ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Τεχνητή Νοημοσύνη

8η διάλεξη (2023-24)

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

- Περισσότερα για την **εξαγωγή συμπερασμάτων** με **προτασιακή λογική**:
  - **Εγκυρότητα** και **ικανοποιησιμότητα**.
  - Εξαγωγή συμπερασμάτων με **αναζήτηση απόδειξης**.
  - Εξαγωγή συμπερασμάτων με τον **κανόνα της ανάλυσης** (resolution).

Ταυτολογική ισοδυναμία

- **α ≡ β** σημαίνει: α ⊨ β και β ⊨ α.
  - Οι προτάσεις α και β αληθεύουν **στα ίδια ακριβώς μοντέλα**.

$$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha) \quad \text{commutativity of } \wedge$$
$$(\alpha \vee \beta) \equiv (\beta \vee \alpha) \quad \text{commutativity of } \vee$$
$$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma)) \quad \text{associativity of } \wedge$$
$$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma)) \quad \text{associativity of } \vee$$
$$\neg(\neg\alpha) \equiv \alpha \quad \text{double-negation elimination}$$
$$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}$$
$$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta) \quad \text{implication elimination}$$
$$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination}$$
$$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta) \quad \text{de Morgan}$$
$$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta) \quad \text{de Morgan}$$
$$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \quad \text{distributivity of } \wedge \text{ over } \vee$$
$$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad \text{distributivity of } \vee \text{ over } \wedge$$

Εγκυρότητα και ικανοποιησιμότητα

- «Η α είναι **ταυτολογία**» ή «**έγκυρη**» σημαίνει: α ≡ True.
  - Η πρόταση α αληθεύει **σε όλα τα μοντέλα**.
- Θεώρημα: **α ⊨ β** ανν **(α ⇒ β) έγκυρη (ταυτολογία)**.
  - Απόδειξη με ορισμό του ⊨ και πίνακα αληθείας του ⇒. *(Άσκηση)*
- «Η α είναι **ικανοποιήσιμη**» σημαίνει υπάρχει μοντέλο στο οποίο η α είναι αληθής.
  - Π.χ. η πρόταση (P ∧ ¬P) είναι *μη* ικανοποιήσιμη.
  - Επομένως **α έγκυρη** ανν **¬α μη ικανοποιήσιμη**.
- **α ⊨ β ανν (α ∧ ¬β) μη ικανοποιήσιμη**
  - Η γνωστή μας «απαγωγή σε άτοπο».
  - α ⊨ β ανν (α ⇒ β) έγκυρη ανν ¬(α ⇒ β) μη ικανοποιήσιμη ανν ¬(¬α ∨ β) μη ικανοποιήσιμη ανν (α ∧ ¬β) μη ικανοποιήσιμη.
  - Δυστυχώς ο έλεγχος ικανοποιησιμότητας στην προτασιακή λογική είναι **NP-πλήρες** πρόβλημα.

Κανόνες εξαγωγής συμπερασμάτων

- Παράγουν συμπεράσματα από τύπους που υπάρχουν ήδη στη ΒΓ.
  - «α, β ⊢ γ» σημαίνει ότι από τους τύπους α και β ο κανόνας παράγει τον τύπο γ.
- $\{(\alpha \Rightarrow \beta), \alpha\} \vdash \beta$ (Modus Ponens)
  - Απόδειξη ορθότητας με πίνακα αληθείας. *(Άσκηση)*
  - Αποδεικνύουμε δηλαδή ότι: $((\alpha \Rightarrow \beta) \wedge \alpha) \models \beta$.
- $\{(\alpha \Rightarrow \beta), \neg\boldsymbol{\beta}\} \vdash \neg\boldsymbol{\alpha}$ (Modus Tollens)
  - Απόδειξη με πίνακα αληθείας. *(Άσκηση)*
- $(\alpha \wedge \beta) \vdash \alpha$ (απαλοιφή σύζευξης)
- $(\alpha \Leftrightarrow \beta) \vdash ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ (απαλοιφή ⇔)
- Μπορούν να χρησιμοποιηθούν και όλες οι ταυτολογικές ισοδυναμίες, προς οποιαδήποτε κατεύθυνση.

Εξαγωγή συμπεράσματος με **αναζήτηση απόδειξης**

- **Αρχική κατάσταση**: η ΒΓ στην αρχική της μορφή.
- **Τελική κατάσταση**: η ΒΓ σε μορφή που να περιέχει τον προς απόδειξη τύπο.
- **Τελεστές μετάβασης**: κανόνες εξαγωγής συμπερασμάτων
  - Προσθέτουν τύπους στη ΒΓ.
- **Απόδειξη**: μονοπάτι από αρχική σε τελική κατάσταση.
- Στην πράξη η αναζήτηση απόδειξης μπορεί να είναι **πιο αποδοτική** από τον εξαντλητικό έλεγχο μοντέλων.
  - Αν θέλουμε να δείξουμε ότι **α** $\models$ **β**, μπορούμε να ελέγξουμε **εξαντλητικά** όλα τα μοντέλα με τον TT-Entails?.
  - Ή να κατασκευάσουμε μια **απόδειξη** (μονοπάτι) από αρχική ΒΓ που περιέχει μόνο το α σε τελική ΒΓ που περιέχει και το β.
- Αλλά πάλι στη χειρότερη περίπτωση, **εκθετικός χρόνος**.
  - Διαφορετικά θα είχαμε τρόπο να λύσουμε το **NP-πλήρες** πρόβλημα της ικανοποιησιμότητας σε πολυωνυμικό χρόνο!

Παράδειγμα στον κόσμο του Wumpus

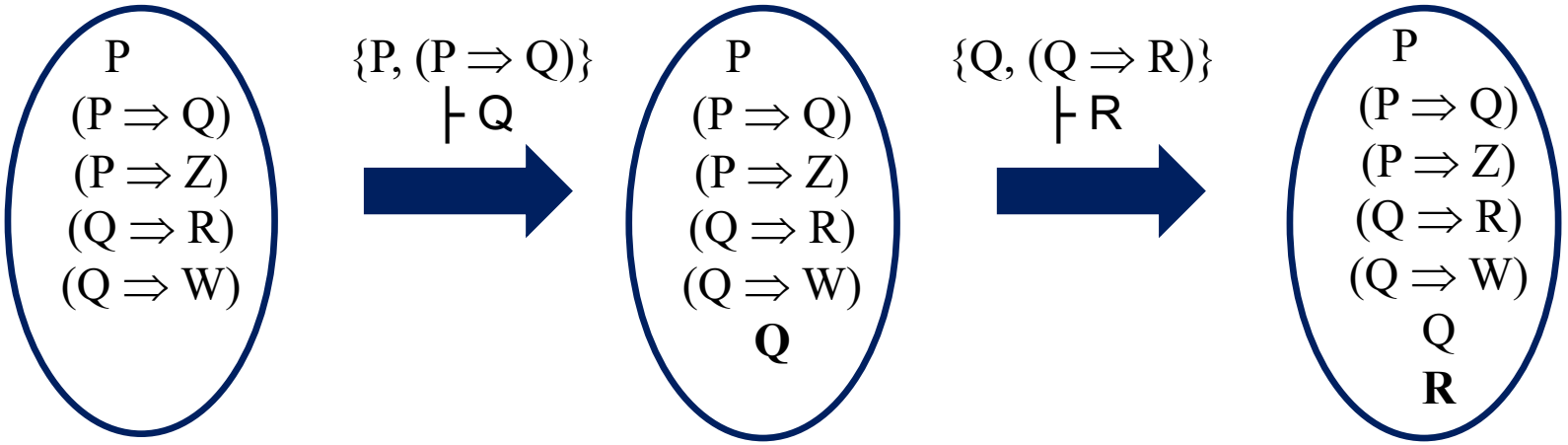
- Έστω ότι η ΒΓ περιέχει αρχικά:
  - R_1: $\neg P_{1,1}$ R_4: $\neg B_{1,1}$ R_5: $B_{2,1}$
  - R_2: $(B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}))$
  - R_3: $(B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1}))$
- Θέλουμε να αποδείξουμε $(\neg P_{1,2} \wedge \neg P_{2,1})$.
- Εφαρμογή απαλοιφής $\Leftrightarrow$ στο R_2:
  - R_6: $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$
- Εφαρμογή απαλοιφής σύζευξης στο R_6:
  - R_7: $((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$
- Εφαρμογή αντιθετοαντιστροφής στο R_7:
  - R_8: $(\neg B_{1,1} \Rightarrow \neg (P_{1,2} \vee P_{2,1}))$
- Modus Ponens με R_8 και R_4:
  - R_9: $\neg (P_{1,2} \vee P_{2,1})$
- Εφαρμογή κανόνα De Morgan στο R_9 δίνει το ζητούμενο.

Η αναζήτηση μπορεί να είναι πιο γρήγορη…

- Έστω ΒΓ: {P, (P ⇒ Q), (P ⇒ Z), (Q ⇒ R), (Q ⇒ W)}
- Θέλουμε να δείξουμε: ΒΓ ⊨ R.
- Ο TT-Entails? θα εξετάσει πίνακα με 2^5 γραμμές:

P	Q	R	Z	W	ΒΓ
T	T	T	T	T	T
F	T	T	T	T	F
…	…	…	…	…	…

- Με ιδανική ευρετική, η αναζήτηση απόδειξης (ή «απόδειξη θεωρημάτων») θα χρειαστεί μόνο 2 βήματα:



Κανονική συζευκτική μορφή

- Κάθε πρόταση της ΠΛ είναι ταυτολογικά ισοδύναμη με μια πρόταση σε **κανονική συζευκτική μορφή** (CNF):
  - $(l_{1,1} \vee \ldots \vee l_{1,k1}) \wedge \ldots \wedge (l_{n,1} \vee \ldots \vee l_{n,kn})$
  - Κάθε $l_{i,j}$ είναι **σύμβολο** (π.χ. P) ή **άρνηση συμβόλου** (π.χ. ¬P).
- Οπότε **και η ΒΓ** μπορεί να γραφτεί σε **CNF**.
  - Μπορούμε να θεωρήσουμε ότι **η ΒΓ** είναι ένας μόνο τύπος, **μία μεγάλη σύζευξη** όλων των τύπων που περιέχει η ΒΓ.
  - Μετατρέπουμε τη μεγάλη σύζευξη σε CNF, οπότε **η ΒΓ γίνεται μια μεγάλη σύζευξη διαζεύξεων**, η κάθε μία της μορφής $(l_{i,1} \vee \ldots \vee l_{i,ki})$.
  - Μπορούμε κατόπιν να θεωρήσουμε ότι η ΒΓ περιέχει ως **ξεχωριστούς τύπους** όλες τις διαζεύξεις $(l_{i,1} \vee \ldots \vee l_{i,ki})$ που προκύπτουν.

Μετατροπή σε CNF

Συμβουλή:
Εφαρμόζετε τα βήματα πάντα **με αυτή τη σειρά** και μην παραλείπετε **παρενθέσεις**!

- **Μέθοδος** μετατροπής σε CNF:
  - Π.χ. ξεκινώντας από $(B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}))$.
- Βήμα 1: **απαλοιφή** $\Leftrightarrow$.
  - $((B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}))$
- Βήμα 2: **απαλοιφή** $\Rightarrow$.
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg(P_{1,2} \vee P_{2,1}) \vee B_{1,1}))$
- Βήμα 3: **μεταφορά** $\neg$ στο εσωτερικό ως τα σύμβολα.
  - Με χρήση κανόνων De Morgan και διπλής άρνησης.
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge ((\neg P_{1,2} \wedge \neg P_{2,1}) \vee B_{1,1}))$
- Βήμα 4: **επιμερισμός** των $\vee$.
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}))$

Ο κανόνας της ανάλυσης (resolution)

- Μετατρέπουμε πρώτα τους τύπους της υπόθεσης (π.χ. της ΒΓ) σε **CNF**.
  - Παίρνουμε ένα σύνολο από διαζευκτικούς τύπους, ο καθένας της μορφής $(l_{i,1} \vee \ldots \vee l_{i,ki})$.
  - Κάθε $l_{i,j}$ («literal») είναι **σύμβολο** (π.χ. P) ή **άρνηση συμβόλου** (π.χ. ¬P).
- **Ο κανόνας της ανάλυσης**:
  - Αν το l_i **είναι η άρνηση του** m_j ή αντίστροφα, τότε:
  - $(l_1 \vee \ldots \vee l_k), (m_1 \vee \ldots \vee m_n) \vdash (l_1 \vee \ldots \vee l_{i-1} \vee l_{i+1} \vee \ldots \vee l_k \vee m_1 \vee \ldots \vee m_{j-1} \vee m_{j+1} \vee \ldots \vee m_n)$
  - Ενώνουμε τις δύο διαζεύξεις και αφαιρούμε τα l_i και m_j.
  - Αφαιρούμε από τη διάζευξη που παράγεται τυχόν **πολλαπλά αντίγραφα** των ίδιων l και m.

Ορθότητα του κανόνα της ανάλυσης

- Πρέπει να **δείξουμε** ότι ο **κανόνας της ανάλυσης**:

  $(l_1 \vee \ldots \vee l_k), (m_1 \vee \ldots \vee m_n) \vdash (l_1 \vee \ldots \vee l_{i-1} \vee l_{i+1} \vee \ldots \vee l_k \vee m_1 \vee \ldots \vee m_{j-1} \vee m_{j+1} \vee \ldots \vee m_n)$

  είναι **ορθός**.

- Δηλαδή ότι **όποτε** (σε όποια μοντέλα) **αληθεύουν** τα $(l_1 \vee \ldots \vee l_k)$ και $(m_1 \vee \ldots \vee m_n)$, τότε αληθεύει και το: $(l_1 \ldots \vee l_{i-1} \vee l_{i+1} \ldots \vee l_k \vee m_1 \vee \ldots \vee m_{j-1} \vee m_{j+1} \vee \ldots \vee m_n)$

Ορθότητα του κανόνα της ανάλυσης

- Έστω ότι βρισκόμαστε σε μοντέλο όπου **αληθεύουν** τα **$(l_1 \vee ... \vee l_k)$** και **$(m_1 \vee ... \vee m_n)$**.
  - Και κάποιο **l_i είναι η άρνηση** κάποιου **m_j**.
- Αν το **l_i είναι αληθές** σε αυτό το μοντέλο,
  - τότε το **m_j είναι ψευδές** (αφού το ένα είναι η άρνηση του άλλου).
  - Αφού όμως αληθεύει $(m_1 \vee ... \vee m_n)$, τότε είναι **αληθές κάποιο από τα υπόλοιπα *m***.
  - Άρα **αληθεύει** το **$(m_1 \vee ... \vee m_{j-1} \vee m_{j+1} \vee ... \vee m_n)$**.
- Αν το **l_i είναι ψευδές** σε αυτό το μοντέλο,
  - τότε αφού αληθεύει $(l_1 \vee ... \vee l_k)$, είναι **αληθές κάποιο από τα υπόλοιπα *l***.
  - Άρα **αληθεύει** το **$(l_1 \vee ... \vee l_{i-1} \vee l_{i+1} \vee ... \vee l_k)$**.
- Άρα **και στις δύο περιπτώσεις αληθεύει** το:
  - **$(l_1 \vee ... \vee l_{i-1} \vee l_{i+1} \vee ... \vee l_k) \vee (m_1 \vee ... \vee m_{j-1} \vee m_{j+1} \vee ... \vee m_n)$**.

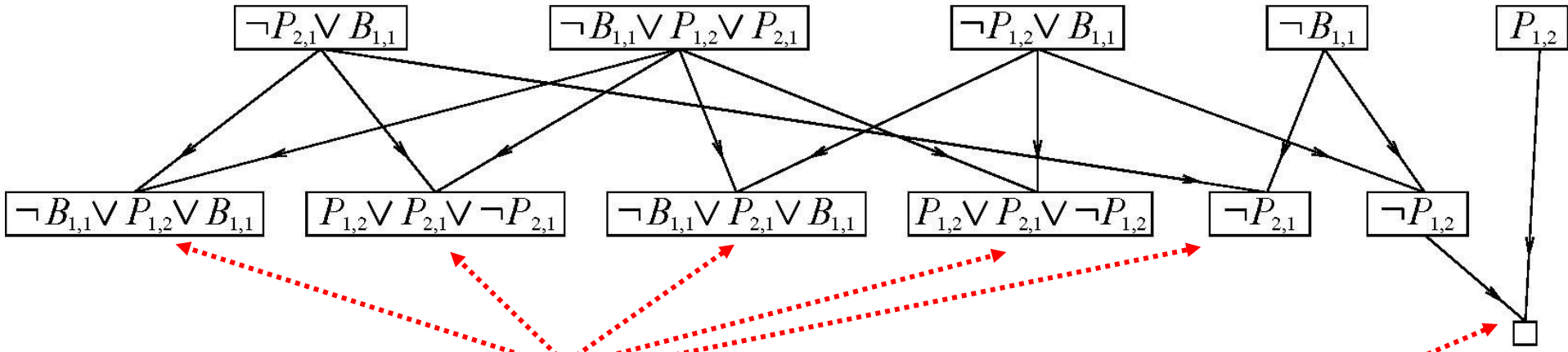
Εξαγωγή συμπερασμάτων με ανάλυση

- Θέλουμε να εξετάσουμε αν **ΒΓ** $\models$ **α**.
- Μετατρέπουμε το **(ΒΓ** $\wedge \neg$**α) σε CNF**.
  - Δηλαδή προσθέτουμε την άρνηση του προς απόδειξη τύπου στη ΒΓ και μετατρέπουμε τη ΒΓ σε CNF.
  - Και «σπάμε» τον τύπο CNF στις διαζεύξεις του.
- Εφαρμόζουμε τον **κανόνα της ανάλυσης όσο είναι δυνατόν**.
  - Μέχρι να μην μπορούμε να παραγάγουμε νέους τύπους.
- **Απαντούμε ΒΓ** $\vdash_i$ **α**, αν παραχθεί η **κενή διάζευξη** $\square$.
  - Η κενή διάζευξη δείχνει ότι καταλήξαμε σε **άτοπο**. Προκύπτει όταν έχουμε στη ΒΓ ένα σύμβολο και την άρνησή του.
- **Απαντούμε ΒΓ** $\not\vdash_i$ **α**, αν **δεν παραχθεί** $\square$.

Παράδειγμα στον κόσμο του Wumpus

Προσοχή: Οι τύποι πρέπει να είναι **διαζεύξεις**. Αν ένας τύπος είναι σύζευξη διαζεύξεων, τον σπάμε σε ξεχωριστούς τύπους-διαζεύξεις.

Προσοχή: Σε κάθε «ζευγάρωμα» τύπων, φεύγει πάντα **ένα μόνο σύμβολο** από τον έναν τύπο **και ένα μόνο σύμβολο** από τον άλλον.



Άχρηστα συμπεράσματα.

Κενή διάζευξη (άτοπο).

Ο αλγόριθμος **τερματίζει πάντα**, γιατί υπάρχει πεπερασμένος αριθμός διαζεύξεων (χωρίς πολλαπλά αντίγραφα του ίδιου όρου) που μπορούν να κατασκευαστούν από τα πεπερασμένα σύμβολα της αρχικής ΒΓ και της προς απόδειξη πρότασης.

Αλγόριθμος απόδειξης με ανάλυση

Οι διαζεύξεις της CNF μορφής.

function PL-RESOLUTION(KB, α) **returns** *true* or *false*

clauses ← the set of clauses in the CNF representation of $KB \wedge \neg\alpha$
new ← { }
loop do
  for each C_i, C_j **in** *clauses* **do**
    resolvents ← PL-RESOLVE(C_i, C_j)
    if *resolvents* contains the empty clause **then return** *true*
    new ← *new* ∪ *resolvents*
  if *new* ⊆ *clauses* **then return** *false*
  clauses ← *clauses* ∪ *new*

Εφαρμογή του κανόνα resolution.

Αν δεν καταφέραμε να παραγάγουμε καμία νέα διάζευξη, σταματάμε και απαντάμε $B\Gamma \not\vdash_i \alpha$.

Ορθότητα και πληρότητα

- **Ορθότητα** (αν ΒΓ $\vdash_i \alpha$, τότε ΒΓ $\models \alpha$).
  - Αν ΒΓ $\vdash_i \alpha$, τότε παρήχθη $\square$ με διαδοχικές εφαρμογές του κανόνα της ανάλυσης, ξεκινώντας από $(\text{ΒΓ} \wedge \neg\alpha)$.
  - Οπότε $(\text{ΒΓ} \wedge \neg\alpha) \models \square$, λόγω της ορθότητας του κανόνα της ανάλυσης. Άρα σε όλα τα μοντέλα όπου αληθεύει η $(\text{ΒΓ} \wedge \neg\alpha)$ αληθεύει και η $\square$. Αλλά η $\square$ δεν αληθεύει ποτέ.
  - Άρα η $(\text{ΒΓ} \wedge \neg\alpha)$ δεν αληθεύει σε κανένα μοντέλο, δηλαδή είναι μη ικανοποιήσιμη. Οπότε, ΒΓ $\models \alpha$.
  - Γιατί, σύμφωνα με τα προηγούμενα, ΒΓ $\models \alpha$ ανν $(\text{ΒΓ} \wedge \neg\alpha)$ μη ικανοποιήσιμη.
- **Πληρότητα** (αν ΒΓ $\models \alpha$, τότε ΒΓ $\vdash_i \alpha$).
  - Δηλαδή αν ΒΓ $\models \alpha$, τότε παράγεται η κενή διάζευξη.
  - Η απόδειξη παραλείπεται.

Βιβλιογραφία

- Russel & Norvig (4η έκδοση): ενότητα 7.5 ως και υπο-ενότητα 7.5.2, χωρίς την υπο-ενότητα «πληρότητα της ανάλυσης»
  - Όσοι ενδιαφέρονται μπορούν να διαβάσουν προαιρετικά και την υπο-ενότητα «πληρότητα της ανάλυσης».
- Βλαχάβας κ.ά: ενότητες 9.1.1, 9.1.2.