Modelling knowledge

- Possible worlds:

Example

Claremont	London	
Sun	Sun	world 1
Sun	Rain	world 2
Rain	Sun	~~world 3~~
Rain	Rain	~~world 4~~

Agent i knows φ, denoted $K_i \varphi$, if φ is true in all worlds considered possible given current information

example: look outside its sunny

K_0 (sunny in claremont)

Let C = claremont
$S(x)$ = sunny at x
L = London

consider

$\neg K_0$ (sunny in London)

$K_0 \neg$(sunny in London)

$\neg K_0 \neg$(sunny in ~~claremont~~ London)

which translates to "it is possible that it is sunny in London

Example Muddy Children

Alice sees that Bob & Charlie have muddy forheads & all others do not

2 possible worlds

$$K_A \left[(M(A) \land M(B) \land M(C)) \lor (\neg M(A) \land M(B) \land M(C)) \right]$$

$$\land \forall i \, (i \notin \{A,B,C\} \Rightarrow \neg M(i))$$

$\in$ children

Syntax of Knowledge Logic

N ~~#~~ Agents : $1 \ldots N$

agent : actor, person, robot, program, who can reason about the world

Φ = primitive propositions (t or f)

$p, q, r, p', q', r' \ldots$

ex: $p \equiv$ rainy in London

q = Sully knows Alice holds $(A, 8)$

Modal operators $K_1 \ldots K_n$

$K_i \varphi \equiv$ "agent i knows φ"

Logical combinations: if ~~φ~~ φ, ψ formulas then so are $\neg \varphi$, $\varphi \land \varphi$, $\varphi \lor \varphi$, $\varphi \Rightarrow \psi$, $K_i \varphi$

Example:

$$K_1 K_2 p \wedge \neg K_2 K_1 K_2 p$$

Possibility: $\neg K_i \neg \varphi \equiv K_i$ believes φ to be possible

A_i does not know that not φ

φ might be T or F

Try this:

Dean doesn't know ~~that~~ whether Nixon knows that Dean knows that Nixon knows that

$p \equiv$ McCord Burgled O'Briens office at Watergate

$$\neg K_1 \neg (K_2 K_1 K_2 p) \wedge \neg K_1 \neg (\neg K_2 K_1 K_2 p)$$

Dean doesn't know $\varphi \equiv$

Dean considers φ possible &

Dean considers $\neg\varphi$ possible

Semantics

Given φ, how to ~~bxx~~ attribute T or F to φ?

Kripke Structures. M is a tuple

$(S, \pi, K_1, \ldots K_n)$

S: set of states (possible worlds)

π: interpretation mapping states to truth assignments of propositions

i.e. $\pi(s): \Phi \to \{T, F\}$

K_i is a binary relation on S

$\pi(s)(p)$ tells us if p is T or F in state s

ex: $p \equiv$ "raining in SF"

$\pi(s)(p)$ = True when it is raining in world s in structure M

K_i captures the possibility relation according to agent i

$(s,t) \in \mathcal{K}_i$ if A_i considers world t possible, given the info available in world s

Assumptions about K_i:

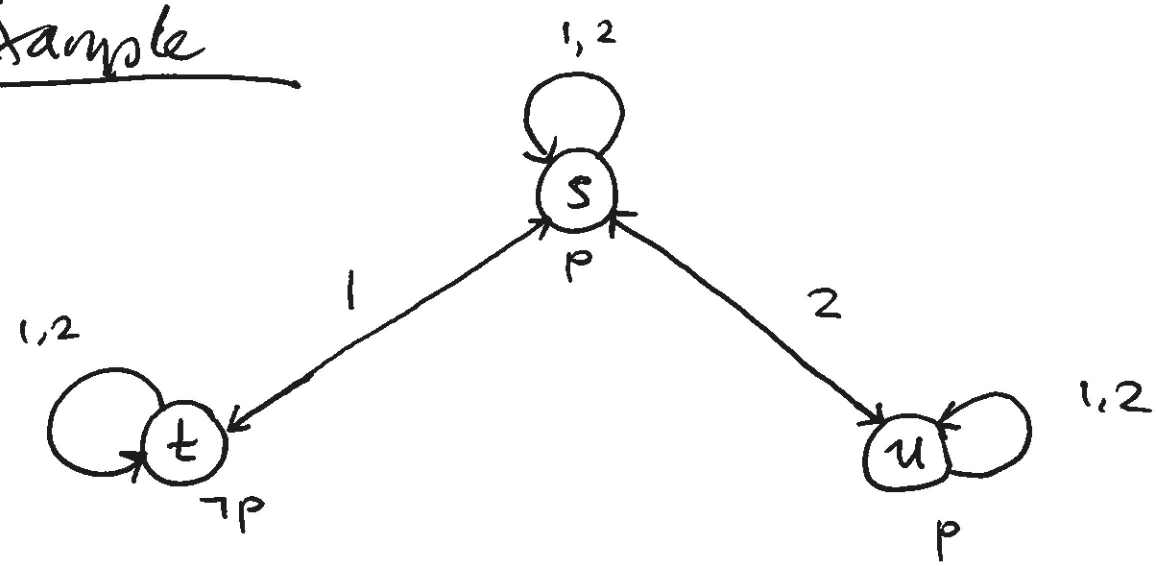
(a) reflexive: $\mathcal{K}_i(s,s)$

(b) symmetric: $\mathcal{K}_i(s,t) \Leftrightarrow \mathcal{K}_i(t,s)$

(c) transitive: $\mathcal{K}_i(s,t) \wedge \mathcal{K}_i(t,u) \Rightarrow \mathcal{K}_i(s,u)$

~~This means that~~ $\mathcal{K}_i(s,t)$ indicates that for A_i s & t are indistinguishable worlds

Example



this means agent 1 cannot distinguish

s from t, so

$\mathcal{K}_1 = \{ \cancel{(s,s),(s,u),(t,t)}$
$(s,s), (s,t), (t,s), (t,t), (u,u)$

Agent 2 cannot distinguish s from u:

$\mathcal{K}_2 = \{(s,s), (s,u), (t,t), (u,s), (u,u)\}$

$p \equiv$ "sunny in SF"

in state s it is sunny in SF.
but agent 1 doesnt know it,
since s & t indistinguishable
knows they are different but
does not have enough info

$\neg K_1 p$

Agent 2 knows it is sunny in SF
since both worlds considered possible
by A_2 true in s & u

$K_2 p$

In state t

Agent 2 knows $\neg p$ $K_2 \neg p$

In state s

A_1 knows that A_2 knows the value of p

$$(M,s) \models p \wedge \neg K_1 p \wedge K_2 p \wedge K_1 (K_2 p \vee K_2 \neg p)$$

~~Also in state s.~~

at state u $K_1 p$

at state s $\neg K_1 p$

at state s A_2 considers a possible

~~therefore: $\neg K_2 \neg K_1 p$~~

Hence $(M,s) \models \neg K_2 \neg K_1 p$

Semantics

$(M, s) \models p$ iff $\pi(s)(p) = \text{true}$

$(M, s) \models \neg p$ if $(M, s) \not\models p$

$(M, s) \models \alpha \text{ op } \beta$ if $(M, s) \models \alpha$ op $(M, s) \models \beta$

$M(s) \models K_i \psi$ iff $(M, t) \models \psi$ $\forall t : (s, t) \in \mathcal{K}_i$

9

Example

Deck of cards $\{A, B, C\}$

A_1 & A_2 each get 1 card,

1 face down

possible world: $(\cancel{x}_1, \cancel{x}_2)$

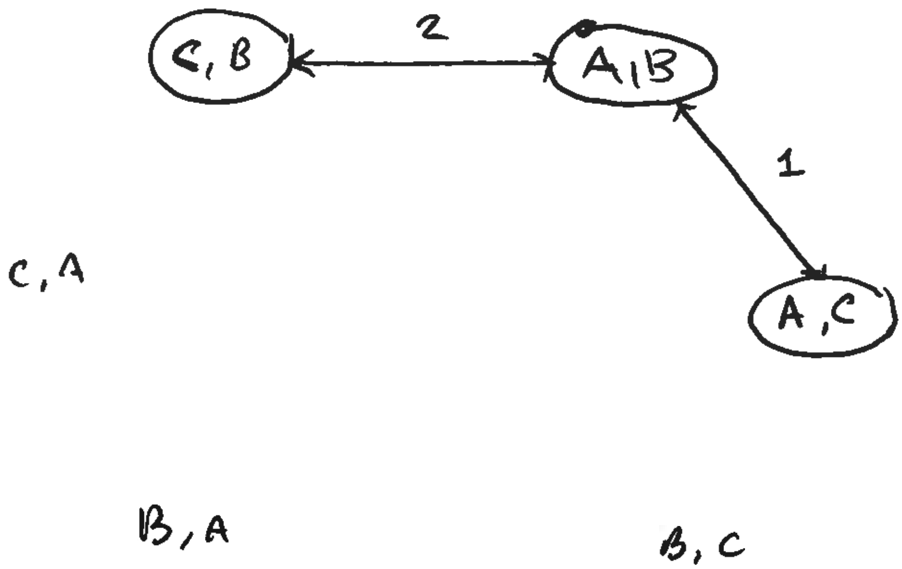
$(\cancel{c_1, c_2})$

card held by A_1

card held by A_2

how many worlds?

what are they?



10

Let $H_1(x) \equiv$ Agent 1 holds card x

$M, (A, B) \models H_1(A) \wedge H_2(B)$

$M, (A, B) \models K_1 (H_2(B) \vee H_2(C))$

$M, (A, B) \models K_1 \neg K_2 H_1(A)$

$\vdots$

properties

$\mathcal{M}, s \models K_i \varphi \Rightarrow \varphi$

$\models K_i \varphi \Rightarrow K_i K_i \varphi$

$\models \neg K_i \varphi \Rightarrow K_i \neg K_i \varphi$

$\models (K_i \varphi \wedge K \varphi \Rightarrow \psi) \Rightarrow K \psi$