιστωθεί στο κεφ. 4), αλλά αν αυτός ήταν ο σκοπός του Penrose, τότε μάλλον απέτυχε η προσπάθειά του διότι ένα μέρος της συζήτησης και πάλι κατέληξε να περιστρέφεται γύρω από την ορθότητα, πλέον, του ανθρώπου, και το αν αυτή είναι αυταπόδεικτη ή απαραίτητη για μία μηχανή που θα μπορούσε να τον αναπαραστήσει κτλ. Για να γίνει όμως πιο σαφές το σκεπτικό σύμφωνα με το οποίο έγινε η συγκεκριμένη επιλογή, παραθέτουμε το σχετικό αποτέλεσμα της μαθηματικής λογικής που εγκαθιδρύει τη σχέση των δύο εννοιών:

Λήμμα⁽¹⁾: Ένα σύστημα Σ είναι συνεπές αν και μόνο αν το Σ είναι ορθό για τις προτάσεις που ανήκουν στην κλάση Π_1^0 .

Έχουμε ήδη κάνει μία νύξη για τη φύση των Π_1^0 προτάσεων στην ιστορική αναδρομή του 1^{ου} μέρους (βλ. σελ. 9) και εύκολα διαπιστώνουμε από τη φύση της Gödel-πρότασης (βλ. Μέρος II, σελ. 50) ότι αυτή αποτελεί όντως μία τέτοια πρόταση. Ο Penrose δηλαδή στην ουσία χρησιμοποιεί για την εξαγωγή των συμπερασμάτων του ένα, ισοδύναμο με το θεώρημα του Gödel, αποτέλεσμα, σύμφωνα με το οποίο: *Αν ένα σύστημα είναι ορθό για τις Π_1^0 προτάσεις, τότε είναι μη-πλήρες για αυτές.*

(1): Με βάση αυτό το λήμμα καταλαβαίνουμε επίσης καλύτερα γιατί, όπως είχε αναφερθεί στην ιστορική αναδρομή χωρίς άλλες λεπτομέρειες, η απόδειξη της συνέπειας των πραγματικών μαθηματικών ισοδυναμεί με την κεντρικότερη απαίτηση που περιγράφεται από το σχήμα της σελίδας 10. Διότι, είναι άμεσα φανερό ότι αν αποδεικνυόταν η συνέπεια των πραγματικών μαθηματικών, τότε οι 'πραγματικές' προτάσεις που αποδεικνυόταν με οποιαδήποτε (ιδεατή) μέθοδο θα ήταν αληθείς και άρα θα αποδεικνυόταν επίσης στα πραγματικά μαθηματικά.

Τα πράγματα όμως, δυστυχώς, δεν είναι τόσο ξεκάθαρα ως προς τη χρήση της έννοιας της ορθότητας στην επιχειρηματολογία του Penrose, κάτι που διαπιστώνεται από πολλούς επικριτές του, και γίνεται αποδεκτό σαν κριτική και από τον ίδιο τον Penrose. Θα δούμε μάλιστα παρακάτω (βλ. σελ. 131-132) ότι ο Penrose φαίνεται γενικά να μην έχει ξεκάθαρη άποψη για την ακριβή μαθηματική φύση των αποτελεσμάτων του Gödel. Αυτό το γεγονός εκ πρώτης όψεως φαίνεται ιδιαίτερα παράδοξο με βάση όσο έχουμε δει μέχρι τώρα, διότι έχουμε διαπιστώσει ότι πολλές αντιρρήσεις υπήρξαν για την μαθηματική μεταχείριση των αποτελεσμάτων αυτών από τον Lucas, άρα θα περίμενε κανείς να είναι ακόμη πιο προσεκτικός ο Penrose στη διατύπωση του δικού του επιχειρήματος.

Όπως και να έχει πάντως το πράγμα, παραμένει αυτός ο κυριότερος υποστηρικτής του ‘Gödel-ικού επιχειρήματος’ όπως συνηθίζεται να λέγεται το αντι-μηχανιστικό επιχείρημα που βασίζεται στο θεώρημα του Gödel (ή τουλάχιστον στην ισοδύναμη διατύπωσή του που χρησιμοποιεί ο Penrose), οπότε αξίζει να δούμε ποιες ακριβώς είναι οι απόψεις του, αλλά και να διαπιστώσουμε την κριτική που έχει γίνει σε αυτές. Όπως έχουμε ήδη αναφέρει, ο Roger Penrose έγραψε δύο βιβλία τα οποία περιέχουν το καθένα μία μορφή επιχειρήματος που αντίκειται στις θεωρίες που θέλουν τον ανθρώπινο νου ισοδύναμο με μία μηχανή. Τα βιβλία αυτά βέβαια περιέχουν πολλές ακόμα ιδέες οι οποίες σχετίζονται έμμεσα με το θέμα μας, αλλά δεν θα παρουσιαστούν εκτενώς στα πλαίσια αυτής της εργασίας, καθώς πρόκειται κυρίως για επιχειρήματα που βασίζονται στη σύγχρονη κβαντική θεωρία της φυσικής οπότε δεν εντάσσονται στο πλαίσιο που μας ενδιαφέρει. Πρέπει να επισημανθεί όμως ότι αυτές οι ιδέες, αν και έχουν κάποια αυτόνομη αξία, στηρίζονται πάνω στην αρχική σκέψη ότι δεν μπορούμε να παράγουμε ένα μηχανοποιημένο μοντέλο για το νου και αυτή η αρχική σκέψη πηγάζει ακριβώς από το θεώρημα μη-πληρότητας του Gödel. Για κάθε βιβλίο του Penrose θα αφιερώσουμε ένα ξεχωριστό κεφάλαιο όπου θα παρουσιάζονται ταυτόχρονα οι απόψεις του και οι κυριότερες αντιρρήσεις που έχουν εκφραστεί. Ο λόγος που δεν θα επεκταθούμε περαιτέρω ίσως ήδη να έχει γίνει φανερός και θα γίνει εμφανέστερος σε όσα ακολουθούν, αρκεί για την ώρα να επισημάνουμε ότι ένα από τα συμπεράσματα που θα προκύψουν παρακάτω θα ασχολείται ιδιαίτερα με το ιστορικό ‘παράδοξο’ που απαντήθηκε κατά την ανάλυση της δεύτερης περιόδου αντιπαράθεσης, πάντα σε σχέση με την πρώτη.

6.1 Το πρώτο επιχείρημα του R. Penrose και οι αδυναμίες του

Στο πρώτο του βιβλίο [The Emperor's New Mind, ENM από εδώ και στο εξής], ο Penrose παραθέτει το 'γυμνό Gödel-ικό επιχείρημα' όπως αργότερα ο ίδιος θα το ονομάσει. Ο όρος 'γυμνό' έχει να κάνει με την πεποίθηση του Penrose ότι *από μόνο του* το θεώρημα του Gödel δείχνει το συμπέρασμα που θέλει τον άνθρωπο να μην συμπεριφέρεται αλγοριθμικά στην αναζήτηση της μαθηματικής αλήθειας, οπότε δεν χρειάζεται κάποια παραπάνω ανάλυση, ακόμα και για την εξαγωγή πιο ισχυρών συμπερασμάτων που έχουν να κάνουν με τη γενικότερη (μη) ισοδυναμία του με μία μηχανή. Το συμπέρασμα αυτό εδραιώνεται με απώτερο στόχο την κατάδειξη της ανικανότητας περιγραφής της μαθηματικής σκέψης μέσω τυπικών συστημάτων και άρα ενός αυστηρού μαθηματικού φορμαλισμού. Γίνεται αμέσως κατανοητό ότι αυτός ο στόχος διαφέρει κατά πολύ από τον αντίστοιχο του Lucas, παρ' όλα αυτά όμως, θα προχωρήσουμε στην ανάλυσή του, διότι ο ίδιος ο Penrose βασίζεται αργότερα σε αυτή για να διατυπώσει ένα πιο συναφές με αυτό του Lucas επιχείρημα. Ο τρόπος με τον οποίο θα γίνει αυτή η μετάβαση μπορεί να προβλεφθεί λαμβάνοντας υπ' όψιν το λήμμα του Craig (βλ. παραπάνω, σελ. 124) και άλλα σχετικά αποτελέσματα.

Το πρώτο επιχείρημα που παρουσιάζεται στο ENM και σχετίζεται με το θεώρημα του Gödel ξεκινά από την αναζήτηση απάντησης στο ερώτημα «*αν οι διαδικασίες της ανθρώπινης σκέψης μπορούν να είναι εντελώς αλγοριθμικές από τη φύση τους*». Ο Penrose, ακολουθώντας τους συλλογισμούς πίσω από την απόδειξη του 1^{ου} θεωρήματος μη-πληρότητας, καταλήγει να διαπιστώσει (όπως αντίστοιχα είχε κάνει και ο Lucas) την αλήθεια της μη-αποκρίσιμης Gödel-πρότασης *μέσω της διαίσθησής του*. Αυτή η διαίσθηση που βρίσκεται *έξω* από το τυπικό σύστημα που μελετάται είναι μέρος μίας διαδικασίας που ονομάζεται γενικότερα 'διαδικασία αναλογισμού' [reflection principle] και η οποία, κατά τον Penrose, έρχεται σε ευθεία αντίφαση με την ιδέα της τυποποιημένης συλλογιστικής. Στη συνέχεια του συγκεκριμένου κεφαλαίου του ENM ο Penrose χρησιμοποιεί το αποτέλεσμά του για να υποστηρίξει την Πλατωνική μαθηματική θεώρηση έναντι της αυστηρά τυποποιημένης (ή φορμαλιστικής) θεώρησης, για να επιστρέψει αργότερα στο θεώρημα του Gödel και να δώσει ένα πιο ολοκληρωμένο επιχείρημα για τις συνέπειές του στην αξιοπιστία της τελευταίας.

Το επιχείρημα αυτό είναι στην ουσία μία εις άτοπο απαγωγή που έχει σαν υπόθεση ότι οι τρόποι με τους οποίους οι άνθρωποι-μαθηματικοί καταλήγουν στα συμπεράσματά τους είναι όντως αλγοριθμικοί. Με αυτή και την επιπλέον υπόθεση ότι η μαθηματική αλήθεια είναι κάτι που μπορεί να μεταφερθεί από τον ένα μαθηματικό στον άλλο χωρίς ουσιαστικό πρόβλημα, μπορούμε να θεωρήσουμε ότι έχουμε ένα τυπικό σύστημα που είναι ισοδύναμο με όλη τη δυνατή μαθηματική (αλγοριθμική) σκέψη. Αυτό όμως το σύστημα *δε μπορεί να αναγνωριστεί ως τέτοιο* διότι αν υπήρχε αυτή η δυνατότητα, η Gödel-πρότασή του θα μας παρείχε μία αλήθεια εκτός συστήματος οπότε θα καταλήγαμε σε άτοπο. Άρα οδηγούμαστε στο συμπέρασμα ότι *ο αλγόριθμος που χρησιμοποιείται για να περιγράψει τη μαθηματική σκέψη είναι τόσο πολύπλοκος που δεν μπορούμε να πιστοποιήσουμε την αξιοπιστία του*. Αυτό όμως το συμπέρασμα έρχεται, κατά τον Penrose, σε ευθεία αντίφαση με τη φύση των ιδίων των μαθηματικών, η οποία πρέπει να είναι σε κάθε έκφασή της απόλυτα ξεκάθαρη και κατανοητή. Διότι, πως θα μπορούσαμε να φανταστούμε έναν μαθηματικό να υποκύπτει σε κανόνες και συμπεράσματα τα οποία δεν μπορεί να επικυρώσει; Έτσι, καταλήγουμε στο επιθυμητό άτοπο και στο επαγόμενο συμπέρασμα ότι «*η μαθηματική αλήθεια δεν είναι κάτι που μπορούμε να εξακριβώσουμε αλγοριθμικά*».

Παρατηρούμε στην πορεία της παραπάνω συλλογιστικής ότι είναι ξεκάθαρη η στροφή της επιχειρηματολογίας προς μία *αποκλειστικά μαθηματική βάση*, όπου σημασία έχουν μόνο οι *μαθηματικές* αλήθειες και το συμπέρασμα είναι η μη-αλγοριθμική φύση αυτών και μόνο των αληθειών. Βέβαια ο στόχος εδώ είναι η κατάρριψη μιας μαθηματικής θεωρίας, οπότε κάτι τέτοιο θα έπρεπε να το περιμέναμε, όμως θα πρέπει να επισημάνουμε ότι αυτή η διάσταση ξεκαθαρίζει ελαφρώς το πεδίο της αντιπαράθεσης. Διότι, μέχρι τώρα, αν και στην ουσία οι μόνες αλήθειες που αναμετριόνταν ήταν μαθηματικές, κανείς δεν είχε επισημάνει το γεγονός, όπως κανείς επίσης δεν είχε αναρωτηθεί αν τα συμπεράσματα τα οποία μπορεί να προκύψουν από την ανάλυση αυτών των αληθειών μπορούν να επεκταθούν και σε όλες τις υπόλοιπες. Αυτό είναι ένα σημαντικό θέμα κατά τη γνώμη μου και θα μας απασχολήσει περαιτέρω στον τομέα των τελικών συμπερασμάτων. Για την ώρα πάντως μία παραπάνω σαφήνεια μας προσφέρει αυτός ο εντοπισμός του πεδίου αντιπαράθεσης.

Το γενικότερο όμως επιχείρημα φαίνεται να επιτελεί την αντίθετη ακριβώς λειτουργία, μπερδεύει δηλαδή ακόμη περισσότερο τα πράγματα. Αυτό φαίνεται κυρίως από τη διατύπωσή του, η οποία στηρίζεται σε μία εις άτοπο απαγωγή που

βασίζεται περισσότερο στις φιλοσοφικές πεποιθήσεις του Penrose παρά σε αντικειμενικά δεδομένα. Αν όμως οι φιλοσοφικές του πεποιθήσεις είναι αντίθετες του φορμαλισμού, πως αποφεύγεται η κυκλικότητα του συλλογισμού, ήτοι το να ξεκινάς από την άρνηση μιας θέσης για να καταλήξεις στην άρνηση της ίδιας θέσης; Χαρακτηριστικό επίσης είναι το γεγονός ότι αμέσως μετά μπερδεύεται μέσα σε όλα και η *συνείδηση*, η οποία είναι «αυτή που επιτρέπει την διαπίστωση της αλήθειας της Gödel-πρότασης», άρα «αυτή είναι το κύριο συστατικό για την κατανόηση της μαθηματικής αλήθειας», οπότε «...με το να μπορούμε να πείσουμε τον εαυτό μας για την εγκυρότητα του θεωρήματος του Gödel, όχι μόνο μπορούμε να το ‘δούμε’, αλλά και...να αποκαλύψουμε την μη-αλγοριθμική φύση της διαδικασίας μέσα από την οποία το ‘βλέπουμε’».

Η αδυναμία του παραπάνω επιχειρήματος, η οποία μέχρι τώρα υπονοείται μόνο από τα συμφραζόμενα, επιβεβαιώνεται από την κριτική του M. Davis, και μάλιστα σε μία καθαρά μαθηματική (άρα συμβατή με τα παραπάνω) βάση. Στο πρώτο κριτικό άρθρο του {#28}, ο Davis επικεντρώνεται σε σύγχρονα αποτελέσματα της μαθηματικής λογικής και πιο συγκεκριμένα σε μία ισοδύναμη διατύπωση του 1^{ου} θεωρήματος μη-πληρότητας σύμφωνα με την οποία: *«Υπάρχει ένας αλγόριθμος, ο οποίος, δεδομένου ενός συνεπούς αξιωματικού συστήματος, έχει σαν εξαγόμενο μία πολυωνυμική (διοφαντική) εξίσωση $P = 0$ ⁽²⁾ η οποία δεν έχει φυσικές λύσεις, αλλά το γεγονός αυτό δεν μπορεί να αναπαραχθεί από τα δεδομένα αξιώματα»*. Με αυτό το αποτέλεσμα δείχνεται ότι στην πραγματικότητα η διαίσθηση του Penrose ανάγεται όντως σε έναν αλγόριθμο, άρα, κατά τον Davis, η διαίσθηση δεν παίζει κάποιο ιδιαίτερο ρόλο στην απόφαση που αφορά την αλήθεια της Gödel-πρότασης. Το πολύ πολύ, όπως παρατηρεί παρακάτω, να παίζει κάποιο ρόλο στην απόφαση που θα κρίνει αν το σύστημα των αξιωμάτων που έχουν επιλεχθεί είναι συνεπές ή όχι.

Ας δούμε όμως λίγο πιο προσεκτικά τι ακριβώς θέλει να πει ο Davis, μέσω και του επόμενου, κριτικού προς τον Penrose, άρθρου του {#29}. Ο λόγος που επανέρχεται στη συζήτηση είναι για να αποσαφηνίσει ότι πίσω από την διαπίστωση:

$$F \text{ ορθο} \Rightarrow G(F) \text{ αληθής},$$

(2): Η αναλογία του αποτελέσματος του Gödel με τις διοφαντικές εξισώσεις έχει ήδη συζητηθεί στο προηγούμενο κεφάλαιο.

όπου F ένα τυπικό σύστημα και $G(F)$ η Gödel- πρότασή του, δεν κρύβεται καμία διαίσθηση, παρά μόνο σύγχρονα αποτελέσματα της μαθηματικής λογικής. Συγκεκριμένα, η κριτική του βασίζεται στο παρακάτω αποτέλεσμα:

Έστω T το σύνολο των αληθών προτάσεων που μπορούν να παραχθούν από το τυπικό σύστημα F . Το (ορθό) F , μέσω αλγορίθμου, μπορεί να παράγει ένα αναδρομικά αριθμήσιμο σύνολο αληθών προτάσεων $T_0 \subseteq T$. Καθότι όμως το T δεν είναι γενικά αναδρομικά αριθμήσιμο, θα υπάρχει ένας αριθμός $q_0 \in T - T_0$ ο οποίος θα κωδικοποιεί ακριβώς μία πρόταση που είναι αληθής αλλά όχι αποδείξιμη (παραγόμενη αλγοριθμικά από το F). Ισχύει όμως ότι το σύνολο T είναι ‘παράξιμο’ [productive], ήτοι υπάρχει αλγόριθμος ο οποίος, δεδομένης μίας μηχανής Turing που μπορεί να ορίσει το σύστημα F σαν εισαγόμενο, να μπορεί να εξάγει τον αριθμό q_0 .

Το αποτέλεσμα αυτό φαίνεται ότι καταρρίπτει άμεσα την προσπάθεια του Penrose εφόσον αυτή στηρίζεται στην κατάδειξη ακριβώς αυτής της διαδικασίας μέσω της οποίας διαπιστώνεται η αλήθεια της Gödel-πρότασης, ως μη-αλγοριθμική. Το ίδιο επιχείρημα έχει χρησιμοποιηθεί και για την κατάρριψη του επιχειρήματος του Lucas {βλ. Bowie - #23} αλλά όχι με την ίδια επιτυχία, διότι η διαλεκτική φύση του επιχειρήματος του τελευταίου μπορεί να απορρίψει την κριτική στη βάση του ότι ο αλγόριθμος αυτός που παράγεται δεν αποτελεί μέρος του αρχικού συστήματος F , και αν ενσωματωθεί σε ένα νέο σύστημα, τότε αυτό το νέο σύστημα θα υπόκειται επίσης στο αντι-μηχανιστικό επιχείρημα κ.ο.κ.

Κατά τα άλλα, στο ίδιο άρθρο, ο Davis επίσης κριτικάρει την προσπάθεια παρουσίασης των απόψεων του ίδιου του Gödel ως τέτοιες οι οποίες συνηγορούν υπέρ του επιχειρήματος του Penrose, κριτική που είναι εύλογη υπό το πρίσμα των όσων έχουμε διαπιστώσει στο προηγούμενο κεφάλαιο όπου ασχοληθήκαμε επιστάμενα με τις απόψεις του Gödel. Τελικά ο Davis καταλήγει, με τη βοήθεια όλων των παραπάνω ότι το ερώτημα της φύσης της μαθηματικής συλλογιστικής μπορεί να είναι ένα κρίσιμο θέμα, αλλά ένα θέμα για το οποίο το θεώρημα του Gödel δεν μπορεί να μας δια φωτίσει ιδιαίτερα.

Αυτή ήταν η πρώτη προσπάθεια του R. Penrose να βγάλει ένα συμπέρασμα, μέσω της διαδικασίας που ακολουθείται στην απόδειξη του θεωρήματος του Gödel, για τη μη-αλγοριθμική φύση της μαθηματικής σκέψης. Η προσπάθειά του στηρίχθηκε αποκλειστικά και μόνο στο θεώρημα του Gödel, χωρίς τίποτα παραπάνω, κάτι στο

οποίο οφείλεται μάλλον και η διαφαινόμενη αδυναμία του. Άλλωστε, εκτός του ότι ακόμα και υποστηρικτές των αντι-μηχανιστικών συνεπειών του θεωρήματος {βλ. S. McCall – #37}, θεωρούν ότι χρειάζεται κάτι παραπάνω από μία απλή αναδιατύπωσή του για να βγει κάποιο εύστοχο συμπέρασμα, και ο ίδιος ο Penrose, στο επόμενο βιβλίο του ανασκευάζει με πιο προσεκτικό (μαθηματικό) τρόπο το επιχείρημά του.

6.2 Το δεύτερο επιχείρημα του R. Penrose και οι σχετικές αντιρρήσεις

Το δεύτερο βιβλίο του R. Penrose στο οποίο ο συγγραφέας καταπιάνεται με τις συνέπειες του θεωρήματος του Gödel [Shadows of the Mind. Oxford University Press. 1994 – SM, από εδώ και στο εξής], αποτελεί στην ουσία μία συνέχεια του ENM και καταπιάνεται με τα ίδια θέματα, αλλά με πιο ολοκληρωμένο και εμπειρισματομένο τρόπο. Έτσι, οι συνέπειες του θεωρήματος του Gödel δεν πηγάζουν πλέον αυτόματα από το θεώρημα αλλά υποστηρίζονται και από μία επιχειρηματολογία που μοιάζει περισσότερο με αυτή του Lucas από αυτή που είδαμε στην περίπτωση του ENM, ενώ και τα υπόλοιπα (κυρίως φυσικά) θέματα αναλύονται με περισσότερη σαφήνεια ως προς τους στόχους τους. Αξίζει εδώ να αναφέρουμε ότι τελικά ο Penrose θέλει να εγκαθιδρύσει ότι η «χαμένη επιστήμη της συνείδησης», (η ανάγκη ύπαρξης της οποίας προκύπτει από τις συνέπειες του θεωρήματος του Gödel) θα αποτελεί μία ριζική μετεξέλιξη της σύγχρονης κβαντικής φυσικής. Σε αυτή του την προσπάθεια εμπλέκονται πολλές έννοιες πέραν των μαθηματικών, όμως εδώ θα ασχοληθούμε *μόνο* με την μεταχείριση των αποτελεσμάτων της μη-πληρότητας, αν και το γενικότερο πλαίσιο παρουσιάζει σίγουρα ένα αυτόνομο ενδιαφέρον.

Πριν όμως δούμε ακριβώς ποιο είναι το νέο επιχείρημα του Penrose πρέπει να διαπιστώσουμε πρώτα μερικά πράγματα σε σχέση με την αρχική διατύπωσή του και τη χρήση των εμπλεκόμενων εννοιών της μαθηματικής λογικής. [Σημ: Για τη διατύπωση αυτή χρησιμοποιούνται σαν ορολογία αποκλειστικά οι νέες έννοιες που περιγράφηκαν στην αρχή του κεφαλαίου]. Η ολοκληρωμένη κριτική του S. Feferman {#32} καταπιάνεται αποκλειστικά με αυτό το θέμα και αποδεικνύεται ιδιαίτερα διαφωτιστική. Συγκεκριμένα ο Feferman παρατηρεί ότι ο Penrose επανειλημμένα συγχέει την έννοια της ορθότητας που αφορά τις Π_1^0 προτάσεις (και είναι ισοδύναμη με τη συνέπεια), με τη γενικότερη έννοια της ορθότητας, και δεν ξεκαθαρίζεται

τελικά σε ποια κατάσταση χρησιμοποιείται η μία και σε ποια η άλλη. Επίσης επισημαίνεται ότι η χρήση του όρου ‘ $\omega(F)$ ’ (που δηλώνει την ω -συνέπεια ενός τυπικού συστήματος F) για να δηλωθεί η Gödel-πρόταση που προκύπτει για το F είναι (προφανώς!) τελείως ατυχής. Ο Feferman επισημαίνει και μία σωρεία μικρότερων λαθών που περιέχονται στην αρχική διατύπωση του επιχειρήματος του Penrose και καταλήγει να σχολιάζει καυστικά την αρτιότητα των γραφόμενων του, χωρίς να επεκταθεί άλλο στο θέμα. Παραδέχεται πάντως ότι και ο ίδιος είναι υπέρ της άποψης ότι η μαθηματική διαίσθηση δεν μπορεί να είναι μία μηχανοποιημένη διαδικασία, απλά πιστεύει (ίσως και δικαιολογημένα, λόγω των παραπάνω) ότι αναλύσεις σαν αυτή του Penrose τελικά περισσότερο κακό παρά καλό κάνουν στην προσπάθειά τους να την υπερασπιστούν.

Δεδομένων αυτών των αρχικών διαπιστώσεων, και για να αποφύγουμε περαιτέρω συγχύσεις, παραθέτουμε το ‘νέο’ επιχειρήμα του Penrose, όχι στην αρχική του μορφή αλλά σε μία ισοδύναμη, η οποία υπάρχει στο άρθρο του D. Chalmers ‘Minds, Machines and Mathematics’ {#33}, διαδικασία που ο Penrose δέχεται ότι όντως αντικατοπτρίζει την ουσία του επιχειρήματός του.

Το Gödel-ικό επιχειρήμα του R. Penrose

1. Έστω ότι ένα τυπικό σύστημα F συμπυκνώνει όλες τις ανθρώπινα προσβάσιμες μεθόδους μαθηματικής απόδειξης. (Για συντομία γράφουμε: ‘Εγώ είμαι F ’).
2. Δεδομένου ότι το F είναι ορθό (εφόσον προϋποτίθεται ότι ‘εγώ’ είμαι ορθός), το σύστημα $F' = (F + \text{‘Εγώ είμαι } F\text{’})$ θα πρέπει να είναι επίσης ορθό, εφόσον έχουμε δεχθεί ότι ισχύει η υπόθεση ‘Εγώ είμαι F ’.
3. Από το θεώρημα του Gödel, ‘εγώ’ ξέρω ότι η Gödel-πρόταση του συστήματος F' , ήτοι η $G(F')$, είναι αληθής. Το F' όμως δεν μπορεί να δει σαν αληθή την $G(F')$.
4. Άρα, από τον ορισμό της ισοδυναμίας εμού και ενός τυπικού συστήματος, μπορώ να συμπεράνω τελικά ότι ‘Εγώ είμαι F' ’. Άτοπο, εφόσον έχει υποτεθεί ότι ‘Εγώ είμαι F ’ και $F \neq F'$.
5. Το συμπέρασμα μπορεί να γενικευτεί για κάθε τυπικό σύστημα που έχει παρόμοιες δυνατότητες με αυτές του F του 1^{ου} βήματος.

Το συμπέρασμα που βγάξει ο Penrose από την παραπάνω συλλογιστική αποτελεί τη μετεξέλιξη αυτού του ENM, και είναι προφανώς πιο κοντά σε όσα έχουμε δει μέχρι τώρα. Η ισοδυναμία του με το αντίστοιχο συμπέρασμα του Lucas είναι επίσης φανερή, δεδομένου και του αποτελέσματος του λήμματος του Craig, το οποίο εγκαθιδρύει τη σχέση ανάμεσα στις μηχανές Turing και τα τυπικά συστήματα. Για να σαφέστερη και η αναλογία της συλλογιστικής των δύο, αρκεί να αναφερθεί ότι ο Penrose σε κάποιο άλλο σημείο του βιβλίου του φέρνει αντιμέτωπους στην ‘αρένα της μαθηματικής αποδεικτικής δύναμης’ έναν άνθρωπο-μαθηματικό και ένα ρομπότ-μαθηματικό και τελικά το ρομπότ, ερχόμενο αντιμέτωπο με την Gödel-πρότασή του που του παρουσιάζεται από τον άνθρωπο-μαθηματικό, τρελαίνεται. [Σημ: Ο Penrose δεν ισχυρίζεται ακριβώς ότι ο άνθρωπος μπορεί να *αποδείξει* την αλήθεια της Gödel-πρότασης, αλλά ότι μπορεί να την ‘πιστέψει αδήριτα’ [unassailably] και με αυτή την έννοια μπορεί να ‘δει’ την Gödel-πρόταση του ρομπότ σαν αληθή ενώ το ίδιο αδυνατεί διότι αντιλαμβάνεται σαν αληθές μόνο κάτι που μπορεί να το αποδείξει]. Έχουμε δηλαδή στην ουσία το ίδιο επιχείρημα με αυτό του Lucas, λίγο πιο γλαφυρά δοσμένο και με μία καινούρια έκφραση που περιγράφει την ικανότητα του ανθρώπου να αντιλαμβάνεται την αλήθεια της μη-αποκρίσιμης Gödel-πρότασης.

Στο ίδιο πνεύμα της αναλογίας των επιχειρημάτων των Penrose και Lucas, παρατηρούμε επίσης ότι η ορθότητα του συστήματος F του Penrose βασίζεται στην ορθότητα του ανθρώπου, όπως αντίστοιχα η συνεπής μηχανή που θα πάρει μέρος στο επιχείρημα του Lucas απαιτείται να είναι συνεπής λόγω της ανθρώπινης συνέπειας. Ίδιος είναι επίσης και ο τρόπος με τον οποίο υπερασπίζονται και οι δύο την ορθότητα (αντ. συνέπεια) του ανθρώπου βασιζόμενοι στο σκεπτικό ότι ο άνθρωπος μπορεί να κάνει λάθη (αντ. να πέφτει σε αντιφάσεις), αλλά αυτά δεν είναι εσκεμμένα και γίνεται πάντα μία προσπάθεια αποβολής τους όπου αυτά εμφανιστούν. Ο λόγος για τον οποίο το ρομπότ του Penrose, όπως αντίστοιχα η μηχανή του Lucas, δεν μπορεί να δει την αλήθεια της Gödel-πρότασης είναι επίσης ο ίδιος ουσιαστικά, εφόσον το ρομπότ του Penrose, ισοδυναμεί με ένα σύστημα το οποίο δεν μπορεί παρά να είναι υπολογιστό-καθορισμένο απόλυτα μέσω αλγορίθμου, σε αντίθεση με διαδικασία αντίληψη της αλήθειας της Gödel-πρότασης, η οποία παραμένει (βλ. προηγούμενη παράγραφο, 6.1) μη-αλγοριθμική. Τελείως ανάλογη επιχειρηματολογία συναντήσαμε και στην ανάλυση του επιχειρήματος του Lucas στο κεφάλαιο 3 του παρόντος μέρους.

Έχοντας διαπιστώσει την ισοδυναμία των δύο επιχειρημάτων, δεν θα πρέπει να αποτελεί έκπληξη το γεγονός ότι οι αντιρρήσεις στα συμπεράσματα του Penrose είναι όμοιες με όσες έχουν ήδη αναλυτικά παρουσιαστεί στο 4^ο κεφάλαιο. Μερικές μάλιστα από αυτές θα μπορούσε κανείς να πει ότι είναι απόλυτα ίδιες! Για να εντοπίσουμε κάποια ουσιώδη διαφορά στα επιχειρήματα της μίας και της άλλης περιόδου θα πρέπει να ανατρέξουμε στην μία και μοναδική ουσιαστική διαφορά των επιχειρημάτων των Lucas και Penrose, η οποία, όπως είναι φανερό, είναι η μονολογικότητα του επιχειρήματος του Penrose έναντι της διαλεκτικής φύσης αυτού του Lucas. Και ενώ θα περίμενε κανείς ότι η μονολογική εκδοχή θα προσθέτετε μία σαφήνεια στα πράγματα, μιας και συνάδει περισσότερο με τη μαθηματική λογική που χρησιμοποιούμε, η σωρεία λογικών λαθών καθώς και η διττή φύση των εννοιών που χρησιμοποιούνται από τον Penrose καθιστά τελικά το επιχείρημά του πολύ πιο ευάλωτο σε κριτική από αυτό του Lucas.

Για να γίνουμε πιο συγκεκριμένοι, εκτός των λογικών λαθών και της διφορούμενης μεταχείρισης της έννοιας της ορθότητας που έχουμε ήδη επισημάνει, επίσης ασαφής είναι και η χρήση της έννοιας του 'πιστεύω κάτι αδήριτα', όπως εύστοχα επισημαίνει ο D. McDermott στο σχετικό κριτικό του άρθρο {#31}. Η απάντηση του Penrose στην ανησυχία για το αν η έννοια αυτή είναι ισοδύναμη με την έννοια της τυπικής απόδειξης ή κάποια άλλη τυπική μέθοδο συλλογισμού ή αν τελικά είναι θέμα υποκειμενικής οπτικής γωνίας, είναι στην ουσία μία αναδιατύπωση της θέσης του. Διαπιστώνεται δηλαδή ότι ακριβώς *δεν μπορεί αυτή η έννοια να είναι τυποποιήσιμη*, διότι τότε θα ήταν υπολογιστή και άρα δεν θα ήταν αυτή που θα επέτρεπε να δούμε σαν αληθή την Gödel-πρόταση. Αυτή η κυκλική αντιμετώπιση όμως, αφήνει ανοιχτό το ερώτημα για το ποιες άλλες αλήθειες είναι τέτοιες που να μπορούμε να τις 'πιστέψουμε αδήριτα' χωρίς να τις αποδείξουμε τυπικά, και ποια κριτήρια είναι τελικά που χρησιμοποιούνται για να γίνει η συγκεκριμένη επιλογή. Συμπεραίνουμε λοιπόν ότι ο τρόπος με τον οποίο ο Penrose θέλει να αποφύγει το πρόβλημα του πως γίνεται να βλέπει κάποιος σαν αληθή την Gödel-πρόταση χωρίς να την αποδεικνύει τυπικά, πρόβλημα που είχε προκύψει και στην πρώτη περίοδο αντιπαράθεσης, μάλλον δεν αποφέρει τα δέοντα.

Όλη αυτή η ασάφεια που διέπει τον τρόπο παρουσίασης του θεωρήματος του Gödel και των πιθανών συνεπειών του από τον Penrose είχε σαν αποτέλεσμα ελάχιστη κριτική να δεχθεί το 'νέο' επιχείρημα του (τουλάχιστον στα πρώτα άρθρα τέτοιου περιεχομένου που αναφέρονταν στην πρώτη έκδοση του βιβλίου και στην

αρχική μορφή του επιχειρήματος που δεν παρουσιάστηκε), απλά και μόνο επειδή σχεδόν κανείς δεν αντιλήφθηκε σε ποιο σημείο του βιβλίου γινόταν η σαφής αναφορά σε αυτό! Έτσι, η κριτική που έχει να κάνει με αυτές καθ' αυτές τις απόψεις του Penrose και όχι με την ασάφεια των γραφόμενών του, επικεντρώνεται κυρίως σε δευτερεύοντα θέματα τα οποία, είτε σχετίζονται έμμεσα με το θεώρημα του Gödel, ή έχουν να κάνουν με τις ιδέες και τα επιχειρήματα που αφορούν τη φυσική και τη βιολογία. Θα δούμε πάντως αμέσως παρακάτω κάποια χαρακτηριστικά κομμάτια κριτικής της ιδέας που μας ενδιαφέρει, αρχίζοντας από αυτά που έχουν να κάνουν αμιγώς με το Gödel-ικό επιχείρημα και συνεχίζοντας με τα υπόλοιπα, από τα οποία τα περισσότερα θα συζητηθούν σύντομα, ένεκα της συνάφειάς τους με αυτά της πρώτης περιόδου αντιπαράθεσης και της εκτενούς και αναλυτικής παρουσίασης της τελευταίας.

Ξεκινάμε από την κριτική του Chalmers {#33}, η οποία είναι μία από τις ελάχιστες που αναφέρεται σε αυτό καθ' εαυτό το νέο επιχείρημα του Penrose. Ο Chalmers κινείται στη γραμμή κριτικής που χάραξε ο Benacerraf (βλ. § 4.2.1), δέχεται δηλαδή την πορεία του επιχειρήματος της σελίδας 132 αλλά εισάγει την επιπλέον υπόθεση στο βήμα (2) ότι 'ξέρω ότι 'Εγώ είμαι F '. Με αυτή την επιπλέον υπόθεση, καταλήγει, αρχικά χωρίς ιδιαίτερη ανάλυση, στο ασθενέστερο συμπέρασμα ότι «δεν μπορώ να ξέρω αν είμαι όμοιος με ένα τυπικό σύστημα F ». Στη συνέχεια του άρθρου και στην προσπάθειά του για την τελική κατάρριψη του επιχειρήματος του Penrose, ο Chalmers παραθέτει μία διαγώνια διαδικασία η οποία στόχο έχει να δείξει ότι η γνώση της συνέπειάς του ανθρώπου *από μόνη της* οδηγεί σε αντίφαση, χωρίς την υπόθεση ότι 'ξέρω ότι είμαι F '. Αυτή η εκτενής ανάλυση είναι που δείχνει με τον πιο δραστικό τρόπο την αναλογία με τη συλλογιστική του Benacerraf. Αρκεί να πούμε ότι οι υποθέσεις των δύο είναι ουσιαστικά οι ίδιες, δηλαδή οι υποθέσεις των 'αντιπάλων' τους και ότι και πάλι το τελικό συμπέρασμα στρέφεται προς την κατάρριψη της υπόθεσης η οποία βγάζει ακριβώς το επιθυμητό αποτέλεσμα. Έτσι, η ανακατασκευή του επιχειρήματος από τον Chalmers δεν προσθέτει ουσιαστικά τίποτα καινούριο και υπόκειται στην ίδια κριτική που έγινε και προς αυτήν του Benacerraf (βλ. σελ. 107).

Η ομοιότητα της ανάλυσης επεκτείνεται φυσιολογικά και στον τομέα των συμπερασμάτων, όπως είναι φανερό από τη γενικότερη θέση του Chalmers: «Αν είμαστε ανάλογοι με τυπικά συστήματα, τότε σίγουρα δεν μπορούμε να ξέρουμε ποιο τυπικό σύστημα είμαστε με βάση μόνο την ενδοσκόπηση». Το σκεπτικό το οποίο

οδηγεί τον Chalmers σε αυτό το συμπέρασμα, πέρα των παραπάνω, είναι ότι για να μπορεί να βγει το συμπέρασμα του Penrose θα πρέπει να υπάρχει ένας τρόπος να αποφασίσουμε ότι το σύστημα F είναι ορθό *ανεξάρτητα* από τη γνώση ότι 'Εγώ είμαι F ', ήτοι να υπάρχει ένας τρόπος, δοθέντος μόνο του F , να μπορούμε να αποφασίσουμε αν αυτό είναι ορθό ή όχι. Είναι σαφές πιστεύω η αναλογία με την απαίτηση του 'μαύρου κουτιού' του Benacerraf (βλ. σελ. 105), οπότε, και με αυτή τη διαπίστωση, γίνεται κατανοητό ότι δεν χρειάζεται να επεκταθούμε σε περαιτέρω λεπτομέρειες των θέσεων του Chalmers, παραπέμπουμε απλά στην παράγραφο 4.2.1, όπου αναλύεται η θέση του Benacerraf και γίνεται η ανάλογη κριτική.

Μία επιπλέον κριτική του επιχειρήματος του Penrose, έρχεται να επιβεβαιώσει την ιστορική επανάληψη της συζήτησης που σκιαγραφήθηκε παραπάνω. Συγκεκριμένα, ο P. Lindström {#39} παίρνει τη θέση του C. Chihara και κριτικάρει τόσο το επιχείρημα του Penrose, όσο και αυτό του Chalmers, όπως ακριβώς ο Chihara (βλ. σελ. 110) κριτικάρει Lucas και Benacerraf αντίστοιχα! Ο Lindström κυρίως καταλήγει να απορρίψει την κατηγορηματικότητα με την οποία υπερασπίζεται το επιχείρημά του ο Penrose, δίνοντας αντιπαραδείγματα συμπερασμάτων που θα μπορούσαν να προκύψουν από την ανάλυση του Gödel-ικού επιχειρήματος, ανεξαρτήτως της ασάφειας που διέπει τη χρήση της ορθότητας στην ανάλυση του Penrose. Το κομμάτι αυτό του άρθρου είναι ιδιαίτερα τεχνικό⁽³⁾ και βασίζεται στην υπόθεση της *τυποποίησης* της ορθότητας σε ένα σύστημα το οποίο περιλαμβάνει τις 'αδήριτα διαπιστωμένες αλήθειες'. Αυτή η ιδέα είναι στην ουσία ίδια με αυτή του Chihara ο οποίος χρησιμοποιεί αντίστοιχα το σύνολο S για να δείξει (μαζί με τον Lindström) ότι προσπαθώντας να 'καλουπώσουμε' έννοιες όπως 'βλέπω σαν αληθές' ή 'διαπιστώνω αδήριτα την αλήθεια', ακόμα και μέσα σε λιγότερο αυστηρά πλαίσια, καταλήγουμε πάντα σε μη-έγκυρα συμπεράσματα, και άρα ότι δεν μπορούν να προκύψουν επιπτώσεις του θεωρήματος του Gödel προς τη συγκεκριμένη κατεύθυνση.

Μέσω της ίδιας συλλογιστικής ο Lindström απορρίπτει και τις επιπτώσεις που προκύπτουν από την επιχειρηματολογία του Chalmers και εγκαθιδρύει τη σχέση του με τις απόψεις του Chihara, άρα δε χρειάζεται να αναλυθεί κάτι παραπάνω από αυτά αναλύθηκαν στην παράγραφο 4.2.3. Αρκεί μόνο να προσθέσουμε τον προβληματισμό του Lindström για το κατά πόσο το γενικότερο ερώτημα αν είμαστε ισοδύναμοι με

(3): Σημειώστε ότι και το αντίστοιχο άρθρο του Chihara είναι ιδιαίτερος τεχνικό (!).

τυπικά συστήματα (ή αντίστοιχα μηχανές) μπορεί τελικά να προκύψει κενό νοήματος ή μάλλον να μην επιδέχεται από τη φύση του ενός ορθού ορισμού. Αυτό είναι ένα ερώτημα που φαίνεται εύλογο και σίγουρα, στα πλαίσια μίας φιλοσοφικής αναζήτησης, οποιοδήποτε ερώτημα μπορεί να θεωρηθεί κάποια στιγμή ότι δεν είναι ‘καλά ορισμένο’. Αυτό όμως δε σημαίνει ότι θα πρέπει να σταματήσει και η προσπάθεια απάντησης ερωτημάτων τα οποία δε είναι τόσο εύληπτα όσο άλλα, διότι αλλιώς ποια θα ήταν η αξία του να φιλοσοφεί κανείς;

Περνώντας τώρα στις κριτικές οι οποίες βασίζονται περισσότερο στα γενικά χαρακτηριστικά της συλλογιστικής του Penrose, παρά στο ακριβές νέο επιχειρήματά του, παρατηρούμε και πάλι ότι κατά κάποιον τρόπο, ‘όλα αλλάζουν κι όλα τα ίδια μένουν’. Συγκεκριμένα, συναντάμε επιχειρήματα που έχουν να κάνουν με τη συνέπεια/ορθότητα και το κατά πόσο αυτή είναι αναγκαία στην περίπτωση των μηχανών {McDermott - #31} ή αν υπάρχει η πιθανότητα να είμαστε τελικά μη-συνεπείς, οπότε η συνέπεια δεν είναι μία απαίτηση που θα πρέπει να έχουμε από τις μηχανές {McCullough - #30}, επιχειρήματα που ήδη έχουν συζητηθεί διεξοδικά. Πέρα όμως από τις γενικές ομοιότητες, υπάρχουν και άλλες, πολύ συγκεκριμένες οι οποίες επιβεβαιώνουν την ιστορική επανάληψη ιδεών. Για παράδειγμα ο McCullough παραθέτει ένα επιχειρήμα με βάση *ακριβώς* την ίδια πρόταση με αυτή που είχε παραθέσει παλιότερα ο Whiteley (βλ. σελ. 96), ενώ ο McDermott αμφισβητεί την ιδέα που έχει ο Penrose για το τι είναι η Τεχνητή Νοημοσύνη και σε ποιες αρχές βασίζεται, όπως αντίστοιχα είχε κάνει ο George (βλ. επίσης σελ. 96).

Γίνεται κατανοητό από όλα τα παραπάνω ότι μία εκτενέστερη παρουσίαση των σύγχρονων απόψεων δεν θα είχε να προσθέσει κάτι το εντυπωσιακά καινούριο σε όσα έχουν ήδη διαπιστωθεί, οπότε η παρουσίαση της δεύτερης περιόδου αντιπαράθεσης σε αυτό το σημείο ουσιαστικά ολοκληρώνεται. Παρατίθενται μόνο παρακάτω κάποια σχόλια σε σχέση με τη σχέση των δύο περιόδων αντιπαράθεσης, όπως αυτή καταγράφηκε σε αυτή την παράγραφο, αμέσως μετά τον πίνακα της επόμενης σελίδας. Σε αυτόν τον πίνακα γίνεται η προσπάθεια καταγραφής όλων των συνδέσεων που παρατηρήθηκαν, ώστε να έχουμε συγκεντρωμένες όλες οι αναλογίες των δύο περιόδων.

	Πρώτη περίοδος αντιπαράθεσης	Δεύτερη περίοδος αντιπαράθεσης
Η θεωρία που υποστηρίζει την ισοδυναμία ανθρώπου-μηχανής	Μηχανισμός	Υπολογιστικισμός / Τεχνητή Νοημοσύνη
Ο βασικός υποστηρικτής των συνεπειών του θεωρήματος του Gödel στις παραπάνω θεωρίες	J.R. Lucas	R. Penrose
Η έννοια από την οποία πηγάζει το αποτέλεσμα του Gödel και το αντίστοιχο επιχείρημα...	Συνέπεια	Ορθότητα
...το οποίο στηρίζεται στο ότι ο άνθρωπος [...] της Gödel-πρότασης	‘βλέπει την αλήθεια’	‘πιστεύει αδήριτα την αλήθεια’
Αντίρρηση στη βάση του ότι ‘πρέπει να ξέρω ότι ξέρω...’. Συμπέρασμα: Δεν μπορώ να ξέρω ποια μηχανή είμαι	P. Benacerraf	D.J. Chalmers
Αντίρρηση στο παραπάνω & συνολική κριτική για το ότι υπάρχουν συνέπειες	C. Chihara	P. Lindström
«Ο Lucas/Penrose δεν μπορεί να δει την αλήθεια αυτής της πρότασης»	C. Whiteley	D. McCullough
Οι παραπάνω πρέπει να παρακολουθήσουν τα σχετικά μαθήματα Τεχνητής Νοημοσύνης	F. George	D. McDermott

Η ‘ιστορική επανάληψη των ιδεών’ η οποία καταγράφεται συνοπτικά σε αυτόν τον πίνακα είναι πράγματι εντυπωσιακή. Αν επιχειρούσαμε να βγάλουμε ένα συμπέρασμα και μόνο από τη διαπίστωση ότι αυτή η επανάληψη υφίσταται, αυτό αναγκαστικά θα ήταν ότι, από ένα σημείο και πέρα, και χωρίς να υπάρξει κάποια θεαματική εξέλιξη στον σχετικό επιστημονικό τομέα, όπως για παράδειγμα η κατάρριψη της θέσης του Church, η συζήτηση του συγκεκριμένου θέματος έχει φτάσει σε ένα κάποιο τέλμα. Οι δύο πλευρές φαίνεται να διατηρούν τις ίδιες θέσεις με επιχειρήματα που μπορεί να αλλάζουν στην διατύπωση αλλά μένουν ίδια στην ουσία. Μία εξήγηση για το γιατί μπορεί να συμβαίνει κάτι τέτοιο επιχειρείται στο τελευταίο κεφάλαιο αυτού του μέρους το οποίο αποτελεί και το κομμάτι των τελικών συμπερασμάτων, για την ώρα μόνο να εκφράσει ο γραφών την (εύλογη) απορία αν τελικά θα έπρεπε να είναι κανείς πιο προσεκτικός όταν διατυπώνει θέσεις οι οποίες έχουν ήδη διατυπωθεί από άλλους. Το σχόλιο βέβαια κυρίως στοχεύει τον R. Penrose, ο οποίος, αν και δημιούργησε αρκετό ‘θόρυβο’ με τα βιβλία του, δεν κατάφερε τελικά να παρουσιάσει ένα καλύτερο επιχείρημα από αυτό του J.R. Lucas. Μάλιστα θα μπορούσε κανείς να πει ότι οι απόψεις του Penrose αποτέλεσαν στην ουσία ένα ‘πισωγύρισμα’, τόσο λόγω της απρόσεκτης (μαθηματικής) διατύπωσής τους η οποία εισήγαγε μία μεγάλη σύγχυση στη συζήτηση, όσο και λόγω της εγκατάλειψης της ιδέας του διαλεκτικού επιχειρήματος του Lucas. Η προσωπική μου πεποίθηση είναι ότι αυτό το επιχείρημα στηρίζει καλύτερα από κάθε άλλο την άποψη ότι το θεώρημα του Gödel έχει συνέπειες στην ευρύτερη μηχανιστική θεώρηση, δεδομένου επίσης και του ότι δεν πρόκειται για ‘στατικό’, αλλά για ένα ‘δυναμικό’ επιχείρημα, και αυτή η δυναμικότητα είναι μία ποιότητα που και ο ίδιος ο Gödel θεωρούσε αναπόσπαστο κομμάτι της φύσης του ανθρώπινου νου.

Τελικά, μπορούμε να πούμε ότι η συζήτηση που έχει να κάνει με τις συνέπειες του θεωρήματος του Gödel στο συγκεκριμένο πεδίο της φιλοσοφίας του νού έχει καλυφθεί με μία πληρότητα η οποία μας επιτρέπει να βγάλουμε τα όποια γενικότερα συμπεράσματα με κάποια σχετική ασφάλεια. Πριν όμως ξεκινήσει αυτή η διαδικασία, αξίζει να αναλύσουμε σύντομα δύο ακόμη άρθρα τα οποία δείχνουν με ακόμη πιο δραστικό τρόπο μέχρι που μπορεί να επεκταθεί η φιλοσοφική σκέψη που ξεκινά από το θεώρημα του Gödel, και τα οποία περιέχουν δύο ιδέες που η κάθε μία από αυτές, για διαφορετικούς λόγους, θα μπορούσε να αποτελεί τη φυσική εξέλιξη της συζήτησης που παρακολουθήσαμε.

6.3 Πιο σύγχρονες και πιο απόμακρες προεκτάσεις του θεωρήματος του Gödel

Σε αυτή την παράγραφο θα σκιαγραφηθούν κάποιες προεκτάσεις του θεωρήματος του Gödel οι οποίες κινούνται στις παρυφές των όσων μέχρι τώρα έχουμε μελετήσει. Πρόκειται στην ουσία για δύο άρθρα των H. Putnam {#25} και J.P. Sullins {#37}, τα οποία αντιμετωπίζουν το όλο θέμα της δυνατότητας ύπαρξης μηχανιστικής αναπαράστασης του νου από μία σκοπιά που ξεφεύγει από τα στενά όρια της αντιπαράθεσης που μέχρι τώρα αναλύσαμε. Δεν συναντάμε δηλαδή σε αυτά θέσεις οι οποίες στοχεύουν στην κατάρριψη ή αποδοχή των όσων έχουν δει, αλλά απόψεις που κινούνται σε διαφορετικά, ξεχωριστά πλαίσια η κάθε μία. Αυτός είναι άλλωστε και ο λόγος που δεν έχει γίνει μέχρι τώρα καμία αναφορά σε αυτά, και αναφερόμαστε κυρίως στο άρθρο του Putnam το οποίο ανήκει μεν *χρονικά* στην πρώτη περίοδο αντιπαράθεσης (καθότι γραμμένο το 1985) αλλά *ουσιαστικά* δεν λαμβάνει μέρος σε αυτή, οπότε δεν είχε νόημα να παρουσιαστεί νωρίτερα στην παρούσα εργασία.

Περνώντας στην ανάλυση των άρθρων ξεκινάμε από την ‘απόμακρη’ προέκταση του τίτλου, όπως αυτή εκφράζεται στο άρθρο του Putnam ‘Reflexive Reflections’. Ο Putnam επιχειρηματολογεί με βάση το θεώρημα του Gödel για να δείξει τα όρια της ενδοσκόπησης του ανθρώπου και αντίστοιχα τα όρια της γνωστικής επιστήμης γενικότερα, όπου σαν γνωστική επιστήμη αντιλαμβανόμαστε αυτή που ασχολείται με τη φύση και τις διαδικασίες επεξεργασίας δεδομένων από τον ανθρώπινο νου. Για να ξεκινήσει την επιχειρηματολογία του, ο Putnam δέχεται την ύπαρξη ενός είδους υπολογιστικού μοντέλου για το νου και εφαρμόζει σε αυτό τη διαδικασία απόδειξης του 1^{ου} θεωρήματος μη-πληρότητας. Η διαφορά με τα όσο έχουμε συναντήσει μέχρι τώρα έγκειται στη φύση του υπολογιστικού μοντέλου. Αυτό που δέχεται ο Putnam δεν προσομοιώνει μόνο την αφαιρετική συλλογιστική όπως τα μέχρι τώρα μοντέλα, αλλά δέχεται και επαγωγικούς συλλογισμούς, οπότε στην ουσία καταργείται η ισοδυναμία του με ένα τυπικό σύστημα, ισοδυναμία η οποία μέχρι τώρα αποτελούσε αναπόσπαστο στοιχείο κάθε ανάλυσης. Το μοντέλο αυτό δείχνεται, τουλάχιστον όσον αφορά τους μαθηματικούς συλλογισμούς, μη-πλήρες με μία παραλλαγή του επιχειρήματος του Gödel η οποία στηρίζεται στην αναδρομικότητα (υπολογιστότητα) των συναρτήσεων που απαρτίζουν το μοντέλο.

Με βάση τα παραπάνω, φυσιολογικά έπεται και το ανάλογο του 2^{ου} θεωρήματος μη-πληρότητας αποτέλεσμα, σύμφωνα με το οποίο, αν υποθέσουμε ότι το μοντέλο που έχει επιλεγεί έχει τη λειτουργία που του έχουμε προσδώσει, τότε προκύπτει ότι *δεν μπορούμε να διαπιστώσουμε αυτό το γεγονός*, δηλαδή δεν μπορούμε να αναγνωρίσουμε το μοντέλο σαν τέτοιο που να προσομοιώνει τη συλλογιστική μας. Ανάλογες ιδέες έχουν προκύψει αρκετές φορές στην ανάλυση του παρόντος μέρους, ο Putnam όμως δίνει στην δική του μία διαφορετική διάσταση, διότι, αφενός αυτή έχει να κάνει με *όλες* τις μεθόδους μαθηματικού συλλογισμού και αφετέρου, όπως ο ίδιος επισημαίνει, εκτός από την αρνητική χροιά που έχει αυτό το αποτέλεσμα, έχει και μία, απελευθερωτική θα έλεγε κανείς, θετική διάσταση, η οποία παρατίθεται αυτούσια: «Δεν είναι καταπληκτικό! Έχουμε πάντα τη δυνατότητα να πάμε παραπέρα από κάθε συλλογιστική που μπορούμε να επιβλέψουμε και να δούμε ότι είναι ορθή. Η ‘σκέψη σκέψης’ [reflexive reflection] μπορεί πάντα να επιβλέπει τον εαυτό της».

Το γενικότερο συμπέρασμα του Putnam για τα όρια την ανθρώπινης (μαθηματικής, έστω) ενδοσκόπησης που σκιαγραφήθηκε παραπάνω, αναπαράγεται από τον H. Gaifman σε ένα από τα πιο σύγχρονα άρθρα κριτικής των επιχειρημάτων των Lucas/Penrose {#38}, όπου διαπιστώνεται ότι τελικά το θεώρημα του Gödel μας δείχνει ότι «η ενδοσκόπηση δεν μπορεί να συμπεριλάβει όλες τις συλλογιστικές μας δυνατότητες, δεν μπορεί δηλαδή να κατανοήσει τον εαυτό της μέσα στα πλαίσια του ορίζοντά της». Και είναι χαρακτηριστικό ότι αυτή η διαπίστωση προέρχεται από την επισκόπηση της όλης συζήτησης, διότι μπορεί να αποτελεί αυτή η ιδέα μία πιθανή κατεύθυνση προς την οποία πρέπει να κινηθεί η όλη συζήτηση για να ξεφύγει από το τέλμα που έχει περιέλθει. Ίσως δηλαδή να πρέπει πρώτα να αναζητήσουμε τις αδυναμίες ή τα όρια της δικής μας διαδικασίας ενδοσκόπησης και να κατανοήσουμε καλύτερα τη φύση της πριν περάσουμε στις αδυναμίες των μηχανιστικών διαδικασιών, οι οποίες έτσι κι αλλιώς παραμένουν ακόμα τόσες πολλές και τόσο ουσιώδεις, ώστε να μην ‘απειλούμαστε’ άμεσα από αυτές.

Περνάμε τώρα στην παρουσίαση της ‘σύγχρονης’ προέκτασης του τίτλου, ‘σύγχρονης’ με την έννοια ότι το άρθρο που θα αναλύσουμε πολύ σύντομα, καταπιάνεται με τις πιθανές επιπτώσεις του θεωρήματος του Gödel σε ένα πολύ νέο και πειραματικό πεδίο της επιστήμης, το πεδίο της Τεχνητής Ζωής [Artificial Life – AL ή TZ]. Να σημειώσουμε έξ’ αρχής ότι οι επιπτώσεις που μπορεί να έχει το θεώρημα του Gödel στον τομέα την Τεχνητής Νοημοσύνης δεν επεκτείνονται

αυτόματα στον αντίστοιχο της TZ, αν και πρόκειται φυσικά για δύο συναφείς περιοχές επιστημονικής αναζήτησης. Αυτό συμβαίνει διότι η TZ δεν έχει σαν στόχο την επίτευξη της μοντελοποίησης της ανθρώπινης νοημοσύνης αλλά περισσότερο ενδιαφέρεται για την μοντελοποίηση των βασικών φυσικών και βιολογικών διεργασιών των ζωντανών οργανισμών. Για να υπάρχει βέβαια νόημα σε μία τέτοια αναζήτηση μία μορφή αποδοχής της μηχανιστικής κοσμοθεώρησης είναι απαραίτητη και στην ύπαρξη αυτής της τάσης βασίζεται ο J.P. Sullins για να επιχειρηματολογήσει βάσει του θεωρήματος του Gödel.

Χωρίς να επεκταθούμε σε περισσότερες λεπτομέρειες που αφορούν το θεωρητικό κομμάτι της αναζήτησης που περιγράφηκε, αναφέρουμε μόνο ότι ο Sullins εγκαθιδρύει ένα αξίωμα της TZ, σύμφωνα με το οποίο *«Υπάρχει ένα ελάχιστο σύνολο τυπικών αξιωμάτων που μπορούν να χρησιμοποιηθούν για να δημιουργήσουμε ένα πλήρες σύστημα τεχνητών φυσικών διεργασιών οι οποίες είναι ικανές να υποστηρίξουν τεχνητή ζωή»*. Με βάση αυτό το αξίωμα, τη γενικότερη φιλοσοφία του Gödel που σκιαγραφήθηκε στο κεφάλαιο 5 και τη χρήση των θεωρημάτων του, δεν αποτελεί έκπληξη ότι ο Sullins καταλήγει στο συμπέρασμα ότι *«...για καθέναν που μπορεί να πειστεί από τον μαθηματικό ρεαλισμό του Gödel, μία επιστήμη μπορεί να είναι ικανή να δημιουργήσει συμπαγείς προσομοιώσεις έμβιων συστημάτων αλλά ποτέ να μη μπορέσει να δημιουργήσει μία προσομοίωση που να μπορεί να εξελιχθεί σε ένα πλήρες στιγμιότυπο ενός ζωντανού συστήματος»*.

Ο Sullins βέβαια δεν είναι καθόλου απόλυτος στα συμπεράσματά του, κάτι που φαίνεται και από την παραπάνω διατύπωση, μάλλον χρησιμοποιεί το αποτέλεσμα του Gödel για να επιστήσει την προσοχή των ερευνητών της TZ στο ότι το 'σύστημα πραγματικότητας' που θα χρησιμοποιηθεί για την δημιουργία τεχνητής ζωής μπορεί να μην προκύψει ποτέ οντολογικά όμοιο με το δικό μας, οπότε ίσως η απαίτηση για την ύπαρξη αυτής της αναλογίας να πρέπει να επανεξεταστεί.

Συνοψίζοντας, διαπιστώνουμε από αυτές τις δύο σύντομες αναφορές, ότι ουσιαστικά δεν υπάρχουν όρια στο βαθμό τον οποίο μπορεί να αναπαραχθεί η βασική ιδέα πίσω από το θεώρημα του Gödel για να φιλοσοφήσει κανείς ακόμα και πάνω σε θέματα που έχουν να κάνουν με την ίδια τη φύση των σκέψεων που μπορούμε να κάνουμε ή της πραγματικότητας μέσα στην οποία ζούμε. Με απόλυτα δεδομένη λοιπόν τη φιλοσοφική αξία του υπέροχου αυτού θεωρήματος, περνάμε στον τομέα των τελικών συμπερασμάτων και την ολοκλήρωση της διπλωματικής αυτής εργασίας.

7. Τελικά συμπεράσματα

Στο παρόν και τελευταίο κεφάλαιο της εργασίας θα γίνει καταρχήν μία κριτική επισκόπηση όλων όσων μέχρι τώρα έχουν γραφεί και στη συνέχεια θα παρουσιαστούν κάποιοι προσωπικοί προβληματισμοί του γραφόντος πάνω στο όλο θέμα που μας απασχολεί. Τα συμπεράσματα τα οποία θα εξαχθούν χαρακτηρίζονται ‘τελικά’ στον τίτλο του κεφαλαίου, διότι ήδη, και καθ’ όλη τη διάρκεια της ανάλυσης, έχουν παρουσιαστεί μικρότερα συμπεράσματα και σκέψεις πάνω στις απόψεις που έχουν εκφραστεί, οπότε στο κεφάλαιο αυτό θα υπάρχει μία πιο συνολική και ίσως περισσότερο φιλοσοφική αντιμετώπιση των πραγμάτων. Θα γίνει δηλαδή η προσπάθεια να επιτευχθεί μία μακροσκοπική γωνία για το θέμα των συνεπειών αλλά και της ίδιας της φύσης του θεωρήματος του Gödel.

Αρχίζουμε από τον τομέα της συνολικής κριτικής των μέχρι τώρα απόψεων και από μία πρώτη διαπίστωση που αφορά το γενικότερο χαρακτήρα της συζήτησης. Αν παρατηρηθούν καλά τα επιχειρήματα που έχουν εκφραστεί σε κάθε μία από τις περιόδους αντιπαράθεσης διαπιστώνεται ότι κάθε συλλογιστική και κάθε επιχειρηματολογία πάσχει από έναν βαθμό κυκλικότητας. Από τη μία έχουμε τους υποστηρικτές της ανωτερότητας/διαφορετικότητας των ανθρώπινων από τις μηχανιστικές διαδικασίες, οι οποίοι, ξεκινώντας από αυτή τους την πεποίθηση, υποστηρίζουν ότι η θέση τους ενισχύεται από το θεώρημα του Gödel με βάση όμως ένα επιχείρημα το οποίο *a priori* βασίζεται στην πεποίθηση ότι υπάρχει κάτι το διαφορετικό (μη-μηχανιστικό) στη φύση μας που μας κάνει να διαπιστώνουμε την αλήθεια της Gödel-πρότασης του προς μελέτη συστήματος. Από την άλλη, οι ανακατασκευές των αντι-μηχανιστικών επιχειρημάτων και η κριτική αυτής της άποψης βασίζεται με τη σειρά της, είτε στην ισοδυναμία της συλλογιστικής του ανθρώπου με ένα τυπικό σύστημα, ή στην εγγενή αδυναμία του να διαπιστώσει αλήθειες πέρα των ορίων που σκιαγραφούνται από το θεώρημα του Gödel.

Ας πάρουμε για παράδειγμα αναλύσεις σαν αυτές των Benacerraf και Chihara οι οποίοι επιχειρηματολογούν ενάντια στις επιπτώσεις του θεωρήματος στο Μηχανισμό ανασκευάζοντάς το σε μία πιο τυπική βάση και ορίζοντας τη διαίσθηση που κρύβεται πίσω από τη διαπίστωση της αλήθειας της Gödel-πρότασης με *τυπικό* τρόπο. Φυσικά, μία τέτοια ανακατασκευή αλλάζει τελείως τη φύση των υποθέσεων

των υποστηρικτών των επιπτώσεων, όσο και να γίνεται προσπάθεια για τη διατήρηση του αρχικού πνεύματος. Διότι, μετά από την εγκόλπωση της αλήθειας σε ένα τυπικό σύστημα, φυσικό είναι ότι, και πάλι μέσω του θεωρήματος του Gödel, θα καταλήξουμε σε αντιφάσεις. Κανείς όμως δεν παραδέχεται σαν λάθος υπόθεση την αλλαγή του επιχειρήματος προς μία πιο τυπική κατεύθυνση. Η όλη προσοχή στρέφεται σε ποια από τις (τυπικές πλέον) υποθέσεις είναι αυτή που οδηγεί στην αντίφαση. Η κυκλικότητα αυτή σχεδόν επισημαίνεται από τον Penrose, ο οποίος όμως την χρησιμοποιεί με μονομερή τρόπο, χωρίς να την επεκτείνει και στη δική του συλλογιστική, η οποία με τη σειρά της προσπαθεί να αποφύγει τα εμπλεκόμενα αποτελέσματα με επιχειρήματα τα οποία αντλούνται από την εξ' αρχής αποδοχή της αδυναμίας τυποποίησης της έννοιας της αλήθειας.

Η παραπάνω κυκλικότητα δεν είναι, κατά τη γνώμη μου, ισοδύναμη (αρνητικής) αξίας στις δύο περιπτώσεις. Πιο πολύ φαίνεται να επιβαρύνει την κριτική η οποία βασίζεται σε ανακατασκευές και αναδιαμορφώσεις των αρχικών επιχειρημάτων. Αυτές μπορεί να είναι περίτεχνες και να χρησιμοποιούν πολύπλοκες εφαρμογές αποτελεσμάτων της μαθηματικής λογικής, όμως *ξεκινούν από τη θεμελιωδώς λάθος υπόθεση ότι υπάρχει μία τέτοια τυποποιημένη αναδιατύπωση που κρατάει ανέπαφη την ουσία του επιχειρήματος*. Άλλωστε, από το ίδιο το θεώρημα του Gödel μπορούμε να συμπεράνουμε τις ελλείψεις που μπορεί να περιέχονται σε κάθε τυποποίηση εννοιών όπως η αλήθεια ή της αποδειξιμότητας, οπότε πού βρίσκεται η έκπληξη της εκ νέου διαπίστωσής τους και από πού πηγάζει η σιγουριά ότι αυτές οι ελλείψεις οδηγούν στην κατάρριψη των συνεπειών του θεωρήματος του Gödel; Άλλωστε, αν το δούμε λιγότερο αυστηρά το θέμα, πως μπορεί κανείς να κρίνει τη σωστή χρήση του θεωρήματος του Gödel σε περιοχές που ξεφεύγουν από τα αυστηρά μαθηματικά πλαίσια, χρησιμοποιώντας μία εξίσου αμφιλεγόμενη μη μαθηματική παραλλαγή του;

Υπό την παραπάνω οπτική γωνία, ίσως τελικά να είναι περισσότερο σωστή η κριτική που γίνεται *στη βάση* των υποθέσεων που οδηγούν στο αντι-μηχανιστικό επιχείρημα, διότι εκεί βρίσκονται οι φιλοσοφικές πεποιθήσεις του καθενός και όχι στο εργαλείο που χρησιμοποιεί για να τις υποστηρίξει. Μερικές βέβαια από αυτές τις αντιρρήσεις μπορούν να επιτελέσουν και την αντίθετη λειτουργία από την επιθυμητή, εφόσον παρέχουν και μία νέα διέξοδο προς την οποία θα μπορούσαν να κινηθούν τα αντι-μηχανιστικά επιχειρήματα. Για παράδειγμα, οι αντιρρήσεις των Good και Webb θα μπορούσαν να μεταλλαχθούν σε ένα επιχείρημα το οποίο θα όριζε τις διαδικασίες

κατασκευής νέων υπερπεπερασμένων διατακτικών αριθμών, ως τις αντικαταστάτριες της Gödel-πρότασης σαν 'διαδικασία που δεν μπορεί να αναπαραχθεί μηχανιστικά'. Οπότε, τελικά, αναρωτιέται κανείς, τι μένει που να μπορεί να καταδείξει την αδυναμία των αντι-μηχανιστικών συμπερασμάτων που βασίζονται στο θεώρημα του Gödel;

Αυτό που μένει είναι κάτι που ξεκινά από την πιο αδύναμη κριτική ανάμεσα σε αυτές που παρουσιάστηκαν μέχρι τώρα, την κριτική που έχει να κάνει με την πρόταση 'Ο τάδε δεν μπορεί να διαπιστώσει την αλήθεια αυτής της πρότασης'. Αυτή η ιδέα την αδυναμίας του ανθρώπου να διαπιστώσει αλήθειες πέρα από κάποιο (αυτοαναφορικό) όριο, μαζί με σκέψεις σαν αυτές της παραγράφου 6.3 μπορεί να οδηγήσει στο εξής επιχείρημα: «Πως μπορείς να καταδείξεις (κ. Lucas/Penrose) την αδυναμία των μηχανιστικών διαδικασιών να σε περιγράψουν βασιζόμενος σε μία αλήθεια που δεν μπορούν να παράγουν, τη στιγμή που κι εσύ ο ίδιος περιορίζεσαι από αλήθειες τέτοιου τύπου που δεν μπορείς να παράγεις τυπικά; Μήπως τελικά μία μηχανή *οφείλει* να είναι μη-πλήρης εφόσον σε αναπαριστά και εσύ είσαι μη-πλήρης;» Παρατηρείστε όμως ότι η κυκλικότητα εμπεριέχεται *και σε αυτό* το επιχείρημα. Διότι, θα μπορούσε να απαντήσει ο Lucas/Penrose: «Μα ακριβώς επειδή δεν έχω αποκρυπτογραφήσει πλήρως τη φύση μου, δεν μπορώ να την αναπαράγω μηχανιστικά, και επιπλέον πιστεύω ότι αυτό ποτέ δεν θα μπορέσει να γίνει, άρα ες αεί θα είμαι διαφορετικός από μία μηχανή». Στην ουσία δηλαδή, μένουμε πάλι και οι δύο στις αρχικές μας θέσεις και η συζήτηση στο τέλμα της.

Καταλήγουμε λοιπόν στο συμπέρασμα ότι *κανένα* επιχείρημα δεν μπορεί να καταρρίψει με *απόλυτο* τρόπο την ευρύτερη αντι-μηχανιστική άποψη των Lucas και Penrose, αλλά αντίστοιχα και *κανένα* επιχείρημα δεν μπορεί απόλυτα να την ισχυροποιήσει. Μπορεί βέβαια το θεώρημα του Gödel να είναι ένα πολύ καλό υποψήφιο μαθηματικό αποτέλεσμα προς αυτή την κατεύθυνση, αλλά δεν φτάνει αυτό και γίνεται φανερό κάτι τέτοιο από την επιχειρηματολογία του Penrose. Αυτή η τελευταία, εκτός του ότι είναι απρόσεκτα διατυπωμένη, κατά τη γνώμη μου υπολείπεται από αυτή του Lucas σε ένα σημείο που έχουμε ήδη επισημάνει, την διαλεκτικότητα του επιχειρήματος του τελευταίου. Αυτή η δυναμική ιδέα βέβαια κανέναν από τους αντιρρησίες δεν πείθει, διότι ακριβώς δεν μπορεί να εγκολπωθεί σε αυστηρά πλαίσια, οπότε κρίνεται ως ασαφής και σχεδόν δεν αντιμετωπίζεται καθόλου, με εξαίρεση την ιδέα του Lewis (βλ. σελ. 115) η οποία όμως, όντας τυποποιημένη, αντιμετωπίζει τα γνωστά προβλήματα που επισημάνθηκαν έως τώρα.

Σε αυτό το σημείο, μιας και γίνεται κριτική στο κομμάτι των κριτικών, να εκφράσω και μία σχετική απορία για την αντίρρηση που εμφανίζεται ως η ‘κλασική’ απόρριψη των αντι-μηχανιστικών επιχειρημάτων. Σε αυτή την αντίρρηση δίνεται η εντύπωση ότι η αλήθεια της Gödel-πρότασης δεν προκύπτει από την ανάλυση του θεωρήματος αλλά από τη συνεπαγωγή ‘Συνέπεια $\Rightarrow$ Gödel-πρόταση’. Αυτή η συνεπαγωγή αποτελεί όντως μία αληθή πρόταση και ένα βήμα προς την απόδειξη του 2^{ου} θεωρήματος μη-πληρότητας. Τι είναι όμως αυτό που κάνει κάποιους να μη δέχονται την επισήμανση του ίδιου του Gödel ότι, εφόσον η μη-αποκρίσιμη πρόταση είναι όντως μη-αποδείξιμη, και εκφράζει *ακριβώς αυτό το γεγονός*, τότε μπορεί να διαπιστωθεί ότι είναι αληθής έξω από τα πλαίσια του συστήματος; Απάντηση δεν υπάρχει φυσικά, εκτός αν πλέον δεν μπορούμε να θεωρήσουμε τίποτα παραπάνω αληθές για την αριθμητική πέρα από της αλήθειες οι οποίες παράγονται τυπικά από συστήματα όπως η αριθμητική του Peano, μία θέση που, υπό το πρίσμα των αποτελεσμάτων της μη-πληρότητας φαντάζει αδύνατο να ισχύει. Είναι λοιπόν απορία άξιον πως αυτή η αντίρρηση θεωρείται κλασική και πως δεν την έχει απορρίψει η ιστορική πορεία της συζήτησης.

Κλείνουμε αυτή την παρένθεση για να δούμε γιατί το διαλεκτικό επιχείρημα του Lucas είναι ίσως το καλύτερο υποψήφιο για να *υποστηρίζει* (όχι εγκαθιδρύσει) την αντι-μηχανιστική κοσμοθεωρία, όπως αυτό φαίνεται να προκύπτει από τα παραπάνω. Έχουμε ήδη επισημάνει (βλ. σελ. 139) πως η δυναμικότητα του επιχειρήματος είναι το μεγαλύτερο προτέρημά της διότι αποτυπώνει την αντίστοιχη δυναμικότητα του ανθρώπινου νου. Ένα ακόμη πλεονέκτημα είναι ότι με την διαλεκτικότητα, όχι μόνο αποφεύγεται η αδυναμία του ανθρώπου να γνωρίζει αλήθειες πέρα από κάποια όρια, αλλά αυτή η ‘αδυναμία’, ουσιαστικά ισχυροποιεί το επιχείρημα αφού παρέχει τελικά μία ανεξάντλητη πηγή αληθειών που προσφέρει κάθε φορά μία νέα Gödel-πρόταση η οποία καταρρίπτει τη θέση ότι ένα μηχανιστικό μοντέλο είναι μοντέλο του νου. Θα μπορούσε βέβαια κάποιος να αντικρούσει (μαζί με τον Good) ότι το σχήμα αντίκρουσης του Lucas μπορεί να ‘γυρίσει ανάποδα’ διότι κάθε καινούρια διαίσθηση του ανθρώπου θα μπορεί να ενσωματωθεί σε μία καινούρια μηχανή. Όμως εφόσον η θέση η οποία κρίνεται είναι η μηχανιστική, αναπόφευκτα θα πρέπει ο Μηχανιστής να έχει τον πρώτο λόγο στην προσπάθειά του να την εγκαθιδρύσει, παράγοντας ένα μοντέλο ισοδύναμο με τις δυνατότητες του ανθρώπινου νου, οπότε τότε, και σε κάθε αντίστοιχη προσπάθεια, θα βρει μπροστά του το διαλεκτικό επιχείρημα του Lucas.

Παρατηρείστε ότι τα παραπάνω επιχειρήματα δεν ξεφεύγουν από την κυκλικότητα που περιγράφηκε εφόσον ο γραφών δεν είναι, όντως, υποστηρικτής μιας μηχανιστικής θεωρίας για το νου, τουλάχιστον όπως αυτή περιγράφεται στο παρόν μέρος της εργασίας. Προς υπεράσπισήν μου, να δηλώσω ότι τα παραπάνω *δεν στοχεύουν στην κατάρριψη του Μηχανισμού*, άλλωστε έχω ήδη δηλώσει πως δεν πιστεύω ότι κάτι τέτοιο μπορεί να επιτευχθεί με χρήση αποκλειστικά και μόνο των επιχειρημάτων που πηγάζουν από το θεώρημα του Gödel. Απλώς επισημαίνεται ότι η προσωπική μου πεποίθηση είναι ότι τελικά το διαλεκτικό επιχείρημα του Lucas μπορεί καλύτερα από κάθε άλλο να υποστηρίξει μία αντι-μηχανιστική κοσμοθεωρία. Αποτελεί δηλαδή το πιο δυνατό επιχείρημα για να πειστεί κάποιος *ο οποίος ήδη δεν πιστεύει στον Μηχανισμό*, ακόμα περισσότερο ότι οι θέσεις του είναι σωστές. Για κάποιον ο οποίος υποστηρίζει το Μηχανισμό, το επιχείρημα, όπως και κάθε άλλο αντίστοιχό του, δεν είναι πειστικό, όμως είναι πειστικές οι αντιρρήσεις που εκφράζονται πάνω σ' αυτό. Το ποια από αυτές είναι η πειστικότερη, αφήνεται να κριθεί από κάποιον υποστηρικτή του.

Με βάση τις παραπάνω σκέψεις και κρατώντας ίσες αποστάσεις από τις δύο αντιμαχόμενες θεωρίες εγκαθιδρύθηκε η ελάχιστη βάση πάνω στην οποία μπορεί το θεώρημα του Gödel να έχει επιπτώσεις στο Μηχανισμό. Θα γίνει τώρα μία απόπειρα να γίνει μία σύνθεση *όλων* των παρατιθέμενων απόψεων, χωρίς να θυσιαστεί αυτό το κεκτημένο. Αυτή η σύνθεση δεν μπορεί να είναι παρά ένα καθαρά φιλοσοφικό κατασκεύασμα όμως ίσως, όπως και κάθε σύνθεση, να βρίσκεται τελικά πιο κοντά στην αλήθεια για το που μπορούν να μας οδηγήσουν οι συνέπειες του θεωρήματος του Gödel στη φιλοσοφία του νου.

Έχουμε διαπιστώσει ότι το θεώρημα του Gödel έχει να κάνει με αλήθειες οι οποίες βρίσκονται έξω από κάποια όρια που έχουμε καθορίσει για ένα τυπικό μαθηματικό σύστημα. Οι αλήθειες αυτές, όσον αφορά τα μαθηματικά, πράγματι ξεπερνούν τα όρια ενός δεδομένου συστήματος και μπορούν να γίνουν αντιληπτές μόνο μέσω ενός ανώτερου συστήματος. Υπάρχει όμως μία εκτενής ανησυχία αν το ίδιο μπορεί να ισχύει και για τον άνθρωπο, ανησυχία που εκφράζεται από τους προβληματισμούς που αφορούν τη συνέπεια του ανθρώπου ή την απουσία απόδειξης

της ή ακόμα και την παντελή απουσία της, προβληματισμοί που όπως είδαμε αποτελούν και το βαθύτερο κομμάτι των συνεπειών του θεωρήματος στη φιλοσοφία του νου. Μέσω των παρακάτω διαπιστώσεων πάνω στο θέμα ψάχνουμε μία δίοδο που θα μας οδηγήσει στην επιθυμητή σύνθεση των απόψεων.

Το βασικό μας θέμα είναι η φύση της απαίτησης για την *απόδειξη* της ανθρώπινης συνέπειας, θέμα που φαίνεται ότι αποτελεί μία υποσυνείδητη προσπάθεια των οπαδών του Μηχανισμού να μας παρασύρουν στην επέκταση αποτελεσμάτων και ορολογίας που αφορά καθαρά την επιστήμη των μαθηματικών σε όλο το φάσμα της ζωής. Αναπαράγουν δηλαδή οι Μηχανιστές (όχι απαραίτητα σκόπιμα) μέσω αυτής της απαίτησης ιδέες της κοσμοθεωρίας τους. Διότι, σύμφωνα με τη μηχανιστική θέση, εφόσον η ισοδυναμία του ανθρώπου-μηχανής υφίσταται με την έννοια που έχει ορίζει ο Putnam (βλ. σελ. 77) και δεν έχει αλλάξει καθ' όλη τη διάρκεια της συζήτησης, τότε, οτιδήποτε μπορεί να αποδείξει ο άνθρωπος μπορεί να το αποδείξει και η μηχανή. Έτσι όμως, είναι φυσιολογικό, δεδομένου του 2^{ου} θεωρήματος μη-πληρότητας να προβάλλεται σαν απαίτηση η απόδειξη της συνέπειας του ανθρώπου, ή σαν επιχείρημα η ανυπαρξία της. *Όμως η προσπάθεια εξίσωσης μαθηματικών αποτελεσμάτων με αποτελέσματα που στοχεύουν σε ολόκληρο το φάσμα της ανθρώπινης ύπαρξης είναι άκρως προβληματική.* Διότι το ερώτημα για την απόδειξη της συνέπειας του ανθρώπου είναι ανάλογο, για παράδειγμα, του ερωτήματος 'Πώς ξέρει κανείς ότι υπάρχει;' Και θεωρείστε μία κατάσταση στην οποία εγώ *ξέρω* ότι υπάρχω. Πως μπορώ να πείσω κάποιον άλλον ότι *αυτός* υπάρχει; Αν αυτός επιμένει ότι εγώ είμαι αποκύημα της φαντασίας του το οποίο σκοπεύει να τον παραπλανήσει πάνω στο θέμα της ύπαρξής του, πώς μπορώ εγώ να τον πείσω για το αντίθετο; Τι *απόδειξη* μπορώ να του παρέχω; Φυσικά καμία, ή τουλάχιστον καμία πειστική γι' αυτόν. Δεν έχουμε να κάνουμε με μαθηματικά όπου μία απόδειξη για τον ένα μαθηματικό γίνεται δεκτή από έναν άλλο χωρίς ουσιαστικά προβλήματα. Έχουμε να κάνουμε με θεμελιώδη ερωτήματα που αφορούν την ανθρώπινη ύπαρξη και μας απασχολούν από τις απαρχές της ζωής. Και αν δεν είμαστε διατεθειμένοι να καταργήσουμε τη Φιλοσοφία στο σύνολό της, τότε θα πρέπει να δεχθούμε ότι για κάποια πράγματα είτε δεν υπάρχει 'απόδειξη', είτε, εάν υπάρχει, η αποδοχή της δεν είναι μία καθόλου τετριμμένη υπόθεση, είναι μάλλον μία υπόθεση που έχει απασχολήσει γενιές και γενιές ανθρώπων.

Θέλω να πιστεύω ότι τα παραπάνω δεν θα εκληφθούν ως μία στείρα απόρριψη του Μηχανισμού, άλλωστε ο στόχος για τον οποίο διατυπώθηκαν δεν ήταν αυτός. Ο στόχος είναι να εγκαθιδρύσουμε ένα ελάχιστο ποσοστό πραγμάτων τα οποία δείχνουν ότι στην προσπάθεια αποσαφήνισης των λειτουργιών σκέψης του ανθρώπου εμπλέκονται έννοιες οι οποίες δεν είναι καθόλου εύκολο να προσομοιωθούν σε οποιοδήποτε σύστημα. Άλλωστε αυτό γίνεται αποδεκτό και από τους θιασώτες του Μηχανισμού όταν επιχειρηματολογούν υπέρ μελλοντικών συστημάτων τα οποία θα είναι ισοδύναμα με τον άνθρωπο, τόσο πολύπλοκα δε, ώστε να μη μπορούμε να αναγνωρίσουμε τη συνέπειά τους ή πολύ παραπάνω, να μη μπορούμε τελικά να τα αναγνωρίσουμε καν ως τέτοια που να μας αναπαριστούν. Μετουσιώνουν όμως τέτοια συστήματα την ουσία του Μηχανισμού όπως τον έχουμε στο μυαλό μας; Ο Lucas ισχυρίζεται πως όχι και ο Gödel συνηγορεί εκφράζοντας την άποψη ότι ο δημιουργός θα πρέπει να γνωρίζει τις ακριβείς ιδιότητες του δημιουργήματός του.

Φαίνεται λοιπόν ότι για να έρθει η λύση στο πρόβλημα της μηχανιστικής αναπαράστασης του νου που θα συγκεντρώνει όλες τις απόψεις και θα συμβιβάζει όλα τα δεδομένα θα πρέπει ο Μηχανισμός να αλλάξει χαρακτήρα και να αφήσει πίσω του μία από τις βασικές απαιτήσεις του, ήτοι την πλήρη κατανόηση των όσων θα πρεσβεύει. Θα πρέπει δηλαδή να ξαναγυρίσουμε στις αρχικές απαιτήσεις του Turing για μηχανές των οποίων δεν θα μπορούμε να προσδιορίσουμε ακριβώς τη λειτουργία, για να υπάρξει η δυνατότητα δημιουργίας μίας μηχανής η οποία θα αναπαριστά πειστικά ένα κομμάτι των λειτουργιών του νου. Αυτή όμως η μηχανή θα λειτουργεί με βάση αρχές οι οποίες δεν θα μπορούν να αναχθούν στην ιδέα που έχουμε σήμερα για το Μηχανισμό, οπότε όλοι οι σημερινοί συζητητές θα βρίσκονται εν αρμονία, οι μεν διότι έχουν το επιθυμητό μοντέλο περιγραφής του νου στα χέρια τους και οι δε διότι η λειτουργία αυτού του μοντέλου θα είναι από τη φύση της αδύνατο να περιγραφεί, οπότε το ίδιο θα ισχύει και για τον ανθρώπινο νου! Και ίσως τελικά η όλη συζήτηση, αλλά και το αποτέλεσμα του Gödel αυτό καθ' ευατό να συνηγορούν προς μία τέτοια κατεύθυνση, κατεύθυνση σύμφωνα με την οποία θα πρέπει να εγκαταλειφθούν οι προσπάθειες αναγωγής της ανθρώπινης σκέψης σε αυστηρά καθορισμένα τυπικά συστήματα ή σε μηχανές διακριτών καταστάσεων. Διότι αυτό που κάνει μοναδικό τον τρόπο με τον οποίο σκεπτόμαστε είναι ακριβώς το γεγονός ότι είναι αδύνατο να χαλιναγωγήσουμε τη σκέψη μας ακόμα και με τα πιο περίτεχνα κατασκευάσματα που η ίδια μπορεί να παράγει. Η σκέψη τα φτιάχνει, η σκέψη τα επεξεργάζεται και η σκέψη τα χαλάει.-

ΒΙΒΛΙΟΓΡΑΦΙΑ

- 01.** Gödel, K. 1931. On formally undecidable propositions of Principia Mathematica and related Systems I. *Collected Works vol. I*.
- 02.** Turing, A.M. 1950. Computing Machinery and Intelligence. *Mind*: 433-460.
- 03 a.** Gödel, K. 1951. Some basic theorems in the foundations of mathematics and their implications. Gibbs Lecture. *Collected Works vol. III*.
- b.** Gödel, K. 1946. Παρατηρήσεις πάνω στα προβλήματα των μαθηματικών στο συνέδριο για τα 200 χρόνια του Princeton. Μία σημείωση του Gödel πάνω στις μηχανιστικές διαδικασίες. Διατίθεται στον τόμο *'The undecidable. Basic Papers on undecidable propositions, unsolvable problems and computable functions'*. Raven Presss, Hewlett, N.Y. 1965.
- 04.** Putnam, H. 1960. Minds and Machines, (πρώτη δημοσίευση στο) *Sidney Hook, ed., Dimensions of Mind. A Symposium*, New York, 1960. [Επίσης διατίθεται στο *'Mind, Language and Reality'* Philosophical Papers vol. 2 (CUP 1975), του ιδίου]
- 05.** Lucas, J.R. 1961. Minds, machines and Gödel. *Philosophy* 36:112-127.
- 06.** Whiteley, C. 1962. Minds, machines and Gödel: A reply to Mr. Lucas. *Philosophy* 37:61-62.
- 07.** George, F. 1962. Minds, machines and Gödel: Another reply to Mr. Lucas. *Philosophy* 37:62-63.
- 08.** Good, I.J. 1967. Human and machine logic. *British Journal for the Philosophy of Science* 18:145-6.
- 09.** Lucas, J.R. 1967. Human and machine logic: a rejoinder. *British Journal for the Philosophy of Science* 19:155-6.
- 10.** Benacerraf, P. 1967. God, the Devil, and Gödel. *Monist* 51:9-32.
- 11.** Lucas, J.R. 1968. Satan stultified: A rejoinder to Paul Benacerraf. *Monist* 52:145-58.
- 12.** Webb, J. 1968. Metamathematics and the philosophy of mind. *Philosophy of Science* 35:156-78.
- 13.** Good, I.J. 1969. Gödel's theorem is a red herring. *British Journal for the Philosophy of Science* 19:357-8.

- 14.** Lewis, D. 1969. Lucas against mechanism. *Philosophy* 44:231-3.
- 15.** Coder, D. 1969. Gödel's theorem and mechanism. *Philosophy* 44:234-7.
- 16.** Lucas, J.R. 1970. Mechanism: A rejoinder. *Philosophy* 45:149-51.
- 17.** Lucas, J.R. 1971. Metamathematics and the philosophy of mind: A rejoinder. *Philosophy of Science* 38:310-13.
- 18.** Chihara, C. 1972. On alleged refutations of mechanism using Gödel's incompleteness results. *Journal of Philosophy* 64:507-26.
- 19.** Hutton, A. 1976. This Gödel is killing me. *Philosophia* 3:135-44.
- 20.** Lucas, J.R. 1976. This Gödel is killing me: A rejoinder. *Philosophia* 6:145-8.
- 21.** Hofstadter, D.R. 1979. *Godel, Escher, Bach: An Eternal Golden Braid*. Basic Books.
- 22.** Lewis, D. 1979. Lucas against mechanism II. *Canadian Journal of Philosophy* 9:373-6.
- 23.** Bowie, G. 1982. Lucas' number is finally up. *Journal of Philosophy Logic* 11:279-85.
- 24.** Lucas, J.R. 1984. Lucas against mechanism II: A rejoinder. *Canadian Journal of Philosophy* 14:189-91.
- 25.** Putnam, H. 1985. Reflexive reflections. *Erkenntnis* 22:143-153.
- 26.** Penrose, R. 1989. *The Emperor's New Mind*. Σελ. 129-146 & 538-541. Oxford University Press.
- 27.** Lucas, J.R. 1990 Mind, machines and Gödel: A retrospect. In (P. Millican & A. Clark, eds) *Machines and Thought*. Oxford University Press.
- 28.** Davis, M. 1990. Is Mathematical Insight Algorithmic?
- 29.** Davis, M. 1993. How Subtle is Gödel's Theorem. More on Roger Penrose.
- 30.** McCullough, D. 1996. Can humans escape Gödel? *Psyche* 2:57-65.
- 31.** McDermott, D. 1996. [Star] Penrose is wrong. *Psyche* 2:66-82.
- 32.** Feferman, S. 1996. Penrose's Gödelian argument. *Psyche* 2:21-32.

33. Chalmers, D.J. 1996. Minds, machines, and mathematics. *Psyche* 2:11-20.
34. Penrose, R. 1996. Beyond the doubting of a shadow. *Psyche* 2:89-129.
35. Lucas, J.R. 1996. The Gödelian Argument: Turn Over the Page. [Μία ομιλία στο συνέδριο BSPS στην Οξφόρδη, διαθέσιμη στο: <http://users.ox.ac.uk/~jrlucas/>].
36. Sullins P.J. 1997. Gödel's Incompleteness Theorems and Artificial Life. *Phil & Tech* 2:3-4.
37. McCall, S. 1999. Can a Turing machine know that the Gödel sentence is true? *Journal of Philosophy* 96:525-32.
38. Gaifman, H. 2000. What Gödel's incompleteness result does and does not show. *Journal of Philosophy* 97:462-471.
39. Lindström P. 2001. Penrose's New Argument. *Journal of Philosophical Logic* 30:241-250.

Άλλες πηγές:

- <http://plato.stanford.edu/> (on-line Εγκυκλοπαίδεια της Φιλοσοφίας του πανεπιστημίου του Stanford).
- Ζάχος, Ε. 1995. *Μοντέλα υπολογισμού, Τυπικές Γλώσσες, Θεωρία αυτομάτων*.
- Celluci, C. 1998. The scope of logic: deduction, abduction, analogy. *Theoria* 64:217-242.
- Δημητρακόπουλος Κ.Ι. 1999. 'Σημειώσεις Μαθηματικής Λογικής'.
- Κολέτσος Γ. 'Σημειώσεις Μαθηματικής Λογικής'.

ΠΑΡΑΡΤΗΜΑ

ΑΠΟΔΕΙΞΗ ΤΟΥ ΘΕΩΡΗΜΑΤΟΣ ΤΟΥ GÖDEL

- Αναδρομικές συναρτήσεις και κατηγορήματα
- Πρωτοβάθμια Αριθμητική
- Αριθμητικοποίηση της Λογικής
- Εφαρμογές
- 2^ο Θεώρημα μη-πληρότητας

Επιμέλεια: κ. Γιώργος Κολέτσος

Μετάφραση: Τζανής Δημήτρης

Επεξεργασία LaTeX: Φραγκίσκος Σωτήρης

1. ΑΝΑΔΡΟΜΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΚΑΙ ΚΑΤΗΓΟΡΗΜΑΤΑ

Σημειογραφία:

- ω = Σύνολο των φυσικών αριθμών. Όλες οι μεταβλητές (συμβ. με x, y, z κτλ.) και οι σταθερές (συμβ. a, b, c , κτλ.), έχουν πεδίο ορισμού το ω .
- Τα F, G, H δηλώνουν συναρτήσεις: $\omega^n \rightarrow \omega$.
- Τα P, Q, R δηλώνουν n -αδικά κατηγορήματα στο ω - δηλαδή υποσύνολα του ω^n .
- Το $\vec{x}$ είναι συντομογραφία του $x_1 \cdots x_n$
- $\forall \vec{x} = \forall x_1 \cdots \forall x_n$

Ορισμός 1.1. Μια συνάρτηση $F : \omega^n \rightarrow \omega$ είναι **υπολογιστή**, αν υπάρχει μια μηχανιστική διαδικασία (αλγόριθμος) $\mathcal{P}_F$ τέτοια ώστε δοθέντος οποιουδήποτε $a_1 \cdots a_n$ ο $\mathcal{P}_F$ να μας δίνει τη δυνατότητα να υπολογίσουμε το $F(a_1 \cdots a_n)$ σε πεπερασμένο χρόνο.

Ορισμός 1.2. Για κάθε κατηγορήμα $P \subseteq \omega^n$, η **χαρακτηριστική συνάρτηση** του P (συμβολίζεται C_P) ορίζεται ως εξής:

$$C_P(\vec{x}) = \begin{cases} 0, & \text{αν } P(\vec{x}) \\ 1, & \text{αν } \neg P(\vec{x}) \end{cases}$$

Λέμε ότι ένα κατηγορήμα P είναι υπολογιστό αν και μόνο αν η C_P είναι υπολογιστή.

Ένα **πρόβλημα απόφασης** είναι ένα πρόβλημα του παρακάτω είδους:

Έστω A ένα σύνολο (συνήθως το ω ή το ω^n) και $B \subseteq A$.

Τότε μια **διαδικασία απόφασης για το B στο A** είναι ένας αλγόριθμος τέτοιος ώστε, δοθέντος οποιουδήποτε $x \in A$ να μπορούμε να καθορίσουμε, μετά από πεπερασμένο αριθμό βημάτων αν $x \in B$ ή $x \notin B$.

Σε σχέση με το παραπάνω, αν $A = \omega$ και $P(x) \leftrightarrow x \in B$, τότε μια διαδικασία απόφασης για το B στο A υπάρχει αν και μόνο αν το P είναι υπολογιστό (δηλαδή η C_P είναι υπολογιστή).

Παραδείγματα προβλημάτων απόφασης:

- (1) Έστω A = σύνολο από τύπους της προτασιακής λογικής (με πεπερασμένο αριθμό συμβόλων) και B = σύνολο των ταυτολογιών.
- (2) Έστω A = σύνολο τύπων του κατηγορηματικού λογισμού και B = σύνολο από λογικά έγκυρους τύπους.
- (3) Έστω A = σύνολο διοφαντικών εξισώσεων και B = σύνολο διοφαντικών εξισώσεων οι οποίες έχουν λύση.

Ορισμός 1.3. Για i, n τέτοια ώστε $1 \leq i \leq n$, ορίζουμε τις συναρτήσεις προβολής ως εξής:

$$I_i^n(x_1 \cdots x_n) = x_i.$$

Έστω Σ μια κλάση συναρτήσεων. Υποθέστε τα ακόλουθα **αξιώματα** για την Σ .

(R1 - Βασικές Συναρτήσεις). $Z, \sigma, +, \cdot, C_<$ και I_i^n (για $1 \leq i \leq n$) όλα $\in \Sigma$, όπου:

- (1) $Z : \omega \longrightarrow \omega$ (η μηδενική συνάρτηση) που ορίζεται ως $Z(x) = 0 \quad \forall x \in \omega$
- (2) $\sigma : \omega \longrightarrow \omega$ (η συνάρτηση του επόμενου) που ορίζεται ως $\sigma(x) = x+1 \quad \forall x \in \omega$

Η πρόσθεση και ο πολλαπλασιασμός ορίζονται: $\omega^2 \longrightarrow \omega$ ως είθισται.

Τέλος, η συνάρτηση διάταξης: $C_< : \omega^2 \longrightarrow \omega$.

Αυτές είναι οι **βασικές συναρτήσεις**.

(R2 - Αξίωμα της Αντικατάστασης). Αν οι συναρτήσεις $G, H_1 \cdots H_R$ ανήκουν στη Σ και η F δίνεται από την

$$F(\vec{x}) = G(H_1(\vec{x}), H_2(\vec{x}), \dots, H_R(\vec{x})),$$

τότε $F \in \Sigma$.

(R3 - Αξίωμα της Αναδρομής). Αν $G, H \in \Sigma$ και η F δίνεται από τον επαγωγικό ορισμό:

$$\begin{cases} F(0, \vec{x}) &= G(\vec{x}) \\ F(y+1, \vec{x}) &= H(f(y, \vec{x}), y, \vec{x}) \end{cases},$$

τότε $F \in \Sigma$.

Ορισμός 1.4. Η κλάση των **πρωταρχικά αναδρομικών συναρτήσεων** (Πρωτ.Αν.), είναι η μικρότερη κλάση Σ που ικανοποιεί τα αξιώματα R1, R2, R3.

Όλες οι "συνηθισμένες" συναρτήσεις βγαίνουν ότι είναι Πρωτ.Αν., λ.χ. η εκθετική συνάρτηση $\exp(x, y) = x^y$ δίνεται από:

$$\begin{cases} \exp(x, 0) = 1 (= \sigma(Z(x))) \\ \exp(x, y+1) = x \cdot \exp(x, y) \end{cases}$$

Λήμμα 1.5. Αν η κλάση Σ ικανοποιεί τα R1, R2 και αν η $\sigma \in \Sigma$, τότε αν $x_1 \cdots x_n$ είναι διακριτές μεταβλητές ($x_i \neq x_j \forall i, j$) και $\{z_1 \cdots z_R\} \subseteq \{x_1 \cdots x_n\}$ (οι $z_1, \dots, z_R$ όχι απαραίτητα διακριτές) και η F δίνεται από τον τύπο:

$$F(x_1 \cdots x_n) = G(z_1 \cdots z_R),$$

τότε $F \in \Sigma$.

[Σημείωση: Μπορεί να έχουμε $R > n$ (αφού μη διακριτά z_i)]

Απόδειξη. Έστω ότι το z_i είναι το $x_{ji} \forall i : 1 \leq i \leq R$. Τότε η F δίνεται από τον τύπο:

$$F(x_1 \cdots x_n) = G(I_{j_1}^n(x_1 \cdots x_n), \dots, I_{j_R}^n(x_1 \cdots x_n))$$

και $I \in \Sigma$ και $G \in \Sigma$ (από R1) $\Rightarrow F \in \Sigma$ (από R2). □

Παράδειγμα: Αν $G \in \Sigma$ και $F(x_1, x_2) = G(x_1, x_2, x_1)$ τότε $F \in \Sigma$

Παρατήρηση: Γενικά μπορούμε να **αναγνωρίσουμε**, να **αντιμεταθέσουμε** και να **προσθέσουμε** μεταβλητές.

Έστω κατηγορήμα $P(\vec{x}, y)$ τέτοιο ώστε $\forall \vec{x} \exists y P(\vec{x}, y)$. Τότε μπορούμε να ορίσουμε τελεστή μ ως ακολούθως:

$$\mu y P(\vec{x}, y) = \text{το ελάχιστο } k, \text{ τέτοιο ώστε } P(\vec{x}, k).$$

Ο μ είναι ο ελαχιστοποιητής τελεστής [μ -τελεστής από δω και στο εξής].

Σημ: Ο μ -τελεστής ορίζεται **μόνο** όταν ο τύπος $\forall \vec{x} \exists y P(\vec{x}, y)$ είναι έγκυρος.

Τώρα υποθέτουμε:

(R4 - μ -Αξίωμα). Αν $G \in \Sigma$ και $\forall \vec{y} \exists x (G(\vec{y}, x) = 0)$, τότε $\mu x (G(\vec{y}, x) = 0) \in \Sigma$.

Ορισμός 1.6. Η κλάση των **αναδρομικών** συναρτήσεων, (Αναδ.), είναι η μικρότερη κλάση Σ που ικανοποιεί τα αξιώματα R1, R2, R4.

Ορισμός 1.7. Ένα κατηγορήμα P είναι **αναδρομικό** (αντίστοιχα Πρωτ.Αν.) αν η χαρακτηριστική του συνάρτηση, C_P , είναι αναδρομική (αντίστοιχα Πρωτ.Αν.).

Ακολουθούν ιδιότητες και ορισμοί για τις αναδρομικές συναρτήσεις και κατηγορήματα (C1 – C10).

C 1. Αν το Q είναι αναδρομικό κατηγορήμα, οι $F_1, \dots, F_R$ είναι αναδρομικές συναρτήσεις και το P είναι κατηγορήμα που δίνεται από τον τύπο

$$P(\vec{x}) \longleftrightarrow Q(F_1(\vec{x}), \dots, F_R(\vec{x})),$$

τότε το P είναι αναδρομικό.

Απόδειξη. Τετριμμένη αφού

$$C_P(\vec{x}) = C_Q(F_1(\vec{x}), \dots, F_R(\vec{x})).$$

□

C 2. Αν το P είναι αναδρομικό κατηγορήμα και $\forall \vec{y} \exists x P(\vec{y}, x)$, τότε η συνάρτηση F που δίνεται από τον τύπο $F(\vec{y}) = \mu x P(\vec{y}, x)$ είναι αναδρομική.

Απόδειξη.

$$F(\vec{y}) = \mu x (C_p(\vec{y}, x) = 0) .$$

□

Μια συνάρτηση ή κατηγορήμα έχει έναν **σαφή ορισμό** σε σχέση με συναρτήσεις $F_1, \dots, F_R$ και κατηγορήματα $P_1, \dots, P_l$ αν μπορεί να οριστεί σε σχέση με αυτές τις συναρτήσεις και κατηγορήματα, χρησιμοποιώντας μόνο αντικατάσταση και τον μ-τελεστή, δηλαδή αν ανήκει στην μικρότερη κλάση Σ τ.ω.

- (i) $F_1, \dots, F_R, P_1, \dots, P_l \in \Sigma$.
- (ii) Αν $Q \in \Sigma$ και $\forall \vec{x} \exists y Q(\vec{x}, y)$ τότε $\mu y Q(\vec{x}, y) \in \Sigma$.
- (iii) Η Σ είναι κλειστή υπό την αντικατάσταση συναρτήσεων σε συναρτήσεις ή σε κατηγορήματα.

Λήμμα 1.8. Αν μια συνάρτηση ή κατηγορήμα έχει ένα σαφή ορισμό σε σχέση με αναδρομικές συναρτήσεις και κατηγορήματα, τότε είναι αναδρομική/ό.

Απόδειξη. Βγαίνει άμεσα από τα C1, C2, R2, R4.

□

C 3. Κάθε σταθερή συνάρτηση είναι αναδρομική.

C 4. Αν τα P, Q είναι αναδρομικά, τότε και τα $\neg P, P \longrightarrow Q, P \vee Q, P \wedge Q$ κλπ. είναι αναδρομικά.

Σημ.: Αν P, Q κατηγορήματα, τότε τα $\neg P, P \longrightarrow Q, P \vee Q$ κλπ. ορίζονται ως εξής:

$$\begin{aligned} (\neg P)(\vec{x}) &\leftrightarrow \neg P(\vec{x}) \\ (P \longrightarrow Q)(\vec{x}) &\leftrightarrow \neg P(\vec{x}) \vee Q(\vec{x}) \\ &\dots \text{ κλπ.} \end{aligned}$$

Απόδειξη. Το δείχνουμε για επαρκές σύνολο συνδέσμων $\{\neg, \vee\}$, ήτοι για το $\neg P$ και το $P \vee Q$. Έχουμε:

$$C_{\neg P}(\vec{x}) = C_{<}(0, C_P(\vec{x})), \quad \text{που είναι αναδρομική από } R1, R2,$$

και

$$C_{P \vee Q}(\vec{x}) = C_P(\vec{x}) \cdot C_Q(\vec{x}), \quad \text{που είναι αναδρομική από } R1, R2.$$

□

C 5. Τα δυαδικά κατηγορήματα $<, \leq, >, \geq, =$ είναι όλα αναδρομικά.

Απόδειξη. Το $<$ είναι αναδρομικό από R1. Τα άλλα έχουν σαφείς ορισμούς:

$$\begin{aligned} x \leq y &\leftrightarrow \neg(y < x) \\ x > y &\leftrightarrow y < x \\ x \geq y &\leftrightarrow \neg(y > x) \\ x = y &\leftrightarrow \leq \wedge y \leq x. \end{aligned}$$

□

Από C3, C4, C5, μπορούμε να σχηματίσουμε αναδρομικές συναρτήσεις και κατηγορήματα χρησιμοποιώντας τα $<$, $\leq$, $=$, σταθερές και προτασιακούς συνδέσμους.

C 6. Ορίζουμε την "πλον" συνάρτηση, $\dot{-}$, ως εξής:

$$x \dot{-} y = \begin{cases} x - y & \text{αν } x \geq y \\ 0 & \text{αλλιώς.} \end{cases}$$

Αυτή έχει ένα σαφή ορισμό, $x \dot{-} y = \mu z(y + z = x \vee x < y)$, άρα είναι αναδρομική.

Ο φραγμένος μ-τελεστής ορίζεται ως ακολούθως: για κάθε κατηγορήμα $P(\vec{y}, x)$

$$\mu x_{x < z} P(\vec{y}, x) \stackrel{\text{ορ.}}{=} \mu x (P(\vec{y}, x) \vee x = z).$$

Προφανώς αυτός είναι καλώς ορισμένος και επιπλέον, αν το P είναι αναδρομικό ($P \in \text{Αναδ.}$) τότε είναι και το $\mu x_{x < z} P(\vec{y}, x)$.

Σημ.: Το $\mu x_{x < z}$ σημαίνει "το ελάχιστο $x (< z)$ αν αυτό υπάρχει ή z αλλιώς, τέτοιο ώστε ..."

C 7. Αν το $P(\vec{y}, x)$ είναι αναδρομικό (x διακριτό από $\vec{y}$), η $H(\vec{y})$ είναι αναδρομική, και η F δίνεται από τον τύπο:

$$F(\vec{y}) = \mu x_{x < H(\vec{y})} P(\vec{y}, x),$$

τότε η F είναι αναδρομική.

Απόδειξη. Από C4 και C5, η $P(\vec{y}, x) \vee x = H(\vec{y})$ είναι αναδρομική και άρα από R4, η $\mu x [P(\vec{y}, x) \vee x = H(\vec{y})]$ είναι αναδρομική.

□

Από το C7 βλέπουμε ότι φραγμένοι ποσοδείκτες δεν οδηγούν έξω από την κλάση των αναδρομικών κατηγορημάτων.

C 8. Αν το $P(\vec{y}, x)$ και η $H(\vec{y})$ είναι αναδρομικά (x διακριτό από το $\vec{y}$), τότε τα κατηγορήματα Q_1 και Q_2 που δίνονται από τους τύπους:

$$Q_1(\vec{y}) \longleftrightarrow \exists x_{x < H(\vec{y})} P(\vec{y}, x),$$

και

$$Q_2(\vec{y}) \longleftrightarrow \forall x_{x < H(\vec{y})} P(\vec{y}, x),$$

είναι αναδρομικά.

Απόδειξη. Για τα Q_1, Q_2 έχω ότι:

$$\begin{aligned} Q_1(\vec{y}) &\longleftrightarrow (\mu x_{x < H(\vec{y})} P(\vec{y}, x)) < H(\vec{y}) \\ Q_2(\vec{y}) &\longleftrightarrow (\mu x_{x < H(\vec{y})} \neg P(\vec{y}, x)) = H(\vec{y}) \end{aligned}$$

και τα δεύτερα μέλη είναι αναδρομικά από C7 και προηγούμενα αποτελέσματα.

□

C 9. (Ορισμός ανά περιπτώσεις:)

Έστω $G_1, \dots, G_R$ αναδρομικές συναρτήσεις και $R_1, \dots, R_R$ αναδρομικά κατηγορήματα τ.ω. για κάθε $\vec{x}$ ένα και μόνο ένα από τα $R_1(\vec{x}), \dots, R_R(\vec{x})$ να ισχύει. Τότε η συνάρτηση F που ορίζεται ως:

$$F(\vec{x}) = \begin{cases} G_1(\vec{x}) & \text{αν } R_1(\vec{x}) \text{ ισχύει} \\ G_2(\vec{x}) & \text{αν } R_2(\vec{x}) \text{ "} \\ \vdots & \vdots \\ G_R(\vec{x}) & \text{αν } R_R(\vec{x}) \text{ "} \end{cases}$$

είναι αναδρομική.

Απόδειξη. Μπορούμε να ορίσουμε σαφώς την F ως εξής:

$$F(\vec{x}) = G_1(\vec{x})C_{\neg R_1}(\vec{x}) + G_2(\vec{x})C_{\neg R_2}(\vec{x}) + \dots + G_R(\vec{x})C_{\neg R_R}(\vec{x}) \quad (\text{προφανώς αναδρομική})$$

Σημ: Έτσι καλύπτονται οι περιπτώσεις της μορφής

$$F(\vec{x}) = \begin{cases} G_1(\vec{x}) & \text{αν } R(\vec{x}) \\ G_2(\vec{x}) & \text{αλλιώς} \\ & (\text{δηλ. } \neg R) \end{cases}$$

□

Ανάλογα (αντί για συναρτήσεις $G_1, \dots, G_R$ έχω τώρα κατηγορήματα $P_1, \dots, P_R$),

C 10. Έστω $P_1, \dots, P_R$ και $R_1, \dots, R_R$ να είναι αναδρομικά τ.ω. για κάθε $\vec{x}$ ένα και μόνο ένα από τα $R_1(\vec{x}), \dots, R_R(\vec{x})$ να ισχύει. Τότε αν το Q δίνεται από τον τύπο:

$$Q(\vec{x}) \longleftrightarrow \begin{cases} P_1(\vec{x}) & \text{αν } R_1(\vec{x}) \\ \vdots & \vdots \\ P_R(\vec{x}) & \text{αν } R_R(\vec{x}) \end{cases}$$

τότε το Q είναι αναδρομική.

Απόδειξη. Απλά θεωρήστε το C_Q σε σχέση με τα $C_{P_1}, \dots, C_{P_R}$.

□

†

Λήμμα 1.9. Η συνάρτηση Ζεύγους, $Pair: \omega^2 \rightarrow \omega$ που ορίζεται από τον τύπο $Pair(x, y) = (x + y)(x + y) + x + 1$ είναι αναδρομική και "1-1".

Λήμμα 1.10. (β-συνάρτηση του Gödel)

Υπάρχει μια δυαδική αναδρομική συνάρτηση β τ.ω.

(i) $\beta(x, y) \leq x \div 1$

(ii) Για οποιοδήποτε n και για όλους τους φυσικούς αριθμούς $a_0, \dots, a_{n-1}$ υπάρχει ένας αριθμός a , τ.ω. $\beta(a, i) = a_i$ για κάθε $i < n$.

Υποθέτουμε την ύπαρξη μιας σταθεροποιημένης [fixed] β .

$$\text{Σημείωση: } \begin{cases} \beta(0, y) = 0 & \forall y \\ \beta(x, y) < x & \text{για } x > 0 \end{cases}$$

Χρησιμοποιώντας την β μπορούμε να ορίσουμε για κάθε σταθεροποιημένο αριθμό n μια αναδρομική "1-1" συνάρτηση: $\omega^n \longrightarrow \omega$.

Ορισμός 1.11. Για κάθε n , η συνάρτηση $\prec \cdots \succ : \omega^n \longrightarrow \omega$ ορίζεται ως:

$$\prec y_0 \cdots y_{n-1} \succ = \mu x [\beta(x, 0) = n \wedge \beta(x, 1) = y_0 \wedge \beta(x, 2) = y_1 \wedge \cdots \wedge \beta(x, n) = y_{n-1}]$$

Το $\prec y_0 \cdots y_{n-1} \succ$ καλείται **αριθμός ακολουθίας** του $\langle y_0 \cdots y_{n-1} \rangle$

Λήμμα 1.12. Οι παρακάτω συναρτήσεις είναι αναδρομικές.

- (i) Το "μήκος του x ", $lh(x) = \beta(x, 0)$.
- (ii) Το " $(i+1)$ οστό στοιχείο του x ", $(x)_i$ που δίνεται από τον τύπο $(x)_i = \beta(x, i+1)$.
- (iii) Το κατηγορημα $Seq(x)$, δηλαδή το " x είναι ένας αριθμός ακολουθίας" που δίνεται από τον τύπο $Seq(x) \leftrightarrow \forall y_{y < x} (lh(y) = lh(x) \longrightarrow \exists i_{i < lh(x)} ((y)_i \neq (x)_i))$.
- (iv) Η δυαδική συνάρτηση $Red(x, i)$ που ορίζεται ως:

$$Red(x, i) = \mu y (lh(y) = i \wedge \forall j_{j < i} ((y)_j = (x)_j)).$$

Η χαρακτηριστική ιδιότητά της είναι:

$$Red(\prec y_0 \cdots y_{n-1} \succ, i) = \prec y_0 \cdots y_{i-1} \succ \quad (\text{όπου } i \leq n)$$

Απόδειξη. Τα (i),(ii),(iii) είναι όλα αναδρομικά αφού η β αναδρομική και ο ορισμός του (iii) περιλαμβάνει μόνο φραγμένους τελεστές.

Από ιδιότητες της β -συνάρτησης, η Red είναι επίσης αναδρομική. □

Σημ: $Seq(x) \wedge lh(x) = n \longrightarrow x = \prec (x)_0 \cdots (x)_{n-1} \succ$

Τώρα για κάθε συνάρτηση $F(y, \vec{x})$ ορίζουμε μια συνάρτηση $\bar{F}(y, \vec{x})$ τέτοια ώστε:

$$\bar{F}(y, \vec{x}) = \prec F(0, \vec{x}), F(1, \vec{x}), \dots, F(y-1, \vec{x}) \succ,$$

επαγωγικά, ως εξής:

$$\begin{cases} \bar{F}(0, \vec{x}) &= \prec \emptyset \succ = 0 \\ \bar{F}(y, \vec{x}) &= \prec F(0, \vec{x}), \dots, F(y-1, \vec{x}) \succ \end{cases}$$

Σημ: Ο αριθμός ακολουθίας της κενής ακολουθίας είναι ίσος με μηδέν, διότι $\prec \emptyset \succ = \mu x (lh(x) = 0)$, αφού $\beta(0, 0) = 0$.

Λήμμα 1.13. Η $F(y, \vec{x})$ είναι αναδρομική αν και μόνο αν η $\bar{F}$ είναι αναδρομική.

Απόδειξη. " $\Rightarrow$ " Η $\bar{F}$ έχει σαφή ορισμό, τον ακόλουθο:

$$\bar{F}(y, \vec{x}) = \mu z (\text{lh}(z) = y \wedge \forall i < y ((z)_i = F(i, \vec{x})))$$

Έτσι, είναι αναδρομική αν είναι και η F .

" $\Leftarrow$ " Η F έχει σαφή ορισμό:

$$F(y, \vec{x}) = (\bar{F}(y + 1, \vec{x}))_y,$$

και άρα είναι αναδρομική αν είναι η $\bar{F}$. □

Θεώρημα 1.14. (Γενική αναδρομή)

Αν η G είναι αναδρομική και η F ορίζεται ως:

$$F(y, \vec{x}) = G(\bar{F}(y, \vec{x}), y, \vec{x}),$$

τότε η F είναι αναδρομική.

Απόδειξη. Μπορούμε να γράψουμε ένα σαφή ορισμό για την F :

$$F(y, \vec{x}) = \mu z [\text{Seq}(z) \wedge \text{lh}(z) = y \wedge \forall i < y ((z)_i = G(\text{Red}(z, i), i, \vec{x}))].$$

Ασχέτως αν είναι σωστός ο ορισμός είναι σίγουρα αναδρομικός. Για να δούμε ότι ορίζει σωστά την F , υποθέτουμε ότι:

$$\bar{F}(y, \vec{x}) = \prec F(0, \vec{x}), \dots, F(y - 1, \vec{x}) \succ = z.$$

Τότε:

$$\begin{aligned} (z)_i &= F(i, \vec{x}) = G(\bar{F}(i, \vec{x}), i, \vec{x}) \\ &= G(\prec F(0, \vec{x}) \cdots F(i - 1, \vec{x}) \succ, i, \vec{x}) \\ &= G(\text{Red}(z, i), i, \vec{x}). \end{aligned}$$

Αυτό δικαιολογεί τον ορισμό της F . Άρα το θεώρημα αποδείχτηκε. □

Πόρισμα 1.15. Το R3 ισχύει για την κλάση $\Sigma = \text{Αναδ.}$, δηλαδή, αν G, H αναδρ. και η F δίνεται από τον επαγωγικό ορισμό:

$$\begin{aligned} F(0, \vec{x}) &= G(\vec{x}) \\ F(y + 1, \vec{x}) &= H(F(y, \vec{x}), y + 1, \vec{x}), \end{aligned}$$

τότε η F είναι αναδρομική (ή το ίδιο αν αντί για $y + 1$ βάλουμε το y).

Απόδειξη. Έχουμε στην ουσία μια ειδική περίπτωση του θεωρήματος 1.14. Μπορούμε να γράψουμε:

$$F(y, \vec{x}) = \begin{cases} G(\vec{x}) & \text{αν } y = 0 \\ H((\bar{F}(y, \vec{x}))_{y-1}, y, \vec{x}) & \text{αλλιώς.} \end{cases}$$

που είναι αναδρομική της $(\bar{F}(y, \vec{x}), y, \vec{x})$, και που άρα ανάγει τον ορισμό στο Θ.1.14. □

†

Για να δείξουμε τώρα ότι η β -συνάρτηση υπάρχει (μέχρι τώρα την έχουμε μόνο υποθέσει), δείχνουμε στη θέση της την ύπαρξη συνάρτησης $\delta(x, y, z)$ με παρόμοιες ιδιότητες:

- (i) $\delta(x, y, z) \leq x$
- (ii) Για κάθε $a_0, \dots, a_{n-1}$, υπάρχουν αριθμοί b, c τ.ω. $\delta(b, c, i) = a_i \quad \forall i < n$.

Για να δείξουμε ότι η ύπαρξη της δ συνεπάγεται την ύπαρξη της β , χρησιμοποιούμε τη **συνάρτηση ζεύγους**.

Ισχυρισμός. Αν η δ υπάρχει, υπάρχει και η β -συνάρτηση

Από Λήμμα 1.9 έχω την: $\text{Pair}: \omega^2 \rightarrow \omega$. Σημειώστε ότι:

$$\text{Pair}(x, y) > \frac{x}{y} \quad (\text{και από τα δύο})$$

Απόδειξη. Ορίζω συνάρτηση l (για το αριστερό) και r (για το δεξί). Την l ως:

$$l(x) = \mu y_{y < x} \exists z_{z < x} (x = \text{Pair}(y, z))$$

που είναι αναδρομικός ορισμός της αριστερής συντεταγμένης του "αριθμού ζεύγους" x (x είναι η ακτίνα δράσης της Pair) και κάτι άλλο, αλλιώς. Όμοια ορίζουμε την r ως:

$$r(x) = \mu y_{y < x} \exists z_{z < x} (x = \text{Pair}(z, y)).$$

Οι l, r είναι αναδρομικές.

Μπορούμε τώρα να ορίσουμε τη β από τη δ ως εξής:

$$\beta(x, i) = \begin{cases} \delta(l(x), r(x), i) & \text{αν } x = \text{Pair}(l(x), r(x)) \\ 0 & \text{αλλιώς.} \end{cases}$$

Αφού η δ είναι αναδρομική, είναι και η β . Επίσης,

$$\beta(x, i) = \delta(l(x), r(x), i) \leq l(x) < x,$$

αν το x είναι αριθμός-ζεύγος,

άρα το ότι: $\beta(x, i) \leq x \dot{-} 1$, ισχύει πάντα.

Επιπλέον η β ικανοποιεί την κύρια συνθήκη (ii) αφού αν b, c έχουν βρεθεί τέτοιες ώστε $\delta(b, c, i) = a_i \quad \forall i < n$, τότε το $a = \text{Pair}(b, c)$ έχει την απαιτούμενη ιδιότητα, δηλαδή $\beta(a, i) = a_i \quad \forall i < n$.

□

Για να δείξουμε ότι υπάρχει η δ .

Λήμμα (a) - Κινέζικο Θεώρημα Υπολοίπων. Έστω οι $d_0, \dots, d_{n-1}$ ανά δύο πρώτοι, και $a_0, \dots, a_{n-1}$ να είναι τέτοιοι ώστε: $a_i < d_i \quad \forall i < n$. Τότε υπάρχει ένα b τέτοιο ώστε $a_i = b \bmod d_i$ για κάθε $i < n$.

Απόδειξη. Έστω $q = \prod_{i < n} d_i$. Για κάθε x έστω ότι το $[x]$ δηλώνει την n -άδα:

$$\langle x \bmod d_0, \dots, x \bmod d_{n-1} \rangle.$$

Τότε υπάρχουν το πολύ q πιθανές τιμές για το $[x]$.

Ισχυρισμός.

$$x, y < q \text{ και } x \neq y \Rightarrow [x] \neq [y].$$

Ισχύει, γιατί αν $[x] = [y]$ τότε το d_i διαιρεί το $|x - y|$ για κάθε $i < n$ που είναι αδύνατο. Άρα το $[x]$ παίρνει όλες τις πιθανές τιμές καθώς το x "τρέχει" πάνω στα $0, 1, \dots, q - 1$. Άρα για κάθε συγκεκριμένα $a_0, \dots, a_{n-1}$ (με $a_i < d_i$) υπάρχει ένα b τέτοιο ώστε $[b] = \langle a_0, \dots, a_{n-1} \rangle$. □

Λήμμα (b). Για κάθε n , οι $n + 1$ αριθμοί:

$$1 + n!, 1 + 2(n!), 1 + 3(n!), \dots, 1 + (n + 1)n!$$

είναι ανά δύο πρώτοι.

Απόδειξη. Υποθέτουμε ότι για κάποια $i, j \in \{1, \dots, (n + 1)\}$ ισχύει ότι οι $1 + i(n!)$ και $1 + j(n!)$ δεν είναι ανά δύο πρώτοι, αλλά έχουν κοινό παράγοντα p . Τότε το p διαιρεί το $|i - j|n!$. Άρα το p διαιρεί το $|i - j|$, αφού το p δε διαιρεί το $n!$. Αυτό όμως είναι δυνατό μόνο αν $i = j$, αφού $|i - j| \leq n < p$ (σημ: $n < p$ διότι $p \nmid n!$) □

Λήμμα (c). Η συνάρτηση δ είναι αναδρομική, όπου $\delta(x, y, z) = x \bmod (1 + (z + 1)y)$ και έχει τις απαιτούμενες ιδιότητες (i) και (ii).

Απόδειξη. Η δ έχει σαφή ορισμό, τον:

$$\delta(x, y, z) = \mu\omega(\exists t_{t < x+1} (x = t(1 + (z + 1)y) + \omega)),$$

άρα είναι αναδρομική.

Δοθέντων $a_0, \dots, a_{n-1}$, έστω n ο μεγαλύτερος από τους αριθμούς $\{a_0, \dots, a_{n-1}, n\}$. Έστω $c = n!$. Τότε από Λήμμα (b) οι αριθμοί $1 + (i + 1) \cdot c$ είναι ανά 2 πρώτοι για $i < n$ ($\leq n$). Τότε από Λήμμα (a) (όπου $d_i = 1 + (i + 1)c$), υπάρχει ένας αριθμός b τ.ώ. $a_i = b \bmod (1 + (i + 1)c) \quad \forall i < n$, δηλαδή $\delta(b, c, i) = a_i \quad \forall i < n$. □

Παραδείγματα εφαρμογών της γενικής αναδρομής:

- (1) Η ακολουθία Fibonacci που δίνεται από:

$$\begin{aligned} u_0 &= u_1 = 1 \\ u_{n+2} &= u_n + u_{n+1}. \end{aligned}$$

Είναι αναδρομική αφού δίνεται από τον τύπο:

$$F(x) = \begin{cases} 1 & \text{αν } x = 0 \vee x = 1 \\ (\overline{F}(x))_x \dot{-} 1 + (\overline{F}(x))_x \dot{-} 2 & \text{αλλιώς,} \end{cases}$$

που είναι αναδρομική συνάρτηση της $\overline{F}(x)$.

- (2) Μπορούμε επίσης να εφαρμόσουμε το 1.14 σε κατηγορήματα. Για παράδειγμα, υποθέτουμε ότι $\mathcal{E}(\overline{C}_p(y, \vec{x}))$ είναι κάποιος σαφής ορισμός ενός κατηγορήματος που περιλαμβάνει τη $\overline{C}_p$ και πιθανώς άλλες συναρτήσεις και κατηγορήματα, τα οποία είναι αναδρομικά. Τότε, αν το P δίνεται ως:

$$P(y, \vec{x}) \Leftrightarrow \mathcal{E}(\overline{C}_p(y, \vec{x})),$$

το P είναι αναδρομικό.

Απόδειξη. Η χαρακτηριστική του συνάρτηση μπορεί να γραφεί ως εξής:

$$C_p(y, \vec{x}) = \begin{cases} 0 & \text{αν } \mathcal{E}(\overline{C}_p(y, \vec{x})) \\ 1 & \text{αλλιώς.} \end{cases}$$

□

Για παράδειγμα, μπορούμε να ορίσουμε το P με ορισμό του είδους:

$$P(y, \vec{x}) \Leftrightarrow \begin{cases} R_1(\vec{x}) & \text{αν } y = 0 \\ R_2(\vec{x}) & \text{αν } y \neq 0 \wedge P(y-1, \vec{x}) \\ R_3(\vec{x}) & \text{αλλιώς,} \end{cases}$$

όπου R_1, R_2, R_3 δοθέντα αναδρομικά κατηγορήματα.

Σημειώστε ότι:

$$P(y-1, \vec{x}) \Leftrightarrow (\overline{C}_p(y, \vec{x}))_y \dot{-} 1 = 0$$

- (3) Μια άλλη μορφή του 1.14 είναι αυτή που η G δίνεται από περιπτώσεις, δηλαδή:

$$F(x, y) = \begin{cases} F(H_1(x), y) & \text{αν } H_1(x) < x \\ H_2(x, y) & \text{αλλιώς,} \end{cases}$$

όπου H_1, H_2 αναδρομικές. (Μπορούμε να αντικαταστήσουμε την πρώτη γραμμή από το $(\overline{F}(y, x))_{H_1(x)}$ αν $H_1(x) < x$)

Ο γενικός κανόνας είναι ότι ένας ορισμός της μορφής $F(y, \vec{x}) = \text{"ένας αναδρομικός ορισμός που περιλαμβάνει την } F\text{"}$, είναι Ο.Κ (δηλαδή αναδρομική) υπό την προϋπόθεση ότι όποτε απαντάται το $F(w, \vec{x})$ στο σχήμα της αναδρομής να μπορούμε να αποδείξουμε ότι $w < y$.

2. ΠΡΩΤΟΒΑΘΜΙΑ ΑΡΙΘΜΗΤΙΚΗ

Η γλώσσα της πρωτοβάθμιας αριθμητικής περιλαμβάνει:

Δύο δυαδικά συναρτησιακά σύμβολα, $(+, \cdot)$

Ένα εναδικό συναρτησιακό σύμβολο, S

Ένα σύμβολο σταθεράς, $\bar{0}$

Ένα δυαδικό σύμβολο κατηγορήματος, $<$

$\mathcal{L}$: Η γλώσσα της πρωτοβάθμιας λογικής των κατηγορημάτων μαζί με το " $=$ ".

Μια δομή $\mathcal{A}$ για την $\mathcal{L}$ είναι της μορφής:

$$\langle |A|, <^A, S^A, +^A, \cdot^A, \bar{0}^A \rangle.$$

Η κανονικοποιημένη δομή που μας ενδιαφέρει είναι η δομή των φυσικών αριθμών:

$$\mathbb{N} = \langle \omega, <, \sigma, +, \cdot, \bar{0} \rangle \quad (\text{Στάνταρ μοντέλο})$$

Δηλώνουμε τα:

$$\begin{array}{ccccccc} S(\bar{0}) & SS(\bar{0}) & SSS(\bar{0}) & \dots \\ \text{σαν} & \bar{1} & \bar{2} & \bar{3} & \dots \bar{n} & \leftarrow \text{αριθμητικά} \end{array}$$

Σύστημα μη-λογικών αξιωμάτων S_0 για την αριθμητική (αριθμητική του Robinson):

- (S1) $S(x) \neq \bar{0}$
- (S2) $S(x) = S(y) \rightarrow x = y$
- (S3) $x + \bar{0} = x$
- (S4) $x + s(y) = s(x + y)$
- (S5) $x \cdot \bar{0} = \bar{0}$
- (S6) $x \cdot s(y) = s(x + y)$
- (S7) $\neg(x < \bar{0})$
- (S8) $x < s(y) \leftrightarrow (x < y \vee x = y)$
- (S9) $x < y \vee x = y \vee y < x$

Τα (S1) – (S9) σχηματίζουν ένα πεπερασμένο αξιωματικό σύστημα.

Αν προσθέσουμε το ακόλουθο **άπειρο** αξίωμα-σχήμα:

(S10) **Επαγωγικό σχήμα :**

Για κάθε τύπο ϕ με ελεύθερη μεταβλητή x ισχύει:

$$[\Phi(\bar{0}) \wedge \forall x(\Phi(x) \rightarrow \Phi(s(x)))] \rightarrow \forall x\Phi(x) \quad [\text{μαθηματική επαγωγή}]$$

Καλούμε το σύστημα $S1 - S10$, S ή **Πρωτοβάθμια αριθμητική του Peano**.

Προφανώς η δομή $\mathbb{N}$ είναι μοντέλο τόσο για το S_0 όσο και για το S .

Μερικά τετριμμένα αποτελέσματα για τη θεωρία S_0 :

Λήμμα 2.1. *Ισχύουν τα ακόλουθα:*

$$\vdash_{S_0} \overline{m} = \overline{n} \text{ αν } m = n$$

$$\text{και } \vdash_{S_0} \overline{m} \neq \overline{n} \text{ αν } m \neq n$$

Απόδειξη. Το πρώτο μέρος τετριμμένο από αξίωμα ισότητας. Το δεύτερο μέρος αποδεικνύεται με επαγωγή στο n . Αρχεί να δειχθεί για όλα τα $m > n$. Για $n = 0$ ισχύει από $S1$. Έστω ότι ισχύει για δοθέν n (και όλα τα $m > n$). Τότε από $S2$, έχω:

$$\vdash_{S_0} (\overline{m+1}) = (\overline{n+1}) \rightarrow \overline{m} = \overline{n},$$

αλλά από υπόθεση έχω ότι:

$$\vdash_{S_0} \overline{m} \neq \overline{n}.$$

Άρα

$$\vdash_{S_0} (\overline{m+1}) \neq (\overline{n+1}).$$

Άρα η πρόταση ισχύει και για $n+1$ αφού οτιδήποτε μεγαλύτερο του $n+1$ είναι της μορφής $m+1$ για κάποιο $m > n$.

□

Λήμμα 2.2. *(Γενικευμένη Επαγωγή) Για κάθε τύπο $\Phi(x)$ και κάθε $n \in \omega$,*

$$\vdash_{S_0} \Phi(\overline{0}) \wedge \Phi(\overline{1}) \cdots \wedge \Phi(\overline{n-1}) \wedge (x < \overline{n}) \rightarrow \Phi(x),$$

Απόδειξη. Με επαγωγή στο n .

Για $n = 0$: Πρέπει να δείξουμε ότι $\vdash_{S_0} x < \overline{0} \rightarrow \Phi(x)$

Αλλά από $S7$: $\vdash_{S_0} \neg(x < \overline{0})$.

Άρα το αποτέλεσμα έπεται.

Έστω ότι ισχύει για κάποιο n . Για το $n+1$ έχω από $S8$:

$$\vdash_{S_0} x < \overline{n+1} \leftrightarrow (x < \overline{n} \vee x = \overline{n}).$$

Από αυτό παίρνουμε:

$$\vdash_{S_0} \Phi(\overline{0}) \wedge \Phi(\overline{1}) \cdots \wedge \Phi(\overline{n-1}) \wedge \Phi(\overline{n}) \wedge (x < \overline{n+1}) \rightarrow (x < \overline{n} \wedge x = \overline{n}).$$

Αφού αμφοτέρωτα

$$\vdash_{S_0} \Phi(\overline{0}) \wedge \cdots \wedge \Phi(\overline{n}) \wedge (x < \overline{n}) \rightarrow \Phi(x)$$

$$\vdash_{S_0} \Phi(\overline{0}) \wedge \cdots \wedge \Phi(\overline{n}) \wedge (x = \overline{n}) \rightarrow \Phi(x)$$

ισχύουν, έπεται ότι:

$$\vdash_{S_0} \Phi(\overline{0}) \wedge \cdots \wedge \Phi(\overline{n}) \wedge (x < \overline{n+1}) \rightarrow \Phi(x).$$

□

Πόρισμα. Για κάθε $n \in \omega$ ισχύει:

$$\vdash_{S_0} x < \bar{n} \rightarrow (x = \bar{0} \vee x = \bar{1} \vee \dots \vee x = \overline{n-1}).$$

Απόδειξη. Το ζητούμενο προκύπτει άμεσα, βάζοντας στη θέση του $\Phi(x)$ του λήμματος 3.2, το $(x = \bar{0} \vee x = \bar{1} \vee \dots \vee x = \overline{n-1})$. $\square$

Λήμμα 2.3. Υποθέτουμε ότι $\vdash_{S_0} \neg\Phi(\bar{i})$ για κάθε $i < n$, και ότι $\vdash_{S_0} \Phi(\bar{n})$. Τότε:

$$\vdash_{S_0} \Phi(x) \wedge \forall y (y < x \rightarrow \neg\Phi(y)) \leftrightarrow x = \bar{n}.$$

Απόδειξη. Δείχνουμε άτυπα ότι το απαιτούμενο γεγονός ισχύει για κάθε μοντέλο του S_0 , και μετά εφαρμόζουμε το θεώρημα πληρότητας. Το γεγονός ότι:

$$\mathcal{A} \models x = \bar{n} \rightarrow \Phi(x) \wedge \forall y (y < x \rightarrow \neg\Phi(y))$$

βγαίνει από υπόθεση και το πιο πάνω πόρισμα. Αντιστρόφως, υποθέτουμε ότι για κάποια $a \in |\mathcal{A}|$:

$$\mathcal{A} \models \Phi[a] \wedge \forall y (y < a \rightarrow \neg\Phi(y)) \quad (*)$$

Τότε, αν $\mathcal{A} \models a \neq \bar{n}$, από αξίωμα S9 έχουμε είτε ότι $\mathcal{A} \models a < \bar{n}$ ή $\mathcal{A} \models \bar{n} < a$. Αλλά και οι δύο αυτές περιπτώσεις είναι αδύνατες. Η πρώτη δίνει $\mathcal{A} \models a = \bar{i}$ για κάποιο $i < n$ και άρα $\mathcal{A} \models \neg\Phi[a]$, το οποίο αντιφάσκει με τη σχέση (*), και η δεύτερη μαζί με την (*) δίνει $\mathcal{A} \models \neg\Phi(\bar{n})$, που είναι άτοπο εξ υποθέσεως. Άρα $\mathcal{A} \models a = \bar{n}$. Άρα ισχύει:

$$\mathcal{A} \models \Phi(x) \wedge \forall y (y < x \rightarrow \neg\Phi(y)) \rightarrow x = \bar{n}.$$

Έτσι από θεώρημα πληρότητας έχω το ζητούμενο:

$$\vdash_{S_0} \Phi(x) \wedge \forall y (y < x \rightarrow \neg\Phi(y)) \leftrightarrow x = \bar{n}.$$

$\square$

†

Ορισμός 2.4. Αν T είναι μια θεωρία στη γλώσσα της αριθμητικής, ένα κατηγορημα $R \subseteq \omega^n$, θα λέμε ότι είναι **αναπαραστάσιμο στην T** αν υπάρχει κάποιος τύπος $\Phi(x_1, \dots, x_n)$ τέτοιος ώστε:

$$R(a_1 \dots a_n) \Rightarrow \vdash_T \Phi(\bar{a}_1, \dots, \bar{a}_n)$$

$$\neg R(a_1 \dots a_n) \Rightarrow \vdash_T \neg\Phi(\bar{a}_1, \dots, \bar{a}_n),$$

όπου τα βέλη μπορούν να αντιστραφούν αν η T είναι συνεπής.

Όμοια, λέμε ότι μια συνάρτηση $F : \omega^n \rightarrow \omega$ είναι **αναπαραστάσιμη στην T** αν υπάρχει ένας τύπος $\Phi(x_1 \dots x_n, y)$ τέτοιος ώστε:

$$F(a_1 \dots a_n) = b \Rightarrow \vdash_T \Phi(\bar{a}_1 \dots \bar{a}_n, y) \leftrightarrow y = \bar{b}.$$

Λήμμα 2.5. Αν T είναι οποιαδήποτε επέκταση της S_0 , και $R \subseteq \omega^n$, τότε το R είναι αναπαραστάσιμο στην T αν και μόνον αν η χαρακτηριστική του συνάρτηση C_R είναι αναπαραστάσιμη στην T .

Απόδειξη. " $\Rightarrow$ " Έστω ότι ο $\Phi(x_1, \dots, x_n)$ αναπαριστά το R στην T . Τότε έστω ο $\Psi(x_1, \dots, x_n)$ να είναι ο $(\Phi(\bar{x}) \wedge y = \bar{0}) \vee (\neg \Phi(\bar{x}) \wedge y = \bar{1})$.

Ισχυρισμός: Ο Ψ αναπαριστά την C_R στην T .

Ισχύει, διότι αν $C_R(a_1, \dots, a_n) = 0$ τότε $R(a_1, \dots, a_n)$ και άρα $\vdash_T \Phi(\bar{a}_1, \dots, \bar{a}_n)$, όπου $\vdash_T \Psi(\bar{a}_1, \dots, \bar{a}_n, y) \leftrightarrow y = \bar{0}$. Όμοια για την περίπτωση $C_R(a_1, \dots, a_n) = 1$.

" $\Leftarrow$ " Αντίστροφα, υποθέστε ότι ο $\Psi(x_1, \dots, x_n, y)$ αναπαριστά την C_R στην T . Έστω ο $\Phi(x_1, \dots, x_n)$ να είναι ο $\Psi(x_1, \dots, x_n, \bar{0})$. Τότε

$$\begin{aligned} R(a_1, \dots, a_n) &\Rightarrow C_R(a_1, \dots, a_n) = 0 \\ &\Rightarrow \vdash_T \Psi(\bar{a}_1, \dots, \bar{a}_n, y) \leftrightarrow y = \bar{0} \\ &\Rightarrow \vdash_T \Psi(\bar{a}_1, \dots, \bar{a}_n, \bar{0}). \end{aligned}$$

$$\begin{aligned} \neg R(a_1, \dots, a_n) &\Rightarrow C_R(a_1, \dots, a_n) = 1 \\ &\Rightarrow \vdash_T \Psi(\bar{a}_1, \dots, \bar{a}_n, y) \leftrightarrow y = T \\ &\Rightarrow \vdash_T \neg \Psi(\bar{a}_1, \dots, \bar{a}_n, \bar{0}), \text{ αφού } \vdash_T \bar{0} \neq T \end{aligned}$$

Άρα ο Φ αναπαριστά το R .

□

Έστω η F αναπαραστάσιμη στην S_0 από τον τύπο $\Phi(x_1, \dots, x_n, y)$. Για να υπολογίσουμε το $F(a_1, \dots, a_n)$, το μόνο που χρειάζεται να κάνουμε είναι να απαριθμήσουμε όλα τα θεωρήματα της S_0 μέχρι να φτάσουμε σε αυτό της μορφής

$$\vdash_{S_0} \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{b}), \text{ για κάποιο } b \in \omega.$$

Τότε σταματάμε και ξέρουμε ότι ισχύει $F(a_1, \dots, a_n) = \bar{b}$.

Λήμμα 2.6. Οι βασικές συναρτήσεις είναι όλες αναπαραστάσιμες στην S_0 .

Απόδειξη. (i) Η μηδενική συνάρτηση $Z(x)$ αναπαριστάται ως $x = x \wedge y = \bar{0}$.

(ii) Από Λήμμα 2.5 για να δείξουμε ότι η συνάρτηση $C_<$ είναι αναπαραστάσιμη αρκεί να δείξουμε ότι το $<$ είναι αναπαραστάσιμο.

Τώρα έστω $\Phi(x_1, x_2)$ να είναι το $x_1 < x_2$.

Υποθέτουμε ότι ο Φ αναπαριστά το $<$.

Πρέπει να δείξουμε ότι:

$$\left. \begin{aligned} a_1 < a_2 &\Rightarrow \vdash_{S_0} \bar{a}_1 < \bar{a}_2 \\ a_1 \geq a_2 &\Rightarrow \vdash_{S_0} \neg(\bar{a}_1 < \bar{a}_2) \end{aligned} \right\} \text{ για όλα τα } a_1, a_2 \in \omega.$$

Το κάνουμε με επαγωγή στο a_2 :

Για $a_2 = 0$, η πρώτη περίπτωση δεν ισχύει και η δεύτερη αληθής από $S7$. Τώρα, από πόρισμα του 2.2, αν $a_2 > 0$ ισχύει:

$$\vdash_{S_0} x < \bar{a}_2 \leftrightarrow (x = \bar{0} \vee x = \bar{1} \vee \cdots \vee x = \overline{a_2 - 1}).$$

Άρα, αν $a_1 < a_2$ έχω $\vdash_{S_0} \bar{a}_1 < \bar{a}_2$.

Αν $a_2 \leq a_1$ τότε

$$\vdash_{S_0} \neg(\bar{a}_1 = \bar{0} \vee \bar{a}_1 = \bar{1} \vee \cdots \vee \bar{a}_1 = \bar{a}_2 - 1) \quad (\text{αφού } \vdash_{S_0} \neg(\bar{m} = \bar{n}) \text{ για } m \neq n)$$

και άρα $\vdash_{S_0} \neg(\bar{a}_1 < \bar{a}_2)$.

(iii) Η συνάρτηση του επόμενου σ αναπαρίσταται ως $S(x) = y$.

Για κάθε $a_1, a_2 \in \omega$

$$\sigma(a_1) = a_2 \Rightarrow \vdash_{S_0} (S(\bar{a}_1) = y) \leftrightarrow y = \bar{a}_2,$$

αφού $S(\bar{a}_1) = \bar{a}_2$ από ορισμό.

(iv) Η συνάρτηση προβολής I_i^n αναπαρίσταται τετριμμένα από το

$$x_1 = x_1 \wedge x_2 = x_2 \wedge \cdots \wedge x_n = x_n \wedge y = x_i.$$

(v) Πρόσθεση: Ισχυριζόμαστε ότι η αναπαρίσταται από τον τύπο

$$\Phi(x_1, x_2, y) : x_1 + x_2 = y.$$

Για να αποδείξουμε ότι $\forall a_1, a_2 \in \omega \quad \vdash_{S_0} (\bar{a}_1 + \bar{a}_2 = y) \leftrightarrow y = \overline{(a_1 + a_2)}$, ή ισοδύναμα ότι:

$$\vdash_{S_0} \bar{a}_1 + \bar{a}_2 = \overline{(a_1 + a_2)} \quad (*)$$

Κάνουμε επαγωγή στο a_2 :

Αν $a_2 = 0$ τότε η (*) είναι αληθής από $S3$.

Υποθέτουμε ότι είναι αληθής για $a_2 = n$.

Τότε ισχύει:

$$\vdash_{S_0} \bar{a}_1 + \bar{n} = \overline{(a_1 + n)}.$$

Οπότε,

$$\vdash_{S_0} s(\bar{a}_1 + \bar{n}) = s(\overline{(a_1 + n)}),$$

δηλαδή

$$\vdash_{S_0} s(\bar{a}_1 + \bar{n}) = \overline{(a_1 + (n + 1))}.$$

Αλλά από $S4$: $\vdash_{S_0} s(\bar{a}_1 + \bar{n}) = \bar{a}_1 + s(\bar{n}) = \bar{a}_1 + \overline{n + 1}$.

Άρα $\vdash_{S_0} \bar{a}_1 + \overline{n + 1} = \overline{a_1 + (n + 1)}$, και το επαγωγικό βήμα αποδείχτηκε.

(vi) Πολλαπλασιασμός - ακριβώς όμοια με την πρόσθεση.

□

Θεώρημα 2.7. Κάθε αναδρομική συνάρτηση ή κατηγορημα είναι αναπαριστάσιμη/ο στο S_0 .

Απόδειξη. Χρειάζεται να θεωρήσουμε μόνο συναρτήσεις. Από Λήμμα 2.6 όλες οι βασικές συναρτήσεις είναι αναπαραστάσιμες. Άρα πρέπει να δείξουμε ότι οι αναπαραστάσιμες συναρτήσεις είναι κλειστές υπό τα $R2, R4$.

Για το $R2$: Υποθέτουμε $H_1, \dots, H_R, G$ όλες αναπαραστάσιμες συναρτήσεις και $F(\vec{x}) = G(H_1(\vec{x}) \dots H_R(\vec{x}))$. Έστω ότι ο $\Phi_i(\vec{x}, y_i)$ αναπαριστά την H_i για κάθε $i \in \{1, \dots, R\}$. Έστω ότι ο $\Psi(y_1, \dots, y_R, z)$ αναπαριστά την G . Τώρα ορίζουμε τον τύπο Θ ως εξής:

$$\Theta(\vec{x}, z) : \exists y_1 \dots y_R [\Phi_1(\vec{x}, y_1) \wedge \dots \wedge \Phi_R(\vec{x}, y_R) \wedge \Psi(y_1 \dots y_R, z)].$$

Για να δείξουμε ότι ο Θ αναπαριστά την F :

Υποθέτουμε $a_1, \dots, a_n \in \omega$, και έστω $F(a_1 \dots a_n) = c$ και $H_i(a_1 \dots a_n) = b_i$ για κάθε $i \in \{1, \dots, R\}$. [Τότε $G(b_1, \dots, b_R) = c$]. Τότε από τον ορισμό των Φ_i έχουμε:

$$\vdash_{S_0} \Phi_i(\bar{a}_1, \dots, \bar{a}_n, y_i) \leftrightarrow y_i = \bar{b}_i$$

Έπεται ότι:

$$\vdash_{S_0} \Theta(\bar{a}_1, \dots, \bar{a}_n, z) \leftrightarrow \exists y_1 \dots y_R (y_1 = \bar{b}_1 \wedge \dots \wedge y_R = \bar{b}_R \wedge \Psi(y_1 \dots y_R, z))$$

και άρα,

$$\vdash_{S_0} \Theta(\bar{a}_1, \dots, \bar{a}_n, z) \leftrightarrow \Psi(\bar{b}_1 \dots \bar{b}_R, z).$$

Αλλά αφού ο Ψ αναπαριστά την G και $G(b_1, \dots, b_R) = c$, έχουμε:

$$\vdash_{S_0} \Psi(\bar{b}_1, \dots, \bar{b}_R, z) \leftrightarrow z = \bar{c},$$

δηλαδή,

$$\vdash_{S_0} \Theta((\bar{a}_1, \dots, \bar{a}_n, z) \leftrightarrow z = \bar{c}.$$

Για το $R4$:

Υποθέτουμε ότι η $G(x_1 \dots x_{n+1})$ είναι αναπαραστάσιμη και $F(x_1 \dots x_n) = \mu x_{n+1} (G(x_1 \dots x_n, x_{n+1}) = 0)$.

Έστω ότι ο $\Phi(x_1 \dots x_{n+1}, y)$ αναπαριστά τη G και ορίζουμε τον $\Psi(x_1, \dots, x_{n+1})$ ως εξής:

$$\Psi : \Phi(x_1 \dots x_n, x_{n+1}, \bar{0}) \wedge \forall z (z < x_{n+1} \rightarrow \neg \Phi(x_1 \dots x_n, z, \bar{0})).$$

Δοθέντος $a_1 \dots a_n \in \omega$ έστω $F(a_1, \dots, a_n) = b$ και $G(a_1, \dots, a_n, i) = c_i \ \forall i \in \omega$. Τότε από ορισμό του Φ έχουμε ότι:

$$\vdash_{S_0} \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{i}, y) \leftrightarrow y = \bar{c}_i \quad \text{για κάθε } i \in \omega.$$

Έτσι,

$$\vdash_{S_0} \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{i}, \bar{0}) \leftrightarrow \bar{0} = \bar{c}_i \quad (*)$$

Τώρα για $i < b$, $c_i \neq 0$, έχουμε από 2.1 ότι $\vdash_{S_0} \bar{c}_i \neq \bar{0} \ \forall i < b$, και άρα:

$$\vdash_{S_0} \neg \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{i}, \bar{0}) \quad \text{για κάθε } i < b \quad (1)$$

Αλλά αφού $c_b = 0$, από (2) παίρνουμε:

$$\vdash_{S_0} \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{b}, \bar{0}) \quad (2)$$

Από τις (1) και (2) και το λήμμα 2.3 έχουμε:

$\vdash_{S_0} [\Phi(\bar{a}_1, \dots, \bar{a}_n, x_{n+1}, \bar{0}) \wedge \forall z (z < x_{n+1} \rightarrow \neg \Phi(\bar{a}_1, \dots, \bar{a}_n, z, \bar{0}))] \leftrightarrow x_{n+1} = \bar{b},$
δηλαδή $\vdash_{S_0} \Psi(\bar{a}_1, \dots, \bar{a}_n, x_{n+1}) \leftrightarrow x_{n+1} = \bar{b}.$
Άρα η Ψ αναπαριστά την F .

□

†

3. ΑΡΙΘΜΗΤΙΚΟΠΟΙΗΣΗ ΤΗΣ ΛΟΓΙΚΗΣ

Ο άμεσος στόχος είναι να εκφράσουμε προτάσεις περί αξιωματικών συστημάτων, όπως η ακόλουθη:

$\exists x$ (x είναι η "απόδειξη" ενός "τύπου", " Ψ ", από "αξιώματα" " $\Phi_1, \dots, \Phi_n$ ")

σαν καθαρά **αριθμοθεωρητικές προτάσεις**.

Υποθέτουμε σε όλα τα παρακάτω ότι η γλώσσα $\mathcal{L}$ που κωδικοποιείται είναι η γλώσσα της αριθμητικής.

Ορισμός 3.1. Αν $*$ είναι ένα βασικό σύμβολο της $\mathcal{L}$, ο **αριθμός-σύμβολο** του $*$ που γράφεται ως $\otimes$ ορίζεται ως εξής:

$$\begin{aligned} \textcircled{x_i} &= 2i && \text{για κάθε μεταβλητή } x_i \ (i \in \omega) \\ \textcircled{5} &= 1, \textcircled{+} = 3, \textcircled{\cdot} = 5, \textcircled{<} = 7, \textcircled{=} = 9, \textcircled{-} = 11, \textcircled{>} = 13 \\ \textcircled{\vee} &= 15, \textcircled{=} = 17 \end{aligned}$$

Ορισμός 3.2. Αν $*$ είναι ένας όρος ή ένας τύπος της $\mathcal{L}$, ορίζουμε τον **Gödel-αριθμό** του $(*)$, που γράφεται ως $\ulcorner * \urcorner$, επαγωγικά:

(1) Για όρους t ,

(α') αν ο t είναι μια μεταβλητή ή το σύμβολο σταθεράς ($\bar{0}$), τότε

$$\ulcorner t \urcorner = \textcircled{t} \succ$$

(β') αν ο t είναι της μορφής $f(t_1, \dots, t_n)$, όπου f είναι ένα σύμβολο συνάρτησης, τότε

$$\ulcorner t \urcorner = \textcircled{f}, \ulcorner t_1 \urcorner, \dots, \ulcorner t_n \urcorner \succ$$

(2) Για τύπους Φ ,

(α') Αν Φ είναι ατομικός τύπος, δηλαδή της μορφής $t_1 = t_2$ ή $t_1 < t_2$, τότε

$$\ulcorner \Phi \urcorner = \textcircled{=}, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \succ, \textcircled{<}, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \succ$$

αντίστοιχα.

(β') Αν ο Φ είναι $\neg\Psi$, τότε

$$\ulcorner \Phi \urcorner = \textcircled{\neg}, \ulcorner \Psi \urcorner \succ$$

Αν ο Φ είναι $\Psi_1 \rightarrow \Psi_2$, τότε

$$\ulcorner \Phi \urcorner = \textcircled{\rightarrow}, \ulcorner \Psi_1 \urcorner, \ulcorner \Psi_2 \urcorner \succ$$

Αν ο Φ είναι $\forall x_i \Psi$, τότε

$$\ulcorner \Phi \urcorner = \textcircled{\forall}, \textcircled{x_i}, \ulcorner \Psi \urcorner \succ$$

Σημείωση 1: Με αναγωγή στο μήκος μιας έκφρασης, διαφορετικές εκφράσεις έχουν διαφορετικούς Gödel-αριθμούς. Υποθέστε ότι κάτι τέτοιο δεν ισχύει. Τότε έστω Φ_1 μια έκφραση με ελάχιστο δυνατό μήκος n , τέτοια ώστε υπάρχει για αυτή μια έκφραση $\Phi_2 \neq \Phi_1$, τ.ω. $\ulcorner \Phi_1 \urcorner = \ulcorner \Phi_2 \urcorner$, δηλαδή οι δύο εκφράσεις να έχουν ίδιους Gödel-αριθμούς. Τότε ο $\ulcorner \Phi_1 \urcorner$ και ο $\ulcorner \Phi_2 \urcorner$ πρέπει αμφότεροι να έχουν τη μορφή:

$$\prec \ominus, \dots \succ, \text{ ή } \prec \ominus, \dots \succ, \text{ ή } \prec \oplus, \oplus_i, \dots \succ.$$

Υποθέστε ότι για παράδειγμα είναι

$$\ulcorner \Phi_1 \urcorner = \prec \ominus, \ulcorner \Psi \urcorner \succ \quad \text{και} \quad \ulcorner \Phi_2 \urcorner = \prec \ominus, \ulcorner \Psi' \urcorner \succ.$$

Τότε θα πρέπει να είναι:

$$\ulcorner \Psi \urcorner = \ulcorner \Psi' \urcorner.$$

Από υπόθεση ελάχιστου μήκους, πρέπει να έχουμε $\Psi = \Psi'$, και άρα $\Phi_1 = \Phi_2$ (άτοπο).

Σημείωση 2: Διαισθητικά, έχουμε μια αποδοτική διαδικασία για να καθορίσουμε το αν οποιοσδήποτε δοσμένος αριθμός a είναι Gödel-αριθμός ή όχι, και αν είναι, να ανακατασκευάσουμε την έκφραση που αυτός αναπαριστά. Για να φανεί αυτό, το υποθέτουμε αληθές για όλους τους $b < a$. Τότε ελέγχουμε να δούμε αν $\text{Seq}(a)$. Αν όχι, τότε ο a δεν είναι Gödel-αριθμός. Αλλιώς, $a = \prec (a)_0 \dots (a)_{n-1} \succ$, όπου $n = \text{lh}(a)$ (μήκος του a). Αν $n \neq 2$ ή 3 τότε ο a δεν είναι Gödel-αριθμός. Αλλιώς, ελέγχουμε αν $(a)_0 = 1, 3, 5, \dots, 13, 15, 17$. Αν όχι, τότε a δεν είναι Gödel-αριθμός. Αν ναι, τότε ελέγχουμε τα $(a)_1, (a)_{n-1}$ για να δούμε αν αυτά είναι αριθμοί Gödel, και αν ναι, τίνος πράγματως είναι αριθμοί Gödel. (Αυτό μπορούμε να το κάνουμε από την υπόθεση ότι $(a)_1 \dots (a)_{n-1} < a$)

Σημείωση 3: Για να ολοκληρωθεί αυτή η αριθμητικοποίηση θα πρέπει να μπορούμε να συσχετίσουμε με μια απόδειξη $\Phi_1 \dots \Phi_n$ έναν αριθμό $\prec \ulcorner \Phi_1 \urcorner \dots \ulcorner \Phi_n \urcorner \succ$ (αυτό θα το κάνουμε αργότερα!)

Αν T είναι οποιαδήποτε θεωρία της $\mathcal{L}$, έστω:

$$\begin{cases} \text{Ax}_T = \{ \ulcorner \Phi \urcorner \mid \Phi \text{ είναι ένα μη-λογικό αξίωμα της } T \} \\ \text{Thm}_T = \{ \ulcorner \Phi \urcorner \mid \vdash_T \Phi \} \end{cases}$$

Ορισμός 3.3.

- (i) Ένα σύνολο $A \subseteq \omega$ είναι αναδρομικό αν και μόνον αν το κατηγορημα $x \in A$ είναι αναδρομικό.
- (ii) Το $P(\vec{x})$ είναι αναδρομικά αριθμήσιμο (r.e) αν υπάρχει ένα αναδρομικό κατηγορημα $R(\vec{x}, y)$ τ.ω. $P(\vec{x}) \leftrightarrow \exists y R(\vec{x}, y)$
- (iii) Ένα σύνολο $A \subseteq \omega$ είναι r.e αν και μόνον αν τυχόν κατηγορημα $x \in A$ είναι r.e

Σημείωση: Κάθε πεπερασμένο σύνολο είναι αναδρομικό.

Ορισμός 3.4. (i) Μια θεωρία T είναι (αναδρομικά) αξιωματοποιημένη αν το Ax_T είναι αναδρομικό.

- (ii) Μια θεωρία T είναι (αναδρομικά) αξιωματοποιήσιμη αν υπάρχει μια αξιωματοποιημένη θεωρία T' , τ.ω. για κάθε Φ να ισχύει:

$$\vdash_T \Phi \Leftrightarrow \vdash_{T'} \Phi.$$

Ορισμός 3.5. Μια θεωρία T είναι **αποκρίσιμη** αν το σύνολο Thm_T είναι αναδρομικό. Αλλιώς η T είναι μη-αποκρίσιμη.

Σχόλιο. Υιοθετώντας τη θέση του Church η T είναι αποκρίσιμη αν και μόνον αν υπάρχει μια διαδικασία απόφασης, για τα θεωρήματά της στο ω .

Τα αποτελέσματα του Gödel μπορούν τώρα να διατυπωθούν ως εξής:

- (1) Κάθε συνεπής επέκταση της S_0 είναι μη-αποκρίσιμη.
- (2) Κάθε αξιωματοποιημένη (συνεπής) επέκταση της S_0 είναι μη-πλήρης.

Ορισμοί για την αριθμητικοποίηση της λογικής

- (1) Για μεταβλητές:

$$\text{Vble}(a) \Leftrightarrow \text{Seq}(a) \wedge \text{lh}(a) = 1 \wedge \exists y_{y < a} ((a)_0 = 2 \cdot y)$$

Το όριο $\exists y_{y < a}$ είναι εντάξει αφού $(a)_0 < a$ (εκτός αν $a = 0$). Άρα το κατηγορήμα Vble είναι αναδρομικό.

- (2) Για όρους:

$$\text{Term}(a) \Leftrightarrow \begin{cases} a = \prec \textcircled{0} \succ, & \text{αν} & a = \prec \textcircled{0} \succ \\ \text{Term}((a)_1), & \text{αν} & a = \prec \textcircled{3}, (a)_1 \succ \\ \text{Term}((a)_1) \wedge \text{Term}((a)_2), & \text{αν} & a = \prec \oplus, (a)_1, (a)_2 \succ \\ \text{Vble}(a), & \text{ή} & \prec \ominus, (a)_1, (a)_2 \succ \\ & & \text{αλλιώς.} \end{cases}$$

Το κατηγορήμα αυτό είναι αναδρομικό από πορίσματα του 1.14, αφού όλα τα κατηγορήματα στο δεξί μέλος είναι αναδρομικά και $(a)_1, (a)_2 < a$. Ο όρος (a) σημαίνει " $a = \ulcorner t \urcorner$ για κάποιο όρο t ".

- (3) Για ατομικούς τύπους:

$$\text{Atfor}(a) \Leftrightarrow \text{Seq}(a) \wedge \text{lh}(a) = 3 \wedge ((a)_0 = \ominus \vee (a)_0 = \otimes),$$

δηλαδή $\text{Atfor}(a) \Leftrightarrow$ " a είναι ο Gödel-αριθμός ενός ατομικού τύπου". Προφανώς το κατηγορήμα Atfor είναι επίσης αναδρομικό.

(4) Για τύπους:

$$For(a) \Leftrightarrow \begin{cases} For(a) \Leftrightarrow For((a)_1), & \text{αν } a = \prec \ominus, (a)_1 \succ \\ For((a)_1) \wedge For((a)_2), & \text{αν } a = \prec \ominus, (a)_1, (a)_2 \succ \\ Vble((a)_1) \wedge For((a)_2), & \text{αν } a = \prec \bigcirc, (a)_1, (a)_2 \succ \\ Atfor, & \text{αλλιώς.} \end{cases}$$

Από πόρισμα του 1.14 το "For" είναι αναδρομικό.

$$For(a) \Leftrightarrow "a \text{ είναι Gödel-αριθμός ενός τύπου }".$$

(5) Υπάρχει μια αναδρομική συνάρτηση $Sub(a, b, c)$ τέτοια ώστε αν Φ είναι ένας τύπος με ελεύθερη μεταβλητή x_i και t είναι κάποιος όρος, τότε $Sub(\ulcorner \Phi \urcorner, \ulcorner x_i \urcorner, \ulcorner t \urcorner) = \ulcorner \Phi_{x_i}(t) \urcorner$

Σημαίνει ότι η x_i είναι αντικαταστάσιμη από τον t $\Phi(x_i)$, και επιπλέον αν t_1 είναι ένας όρος που περιλαμβάνει το x_i και t_2 είναι οποιοσδήποτε άλλος όρος, τότε:

$$Sub(\ulcorner t_1 \urcorner, \ulcorner x_i \urcorner, \ulcorner t_2 \urcorner) = \ulcorner t_1 \urcorner_{x_1}(\ulcorner t_2 \urcorner)$$

$$Sub(a, b, c) = \begin{cases} c, & \text{αν } Vble(a) \wedge a = b \\ \prec (a)_0, Sub((a)_1, b, c) \succ, & \text{αν } Seq(a) \wedge lh(a) = 2 \\ \prec (a)_0, Sub((a)_1, b, c), Sub((a)_2, b, c) \succ, & \text{αν } Seq(a) \wedge lh(a) = 3 \wedge (a)_0 \neq \bigcirc \\ \prec (a)_0, (a)_1, Sub((a)_2, b, c) \succ, & \text{αν } Seq(a) \wedge lh(a) = 3 \wedge (a)_0 = \bigcirc \wedge (a)_1 \neq b \\ a, & \text{αλλιώς.} \end{cases}$$

Προφανώς η Sub είναι αναδρομική από 1.14.

Σημ: Στη 2η παραπάνω περίπτωση a είναι της μορφής $\ulcorner S(t) \urcorner$, αν $\ulcorner \neg \Phi \urcorner$.

Η 3η παραπάνω περίπτωση καλύπτει τα a να είναι της μορφής:

- $\ulcorner t_1 + t_2 \urcorner$
- $\ulcorner t_1 \cdot t_2 \urcorner$
- $\ulcorner \Phi_1 \longrightarrow \Phi_2 \urcorner$
- $\ulcorner t_1 = t_2 \urcorner$
- $\ulcorner t_1 < t_2 \urcorner$

Στην 4η περίπτωση $a = \ulcorner \forall x_i \Phi \urcorner$, όπου $\ulcorner x_i \urcorner \neq b$

Στην 5η περίπτωση $a = \ulcorner \forall x_i \Phi \urcorner$, όπου $\ulcorner x_i \urcorner = b$, και όλες τις άλλες περιπτώσεις.

(6) Όμοια μπορούμε να δείξουμε ότι υπάρχει ένα αναδρομικό κατηγορήμα $Subtl(a, b, c)$ τέτοιο ώστε:

$Subtl(\ulcorner \Phi \urcorner, \ulcorner x_i \urcorner, \ulcorner t \urcorner)$ αν και μόνον αν ο t είναι αντικαταστάσιμος από το x_i στον Φ .

(7) Μπορούμε να ορίσουμε ένα αναδρομικό κατηγορήμα $Fr(a, b)$ τέτοιο ώστε:

$$Fr(\ulcorner \Phi \urcorner, \ulcorner x_i \urcorner) \text{ αν } x_i \text{ εμφανίζεται ελεύθερη στον } \Phi.$$

- (8) Το κατηγορήμα $\text{Lax}(a)$ αν και μόνον αν " a είναι ο Gödel-αριθμός ενός λογικού αξιώματος", είναι αναδρομικό.

$$\text{Lax}(a) \Leftrightarrow \begin{cases} \exists b_{b < a} \exists c_{c < a} [\text{For}(b) \wedge \text{For}(c) \wedge a = \prec \ominus, b, \prec \ominus, c, b \succ \succ] & (A1) \\ \vee \dots\dots\dots & (A2) \\ \vee \dots\dots\dots & (A3) \\ \vee \exists b_{b < a} \exists c_{c < a} \exists d_{d < a} [\text{For}(b) \wedge \text{Var}(c) \wedge \text{Term}(d) \wedge \text{Subtl}(b, c, d) \wedge \\ \wedge a = \prec \ominus, \prec \oplus, c, b \succ, \text{Sub}(b, c, d) \succ] & (A4) \\ \vee \dots\dots\dots & (A5) \end{cases}$$

- (9) Υπάρχουν αναδρομικά κατηγορήματα $\text{MP}(a, b, c)$ και $\text{Gen}(a, b, c)$ τέτοια ώστε

$$\text{MP}(\ulcorner \Phi \urcorner, \ulcorner \Psi \urcorner, \ulcorner X \urcorner) \Leftrightarrow X \text{ επάγεται από τους } \Phi \text{ και } \Psi \text{ μέσω του κανόνα MP.}$$

και

$$\text{Gen}(\ulcorner \Phi \urcorner, \ulcorner \Psi \urcorner) \Leftrightarrow \text{Ο } \Psi \text{ είναι της μορφής } \forall x_i \Phi.$$

Τα MP και G ορίζονται ως:

$$\begin{aligned} \text{MP}(a, b, c) &\Leftrightarrow \text{For}(a) \wedge \text{For}(b) \wedge \text{For}(c) \wedge (a = \prec \ominus, b, c \succ \vee b = a \ominus, a, c \succ) \\ \text{Gen}(a, b) &\Leftrightarrow \text{For}(a) \wedge \text{For}(b) \wedge \exists c_{c < b} (\text{Var}(c) \wedge b = \prec \oplus, c, a \succ) \end{aligned}$$

- (10) Υπάρχει μια αναδρομική συνάρτηση $\text{Cl}(a)$ τέτοια ώστε:

$$\text{Cl}(\ulcorner \Phi \urcorner) = \ulcorner \Phi^* \urcorner,$$

όπου Φ^* ορίζει το κλείσιμο του Φ .

- (11) Σε κάθε πεπερασμένη ακολουθία από τύπους $\Phi_1, \dots, \Phi_n$ αναθέτουμε τώρα έναν αριθμό $\prec \ulcorner \Phi_1 \urcorner, \dots, \ulcorner \Phi_n \urcorner \succ$. Προφανώς αυτός ο ορισμός είναι 1-1.

Έστω T μια αξιωματοποιημένη θεωρία (δηλ. το A_{XT} είναι αναδρομικό). Τότε μπορούμε να δείξουμε ότι το κατηγορήμα:

$$\text{Prf}_T(a) \Leftrightarrow \begin{aligned} &\text{υπάρχει μια ακολουθία τύπων } \Phi_1, \dots, \Phi_n, \text{ η οποία αποτελεί} \\ &\text{μια απόδειξη του } \Phi_n \text{ από τα αξιώματα της } T \text{ τέτοια ώστε} \\ &a = \prec \ulcorner \Phi_1 \urcorner, \dots, \ulcorner \Phi_n \urcorner \succ \end{aligned}$$

είναι αναδρομικό. Αυτό ισχύει διότι το Prf_T γράφεται ως εξής:

$$\begin{aligned} \text{Prf}_T(a) \Leftrightarrow \text{Seq}(a) \wedge \text{lh}(a) \neq 0 \wedge \bigvee_{i < \text{lh}(a)} \forall i \left[A_{XT}((a)_i) \vee \text{Lax}((a)_i) \vee \right. \\ \left. \vee \exists j_{j < i} \exists k_{k < i} \left(\text{MP}((a)_j, (a)_k, (a)_i) \vee \text{Gen}((a)_j, (a)_i) \right) \right] \end{aligned}$$

(12) Για ευκολία ορίζουμε

$$\text{Pf}_T(a, b) \Leftrightarrow \text{PrF}_T(b) \wedge a = (b)_{lh(b)} \dot{-} 1,$$

δηλαδή, "b είναι ο αριθμός απόδειξης μιας απόδειξης του τύπου ο οποίος έχει Gödel-αριθμό τον a."

Φανερά το Pf_T είναι αναδρομικό αν η T είναι αξιωματοποιημένη.

†

Θεώρημα 3.6. *Αν η T είναι αξιωματοποιημένη, τότε το Thm_T είναι αναδρομικά αριθμήσιμο.*

Απόδειξη. $\text{Thm}_T(a) \leftrightarrow \exists b \text{Pf}_T(a, b)$.

Άρα από το (12) παραπάνω, το Thm_T είναι αναδρομικά αριθμήσιμο, αν η T είναι αξιωματοποιημένη. □

Ορισμός: Ορίζουμε τη συνάρτηση ενός ορίσματος Num ως εξής:

$$\begin{aligned} \text{Num}(0) &= \prec \textcircled{0} \succ \\ \text{Num}(a+1) &= \prec \textcircled{S}, \text{Num}(a) \succ \end{aligned}$$

Φανερά, $\text{Num}(a) = \ulcorner \bar{a} \urcorner$ και η $\text{Num}(a)$ είναι αναδρομική (από R3).

Θεώρημα 3.7. (Θεώρημα του Church) *Κάθε συνεπής επέκταση του S_0 είναι μη-αποκρίσιμη.*

Απόδειξη. Σταθεροποιήστε κάποια μεταβλητή της $\mathcal{L}$, έστω τη x_1 . Έστω $T \supseteq S_0$ να είναι συνεπής. Πρέπει να δείξουμε ότι το Thm_T δεν είναι αναδρομικό. Έστω ένα κατηγορήμα P που ορίζεται ως εξής:

$$P(a) \Leftrightarrow \neg \text{Thm}_T(\text{Sub}(a, \ulcorner x_1 \urcorner, \text{Num}(a))).$$

Υποθέστε ότι το Thm_T είναι αναδρομικό. Τότε το P είναι αναδρομικό. Τότε το P είναι αναπαριστάσιμο στην T από κάποιο τύπο $\Phi(x_1)$ (μπορούμε να επιλέξουμε την ελεύθερη μεταβλητή ώστε να είναι η x_1).

Έστω $k = \ulcorner \Phi(x_1) \urcorner$. Τότε

$$\begin{aligned} P(k) &\Leftrightarrow \neg \text{Thm}_T(\ulcorner \Phi(\bar{k}) \urcorner) \\ &\Leftrightarrow \not\vdash_T \Phi(\bar{k}) \end{aligned}$$

Αλλά αφού ο Φ αναπαριστά το P , έχουμε:

$$\begin{aligned} P(k) &\Rightarrow \vdash_T \Phi(\bar{k}) \\ \text{και } \neg P(k) &\Rightarrow \vdash_T \neg \Phi(\bar{k}) \end{aligned}$$

Έτσι το $P(k)$ δε μπορεί να ισχύει, και αν το $\neg P(k)$ ισχύει, τότε η T είναι μη συνεπής. Άρα το Thm_T δε μπορεί να είναι αναδρομικό, και έπεται το αποτέλεσμα. $\square$

Λήμμα 3.8. Ένα κατηγορημα P είναι αναδρομικό αν και μόνον αν το P και το $\neg P$ είναι αναδρομικά αριθμήσιμα.

Θεώρημα 3.9. Αν η T είναι αξιωματοποιημένη και πλήρης, τότε η T είναι αποκρίσιμη.

Απόδειξη. Υποθέτουμε την T συνεπή, αλλιώς τετριμμένο. Από λήμμα 3.8 αρκεί να αποδείξουμε ότι τα Thm_T και $\neg \text{Thm}_T$ είναι αναδρομικά αριθμήσιμα. Ξέρουμε ότι το Thm_T είναι αναδρομικά αριθμήσιμο αφού η T είναι αξιωματοποιημένη. Τώρα για κάθε τύπο Φ έχουμε

$$\not\models_T \Phi \Leftrightarrow \not\models_T \Phi^* \Leftrightarrow \vdash_T \neg \Phi^*,$$

όπου Φ^* είναι το κλείσιμο του Φ (αφού T είναι πλήρης). Άρα,

$$\neg \text{Thm}_T(a) \leftrightarrow \neg \text{For}(a) \vee \exists y \text{Pf}_T(\prec \ominus, \text{Cl}(a) \succ, y),$$

οπότε φανερά το $\neg \text{Thm}_T$ είναι αναδρομικά αριθμήσιμο και το αποτέλεσμα έπεται. $\square$

Θεώρημα 3.10. (1^ο θεώρημα μη-πληρότητας του Gödel, 1931)

Κάθε συνεπής αξιωματοποιημένη επέκταση του S_0 είναι μη-πλήρης.

Σημείωση 1: Προφανώς, το "αξιωματοποιημένη" μπορεί να αντικατασταθεί από το "αξιωματοποιήσιμη".

Σημείωση 2: Συγκεκριμένα, η S (δηλ. η αριθμητική του Peano) είναι μη-πλήρης.

Θεώρημα 3.11. (Λήμμα της Αυτο-Αναφοράς)

Έστω $\Phi(x)$ ένας τύπος με μια ελεύθερη μεταβλητή. Τότε μπορούμε να κατασκευάσουμε μια πρόταση Ψ τέτοια ώστε

$$\vdash_{S_0} \Psi \leftrightarrow \Phi(\ulcorner \Psi \urcorner)$$

[**Σημείωση :** Μας ενδιαφέρει η περίπτωση όπου ο $\Phi(x)$ "λέει" ότι, "το x δεν είναι Gödel-αριθμός ενός θεωρήματος της S_0 "]

Απόδειξη. Έστω F να ορίζεται ως εξής:

$$F(a) = \text{Sub}(a, \ulcorner x_1 \urcorner, \text{Num}(a))$$

Αφού η F είναι αναδρομική, είναι αναπαραστάσιμη στο S_0 από κάποιο τύπο $\Theta(x_1, x_2)$. Θεωρούμε τώρα τον τύπο $\forall x_2[\Theta(x_1, x_2) \rightarrow \Phi(x_2)]$. Δηλώνουμε τον Gödel-αριθμό αυτού του τύπου ως k . Τώρα έστω Ψ να είναι η πρόταση: $\forall x_2[\Theta(\bar{k}, x_2) \rightarrow \Phi(x_2)]$. Σημειώνουμε ότι $F(k) = \ulcorner \Psi \urcorner$. Τώρα, αφού ο Θ αναπαριστά την F στην S_0 έχω:

$$\vdash_{S_0} \Theta(\bar{k}, x_2) \leftrightarrow x_2 = \ulcorner \Psi \urcorner \quad (*)$$

Επίσης,

$$\Psi \vdash_{S_0} \Theta(\bar{k}, \ulcorner \Psi \urcorner) \rightarrow \Phi(\ulcorner \Psi \urcorner) \quad (\text{από ορισμό της } \Psi)$$

Άρα η $(*)$ δίνει $\Psi \vdash_{S_0} \Phi(\ulcorner \Psi \urcorner)$, άρα

$$\vdash_{S_0} \Psi \rightarrow \Phi(\ulcorner \Psi \urcorner) \quad (1)$$

Αντίστροφα, από $(*)$ έχω:

$$\vdash_{S_0} \Phi(\ulcorner \Psi \urcorner) \rightarrow \forall x_2(\Theta(\bar{k}, x_2) \rightarrow \Phi(x_2)),$$

δηλαδή,

$$\vdash_{S_0} \Phi(\ulcorner \Psi \urcorner) \rightarrow \Psi \quad (2)$$

Από τις σχέσεις (1) και (2) το αποτέλεσμα του λήμματος έπεται. $\square$

Ορισμός 3.12. Ένα κατηγορήμα $P \subseteq \omega^n$ καλείται **αριθμητικό** αν υπάρχει ένας τύπος $\Phi(x_1, \dots, x_n)$ τέτοιος ώστε $\forall a_1 \dots a_n \in \omega, P(a_1, \dots, a_n) \Leftrightarrow \mathbb{N} \models \Phi(\bar{a}_1, \dots, \bar{a}_n)$. Ένα σύνολο A είναι αριθμητικό αν και μόνον αν το κατηγορήμα $x \in A$ είναι αριθμητικό.

Θεώρημα 3.13. Κάθε αναδρομικά αριθμήσιμο (κατά μείζονα λόγο αναδρομικό) κατηγορήμα P είναι αριθμητικό.

Απόδειξη. Απλά βασίζεται στο θεώρημα αναπαραστασιμότητας.

Υποθέτουμε $P(\vec{x}) \Leftrightarrow \exists t R(\vec{x}, y)$, όπου το R είναι αναδρομικό, και έστω ότι ο $\Psi(\vec{x}, y)$ αναπαριστά το R στην S_0 . Τότε:

$$\begin{aligned} P(a_1, \dots, a_n) &\Leftrightarrow \exists y R(a_1, \dots, a_n, y) \\ &\Leftrightarrow R(a_1, \dots, a_n, b) \text{ για κάποιο } b \in \omega \\ &\Leftrightarrow \vdash_{S_0} \Psi(\bar{a}_1, \dots, \bar{a}_n, \bar{b}) \text{ για κάποιο } b \in \omega \\ &\Leftrightarrow \mathbb{N} \models \Psi(\bar{a}_1, \dots, \bar{a}_n, \bar{b}) \text{ για κάποιο } b \in \omega \\ &\Leftrightarrow \mathbb{N} \models \exists y \Psi(\bar{a}_1, \dots, \bar{a}_n, y) \end{aligned}$$

Άρα, φανερά το $\exists y \Psi(\vec{x}, y)$ ορίζει το $P(\vec{x})$ στη δομή $\mathbb{N}$. Άρα το P είναι αριθμητικό.

□

Σχόλιο. Γενικοποιώντας τα παραπάνω, μπορούμε να δείξουμε ότι ένα κατηγορημα $P(\vec{x})$ είναι αριθμητικό αν και μόνον αν είναι ισοδύναμο με μια από τις μορφές:
 $\Box y_1 \cdots \Box y_n R(\vec{x}, y_1, \dots, y_n)$, όπου το R είναι αναδρομικό και τα $\Box$ είναι εναλλασσόμενοι $\exists$ και $\forall$ ποσοδείκτες.

Θεώρημα 3.14. (Θεώρημα απροσδιοριστίας του Tarski)

Το σύνολο Tr των Gödel-αριθμών αληθών τύπων της $\mathbb{N}$ δεν είναι αριθμητικό.

Απόδειξη. Έστω το σύνολο: $Tr = \{\ulcorner \Phi \urcorner \mid \mathbb{N} \models \Phi\}$.

Υποθέτουμε ότι το Tr είναι όντως αριθμητικό, και έστω ότι για κάποιο $\Phi(x)$ ισχύει:

$$Tr = \{a \mid a \in \omega \wedge \mathbb{N} \models \Phi(\bar{a})\}$$

Τότε από 3.11 (λήμμα αυτο-αναφοράς), υπάρχει μια πρόταση Ψ τέτοια ώστε:

$$\mathbb{N} \models \Psi \leftrightarrow \neg \Phi(\ulcorner \Psi \urcorner).$$

Τότε,

$$\mathbb{N} \models \Psi \Leftrightarrow \mathbb{N} \models \neg \Phi(\ulcorner \Psi \urcorner) \Leftrightarrow \mathbb{N} \not\models \Psi,$$

από ιδιότητα ορισμού της Φ .

□

Απόδειξη 1^{ου} θεωρήματος μη-πληρότητας (4.11).

Τώρα εφαρμόζουμε το λήμμα της αυτοαναφοράς 3.11 για να κατασκευάσουμε μια μη-αποκρίσιμη πρόταση για μια δοθείσα συνεπή αξιωματοποιημένη επέκταση T της S_0 . Για απλότητα θα υποθέσουμε ότι $T \subseteq Th(\mathbb{N})$. Αφού η T είναι αξιωματοποιημένη, το $Pf_T(x, y)$ είναι αναδρομικό και άρα αναπαρίσταται στην T από κάποιο τύπο $\Theta(x, y)$. Έστω τώρα ο τύπος $\Phi(x)$ να είναι ο $\forall y \neg \Theta(x, y)$. Τότε από 3.11 μπορούμε να κατασκευάσουμε μια πρόταση Ψ τέτοια ώστε:

$$\vdash_T \Psi \leftrightarrow \forall y \neg \Theta(\ulcorner \Psi \urcorner, y).$$

Ισχυρισμός. Η Ψ είναι μη αποκρίσιμη στην T .

Αρκεί να δείξουμε ότι δε μπορούμε να αποδείξουμε στην T ούτε την Ψ , ούτε την $\neg \Psi$. Για την Ψ έχουμε:

$$\begin{aligned} \vdash_T \Psi &\Rightarrow \vdash_T \forall y \neg \Theta(\ulcorner \Psi \urcorner, y) \\ &\Rightarrow \vdash_T \neg \Theta(\ulcorner \Psi \urcorner, \bar{a}) \quad \forall a \in \omega \\ &\Rightarrow \neg Pf_T(\ulcorner \Psi \urcorner, a) \quad \forall a \in \omega \\ &\Rightarrow \not\vdash_T \Psi \quad \text{Άτοπο.} \end{aligned}$$

xxx

Για τη $\neg\Psi$ έχουμε:

$$\begin{aligned}
\vdash_T \neg\Psi &\Rightarrow \vdash_T \exists y \Theta(\ulcorner \Psi \urcorner, y) \\
&\Rightarrow \mathbb{N} \models \exists y \Theta(\ulcorner \Psi \urcorner, y), \quad \text{αφού } T \subseteq \text{Th}(\mathbb{N}) \\
&\Rightarrow \mathbb{N} \models \Theta(\ulcorner \Psi \urcorner, \bar{a}), \quad \text{για κάποιο } a \in \omega \\
&\Rightarrow \not\vdash_T \neg\Theta(\ulcorner \Psi \urcorner, \bar{a}), \quad \text{αφού } T \subseteq \text{Th}(\mathbb{N}) \\
&\Rightarrow \text{Pf}_T(\ulcorner \Psi \urcorner, a), \quad \text{αφού το } \Theta \text{ αναπαριστά το } \text{Pf}_T \\
&\Rightarrow \vdash_T \Psi \quad \text{και πάλι άτοπο.}
\end{aligned}$$

Άρα τελικά η Ψ είναι όντως μη-αποκρίσιμη, οπότε η T είναι μη-πλήρης. □

Θεώρημα 3.15. *Αν T είναι μια αξιωματοποιημένη επέκταση της S_0 , τότε οι συναρτήσεις και τα κατηγορήματα που αναπαρίστανται στην T είναι ακριβώς οι αναδρομικές συναρτήσεις και κατηγορήματα.*

Απόδειξη. Από προηγούμενα αποτελέσματα χρειαζόμαστε μόνο να θεωρήσουμε συναρτήσεις. Επίσης χρειάζεται μόνο να αποδείξουμε ότι κάθε αναπαριστάσιμη συνάρτηση είναι αναδρομική. Έστω η $F(x_1, \dots, x_n)$ που μπορεί να παρασταθεί στην T από τον τύπο $\Phi(x_1, \dots, x_n, x_{n+1})$. Έστω $\text{Sub}^{n+1}(z, x_1 \dots x_{n+1}, y_1 \dots y_{n+1})$ να ορίζει το εξής:

$$\underbrace{\text{Sub}(\dots (\text{Sub}(z, x_1, y_1), x_2, y_2), x_3, y_3) \dots, x_{n+1}, y_{n+1})}_{n+1 \text{ φορές}} \dots$$

Προφανώς η Sub^{n+1} είναι αναδρομική, και έχει την ιδιότητα:

$$\text{Sub}^{n+1}(\ulcorner \Phi \urcorner, x_1 \dots x_{n+1}, \text{Num}(a_1), \dots \text{Num}(a_n), \text{Num}(b)) = \ulcorner \Phi(\bar{a}_1 \dots \bar{a}_n, \bar{b}) \urcorner,$$

για οποιοδήποτε $a_1 \dots a_n, b \in \omega$.

Τώρα, εάν $F(a_1, \dots, a_n) = b$ τότε $\exists z \text{Pf}_T(\ulcorner \Phi(\bar{a}_1 \dots \bar{a}_n, \bar{b}) \urcorner, z)$. Επίσης σημειώνεται ότι για οποιαδήποτε $b' \neq b$, $\neg \exists z \text{Pf}_T(\ulcorner \Phi(\bar{a}_1 \dots \bar{a}_n, \bar{b}') \urcorner, z)$.

Έτσι, για οποιαδήποτε $a_1, \dots, a_n$ υπάρχει z και **μοναδικό** b τ.ώ.

$$\text{Pf}_T(\ulcorner \Phi(\bar{a}_1, \dots, \bar{a}_n, \bar{b}) \urcorner, z)$$

και σε αυτή την περίπτωση $F(\bar{a}_1, \dots, \bar{a}_n) = b$. Έτσι,

$$\begin{aligned}
F(a_1, \dots, a_n) &= (\mu y \text{Pf}_T(\ulcorner \Phi(\bar{a}_1, \dots, \bar{a}_n, \overline{(y)}_0) \urcorner, (y)_1))_0 \\
&= [\mu y \text{Pf}_T(\text{Sub}^{n+1}(\ulcorner \Phi \urcorner, \ulcorner x_1 \urcorner \dots \ulcorner x_{n+1} \urcorner, \text{Num}(a_1), \dots \text{Num}(a_n), \text{Num}((y)_0), (y)_1)))]_0,
\end{aligned}$$

η οποία είναι προφανώς αναδρομική. □

4. ΕΦΑΡΜΟΓΕΣ

Υποθέτουμε ότι έχουμε μια γλώσσα $\mathcal{L}$ η οποία έχει πεπερασμένες μόνο μη-λογικές σταθερές και έτσι υπόκειται σε αριθμητικοποίηση κατά Gödel.

Λήμμα 4.1. Έστω T και T' θεωρίες (πιθανόν σε διαφορετικές γλώσσες) τέτοιες ώστε για κάθε Φ της $\mathcal{L}_T$ να υπάρχει ένας τύπος Φ' της $\mathcal{L}_{T'}$ τ.ώ. $\vdash_T \Phi \Leftrightarrow \vdash_{T'} \Phi'$ και υποθέστε ότι αυτή η χαρτογράφηση είναι αναδρομική, δηλαδή υπάρχει μια αναδρομική συνάρτηση F τέτοια ώστε $F(\ulcorner \Phi \urcorner) = \ulcorner \Phi' \urcorner$. Τότε, αν η T είναι μη-αποκρίσιμη, είναι και η T' .

Απόδειξη. Υποθέτουμε ότι η T' είναι αποκρίσιμη. Τότε το $\text{Thm}_{T'}$ είναι αναδρομικό. Αλλά μπορούμε να ορίσουμε το Thm_T ως εξής:

$$\text{Thm}_T(a) \Leftrightarrow \underbrace{\text{Thm}_{T'}[F(a)]}_{\text{αναδρομικό}} \wedge \text{For}_T(a).$$

Άρα το Thm_T είναι αναδρομικό, άτοπο (αφού η T μη αποκρίσιμη). □

Ορισμός 4.2. Η T' είναι μια πεπερασμένη επέκταση της T αν $T' = T \cup \{\Phi_1, \dots, \Phi_n\}$ για κάποια $\Phi_1, \dots, \Phi_n$.

Θεώρημα 4.3. Αν T είναι πεπερασμένη επέκταση της T' , και T είναι μη αποκρίσιμη, τότε και η T' είναι μη-αποκρίσιμη.

Απόδειξη. Έστω $T = T' \cup \{\Phi \cdots \Phi_n\}$ και έστω ο τύπος: $\Psi = \Phi_1^* \wedge \Phi_2^* \wedge \cdots \wedge \Phi_n^*$. Τότε από θεώρημα πληρότητας έχω:

$$\vdash_T \Phi \Leftrightarrow \vdash_{T'} \Psi \rightarrow \Phi.$$

Η συνάρτηση $F(x) = \ulcorner \ominus, \ulcorner \Psi \urcorner, x \urcorner$ είναι αναδρομική και άρα από 4.1 αν η T είναι μη-αποκρίσιμη, είναι και η T' . □

Πόρισμα 4.4. Η κενή θεωρία της γλώσσας της αριθμητικής είναι μη αποκρίσιμη. Για αυτό ο κατηγορηματικός λογισμός είναι μη-αποκρίσιμος.

Απόδειξη. Η S_0 είναι πεπερασμένα αξιωματοποιημένη και μη-αποκρίσιμη. □

Αποτέλεσμα 4.5. Ο κατηγορηματικός λογισμός για μια γλώσσα με ένα μόνο δυαδικό κατηγορικό γράμμα (και χωρίς την ισότητα) είναι μη-αποκρίσιμος. Αυτό είναι το καλύτερο δυνατό αποτέλεσμα διότι:

Αποτέλεσμα 4.6. Ο κατηγορηματικός λογισμός για μια γλώσσα με μόνο εναδικά κατηγορήματα (και χωρίς συναρτήσεις) είναι αποκρίσιμος.

Απόδειξη. Υπάρχει διαδικασία απόφασης, η οποία είναι η εξής:

Αν $P_1 \cdots P_k$ είναι τα διακεκριμένα κατηγορήματα τα οποία απαντώνται σε ένα τύπο Φ αυτής της γλώσσας, τότε ισχύει ότι:

$\vdash \Phi \Leftrightarrow$ ο Φ ισχύει σε όλες τις ερμηνείες των $(P_1 \cdots P_k)$ με λιγότερα από 2^k στοιχεία.

Διότι, αν $\mathcal{A}$ είναι **οποιαδήποτε** ερμηνεία της γλώσσας με σύμβολα κατηγορημάτων $P_1 \cdots P_k$, τότε μπορούμε να ανάγουμε την $\mathcal{A}$ σε μια ερμηνεία $\mathcal{A}'$ ώστε:

$$\mathcal{A} \models \Phi \Leftrightarrow \mathcal{A}' \models \Phi \quad \text{και} \quad |\overline{\mathcal{A}'}| \leq 2^k,$$

παίρνοντας ισοδύναμες κλάσεις από στοιχεία του $|\mathcal{A}|$ υπό τη σχέση $\sim$, ως εξής:

$$a \sim b \Leftrightarrow [P_i^{\mathcal{A}}(a) \Leftrightarrow P_i^{\mathcal{A}}(b)] \text{ για κάθε } i \in \{1, \dots, k\}$$

$$P_i^{\mathcal{A}'}([a]) \Leftrightarrow P_i^{\mathcal{A}}(a).$$

Τότε με επαγωγή στην πολυπλοκότητα του Φ έχω

$$\mathcal{A} \models \Phi \Leftrightarrow \mathcal{A}' \models \Phi.$$

Για αυτό αν ο Φ δεν είναι πλέον έγκυρος, τότε $\mathcal{A} \not\models \Phi$ για κάποιες $\mathcal{A}$ με λιγότερα από 2^k στοιχεία.

Άρα αποδείχτηκε.

Τώρα υπάρχουν μόνο ένας πεπερασμένος αριθμός από μη-ισομορφικές ερμηνείες με λιγότερα από 2^k στοιχεία στον τομέα (για την ακρίβεια υπάρχουν λιγότερα από $\sum_{r=1}^{2^k} (2^r)^k$ από αυτά).

Καθεμία από αυτές πρέπει να εξεταστεί για να καθορίσουμε αν ο Φ ισχύει σ' αυτήν ή όχι. Αν ο Φ ισχύει σε όλες, τότε ο Φ είναι λογικά έγκυρος (αλλιώς όχι).

Ορισμός. Μια δομή $\mathcal{A}$ είναι **ισχυρά μη-αποκρίσιμη** αν κάθε θεωρία T που έχει την $\mathcal{A}$ σαν μοντέλο είναι μη-αποκρίσιμη.

□

Θεώρημα 4.7. Η δομή $\mathbb{N}$ είναι ισχυρά μη αποκρίσιμη.

Απόδειξη. Υποθέστε ότι $T \subseteq \text{Th}(\mathbb{N})$. Τότε η $\mathbb{N}$ είναι μοντέλο για την $S_0 \cup T$. Αλλά η $S_0 \cup T$ είναι μη αποκρίσιμη από θεώρημα του Church. Άρα η T είναι μη αποκρίσιμη από θεώρημα 4.3.

□

Ορισμός 4.8. Αν $\mathcal{A}$ και $\mathcal{B}$ είναι δομές (πιθανώς διαφορετικού τύπου) θα λέμε ότι η $\mathcal{A}$ είναι **ορίσιμη στην $\mathcal{B}$** αν $|\mathcal{A}| \subseteq |\mathcal{B}|$ και:

- (1) Υπάρχει ένας τύπος $\Phi_{\mathcal{A}}(x)$ της $\mathcal{L}_{\mathcal{B}}$ τέτοιος ώστε:

$$|\mathcal{A}| = \{b \in |\mathcal{B}| \mid \mathcal{B} \models \Phi_{\mathcal{A}}[b]\}$$

- (2) Για κάθε n -αδικό κατηγορήμα P^A της $\mathcal{A}$ υπάρχει ένας τύπος $\Phi_P(x_1 \cdots x_n)$ της $\mathcal{L}_B$ τ.ώ.

$$P^A(a_1 \cdots a_n) \Leftrightarrow B \models \Phi_P[a_1 \cdots a_n], \quad \forall a_1 \cdots a_n \in |A|$$

- (3) Για κάθε n -αδική συνάρτηση f^A της $\mathcal{A}$, υπάρχει ένας τύπος $\Phi_f(x_1 \cdots x_n, x_{n+1})$ της $\mathcal{L}_B$ τ.ώ.

$$f^A(a_1 \cdots a_n) = b \Leftrightarrow B \models \Phi_f[a_1 \cdots a_n, b] \quad \forall a_1 \cdots a_n, b \in |A|$$

- (4) Για κάθε σταθερά c^A της $\mathcal{A}$, υπάρχει ένας τύπος $\Phi_c(x)$ της $\mathcal{L}_B$ τ.ώ.

$$a = c^A \Leftrightarrow B \models \Phi_c[a] \quad \text{για κάθε } a \in |A|$$

Παράδειγμα: Η $\mathbb{N}$ είναι ορίσιμη στη δομή $\langle \omega, +, \cdot \rangle = \mathbb{N}^-$.
Διότι, $|\mathbb{N}| = |\mathbb{N}^-| = \omega$, άρα το (1) ισχύει τετρωμένα. Το μοναδικό κατηγορήμα $<$ της $\mathbb{N}$ μπορεί να οριστεί στην $\mathbb{N}^-$ ως εξής:

$$\Phi_{<}(x_1, x_2) \leftrightarrow \exists z(x_1 + z = x_2) \wedge x_1 \neq x_2.$$

Για το (3), η συνάρτηση του επόμενου σ μπορεί να οριστεί από τη σχέση:

$$\Phi_{\sigma}(x_1, x_2) \leftrightarrow \Phi_{<}(x_1, x_2) \wedge \neg \exists z(\Phi_{<}(x_1, z) \wedge \Phi_{<}(z, x_2))$$

Τέλος, για το (4) έχω ότι το 0 μπορεί να οριστεί ως εξής:

$$\Phi_0(x) \leftrightarrow \neg \exists z \Phi_{<}(z, x)$$

Θεώρημα 4.9. (Tarski)

Αν η $\mathcal{A}$ είναι ορίσιμη στη B και η $\mathcal{A}$ είναι ισχυρά μη-αποκρίσιμη, τότε είναι και η B .

Σκιαγράφηση απόδειξης. Πρώτα σημειώνουμε ότι για κάθε τύπο Φ της $\mathcal{L}_A$ μπορούμε να κατασκευάσουμε ένα τύπο Φ^+ της $\mathcal{L}_B$ τ.ώ. $A \models \Phi \Leftrightarrow B \models \Phi^+$ (και επιπλέον η χαρτογράφηση $\Phi \rightarrow \Phi^+$ είναι αναδρομική).

Για παράδειγμα, αν ο Φ είναι ο $\forall x_1(\neg \forall x_2 P(x_1, x_2) \vee Q(x_1))$, τότε ο Φ^+ θα είναι ο:

$$\forall x_1 [\Phi_A(x_1) \rightarrow (\neg \forall x_2 (\Phi_A(x_2) \rightarrow \Phi_P(x_1, x_2) \vee \Phi_Q(x_1)))]$$

Έστω $\Sigma \subseteq \text{Th}(B)$. Θα δείξουμε ότι η Σ είναι μη-αποκρίσιμη. Τώρα έστω

$$T = \{\Phi \mid \Phi \in \mathcal{L}_A \wedge \vdash_{\Sigma} \Phi^+\}$$

Τότε είναι σχετικά εύκολο να δείξουμε ότι:

$$\vdash_T \Phi \Leftrightarrow \vdash_{\Sigma} \Phi^+ \quad (*)$$

[Το " $\Leftarrow$ " βγαίνει κατευθείαν από τον ορισμό, ενώ η άλλη κατεύθυνση (" $\Rightarrow$ ") είναι συνέπεια της φύσης της χαρτογράφησης του Φ στο Φ^+ . Γιατί το $\vdash_T \Phi \Rightarrow \Phi^+$ ισχύει σε κάθε μοντέλο της T και έστω B_0 να είναι τυχαίο μοντέλο της Σ . Τότε μπορούμε να ορίσουμε μια δομή A_0 τ.ώ. $|A_0| \subseteq |B_0|$ με τον προφανή τρόπο. Τότε θα έχουμε ότι $A_0 \models \Phi$ και λόγω αυτού, ότι $B_0 \models \Phi^+$. Άρα $\vdash_{\Sigma} \Phi^+$.]

Τέλος, η $\mathcal{A}$ είναι ένα μοντέλο της T . Διότι:

$$\begin{aligned}\Phi \in T \Rightarrow \vdash_{\Sigma} \Phi^+ &\Rightarrow \mathcal{B} \models \Phi^+ \text{ αφού } \mathcal{B} \text{ είναι μοντέλο της } \Sigma \\ &\Rightarrow \mathcal{A} \models \Phi.\end{aligned}$$

Άρα η T είναι μη-αποκρίσιμη από υπόθεση για την $\mathcal{A}$. Άρα από (*) και λήμμα 4.1 έχουμε ότι η Σ είναι μη-αποκρίσιμη και άρα αποδείχτηκε το ζητούμενο. $\square$

Πόρισμα 4.10. Η δομή $\langle \omega, +, \cdot \rangle$ είναι ισχυρά μη-αποκρίσιμη.

Θεώρημα 4.11. (Julia Robinson, 1949)

Ο δακτύλιος των ακεραίων $\langle \mathbb{Z}, +, \cdot \rangle$ είναι ισχυρά μη-αποκρίσιμος.

Απόδειξη. Αρχεί να δείξουμε ότι υπάρχει ένας τύπος $\Phi_{\omega}(x)$ που ορίζει τους φυσικούς αριθμούς πάνω στο δακτύλιο $\langle \mathbb{Z}, +, \cdot \rangle$. Τότε απλά εφαρμόζουμε το θεώρημα Tarski και το 4.10.

Έστω ότι αυτός είναι ο τύπος:

$$\Phi_{\omega}(x) \leftrightarrow \exists y \exists z \exists s \exists t (x = y \cdot y + z \cdot z + s \cdot s + t \cdot t).$$

Τότε από το θεώρημα Lagrange (κάθε φυσικός είναι άθροισμα 4 τετραγώνων), ο $\Phi_{\omega}(x)$ ορίζει το ω πάνω στον $\langle \mathbb{Z}, +, \cdot \rangle$. $\square$

Πόρισμα 4.12. Οι πρωτοβάθμιες θεωρίες δακτυλίων, μεταβατικών δακτυλίων, ολοκληρωτικών πεδίων είναι όλες μη-αποκρίσιμες.

Το σύνολο $\text{Th}(\langle \mathbb{R}, +, \cdot \rangle)$ είναι αποκρίσιμο αφού είναι αξιωματοποιήσιμο και πλήρες.

Πόρισμα 4.13. Οι φυσικοί αριθμοί δεν είναι ορίσιμοι από κανένα πρωτοβάθμιο τύπο πάνω στο $\langle \mathbb{R}, +, \cdot \rangle$. Επίσης, οι ακέραιοι μπορούν να οριστούν με πρωτοβάθμιο τύπο πάνω στους ρητούς $\langle \mathbb{Q}, +, \cdot \rangle$.

Από αυτό το γεγονός, έπεται το:

Πόρισμα 4.14. Οι πρωτοβάθμιες θεωρίες των πεδίων, των πεδίων με χαρακτήρες 0, κτλ. είναι όλες μη-αποκρίσιμες.

5. 2^ο ΘΕΩΡΗΜΑ ΜΗ-ΠΛΗΡΟΤΗΤΑΣ ΤΟΥ GÖDEL

Θεωρήστε την αριθμητική του Peano (αξιωματικό σύστημα S).

Έστω Ψ_S να είναι η Gödel πρόταση που αντιστοιχεί στο S . Ένα απόσπασμα από το 1^ο θεώρημα μη πληρότητας:

$$\text{Αν η } S \text{ είναι συνεπής τότε } \not\vdash_S \Psi_S. \quad (*)$$

Η $(*)$ είναι μια μετα-θεωρητική πρόταση η οποία μπορεί να μεταφραστεί σε μια πρόταση της πρωτοβάθμιας αριθμητικής ως εξής:

Έστω $\Theta_S(x, y)$ να είναι ο τύπος που αναπαριστά το $\text{Pf}_S(x, y)$ στην S .

Ας πάρουμε το:

$$\underbrace{\forall y \neg \Theta(\overline{\Gamma \bar{0} = \bar{1}}, y)}_{\text{Con}_S} \rightarrow \forall y \neg \Theta_S(\overline{\Gamma \Psi_S}, y) \quad (**)$$

Εκτελώντας τα επιχειρήματα του 1^{ου} θεωρήματος μη-πληρότητας στην τυπική αριθμητική, μπορούμε να πάρουμε την απόδειξη στο S της $(**)$. Έστω το Con_S να δηλώνει το 1^ο μέρος της $(**)$. Έπεται από τα παραπάνω ότι:

Θεώρημα 5.1. (2^ο θεώρημα μη-πληρότητας του Gödel)

Αν το S είναι συνεπές, τότε δε μπορούμε να αποδείξουμε τη συνέπειά του σε αυτό, δηλ. $\not\vdash_S \text{Con}_S$.

Απόδειξη. Αν $\vdash_S \text{Con}_S$, τότε, αφού αποδεικνύεται η $\vdash_S (**)$, παίρνουμε ότι

$$\vdash_S \forall y \neg \Theta_S(\overline{\Gamma \Psi_S}, y).$$

Τώρα από την κατασκευή της Ψ_S , έχω $\vdash_S \Psi_S \leftrightarrow \forall y \neg \Theta_S(\overline{\Gamma \Psi_S}, y)$, άρα η Ψ_S αποδεικνύεται στο S , πράγμα αδύνατο αν το S είναι συνεπές, από επιχείρημα του 1^{ου} θεωρήματος μη-πληρότητας. □

Το θεώρημα 5.1 ισχύει γενικότερα σε κάθε αξιωματοποιημένη θεωρία T η οποία είναι "αρκετά ισχυρή", όπου μπορούμε να ορίσουμε αυτή την έννοια ως εξής:

- (1) Υπάρχει μια αναδρομική 1-1 αντιστοιχία τύπων της $\mathcal{L}_{\mathbb{N}}$ σε τύπους της $\mathcal{L}_T$, η $\Phi \rightarrow \Phi^+$ τέτοια ώστε
- (2) $\vdash_S \Phi \Rightarrow \vdash_T \Phi^+$

Συγκεκριμένα, η θεωρία συνόλων ZF είναι αρκετά ισχυρή.

Γι' αυτό, αν η ZF είναι συνεπής, τότε δε μπορεί να αποδείξει η ίδια τη συνέπειά της.

Από ένα αποτέλεσμα του Yuri Matiyasevich (1970), κάθε αναδρομικό κατηγορημα $P(\vec{x})$ μπορεί να αναπαρασταθεί (π.χ. στην S) από ένα τύπο της μορφής:

$$\exists \vec{y} \quad \underbrace{(t_1(\vec{x}, \vec{y}) = t_2(\vec{x}, \vec{y}))}_{\text{Αυτό είναι απλά ένα πολυώνυμο στα } \vec{x}, \vec{y}}$$

Για παράδειγμα, υποθέτουμε το $P(x)$ εναδικό και αναδρομικό κατηγορημα. Τότε για κάποια διοφαντική εξίσωση είναι $x, y_1, \dots, y_n$, ας πούμε

$\forall x \in \omega \quad P(x) \leftrightarrow \exists y_1 \dots y_n \quad (\text{δηλαδή η διοφαντική εξίσωση στο } (y_1 \dots y_n, x) \text{ ισχύει.})$

Γι'αυτό η Con_{ZF} , η Con_S , κτλ, είναι όλες προτάσεις οι οποίες μπορούν να εκφραστούν στη μορφή:

$$\forall x \neg \exists \vec{y} (t_1(x, \vec{y}) = t_2(x, \vec{y})),$$

δηλαδή

$$\forall x \forall \vec{y} \neg (t_1(x, \vec{y}) = t_2(x, \vec{y})),$$

δηλαδή η διαφαντική εξίσωση $t_1(x, \vec{y}) = t_2(x, \vec{y})$ δεν έχει λύση για φυσικούς αριθμούς.

Για να δείξουμε την επέκταση των αναπόφευκτων διασυνδέσεων ανάμεσα στη θεωρία συνόλων και τη θεωρία αριθμών, μπορούμε να δείξουμε ότι η πρόταση Con_{ZF} υπονοείται από την ύπαρξη κάποιου είδους μεγάλων πληθικών αριθμών. Δουλεύοντας στη θεωρία συνόλων ZF με το αξίωμα της επιλογής και αναγνωρίζοντας σαν πληθικούς αριθμούς τους αρχικούς διατακτικούς, ορίζουμε:

Ένας πληθικός k είναι μη-προσβάσιμος αν:

- (i) Αν $\bar{A}_i < k$ για κάθε $i \in I$ και $\bar{I} < k$, τότε $\overline{\bigcup_{i \in I} A_i} < k$, και
- (ii) $\bar{A} < k \Rightarrow \overline{\mathcal{P}(A)} < k$

Αυτά είναι ισοδύναμα με:

- (i') Αν πληθικοί $\mu < k$ και $\alpha_\beta < k$, όπου α_β διατακτικοί και k πληθικός για κάθε $\beta < \mu$, τότε $\bigcup_{\beta < \mu} \alpha_\beta < k$.
- (ii') Για κάθε άλλο πληθικό λ ισχύουν: $\lambda < k$ και $2^\lambda < k$.

Προφανώς ο $\aleph_0$ είναι μη προσβάσιμος, αλλά είναι εύκολο να δούμε ότι οι $\aleph_1, \aleph_2, \dots, \aleph_\omega$ και $2^{\aleph_0}, 2^{2^{\aleph_0}}, \dots$ δεν είναι μη προσβάσιμοι. Μάλλον, κανείς δεν ξέρει αν κάποιος μεγαλύτερος μη-προσβάσιμος πληθικός από τον $\aleph_0$ υπάρχει, αν και φαίνεται αρκετά πιθανό να συμβαίνει. Παρόλα αυτά, δεν είναι δύσκολο να δείξουμε ότι η ZF αποδεικνύει πως:

”Αν ο K είναι μη προσβάσιμος και μεγαλύτερος του $\aleph_0$, τότε το $\langle R(k), \in \upharpoonright R(k) \rangle$ είναι ένα μοντέλο της ZF .”

Τώρα, αφού ισχύει ότι:

$$\vdash_{ZF} \text{”}ZF \text{ έχει μοντέλο} \rightarrow \text{Con}_{ZF}\text{”}$$

παίρνουμε ότι $\vdash_{ZF} \exists k (k \text{ μη-προσβάσιμο} \wedge k > \aleph_0) \rightarrow \text{Con}_{ZF}$.

Έτσι από το 2^ο θεώρημα μη πληρότητας, αν η ZF είναι συνεπής δε μπορούμε να αποδείξουμε την ύπαρξη μη προσβάσιμων, μεγαλύτερων του $\aleph_0$ στην ZF . Από την άλλη, αν υποθέσουμε ότι υπάρχει ένας (σαν αξίωμα), τότε παίρνουμε μια απόδειξη ότι μια συγκεκριμένη διαφαντική εξίσωση (ισοδύναμη με την Con_{ZF} , βλ. αποτέλεσμα του Matiyasevich) δεν έχει λύση - ένα θεώρημα το οποίο δε μπορούμε να ελπίζουμε ότι θα το αποδείξουμε στη ZF μόνο (αν αυτή είναι συνεπής) ! Αυτό το αποτέλεσμα δείχνει με δραματικό τρόπο τις καταπληκτικές διασυνδέσεις ανάμεσα στη θεωρία αριθμών και τις πιο απόμακρες περιοχές της θεωρίας συνόλων.