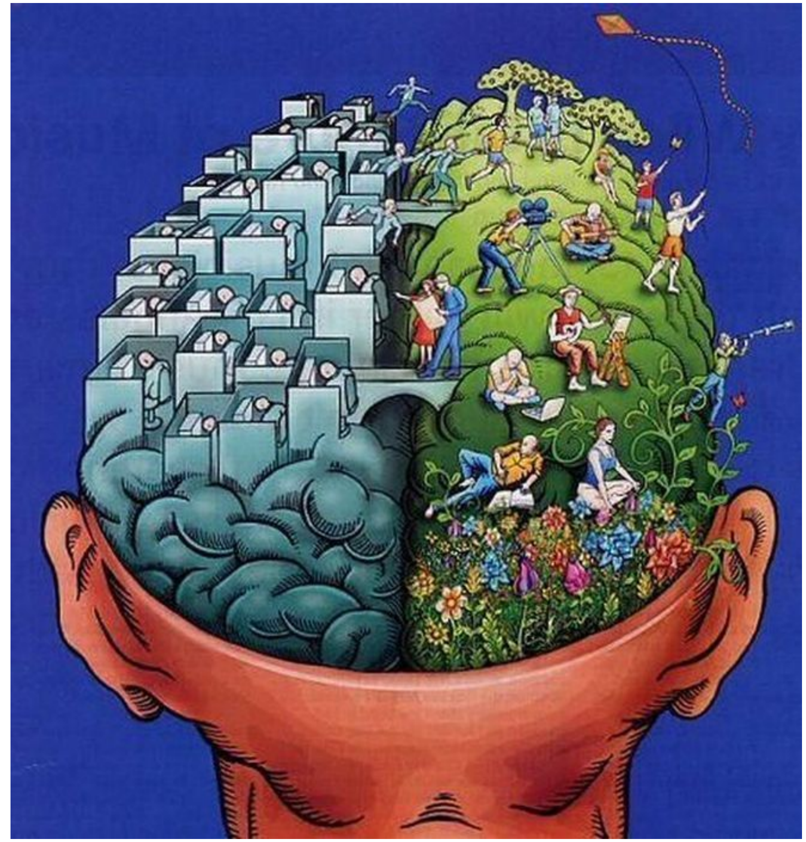
# Mesterséges Intelligencia MI

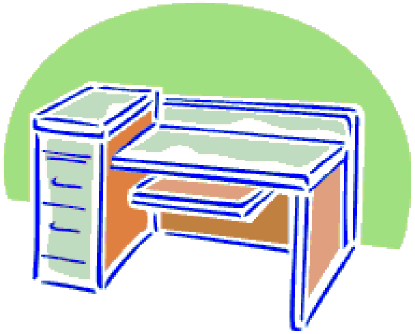
## Logikai ágens cselekvésben - szituációkalkulustól tervkészítésig

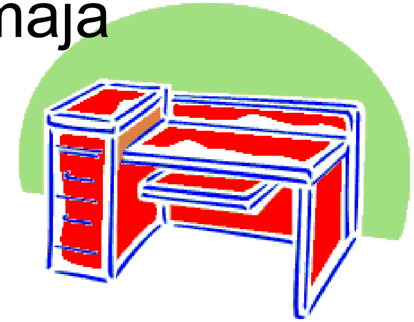



Dobrowiecki Tadeusz
Eredics Péter, és mások

BME I.E. 437, 463-28-99
dobrowiecki@mit.bme.hu,
http://www.mit.bme.hu/general/staff/tade

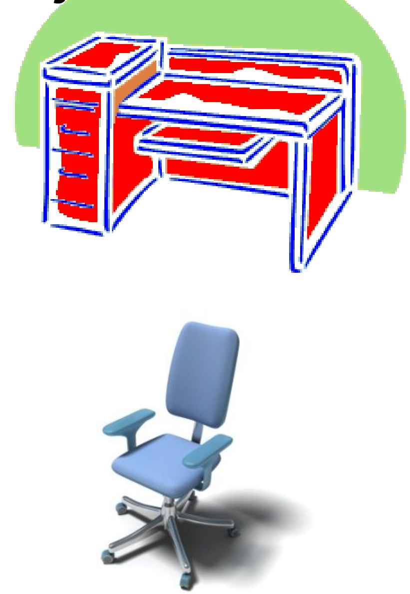
# A festő ágens problémája




Predikátum kalkulus hogyan vethető be a környezetüket alakító intelligens rendszerek leírására? Meg kell oldani annak ábrázolását, hogy az idő múlik, az ágens cselekszik, és ennek hatására változik az ágens környezete.
Az ágens világában egy asztal és egy szék van. Ágens bútort kizárólag pirosra tud festeni. Egyik bútor sem piros, de csak az asztalt szeretnénk pirosnak.
Hogyan írjuk le logikával, hogy milyen világgal szembesül az ágens, és milyen világot hagy maga után? Ráadásul legyen ez a leírás kellően általános is.

# A festő ágens problémája

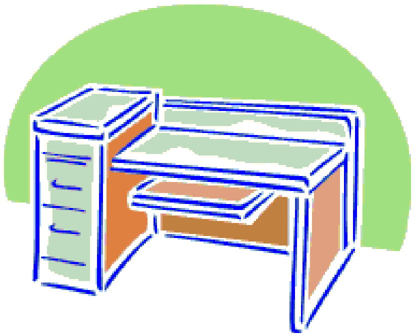


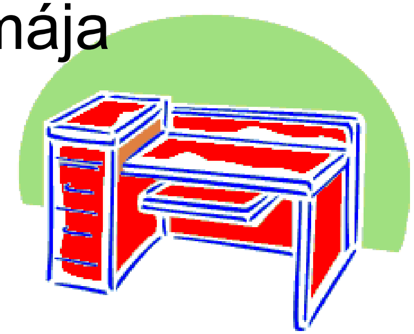
$\neg$ piros(Asztal) piros(Asztal)

volt? lesz?

Ellentmondás!

# A festő ágens problémája



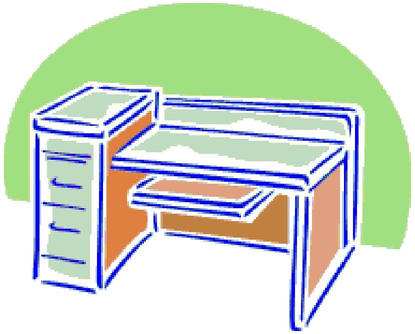


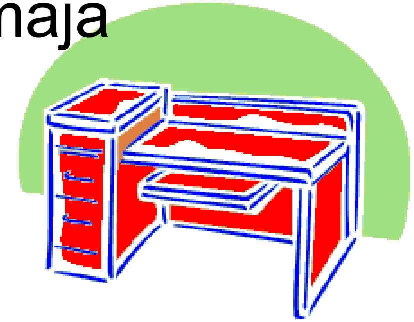
$\neg$ piros(Asztal) piros(Asztal)

$\neg$ piros(Asztal,S1) piros(Asztal,S2)

Már nincs
ellentmondás!

# A festő ágens problémája





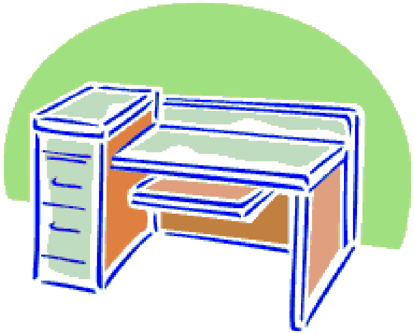
$\neg$ piros(Asztal) piros(Asztal)

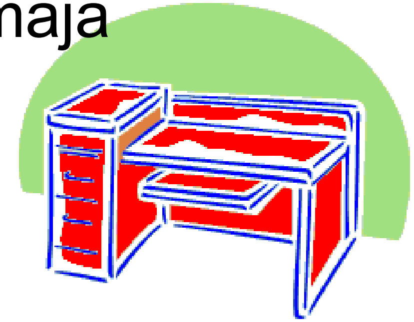
$\neg$ piros(Asztal,S1) piros(Asztal,S2)

S1 objektum $\rightarrow$ S2 objektum

Hogyan kapcsoljuk össze?

# A festő ágens problémája





¬ piros(Asztal)  piros(Asztal)

¬ piros(Asztal,S1)  piros(Asztal,S2)

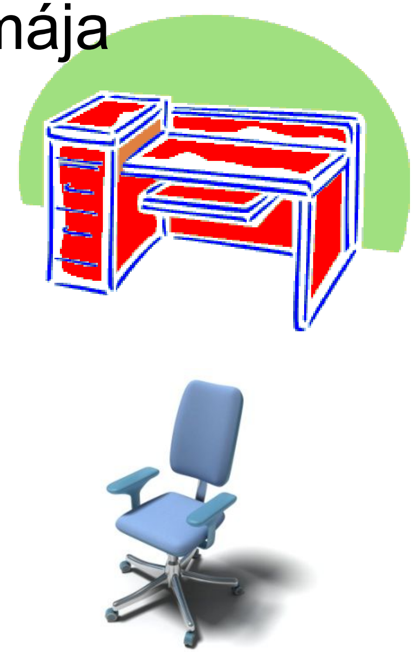
S1  S2

objektum  objektum

S2 = eredmény(Átfest, S1)

Hogyan kapcsoljuk össze?

# A festő ágens problémája

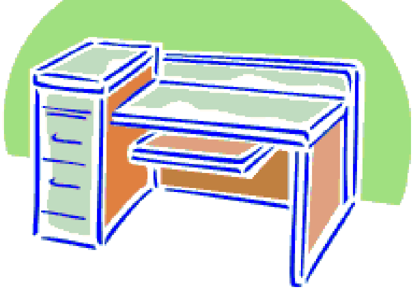


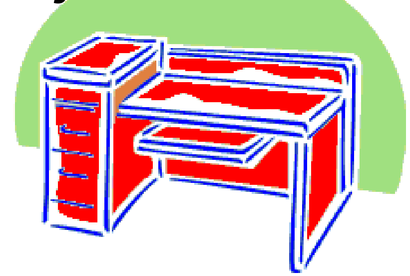
kék(Szék) kék(Szék)

volt lesz?

Nincs ellentmondás!

# A festő ágens problémája



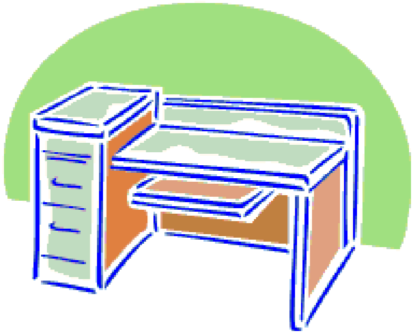


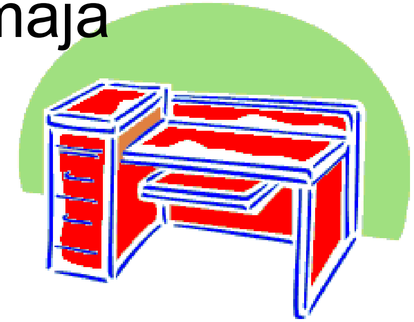
kék(Szék) kék(Szék)

kék(Szék,S1) kék(Szék,S2)

De van helyette probléma!

# A festő ágens problémája





kék(Szék) kék(Szék)

kék(Szék,S1) kék(Szék,S2)

S1 S2

objektum objektum

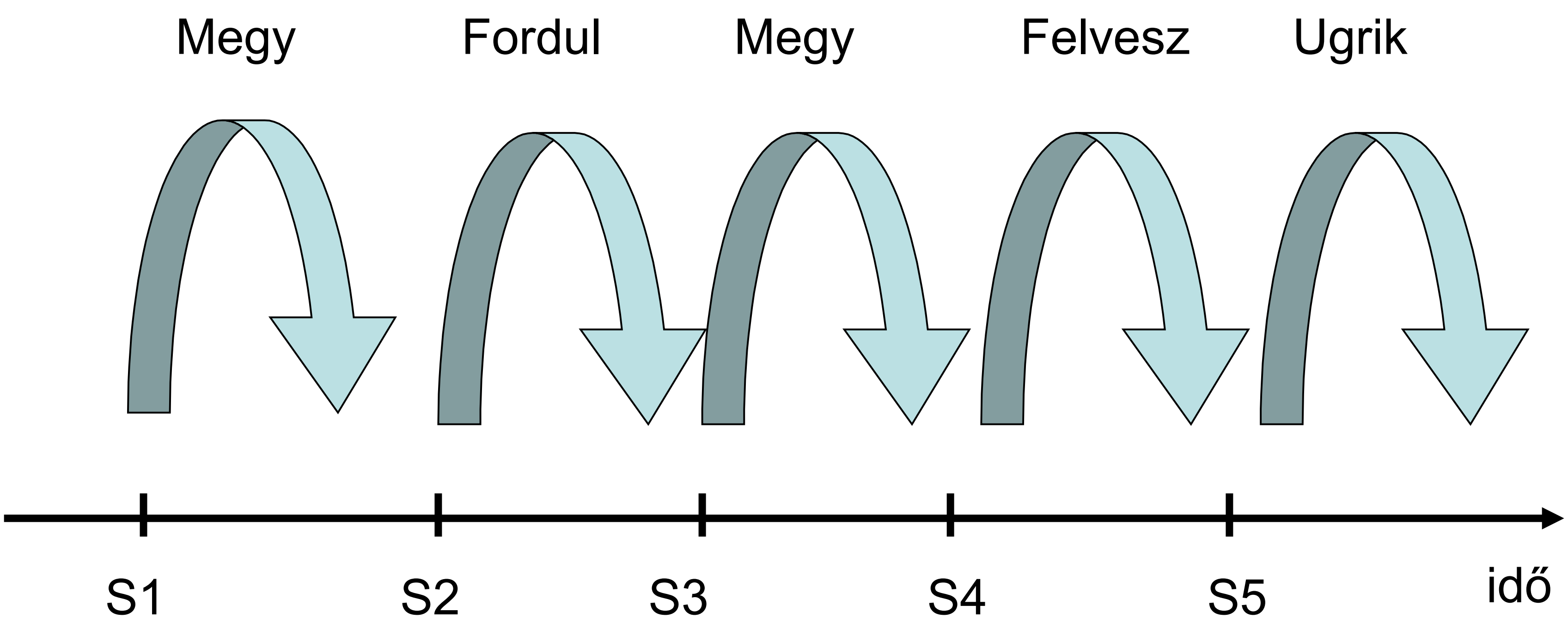
S2 = eredmény(Átfest, S1)

Hogyan kapcsoljuk össze?

# Szituáció kalkulus

- a változások leírásának egy módja az elsőrendű logikában.
  - a világ a **szituációk** sorozatából áll
  - mindegyike egy „pillanat felvétel” a világ állapotáról
  - egy-egy szituációban egy tény igaz, vagy hamis, **változhat**!

  **fluent** = „**folyékony esemény**” (pl. piros(Szék, σ))



# Hatás axiómák (Wumpus probléma esetére)

Pl. az arany nyomon követése: állítani kell bármely szituációban, hogy ha az arany ott van-e, és az ágens **Megfogta**, akkor az adódó szituációban (és majd később is) birtokolni fogja az aranyat:

$\forall$ s ott(Arany,s) $\wedge$ vihető(Arany) $\rightarrow$ **birtokol**(Arany,eredmény(Megfogás,s))

Hasonlóan az ágens nem birtokol semmit az ***Elenged*** cselekvése után:

$\forall$ x, s $\neg$ **birtokol** (x, eredmény (Elenged, s))

**Sajnos ez nem elég! Szükség van még …..**

# Keret axiómák

Hogyha az ágens birtokol valamit és nem engedi el, akkor a következő szituációban is fogja birtokolni. Ha az ágens nem birtokol valamit és nem tudja megfogni, akkor a következő állapotban sem fogja birtokolni:

$\forall$ a, x, s **birtokol** (x, s) $\wedge$ (a $\neq$ Elenged) $\rightarrow$ **birtokol** (x, eredmény(a, s))

$\forall$ a, x, s $\neg$**birtokol** (x, s) $\wedge$ (a $\neq$ Megfogás $\vee$ $\neg$ (ott(x, s) $\wedge$ vihető(x))
$\rightarrow$ $\neg$ **birtokol** (x, eredmény (a, s))

## Egy elegánsabb reprezentáció:

**hatás axióma** és **keret axióma** egyetlen axiómába:

**Igaz utána** $\Leftrightarrow$ **[bármely cselekvés, amely igazzá tette**
$\vee$ **már igaz volt és nem volt olyan cselekvés,**
**ami hamissá tette volna]**

A *Birtokol* esetében az utód-állapot axióma a következő:

$\forall$ a, x, s **birtokol** (x,eredmény (a, s)) $\Leftrightarrow$
[(a = Megfogás $\wedge$ ott (x,s) $\wedge$ vihető(x)) $\vee$ (**birtokol** (x,s) $\wedge$ a $\neq$ Elenged)]

# A világ rejtett tulajdonságainak a levezetése

## Ok-okozati szabályok:

a világ ok-okozatiságának feltételezett iránya:

a világ rejtett tulajdonságai okozzák bizonyos érzetek keletkezését.

$\forall$ l1,l2,s hely(Wumpus, l1,s) $\wedge$ szomszédos(l1,l2) $\rightarrow$ büdös(l2)
$\forall$l1,l2,s hely(Csapda, l1,s) $\wedge$ szomszédos(l1,l2) $\rightarrow$ szellős(l2)

(**modell-alapú következtető** rendszer) **("méret" független)**

## Diagnosztikai szabályok:

rejtett tulajdonságok meglétére következtetnek, közvetlenül az érzetből származó információk felhasználásával (abdukció kikerülése, antikauzális).

$\forall$l1,s büdös(l1) $\rightarrow$ ( $\exists$ l2 hely(Wumpus,l2,s) $\wedge$ (l2 = l1 $\vee$ szomszédos(l1,l2))

# Egy cél-orientált ágens

Ha az ágens aranyat megtalálta, a cselekvési stratégiát radikálisan meg kell változtatnia.
Az új cél most az, hogy kimeneküljön olyan gyorsan, amilyen gyorsan csak lehet. Most azt kellene kikövetkeztetni, hogy az ágens **célja** az [1,1] pozíció:

$\forall$s birtokol(Arany, s) $\rightarrow$ célpozíció([1,1], s)

Egy explicit cél jelenléte lehetővé teszi az ágens számára, hogy elkészítsen egy cselekvések sorozatából álló mondatot (tervet), amellyel elérheti a célt (és hajtja is végre – ezt a viselkedést kell rendszer specifikációban megadni és így implementálni).

Legalább három módja van egy ilyen mondat megtalálásának:

Kereséssel          már láttuk
**Következtetéssel    lássuk!**
**Lenne más is …   ?**

# Tervkészítés következtetéssel szituációkalkulusban

**Legyen a feladat nyelvezete:**

Ágens szobában van, *S* szituációban: *szoba(Ágens, S)*
asztal színe piros, *S* szituációban: *piros(Asztal, S)*

Mivel más objektum nincs is, le lehet rövidíteni:
ágens szobában van, *S* (kezdeti) szituációban: *szoba*(*S*)
asztal színe piros, *S* (kezdeti) szituációban: *piros(S)*

Ágens cselekvései legyenek: "**Bemegy**", "**Kimegy**", "**Átfest**".

**Jelen helyzet:** 1. $\neg$*szoba(S)*
2. $\neg$*piros(S)*

azaz az ágens szobán kívül van, szobában az asztal nincs pirosra átfestve.

**Probléma:** a kívánt helyzet az, hogy az asztal piros legyen és az ágens szobán kívül legyen. Az ágens képes ezt megvalósítani?
**Létezik egyáltalán egy ilyen helyzet?**

**Hatás axiómák** (avagy az ágens kényszercselekedetei)

$\forall\sigma.\ \neg\ szoba(\sigma) \land \neg\ piros(\sigma) \rightarrow szoba(eredmény(Bemegy,\ \sigma))$

$\forall\sigma.\ szoba(\sigma) \land \neg\ piros(\sigma) \rightarrow piros(eredmény(Atfest,\ \sigma))$

$\forall\sigma.\ szoba(\sigma) \land piros(\sigma) \rightarrow \neg\ szoba(eredmény(Kimegy,\ \sigma))$

$\forall\sigma.\ szoba(\sigma) \rightarrow \neg\ szoba(eredmény(Kimegy,\ \sigma))$

$\forall\sigma.\ \neg szoba(\sigma) \rightarrow szoba(eredmény(Bemegy,\ \sigma))$

$\forall\sigma.\ \neg piros(\sigma) \rightarrow piros(eredmény(Atfest,\ \sigma))$

**Mitől lesz ilyen? Ha ilyennek megtervezzük és implementáljuk**

**Keret axiómák** (buta formában)

$\forall\sigma.\ szoba(\sigma) \rightarrow szoba(eredmény(Atfest,\ \sigma))$

$\forall\sigma.\ piros(\sigma) \rightarrow piros(eredmény(Bemegy,\ \sigma))$

$\forall\sigma.\ piros(\sigma) \rightarrow piros(eredmény(Kimegy,\ \sigma))$

$\forall\sigma.\ \neg piros(\sigma) \rightarrow \neg\ piros(eredmény(Kimegy,\ \sigma))$

$\forall\sigma.\ \neg piros(\sigma) \rightarrow \neg\ piros(eredmény(Bemegy,\ \sigma))$

…

**Probléma:** a kívánt helyzet az, …. **létezik egyáltalán?**

$\exists\sigma.\ \neg szoba(\sigma) \land piros(\sigma)$ ?

**Legyen egy további rövidítés:** $B(s) = eredmény(Bemegy, s)$
$K(s) = eredmény(Kimegy, s)$
$A(s) = eredmény(Átfest, s)$

Ne felejtsük, hogy a beszédes logikai nevek nem a gépnek szólnak, hanem az embereknek, és a gépi eljárásra nincsenek hatással.

**Átírással klóz formára:**

1. $\neg szoba(S)$
2. $\neg piros(S)$
3. $szoba(\sigma 1) \vee piros(\sigma 1) \vee szoba(B(\sigma 1))$
4. $\neg szoba(\sigma 2) \vee piros(\sigma 2) \vee piros(A(\sigma 2))$
5. $\neg szoba(\sigma 3) \vee \neg piros(\sigma 3) \vee \neg szoba(K(\sigma 3))$
6. $\neg szoba(\sigma 4) \vee szoba(A(\sigma 4))$
7. $\neg szoba(\sigma 5) \vee \neg szoba(K(\sigma 5))$
8. $szoba(\sigma 6) \vee szoba(B(\sigma 6))$
9. $\neg piros(\sigma 7) \vee piros(B(\sigma 7))$
10. $\neg piros(\sigma 8) \vee piros(K(\sigma 8))$
11. $piros(\sigma 9) \vee \neg piros(K(\sigma 9))$
12. $piros(\sigma 10) \vee \neg piros(B(\sigma 10))$
13. $piros(\sigma 11) \vee piros(A(\sigma 11))$
14. $szoba(\sigma 12) \vee \neg piros(\sigma 12)$

és még egy egyszerűsítés (diszjunkció explicite nem szükséges):

1. $\neg sz(S)$
2. $\neg p(S)$
3. $sz(\sigma 1)$ $p(\sigma 1)$ $sz(B(\sigma 1))$
4. $\neg sz(\sigma 2)$ $p(\sigma 2)$ $p(A(\sigma 2))$
5. $\neg sz(\sigma 3)$ $\neg p(\sigma 3)$ $\neg sz(K(\sigma 3))$
6. $\neg sz(\sigma 4)$ $sz(A(\sigma 4))$
7. $\neg sz(\sigma 5)$ $\neg sz(K(\sigma 5))$
8. $sz(\sigma 6)$ $sz(B(\sigma 6))$
9. $\neg p(\sigma 7)$ $p(B(\sigma 7))$
10. $\neg p(\sigma 8)$ $p(K(\sigma 8))$
11. $p(\sigma 9)$ $\neg p(K(\sigma 9))$
12. $p(\sigma 10)$ $\neg p(B(\sigma 10))$
13. $p(\sigma 11)$ $p(A(\sigma 11))$
14. $sz(\sigma 12)$ $\neg p(\sigma 12)$

1. $\neg sz(S)$
2. $\neg p(S)$
3. $sz(\sigma 1) \quad p(\sigma 1) \quad sz(B(\sigma 1))$
4. $\neg sz(\sigma 2) \quad p(\sigma 2) \quad p(A(\sigma 2))$
5. $\neg sz(\sigma 3) \; \neg p(\sigma 3) \quad \neg sz(K(\sigma 3))$
6. $\neg sz(\sigma 4) \quad sz(A(\sigma 4))$
7. $\neg sz(\sigma 5) \; \neg sz(K(\sigma 5))$
8. $sz(\sigma 6) \quad sz(B(\sigma 6))$
9. $\neg p(\sigma 7) \quad p(B(\sigma 7))$
10. $\neg p(\sigma 8) \quad p(K(\sigma 8))$
11. $p(\sigma 9) \quad \neg p(K(\sigma 9))$
12. $p(\sigma 10) \; \neg p(B(\sigma 10))$
13. $p(\sigma 11) \quad p(A(\sigma 11))$
14. $sz(\sigma 12) \; \neg p(\sigma 12)$

2. és 12. -ből

15. $\neg\, p(B(S)) \quad \sigma 10/S$

1. $\neg sz(S)$
2. $\neg p(S)$
3. $sz(\sigma 1) \quad p(\sigma 1) \quad sz(B(\sigma 1))$
4. $\neg sz(\sigma 2) \quad p(\sigma 2) \quad p(A(\sigma 2))$
5. $\neg sz(\sigma 3) \ \neg p(\sigma 3) \ \neg sz(K(\sigma 3))$
6. $\neg sz(\sigma 4) \quad sz(A(\sigma 4))$
7. $\neg sz(\sigma 5) \ \neg sz(K(\sigma 5))$
8. $sz(\sigma 6) \quad sz(B(\sigma 6))$
9. $\neg p(\sigma 7) \quad p(B(\sigma 7))$
10. $\neg p(\sigma 8) \quad p(K(\sigma 8))$
11. $p(\sigma 9) \quad \neg p(K(\sigma 9))$
12. $p(\sigma 10) \ \neg p(B(\sigma 10))$
13. $p(\sigma 11) \quad p(A(\sigma 11))$
14. $sz(\sigma 12) \ \neg p(\sigma 12)$
15. $\neg p(B(S))$

1., 2, és 3. -ből

16. $sz(B(S)) \qquad \sigma 1/S$

3. $sz(\sigma1)$ $p(\sigma1)$ $sz(B(\sigma1))$
4. $\neg sz(\sigma2)$ $p(\sigma2)$ $p(A(\sigma2))$
5. $\neg sz(\sigma3)$ $\neg p(\sigma3)$ $\neg sz(K(\sigma3))$
6. $\neg sz(\sigma4)$ $sz(A(\sigma4))$
7. $\neg sz(\sigma5)$ $\neg sz(K(\sigma5))$
8. $sz(\sigma6)$ $sz(B(\sigma6))$
9. $\neg p(\sigma7)$ $p(B(\sigma7))$
10. $\neg p(\sigma8)$ $p(K(\sigma8))$
11. $p(\sigma9)$ $\neg p(K(\sigma9))$
12. $p(\sigma10)$ $\neg p(B(\sigma10))$
13. $p(\sigma11)$ $p(A(\sigma11))$
14. $sz(\sigma12)$ $\neg p(\sigma12)$
15. $\neg p(B(S))$
16. $sz(B(S))$

4., 15., és 16. -ből

17. $p(A(B(S))))$ $\sigma2/ B(S))$

3. $sz(\sigma 1)$ $p(\sigma 1)$ $sz(B(\sigma 1))$

4. $\neg sz(\sigma 2)$ $p(\sigma 2)$ $p(A(\sigma 2))$

5. $\neg sz(\sigma 3)$ $\neg p(\sigma 3)$ $\neg sz(K(\sigma 3))$

6. $\neg sz(\sigma 4)$ $sz(A(\sigma 4))$

7. $\neg sz(\sigma 5)$ $\neg sz(K(\sigma 5))$

8. $sz(\sigma 6)$ $sz(B(\sigma 6))$

9. $\neg p(\sigma 7)$ $p(B(\sigma 7))$

10. $\neg p(\sigma 8)$ $p(K(\sigma 8))$

11. $p(\sigma 9)$ $\neg p(K(\sigma 9))$

12. $p(\sigma 10)$ $\neg p(B(\sigma 10))$

13. $p(\sigma 11)$ $p(A(\sigma 11))$

14. $sz(\sigma 12)$ $\neg p(\sigma 12)$

15. $\neg p(B(S))$

16. $sz(B(S))$

17. $p(A(B(S)))$

6. és 16. -ből

18. $sz(A(B(S)))$ $\sigma 4/ B(S)$

3. $sz(\sigma 1)$ $p(\sigma 1)$ $sz(B(\sigma 1))$
4. $\neg sz(\sigma 2)$ $p(\sigma 2)$ $p(A(\sigma 2))$
5. $\neg sz(\sigma 3)$ $\neg p(\sigma 3)$ $\neg sz(K(\sigma 3))$
6. $\neg sz(\sigma 4)$ $sz(A(\sigma 4))$
7. $\neg sz(\sigma 5)$ $\neg sz(K(\sigma 5))$
8. $sz(\sigma 6)$ $sz(B(\sigma 6))$
9. $\neg p(\sigma 7)$ $p(B(\sigma 7))$
10. $\neg p(\sigma 8)$ $p(K(\sigma 8))$
11. $p(\sigma 9)$ $\neg p(K(\sigma 9))$
12. $p(\sigma 10)$ $\neg p(B(\sigma 10))$
13. $p(\sigma 11)$ $p(A(\sigma 11))$
14. $sz(\sigma 12)$ $\neg p(\sigma 12)$
15. $\neg p(B(S))$
16. $sz(B(S))$
17. $p(A(B(S)))$
18. $sz(A(B(S)))$

5., 17., és 18. -ből

19. $\neg sz(K(A(B(S))))$ $\sigma 3/ A(B(S))$

5. $\neg sz(\sigma 3)$ $\neg p(\sigma 3)$ $\neg sz(K(\sigma 3))$

6. $\neg sz(\sigma 4)$ $sz(A(\sigma 4))$

7. $\neg sz(\sigma 5)$ $\neg sz(K(\sigma 5))$

8. $sz(\sigma 6)$ $sz(B(\sigma 6))$

9. $\neg p(\sigma 7)$ $p(B(\sigma 7))$

10. $\neg p(\sigma 8)$ $p(K(\sigma 8))$

11. $p(\sigma 9)$ $\neg p(K(\sigma 9))$

12. $p(\sigma 10)$ $\neg p(B(\sigma 10))$

13. $p(\sigma 11)$ $p(A(\sigma 11))$

14. $sz(\sigma 12)$ $\neg p(\sigma 12)$

15. $\neg p(B(S))$

16. $sz(B(S))$

17. $p(A(B(S))))$

18. $sz(A(B(S)))$

19. $\neg sz(K(A(B(S)))$

9. és 16. -ből

20. $p(K(A(B(S)))))$ $\sigma 8/A(B(S)))$

5. $\neg sz(\sigma 3)\ \neg p(\sigma 3)\ \neg sz(K(\sigma 3))$

6. $\neg sz(\sigma 4)\quad sz(A(\sigma 4))$

7. $\neg sz(\sigma 5)\ \neg sz(K(\sigma 5))$

8. $sz(\sigma 6)\quad sz(B(\sigma 6))$

9. $\neg p(\sigma 7)\quad p(B(\sigma 7))$

10. $\neg p(\sigma 8)\quad p(K(\sigma 8))$

11. $p(\sigma 9)\quad \neg p(K(\sigma 9))$

12. $p(\sigma 10)\ \neg p(B(\sigma 10))$

13. $p(\sigma 11)\quad p(A(\sigma 11))$

14. $sz(\sigma 12)\ \neg p(\sigma 12)$

15. $\neg p(B(S))$

16. $sz(B(S))$

17. $p(A(B(S))))$

18. $sz(A(B(S)))$

19. $\neg sz(K(A(B(S))))$

20. $p(K(A(B(S)))))$

14. és 20. -ből

21. $sz(K(A(B(S))))\qquad \sigma 12/K(A(B(S)))$

5. $\neg sz(\sigma 3)\ \neg p(\sigma 3)\ \neg sz(K(\sigma 3))$
6. $\neg sz(\sigma 4)\quad sz(A(\sigma 4))$
7. $\neg sz(\sigma 5)\ \neg sz(K(\sigma 5))$
8. $sz(\sigma 6)\quad sz(B(\sigma 6))$
9. $\neg p(\sigma 7)\quad p(B(\sigma 7))$
10. $\neg p(\sigma 8)\quad p(K(\sigma 8))$
11. $p(\sigma 9)\quad \neg p(K(\sigma 9))$
12. $p(\sigma 10)\ \neg p(B(\sigma 10))$
13. $p(\sigma 11)\quad p(A(\sigma 11))$
14. $sz(\sigma 12)\ \neg p(\sigma 12)$
15. $\neg p(B(S))$
16. $sz(B(S))$
17. $p(A(B(S)))$)
18. $sz(A(B(S)))$
19. $\neg sz(K(A(B(S))))$
20. $p(K(A(B(S)))))$
21. $sz(K(A(B(S))))$

19. és 21. -ből: Ø

És az eredmény: Az ágens igenis képes megvalósítani a feladatot, ráadásul $\sigma 12$ állapotban. Mi is ez az állapot?

$\sigma$12 = K(A(B(S))) =
eredmeny(**Kimegy**,
eredmeny(**Átfest**,
eredmeny(**Bemegy**, **S**)))

Megvalósítja (akkor meglesz a kívánt állapot), ha:

**Bemegy** $\rightarrow$ **Átfest** $\rightarrow$ **Kimegy**
S --------------- S1 ------------- S3 ----------------- S4 (szituációk)

cselekvési sorozatot visz véghez.

A szükséges cselekvési sorozatot tehát előre logikailag **kitervelte**!

Prover9

```
-szoba(s0).
-piros(s0).
all x (-szoba(x) -> szoba(er(be,x))).
all x (szoba(x) & -piros(x) -> szoba(er(at,s)) & piros(er(at,x))).
all x (szoba(x) -> -szoba(er(ki,x))).
all x (szoba(x) -> szoba(er(at,x))).
all x (-piros(x) -> -piros(er(be,x))).
all x (piros(x) -> piros(er(be,x))).
all x (-piros(x) -> -piros(er(ki,x))).
all x (piros(x) -> piros(er(ki,x))).
szoba(y) | -piros(y)     # answer(y).
```

```
=================== PROOF ========================
% Proof 1 at 0.01 (+ 0.03) seconds: er(ki,er(at,er(be,s0))).
...
1 (all x (-szoba(x) -> szoba(er(be,x)))) # label(non_clause).  [assumption].
2 (all x (szoba(x) & -piros(x)
     -> szoba(er(at,s)) & piros(er(at,x)))) # label(non_clause).  [assumption].

11 szoba(x) | szoba(er(be,x)).  [clausify(1)].
...
76 $F # answer(er(ki,er(at,er(be,s0)))).  [resolve(70,a,20,b),unit_del(a,75)].
========================== end of proof ==================
```