关于四类信念的逻辑研究

Logics of (In)sane and (Un)reliable beliefs

范杰
中科院哲学所
中国科学院大学人文学院
jiefan@ucas.ac.cn

2021 年全国数理逻辑年会

南开大学 2021 年 7 月 9 日-11 日

1. 四类信念的由来
2. 与文献中某些逻辑的联系：表达力对比
3. 这四类信念之间的关系：对当方阵
4. 信念的组合
5. 一种 unification 方法

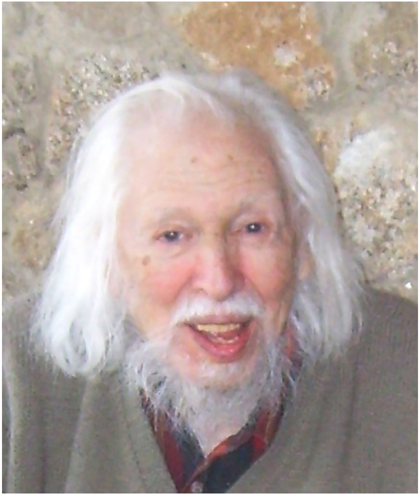
1 四类信念的由来

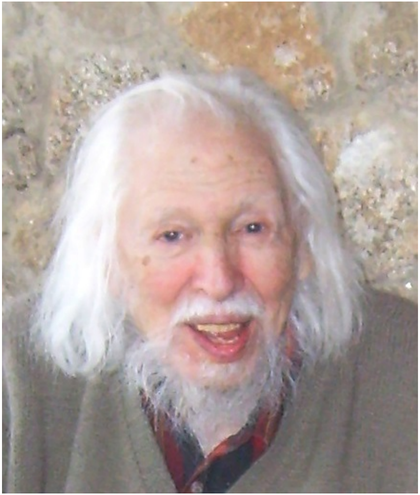
2 与文献中某些逻辑的联系：表达力对比

3 这四类信念之间的关系：对当方阵

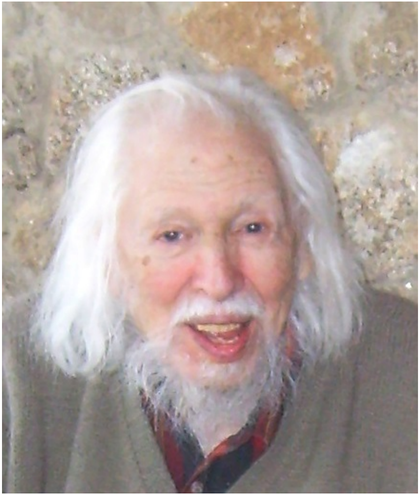
4 信念的组合

5 一种 unification 方法



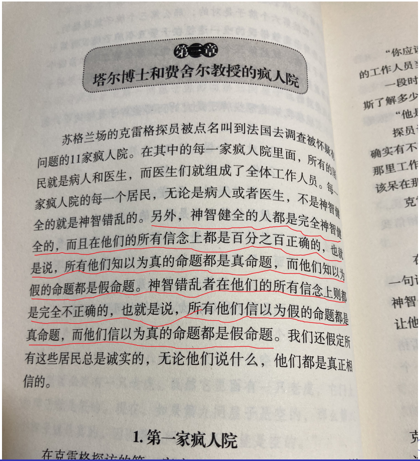


- Raymond M. Smullyan in 2008（雷蒙德·斯穆里安）
- 1919 年 5 月 25 日—2017 年 2 月 6 日
- 美国数学家、逻辑学家、哲学家、魔术师、钢琴家、道家
- What is the name of this book?
- The Lady or the Tiger?

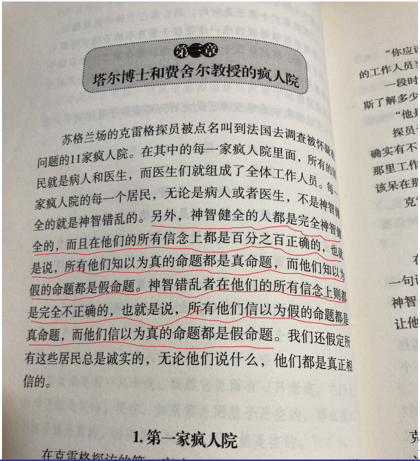


- Raymond M. Smullyan in 2008 （雷蒙德·斯穆里安）
- 1919 年 5 月 25 日—2017 年 2 月 6 日
- 美国数学家、逻辑学家、哲学家、魔术师、钢琴家、道家
- What is the name of this book?
- The Lady or the Tiger?

Introduction：从 Smullyan 开始谈起



Introduction：从 Smullyan 开始谈起



第二章

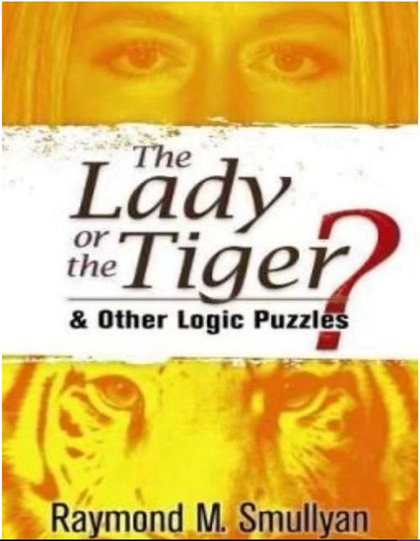
塔尔博士和费舍尔教授的疯人院

苏格兰场的克雷格探员被点名叫到法国去调查被怀疑有问题的11家疯人院。在其中的每一家疯人院里面，所有的居民就是病人和医生，而医生们就组成了全体工作人员。每一家疯人院的每一个居民，无论是病人或者医生，不是神智健全的就是神智错乱的。另外，神智健全的人都是完全神智健全的，而且在他们的所有信念上都是百分之百正确的，也就是说，所有他们知以为真的命题都是真命题，而他们知以为假的命题都是假命题。神智错乱者在他们的所有信念上则都是完全不正确的，也就是说，所有他们信以为假的命题都是真命题，而他们信以为真的命题都是假命题。我们还假定所有这些居民总是诚实的，无论他们说什么，他们都是真正相信的。

1. 第一家疯人院

“另外，神志健全的人都是完全神志健全的，而且在他们的所有信念上都是百分之百正确的，也就是说，所有他们知以为真的命题都是真命题，而他们知以为假的命题都是假命题。神志错乱者在他们的所有信念上则都是完全不正确的，也就是说，所有他们信以为假的命题都是真命题，而他们信以为真的命题都是假命题。”

“另外，神志健全的人都是完全神志健全的，而且在他们的所有信念上都是百分之百正确的，也就是说，所有他们知以为真的命题都是真命题，而他们知以为假的命题都是假命题。神志错乱者在他们的所有信念上则都是完全不正确的，也就是说，所有他们信以为假的命题都是真命题，而他们信以为真的命题都是假命题。”



◆ 3 ◆

The Asylum of Doctor Tarr and Professor Fether

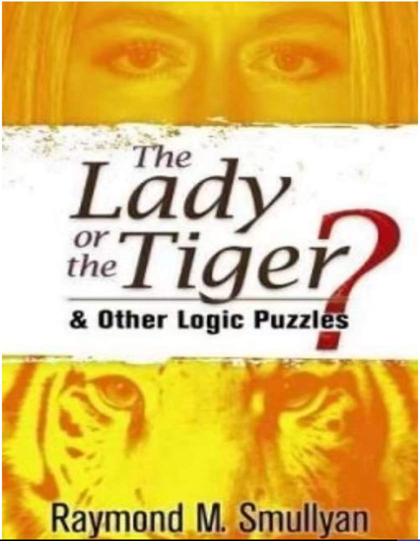
Inspector Craig of Scotland Yard was called over to France to investigate eleven insane asylums where it was suspected that something was wrong. In each of these asylums, the only inhabitants were patients and doctors—the doctors constituted the entire staff. Each inhabitant of each asylum, patient or doctor, was either sane or insane. Moreover, the sane ones were *totally* sane and a hundred percent accurate in all their beliefs; all true propositions they knew to be true and all false propositions they knew to be false. The insane ones were totally inaccurate in their beliefs; all true propositions they believed to be false and all false propositions they believed to be true. It is to be assumed also that all the inhabitants were always honest—whatever they said, they really believed.

1 ◆ The First Asylum

In the first asylum Craig visited, he spoke separately to two inhabitants whose last names were Jones and Smith.

"Tell me," Craig asked Jones, "what do you know about Mr. Smith?"

29



◆ 3 ◆

The Asylum of Doctor Tarr and Professor Fether

Inspector Craig of Scotland Yard was called over to France to investigate eleven insane asylums where it was suspected that something was wrong. In each of these asylums, the only inhabitants were patients and doctors—the doctors constituted the entire staff. Each inhabitant of each asylum, patient or doctor, was either sane or insane. Moreover, the sane ones were *totally* sane and a hundred percent accurate in all their beliefs; all true propositions they knew to be true and all false propositions they knew to be false. The insane ones were totally inaccurate in their beliefs; all true propositions they believed to be false and all false propositions they believed to be true. It is to be assumed also that all the inhabitants were always honest—whatever they said, they really believed.

1 ◆ The First Asylum

In the first asylum Craig visited, he spoke separately to two inhabitants whose last names were Jones and Smith.

"Tell me," Craig asked Jones, "what do you know about Mr. Smith?"

29

"... *Moreover, the sane ones were totally sane and a hundred percent accurate in all their beliefs; all true propositions they knew to be true and all false propositions they knew to be false. The insane ones were totally inaccurate in their beliefs; all true propositions they believed to be false and all false propositions they believed to be true.* ..." *(R. M. Smullyan,* The Lady or the Tiger?, *1982, p. 29)*

- all true propositions they knew to be true.
- all false propositions they knew to be false.
- all true propositions they believed to be false.
- all false propositions they believed to be true.

"· · · *Moreover, the sane ones were totally sane and a hundred percent accurate in all their beliefs; all true propositions they knew to be true and all false propositions they knew to be false. The insane ones were totally inaccurate in their beliefs; all true propositions they believed to be false and all false propositions they believed to be true. . . ." (R. M. Smullyan,* The Lady or the Tiger?, *1982, p. 29)*

- all true propositions they knew to be true.
- all false propositions they knew to be false.
- all true propositions they believed to be false.
- all false propositions they believed to be true.

1a. All true propositions they knew to be true.

1b. All true propositions they knew to be true were things they had learned at school.

2a. Every small dog, Harry likes.

2b. Every small dog Harry likes, he pats.

2a. Every small dog (is something that) Harry likes.

2b. Every small dog (that) Harry likes (is a thing) he pats.

1a. All true propositions (are something that) they knew to be true. 结合上下文：. . . believed . . . .

1b. All true propositions (that) they knew to be true were things they had learned at school.

1a. All true propositions they knew to be true.

1b. All true propositions they knew to be true were things they had learned at school.

2a. Every small dog, Harry likes.

2b. Every small dog Harry likes, he pats.

2a. Every small dog (is something that) Harry likes.

2b. Every small dog (that) Harry likes (is a thing) he pats.

1a. All true propositions (are something that) they knew to be true. 结合上下文：... believed ....

1b. All true propositions (that) they knew to be true were things they had learned at school.

1a. All true propositions they knew to be true.

1b. All true propositions they knew to be true were things they had learned at school.

2a. Every small dog, Harry likes.

2b. Every small dog Harry likes, he pats.

2a. Every small dog (is something that) Harry likes.

2b. Every small dog (that) Harry likes (is a thing) he pats.

1a. All true propositions (are something that) they knew to be true. 结合上下文：... believed ....

1b. All true propositions (that) they knew to be true were things they had learned at school.

1a. All true propositions they knew to be true.
1b. All true propositions they knew to be true were things they had learned at school.

2a. Every small dog, Harry likes.
2b. Every small dog Harry likes, he pats.

2a. Every small dog (is something that) Harry likes.
2b. Every small dog (that) Harry likes (is a thing) he pats.

1a. All true propositions (are something that) they knew to be true. 结合上下文：... believed ....
1b. All true propositions (that) they knew to be true were things they had learned at school.

1a. All true propositions they knew to be true.
1b. All true propositions they knew to be true were things they had learned at school.

2a. Every small dog, Harry likes.
2b. Every small dog Harry likes, he pats.

2a. Every small dog (is something that) Harry likes.
2b. Every small dog (that) Harry likes (is a thing) he pats.

1a. All true propositions (are something that) they knew to be true. 结合上下文：... believed ....
1b. All true propositions (that) they knew to be true were things they had learned at school.

https://math.stackexchange.com/questions/1802666/a-logic-riddle-from-the-lady-or-the-tiger-by-raymondsmullyan

Just to clarify, Case 3 and Case 4 must have flawed reasoning in order to reconcile my proof with the author's.

4

I have been having a problem with a particular riddle from Raymond Smullyan and I can't seem to reconcile my proof with his solution. I am more inclined to think I am wrong, but to myself my proof looks quite convincing. Maybe you'll be able to point out the flaw in my logic.

Background: There are natives on the Isle of Questioners that are either of Type A or Type B. People of Type A may only ask questions to which the correct answer is "Yes" such as "Does 2 + 2 = 4?". People of Type B may only ask questions to which the correct answer is "No" such as "Does 2 + 2 = 5?". Furthermore there are people who have wandered onto the island who are either sane or insane. People who are sane are completely sane and believe all true things, they will always answer honestly and accurately. People who are insane are completely insane and believe all untrue things, they will always answer honestly and inaccurately. And all parties involved have perfect knowledge of the universe, but whether they are correct or incorrect about it depends on their sanity.

Problem: You meet a native and a sane or insane person named Thomas. The native asks Thomas, "Do you believe I am the type who could ask you whether you are insane?" What can be deduced about the native, and what can be deduced about Thomas?

People who are sane are completely sane and believe all true things, they will always answer honestly and accurately. People who are insane are completely insane and believe all untrue things, they will always answer honestly and inaccurately.

What Smullyan is saying is that

- for the sane group, every true proposition is something they believe to be true and every false proposition is something they believe to be false.

$$\forall\varphi[(\varphi \to B\varphi) \wedge (\neg\varphi \to B\neg\varphi)]$$

- for the insane group, every true proposition is something they believe to be false, and every false proposition is something they believe to be true.

$$\forall\varphi[(\varphi \to B\neg\varphi) \wedge (\neg\varphi \to B\varphi)]$$

What Smullyan is saying is that

- for the sane group, every true proposition is something they believe to be true and every false proposition is something they believe to be false.

$$\forall\varphi[(\varphi \to B\varphi) \wedge (\neg\varphi \to B\neg\varphi)]$$

- for the insane group, every true proposition is something they believe to be false, and every false proposition is something they believe to be true.

$$\forall\varphi[(\varphi \to B\neg\varphi) \wedge (\neg\varphi \to B\varphi)]$$

Two modalities: sane belief $\mathcal{S}$ and insane belief $\mathcal{I}$

- $\mathcal{S}\varphi$ says that the agent believes sanely in respect of whether φ, that is, if φ is true then the agent believes that φ is true, and if φ is false then the agent believes that φ is false, which can be symbolized as $(\varphi \to B\varphi) \land (\neg\varphi \to B\neg\varphi)$;
- $\mathcal{I}\varphi$ says that the agent believes insanely in respect of whether φ, that is, if φ is true then the agent believes that φ is false, and if φ is false then the agent believes that φ is true, which can be symbolized as $(\varphi \to B\neg\varphi) \land (\neg\varphi \to B\varphi)$.

Two extra modalities: reliable belief $\mathcal{R}$ and unreliable belief $\mathcal{U}$

- $\mathcal{R}\varphi$ says that the agent believes reliably in respect of whether φ, that is, if the agent believes that φ is true then φ is true, and if the agent believes that φ is false then φ is false, which can be formalized as $(B\varphi \to \varphi) \land (B\neg\varphi \to \neg\varphi)$;
- $\mathcal{U}\varphi$ says that the agent believes unreliably in respect of whether φ, that is, if the agent believes that φ is true then φ is false, and if the agent believes that φ is false then φ is true, which can be formalized as $(B\varphi \to \neg\varphi) \land (B\neg\varphi \to \varphi)$.

- 理智的信念（sane belief）：称 φ 是理智的信念，如果 φ 为真，那么主体相信它为真，并且如果 φ 为假，那么主体相信它为假（事实都为主体相信）。
- 不理智的信念（insane belief）：称 φ 是不理智的信念，如果 φ 为真，那么主体相信它为假，并且如果 φ 为假，那么主体相信它为真（事实的反面都为主体相信）。
- 可靠的信念（reliable belief）：称 φ 是可靠的信念，如果主体相信 φ 为真，那么它为真，并且如果主体相信 φ 为假，那么它为假（主体相信的都是事实）。
- 不可靠的信念（unreliable belief）：称 φ 是不可靠的信念，如果主体相信 φ 为真，那么它为假，并且如果主体相信 φ 为假，那么它为真（主体相信的都是事实的反面）。

1 四类信念的由来

2 与文献中某些逻辑的联系：表达力对比

3 这四类信念之间的关系：对当方阵

4 信念的组合

5 一种 unification 方法

The languages of the logics of sane belief $\mathcal{L}(\mathcal{S})$, insane belief $\mathcal{L}(\mathcal{I})$, reliable belief $\mathcal{L}(\mathcal{R})$, unreliable belief $\mathcal{L}(\mathcal{U})$ are defined recursively as follows, where $p \in \mathbf{P}$:

$$\begin{array}{ll}
\mathcal{L}(\mathcal{S}) & \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{S}\varphi \\
\mathcal{L}(\mathcal{I}) & \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{I}\varphi \\
\mathcal{L}(\mathcal{R}) & \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{R}\varphi \\
\mathcal{L}(\mathcal{U}) & \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{U}\varphi
\end{array}$$

Intuitively, $\mathcal{S}\varphi$ means that the agent believes sanely in respect of whether φ, $\mathcal{I}\varphi$ means that the agent believes insanely in respect of whether φ, $\mathcal{R}\varphi$ means that the agent believes reliably in respect of whether φ, and $\mathcal{U}\varphi$ means that the agent believes unreliably in respect of whether φ.

$$
\begin{array}{lcl}
\mathcal{M}, s \models p & \iff & s \in V(p) \\
\mathcal{M}, s \models \neg\varphi & \iff & \mathcal{M}, s \not\models \varphi \\
\mathcal{M}, s \models \varphi \wedge \psi & \iff & \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi \\
\mathcal{M}, s \models \mathcal{S}\varphi & \iff & (\text{if } \mathcal{M}, s \models \varphi, \text{ then } R(s) \models \varphi) \text{ and} \\
 & & (\text{if } \mathcal{M}, s \models \neg\varphi, \text{ then } R(s) \models \neg\varphi) \\
\mathcal{M}, s \models \mathcal{I}\varphi & \iff & (\text{if } \mathcal{M}, s \models \varphi, \text{ then } R(s) \models \neg\varphi) \text{ and} \\
 & & (\text{if } \mathcal{M}, s \models \neg\varphi, \text{ then } R(s) \models \varphi) \\
\mathcal{M}, s \models \mathcal{R}\varphi & \iff & (\text{if } R(s) \models \varphi, \text{ then } \mathcal{M}, s \models \varphi) \text{ and} \\
 & & (\text{if } R(s) \models \neg\varphi, \text{ then } \mathcal{M}, s \models \neg\varphi) \\
\mathcal{M}, s \models \mathcal{U}\varphi & \iff & (\text{if } R(s) \models \varphi, \text{ then } \mathcal{M}, s \models \neg\varphi) \text{ and} \\
 & & (\text{if } R(s) \models \neg\varphi, \text{ then } \mathcal{M}, s \models \varphi)
\end{array}
$$

$\mathcal{L}(W)$	$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid W\varphi$	(Steinsvold, 2011)
$\mathcal{L}(\bullet)$	$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \bullet\varphi$	(Marcos, 2005)
$\mathcal{L}(\Delta)$	$\varphi :: p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Delta\varphi$	(Montgomery & Routley, 1966)
$\mathcal{L}(\Box)$	$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi$	

$W\varphi$ is read "the agent is wrong about φ", that is, "the agent believes that φ but φ is false", $\bullet\varphi$ is read "it is accidentally true that φ" (and read "φ is an unknown truth" in an epistemic setting), $\Delta\varphi$ is read "it is noncontingent that φ" (and read "the agent knows whether φ" in an epistemic setting and "the agent is opinionated as to whether φ" in a doxastic setting), and $\Box\varphi$ is read "it is necessary that φ" (and read "the agent believes that φ" in a doxastic setting and "the agent knows that φ" in an epistemic setting).

$$\mathcal{M}, s \vDash W\varphi \iff R(s) \vDash \varphi \text{ and } \mathcal{M}, s \nvDash \varphi$$
$$\mathcal{M}, s \vDash \bullet\varphi \iff \mathcal{M}, s \vDash \varphi \text{ and } R(s) \nvDash \varphi$$
$$\mathcal{M}, s \vDash \Delta\varphi \iff R(s) \vDash \varphi \text{ or } R(s) \vDash \neg\varphi$$
$$\mathcal{M}, s \vDash \Box\varphi \iff R(s) \vDash \varphi.$$

Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two logical languages that are interpreted on the same class $\mathbb{M}$ of models,

- $\mathcal{L}_2$ is *at least as expressive as* $\mathcal{L}_1$, notation: $\mathcal{L}_1 \preceq \mathcal{L}_2$, if for every formula $\varphi_1 \in \mathcal{L}_1$ there is a formula $\varphi_2 \in \mathcal{L}_2$ such that for all $\mathcal{M}$ in $\mathbb{M}$ and all s in $\mathcal{M}$ we have $\mathcal{M}, s \vDash \varphi_1$ iff $\mathcal{M}, s \vDash \varphi_2$.
- $\mathcal{L}_1$ and $\mathcal{L}_2$ are *equally expressive*, notation: $\mathcal{L}_1 \equiv \mathcal{L}_2$, if $\mathcal{L}_1 \preceq \mathcal{L}_2$ and $\mathcal{L}_2 \preceq \mathcal{L}_1$.
- $\mathcal{L}_1$ is *less expressive than* $\mathcal{L}_2$, notation: $\mathcal{L}_1 \prec \mathcal{L}_2$, if $\mathcal{L}_1 \preceq \mathcal{L}_2$ and $\mathcal{L}_2 \not\preceq \mathcal{L}_1$.
- $\mathcal{L}_1$ and $\mathcal{L}_2$ are *incomparable* (in expressivity), notation: $\mathcal{L}_1 \asymp \mathcal{L}_2$, if $\mathcal{L}_1 \not\preceq \mathcal{L}_2$ and $\mathcal{L}_2 \not\preceq \mathcal{L}_1$.

Proposition

On any class of models, $\mathcal{L}(\mathcal{R})$ and $\mathcal{L}(W)$ are equally expressive.

Proof.

$\vDash \mathcal{R}\varphi \leftrightarrow \neg W\varphi \wedge \neg W\neg\varphi$ and $\vDash W\varphi \leftrightarrow \neg\mathcal{R}\varphi \wedge \neg\varphi$. □

Proposition

Over any class of models, $\mathcal{L}(\mathcal{I})$ and $\mathcal{L}(W)$ are equally expressive.

Consequently, $\mathcal{L}(\mathcal{I})$ and $\mathcal{L}(\mathcal{R})$ are equally expressive.

Proof.

$\vDash W\varphi \leftrightarrow \mathcal{I}\varphi \wedge \neg\varphi$ and $\vDash \mathcal{I}\varphi \leftrightarrow (\varphi \to W\neg\varphi) \wedge (\neg\varphi \to W\varphi)$. □

Proposition

On any class of models, $\mathcal{L}(\mathcal{U})$ and $\mathcal{L}(\bullet)$ are equally expressive. As a result, $\mathcal{L}(\mathcal{U})$ and $\mathcal{L}(\mathcal{S})$ are equally expressive on any class of models.

Proof.

$\vDash \mathcal{U}\varphi \leftrightarrow (\varphi \to \bullet\varphi) \wedge (\neg\varphi \to \bullet\neg\varphi)$ and $\vDash \bullet\varphi \leftrightarrow \mathcal{U}\varphi \wedge \varphi$. □

Corollary

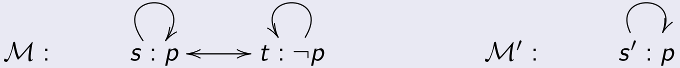
$\mathcal{L}(\mathcal{U})$ is less expressive than $\mathcal{L}(\Box)$ and incomparable to $\mathcal{L}(\Delta)$ over the class of nonreflexive models, but equally expressive as $\mathcal{L}(\Box)$ and $\mathcal{L}(\Delta)$ over the class of reflexive models.

Proposition

$\mathcal{L}(\mathcal{R}) \prec \mathcal{L}(\Box)$ on the class of d-models, t-models, b-models, 4-models, 5-models. Therefore, $\mathcal{L}(\mathcal{I})$ and $\mathcal{L}(W)$ are both less expressive than $\mathcal{L}(\Box)$ on the class of any such models.

Proof.

$\vDash \mathcal{R}\varphi \leftrightarrow (\Box\varphi \to \varphi) \land (\Box\neg\varphi \to \neg\varphi)$.



□

Corollary

$\mathcal{L}(\mathcal{U}) \not\preceq \mathcal{L}(\mathcal{R})$ and $\mathcal{L}(\Delta) \not\preceq \mathcal{L}(\mathcal{R})$ on the class of d-models, t-models, b-models, 4-models, 5-models.

Proposition

$\mathcal{L}(\mathcal{R}) \not\preceq \mathcal{L}(\mathcal{U})$ *on the class of b-models, 4-models, 5-models.*

Proof.

$\mathcal{M}: \quad s : p$ $\qquad\qquad \mathcal{M}': \quad s' : p$

□

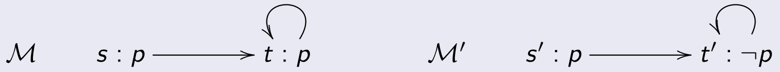
Proposition

$\mathcal{L}(\mathcal{R}) \not\preceq \mathcal{L}(\Delta)$ *on the class of b-models, 4-models, 5-models.*

Proposition

$\mathcal{L}(\mathcal{R}) \not\preceq \mathcal{L}(\Delta)$ *on the class of d-models.*

Proof.

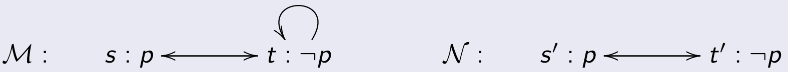


□

Proposition

$\mathcal{L}(\mathcal{R}) \not\preceq \mathcal{L}(\mathcal{U})$ *on the class of d-models.*

Proof.



Corollary

$\mathcal{L}(\mathcal{R})$ *and* $\mathcal{L}(\mathcal{U})$ *are incomparable on the class of d-models, b-models, 4-models, 5-models. Consequently,* $\mathcal{L}(\mathcal{I})$ $(\mathcal{L}(W))$ *and* $\mathcal{L}(\mathcal{S})$ $(\mathcal{L}(\bullet))$ *are incomparable on any class of models containing any of these classes as a subclass.*

Corollary

$\mathcal{L}(\mathcal{R})$ *and* $\mathcal{L}(\Delta)$ *are incomparable on the class of d-models, b-models, 4-models, 5-models. Consequently,* $\mathcal{L}(\mathcal{I})$ *and* $\mathcal{L}(W)$ *are both incomparable to* $\mathcal{L}(\Delta)$ *on any class of models containing any of these classes as a subclass.*

Corollary

$\mathcal{L}(\mathcal{R}) \prec \mathcal{L}(\mathcal{U}) \equiv \mathcal{L}(\Delta) \equiv \mathcal{L}(\Box)$ *on the class of t-models.*

表达力对比的结果

$$\mathcal{L}(\mathcal{S}) \equiv \mathcal{L}(\mathcal{U}) \equiv \mathcal{L}(\bullet)$$
$$\mathcal{L}(\mathcal{I}) \equiv \mathcal{L}(\mathcal{R}) \equiv \mathcal{L}(W) \prec \mathcal{L}(\Box)$$
$$\mathcal{L}(\mathcal{R}) \prec_t \mathcal{L}(\mathcal{U}) \equiv_t \mathcal{L}(\Delta) \equiv_t \mathcal{L}(\Box)$$
$$\mathcal{L}(\mathcal{R}) \asymp_{\bar{t}} \mathcal{L}(\mathcal{U}), \mathcal{L}(\mathcal{R}) \asymp_{\bar{t}} \mathcal{L}(\Delta), \mathcal{L}(\mathcal{U}) \asymp_{\bar{t}} \mathcal{L}(\Delta)$$
$$\mathcal{L}(\mathcal{U}) \prec_{\bar{t}} \mathcal{L}(\Box)$$

The equal expressivity results also indicate that some known logics in the literature can be translated into a non-contingency logic, that is, a logic which has the validity $\mathcal{O}\varphi \leftrightarrow \mathcal{O}\neg\varphi$ for the primitive $\mathcal{O}$ in that language.

1 四类信念的由来

2 与文献中某些逻辑的联系：表达力对比

3 这四类信念之间的关系：对当方阵

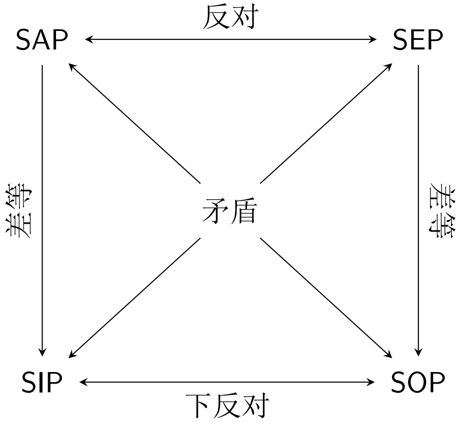
4 信念的组合

5 一种 unification 方法

Given our talk of four kinds of belief operators (sane/insane belief, reliable/unreliable belief), it is natural to ask what are the relationships among them.

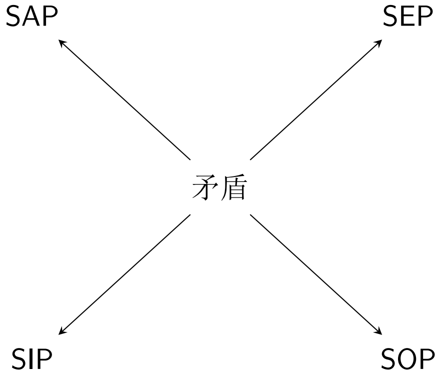
It turns out that there is an interesting property: under the assumption that the agent’s beliefs are consistent, the four operators constitute a square of opposition.

回顾：传统对当方阵（假定主项存在）



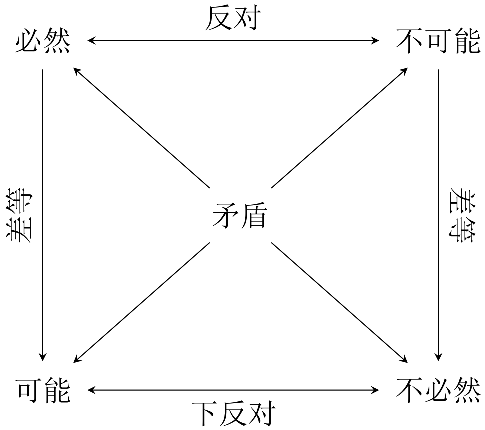
- 反对关系：不能同真，可以同假
- 下反对关系：可以同真，不能同假
- 矛盾关系：不能同真，也不能同假
- 差等/从属关系：全称真，则特称真

回顾：现代对当方阵（不假定主项存在）

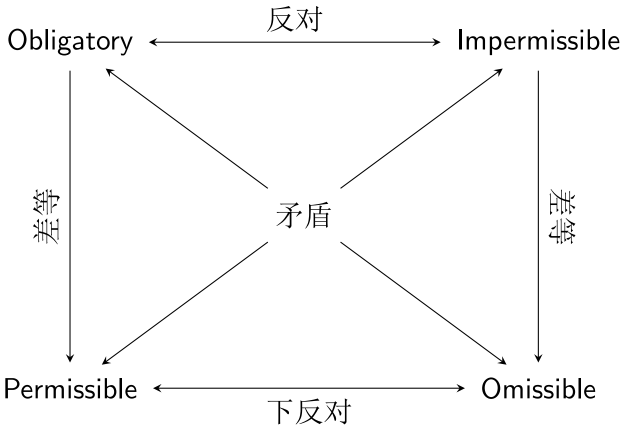


- 反对关系：不能同真，可以同假
- 下反对关系：可以同真，不能同假
- 矛盾关系：不能同真，也不能同假
- 差等/从属关系：全称真，则特称真

回顾：模态对当方阵



回顾：道义（对当）方阵


反对
Obligatory
Impermissible
差等
矛盾
差等
Permissible
Omissible
下反对

回顾

$\mathcal{M}, s \vDash \mathcal{S}\varphi$	$\iff$	(if $\mathcal{M}, s \vDash \varphi$, then $R(s) \vDash \varphi$) and
		(if $\mathcal{M}, s \vDash \neg\varphi$, then $R(s) \vDash \neg\varphi$)
$\mathcal{M}, s \vDash \mathcal{I}\varphi$	$\iff$	(if $\mathcal{M}, s \vDash \varphi$, then $R(s) \vDash \neg\varphi$) and
		(if $\mathcal{M}, s \vDash \neg\varphi$, then $R(s) \vDash \varphi$)
$\mathcal{M}, s \vDash \mathcal{R}\varphi$	$\iff$	(if $R(s) \vDash \varphi$, then $\mathcal{M}, s \vDash \varphi$) and
		(if $R(s) \vDash \neg\varphi$, then $\mathcal{M}, s \vDash \neg\varphi$)
$\mathcal{M}, s \vDash \mathcal{U}\varphi$	$\iff$	(if $R(s) \vDash \varphi$, then $\mathcal{M}, s \vDash \neg\varphi$) and
		(if $R(s) \vDash \neg\varphi$, then $\mathcal{M}, s \vDash \varphi$)

Two operators $\mathcal{O}$ and $\mathcal{O}'$ are said to be in the following relations for a class C of frames:

- *Contrary* if $C \vDash \neg(\mathcal{O}\varphi \wedge \mathcal{O}'\varphi)$ for all φ and $C \nvDash \mathcal{O}\varphi \vee \mathcal{O}'\varphi$ for some φ; (that is, they cannot be true together, but they can be false together)
- *Subcontrary* if $C \vDash \mathcal{O}\varphi \vee \mathcal{O}'\varphi$ for all φ and $C \nvDash \neg(\mathcal{O}\varphi \wedge \mathcal{O}'\varphi)$ for some φ; (that is, they cannot be false together, but they can be true together)
- *Contradictory* if $C \vDash \neg(\mathcal{O}\varphi \wedge \mathcal{O}'\varphi)$ and $C \vDash \mathcal{O}\varphi \vee \mathcal{O}'\varphi$ for all φ; (that is, they cannot be true together, and they cannot be false together)
- *in subalternation* if $C \vDash \mathcal{O}\varphi \to \mathcal{O}'\varphi$ for all φ and $C \nvDash \mathcal{O}'\varphi \to \mathcal{O}\varphi$ for some φ. ($\mathcal{O}$ is called the *superaltern* and $\mathcal{O}'$ the *subaltern*)

反对关系

Under the assumption that the agent's beliefs are consistent, sane belief and insane belief are contrary: if the agent believes sanely in respect of whether φ, then the agent does not believe insanely in respect of whether φ, but the converse does not hold.

Proposition

$C_d \vDash \neg(\mathcal{S}\varphi \wedge \mathcal{I}\varphi)$ *for all* φ, *but* $C_d \nvDash \mathcal{S}\varphi \vee \mathcal{I}\varphi$ *for some* φ.

矛盾关系

Sane belief and unreliable belief are contradictory, i.e. if the agent believes sanely in respect of whether φ, then the agent does not believe unreliably in respect of whether φ, and vice versa. Also, reliable belief and insane belief are contradictory, that is, if the agent believes reliably in respect of whether φ, then the agent does not believe insanely in respect of whether φ, and vice versa. Note that we do not need to presume the consistency of the agent's beliefs.

Proposition

$\vDash \neg(\mathcal{S}\varphi \wedge \mathcal{U}\varphi)$ *and* $\vDash \mathcal{S}\varphi \vee \mathcal{U}\varphi$. $\vDash \neg(\mathcal{R}\varphi \wedge \mathcal{I}\varphi)$ *and* $\vDash \mathcal{R}\varphi \vee \mathcal{I}\varphi$.

下反对关系

Under the assumption that the agent’s beliefs are consistent, reliable belief and unreliable belief are subcontrary: if the agent does not believe reliably in respect of whether φ, then the agent believes unreliably in respect of whether φ, but the converse does not hold.

Proposition

$C_d \vDash \mathcal{R}\varphi \vee \mathcal{U}\varphi$ for all φ, and $C_d \nvDash \neg(\mathcal{R}\varphi \wedge \mathcal{U}\varphi)$ for some φ.

差等/从属关系

Under the assumption that the agent’s beliefs are consistent, sane belief and reliable belief are in subalternation: if the agent believes sanely in respect of whether φ, then the agent believes reliably in respect of whether φ, but the converse does not hold.

Proposition

$C_d \vDash \mathcal{S}\varphi \to \mathcal{R}\varphi$ for all φ, and $C_d \nvDash \mathcal{R}\varphi \to \mathcal{S}\varphi$ for some φ.

差等/从属关系

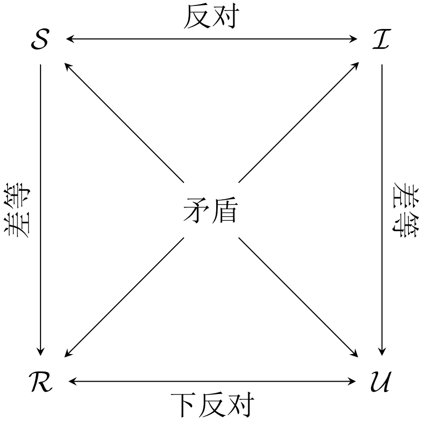
Under the assumption that the agent’s beliefs are consistent, insane belief and unreliable belief are in subalternation: if the agent believes insanely in respect of whether φ, then the agent believes unreliably in respect of whether φ, but the converse does not hold.

Proposition

$C_d \vDash \mathcal{I}\varphi \to \mathcal{U}\varphi$ *for all* φ, *and* $C_d \nvDash \mathcal{U}\varphi \to \mathcal{I}\varphi$ *for some* φ.

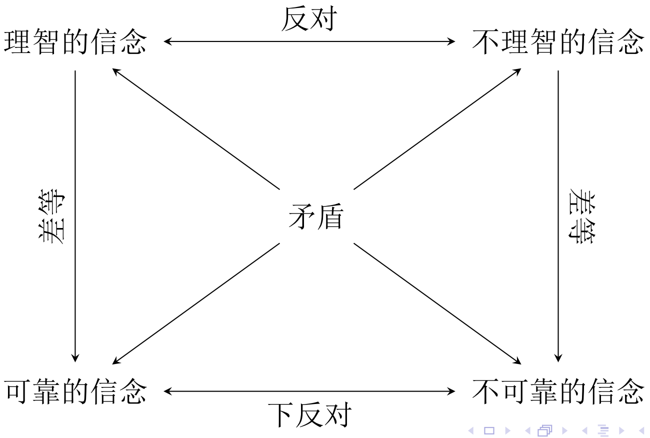
信念对当方阵 (belief square of opposition)

Under the assumption that the agent’s beliefs are consistent,

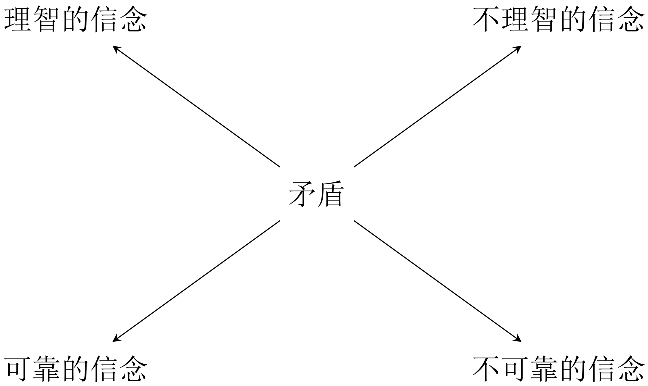


信念对当方阵 (belief square of opposition)

Under the assumption that the agent’s beliefs are consistent,



If we do not assume that the accessibility relation is serial (i.e. if we permit the inconsistency of beliefs), then only the contradictory relation holds.



1 四类信念的由来

2 与文献中某些逻辑的联系：表达力对比

3 这四类信念之间的关系：对当方阵

4 信念的组合

5 一种 unification 方法

$\mathcal{L}(\mathcal{R})$ (equi-expressive as $\mathcal{L}(\mathcal{I})$) and $\mathcal{L}(\mathcal{U})$ (equi-expressive as $\mathcal{L}(\mathcal{S})$) are less expressive than $\mathcal{L}(\Box)$ over various classes of models, and incomparable to each other over various classes of models. It is then natural to combine these logics together to increase the expressivity. In this section, we do this job, and in this way we obtain a bimodal logic. We choose to combine $\mathcal{L}(\mathcal{R})$ and $\mathcal{L}(\mathcal{U})$ together, but we can also combine $\mathcal{L}(\mathcal{I})$ and $\mathcal{L}(\mathcal{S})$.

Definition

The language of reliable and unreliable belief, denoted $\mathcal{L}(\mathcal{R},\mathcal{U})$, is defined recursively as follows:

$$\varphi ::= p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \mathcal{R}\varphi \mid \mathcal{U}\varphi.$$

Proposition

On any class of models, $\mathcal{L}(\mathcal{R},\mathcal{U})$ and $\mathcal{L}(\Box)$ are equally expressive.

Proof.

$\mathcal{L}(\mathcal{R},\mathcal{U}) \preceq \mathcal{L}(\Box)$: $\vDash \mathcal{R}\varphi \leftrightarrow (\Box\varphi\to\varphi)\wedge(\Box\neg\varphi\to\neg\varphi)$, and

$\vDash \mathcal{U}\varphi \leftrightarrow (\Box\varphi\to\neg\varphi)\wedge(\Box\neg\varphi\to\varphi)$

$\mathcal{L}(\Box) \preceq \mathcal{L}(\mathcal{R},\mathcal{U})$: $\vDash \Box\varphi \leftrightarrow (\neg\mathcal{R}\varphi\wedge\neg\varphi)\vee(\neg\mathcal{U}\varphi\wedge\varphi)$ $\Box$

The minimal system of $\mathcal{L}(\mathcal{R},\mathcal{U})$, denoted $\mathcal{RU}\mathbf{K}$, consists of the following axiom schemas and inference rules, plus PC and MP.

$$
\begin{array}{ll}
\mathcal{R}\texttt{Equ} & \mathcal{R}\varphi \leftrightarrow \mathcal{R}\neg\varphi \\
\mathcal{U}\texttt{Top} & \neg\mathcal{U}\top \\
\mathcal{R}\texttt{Con} & \mathcal{R}(\varphi\wedge\psi) \to \mathcal{R}\varphi \vee \varphi \vee \mathcal{R}\psi \vee \psi \\
\mathcal{U}\texttt{Equ} & \mathcal{U}\varphi \leftrightarrow \mathcal{U}\neg\varphi \\
\mathcal{U}\texttt{Con} & \mathcal{U}(\varphi\wedge\psi) \to \mathcal{U}\varphi \vee \neg\varphi \vee \mathcal{U}\psi \vee \neg\psi \\
\mathcal{RU} & \mathcal{R}(\varphi\wedge\psi) \to \mathcal{R}\varphi \vee \varphi \vee \mathcal{U}\psi \vee \neg\psi \\
\texttt{R}\mathcal{R} & \dfrac{\varphi\to\psi}{\mathcal{R}\psi\wedge\neg\psi\to\mathcal{R}\varphi} \\
\texttt{R}\mathcal{U} & \dfrac{\varphi\to\psi}{\neg\mathcal{U}\varphi\wedge\varphi\to\neg\mathcal{U}\psi} \\
\texttt{R}\mathcal{RU}1 & \dfrac{\varphi\to\psi}{\mathcal{R}\psi\wedge\mathcal{U}\psi\to\mathcal{R}\varphi\vee\varphi} \\
\texttt{R}\mathcal{RU}2 & \dfrac{\varphi\to\psi}{\mathcal{U}\psi\wedge\psi\to\mathcal{R}\varphi\vee\varphi}
\end{array}
$$

Definition

The canonical model of $\mathcal{RU}\mathbf{K}$ is a triple $\mathcal{M}^c = \langle S^c, R^c, V^c \rangle$, where

- $W^c = \{s \mid s \text{ is a maximal } \mathcal{RU}\mathbf{K}\text{-consistent set}\}$,
- sR^ct iff $\Omega(s) \subseteq t$, where $\Omega(s) = \{\chi \mid (\neg\mathcal{R}\chi \wedge \neg\chi) \vee (\neg\mathcal{U}\chi \wedge \chi) \in s\}$,
- $V^c(p) = \{s \in S^c \mid p \in s\}$.

Theorem

$\mathcal{RU}\mathbf{K}$ is sound and strongly complete with respect to the class of all frames.

If the axioms and rules of the minimal normal modal logic **K** of $\mathcal{L}(\Box)$ are simply translated into the bimodal language, the resulting proof system would be incomplete.

K: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

RN: $\dfrac{\varphi}{\Box\varphi}$, PC, MP

With the translation from $\mathcal{L}(\Box)$ into $\mathcal{L}(\mathcal{R},\mathcal{U})$ due to $\vDash \Box\varphi \leftrightarrow (\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)$, we obtain a proof system of $\mathcal{L}(\mathcal{R},\mathcal{U})$, where *K* and *RN* are replaced by the following axiom and rule of inference.

$$[(\neg\mathcal{R}(\varphi \to \psi) \wedge \neg(\varphi \to \psi)) \vee (\neg\mathcal{U}(\varphi \to \psi) \wedge (\varphi \to \psi))] \to [((\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)) \to ((\neg\mathcal{R}\psi \wedge \neg\psi) \vee (\neg\mathcal{U}\psi \wedge \psi))]$$

$$\frac{\varphi}{(\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)}$$

If the axioms and rules of the minimal normal modal logic **K** of $\mathcal{L}(\Box)$ are simply translated into the bimodal language, the resulting proof system would be incomplete.

K: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

RN: $\dfrac{\varphi}{\Box\varphi}$, PC, MP

With the translation from $\mathcal{L}(\Box)$ into $\mathcal{L}(\mathcal{R},\mathcal{U})$ due to $\vDash \Box\varphi \leftrightarrow (\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)$, we obtain a proof system of $\mathcal{L}(\mathcal{R},\mathcal{U})$, where *K* and *RN* are replaced by the following axiom and rule of inference.

$$[(\neg\mathcal{R}(\varphi \to \psi) \wedge \neg(\varphi \to \psi)) \vee (\neg\mathcal{U}(\varphi \to \psi) \wedge (\varphi \to \psi))] \to [((\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)) \to ((\neg\mathcal{R}\psi \wedge \neg\psi) \vee (\neg\mathcal{U}\psi \wedge \psi))]$$

$$\frac{\varphi}{(\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)}$$

If the axioms and rules of the minimal normal modal logic **K** of $\mathcal{L}(\Box)$ are simply translated into the bimodal language, the resulting proof system would be incomplete.

K: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

RN: $\dfrac{\varphi}{\Box\varphi}$, $\mathtt{PC}$, $\mathtt{MP}$

With the translation from $\mathcal{L}(\Box)$ into $\mathcal{L}(\mathcal{R},\mathcal{U})$ due to $\vDash \Box\varphi \leftrightarrow (\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)$, we obtain a proof system of $\mathcal{L}(\mathcal{R},\mathcal{U})$, where *K* and *RN* are replaced by the following axiom and rule of inference.

$$[(\neg\mathcal{R}(\varphi \to \psi) \wedge \neg(\varphi \to \psi)) \vee (\neg\mathcal{U}(\varphi \to \psi) \wedge (\varphi \to \psi))] \to$$
$$[((\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)) \to ((\neg\mathcal{R}\psi \wedge \neg\psi) \vee (\neg\mathcal{U}\psi \wedge \psi))]$$

$$\frac{\varphi}{(\neg\mathcal{R}\varphi \wedge \neg\varphi) \vee (\neg\mathcal{U}\varphi \wedge \varphi)}$$

This system is *not* complete with respect to the given semantics, because the formula $\mathcal{U}\varphi \leftrightarrow \mathcal{U}\neg\varphi$ (that is, axiom $\mathcal{U}\mathtt{Equ}$) is valid on the given semantics, but it is not provable in the system. To see the latter, consider an auxiliary semantics in which all formulas of the form $\mathcal{R}\varphi$ are interpreted as φ, and all formulas of the form $\mathcal{U}\varphi$ are interpreted as $\neg\varphi$. One may check that the proof system is sound with respect to the auxiliary semantics, but the formula is not. Note that the incompleteness result also holds if all formulas of the form $\mathcal{R}\varphi$ are interpreted as *true* instead of φ (while those of the form $\mathcal{U}\varphi$ continue to be interpreted as $\neg\varphi$).

Let $\mathcal{RU}\mathbf{D} = \mathcal{RU}\mathbf{K} + \mathcal{R}\bot$.

Theorem

$\mathcal{RU}\mathbf{D}$ is sound and strongly complete with respect to the class of serial frames.

1 四类信念的由来

2 与文献中某些逻辑的联系：表达力对比

3 这四类信念之间的关系：对当方阵

4 信念的组合

5 一种 unification 方法

Unifying reliable and unreliable beliefs

- Although $\mathcal{L}(\mathcal{R},\mathcal{U})$ can handle $\mathcal{R}$ and $\mathcal{U}$, its minimal axiomatization is rather complex, in which $\mathcal{R}$ and $\mathcal{U}$ appear together and interact.
- But there is also something else we might consider doing which involves both $\mathcal{R}$ and $\mathcal{U}$, though not appearing together in a single language.
- Instead of looking at the *interactions* between these operators, we can turn our attention to the *resemblances* between them. We thus consider the project of constructing what might be called a *unified* treatment of the two operators.

Unifying reliable and unreliable beliefs

- Although $\mathcal{L}(\mathcal{R},\mathcal{U})$ can handle $\mathcal{R}$ and $\mathcal{U}$, its minimal axiomatization is rather complex, in which $\mathcal{R}$ and $\mathcal{U}$ appear together and interact.
- But there is also something else we might consider doing which involves both $\mathcal{R}$ and $\mathcal{U}$, though not appearing together in a single language.
- Instead of looking at the *interactions* between these operators, we can turn our attention to the *resemblances* between them. We thus consider the project of constructing what might be called a *unified* treatment of the two operators.

Unifying reliable and unreliable beliefs

- Although $\mathcal{L}(\mathcal{R},\mathcal{U})$ can handle $\mathcal{R}$ and $\mathcal{U}$, its minimal axiomatization is rather complex, in which $\mathcal{R}$ and $\mathcal{U}$ appear together and interact.
- But there is also something else we might consider doing which involves both $\mathcal{R}$ and $\mathcal{U}$, though not appearing together in a single language.
- Instead of looking at the *interactions* between these operators, we can turn our attention to the *resemblances* between them. We thus consider the project of constructing what might be called a *unified* treatment of the two operators.

Now we unify the operators $\mathcal{R}$ and $\mathcal{U}$ and obtain a schematic modality which we denote G. Adding this operator to the language of propositional logic, we obtain a logic $\mathcal{L}(G)$. In symbol,

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid G\varphi,$$

where $p \in \mathbf{P}$, and G is a schematic symbol in the metalanguage (like the schematic formula φ), which can and only can be instantiated as $\mathcal{R}$ or $\mathcal{U}$. The resemblances or similarities of $\mathcal{R}$ and $\mathcal{U}$ are brought out with the aid of G, as will be indicated in the semantics and proof systems below.

Q: What about its semantics?

Now we unify the operators $\mathcal{R}$ and $\mathcal{U}$ and obtain a schematic modality which we denote G. Adding this operator to the language of propositional logic, we obtain a logic $\mathcal{L}(G)$. In symbol,

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid G\varphi,$$

where $p \in \mathbf{P}$, and G is a schematic symbol in the metalanguage (like the schematic formula φ), which can and only can be instantiated as $\mathcal{R}$ or $\mathcal{U}$. The resemblances or similarities of $\mathcal{R}$ and $\mathcal{U}$ are brought out with the aid of G, as will be indicated in the semantics and proof systems below.

Q: What about its semantics?

Recall that

$$\mathcal{M}, s \vDash \mathcal{R}\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash \varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash \neg\varphi)$$

$$\mathcal{M}, s \vDash \mathcal{U}\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash \neg\varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash \varphi)$$

$$\mathcal{M}, s \vDash G\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^n\varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^{1-n}\varphi).$$

where $n \in \{0, 1\}$ such that $n = 0$ if $G = \mathcal{R}$, and $n = 1$ if $G = \mathcal{U}$. Thus we now have two choices for instantiating G (namely, as $\mathcal{R}$ and as $\mathcal{U}$) with a matching pair of instantiations for n (0 and 1).

Recall that

$$\mathcal{M}, s \vDash \mathcal{R}\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash \varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash \neg\varphi)$$

$$\mathcal{M}, s \vDash \mathcal{U}\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash \neg\varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash \varphi)$$

$$\mathcal{M}, s \vDash G\varphi \iff (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^n\varphi) \text{ and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^{1-n}\varphi).$$

where $n \in \{0, 1\}$ such that $n = 0$ if $G = \mathcal{R}$, and $n = 1$ if $G = \mathcal{U}$. Thus we now have two choices for instantiating G (namely, as $\mathcal{R}$ and as $\mathcal{U}$) with a matching pair of instantiations for n (0 and 1).

- Both G and $(\neg)^n$ are schematic symbols, but they are not independent, in this sense.
- An instance of a schema involving them as well as the usual schematic letter φ, ψ etc. for formula is the result of putting formulas (uniformly) in place of the occurrences of φ, ψ etc., and putting either $\mathcal{R}$ or $\mathcal{U}$ (again uniformly) in place of G but specifically taking the n in $(\neg)^n$ as 0 when G is instantiated as $\mathcal{R}$, and as 1 when G is instantiated as $\mathcal{U}$. (So, we simultaneously choose the appropriate instantiation in the object language of the schematic symbols G and n, rather than just interpreting them independently, as we do for, say φ and ψ.)

Definition

The minimal logic **GK** consists of the following axiom schemas and inference rules, plus PC and MP (where again, $n = 0$ if $G = \mathcal{R}$, and $n = 1$ if $G = \mathcal{U}$):

G$\top$	$G\top \to (\neg)^n\top$
GEqu	$G\varphi \leftrightarrow G\neg\varphi$
GCon	$G(\varphi \wedge \psi) \to G\varphi \vee (\neg)^n\varphi \vee G\psi \vee (\neg)^n\psi$
RG	$\dfrac{\varphi \to \psi}{G\psi \wedge (\neg)^{1-n}\psi \to G\varphi \vee (\neg)^n\varphi}$

- We can see that **GK** is indeed simpler than $\mathcal{RU}$**K**, at least now we do not have the interaction axiom $\mathcal{RU}$ and interaction rules R$\mathcal{RU}$1 and R$\mathcal{RU}$2.
- It is not hard to find the minimal axiomatizations of $\mathcal{L}(\mathcal{R})$ and $\mathcal{L}(\mathcal{U})$ from the proof system **GK**, by instantiating G as $\mathcal{R}$ (thus n as 0) and as $\mathcal{U}$ (thus n as 1), respectively.

[Soundness] **GK** is sound with respect to the class of all frames.

Completeness of GK

Definition

The *canonical model* for **GK** is a triple $\mathcal{M}^c = \langle S^c, R^c, V^c \rangle$, where

- $S^c = \{s \mid s$ is a maximal **GK**-consistent set$\}$,
- $sR^c t$ iff $\eta(s) \subseteq t$, where $\eta(s) = \{\chi \mid \neg G\chi \wedge (\neg)^{1-n}\chi \in s\}$ (where again, $n = 0$ if $G = \mathcal{R}$, and $n = 1$ if $G = \mathcal{U}$),
- $V^c(p) = \{s \in S^c \mid p \in s\}$.

Theorem

GK *is sound and strongly complete with respect to the class of all frames.*

Serial logic

Denote **GD** $= \mathbf{GK} + \mathtt{GD}$, where GD is $(\neg)^{1-n}\bot \to G\bot$, and again $n = 0$ if $G = \mathcal{R}$ and $n = 1$ if $G = \mathcal{U}$. It is not hard to show that GD is interderivable with the following rule, denoted RGD:

$$\frac{\chi \to \bot}{\neg G\chi \wedge (\neg)^{1-n}\chi \to \bot}.$$

Thus we will use GD and RGD interchangeably.

Theorem

GD *is sound and strongly complete with respect to the class of serial frames.*

Unifying sane and insane beliefs

- We now extend the line of research in the previous section to unify sane and insane beliefs.
- In detail, we introduce a schematic modality $\mathcal{G}$ and add it to the language of propositional logic to obtain a logic $\mathcal{L}(\mathcal{G})$,
- In symbol,

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{G}\varphi,$$

where $p \in \mathbf{P}$, and $\mathcal{G}$ is a schematic symbol in the metalanguage (like the schematic formula φ), which can and only can be instantiated as $\mathcal{S}$ or $\mathcal{I}$.

Recall that

$$\mathcal{M}, s \vDash \mathcal{S}\varphi \iff (\text{if } \mathcal{M}, s \vDash \varphi, \text{ then } R(s) \vDash \varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash \neg\varphi, \text{ then } R(s) \vDash \neg\varphi)$$

$$\mathcal{M}, s \vDash \mathcal{I}\varphi \iff (\text{if } \mathcal{M}, s \vDash \varphi, \text{ then } R(s) \vDash \neg\varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash \neg\varphi, \text{ then } R(s) \vDash \varphi)$$

Semantically, $\mathcal{G}$ is interpreted as follows.

$$\mathcal{M}, s \vDash \mathcal{G}\varphi \iff (\text{if } \mathcal{M}, s \vDash (\neg)^n\varphi, \text{ then } R(s) \vDash \varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash (\neg)^{1-n}\varphi, \text{ then } R(s) \vDash \neg\varphi),$$

where $n \in \{0, 1\}$ such that $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$. Thus we now have two choices for instantiating $\mathcal{G}$ (namely, as $\mathcal{S}$ and as $\mathcal{I}$) with a matching pair of instantiations for n (0 and 1).

Recall that

$$\mathcal{M}, s \vDash \mathcal{S}\varphi \iff (\text{if } \mathcal{M}, s \vDash \varphi, \text{ then } R(s) \vDash \varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash \neg\varphi, \text{ then } R(s) \vDash \neg\varphi)$$

$$\mathcal{M}, s \vDash \mathcal{I}\varphi \iff (\text{if } \mathcal{M}, s \vDash \varphi, \text{ then } R(s) \vDash \neg\varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash \neg\varphi, \text{ then } R(s) \vDash \varphi)$$

Semantically, $\mathcal{G}$ is interpreted as follows.

$$\mathcal{M}, s \vDash \mathcal{G}\varphi \iff (\text{if } \mathcal{M}, s \vDash (\neg)^n\varphi, \text{ then } R(s) \vDash \varphi) \text{ and } (\text{if } \mathcal{M}, s \vDash (\neg)^{1-n}\varphi, \text{ then } R(s) \vDash \neg\varphi),$$

where $n \in \{0, 1\}$ such that $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$. Thus we now have two choices for instantiating $\mathcal{G}$ (namely, as $\mathcal{S}$ and as $\mathcal{I}$) with a matching pair of instantiations for n (0 and 1).

- Similar to the previous section, here both $\mathcal{G}$ and $(\neg)^n$ are schematic symbols, but they are not independent, in this sense.
- An instance of a schema involving them as well as the usual schematic letter φ, ψ etc. for formula is the result of putting formulas (uniformly) in place of the occurrences of φ, ψ etc., and putting either $\mathcal{S}$ or $\mathcal{I}$ (again uniformly) in place of $\mathcal{G}$ but specifically taking the n in $(\neg)^n$ as 0 when $\mathcal{G}$ is instantiated as $\mathcal{S}$, and as 1 when $\mathcal{G}$ is instantiated as $\mathcal{I}$. (So, we simultaneously choose the appropriate instantiation in the object language of the schematic symbols $\mathcal{G}$ and n, rather than just interpreting them independently, as we do for, say φ and ψ.)

Definition

The minimal logic $\mathcal{G}\mathbf{K}$ consists of the following axiom schemas and inference rules, plus PC and MP (where again, $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$):

$\mathcal{G}\top$	$(\neg)^n\top \to \mathcal{G}\top$
$\mathcal{G}$Equ	$\mathcal{G}\varphi \leftrightarrow \mathcal{G}\neg\varphi$
$\mathcal{G}$Con	$\mathcal{G}\varphi \wedge (\neg)^n\varphi \wedge \mathcal{G}\psi \wedge (\neg)^n\psi \to \mathcal{G}(\varphi \wedge \psi)$
R$\mathcal{G}$	$\dfrac{\psi \to \varphi}{\mathcal{G}\psi \wedge (\neg)^n\psi \to \mathcal{G}\varphi \vee (\neg)^{1-n}\varphi}$

[Soundness] $\mathcal{G}\mathbf{K}$ is sound with respect to the class of all frames.

Definition

The canonical model of $\mathcal{G}\mathbf{K}$ is a triple $\mathcal{M}^c = \langle S^c, R^c, V^c \rangle$, where

- $S^c = \{s \mid s \text{ is a maximal } \mathcal{G}\mathbf{K}\text{-consistent set}\}$,
- $sR^c t$ iff $\mu(s) \subseteq t$, where $\mu(s) = \{\psi \mid \mathcal{G}\psi \wedge (\neg)^n \psi \in s\}$ (where again, $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$),
- $V^c(p) = \{s \in S^c \mid p \in s\}$.

Theorem

$\mathcal{G}\mathbf{K}$ is sound and strongly complete with respect to the class of all frames.

Serial logic

Let $\mathcal{G}\mathbf{D}$ denote $\mathcal{G}\mathbf{K} + \mathcal{G}\mathtt{D}$, where $\mathcal{G}\mathtt{D}$ is $\mathcal{G}\bot \to (\neg)^{1-n}\bot$ (where again, $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$). It should be straightforward to see that $\mathcal{G}\mathtt{D}$ is interderivable with the following rule, denoted $\mathtt{R}\mathcal{G}\mathtt{D}$:

$$\frac{\psi \to \bot}{\mathcal{G}\psi \wedge (\neg)^n\psi \to \bot.}$$

Therefore we will use them interchangeably.

Theorem

$\mathcal{G}\mathbf{D}$ is sound and strongly complete with respect to the class of serial frames.

Conclusion

1. 四类信念的由来
2. 与文献中某些逻辑的联系：表达力对比
3. 这四类信念之间的关系：对当方阵
4. 信念的组合
5. 一种 unification 方法

Future work

- Extend the strategy to unify all of the four novel modalities; in other words, to unify G and $\mathcal{G}$. Recall that the semantics of G and $\mathcal{G}$ are as follows:

$$\begin{array}{lcl} \mathcal{M}, s \vDash G\varphi & \iff & (\text{if } R(s) \vDash \varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^n \varphi) \\ & & \text{and } (\text{if } R(s) \vDash \neg\varphi, \text{ then } \mathcal{M}, s \vDash (\neg)^{1-n} \varphi), \\ \mathcal{M}, s \vDash \mathcal{G}\varphi & \iff & (\text{if } \mathcal{M}, s \vDash (\neg)^n \varphi, \text{ then } R(s) \vDash \varphi) \\ & & \text{and } (\text{if } \mathcal{M}, s \vDash (\neg)^{1-n} \varphi, \text{ then } R(s) \vDash \neg\varphi), \end{array}$$

  where in the definition of G, $n = 0$ if $G = \mathcal{R}$, and $n = 1$ if $G = \mathcal{U}$; in the definition of $\mathcal{G}$, $n = 0$ if $\mathcal{G} = \mathcal{S}$, and $n = 1$ if $\mathcal{G} = \mathcal{I}$. Also, $\vDash \mathcal{S}\varphi \leftrightarrow \neg\mathcal{U}\varphi$ and $\vDash \mathcal{R}\varphi \leftrightarrow \neg\mathcal{I}\varphi$. This leads to the non-triviality of the unification work, since the semantics of G and $\mathcal{G}$ seems hard to be unified.

- Apply the strategy to unify other operators, e.g. contingency and accident operators.

进一步阅读

J. Fan. **Logics of (in)sane beliefs and (un)reliable beliefs**. *Logic Journal of the IGPL*, https://doi.org/10.1093/jigpal/jzaa052, Sep. 2020.

Logics of (In)sane and (Un)reliable Beliefs

JIE FAN*, *Institute of Philosophy, Chinese Academy of Sciences; School of Humanities, University of Chinese Academy of Sciences, Beijing 100049, China.*

Abstract

Inspired by an interesting quotation from the literature, we propose four modalities, called 'sane belief', 'insane belief', 'reliable belief' and 'unreliable belief', and introduce logics with each operator as the modal primitive. We show that the four modalities constitute a square of opposition, which indicates some interesting relationships among them. We compare the relative expressivity of these logics and other related logics, including a logic of *false beliefs* from the literature. The four main logics are all less expressive than the standard modal logic over various model classes, and the logics of sane and insane beliefs are, respectively, equally expressive as the logics of unreliable and reliable beliefs on any class of models. The logics of reliable and unreliable beliefs are then combined into a bimodal logic, which turns out to be equally expressive as the standard modal logic. Despite this, we cannot obtain a complete axiomatization of the minimal bimodal logic, by simply translating the axioms and rules of the minimal modal logic **K** into the bimodal language. We then introduce a schematic modality which unifies reliable and unreliable beliefs and axiomatize it over the class of all frames and also the class of serial frames. This line of research is finally extended to unify sane and insane beliefs and some axiomatizations are given.

Keywords: Sane/insane belief, reliable/unreliable belief, square of opposition, expressivity, axiomatization.

Thank you for your attention!