1 A Logical Problem

A prison warden has three prisoners summoned and announces to them the following:

" I have here five discs differing only in color: three white and rwo black. Without letting you know which I have chosen,I shall fasten one of them to each of you between his shoulders; outside, that is, your direct visual field.
At that point, you will be left at your leisure to consider your companions and their respective discs. The first to be able to deduce his own color will be the one to benefit. His conclusion, moreover, must be founded upon logical *and not simply probabilistic* reasons. "

How can the subiects solve the problem?

The Classical Solution

Case 1 Two blacks one white: A is white, **B, C are black**

Step 1.1 A *announces 'I know my color'* *Reasoning*: _____

Step 1.2 B, C *announce 'I know my color'* *Reasoning*: _____

Case 2 Two whites one black: **A is black**, B, C are white

Step 2.1 A, B, C *announce 'I do not know my color'*

Step 2.2 B, C *announce 'I know my color'*

Reasoning of B: If I were black, we would be in Case 1, therefore C would have announced *'I know my color'* at Step 2.1.

Reasoning of C: _____

Step 2.3 A *announces 'I know my color'*

Reasoning of A: If I were either white or black, my announcement at Step 2.1 would have been *'I do not know my color'* ; and the same holds for B and C. Therefore, if we are all white we would all have, after Step 2.1, the same information we had at the beginning. So B, C could not have announced *'I know my color'* at Step 2.2.

Alternative reasoning of A:
After Step 2.1 everybody knows that at most one is black (by Case 1).
Since B, C announce *'I know my color'* at Step 2.2. they must have used that information to reach their conclusion; if I were white no conclusion would be reached, therefore I must be black.

Case 3 Three whites: **A, B, C are white**

Step 3.1 A, B, C *announce 'I do not know my color'*

Step 3.2 A, B, C *announce 'I do not know my color'*

Step 3.3 A, B, C *announce 'I know my color'*

Reasoning of A: If I were black, we would be in Case 2 so B, C could not have announced *'I do not know my color'* at Step 2.2.

Reasoning of B: _____ *Reasoning of C*: _____

Σχετικό εκπαιδευτικό υλικό

Fagin – Halpern – Moses – Vardi *Reasoning About Knowledg*e

Ενότητα 1.1. The Muddy Children Puzzle

Προτεινόμενες ασκήσεις

1 Συμπληρώστε τον συλλογισμό των B, C στο Step 1.2.

2 α Εξηγείστε γιατί οι παίκτες δεν είναι σε θέση να βρούν το χρώμα τους μετά το Step 3.1.

β Εξηγείστε για ποιό λόγο οι παίκτες είναι σε θέση να βρούν το χρώμα τους μετά το Step 3.2, παρόλο που η πληροφορία που πήραν σε αυτό το βήμα (A, B, C do not know their color) μπορεί να θεωρηθεί ως ήδη γνωστή, μετά το Step 3.1.

3 *Reasoning About Knowledg*e Exercise 1.3.

4 Εξετάστε το Muddy Children Puzzle όταν k = 2 (δύο λερωμένα παιδάκια).
Πόσες επαναλήψεις της ερώτησης του πατέρα χρειάζονται για να καταλάβουν όλα τα παιδάκια άν είναι λερωμένα; Ποιούς συλλογισμούς κάνουν;

2 Possible worlds semantics (Kripke models) για το δίλημμα των τριών φυλακισμένων

States - Primitive propositions:

Οι *ατομικές προτάσεις* (primitive propositions) είναι: AisWh , BisWh , CisWh

Κάθε ατομική πρόταση θεωρείται ως προτασιακό γράμμα.

Κάθε *κατάσταση* (state) θεωρείται ως ***απόδοση τιμών αλήθειας*** στις ατομικές προτάσεις.

AisWh	is true at state s (συμβολισμός: s(AisWh) = true)	if	A is white at s
BisWh	is true at state s (συμβολισμός: s(BisWh) = true)	if	B is white at s
CisWh	is true at state s (συμβολισμός: s(CisWh) = true)	if	C is white at s

Possibility relations:

Οι καταστάσεις sj , sk θεωρούνται *ισοδύναμες για τον παίκτη* X (συμβολισμός: $sj \approx_X sk$) όταν τα χρώματα των άλλων παικτών είναι τα ίδια στις sj , sk .

Για κάθε παίκτη X: $sj \approx_X sj$, $j = 1,2,3,$ επειδή sj(YisWh) = sj(YisWh) , για κάθε άλλο παίκτη Y.

Σύμφωνα με το Kripke model:

Ο παίκτης Y ***γνωρίζει το χρώμα του*** στην κατάσταση u, όταν:

η φόρμουλα $(K_Y\ YisWh) \vee (K_Y\ \neg YisWh))$ αληθεύει στην u.

Ο παίκτης Y ***γνωρίζει την (ιδιότητα που εκφράζει η) φόρμουλα φ*** στην κατάσταση u, όταν:

η φόρμουλα $(K_Y\ \varphi)$ αληθεύει στην u.

Σχετικό εκπαιδευτικό υλικό

Fagin – Halpern – Moses – Vardi *Reasoning About Knowledg*e

Ενότητα 2.1. The Possible-Worlds Model

Προτεινόμενες ασκήσεις

1 *Reasoning About Knowledg*e Exercise 2.1.

2 Κατασκευάστε ένα μοντέλο Kripke για την Exercise 1.3 - *Reasoning About Knowledg*e.

3 Έστω ότι η φόρμουλα $(K_Y\ (\varphi \wedge \psi))$ αληθεύει στην κατάσταση s ενός μοντέλου Kripke M. Εξηγείστε με βάση τους γενικούς ορισμούς, ότι η φόρμουλα $((K_Y\ \varphi) \wedge (K_Y\ \psi))$ θα αληθεύει στην κατάσταση s του M.

3 Σημασιολογική ανάλυση του διλήμματος των τριών φυλακισμένων με Kripke models

Case 1 : *το μοντέλο M1*

States: s1 A:w - B:b - C:b (η ‘πραγματική’ κατάσταση)

s2 A:w - B:w - C:b s3 A:w - B:b - C:w

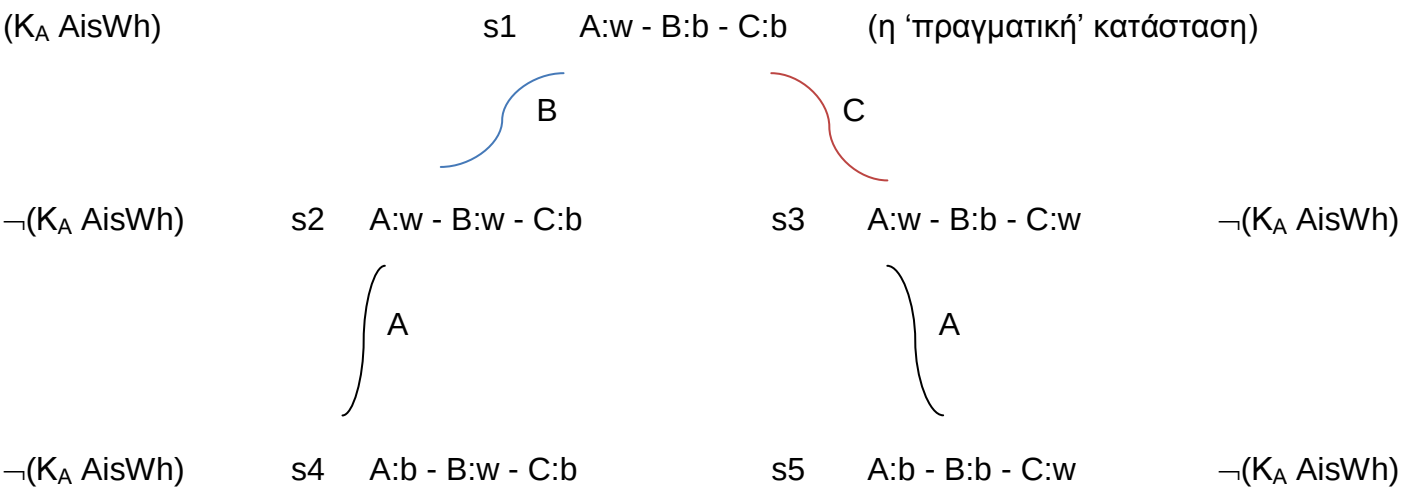
s4 A:b - B:w - C:b s5 A:b - B:b - C:w

	AisWh	BisWh	CisWh
s1	T	F	F
s2	T	T	F
s3	T	F	T
s4	F	T	F
s5	F	F	T

Possibility relations:

Για τον A: $s4 \approx_A s2$, $s5 \approx_A s3$ Για τον B: $s1 \approx_B s2$ Για τον C: $s1 \approx_C s3$

M1



Ανάλυση ***από την σκοπιά των*** B , C

1 (K_A AisWh) αληθεύει στην s1.

(K_A AisWh) δεν αληθεύει στις s2 , s3 , s4 , s5 .

(K_B (K_A AisWh)) δεν αληθεύει στην s1.

Για τον B οι καταστάσεις s1, s2 είναι ισοδύναμες ως προς τα χρώματα των δύο άλλων παικτών, αλλά είναι διακριτές ως προς *την γνώση του* A *για το χρώμα του* (τιμή αλήθειας της (K_A AisWh)) .

(K_C (K_A AisWh)) δεν αληθεύει στην s1.

Για τον C οι καταστάσεις s1, s3 είναι ισοδύναμες ως προς τα χρώματα των δύο άλλων παικτών, αλλά είναι διακριτές ως προς *την γνώση του* A *για το χρώμα του* (τιμή αλήθειας της (K_A AisWh)) .

2 Μετά το **Step 1.1:** οι B , C γνωρίζουν ότι αληθεύει η φόρμουλα (K_A AisWh) .

Επομένως έχουν πληροφορηθεί (λόγω του **1**) ότι η κατάσταση s1 είναι η μόνη δυνατή.

Σχετικό εκπαιδευτικό υλικό

Fagin – Halpern – Moses – Vardi *Reasoning About Knowledge*

Ενότητα 2.1. The Possible-Worlds Model

Προτεινόμενες ασκήσεις

1 Με βάση τους ορισμούς, εξηγείστε ότι στο μοντέλο *M1*:

Η φόρμουλα (K_A AisWh) αληθεύει *μόνο* στην κατάσταση s1 .

Οι φόρμουλες (K_B (K_A AisWh)) και (K_C (K_A AisWh)) δεν αληθεύουν στην s1 .

2 Εξηγήστε, χρησιμοποιώντας το μοντέλο *M1*, το λόγο που οι B, C δεν μπορούν να καταλάβουν το χρώμα τους στο **Step 1.1** του παιχνιδιού.

3 Έστω το παρακάτω μοντέλο *N1*

States: s1 A:w - B:b - C:b (η ‘πραγματική’ κατάσταση)
s2 A:w - B:w - C:b s3 A:w - B:b - C:w

	AisWh	BisWh	CisWh
s1	T	F	F
s2	T	T	F
s3	T	F	T

α Βρείτε τις Possibility relations του μοντέλου *N1*, για καθέναν από τους παίκτες.

β Με βάση τους ορισμούς, βρείτε σε ποιές καταστάσεις του μοντέλου *N1* αληθεύει καθε μία από τις φόρμουλες: $(K_A\ AisWh)$ $(K_B\ \neg BisWh)$ $(K_C\ \neg CisWh)$
$(K_B\ (K_A\ AisWh))$ $(K_C\ (K_A\ AisWh))$.

N1

$(K_A\ AisWh)$
$\neg(K_B\ \neg BisWh)$
$\neg(K_C\ \neg CisWh)$

s1 A:w - B:b - C:b (η ‘πραγματική’ κατάσταση)

$(K_A\ AisWh)$ $\neg(K_B\ \neg BisWh)$ $(K_C\ \neg CisWh)$ s2 A:w - B:w - C:b

s3 A:w - B:b - C:w

$(K_A\ AisWh)$ $(K_B\ \neg BisWh)$ $\neg(K_C\ \neg CisWh)$

γ Μπορείτε να εξηγήσετε, χρησιμοποιώντας το μοντέλο *N1* , ότι οι B, C καταλαβαίνουν το χρώμα τους στο **Step 1.2** του παιχνιδιού;

Case 2 : *το μοντέλο N2*

States:

s1 A:b - B:w - C:w (η ‘πραγματική’ κατάσταση)

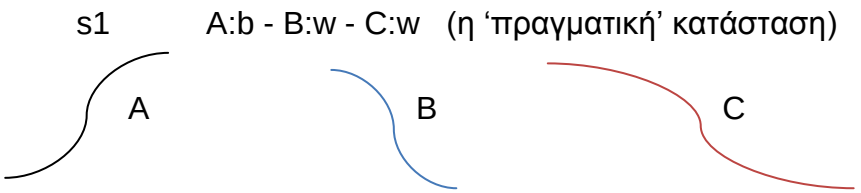
s2 A:w - B:w - C:w s3 A:b - B:b - C:w s4 A:b - B:w - C:b

	AisWh	BisWh	CisWh
s1	F	T	T
s2	T	T	T
s3	F	F	T
s4	F	T	F

Possibility relations:

Για τον A: s1 $\approx_A$ s2 Για τον B: s1 $\approx_B$ s3 Για τον C: s1 $\approx_C$ s4

N2



Ανάλυση ***από την σκοπιά των*** B , C ***και του*** A

1 (K_A ¬AisWh) δεν αληθεύει στις s1, s2 , αληθεύει στις s3 , s4 .

(K_B BisWh) δεν αληθεύει στις s1, s3 , αληθεύει στις s2 , s4 .

(K_C CisWh) δεν αληθεύει στις s1, s4 , αληθεύει στις s2 , s3 .

¬(K_B BisWh) ∧ ¬(K_C CisWh)) αληθεύει *μόνο* στην s1.

2 Μετά το **Step 2.1:** οι A , B , C γνωρίζουν ότι αληθεύει η φόρμουλα

¬(K_B BisWh) ∧ ¬(K_C CisWh).

Επομένως έχουν πληροφορηθεί (λόγω του **1**) ότι η κατάσταση s1 είναι η μόνη δυνατή.

Case 2 : *το μοντέλο M2*

States:

s1 A:b - B:w - C:w (η ‘πραγματική’ κατάσταση)

s2 A:w - B:w - C:w s3 A:b - B:b - C:w s4 A:b - B:w - C:b

s5 A:w - B:b - C:w

s6 A:w - B:w - C:b

	AisWh	BisWh	CisWh
s1	F	T	T
s2	T	T	T
s3	F	F	T
s4	F	T	F
s5	T	F	T
s6	T	T	F

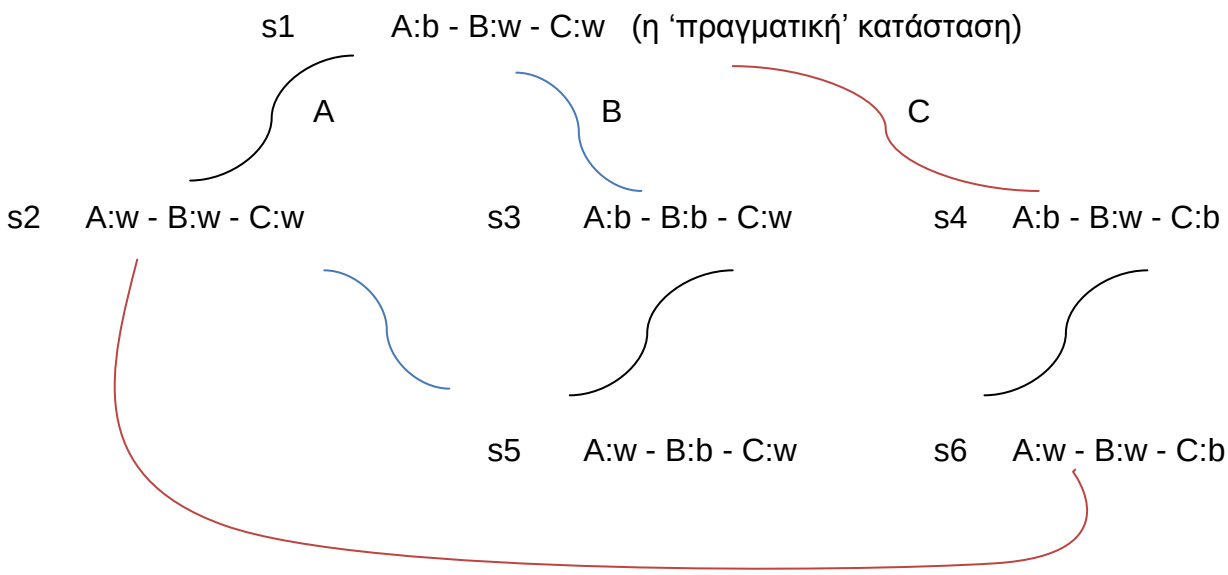
Possibility relations:

Για τον A: $s1 \approx_A s2$, $s3 \approx_A s5$, $s4 \approx_A s6$

Για τον B: $s1 \approx_B s3$, $s2 \approx_B s5$

Για τον C: $s1 \approx_C s4$, $s2 \approx_C s6$

M2



Ανάλυση ***από την σκοπιά των*** B , C ***και του*** A

1 (K_A $\neg$AisWh) δεν αληθεύει, σε οποιαδήποτε κατάσταση .

(K_B BisWh) αληθεύει στις s4, s6 , δεν αληθεύει σε οποιαδήποτε άλλη κατάσταση

(K_C CisWh) αληθεύει στις s3, s5 , δεν αληθεύει σε οποιαδήποτε άλλη κατάσταση .

$\neg$ (K_A $\neg$AisWh) $\wedge$ $\neg$(K_B BisWh) $\wedge$ $\neg$(K_C CisWh)) αληθεύει στις s1, s2 *μόνο*.

2 Μετά το **Step 2.1:** οι A , B , C γνωρίζουν ότι αληθεύει η φόρμουλα

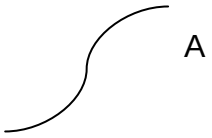
$\neg$ (K_A $\neg$AisWh) $\wedge$ $\neg$ (K_B BisWh) $\wedge$ $\neg$ (K_C CisWh) .

Επομένως έχουν πληροφορηθεί (λόγω του **1**) ότι οι δυνατές καταστάσεις είναι οι s1, s2 .

Το μοντέλο τώρα είναι:

R1

s1 A:b - B:w - C:w (η 'πραγματική' κατάσταση)



s2 A:w - B:w - C:w

Σχετικό εκπαιδευτικό υλικό

Fagin – Halpern – Moses – Vardi *Reasoning About Knowledg*e

Ενότητα 2.1. The Possible-Worlds Model

Προτεινόμενες ασκήσεις

1 Με βάση τους ορισμούς, εξηγείστε ότι η φόρμουλα

$\neg$(K_A $\neg$AisWh) $\wedge$ $\neg$(K_B BisWh) $\wedge$ $\neg$(K_C CisWh)

αληθεύει *μόνο* στην κατάσταση s1 του *N2*, και αληθεύει *μόνο* στις καταστάσεις s1, s2 του *M2* ..

2 Άν οι B, C αγνοήσουν την ανακοίνωση του A στο **Step 2.1,** θα μπορέσουν να καταλάβουν το χρώμα τους;

3 Είναι δυνατόν να εξηγηθεί, χρησιμοποιώντας το μοντέλο *N2* είτε χρησιμοποιώντας το μοντέλο *M2*, ο λόγος που ο A δεν έχει καταλάβει το χρώμα του στο **Step 2.2** του παιχνιδιού;

4 Είναι δυνατόν να εξηγηθεί, χρησιμοποιώντας το μοντέλο *M2*, ο λόγος που ο A έχει καταλάβει το χρώμα του στο **Step 2.3** του παιχνιδιού;

Case 2 και Case 3 : *Το γενικό μοντέλο P*

States: *Όλες* οι δυνατές αναθέσεις χρωμάτων στους τρείς παίκτες.

s1 A:w - B:w - C:w (η 'πραγματική' κατάσταση στην **Case 3**)

s2 A:b - B:w - C:w (η 'πραγματική' κατάσταση στην **Case 2**)

s3 A:w - B:b - C:w s4 A:w - B:w - C:b

s5 A:b - B:b - C:w s6 A:w - B:b - C:b s7 A:b - B:w - C:b

	AisWh	BisWh	CisWh
s1	T	T	T
s2	F	T	T
s3	T	F	T
s4	T	T	F
s5	F	F	T
s6	T	F	F
s7	F	T	F

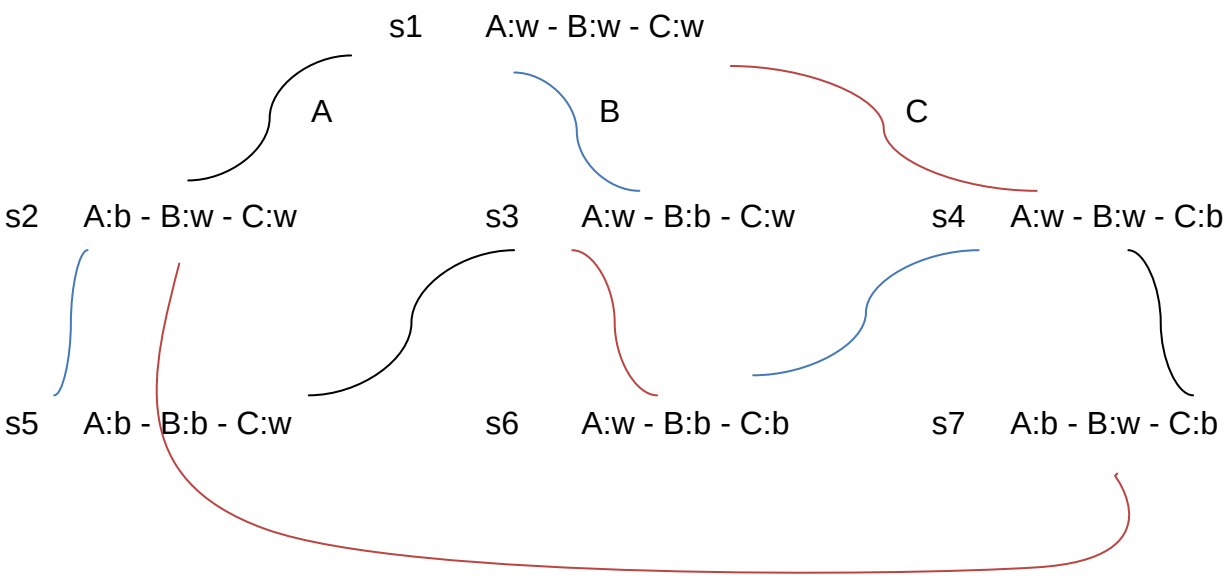
Possibility relations:

Για τον A: $s1 \approx_A s2$, $s3 \approx_A s5$, $s4 \approx_A s7$

Για τον B: $s1 \approx_B s3$, $s2 \approx_B s5$, $s4 \approx_B s6$

Για τον C: $s1 \approx_C s4$, $s2 \approx_C s7$, $s3 \approx_C s6$

P



Case 2 **A is black, B, C are white**

Ανάλυση *στο μοντέλο P* ***από την σκοπιά των*** B , C ***και του*** A

1 (K_A ¬AisWh) δεν αληθεύει σε καμμία κατάσταση, επομένως

¬(K_A ¬AisWh) αληθεύει σε *όλες* τις καταστάσεις

(K_B BisWh) αληθεύει στην s7 *μόνο*, (K_C CisWh) αληθεύει στην s5 *μόνο*.

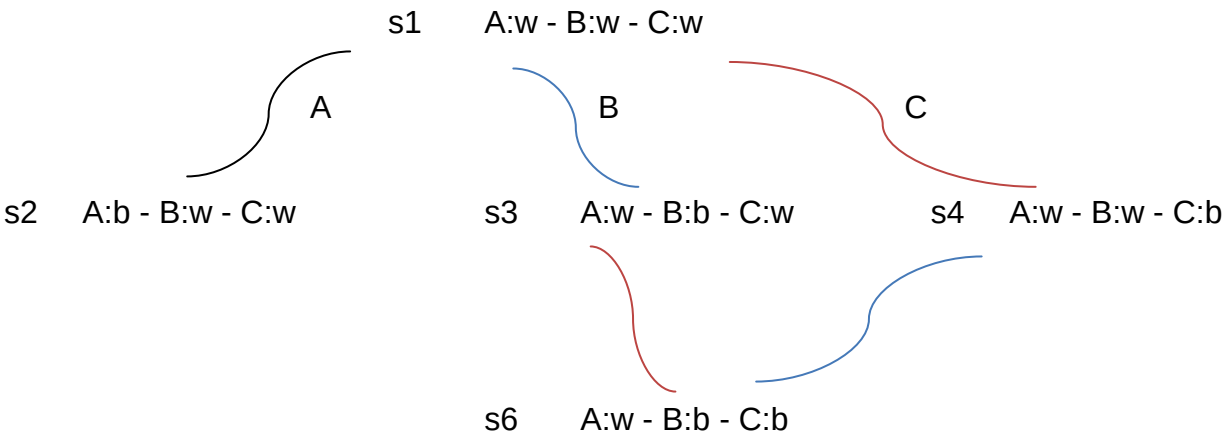
2 Μετά το **Step 2.1:** οι A , B , C γνωρίζουν ότι αληθεύει η φόρμουλα

¬(K_A ¬AisWh) ∧ ¬(K_B BisWh) ∧ ¬(K_C CisWh)) .

Επομένως έχουν πληροφορηθεί (λόγω του **1**) ότι οι καταστάσεις s5 , s7 είναι *αδύνατες*.

Το μοντέλο τώρα είναι:

R2



Ανάλυση *στο μοντέλο R2*

3 (K_B BisWh) αληθεύει στην s2 *μόνο*, (K_C CisWh) αληθεύει στην s2 *μόνο*.

Μετά το **Step 2.2:** ο A γνωρίζει ότι αληθεύει η φόρμουλα

(K_B BisWh) ∧ (K_C CisWh) .

Επομένως έχει πληροφορηθεί (λόγω του **3**) ότι η μόνη δυνατή κατάσταση είναι η s2 .

Το μοντέλο τώρα είναι:

States:

s2	A:b - B:w - C:w		AisWh	BisWh	CisWh
		s2	F	T	T

Ανάλυση (K_A ¬AisWh) αληθεύει στην s2

Case 3 **A, B, C are white**

Ανάλυση *στο μοντέλο P*

1 (K_A AisWh) αληθεύει στην s6 *μόνο*, (K_B BisWh) αληθεύει στην s7 *μόνο*,

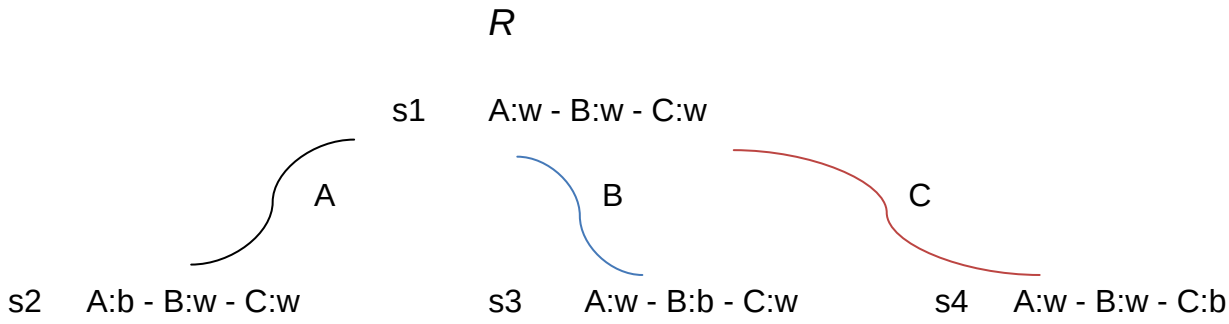
(K_C CisWh) αληθεύει στην s5 *μόνο*.

2 Μετά το **Step 3.1 :** οι A , B , C γνωρίζουν ότι αληθεύει η φόρμουλα

$\neg(K_A \text{ AisWh}) \wedge \neg(K_B \text{ BisWh}) \wedge \neg(K_C \text{ CisWh}))$.

Επομένως έχουν πληροφορηθεί (λόγω του **1**) ότι οι καταστάσεις s5 , s6 , s7 είναι *αδύνατες*.

Το μοντέλο τώρα είναι:



Ανάλυση *στο μοντέλο R*

3 (K_A AisWh) αληθεύει στις s3 , s4 *μόνο*,

(K_B BisWh) αληθεύει στις s2 , s4 *μόνο*,

(K_C CisWh) αληθεύει στις s3 , s2 *μόνο*.

4 Μετά το **Step 3.2 :** οι A , B , C γνωρίζουν ότι αληθεύει η φόρμουλα

$\neg(K_A \text{ AisWh}) \wedge \neg(K_B \text{ BisWh}) \wedge \neg(K_C \text{ CisWh}))$.

Επομένως έχουν πληροφορηθεί (λόγω του **3**) ότι οι καταστάσεις s2 , s3 , s4 είναι *αδύνατες*.

Το μοντέλο τώρα είναι:

States:

s1	A:w - B:w - C:w		AisWh	BisWh	CisWh
		s1	T	T	T

Ανάλυση (K_A AisWh) , (K_B BisWh) , (K_C CisWh) αληθεύουν στην s1 .

Σχετικό εκπαιδευτικό υλικό

Fagin – Halpern – Moses – Vardi *Reasoning About Knowledg*e

Ενότητα 2.3. The Muddy Children Revisited

Προτεινόμενες ασκήσεις

1 Χρησιμοποιείστε ένα μοντέλο Kripke με όλες τις δυνατές καταστάσεις, για να αναλύσετε το πρόβλημα στην Exercise 1.3 - *Reasoning About Knowledg*e.

2 Χρησιμοποιείστε ένα μοντέλο Kripke με όλες τις δυνατές καταστάσεις, για να αναλύσετε το Muddy Children Puzzle όταν $k = 2$ (δύο λερωμένα παιδάκια).

3 Να εξηγηθεί, χρησιμοποιώντας το μοντέλο P , ο λόγος που οι παίκτες B , C έχουν καταλάβει το χρώμα τους στο **Step 2.2** του παιχνιδιού.

Να εξηγηθεί χρησιμοποιώντας το μοντέλο *P*, ο λόγος που ο Α δεν έχει καταλάβει το χρώμα του στο **Step 2.2** του παιχνιδιού.

4 Μπορείτε να αναλύσετε την **Case 2** χρησιμοποιώντας ένα μοντέλο Kripke με λιγώτερες από επτά καταστάσεις;

Μπορείτε να αναλύσετε την **Case 3** χρησιμοποιώντας ένα μοντέλο Kripke με λιγώτερες από επτά καταστάσεις;