

知识更新的两类模型

王轶

ynw@zju.edu.cn

复旦大学光华楼西主楼 2401 室

2016 年 06 月 08 日

(一)

背景知识

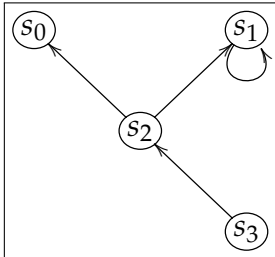


经典模态逻辑 (1)

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box\varphi$

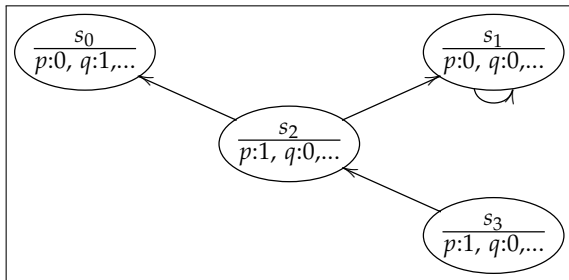
经典模态逻辑 (1)

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box\varphi$
- 框架: 有向图
 - 每个点代表一个可能世界
 - 箭头代表可能世界之间的可通达关系



经典模态逻辑 (2)

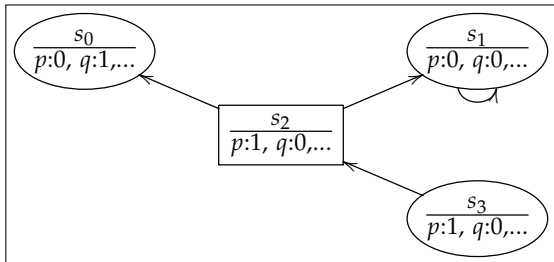
- 模型：在框架的基础上为所有命题变元指定所有可能世界上的真值





经典模态逻辑 (3)

指定真实世界的可能世界模型



- 以方框表示真实世界/当前世界，上例中为 s_2
- 一个世界通过单个箭头所连接的世界称为该结点**可通达**的世界
 - s_0 没有可通达的世界
 - s_2 可通达 s_0 和 s_1
 -



经典模态逻辑 (4): 可能世界语义学

给定模型, 以 “ $s \models \varphi$ ” 表示 “ φ 在 s 上真”

命题变元在 s 上的真值由模型给定

$$s \models \neg \varphi \quad \Leftrightarrow \quad \text{并非 } s \models \varphi$$

$$s \models \varphi \rightarrow \psi \quad \Leftrightarrow \quad \text{如果 } s \models \varphi \text{ 则 } s \models \psi$$

$$s \models \Box \varphi \quad \Leftrightarrow \quad \text{所有 } s \text{ 通达的世界 } t: t \models \varphi$$

$$s \models \Diamond \varphi \quad \Leftrightarrow \quad \text{存在 } s \text{ 通达的世界 } t: t \models \varphi$$

经典模态逻辑 (4): 可能世界语义学

给定模型, 以 “ $s \models \varphi$ ” 表示 “ φ 在 s 上真”

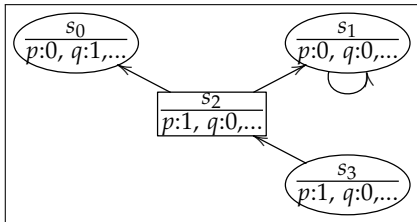
命题变元在 s 上的真值由模型给定

$$s \models \neg \varphi \quad \Leftrightarrow \quad \text{并非 } s \models \varphi$$

$$s \models \varphi \rightarrow \psi \quad \Leftrightarrow \quad \text{如果 } s \models \varphi \text{ 则 } s \models \psi$$

$$s \models \Box \varphi \quad \Leftrightarrow \quad \text{所有 } s \text{ 通达的世界 } t: t \models \varphi$$

$$s \models \Diamond \varphi \quad \Leftrightarrow \quad \text{存在 } s \text{ 通达的世界 } t: t \models \varphi$$



$$s_0 \models \Box p$$

$$s_2 \models \Diamond \Box p$$

$$s_0 \models p \rightarrow q$$

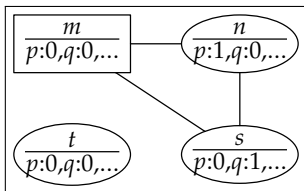
$$s_1 \models p \rightarrow q$$

$$s_2 \models \Box(p \rightarrow q)$$

知识模型与经典认知逻辑

□ 被理解为“知道”

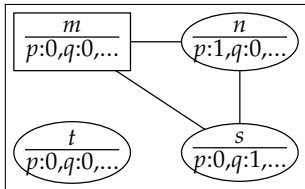
在知识模型中，箭头都是等价关系并以线段表示，省略自反箭头
线段连接的阶段可理解为“不可区分的可能世界”



知识模型与经典认知逻辑

□ 被理解为 “知道”

在知识模型中，箭头都是等价关系并以线段表示，省略自反箭头
 线段连接的阶段可理解为 “不可区分的可能世界”



(T) $\Box\phi \rightarrow \phi$

“所知皆真”

自反性

(4) $\Box\phi \rightarrow \Box\Box\phi$

“正自省”

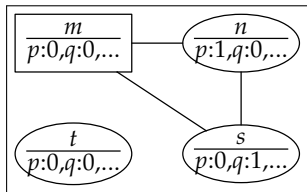
传递性

(5) $\neg\Box\phi \rightarrow \Box\neg\Box\phi$

“反自省”

欧性

视野缩小意味着知识增加



$$m \not\equiv \Box p$$

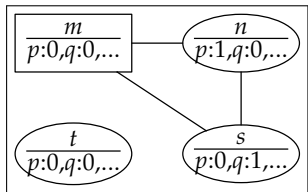
$$m \not\equiv \Box q$$

$$m \not\equiv \Box(p \wedge q)$$

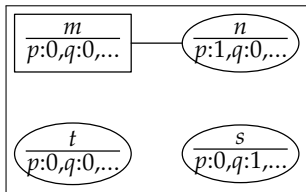
$$m \not\equiv \Box\Box p$$

...

视野缩小意味着知识增加



\Rightarrow



$$m \not\models \Box p$$

$$m \not\models \Box q$$

$$m \not\models \Box(p \wedge q)$$

$$m \not\models \Box \Box p$$

...

$$m \not\models \Box p$$

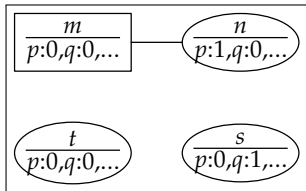
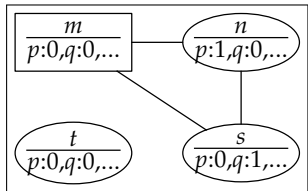
$$m \models \Box q$$

$$m \models \Box(p \wedge q)$$

$$m \not\models \Box \Box p$$

...

视野缩小意味着知识增加



$$m \not\models \Box p$$

$$m \not\models \Box q$$

$$m \not\models \Box(p \wedge q)$$

$$m \not\models \Box\Box p$$

...

$$m \not\models \Box p$$

$$m \models \Box q$$

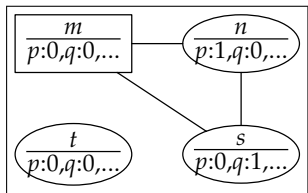
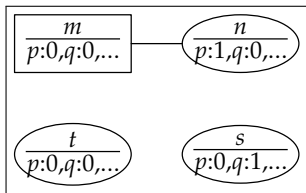
$$m \models \Box(p \wedge q)$$

$$m \not\models \Box\Box p$$

...

注意：“无知”则减少！

视野缩小意味着知识增加


 \Rightarrow


$$m \not\models \Box p$$

$$m \not\models \Box q$$

$$m \not\models \Box(p \wedge q)$$

$$m \not\models \Box \Box p$$

$$m \models \neg \Box q$$

$$m \models \Box \neg \Box q$$

$$m \not\models \Box p$$

$$m \models \Box q$$

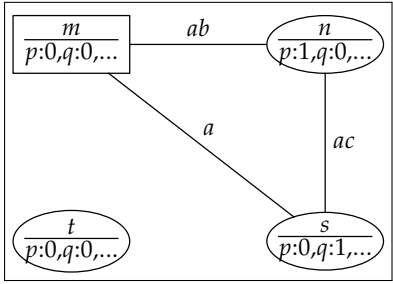
$$m \models \Box(p \wedge q)$$

$$m \not\models \Box \Box p$$

$$m \models \neg \Box \neg \Box q$$

多主体认知逻辑

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \square_i\varphi$
- 模型: 每个主体对应一个等价关系 (在图中以加标的线段表示)



(二)

可能世界模型 下的知识更新

华罗庚帽子



华罗庚, 1963, 《数学归纳法》, 上海教育出版社, 第 16 页

有一位老师, 想辨别出他的三个得意门生中哪一个更聪明一些, 他采用了以下的方法。事先准备好 5 顶, 其中 3 顶是白的, 2 顶是黑的。在试验时, 他先把这些帽子让学生们看了一看, 然后要他们闭上眼睛, 替每个学生戴上一顶白色的帽子, 并且把 2 顶黑帽子藏了起来, 最后再让他们睁开眼睛, 请他们说出自己头上戴的帽子, 究竟是哪一种颜色。

三个学生相互看了一看, 踌躇了一会儿, 然后他们异口同声地说, 自己头上戴的是白色的帽子。



华罗庚帽子

华罗庚，1963，《数学归纳法》，上海教育出版社，第16页

有一位老师，想辨别出他的三个得意门生中哪一个更聪明一些，他采用了以下的方法。事先准备好5顶，其中3顶是白的，2顶是黑的。在试验时，他先把这些帽子让学生们看了一看，然后要他们闭上眼睛，替每个学生戴上一顶白色的帽子，并且把2顶黑帽子藏了起来，最后再让他们睁开眼睛，请他们说出自己头上戴的帽子，究竟是哪一种颜色。

三个学生相互看了一看，踌躇了一会儿，然后他们异口同声地说，自己头上戴的是白色的帽子。

简单情形：2位同学，2白1黑



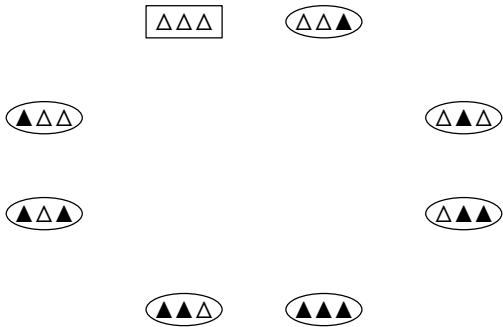
模型与更新

三位同学的情形已经比较复杂了

使用形式化方法可以检验有时不那么靠谱的
直观 😊

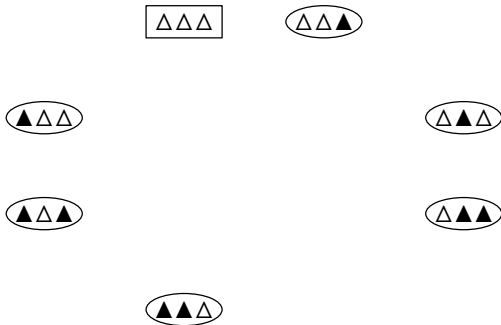
模型与更新

三位同学：a、b、c；以 $\triangle \blacktriangle \triangle$ 表示 $\begin{matrix} \triangle \blacktriangle \triangle \\ a \quad b \quad c \end{matrix}$



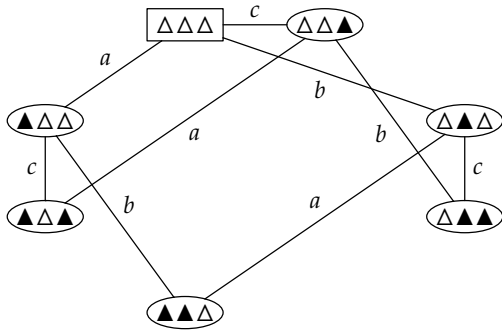
模型与更新

只有两顶黑帽子



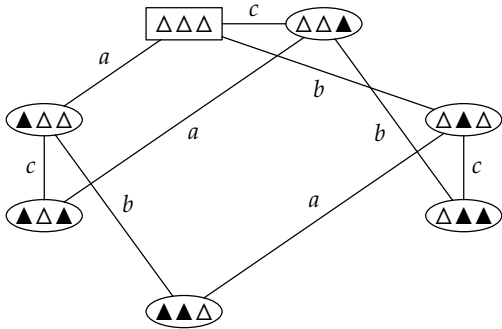
模型与更新

每位同学都看不见自己头上帽子的颜色



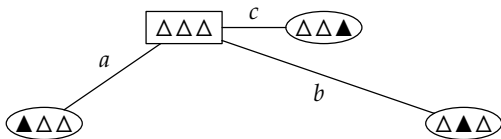
模型与更新

踌躇第 1 回合开始——没人知道答案：



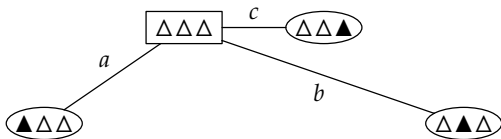
模型与更新

踌躇第 1 回合结束：



模型与更新

踌躇第 2 回合开始——没人知道答案：





模型与更新

踌躇第 2 回合结束：



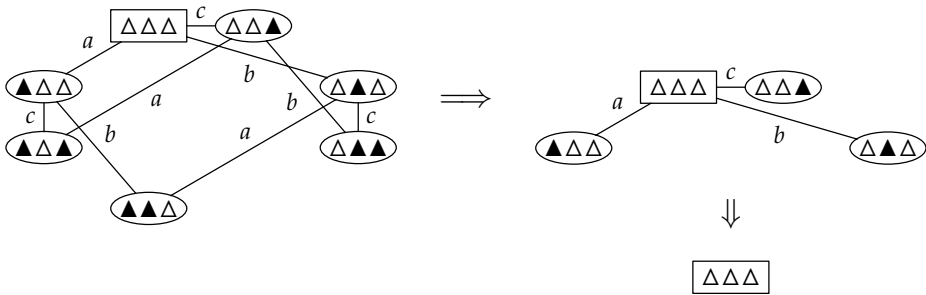


模型与更新

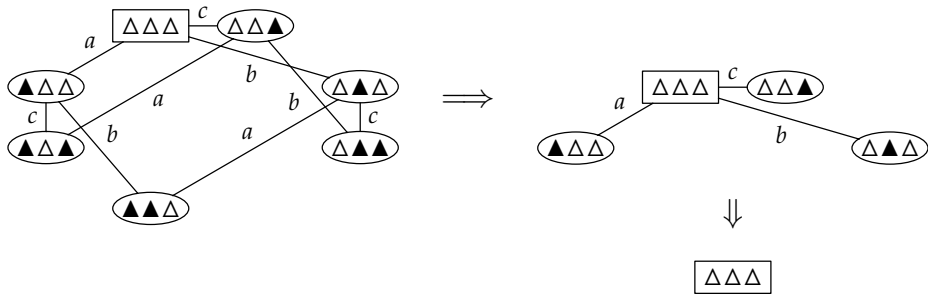
得到答案“都是白帽子”，回答之后模型无变化。



模型更新情况



模型更新情况



- 将信息交流视为模型更新算子（函数）
 - 不满足条件的可能世界被删除
 - 更新过程中应保留真实世界



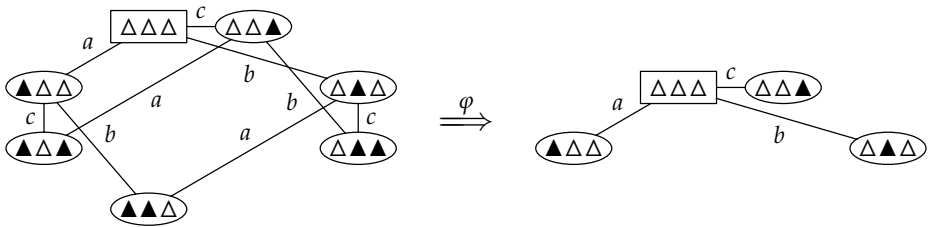
公开宣告逻辑 (Plaza, 1989)

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box_i\varphi \mid [\varphi]\varphi$
- $[\varphi]\psi$: “ φ 被公开宣告后 ψ 成立”

公开宣告逻辑 (Plaza, 1989)

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box_i\varphi \mid [\varphi]\varphi$
- $[\varphi]\psi$: “ φ 被公开宣告后 ψ 成立”

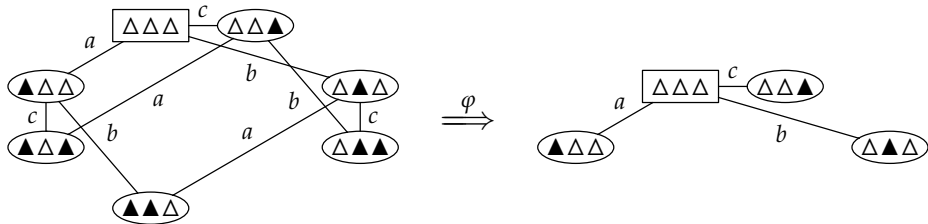
前例中以 φ 表示: “三个人都不知道自己头上帽子的颜色”



公开宣告逻辑 (Plaza, 1989)

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box_i\varphi \mid [\varphi]\varphi$
- $[\varphi]\psi$: “ φ 被公开宣告后 ψ 成立”

前例中以 φ 表示: “三个人都不知道自己头上帽子的颜色”



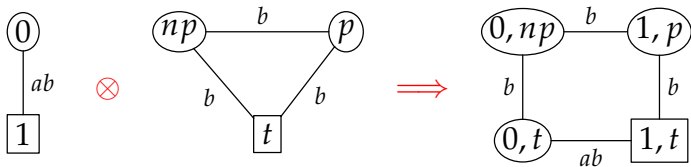
类似的例子 “泥孩难题 (muddy children puzzle)” 可以更好地呈现上述更新过程



模型的更新是否一定变小?

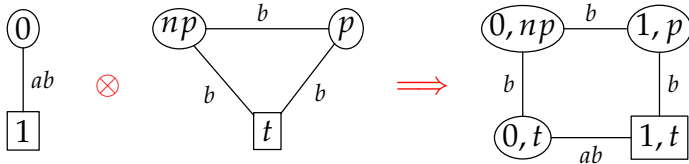
模型的更新是否一定变小?

- 行动模型逻辑 (Baltag et al., 1998)

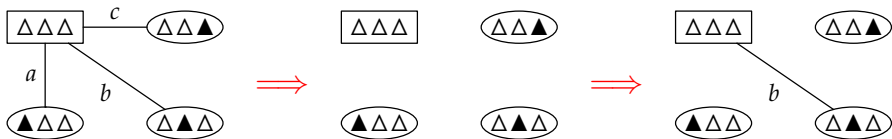


模型的更新是否一定变小?

- 行动模型逻辑 (Baltag et al., 1998)



- 遗忘



(三)

与群体有关的更新



群体

群体是否“有知识”？



群体

群体是否“有知识”？

- 普遍知识: $E_{\{1\dots n\}}\varphi \leftrightarrow (K_1\varphi \wedge \dots \wedge K_n\varphi)$
- 公共知识: $C_G\varphi \leftrightarrow (E_G\varphi \wedge E_GE_G\varphi \wedge \dots)$



群体

群体是否“有知识”？

- 普遍知识: $E_{\{1\dots n\}}\varphi \leftrightarrow (K_1\varphi \wedge \dots \wedge K_n\varphi)$
- 公共知识: $C_G\varphi \leftrightarrow (E_G\varphi \wedge E_GE_G\varphi \wedge \dots)$

如何获取公共知识？



群体

群体是否“有知识”？

- 普遍知识: $E_{\{1\dots n\}}\varphi \leftrightarrow (K_1\varphi \wedge \dots \wedge K_n\varphi)$
- 公共知识: $C_G\varphi \leftrightarrow (E_G\varphi \wedge E_GE_G\varphi \wedge \dots)$
如何获取公共知识?
- 分布式知识: $D_G\varphi$



群体

群体是否“有知识”？

- 普遍知识: $E_{\{1\dots n\}}\varphi \leftrightarrow (K_1\varphi \wedge \dots \wedge K_n\varphi)$
- 公共知识: $C_G\varphi \leftrightarrow (E_G\varphi \wedge E_GE_G\varphi \wedge \dots)$

如何获取公共知识？

- 分布式知识: $D_G\varphi \not\leftrightarrow \bigvee_{i \in G} K_i\varphi$
 - 群体内成员采用任何可能的交流途径所能获取的知识
 - “A group has distributed knowledge of a fact φ if the knowledge of φ is distributed among its members, so that by pooling their knowledge together the members of the group can deduce φ , even though it may be the case that no member of the group individually knows φ .” (Fagin et al., 1995)

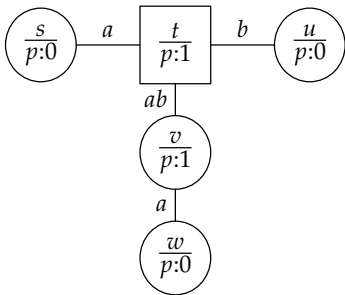


分布式知识的定义所面临的问题

语义定义： $m \models D_{ab}\varphi$ 当且仅当对所有由 m 经 ab 线段可通达的世界 m' ，有 $m' \models \varphi$

分布式知识的定义所面临的问题

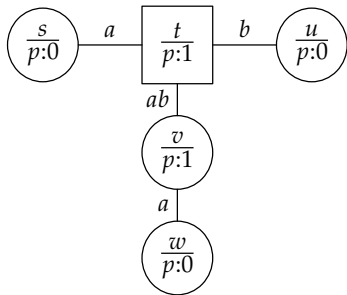
语义定义： $m \models D_{ab}\varphi$ 当且仅当对所有由 m 经 ab 线段可通达的世界 m' ，有 $m' \models \varphi$



$$t \models D_{ab}(p \wedge \neg K_a p)$$

分布式知识的定义所面临的问题

语义定义： $m \models D_{ab}\varphi$ 当且仅当对所有由 m 经 ab 线段可通达的世界 m' ，有 $m' \models \varphi$



$$t \models D_{ab}(p \wedge \neg K_a p)$$

既然 a 和 b 的分布式知识是经过充分讨论后得到的知识，那就不可能出现 $p \wedge \neg K_a p$ 是分布式知识的情况。**因此定义有误!**



知识聚集

- 分布式知识算子 D_G 刻画的不是“群体知识聚集”；实际使用中的分布式知识的真实定义如下：

“A group has distributed knowledge of a fact φ if the knowledge of φ is distributed among its members, so that by pooling their knowledge together the members of the group ~~can deduce φ ...~~ know that φ was true before they did that.”



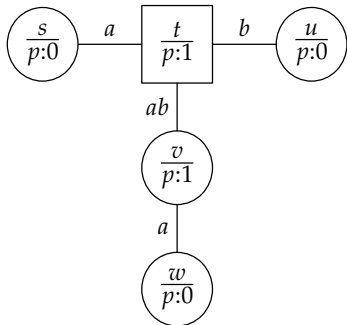
知识聚集

- 分布式知识算子 D_G 刻画的不是“群体知识聚集”；实际使用中的分布式知识的真实定义如下：

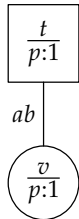
“A group has distributed knowledge of a fact φ if the knowledge of φ is distributed among its members, so that by pooling their knowledge together the members of the group ~~can deduce φ ...~~ know that φ was true before they did that.”

- 将满足先前定义的“群体知识聚集”用算子 R 进行刻画。

知识聚集模型更新

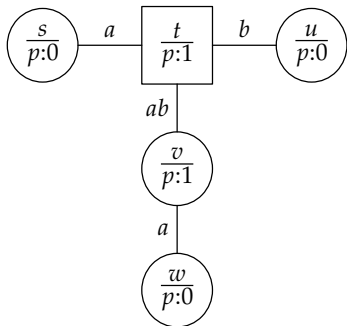


Resolving \Rightarrow

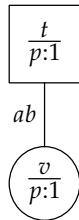


"communication core"
(van Benthem, 2011)

知识聚集模型更新



Resolving \Rightarrow



$$D_{ab}(p \wedge \neg K_a p)$$

$$D_{ab}(p \wedge \neg K_b p)$$

$$R_{ab}(p \wedge K_a p)$$

$$R_{ab}(p \wedge K_b p)$$



分布式知识如何转化为公共知识

- 对任意 $G \cap H = \emptyset$

$$R_G C_H \varphi \leftrightarrow C_H R_G \varphi$$



分布式知识如何转化为公共知识

- 对任意 $G \cap H = \emptyset$

$$R_G C_H \varphi \leftrightarrow C_H R_G \varphi$$

- 对任意 $G \subseteq H$:

$$R_G C_H \varphi \leftrightarrow D_G R_G \varphi$$

特别地:

$$R_G C_G \varphi \leftrightarrow D_G R_G \varphi$$

讨论之后 φ 是公共知识 当且仅当 讨论之后可得:

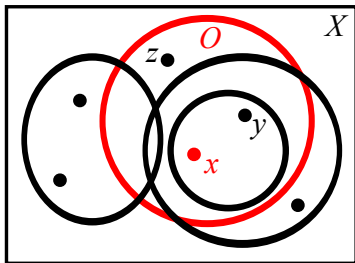
在讨论前有“讨论之后可得 φ ”

(四)

子集空间逻辑

子集空间逻辑

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \square\varphi \mid \boxplus\varphi$
- 子集空间:



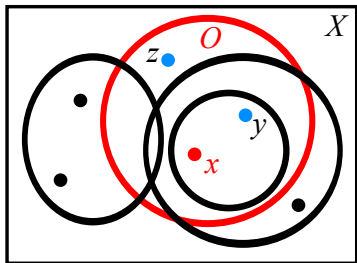
x : 可能世界

O : 认知域

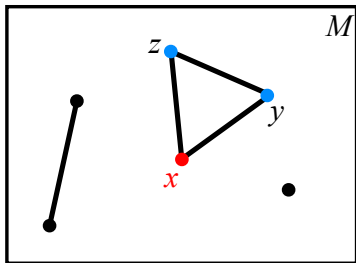
(x, O) : 认知情境

子集空间逻辑

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box\varphi \mid \boxplus\varphi$
- 子集空间:



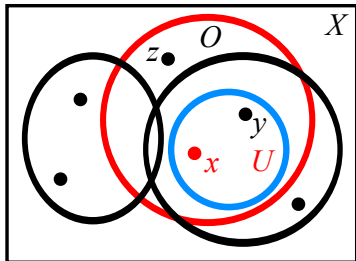
$(x, O) \models \Box\varphi$ 的条件?



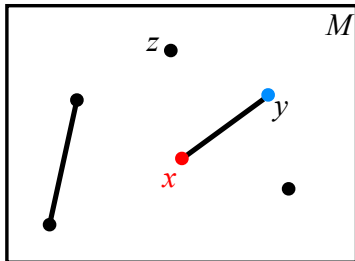
$x \models \Box\varphi$ 的条件?

子集空间逻辑

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box\varphi \mid \boxplus\varphi$
- 子集空间:



$(x, O) \models \boxplus\varphi$ 的条件?



$x \models \boxplus\varphi$ 的条件?



子集空间逻辑的动静原则

- 子集空间逻辑刻画知识更新的方法：
 - **观念上**：刻画“知识更新”这一动态性质
 - **形式上**：模型不变，当前认知情境改变
 - 语义中不考虑模型或框架改变的逻辑在形式上称为**静态逻辑**
- 动静原则：观念上动态，形式上静态



子集语义学下对公开宣告的两类刻画方式

- 语言: $\varphi ::= p \mid \neg\varphi \mid (\varphi \rightarrow \varphi) \mid \Box\varphi \mid \boxplus\varphi \mid [\varphi]\varphi$
- 语义 1 (Balbiani et al., 2013):

$$\mathfrak{M}, (x, O) \models [\varphi]\psi \quad \text{iff} \quad \mathfrak{M}, (x, O) \models \varphi \Rightarrow \mathfrak{M}|_{\varphi}, (x, O) \models \psi$$

不遵循动静原则

- 语义 2 (Wáng & Ågotnes, 2013):

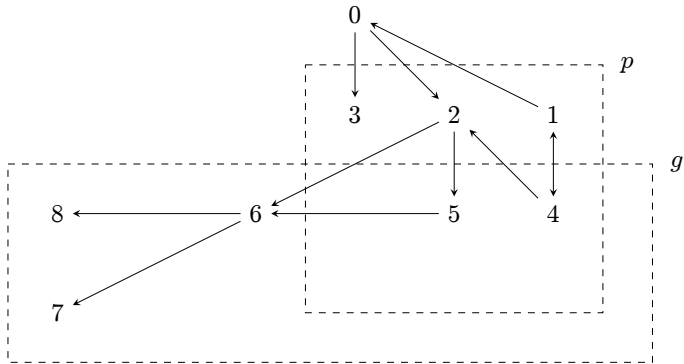
$$\mathfrak{M}, (x, O) \models [\varphi]\psi \quad \text{iff} \quad \mathfrak{M}, (x, O) \models \text{pre}(\varphi) \Rightarrow \mathfrak{M}, (x, O^\varphi) \models \psi$$

遵循动静原则



子集模型用于刻画偏好

(Wáng & Ågotnes, 2013b)



将认知域视为偏好对比的范围（焦点）



子集空间逻辑刻画知识更新的优势和劣势

- 劣势：
 - 直观呈现弱于关系语义学
 - 点集结构作为“状态”较复杂，相关模型论结论较少



子集空间逻辑刻画知识更新的优势和劣势

- 劣势：

- 直观呈现弱于关系语义学
- 点集结构作为“状态”较复杂，相关模型论结论较少

- 优势：

- 基于历史的模型：保留更新前的可能世界，易于刻画遗忘
- 与基于历史/路径的逻辑相比，模型较小，定义较简单



参考文献 I

- Ågotnes, T. and Wáng, Y. N. (2015) Resolving Distributed Knowledge. *Proceedings of the 15 th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*
- Balbiani, P.; van Ditmarsch, H. and Kudinov, A. (2013) Subset space logic with arbitrary announcements. *Proceedings of ICLA*
- van Benthem, J. F. A. K. (2011) *Logical Dynamics of Information and Interaction*. Cambridge University Press
- Fagin, R.; Halpern, J. Y.; Moses, Y. and Vardi, M. Y. (1995) *Reasoning about Knowledge* Cambridge, MA: The MIT Press



参考文献 II

- Plaza, J. A. (1989) Logics of Public Communications. *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (ISMIS '89)*, 201–216
- Wáng, Y. N. and Ågotnes, T. (2013) Subset Space Public Announcement Logic. *Proceedings of ICLA*, Springer, 2013, 7750, 245–257
- Wáng, Y. N. and Ågotnes, T. (2013b) Preference Logic of Focus Change. *Agreement Technologies*, Springer, 2013, LNAI 8068, 225–239