Τεχνητή Νοημοσύνη

7η διάλεξη (2024-25)

Ίων Ανδρουτσόπουλος

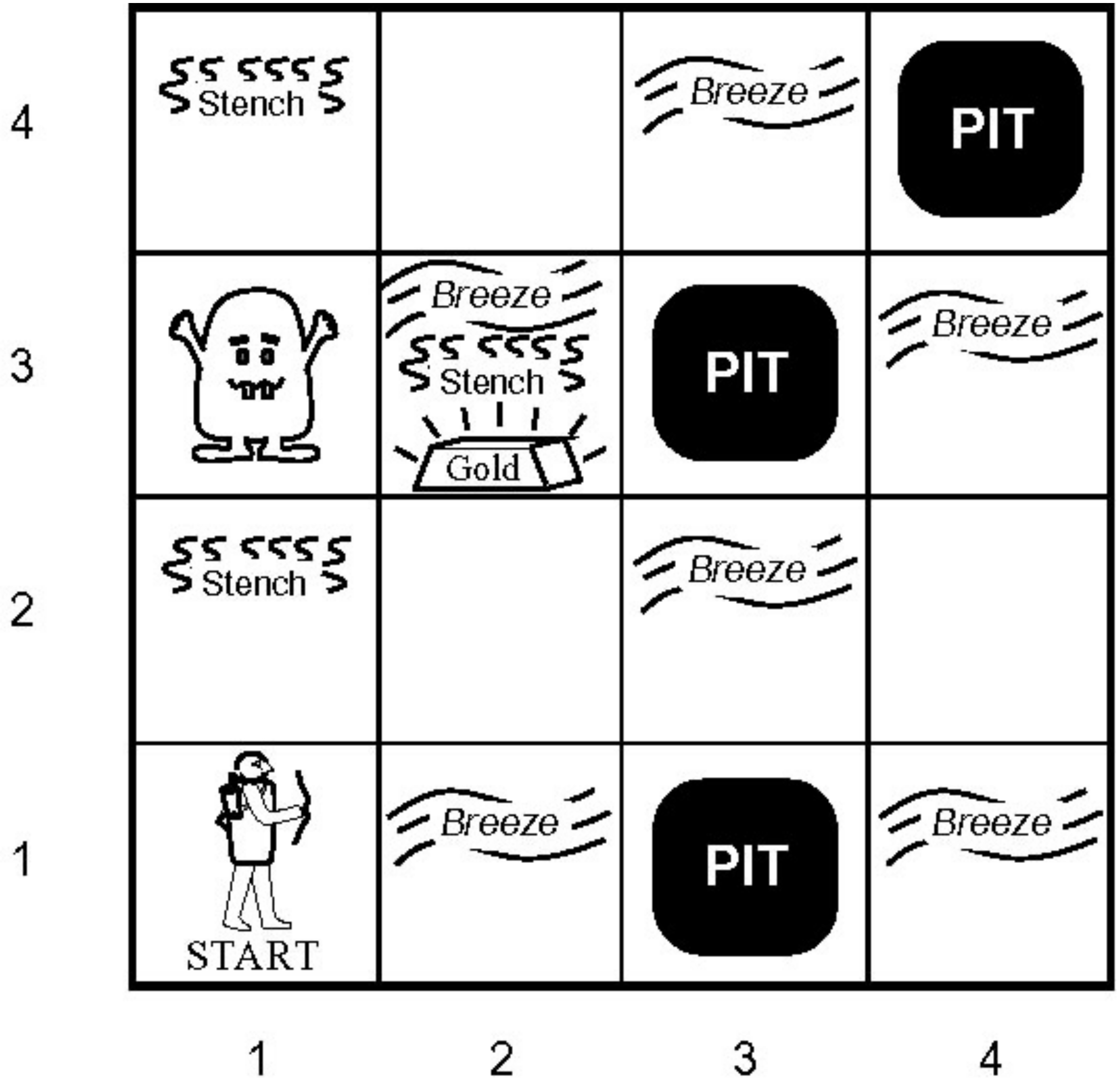
http://www.aueb.gr/users/ion/

Οι διαφάνειες αυτής της διάλεξης βασίζονται στο βιβλίο *Artificial Intelligence – A Modern Approach* των S. Russel και P. Norvig, 2η και 4η έκδοση, Prentice Hall, 2003 και 2020. Τα περισσότερα σχήματα των διαφανειών προέρχονται από αντίστοιχες διαφάνειες του ίδιου βιβλίου.

Τι θα ακούσετε σήμερα

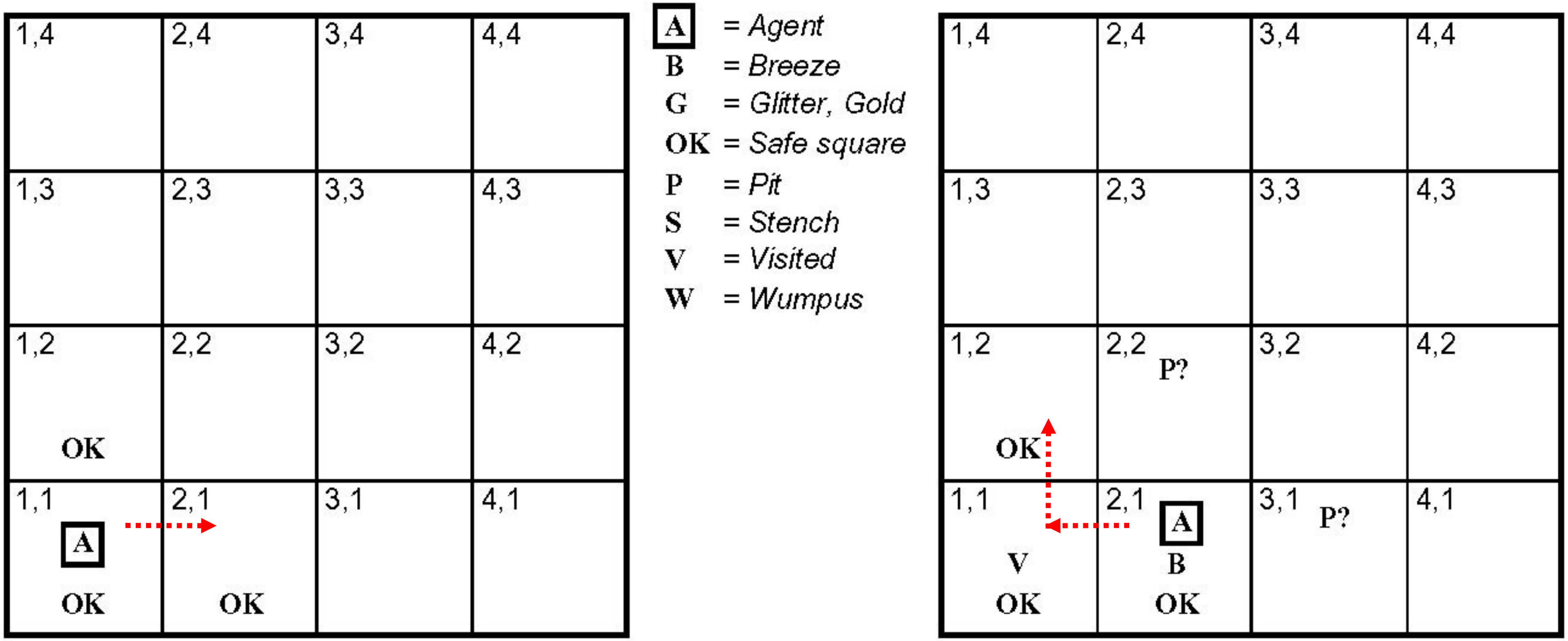
- Προτασιακή λογική:
  - Σύνταξη και σημασιολογία.
  - Μοντέλα και ταυτολογική συνεπαγωγή.
  - Ορθότητα και πληρότητα αλγορίθμων εξαγωγής συμπερασμάτων.
  - Εξαγωγή συμπερασμάτων με αναζήτηση μοντέλων.

Ο κόσμος του Wumpus



- Δίπλα στο Wumpus μυρίζει.
- Δίπλα στα ορύγματα υπάρχει ρεύμα αέρος.
- Ο χρυσός λάμπει στο τετράγωνό του.
- Κάθε φορά ξεκινάμε με τυχαία θέση του Wumpus, ορυγμάτων, χρυσού.
- Σε τετράγωνα ζωντανού Wumpus και ορυγμάτων ο πράκτορας πεθαίνει.
- Ενέργειες: κίνηση ένα τετράγωνο εμπρός, στροφή 90^{o} αριστερά ή δεξιά, συλλογή αντικειμένου, εκτόξευση βέλους εμπρός.
- Το βέλος (ή ντομάτα) σκοτώνει (λερώνει) το Wumpus, οπότε ακούγεται κραυγή παντού (αποσύρεται).
- Ένα μόνο βέλος (ντομάτα).

Γνώσεις και κινήσεις του πράκτορα



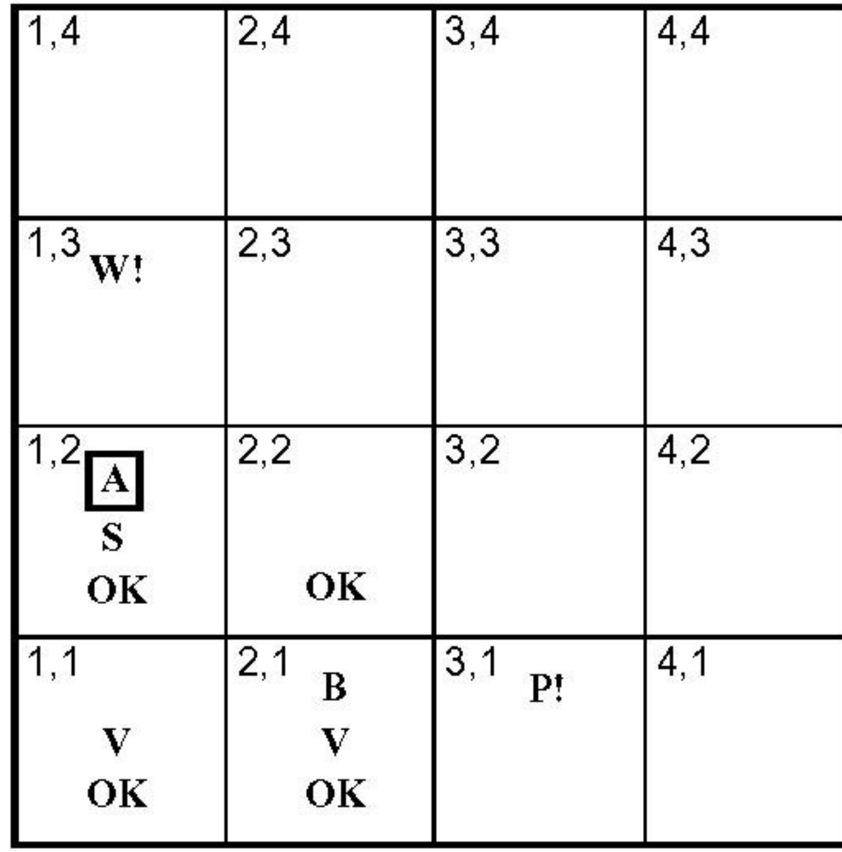
(a) (b)

Δεν υπάρχει ούτε μυρωδιά ούτε ρεύμα στο (1,1). Άρα τα (1,2) και (2,1) είναι ασφαλή. Πηγαίνει π.χ. στο (2,1).

Στο (2,1) υπάρχει ρεύμα. Άρα όρυγμα στο (2,2) ή το (3,1). Υπάρχει μόνο ένα σίγουρο τετράγωνο στο οποίο δεν έχει πάει, το (1,2). Πηγαίνει εκεί μέσω σίγουρων τετραγώνων.

Γνώσεις και κινήσεις του πράκτορα

- Δεν υπάρχει ρεύμα στο (1,2). Άρα δεν μπορεί να υπάρχει όρυγμα στο (2,2). Ξέρουμε, όμως, ότι υπάρχει όρυγμα στο (2,2) ή το (3,1). **Άρα το όρυγμα είναι στο (3,1).**
- Στο (1,2) μυρίζει. Άρα το Wumpus βρίσκεται στο (1,3) ή στο (2,2). Αλλά αν βρισκόταν στο (2,2), θα το είχαμε μυρίσει στο (2,1). **Άρα το Wumpus είναι στο (1,3).**



Γνώσεις και εξαγωγή συμπερασμάτων

- **Βάση γνώσης** (ΒΓ) του πράκτορα.
  - Περιέχει **τύπους** μια **γλώσσας παράστασης γνώσεων** (π.χ. προτάσεις προτασιακής λογικής).
  - Το **συντακτικό** της γλώσσας καθορίζει τους τύπους που ανήκουν στη γλώσσα και άρα μπορούν να χρησιμοποιηθούν.
  - Οι τύποι παριστάνουν τις γνώσεις του πράκτορα.
- Αλγόριθμοι **εξαγωγής συμπερασμάτων**.
  - Παράγουν νέους τύπους από αυτούς που βρίσκονται στη ΒΓ.
  - Οι νέοι τύποι παριστάνουν συμπεράσματα.
- **Προσθήκη γνώσεων** στη ΒΓ.
  - Νέα δεδομένα (π.χ. αισθητήρες) ή συμπεράσματα.
- **Ερωτήσεις** προς τη ΒΓ.
  - Π.χ. είναι ασφαλές να πάω στο (2,2);

Σημασιολογία και μοντέλο

- Η **σημασιολογία** μιας γλώσσας ορίζει τη σημασία κάθε τύπου της γλώσσας σε κάθε **κόσμο**.
  - Στην κλασική **λογική**, σε έναν κόσμο η σημασία ενός τύπου μπορεί να είναι ***αληθές*** ή ***ψευδές***.
  - Π.χ. έστω ότι χρησιμοποιούμε τα **x** και **y** για να παραστήσουμε τους αριθμούς των φοιτητών και φοιτητριών, αντίστοιχα, και ότι τα + και = έχουν τη συνηθισμένη αριθμητική σημασία.
  - Για να αποφανθούμε αν ο τύπος **x + y = 4** είναι αληθής σε έναν κόσμο, πρέπει να ξέρουμε πόσοι είναι οι φοιτητές και οι φοιτήτριες στον κόσμο αυτό (τις **τιμές των x και y**).
  - Δεδομένων των τιμών των x και y σε έναν κόσμο, η **σημασιολογία** της αριθμητικής γλώσσας καθορίζει αν ο τύπος x + y = 4 είναι αληθής ή ψευδής σε αυτόν τον κόσμο.
- **Μοντέλο** του κόσμου.
  - **Αφηρημένη παράσταση του κόσμου**, που παρέχει τις πληροφορίες που μας χρειάζονται για να υπολογίσουμε τη σημασία κάθε τύπου (στο παράδειγμα, τις τιμές των x και y).

Ταυτολογική συνεπαγωγή

- **α ⊨ β** σημαίνει: σε κάθε μοντέλο όπου είναι αληθής ο τύπος α, είναι αληθής και ο τύπος β.
  - Όποτε είναι αληθής ο α, είναι και ο β.
  - $(x + y = 4) \models (4 = x + y)$, γιατί σε οποιοδήποτε μοντέλο στο οποίο αληθεύει $(x + y = 4)$, αληθεύει και $(4 = x + y)$.
- **ΒΓ ⊨ β** σημαίνει: σε κάθε μοντέλο όπου είναι αληθής η ΒΓ, είναι αληθής και ο τύπος β.
  - Μπορούμε να θεωρήσουμε ότι όλοι οι τύποι της ΒΓ μαζί αποτελούν ένα μεγάλο τύπο (μια μεγάλη σύζευξη).
  - Όποτε είναι αληθής η ΒΓ, είναι και ο β.
  - Άρα αν η ΒΓ είναι αληθής στον κόσμο όπου βρίσκομαι, είναι αληθής και ο τύπος β.
  - Άρα αν πιστεύω ότι είναι αληθής η ΒΓ στον κόσμο όπου βρίσκομαι, πρέπει να πιστεύω ότι είναι αληθής και ο β.

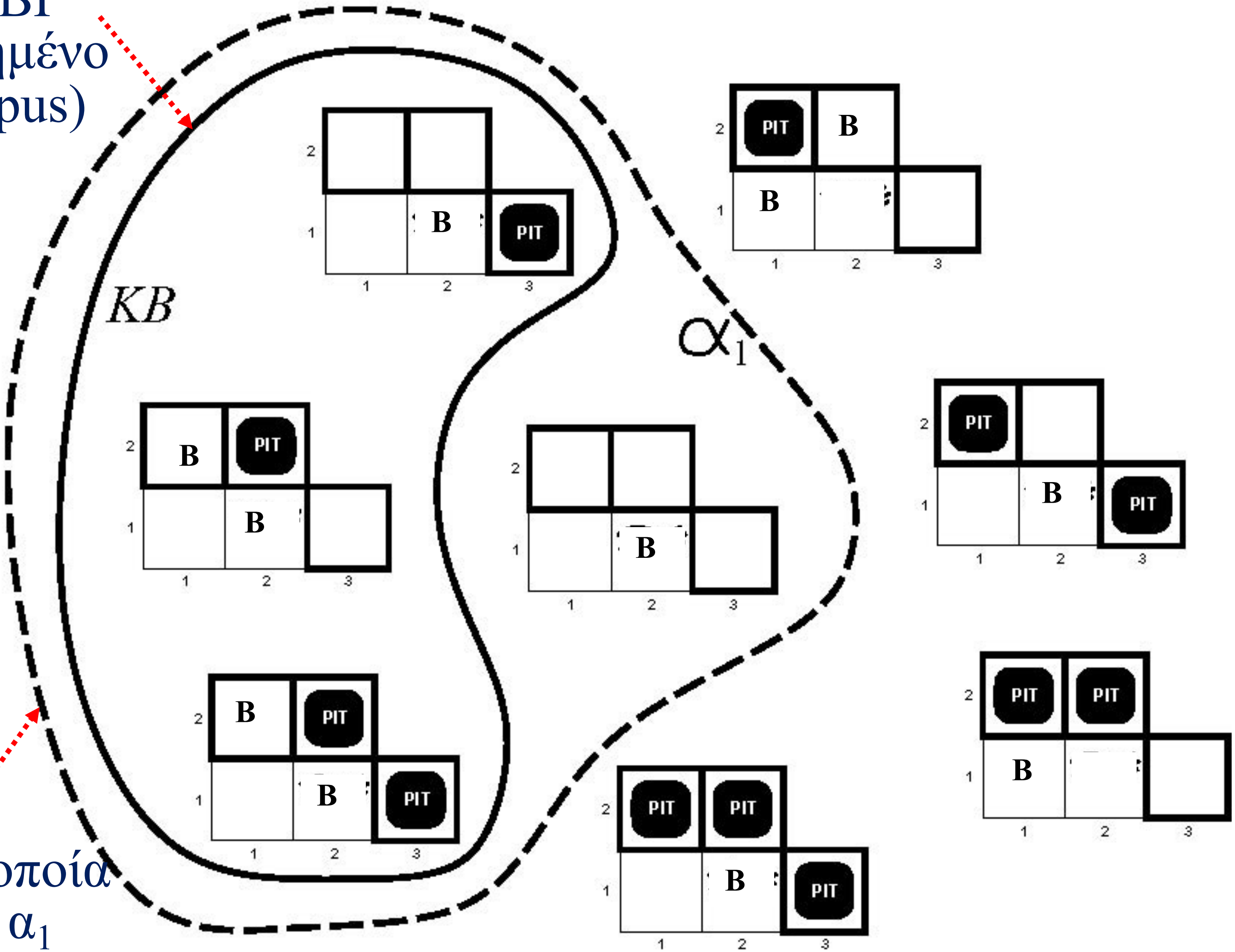
Παράδειγμα

- Έστω ότι **η ΒΓ περιέχει**:
  - Όχι ρεύμα ούτε όρυγμα στο (1,1).
  - Ρεύμα στο (2,1).
  - Τους κανόνες του κόσμου (π.χ. πάντα ρεύμα στα γειτονικά τετράγωνα ορυγμάτων).
- **Έπονται ταυτολογικά** από τη **ΒΓ** τα **ακόλουθα;**
  - $α_1$: Όχι όρυγμα στο (1,2).
  - $α_2$: Όχι όρυγμα στο (2,2).
- Προκύπτει ότι: ΒΓ $\models α_1$ αλλά ΒΓ $\not\models α_2$.
  - Με **έλεγχο όλων των μοντέλων** (πιθανών κόσμων).
  - Σε όλα τα μοντέλα όπου αληθεύει η ΒΓ αληθεύει και ο $α_1$.
  - Υπάρχουν όμως μοντέλα όπου αληθεύει η ΒΓ αλλά όχι ο $α_2$.

$ΒΓ \models α_1$

Όχι όρυγμα στο (1,2).

μοντέλα στα οποία αληθεύει η ΒΓ (για απλοποιημένο κόσμο Wumpus)



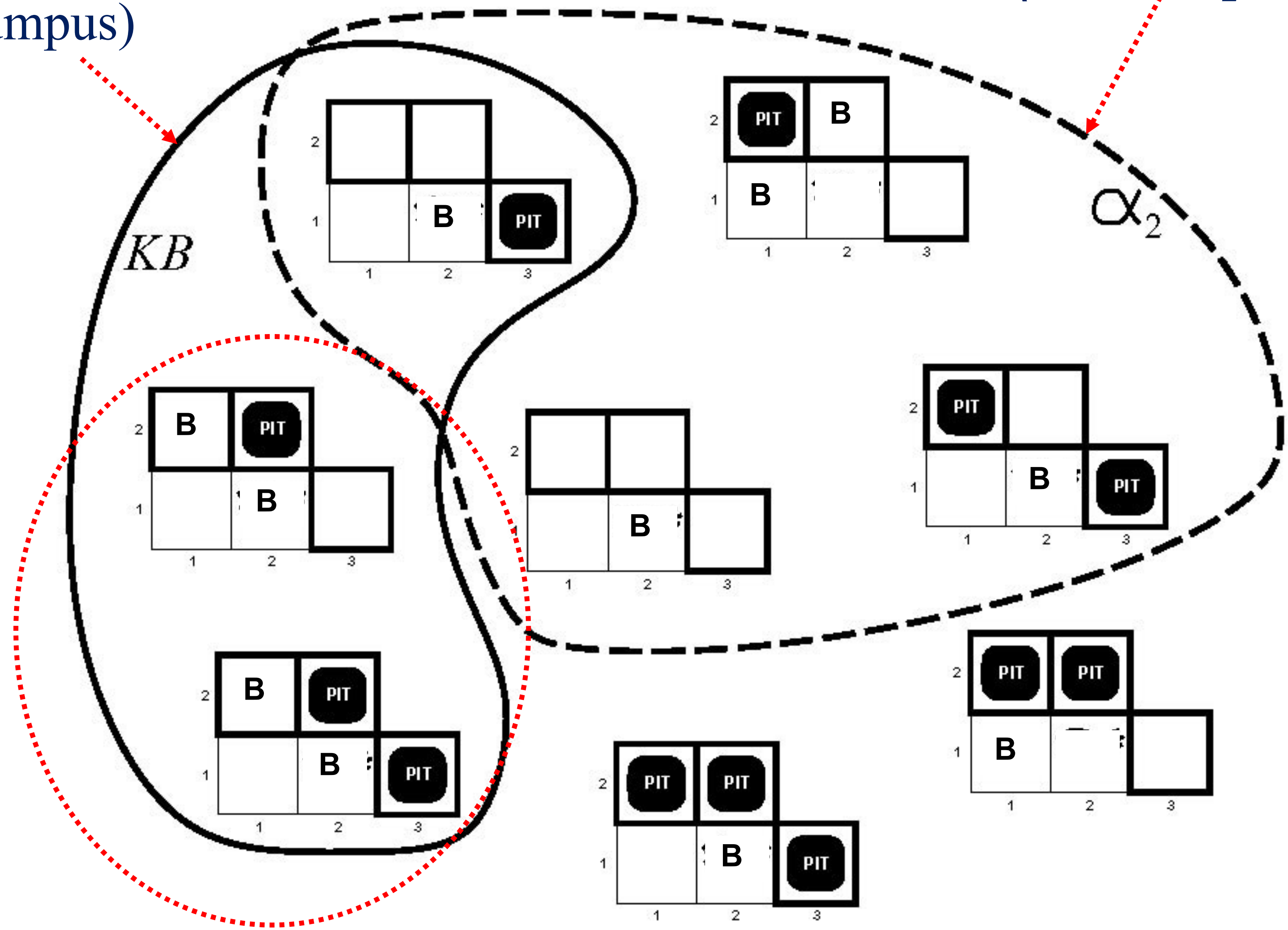
μοντέλα στα οποία αληθεύει ο $α_1$

$$\text{ΒΓ} \not\models \alpha_2$$

Όχι όρυγμα στο (2,2).

μοντέλα στα οποία αληθεύει η ΒΓ (για απλοποιημένο κόσμο Wumpus)

μοντέλα στα οποία αληθεύει ο α_2



Αλγόριθμοι εξαγωγής συμπερασμάτων

- **$\alpha \vdash_i \beta$** : Ο αλγόριθμος i «παράγει» από τον τύπο α τον β.
  - Μπορεί να δίνουμε συγκεκριμένους τύπους α και β και να ρωτάμε αν $\alpha \vdash_i \beta$ (**ερώτηση**).
  - Μπορεί να δίνουμε συγκεκριμένο τύπο α και να θέλουμε να μας **απαριθμήσει** όλους τους τύπους β για τους οποίους $\alpha \vdash_i \beta$.
- «Ο i είναι **ορθός**»: Αν $\alpha \vdash_i \beta$, τότε $\alpha \models \beta$.
  - Οι τύποι που παράγει έπονται **πράγματι** ταυτολογικά από τον α.
- «Ο i είναι **πλήρης**»: Αν $\alpha \models \beta$, τότε $\alpha \vdash_i \beta$.
  - Παράγει **όλους** τους τύπους που έπονται ταυτολογικά από τον α.

Συντακτικό προτασιακής λογικής

Ορισμός με γραμματική χωρίς συμφραζόμενα:

πρόταση → ατομική-πρόταση | σύνθετη-πρόταση

ατομική-πρόταση → True | False | σύμβολο

σύμβολο → B | P | S | … | $B_{1,1}$ | $P_{2,1}$ | …

σύνθετη-πρόταση → $\neg$ πρόταση

| (πρόταση $\wedge$ πρόταση)

| (πρόταση $\vee$ πρόταση)

| (πρόταση $\Rightarrow$ πρόταση)

| (πρόταση $\Leftrightarrow$ πρόταση)

Τα σύμβολα « $\vdash_i$» και « $\models$ » δεν είναι σύμβολα της γλώσσας. Δεν περιέχονται ποτέ σε τύπους της γλώσσας. Τα χρησιμοποιούμε εμείς για να συζητάμε για τους τύπους της γλώσσας (ανήκουν σε μια μετα-γλώσσα).

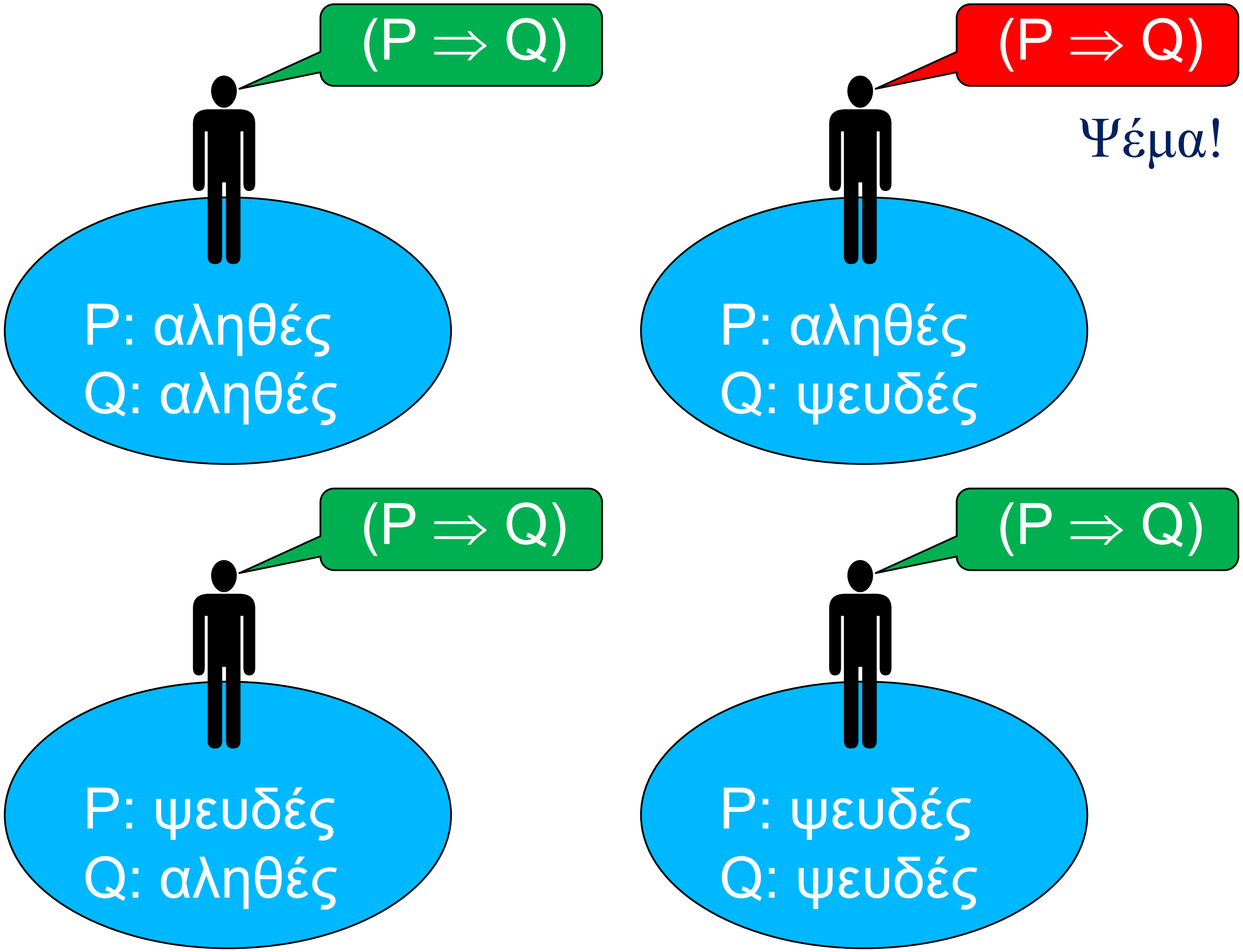
Σημασιολογία προτασιακής λογικής

- Η σημασία των True και False είναι πάντα *αληθές* και *ψευδές* αντίστοιχα.
- Η σημασία κάθε **συμβόλου** καθορίζεται από το **μοντέλο**.
- Η σημασία των **σύνθετων προτάσεων** καθορίζεται από τη σημασία των συστατικών τους και τον πίνακα:

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
$false$	$false$	$true$	$false$	$false$	$true$	$true$
$false$	$true$	$true$	$false$	$true$	$true$	$false$
$true$	$false$	$false$	$false$	$true$	$false$	$false$
$true$	$true$	$false$	$true$	$true$	$true$	$true$

Διαισθητικά ο τύπος λέει «Αν η σημασία του P στο συγκεκριμένο κόσμο είναι *αληθές*, τότε και η σημασία του Q στο συγκεκριμένο κόσμο είναι *αληθές*». Αυτή η δήλωση είναι αληθής σε όλους τους κόσμους, εκτός από εκείνους όπου P αληθές και Q ψευδές (τρίτη γραμμή).

Πότε λέει ψέματα;



Παράδειγμα

- Ας θεωρήσουμε τον φ: $((\neg B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \Rightarrow \neg P_{1,2})$

α/α	$\mathbf{B_{1,1}}$	$\mathbf{P_{1,2}}$	$\mathbf{P_{2,1}}$	$\mathbf{(\neg B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})))}$	$\mathbf{\neg P_{1,2}}$	**φ**
1	T	T	T	F	F	T
2	T	T	F	F	F	T
3	T	F	T	F	T	T
4	T	F	F	F	T	T
5	F	T	T	T	F	F
6	F	T	F	T	F	F
7	F	F	T	T	T	T
8	F	F	F	T	T	T

- Ο φ δεν αληθεύει στα μοντέλα των γραμμών 5, 6, όπου αληθεύει το αριστερό αλλά όχι το δεξί. Αληθεύει σε όλα τα άλλα μοντέλα.
- Επίσης, $(\neg B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \not\models \neg P_{1,2}$, γιατί στα μοντέλα των γραμμών 5, 6 αληθεύει το αριστερό αλλά όχι το δεξί.

Άλλο παράδειγμα

- Ας θεωρήσουμε ψ: $((B_{1,1} \land (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1}))) \Rightarrow (P_{1,2} \lor P_{2,1}))$

α/α	$B_{1,1}$	$P_{1,2}$	$P_{2,1}$	$(B_{1,1} \land (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})))$	$(P_{1,2} \lor P_{2,1})$	ψ
1	T	T	T	T	T	T
2	T	T	F	T	T	T
3	T	F	T	T	T	T
4	T	F	F	F	F	T
5	F	T	T	F	T	T
6	F	T	F	F	T	T
7	F	F	T	F	T	T
8	F	F	F	F	F	T

- Ο ψ αληθεύει σε όλα τα μοντέλα (γραμμές). Είναι «ταυτολογία».
- Επίσης, $(B_{1,1} \land (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1}))) \models (P_{1,2} \lor P_{2,1})$, γιατί σε όσες γραμμές (μοντέλα) αληθεύει το αριστερό, αληθεύει και το δεξί.
- Γενικότερα **α** $\models$ **β** ανν ο τύπος **(α ⇒ β) είναι αληθής σε όλα τα μοντέλα** (ανν ο τύπος είναι ταυτολογία).

Η σχέση των $\models$ και $\Rightarrow$

- Παρατηρήστε ότι παρ' όλο που ο τύπος
  $((\neg B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \Rightarrow \neg P_{1,2})$
  αληθεύει σε <u>μερικά</u> μοντέλα, έχουμε:
  $(\neg B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \not\models \neg P_{1,2}$,
  που συμφωνεί με τη διαίσθησή μας ότι το $\neg P_{1,2}$ **δεν προκύπτει ως συμπέρασμα** από το αριστερό μέρος.
- Ενώ ο $((B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \Rightarrow (P_{1,2} \vee P_{2,1}))$ αληθεύει σε <u>όλα</u> τα μοντέλα (ταυτολογία) και επίσης:
  $(B_{1,1} \wedge (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))) \models (P_{1,2} \vee P_{2,1})$,
  που συμφωνεί με τη διαίσθησή μας ότι το $(P_{1,2} \vee P_{2,1})$ **προκύπτει ως συμπέρασμα** από το αριστερό μέρος.

Παράδειγμα στον κόσμο του Wumpus

- $\neg P_{1,1}$
  - Δεν υπάρχει όρυγμα στο (1,1).
- $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1}))$
  - Αν ρεύμα στο (1,1), τότε όρυγμα στα (1,2) ή (2,1).
- $(B_{2,1} \Rightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1}))$
  - Χρειάζεται αντίστοιχη πρόταση για κάθε τετράγωνο!
- $\neg B_{1,1}$
  - Δεν υπάρχει ρεύμα στο (1,1).
- $B_{2,1}$
  - Υπάρχει ρεύμα στο (2,1).

μερικοί από τους τύπους της Βάσης Γνώσης («ΒΓ», «KB»)

- **Συμπεράσματα** για ορύγματα στα (1,2) και (2,2);
  - Ισχύουν τα $ΒΓ \models P_{1,2}$, $ΒΓ \models \neg P_{1,2}$, $ΒΓ \models P_{2,2}$, $ΒΓ \models \neg P_{2,2}$;

Συμπεράσματα με αναζήτηση μοντέλων

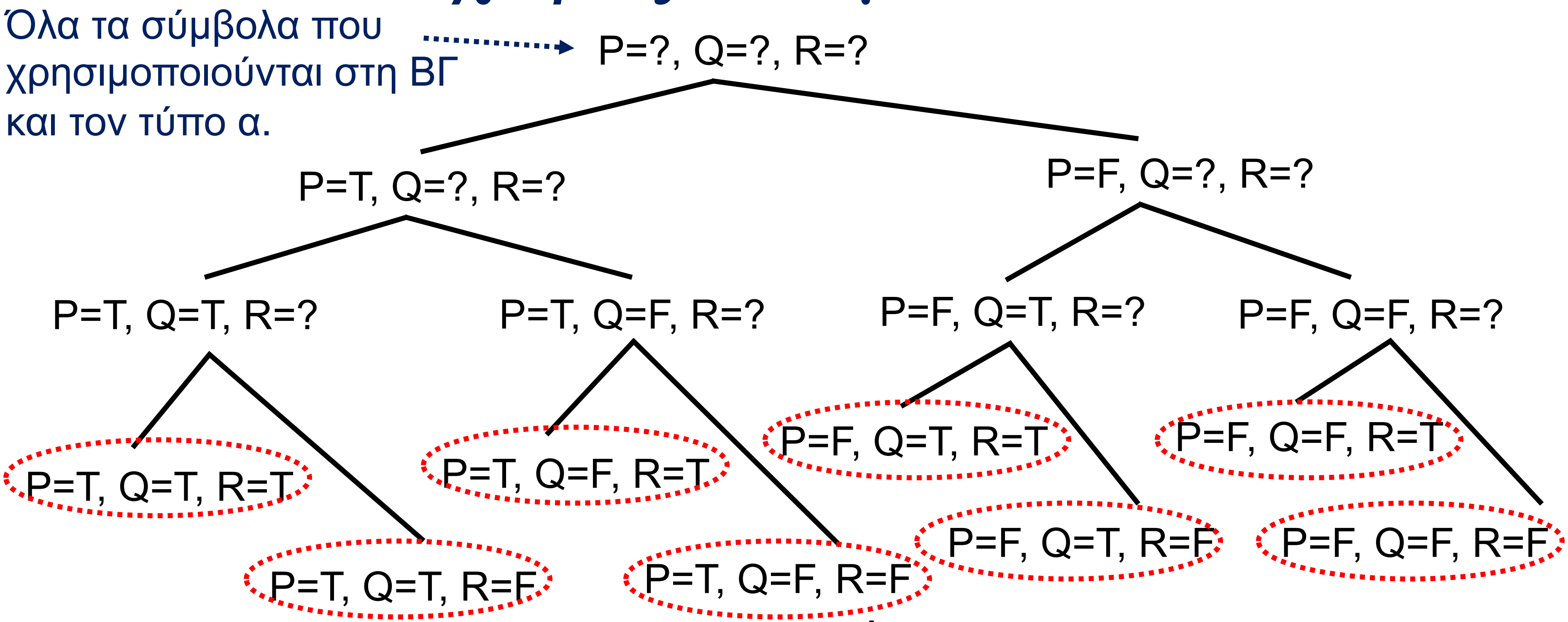
– Ισχύουν τα ΒΓ $\models P_{1,2}$, ΒΓ $\models \neg P_{1,2}$, ΒΓ $\models P_{2,2}$, ΒΓ $\models \neg P_{2,2}$;

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	KB
$false$	$false$	$false$	$false$	$false$	$false$	$false$	$false$
$false$	$false$	$false$	$false$	$false$	$false$	$true$	$false$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$false$	$true$	$false$	$false$	$false$	$false$	$false$	$false$
$false$	$true$	$false$	$false$	$false$	$false$	$true$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$false$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$true$	$\underline{true}$
$false$	$true$	$false$	$false$	$true$	$false$	$false$	$false$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$true$	$true$	$true$	$true$	$true$	$true$	$true$	$false$

Συμπεράσματα με αναζήτηση μοντέλων

- Σε όλα τα μοντέλα όπου αληθεύει η ΒΓ, αληθεύει και ο $\neg P_{1,2}$. Επομένως ΒΓ $\models \neg P_{1,2}$ και ΒΓ $\not\models P_{1,2}$.
- Για το $P_{2,2}$ δεν μπορούμε να εξαγάγουμε συμπέρασμα.
  - Δεν ισχύει ούτε ΒΓ $\models P_{2,2}$ ούτε ΒΓ $\models \neg P_{2,2}$.
  - Δηλαδή ΒΓ $\not\models P_{2,2}$ και ΒΓ $\not\models \neg P_{2,2}$.
- Αν χρησιμοποιούμε μόνο 7 σύμβολα, υπάρχουν $2^7 = 128$ μοντέλα.
  - Γενικότερα χρειάζεται να ελέγξουμε **$O(2^n)$** μοντέλα (γραμμές του πίνακα), όπου n ο αριθμός των συμβόλων που χρησιμοποιούνται.

Ο χώρος των μοντέλων



- Προσπαθούμε να ελέγξουμε αν **ΒΓ** $\models$ **α**.
  - Έστω P, Q, R όλα τα σύμβολα που χρησιμοποιούνται στη ΒΓ και τον α.
- Επιτρέπουμε και **ημιτελή μοντέλα** (στους κόμβους που δεν είναι φύλλα), τα οποία δεν δίνουν τιμές σε όλα τα σύμβολα.
- Θέλουμε να εξετάσουμε όλα τα **πλήρη μοντέλα** (τα φύλλα).
  - Πρέπει σε κάθε φύλλο είτε (i) **να μην αληθεύει η ΒΓ** είτε (ii) **να αληθεύει και η ΒΓ και ο α**.

Αναζήτηση στο χώρο των μοντέλων με DFS

- **Αρχική κατάσταση**: ημιτελές μοντέλο που δε δίνει καμία τιμή.
- **Μεταβάσεις**: αποκτά τιμή το αριστερότερο σύμβολο που δεν έχει.
- **Τελικές καταστάσεις**: όλα τα σύμβολα έχουν τιμές.
- Εξερευνούμε το **δέντρο με DFS**.
  – Δεν σταματάμε όταν συναντούμε τελική κατάσταση.
- **Ο αλγόριθμος** της επόμενης διαφάνειας καλείται αναδρομικά με **όρισμα έναν κόμβο** του δέντρου.
  – Δηλαδή ένα μοντέλο, πιθανώς ημιτελές, που αρχικά είναι η ρίζα.
  – Επιστρέφει **true ανν σε κάθε φύλλο του υπο-δέντρου** με ρίζα τον κόμβο ισχύει ότι: (i) δεν αληθεύει η ΒΓ **ή** (ii) αληθεύει και η ΒΓ και ο τύπος α.

Έλεγχος μοντέλων με DFS

function TT-ENTAILS?(KB, α) **returns** *true* or *false*

σύμβολα στα οποία δεν δίνει τιμή το (ημιτελές) μοντέλο

symbols ← a list of the proposition symbols in KB and α
return TT-CHECK-ALL(KB, α, *symbols*, [ ])

το (ημιτελές) μοντέλο

---

function TT-CHECK-ALL(KB, α, *symbols*, *model*) **returns** *true* or *false*
if EMPTY?(*symbols*) **then**
  if PL-TRUE?(KB, *model*) **then return** PL-TRUE?(α, *model*)
  else return *true*

Αληθεύει η ΒΓ στο μοντέλο;

else do
  P ← FIRST(*symbols*); *rest* ← REST(*symbols*)

έλεγχος των δύο παιδιών

  return TT-CHECK-ALL(KB, α, *rest*, EXTEND(P, *true*, *model*) **and**
    TT-CHECK-ALL(KB, α, *rest*, EXTEND(P, *false*, *model*)

- Καλούμε την TT-Check-All για ένα συγκεκριμένο κόμβο (ημιτελές μοντέλο), αρχικά τη ρίζα.
- Επιστρέφουμε Τ ανν σε κάθε φύλλο του υποδέντρου (πλήρες μοντέλο), (i) δεν αληθεύει η ΒΓ <u>ή</u> (ii) αληθεύουν η ΒΓ και ο α.

Χαρακτηριστικά του TT-Entails?

- Πολυπλοκότητα **χρόνου $O(2^n)$**, όπου n ο αριθμός των συμβόλων της ΒΓ και της πρότασης α.
  - Παράγουμε $2^1 + 2^2 + \ldots + 2^n = O(2^n)$ κόμβους.
- Πολυπλοκότητα **χώρου O(n)**.
  - Ψάχνουμε το χώρο των ημιτελών μοντέλων με DFS. Έχουμε πάντα b = 2 και μέγιστο βάθος n.
- **Ορθός** (αν ΒΓ $\vdash_i$ α, τότε ΒΓ $\models$ α).
  - Για να απαντήσει «ΒΓ $\vdash_i$ α», έχει βεβαιωθεί πρώτα ότι σε όλα τα φύλλα όπου αληθεύει η ΒΓ αληθεύει και η πρόταση α.
- **Πλήρης** (αν ΒΓ $\models$ α, τότε ΒΓ $\vdash_i$ α).
  - Με την έννοια ότι αν του δώσουμε συγκεκριμένη ΒΓ και α με ΒΓ $\models$ α, θα απαντήσει «ΒΓ $\vdash_i$ α». Θα διαπιστώσει ότι σε όλα τα φύλλα όπου αληθεύει η ΒΓ, αληθεύει και η α, οπότε θα απαντήσει «ΒΓ $\vdash_i$ α».
  - Δεν μπορεί, όμως, να απαριθμήσει όλα τα α για τα οποία ισχύει ότι ΒΓ $\models$ α.

Βιβλιογραφία

- Russel & Norvig (4η έκδοση): κεφ. 7 ως και ενότητα 7.4.
- Βλαχάβας κ.ά.: εισαγωγή κεφ. 9, ενότητα 9.1 (εκτός των υπο-ενοτήτων 9.1.1, 9.1.2).