# 1 合情推理

现实的科学逻辑目前只是熟悉那些确定的和不可能的，或者说是完全怀疑的，而没有一个是那些我们必须的推理基础（庆幸的是我们以此为基础）。所以，这个世界上的真实逻辑是概率的积分，也就是考量概率的大小，也应该是理性人的思维。

James Clerk Maxwell (1850)

设想在一个漆黑的夜晚，一个警察走在一条街道上，漫无目的的走着。突然之间，他听到盗铃响，透过街道看到一家珠宝店的窗户破了。这时，有个带着面具的男士正从那个破窗里爬出来，拿着一个袋子里面像是装满了贵重的珠宝首饰。警察丝毫也没有犹豫就毅然断定这个男士是小偷。但是他是通过怎样的推理得到如此结论？让我们首先以一种轻松的方式看这类问题的一般本质。

## 1.1 演绎推理和合情推理

想一会我们就可以清楚，警察的结论不是根据证据得到的演绎逻辑结果，因为可能存在为了表明清白任何可能的解释。比如说，有可能是这样，这个嫌疑人是这家珠宝店的拥有者，同时他正从一个假面舞会往家返，而且身上没有带钥匙。然而，就在他路过自己的珠宝店时，一个路过的卡车击起一块石头打向了店玻璃，于是他不得不保护自己的财产。

现在情况是，即使警察的推理过程不是演绎推理，我们也可能保证其很大程度上是正确的。证据并不能给嫌犯定罪，但是它确实使犯罪变得非常合情。这是一类推理的一个例子，我们学习数学理论很久后，在推理上很快就会变得或多或少的专业。我们很难能够在一个小时的行走过程中不需要面对一些情景（比如，下雨还是不会下雨？），对于这些情景，我们没有足够的信息来做出演绎推理，但是我们还是需要立即做出决定该如何做。

尽管我们对合情推理非常熟悉，得到合情结果的过程确实非常微妙的。尽管超过24个世纪的历史纪录都对它进行过讨论，但是可能还没有人曾经对此过程进行过足够的分析，同时能给其他人一个满意的答案。在本书中，我们将能够给出一些有用而令人鼓舞的新进展，这些进展包括确定的命题将取代矛盾的直觉判断，和临时特定的过程将被一些唯一确定的基础理性标准原则所替代。

所有这些问题的讨论将从演绎推理和合情推理对比的例子开始。演绎推理(证明)被普遍归功于亚里士多德（公元前四世纪）[[1]](#footnote-1)的推理工具，其最终可以分为以下两个被反复应用的强论断：

如果命题A为真，则命题B为真

命题A为真

所以，命题B为真

(1.1)

及其逆论断：

如果命题A为真，则命题B为真

命题B为假

所以，命题A为假

(1.2)

这就是我们一直喜欢用的那类推理，但是就像我们所提到的，我们所面对的几乎所有情形是，我们没有足够正确的信息来进行这类推理。我们会依靠消弱的论断（归纳）：

如果命题A为真，之后命题B为真

命题为真

所以，命题A为真变得更加合情

(1.3)

证据并没有证明命题A为真，但是论断中某一个结果的验证却使得我们对命题A为真更有信心。例如，令

最晚，早上10点可能会开始下雨；

早上10点前天空会变得乌云密布。

早上9点45分观察到阴天并不能给我们一会要下雨的逻辑必然，但是我们的尝试根据弱论断会是我们改变我们的计划和行为，如果阴天到足够程度，如同我们相信将要下雨。

这个例子也给出了一个重要前提，如果A则B表示命题B仅是命题A的一个逻辑结果，而且没有必要是物理上的因果关系，该物理的因果关系会在以后起作用。‘早上10点下雨’不是‘9点45分阴天’的物理成因。而且，恰当的逻辑联系不在不确定的因果关系(乌云⇒下雨)中，而是在确定但非因果关系的(下雨⇒乌云)中。

我们在开始就强调这里涉及的是逻辑关系，因为一些关于推断的讨论和应用出现了严重的错误，他们没有搞清楚逻辑关联与物理成因之间的关系。二者之间的区别在Simon & Rescher (1966)的著作中被深入探讨，他们指出，所有尝试将逻辑关联说出物理成因的人是漏掉了第二个论断(1.2)的对照表述。也就是说，如果我们努力将主要前提解释为‘命题A是命题B的物理成因’，则我们会很难接受‘命题非B是命题非A的物理成因’。在第三章中，我们会看到尝试将合情推断用物理成因来混淆不会有任何好处。

基于前面相同的假设前提，另一个弱论断是

如果命题A为真，之后命题B为真

命题A为假

所以，命题B为假变得更合情

(1.4)

在此情形下，证据并不能证明命题B是假的，但是它为真一个可能原因被剔除了，然后我们就对命题B为真的信心有所降低。科学家对他理论的拒绝和接受的推理，包括几乎所有的第二和第三个弱论断。

现在，警察的推理不是上面中的任何一个类型。该类型可以用另一个更弱的论断来很好的描述：

如果命题A为真，之后命题B为真变得更加合情

命题B为真

所以，命题A为真变得更加合情

(1.5)

但是，当用命题A和B来说明时，不管这个论断的表面较弱，我们都认为警察的结论具有很强的说服力。有些事情让我们相信，特别是在一些特殊情况下，警察的论证几乎具有演绎推理的说服力。

这些例子表明，大脑在进行合情推理时，不仅决定事件发生变得是否更合情或不合情，而且还以一定方式来估计合情的程度。早上10点下雨的合情程度在很大程度上取决于9点45分的阴天程度。同时，大脑还会用一些旧信息，这和用有关问题的特定新信息是一样的。我们在决定去做什么的时候，会努力回想过去有关下雨和阴天的经验，和昨天晚上的天气预报。

为了说明警察也会用自己过去的经验，我们需要仅仅改变下过去经验。假设那样的事情（有假面舞会的那种情形）对于每个警察来说每天晚上都会发生好几次，在任何情况下疑犯都可能被证明是清白的。警察会很快的学会忽略此类不足道的事情。

于是，在我们的推理过程中，我们很多事基于先验信息来估计新问题的合情程度。这样的推理过程是几乎无时滞得通过潜意识进行的，而且我们称其为常识以掩盖实际的复杂性。

数学家George P´olya (1945, 1954)关于合情推理曾写过三本书，书中给出了许多有趣的例子，也揭示出我们进行合情推理是有确定的准则的。上面提到的弱论断出现在他的第三卷书中。阅读者被强有力地驱使去研究P´olya的解释，这也是本书背后许多思想的源泉。我们在下面给出P´olya的原则是如何被定量化及得到有用应用的结果。

明显的，上面所说的演绎推理具有这样的性质，即我们能走过长长的(1.1)和(1.2)类型推理链条，且结论会和前提一样的确定。在其他类型的推理(1.3)-(1.5)过程中，结论的可信程度会随着我们走过的每个阶段而发生变化。但是，在定量形式中，我们可以找到许多情形下结论依然可以达到演绎推理那种确定性（就如警察引导我们去预期的例子）。P´olya指出即使纯粹的数学家实际上大部分时间会用到推理的弱论断。当然，对于发表新的定理，数学家会努力去找论证而用到的是前者类型，但得到定理的推理过程几乎总是包括其中一种弱推断（比如，基于下面通过类比的推理方式）。S. Banach（被S.Ulam, 1957引用）有句话表达了同样的思想观点：

*好的数学家会类比定理之间的不同；伟大的数学家会类比不同之间的不同。*

然后，让我们注意一些对另一个领域很有启发的比喻，这些比喻是其本身基于合情推理的最终分析。

## 1.2 类比物理学理论

在物理学中，我们很快就知道，这个世界复杂到我们很难一下子就能分析所有。我们只有将物理学问题变成一小块一小块后一点一点的研究，才能取得进步。有时，我们能够创建数学模型，让数学模型来塑造那些分块的某些特征，此时我们觉察到进步已经在此过程中发生了。那些模型被叫做物理理论。当我们的知识进步之后，就能够创建越来越好的模型，这些模型塑造越来越多现实世界的越来越多的特性，也会越来越精确。没有人知道是否存在进步的重点，也没有人知道此过程会走向何方。

为了懂得常识。我们需要走类似的路。我们不试图去一下子就懂得它，但是我们会感觉到，如果我们能够构建塑造一些特征的理想模型，就会有进步。我们期望我们现在能够构建的任何模型，会被更完美的模型在未来的某一天被取代，我们也不知道是否存在进步的重点。

对物理理论的对比要比对方法的简单对比要有理论深度。我们最熟悉的事物常常被证明是我们最难于理解的。那些存在的不被大多数人所知道的现象，能够用详尽的数学细节来解释（如铁元素和镍元素紫外线色谱的不同），但是当所有的现代科学面对司空见惯的草叶生长事实时，就变得不再那么实用了。因此，我们不会对我们的期望太多，我们应该准备去发现一些心理活动的最熟悉的特征，他们可能是那些构建恰当模型中最难的部分。

对比还有很多。我们已经习惯在物理学中，寻找那些任何知识的进步就会带来巨大的应用价值的结果，而且是不可以预测的。R¨ontgen对x射线的发现带来了医疗诊断上的重要可能；Maxwell在等式中加入旋度H的发现带来了全世界即时通讯的应用。

常识的数学模型也可以给出一些实践应用的特征。任何成功的模型，即使它只是塑造了常识的一点点特性，也会被证明是一些应用领域的有利拓展工具。在此领域里，模型使我们有能力去解决推断问题，而推断问题自身有太多的复杂细节，以致于我们不会在没有数学模型帮助的情况下尝试去解决他们。

## 1.3 在思考的计算机

模型有非常不同的实际应用。多数人乐于谈到：‘人类不会设计出一个可以替代人类大脑的机器，因为大脑可以做的一些事情机器是无法做到的’。在1948年普林斯顿大学的一次计算机讨论中，J. von Neumann给出了一个漂亮的回答，当时作者也很荣幸参加了。他对听众的这个简洁问题是这样来回答的：

*你坚持认为有一些事情机器不可能做。如果你能确切得告诉我机器不能够做什么事，那么我就能够做出个机器来做你所提到的那件事。*

原则上，机器不能演示的程序是那些我们不能用细节将其描写的程序，或者说是那些不能用有限的步骤完成的程序。当然，一些人会想到哥德尔不完备性、悬而未决、永不停息的图灵机等等。但是为了解答所有此类疑惑，我们仅需要指向人类大脑的存在及如何形成的。就如von Neumann所说，制造‘会思考的机器’的唯一实际限制是我们自己不知道‘思考’确切包括什么的限制。

但是在我们对常识的研究中，我们将不得使用一些关于思考机制非常简洁的构思。每一次我们构建数学模型中，都通过描述一些运算的确定集合来塑造部分常识，这样可以给我们展示如何‘建造一个机器’（即写一个电脑程序），该机器在不完全的信息条件下运行，而且用到上面几个弱论断的定量形式进行合情推理，而不是演绎推理。

实际上，基于特定推断问题而进行的此类电脑程序设计，是此领域内最活跃和最实用的趋势之一。于是需要解决的一类问题可能是：给定包括10000离散的大量观察数据，确定数据的纯度和手上的先验信息，关于原因的不同假设可能有100个相对合情理。

没有外援的常识可能足以在两个结果很不相同的假设之间做出选择，但是在解决100个区别不大的假设时，如果没有计算机和一个很好的数学理论告我们如何编程时，我们会很无助。在警察的论断(1.5)中，到底是什么决定了命题A的合情程度增加很多后使其几乎确定了，还有到底是什么决定仅仅很小可能忽略的量使得数据B几乎不相关了？本书的目的就是要构建数学理论来解答此类问题，并用可能是当下最有深度和最具有一般性的方式。

虽然我们期望一个数学理论对于编程计算机是有用的，但在心理学上在思考的计算机的想法对数学理论的构建也会有用。实际大脑所用推理过程的问题，充斥着情感和荒诞误解。如果不能够深入讨论问题，不能谈与此有关的任何东西，那时不仅我们现有的知识状态不能确定，而且和我们这里的目的也是无关的。

显然，实际大脑的运作时很复杂的，以致于我们不能对解释它的神秘进行伪装，而且我们不是尝试解释任何的事情，很少塑造那些人类大脑中的反常和不一致性表现。那些是让人感兴趣和重要的课题，但是不是我们这里研究的对象。我们的主题是规范的逻辑准则，且不是心理学和神经生理学的准则。

为了强调这一点，让我们提出这样的问题‘我们可否构造一个人类常识的数学模型，在清晰定义表达理想常识的准则之后，其将进行有用的合情推理？’，而不是该问题‘我们是如何能够构造一个人类常识的数学模型？’。

## 1.4 机器人介绍

为了将注意力集中到有意义的事情上而远离无关的争议琐事上，我们将创建一个想象中的存在物。它的大脑有我们来设计，所以它推理将遵循严格定义的准则。这些准则将由一些简单的愿景推导出，这些愿景表明看上去将是我们大脑所想望的，即：我们认为，一个理性的人发现他违背了任何一个愿景后其会据此来修正自己的想法。

原则上，我们可以自由的采用任何我们喜欢的准则，这就是我们定义要研究的机器人的方式。将机器人的推理和你的进行比较，如果你找不到类似相似之处，你可以随意去反对我们的机器人并可以升级一个你自己喜欢的。但是如果你找到一个非常相似的，并下决定去相信和去寻求它帮助你解决自己的推断问题，那么那将是该理论的完善，而不是一个前提。

我们设计的机器人将进行有关命题的推理。就如上面早已标示的那样，我们会用斜体的大写字母来标记不同的命题。对于我们的机器人，眼下我们需要所有的命题必须有无歧义的意思，还有必须是简单，必须是非真即假的定义良好的逻辑类型。那也就是说，在没有其他说明的时候，我们将要涉及的只是二元逻辑或亚里士多德逻辑。我们不需要‘亚里士多德的命题’的真假被可行的调查来确定。实际上，我们没有能力做到这一点，原因常常就是我们需要机器人帮助的理由。比如，作者个人认为下面的两个命题都是真的：

Beethoven 和 Berlioz从来没见过面；

Beethoven的音乐比Berlioz的更耐听，虽然Berlioz使劲全力才达到一般水平。

命题B对于我们的机器人目前的思考来说是不允许，尽管命题A是允许的，虽然命题A不像是，但其真假如今可以被定义出来[[2]](#footnote-2)。我们的理论被创建之后，就会有兴致去试着放松对‘亚里士多德的命题’（如命题A）的限制，进而机器人就可能会帮助我们处理诸如命题B这样的模糊命题了（见第18章的分布）[[3]](#footnote-3)。

## 1.5 布尔代数

为了将观点阐述的更加正式些，我们引入一些常用的逻辑符号，或称为布尔代数，是因为George Boole (1854)给出的和下文类似的符号。当然，演绎逻辑的原理早在Boole之前的几个世纪里就已经被理解了。我们将会看到所有布尔代数后面所包括的这些结果，都将作为Laplace (1812)给出的合情推理准则的一些特例。符号

(1.6)

被称为逻辑的积或交集，表示命题‘命题A和命题B两个都为真’。二者交换位置明显不会产生影响；和说的是同一个东西。表达式

(1.7)

被叫做逻辑的和或并集，表示命题‘命题A和命题B两个之中至少有一个为真’，和的意思一样。这些符号只是所写命题的简记方式，并不代表数值。

给定两个命题A、B，可能会发生有且仅有一个为真时另一个为真，则我们称他们有相同的真值。这可能只是一个简单的赘述（即命题A和B是两个陈述语句，其明显说的是同样的东西），或者它仅可能是在大量数学工作之后终于被证明命题A是命题B的充分必要条件。从逻辑的视角来看这并不重要，上述关系一旦成立，就意味着命题A和B有相同的真值，则他们在逻辑上市相等的。在某种意义上，任何有关一个为真的证据同样适用于另一个为真，并且他们有为进一步推理相同的含义。

明显的有，两个命题有相同的真值就有同样合情，这必然是合情推理最基本的公里。这可能看起来太微不足道了，要不是Boole(Boole, 1854, p. 286)自己在这一点上出错的事实，将两个不同的的命题错误得认定相同了，而且没有看到二者间的矛盾之处。三年过后，Boole (1857)给出了一个修正的理论，该理论接替了他早期版本书中的理论。有关此事的更多评论，请看Keynes (1921, pp. 167–168); Jaynes (1976, pp. 240–242)。

在布尔代数中，相等符号被用来表示相同的真值，而不是相同的数值：。而且，布尔代数的‘等式’于是明确肯定命题的左边与右边有相同的真值。符号‘≡’的通常意思是‘来定义的等式’。

我们用小括号来表示复合的命题，这和通常代数里的用法一致，即为了表明命题是怎样一个顺序被复合在一起的（有时，我们会用他们表达上会清楚些，虽然他们不是特别有必要）。在他们出现的地方，我们能看到算数优先级的规则，对那些手算的人来说会熟悉些，于是表示，而不是。

命题的逆用横线来表示：

(1.9)

(1.8)

为假

与间的关系是互补的：

为假

不管哪个有没有横线都无关紧要。需要注意的是，横线无歧义的用法。比如，基于上面的内容，

(1.11)

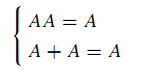
(1.10)

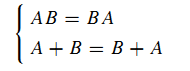
为假

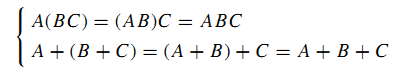
A和B都为假

这些命题是非常不同的，实际上，不是逻辑积，而是逻辑和：。

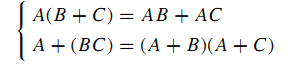
为了理解这些，布尔代数的特点是用一些琐碎而明显的基本等式表达一些特性：

幂等性：

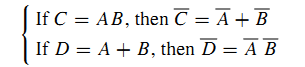
结合律：

交换律：

(1.12)

分布律：

则

对偶性：

则

通过他们的应用能够证明任何数量的进一步的关系和一些非常重要的关系。比如，我们将要用的一个基本命题：

如果，则和

(1.13)

##### 1.5.1 蕴含

命题

(1.14)

被读作‘A蕴含B’，并不是声明A或者B为真，它的意思只是是假，或者。还可以写成逻辑方程。给定(1.14)式，也就是说，如果A为真则B也必然为真；或者，如果B为假则A必然为假。这也是在强论断(1.1)与(1.2)里所阐述的。

另一方面，如果A为假，(1.14)式没有说到B的任何事情；同样，如果B为真，(1.14)式没有说到A的任何事情。但是，在弱论断(1.3)与(1.4)中的情形下，(1.14)式就会说到一些事情。一方面，‘弱论断’这个词容易让人误解。弱论断基础上的合情推理理论不是逻辑的一个‘虚弱’的形式；它是有新内容逻辑的拓展，并没有在传统演绎逻辑里面出现。在下一个章节里（见(2.69)与(2.70)）就会比较清晰的看到我们的准则将演绎逻辑囊括为一个特例。

##### 1.5.2 复杂之处

需要仔细注意的是，‘A蕴含B’在平时的语句中会被理解为是：命题B是命题A的演绎推理。但是，在正式的逻辑中，‘A蕴含B’仅表示命题AB与命题A有相同的真值。通常，命题B是否为命题A的演绎推理，不仅取决于命题B与命题A，还取决于所有的命题被接受为真，这才可以被认为是演绎推理。Devinatz (1968, p. 3)和Hamilton (1988, p. 5) 蕴含作为二元运算的真值表格，讲到仅当命题A为真且命题B为假时为假，在其他情况下为真。

这第一眼看起来使人惊讶不已，然而，需要注意的是，如果命题A与命题B确实都为真，则，即为真。用正式的逻辑任何为真的命题声明都蕴涵着任何一个其它的命题声明。换句话说，如果命题A为假，则对于所有的Q也是假，那么和都为真，所以就有和都为真，即一个错误的命题蕴含所有的命题。如果我们试着将此解释为演绎逻辑（即命题和都是命题A的演绎推理结果），那就会有任何为假的命题都是逻辑矛盾的。还有，命题‘Beethoven活的比Berlioz要久’是假的，但是几乎没有逻辑矛盾（因为Beethoven确实比那些和Berlioz同样岁数的许多人活的要久）。

仅仅知道命题A与命题B都为真，明显不能提供足够的信息来确定哪个是哪个的演绎逻辑，需要增加一些其它命题未定义的‘工具箱’。从一些其他命题里得到一个命题的演绎逻辑的问题，将在第二章最后的哥德尔定理重点讨论。普通言语中词语‘蕴含’的意思与正式逻辑中的复杂之处之间的区别，如果不能正确的理解会导致严重的错误。此处的‘蕴含’看来是不好的用词，在传统的逻辑表述中这一点并没有得到足够重视。

## 1.6 运算的完备集合

我们提到演绎逻辑的某些特性将用于机器人的设计。我们定义了四种运算，或‘关联性’，通过他们其他运算也可能被定义，以命题A和命题B开始：逻辑积或交，逻辑和或并，蕴含关系，和逆。将这些运算以任何可能的形式进行组合，就能够得到任意数量的新命题，比如：



(1.15)

许多问题在此产生了：新产生类型的命题有多大？它是无限大，还是在这些运算限制下有限?任何被命题A和命题B定义的命题都可以表达出来吗？或是还需要其他的运算关系？或是这四个运算符已经多余并需要剔除一些？足以被命题A和命题B定义的最小运算符集是什么？如果我们有任意数量的命题，而不是只有命题A和命题B，这些运算符是否足以产生命题的所有逻辑方程？

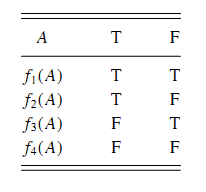
用逻辑结果、概率理论和计算机设计可以容易的回答所有的这些问题。概括的说，我们在问是否可以从现在的优势状态，一是增加方程的个数，二是减少运算符的个数。第一问用简写来应对：即使两个命题被写成(1.15)的形式看起来完全不同，但是如果他们有相同的真值，则他们在逻辑上是相同的。比如，读者自己可以去证明(1.15)中的命题C在逻辑上和蕴含是同一命题。

在此阶段，我们已经将我们的注意力放到亚里士多德的命题上了。任何逻辑方程，诸如(1.15)式只有真或假这两个可能的值。同样‘独立变量’命题A和命题B也只有两个取值。

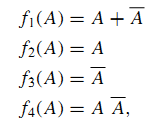
逻辑学家在此处可能会反对我们的标识符，他会说符号A已经被定义为代表一些命题，其真值不能改变。所以，如果我们考虑逻辑方程，则我们将给出新的运算符来代替，写为，此处的x、y、z是替代不同命题A、B、C的‘声明变量’。但是如果命题A代表一些固定但不特定的命题时，它既可能是真有可能是假。我们通过了解(1.15)那样的方程后，会做出同样的灵活处理，其中(1.15)那样的方程对于定义命题A和命题B的所有方式，其定义的逻辑方程为真，即我们用一个变量的声明来替代一个声明的变量。

在关系式中，我们所关系的逻辑返程被定义在仅包括个点的离散空间S上，并分别给那些基于命题A和命题B的命题赋值。而且，在每一个点上，方程独立有两值中的一个。所以，只有个不同的逻辑方程。包括n个命题的表达式是一个在包括个点的空间S上的逻辑方程，同时有个这样的方程。

在的情况下，有4个逻辑方程，其可以通过枚举法定义。将他们所有可能的取值列表如下：



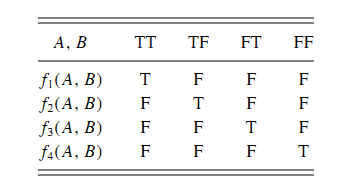
但是，通过查看就很明显的知道他们就是：



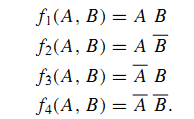
(1.16)

所以，我们通过枚举证明这三个运算符：交、并和逆就足以产生一个命题的所有方程。

对于一般情况n时，首先考虑特定方程，其每一个为真记一，且S中只有一个点。对于，有个这样的方程，

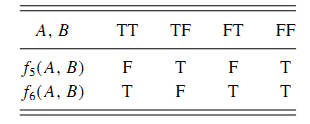


通过查看可以看到只有4个基本的交运算，

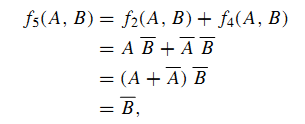


(1.17)

现在考虑任意逻辑方程在S上的确定点为真，如和，其定义为

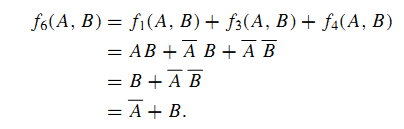


我们声明每一个这样的方程都是交运算(1.17)的逻辑和，其在同样的点上为真（这个值得注意；细心的读者可以证明这一点）。于是有



(1.18)

同样的有



(1.19)

那就是上面讨论的真值表格中的，蕴含。任何逻辑方程在空间S上至少有一点为真，能以基本交运算(1.17)的逻辑和的方式来构建。这类方程有。对于剩下的方程都为假，足以得到矛盾。

此法（在逻辑教科书中被叫做‘简化为并的一般形式’）对于任何的n都适用。比如，在的情况下，有个基本并，



(1.20)

同时有个不同的逻辑方程。其中，有个可以被写为基本交运算的逻辑和，除了一个矛盾项



(1.21)

于是可以通过‘思维的架构’证明这三个运算符

(1.22)

，即

足以产生所有可能的逻辑方程，或者，更精确的说，他们组成了一个充足的集合。

(1.12)的对偶性质表明一个小的集合就足以。对于A与B的并运算与他们都是假的逆运算相同：



(1.23)

所以，这两个运算符（且，否）对于演绎逻辑[[4]](#footnote-4)总是可以产生足够的集合。这个事实在断定什么时候我们会有足够合情推理的准则集合（见第二章）是非常重要的。

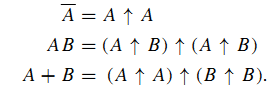
我们明显是不能够将这些运算符中的任一个移除而剩下其他的。也就是说，‘且’运算符不能被简化为逆，而且逆不能被任何数量的‘且’运算符所替代。但是还是有种可能，就是交运算和逆运算可能会被化简为某个还没被介绍的第三个运算符，这样一个逻辑运算符就构成了一个充足集合。

让人惊喜的发现是不是仅有一个而是有两个这样的运算符。运算符‘逆并’被定义为‘并’的逆：



(1.24)

被我们读作‘A逆并B’。但是我们马上会有



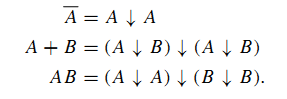
(1.25)

所以，任何一个逻辑方程能用逆并运算符一个就可以构建。同样的，运算符逆或定义为



(1.26)

也同样足以强大到可以产生所有的逻辑方程



(1.27)

利用这些运算符的优势可以来设计电脑和逻辑电路。一个‘逻辑门’是一个除了公共基准之外，具有两个输入端和一个输出端的电路。任何那样的终端上，相对于基准的电压只能够有两个值：伏或‘上’，代表‘真’，和0伏或‘下’，代表‘假’。一个逆并门就是当且仅当至少有一个输入为下时输出为上的门，或者当且仅当至少有一个输入为上时输出为下的门。然而，一个逆或门是当且仅当两个输入是下时输出为上的门。

一个逻辑电路的标准构成是‘四倍逆且门’，在一个半导体碎片上，一个完整的电路包括四个独立的逆且门。没有其他的电路构成，给定足够数量的这些逆且门，通过将他们以各种形式连到一起，可能会产生任意需要数量的逻辑方程。

对演绎逻辑的些许离题直到我们为了达到目的需要了为止。更深入的分析会在许多教科书中找到，比如，Copi (1994)对亚里士多德逻辑的现代处理。对于非亚里士多德的逻辑形式，特别是哥德尔不可能定理、可计算性、可决定性、图灵机等等，见Hamilton (1988)。

我们现在转向逻辑的延伸，它是在下面讨论的条件下给出。我们叫他们‘愿景’，而不是‘公理’，因为他们不会确定任何东西是‘真的’，而只是声明什么看起来是值得达到的目的。这些目的是不是没有矛盾就能达到，和他们是否决定了任何唯一的逻辑延伸，第二章的数学分析所关心的问题。

## 1.7 基本愿景

对于每一个机器人要进行推理的命题，其必然根据我们给它的证据，给命题赋一个合情度。并且，当它得到新证据时，其必须考虑新的证据来修正该赋值。为了实现这些赋值在大脑的电路中能被储存和修改，它们必须和一些确定的物理量相联系，比如电压、脉冲时间或二元编码的数字，等等。然而，我们的设计师想设计那些细节。为了达到这个目的，这就意味着合情度和数字之间应该有某种联系：

(1.28)

（Ⅰ）合情度由实数表示。

实用上，愿景（Ⅰ）是强加给我们的，这要求机器人的大脑必须通过运行一些确定的物理过程来运作。然而，看起来它同样需要理论（附录A）。我们趋势没有看到任何可能的一致性理论没有在功能上和愿景（Ⅰ）一样的性质。

我们采用一种非传统的而自然的方式：大的合情度要与大的数值相对应。为了更方便还需要连续性的假设前提，其在这还难以阐述清楚。从直觉上来说，一个无限小的合情度增加会带来一个无限小的数值增加。

机器人给一些命题A赋予的合情度，通常会依赖于我们所告诉他的其他一些命题B的真假。与Keynes (1921)和Cox (1961)的标识相同，我们使用表示符号

(1.29)

被我们叫做‘给定命题B为真时，’或就叫‘给定B时的A’。它代表一些实际数字。于是，举个例子

(1.30)

(被我们叫做‘给定BC时的A’)表示：给定命题B和命题C为真时，命题A为真的合情度。或

(1.31)

表示：给定命题C和命题D为真时，命题A和命题B中至少有一个为真的合情度，等等。我们已经决定用一个大点的数值代表一个大点的合情度，所以有

(1.32)

即，给定命题B为真时，命题A为真的合情度大于命题C的。在这个标记中，合情度的符号只是没有小括号的形式，我们常常为了表述的清楚性而加上小括号。因此，(1.32)和

(1.33)

是一个意思，但是第一个式子的意思看上去更明了。

为了避免发生不可的问题，不去问机器人来处理从不可能或自相矛盾的命题推理的棘手问题。因为那里不可能有‘正确’的答案。所以，当命题B和命题C相互矛盾时，我们不尝试去定义。每当此类符号出现，表明命题B和命题C是可以相容的。

还有，我们不想让机器人去以一种和你我思考方式不同的方式来思考。所以，我们要将它设计成和人类尝试推理的方式至少定性上一样的推理方式，就如上面的弱论断和一些其他类似的论断。

因此，如果机器人拥有的旧信息C升级为新信息C’，使得命题A的合情度增加：

(1.34)

但是给定命题A时命题B的合情度没有发生变化：

(1.35)

这必然会使得命题A和命题B二者为真的合情度增加，而不是减少：

(1.36)

而且这也必然会使得命题A为假的合情度减少：

(1.37)

这个定性条件简单给出了机器人推理要去的‘感知方向’，没有说合情度改变的程度大小。除了联系性前提（这也是和常识相一致的定性条件）正需要：如果变化无限小，它仅仅会造成和变化无限小。我们以这些特别的方式所用到的这些特定将在下一章给出，在那里我们就可以看到我们为什么需要他们。现今我们将他们简单总结如下：

(1.38)

（Ⅱ）定性与常识相一致。

最终，我们想给出机器人的另一个期望的性质，即人们努力没有总是达到的性质：机器人总是推理一致的。基于此，我们用三个普通口语的意思来解释词语‘一致性’：

（Ⅲa）如果一个结论可以由多种方式得出，则每一种可能方式都会导致相同结果。

(1.39a)

（Ⅲb）该代理人总是考虑到所有与问题相关的证据。它并不能随意忽略其中的一些信息，其结论只能在此基础之上。

(1.39b)

（Ⅲc）通过赋予相同的合情度，代理人总是表现出相同的知识状态。也就是说，如果代理人在两个问题的知识状态相同（除了可能被标记的命题），则它们被分配相同的合情度。

(1.39c)

愿景（Ⅰ）、（Ⅱ）和（Ⅲa）是代理人大脑内部运行的基本“结构”条件，而愿景（Ⅲb）和（Ⅲc）是代理人行为与外界关联的“界面”条件，该条件描述了机器人的行为是如何和外界相联系的。

多数学生到了这里会惊讶的知道我们对愿景的寻找已经结束了。已经表明上面的条件唯一决定了我们的机器人推理的准则，即：只有唯一的运算符集合来实现具有所有以上性质的合情度。这些准则在第二章中给出。

（在多数章节的最后，我们会插入一些非正式的评论部分，该部分会收集各种旁论和背景资料等。读者可以跳过而不会错过论证的主要线索）

## 1.8 评论

就如政客、广告商、销售人员以及为各种政治、经济、道德、宗教、心理、环境、饮食和艺术教条主义立场所布道者们很清楚得知道，容易犯错的人类思维是很容易被聪明的空话所骗，而违背上述的愿景条件。我们会设法确保他们在机器人面前不能得逞。

我们强调人类大脑与机器人之间的另一种对比。由愿望一可知，机器人有关任何命题的心理状态是由一个实数来表示。现在很明显，我们对任何给定命题的态度可能会多了一个‘维度’。你和我同时对于一个命题的判断形成，不只看它是否合情合理，还要看它是否值得、是否重要、是否有用、是否有趣、是否娱乐、是否道德等等。如果我们假设每个这样的判断可由一个数值表示，则人类心理状态的充分描述就可由相当数量维度的空间中的向量来表示。

并不是所有的命题需要如此。例如，‘水的折射率是小于1.3’的命题不会对情绪有任何作用。因此，它产生的精神状态具有很少的维度。另一方面，命题‘你的婆婆刚刚毁了你的新车’生成一种具有很多维度的精神状态。日常生活中相当一般的情况，都是那些涉及许多维度的。我们建议，只是出于这样一个原因，即最熟悉的例子的心理活动往往是最难用模型重现的。或许在这里，有为什么科学和数学中是最成功的人类活动的原因：他们处理产生所有精神状态的最简单命题。这样的一些状态的将会是那些最少被一定的人类思维缺陷所干扰的状态。

当然许多目的下，我们不希望机器人采取如此更多，从其它的维度所产生的任何一个‘人’的特征。事实上只是，电脑不被情感因素所困扰，不会厌倦冗长的问题，追求的不是反对我们的隐藏动机，使得他们比人类更安全的去执行某些任务。

插入的这些话指出，该理论在这里有许多未探索的扩展和广义的可能。这也许可激发别人尝试发展心理活动的‘多维理论’，将更多类似于实际中不都是不可取的人类大脑行为。如果这样一个理论成功的话，可能会有超出目前所能想象的重要性[[5]](#footnote-5)。

然而到目前为止，我们将不得不满足于一种更加中性的处理。是否可能采用一种一维的合情推理模型？如果我们能够用一个单独的实数来表示合情度，并忽略刚刚提到的其他维度，很明显，我们的问题将是最简单的。

需要强调的是，我们实在是没有办法断言实际人类心目中的合情度会有一个唯一的数值来测度。我们的工作并不是要假设或确实猜测任何的此类事情，而是探讨在机器人中，是否有可能，没有矛盾得建立此类关系。

但对一些人来说，可能我们看上去已经做了必要假定之外的假设，从而无端的限制我们的理论一般性。为什么我们必须由实数表示合情度呢？基于定性序数关系系统的‘对比’理论，诸如，不足以表示吗？这一点将在附录A中被进一步讨论。在附录A中，我们描述概率理论的其他方法，并且指出为了发展‘对比’的理论所作出的一些尝试，此处的‘对比’理论被认为将是逻辑上比较简单或更普遍的理论。但这最终不是这样的，所以即使它很有可能以其他的方式建立的还要基础，但最终的结果是相同的。

##### 1.8.1 不同语言VS.正式逻辑

我们应该注意到形式逻辑语句和那些普通语言之间的区别。人们可能会认为，后者只是一个不太精确的表达式形式，但详细的考察中这种关系会变得不同。在我们看来，对那普通的语言的小心使用，不需要比形式逻辑少些精确。但普通语言的规则更复杂，也因此比我们在形式逻辑中所允许的表达，其更具丰富的可能性。

特别的是，普通语言而不是逻辑语言，为了其他目的被经常使用，演化出了微妙细节-暗示的东西而无需实际声明它-形式逻辑中却没有。A先生为了证明他的客观性，说‘我相信自己看到了什么’。B先生反驳说: 他看不到他不相信的。从形式逻辑的角度来看，似乎他们说的是同样事情，然而从日常语言的角度来看，这些表达却有传达相反意义的意图和效果。

这里有一个取自数学教材的重要例子。让L是平面上的一条直线，S是该平面上一个点的无穷集合，该集合的每一点都映射到直线L上。现在考虑下面的语句：

(I)极限的映射是映射的极限。

(II)映射的极限是极限的映射。

这些语句的语法结构是‘A是B’和‘B是A’，所以他们可能会出现逻辑上的相等。然而在教科书中，一般情况下(I)被认为是真的而(II)不是真的，原因是当集合的极限不存在时该映射的极限可能存在。

正如在这里所看到的，在普通语言中，甚至在数学教科书中，我们已经学会了阅读微妙细节的确切表述意义，只是当时可能没有意识到这一点，直到这样的例子点了出来。我们解释‘A是B’时，就已首先将A存在作为一类重要前提。而且，其余部分的陈述被认为是在建立在这一前提条件下的。换一种说法，普通语法中的动词‘是’意味着主语和谓语之间的区别，而在形式逻辑中或传统数学中的‘=’却没有这样的区别。（然而，在计算机语言中，我们遇到诸如‘’类的语句，每个人都对此似乎理解，但在其中的‘=’号毕竟现在已经具有那中隐含的区别了。）

另一个有趣的例子是格言‘知识就是力量’，其在热力学和人类关系中是一句很有说服力的真理。一位化工贸易杂志的广告作家将此句错用为‘权力就是知识’，显得荒谬、很下流、虚假。

这些例子提醒我们，动词‘是’像任何其他动词一样有一个主语和一个谓语，但它很少被注意到这个动词有两个完全不同的含义。母语是英语的人可能需要一些努力才看出‘The room is noisy’和‘There is noise in the room’中的不同含义。但在土耳其，这些意思通过不同词加以区别清楚，使得使用错误语句的访客不被理解。后者是本体论，断言事物的物理存在，而前者是认识论，表达的只是说话者的个人看法。

普通语言，或者至少英语语言，有几乎统一的趋势将它们放入一种语法形式来掩饰认识论语句，以此来表示不被提防的本体论语句。当前概率论中的错误，主要是因为不假思索而未能察觉到这一点。为了解释第一种本体论意义上的语句，要断定个人自己的想法和感觉是外在现实自然存在的。我们把这称为‘心灵投影的谬论’，并且点到它接连多次所引起的麻烦。但这种麻烦几乎不局限于概率理论，当量子理论被指出时，明显就会有很多哲学家和格式塔心理学家的布道和物理学家们试图解释它，通过作者不断的重复陷入心灵投影谬误，其将沦为无意义的话。

这些例子说明，当我们尝试将复杂的普通语言语句翻译成较简单的逻辑语句时，需要小心为上。当然，普通语言往往比在形式逻辑中我们想要的精确性要差，但每个人都希望如此，而且对此事有警惕的，所以它的危险性降低了。

普通人得花大约20年才能掌握普通语言中的所有微妙细节，让我们的机器人掌握这些，实在是对它期望过高了。从这方面来看，我们的机器人依然会像个小孩一样，它将从字面上解释所有的语句，而且会将真相脱口而出，而不去想这可能会得罪到别人。

对于作者来说，去设计一个能够认识到这些更微妙隐晦意义的新模型机器人，目前尚不清楚写作的难度，甚至更不清楚如何得到。通过人类大脑的存在性，一下就可以处置原则的问题。但是，在实践中，von Neumann的原则是适用的。我们设计的机器人不能做到，直到有人创建了‘微妙识别’的理论，该理论减少到一个确定运算集合的过程。我们很高兴地把这个留给别人来解决。

在任何情况下，我们目前的机器人模型是非常真实的，因为现今几乎普遍真理的是，任何重要的概率运算都可由一台计算机来运行。设计该计算机的人，不论他们是否有那样的想法，一定是根据机器人应该怎样行为的一些先入为主的概念，来设计机器人大脑的一部分。但现在所使用的计算机程序很少能满足所有的愿景。事实上，大多数是直观临时设定的，都未选择思维中任何定义良好的愿景。

任何此类的临时设定在一些特殊的应用中大概是可用的-这是选择它的标准-但就如第二章的证明将显示，那些与概率理论规则相冲突的任何临时设定，当我们尝试在限制领域应用时，都有明显不一致的地方。我们的目标是通过建立推理的一般原则，直接从一致性的要求，来一劳永逸的避免这种情况发生，并以一种的形式，适用于以一种足够明确方式所形成的合情推理中的任何问题。

##### 1.8.2 吹毛求疵

如上可以看出，在当前的工作中，我们使用术语布尔代数，用它建立已久以来的意思，其指的是像符号‘’来表示命题这样的二值逻辑。喜欢吹毛求疵的人已向我们抱怨到，一些数学家赋予略有不同的含义来使用该术语，其中，‘’可以指一组命题。但两种用法并不冲突。我们认识到更广的含义，但只是没有找到合适的理由让自己来用它。

已被我们称为布尔代数的规则和符号集有时被称为‘命题演算’。该术语似乎仅用于加法的目的，我们还需要另一套称为‘谓词演算’的规则和符号。然而，这些新的符号会成为仅仅短而熟识短语的缩写。‘universal quantiﬁer’仅是‘for all’的简称；存在‘existential quantiﬁer’是‘there is a’的简称。如果我们只是用直白的英语陈述，我们会自动使用所有的谓词演算，我们为了目的而需要并做的更加清楚明了。

第二个强段论的有效性（在两值逻辑中）有时会被质疑。然而，在当前数学中，给定的定理可通过给出一个反例来反驳，如果我们能够从一组陈述中推出矛盾那么该组陈述就被认为是非一致性的，一个命题可以通过反证法来建立并从其否定中产生矛盾，这几个说法似乎仍然被认为是正确的推理。这对于我们来说已经足够了，且我们是很乐意跟随这悠久的传统。我们在这一立场上的安全感源自于承认：当逻辑可能会在未来向前进时，它几乎不可能往后退步。一个新逻辑可能会导致亚里士多德逻辑什么都没有说的新结果，这就是我们在此处正试图创建的。当然，如果一个新逻辑被发现在一些地方与亚里士多德逻辑相冲突，而亚里士多德逻辑在那些地方却是是适用的，我们会认为此异议对新逻辑是致命的。

因此，对于那些感到被两值演绎逻辑所局限的人，我们可能只有说:‘如果你愿意，通过一切手段调查其他的可能性；当你找到一个新的结果并不包含在二值逻辑或我们对它的扩展逻辑中，并且在科学推理中很有用的时候，请让尽快我们知道它。其实，文献中已经有很多不同且相互不一致的多值逻辑。

但在附录A中，我们引用论点：他们已经不是在两个值逻辑中，表明他们没有实质的内容。也就是说，被应用到一组命题的n值逻辑，要么等同于两值逻辑被应用到一个拓展的集，要么它包含内部不一致性。

我们的经验与推断是一致的。在实践中，多值逻辑似乎用不到找到新的有用结果，但宁可去尝试删除二值逻辑的所谓困难，特别是在量子理论、模糊集和人工智能中。但仔细研究后，我们所知的所有这些困难已经被证明是头脑投影谬误的例子，其呼吁进行概念的直接修改，而不是一个新的逻辑。

1. 今天，对亚里士多德的实际贡献有许多不同的看法。该类问题和本书引用的目的没有关系，但是有兴趣的读者可以在Lukasiewicz (1957)的书中找到相关的一些讨论。 [↑](#footnote-ref-1)
2. 他们的碰面在时间上是有可能的，因为他们生命中有24年是重合的。我对此怀疑的理由是Berlioz在他的回忆中并没有提及任何见面的事情。另一方面，他也没有说他们确实见过面。 [↑](#footnote-ref-2)
3. 如何设计一个机器具有概念意义的‘察觉’意识，就如命题A对于人类，此问题看起来太难了。许多人工智能的目的是创建临时特定的工具来解决此类问题。然而，我们会在第4章发现，该问题几乎不会存在，合情推理的准则自然而然的就会提供一些均值来做出它的数学等式。 [↑](#footnote-ref-3)
4. 你可以回想：这两个命令是写任何电脑程序的仅有两个吗？ [↑](#footnote-ref-4)
5. 事实上，一些心理学家认为，五个维度就足以描述一个人的性格。即，我们之间的区别是五大基本人格特质的不同搭配所决定的。但我们看来，这必然是被严重简化的结果。可辨认的化学因素会在空间和时间上不断变化进而影响心理活动（如脑中葡萄糖代谢的分布），而不能仅用五个维度的空间来表示清楚。然而，五个维度可能为了多种用途，足以捕获有益的事实。 [↑](#footnote-ref-5)