



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

ΤΜΗΜΑ ΙΣΤΟΡΙΑΣ ΚΑΙ ΦΙΛΟΣΟΦΙΑΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΨΥΧΟΛΟΓΙΑΣ

Γεωργία Δέδε

A.M. 15Μ03

**Γνωστικές Προκαταλήψεις: Σκέψεις και ιδέες με βάση την προσέγγιση της  
Δυνητικής Λογικής**

Διπλωματική εργασία

για τη λήψη μεταπτυχιακού διπλώματος ειδίκευσης από το  
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών στη Γνωστική Επιστήμη

Συμβουλευτική επιτροπή:

Ιζαμπώ Καράλη, Επ. Καθηγήτρια, ΕΚΠΑ

Κωνσταντίνος Δημητρακόπουλος, Καθηγητής, ΕΚΠΑ

Κωνσταντίνος Μουτούσης, Καθηγητής, ΕΚΠΑ

Αθήνα

Νοέμβριος 2021

Εγκρίνεται η διπλωματική εργασία

Ιζαμπώ Καράλη .....

Κωνσταντίνος Δημητρακόπουλος .....

Κωνσταντίνος Μουτούσης .....

© 2021 Georgia Dede

Διαθέσιμη στο διαδίκτυο με διατήρηση πνευματικών δικαιωμάτων δημιουργού  
(άδεια χρήσης τύπου Creative Commons Attribution 4.0 International, CC-BY 4.0)

<https://creativecommons.org/licenses/by/4.0/deed.el>

## Σύνοψη

Το φαινόμενο των γνωστικών προκαταλήψεων παραμένει αρκετά σκοτεινό πεδίο της ανθρώπινης νόησης. Αρκετά πειράματα έχουν δείξει ότι κατά τη διαδικασία συμπερασμού είναι αρκετά συχνό φαινόμενο ένα έλλειμμα ορθολογικότητας και ταυτόχρονα μια τάση επιμονής στο σύνολο των παγιωμένων ατομικών πεποιθήσεων ανεξάρτητα από τα δεδομένα που είναι διαθέσιμα. Θα προσπαθήσουμε να προσεγγίσουμε τα φαινόμενα αυτά μέσα από τη θεωρία της Δυνητικής Λογικής χρησιμοποιώντας τις βασικές έννοιες των μέτρων δυνατότητας και αναγκαιότητας και διατυπώνοντας την υπόθεση ότι οι γνωστικές προκαταλήψεις έχουν έναν αυξημένο βαθμό βεβαιότητας έναντι της νέας πληροφορίας καθώς και ότι η αναθεώρηση πεποιθήσεων δεν είναι μια γραμμική διαδικασία, δεδομένου ότι στο εκάστοτε σύστημα πεποιθήσεων του υποκειμένου μπορούν περισσότερα από ένα γεγονότα να είναι ταυτόχρονα αληθή.

# **Cognitive Biases: A possibilistic approach**

## **Abstract**

Cognitive biases remain a dark field when studying the human mind. Many experiments over time indicate that during inference people tend to have a lack of rationality and at the same time persistence in prior beliefs regardless of the available data. We will try to approach these phenomena using the Possibilistic Logic theory and the main notions of necessity and possibility measures. We make a hypothesis that cognitive biases stand as preconditions with a certainty degree higher than the incoming information and that revising and changing beliefs is not a linear procedure, since in reality, more than one fact can be true at the same time.

## Ευχαριστίες

Θα ήθελα να πω ένα μεγάλο ευχαριστώ στους καθηγητές μου καθ' όλη τη διάρκεια του προγράμματος για όλες τις γνώσεις που πήρα για να φτάσω ως εδώ.

Ιδιαίτερες ευχαριστίες οφείλω στα μέλη της τριμελούς συμβουλευτικής επιτροπής για τη συνεργασία που είχαμε και ιδιαίτερα στην κυρία Καράλη γιατί οι συζητήσεις μας ήταν για μένα πηγή έμπνευσης και δύναμης.

Χρωστάω απέραντη ευγνωμοσύνη σε όλους αυτούς που στάθηκαν δίπλα μου αυτήν τη δύσκολη περίοδο και με στήριξαν για να ολοκληρώσω αυτό το πόνημα. Τους ευχαριστώ όλους προσωπικά – οικογένεια, φίλους και μη.

Τέλος, θέλω να αφιερώσω αυτό το επίτευγμα στο Δημήτρη, γιατί του οφείλω αυτό που είμαι σήμερα...

# Περιεχόμενα

Σύνοψη.....	2
<b>Cognitive Biases: A possibilistic approach .....</b>	<b>3</b>
<b>Abstract.....</b>	<b>3</b>
<b>Ευχαριστίες .....</b>	<b>4</b>
<b>Εισαγωγή .....</b>	<b>6</b>
<b>Τελικά είμαστε ορθολογικά όντα ή μήπως όχι;.....</b>	<b>7</b>
<b>Doxastic conservativeness: Προσπάθεια διατήρησης ομοιόστασης; .....</b>	<b>12</b>
<b>Η προσέγγιση της Δυνητικής Λογικής.....</b>	<b>13</b>
<b>Σύντομη εισαγωγή στις έννοιες της Δυνητικής Λογικής .....</b>	<b>16</b>
<b>Σύντομη εισαγωγή στη Θεωρία Αναθεώρησης Πεποιθήσεων (Belief Revision).....</b>	<b>22</b>
<b>Οι συλλογιστικές προκαταλήψεις μέσα από την έρευνα των Nonaes και Veluwenkamp .....</b>	<b>32</b>
<b>Τα φαινόμενα overgeneration και undergeneration ως συνέπειες των προκαταλήψεων.....</b>	<b>33</b>
<b>Σκέψεις και Ιδέες για την εφαρμογή της Δυνητικής Λογικής και της Αναθεώρησης Πεποιθήσεων στις γνωστικές προκαταλήψεις.....</b>	<b>36</b>
<b>Οι γνωστικές προκαταλήψεις ως καταστάσεις με αυξημένο βαθμό αναγκαιότητας .....</b>	<b>36</b>
<b>Βάσεις πεποιθήσεων και νοητικά μοντέλα αναπαράστασης του κόσμου .....</b>	<b>38</b>
<b>Το παράδοξο της Λίντα .....</b>	<b>41</b>
<b>Συμπεράσματα και ιδέες για μελλοντικές έρευνες.....</b>	<b>45</b>
<b>Βιβλιογραφία.....</b>	<b>47</b>

## Εισαγωγή

Η ανθρώπινη συλλογιστική έχει αποτελέσει πεδίο ενδιαφέροντος ήδη από τους φιλοσόφους της Αρχαιότητας και έχουν γίνει προσπάθειες διερεύνησης και προσέγγισης της από διάφορους επιστημονικούς κλάδους, όπως η Ψυχολογία και η Τεχνητή Νοημοσύνη (ΤΝ). Η ικανότητά μας να διατυπώνουμε συλλογισμούς, να εκτελούμε πράξεις σύγκρισης, αφαίρεσης και σύνθεσης μεταξύ αυτών και να οδηγούμαστε σε συμπεράσματα θεωρείται ένα από τα κεντρικά χαρακτηριστικά της ανθρώπινης σκέψης και σημαντικό σημείο διαχωρισμού του ανθρώπου από τα υπόλοιπα ζώα. Πολλές προσπάθειες έχουν γίνει κυρίως από περιοχές της Πειραματικής Ψυχολογίας και της Τεχνητής Νοημοσύνης να αποκρυπτογραφήσουμε τους μηχανισμούς της ανθρώπινης συλλογιστικής και να προσεγγίσουμε τους πιθανούς «αλγορίθμους» της. Ταυτόχρονα, η ανθρώπινη συλλογιστική έχει τεθεί υπό κρίση όσον αφορά στην ορθότητά της και στην εγγενή δυνατότητά της να παράγει ορθά συμπεράσματα. Συνεπώς, οι επιστημονικές προσπάθειες προσέγγισης έχουν πολλαπλούς στόχους – να ανακαλύψουμε αν υπάρχει ένας κοινός τρόπος συλλογιστικής, ποιοι είναι οι υφιστάμενοι μηχανισμοί, αν υπάρχουν συγκεκριμένα νοητικά μοντέλα που μπορούν να προσεγγίσουν την ανθρώπινη συλλογιστική και να την αναπαραγάγουν και αν οι τρόποι συλλογισμού που χρησιμοποιούμε είναι σωστοί, υπό την έννοια ότι μας οδηγούν σε βέλτιστα και ορθά συμπεράσματα.

Η Κλασική Λογική ως διαδικασία συμπερασμού είχε ήδη θεμελιωθεί από τον Αριστοτέλη. Ήταν και είναι σαφές ότι αυτή η διαδικασία μας οδηγεί σε αδιαμφισβήτητα συμπεράσματα και ουσιαστικά στην οδό της αλήθειας. Εντούτοις, η πραγματικότητα μας έχει δείξει ότι δε σκεφτόμαστε πάντα χρησιμοποιώντας αυστηρούς λογικούς κανόνες παραγωγής συμπερασμάτων. Τις περισσότερες φορές τα δεδομένα που έχουμε δεν επαρκούν για να διατυπώσουμε ακριβείς και πλήρεις προκείμενες που θα μας οδηγήσουν, μέσω της εφαρμογής των κανόνων, σε αλάνθαστα συμπεράσματα. Επίσης, ζώντας σε ένα ανοιχτό σύστημα, ο κόσμος των δυνατών ενδεχομένων είναι επίσης ανοιχτός και τα δεδομένα μπορούν να αλλάξουν από τη μια στιγμή στην άλλη. Συνεπώς, οδηγούμαστε στην αναζήτηση άλλων προσεγγίσεων εκτός της Κλασικής Λογικής προκειμένου να

δημιουργήσουμε κάποιο πιθανό εξηγητικό πλαίσιο που θα μπορεί να ενσωματώσει την ανθρώπινη συλλογιστική.

## **Τελικά είμαστε ορθολογικά όντα ή μήπως όχι;**

Οι γνωστικές προκαταλήψεις, θεωρούμενες ως απόκλιση από τον ορθολογικό τρόπο σκέψης, έχουν μελετηθεί κατά καιρούς στο πλαίσιο της κατανόησης της ανθρώπινης σκέψης. Είναι σημαντικό να σκεφτούμε ότι γνωστικές προκαταλήψεις υπήρχαν ανέκαθεν και θα συνεχίσουν να υπάρχουν. Πριν μιλήσουμε για τις γνωστικές προκαταλήψεις, θα ήταν χρήσιμο να διευκρινίσουμε τί ορίζουμε ως απόκλιση από τον ορθό λόγο.

Ως ορθολογική διαδικασία συμπερασμού τείνουμε να θεωρούμε κυρίως τη διαδικασία της παραγωγής (deduction). Τα παραγωγικά ορίσματα μπορούν να διατυπωθούν με μαθηματικούς φορμαλισμούς βάσει των οποίων το συμπέρασμα προκύπτει από τις προκειμένες και η ορθότητά του μπορεί να ελεγχθεί λαμβάνοντας υπ' όψιν τον πίνακα αληθοτιμών των προκειμένων. Αντιλαμβανόμαστε ότι η διαδικασία αυτή έχει μεγάλη αποδεικτική αξία παράγοντας ασφαλή αληθή συμπεράσματα και πιθανώς αυτός είναι και ο λόγος που έχει ταυτιστεί με την ορθολογική σκέψη. Ως απόκλιση, λοιπόν, συχνά θεωρούμε μια διαδικασία συμπερασμού που διαφέρει από αυτήν τη συγκεκριμένη διαδικασία.

Για παράδειγμα, για την Παραγωγική Συλλογιστική (Deductive Reasoning) ισχύει:

Προκειμένες:

1. Όλα τα A είναι B. ( $A \rightarrow B$ )
2. Όλα τα B είναι Γ. ( $B \rightarrow \Gamma$ )

Συμπέρασμα:

Όλα τα A είναι Γ. ( $A \rightarrow \Gamma$ )

Η αλήθεια είναι, όμως, ότι πολύ συχνά δε σκεφτόμαστε έτσι – ορθολογικά. Για παράδειγμα, αν τώρα είναι πρωί, είναι εύλογο να σκεφτώ ότι σε μερικές ώρες θα είναι βράδυ, χωρίς να έχω καμιά σχετική πληροφορία. Μου αρκεί απλώς η παρατήρηση ότι



αυτό συμβαίνει κατά κανόνα στη φύση και γενικεύω τη γνώση αυτή υποθέτοντας – και μάλιστα με απόλυτη βεβαιότητα – ότι αυτό θα συμβεί και σήμερα και θα συνεχίσει να συμβαίνει για όσο υπάρχει ο κόσμος με τη μορφή που γνωρίζουμε. Αλλά ας πούμε ότι αυτό το παράδειγμα δεν είναι πολύ ενδεικτικό, διότι αυτή η παρατήρηση του κόσμου στηρίζεται ταυτόχρονα σε επιστημονικά δεδομένα που με τη σειρά τους οδήγησαν στη δημιουργία μοντέλων που είναι σε θέση με ακρίβεια και σε μεγάλο βάθος χρόνου να υπολογίσουν αρκετά από τα φυσικά φαινόμενα. Συνεπώς, η πεποίθησή μου ότι σε μερικές ώρες θα είναι βράδυ θα μπορούσε να θεωρηθεί ορθολογική και μάλιστα επιστημονικά τεκμηριωμένη.

Ας θεωρήσουμε μια άλλη συνθήκη ως παράδειγμα:

Ο άνθρωπος X φτάνει στη στάση του λεωφορείου μιας πολυσύχναστης γραμμής σε ώρα αιχμής και δε βλέπει κανέναν άλλον να περιμένει. Ας υποθέσουμε, επίσης, ότι δε συντρέχει κάποια εξαιρετική συνθήκη, όπως μια απεργία, για παράδειγμα. Με αρκετή βεβαιότητα μπορεί να υποθέσει ότι το λεωφορείο έχει περάσει και μάλιστα πριν από λίγη ώρα. Αν ο ίδιος άνθρωπος φτάσει σε ώρα που δεν έχει πολλή κίνηση και δεν υπάρχουν άλλοι άνθρωποι που περιμένουν, ο βαθμός βεβαιότητας για να συμπεράνει αν όντως έχει περάσει το λεωφορείο μειώνεται κατά πολύ.

Ας πούμε τώρα ότι ο ίδιος άνθρωπος X βιάζεται και πρέπει να αποφασίσει τί θα κάνει σε κάθε περίπτωση – αν, δηλαδή, θα περιμένει το λεωφορείο ή θα επιλέξει άλλο μέσο. Όταν πρόκειται για ώρα αιχμής, δίνει μεγαλύτερο βαθμό βεβαιότητας στο ενδεχόμενο να έχει μόλις περάσει το λεωφορείο, οπότε, το πιθανότερο είναι να πάρει ένα άλλο μέσο. Όταν δεν είναι ώρα αιχμής, είναι περίπου ισοπίθανο να επιλέξει άλλο μέσο ή να περιμένει κάποια ώρα μήπως έρθει το λεωφορείο.

Το παράδειγμα αυτό μας οδηγεί σε διάφορες σκέψεις. Είναι ξεκάθαρο ότι ο X σε καμία από τις δύο συνθήκες δεν έχει σαφή πληροφόρηση για το τί συμβαίνει με το λεωφορείο – αν έχει περάσει ή όχι. Σήμερα θα μπορούσε να έχει πληροφόρηση χρησιμοποιώντας μια εφαρμογή στο κινητό του τηλέφωνο. Τί γινόταν όμως σε προηγούμενες δεκαετίες; Το σίγουρο είναι ότι η ζωή δε σταματούσε περιμένοντας να έχουμε επαρκή πληροφόρηση. Πολύ συχνά, λοιπόν, οι άνθρωποι καλούνταν να πάρουν αποφάσεις βάσει της εμπειρίας

τους, βάσει του εσωτερικού τους μοντέλου του κόσμου. Και, όπως φαίνεται, δεν τα πήγαμε και άσχημα ως είδος, όσον αφορά στο κομμάτι της επιβίωσης τουλάχιστον.

Ο τρόπος συμπερασμού που δε βασίζεται μόνο σε πραγματικά και συγκεκριμένα δεδομένα, αλλά και σε γενικεύσεις ή πεποιθήσεις που προκύπτουν από τη δική μας ατομική ή συλλογική εμπειρία φαίνεται να είναι αρκετά αποτελεσματικός. Μια πιθανή υπόθεση αναφορικά με το τί πραγματικά συμβαίνει είναι ότι χρησιμοποιούμε κάποιου είδους ευρετικές (heuristics) νοητικές τεχνικές οι οποίες δεν είναι προϊόντα αυστηρών μαθηματικών υπολογισμών και οι οποίες καταφέρνουν να μειώσουν τελικά το χρόνο που χρειαζόμαστε για επίλυση προβλημάτων και λήψη αποφάσεων στην καθημερινή ζωή. Οι τεχνικές αυτές οδηγούν στην εδραίωση κάποιων πεποιθήσεων που είναι πιθανό να αποτελέσουν πυλώνες στον τρόπο σκέψης μας και που δεν είναι απαραίτητα σωστές ή λανθασμένες, αλλά φαίνεται πως είναι αρκετά συχνές μέσα στο ανθρώπινο είδος και πιθανότατα αρκετά χρήσιμες. Φυσικά περιμένουμε ότι με τις νέες πληροφορίες που αποκτούμε οι πεποιθήσεις μας επικαιροποιούνται, ή τουλάχιστον έτσι θα πρέπει να γίνεται. Μπορούμε, επίσης, να υποθέσουμε ότι όσο πιο άκαμπτες είναι αυτές οι πεποιθήσεις, τόσο δυσχεραίνεται η προσαρμογή του εκάστοτε οργανισμού. Παρ' όλα αυτά πολύ συχνά βλέπουμε ότι προκείμενες που συμφωνούν με τις πεποιθήσεις μας να υπερισχύουν στη διαδικασία του συμπερασμού συγκριτικά με προκείμενες που είναι μεν αληθείς αλλά δε συνάδουν με τις πεποιθήσεις μας, όπως φαίνεται και από τη μελέτη των Novaes και Veluwenkamp. (Novaes et al, 2017)

Οι γνωστικές προκαταλήψεις (cognitive biases) έχουν προβληματίσει αρκετά τους επιστήμονες που ασχολούνται με τη νόηση. Φαίνεται ότι διαδραματίζουν σημαντικό ρόλο στις ευρετικές τεχνικές που χρησιμοποιούμε κατά την επίλυση προβλημάτων και τη λήψη αποφάσεων χωρίς να είναι ξεκάθαρο αν είναι προϊόντα ή γενεσιουργός αιτία αυτών ή και τα δύο. Γιατί όμως να καταλήγουμε σε τέτοιες διαδικασίες συμπερασμού και όχι στην παραγωγή, που η ορθότητα των προκειμένων θα μας διασφαλίσει την ορθότητα του συμπεράσματος; Μήπως τελικά δεν είμαστε ορθολογικά όντα; Όσο φιλοσοφικό και αν ακούγεται αυτό το ερώτημα, έχουν υπάρξει προσπάθειες να δοθούν επιστημονικά τεκμηριωμένες απαντήσεις.

Μια απάντηση δίνεται στο άρθρο «Heuristics Made Easy: An Effort-Reduction Framework» (Shah et al, 2008) όπου προτείνεται ότι η χρήση ευρετικών μηχανισμών προκύπτει από την ανάγκη μείωσης των απαιτούμενων πόρων που σχετίζονται με μια νοητική διεργασία. Κατά τους Shah και Oppenheimer για τη διαδικασία της λήψης απόφασης, σύμφωνα με τον αλγόριθμο που προτείνεται από τον σταθμισμένο προσθετικό κανόνα (weighted additive rule), οι άνθρωποι πρέπει να δαπανήσουν πόρους για πέντε προαπαιτούμενα βήματα: 1) γνώση όλων των παραμέτρων που σχετίζονται με το ζήτημα, 2) ανάκληση και αποθήκευση (είτε στη μνήμη είτε εξωτερικά) όλων των τιμών αυτών των παραμέτρων, 3) ανάθεση τιμών βάρους σε κάθε παράμετρο ανάλογα με τη σημαντικότητά της στο πρόβλημα, 4) συνδυασμός παραμέτρων προκειμένου να καθοριστούν όλες οι πιθανές εναλλακτικές και 5) σύγκριση των εναλλακτικών ώστε να γίνει επιλογή της πιο συμφέρουσας. Προφανώς, τα στάδια αυτά είναι κοστοβόρα – απαιτούν και υπολογιστικούς πόρους και χρόνο. Οι υπολογιστικοί πόροι του ανθρώπινου εγκεφάλου έχουν τους προβλεπόμενους φυσικούς περιορισμούς ενώ ο χρόνος υπόκειται επίσης σε περιορισμούς με βάση τις συνθήκες. Ας αναλογιστούμε, για παράδειγμα, ότι μια ομάδα πρωτόγονων ανθρώπων έβλεπαν για πρώτη φορά κατά τη διάρκεια του κυνηγιού ένα ζώο πολύ μεγαλύτερο σε μέγεθος από το μέσο άνθρωπο. Επειδή ήταν η πρώτη φορά, δεν είχαν την πληροφόρηση αν το ζώο είναι επιθετικό και δυνητικά απειλητικό για τη ζωή τους και σίγουρα δεν είχαν το χρόνο να συγκεντρώσουν τα απαιτούμενα δεδομένα πριν δράσουν. Η προσπάθεια διαφυγής, ακόμα κι αν ήταν λάθος στην περίπτωση που το ζώο ήταν φιλικό, διασφάλιζε τουλάχιστον την επιβίωσή τους. Με διασφαλισμένη την επιβίωση, θα μπορούσαν σε δεύτερο χρόνο να διερευνήσουν περαιτέρω το συγκεκριμένο ζώο και τη συμπεριφορά του, ώστε είτε να συνεχίσουν να το αποφεύγουν είτε να το κυνηγήσουν είτε να το χρησιμοποιήσουν με άλλους τρόπους.

Επιστρέφοντας στο ερώτημα περί ορθολογικότητας του ανθρώπου, μια προφανής απάντηση είναι η ανάπτυξη της επιστήμης. Αν δεν είμαστε ορθολογικά όντα, τί μπορούμε να ισχυριστούμε για την επιστήμη που έχουμε δημιουργήσει και αυτή αποτελεί το σημείο αναφοράς μας καθώς ορίζουμε τον ορθό λόγο; Είναι άραγε δυνατό για μη ορθολογικά όντα να παραγάγουν ορθό επιστημονικό λόγο; Μήπως οι επιστήμονες διαφέρουν νοητικά από τον υπόλοιπο πληθυσμό; Από τη σκοπιά της μοναδικότητας του κάθε εγκεφάλου, όλοι είμαστε διαφορετικοί. Όμως, η θεωρία του κοινού νου μας αποτρέπει από

διαχωρισμούς των ανθρώπων βάσει νοητικών λειτουργιών. Ταυτόχρονα, όπως πολύ εύστοχα επισημαίνει ο Philip Johnson-Laird, αν δεν είμαστε σε θέση να κάνουμε αυστηρά παραγωγικούς λογικούς συμπερασμούς, δεν μπορεί να εξηγηθεί η αγάπη πολλών ανθρώπων, οι οποίοι μάλιστα δεν έχουν λάβει κάποια εκπαίδευση στη λογική, για παιχνίδια όπως το Sudoku, που βασίζονται σε αυστηρά λογικούς κανόνες. (Johnson-Laird, 2010)

Αξιοσημείωτη είναι μια φράση του Simon όπως παρατίθεται στο άρθρο των Dunbar και Klahr “Scientific Thinking and Reasoning”:

*«It is understandable, if ironic, that ‘normal’ science fits ... the description of expert problem solving, while ‘revolutionary’ science fits the description of problem solving by novices.» (Simon, Langley, & Bradshaw, 1981, p. 5)*

Κατά το Simon, η «τυπική» επιστήμη απαιτεί τον τρόπο σκέψης μιας αυθεντίας, αλλά η ρηξικέλευθη τον τρόπο σκέψης ενός αρχαρίου.

Οι Dunbar και Klahr διαχωρίζουν δύο τύπους επιστημονικής σκέψης: ο ένας τύπος είναι η σκέψη σχετικά με το περιεχόμενο της επιστήμης (π.χ. σκέψη για τα φυσικά μεγέθη, τη μάζα, τη δύναμη, τα πεδία, ο σχεδιασμός ενός πειράματος κλπ) και ο άλλος τύπος αφορά στις νοητικές διαδικασίες που ακολουθούνται για να παραχθεί επιστήμη. Για τους συγγραφείς οι μεγάλες ανακαλύψεις δεν προκύπτουν από εξαιρετικές νοητικές διεργασίες, αλλά από κοινές, όπως είναι η επαγωγή, η απαγωγή και ο τρόπος επίλυσης προβλημάτων που χρησιμοποιούμε στην καθημερινή μας ζωή. Άλλωστε, «*το μέγεθος μιας επιστημονικής ανακάλυψης προσδιορίζεται από αυτό που έχει ανακαλυφθεί*». (Dunbar et al., 2012)

Η προσέγγιση αυτή της επιστημονικής σκέψης καταρχήν μας κάνει να σκεφτούμε ότι δεν μπορούμε να απορρίψουμε κάποια από τις μεθόδους συμπερασμού ως άχρηστη ή ακατάλληλη, γιατί το ζητούμενο δεν είναι πάντα η υψηλή αποδεικτική αξία ενός συμπερασμού. Μπορεί μεν η εγκυρότητα ενός συλλογισμού να σχετίζεται με την αποδεικτική του αξία, αλλά πολύ συχνά, και στην επιστήμη αλλά και στην καθημερινή ζωή, η ανάγκη, για παράδειγμα, ομαδοποίησης οντοτήτων βάσει κοινών χαρακτηριστικών είναι το πρώτο απαιτούμενο στάδιο πριν φτάσουμε στη διατύπωση μιας υπόθεσης, εν συνεχεία σε γνώση και τελικά στη διαμόρφωση κανόνων που

παράγουν ασφαλή συμπεράσματα. Ας σκεφτούμε το παράδειγμα με τους ανθρώπους των προϊστορικών χρόνων. Αν συναντούσαν ένα άγνωστο πολύ μεγαλόσωμο ζώο, η γενίκευση ότι κάθε πολύ μεγαλόσωμο ζώο είναι σε θέση να νικήσει τον άνθρωπο σε μια μάχη μπορεί να αποβεί σωτήρια. Πιθανότατα δεν ισχύει για όλα τα μεγαλόσωμα ζώα, αλλά μέχρι να αποκτήσει τα απαιτούμενα δεδομένα για να τα διακρίνει, μπορεί να κινδυνεύσει ακόμη και η ζωή του. Συνεπώς, μια γενίκευση πολύ συχνά μπορεί να είναι πολύ πιο χρήσιμη από την ανάγκη ορθού συμπερασμού έστω και αν το συμπέρασμα στο οποίο έχουμε καταλήξει δεν είναι απολύτως βέβαιο, αλλά έχει έναν, δεδομένων των συνθηκών, ικανοποιητικό βαθμό βεβαιότητας. Μεταφέροντας ανάλογα τη ρήση των Dunbar και Klahr, μπορούμε να πούμε ότι η αξία του είδους του συμπερασμού σχετίζεται με την ελαχιστοποίηση του πιθανού κόστους που απορρέει από τη χρήση του. Αν η προσπάθεια επίτευξης ενός έγκυρου συμπερασμού είναι πιθανό να κοστίσει την επιβίωση, τότε είναι προτιμότερος ένας μη έγκυρος.

### **Doxastic conservatism: Προσπάθεια διατήρησης ομοιότητας;**

Όπως συζητήσαμε, φαίνεται ότι κανένα από τα διαφορετικά είδη συμπερασμού που χρησιμοποιούνται από τους ανθρώπους δεν μπορεί να απορριφθεί ως άχρηστο. Μια υπόθεση που μπορούμε να διατυπώσουμε είναι ότι συχνά ίσως το στάδιο της παραγωγής προηγείται της παραγωγής, σκεπτόμενοι πως πρώτα παρατηρήσαμε κανονικότητες στον κόσμο που μας περιβάλλει και στη συνέχεια οδηγηθήκαμε σε νόμους που φαίνεται ότι τον εξηγούν με έναν απaráβατο και συνεπή τρόπο. Επίσης, σύμφωνα με τον Philip N. Johnson-Laird, η επαγωγική διαδικασία, συγκριτικά με την έγκυρη παραγωγική διαδικασία, αυξάνει την πληροφορία γιατί δίνει τη δυνατότητα να προσθέσουμε προϋπάρχουσα γνώση και να πάμε πέρα από τα αυστηρά πλαίσια των προκειμένων. (Johnson-Laird, 2010) Ας μην ξεχνάμε ότι η προσπάθεια εύρεσης κοινών μοτίβων μεταξύ φαινομένων, αντικειμένων κλπ με στόχο την τοποθέτησή τους σε κλάσεις ώστε να δημιουργηθεί ένα ενιαίο εξηγητικό πλαίσιο είναι απαραίτητο στάδιο της εξέλιξης της επιστήμης. Συνεπώς, το ότι η παραγωγική διαδικασία συμπερασμού έχει υψηλή αποδεικτική αξία δε σημαίνει ότι όποτε εφαρμόζεται κάποια άλλη διαδικασία

συμπερασμού, τότε αυτή είναι λανθασμένη. Αυτά όμως όσον αφορά σε πρωταρχικές απόπειρες συμπερασμού και ελλείπει απαιτούμενων δεδομένων. Τί γίνεται όμως όταν έχουμε τα σωστά δεδομένα, αλλά τα αγνοούμε οδηγούμενοι σε λανθασμένα συμπεράσματα;

Φαίνεται ότι η ανθρώπινη τάση είναι να διατηρούμε τις πεποιθήσεις μας, φαινόμενο που απαντούμε στη βιβλιογραφία με τον όρο «doxastic conservativeness». Αυτό εκφράζεται με μια τάση είτε να εστιάζουμε σε γεγονότα που τις επιβεβαιώνουν είτε να παραβλέπουμε αυτά που τις διαψεύδουν. (Novaes et al., 2017) Μια πιθανή υπόθεση που μπορούμε να διατυπώσουμε για τη συμπεριφορά αυτή είναι η προσπάθεια διατήρησης της ομοιόστασης. Ας μην ξεχνάμε ότι ο άνθρωπος είναι ένα φυσικό σύστημα που τείνει να διατηρήσει την ισορροπία της κατάστασης στην οποία βρίσκεται. Αυτό όμως δεν αφορά μόνο στα φυσικά μεγέθη, όπως για παράδειγμα τη θερμοκρασία του. Ας σκεφτούμε ότι ένα άτομο είναι ένα φυσικό σύστημα που βρίσκεται σε ένα φυσικό περιβάλλον και συνεπώς αλληλεπιδρά με άλλα φυσικά συστήματα. Συνέπεια της αλληλεπίδρασης αυτής είναι να έχει το άτομο τη «θέση» του στο σύστημα μέσα στο οποίο ανήκει και μάλιστα να βρίσκεται σε ισορροπία με το σύστημα αυτό, γιατί, στην αντίθετη περίπτωση, το άτομο κινδυνεύει είτε να δεχθεί εχθρική συμπεριφορά από το σύστημα είτε να τεθεί εκτός συστήματος. Το να υιοθετήσει το άτομο ένα εξηγητικό πλαίσιο για το ευρύτερο σύστημα στο οποίο ανήκει δεν εξυπηρετεί μόνο τις δικές του λειτουργίες, αλλά και την αποκατάσταση της ισορροπίας με το περιβάλλον. Καταλαβαίνουμε, λοιπόν, ότι η διατήρηση των ατομικών πεποιθήσεων εξυπηρετεί ακριβώς τη διασφάλιση της ισορροπίας μεταξύ του ατόμου και των συστημάτων στα οποία ανήκει και, εν τέλει, την ισορροπία του ευρύτερου συστήματος.

## **Η προσέγγιση της Δυνητικής Λογικής**

Η ανάγκη να συνυπάρξουμε αρμονικά με τα υπόλοιπα φυσικά συστήματα με τα οποία μοιραία ερχόμαστε σε αλληλεπίδραση μας οδήγησε στην υιοθέτηση ενός νοητικού πλαισίου κατανόησης του εξωτερικού κόσμου. Οι πληροφορίες εσωτερικεύονται, οργανώνονται και συστηματοποιούνται και τελικά διαμορφώνουν τη γνώση μας για τον

κόσμο. Φυσικά, αυτός ο κόσμος έχει κανονικότητες αλλά είναι και διαρκώς μεταβαλλόμενος ταυτόχρονα. Συνεπώς, οι άνθρωποι, ως συστήματα, βομβαρδιζόμαστε διαρκώς από νέες πληροφορίες τις οποίες καλούμαστε να εντάξουμε στο ήδη υπάρχον σύστημα πεποιθήσεών μας και να το επικαιροποιήσουμε. Πολύ συχνά οι πληροφορίες αυτές είναι ελλιπείς, αντικρουόμενες ή ασαφείς και τελικά καλούμαστε να διαχειριστούμε την αβεβαιότητα που απορρέει από την έλλειψη πληροφοριών.

Κατανοούμε ότι η έλλειψη πληροφοριών στην Κλασική Λογική αποτυπώνεται με προκείμενες που είναι λιγότερες από τις απαιτούμενες ώστε να προκύψει ασφαλές συμπέρασμα. Στην καθημερινότητα τα πράγματα δεν είναι τόσο ξεκάθαρα. Συχνά συμβαίνει οι προκείμενες να μην έχουν καθολική ισχύ, να είναι αντικρουόμενες μεταξύ τους ή και να μην είναι ξεκάθαρες δίνοντάς μας τη «μισή αλήθεια», γιατί το σύστημα στο οποίο αναφέρονται δεν είναι κλειστό και ταυτόχρονα η απλή προσθήκη μιας επιπλέον πληροφορίας μπορεί να αλλάξει ουσιαστικά το συμπέρασμα. Σε ένα κλειστό σύστημα, η αλήθεια μιας ερμηνείας τυπικής γλώσσας θα πρέπει να παραμένει σταθερή προκειμένου αυτό να είναι συνεπές. Ταυτόχρονα, ακόμα και ο χώρος των δυνατών ενδεχομένων, ο κόσμος στον οποίο αναφερόμαστε, δεν μπορεί να προσδιοριστεί εύκολα σε ένα ανοιχτό σύστημα που δεν είναι ξεκάθαρο το ποια ακριβώς είναι τα δυνατά γεγονότα. Κατά συνέπεια, για να χειριστούμε την αβεβαιότητα, δεν επαρκεί ένα πιθανοτικό μοντέλο Μπεϋζιανής Στατιστικής γιατί η πραγματικότητά μας δεν είναι απόλυτα και αυστηρά ορισμένη. Η αλήθεια μιας πρότασης μπορεί να περιέχει ασάφεια (ποσοστό αλήθειας πρότασης) αλλά ταυτόχρονα η άρνηση μιας πρότασης είναι συχνό φαινόμενο να μη μας δίνει την πληροφορία αν ισχύει η αντίθετη πρόταση ή αν δεν ξέρουμε τι πραγματικά ισχύει.

Η Δυνητική Λογική είναι μια προσέγγιση που παρέχει ένα αρκετά επαρκές εξηγητικό πλαίσιο που μπορεί να ενσωματώσει όχι μόνο την αβεβαιότητα του συνόλου των καταστάσεων και τις πιθανές αντιφάσεις του συστήματος πεποιθήσεων, αλλά και, όπως θα δούμε παρακάτω, να προσεγγίσει ικανοποιητικά την παρουσία γνωστικών προκαταλήψεων κατά τη συλλογιστική διαδικασία, τη διαδικασία αναθεώρησης πεποιθήσεων και τις εκφράσεις της όπως είναι η προσπάθεια διατήρησης άθικτου του συστήματος πεποιθήσεων, οι γνωστικές προκαταλήψεις και τα συστηματικά λάθη που κάνουμε κατά τη διαδικασία συμπερασμού έχοντας την τάση να επαληθεύουμε

συμπεράσματα που προκύπτουν από λανθασμένη συλλογιστική αλλά είναι συνεπή με τις πεποιθήσεις μας (overgeneration) και να ακυρώνουμε συμπεράσματα που είναι μεν προϊόντα ορθού συμπερασμού αλλά δε συμμορφώνονται με τις πεποιθήσεις μας (undergeneration). (Dubois & Prade, 2014 και Novaes & Veluwenkamp, 2017)



## Σύντομη εισαγωγή στις έννοιες της Δυνητικής Λογικής

Η θεωρία δυνατοτήτων είναι μια θεωρία χειρισμού αβεβαιότητας που προέκυψε από την ανάγκη διαχείρισης ελλιπούς πληροφόρησης. Τις περισσότερες φορές οι γνώσεις μας για τον κόσμο και τα γεγονότα δεν είναι απόλυτες και μονοσήμαντα ορισμένες είτε επειδή έχουμε άγνοια είτε επειδή έχουμε αντικρουόμενες πληροφορίες. Η Μπεϋζιανή πιθανοτική προσέγγιση αλλά και η Κλασική Λογική έχουν αποδειχθεί χρήσιμα εργαλεία αναπαράστασης γνώσης, αλλά δε μας δίνουν τη δυνατότητα να αναπαραστήσουμε αυτήν την αβεβαιότητα. Για παράδειγμα, αν έχουμε ένα ενδεχόμενο  $A$  και το συμπληρωματικό του  $A'$  δεν μπορούμε να διαχωρίσουμε την περίπτωση κατά την οποία το  $A'$  συμβολίζει ότι ισχύει η άρνηση του  $A$  ή ότι δεν γνωρίζουμε αν ισχύει το  $A$ . Η ανάγκη αυτή οδήγησε στην προσπάθεια να προσεγγίσουμε την ανθρώπινη συλλογιστική μέσω μη μονότονων λογικών και μη δυϊκών τιμών (multi-valued logic) για τα ενδεχόμενα εισάγοντας έννοιες όπως ο βαθμός εμπιστοσύνης που ορίζει πόσο «εφικτό» είναι ένα γεγονός, τιμές όπως «άγνωστο» ή/και «αντιφατικό», εκτός από τις έννοιες «αληθές» και «ψευδές», για ένα γεγονός καθώς και έννοιες όπως η «πίστη» σε ένα γεγονός. (Dubois & Prade, 2018)

**Βασικές έννοιες της Θεωρίας Δυνατότητας (Possibility Theory)**, όπως διατυπώθηκαν από τους Dubois και Prade στο βιβλίο “Possibilistic Logic – An overview”:

Έστω  $\pi: U \rightarrow S$  ( $U$  το σύνολο των καταστάσεων και  $S = \{0 = \lambda_1, \lambda_2, \dots, \lambda_n = 1\}$ ) συνάρτηση κατανομής δυνατοτήτων που αναπαριστά την επιστημική κατάσταση (epistemic state) του υποκειμένου για την οποία ισχύει:

- $\pi(u) = 1$ , το  $u$  είναι απόλυτα δυνατό
- $\pi(u) = 0$ , το  $u$  είναι εντελώς αδύνατο

Δηλώνουμε:

- Απόλυτη γνώση (complete knowledge): για κάποιο  $u_0$ ,  $\pi(u_0) = 1$  and  $\pi(u) = 0$ ,  $\forall u \neq u_0$ , (μόνο το  $u_0$  είναι δυνατό)
- Απόλυτη άγνοια (complete ignorance):  $\pi(u) = 1 \forall u \in U$ , (όλες οι καταστάσεις είναι δυνατές)

Η συνάρτηση πιθανοκρατικά αναπαριστά την κατάσταση γνώσης του υποκειμένου σχετικά με το σύνολο των γεγονότων και διαχωρίζει τι από αυτά είναι εφικτό ή αναμενόμενο και τι όχι.

Ισχύει ότι περισσότερες από μία καταστάσεις μπορούν να είναι δυνατές ταυτόχρονα και δεν μπορούμε να εξαιρέσουμε κάποια οποιαδήποτε υπόθεση αν δεν είναι βέβαιο ότι είναι αδύνατη. Επίσης, όσο μικρότερος είναι ο βαθμός δυνατότητας, τόσο λιγότερες πληροφορίες έχουμε. Συνεπώς, ο στόχος είναι να προσπαθούμε να τον αυξήσουμε, πάντοτε δεδομένων των περιορισμών. Ταυτόχρονα, αν  $\pi(A)=0$ , τότε γνωρίζουμε ότι το  $A$  είναι αδύνατο, ενώ αν  $\pi(A)=1$ , γνωρίζουμε μόνο ότι το  $A$  είναι μια από τις πιο εφικτές ερμηνείες αλλά δεν έχουμε καμία πληροφορία αν είναι η σωστή ερμηνεία.

Στη Θεωρία Δυνατότητας (Possibility Theory) ισχύει η αρχή της ελάχιστης ιδιαιτερότητας (minimal specificity) σύμφωνα με την οποία οποιαδήποτε υπόθεση δεν είναι γνωστό ότι είναι αδύνατη δεν μπορεί να εξαιρεθεί.

**Θεμελιώδεις έννοιες** στη Δυνητική Λογική είναι τα μέτρα δυνατότητας και αναγκαιότητας για τα οποία ισχύει ότι όσο πιο σίγουρο είναι ένα γεγονός  $A$ , τόσο μικρότερη είναι η δυνατότητα του συμπληρωματικού του,  $A'$ .

- Μέτρο δυνατότητας (possibility measure):  $\Pi(A)=\sup_{u \in A} \pi(u)$ , όταν η κατανομή δυνατότητας ανήκει στο  $S=[0,1]$  και  $\Pi(A)=1$ ,  $A \cap E \neq \emptyset$  και ο διαφορετικά, για τη Boolean συνθήκη,
- Μέτρο αναγκαιότητας (necessity measure):  $N(A)=\inf_{u \notin A} (1 - \pi(u))$ , όταν η κατανομή δυνατότητας ανήκει στο  $S=[0,1]$  και  $N(A)=1$ ,  $E \subseteq A$  και ο διαφορετικά, για τη Boolean συνθήκη,

όπου  $A$  είναι ένα γεγονός (το σύνολο των ερμηνειών για τις οποίες αληθεύει η πρόταση  $\alpha$ ),  $E$  η επιστημική κατάσταση του υποκειμένου και  $U$  το σύνολο των καταστάσεων.

Ισχύουν οι σχέσεις  $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$  και  $N(A \cap B) = \min(N(A), N(B))$  και γενικά  $\Pi(U) = N(U) = 1$  and  $\Pi(\emptyset) = N(\emptyset) = 0$ .

Συνέπεια της κανονικοποίησης του  $\pi$  είναι ότι ισχύει  $\max(\Pi(A), \Pi(A'))=1$  και  $\min(N(A), N(A'))=0$  (όπου  $A'$  το συμπληρωματικό του  $A$ ) ή ισοδύναμα  $\Pi(A)=1$  όταν  $N(A)>0$ , το οποίο διαισθητικά σημαίνει ότι για να είναι κάτι βέβαιο πρέπει καταρχήν να είναι απολύτως δυνατό ή ότι κάποιος δεν μπορεί να είναι ταυτόχρονα βέβαιος και για το  $A$  και για το  $A'$  χωρίς να είναι ασυνεπής.

Ισχύει επίσης ότι  $N(A \cup B) \geq \max(N(A), N(B))$  το οποίο μας φαίνεται φυσικό αν σκεφτούμε ότι μπορεί να γνωρίζουμε τί ισχύει για το γεγονός  $A \cup B$ , αλλά όχι για τα  $A$  και  $B$  μεμονωμένα.

Για το μέτρο αναγκαιότητας ισχύει η σχέση:  $\forall p, N(p) = 1 - \Pi(\neg p)$ , δηλαδή, αυτό που αντιλαμβανόμαστε και διαισθητικά, είμαστε σίγουροι για κάτι όσο μικραίνει η δυνατότητα του αντιθέτου του.

Για να δηλώσουμε γνώση για το γεγονός  $A$  με κάποιο βαθμό σιγουριάς  $\alpha$ , πρέπει να ισχύει  $N(A) > \alpha$ . Η κατανομή δυνατότητας ορίζεται ως:

$$\pi_{(A, \alpha)}(u) = \begin{cases} 1, & \text{αν } u \in A \\ 1 - \alpha, & \text{διαφορετικά} \end{cases}$$

Αν  $\alpha=1$  παίρνουμε τη χαρακτηριστική συνάρτηση του  $A$ . Αν  $\alpha=0$  έχουμε απόλυτη άγνοια για το γεγονός.

Η Θεωρία Δυνατότητας μπορεί να είναι ποσοτική ή ποιοτική:

- Όταν είναι ποσοτική, τα μέτρα δυνατότητας και αναγκαιότητας είναι αντίστοιχα το πάνω και το κάτω όριο από ασθενώς γνωστές πιθανότητες ή/και ειδικές περιπτώσεις συναρτήσεων αληθοφάνειας (plausibility) ή πεποίθησης (belief).
- Όταν είναι ποιοτική, παρέχει μια φυσική προσέγγιση στη διαβάθμιση δυνατότητας και αναγκαιότητας σε πεπερασμένες κανονικές κλίμακες.

Δύο επιπλέον μέτρα της Θεωρίας Δυνατότητας είναι:

- Το μέτρο της εγγυημένης ή ισχυρής δυνατότητας (guaranteed or strong possibility measure):  $\Delta(A) = \inf_{u \in A} \pi(u)$ , που εκτιμά ως ποιο βαθμό όλες οι καταστάσεις του  $A$  είναι δυνατές δεδομένων των στοιχείων. Αβέβαιες δηλώσεις της μορφής «το  $A$  είναι δυνατό στο βαθμό  $\beta$ » μπορούν να αναπαρασταθούν από τη σχέση  $\Delta(A) \geq \beta$ .

- Το μέτρο της ασθενούς αναγκαιότητας (weak necessity measure):  $\nabla(A) = 1 - \Delta(A) = \inf_{u \in A} (1 - \pi(u))$ , που εκτιμά το βαθμό της ασθενούς αναγκαιότητας του A και είναι 1 μόνο αν κάποια κατάσταση s του A είναι αδύνατη.

Για τα δύο πρόσθετα μέτρα ισχύουν οι σχέσεις  $\Delta(A \cup B) = \min(\Delta(A), \Delta(B))$  και  $\nabla(A \cap B) = \max(\nabla(A), \nabla(B))$ .

Μια βασική φόρμουλα της Δυνητικής Λογικής είναι το ζεύγος  $(a, \alpha)$ , όπου a είναι μια καλά ορισμένη φόρμουλα της Κλασικής Λογικής εφοδιασμένη με ένα μέτρο βεβαιότητας  $\alpha \in (0, 1]$  και το οποίο προσδιορίζει το κάτω όριο του επιπέδου αναγκαιότητας, δηλαδή  $N(a) > \alpha$ .

Θεωρούμε την Προτασιακή Δυνητική Λογική PL ως ένα σύνολο από φόρμουλες της μορφής  $(a, \alpha)$ , όπου a είναι μια φόρμουλα προτασιακής γλώσσας. Τα αξιώματα της PL είναι τα ίδια με εκείνα της Προτασιακής Λογικής PL.

Μια βάση  $\Gamma = \{(a_i, \alpha_i), i=1, \dots, m\}$  της Δυνητικής Λογικής (δηλαδή ένα σύνολο από φόρμουλες) έχει ένα **επίπεδο ασυνέπειας** (inconsistency level)  $inc-l(\Gamma)$  τέτοιο ώστε ένα υποσύνολο από φόρμουλες με επίπεδο βεβαιότητας αυστηρά μεγαλύτερο να ορίζουν ένα συνεπές υποσύνολο και ισχύει ότι:

$$inc-l(\Gamma) = \max \{ \alpha \mid \Gamma \vdash (\perp, \alpha) \}.$$

Μια δυνητική λογική βάση είναι σημασιολογικά ισοδύναμη με μια κατανομή δυνατότητας που περιορίζει το σύνολο των ερμηνειών ώστε να είναι συμβατές με τη βάση. Ενδιαφέρον παρουσιάζει το γεγονός ότι έχουμε ένα ασαφές σύνολο μοντέλων, εν αντιθέσει με την Κλασική Λογική, και μια παράβαση της ερμηνείας μιας φόρμουλας που δεν απόλυτα βέβαια δεν μπορεί να οδηγήσει σε εξαίρεση της ερμηνείας.

Οι δυνητικές φόρμουλες των οποίων το επίπεδο βεβαιότητας είναι μεγαλύτερο από  $inc-l(\Gamma)$  δημιουργούν ένα συνεπές σύνολο,  $inc-l(\{(a_i, \alpha_i) \mid (a_i, \alpha_i) \in \Gamma \text{ και } \alpha_i > inc-l(\Gamma)\}) = 0$  και  $\Gamma(\alpha)^*$  είναι συνεπές.

### Βασικοί κανόνες συμπερασμού της Δυνητικής Λογικής:

- Αν  $\beta \leq \alpha$ , τότε  $(a, \alpha) \vdash (b, \beta)$ , (certainty weakening)
- $(\neg a \vee b, \alpha), (a, \alpha) \vdash (b, \alpha), \forall \alpha \in (0, 1]$ , (modus ponens)

Και δευτερεύοντες:

- $(\neg a \vee b, \alpha), (a \vee c, \alpha) \vdash (b \vee c, \alpha), \forall \alpha \in (0, 1]$ , (resolution)
- $(\neg a \vee b, \alpha), (a \vee c, \beta) \vdash (b \vee c, \min(\alpha, \beta))$ , (weakest link resolution)
- Αν  $a \vdash b$  τότε  $(a, \alpha) \vdash (b, \alpha), \forall \alpha \in (0, 1]$ , (formula weakening)

Η **σημασιολογία** μιας φόρμουλας  $(a, \alpha)$  στην Προτασιακή Δυνητική Λογική δίνεται από τη συνάρτηση κατανομής δυνατότητας  $\pi_{(a, \alpha)}$ :

$$\pi_{(a, \alpha)}(\omega) = 1 \text{ αν } \omega \models a \text{ και } \pi_{(a, \alpha)}(\omega) = 1 - \alpha \text{ αν } \omega \models \neg a$$

το οποίο διαισθητικά σημαίνει ότι οποιοδήποτε μοντέλο του  $a$  πρέπει να είναι απόλυτα δυνατό ενώ οποιοδήποτε αντι-μοντέλο (counter-model) είναι τόσο αδύνατο όσο μεγαλύτερο είναι το επίπεδο βεβαιότητας  $\alpha$ .

Μια βάση γνώσης  $\Gamma$  της ΠΛ τέτοια ώστε  $\Gamma = \{(a_i, \alpha_i), i=1, \dots, m\}$  συνδέεται σημασιολογικά με τη συνάρτηση κατανομής δυνατότητας:

$$\pi_{\Gamma}(\omega) = \min_{i=1, \dots, m} \pi_{(a_i, \alpha_i)}(\omega) = \min_{i=1, \dots, m} \max([a_i](\omega), 1 - \alpha_i),$$

όπου  $[a_i]$  η χαρακτηριστική συνάρτηση των μοντέλων  $a_i$ , για την οποία ισχύει ότι  $[a_i](\omega) = 1$  αν  $\omega \models a_i$  και  $[a_i](\omega) = 0$  διαφορετικά.

Συνεπώς, μια βάση γνώσης είναι ένα σύνολο από περιορισμούς  $N(a_i) \geq \alpha_i$  και το σύνολο με τις συνεπαγόμενες κατανομές δυνατότητας  $\pi$  που σχετίζονται με το  $N$  και είναι συμβατές με τους περιορισμούς έχει ένα μέγιστο στοιχείο  $\pi_{\Gamma}$  για το οποίο ισχύει ότι  $\forall \omega, \pi(\omega) < \min_{i=1, \dots, m} \pi_{(a_i, \alpha_i)} = \pi_{\Gamma}(\omega)$ . Άρα κάθε συνάρτηση  $\pi$  συμφωνεί σημασιολογικά με τη βάση  $\Gamma$  αν ισχύει  $\pi \leq \pi_{\Gamma}$  και γράφουμε  $\pi \models \Gamma$ .

Η σημασιολογική συνεπαγωγή ορίζεται από τη σχέση:

$\Gamma \models (a, a)$  αν και μόνο αν  $\forall \omega \pi_{\Gamma}(\omega) \leq \pi_{\{(a, a)\}}(\omega)$

και ταυτόχρονα ισχύει ότι  $inc-l(\Gamma) = 1 - \max_{\omega \in \Omega} \pi_{\Gamma}(\omega)$ .

Για τη δεσμευμένη δυνατότητα ισχύει εξίσωση παρόμοια με τη Μπεϋζιανή που έχει τη μορφή:

$$\Pi(B \cap A) = \Pi(B|A) * \Pi(A)$$

όπου  $\Pi(A) > 0$  και  $*$  μπορεί να είναι είτε το γινόμενο είτε η συνάρτηση ελαχίστου. Στην αριθμητική συνθήκη  $*$  είναι το γινόμενο, που διατηρεί τη συνέχεια, έτσι ώστε  $\Pi(B|A) = \frac{\Pi(B \cap A)}{\Pi(A)}$  που μοιάζει με την αντίστοιχη πιθανοτική σχέση.

Για το μέτρο αναγκαιότητας στην υπό όρους συλλογιστική ισχύει  $N(B|A) = 1 - \Pi(B'|A)$ , όπου  $B'$  το συμπληρωματικό του  $B$ .

Για δύο γεγονότα  $A$  και  $B$  ισχύει ότι  $\Pi(A \cap B) = \min(\Pi(A), \Pi(B))$ . Η σχέση είναι συμμετρική αλλά ευαίσθητη στην άρνηση. Όταν δεν ισχύει αυτή η σχέση (ασυσχέτιστα γεγονότα) δηλώνει την επιστημική κατάσταση κατά την οποία τα δύο γεγονότα είναι αμοιβαία αποκλειόμενα ενώ, όταν ισχύει για δύο οποιαδήποτε ζεύγη  $A$  και  $B$  και τα συμπληρωματικά τους, δηλώνει την κατάσταση λογικής ανεξαρτησίας.

Για δύο ανεξάρτητα γεγονότα  $A$  και  $B$  ισχύει ότι  $\Pi(B|A) = \Pi(B)$ . Η σχέση είναι πιο ισχυρή από την προηγούμενη, δεν είναι συμμετρική και είναι ευαίσθητη στην άρνηση. Επίσης, δεν ισχύει ότι  $N(B|A) = N(B)$ . Δηλώνει ότι η γνώση για το  $A$  δεν επηρεάζει την αληθοφάνεια του  $B$ .

## Σύντομη εισαγωγή στη Θεωρία Αναθεώρησης Πεποιθήσεων (Belief Revision)

Δρώντας σε ένα δυναμικό περιβάλλον υπάρχουν κυρίως δύο λόγοι για τους οποίους κάποιος πρέπει να τροποποιήσει το σύνολο των πεποιθήσεών του: είτε γιατί αλλάζει ο κόσμος είτε γιατί αλλάζει η γνώση του για τον κόσμο. Αυτό σημαίνει ότι η τρέχουσα κατάσταση πεποιθήσεων πρέπει να αναθεωρηθεί, προσθέτοντας και αφαιρώντας νέα δεδομένα. Η διαδικασία της αναθεώρησης δεν είναι απλή, γιατί οι πεποιθήσεις διέπονται από σχέσεις πιθανής εξάρτησης μεταξύ τους, οπότε, αφαιρώντας μία πεποίθηση πρέπει να αφαιρεθούν και όσες σχετίζονται με αυτήν, έτσι ώστε να διατηρηθεί η συνοχή του συνόλου. Επίσης, η προσθήκη μιας νέας πληροφορίας στο σύνολο επιφέρει την προσθήκη και όλων εκείνων που σχετίζονται με αυτή.

Στη θεωρία Αναθεώρησης Πεποιθήσεων ένας πιθανός κόσμος είναι ένα μέγιστο συνεπές υποσύνολο της γλώσσας και μία πρόταση (proposition) είναι ένα σύνολο πιθανών κόσμων. Υπάρχει μία ένα-προς-ένα αντιστοιχία μεταξύ προτάσεων και συνόλων πεποιθήσεων. Κατά τον Isaac Levi ένα σύνολο πεποιθήσεων (belief set) αποτελείται από τις πεποιθήσεις εκείνες που δεσμεύεται να πιστεύει κάποιος. (Levi, 1991) Αυτό σημαίνει ότι αν αναπαραστήσουμε τις πεποιθήσεις με προτάσεις σε τυπική γλώσσα και με τους συνήθεις λογικούς συνδέσμους ( $\neg$ ,  $\&$ ,  $\vee$ ,  $\rightarrow$ ,  $\leftrightarrow$ ,  $\perp$ ,  $\top$ ), τότε το σύνολο πεποιθήσεων ορίζεται από όλες τις δυνατές παραγόμενες προτάσεις, ακόμα και αν αυτές δεν έχουν νόημα για μια πιθανή ερμηνεία. Για παράδειγμα, η πεποίθηση ότι «το πιάνο είναι μουσικό όργανο» συνεπάγεται ότι στο σύνολο πεποιθήσεων υπάρχει και η πεποίθηση ότι «το πιάνο δεν είναι όργανο γυμναστικής».

Στην πραγματικότητα αυτό δε φαίνεται να συμβαίνει – το σύνολο των πεποιθήσεων που δεσμεύεται να πιστεύει το άτομο είναι πολύ ευρύτερο από το σύνολο των πραγματικών του πεποιθήσεων. Η ανθρώπινη νόηση προσεγγίζεται καλύτερα από τις βάσεις πεποιθήσεων (belief bases) που αντιπροσωπεύουν επιστημικές καταστάσεις του ατόμου χωρίς να είναι λογικά κλειστές (εκτός από την οριακή κατάσταση), δηλαδή χωρίς να δεσμεύουν το άτομο σε όλες τις λογικές συνέπειες των προτάσεων, και χωρίς να είναι κατ' ανάγκη συνεπείς. Συνεπώς, ενώ υπάρχει μόνο ένα ασυνεπές σύνολο πεποιθήσεων

(ολόκληρη η γλώσσα που σχηματίζουν οι προτάσεις) υπάρχουν πολλές ασυνεπείς βάσεις πεποιθήσεων που προσεγγίζουν επιστημικές καταστάσεις του ατόμου για τις οποίες δεν υπάρχει απαίτηση συνέπειας μεταξύ τους. (Stanford Encyclopedia)

Η αναθεώρηση πεποιθήσεων είναι παρόμοια για τα σύνολα και για τις βάσεις. Ακολουθούν κάποιοι κανόνες όπως καταγράφονται στην ηλεκτρονική έκδοση της εγκυκλοπαίδειας του Stanford και στο άρθρο των Gärdenfors και Rott (1995). Για λόγους ακρίβειας έχει διατηρηθεί η αγγλική ορολογία.

### **Βασικά στοιχεία της Αναθεώρησης Πεποιθήσεων:**

Για ένα σύνολο  $A$  ισχύουν τα εξής:

*Inclusion:*

$$A \subseteq Cn(A)$$

*Monotony:*

$$\text{Αν } A \subseteq B, \text{ τότε } Cn(A) \subseteq Cn(B)$$

*Iteration:*

$$Cn(A) = Cn(Cn(A))$$

Όπου  $Cn$  είναι το σύνολο των λογικών συνεπειών των προτάσεων της γλώσσας, δηλαδή  $Cn(A)$  είναι το σύνολο των λογικών συνεπειών των προτάσεων του συνόλου  $A$ .

Το  $A$  είναι ένα σύνολο πεποιθήσεων αν ισχύει  $A = Cn(A)$  και  $Cn(\emptyset)$  είναι το σύνολο των ταυτολογιών.

Αν επεκτείνουμε ένα σύνολο  $A$  κατά μία πρόταση  $p$  χωρίς να προσθέσουμε τίποτε άλλο, γράφουμε:  $A+p = Cn(A \cup \{p\})$ .

Για οποιοδήποτε σύνολο  $A$  και πρόταση  $p$  ορίζουμε ως υπόλοιπο (remainder) το σύνολο των μέγιστων υποσυνόλων του  $A$  που δε συνεπάγονται την  $p$  και γράφουμε  $A \perp p$ . Δηλαδή, το σύνολο  $B$  είναι στοιχείο του  $A \perp p$  αν και μόνο αν δε συνεπάγεται την  $p$  και δεν υπάρχει  $B'$  τέτοιο ώστε  $B \subset B' \subseteq A$ .



Η **συστολή (contraction)** ενός συνόλου  $K$  από το στοιχείο  $p$  είναι ένα υποσύνολο του  $K$  που δε συνεπάγεται το  $p$ , χωρίς να έχουν γίνει περιττές αφαιρέσεις στοιχείων και μειώνοντας ταυτόχρονα την απώλεια πληροφορίας. Η **μερική συστολή (partial meet contraction)**  $K \dot{\div} p$  προκύπτει με ασφάλεια από την τομή των υπολοίπων του  $K$ .

Ισχύουν οι σχέσεις:

*Closure:*

$$K \dot{\div} p = \text{Cn}(K \dot{\div} p)$$

*Success:*

$$\text{Αν } p \notin \text{Cn}(\emptyset), \text{ τότε } p \notin \text{Cn}(K \dot{\div} p)$$

*Inclusion:*

$$K \dot{\div} p \subseteq K$$

*Vacuity:*

$$\text{Αν } p \notin \text{Cn}(K), \text{ τότε } K \dot{\div} p = K.$$

*Extensionality:*

$$\text{Αν } p \leftrightarrow q \in \text{Cn}(\emptyset), \text{ τότε } K \dot{\div} p = K \dot{\div} q.$$

*Recovery:*

$$K \subseteq (K \dot{\div} p) + p$$

Στην περίπτωση που το  $K$  πρέπει να συσταλεί κατά τη σύζευξη δύο προτάσεων  $p \& q$ , το άτομο καλείται να αφαιρέσει είτε την  $p$  είτε την  $q$  είτε και τις δύο. Αν αφαιρέσει μόνο τη μία, έστω την  $p$ , θα πρέπει να αφαιρέσει και όσες προτάσεις συνδέονται λογικά με αυτήν.

Ισχύει ότι:

*Conjunctive inclusion:*

$$\text{Αν } p \notin K \dot{\div} (p \& q), \text{ τότε } K \dot{\div} (p \& q) \subseteq K \dot{\div} p.$$

Ισχύει επίσης πως ό,τι είναι ταυτόχρονα στοιχείο του  $K \div p$  και του  $K \div q$  είναι και του  $K \div (p \& q)$ :

*Conjunctive overlap:*

$$(K \div p) \cap (K \div q) \subseteq K \div (p \& q)$$

Κατά την αναθεώρηση πεποιθήσεων δεν είναι όλες οι πεποιθήσεις εξίσου ισοδύναμες υπό αναθεώρηση και συνεπώς κάποιες είναι πιο οχυρωμένες από άλλες. Για παράδειγμα, οι φυσικοί νόμοι έχουν υψηλή **επιστημική οχύρωση** και δεν είναι σε προτεραιότητα αναθεώρησης.

Ισχύουν οι σχέσεις:

$p \leq q$ :  $p$  είναι ίσο ή λιγότερο οχυρωμένο από το  $q$ .

$p < q$ :  $p$  είναι λιγότερο οχυρωμένο από το  $q$ .  $((p \leq q) \& \neg (q \leq p))$

$p \equiv q$ :  $p$  και το  $q$  είναι το ίδιο οχυρωμένα.  $((p \leq q) \& (q \leq p))$

Σύμφωνα με τον Gärdenfors για την επιστημική οχύρωση ισχύουν οι εξής απαιτήσεις:

*Transitivity:*

Αν  $p \leq q$  και  $q \leq r$ , τότε  $p \leq r$ .

*Dominance:*

Αν  $p \vdash q$ , τότε  $p \leq q$ .

*Conjunctiveness:*

Είτε  $p \leq (p \& q)$  είτε  $q \leq (p \& q)$ .

*Minimality:*

Αν το σύνολο πεποιθήσεων  $K$  είναι συνεπές, τότε  $p \notin K$  αν και μόνο αν  $p \leq q$  για όλα τα  $q$ .

*Maximality:*

Αν  $q \leq p$  για όλα τα  $q$ , τότε  $p \in Cn(\emptyset)$ .

Η σχέση οχύρωσης οδηγεί σε συστολή συνόλου σύμφωνα με τη σχέση:  
 $q \in K \div p$  αν και μόνο αν  $q \in K$  και είτε  $p < (p \vee q)$  είτε  $p \in Cn(\emptyset)$ .

Η **αναθεώρηση** (revision) πεποιθήσεων, που συμβολίζεται με τον τελεστή  $*$ , επιτελεί δύο βασικές λειτουργίες: προσθέτει την πεποίθηση  $p$  στο σύνολο  $K$  και διασφαλίζει ότι το παραγόμενο σύνολο  $K * p$  είναι συνεπές, υπό την προϋπόθεση ότι η  $p$  δεν είναι ασυνεπής. Αν ένα σύνολο δε συνεπάγεται το  $\neg p$ , τότε η  $p$  μπορεί να προστεθεί χωρίς βλάβη της συνέπειάς του.

Ισχύει η ταυτότητα του Levi:  $K * p = (K \div \neg p) + p$ .

Αν η  $K \div \neg p$  είναι μερική συστολή, τότε η  $K * p$  είναι **μερική αναθεώρηση** (partial meet revision).

Μια πράξη  $*$  είναι μερική αναθεώρηση πεποιθήσεων αν ισχύουν οι ακόλουθες *βασικές απαιτήσεις του Gärdenfors* για τη μερική αναθεώρηση:

*Closure:*

$$K * p = Cn(K * p)$$

*Success:*

$$p \in K * p$$

*Inclusion:*

$$K * p \subseteq K + p$$

*Vacuity:*

Αν  $\neg p \notin K$ , τότε  $K * p = K + p$ .

*Consistency:*

$K * p$  είναι συνεπές αν  $p$  είναι συνεπές.

*Extensionality:*

Αν  $(p \leftrightarrow q) \in Cn(\emptyset)$ , τότε  $K^*p = K^*q$ .

Ισχύουν επίσης και δύο συμπληρωματικές:

*Superexpansion:*

$K^*(p \& q) \subseteq (K^*p) + q$

*Subexpansion:*

Αν  $\neg q \notin Cn(K^*p)$ , τότε  $(K^*p) + q \subseteq K^*(p \& q)$ .

Ακολουθούν οι αντίστοιχοι κανόνες για τις βάσεις πεποιθήσεων όπως αναφέρονται στην ηλεκτρονική έκδοση της εγκυκλοπαίδειας του Stanford και στο άρθρο των Gärdenfors και Rott (1995).

Μια πράξη  $\div$  είναι μια μερική συστολή ενός συνόλου  $A$  αν ικανοποιούνται οι ακόλουθες απαιτήσεις:

*Success:*

Αν  $p \notin Cn(\emptyset)$ , τότε  $p \notin Cn(A \div p)$ .

*Inclusion:*

$A \div p \subseteq A$

*Relevance:*

Αν  $q \in A$  και  $q \notin A \div p$ , τότε υπάρχει σύνολο  $A'$  τέτοιο ώστε  $A \div p \subseteq A' \subseteq A$  και  $p \notin Cn(A')$  αλλά  $p \in Cn(A' \cup \{q\})$ .

*Uniformity:*

Αν για όλα τα υποσύνολα  $A'$  του  $A$  ισχύει ότι  $p \in Cn(A')$  αν και μόνο αν  $q \in Cn(A')$ , τότε  $A \div p = A \div q$ .

Για τις βάσεις πεποιθήσεων υπάρχουν δύο διακριτοί τρόποι αναθεώρησης από τη συστολή και την επέκταση:

*Internal revision:*

$$A * p = (A \div \neg p) + 'p$$

*External revision:*

$$A * p = (A + 'p) \div \neg p$$

Διαισθητικά η εξωτερική αναθεώρηση (external revision) περιλαμβάνει ένα ενδιάμεσο ασυνεπές βήμα κατά το οποίο και η  $p$  και η  $\neg p$  είναι ταυτόχρονα πιστευτές. Αντίστοιχα, η εσωτερική αναθεώρηση (internal revision) περιλαμβάνει ένα ενδιάμεσο μη δεσμευτικό βήμα κατά το οποίο καμία από τις  $p$  και  $\neg p$  δεν είναι πιστευτή.

Η συστολή που πραγματοποιείται σε μια βάση πεποιθήσεων προκαλεί μια επαγόμενη συστολή στο αντίστοιχο σύνολο πεποιθήσεων. Έστω  $A$  μια βάση πεποιθήσεων και  $-$  μια πράξη συστολής σε αυτήν και  $K = Cn(A)$  το αντίστοιχο σύνολο πεποιθήσεων με  $\div$  την παραγόμενη από τη βάση πράξη συστολής έτσι ώστε για όλες τις προτάσεις  $p$  να ισχύει  $K \div p = Cn(A - p)$ . Θα πρέπει να ισχύουν οι εξής απαιτήσεις:

*Closure:*

$$K \div p = Cn(K \div p)$$

*Success:*

$$\text{Αν } p \notin Cn(\emptyset), \text{ τότε } p \notin Cn(K \div p).$$

*Inclusion:*

$$K \div p \subseteq K$$

*Vacuity:*

$$\text{Αν } p \notin Cn(K), \text{ τότε } K \div p = K.$$

*Extensionality:*

$$\text{Αν } p \leftrightarrow q \in Cn(\emptyset), \text{ τότε } K \div p = K \div q.$$

*Finitude:*

Υπάρχει ένα ορισμένο σύνολο  $X$  τέτοιο ώστε για κάθε πρόταση  $p$ ,  $K \div p = Cn(X')$  για κάποιο  $X' \subseteq X$ .

*Symmetry:*

Αν ισχύει για όλα τα  $r$  ότι  $K \div r \vdash p$  αν και μόνο αν  $K \div r \vdash q$ , τότε  $K \div p = K \div q$ .

*Conservativity:*

Αν  $K \div q$  δεν είναι υποσύνολο του  $K \div p$ , τότε υπάρχει κάποιο  $r$  τέτοιο ώστε  $K \div p \subseteq K \div r \neq p$  και  $K \div r \cup K \div q \vdash p$ .

Αν μια βάση είναι ασυνεπής, μπορεί να γίνει συνεπής αν αφαιρέσουμε ένα ή περισσότερα περιττά στοιχεία. Η διαδικασία αυτή ονομάζεται **consolidation**, συμβολίζεται  $A!$  και μπορεί να επιτευχθεί με συστολή της βάσης κατά τις αντιφάσεις, δηλαδή  $A! = A \div \perp$ .

Στην αλλαγή πεποιθήσεων χωρίς προτεραιότητα, η νέα πληροφορία αξιολογείται σε σχέση με τις παλιές χωρίς να έχει κάποια προτεραιότητα λόγω καινοτομίας. Η διαδικασία αυτή ονομάζεται **ημι-αναθεώρηση (semi-revision)**. Η ημι-αναθεώρηση ενός συνόλου  $K$  κατά την πρόταση  $p$  συμβολίζεται  $K \circ p$ . Αν η νέα πληροφορία έρχεται σε αντίφαση με κάποια από τις ήδη υπάρχουσες, αξιολογείται η επιστημική της αξία ώστε να διαγραφεί είτε η ίδια είτε εκείνες που έρχονται σε αντίφαση με την  $p$ .

Η **εκλεκτική αναθεώρηση (selective revision)** είναι μια ειδική μορφή της ημι-αναθεώρησης. Στην ημι-αναθεώρηση η νέα πληροφορία είτε γίνεται αποδεκτή είτε απορρίπτεται εξ ολοκλήρου. Στην εκλεκτική αναθεώρηση μπορεί να γίνει αποδεκτό μόνο ένα μέρος της νέας πληροφορίας. Ισχύει ότι  $K \circ p = K * f(p)$ , όπου  $\circ$  είναι η εκλεκτική αναθεώρηση,  $*$  η τυπική αναθεώρηση και  $f(p)$  συνάρτηση που δεν περιέχει πληροφορίες που δεν υπάρχουν στην  $p$  διασφαλίζοντας έτσι ότι είναι λογική συνέπεια της  $p$ .

Θεωρητικά, στη συστολή όλες οι πεποιθήσεις που δεν είναι ταυτολογικές θα πρέπει να είναι αναιρέσιμες. Στην πραγματικότητα αυτό δεν ισχύει και βλέπουμε ότι υπάρχουν μη λογικές πεποιθήσεις που όμως δεν μπορούν να αφαιρεθούν. Η **θωρακισμένη συστολή (shielded contraction)** βασίζεται στην τυπική συστολή  $\div$  και ένα σύνολο  $R$  αναιρέσιμων προτάσεων έτσι ώστε: αν  $p \in R$ , τότε  $K - p = K \div p$ , αλλιώς,  $K - p = K$ .

Με την **αντικατάσταση (replacement)** επιτυγχάνονται δύο διαδικασίες – η αφαίρεση της πρότασης  $p$  από ένα σύνολο πεποιθήσεων και η προσθήκη της πρότασης  $q$ . Συνεπώς, το σύνολο  $K[p/q]$  περιλαμβάνει την  $q$  αλλά όχι την  $p$  και ισχύουν:

*Contractive success:*

Αν  $p \notin Cn(\{q\})$ , τότε  $p \notin Cn(K[p/q])$ .

*Revision success:*

$q \in K[p/q]$ .

Αν ένα σύνολο  $K$  έχει μια πεπερασμένη βάση, τότε μπορεί να αναπαρασταθεί από μία πρόταση  $k$  η οποία είναι η σύζευξη όλων των στοιχείων της βάσης του. Συνεπώς, η αναθεώρηση  $K * p$  μπορεί να αντικατασταθεί από την  $k * p$ . Η διαδικασία ονομάζεται **συγχώνευση (merging)** δύο αναπαραστάσεων πεποιθήσεων και μπορεί να χρησιμοποιηθεί στη διευθέτηση σύγκρουσης μέσω συνδυασμού στην περίπτωση που υπάρχουν πληροφορίες από δύο διαφορετικές πηγές.

Με τον όρο **πολλαπλή συστολή (multiple contraction)** εννοούμε τη συστολή περισσότερων από μία προτάσεων ταυτόχρονα. Υπάρχουν δύο τύποι πολλαπλής συστολής, η **package contraction**, όπου όλα τα εισερχόμενα στοιχεία αναιρούνται μαζί σαν πακέτο, και η **choice contraction**, όπου αρκεί να αναιρεθεί τουλάχιστον ένα από τα εισερχόμενα στοιχεία.

Ισχύουν:

*Package success:*

Αν  $B \cap Cn(\emptyset) = \emptyset$  τότε  $B \cap Cn(A \div B) = \emptyset$

*Choice success:*

Αν  $B$  δεν είναι υποσύνολο του  $Cn(\emptyset)$ , τότε το  $B$  δεν είναι υποσύνολο του  $Cn(A \div B)$ .

*Negation of a set:*

$p \in \neg B$  αν και μόνο αν  $p$  είναι είτε άρνηση ενός στοιχείου του  $B \cup \{T\}$  είτε διάζευξη αρνήσεων στοιχείων του  $B \cup \{T\}$ .

Η πολλαπλή αναθεώρηση μπορεί να οριστεί από την package multiple contraction και την επέκταση μέσω της γενικευμένης εξίσωσης του Levi:

$$K*B=(K\div\neg B)+B$$



## **Οι συλλογιστικές προκαταλήψεις μέσα από την έρευνα των Novaes και Veluwenkamp**

Η έρευνα των Novaes και Veluwenkamp εστιάζει στη μελέτη των προκαταλήψεων, ενός φαινομένου που έχει απασχολήσει αρκετά την Ψυχολογία και τη διερεύνηση της ανθρώπινης συλλογιστικής. Η προσέγγισή τους γίνεται χρησιμοποιώντας δύο μοντέλα, το ένα από τη σκοπιά της μη μονότονης Προτιμησιακής Λογικής (Preferential Logic) και το άλλο από τη σκοπιά της Αναθεώρησης Πεποιθήσεων (Belief Revision).

Για να διερευνήσουμε αν είναι ακριβή τα (τυπικά) μοντέλα που έχουν κατά καιρούς προταθεί για να εξηγήσουν/αναπαραστήσουν την ανθρώπινη συλλογιστική (human reasoning) πρέπει να απομονωθούν κάποιες βασικές αρχές της λειτουργίας του συστήματος (ανθρώπινος νους) και να σχεδιαστούν πειράματα που θα μετρούν σε ποιο βαθμό οι λειτουργίες του συστήματος προβλέπονται από το μοντέλο. Ένας άλλος τρόπος είναι να σχεδιαστούν μοντέλα και να διερευνηθεί αν και κατά πόσο τα αποτελέσματα της μοντελοποίησης είναι συμβατά με πειραματικά δεδομένα του πραγματικού συστήματος. Στην παρούσα έρευνα χρησιμοποιείται η δεύτερη προσέγγιση. Η βασική πειραματική διαδικασία που ακολουθήθηκε ήταν η αξιολόγηση ενός συνόλου συλλογισμών ως προς την εγκυρότητα και την ορθότητά τους.

Η κύρια παρατήρηση που προκύπτει από την παρούσα έρευνα είναι η ανθρώπινη τάση να συλλογιζόμαστε προς την κατεύθυνση διατήρησης των πεποιθήσεών μας (doxastic conservatism). Αυτό σημαίνει ότι απορρίπτουμε τις εισερχόμενες πληροφορίες που δεν είναι συμβατές με τα πιστεύω μας και πολύ συχνά το κατά πόσο είναι πιστευτό ή όχι ένα συμπέρασμα μάς επηρεάζει στην κρίση μας για την ορθότητα των προκειμένων (belief bias). Αυτό συχνά οδηγεί σε συστηματικά σφάλματα κατά τη διαδικασία συμπερασμού. Από τα πειράματα φαίνεται ότι οι συμμετέχοντες συχνά δεν πραγματοποιούν παραγωγή (deduction) αλλά μια αναιρέσιμη συλλογιστική (defeasible reasoning) και κάποιες φορές φαίνεται ότι προβαίνουν σε κάποιου είδους διαχείριση πεποιθήσεων (belief management) έτσι ώστε να αξιολογήσουν την εισερχόμενη πληροφορία σε σχέση με το αν συμφωνεί με τις ήδη υπάρχουσες πεποιθήσεις τους και

ταυτόχρονα σε σχέση με την αξιοπιστία της πηγής προκειμένου να αποφασίσουν τελικά αν θα επικυρώσουν ή όχι το συλλογισμό.

Αρχικά η έννοια της προκατάληψης ταυτιζόταν με την απόκλιση από την Κλασική Λογική. Όμως, στην πορεία, η Κλασική Λογική άρχισε να τίθεται υπό αμφισβήτηση ως ο μόνος τρόπος αναπαράστασης της συλλογιστικής και η έννοια των προκαταλήψεων ως «συλλογιστικά σφάλματα» αμφισβητήθηκε επίσης. Όπως αναφέρουν οι Novaes και Veluwenkamp, κατά τον Klayman (1995) μια προκατάληψη μπορεί να έχει (τουλάχιστον) τρεις εκδοχές:

- Τάση ή κλίση (ουδέτερη)
- Τάση ελαττωματικού συμπερασμού (αρνητική)
- Περιορισμένος ορθολογισμός: οι άνθρωποι μπορούν να αποκλίνουν συστηματικά από τα θεωρητικά πρότυπα, αλλά συμπεριφέρονται βέλτιστα σε ευρύτερο πλαίσιο, λαμβάνοντας υπ' όψιν τους περιορισμούς. (Novaes et al., 2016)

Από τα πειράματα φαίνεται απόκλιση από την παραγωγική διαδικασία συμπερασμού και μια τάση επαλήθευσης και διατήρησης των πεποιθήσεων. Πιο συγκεκριμένα, τα πιστευτά συμπεράσματα φαίνεται ότι γίνονταν περισσότερο αποδεκτά από τους συμμετέχοντες σε σχέση με τα μη πιστευτά, ανεξάρτητα από την εγκυρότητα του συμπερασμού. Το φαινόμενο γίνεται εντονότερο όταν στο συμπέρασμα χρησιμοποιούνται λέξεις της φυσικής γλώσσας. Αντίθετα, όταν χρησιμοποιούνται κατασκευασμένες λέξεις (χωρίς νόημα για τη φυσική γλώσσα) ακόμα και με την ίδια σύνταξη προτάσεων, οι συμμετέχοντες φαίνεται να τα πηγαίνουν καλύτερα στη συλλογιστική και να είναι σε θέση να παραγάγουν ορθά συμπεράσματα.

### **Τα φαινόμενα overgeneration και undergeneration ως συνέπειες των προκαταλήψεων**

Υπάρχουν δύο φαινόμενα που επιβεβαιώνονται ισχυρά από τα πειραματικά δεδομένα στην έρευνα των Novaes και Veluwenkamp – το overgeneration και το undergeneration.

**Overgeneration** παρατηρείται όταν το συμπέρασμα είναι αποδεκτό χωρίς όμως να προκύπτει αυστηρά και με λογική συνέπεια από τις προκειμένες. Ο λόγος που γίνεται αποδεκτό είναι επειδή φαίνεται πιο πιστευτό – συνάδει με τις ήδη υπάρχουσες πεποιθήσεις του υποκειμένου.

**Undergeneration** παρατηρείται όταν το συμπέρασμα δε γίνεται αποδεκτό, παρά το γεγονός ότι είναι λογικό επακόλουθο των προκειμένων. Ο λόγος για αυτό είναι γιατί αντιβαίνει προς τις παγιωμένες αντιλήψεις του υποκειμένου και κατά συνέπεια δεν είναι πιστευτό.

Στη συγκεκριμένη έρευνα οι συμμετέχοντες κλήθηκαν να αξιολογήσουν προδιατυπωμένους συμπερασμούς ως προς την ορθότητά τους και όχι να παραγάγουν οι ίδιοι το συμπέρασμα με βάση τις προκειμένες. Κατά συνέπεια, αυτό δεν μπορεί να θεωρηθεί αμιγώς ένα φαινόμενο συλλογιστικής προκατάληψης (reasoning bias). Εντούτοις όμως φαίνεται ότι οι υπάρχουσες αντιλήψεις επιδρούν καθοριστικά κατά τη διαδικασία του συλλογισμού. Φαίνεται επίσης ότι οι άνθρωποι επιδιώκουν τη διατήρηση του ήδη υπάρχοντος συστήματος πεποιθήσεων είτε αναζητώντας τα στοιχεία εκείνα που τις επιβεβαιώνουν είτε παραβλέποντας όσα στοιχεία τις απορρίπτουν.

Μια πολύ σημαντική παρατήρηση πάνω στα αποτελέσματα του πειράματος είναι το γεγονός ότι οι συμμετέχοντες κάνουν μια διαχείριση της εισερχόμενης πληροφορίας. Αυτό καθώς φαίνεται δεν μπορεί να οδηγήσει κατ' ανάγκην σε παραγωγικό συμπέρασμα, αλλά ούτε και σε αναιρέσιμη συλλογιστική. Αυτό σημαίνει ότι, δεδομένου ότι σε ένα δυναμικό σύστημα η περιπτωσιολογία μπορεί να είναι διευρυμένη, τείνουμε να κρίνουμε και να οδηγούμαστε σε συμπεράσματα σύμφωνα με αυτό που εν γένει συμβαίνει. Συνεπώς, όπως φάνηκε από το πείραμα, στις περιπτώσεις που οι συμμετέχοντες πραγματοποιούσαν εσφαλμένη παραγωγική διαδικασία συμπερασμού, στην πραγματικότητα υιοθετούσαν μια αναιρέσιμη συλλογιστική. Μπορούμε να υποθέσουμε ότι αυτό που εν γένει συμβαίνει είναι αυτό ακριβώς που έχει οδηγήσει στη διαμόρφωση ενός συστήματος προϋπαρχουσών πεποιθήσεων και οι οποίες φαίνεται ότι κυριαρχούν στη συλλογιστική μας διαδικασία. Εντούτοις, σε άλλες περιπτώσεις, οι συμμετέχοντες φάνηκε να μην υιοθετούν ούτε την αναιρέσιμη συλλογιστική αλλά να αξιολογούν την εισερχόμενη πληροφορία και είτε να την ενστερνίζονται είτε να την απορρίπτουν. Και

πάλι εδώ παρατηρούμε μια προσπάθεια διαχείρισης της εισερχόμενης πληροφορίας που ούτως ή άλλως φαίνεται να είναι αναγκαία και σύμφωνα με το περιεχόμενο της πληροφορίας και σύμφωνα με την αξιοπιστία της με βάση κυρίως την πηγή από την οποία προέρχεται.

Ένα επίσης πολύ σημαντικό εύρημα του πειράματος είναι το γεγονός ότι αυτό το συλλογιστικό μοτίβο της διατήρησης πεποιθήσεων δε φαίνεται να παρεισφρέει όταν οι συλλογισμοί που δόθηκαν στους συμμετέχοντες ήταν ίδιοι στη δομή με τους αρχικούς αλλά αφορούσαν σε έννοιες που δεν υπάρχουν σε κάποια φυσική γλώσσα. Οι συμμετέχοντες έπρεπε να αποφανθούν για συμπερασμούς σχετιζόμενους με έννοιες για τις οποίες δεν είχαν προηγούμενη γνώση και, συνεπώς, παγιωμένες, σε κάποιο βαθμό, αντιλήψεις. Στη συνθήκη αυτή, τα ευρήματα έδειχναν σαφώς καλύτερη επίδοση από τους συμμετέχοντες όσον αφορά στην επιλογή των ορθών και έγκυρων συμπερασμών ανεξάρτητα από το πόσο πιστευτό είναι το συμπέρασμα. Το εύρημα αυτό δείχνει κατ' αρχήν ότι ο ανθρώπινος νους έχει την ικανότητα παραγωγής λογικών συμπερασμάτων. Επίσης, φαίνεται ότι η παρεμβολή στη συλλογιστική διαδικασία είναι εμφανής όταν πρόκειται για έννοιες που είναι ήδη γνωστές και συνεπώς για προτάσεις της γλώσσας που ανήκουν στο σύνολο των πεποιθήσεων του ατόμου.

## **Σκέψεις και Ιδέες για την εφαρμογή της Δυνητικής Λογικής και της Αναθεώρησης Πεποιθήσεων στις γνωστικές προκαταλήψεις**

Όπως ήδη έχουμε θίξει, έχει διαπιστωθεί ότι η Κλασική Λογική αποτυγχάνει να αναπαραστήσει την ανθρώπινη συλλογιστική. Από πειράματα έχει διαπιστωθεί ότι οι άνθρωποι είμαστε σε θέση να σκεφτόμαστε με βάση τους κανόνες της Κλασικής Λογικής και μάλιστα χωρίς να έχουμε τη σχετική εκπαίδευση, αλλά φαίνεται ότι δεν είναι αυτό το σύννηθες συλλογιστικό μοτίβο μας. Ο κύριος λόγος για αυτό είναι ότι σχεδόν ποτέ δεν έχουμε εξαρχής συγκεκριμένα και σαφώς καθορισμένα δεδομένα τα οποία είναι βέβαιο ότι δεν πρόκειται να αλλάξουν και μοιραία είμαστε αναγκασμένοι να βρούμε άλλες συλλογιστικές πρακτικές, οι οποίες να μπορούν να μας εφοδιάσουν με σταθερότητα και απαντήσεις για να μπορούμε να επιβιώνουμε μέσα σε ένα διαρκώς μεταβαλλόμενο κόσμο.

### **Οι γνωστικές προκαταλήψεις ως καταστάσεις με αυξημένο βαθμό αναγκαιότητας**

Θεωρητικά, σε ένα αυστηρά καθορισμένο πλαίσιο, μια μονότονη λογική μπορεί να βελτιώσει, εμπλουτίσει και επεκτείνει ένα σύνολο πεποιθήσεων υπό τη μορφή προτάσεων σε κάποια τυπική γλώσσα. Οι προκείμενες είναι σαφείς και καθορισμένες και η σωστή εφαρμογή των κανόνων οδηγεί σε συμπεράσματα που δεν έχουν αντιφάσεις ούτε με τις προκείμενες αλλά ούτε με τις υπόλοιπες πεποιθήσεις του συνόλου και η παραγωγική συλλογιστική (deductive reasoning) είναι η ασφαλέστερη διαδικασία για να έχουμε έγκυρα και αληθή συμπεράσματα.

Πρακτικά αυτό δεν είναι εφικτό. Η αδυναμία να ενστερνιστούμε μια μονότονη λογική δεν οφείλεται στο γεγονός ότι η πραγματικότητα δεν είναι μονοσήμαντα ορισμένη, αλλά στο ότι η γνώση μας γι' αυτή δεν είναι μονοσήμαντα ορισμένη και πολλά πράγματα δεν είναι

ξεκάθαρα. Συνεπώς, οι αποφάσεις μας πρέπει με κάποιον τρόπο να ενσωματώσουν αυτήν την αβεβαιότητα. Οι πληροφορίες μας για τον κόσμο συνήθως δεν είναι ξεκάθαρες, μπορούν να είναι ελλιπείς ή αντικρουόμενες και φαίνεται ότι προσπαθούμε να συμπληρώσουμε τα «κενά» βασιζόμενοι στο διαθέσιμο οπλοστάσιό μας – την πρότερη γνώση μας. Μάλιστα, όπως αναφέρουν οι Novaes και Veluwenkamp, φαίνεται ότι η πρότερη γνώση μας έχει προτεραιότητα έναντι της εισερχόμενης πληροφορίας, μια πρακτική που δε μας δίνει απόλυτα σωστά συμπεράσματα, αλλά φαίνεται ότι είναι αρκετά αξιόπιστη. (Novaes et al., 2016)

Η προτεραιότητα που δίνουμε στις ήδη υπάρχουσες πεποιθήσεις μας συχνά οδηγεί σε συστηματικά συλλογιστικά σφάλματα, γιατί φαίνεται να παραβλέπουμε την εισερχόμενη πληροφορία. Αυτό που σίγουρα παίζει σπουδαίο ρόλο και δεν μπορούμε να αγνοήσουμε είναι η σημαντικότητα της εισερχόμενης πληροφορίας. Η θεωρία Αναθεώρησης Πεποιθήσεων δίνει προβάδισμα στη νεοεισερχόμενη πληροφορία, αλλά πιθανώς δεν είναι αυτή η ισχύουσα κατάσταση, κυρίως σε μια πολυπρακτορική (multi-agent) αλληλεπίδραση. Αυτό που φαίνεται πιο ρεαλιστικό είναι η προτεραιότητα της προς επεξεργασία πληροφορίας να εξαρτάται από την αξιοπιστία της πηγής από την οποία προέρχεται. (da Costa Pereira et al., 2011) Αν γενικεύσουμε λίγο αυτήν την άποψη, μπορούμε να πούμε ότι, αν η αξιοπιστία της πηγής επηρεάζει την προτεραιότητα της πληροφορίας, είναι εύλογο για μια ήδη υπάρχουσα πεποίθηση, δεδομένου ότι η πηγή της είναι το ίδιο το υποκείμενο, να ισοδυναμεί με πληροφορία αυξημένης προτεραιότητας όχι τόσο ως προς τη σειρά επεξεργασίας αλλά ως προς τη σημαντικότητά της και την επιστημική της οχύρωση.

Η τάση που έχουμε να διατηρούμε άθικτο το σύνολο των πεποιθήσεών μας έχει φανεί πειραματικά και μάλλον καταδεικνύει με έναν τρόπο την αυξημένη βαρύτητα των προϋπαρχουσών πεποιθήσεων. Κι όσο κι αν ένα σταθερό σύνολο αντιλήψεων μας δίνει ένα σταθερό και ασφαλές εξηγητικό πλαίσιο για να διασφαλίζεται η ισορροπία στην αλληλεπίδρασή μας με το φυσικό κόσμο, όπως ήδη διατυπώθηκε στην αρχή, η τάση να το διατηρούμε αναλλοίωτο είναι εκείνη που οδηγεί σε συστηματικά συλλογιστικά σφάλματα. Συνεπώς, βλέποντάς το από τη σκοπιά μιας θεωρίας διαχείρισης αβεβαιότητας, όπως η Δυνητική Λογική, μπορούμε να υποθέσουμε ότι οι γνωστικές προκαταλήψεις υπεισέρχονται στη συλλογιστική διαδικασία με τη μορφή προκειμένων

αυξημένου βαθμού αναγκαιότητας, χωρίς αυτός ο βαθμός να ορίζεται ρητά στις προκείμενες ή να υπονοείται με κάποιον τρόπο στο προς επίλυση πρόβλημα.

## **Βάσεις πεποιθήσεων και νοητικά μοντέλα αναπαράστασης του κόσμου**

Οι πεποιθήσεις μας προκύπτουν από την εσωτερικευση του κόσμου που μας περιβάλλει, από νοητικές του αναπαραστάσεις. Μπορούμε να παρατηρήσουμε τις διαφορές που υπάρχουν στις πεποιθήσεις των ανθρώπων ανά τους αιώνες και ανάλογα με το επίπεδο της κοινωνίας στην οποία ζουν και να συμπεράνουμε ότι η αποτελεσματική αλληλεπίδραση με το περιβάλλον προσδιορίζει σίγουρα το περιεχόμενο των αντιλήψεών μας αλλά και την προσήλωσή μας σε αυτές.

Το γεγονός ότι το περιεχόμενο των αντιλήψεων μπορεί να μας εξασφαλίσει την επιβίωση είναι μάλλον προφανές. Μια λανθασμένη πεποίθηση, όπως για παράδειγμα για την επικινδυνότητα ενός δηλητηρίου, μπορεί να οδηγήσει ακόμα και στο θάνατο. Επίσης, προκειμένου να συμβιώνουμε αρμονικά σε μια κοινωνία πρέπει να μοιραζόμαστε κοινές πεποιθήσεις, οι οποίες συνήθως έχουν τη μορφή νόμων, αρχών και κοινωνικών αξιών. Στην αντίθετη περίπτωση, η συμβίωση μπορεί να γίνει δυσλειτουργική σε διάφορες διαβαθμίσεις. Η επιμονή στη διατήρηση των πεποιθήσεων είναι αυτή που χρειάζεται προκειμένου να βιώνουμε τον κόσμο με συνέπεια και χωρίς αντιφάσεις, έτσι ώστε να είμαστε σε θέση να προβλέψουμε το μέγεθος των συνεπειών που μπορεί να έχει μια αλλαγή και να μπορούμε να αντιδράσουμε ανάλογα ώστε να προστατευθούμε. Η επιμονή αυτή δεν ταυτίζεται με αδιαλλαξία, διότι έχουμε την ικανότητα να αξιολογούμε τις πεποιθήσεις μας και να τις αναθεωρούμε ανάλογα με τις συνθήκες και η ανάγκη της επιβίωσης επιβάλλει τέτοιες ενέργειες.

Η θεωρία Αναθεώρησης Πεποιθήσεων προσφέρει ένα θεωρητικό πλαίσιο αναπαράστασης των πεποιθήσεων και των διαδικασιών αλλαγών αυτών. Φορμαλιστικά χρησιμοποιείται ένα σύνολο προτάσεων σε τυπική γλώσσα για την αναπαράσταση των πεποιθήσεων. Το σύνολο αυτό πρέπει να μην έχει αντιφάσεις. Αυτό σημαίνει ότι οι λογικές συνέπειες των προτάσεων του συνόλου ανήκουν επίσης στο σύνολο. Συνεπώς, η αφαίρεση ή η πρόσθεση μιας πεποίθησης συνεπάγεται αφαίρεση ή πρόσθεση όλων των

λογικών συνεπαγωγών, ώστε το σύνολο να παραμείνει συνεπές. Πιο συγκεκριμένα, αν στο σύνολο υπάρχει η πεποίθηση «η Αθήνα είναι πρωτεύουσα της Ελλάδας», αυτό σημαίνει ότι υπάρχουν και οι πεποιθήσεις «η Λάρισα δεν είναι πρωτεύουσα της Ελλάδας», «το Παρίσι δεν είναι πρωτεύουσα της Ελλάδας» κοκ.

Όμως, έχουμε την τάση να αποθηκεύουμε στη μνήμη μας μόνο τις πληροφορίες που είναι αληθείς, εις βάρος των πολυπληθέστερων αντίστοιχων ψευδών πληροφοριών. Αυτή η αρχή της αλήθειας, οδηγεί σε νοητικά μοντέλα αναπαράστασης του κόσμου που απαιτούν λιγότερους πόρους, αλλά οδηγεί και σε συστηματικά σφάλματα κατά τον παραγωγικό συλλογισμό. (Johnson-Laird, 2010) Πιθανότατα, η αρχή της αλήθειας ευθύνεται και για το γεγονός ότι η ρεαλιστική αναπαράσταση των πεποιθήσεων γίνεται με τη μορφή βάσεων και όχι συνόλων. Ας σκεφτούμε το παράδειγμα που δώσαμε πριν για την πρωτεύουσα της Ελλάδας. Αν κρατήσουμε στη μνήμη μας μόνο την αληθή πληροφορία «η Αθήνα είναι πρωτεύουσα της Ελλάδας», στη βάση πεποιθήσεων δε χρειάζεται να συμπεριληφθούν όλες οι συνέπειές της για όλες τις υπόλοιπες πόλεις του κόσμου.

Χρησιμοποιώντας τις βάσεις πεποιθήσεων και εισάγοντας μέσα στη θεωρία Αναθεώρησης Πεποιθήσεων την έννοια του μέτρου αναγκαιότητας της Δυνητικής Λογικής, όπως προτείνεται από τους Dubois και Prade (2014) μπορούμε να φτιάξουμε πολύ ρεαλιστικά μοντέλα αναπαράστασης του συστήματος πεποιθήσεων.

Κατ' αρχήν ας σκεφτούμε ότι στην πραγματικότητα έχουμε αντιφάσεις στον τρόπο σκέψης μας. Για παράδειγμα, μπορεί κάποιος να έχει την πεποίθηση «όλα τα ζώα είναι καλά», αλλά αν δει ένα ποντίκι να το σκοτώσει. Η πράξη έρχεται σε αντίφαση με την πεποίθηση, αλλά αυτό συμβαίνει στη φύση και μάλιστα συχνά, χωρίς απαραίτητα να σημαίνει ότι διαγράφηκε η πεποίθηση από τη βάση. Μπορούμε να αναπαραστήσουμε τη βάση με τέτοιο τρόπο ώστε και να ισχύει η αρχή της αλήθειας και να ενσωματώνεται η ενδεχόμενη περιπτώσιολογία.

Άρα μπορούμε να έχουμε μια βάση  $A = \{(a_i, \alpha_i), i=1, \dots, m\}$ , όπου  $a_i$  να αναπαριστούμε το γεγονός «ένα ζώο  $i$  είναι καλό» με βαθμό αναγκαιότητας  $\alpha_i$ .

Κάτι ανάλογο μπορούμε να υποθέσουμε ότι συμβαίνει στις πεποιθήσεις μας μετά από ένα τραυματικό γεγονός. Μπορεί να θεωρώ τα αεροπορικά ταξίδια αρκετά ασφαλή, αλλά μετά από ένα άσχημο ταξίδι να φοβάμαι να ξαναμπώ σε αεροπλάνο.



Η εισαγωγή του μέτρου αναγκαιότητας μέσα στην Αναθεώρηση Πεποιθήσεων μας βοηθά όχι μόνο να μοντελοποιήσουμε όλους τους δυνατούς κόσμους – και συνεπώς όλους τους δυνατούς νόες – αλλά και τη δυναμικότητα αυτών των κόσμων. Αυτό σημαίνει ότι αν έχουμε ένα σύνολο προτάσεων σε τυπική γλώσσα που αναπαριστούν ένα σύνολο πεποιθήσεων, οποιαδήποτε νέα πληροφορία ή αλλαγή του κόσμου δεν οδηγεί σε αφαίρεση ή προσθήκη ενός στοιχείου και όλων των λογικών συνεπαγωγών του στο σύνολο, αλλά αλλαγή στο βαθμό αναγκαιότητας. Με την προσέγγιση αυτή είναι εφικτή και η αναίρεση μιας αλλαγής στο σύνολο των πεποιθήσεων, αν η πρώτη αλλαγή ήταν λάθος, ή, για παράδειγμα, αν η τραυματική εμπειρία που την προκάλεσε ξεχάστηκε ή θεραπεύτηκε.

Αναθέτοντας στις πεποιθήσεις ένα βαθμό αναγκαιότητας, καταλαβαίνουμε ότι όσο μεγαλύτερος είναι ο βαθμός αυτός, τόσο ισχυρότερη οχύρωση έχει η κάθε πεποίθηση. Προφανώς οι άνθρωποι είμαστε δυναμικά συστήματα και οι πεποιθήσεις μας αλλάζουν βαθμό αναγκαιότητας με βάση την ηλικία, την εκπαίδευση και τις εμπειρίες μας και πάντα κάποιες είναι πιο οχυρωμένες από άλλες και δεν αναιρούνται εύκολα, όπως για παράδειγμα οι φυσικοί νόμοι. Οι συλλογιστικές προκαταλήψεις είναι παραδείγματα πεποιθήσεων με αυξημένη οχύρωση έναντι της αλλαγής.

Οι μηχανισμοί μέσω των οποίων μια πεποίθηση αποκτά το βαθμό αναγκαιότητας που της αντιστοιχεί δε μας είναι γνωστοί. Είναι οξύμωρο που μας φαίνεται εύλογο οι φυσικοί νόμοι να έχουν μεγάλο βαθμό αναγκαιότητας και ταυτόχρονα έχουμε την τάση να δαιμονολογούμε τις προκαταλήψεις, ενώ είναι και τα δύο προϊόντα της ίδιας νοητικής διεργασίας. Επίσης, σε μια συλλογιστική διαδικασία, είτε είναι πρωτογενής συμπερασμός είτε αξιολόγηση συμπερασμάτων, όπως στο πείραμα που αναφέραμε, ακριβώς επειδή έχουμε συνηθίσει να σκεφτόμαστε σε συνθήκες ελλιπούς πληροφόρησης, μοιραία θα έχουν συνεισφορά όλα τα είδη των πεποιθήσεών μας - επιστημονικά ή μη.

## Το παράδοξο της Λίντα

Ένα από τα κλασικά παράδοξα της βιβλιογραφίας, όπου φαίνεται χαρακτηριστικά η επιρροή των προϋπαρχουσών πεποιθήσεων στη συλλογιστική διαδικασία, είναι το λεγόμενο παράδοξο της Λίντα ή αλλιώς conjunction fallacy.

Η περιγραφή του παραδόξου όπως αναφέρεται στο διαδίκτυο:

*Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.*

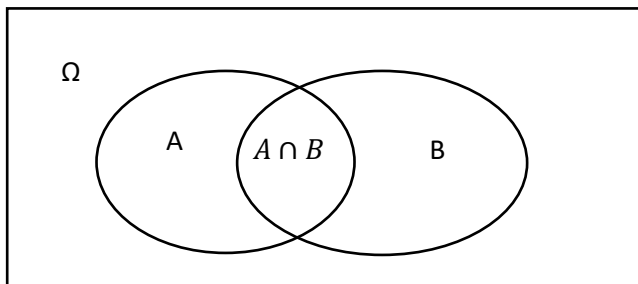
*Which is more probable?*

- *Linda is a bank teller.*
- *Linda is a bank teller and is active in the feminist movement.*

(πηγή: [https://en.wikipedia.org/wiki/Conjunction\\_fallacy](https://en.wikipedia.org/wiki/Conjunction_fallacy))

Η πλειοψηφία των ερωτηθέντων συμφωνούσε με τη δεύτερη πρόταση – η Λίντα είναι ταμίας τραπέζης και φεμινίστρια.

Με αυστηρά μαθηματικούς υπολογισμούς και στο πλαίσιο της Κλασικής Λογικής ισχύει ότι:



$$P(A \cap B) \leq P(A) \text{ και } P(A \cap B) \leq P(B)$$

Η πιθανότητα της τομής δύο ενδεχομένων είναι μικρότερη από την πιθανότητα του καθενός ξεχωριστά. Άρα, η πιθανότητα να είναι και ταμίας τραπέζης και φεμινίστρια είναι μικρότερη από την πιθανότητα να είναι μόνο ταμίας τραπέζης ή μόνο φεμινίστρια.

Γιατί όμως οι περισσότεροι άνθρωποι επιλέγουν ως πιθανότερη τη δεύτερη επιλογή; Δείχνει αυτό ότι είμαστε ανίκανοι να παραγάγουμε ορθά συμπεράσματα ή μήπως να κατανοήσουμε μαθηματικές έννοιες;

Ας δούμε κάποια πολύ σημαντικά στοιχεία του παραδόξου:

Το πρώτο αξιοσημείωτο είναι ότι τα δύο προτεινόμενα συμπεράσματα δεν αναφέρονται πουθενά στις προκειμένες. Στην περιγραφή της Λίντα δεν υπάρχει καμία ρητή και ξεκάθαρη αναφορά στο επάγγελμά της ή στη θέση που έχει αναφορικά με τις σχέσεις των δύο φύλων. Άρα, προκειμένου να απαντήσουμε είναι απαραίτητο να χρησιμοποιήσουμε τη γνώση που έχουμε για τον κόσμο. Αυτό σημαίνει ότι καλούμαστε να συμπληρώσουμε τα «κενά» στην πληροφορία με βάση τις πεποιθήσεις μας. Αν το ίδιο παράδοξο δινόταν σε μια φυλή που δεν γνωρίζει αυτές τις έννοιες, οι συμμετέχοντες δε θα μπορούσαν να απαντήσουν τίποτα γιατί ούτε θα το κατανοούσαν ούτε θα είχαν αντίστοιχες πολιτισμικές αναφορές για να τις χρησιμοποιήσουν.

Το δεύτερο αξιοσημείωτο είναι ότι το επάγγελμα που της έχει αποδοθεί στην πραγματικότητα είναι μια dummy τιμή. Ανάλογα με την εποχή θα μπορούσε η Λίντα, με το ίδιο προφίλ, να εργάζεται ως μοντέλο, να έχει αναλάβει την επιχείρηση των γονιών της, να είναι άνεργη ή να είναι αναγκασμένη να δουλεύει ως τηλεφωνήτρια γιατί η αγορά εργασίας διέρχεται κρίση. Άρα, το επάγγελμα που προτείνεται από τις διαθέσιμες επιλογές είναι δεν έχει επαρκές λογικό στήριγμα.

Θα προσπαθήσουμε να προσεγγίσουμε το παράδοξο με τη βοήθεια της Δυνητικής Λογικής.

Γιατί οι απαντήσεις στη συντριπτική πλειοψηφία δείχνουν πως οι άνθρωποι αγνοούν τους νόμους των μαθηματικών;

Αναπαριστούμε:

BT: η Λίντα είναι ταμίας σε τράπεζα

F: η Λίντα είναι φεμινίστρια

Prf: η Λίντα έχει το συγκεκριμένο προφίλ που περιγράφεται

Από τις διαθέσιμες επιλογές εξετάζουμε αν είναι ταμίας σε τράπεζα έχει το συγκεκριμένο προφίλ και ταυτόχρονα είναι και φεμινίστρια.

Υπολογίζουμε τη δυνατότητα του ενδεχομένου που επιλέγει η πλειοψηφία.

Για τον υπολογισμό του μέτρου αναγκαιότητας της δεύτερης επιλογής έχουμε τα εξής:

$$N(BT \wedge Prf \rightarrow F) = N(\neg(BT \wedge Prf) \vee F) \geq \max(N(\neg(BT \wedge Prf)), N(F)) \quad (1)$$

$$N(\neg(BT \wedge Prf)) = N(\neg BT \vee \neg Prf) \geq \max(N(\neg BT), N(\neg Prf)) \quad (2)$$

$$\text{Αλλά } \Pi(Prf) = 1 \text{ και άρα } N(\neg Prf) = 0 \quad (3)$$

$$\text{Επίσης, από τις προκείμενες έχω άγνοια για το BT: } \Pi(BT) = \Pi(\neg BT) = 1 \text{ και άρα } N(\neg BT) = N(BT) = 0 \quad (4)$$

$$\text{Από τις (2), (3) και (4) η (1) γίνεται } \mathbf{N(BT \wedge Prf \rightarrow F) \geq N(F)}$$

Υπολογίζουμε το μέτρο αναγκαιότητας της πρώτης επιλογής:

$$N(BT \wedge Prf) = \min(N(BT), N(Prf))$$

$$\text{Αλλά ισχύει ότι } N(\neg BT) = N(BT) = 0$$

$$\text{Συνεπώς, } \mathbf{N(BT \wedge Prf) = 0}$$

Ή εναλλακτικά υπολογίζουμε το μέτρο αναγκαιότητας στη συνθήκη «αν έχει αυτό το προφίλ, τότε είναι ταμίας σε τράπεζα»:

$$N(Prf \rightarrow BT) = N(\neg Prf \vee BT) \geq \max(N(\neg Prf), N(BT))$$

$$\text{Αλλά ισχύει ότι } N(\neg BT) = N(BT) = 0$$

$$\text{Επίσης, Αλλά } \Pi(Prf) = 1 \text{ και άρα } N(\neg Prf) = 0$$

Συνεπώς,  $\mathbf{N}(\mathbf{Prf} \rightarrow \mathbf{BT}) \geq \mathbf{0}$  (τετριμμένη συνθήκη)

$$N(\text{Prf} \rightarrow \text{BT}) = N(\neg \text{Prf} \vee \text{BT}) = N(\neg(\neg \text{Prf}) \wedge \neg \text{BT}) = N(\text{Prf} \wedge \neg \text{BT}) = \min(N(\text{Prf}), N(\neg \text{BT}))$$

Αλλά,  $N(\neg \text{BT}) = 0$ , όπως έχουμε ήδη δει.

Συνεπώς,  $\mathbf{N}(\mathbf{Prf} \rightarrow \mathbf{BT}) = \mathbf{0}$

Άρα, για την πρώτη επιλογή έχουμε μηδενικό μέτρο αναγκαιότητας, κάτι που συμφωνεί με την τάση να μην καθοδηγείται η προτίμηση του κοινού προς αυτή.

Υπολογίζουμε το μέτρο αναγκαιότητας στη συνθήκη «αν έχει αυτό το προφίλ, τότε είναι φεμινίστρια»:

$$N(\text{Prf} \rightarrow \text{F}) = N(\neg \text{Prf} \vee \text{F}) \geq \max(N(\neg \text{Prf}), N(\text{F}))$$

Αλλά  $\Pi(\text{Prf}) = 1$  και άρα  $N(\neg \text{Prf}) = 0$

Συνεπώς,  $\mathbf{N}(\mathbf{Prf} \rightarrow \mathbf{F}) \geq \mathbf{N}(\mathbf{F})$

Παρατηρούμε, λοιπόν, ότι το μέτρο αναγκαιότητας για την πρώτη συνθήκη είναι μηδέν. Άρα, δεδομένου του προφίλ που έχει δοθεί στην περιγραφή της Λίντα, δεν προκύπτει λογική σύνδεση με το επάγγελμα του ταμιά τραπέζης. Αντίθετα, είναι το μέτρο αναγκαιότητας της φεμινίστριας που είναι μη μηδενικό και κληροδοτεί την τιμή του στη δεύτερη συνθήκη, αυτή που επιλέγουν οι περισσότεροι ερωτηθέντες.

Πού αποτυγχάνει η ερμηνεία με βάση την Κλασική Λογική, όμως, και γιατί η πλειοψηφία την απορρίπτει;

Στην προσέγγιση με βάση την Κλασική Λογική, φαίνεται ότι έχουμε ένα κλειστό σύμπαν που αποτελείται από δύο σύνολα – αυτό των ταμιών τραπέζης και αυτό των φεμινιστριών. Στις προκείμενες, όμως, πουθενά δεν οριοθετείται το σύμπαν των διαθέσιμων επιλογών. Αντίθετα, οι συμμετέχοντες άρρητα καλούνται να συνεισφέρουν την προσωπική τους γνώση για τον κόσμο που ζουν και με βάση αυτή να αποφασίσουν. Τελικά, υπάρχει μια άρρητη απαίτηση από τους συμμετέχοντες να εμπλέξουν τις πεποιθήσεις - και ενδεχομένως τις προκαταλήψεις τους - στη συλλογιστική διαδικασία προκειμένου να λειτουργήσουν ορθολογικά.

## Συμπεράσματα και ιδέες για μελλοντικές έρευνες

Μέσα από την παρούσα έρευνα ενισχύεται η πεποίθηση ότι υπάρχει ανάγκη προσέγγισης των νοητικών διεργασιών μέσα από διαδικασίες διαφορετικές από την Κλασική Λογική, η οποία φαίνεται να είναι αρκετά περιοριστική σε μια ρεαλιστική μοντελοποίηση.

Οι άνθρωποι τις περισσότερες φορές καλούμαστε να πάρουμε αποφάσεις υπό αβεβαιότητα και ελλιπή πληροφόρηση. Άρα, ακόμα και ο προσδιορισμός του συνόλου των καταστάσεων δυσχεραίνεται ή καθίσταται αδύνατος. Η ορθολογική συμπεριφορά είναι να αξιοποιήσουμε την προϋπάρχουσα γνώση μας για να δούμε τί συμβαίνει εν γένει, ώστε να καθοδηγηθούμε στη συλλογιστική διαδικασία. Παρά το γεγονός ότι αυτή η συμπεριφορά είναι χρήσιμη και αποτελεσματική, συχνά οδηγεί σε σφάλματα συμπερασμού, γιατί οι ήδη διαμορφωμένες αντιλήψεις μας παρεισφρέουν στο συλλογισμό και διαμορφώνουν το αποτέλεσμα.

Το πώς επιμένουν οι προϋπάρχουσες αντιλήψεις να υπεισέρχονται και να επηρεάζουν την εγκυρότητα ενός συλλογισμού θεωρούμε ότι εξηγείται από το φαινόμενο διατήρησης της ομοιότητας. Η τάση να διατηρούμε άθικτο το σύνολο των πεποιθήσεών μας φαίνεται ότι είναι ορθολογική, γιατί έτσι διατηρούνται άθικτα τα νοητικά μοντέλα αναπαράστασης του κόσμου πάνω στα οποία θεμελιώνεται η στάση και η δράση μας. Άρα, το να διατηρούμε τις πεποιθήσεις μας μάς βοηθά να βρισκόμαστε σε ισορροπία με το περιβάλλον.

Συνέπεια της τάσης αυτής είναι η υιοθέτηση αντιλήψεων, με τη μορφή γνωστικών προκαταλήψεων, οι οποίες φαίνεται να κυριαρχούν στη συλλογιστική διαδικασία. Οι προκαταλήψεις αυτές μπορούν να προσεγγιστούν από τη σκοπιά της Δυνητικής Λογικής ως προτάσεις με αυξημένο βαθμό αναγκαιότητας, ο οποίος δε δικαιολογείται απόλυτα από τις εκάστοτε συνθήκες.

Η Δυνητική Λογική σε συνδυασμό με τη θεωρία Αναθεώρησης Πεποιθήσεων φαίνεται ότι μπορούν να μας δώσουν μια ρεαλιστική μοντελοποίηση της ανθρώπινης συλλογιστικής και να μας βοηθήσουν να προσεγγίσουμε φαινόμενα όπου είναι έκδηλη η επιρροή των προκαταλήψεων, όπως το conjunction fallacy.

Θα είχε ενδιαφέρον να δούμε πώς μπορούμε να αναστρέψουμε την επιρροή των προκαταλήψεων σε τέτοιου είδους φαινόμενα, αξιοποιώντας την Αναθεώρηση Πεποιθήσεων.

Επίσης, χρειάζεται να γίνει έρευνα αν και κατά πόσο η συγκεκριμένη μοντελοποίηση μπορεί να φανεί χρήσιμη στην ερμηνεία άλλων ειδών προκαταλήψεων, όπως τα anchoring biases, framing effect, dread aversion κλπ που εμπλέκονται μεν στη διαδικασία λήψης απόφασης και μπορούν να οδηγήσουν σε μη βέλτιστες αποφάσεις, αλλά δεν είναι σφάλματα συμπερασμού με τη στενή έννοια του όρου.

## Βιβλιογραφία

- 1) Conjunction Fallacy. (2021, June 25). In Wikipedia.  
[https://en.wikipedia.org/w/index.php?title=Conjunction\\_fallacy&oldid=1030327034](https://en.wikipedia.org/w/index.php?title=Conjunction_fallacy&oldid=1030327034)
- 2) Da Costa Pereira, C., Tettamanzi, A., Villata, S. (2011). Changing One's Mind: Erase or Rewind? Possibilistic Belief Revision with Fuzzy Argumentation based on Trust. *International Joint Conference on Artificial Intelligence*. HAL Id: hal-01328696, <https://hal.archives-ouvertes.fr/hal-01328696>. DOI:10.5591/978-1-57735-516-8/IJCAI11-039
- 3) Dubois, D., Prade, H. (1991). Possibilistic logic, preferential models, non-monotonicity and related issues. *Proceedings of the 12<sup>th</sup> international joint conference on Artificial Intelligence*, 1: 419-424.
- 4) Dubois, D., Prade, H. (1996). Belief Revision with Uncertain inputs in the Possibilistic Setting. *Proceedings of the 12<sup>th</sup> international joint conference on Uncertainty in Artificial Intelligence*: 236-243.
- 5) Dubois, D., Prade, H. (2014). Possibilistic Logic – An overview. *Handbook of the History of Logic*, 9: 197-255. doi: 10.1016/B978-0-444-51624-4.50007-1
- 6) Dubois, D., Prade, H. (2018). A crash course on Generalized Possibilistic Logic. In: Ciucci D., Pasi G., Vantaggi B. (eds) *Scalable Uncertainty Management. SUM 2018. Lecture Notes in Computer Science, vol 11142*. Springer, Cham.  
[https://doi.org/10.1007/978-3-030-00461-3\\_1](https://doi.org/10.1007/978-3-030-00461-3_1)
- 7) Dunbar, K., Klahr, D. (2012). Scientific Thinking and Reasoning. *Oxford Handbooks Online: Nov-12, Subject: Psychology, Cognitive Psychology, Cognitive Neuroscience*. DOI: 10.1093/oxfordhb/9780199734689.001.0001
- 8) Gärdenfors, P. (1990). Belief Revision and Nonmonotonic Logic: Two sides of the same coin? *Proceedings of ECAI 90*: 768-773. DOI:10.1007/BFb0018432
- 9) Gärdenfors, P., Rott, H. (1995). Belief Revision. *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, *Epistemic and Temporal Reasoning*: 55-132.



- 10) Hansson, Sven Ove. (2017). "Logic of Belief Revision", *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, Edward N. Zalta (ed.).  
<https://plato.stanford.edu/archives/win2017/entries/logic-belief-revision/>
- 11) Johnson-Laird, P. (2010). Mental models and human reasoning. *Special series of Inaugural Articles by members of the National Academy of Sciences elected in 2007*. [www.pnas.org/cgi/doi/10.1073/pnas.1012933107](http://www.pnas.org/cgi/doi/10.1073/pnas.1012933107)
- 12) Levi, I., (1991). *The Fixation of Belief and its Undoing: Changing Beliefs through Inquiry*. Cambridge University Press. doi:10.1017/CBO9780511663819
- 13) Novaes C. D., Veluwenkamp H. (2017). Reasoning Biases, Non-Monotonic Logics and Belief Revision. *Theoria*, 83: 29-52. doi: 10.1111/theo.12108
- 14) Shah, A., Oppenheimer, D. (2008). Heuristics Made Easy: An Effort-Reduction Framework. *Psychological Bulletin*, Volume 134, No. 2: 207-222. DOI: 10.1037/0033-2909.134.2.207