信念，觉知与二维逻辑

刘 虎 撰

分类号：　　密级：

UDC　　编号：

中山大學

博士学位论文

信念，觉知与二维逻辑

学位申请人：刘虎

导师姓名及职称：鞠实儿　教授

专业名称：逻辑学

2003年5月6日

摘 要

信念逻辑是人工智能研究的一个重要方向。Hintikka[1]于本世纪六十年代初期提出了使用 Kripke 的可能世界语义建立的信念模型。Hitikka 的方法已经成为被广泛采用的建立信念模型的标准方法。但是，基于这种信念模型的逻辑系统具有一些不合理的性质；其中，最主要的是假定了 agent 相信其当前信念的所有后承；这就是所谓的逻辑全知假定。但是，对于一个占有资源有限的 agent 来说，显然不具有这样的特性。

虽然 Hintikka 本人已经认识到了逻辑全知假定引起的问题，但在其后的二十年间，这一问题并没有为研究者重视。直到本世纪八十年代，信念逻辑成为了人工智能研究的一个重要方向，这一问题才重新引起广泛关注。因为，人工智能所考虑的只能是资源有限的 agent，对于这样的 agent，逻辑全知假定必须被排除。

研究者们已经提出了很多种方法，通过修正 Hintikka 的信念模型来克服逻辑全知问题。其中影响最大的是由 Fagin 和 Halpern[2]提出的（广义）觉知逻辑。简言之，觉知逻辑是使用觉知(awareness)来限制 agent 的信念，从而使得 agent 不可能具有逻辑全知性质。本文的研究主要集中在觉知逻辑方向。

首先，本文将给出一种建立 agent 的觉知－信念模型的新方法。主要做法是把觉知作为信念的预设，使用二维逻辑来建立觉知－信念模型。在文中将表明，这种二维逻辑的方法具有相当强的表达力，使得我们可以根据不同条件，灵活地建立起各种不同的觉知－信念模型。

其次，本文将使用这种方法给出三种不同的觉知－信念模型。第一个是所谓二维广义觉知逻辑模型，它建立在广义觉知逻辑的直观背景下，采用二维逻辑替代经典逻辑给出它的命题语义；这种做法使得我们在得到的逻辑中能够表达关于觉知－信念关系的更多信息。第二个是所谓二维严格觉知逻辑模型，其中 agent 采用辩护策略来决定命题的合理性，只有被判定为合理的命题才有可能成为它的信念。第三，本文给出一种二维复合语义逻辑模型，使得经典语义和二维逻辑的语义被统一到同一个模型中，从而使得“客观真”和“主观真”在同一个系统中被统一地处理。

最后，在上述三个模型的基础上，本文分别给出三个形式公理系统；并且证明了相应的元定理，其中包括有效性定理和完全性定理。

关键词：觉知 明晰信念 潜在信念 二维逻辑

ABSTRACT

Doxastic logic has been an important topic in artificial intelligence. Hintikka[1] used the possible worlds model to provide an intuitive semantics for these logics, which is generally accepted as the foundation of doxastic logics. But it also commits us to the problem of logical omniscience. Logical omniscience says that an agent believes all consequence of its current beliefs. It is clear that a reasonable agent should not have so powerful ability that it can derive all logical consequence from its current beliefs.

Hintikka himself has already discovered logical omniscience. However, there was not so much research on this topic until the eighties of this century, when doxastic logics became an important topic in artificial intelligence, where the resource-bounded agents has to be considered.

A number of logics based on Hitikka's model have been introduced to circumvent the problem. Of particular interest is the logic of awareness originated from Fagin and Halpern[2]. The idea of Fagin and Halpern is to use awareness as a limitation of belief so that agents cannot believe so much. This thesis is based on their logic.

We present a new proposal to model awareness and belief. The underlying logics are two-dimensional logics, in which awareness is considered as the presupposition of belief. We will show that by this way we are able to flexibly construct belief models according to different conditions.

Three models of belief will be given using two-dimensional logics. The first is essentially a limited variation of Fagin and Halpern's logic of general awareness, in which we substitute two-dimensional logic for classical logic as its propositional semantics. We call it two-dimensional logic of general awareness. In the second logic, which is called two-dimensional rigorous awareness logic, an agent justifies a proposition so as to make it to be its belief. Third, we give hybrid semantics such that two-dimensional and classical semantics are combined into one system.

Each belief models presented in this thesis is completely formalized as an axiom system.

Key words: awareness, explicit belief, implicit belief, two-dimensional logic.

目 录

引 言

信念自古希腊始就是哲学研究中的一个重要概念。十九世纪后半叶，在对数学基础的问题的研究中，逻辑学的现代形式－以形式公理化为特征的数理逻辑－诞生了，从而使得用形式方法分析哲学概念成为可能。哲学家和逻辑学家开发出相当多的逻辑系统来刻划不同的哲学概念，例如，必然性（模态逻辑），时间（时态逻辑），义务（道义逻辑）等。本世纪六十年代以来，为信念和知识等认识论概念寻找合适的语义表达和逻辑构造，也成为哲学和逻辑学研究的一个重要方向。这一领域通常被称为认知逻辑（epistemic logic），信念逻辑（doxastic logic）是其主要分支，也是本文讨论的主要对象。

认知逻辑公认的开创者是 Hintikka。在 1962 年他出版了影响深远的著作：知识和信念[1]。其中，Hintikka 给出了基于 Kripke 的可能世界模型的信念模型，认知逻辑被作为一种特殊的模态逻辑来研究。Hintikka 的方法已经成为认知逻辑领域的标准做法，奠定了认知逻辑研究的基础。

Hittika 所建立的认知模型是相当成功的，他给出了一个令人满意的逻辑系统和相当直观的解释。正因为如此，在 Hitikka 建立认知逻辑后的二十余年间，由于缺乏进一步研究的空间，认知逻辑的研究基本处于停滞阶段。

进入本世纪八十年代，人工智能开始吸引了越来越多的研究者的目光。逻辑方法在人工智能研究中被大量应用。一些原来产生于哲学研究的逻辑结果被应用于人工智能的研究。认知逻辑就是被他们“重新发现”的哲学逻辑之一。他们使用认知逻辑来形式刻画分布式系统和智能系统中 agent 的知识和信念，认知逻辑被广泛使用于知识表达，协议验证，形式推理方法等人工智能的方向。在这样的背景下，对认知逻辑本身的研究提出的问题是：在资源有限的条件下，agent 的信念（或知识）应该具有怎样的逻辑结构。

智能 agent 的设计和构造一直是人工智能研究的主要方向之一。对于智能 agent 有多种不同方式的定义，其中被通常使用的标准定义是：“一个系统，具有初始的信念，能够感知其周围环境，并能够通过影响其周围环境从而达到它想达到的目标。”[3]。依据上述定义，agent 对其环境和其自身状态应该具有明确的意识，它具有关于它们的信念。并且，agent 的信念将会根据环境及其自身的变化而变化。

显然，这个方向研究的第一步应该是给出合理的方式来表达 agent 的信念。自然，这也成为该方向研究的一个重要的子方向。只有首先给出 agent 的信念模型，才能进一步地讨论它的信念更新问题。本文的讨论也将集中于信念模型，而

暂不考虑 agent 的信念更新问题。

agent 基于信念的推理模式的形式化方法，大致可归为两大类：句法的方法和模态的方法。所谓句法方法，是将信念作为一个谓词“*bel*”，使用谓词逻辑来处理。例如，句子“A 相信某些东西”可形式化为“$\exists x$（*bel*（A, x））”。这种方法的优点在于它具有很强的表达力。事实上，上面那句话就不能单独使用模态方法来给出形式化表达。但是，基于句法的方法也有其内在的缺点。首先，它所使用的一阶语言相当复杂。在这种语言中，公式本身可以被作为项（term），例如，在上例中，项 x 可以是某个公式。其次，使用这种方法在具体计算时，需要诉诸传统的定理证明机制，这样导致其计算复杂性较高。由于有这些缺陷，相对于模态方法而言，这种基于句法方法的研究，不被研究者所重视。本文的工作也是属于模态方法的范围，下文中将不再涉及句法方法，有兴趣的读者可以参考 Konolige[4]以及 Reichgelt[5]。

使用模态逻辑来研究信念，已经成为认知逻辑的主流。使用这种方法，在技术层面上，认知逻辑被作为一种特殊的模态逻辑来处理，而信念模型也就是一种特殊的 Kripke 可能世界模型。模态方法的最大好处就是给予信念公式既自然又直观的解释。下一章将详细地介绍这种方法。

使用模态逻辑和可能世界语义同样具有其不利之处。这一方向的开创者 Hintikka 直接使用标准的可能世界语义来给出信念模型，由此产生了信念的不合理的性质，就是所谓的“逻辑全知问题（logical omniscience）”。Hintikka 本人已经注意到了这一问题。事实上，“logical omniscience”正是 Hitikka 首先使用的。

所谓逻辑全知问题，是指 agent 的信念的后承同样也是它的信念，并且，所有重言式都是信念。显然，对于人工智能来说，这个性质是应该尽量被避免的。因为在这种背景下，所考虑的只能是资源有限的（resource－bounded）agent，这种 agent 的推理能力局限于它所占有的资源，如内存，cpu 等。这样，即使一个命题α理论上有可能从其当前信念中被推出，α也不能当然成为该 agent 事实上拥有的信念，因为该 agent 可能没有足够的资源和推理能力来得到α。然而，直接使用标准的可能世界语义来建立信念模型，就会导致这样一个不合理的结果。

建立非逻辑全知者的 agent 的信念模型，是认知逻辑研究的主要方向之一。在短短十几年间，在这一方向上发表了大量的论文。研究者们试图修正或者扩充 Hittika 给出的标准的信念模型，来克服逻辑全知问题。限于篇幅，本文不可能给予一一介绍，仅列举其中较有影响的一些理论如下，有兴趣的读者可以查阅列出的文献：觉知逻辑（awareness logic）[2][6][7]；复合筛系统（hybrid sieve systems）[8][9]；不可能世界模型（impossible worlds models）[10]；以可能算子作为信念算子的模型（belief as possibility）[11]；非标准结构（non-standard structures）[12][13]；

明晰信念与潜在信念模型（explicit and implicit belief）[14]；多 agent 的嵌套信念模型（nested beliefs）[15]；近似知识模型（approximate knowledge）[16][17]；原理与潜在信念模型（principles and implicit belief）[11][18]；局部推理模型（logic of local reasoning）[5][7]；混合模型（fusion models）[19][20]；信念的内涵逻辑（intensional logic of beliefs）[21]；信念世界语义（belief worlds）[22][23]；动态认知逻辑（dynamic epistemic logic）[24][25]；多值认知逻辑（multi-valued epistemic logic）[26][27]。

逻辑全知所带来的问题是，它使得 agent 相信的东西过多：某些超出它的推理能力限制之外的命题也成为了它的信念。因此，一种显见的克服逻辑全知问题的办法，就是直接限制 agent 的信念集，从其中排除掉过多的信念。Fagin 和 Halpern[5][7]采用这种方案提出了所谓的觉知逻辑。直观上，他们的觉知逻辑就是使用“觉知”（awareness）作为对信念的限制：只有当 agent 觉知到一个命题，这个命题才有可能成为它的信念。Fagin 和 Halpern 的觉知逻辑具有简洁而且表达力很强的逻辑构造，同时明确地与直观相符。这种觉知逻辑不仅是认知逻辑研究中的一个重要成果，而且开创了一个新的研究方向。（在 2001 年版的哲学逻辑手册中，觉知逻辑被作为单独的一章介绍）。

本文的主要工作正是在这种觉知逻辑的基础上发展起来的。根据 Fagin 和 Halpern 的做法，可能世界模型中的每个可能世界都被赋予一个觉知集（awareness set），所有不在该觉知集中的命题，都不能成为 agent 的信念。他们直接把觉知作为信念的限制加在信念集上。我们保留他们的基本直觉，但是将采用不同的方式来实现这种由觉知而来的限制。简言之，我们的做法是把觉知作为信念存在的预设。严格地说，命题“A 觉知到 x”将被作为命题“A 相信 x”的预设。这样，我们就可以使用在理论语言学中对预设研究的已有成果来处理信念逻辑中的问题。

预设是在理论语言学和应用逻辑学中被广泛讨论的一个概念。本文的主旨当然不是讨论关于预设本身的问题。我们所做的，是把觉知处理为信念的预设，从而使用关于预设研究的已有的成果来构造信念模型。具体地，我们将采用语言学家 Bergmann 提出的二维逻辑来建立信念模型，我们最后得到的逻辑将是一种二维的模态逻辑。从下文中将会看到，使用二维逻辑使得我们可以根据不同的看待觉知与信念关系的观点，灵活地构造出不同的信念模型。不仅如此，由于在这种二维逻辑中，命题的觉知条件与命题本身的真值被相对独立地处理，使得所得到的信念模型能够表达关于觉知与信念关系的更多的信息（觉知条件的定义将在文中给出）。

Thijsse 基于部分逻辑提出的复合筛系统（hybrid sieve systems）[8][9]是觉知逻辑研究中的一个重要的成果。它使得我们可以在一个逻辑系统中统一地处理经典

逻辑和非经典逻辑。从直观上说，这种方法在一个逻辑系统中区分出两部分：经典逻辑部分，它处理的是“客观真”；非经典逻辑部分，它处理的是“主观真”。而把这两部分连接在一起的就是信念算子。在本文中将把这种方法推广使用到二维信念逻辑中。

在本文中给出的逻辑模型，相应地都给出其形式化公理系统。并且证明包括完全性定理在内的一系列系统的元定理。

本文的主要工作和贡献是：给出了三种信念模型，构造了其相应的形式系统；更重要地，我们提出了一种基于二维逻辑的研究信念逻辑的方法，它使得灵活地建立起各种不同的信念逻辑成为可能。

本文的结构如下：

第一章中介绍信念的可能世界模型及其形式化系统 KD45，从而引出由这种可能世界语义构造的信念逻辑将会遇到的问题，即所谓的逻辑全知问题。然后，我们讨论这一问题的不同形式，与人工智能的关系，以及逻辑全知性质的不合理性。

第二章详细分析 Fagin 和 Halpern 提出的广义觉知逻辑，它的语义模型以及相应的形式化公理系统；讨论这种广义觉知逻辑的优点和缺点；进而探讨广义知觉逻辑的各种修订版。

从第三章开始进入本文的主体。我们首先介绍预设问题，以及建立预设语义的已有成果。特别地，我们介绍使用二维逻辑来处理预设问题的方法。然后，我们借用这一方法来建立二维信念模型，并对这种方法给予合理性辩护。然后，我们将使用二维逻辑建立基于广义觉知逻辑的二维广义觉知逻辑，并讨论该逻辑的性质，然后给出其完全的公理系统。

第四章中我们给出另外一种使用二维逻辑建立的信念模型。这种信念模型与二维广义觉知逻辑相比，刻画的是实施更谨慎策略的 agent 的信念，在其中“觉知”被理解为 agent 的推理能力。由这种模型的建立，能够看出使用二维逻辑的最大优点－使得我们能够灵活地建立觉知－信念模型。本章还将讨论它的性质并给出其完全的公理化系统。

第五章我们引入另外一种建立信念模型的方法－复合逻辑。在这种逻辑中，经典语义和非经典语义统一在一个模型中，从而使得我们能够在一个模型中统一地处理“客观真”和“主观真”。我们把这种方法引入到二维信念逻辑中，基于上一章给出的二维信念模型，给出一种二维的复合信念模型。同样地，我们将给出其完全的公理化。

论文的最后，我们在第六章对本论文作一总结，以及提出对进一步研究的展望。参考文献附于文后。

第 1 章　标准信念逻辑与逻辑全知问题

1．1　标准的信念逻辑：KD45 系统

1．1．1 模态信念逻辑

模态逻辑发端于 Lewis1912 年的文章[28]。目前，它已经成为最重要的，应用最广泛的非经典逻辑。命题模态逻辑是在命题演算的基础上，加上两个一元算子：□，◇，分别称为必然算子和可能算子。命题模态逻辑的形成规则如下：

（1） 所有命题演算的形成规则；

（2） 如果 A 是公式，则□A 与◇A 也是公式。

模态逻辑之所以如此流行，主要是因为它具有强大的表达能力。通过给予两个一元模态算子以不同的解释，可以得到不同的逻辑。如：

（1） 信念逻辑：□A 解释为“A 被相信”(大多数信念逻辑都不是用算子◇)。

（2） 缺省逻辑：□A 解释为“通常情况下 A 成立”。

（3） 道义逻辑：□A 解释为“A 是必需做的”，◇A 解释为“A 是允许的”。

（4） 证明逻辑：□A 解释为“A 是可证的”，◇A 解释为“A 是一致的”。

（5） 时态逻辑：□A 解释为“A 将始终为真”，◇A 解释为“A 将为真”。

本文将要讨论的是信念逻辑，必然算子总是代表着相信算子，下文中不再一一指明。

在信念逻辑的文献中，相信算子通常不使用符号“□”标记，而是使用“L”或者“B”。如果要考虑的是多 agent 的模型，则使用一组加下标的算子，如“L_1，……，L_m”。公式 $B_i\varphi$解释为“第 i 个 agent 相信φ”。多 agent 信念逻辑的语言通常包括一族原子命题 $P=\{p_1, \cdots\cdots, p_n, \cdots\cdots\}$；基本的逻辑连接词~，$\rightarrow$，$\wedge$，$\vee$，$\leftrightarrow$；以及模态相信算子 L_1，……，L_m。公式由原始命题，连接词，模态算子按照形成规则形成。

本世纪五十年代之前，模态逻辑的研究集中于它的句法和代数性质上。五十年代以后至今，研究的重点转向了它的模型论。

1．1．2 信念逻辑的 Kripke 语义

可能世界模型通常被作为信念逻辑的语义模型。可能世界模型假定一个可能世界的集合，当 agent 位于其中一个可能世界时，对它而言，有一族可能世界是

可达的，或称认知可达的。直观上，一个可能世界代表着一种可能的状态，从一个可能世界 s_1 到另一个可能世界 s_2 是认知可达的，意思就是当 agent 处在 s_1 时，s_2 是它能够想象到的一种可能的状态。需要注意的是，在这里可达关系不像在时态逻辑中代表着时间上的先后关系，而是和 agent 的认知相关的关系。这种可能世界模型刻画这样一种直观：agent 可能没有关于当前状态的完整的信息，它只能根据自己可以认知到的状态来决定自己的信念。

对于上一节给出的认知逻辑的语言，它的语义是建立在这种可能世界模型基础上的 Kripke 语义。在这种语义下，agent 相信一个命题，当且仅当，该命题在所有它认知可达的可能世界(状态)中都成立。下面给出形式定义。

定义 1.1 一个 Kripke 信念模型是一个多元组 $M=(S, \pi, R)$, 其中 S 是一个可能世界的集合；π是一个真值指派，在每一个可能世界中，赋予每个原子命题一个真值 1（真）或者 0（假）；R 是定义在 S 上的一个二元关系，代表可能世界间的认知可达关系。

给定一个信念模型 M，真值关系 $M, s|=\varphi$，读作“在模型 M 中，φ 在可能世界 s 中为真（或被满足），其中 s 是 M 中的一个可能世界”，有以下的递归定义：

（1） $M, s|= p$，其中 p 是原子命题，当且仅当，$\pi(s, p)=1$；

（2） $M, s|= \sim\varphi$，当且仅当，$M, s|= \varphi$ 不成立；

（3） $M, s|=\varphi\rightarrow\psi$，当且仅当，如果 $M, s|= \varphi$，则 $M, s|= \psi$；

（3） $M, s|=\varphi\wedge\psi$，当且仅当，$M, s|= \varphi$ 并且 $M, s|= \psi$；

（4） $M, s|=\varphi\vee\psi$，当且仅当，$M, s|= \varphi$ 或者 $M, s|= \psi$；

（5） $M, s|=\varphi\leftrightarrow\psi$，当且仅当，$M, s|=\varphi\rightarrow\psi$ 并且 $M, s|=\psi\rightarrow\varphi$；

（6） $M, s|=L\varphi$，当且仅当，对于任意可能世界 t，如果 sRt，则 $M, t|= \varphi$。

定义 1.1（6）反映了上文所说的可能世界语义下的信念概念：在可能世界 s 中，agent 没有足够的关于命题 φ 是否为真的信息。在它看来，某些状态是可能的，即那些由 s 认知可达的可能世界。如果在 agent 所认为可能的所有状态中，命题 φ 总是真的，那么 agent 将相信 φ。

我们用一个简单的例子来说明这种语义。假设我现在在广州，我不知道是否“北京在下雨”(p)，也不知道是否“上海在下雨”(q)。那么我将面临四种可能的状态：p 和 q 都是真的；p 真 q 假；p 假 q 真；p 和 q 都是假的。再进一步假设，事实情况是北京没有下雨而上海在下雨。这个例子可以使用 Kripke 语义建立如图 1.1 所示的信念模型。

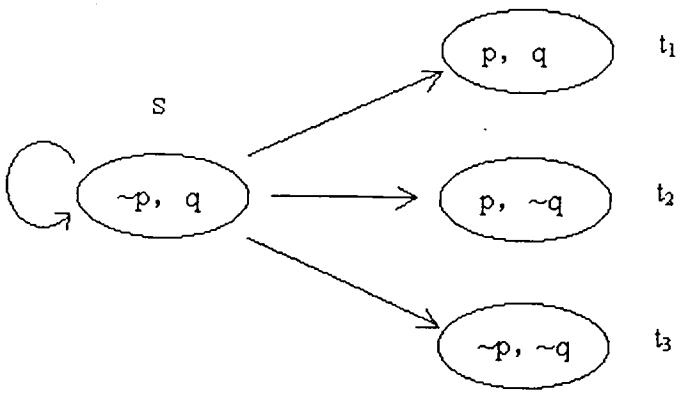

图 1.1

图 1.1 中，s 是我所在的可能世界，在其中北京没有下雨（~p）而上海在下雨（q）。该模型中共有四个可能世界（可能的状态）：s，t_1，t_2，t_3。并且从 s 到每个可能世界都是可达的，因为这四种状态都是我认为可能出现的。按照定义 1.1，只有在这种情况下我才会相信 p（q）：在 s 可达的可能世界中，p（q）总是真的。因此，我不相信 p，因为由 s 认知可达 s 和 t_3，而 p 在这两个可能世界为假；我不相信 q，因为由 s 认知可达 t_2 和 t_3，而 q 在这两个可能世界为假。形式上，即，M, s|= Lp 不成立，M, s|= Lq 不成立。

考虑这样的情况，我通过某种方式得知上海确实在下雨。那么现在对我来说，可能的状态只有两种：p 真 q 真；p 假 q 真。也就是说，从 s 认知可达 s 和 t_1，而从 s 到 t_2 和 t_3 都不可达。这时的信念模型如图 1.2 所示。

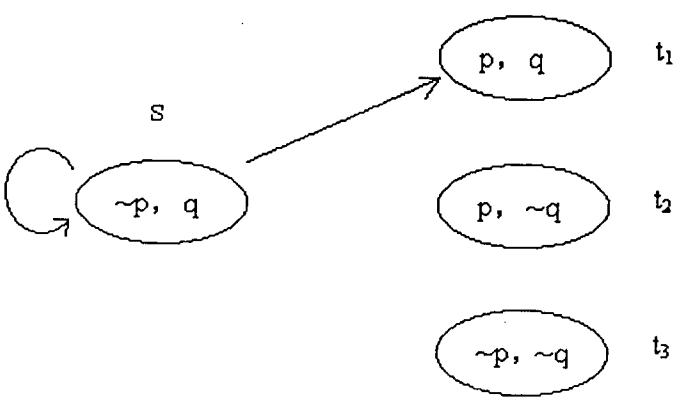

图 1.2

在该模型中有 M, s|=Lq，因为在 s 可达的世界 s 和 t_1 中，q 都为真。而 M, s|= Lp 不成立，因为 s 到其自身可达并且 p 在 s 为假。

1.1.3 KD45 系统

如果一个公式在任意模型,任意可能世界中都为真,那么称这个公式是有效的。在上一节中所定义的语义中的所有有效公式，可以被公理化为如下的系统，这个系统由公理 P，K 以及分离规则（modus ponens，记为 MP）与必然规则（necessitation rule 记为 RN）组成：

（P） 所有的命题演算的公理
（K） $L(\varphi\rightarrow\psi)\rightarrow(L\varphi\rightarrow L\psi)$
（MP） φ，$\varphi\rightarrow\psi$ / ψ
（RN） φ / $L\varphi$

在模态逻辑的文献中，这个公理系统通常被称为 K 系统。可以证明，K 系统相对于一般的可能世界模型来说是一致的和完全的，即是说，K 系统的定理集恰好包括在所有的的可能世界模型中都为真的公式。在这里我们省略了这个证明，它可以在很多关于模态逻辑的教科书中找到，例如 Chellas[29]，Chagrov[30]。

K 系统也被称为是最小模态逻辑，K 系统的定理是在所有可能世界模型中都为真的公式。这里的关键一点在于模型中的认知可达关系 R：不管 R 是什么样的二元关系，K-定理总是真的。而通过限制关系 R，可以使得原来不是有效的公式成为有效公式，模态逻辑最大的魅力一它的表达能力一正是体现于此：我们希望得到的系统具有某些性质，那么我们只需要通过限制关系 R，使得反映这些性质的公式成为有效公式。把这些公式加入到 K 系统中，我们就得到了想要的公理系统中。得到的公理系统是 K 系统的扩充。这也是 K 系统被称为最小模态逻辑的原因：其他模态逻辑系统都包含 K 系统作为子系统。

通过限制关系 R，得到不同的模态逻辑，这是模态逻辑研究的重要组成部分。在文献中可见到大量的出于不同目的的这种限制，以及由此得到的形式系统。对于信念逻辑来讲，问题就是：信念应该具有什么样的性质？

显然，要回答这个问题不可避免地将陷入哲学争论中。一个严格的哲学家将会指出，对这个问题不可能达到完全的共识，使得某个性质确定无疑地就是信念应该具有的性质。对于逻辑学家而言，他不可避免地要假定对信念的某种理解，并将它作为其构造逻辑系统的出发点。但是，这种理解或多或少地是对“信念”

概念的一种抽象。而认知逻辑发展的标志之一就是构造精细的逻辑系统，尽可能完整的描述事实上已在运行的信念系统。接受某种假定并不排斥逻辑学家对逻辑学的哲学基础采取开放态度。本文自始至终将采取上述方式处理涉及到的有争议的哲学概念。下文中将要介绍的，是在认知逻辑的研究中被广泛接受的关于信念性质的假设。在它们基础上建立起来的公理系统 KD45 事实上已经成为人工智能研究的约定的前提。本文的工作也将从这些假设为前提出发，接受 KD45 作为标准的认知逻辑系统。

首先，agent 的信念具有正自省的性质，agent 掌握着关于它自身状态的信息。这就是说，如果 agent 相信一件事情，那么它相信它是相信它的。例如，如果我相信明天将会下雨，那么我相信我具有这样一个信念：明天将会下雨。也许有些人觉得这种正自省性质并不像它看起来的那么显然。他们可能会有这样的反驳：一个人相信一个命题，但是也许他的这个信念是处于其潜意识中的，他本身并没有意识到自己相信这个命题。这种争论是有意义的，但是就人工智能研究者当前面临的任务和掌握的工具而言，把正自省作为信念的性质之一已经足够了。

下面的公式表达正自省性，传统上称其为“4-公理”：

（4） $L\varphi \rightarrow L\,L\varphi$

在上一节定义的可能世界模型中，对于可达关系 R 没有作任何限制。在这样的模型中，4-公理并不是有效的。要想使 4-公理成为有效的，必须对可能世界模型加以限制。如果可能世界模型中的可达关系 R 是传递的（transitive），即 R 满足：

$R(s, t) \,\&\, R(t, u) \Rightarrow R(s, u)$，其中 s，t，u 为任意可能世界，

那么 4-公理相对于这样的模型类是有效的。将 4-公理作为附加公理加入 K 系统，得到的系统传统上称之为 K4 系统。K4 系统相对于传递的可能世界模型是一致的和完全的。

第二个要考虑的性质是反自省性。这个性质同样来源于这样的假设：agent 掌握着关于它自身状态的信息。反自省性指的是，如果 agent 不相信一件事情，那么它相信它是不相信它的。既是说，agent 既明确地觉察哪些是它的信念，又明确地觉察哪些不是它的信念。

下面的公式表达反自省性，传统上称其为“5-公理”：

（5） $\sim L\varphi \rightarrow L \sim L\varphi$

同样地，相对于可达关系 R 无限制的可能世界模型类，5-公理并不是有效的。要想使 5-公理成为有效的，必须对可能世界模型加以限制。如果可能世界模型中的可达关系 R 是欧几里得的（Euclidean），即 R 满足：

R（s，t）& R（s，u）⇒ R（t，u），其中 s，t，u 为任意可能世界，

那么 5-公理相对于这样的模型类是有效的。将 5-公理作为附加公理加入 K4 系统，得到的系统传统上称之为 K45 系统。K45 系统相对于传递的并且欧几里得的可能世界模型是一致的和完全的。

第三个要考虑的性质是一致性。Agent 不应该相信一个矛盾命题。直观上，类似下面的话是假的："我相信明天下雨并且不下雨"。"下雨并且不下雨"显然是一种不可能出现的状态，一个理性的 agent 不应该相信这样的命题。

形式上，一致性表达为下面的公式，传统上称之为"D-公理"

（D） ~L（φ∧~φ）

同样地，相对于可达关系 R 无限制的可能世界模型类，D-公理并不是有效的。要想使 D-公理成为有效公式，必须对可能世界模型加以限制。如果可能世界模型中的可达关系 R 是序列的（serial），即 R 满足：

对任意可能世界 s，存在可能世界 t，使得 R（s，t）

那么 D-公理相对于这样的模型类是有效的。将 D-公理作为附加公理加入 K45 系统，得到的系统传统上称之为 KD45 系统。KD45 系统相对于传递的，欧几里德的，序列的可能世界模型是一致的和完全的。

在上面的论述中，有关的元定理的证明都被省略，这些证明可以在文献 Fagin[31]，Meyer[32]找到。

在认知逻辑研究中，KD45 系统被作为标准的信念逻辑系统。当然，进一步讨论上述三种性质外的其它性质，也是有意义的工作。但是本文将接受人工智能研究的约定，接受 KD45 系统作为我们讨论的基础。

最后，我们说明认知逻辑中是如何区分知识和信念的？标准的认知逻辑所采取的认识论立场，可以追溯到柏拉图的一种传统的观点：知识就是合理的真信念（justified true belief）。知识都是信念。信念可以是事实上的假命题，但是知识总

是真的。根据知识与信念的这种关系，若形式上用模态算子 K 代表知识算子，则知识算子 K 总是满足所有信念算子的性质，如有 K 算子的 D 公理：~K(φ∧~φ)。除此之外，K 算子比 L 算子多满足一个性质，传统上称其为 T-公理：

（T） $K\varphi \rightarrow \varphi$

这个公式表达的意思就是：如果知道 φ，那么 φ 是真的。相对于 T-公理的模型性质是可达关系 R 的自反性，即，

对任意可能世界 s 和 t，R（s，t）⇒R（t，s）。

将 T-公理加入到 KD45 系统作为附加公理，得到的系统称之为 S5 系统。可以证明，如果一个二元关系 R 是传递的，欧几里德的，序列的和自反的，那么 R 是一个等价关系。实际上，在上述条件中，序列性是不必要的，它可以从自反性中推出。从句法上来说，就是 D-公理可以从 T-公理中推出。因此，我们可以从 S5 系统中去掉 D-公理而不会有任何损失。

知识模型与信念模型的差别仅在于是否满足自反性，可以想见，知识逻辑与信念逻辑的差别也是不大的。从技术上讲，由于知识模型是一个等价的模型，它相对于信念模型来讲更简单一些。由于知识模型和信念模型的这种相似性，对二者的研究也是相似的，很多信念逻辑的研究结果经过简单地修改就可以成为知识逻辑的结果。事实上，在人工智能的研究中，以信念为出发点的研究更多。本文讨论的也将是信念而不是知识。因此，我们考虑的系统也是 KD45 而不是 S5。

1．2 逻辑全知问题

上文中我们讨论了信念的可能世界模型，以及标准的信念逻辑系统 KD45。接下来一个自然的问题就是：这种信念逻辑是否合理地刻画了信念这个概念的特征？

首先，它看起来是相当成功的。它具有清楚的直观背景以及简单明了的句法结构。但是，它蕴含着一个相当不合理的结果，即所谓的逻辑全知问题。Agent 是逻辑全知的，简单地说，就是 agent 具有无限的资源和推理能力，使得它可以推出所有能够推出的结论。显然，这个性质对于实际的 agent 来说过于理想化了，实际的 agent 所占有的资源总是有限的，它的推理能力也要受其有限的资源的限制。