Introduction to Logics of Knowledge and Belief

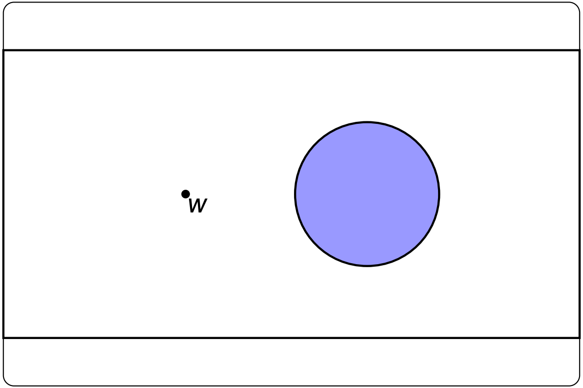
Eric Pacuit

University of Maryland

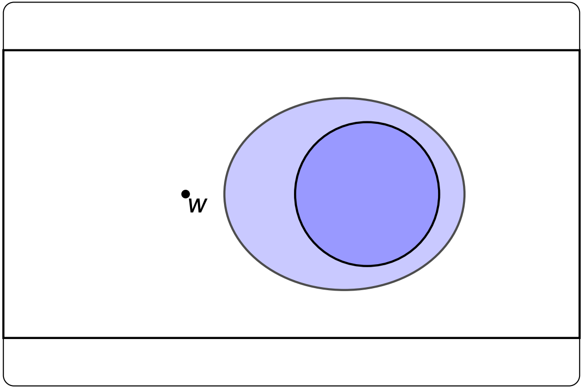
pacuit.org

epacuit@umd.edu

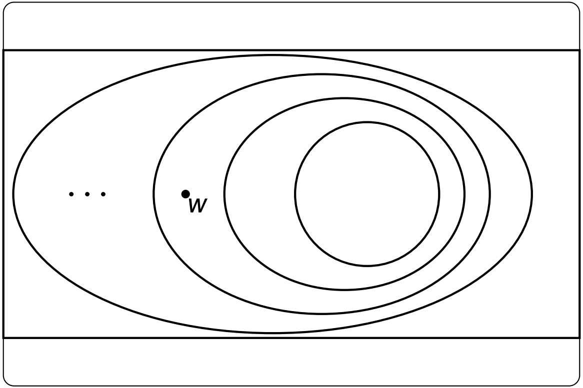
April 22, 2019



- The agent's (hard) information (i.e., the states consistent with what the agent knows)
- The agent's beliefs (soft information—-the states consistent with what the agent believes)



- The agent's beliefs (soft information—-the states consistent with what the agent believes)
- The agent's "contingency plan": when the stronger beliefs fail, go with the weaker ones.



- The agent's beliefs (soft information—-the states consistent with what the agent believes)
- The agent's "contingency plan": when the stronger beliefs fail, go with the weaker ones.

Plausibility Models

Epistemic Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in$ At)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Plausibility Models

Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Plausibility Models

Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

Plausibility Models

Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

Assumptions:

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$.

Plausibility Models

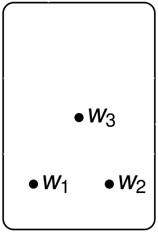
Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V \rangle$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ if for each $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$
  $[w]_i = \{v \mid w \sim_i v\}$ is the agent's **information cell.**

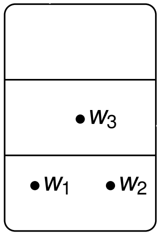
Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$



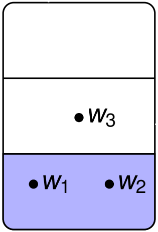
Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

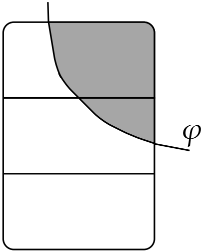


Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

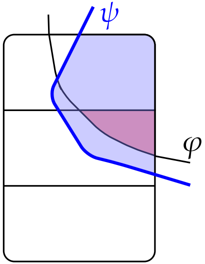


Beliefs via Plausibility



Conditional Belief: $B^{\varphi}\psi$

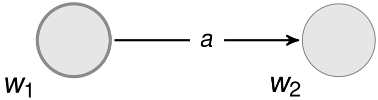
Beliefs via Plausibility



Conditional Belief: $B^{\varphi}\psi$

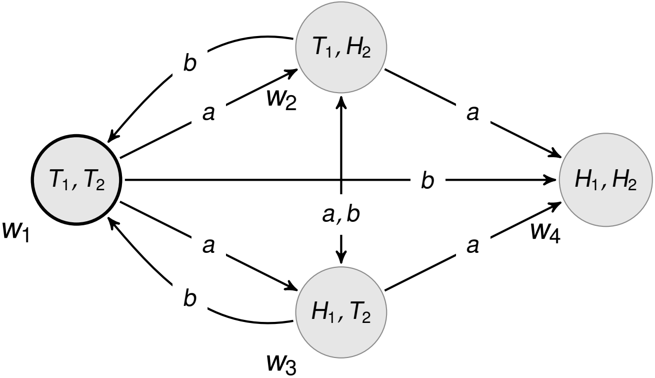
$$Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}}) \subseteq \llbracket\psi\rrbracket_{\mathcal{M}}$$

Example

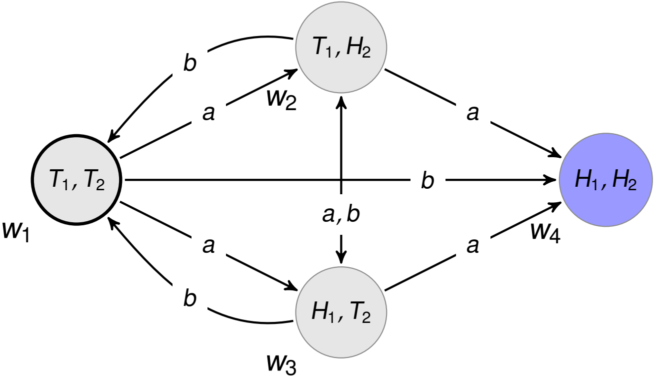


$$w_2 \preceq_a w_1$$

Example

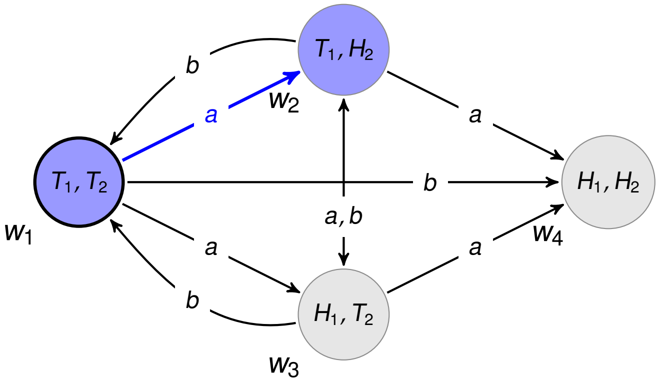


Example



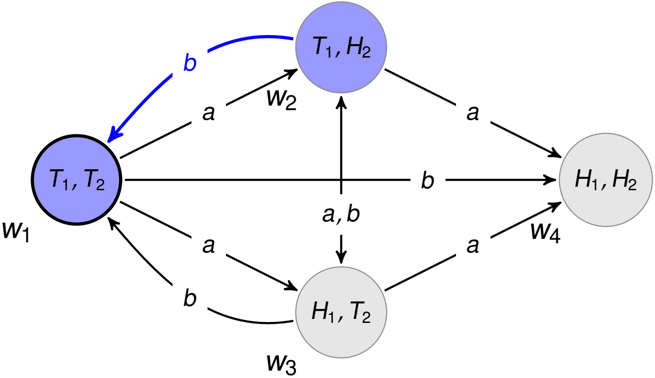
- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$

Example



- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$
- $w_1 \models B_a^{T_1} H_2$

Example



- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$
- $w_1 \models B_a^{T_1} H_2$
- $w_1 \models B_b^{T_1} T_2$

Group Knowledge

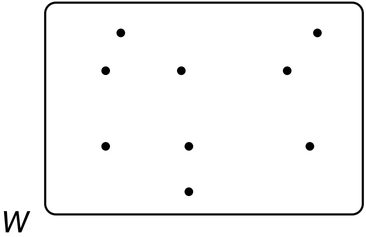
P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
http://plato.stanford.edu/entries/common-knowledge/.

The “Standard” Account

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

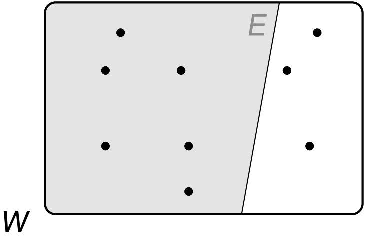
R. Fagin, J. Halpern, Y. Moses and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

The “Standard” Account



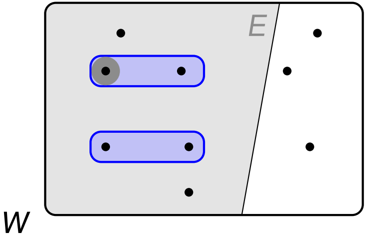
W is a set of **states** or **worlds**.

The "Standard" Account



An **event**/**proposition** is any (definable) subset $E \subseteq W$

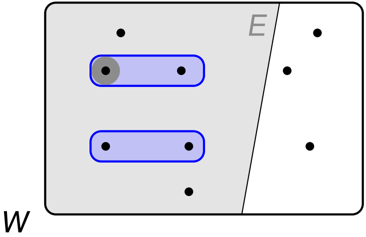
The "Standard" Account



At each state, agents are assigned a set of states they *consider possible* (according to their information).

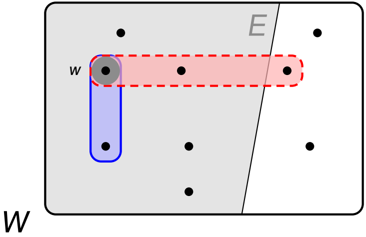
The information may be (in)correct, partitional, ...

The “Standard” Account



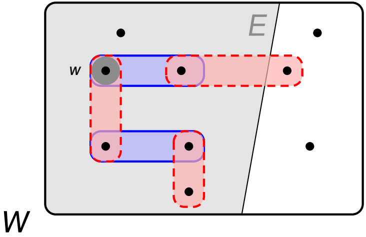
Knowledge Function: $K_i : \wp(W) \to \wp(W)$ where $K_i(E) = \{w \mid R_i(w) \subseteq E\}$

The “Standard” Account



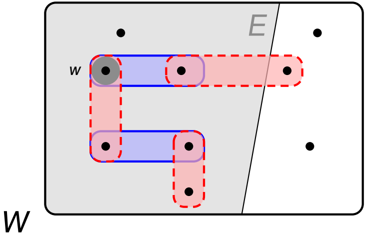
$w \in K_A(E)$ and $w \notin K_B(E)$

The “Standard” Account



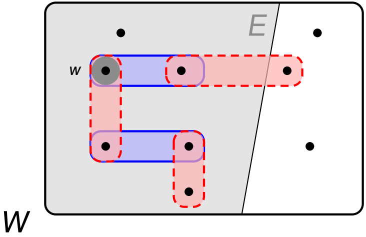
The model also describes the agents’ **higher-order knowledge/beliefs**

The "Standard" Account



Everyone Knows: $K(E) = \bigcap_{i \in \mathcal{A}} K_i(E)$, $K^0(E) = E$, $K^m(E) = K(K^{m-1}(E))$

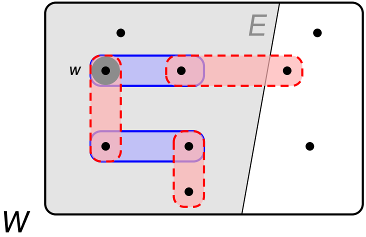
The "Standard" Account



Common Knowledge: $C : \wp(W) \rightarrow \wp(W)$ with

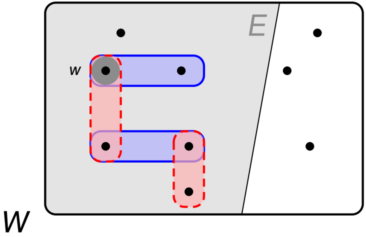
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$

The “Standard” Account



$w \in K(E)$ $\qquad$ $w \notin C(E)$

The "Standard" Account



$w \in C(E)$

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). *(Aumann, pg. 271, footnote 8)*

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_iC(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. $w \in C(E)$ if every finite path starting at w ends in a state in E

The following axiomatize common knowledge:

- $C(\varphi \to \psi) \to (C\varphi \to C\psi)$
- $C\varphi \to (\varphi \wedge EC\varphi)$ (Fixed-Point)
- $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$ (Induction)

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

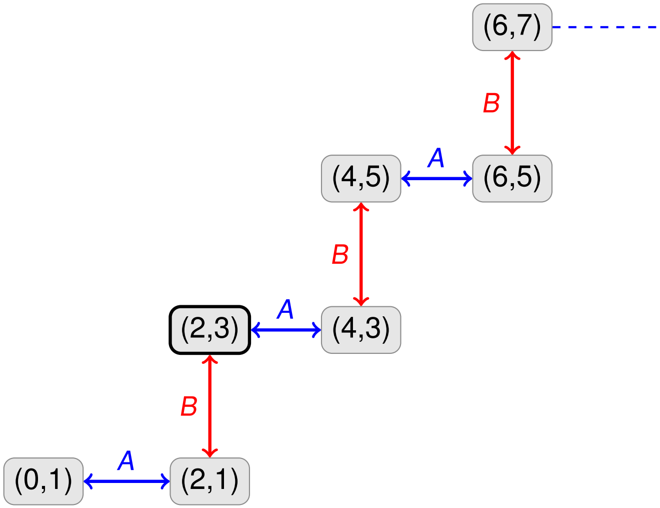
An Example

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

Is it common knowledge that their numbers are less than 1000?


(6,7)
B
(4,5)
A
(6,5)
B
(2,3)
A
(4,3)
B
(0,1)
A
(2,1)

Some Issues

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

Some Issues

- What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

- Where does common knowledge come from?

Some Issues

✓ What *does* a group know/believe/accept? vs. what *can* a group (come to) know/believe/accept?

C. List. *Group knowledge and group rationality: a judgment aggregation perspective*. Episteme (2008).

- Other "group informational attitudes": distributed knowledge, common belief, ...

- Where does common knowledge come from?

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i\in G} R_i(w)\right) \subseteq E\}$$

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i \in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \rightarrow q) \rightarrow D_{A,B}(q)$
- $D_G(\varphi) \rightarrow \bigwedge_{i \in G} K_i\varphi$

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i\in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \rightarrow q) \rightarrow D_{A,B}(q)$
- $D_G(\varphi) \rightarrow \bigwedge_{i\in G} K_i\varphi$

F. Roelofsen. *Distributed Knowledge*. Journal of Applied Nonclassical Logic (2006).

Distributed Knowledge

$$D_G(E) = \{w \mid \left(\bigcap_{i\in G} R_i(w)\right) \subseteq E\}$$

- $K_A(p) \wedge K_B(p \rightarrow q) \rightarrow D_{A,B}(q)$
- $D_G(\varphi) \rightarrow \bigwedge_{i\in G} K_i\varphi$

F. Roelofsen. *Distributed Knowledge*. Journal of Applied Nonclassical Logic (2006).

$w \in K_G(E)$ iff $R_G(w) \subseteq E$ (without necessarily $R_G(w) = \bigcap_{i\in G} R_i(w)$)

A. Baltag and S. Smets. *Correlated Knowledge: an Epistemic-Logic view on Quantum Entanglement*. Int. Journal of Theoretical Physics (2010).

Ingredients of a Logical Analysis of Rational Agency

- $\Rightarrow$ informational attitudes (eg., knowledge, belief, certainty)
- $\Rightarrow$ time, actions and ability
- $\Rightarrow$ motivational attitudes (eg., preferences)
- $\Rightarrow$ group notions (e.g., common knowledge and coalitional ability)
- $\Rightarrow$ normative attitudes (eg., obligations)

Ingredients of a Logical Analysis of Rational Agency

✓ informational attitudes (eg., knowledge, belief, certainty)

$\Rightarrow$ time, actions and ability

$\Rightarrow$ motivational attitudes (eg., preferences)

✓ group notions (e.g., common knowledge)

$\Rightarrow$ normative attitudes (eg., obligations)

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge in the group of the posterior probabilities, then the posteriors must be equal.

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge in the group of the posterior probabilities, then the posteriors must be equal.

S. Morris. *The common prior assumption in economic theory*. Economics and Philosophy, 11, pgs. 227 - 254, 1995.

Generalized Aumann's Theorem

Qualitative versions: *like-minded individuals cannot agree to make different decisions*.

M. Bacharach. *Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge*. Journal of Economic Theory (1985).

J.A.K. Cave. *Learning to Agree*. Economic Letters (1983).

D. Samet. *Agreeing to disagree: The non-probabilistic case*. Games and Economic Behavior, Vol. 69, 2010, 169-174.

The Framework

Knowledge Structure: $\langle W, \{\Pi_i\}_{i\in\mathcal{A}}\rangle$ where each Π_i is a partition on W ($\Pi_i(w)$ is the cell in Π_i containing w).

Decision Function: Let D be a nonempty set of **decisions.** A decision function for $i \in \mathcal{A}$ is a function $\mathbf{d}_i : W \to D$. A vector $\mathbf{d} = (d_1, \ldots, d_n)$ is a decision function profile. Let $[\mathbf{d}_i = d] = \{w \mid \mathbf{d}_i(w) = d\}$.

The Framework

Knowledge Structure: $\langle W, \{\Pi_i\}_{i\in\mathcal{A}}\rangle$ where each Π_i is a partition on W ($\Pi_i(w)$ is the cell in Π_i containing w).

Decision Function: Let D be a nonempty set of **decisions.** A decision function for $i \in \mathcal{A}$ is a function $\mathbf{d}_i : W \to D$. A vector $\mathbf{d} = (d_1, \ldots, d_n)$ is a decision function profile. Let $[\mathbf{d}_i = d] = \{w \mid \mathbf{d}_i(w) = d\}$.

(A1) Each agent knows her own decision:

$$[\mathbf{d}_i = d] \subseteq K_i([\mathbf{d}_i = d])$$

Comparing Knowledge

$[j \succeq i]$: agent j is at least as knowledgeable as agent i.

$$[j \succeq i] := \bigcap_{E \in \wp(W)} (K_i(E) \Rightarrow K_j(E)) = \bigcap_{E \in \wp(W)} (\neg K_i(E) \cup K_j(E))$$

Comparing Knowledge

$[j \succeq i]$: agent j is at least as knowledgeable as agent i.

$$[j \succeq i] := \bigcap_{E \in \wp(W)} (K_i(E) \Rightarrow K_j(E)) = \bigcap_{E \in \wp(W)} (\neg K_i(E) \cup K_j(E))$$

$w \in [j \succeq i]$ then j knows at w every event that i knows there.

Comparing Knowledge

$[j \succeq i]$: agent j is at least as knowledgeable as agent i.

$$[j \succeq i] := \bigcap_{E \in \wp(W)} (K_i(E) \Rightarrow K_j(E)) = \bigcap_{E \in \wp(W)} (\neg K_i(E) \cup K_j(E))$$

$w \in [j \succeq i]$ then j knows at w every event that i knows there.

$[j \sim i] = [j \succeq i] \cap [i \succeq j]$

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant.

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew *that the Democratic candidate were going to win, and decides that he would.*

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew *that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he* knew *a Republican candidate were going to win, and again he finds that he would.*

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew *that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he* knew *a Republican candidate were going to win, and again he finds that he would. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, as we would ordinarily say.* *(Savage, 1954)*

The sure-thing principle cannot appropriately be accepted as a postulate...because it would introduce new undefined technical terms referring to knowledge and possibility that would render it mathematically useless without still more postulates governing these terms. It will be preferable to regard the principle as a loose one that suggests certain formal postulates well articulated with P1 [the transitivity of preferences] (Savage, 1954)

Sure-Thing Principle

Should I study or have a beer?

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam.

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam. If I pass, it is better to drink and pass, so I should drink. If I fail, it is better to drink and fail, so I should drink.

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam. If I pass, it is better to drink and pass, so I should drink. If I fail, it is better to drink and fail, so I should drink. I should drink in either case, so I should have a drink.

Sure-Thing Principle

It is not the logical principle $\varphi \to \chi$ and $\psi \to \chi$ then $\varphi \vee \psi \to \chi$.

Sure-Thing Principle

R. Aumann, S. Hart and M. Perry. *Conditioning and the Sure-Thing Principle*. manuscript, 2005.

J. Pearl. *The Sure-Thing Principle*. Journal of Causal Inference, Causal, Casual, and Curious Section, 4(1):81-86, 2016.

Branden Fitelson. *Confirmation, Causation, and Simpson's Paradox*. Episteme, 2017.

"Change Savage's example to make the election be merely for the office of mayor, and suppose that the businessman thinks—perhaps correctly, and perhaps with excellent reason—that his buying the property would improve the Democratic contender's chances of winning."

“Change Savage’s example to make the election be merely for the office of mayor, and suppose that the businessman thinks—perhaps correctly, and perhaps with excellent reason—that his buying the property would improve the Democratic contender’s chances of winning.”

Imagine that the businessman believes that the Democratic candidate, if elected mayor, would be a disaster to the city, regardless of whether he buys the property of not. Under such circumstances, it is quite reasonable that buying the property would be a good post-election deal, regardless of which candidate wins, yet a terrible pre-election deal, prior to knowing the winner, in blatant violation of the sure-thing principle.

The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove.

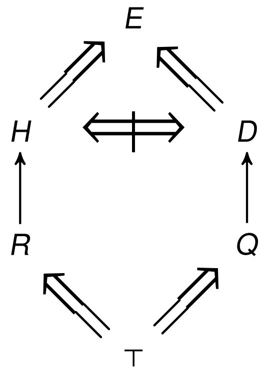
The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove. Either being a Hawk or a Dove implies having extreme political views.

The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove. Either being a Hawk or a Dove implies having extreme political views. Should you conclude that Nixon has extreme political views?

Floating Conclusions



J. Horty. *Skepticism and floating conclusions*. Artificial Intelligence, 135, pp. 55 - 72, 2002.

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M).

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father).

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother).

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother). You want to buy a yacht which requires a large deposit and you can only afford it provided you inherit the money.

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother). You want to buy a yacht which requires a large deposit and you can only afford it provided you inherit the money. Should you make a deposit on the yacht?

Interpersonal Sure-Thing Principle (ISTP)

For any pair of agents i and j and decision d,

$$K_i([j \succeq i] \cap [\mathbf{d}_j = d]) \subseteq [\mathbf{d}_i = d]$$

Interpersonal Sure-Thing Principle (ISTP): Illustration

Suppose that Alice and Bob, two detectives who graduated the same police academy, are assigned to investigate a murder case.

Interpersonal Sure-Thing Principle (ISTP): Illustration

Suppose that Alice and Bob, two detectives who graduated the same police academy, are assigned to investigate a murder case. If they are exposed to different evidence, they may reach different decisions.

Interpersonal Sure-Thing Principle (ISTP): Illustration

Suppose that Alice and Bob, two detectives who graduated the same police academy, are assigned to investigate a murder case. If they are exposed to different evidence, they may reach different decisions. Yet, being the students of the same academy, the method by which they arrive at their conclusions is the same.

Interpersonal Sure-Thing Principle (ISTP): Illustration

Suppose that Alice and Bob, two detectives who graduated the same police academy, are assigned to investigate a murder case. If they are exposed to different evidence, they may reach different decisions. Yet, being the students of the same academy, the method by which they arrive at their conclusions is the same. Suppose now that detective Bob, a father of four who returns home every day at five oclock, collects all the information about the case at hand together with detective Alice.

Interpersonal Sure-Thing Principle (ISTP): Illustration

However, Alice, single and a workaholic, continues to collect more information every day until the wee hours of the morning — information which she does not necessarily share with Bob.

Interpersonal Sure-Thing Principle (ISTP): Illustration

However, Alice, single and a workaholic, continues to collect more information every day until the wee hours of the morning — information which she does not necessarily share with Bob. Obviously, Bob knows that Alice is at least as knowledgeable as he is.

Interpersonal Sure-Thing Principle (ISTP): Illustration

However, Alice, single and a workaholic, continues to collect more information every day until the wee hours of the morning — information which she does not necessarily share with Bob. Obviously, Bob knows that Alice is at least as knowledgeable as he is. Suppose that he also knows what Alices decision is.

Interpersonal Sure-Thing Principle (ISTP): Illustration

However, Alice, single and a workaholic, continues to collect more information every day until the wee hours of the morning — information which she does not necessarily share with Bob. Obviously, Bob knows that Alice is at least as knowledgeable as he is. Suppose that he also knows what Alices decision is. Since Alice uses the same investigation method as Bob, he knows that had he been in possession of the more extensive knowledge that Alice has collected, he would have made the same decision as she did. Thus, this is indeed his decision.

Implications of ISTP

Proposition. If the decision function profile **d** satisfies ISTP, then

$$[i \sim j] \subseteq \bigcup_{d \in D} ([\mathbf{d}_i = d] \cap [\mathbf{d}_j = d])$$

ISTP Expandability

Agent i is an **epistemic dummy** if it is always the case that all the agents are at least as knowledgeable as i. That is, for each agent j,

$$[j \geq i] = W$$

A decision function profile $\mathbf{d}$ on $\langle W, \Pi_1, \ldots, \Pi_n \rangle$ is **ISTP expandable** if for any expanded structure $\langle W, \Pi_1, \ldots, \Pi_{n+1} \rangle$ where $n+1$ is an epistemic dummy, there exists a decision function $\mathbf{d}_{n+1}$ such that $(\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{n+1})$ satisfies ISTP.

ISTP Expandability: Illustration

Suppose that after making their decisions, Alice and Bob are told that another detective, one E.P. Dummy, who graduated the very same police academy, had also been assigned to investigate the same case.

ISTP Expandability: Illustration

Suppose that after making their decisions, Alice and Bob are told that another detective, one E.P. Dummy, who graduated the very same police academy, had also been assigned to investigate the same case. In principle, they would need to review their decisions in light of the third detectives knowledge: knowing what they know about the third detective, his usual sources of information, for example, may impinge upon their decision.

ISTP Expandability: Illustration

But this is not so in the case of detective Dummy. It is commonly known that the only information source of this detective, known among his colleagues as the couch detective, is the TV set.

ISTP Expandability: Illustration

But this is not so in the case of detective Dummy. It is commonly known that the only information source of this detective, known among his colleagues as the couch detective, is the TV set. Thus, it is commonly known that every detective is at least as knowledgeable as Dummy.

ISTP Expandability: Illustration

But this is not so in the case of detective Dummy. It is commonly known that the only information source of this detective, known among his colleagues as the couch detective, is the TV set. Thus, it is commonly known that every detective is at least as knowledgeable as Dummy. The news that he had been assigned to the same case is completely irrelevant to the conclusions that Alice and Bob have reached. Obviously, based on the information he gets from the media, Dummy also makes a decision.

ISTP Expandability: Illustration

But this is not so in the case of detective Dummy. It is commonly known that the only information source of this detective, known among his colleagues as the couch detective, is the TV set. Thus, it is commonly known that every detective is at least as knowledgeable as Dummy. The news that he had been assigned to the same case is completely irrelevant to the conclusions that Alice and Bob have reached. Obviously, based on the information he gets from the media, Dummy also makes a decision. We may assume that the decisions made by the three detectives satisfy the ISTP, for exactly the same reason we assumed it for the two detectives decisions

Generalized Agreement Theorem

If **d** is an ISTP expandable decision function profile on a partition structure $\langle W, \Pi_1, \ldots, \Pi_n \rangle$, then for any decisions $d_1, \ldots, d_n$ which are not identical, $C(\bigcap_i [\mathbf{d}_i = d_i]) = \emptyset$.

Dynamic characterization of Aumann's Theorem

- How do the posteriors *become* common knowledge?

J. Geanakoplos and H. Polemarchakis. *We Can't Disagree Forever*. Journal of Economic Theory (1982).

Dynamic characterization of Aumann's Theorem

- How do the posteriors *become* common knowledge?

J. Geanakoplos and H. Polemarchakis. *We Can't Disagree Forever*. Journal of Economic Theory (1982).

- What happens when communication is not the the whole group, but pairwise?

R. Parikh and P. Krasucki. *Communication, Consensus and Knowledge*. Journal of Economic Theory (1990).

Dynamic Logics of Knowledge and Belief

Fitch's Paradox

Fitch (1963) derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*:

Fitch's Paradox

Fitch (1963) derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*:

(VT) $q \to \Diamond Kq$,

where $\Diamond$ is a *possibility* operator (more on this later).

Fitch's Paradox

Fitch (1963) derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*:

(VT) $q \rightarrow \Diamond Kq$,

where $\Diamond$ is a *possibility* operator (more on this later).

Fitch make two modest assumptions for K, $K\varphi \rightarrow \varphi$ (T) and $K(\varphi \wedge \psi) \rightarrow (K\varphi \wedge K\psi)$ (M), and two modest assumptions for $\Diamond$:

Fitch's Paradox

Fitch (1963) derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*:

(VT) $q \rightarrow \Diamond Kq$,

where $\Diamond$ is a *possibility* operator (more on this later).

Fitch make two modest assumptions for K, $K\varphi \rightarrow \varphi$ (T) and $K(\varphi \wedge \psi) \rightarrow (K\varphi \wedge K\psi)$ (M), and two modest assumptions for $\Diamond$:

- $\Diamond$ is the dual of $\Box$ for *necessity*, so $\neg\Diamond\varphi$ follows from $\Box\neg\varphi$.

Fitch's Paradox

Fitch (1963) derived an unexpected consequence from the thesis, advocated by some anti-realists, that *every truth is knowable*:

(VT) $q \rightarrow \Diamond Kq$,

where $\Diamond$ is a *possibility* operator (more on this later).

Fitch make two modest assumptions for K, $K\varphi \rightarrow \varphi$ (T) and $K(\varphi \wedge \psi) \rightarrow (K\varphi \wedge K\psi)$ (M), and two modest assumptions for $\Diamond$:

- $\Diamond$ is the dual of $\Box$ for *necessity*, so $\neg\Diamond\varphi$ follows from $\Box\neg\varphi$.
- $\Box$ obeys the rule of Necessitation: if φ is a theorem, so is $\Box\varphi$.

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge \neg Kp)$ from (1) and (2) by PL

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \to (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \to \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \to (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

(5) $\Box\neg K(p \wedge \neg Kp)$ from (4) by $\Box$-Necessitation

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

(5) $\Box\neg K(p \wedge \neg Kp)$ from (4) by $\Box$-Necessitation

(6) $\neg\Diamond K(p \wedge \neg Kp)$ from (5) by $\Box$ - $\Diamond$ Duality

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \to (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \to \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \to (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

(5) $\Box\neg K(p \wedge \neg Kp)$ from (4) by $\Box$-Necessitation

(6) $\neg\Diamond K(p \wedge \neg Kp)$ from (5) by $\Box$ - $\Diamond$ Duality

(7) $\neg(p \wedge \neg Kp)$ from (0) by PL

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

(5) $\Box\neg K(p \wedge \neg Kp)$ from (4) by $\Box$-Necessitation

(6) $\neg\Diamond K(p \wedge \neg Kp)$ from (5) by $\Box$ - $\Diamond$ Duality

(7) $\neg(p \wedge \neg Kp)$ from (0) by PL

(8) $p \rightarrow Kp$ from (7) by classical PL

Fitch's Paradox

For an arbitrary p, consider the following instance of (VT):

(0) $(p \wedge \neg Kp) \rightarrow \Diamond K(p \wedge \neg Kp)$

Here is the proof for Fitch's Paradox:

(1) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$ instance of M axiom

(2) $K\neg Kp \rightarrow \neg Kp$ instance of T axiom

(3) $K(p \wedge \neg Kp) \rightarrow (Kp \wedge \neg Kp)$ from (1) and (2) by PL

(4) $\neg K(p \wedge \neg Kp)$ from (3) by PL

(5) $\Box\neg K(p \wedge \neg Kp)$ from (4) by $\Box$-Necessitation

(6) $\neg\Diamond K(p \wedge \neg Kp)$ from (5) by $\Box$ - $\Diamond$ Duality

(7) $\neg(p \wedge \neg Kp)$ from (0) by PL

(8) $p \rightarrow Kp$ from (7) by classical PL

Since p was arbitrary, we have shown that *every truth is known*.

The Question

Fitch's Paradox leaves us with **the question**: what must we require in addition to the truth of φ to ensure the knowability of φ?

The Question

Fitch's Paradox leaves us with **the question**: what must we require in addition to the truth of φ to ensure the knowability of φ?

There is a fairly large literature on knowability and related issues. See, e.g.:

J. Salerno. 2009. *New Essays on the Knowability Paradox*, OUP

J. van Benthem. 2004. "What One May Come to Know," *Analysis*.

P. Balbiani et al. 2008. "'Knowable' as 'Known after an Announcement,'" *Review of Symbolic Logic*.

Dynamic Epistemic Logic

The key idea of dynamic epistemic logic is that we can represent changes in agents' epistemic states by *transforming models*.

Dynamic Epistemic Logic

The key idea of dynamic epistemic logic is that we can represent changes in agents' epistemic states by *transforming models*.

In the simplest case, we model an agent's acquisition of knowledge by the elimination of possibilities from an initial epistemic model.

Finding out that φ

$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$$

$$\Downarrow$$

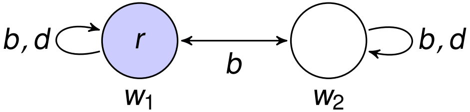
Find out that φ

$$\Downarrow$$

$$\mathcal{M}' = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V|_{W'}\rangle$$

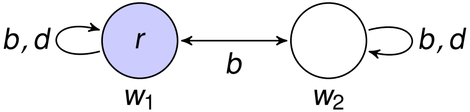
Example: College Park and Amsterdam

Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:



Example: College Park and Amsterdam

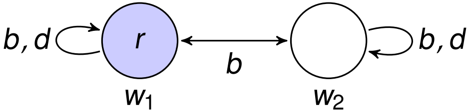
Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:



What happens when the Amsterdam agent calls the College Park agent on the phone and says, "It's raining in Amsterdam"?

Example: College Park and Amsterdam

Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:

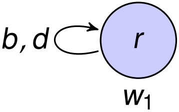


What happens when the Amsterdam agent calls the College Park agent on the phone and says, "It's raining in Amsterdam"?

We model the change in b's epistemic state by eliminating all epistemic possibilities in which it's *not* raining in Amsterdam.

Example: College Park and Amsterdam

Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:



What happens when the Amsterdam agent calls the College Park agent on the phone and says, "It's raining in Amsterdam"?

We model the change in *b*'s epistemic state by eliminating all epistemic possibilities in which it's *not* raining in Amsterdam.

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model s.th.:

$$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\};$$

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model s.th.:

$$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\};$$

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model s.th.:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

$V_{|\varphi}(p)$ is the intersection of $V(p)$ and $W_{|\varphi}$.

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model s.th.:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

$V_{|\varphi}(p)$ is the intersection of $V(p)$ and $W_{|\varphi}$.

In the single-agent case, this models the agent learning φ. In the multi-agent case, this models all agents *publicly* learning φ.

Public Announcement Logic

One of the **big ideas** of dynamic epistemic logic is to add to our formal language operators that can describe the kinds of model updates that we just saw for the College Park and Amsterdam example.

Public Announcement Logic

One of the **big ideas** of dynamic epistemic logic is to add to our formal language operators that can describe the kinds of model updates that we just saw for the College Park and Amsterdam example.

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Public Announcement Logic

One of the **big ideas** of dynamic epistemic logic is to add to our formal language operators that can describe the kinds of model updates that we just saw for the College Park and Amsterdam example.

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Public Announcement Logic

One of the **big ideas** of dynamic epistemic logic is to add to our formal language operators that can describe the kinds of model updates that we just saw for the College Park and Amsterdam example.

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle!\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

- $\mathcal{M}, w \models \langle !\varphi\rangle\psi$ iff $\mathcal{M}, w \models \varphi$ *and* $\mathcal{M}_{|\varphi}, w \models \psi$.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

- $\mathcal{M}, w \models \langle !\varphi\rangle\psi$ iff $\mathcal{M}, w \models \varphi$ *and* $\mathcal{M}_{|\varphi}, w \models \psi$.

Big Idea: we evaluate $[!\varphi]\psi$ and $\langle !\varphi\rangle\psi$ not by looking at *other worlds in the same model*, but rather by looking at a new model.

Public Announcement Logic

Suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ is a multi-agent Kripke Model

$$\mathcal{M}, w \models [\psi]\varphi \text{ iff } \mathcal{M}, w \models \psi \text{ implies } \mathcal{M}|_\psi, w \models \varphi$$

where $\mathcal{M}|_\psi = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V'\rangle$ with

- $W' = W \cap \{w \mid \mathcal{M}, w \models \psi\}$
- For each i, $\sim_i' = \sim_i \cap (W' \times W')$
- For each i, $\preceq_i' = \preceq_i \cap (W' \times W')$
- for all $p \in \text{At}$, $V'(p) = V(p) \cap W'$

Public Announcement Logic

$$[\psi]p \quad \leftrightarrow \quad (\psi \rightarrow p)$$

Public Announcement Logic

$$\begin{array}{rcl} [\psi]p & \leftrightarrow & (\psi \to p) \\ [\psi]\neg\varphi & \leftrightarrow & (\psi \to \neg[\psi]\varphi) \end{array}$$

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi)
\end{array}
$$

Public Announcement Logic

$$\begin{aligned}
[\psi]p &\leftrightarrow (\psi \to p) \\
[\psi]\neg\varphi &\leftrightarrow (\psi \to \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) &\leftrightarrow ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi &\leftrightarrow [\psi \wedge [\psi]\varphi]\chi
\end{aligned}$$

Public Announcement Logic

$$\begin{aligned}
[\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) &\leftrightarrow ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi &\leftrightarrow [\psi \wedge [\psi]\varphi]\chi \\
[\psi]K_i\varphi &\leftrightarrow (\psi \rightarrow K_i(\psi \rightarrow [\psi]\varphi))
\end{aligned}$$

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi & \leftrightarrow & [\psi \wedge [\psi]\varphi]\chi \\
[\psi]K_i\varphi & \leftrightarrow & (\psi \rightarrow K_i(\psi \rightarrow [\psi]\varphi))
\end{array}
$$

Theorem Every formula of Public Announcement Logic is equivalent to a formula of Epistemic Logic.

- $[q]Kq$

- $[q]Kq$

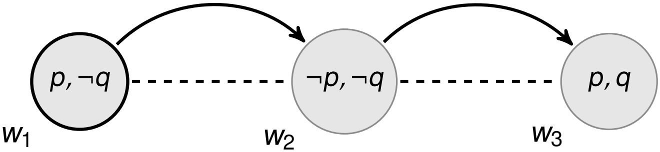
- $Kp \rightarrow [q]Kp$

- $[q]Kq$
- $Kp \rightarrow [q]Kp$
- $B\varphi \rightarrow [\psi]B\varphi$

- $[q]Kq$

- $Kp \rightarrow [q]Kp$

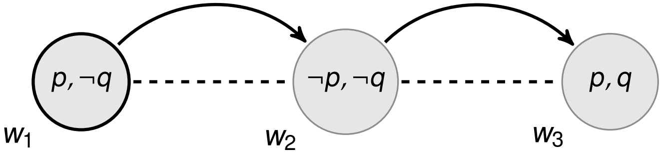
- $B\varphi \rightarrow [\psi]B\varphi$



- $[q]Kq$

- $Kp \rightarrow [q]Kp$

- $B\varphi \rightarrow [\psi]B\varphi$



- $[\varphi]\varphi$

Public Announcement vs. Conditional Belief

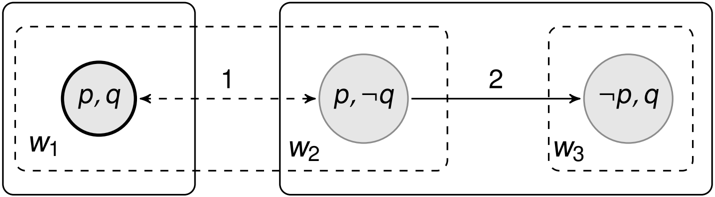
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different?

Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!

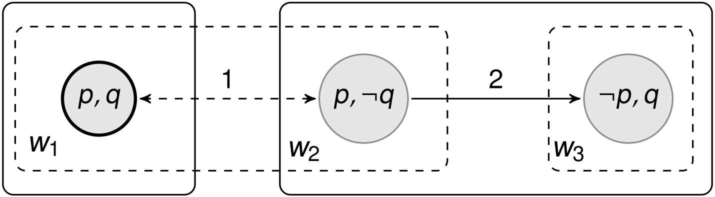
Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



Public Announcement vs. Conditional Belief

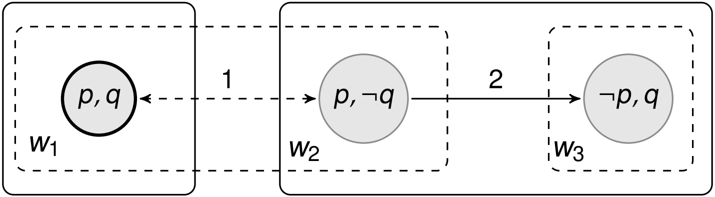
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$

Public Announcement vs. Conditional Belief

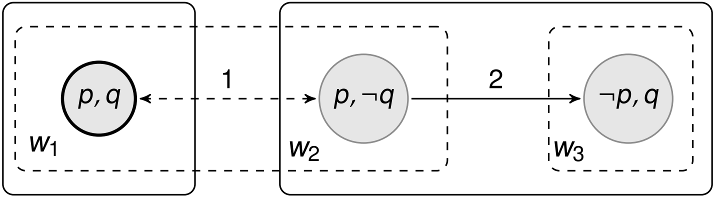
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$

Public Announcement vs. Conditional Belief

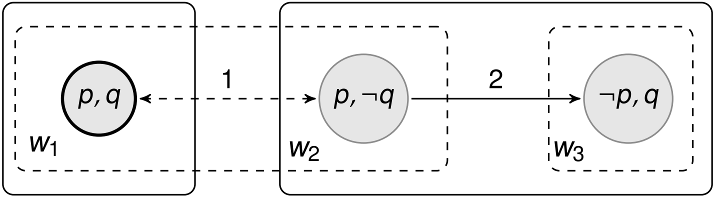
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [p]\neg B_1 B_2 q$

Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [p]\neg B_1 B_2 q$
- More generally, $B_i^p(p \wedge \neg K_i p)$ is satisfiable but $[p]B_i(p \wedge \neg K_i p)$ is not.

The Logic of Public Observation

- $[\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [\varphi]\psi))$

The Logic of Public Observation

- $[\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [\varphi]\psi))$

- $[\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [\varphi]\psi))$

The Logic of Public Observation

- $[\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [\varphi]\psi))$

- $[\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [\varphi]\psi))$

- **Belief**: $[\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [\varphi]\psi))$

The Logic of Public Observation

- $[\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [\varphi]\psi))$

- $[\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [\varphi]\psi))$

- **Belief**: $[\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [\varphi]\psi))$

  $[\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[\varphi]\psi)$

The Logic of Public Observation

- $[\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [\varphi]\psi))$

- $[\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [\varphi]\psi))$

- **Belief**: $[\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [\varphi]\psi))$

  $[\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[\varphi]\psi)$
  $[\varphi]B^{\alpha}\psi \leftrightarrow (\varphi \rightarrow B^{\varphi\wedge[\varphi]\alpha}[\varphi]\psi)$

Finding out that φ

$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$$

$$\Downarrow$$

Find out that φ

$$\Downarrow$$

$$\mathcal{M}' = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V|_{W'}\rangle$$

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

- “Finding out that φ”
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...
- *How* did you find out that φ?
  - Directly observed φ
  - Indirectly observed φ
  - Told 'φ' (by an epistemic peer, by an expert, by a trusted individual)
  - ...
- Belief change over time

The Theory of Belief Revision

C. Alchourrón, P. Gärdenfors and D. Makinson. *On the logic of theory change: Partial meet contraction and revision functions*. Journal of Symbolic Logic, 50, 510 - 530, 1985.

Hans Rott. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford University Press, 2001.

A.P. Pedersen and H. Arló-Costa. *"Belief Revision."*. In Continuum Companion to Philosophical Logic. Continuum Press, 2011.

$$\mathcal{B} * \varphi$$

$\mathcal{B} * \varphi$

Initial set of beliefs

$\mathcal{B} * \varphi$

Initial set of beliefs

New evidence φ

Revision operator: $* : \mathcal{B} \times \mathcal{L} \to \mathcal{B}$

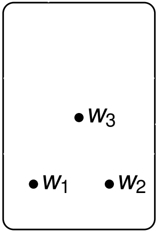
$$\mathcal{B} * \varphi$$

Initial set of beliefs

New evidence φ

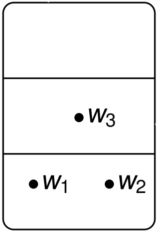
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$



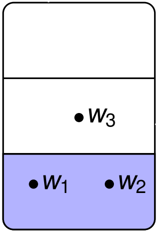
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

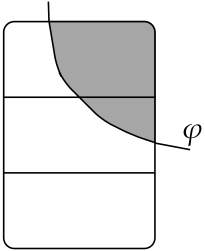


Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

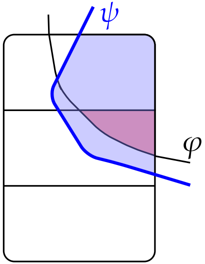


Belief Revision via Plausibility



Conditional Belief: $B^{\varphi}\psi$

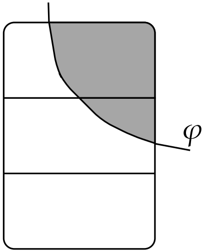
Belief Revision via Plausibility



Conditional Belief: $B^{\varphi}\psi$

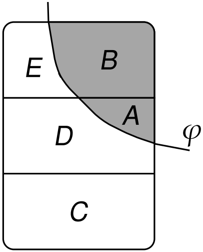
$$Min_{\leq}([\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

Belief Revision via Plausibility



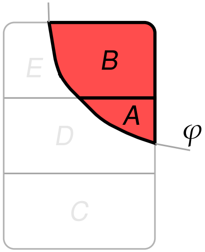
Incorporate the new information φ

Belief Revision via Plausibility



Incorporate the new information φ

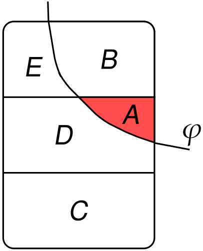
Belief Revision via Plausibility



Public Announcement: Information from an infallible source

$(!\varphi)$: $A \prec_i B$

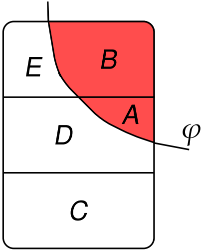
Belief Revision via Plausibility



Public Announcement: Information from an infallible source
$(!\varphi)$: $A \prec_i B$

Conservative Upgrade: Information from a trusted source
$(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

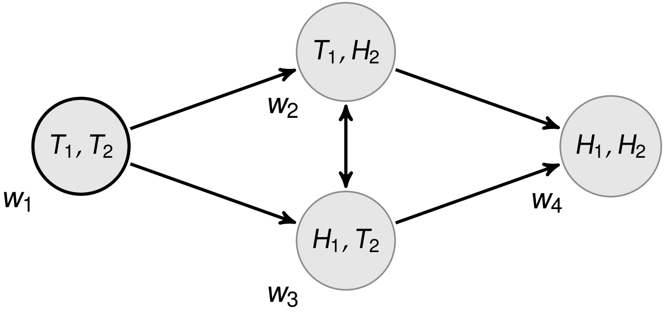
Belief Revision via Plausibility



Public Announcement: Information from an infallible source
$(!\varphi)$: $A \prec_i B$

Conservative Upgrade: Information from a trusted source
$(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

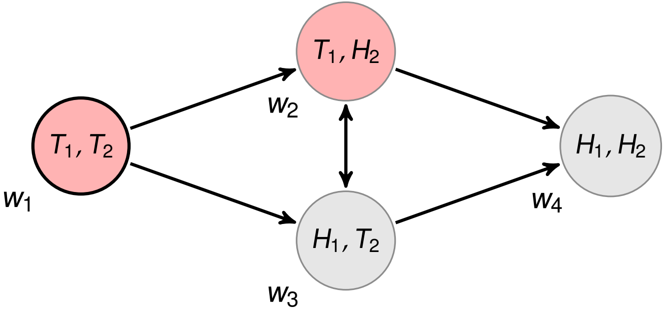
Radical Upgrade: Information from a strongly trusted source
$(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$



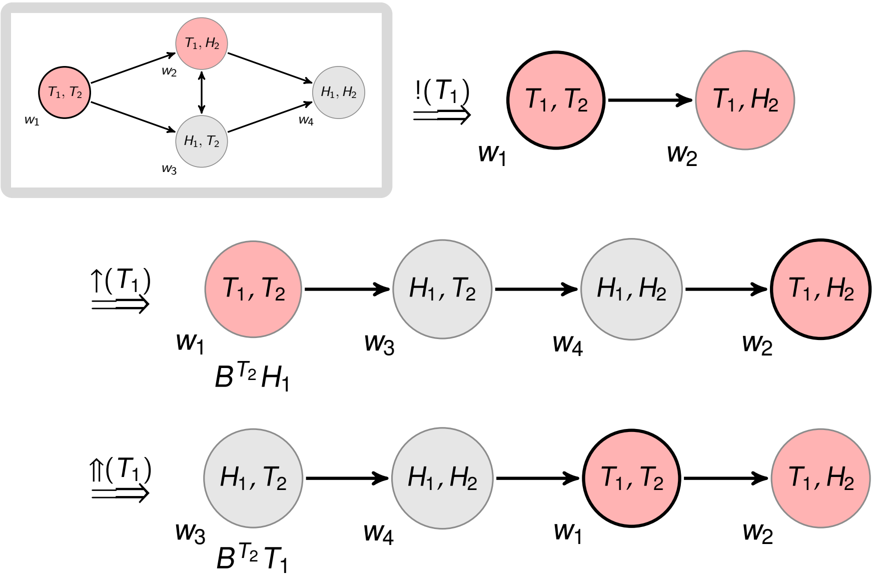
$Min_{\leq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

$Min_{\leq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_2\}$, so $w_1 \models B^{T_1} H_2$

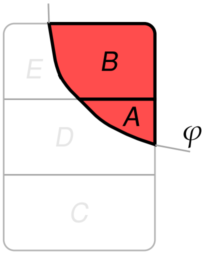
$Min_{\leq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_3\}$, so $w_1 \models B^{T_2} H_1$



Suppose the agent *finds out that* T_1 *is true*.


T_1, T_2
w_1
T_1, H_2
w_2
H_1, T_2
w_3
H_1, H_2
w_4
$\stackrel{!(T_1)}{\Longrightarrow}$
T_1, T_2
w_1
T_1, H_2
w_2
$\stackrel{\uparrow(T_1)}{\Longrightarrow}$
T_1, T_2
w_1
$B^{T_2} H_1$
H_1, T_2
w_3
H_1, H_2
w_4
T_1, H_2
w_2
$\stackrel{\Uparrow(T_1)}{\Longrightarrow}$
H_1, T_2
w_3
$B^{T_2} T_1$
H_1, H_2
w_4
T_1, T_2
w_1
T_1, H_2
w_2

Informative Actions

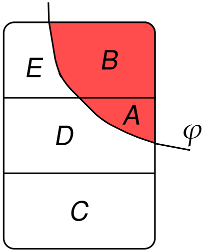


Public Announcement: Information from an infallible source

(!φ): $A \prec_i B$ $\quad\quad \mathcal{M}^{!\varphi} = \langle W^{!\varphi}, \{\sim_i^{!\varphi}\}_{i \in \mathcal{A}}, V^{!\varphi} \rangle$

$W^{!\varphi} = [\![\varphi]\!]_{\mathcal{M}}$
$\sim_i^{!\varphi} = \sim_i \cap (W^{!\varphi} \times W^{!\varphi})$
$\preceq_i^{!\varphi} = \preceq_i \cap (W^{!\varphi} \times W^{!\varphi})$

Informative Actions



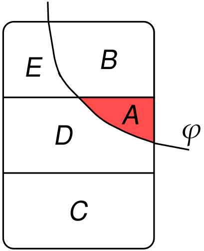
Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,

$\mathcal{M}^{\Uparrow\varphi} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathcal{A}}, V\rangle$

Let $[\![\varphi]\!]_i^w = \{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i$

- for all $x \in [\![\varphi]\!]_i^w$ and $y \in [\![\neg\varphi]\!]_i^w$, set $x \prec_i^{\Uparrow\varphi} y$,
- for all $x, y \in [\![\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$, and
- for all $x, y \in [\![\neg\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$.

Informative Actions



Conservative Upgrade: $(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

Conservative upgrade is radical upgrade with the formula

$$best_i(\varphi, w) := Min_{\preceq_i}([w]_i \cap \{x \mid \mathcal{M}, x \models \varphi\})$$

1. If $v \in best_i(\varphi, w)$ then $v \prec_i^{\uparrow\varphi} x$ for all $x \in [w]_i$, and
2. for all $x, y \in [w]_i - best_i(\varphi, w)$, $x \preceq_i^{\uparrow\varphi} y$ iff $x \preceq_i y$.

Recursion Axioms

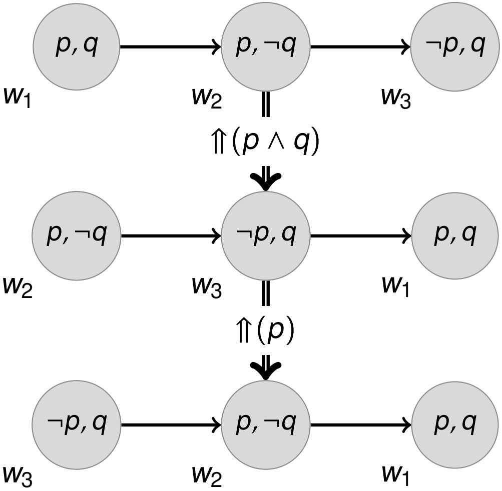
$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

Recursion Axioms

$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

$$[\uparrow\varphi]B^{\psi}\chi \leftrightarrow (B^{\varphi}\neg[\uparrow\varphi]\psi\wedge B^{[\uparrow\varphi]\psi}[\uparrow\varphi]\chi)\vee(\neg B^{\varphi}\neg[\uparrow\varphi]\psi\wedge B^{\varphi\wedge[\uparrow\varphi]\psi}[\uparrow\varphi]\chi)$$

Composition



Iterated Updates

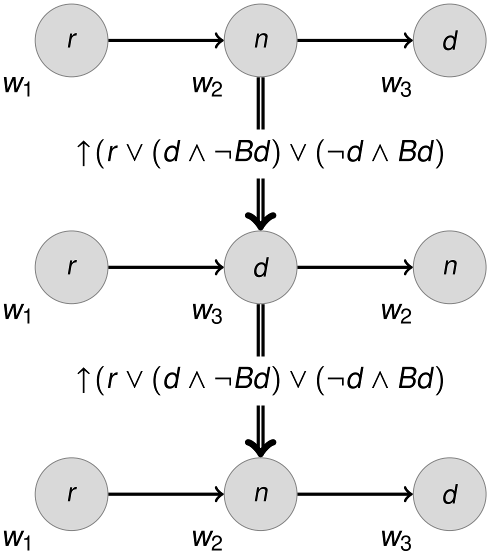
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \ \Uparrow \neg p \ \Uparrow p \cdots$
Contradictory beliefs leads to oscillations

$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

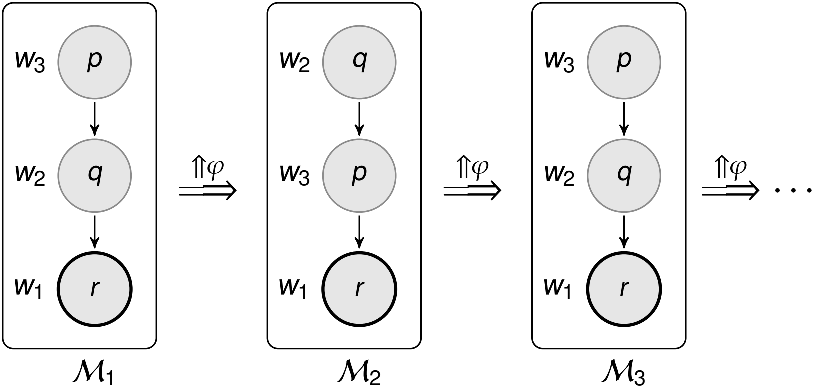
$\Uparrow\varphi, \Uparrow\varphi, \ldots$
Simple beliefs stabilize, but conditional beliefs do not

A. Baltag and S. Smets. *Group Belief Dynamics under Iterated Revision: Fixed Points and Cycles of Joint Upgrades*. TARK, 2009.



r
n
d
w_1
w_2
w_3
$\uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$
r
d
n
w_1
w_3
w_2
$\uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$
r
n
d
w_1
w_2
w_3

Let φ be $(r \vee (B^{\neg r} q \wedge p) \vee (B^{\neg r} p \wedge q))$



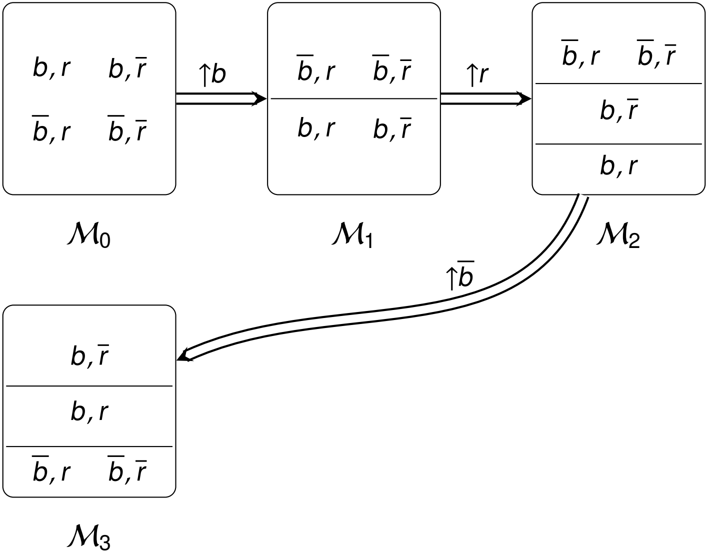
Suppose that you are in the forest and happen to a see strange-looking animal.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red. So, you also update your beliefs with that fact.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red. So, you also update your beliefs with that fact. Now, suppose that an expert (whom you trust) happens to walk by and tells you that the animal is, in fact, not a bird.

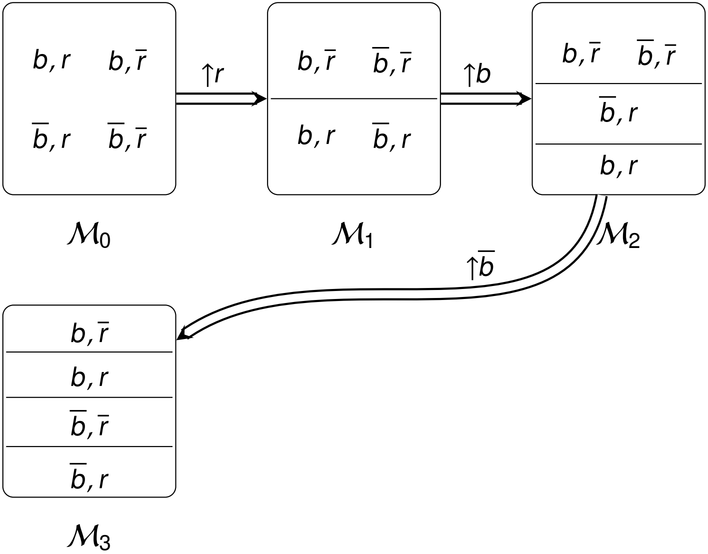


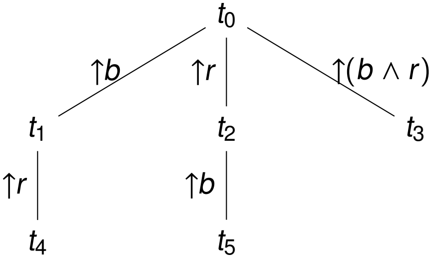
b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow b$
$\overline{b}, r$ $\overline{b}, \overline{r}$
b, r $b, \overline{r}$
$\mathcal{M}_1$
$\uparrow r$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_3$

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red.

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with the information that the animal is a bird.

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with the information that the animal is a bird. In particular, note that the following sequence of updates is not problematic:


b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\Uparrow r$
$b, \overline{r}$ $\overline{b}, \overline{r}$
b, r $\overline{b}, r$
$\Uparrow b$
$b, \overline{r}$ $\overline{b}, \overline{r}$
$\overline{b}, r$
b, r
$\mathcal{M}_0$
$\mathcal{M}_1$
$\mathcal{M}_2$
$\Uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, \overline{r}$
$\overline{b}, r$
$\mathcal{M}_3$


t_0
$\uparrow b$
$\uparrow r$
$\uparrow(b \wedge r)$
t_1
t_2
t_3
$\uparrow r$
$\uparrow b$
t_4
t_5

R. Stalnaker. *Iterated Belief Revision*. Erkenntnis 70, pgs. 189 - 209, 2009.

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

- Postulate I1 demands if $\varphi \to \psi$ is a theorem (with respect to the background theory), then first learning ψ followed by the more specific information φ is equivalent to directly learning the more specific information φ.

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

- Postulate I1 demands if $\varphi \rightarrow \psi$ is a theorem (with respect to the background theory), then first learning ψ followed by the more specific information φ is equivalent to directly learning the more specific information φ.

- Postulate I2 demands that first learning φ followed by learning a piece of information ψ incompatible with φ is the same as simply learning ψ outright. So, for example, first learning φ and then $\neg\varphi$ should result in the same belief state as directly learning $\neg\varphi$.

Stalnaker's Counterexample to I1

UUU	*DDD*
UUD	*DDU*
UDU	*DUD*
UDD	*DUU*

- Three switches wired such that a light is on iff all three switches are up or all three are down.

Stalnaker's Counterexample to I1

UUU	*DDD*
UUD	*DDU*
UDU	*DUD*
UDD	*DUU*

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is *U*, Bob says switch 2 is *D* and Carla says switch 3 is *U*.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should I believe?

Stalnaker's Counterexample to I1

UUU	*DDD*
UUD	*DDU*
UDU	*DUD*
UDD	*DUU*

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is *U*, Bob says switch 2 is *D* and Carla says switch 3 is *U*.
- I receive the information that the light is on. What should I believe?
- Cautious: *UUU*, *DDD*; Bold: *UUU*

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \to L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.
- Now, after learning L_1, the only rational thing to believe is that all three switches are up.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.
- Now, after learning L_1, the only rational thing to believe is that all three switches are up.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- So, $L_2 \in Cn(\{L_1\})$ but (potentially) $(K * L_2) * L_1 \neq K * L_1$.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.
- Finally, Elmer, a third witness considered the most reliable overall, reports that the coin in box 1 is lying heads up.

H_i (T_i): the coin in box i facing heads (tails) up.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

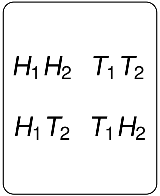
- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

  Yet, since $H_1 \wedge H_2 \in K'$ and H_1 is consistent with H_2, we must have $H_1 \wedge H_2 \in K' * H_1$, which yields a conflict with the assumption that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.

...[Postulate I2] directs us to take back the totality of any information that is overturned. Specifically, if we first receive information α, and then receive information that conflicts with α, we should return to the belief state we were previously in, before learning α. But this directive is too strong. Even if the new information conflicts with the information just received, it need not necessarily cast doubt on all *of that information.*

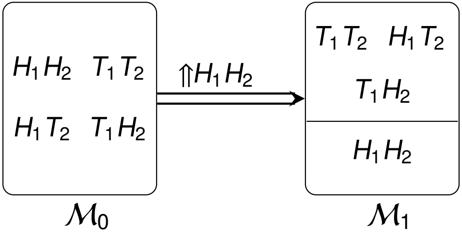
(Stalnaker, pg. 207–208)

Heuristic Diagnosis of Stalnaker’s Example

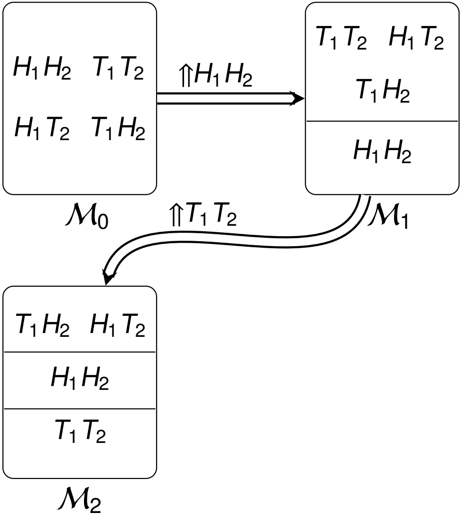


$\mathcal{M}_0$

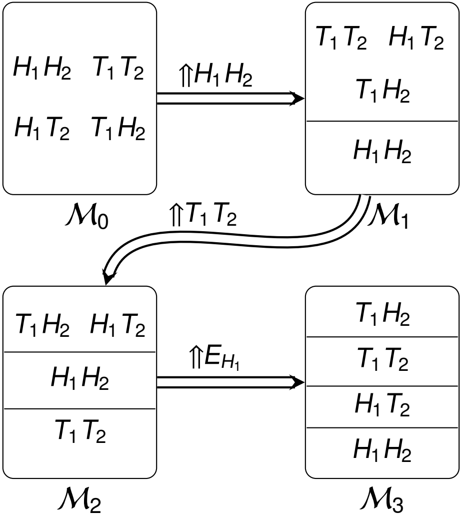
Heuristic Diagnosis of Stalnaker's Example



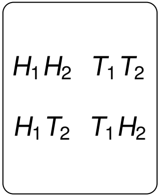
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

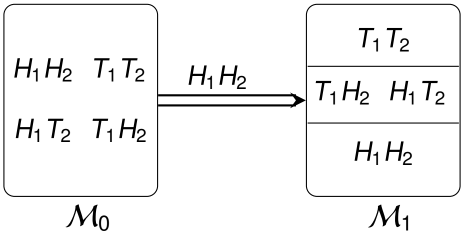


Heuristic Diagnosis of Stalnaker's Example

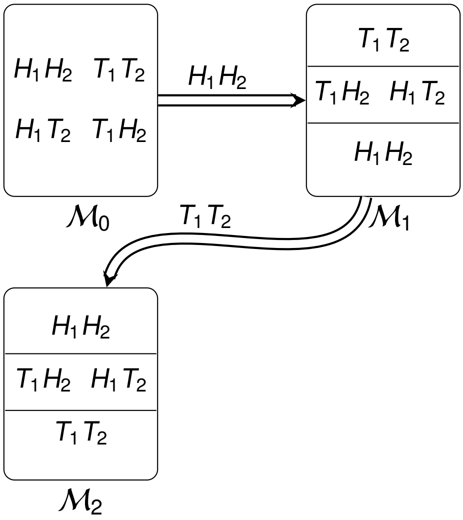


$\mathcal{M}_0$

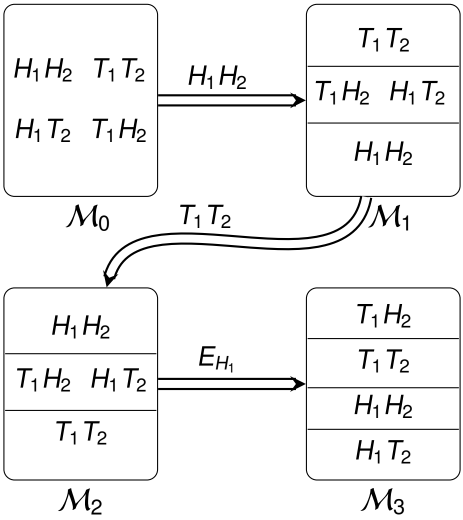
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



A key aspect of any formal model of a (social) interactive situation or situation of rational inquiry is the way it accounts for the

> *...information about how I learn some of the things I learn, about the sources of my information, or about what I believe about what I believe and don't believe. If the story we tell in an example makes certain information about any of these things relevant, then it needs to be included in a proper model of the story, if it is to play the right role in the evaluation of the abstract principles of the model.* (Stalnaker, pg. 203)

R. Stalnaker. *Iterated Belief Revision*. Erkentnis 70, pgs. 189 - 209, 2009.

Discussion, I

A proper conceptualization of the event and report structure is crucial (the event space must be ‘rich enough’): A theory must be able to accommodate the conceptualization, but other than that it hardly counts in favor of a theory that the modeler gets this conceptualization right.

Discussion, II

There seems to be a trade-off between a rich set of states and event structure, and a rich theory of ‘doxastic actions’.

How should we resolve this trade-off when analyzing counterexamples to postulates of belief changes over time?

meta-information: information about how “trusted” or “reliable” the sources of the information are.

meta-information: information about how “trusted” or “reliable” the sources of the information are.

This is particularly important when analyzing how an agent’s beliefs change over an extended period of time. For example, rather than taking a stream of contradictory incoming evidence (i.e., the agent receives the information that p, then the information that q, then the information that $\neg p$, then the information that $\neg q$) at face value (and performing the suggested belief revisions), a rational agent may consider the stream itself as evidence that the source is not reliable

procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

A *protocol* describes what the agents “can” or “cannot” do (say, observe) in a social interactive situation or rational inquiry.

meta-information: information about how "trusted" or "reliable" the sources of the information are.

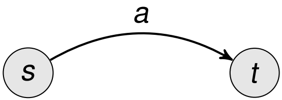
procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

Actions

1. Actions as *transitions between states, or situations*:

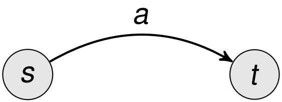
Actions

1. Actions as *transitions between states, or situations*:

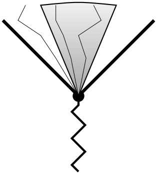


Actions

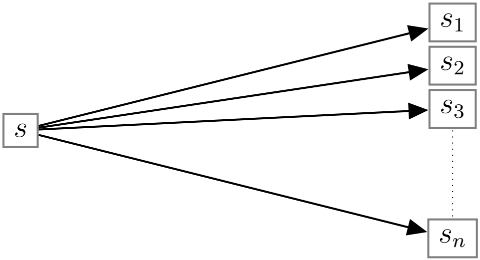
1. Actions as *transitions between states, or situations*:



2. Actions *restrict* the set of possible future histories.



J. van Benthem, H. van Ditmarsch, J. van Eijck and J. Jaspers. *Chapter 6: Propositional Dynamic Logic*. Logic in Action Online Course Project, 2011.


s
s_1
s_2
s_3
s_n

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{P}\}, V\rangle$ where for each $a \in \mathsf{P}$, $R_a \subseteq W \times W$ and $V : \mathsf{At} \rightarrow \wp(W)$

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathrm{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathrm{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathrm{P}\}, V\rangle$ where for each $a \in \mathrm{P}$, $R_a \subseteq W \times W$ and $V : \mathrm{At} \rightarrow \wp(W)$

$[\alpha]\varphi$ means "after doing α, φ will be true"

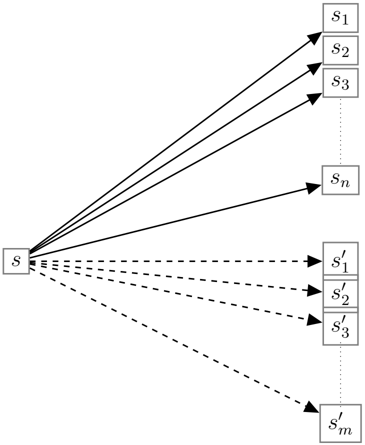
$\langle\alpha\rangle\varphi$ means "after doing α, φ may be true"

$\mathcal{M}, w \models [\alpha]\varphi$ iff for each v, if $wR_\alpha v$ then $\mathcal{M}, v \models \varphi$

$\mathcal{M}, w \models \langle\alpha\rangle\varphi$ iff there is a v such that $wR_\alpha v$ and $\mathcal{M}, v \models \varphi$

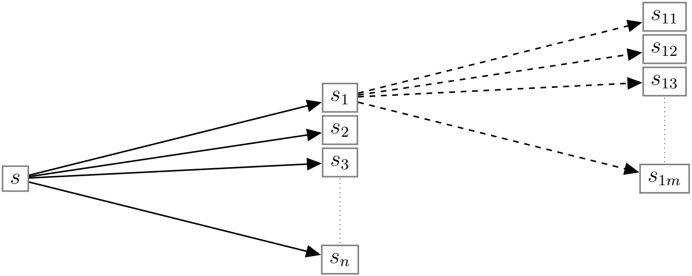
Union

$$R_{\alpha \cup \beta} := R_\alpha \cup R_\beta$$



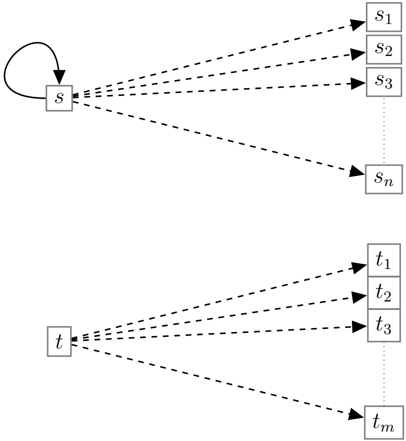
Sequence

$$R_{\alpha;\beta} := R_\alpha \circ R_\beta$$



Test

$$R_{\varphi?} = \{(w, w) \mid \mathcal{M}, w \models \varphi\}$$



Iteration

$$R_{\alpha^*} := \cup_{n \geq 0} R_\alpha^n$$

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha;\beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$
8. Modus Ponens and Necessitation (for each program α)

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha;\beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$ (Induction Axiom)
8. Modus Ponens and Necessitation (for each program α)

Actions and Ability

An early approach to interpret PDL as logic of actions was put forward by Krister Segerberg.

Segerberg adds an "agency" program to the PDL language δA where A is a formula.

K. Segerberg. *Bringing it about*. JPL, 1989.

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

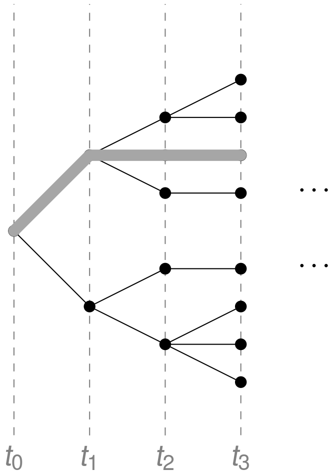
3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

1. $[\delta A]A$
2. $[\delta A]B \rightarrow ([\delta B]C \rightarrow [\delta A]C)$

Actions and Agency in Branching Time

Alternative accounts of agency do not include explicit description of the actions:



STIT

- Each node represents a choice point for the agent.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)
- The key modality is $[i\ stit]\varphi$ which is intended to mean that the agent i can "see to it that φ is true".
    - $[i\ stit]\varphi$ is true at a history moment pair provided the agent can choose a (set of) branch(es) such that every future history-moment pair satisfies φ

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[i\ \mathit{stit}]\varphi$: "the agent has the ability to bring about φ".

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \rightarrow \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \rightarrow \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

- $\langle T, <\rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let $Hist$ be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \rightarrow \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \rightarrow \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i\in\mathcal{A}} s_t(i) \neq \emptyset$
- $V : At \rightarrow \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

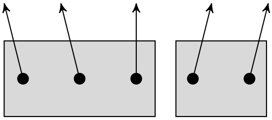
STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$
- $V : At \to \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

Many Agents

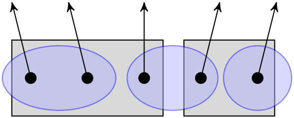
The previous model assumes there is *one* agent that “controls” the transition system.



Many Agents

The previous model assumes there is *one* agent that “controls” the transition system.

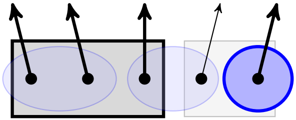
What if there is more than one agent?



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

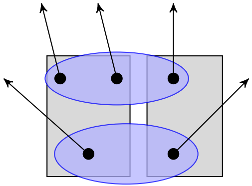
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

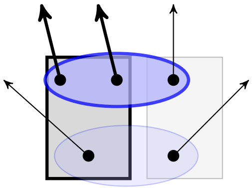
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent? *Independence of agents*



STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$

STIT Language

$$\varphi \ = \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$

STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$

STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i\ stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$

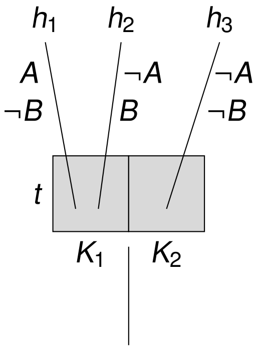
STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i\ stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$
- $\mathcal{M}, t/h \models [i\ dstit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$ and there is a $h'' \in H_t$ such that $\mathcal{M}, t/h \models \neg\varphi$

STIT: Example

The following are false: $A \rightarrow \Diamond[stit]A$ and $\Diamond[stit](A \vee B) \rightarrow \Diamond[stit]A \vee \Diamond[stit]B$.



J. Horty. *Agency and Deontic Logic*. 2001.

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ \mathit{stit}]$: $[i\ \mathit{stit}](\varphi \to \psi) \to ([i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \to \varphi$, $[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for [*i stit*]: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ stit]$: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$
- $(\bigwedge_{i\in\mathcal{A}} \Diamond[i\ stit]\varphi_i) \to \Diamond(\bigwedge_{i\in\mathcal{A}}[i\ stit]\varphi_i)$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ \mathit{stit}]$: $[i\ \mathit{stit}](\varphi \to \psi) \to ([i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \to \varphi$, $[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$
- $\Box\varphi \to [i\ \mathit{stit}]\varphi$
- $(\bigwedge_{i\in\mathcal{A}} \Diamond[i\ \mathit{stit}]\varphi_i) \to \Diamond(\bigwedge_{i\in\mathcal{A}}[i\ \mathit{stit}]\varphi_i)$
- Modus Ponens and Necessitation for $\Box$

M. Xu. *Axioms for deliberative STIT*. Journal of Philosophical Logic, Volume 27, pp. 505 - 552, 1998.

P. Balbiani, A. Herzig and N. Troquard. *Alternative axiomatics and complexity of deliberative STIT theories*. Journal of Philosophical Logic, 37:4, pp. 387 - 406, 2008.

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ stit]\varphi$: the agent can "see to it that" φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ \mathit{stit}]\varphi$: the agent can "see to it that" φ is true
- $\Diamond[i\ \mathit{stit}]\varphi$: the agent has the ability to bring about φ

Epistemizing logics of action and ability

Knowledge, action, abilities

A. Herzig. *Logics of knowledge and action: critical analysis and challenges*. Autonomous Agent and Multi-Agent Systems, 2014.

J. Broeresen, A. Herzig and N. Troquard. *What groups do, can do and know they can do: An analysis in normal modal logics*. Journal of Applied and Non-Classical Logics, 19:3, pgs. 261 - 289, 2009.

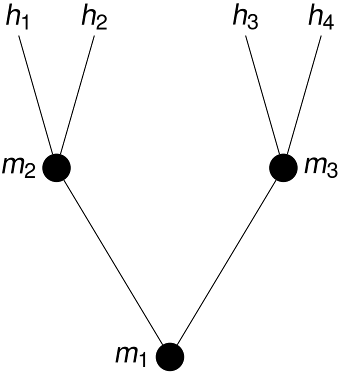
W. van der Hoek and M. Wooldridge. *Cooperation, knowledge and time: Alternating-time temporal epistemic logic and its applications*. Studia Logica, 75, pgs. 125 - 157, 2003.

Epistemic stit model

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, V \rangle$$

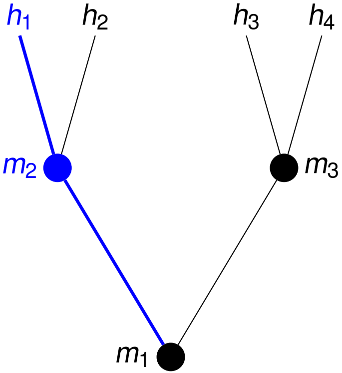
Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



Epistemic stit model

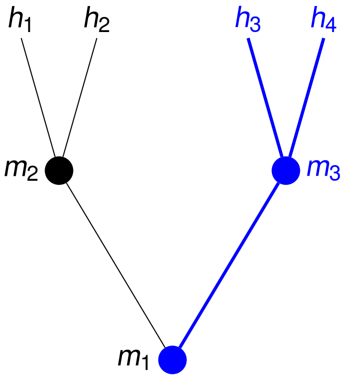
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V \rangle$$



m/h denotes (m, h) with $m \in h$ is called an **index**

Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$

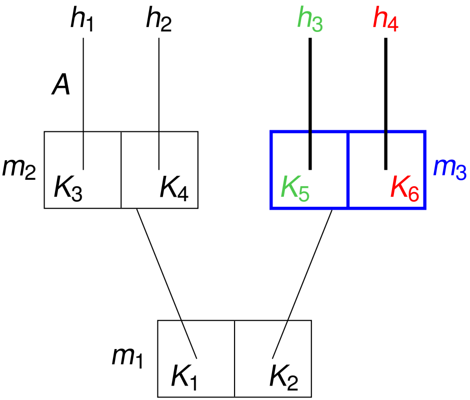


m/h denotes (m, h) with $m \in h$ is called an **index**

$H^m = \{h \mid m \in h\}$

Epistemic stit model

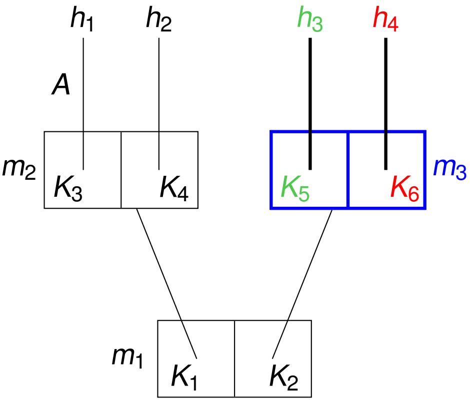
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



For $\alpha \in Agent$, $Choice_\alpha^m$ is a partition on H^m

Epistemic stit model

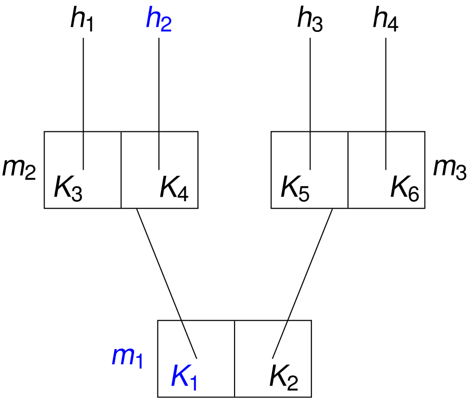
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V \rangle$$



For $\alpha \in Agent$, $Choice^m_\alpha$ is a partition on H^m

Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V \rangle$$

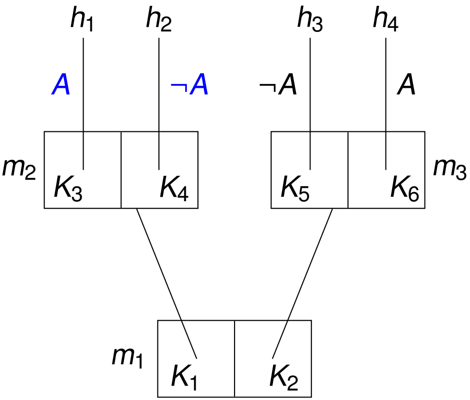


For $\alpha \in Agent$, $Choice^m_\alpha$ is a partition on H^m

$Choice^m_\alpha(h)$ is the particular action at m that contains h

Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V \rangle$$

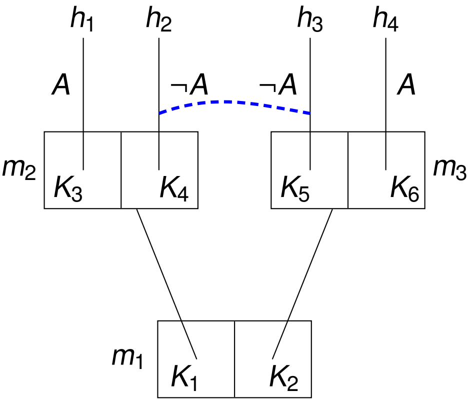


V assigns sets of indices to atomic propositions.

$m_2/h_1 \models A$ $\qquad$ $m_2/h_2 \not\models A$

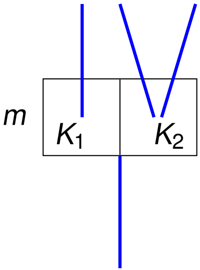
Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$

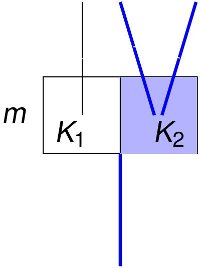


$\sim_\alpha$ is an (equivalence) relation on indices

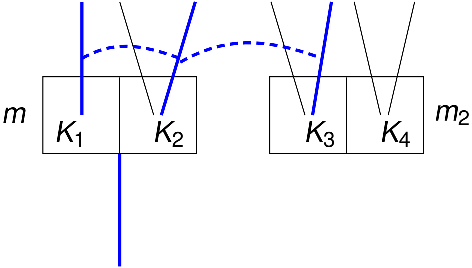
$m/h \sim_\alpha m'/h'$: everything α knows at m/h is true at m'/h', α cannot distinguish m/h and m'/h', ...



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$ for all $h' \in H^m$,



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$,
- $\mathcal{M}, m/h \models [\alpha\ stit\colon A]$ if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$,



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$,
- $\mathcal{M}, m/h \models [\alpha\ stit\colon A]$ if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$,
- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$, then $\mathcal{M}, m'/h' \models A$

Action labels

Let *Type* $= \{\tau_1, \tau_2, \ldots, \tau_n\}$ be a set of action types—general kinds of action, as opposed to the concrete action tokens.

An action type τ is interpreted as a partial function mapping each agent α and moment m into the particular action token $[\tau]_\alpha^m$ that results when τ is executed by α at m (so, $[\tau]_\alpha^m \in Choice_\alpha^m$)

Labeled stit frames

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, \mathit{Type}, \mathit{Label}, V\rangle,$$

Label maps each action token $K \in \mathit{Choice}^m_\alpha$ to a particular action type $\mathit{Label}(K) \in \mathit{Type}$.

1. If $K \in \mathit{Choice}^m_\alpha$, then $[\mathit{Label}(K)]^m_\alpha = K$,
2. If $\tau \in \mathit{Type}$ and $[\tau]^m_\alpha$ is defined, then $\mathit{Label}([\tau]^m_\alpha) = \tau$.

Labeled stit frames

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, \mathit{Type}, \mathit{Label}, V \rangle,$$

Label maps each action token $K \in \mathit{Choice}^m_\alpha$ to a particular action type $\mathit{Label}(K) \in \mathit{Type}$.

1. If $K \in \mathit{Choice}^m_\alpha$, then $[\mathit{Label}(K)]^m_\alpha = K$,
2. If $\tau \in \mathit{Type}$ and $[\tau]^m_\alpha$ is defined, then $\mathit{Label}([\tau]^m_\alpha) = \tau$.

$$\mathit{Type}^m_\alpha = \{\mathit{Label}(K) : K \in \mathit{Choice}^m_\alpha\}$$

$$\mathit{Type}^m_\alpha(h) = \mathit{Label}(\mathit{Choice}^m_\alpha(h))$$

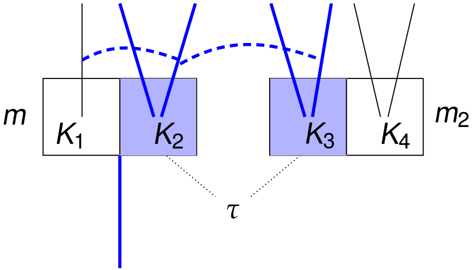
Frame properties

- If $m/h \sim_\alpha m'/h'$, then $m/h'' \sim_\alpha m'/h'''$ for each $h'' \in H^m$ and $h''' \in H^{m'}$.
- For all m/h, $Know_\alpha(m/h) \subseteq H^m$.
- If $m/h \sim_\alpha m'/h'$, then $Type_\alpha^m = Type_\alpha^{m'}$.
- If $m/h \sim_\alpha m'/h'$, then $Type_\alpha^m(h) = Type_\alpha^{m'}(h')$.

kstit

- $\mathcal{M}, m/h \models [\alpha\ \textit{kstit}\colon A]$ if and only if $[\textit{Type}^m_\alpha(h)]^{m'}_\alpha \subseteq |A|^{m'}_{\mathcal{M}}$ for all m'/h' such that $m'/h' \sim_\alpha m/h$.

kstit

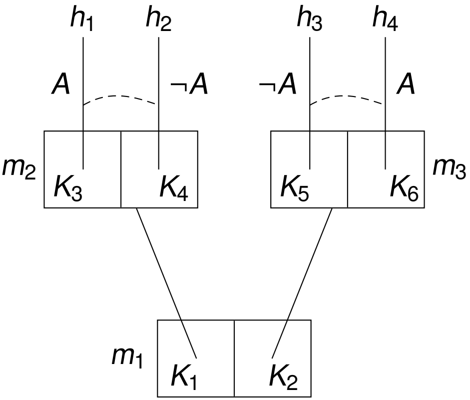


- $\mathcal{M}, m/h \models [\alpha\ \mathit{kstit}: A]$ if and only if $[Type_\alpha^m(h)]_\alpha^{m'} \subseteq |A|_{\mathcal{M}}^{m'}$ for all m'/h' such that $m'/h' \sim_\alpha m/h$.

Causal vs. epistemic ability

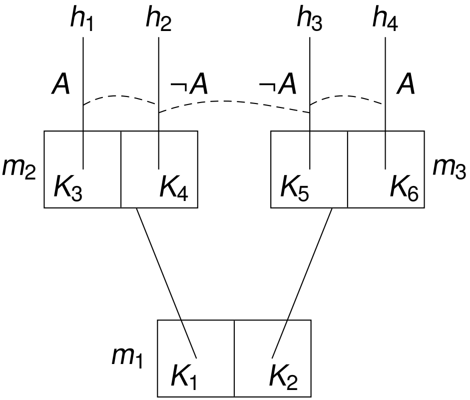
$$\Diamond[\alpha\ \textit{stit}\colon A]$$

Causal vs. epistemic ability



$\Diamond[\alpha$ *stit*: $A]$ is settled true at m_2

Causal vs. epistemic ability



$\Diamond[\alpha$ *stit*: $A]$ is settled true at m_2

Causal vs. epistemic ability

$$\Diamond[\alpha\ \mathit{stit}\colon A]$$

$$\mathsf{K}_\alpha\Diamond[\alpha\ \mathit{stit}\colon A]$$

$$\Diamond\mathsf{K}_\alpha[\alpha\ \mathit{stit}\colon A]$$

$$\Diamond[\alpha\ \mathit{kstit}\colon A]$$

Ex ante vs. *ex interim* knowledge

- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$, then $\mathcal{M}, m'/h' \models A$
- $\mathcal{M}, m/h \models \mathsf{K}^{\mathrm{act}}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$ and $h' \in [\mathit{Type}^m_\alpha(h)]^{m'}_\alpha$, $\mathcal{M}, m'/h' \models A$

Discussion

- Language/validities

$$\Box A \supset [\alpha \textit{ stit}: A]$$
$$\mathsf{K}_\alpha \Box A \supset [\alpha \textit{ kstit}: A]$$
$$[\alpha \textit{ kstit}: A] \equiv \mathsf{K}_\alpha^{\mathsf{act}}[\alpha \textit{ stit}: A]$$
$$\dots$$

- What do the agents know vs. What do the agents know *given what they are doing*.

- Equivalence between labeled stit models (cf. Thompson transformations specifying when two imperfect information games reduce to the same Normal form)