逻辑学简短入门

作者：Graham Priest

译者：wxflogic

重译说明

Graham Priest 的 Logic: A Very Short Introduction 是牛津通识系列中的一本。该书在众多逻辑学入门书中独树一帜，并不试图完整介绍逻辑学的理论，而是通过一些哲学难题或逻辑谜题引入解决这些问题的逻辑理论和方法，在介绍逻辑知识的同时展示逻辑可以如何来用。虽然篇幅简短，但涵盖了从演绎逻辑到归纳逻辑，从经典逻辑到非经典逻辑，从理论推理到实践推理等几乎全部逻辑学分支。与普通教科书不同，该书每章都留下了开放的问题供读者进一步思考和探索。当然，作者已经声明，这并不是一本教科书。因而该书只能起到引领和一窥门径的作用。想要系统学习和掌握逻辑学的读者在看完本书后还需找专门的教科书来阅读 [1] 。

该书中译本《简明逻辑学》由译林出版社出版。翻译总体流畅，不会出现英式中文，但有的地方翻译得也比较随意，虽不失原意，但多有出入。更严重的是，由于译者并非专业出身，对一些基本逻辑学术语的翻译都不准确（如将 validity 译为效度）。还有一些地方，译者囿于学识未能理解作者所讲的内容，存在不少误译。虽然该书英文版通俗易懂，但对于完全不懂逻辑又希望能迅速入门的读者而言，有一个准确可靠又能方便获取的中译本还是有益的。

我本人虽是逻辑学专业出身，在术语翻译方面自信不会出错，但并未从事过文字翻译工作，在其他地方难免错译。故在这里公开译本，以供大家批评指正。

更新：新增了原书第 2 版新增两章的翻译。

想要 Fork 该译本的请移步至我的 Github 仓库（若已在 Github 请忽略）。在线阅读地址（Gitbook改版后不能下载了）

前言

逻辑学是最古老也是最现代的学科之一。它始于公元前4世纪。比它更古老的学科只有哲学和数学，二者与逻辑学一直有着密切联系。20世纪之交，通过应用新的数学技巧，逻辑学发生了变革；在过去半个世纪里，逻辑在计算和信息处理中，起到了全新而重要的作用。因而它在人类许多思想和行为中成为中心主题。

本书按当代逻辑学家现在如何理解该主题对逻辑学进行介绍。但它并不试图成为一本教科书，这样的书如今比比皆是。本书的目的在于探索逻辑的根源，它潜藏于哲学，一些形式逻辑会顺带进行讲解。

每一章我从某个特定的哲学难题或逻辑谜题出发，然后讲解一种解决它的方案，通常是相当标准的方案。但在某些领域却没有标准答案：逻辑学家之间仍然存在分歧。在这种情况下，我就只是选取一种有趣的方案。几乎所有的方案，无论是否标准，都可以质疑。每章结束时，我会对讲解的方案指出一些存在的问题。有时这些问题是标准的，有时不是。有时这些问题有简单的答案，有时不一定。这样做的目的是激励读者自己对这些问题进行思考。

现代逻辑学是高度数学化的学科。我尝试以避免几乎所有数学的方式来写作这些内容。最多在最后几章要求一点高中代数的知识。当然，你要下决心来掌握一些对你而言可能是新的符号，但这比初步掌握任何一门语言要容易得多。并且，符号给困难问题带来的明晰性，使得任何掌握这些符号可能遇到的麻烦都是值得的。但有一点要注意，读一本逻辑或哲学方面的书不同于读一本小说，你不得不时常慢慢和仔细地阅读。有时你可能还得停下来想一想。如果必要的话，你还应该准备回过头重读某个段落。

本书最后一章是关于逻辑学的发展的。在那里，我试图将本书涉及的问题放入一种历史视角中，以表明逻辑学是一个活生生的学科，一直在逐步发展，并将继续如此。这一章还包含一些进阶阅读的建议。

本书有两个附录。第一个是术语和符号表。如果你忘记了某个词或符号的意义，你可能需要查阅它。第二个附录包含每一章的一道相关题目，你可以用来检验对这一章主要思想的理解。

本书追求广度而不是深度。每章内容都可以轻易写成一本书——这些书其实已经写出来了。即便如此，还有一些逻辑学中非常重要的问题，我在这里并未触及。但如果你坚持读完本书，你就会对现代逻辑的基本内容有一个很好的了解，并且会理解，为什么人们会认为逻辑是值得思考的。

第 1 章：有效性：从什么可以推出什么？

大多数人喜欢自认为是有逻辑性的。告诉某人“你没有逻辑性”，通常是一种批评。不合逻辑意味着混乱、糊涂和不理性。但什么是逻辑呢？在刘易斯·卡罗尔的《镜中世界》中，爱丽丝遇见了一对强词夺理讲逻辑的兄弟——特伟哥和特伟弟。在爱丽丝一时无语时，他们开始攻击爱丽丝：

> “我知道你在想什么”，特伟哥说：“但现在并不是这样的，决不是”。
>
> “反过来说”，特伟弟接着说：“如果过去是这样的，那么可能是这样的；假如现在是这样的，那就是这样的。但现在不是这样的，所以不是这样的。这就是逻辑。”

特伟弟所做的，至少在卡罗尔的戏仿作品中，是推理。而正如他所说，逻辑就是关于推理的。

我们都进行推理。我们根据已知进行推理，设法弄清什么是这样的。我们试图通过给出理由，来说服他人某事就是这样的。逻辑学就是研究什么可以算作什么的好的理由，以及为何如此。不过，你得以特定方式来理解这一论断。这里有两条推理：逻辑学家称之为**推断**（*inferences*）：

1. 罗马是意大利的首都，这架飞机在罗马着陆；因此，这架飞机在意大利着陆。
2. 莫斯科是美国的首都；因此，你无法去了莫斯科而没有去美国。

在每条推理中，“因此”之前的论断——逻辑学家称之为前提——在给出理由；“因此”之后的论断——逻辑学家称之为结论——被认为是由这些理由所支持的。第1条推理是好的；但第2条就很成问题，也不会说服任何有地理常识的人：莫斯科是美国首都这一前提就是假的。不过，要注意，假如前提是真的——比如说美国已买下了整个俄罗斯（而不只是阿拉斯加）并把白宫搬到莫斯科，从而离欧洲的权力中心更近——结论就会确实为真了。它会由前提推出；而这就是逻辑学所关心的。它不关心推断的前提是真是假，那是其他人的事（在这个例子中，是地理学家的事）。逻辑学家把结论确实可以从前提推出的推断，称为**有效的**（*valid*）。因此逻辑学的中心目标就是理解**有效性**（*validity*）。

你或许认为这很无聊——一种吸引力还不如解纵横字谜的智力活动。但结果证明，这不但是一件很困难的事，而且与若干重要的（有时还是深刻的）哲学问题密不可分。我们接下来会看到其中一些。现在，让我们了解更多一些有关有效性的基本事实。

首先，一般区分两种不同的有效性。为了理解这一点，考虑下面三条推断：

1. 假如窃贼是从厨房破窗而入，就会在外面留下脚印；但并没有脚印；因此，窃贼不是从厨房破窗而入。
2. 琼斯的手指有尼古丁痕渍；因此，琼斯抽烟。
3. 琼斯每天买两包烟；因此，有人在厨房窗外留下了脚印。

第1条是很直接的推理。如果前提为真，结论也一定为真。或者换一种方式说，结论不真，前提也不能为真。逻辑学家把这样的推断称为**演绎有效的**（*deductively valid*）。第2条推理有所不同。前提明显对结论给出了好的理由，但结论并非确凿无疑。毕竟，琼斯完全有可能故意把手弄脏，好让别人认为他抽烟。因此该推断不是演绎有效的。这样的推断通常称为**归纳有效的**（*inductively valid*）。相比之下，第3条推断看上去按任何标准都无可救药。前提似乎完全没有为结论提供任何理由。它是无效的——不论在演绎上还是归纳上。事实上，既然大家都不是白痴，如果有人真的像这样给出理由，人们就会假定，还有一些无需告知的额外前提（也许有人通过厨房窗户把烟递给琼斯）。

归纳有效性是个非常重要的概念。我们无时不在进行归纳推理；比如，在设法解决汽车为何抛锚，人为何生病，或者谁是罪犯这些问题时。虚构的逻辑学家夏洛克·福尔摩斯就是归纳推理的大师。尽管如此，历史上更多的努力都投入到理解演绎有效性上——或许逻辑学家往往也是哲学家或数学家（在他们的研究中，演绎有效性至关重要），而不是医生或侦探。本书晚一点我们再回到归纳这个概念。现在，让我们对演绎有效性作更多思考。（认为演绎有效性是更简单的概念是很自然的，因为演绎有效的推断更为确定不变（cut-and-dried）。因此，试图先理解这个概念是个不错的主意。而我们会看到，即使做到这一点也并不容易。）除非特别说明，“有效”以后就指“演绎有效”。

那么什么是有效的推断呢？前面我们看到了，就是结论不真前提也不能为真的推断。但这是什么意思呢？特别的，“不能”是什么意思呢？一般而言，“不能”可以意指很多不同的事情。比如，考虑：“玛丽能弹钢琴，但约翰不能”；这里我们谈论的是人的能力。试比较：“你不能进入这里：你需要许可”；这里我们谈论的是，什么是某些规则所允许的。

这样来理解与当前问题相关的“不能”是自然的：说结论不真前提就不能为真，就是说，在所有前提为真的情形，结论都为真。目前为止，都很好；但到底什么是一个情形呢？它由什么样的事物组成，这些事物又如何相互关联呢？什么叫为真（*true*）呢？特伟弟可能会说，现在，你有个哲学问题要回答。

不久我们就要关心这些问题；但让我们暂时将其放在一边，先解决另一件事。不要轻易认为，我刚才给出的演绎有效性的解释是毫无问题的。（在哲学中，所有有趣的论断都是有争议的。）这里就有一个难题。假定这个解释是正确的，知道一个论断是演绎有效的，就是知道不存在前提为真而结论不为真的情形。现在，在什么是一个情形的任何合理的理解下，都会有多得可怕的情形：关于遥远恒星的行星上的事物的情形；关于宇宙有任何生命之前的事件的情形；虚构作品中描述的情形；梦想家想象的情形。人们如何才能知道在所有情形中哪些是成立的呢？更糟的是，似乎有无穷多的情形（一年后的情形，两年后的情形，三年后的情形，......）。所以即使在原则上，也不可能考察所有情形。这样一来，如果这种关于有效性的解释是正确的，而且假定我们能辨认推理是有效还是无效的（至少在很多情况下如此），我们就要有某种特别来源的洞见。什么来源呢？

我们需要借助某种神秘直觉吗？不一定。考虑一个类似的问题。我们都能区分母语中符合语法与不合语法的词语串，这没有太大问题。比如，任何母语为中文的人都知道，“这是一把椅子”是符合语法的，但“一把椅子是是一把”就不是。但符合语法和不合语法的句子似乎都有无穷多。（例如，“1 是一个数”，“2 是一个数”，“3 是一个数”，......都是符合语法的句子。我们也很容易造出杂乱无章的句子）。那么我们是如何做到这一点的呢？也许就像最有影响力的现代语言学家诺姆·乔姆斯基所认为的那样，我们能做这些是因为，无穷的句子由有穷的规则所囊括，而这些有穷的规则已经固化到我们大脑里；生物进化让我们天生具有语法知识。逻辑也是同样如此吗？逻辑规则也以同样的方式固化到我们大脑里了吗？

本章要点

- 一个有效的推断就是结论可以从前提推出的推断。
- 一个演绎有效的推断就是不存在前提为真而结论不为真的情形的推断。

第 2 章（上）：真值函数——亦或不是？

无论关于有效性的规则是否固化到我们大脑里，我们对于各种推断是否有效都有很强的直觉。比如，以下推断的有效性是没有多少争议的：“她是女人且是银行家；因此她是银行家”。以下推断是无效的：“他是木匠；因此他是木匠且打棒球”。

但我们的直觉有时会让我们陷入麻烦。你怎么看下面这个推断？横线上方是两个前提，横线

下方是结论。

```
女王富有。女王不富有。
————————————————————
      猪会飞。
```

它看上去当然不是有效的。女王的财富——无论多少——似乎都和猪的飞行能力无关。

但你认为以下两个推断怎么样？

```
    女王富有。
————————————————
女王富有或猪会飞。
```

```
女王富有或猪会飞。女王不富有。
——————————————————————————
          猪会飞
```

第 1 个推断似乎是有效的。考虑它的结论，逻辑学家称这样的语句为析取式（*disjunction*），称“或”两边的子句为析取项（*disjuncts*）。那么，一个析取式怎样为真呢？只要其中一个析取项为真即可。因此对于第 1 个推断，在任何前提为真的情形，结论也为真。第 2 个推断似乎也是有效的。如果两个论断中的一个或另一个为真，且其中一个不为真，那么另一个一定为真。

现在的麻烦是，把两个明显有效的推断串连在一起，我们会得到明显无效的推断，像这样：

```
    女王富有。
————————————————
女王富有或猪会飞。 女王不富有。
———————————————————————————
           猪会飞
```

这不可能是对的。将两个有效的推断以这种方式连接起来不可能得到一个无效的推断。（译者注：对于有效的论证）任何情形下如果所有前提为真，那么它们的所有结论也为真，由这些结论推得的结论也为真，如此下去，直到我们得到最终的结论。问题出在哪儿呢？

为了给这个问题一个正统的解答，让我们更详细地考察一下。首先，我们把语句“猪会飞”记作 p，把语句“女王富有”记作 q，这让事情更紧凑些，但不止于此：如果你思考一下就会发现，在这个例子中使用什么具体语句是无关紧要的，我完全可以使用任意两个语句来构造这种形式的推断，因此我们可以忽略其内容，这正是我们将语句记作单个字母时所做的。

语句“女王富有或猪会飞”现在就变成“ q 或 p ”，逻辑学家通常将其记作 $q \vee p$。“女王不富有”怎么记呢？让我们先把这句话改写为“并非女王富有”，将否定词挪到句子前面。这样，这句话就变成“并非 q”，逻辑学家通常将其记作 $\neg q$，称为 q 的否定（*negation*）。“女王富有且猪会飞”，也就是“ q 且 p ”这句话怎么记呢？逻辑学家通常将其记作 $q \wedge p$，称为 q 和 p 的合取

（*conjunction*）。有了这套改写机制，我们就可以把上面那个串连起来的推断写成如下形式：

$$\frac{\dfrac{q}{q \vee p} \qquad \neg q}{p}$$

关于这个推断我们有什么要说的呢？

语句可以为真，也可以为假。让我们用 T 表示真，用 F 表示假。自现代逻辑的奠基者之一，德国哲学家和数学家弗雷格之后，它们通常被称为**真值**（*truth values*）。给定任何语句 a，a 的真值与其否定 $\neg a$ 的真值之间有何联系呢？一个自然的回答是，如果一个为真，另一个就为假，反之亦然。这样，如果“女王富有”为真，则“女王不富有”就为假，反之亦然。我们可以将这一联系记录如下：

- $\neg a$ 具有真值 T，当且仅当 a 具有真值 F。
- $\neg a$ 具有真值 F，当且仅当 a 具有真值 T。

逻辑学家将其称为否定的**真值条件**（*truth conditions*）。如果我们假定每个语句或者为真或者为假，但不会既真又假，我们就可以用下面的表格来描绘这些条件，逻辑学家称之为**真值表**（*truth table*）：

a	$\neg a$
T	F
F	T

如果 a 具有它下面那一列的真值，$\neg a$ 就具有其右边对应的的真值。析取 $\vee$ 的真值条件呢？正如我已提到的，一个自然的假设是，一个析取式 $a \vee b$ 为真，若 a 和 b 中的其中一个（或两个都）为真，否则该析取式就为假。我们可以将这一联系记录在析取的真值条件中：

- $a \vee b$ 具有真值 T，当且仅当 a 和 b 中至少有一个具有真值 T。
- $a \vee b$ 具有真值 F，当且仅当 a 和 b 都具有真值 F。

这些条件可以用真值表描绘如下：

a	b	$a \vee b$
T	T	T
T	F	T
F	T	T
F	F	F

除了第 1 行的表头，每一行记录了一种 a（第 1 列） 与 b（第 2 列） 的真值的可能组合。共有 4 种这样可能的组合，因而有 4 行。对每一种组合，右边都给出了$a \vee b$ 对应的真值（第 3 列）。

同样，既然谈到这里，a 和 b 的真值与 $a \wedge b$ [2] 的真值之间有何联系呢？一个自然的假定是，$a \wedge b$ 为真，若 a 和 b 都为真，否则 $a \wedge b$ 为假。这样，比如，“约翰 35 岁且有棕色的头发”为真，当且仅当“约翰 35 岁”和“约翰有棕色的头发”都为真。我们可以将这一联系记录在合取的真值条件中：

- $a \wedge b$ 具有真值 T，当且仅当 a 和 b 都具有真值 T。
- $a \wedge b$ 具有真值 F，当且仅当 a 和 b 中至少有一个具有真值 F。

这些条件可以用真值表描绘如下：

a	b	$a \wedge b$
T	T	T
T	F	F
F	T	F
F	F	F

第 2 章（下）：真值函数——亦或不是？

那么，所有这些与我们一开始提出的问题有何关系呢？让我们回到我在上一章结束时提出的问题：什么是一个情形？一个自然的想法是，不管情形是什么，它确定了每个语句的真值。因此，比如，在某个特定情形中，女王富有为真而猪会飞为假。在另一个情形中，女王富有为假而猪会飞为真。（注意这些情形可以纯粹是假设性的！）换言之，一个情形确定了每个相关语句为真或为假。这里的相关语句不包含“且”、“或”或“非”的任何出现。给定关于某个情形的基本信息，包含这些词的语句的真值我们可以利用真值表算出。

例如，假设我们有下面的情形：

$p : T$
$q : F$
$r : T$

（r 可能是语句“大黄有营养”，“$p : T$”表示 p 被指派真值 T，等等。）那么，比方说 $p \wedge (\neg r \vee q)$ 的真值是多少呢？我们计算这一真值的方法，和用乘法表与加法表计算 $3 \times (-6 + 2)$ 的数值完全一样。于是，$\neg$ 的真值表告诉我们，$\neg r$ 的真值是 F。又由于 q 的真值也是 F，$\vee$ 的真值表告诉我们，$\neg r \vee q$ 的真值为 F。由于 p 的真值是 T（译者注：这里其实不需

要考虑这一条件），$\wedge$ 的真值表告诉我们，$p \wedge (\neg r \vee q)$ 的真值为 F。用这种逐步计算的方法，我们可以算出任何包含 $\wedge, \vee$ 和 $\neg$ 出现的公式的真值。

现在，回想一下，上一章我们说一个推断是有效的，只要不存在使得所有前提都为真而结论不为真（为假）的情形。也就是说，一个推断是有效的，只要不存在对相关语句的真值指派，使得所有前提的真值都为 T 而结论的真值为 F。比如，考虑我们前面见过的推断：$q/q \vee p$。（我把它写成一行是为了给牛津大学出版社省点钱。）这里的相关语句是 q 和 p。有 4 种真值组合，对每种组合我们都能算出前提和结论的真值。我们可以将结果表示如下：

q	p	q	$q \vee p$
T	T	T	T
T	F	T	T
F	T	F	T
F	F	F	F

前两列给出了 q 和 p 的真值的所有可能组合。后两列分别给出了前提和结论在每种组合下对应的真值。第 3 列与第 1 列相同，这是本例的一个巧合，之所以如此是因为，在这个特定的例子中，前提恰好是其中一个相关语句。第 4 列可以由析取的真值表读出。给定这些信息，我们就能看出这个推断是有效的，因为没有一行前提 q 为真而结论 $q \vee p$ 不为真。

推断 $q \vee p, \neg q/p$ 的有效性如何呢？用同样的方法，我们得到：

q	p	$q \vee p$	$\neg q$	p
T	T	T	F	T
T	F	T	F	F
F	T	T	T	T
F	F	F	T	F

这次有 5 列，因为有两个前提。前提和结论的真值可以由析取和否定的真值表读出。同样，没有一行两个前提都为真而结论不为真。因此，该推断是有效的。

我们一开始提到的那个推断 $q, \neg q/p$ 的有效性如何呢？用之前的方法，我们得到：

q	p	q	$\neg q$	p
T	T	T	F	T
T	F	T	F	F
F	T	F	T	T

F	F	F	T	F

同样，这个推断是有效的；而且现在我们明白了为什么它是有效的：没有一行两个前提都为真而结论为假。事实上，没有一行两个前提都为真。结论完全不起作用！有时逻辑学把推断的这种情形描述为空洞（*vacuously*）有效，只是因为前提永远也无法同时为真（译者注：所以无论结论是什么推断都有效）。

这就是我们最开始提出的问题的一种解决方案。根据这种解释，我们最初关于这个推断的直觉是错误的。毕竟，人们的直觉经常是误导性的。对每个人来说，地球似乎显然是不动的，直到他们学了一门物理课，才发现地球实际上在飞速穿越太空。我们甚至可以为我们的逻辑直觉为何出错提供解释。我们实际遇到的推断多数都不是空洞的那种。我们的直觉只在这类场合下得到发展，因而不是普遍适用的，正如你在学走路时养成的习惯（比如，不朝一边倾斜）在其他场合并不总是管用（比如当你学骑车时）。

在后面某一章里，我们会再回到这个问题。但让我们简要看一下我们所用逻辑工具的恰当性（adequacy），以结束本章。这里的情况并没有人们原本希望的那么简单。根据这种解释，语句 $\neg a$ 的真值完全由语句 a 的真值确定。类似的，语句 $a \vee b$ 和 $a \wedge b$ 的真值也完全由 a 和 b 的真值确定。逻辑学家称像这样作用的运算为**真值函数**（*truth functions*）。但有很好的理由让我们认为，汉语中出现的“或”和“且”并不是真值函数，至少不总是如此。例如，根据 $\wedge$ 的真值表，“a 且 b”与“b 且 a”总是有相同的真值，即它们在a 和 b 都为真时也都为真，否则都为假。但考虑这两句话：

1. 约翰撞了头，且跌倒了。
2. 约翰跌倒了，且撞了头。

第 1 句说的是，约翰撞了头，然后跌倒了。第 2 句说的是，约翰跌倒了，然后撞了头。很明显，当第二句为假时第一句也可以为真，反之亦然。 因此，不仅合取项的真值是重要的，哪一个引起了哪一个也是重要的。

“或”也有类似的问题。根据我们前面的解释，“a 或 b 为真”若 a 和 b 中的某一个为真。但假设有个朋友对你说：

你现在就来，或者我们会迟到；[3]

于是你来了。根据 $\vee$ 的真值表，这个析取式为真。但假如你发现你的朋友是在跟你开玩笑：你完全可以半小时后再出发也来得及。在这种情况下，你肯定会说你的朋友说谎了：他所说的是假的。同样，不仅析取项的真值是重要的，析取项之间某种联系的存在也是重要的。

这些问题我留给读者自己思考。我们已经考察的材料至少初步解释了某些逻辑工具是如何使用的。在后续章节里我们还会继续利用这些工具，除非有些章节中的观点明确将其推翻——有时会出现这种情况。

本章给出的逻辑工具只涉及某些类型的推断，还有很多其他类型的推断。我们只是刚开了个头。

本章要点

- 在一个情形中，一个唯一的真值（T 或 F）被指派给每个相关语句。
- $\neg a$ 为 T 当且仅当 a 为 F。
- $a \vee b$ 为真当且仅当 a 和 b 中至少一个为 T。
- $a \wedge b$ 为真当且仅当 a 和 b 两个都为真。

第 3 章：名称与量词：空无一物是某物吗?

上一章我们考察的推断包含“或”和“并非”这样的短语，它们添加到完整句上或者连接完整句后构成其他的完整句。但有许多推断看上去以完全不同的方式工作。比如，考虑下面这个推断：

```
Marcus给了我一本书。
————————————————
 有人给了我一本书。
```

前提和结论都没有一个部分能独自构成一个完整句。如果该推断是有效的，那么其有效性是源自完整句的**内部结构**。

传统语法告诉我们，最简单的完整句由一个**主语**和一个**谓语**构成。这样，考虑下面这些例子：

1. Marcus 看见了那头大象。
2. Annika 睡着了。
3. 有人打我。
4. 没有人来我的派对。

每句话的第一个词是句子的主语，告诉我们这句话是关于什么的；其余的是谓语，告诉我们关于主语说了什么。那么，这样的句子什么时候为真呢？拿第 2 个例子来说，它为真，如果主语“Annika”指称的对象具有谓语表达的性质，即睡着了。

都很好。但第 3 句话的主语指称什么呢？那个打我的人？但也许没有人打我。没有人说它是一个真的语句。第 4 句话的情况更糟。“没有人”指称谁呢？《镜中世界》里，恰在爱丽丝遭遇狮子和独角兽之前，她偶然碰见正在等信使的白方国王。（由于某种原因，当信使出现时，它看上去像一只兔子一样仓皇失措。）当国王遇见爱丽丝时，他说：

> “朝路上瞧瞧看，然后告诉我你是否看见了……[信使]。”
>
> “我看没人在路上”，爱丽丝说。
>
> “真希望我也有双这样的眼睛，”国王以烦躁不安的语气评论说，"能看见没人！还是在那么远的距离！啊哟，在这个距离这样的光线下，我最多只能看见真人。”

卡罗尔在这里开了个逻辑玩笑，正如他经常所为。当爱丽丝说她能看见没人时，她不是在说她能看见某个人——不管是不是真实的人。“没人”不指称某个人，或任何其他事物。

像“没人”、“有人”、“每人”这样的词，现代逻辑学家称之为**量词**（*quantifiers*），它们有别于像“Marcus”和“Annika”这样的名称。我们刚看到的是，即使量词和名称都能充当句子的主语，它们一定以相当不同的方式在起作用。那么，量词是如何工作的呢？

这里是一个标准的现代回答。一个情形由一集对象所装备。在上面这个例子里，相关对象是所有的人。我们关于这个情形的推理中出现的所有名称，都指称这集对象中的某一个。因此，如果我们把“Marcus”记作 m，m 指称的就是这些对象中的某一个。如果我们把“很高兴”记作 H，那么语句 mH 在该情形为真，当且仅当由 m 所指称的对象具有由 H 所表达的性质。（出于他们自己的乖张理由，逻辑学家通常颠倒次序，把这句话记作 Hm 而不是 mH。这只是约定问题。）

现在考虑语句“某人很高兴”。这句话在该情形为真，当且仅当在这集对象中有某个对象很高兴，即集合中某个对象，称它为 x，使得 x 高兴。让我们把“某个对象 x 使得”记作 $\exists x$，那么我们可以把这句话记作：“$\exists x$ x 很高兴”，或：$\exists x\ xH$——如果记得我们可以把“很高兴”记作 H。逻辑学家有时把 $\exists x$ 称作**特称量词**（*particular quantifier*）。

那么“每人都很高兴”呢？这句话在某个情形为真，若相关集合中的每个对象都很高兴，即，集合中的每个对象，x，使得 x 很高兴。如果我们把“每个对象 x 使得”记作 $\forall x$，那么我们可以把它记作 $\forall x\ xH$。逻辑学家通常把 $\forall x$ 称作**全称量词**（*universal quantifier*）。

现在不难猜出我们会如何理解“没人很高兴”了。这句话就是指，在相关集合里没有对象，x，使得 x 很高兴。我们可以用一个特殊符号意指“没有对象，x，使得”，但事实上，逻辑学家一般不这么做。因为说没人高兴就是说并非有人高兴。所以我们可以把它记作 $\neg\exists x\ xH$。

这种对量词的分析向我们表明，名称和量词起作用的方式是相当不同的。特别的，“Marcus很高兴”和“某人很高兴”被记作截然不同的 mH 和 $\exists x\ xH$，这一事实就表明了这一点。此外，它还向我们表明，表面简单的语法形式可能会产生误导。不是所有主语都是等同的。这种解释还顺便向我们解释了，为什么本章开头那个推断是有效的。让我们把“给了我那本书”记作 G，则这个推断就是：

$$\frac{mG}{\exists x\ xG}$$

很明显，如果在某个情形由名称 m 指称的对象给了我那本书，那么在相关集合中就有某个对象给了我那本书。作为对照，白方国王由爱丽丝看见没人，推出她看见某人（即“没人”）。如果我们把“被爱丽丝看见”记作 A，那么国王的推断就是：

$$\frac{\neg\exists x\ xA}{\exists x\ xA}$$

这明显是无效的。如果在相关论域里没有对象被爱丽丝看见，在相关论域里有某个对象被她看见就显然为假。

你可能会认为，这是毫无意义的小题大做——实际上，这种做法只是糟蹋了一个好笑话。但并非如此，它要严肃得多，因为量词在数学和哲学中的许多重要论证中都起着核心作用。下面是一个哲学论证的例子。一个很自然的假定是，没有什么事会毫无理由地发生：人们不会无缘无故生病，汽车不会没有故障就抛锚。于是，每件事都有原因。但引起每件事的原因可能是什么呢？显然不能是任何物理对象，比如某个人，甚或也不是宇宙大爆炸之类。这些东西本身也是有原因的。因此，它必定是某种形而上学的东西。上帝就是那个显而易见的候选者。

这是一种版本的证明上帝存在的论证，常被称作**宇宙论论证**（*Cosmological Argument*）。人们可能会以各种方式反对这个论证。但其核心是，该论证有一个巨大的逻辑谬误。语句“每件事都有原因”是有歧义的。它可以指，每件发生的事都有这样或那样的原因，即对任何 x，都有一个 y，使得 x 由 y 引起。它也可以指，有某个东西是所有事的原因，即有某个 y，使得对任何 x，x 由 y 引起。假设我们把相关论域的对象考虑为所有的原因和结果，把“x 由 y 引起”记作 xCy，那么我们可以把这两种意思分别写成：

1. $\forall x\, \exists y\, xCy$
2. $\exists y\, \forall x\, xCy$

现在，这两个不是逻辑等价的。第 1 个可以由第 2 个推出。如果有某个东西是**所有**事的原因，那么当然，每件发生的事都有**这样或那样**的原因。但如果每件事有这样或那样的原因，并不能推出有某个相同的东西是所有事的原因。（试比较：每个人都有母亲，这不能推出有人是所有人的母亲。）

这种版本的宇宙论论证利用了这一歧义。在谈论疾病和汽车时，所确立的是 1。但该论证马上就接着问那个原因是什么，此时假定 2 是已经确立的。而且，这一滑移是隐蔽的，因为在汉语中，“每件事都有原因”可以用来表达 1，也可以用来表达 2。还要注意，如果把量词换成名称就没有歧义了。“宇宙背景辐射是由大爆炸引起的”就毫无歧义。很可能，未能区分名称与量词是人们无法看出歧义的一个更深层的原因。

因此，正确理解量词是很重要的——不仅对逻辑学如此。“某物”、“空无一物”等等这样的词并不代表对象，而是以完全不同的方式起作用。亦或是，至少它们可以代表对象——事情不完全是那么简单。再考虑一下宇宙。要么它在时间上向过去无限延伸，要么它开始于某个特定的时间点。事实上，关于这件事的真相，物理学在不同时期告诉我们不同答案。不过，不用担心这一点，让我们仅仅考虑第二种可能性。在这种情况下，宇宙是从空无一物中形成的，或者至少是物理上的空无一物，因为不管怎样，宇宙就是所有物理对象的总和。现在考虑“宇宙由空无一物形成”这个句子。用 c 表示宇宙，将“x 由 y 形成”记作 xEy。那么，根据我们对量词的理解，这句话就应该意指 $\neg\exists x\, cEx$。但它不是这个意思，因为在第一种供选择的宇宙论解释下，这个公式同样为真。在这一解释下，因为宇宙在过去的时间上是无限的，就完全不会形成。特别的，它并非由某物形成。当我们说在第二种宇宙论下，宇宙由空无一物形成，我们是指它由“空无”这种东西形成。因此，空无一物可以是某个东西。白方国王毕竟也没有那么傻。

本章要点

- 语句 nP 在某个情形为真，若在该情形中由 n 指称的对象具有 P 表达的性质。
- $\exists x\, xP$ 在某个情形为真，当且仅当该情形中有某个对象 x 使得 xP 成立。

- $\forall x\, xP$ 在某个情形为真，当且仅当该情形中的每个对象 x 都有 xP 成立。

第 4 章（上）：摹状词与存在：古希腊人崇拜宙斯吗?

当我们讨论主谓语这一话题时，有一种特定的可以充当句子主语的短语，我们还没有论及。逻辑学家称之为**限定摹状词**（*definite descriptions*），有时简称为**摹状词**（*descriptions*）——要注意，这是一个专业术语。摹状词就是像“第一个登上月球的人”和“从太空中唯一可见的地球人造物”这样的短语。一般而言，摹状词具有这样的形式：**那个满足某某条件的东西**（*the thing satisfying such and such a condition*）。按照英国哲学家和数学家，现代逻辑的奠基人之一，伯特兰·罗素的做法，我们可以将摹状词写成如下形式。将“第一个登上月球的人”重写为“那个对象 x，使得 x 是人且 x 第一个登上月球。”现在将“那个对象 x，使得”记作 ιx”，那它就变成“$\iota x(x$是人且 x 第一个登上月球)”。如果我们把“是人”记作 M，把“第一个登上月球”记作 F，那么我们就得到：$\iota x(xM \wedge xF)$。一般而言，一个摹状词就是某种形如 $\iota x c_x$ 的东西，其中 c_x 是某个包含 x 的出现的条件。（这就是下标 x 在那里要提醒你的。）

由于摹状词是主语，它们可以与谓语结合构成完整句。因此，如果我们把“出生于美国”记作 U，那么“第一个登上月球的人出生于美国”就是：$\iota x(xM \wedge xF)U$。让我们将 $\iota x(xM \wedge xF)$ 缩写为 μ。（我用希腊字母提醒你，它的确是一个摹状词）。这样，这句话就是 μU。类似的，“第一个登上月球的人是人且他第一个登上月球”就是 $\mu M \wedge \mu F$。

依据上一章的区分，摹状词是名称，不是量词。即，它们指称对象——如果我们幸运的话：这一点我们回头再谈。这样，“第一个登上月球的人出生于美国”，μU，为真，当且仅当，由短语 μ 指称的那个特定的人，具有由 U 表达的性质。

但摹状词是一类特殊的名称。不像**专名**（*proper names*）（如“Annika”和“大爆炸”），摹状词携带着关于所指对象的信息。比如，“第一个登上月球的人”携带的信息是，所指对象具有是人和第一个登上月球这些性质。这一点似乎平凡无趣，显而易见。然而，事情不像它看上去那么简单。由于摹状词以这种方式携带信息，它们经常在数学和哲学的重要论证中发挥核心作用。一种理解其中某些复杂性的方法，就是来看一个这种论证的例子。这是另一个证明上帝存在的论证，常被称作**本体论论证**（*Ontological Argument*）。该论证有多种版本，这是其中一个简单版本：

> 上帝是拥有所有至善之物。
>
> 而存在性也是一种至善。
>
> 因此，上帝具有存在性。

即，上帝存在。如果你以前没有见过这个论证的话，它会显得相当费解。首先，什么是至善？宽泛地说，至善就是全知（知道一切可知的之事）、全能（能做一切可做之事）和道德完美（总是以尽可能最好的方式行事）之类的东西。更一般的，至善就是一个美好事物所具有的全部性质。现在，第二个前提说，存在性是一种至善。究竟为何如此呢？理由相当复杂，其哲学思想

渊源可追溯到古希腊两个最有影响力的哲学家之一，柏拉图。幸运的是，我们可以解决这个问题。我们可以列一个清单，清单上包含像全知、全能等这样的性质，将存在性也包括在清单上，然后让“至善”就指清单上的任何一个性质。此外，我们可以令“上帝”与某个特定的摹状词同义，即“那个拥有所有至善（即，清单上的所有性质）之物”。现在，根据定义，本体论论证的两个前提都为真，因而可以不用考虑。这个论证可以简化为一句俏皮话：

> 那个全知、全能、道德完美……且存在的对象，存在。

——我们还可以加上，是全知、全能、道德完美的，如此等等。这句话看上去当然为真。为了让事情更加明晰，假设我们把上帝具有的清单上的性质记为 $P_1, P_2, \ldots, P_n$，其中最后一个，P_n， 是存在性。“上帝”的定义是：$\iota x(xP_1 \wedge \ldots \wedge xP_n)$。让我们把它记作 γ。这样，那句俏皮话就是 $\gamma P_1 \wedge \ldots \wedge \gamma P_n$（由此可以推出$\gamma P_n$）。

这是一个更一般结论的特例：**那个满足某某条件的东西满足恰好那个条件**。这常被称作刻画原则（*Characterization Principle*）（事物具有刻画它的那些性质）。我们将该原则缩写为 CP。我们已经见过 CP 的一个例子：“第一个登上月球的人是人且他第一个登上月球”，$\mu M \wedge \mu F$。更一般的，如果我们取某个摹状词 ιxc_x，然后把条件 c_x 中 x 的每一次出现都替换为该摹状词，我们就得到了 CP 的一个特例。

第 4 章（下）：摹状词与存在：古希腊人崇拜宙斯吗？

现在，在各方面，CP 据定义看上去都是真的。事物当然具有被刻画为具有的那些性质。不幸的是，一般而言，它是假的，因为由它推出的许多结论都是毫无争议不为真的。

首先，我们可以用 CP 推导出所有各种实际上不存在的事物的存在性。考虑（非负）整数：$0, 1, 2, 3, \ldots$，没有最大的整数。但使用 CP，我们可以表明最大整数是存在的。令 c_x 为条件“x 是最大的整数且 x 存在”。令 δ 为 ιxc_x。那么， CP 让我们得到“δ 是最大整数且 δ 存在”。荒谬还不止于此。考虑某个未婚的人，比方说罗马教皇。令 c_x 为条件“x 嫁给了罗马教皇”。令 δ 为摹状词 ιxc_x。CP 让我们得到“δ 嫁给了罗马教皇”。因此，某人嫁给了罗马教皇，即，罗马教皇是结了婚的。

关于所有这些有什么要说的呢？下面是一个相当标准的现代回答。考虑摹状词 ιxc_x。如果在某个情形存在唯一对象满足条件 c_x，那么该摹状词就指称它。否则，该摹状词不指称任何事物：它是一个“空名”。比如，存在唯一对象 x，使得 x 是人且 x 第一个登上月球，即阿姆斯特朗。因此，“那个使得 x 是人且 x 第一个登上月球的 x”指称阿姆斯特朗。类似的，存在唯一最小（非负）整数，即 0，因此，摹状词“那个是最小（非负）整数的对象”指的就是 0。但由于不存在最大整数，“那个是最大整数的对象”就不指称任何事物。同样，摹状词“那个人口超过一百万的澳大利亚城市”也不指称任何事物，这次不是因为没有这样的城市，而是因为有好几个这样的城市。

这和 CP 有什么关系呢？如果在某个情形下存在唯一对象满足 c_x，那么 $\iota x c_x$ 便指称它。因此，CP 关于 c_x 的实例就是真的：$\iota x c_x$ 就是其中一个——事实上是唯一的一个——满足 c_x 的事物。特别的，那个最小（非负）整数，（的确）是最小的（非负）整数；那个是澳大利亚首都的城市，的确是澳大利亚的首都，等等。因此，CP 的某些实例是成立的。

但如果没有唯一对象满足 c_x 呢？如果 n 是一个名称，P 是一个谓词，那么语句 nP 为真，当且仅当存在一个 n 指称的对象，且它具有 P 表达的性质。因此，如果 n 不指任何对象，nP 一定为假。这样，如果没有唯一对象具有性质 P，（比如，若 P 是“是有翅膀的马”）$(\iota x\ xP)P$ 就为假。正如所料，在这些条件下，CP 会不成立。

那么，所有这些与那个本体论论证有何关系呢？回想一下，那里由 CP 导出的实例是 $\gamma P_1 \wedge \ldots \wedge \gamma P_n$，其中 γ 是摹状词 $\iota x(xP_1 \wedge \ldots \wedge xP_n)$。要么存在某物满足 $xP_1 \wedge \ldots \wedge xP_n$，要么不存在。如果存在，它一定是唯一的。（不可能有两个全能的对象：如果我是全能的，我就能阻止你做事，因此你就不能是全能的。）因此，γ 指称此物，且 $\gamma P_1 \wedge \ldots \wedge \gamma P_n$ 为真。如果不存在，则 γ 不指称任何事物；因此，$\gamma P_1 \wedge \ldots \wedge \gamma P_n$ 的每个合取项均为假；因而整个合取式也为假。换言之，如果上帝存在，那么该论证使用的 CP 实例就足以为真；但如果上帝不存在，它就为假。因此，如果一个人要论证上帝的存在，就不能仅仅调用这个 CP 的实例，那恰恰是在假定自己要证明的东西。哲学家称这样的论证为乞题（*begs the question*），即在论证时求助于恰好有待讨论的结论被承认。一个乞题的论证显然是不起作用的。

关于本体论论证就说到这里。让我们这样结束本章：我们会看到，在某些方面，我所讲解的关于摹状词的解释，本身是有问题的。根据这一解释，如果语句 δP 中的摹状词 δ 不指称任何事物，那么该语句就为假。但这似乎并不总是对的。比如，以下这些似乎为真：古希腊诸神中最强大的神被称作“宙斯”，他住在奥林匹斯山，被古希腊人所崇拜，等等。但实际上，没有古希腊诸神。他们实际上并不存在。如果这是对的，那么摹状词“那个古希腊诸神中最强大的”就不指称任何事物。但假如那样的话，就有真的主/谓句，其中主语词项不指称任何事物，如“古希腊诸神中最强大的神被古希腊人所崇拜”。说得更强烈一点，毕竟有关于不存在对象的真陈述。

本章要点

- $\iota x c_x P$ 在某个情形下为真，当且仅当在该情形中，存在唯一对象 a 满足 c_x 且 aP。

第 5 章：自指：本章是关于什么的？

常常，当人们思考寻常情形时，事情似乎很简单。但这可能是有迷惑性的。当人们考虑更不寻常的情形时，这种简单性就会完全消失。指称问题就是如此。上一章我们看到，一旦人们考虑到有些名称可以不指称任何事物这一事实，事情就不像人们原本想的那么简单了。当我们考虑另一种不寻常的情形——自指时，进一步的复杂性就出现了。

一个名称完全有可能指称包含它自身的对象。比如，考虑“本语句有七个字”这句话，作为主语的名称“本语句”指称整个语句，而该名称是整个语句的一部分。类似的情况也发生在一组规章里，其中有这样的条款：“这些规章可以经哲学系多数人的决定予以修订”，或者当一个人在思考下面这句话时：“如果我在思考本思考，我就必定是有意识的”。

这些都是相对没有问题的自指。还有些情况就大不相同了。比如，假设某人说：

> 我正在说的这句话是假的。

称这句话为 λ。λ 是真是假呢？如果它为真，那么它所说的就是实际情况，因此 λ 为假。但如果它为假，由于这恰好就是它所声称的，它就为真。不管哪种情况，λ 似乎既真又假。该语句就像一条莫比乌斯带，这种拓扑结构由于一个扭转，使得带子的内部就是外部，外部就是内部，而在这里，真就是假，假就是真。

或者假设某人说：

> 我正在说的这句话是真的。

它是真是假呢？如果它为真，它就为真，因为这就是它所说的。如果它为假，它就为假，因为它说自己为真。因此，假定它为真和假定它为假似乎都是一致的。此外，似乎没有其他事实可以解决其真值问题。并不是它有某个我们不知道，甚或无法知道的值，而是似乎完全没有什么东西能确定其为真或为假。它似乎既不真又不假。

这两个悖论非常古老。其中第一个似乎是由古希腊哲学家欧布里德首先发现的，常被称作**说谎者悖论**（*liar paradox*）。近来有越来越多同类型的悖论出现，其中一些在数学推理的核心部分起着至关重要的作用。这里是另一个例子。一个集合就是一组对象的聚集。比如，我们有所有人构成的集合，所有数构成的集合，所有抽象观念构成的集合。集合也可以是其他集合的成员[4]。比如，一间房子里的所有人构成的集合，就是由所有集合构成的集合的一个成员。有些集合甚至可以是自身的成员：本页提到的所有对象构成的集合，就是本页提到的一个对象（我刚刚提到），因此是它自身的成员。由所有对象构成的集合，是一个集合，因此是它自身的成员。还有些集合无疑不是它们自身的成员：所有人构成的集合并不是人，因此它不是所有人构成的集合的成员。

现在，考虑由所有不是自身成员的集合构成的集合。称这个集合为 R。R 是不是自身的成员呢？如果它是自身的成员，那它就是那些不是自身成员的对象中的一个，因此它就不是自身的成员。另一方面，如果它不是自身的成员，那它就是那些不是自身成员的集合中的一个，因此它就是自身的成员。R 似乎既是又不是自身的成员。

这个悖论是由伯特兰·罗素发现的——上一章我们已经提到他，因而该悖论被称作**罗素悖论**（*Russell's paradox*）。同说谎者悖论一样，它也有一个表亲。考虑所有是自身成员的集合构成的集合会怎么样呢？它是否是自身的成员呢？如果它是，它就是；如果它不是，它就不是。同样，似乎没有任何东西能确定它是还是不是。

这类例子所做的，是在挑战我们在第2章所做的假设，即每个语句或者为真，或者为假，但不会既真又假。“本语句为假”和“R 不是自身成员”似乎既真又假，而它们的表亲似乎既不真又不假。

怎么才能容纳这一观念呢？将这些其他的可能性考虑进来即可。假设在任何情形下，每个语句只真不假，只假不真，既真又假，或者既不真又不假。回想一下第 2 章关于否定、合取和析取的真值条件。在任何情形下：

- $\neg a$ 具有真值 T，当且仅当 a 具有真值 F。
- $\neg a$ 具有真值 F，当且仅当 a 具有真值 T。
- $a \wedge b$ 具有真值 T，当且仅当 a 和 b 都具有真值 T。
- $a \wedge b$ 具有真值 F，当且仅当 a 和 b 中至少有一个具有真值 F。
- $a \vee b$ 具有真值 T，当且仅当 a 和 b 中至少有一个具有真值 T。
- $a \vee b$ 具有真值 F，当且仅当 a 和 b 都具有真值 F。

利用这些信息，不难算出新系统下语句的真值。例如：

- 假设 a 只为 F 不为 T。那么，由于 a 为 F，$\neg a$ 就为 T（据否定的第一个条件）。又由于 a 不为 T，$\neg a$ 不为 F（据否定的第二个条件）。因此，$\neg a$ 只为 T 不为 F。
- 假设 a 为 T 和 F，b 只为 T。那么，a 和 b 都为 T，所以 $a \wedge b$ 为 T（据合取的第一个条件）。但，因为 a 为 F，故 a 和 b 中至少一个为 F，所以 $a \wedge b$ 为 F（据合取的第二个条件）。因此，$a \wedge b$ 既为 T 也为 F。
- 假设 a 只为 T，b 既不为 T 也不为 F。那么，由于 a 为 T，故 a 和 b 中至少有一个为 T，因此 $a \vee b$ 为 T（据析取的第一个条件）。但由于 a 不为 F，故并非 a 和 b 都为 F，所以 $a \vee b$ 不为 F（据析取的第二个条件）。因此，$a \vee b$ 只为 T。

关于有效性，这告诉我们什么呢？一个有效的论证仍然是没有情形使得其前提为真而结论不为真的论证。一个情形仍然是对每个相关语句给出真值的东西。只是现在，这个情形可以对一个语句给出一个真值，两个真值或没有真值。那么，考虑推断 $q/q \vee p$。在任何 q 具有真值 T 的情形中，$\vee$ 的真值条件向我们保证了 $q \vee p$ 也具有真值 T。（它也可能还有真值 F，但没关系。）这样，如果前提具有真值 T，结论也具有真值 T。该推断是有效的。

此刻，有必要回到我们第 2 章开头那个推断：$q, \neg q/p$。我们在那一章已看到，在那里给出的假设下，该推断是有效的。但在新的假设下，情况就不同了。要知道为什么，只需考虑一个 q 具有真值 T 和 F，而 p 只具有真值 F 的情形。由于 q 既为 T 又为 F，$\neg q$ 也既为 T 又为 F。因此，两个前提都为 T（也都为 F，但那不相干），而结论 p 不为 T。这给了我们另一个诊断，解释了为什么我们发现这个推断直觉上是无效的。它确实是无效的。

不过，这还不是本章的结束。我们在第 2 章看到，这个推断可以由其他两个推断推出。第一个（$q/q \vee p$）我们已经看到，在目前的解释下是有效的。所以另一个一定是无效的，它确实如此。另一个推断是：

$$\frac{q \vee p, \neg q}{p}$$

现在考虑一个 q 具有真值 T 和 F 而 p 只具有真值 F 的情形。很容易验证两个前提都有真值 T（也有 F），但结论没有真值 T。因此，该推断是无效的。

在第 2 章我说过，这个推断在直觉上确乎是有效的。因此，在新的解释下，我们关于它的直觉一定是错的。然而，我们可以对这一事实提供一种解释。该推断看上去有效是因为，若 $\neg q$ 为真，则它似乎排除了 q 为真，从而留下 p 给我们。但根据现在的解释，$\neg q$ 的真并不排除 q 的真。只有在某个东西不能既真又假时，才是如此。当我们认为该推断有效时，我们或许忘记了这种可能性，而它会在自指这种不寻常的情形下出现。

哪种解释更好呢？是我们在第 2 章结束时给出的解释，还是我们现在给出的？这个问题我留给读者自己思考。作为替代，我们以提出如下问题结束本章：像往常一样，人们可以挑战这个新解释所依赖的某些观念。考虑说谎者悖论及其表亲。先考虑后者。“本语句为真”这个句子被认为是既不真又不假的例子。让我们假设它就是如此。

那么，特别的，它不是真的。但它本身说的是，它是真的。所以，它一定是假的，与我们开始假设它既不真又不假相反。我们似乎最终陷入了矛盾。或者考虑说谎者语句：“本语句为假”。它被认为是一个句子既真又假的例子。让我们稍加变形，考虑“本语句不为真”这句话。它的真值是什么呢？如果它为真，那么它所说的就是实际情况，所以它就不为真。但如果它不是真的，那么由于这就是它所说的，它就为真。同样，我们得到了一个矛盾。矛盾不只是一个句子取值可以既是 T 又是 F，而是，一个句子可以既是 T 又不是 T。

正是这类情形，使得自指是自欧布里德以来一直引起争论的主题。它的确是一个非常棘手的问题。

本章要点

- 句子可以为真，为假，既真又假，或既不真又不假。

第 6 章：必然与可能：什么会是一定如此的？

我们不仅经常声称某事如此，而且还说它一定如此。我们说：“一定会下雨”，“不可能不下雨”，“必然地，会下雨”。我们也有很多不同方式说，可能是，尽管有时实际并非如此。我们说：“明天可能会下雨”，“明天下雨是可能的”，“明天下雨不是不可能的”。若 a 是任一语句，逻辑学家通常将论断“a 一定为真” 记作 $\Box a$，将论断“a 可能为真”记作 $\Diamond a$。

$\Box$ 和 $\Diamond$ 称作模态算子（*modal operators*），因为它们表达了事情为真或为假（必然地，可能地）的模式。这两个算子其实是相互联系的。说某事一定如此，就是说它不可能不如此。即，$\Box a$ 和 $\neg\Diamond\neg a$ 有相同的意思。类似的，说某事可能如此，就是说它不必然为假。即，$\Diamond a$ 和 $\neg\Box\neg a$ 有相同的意思。另外，我们可以将不可能 a 为真这一事实，无差别地表达为 $\neg\Diamond a$（并非 a 是可能的），或者表达为 $\Box\neg a$（a 必然为假）。

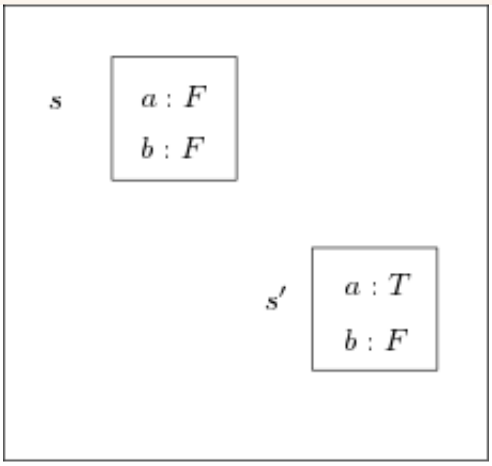
与我们见过的算子不同，$\Box$ 和 $\Diamond$ 不是真值函数。我们在第2章看到，当你知道 a 的真值，你就可以算出 $\neg a$ 的真值。类似的，当你知道 a 和 b 的真值，你就能算出 $a \vee b$ 和 $a \wedge b$ 的真值。但你不能仅由知道 a 的真值就推断出 $\Diamond a$ 的真值。比如，令 r 为语句“我会在明天上午7点前起床”。r 实际上为假。但它当然可以为真：我可以定好闹钟早点起床。因此，$\Diamond r$ 为真。作为对比，令 j 为语句“我会从床上跳起并且悬浮在离地2米的空中”，同 r 一样，j 也为假。但与 r 不同，它甚至不可能为真。那将违反万有引力定理。因此，$\Diamond j$ 为假。所以，语句 a 的真值，并不决定 $\Diamond a$ 的真值：r 和 j 都为假，但 $\Diamond r$ 为真而 $\Diamond j$ 为假。同样，a [5]的真值也不决定 $\Box a$ 的真值。现在令 r 为语句“我会在明天上午8点前起床”。这事实上是真的，但并不必然为真。我可以赖在床上不起来。现在令 j 为“如果我明早从床上跳起，我就会发生移动”。这句话也为真，但不可能为假。它必然为真。因此，r 和 j 都为真，但一个必然为真，另一个不是。

因此，模态算子是一类与我们见过的完全不同的算子。它们也是非常重要且常常令人困惑的算子。为了表明这一点，这有一个宿命论论证，它由古希腊两大哲学家中的另一位，亚里士多德提出。

宿命论是这样一种观点，任何发生的事情都必定会发生，它不可能避免。当一个事故出现，或一个人死了，没有什么可以阻止它发生。宿命论是一种对某些人很有吸引力的观点。当事情出现差错时，人们便可以从中获得些许安慰：事情不可能是其他样子的。但是，宿命论蕴含我无力改变什么会发生，而这显然是假的。如果我今天遇上车祸，我完全可以通过选择不同路线来避免它。[6] 那么，亚里士多德是如何论证的呢？它是这样进行的。（暂时先忽略粗体部分，我们等会儿再讨论它。）

为了举例说明，考虑任何你喜欢的论断，比如，你明天会遇上车祸。现在，我们也许不知道它是否为真。但我们知道，我明天要么遇上车祸，要么不会。假设第一种情况为真，那么，事实上我会遇上车祸。**如果说我会遇上车祸是真的，那么我遇上车祸就不能不是如此**。即，我遇上车祸是必定如此的。另一方面，假设我事实上明天不会遇上车祸，那么，说我不会遇上车祸就为真。如果是这样，我不会遇上车祸就不可能不如此。即，我不会遇上车祸是必定如此的。不管这两种情况哪一个确实发生，它一定发生。这就是宿命论。

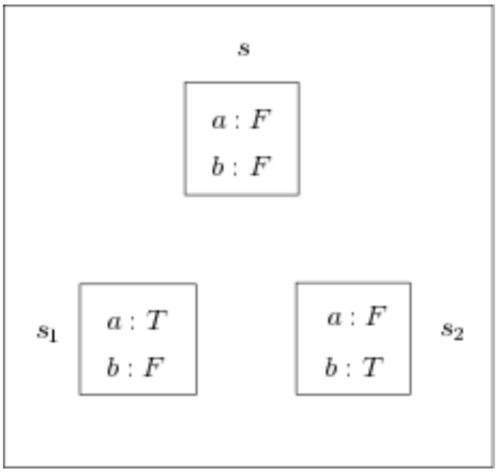
关于这个论证有什么可说的呢？为了回答这个问题，让我们看一下对模态算子的一种标准的现代理解。我们假设每个情形，s，都由一组可能性所装备，即相对于 s 可能的那些情形——为了确定起见，让我们假设是不违反物理定律的情形。这样，如果 s 是我现在所处（在澳大利亚）的情形，我一周内在伦敦就是一个可能的情形；而我一周内在半人马座阿尔法星（超过4光年之遥）就不是一个可能的情形。沿用17世纪哲学家和逻辑学家莱布尼茨的用法，逻辑学家常把这些可能的情形形象地称为**可能世界**（*possible world*）。现在，说 $\Diamond a$（a 是可能的）在 s 中为真，就是说 a 在至少一个与 s 关联的可能世界中实际为真。说 $\Box a$（a 是必然的）在 s 中为真，就是说 a 在所有与 s 关联的可能世界中都为真。这就是为什么 $\Box$ 和 $\Diamond$ 不是真值函数。因为 a 和 b 可能在 s 中有相同的真值，比方说 F，但可能在与 s 关联的世界中有不同真值。比如，a 在其中一个为真（比方说 s'），而 b 在其中一个也不真，如图：



这种解释给了我们一种对含模态算子的推断进行分析的方法。比如，考虑推断：

$$\frac{\Diamond a \quad \Diamond b}{\Diamond(a \wedge b)}$$

它是无效的。要知道为什么，假设与 s 关联的情形是 s_1 和 s_2，且真值指派如下：



a 在 s_1 为 T，因此，$\Diamond a$ 在 s 中为真。类似的，b 在 s_2 为 T，因此，$\Diamond b$ 在 s 中为真。但没有一个关联世界中 $a \wedge b$ 为真，因此 $\Diamond(a \wedge b)$ 在 s 中不为真。

相比之下，下面的推断是有效的：

$$\frac{\Box a \quad \Box b}{\Box(a \wedge b)}$$

因为如果前提在某个情形 s 中为真，那么 a 和 b 在所有与 s 关联的世界中都为真。这样 $a \wedge b$ 在所有这些世界中也为真，即 $\Box(a \wedge b)$ 在 s 中为真。

在回到这与亚里士多德的论证有何关联这个问题之前，我们需要简要谈一下另一个我们还没见过的逻辑算子。让我们把"如果 a 那么 b"记作 $a \to b$。这种形式的语句称为**条件句**（*conditionals*），我们在下一章会专门讨论。目前我们只需要知道，条件句涉及的主要推断如下：

$$\frac{a \quad a \to b}{b}$$

（比如，“如果她定期健身，她就会身材好。她确实定期健身；因此，她身材好。”现代逻辑学家通常用中世纪逻辑学家标记它的名称 modus ponens，来称呼这种推断。它的字面意思是“假设的方法”。[7]（别问我为什么。）

现在，对于亚里士多德的论证，我们需要稍微考虑一下这种形式的条件句：

> 如果 a，那么不可能不是 b。

这样的句子实际上是有歧义的。一个意思是，如果 a 实际上为真，那么 b 必然为真。即，如果 a 在我们谈论的情形 s 中为真，那么 b 在所有与 s 关联的情形中都为真。我们可以把它记作 $a \to \Box b$。在说下面这样的话时，我们就是在这个意思上使用该语句：“你无法改变过去。如果某事在过去为真，它现在就不可能不为真。你没有办法让它变成其他样子：它是不可改变的。”

形如“如果 a，那么不可能不是 b”的条件句的第二个意思完全不同。我们经常用这种形式的文字来表达 b 可以由 a 推出这一事实。在说下面这样的话时，我们就是在这个意思上使用该语句：“如果弗雷德打算离婚，那么他不可能不是结了婚的。”我们不是在说，如果弗雷德打算离婚，那么他结了婚这件事是不可改变的。我们是在说，除非你结了婚，否则你不可能离婚。没有一个可能情形使得一个成立而另一个不成立。也就是说，在任何可能的情形中，只要一个为真，另一个也要为真。即，$\Box(a \to b)$ 为真。

现在，$a \to \Box b$ 和 $\Box(a \to b)$ 意指完全不同的东西。而且第一个无疑不能从第二个推出。仅 $a \to b$ 在每个与 s 关联的情形都为真这一事实，并不意味着 $a \to \Box b$ 在 s 中为真。a 可以在 s 中为真，而 $\Box b$ 不为真：b 和 a 在某些关联世界中可以都不为真。或者，给一个具体的反例吧：如果约翰正在离婚，那么他是结了婚的，这是必然真的。但，如果约翰正在离婚，那么他是必然（不可改变地）结了婚的，这无疑不为真。

最后回到亚里士多德的论证。考虑我前面加粗的句子：“如果说我会遇上车祸是真的，那么我遇上车祸就不能不是如此”。这恰好是我们刚讨论过的形式。所以它是有歧义的。而且，这个论证利用了这种歧义。如果 a 是语句“说我会遇上车祸是真的”，b 是语句“我会遇上车祸”，那么粗体的条件句在下面的意义上为真：

1. $\Box(a \to b)$

必然地，如果说某事为真，那么它的确如此。但需要确立的是：

2. $a \to \Box b$

毕竟，该论证的下一步正好是根据分离规则（*modus ponens*）[8] 由 a 得到 $\Box b$。但我们已经看到，2 根本不能由 1 推出。因此，亚里士多德的论证是无效的。另外，同样的问题出现在该论证的第二部分，它使用了条件句“如果说我不会遇上车祸为真，那么我不遇上车祸就不可能不如此”。

这似乎是对亚里士多德论证的一个令人满意的回答。但有一个与之密切相关的论证却无法这么轻易得到回答。确乎为真的是，如果某个关于过去的陈述为真，那么它现在必然为真。不可能在现在使得其为假。黑斯廷斯之战发生在 1066 年，现在没有什么办法可以让它发生在 1067 年。因此，如果 p 是一个关于过去的陈述，那么 $p \to \Box p$。

现在考虑某个关于未来的陈述。比方说还是这个论断：我会在明天遇上车祸。假设它为真。那么如果有人在 100 年前说了这句话，他们说的就是真的。即使实际上没人说过这句话，只要有人真的说过，他们说的就是真的。这样，我会在明天遇上车祸就在 100 年前为真。这一陈述（p）当然是关于过去的陈述，因而，既然它为真，它必然为真（$\Box p$）。因此，我明天遇上车祸就一定必然为真。但这只是一个例子，同样的论证可以应用于任何事。这样，任何发生之事，必然发生。这个宿命论论证并没有犯和我前面给出的第一个论证相同的谬误（即，使用同样的无效论证）。所以，宿命论终究是真的吗?

本章要点

- 每个情形都有一集与之关联的可能情形。
- $\Box a$ 在某个情形 s 中为真，若 a 在每个与 s 关联的情形中为真。
- $\Diamond a$ 在某个情形 s 中为真，若 a 在某个与 s 关联的情形中为真。

第 7 章：条件句："如果"中有什么？

本章我们将讨论在上一章附带引入的逻辑算子——条件句。回想一下，条件句就是形如"如果 a 那么 c"的语句，我们现在记作 $a \to c$。逻辑学家称 a 为该条件句的前件（*antecedent*），c 为该条件句的后件（*consequent*）。我们还提到，关于条件句的一个最基本的推断是分离规则：$a, a \to c / c$。条件句是我们许多推理的重要基础。上一章就举过一个这方面的例子。不过，条件句却十分令人困惑。它们自逻辑学最早期开始就得到了研究。事实上，据一位古代评论家（Callimachus）说，甚至屋顶上的乌鸦也曾叫嚷着条件句。

让我们看看为什么——或者至少一个理由为什么——条件句是令人困惑的。如果你知道 $a \to c$，似乎你就能推出 $\neg(a \wedge \neg c)$（并非 a 且 $\neg c$）。比如，假设有人通知你，如果你错过巴士，你就会迟到。你就能从中推出，你错过巴士而又不迟到为假。反过来，如果你知道 $\neg(a \wedge \neg c)$，似乎你就能从中推出 $a \to c$。比如，假设有人告诉你，你不会去看电影而不花钱（并非你去看电影且不花钱）。你就能推出，如果你去看电影，你就要花钱。

$\neg(a \wedge \neg c)$ 常被记作 $a \supset c$，称为实质条件句（*material conditional*）。这样，似乎 $a \to c$ 和 $a \supset c$ 意指同样的事。特别的，使用第 2 章的逻辑工具，它们有相同的真值表。作为一个简单练习，我请读者自行验证，该真值表如下：

a	c	$a \supset c$
T	T	T
T	F	F
F	T	T
F	F	T

但这很奇怪。它意味着若 c 在某个情形为真（第 1 和第 3 行），则 $a \to c$ 也为真。这不可能是对的。比如，堪培拉是澳大利亚首都，这是真的，但条件句“如果堪培拉不是澳大利亚首都，那么堪培拉是澳大利亚首都”似乎明显为假。类似的，该真值表向我们表明，若 a 为假（第 3 和第 4 行），则 $a \to c$ 为真。但这也不可能是对的。条件句“如果悉尼是澳大利亚首都，那么布里斯班是澳大利亚首都”似乎也明显为假。哪里出问题了呢？

这些例子似乎表明，$\to$ 不是真值函数：$a \to c$ 的真值不由 a 和 c 的真值确定。“罗马在法国”和“北京在法国”都为假，但下面这句话为真：

> 如果意大利是法国的一部分，那么罗马在法国。

而下面这句话为假：

> 如果意大利是法国的一部分，那么北京在法国。

那么，条件句是如何工作的呢？

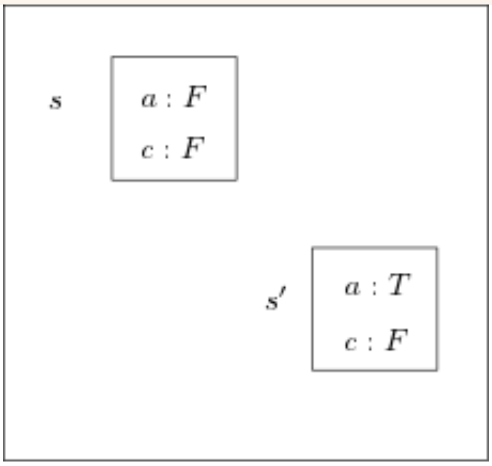
一种回答是使用上一章给出的可能世界工具。考虑上面两个条件句。在任何意大利真的被并入法国的可能情形中，罗马的确会在法国。但存在意大利被并入法国的可能情形，而这对中国毫无影响。因此，北京仍然不在法国。这提示我们，条件句 $a \to c$ 在某个情形 s 中为真，只要 c 在所有与 s 关联且 a 为真的情形中都为真；它为假，若 c 在某个与 s 关联且 a 为真的情形中为假。

这给了我们 $\to$ 一个看上去合理的解释。例如，它表明了为什么分离规则是有效的——至少在某个假设下。该假设是，我们把 s 自身算作一个与 s 相关的情形。这似乎是合理的，s 中任何实际的情况当然是可能的情况。现在，假设 a 和 $a \to c$ 在某个情形 s 为真。那么，c 在所有与 s 关联且 a 为真的情形中为真。但 s 就是这样的一个情形（与 s 关联且 a 在其中为真），因此 c 在 s 中为真，这正是我们要证的结果。

回到我们开头那个论证。我们现在就能看出为什么它不成立。该论证所依赖的推断是：

$$\frac{\neg(a \wedge \neg c)}{a \to c}$$

而这个推断不是有效的。例如，如果 a 在某个情形 s 为 F，这足以使得前提在 s 为真。但这并没有告诉我们在与 s 关联的可能情形中 a 和 c 会如何表现。完全有可能在其中一个可能情形，比方说 s' 中，a 为真而 c 不为真，如图：



因此，$a \to c$ 在 s 不为真。

我们早先谈到的那个例子怎么样呢（你被告知你不会去电影院而不花钱）？那个推断看上去难道不是有效的吗？假设你知道你不会去看电影而不花钱：$\neg(g \wedge \neg m)$。你真的就能得出结论说，如果你去看电影你就要花钱吗：$g \to m$？不一定。假设你不打算去看电影，即使那天晚上电影是免费的。（电视上有个节目有趣得多。）那么你知道，你会去看电影不为真（$\neg g$），于是你会去看电影且不花钱也不为真：$\neg(g \wedge \neg m)$[9]。现在你能推出如果你去看电影就要花钱吗？当然不能：那天可能是免费之夜。

要注意的是，当你处在经由被告知而得知前提为真的情形时，其他因素通常也在起作用。注意到这一点很重要。当有人告诉你 $\neg(g \wedge \neg m)$[10] 这样的事的时，他们一般不会基于他们知道 $\neg g$ 为真而这样做。（如果他们知道的话，告诉你关于这种情形的任何事通常都是没有意义的。）如果他们告诉你这样的事，那就是基于 g 和 m 之间有某种联系：你不可能有 g 为真而 m 不为真，而这正是那个条件句为真所需要的。因此，在你被告知这个前提的情况下，推出 $g \to m$ 通常是合理的；但并不是由所说的内容推出，而是由它**被说出来了**这一事实推理得到。

事实上，我们经常不加思索地正确作出这种推断。例如，假设我问别人，如何让我的电脑做这样那样的事，然后他们回答说，“书架上有一本手册”。我就可以推断出，那是一本电脑手册。这并不能从实际所说的推出。但除非那本手册是电脑手册，否则那个回答就是不相干的了，而人们一般不说不相干的话。这个推断并不是演绎推理。毕竟，那个人可以说那句话，而指的不是电脑手册。但这个推断仍然是一个极好的归纳推理，是通常被称作**会话蕴含**（*conversational implicature*）的一种。

我们刚看到的对条件句的解释似乎很成功——至少就我们看到的而言。不过，它也面临一些问题。这里是一个。考虑如下推断：

> 如果你去罗马，你就会在意大利。
>
> 如果你在意大利，你就在欧洲。
>
> 因此，如果你去罗马，你就会在欧洲。
>
> ---
>
> 如果 x 大于 10，那么 x 大于 5。

因此，如果 x 大于 10 且小于 100，那么 x 大于 5。

这些推断似乎完全是有效的。在目前的解释下，它们也确实有效。我们可以把第 1 个推断写成

1. $\dfrac{a \to b \quad b \to c}{a \to c}$

要看出它是有效的，假设前提在某个情形 s 为真。那么 b 在每个与 s 关联且 a 为真的情形为真；同样，c 在每个与 s 关联且 b 为真的情形为真。因此，c 在每个与 s 关联且 a 为真的情形为真。即，$a \to c$ 在 s 为真。

我们可以把第 2 个推断写成：

2. $\dfrac{a \to c}{(a \wedge b) \to c}$

要看出它是有效的，假设前提在某个情形 s 为真。那么 c 在每个与 s 关联且 a 为真的情形为真。现在，假设 $a \wedge b$ 在某个与 s 关联的情形为真，那么 a 在该情形当然也为真，因此 c 也为真。因此，$(a \wedge b) \to c$ 在 s 为真。

目前都很好。问题在于，有些推断在形式上与前面两个推断完全相同，却似乎是无效的。例如，假设有一场首相选举，只有两个候选人，琼斯和现任首相史密斯。现在考虑如下推断：

如果史密斯在选举前死了，琼斯就会赢得选举。

如果琼斯赢得选举，史密斯就会退休然后领取退休金。

因此，如果史密斯在选举前死了，她就会退休然后领取退休金。

这是一个恰好形如 1 的推断。但似乎很明显存在某个情形使得两个前提都为真，而结论不为真——除非我们在考虑某个怪异的情形，其中政府会在人死之后继续发放退休金！

或者考虑下面这个关于史密斯的推断：

如果史密斯从很高悬崖的顶上跳下，她就会摔死。

因此，如果史密斯从很高悬崖的顶上跳下且背了降落伞，她就会摔死。

这是一个形如 2 的推断。不过同样，似乎很明显存在前提为真而结论不为真的情形。

关于这种状况有什么要说的呢？我留给读者自己思考。尽管条件句对我们如何进行大多数推理至关重要，但条件句逻辑仍然是逻辑学中最有争议的领域之一。即便鸟儿不再吵嚷着条件句，逻辑学家肯定还在为此争论不休。

本章要点

- $a \to b$ 在某个情形 s 中为真，当且仅当 b 在每个与 s 关联且 a 为真的情形中为真。

第 8 章：将来和过去：时间是真实的吗？

时间是我们都很熟悉的东西。我们计划将来之事，我们记得过去之事，有时我们则只是享受现在。一种了解时间的办法就是进行有关时间的推断。比如，下面两个推断直觉上是有效的：[11]

```
 正在下雨。            过去一直在下雨将是真的。
__________           ________________________
将要下过雨。                正在下雨。
```

所有这些似乎都很基本。

但一旦人们开始思考时间，似乎就会陷入纠缠不清。正如奥古斯丁说过的，如果没有人问我时间是什么，我知道得很清楚；但当有人问我时，我就什么也不知道了。时间最令人困惑的一个地方就在于，它似乎是流动的。现在似乎在移动，先是今天，接着是明天，如此等等。但时间怎么能变化呢？时间是衡量其**他万物**变化速度的尺度。这个问题处在好几个有关时间之谜的核心。其中一个在 20 世纪早期由英国哲学家 John McTaggart Ellis McTaggart（没错，就是这个名字）提出。同许多哲学家一样，McTaggart 被时间是不真实的这一观点所吸引——即，位于终极秩序中的时间不过是一种幻觉。

为了解释 McTaggart 对这一观点的论证，有必要引入一些符号。考虑一个过去时的句子，比如“太阳那时在照耀”（The Sun was shining）。我们可以（若稍有别扭的话）等价地将其表达为，“过去的情况是：太阳在照耀”。让我们把“过去的情况是”记作 **P**（表示 Past）。那么我们就可以把这句话记作“**P** 太阳在照耀”，或者，若将“太阳在照耀”记为 s 的话，就可以简写为 $\mathbf{P}s$。类似的，考虑任意一个将来时的句子，比方说“太阳将要照耀”。（严格来说，语法学家会告诉你，不像法语或拉丁语，英语没有真正的将来时。不过你知道我的意思。）我们可以把它记作“将来的情况是：太阳在照耀”。如果我们把“将来的情况是”记作 **F**（表示 Future），那么我们可以把它记作 $\mathbf{F}s$（不要混淆 **F** 和表示真值的 F。）

P 和 **F** 是像 □ 和 ◊ 一样的算子，它们加在完整句的前面构成新的完整句。并且，同 □ 和 *Diamond* 一样，它们也不是真值函数。“现在是下午 4 点”和“现在是 2017 年 9 月 10 日下午 4 点”都为真（在我写下这两句话时）。“将会是下午 4 点”也为真（在当前时刻）——每天都有一次下午 4 点——但“将会是 2017 年 9 月 10 日下午 4 点”则不为真。逻辑学家称 **P** 和 **F** 为时态算子（*tense operators*）。时态算子可以叠加或复合。例如，我们可以说“太阳将照耀过”，即“将来的情况是过去的情况是：太阳在照耀”：$\mathbf{FP}s$。或者我们也可以说“太阳那时照耀过”[12]，即，“过去的情况是过去的情况是：太阳在照耀”：$\mathbf{PP}s$。（我们上上一章[13]见过的模态算子也可以像这样叠加，不过我们在那里并没有考虑。）并不是所有时态算子的叠加在英语中都有简洁明快的表达。比如，除了蹩脚的“将来的情况是过去的情况是将来的情况是：太阳在照耀”，没有更好的办法来表达 $\mathbf{FPF}s$。尽管如此，这样的叠加在语法上是完全说得通的。我们可以把像 $\mathbf{FP}, \mathbf{PP}, \mathbf{FFP}$ 这样的 **P** 和 **F** 的叠加称为复合时态（*compound tenses*）。

现在回到 McTaggart 的论证。他的推理是，如果没有过去和将来，就没有时间，因为过去和将来是时间之本。他论证说，过去性和将来性是内在矛盾的，因此没有什么现实的东西与它们对应。也许是吧。但为什么说过去和将来是矛盾的呢？首先，过去和将来是不相容的。如果某个瞬时事件是过去的，它就不是将来的，反之亦然。令 e 为某个瞬时事件。它可以是任何你喜欢的事，但让我们假设它是俄国革命时第一个子弹从尼古拉沙皇的心脏穿过。令 h 为语句“e 在发生”。那么我们有：

$$\neg(\mathbf{P}h \wedge \mathbf{F}h)$$

但 e 就像所有事件一样，是过去的也是将来的。因为时间在流逝，所有的事件既有成为将来的性质（在它们发生前）也有成为过去的性质（在他们发生后）。即：

$$\mathbf{P}h \wedge \mathbf{F}h$$

这个论证不太可能说服任何人很久。一个事件不可能同时既是过去的又是将来的。子弹穿过沙皇心脏的瞬间是在不同时刻成为过去和将来。它一开始是未来之事，在某个痛苦的瞬间成为现在，然后成为过去。但是，现在——这正是 McTaggart 论证的狡猾之处——我们正在说的是什么呢？我们在对 h 使用复合时态。我们在说，过去的情况是：该事件是未来的；然后是，过去的情况是：该事件是过去的。而很多复合时态，就像简单时态一样，是不相容的。例如，如果任何事件将要成为将来，那么并非它在过去会成为过去，即：

$$\neg(\mathbf{PP}h \wedge \mathbf{FF}h)$$

但正如简单时态那样，时间的流逝足以确保所有事件也具有所有的复合时态。在过去，$\mathbf{F}h$，因此在更遥远的过去 $\mathbf{FF}h$。在将来，$\mathbf{P}h$，因此在更远的将来，$\mathbf{PP}h$。即：

$$\mathbf{PP}h \wedge \mathbf{FF}h$$

那些保持机警之人会像前面一样回应说，h 是在不同的时刻具有那些复合时态。过去的情况是 $\mathbf{FF}h$；然后，过些时候，过去的情况是 $\mathbf{PP}h$。但这里我们正在说什么呢？我们在对 h 应用更复杂的复合时态：$\mathbf{PFF}h$ 和 $\mathbf{PPP}h$；然后我们可以进行和前面完全一样的论证：这些复合时态并不都是相互一致的，但时间的流逝确保 h 可以具有所有这些时态。我们也许可以进行同样的回应，但还是会遇到相同的反驳。每当我们试图用一组时态来避免矛盾，我们只不过是用其他同样矛盾的时态来描述事情。因此，我们永远也摆脱不了矛盾。这就是 McTaggart 的论证。

关于这个论证有什么要说的呢？为了回答这个问题，让我们看一下有关时态推断的有效性。为了解释这一点，我们假设每个情形 s_0 都与一组其他的情形一起出现——这次的情形不是代表与 s_0 关联的可能情形（像模态算子那样），而是要么在 s_0 之前要么在 s_0 之后的情形。假设（正如我们通常所做的）时间是一维的，且在过去和将来两个方向上都是无限的，我们就可以把这些情形用常见的方式表示为：

$$\ldots s_{-3} \quad s_{-2} \quad s_0 \quad s_1 \quad s_2 \quad s_3 \ldots$$

左边是更早的情形，右面是更晚的。像以前一样，每个 s 对每个不含时态算子的语句提供一个真值。[14] 含时态算子的语句真值如何呢？$\mathbf{P}a$ 在情形 s 为 T，当且仅当 a 在 s 左边某个情形为真；$\mathbf{F}a$ 在 s 为真，当且仅当 a 在 s 右边某个情形为真。

在做所有这些的时候，我们还可以增加两个新的时态算子，**G** 和 **H**。**G** 可以读作“将要永远是”。**G**a 在情形 s 为真，当且仅当 a 在 s 右边的所有情形都为真。**H** 可以读作“过去一直是”。**H**a 在 情形 s 为真，当且仅当 a 在 s 左边的所有情形都为真。（**G** 和 **H** 分别与 **F** 和 **P** 对应，就像 $\Box$ 与 $\Diamond$ 的对应。）

这套工具向我们表明，为什么本章开头的两个推断是有效的。应用时态算子，这两个推断可以分别写成：

$$\frac{r}{\mathbf{FP}r} \qquad \frac{\mathbf{FH}r}{r}$$

第 1 个推断是有效的，因为如果 r 在某个情形 s_0 为真，那么在 s_0 右边的任何情形，比如 s_1，**P**r 都为真（因为 s_0 在它左边，[而 r 在 s_0 为真——译者补注]）。于是 **FP**r 在 s_0 为真，因为 s_1 在它右边。我们可以把这里的情况描绘如下：

...	s_{-3}	s_{-2}	s_{-1}	s_0	s_1	s_2	s_3	...
				r				
					$\mathbf{P}r$			
				$\mathbf{FP}r$				

第 2 个推断是有效的，因为如果 **FH**r 在 s_0 为真，那么在 s_0 右边的某个情形，比方说 s_2，**H**r 为真。于是在 s_2 左边的所有情形，特别的，在 s_0，r 为真：

...	s_{-3}	s_{-2}	s_{-1}	s_0	s_1	s_2	s_3	...
				FHr				
						$\mathbf{H}r$		
	r	r	r	r	r			

并且，正如所料，某些时态的组合是不可能的。例如，若 h 是仅在一个情形，比方说 s_0 为真的语句，则 **P**$h \wedge$ **F**h 在每个 s 均为假：每个合取项在 s_0 都为假；第一个合取项在 s_0 左边的所有情形均为假；第二个合取项在 s_0 右边的所有情形均为假。类似的，**PP**$h \wedge$ **FF**h 在每个情形均为假。详细验证留给读者完成。

现在，所有这些和 McTaggart 的论证有何关系呢？回想一下，McTaggart 论证的要点在于，给定 h 具有所有可能的时态，矛盾就永远不可避免：对复合时态的矛盾在一个层面进行消解，只不过让它们又在另一个更复杂的层面出现。我们刚才对时态算子的解释将表明，这并不成立。假设 h 只在 s_0 为真。那么关于 h 的任何复合时态陈述都在某个地方为真。比如，考虑 **FPPF**h，则它在 s_{-2} 为真，如下表所示：

...	s_{-3}	s_{-2}	s_{-1}	s_0	s_1	s_2	s_3	...

				h				
			$\mathbf{F}h$					
					$\mathbf{PF}h$			
						$\mathbf{PPF}h$		
		$\mathbf{FPPF}h$						

显然，只要之字形地按要求向左向右进行推断，对每个由 **F** 和 **P** 构成的复合时态我们都可以做同样的事。而所有这些都是完全一致的。无穷多的不同情形让我们可以对 h 在合适的地方指派所有的复合时态，而不会违反它们之间的各种不相容性，比如让 $\mathbf{F}h$ 和 $\mathbf{P}h$ 在相同的情形为真。因此，McTaggart 的论证并不成立。

对那些愿意相信时间现实性的人来说，这是个令人高兴的结果。但那些赞同 McTaggart 的人或许仍然不会被我们的考虑所说服。假设我给你一组造房子的具体要求：这里是前门，那里是一扇窗户……，你怎么知道所有这些要求是相互一致的呢？你怎么知道当你进行建造时，一切都没问题；你不会被要求，比方说，把门安在一个不相容的地方？一种判定它的方法就是根据所有要求建一个比例模型。如果这个比例模型能建成，那么那些要求就是一致的。这也正是我们在讨论时态问题时所做的。时间的模型是一些情形构成的序列，加上一种对时态语句指派真值的方法。它比房子的模型要稍微抽象一点，但原理本质上是一样的。

不过，人们也可以反对一个模型。有时，模型会忽略一些重要的东西。例如在房子的比例模型中，横梁也许不会垮，因为它受到的压力远比在全尺寸建筑中小得多。实际的横梁也许被要求承担不可能的重量，从而使得整个建筑也不可能——尽管模型是可能的。类似的，我们的时间模型也可能被认为忽略了重要的东西。毕竟，我们所做的是对时间给出了一个空间模型（左、右等等）。空间不像时间那样流动（不管它实际是什么意思）。现在，正是时间的流动产生了 McTaggart 指出的所谓矛盾。难怪矛盾在我们的模型中不会出现。那么，我们的模型中到底漏掉了什么呢？一旦把它考虑进去，矛盾会重新出现吗？

本章要点

- 每个情形都关联一些更早和更晚的情形。
- $\mathbf{F}a$ 在某个情形为真，若a 在某个更晚的情形为真。
- $\mathbf{P}a$ 在某个情形为真，若 a 在某个更早的情形为真。
- $\mathbf{G}a$ 在某个情形为真，若 a 在每个更晚的情形为真。
- $\mathbf{H}a$ 在某个情形为真，若 a 在每个更早的情形为真。

第 9 章：同一性与变化：有什么是一成不变的吗？

我们还没有结束对时间的讨论。各种其他难解之谜都牵涉到时间。本章我们将考察其中一种。它与事物变化时产生的问题有关，特别是历时变化的对象的同一性问题。

这里有一个例子。我们都认为对象可以在经历变化之后得以保存。例如，当我油漆碗柜时，尽管其颜色发生了改变，但它还是同一个碗柜。或者当你改变发型时，或不幸失去一条胳膊或一条腿时，你还是你。但怎么会有东西在经历变化之后仍然保存呢？毕竟，当你改变发型后，你和原来是有差异的，一点也不会和原来相同。而如果一个人是不同的，那就是一个不同的人。用完全相同的方式，我们可以论证，无论经历什么样的变化，没有什么对象能在经历变化之后仍然得以保持。因为任何变化都意味着原来的对象不复存在，它被一个完全不同的对象所取代。

这样的论证在哲学史上的不同地方都出现过，但现在逻辑学家一般认为它们是错误的，是基于一个简单的歧义性。我们必须区分一个对象和它的性质。当我们说，你有个不同的发型，是不同的，我们是在说你有不同的性质。这推不出你也成了一个不同的人，就像你和我是不同的人一样。

人们不能区分是某个对象和具有某些性质的一个原因是，汉中的动词“是”可以用来表达这两方面的意思。（其他语言的类似词语也同样如此。）如果我们说“这张桌子是红的”，“你的头发现在是短的”，以及类似的句子，我们就是在赋予一个性质给某个对象。但如果有人说“我是Graham Priest”，“那个赢了比赛的人和去年那个赢了比赛的人是同一个人。”，等等，那他们就是在以某种方式确认一个对象。即，他们在陈述其同一性。

逻辑学家称“是”的第一种用法为**谓述之“是”**（*'is' of predication*），称“是”的第二种用法为**同一之“是”**（*'is' of identity*）。由于它们有稍微不同的性质，逻辑学家把它们写成不同的形式。在第 3 章我们已经见过谓述之“是”。“约翰是红的”通常记作 jR（实际上，正如我在第 3 章指出的，更常见的是反过来记作 Rj。同一之“是”记作中学数学就常见的 $=$。这样，“约翰是那个赢了比赛的人”就记作 $j = w$。（名称 w 这里是摹状词，但对目前的问题不重要。）这样的句子被称作恒等句（*identities*）。

同一性具有什么性质呢？首先，它是一种关系。关系就是连接两个对象的东西。比如，**看见**就是一种关系。如果我们说“约翰看见玛丽”，我们就是在陈述二者之间的一种关系。由关系连接起来的对象不必是不同的。我们说“约翰看见他自己”（也许是在镜子里），我们就是在陈述一种约翰和约翰之间的关系。现在，同一性是一种非常特殊的关系。它是一种每个对象都和自己而不和任何其他对象有关系的关系。

你也许会认为这会使得同一性是一种毫无用处的关系，但事实并非如此。例如，如果我说“约翰是那个赢了比赛的人”，我就是在说由“约翰”指称的对象与由“那个赢了比赛的人”之间的同一关系，换言之，这两个名称指称同一个人。这可以是非常重要的一条信息。

不过，有关同一性最重要的是涉及它的推断。这里是一个例子：

> 约翰是那个赢了比赛的人。
>
> 那个赢了比赛的人得了奖。
>
> 因此，约翰得了奖。

我们可以把这个推断写成：

$$\frac{j = w \quad wP}{jP}$$

这个推断之所以有效是因为，对任何两个对象，x 和 y，如果 $x = y$，那么 x 具有所有 y 具有的性质，反之亦然。同一个对象要么具有一种所说的性质，要么不具有。这通常称作**莱布尼茨律**（*Leibniz's Law*），以莱布尼茨命名（我们在第 6 章见过此人）。在应用莱布尼茨律时，一个前提是同一性陈述，比如 $m = n$，第二个前提是包含等号两边某个名称，比如 m的语句，而结论由在该语句中用 n 替换 m 得到。

莱布尼茨律是一条非常重要的定律，也有许多毫无问题的应用。例如，高中代数向我们保证 $(x + y)(x - y) = x^2 - y^2$。因此，如果你在解题时得到 $x^2 - y^2 = 3$，你就可以应用莱布尼茨律推断得到 $(x + y)(x - y) = 3$。不过，这一具有迷惑性的简单性隐藏了众多问题。特别的，莱布尼茨律似乎存在着许多反例。比如，考虑下面这个推断：

> 约翰是那个赢了比赛的人。
>
> 玛丽知道那个赢了比赛的人得了奖。
>
> 因此，玛丽知道约翰得了奖。

这看上去像是一个莱布尼茨律的应用，因为结论是由第二个前提用“约翰”替换“那个赢了比赛的人”得到的。然而，显然完全有可能前提为真而结论不为真：玛丽也许不知道约翰就是那个赢了比赛的人。这违反了莱布尼茨律吗？未必。莱布尼茨律说的是，如果 $x = y$，那么 x 的任何性质也是 y 的性质。那么，条件“玛丽知道 x 得了奖”表达了 x 的某个性质吗？不见得：相反，它表达的似乎是玛丽的某个性质。假如玛丽突然不存在了，这不会对 x 有任何改变！（逻辑学中关于“知道”的逻辑仍然悬而未决。）

另一类问题如下。这里有一条路，它是一条柏油路，称它为 t。这里有一条路，它是一条泥土路，称为为 d。不过，这两条路是同一条路，$t = d$。只不过柏油往路的尽头逐渐耗尽，于是成了泥土路。于是莱布尼茨律告诉我们，t 是泥土路而 d 是柏油路——而它们并不是。哪里出问题了呢？我们不能说，是泥土的和是柏油的不是路的真正性质。它们当然是。（可以认为）出问题的地方在于：我们在性质的具体描述上不够精确。相关性质应为，在某某地方是柏油的，在某某地方是泥土的。由于 t 和 d 是同一条路，它们两个性质都具有，因而我们并未违反莱布尼茨律。

目前为止都很好。这些问题都相对容易。现在让我们看一个不那么容易的。这里，时间又回到问题中来了。为了解释这个问题，有必要应用上一章介绍的时态算子，特别是 $\mathbf{G}$ 算子（“将要永远是”）。令 x 是任何你喜欢的东西，比如一棵树，某个人，然后考虑 $x = x$ 这一陈述。它说的是 x 具有同一于 x 的性质。这显然为真：这正是同一性含义的一部分。而且，该陈述为真不会受到时间的影响。它现在为真，将来所有时间也都为真。于是，特别的，$\mathbf{G}\, x = x$。现在，这里是莱布尼茨律的一个实例：

$$\frac{x = y \quad \mathbf{G}\, x = x}{\mathbf{G}\, x = y}$$

（不要被我们在第二个前提中用 y 只替换了 x 的一次出现这一事实所迷惑。如此应用莱布尼茨律是完全说得通的。只需考虑：“约翰就是那个赢了比赛的人；约翰看见约翰；因此，约翰看见那个赢了比赛的人”。）该推断表明的是，若 x 同一于 y，且 x 具有在将来所有时间都同一于 x 的性质，那么 y 也具有这样的性质[15]。我们已经指出，第二个前提为真。由此可以推出，只要两个东西是同一的，它们就会永远是同一的。

这个结论如何呢？简单来说，它似乎并不总是为真。例如，考虑一只变形虫。变形虫是单细胞水生物，靠细胞分裂进行繁殖：一只变形虫会从中间分裂变成两只变形虫。现在，考虑某个变形虫 A，它分裂成两个变形虫 B 和 C。在分裂前，B 和 C 都是 A。因此在分裂前，$B = C$。但在分裂后，B 和 C 是不同的变形虫，$\neg B = C$。因此，即使两个东西现在相同，也不必然推出它们永远相同。

我们无法使用前面的方法避免这个问题。在将来所有时间都同一于 x 当然是 x 的一个性质。而且似乎该性质也并非不够精细。我们似乎没有办法能够使它更精确而避免问题。

还有什么要说的呢？一个自然的想法是，在分裂前 B 不是 A：它只是 A 的一部分。但 B 是一只变形虫，而 A 是一个单细胞生物：它没有任何部分是变形虫。因此 B 不可能是 A 的一部分。

更极端地，人们也许会认为，B 和 C 在在分裂前并没有真正存在，它们是在分裂时才产生的。如果分裂前它们并不存在，那么分裂前它们也不是 A。因此分裂前并没有 $B = C$。但这似乎也是错的。B 并不是一只新的变形虫，它就是 A，只不过它的一些性质发生了改变。如果这一点不够清楚的话，只需设想 C 在分裂时死掉了。在这种情况下，我们会毫不犹豫地说 B 就是 A。（这就像蛇蜕皮一样。）现在，事物的同一性不会受到其周围是否还有其他事物而影响。所以，A 就是 B。同样，A 也是 C。

当然，人们也许会坚持，正因为 A 具有了新的性质，严格来说，它就是一个新的对象了，而不只是旧的对象具有了新的性质。因此，B 并不真的就是 A。C 也一样。但这样一来，我们又回到了本章开头那个问题。

本章要点

- $m = n$ 为真，当且仅当 m 和 n 指称相同的对象。
- 如果两个对象相同，那么一个对象的任何性质也是另一个对象的性质（莱布尼茨律）。

第 10 章：模糊性：如何在滑坡上停止下滑？

当我们讨论同一性这一主题时，这里有另一个相关的问题。所有东西都会随时间而损耗。有时，零件会被更换。摩托车和汽车会换新的离合器，房子会换新的屋顶；甚至人体内的单个细胞也会随着时间而更换。这样的变化并不会影响上述对象的同一性。当我更换摩托车的离合器时，它还是原来那辆车。现在假设经过几年时间后，我更换了我的摩托车黑霹雳的每一个零件。作为

一个细心的人，我保留了所有旧的零件。当全部零件更换完后，我把旧零件重新组装回去造出原来那辆车。但我是从黑霹雳开始的，而更换摩托车的一个零件并不会影响其同一性：它还是原来那辆车。因此，每次更换之后，得到的机器仍然是黑霹雳；直到最后，它也还是黑霹雳。但我们知道这不可能是对的。黑霹雳现在正在车库里它的旁边放着呢。

下面是相同问题的另一个例子。一个 5 岁的人（在生物学上）是个小孩。如果某人是个小孩，一秒钟之后他（她）还是小孩。这样，再过一秒，他（她）还是小孩，再过一秒，再过一秒，……。因此，过了 630720000 秒后，他（她）还是小孩。但那时他（她）已经 25 岁了！

这样的论证被认为是由欧布里德（同样是他创造了第 5 章中的说谎者悖论）创造的。它们现在被称为**堆积悖论**（*sorites paradox*）。（它的一个标准形式大意是，每次增加一粒沙，永远也形成不了沙堆；“sorites”来自于希腊语“soros”，意思是“堆”。）这些是逻辑学中最让人头痛的一些悖论。当所使用的谓词（“是黑霹雳”，“是小孩”）在某种意义上是**模糊**（*vague*）的时候，悖论就会产生。模糊谓词是指，其适用性可以容忍非常小的变化：如果它适用于某个对象，那么该对象一个非常小的变化不会更改这一事实。我们日常对话中使用的几乎所有谓词在这个意义上都是模糊的：“是红的”、“是醒的”、“是高兴的”、“是醉的”，甚至“是死的”（死也需要时间）。因此，堆积悖论这类滑坡论证在我们的推理中可能极为普遍。

为了集中探讨关于它们的问题，让我们更仔细地考察其中一个论证。令杰克是那个 5 岁的小孩。令 a_0 为语句“杰克 0 秒后是小孩”。令 a_1 为语句“杰克 1 秒后是小孩”，如此等等。如果 n 是任一自然数，a_n 就是语句“杰克 n 秒后是小孩”。令 k 为某个巨大的自然数，至少不小于 630720000。我们知道 a_0 为真。（0 秒过后，杰克仍然是 5 岁。）对每个自然数 n，我们知道 $a_n \to a_{n+1}$（在任何时刻如果杰克是个小孩，一秒后他还是小孩。）我们可以通过使用一系列分离规则把所有这些前提串连在一起，如下所示：

$$\dfrac{\dfrac{a_0 \quad a_0 \to a_1}{a_1} \quad a_1 \to a_2}{a_2}$$

$$\ddots$$

$$\dfrac{a_{k-1} \quad a_{k-1} \to a_k}{a_k}$$

最终结论是 a_k，而我们知道它不为真。一定是哪里出错了，而且似乎没有多少回旋的余地。那么我们有什么要说的呢？这里是一个解答，有时称为**模糊逻辑**（*fuzzy logic*）。是小孩似乎是逐渐消失的，正如（在生物学上）是成人也是逐渐显现的。假定“杰克是小孩”的真值也由真渐变为假似乎是自然的。于是，真是有程度的。假设我们用 0 和 1 之间的数字来衡量真的程度，1 表示完全为真，0 表示完全为假。那么，每个情形都对基本语句指派一个这样的数。

包含否定和合取这样算子的语句的真值如何呢？随着杰克长大，“杰克是小孩”的真值就变小。“杰克不是小孩”的真值似乎也相应变大。这提示 $\neg a$ 的真值为 1 减去 a 的真值。假设我们把 a 的真值记作 $|a|$，那么我们有：

$$|\neg a| = 1 - |a|$$

下表是一些真值样本：

a	$\neg a$
1	0
0.75	0.25
0.5	0.5
0.25	0.75
0	1

合取式的真值呢？一个合取式最多只能和它最差的部分一样好。因此假设 $a \wedge b$ 的真值是 $|a|$ 和 $|b|$ 的**极小值**（*minimum*）（更小的）是自然的：

$$|a \wedge b| = \min(|a|, |b|)$$

下表是一些真值样本：

$a \wedge b$	**1**	**0.75**	***0.5***	**0.25**	**0**
1	1	0.75	*0.5*	0.25	0
0.75	0.75	0.75	*0.5*	0.25	0
0.5	0.5	0.5	*0.5*	0.25	0
0.25	*0.25*	*0.25*	***0.25***	*0.25*	*0*
0	0	0	*0*	0	0

a 的真值在最左边一列，b 的真值在最上一行。$a \wedge b$ 的对应真值在相应行和列相交的地方。比如，如果我们想找到 $a = 0.25$ 和 $b = 0.5$ 时 $a \wedge b$ 的真值，我们就看斜体的行和列相交的地方。结果是粗体显示的那个值。

类似的，析取式的值是析取项值的**极大值**（*maximum*）（更大的）：

$$|a \vee b| = \max(|a|, |b|)$$

读者可以自行构造一个它的真值样本表。注意，根据上面的假定，$\neg$，$\wedge$ 和 $\vee$ 仍然是真值函数。即，比如 $a \wedge b$ 的真值由 a 和 b 的真值确定。只是现在这些值不再是 T 和 F，而是 0 和 1 之间的数。（不过，或许值得一提的是，如果我们把 1 看作 T，把 0 看做 F，在只涉及 1 和 0 时的结果和第 2 章中的真值函数是相同的。这一点读者可以自行验证。）

条件句的真值呢？我们在第 7 章中看到，有很好的理由认为 $\rightarrow$ 不是真值函数。但让我们把那些顾虑暂时放在一边。如果它是一个真值函数，当考虑真的程度时，它会是哪个函数？似乎没有很明显的答案。这里是一个（相当标准的）建议，至少看上去给出了**某种**正确的结果。

- 若 $|a| \leq |b|$，则 $|a \rightarrow b| = 1$

- 若 $|b| < |a|$，则 $|a \to b| = 1 - (|a| - |b|)$

（$<$ 指“小于”；$\leq$ 指“小于或等于”。）

这样，如果前件没有后件真，该条件句就完全真。如果前件比后件更真，该条件句的真值就是最大真值减去二者真值之差。下表是一些真值样本：

$a \wedge b$	**1**	**0.75**	**0.5**	**0.25**	**0**
1	1	0.75	0.5	0.25	0
0.75	1	1	0.75	0.5	0.25
0.5	1	1	1	0.75	0.5
0.25	1	1	1	1	0.75
0	1	1	1	1	1

（回想一下，a 的值在最左一列，b 的值在最上一行。）

有效性如何呢？一个推断是有效的，若结论在前提成立的每个情形都成立。但现在什么是在某个情形成立呢？在足够真的时候。但多真是足够真呢？这要看语境。例如，“是一辆新自行车”是一个模糊谓词。如果你去买自行车，经销商告诉你某辆车是新的，你会期望它是从未用过的。即，你期望“这是一辆新自行车”具有真值 1。另一方面，假设你去参加自行车公路赛，被要求选新车，你会挑选那些不超过一年的车。换言之，你可接受的新车标准要更宽松。“这是一辆新自行车”只需要比方说 0.9 或更高的真值。

因此，我们假设存在某个由语境确定的可接受性水平。这是一个介于 0 和 1 之间的数字——在极端情况下也许就是 1 本身。让我们把这个数字记为 ε。于是一个推断对某个语境是有效的，当且仅当结论的真值在前提的真值都不小于 ε 的每个情形也都不小于 ε。

那么，所有这些和堆积悖论有什么关系呢？假设我们有一个堆积的序列。像前面一样，令 a_n 为语句“杰克 n 秒后是小孩”；但为了让事情更容易处理，让我们假设杰克 4 秒后就长大了！那么一个真值记录表可以是：

a_0	a_1	a_2	a_3	a_4
1	0.75	0.5	0.25	0

$a_0 \to a_1$ 具有真值 0.75（ = (1 - (1 - 0.75))）； $a_1 \to a_2$ 的真值也是 0.75；事实上，每个形如 $a_n \to a_{n+1}$ 的条件句的真值都是 0.75。

关于堆积悖论这些告诉了我们什么，依赖于可接受性水平，而正是这一点在起作用。假设语境要求最高的可接受性水平，即 1。在这种情况下，分离规则是有效的。因为假设 $|a| = 1$ 且 $|a \to b| = 1$，由于 $|a \to b| = 1$，我们必有 $|a| \leq |b|$。由此可得 $|b| = 1$。这样，堆积论证是有效的。不过，在这种情况下，每个条件句前提具有真值 0.75，而这（在该语境下）是不可接受的。

另一方面，如果我们设定可接受性水平低于 1，那么分离规则就变得无效。为了解释方便，假设 ε 为 0.75。正如我们已看到的，a_1 和 $a_{\to}a_2$ 都具有真值 0.75，但 a_2 的真值是 0.5，小于 0.75。

于是，无论你以哪种方式看，该论证都不成立。要么某个前提不可接受，要么如果它们是可接受的，结论却不能有效地推出。为什么我们这么容易就被堆积论证所欺骗呢？也许因为我们混淆了完全的真和近乎完全的真。不能对此作出区分通常不会造成什么差异。但如果你一而再、再而三如此，结果就不同了。

以上是对该问题的一种诊断。但对于模糊性问题，没有什么是简单明了的。说“杰克是个小孩”完全真，直到某个特定时间点，它变成了完全假，这有什么问题呢？问题在于，就没有这样的点。在任何地方划线都是完全任意的；它最多只能是一种约定。但现在，在杰克长大的哪个时间点他不再是 100% 的小孩呢；即在哪个时间点“杰克是个小孩”的值从完全为 1 变成低于 1 的值呢？在任何地方划线似乎都和前面一样是任意的。如果这是对的，我们就没有真正解决模糊性最根本的问题：我们只是转移了问题。

本章要点

- 真值是介于 0 和 1（含）之间的数字。
- $|\neg a| = 1 - |a|$
- $|a \vee b| = \max(|a|, |b|)$
- $|a \wedge b| = \min(|a|, |b|)$
- $|a \to b| = 1$，若$|a| \leq |b|$；否则，
- $|a \to b| = 1 - (|a| - |b|)$
- 一个语句在某个情形为真当且仅当其真值至少达到（由语境确定的）可接受性水平。

第 11 章：概率：缺少参照类的奇怪情形

前面各章让我们初步认识了哪些推断是演绎有效的，以及为什么。现在该回到归纳有效性这个问题了，它是这样一些推断的有效性：前提对结论提供了根据，但即使在某个情形下前提都为真，结论仍然可以为假。

正如我在第 1 章中提到的，夏洛克·福尔摩斯很擅长这类推断。让我们以他的一个推断作为例子开头。《红发会》之谜是这样开始的：福尔摩斯和华生博士受到了来自 Jabez Wilson 先生的拜访。当 Wilson 进来后，华生想看看福尔摩斯对他作了哪些推断：

> “他干过一段时间体力活，吸鼻烟，是共济会成员，去过中国，最近写过不少东西，除了这些显而易见的事实外，我推不出什么别的。”
>
> Jabez Wilson 从他的椅子上突然站起，手指仍然压着报纸，但眼睛已经转向我的同伴。
>
> “我的天呐，福尔摩斯先生，你怎么知道所有这些的？”他问道。

福尔摩斯高兴地进行了解释。比如，关于写作：

> “你右手袖子上足有五寸长的地方闪闪发光，而左袖靠近手腕经常贴在桌面上的地方打了个整洁的补丁。这还能说明什么别的吗？”

尽管福尔摩斯习惯把这样的推断称为演绎，它事实上是一个归纳推断。完全有可能 Wilson 的外套会呈现上述特征但他并没有从事过大量写作。比如，他可以从某个从事过大量写作的人那里偷来这件外套。但是，福尔摩斯的推断无疑是一个很好的推断。是什么使得这个推断，以及这类推断是好的呢？一个合理的回答是使用概率进行解释。因此，让我们先来谈一下概率，然后再回到这个问题。

概率就是指派给语句的一个数字，用来度量该语句为真有多大可能（在某种意义上）。让我们把 a 的概率记作 $pr(a)$。按惯例，我们用 0 和 1之间的数值来度量概率。若 $pr(a) = 0$，则 a 确定为假；随着 $pr(a)$ 增加，a 为真的可能性就越大，直到 $pr(a) = 1$，a 确定为真。

关于这些数字还有什么其他可说的呢？让我们用一个简单的例子来解释。假设我们考虑某个特定星期的那些天。令 w 为每天要么为真要么为假的语句，比如“今天暖和”，令 r 是另一个这样的语句，比如“今天下雨”。相关信息由下表给出：

	周一	周二	周三	周四	周五	周六	周日
w			✓	✓		✓	✓
r		✓	✓			✓	

打勾表示语句在那一天为真，空白表示不为真。

现在，如果我们谈论的是这个特定的星期，随机选取的一天是暖和的概率是多少呢？有 4 天暖和，总共有 7 天。因此概率是 4/7。类似的，有 3 天下雨，因此下雨的概率是 3/7：

$$pr(w) = 4/7$$

$$pr(r) = 3/7$$

一般地，如果我们将 a 为真的天数记为 $\#a$，将总天数记为 N，则：

$$pr(a) = \#a/N$$

概率如何与否定、合取和析取相联系呢？首先看否定。$\neg w$ 的概率是多少？有 3 天不是暖和的，因此 $pr(\neg w) = 3/7$。注意到 $pr(w)$ 和 $pr(\neg w)$ 加起来等于 1。这并非偶然。我们有：

$$\#w + \#\neg w = N$$

两边除以 N 得到：

$$\frac{\#w}{N} + \frac{\#\neg w}{N} = 1$$

即，$pr(w)+pr(\neg w)=1$。

合取和析取的情况如下：有 2 天既暖和又下雨，因此 $pr(w\wedge r)=\#(w\wedge r)/N=2/7$。有 5 天暖和或者下雨，因此 $pr(w\vee r)=\#(w\vee r)/N=5/7$。这两个数字之间有什么关系呢？为了找出 $w\vee r$ 为真的天数，我们可以先把 w 为真的天数和 r 为真的天数加起来。这还不完全对，因为有些天被计算了两次：周三和周六。这两天既下雨又暖和。因此为了得到正确的数字，我们得减去二者都为真的天数：

$$\#(w\vee r)=\#w+\#r-\#(w\wedge r)$$

两边除以 N 得到：

$$\frac{\#(w\vee r)}{N}=\frac{\#w}{N}+\frac{\#r}{N}-\frac{\#(w\wedge r)}{N}$$

即：

$$pr(w\vee r)=pr(w)+pr(r)-pr(w\wedge r)$$

这是合取式和析取式概率的一般关系。

上一章我们看到，真的程度也可以用 0 和 1 之间的数来度量。因而我们也许会自然地认为，真的程度与概率是一样的。它们并不一样。特别的，合取和析取以完全不同的方式运算。对真的程度，析取是一个真值函数。具体而言，$|w\vee r|$ 是 $|w|$ 和 $|r|$ 的极大值。但我们刚才已看到，$pr(w\vee r)$ 并不由 $pr(w)$ 和 $pr(r)$ 单独决定。特别的，对我们的 w 和 r，$pr(w)=4/7$，$pr(r)=3/7$，而 $pr(w\vee r)=5/7$。但如果 $|w|=4/7$ 且 $|r|=3/7$，则 $|w\vee r|=4/7$ 而不是 $5/7$。

在我们回到归纳推断之前，我们还需要一点关于概率的知识。给定我们的样本星期，随机选取的某天下雨的概率为 3/7。但假设你知道选取的那天是暖和的，那么那天下雨的概率是多少呢？有 4 天是暖和的，但其中只有 2 天下雨，因此概率为 2/4。这个数字称为条件概率（*conditional probability*），记作 $pr(r|w)$，表示给定 w 的条件下 r 的概率。如果我们稍加思考，就能给出计算条件概率的一般公式。我们是怎么得到 2/4 这个数的呢？首先，我们把自己限制在 w 为真的那些天，然后用其中 r 为真的天数（即 w 和 r 都为真的天数）除以这个天数。换言之：

$$pr(r|w)=\#(w\wedge r)\div\#w$$

运用一点代数知识，这等于

$$\frac{\#(w\wedge r)}{N}\div\frac{\#w}{N}$$

而这就是 $pr(w\wedge r)\div pr(w)$。

因此，下面是计算条件概率的一般公式：

CP：$pr(w|r)=pr(w\wedge r)/pr(w)$

应用该公式要小心一点。除以 0 是没有意义的。比如，3/0 没有值。数学家称这种比率是**未定义的**（*undefined*）。在计算 $pr(w|r)$ 的公式中，我们用 $pr(w)$ 作除数，这只有在它不等于 0 时才有意义，即只有在 w 至少有时为真时才有意义。否则，这个条件概率就是未定义的。

现在我们终于能回到归纳推断上来。一个推断是归纳有效的是指什么呢？就是前提使得结论比其否定更加可能。即，给定前提（或多个前提的合取）p 的条件下，结论 c 的条件概率大于 c 的否定的条件概率：

$$pr(c|p) > pr(\neg c|p)$$

因此，如果我们对我们举例说明的星期进行推理，以下推断：

> 这天下雨；因此这天暖和

就是归纳有效的。因为容易验证，$pr(w|r) = 2/3$ 而 $pr(\neg w|r) = 1/3$。

这个分析可以用来表明，为什么我们开头那个福尔摩斯的推断是有效的。 福尔摩斯得出结论说 Jabez Wilson 从事过大量写作（c）。他的前提大致意思是，Wilson 的夹克上有一些特定的穿着标记（p）。现在，假如我们回到福尔摩斯时代的伦敦，聚集所有那些有上述磨损袖口的人，那么其中大多数会是职员，工作时间都花在写作上——或者我们可以假设如此。这样，给定 Jabez 的外套具有那些标记的条件下，Jabez 从事过大量写作的概率就比没有更高。福尔摩斯的推断的确是归纳有效的。

我将以我刚给出的这套工具产生的一个难题结束本章。我们已看到，概率可以计算为一个比率：我们选取特定的参照类，计算其中各组的数字，然后做些除法。但我们使用哪个参照类呢？在关于天气的示例中，我一开始限定了所考虑的参照类：某个特定星期的那些天。但现实生活问题不会以这种方式提出。回到 Jabez Wilson 的例子。要计算这种情况的相关概率，我建议采用的参照类是福尔摩斯时代的伦敦居民。但为什么是这个呢？为什么不是整个英国的居民，或者欧洲的居民，或者只是伦敦居民中的男性，或者只是那些能来拜访福尔摩斯的人？也许，在这里的某些情形，并没有太大差别。但在其他情形当然会有区别。比如，来拜访福尔摩斯的人都相对富有，不太可能穿二手外套。在更大的人群中情况会完全不同。那么，什么才应该是恰当的参照类呢？这是某种程度上让精算师（为保险公司计算风险系数的人）彻夜不眠的问题。

在上一个分析里，最精确的参照类似乎就只包含 Wilson 本人。毕竟，关于其他人的事实最终和他有多大关系呢？但这样一来，他要么从事过大量写作，要么没有。在第一种情况下，给定他有闪亮袖口的条件下他从事过大量写作的概率为 1，因而推断有效；在第二种情况下，概率为 0，因而推断无效。换言之，该推断的有效性完全依赖于结论的真假。因此，你无法为了确定结论的真假而使用这个推断。如果我们走到这个地步，那么给出的有效性概念就毫无用处了。

> 本章要点
>
> - 一个陈述的概率是它为真的情形的数量除以参照类中情形的数量。
> - $pr(\neg a) = 1 - pr(a)$
> - $pr(a \vee b) = pr(a) + pr(b) - pr(a \wedge b)$
> - $pr(a|b) = pr(a \wedge b)/pr(b)$

- 一个推断是归纳有效的，当且仅当给定前提（的合取）的条件下，结论的条件概率大于其否定在给定前提（的合取）的条件下的条件概率。

第 12 章：互逆概率：你无法忽略其差别！16

前一章我们对概率及其在归纳推断中可能起到的作用，有了一个基本理解。本章我们将对之作进一步考察。作为开始，让我们考虑一个非常著名的归纳推断。

物理宇宙不是杂乱无章的。它表现出非常独特的模式：物质组织成星系，星系又组织成恒星和行星系统，在其中一些行星系统上，物质又按某种方式组织产生了像你我这样的生物。我们如何解释这一点呢？你可能会说，物理学和生物学提供了解释。但为什么物理学和生物学的定律是那样的呢？毕竟，它们也可以是完全不同的样子。比如，重力可以是一种排斥力，而不是吸引力。那样的话，就永远不会有稳定的物质块，我们所知的生命在宇宙任何地方都不可能存在。这难道没有给我们极好的理由让我们相信，存在一个宇宙的造物者吗？一个为了某种目的创造了宇宙以及物理学和生物学定律的智慧之物。

这个论证常被称作“由于设计的论证（Argument from Design）”（为了证明上帝存在）。也许更好的叫法是，为了设计的论证（Argument *to* Design），但这不重要。让我们更仔细地考虑一下这个论证。该论证的前提 o 大致是这样的陈述：宇宙以某种方式形成秩序。结论 g 断言，存在一个造物者上帝。除非 g 为真，否则 o 是极不可能的；因此，该论证进一步得到：既然有 o，g 是很可能的。

现在，显然为真的是，给定 g 为真的条件下 o 的条件概率远大于给定 g 不为真的条件下 o 的条件概率：

1. $pr(o|g) > pr(o|\neg g)$

但这并没有给出我们想要的。o 要成为 g 的一个好的归纳理由，我们需要的是：给定 o 的条件下 g 的条件概率大于其否定的条件概率：

2. $pr(g|o) > pr(\neg g|o)$

而 $pr(o|g)$ 是高概率并不必然意味着 $pr(g|o)$ 也是高概率。比如，给定你在野外看见袋鼠的条件下，你在澳大利亚的概率会很高。（任何其他地方，它只能是从动物园跑出来的。）但给定你在澳大利亚的条件下，你在野外看见袋鼠的概率却很低。（我在澳大利亚住了 10 年才看见一只。）

$pr(o|g)$ 和 $pr(g|o)$ 称为（互）逆概率（*inverse probabilities*）。我们已经看到的是，设计论证要成立的话，二者之间的关系必须能让我们从 1 得到 2。是否如此呢？事实上，互逆概率之间存在一个非常简单的关系。回想一下上一章的公式 **CP**，根据定义：

$$pr(a|b) = pr(a \wedge b)/pr(b)$$

因此：

3. $pr(a|b) \times pr(b) = pr(a \wedge b)$

类似的：

$$pr(b|a) = pr(b \wedge a)/pr(a)$$

因此：

4. $pr(b|a) \times pr(a) = pr(b \wedge a)$

但 $pr(a \wedge b) = pr(b \wedge a)$（因为 $a \wedge b$ 和 $b \wedge a$ 为真的情形完全一样）。这样，由 3 和 4 可得：

$$pr(a|b) \times pr(b) = pr(b|a) \times pr(a)$$

假定 $pr(b)$ 不为 0——下面作这类假定时不再说明——我们可以将等式变形得到：

Inv：$pr(a|b) = pr(b|a) \times pr(a)/pr(b)$

这就是互逆概率之间的关系。（要记住这个公式，注意到下面的顺序可能会有所帮助：等式右边先是 b，后面跟着 a，然后是 a，后面跟着 b。）

使用 **Inv** 对 1 中的互逆概率进行改写得到：

$$pr(g|o) \times \frac{pr(o)}{pr(g)} > pr(\neg g|o) \times \frac{pr(o)}{pr(\neg g)}$$

两边消去 $pr(o)$ 得到：

$$\frac{pr(g|o)}{pr(g)} > \frac{pr(\neg g|o)}{pr(\neg g)}$$

或者，变形得到：

5. $\dfrac{pr(g|o)}{pr(\neg g|o)} > \dfrac{pr(g)}{pr(\neg g)}$

回想一下，设计论证要成立的话我们必须得到 2，而它等价于：

$$\frac{pr(g|o)}{pr(\neg g|o)} > 1$$

而要由 5 得到 2，合理的假设似乎只有 $\dfrac{pr(g)}{pr(\neg g)} \geq 1$，即：

$$pr(g) \geq pr(\neg g)$$

$pr(g)$ 和 $pr(\neg g)$ 的值称作先验概率（*prior probabilities*），即，先于使用任何证据（比如 o）时 g 和 $\neg g$ 的概率。因此，要使得该论证通过的话，我们似乎需要的是，存在造物者上帝的先验概率大于（或等于）不存在的概率。

是否如此呢？不幸的是，没有理由认为是这样。事实上，情况似乎正好相反。假设你不知道今天星期几。令 m 为假设：今天星期一。那么 $\neg m$ 就是假设：今天不是星期一。哪一个更有可能呢？ m 还是 $\neg m$？ 当然是 $\neg m$：因为比起今天是星期一，有更多的方式使得今天不是星期一。（今天可以是星期二，星期三，星期四……）关于上帝也是类似的。可以设想，宇宙可以有很多不同的方式。直觉上，其中有显著秩序的要相对少得多：秩序是某种特殊的东西。毕竟，这就是设计论证的前提。但这样一来，其中有秩序制定者（*orderer*）的可能宇宙也要相对少得多。由此推理可得，没有造物主的可能性比有造物主的可能性要大得多。

于是我们看到，设计论证并不成立。它之所以有诱惑力，是因为人们混淆了概率和它的逆概率，因而略过了论证的关键部分。

许多归纳论证要求我们对互逆概率进行推理。设计论证在这方面并不特殊。但许多论证在这方面做得更成功。让我举例说明。假设你去当地赌场。有两个轮盘赌的轮盘。称它们为 A 和 B。一个朋友告诉你，其中一个轮盘被做了手脚——不过朋友不能告诉你究竟是哪一个。一个公平的轮盘在一半时间停在红色，一半时间停在黑色，而被做过手脚的轮盘 3/4 的时间停在红色，1/4 的时间停在黑色。（严格说来，真实的轮盘还会偶尔停在绿色，但为了简单起见让我们忽略这一点。）现在，假设你观察其中一个轮盘，比方说 A，经过 5 次连续旋转后结果如下：

R, R, R, R, B

（R 表示红色，B 表示黑色）。你有理由推断这就是那个被做过手脚的轮盘吗？换言之，令 c 是大致这样的陈述：这个特定的序列出现，令 f 为陈述：轮盘 A 被做过手脚，那么从 c 到 f 的推断是个好的归纳推断吗？

我们需要知道 $pr(f|c) > pr(\neg f|c)$ 是否成立。使用公式 **Inv** 将其转换为逆概率之间的关系，我们得到：

$$pr(c|f) \times \frac{pr(f)}{pr(c)} > pr(c|\neg f) \times \frac{pr(\neg f)}{pr(c)}$$

两边同时乘以 $pr(c)$ 得到：

$$pr(c|f) \times pr(f) > pr(c|\neg f) \times pr(\neg f)$$

它是否为真呢？首先，f 和 $\neg f$ 的先验概率是多少呢？我们知道要么 A 要么 B 被做了手脚。我们没有更多的理由相信它就是 A 而不是 B，反过来也是如此。因此该轮盘是 A 的概率是 1/2，是 B 的概率也是 1/2。换言之，$pr(f) = 1/2$ 且 $pr(\neg f) = 1/2$。因此，我们可以在上面的等式两边消去它们，得到：

$$pr(c|f) > pr(c|\neg f)$$

给定轮盘像上面描述的那样被做了手脚的条件下，观察到由 c 陈述的序列的概率 $pr(c|f)$ 为：$(3/4)^4 \times (1/4)$（如果你不知道为什么，不要紧，相信我的计算），也就是 $81/4^5$，算出来等于 0.079。给定轮盘没有被做过手脚因而是公平的条件下，观察到该序列的概率 $pr(c|\neg f)$ 为 $(1/2)^5$（同样，请相信我的计算），算出来等于 0.031，小于 0.079。因此，推断是有效的。

这里我们计算先验概率的方法值得注意。我们有两种可能性：要么轮盘 A 被做了手脚，要么

轮盘 B 被做了手脚。而我们没有任何信息能区分这两种可能性。因此我们指派相同的概率给它们。这是一种被称作无差别原则（*Principle of Indifference*）的应用。该原则告诉我们，当我们有若干可能性，它们之间没有任何相关差别时，它们就有相同的概率。这样，如果总共有 N 个可能性，每一个的概率就是 $1/N$。无差别原则是一种**对称性**原理。

注意，我们不能在设计论证中应用无差别原则。在轮盘赌的例子中，有两个完全对称的可能情形：轮盘 A 被做了手脚，轮盘 B 被做了手脚。在设计论证中，有两个情形：造物者上帝存在，造物者上帝不存在。但这两个情形并不比今天是星期一和今天不是星期一更加对称。正如我们看到的，直觉上，没有造物主的可能性要远远大于有造物主的可能性。

无差别原则是对概率进行直觉推理的一个重要部分。但它并不是没有问题的，我们以指出其问题而结束本章。众所周知的是，它会在某些应用中导致悖论。这里便是一个。

假设一辆车在正午离开布里斯班，驶向 300 公里之外的某个城市。车以 50 至 100 公里每小时之间的某个恒定速度行驶。关于其到达时间的概率我们能说些什么呢？如果它以 100 公里每小时的速度行驶，那么它会在下午 3 点达到；如果它以 50 公里每小时的速度行驶，那么它会在下午 6 点到达。因此，它在这两个时刻之间到达。这两个时刻的中点是下午 4 点 30 分。那么根据无差别原则，这辆车在下午 4 点 30 分之前和之后到达的可能性一样大。但现在，50 公里每小时和 100 公里每小时之间的中点是 75 公里每小时。那么同样根据无差别原则，车以超过和低于 75 公里每小时行驶的可能性一样大。如果车以 75 公里每小时行驶，那么它会在下午 4 点到达。因此，在下午 4 点之前和之后到达的可能性一样大。特别的，在下午 4 点 30 分之前比之后到达的可能性要更大（因为多出半小时）。

我请读者自己思考这个例子。我们已经在一章之内讨论够多概率问题了。

本章要点

- $pr(a|b) = pr(b|a) \times \dfrac{pr(a)}{pr(b)}$
- 给定若干可能性，若它们之间没有任何相关差别，则它们有相同的概率（无差别原则）。

第 13 章：决策论：远大期望 [17]

让我们看一下关于归纳推理的一个最后问题。这个主题有时也称作**实践推理**（*practial reasoning*），因为它是对人们应该如何行动进行推理。下面是一条著名的实践推理。

你可以选择相信（一个基督教的）上帝存在，也可以选择不相信。让我们假设你选择相信。上帝要么存在要么不存在。如果上帝存在，一切好说。如果不存在，那么你的信念会带来小小的不便：它意味着你要浪费一点时间在教堂里，也许还要做少量其他你本来不想做的事；但这些都不是灾难性的。另一方面，现在假设你选择不相信上帝存在。同样，上帝要么存在要么不存在。如果上帝不存在，一切好说。但如果上帝**真的**存在，老兄你就有麻烦了！你死后会遭受很多痛苦；如果得不到一点宽恕的话，也许会永世不得翻身。因此，任何明智之人都应该选择相信上帝

存在。这是唯一谨慎的选择。

这个论证现在通常称作帕斯卡赌注（*Pascal's Wager*），以首次提出它的 17 世纪哲学家布莱斯·帕斯卡命名。关于这个赌注有什么要说的呢？

让我们稍微思考一下这类推理是如何进行的。我们先从一个不那么有争议的例子开始。当我们实施行动时，我们往往不确定最后的结果，它们可能不完全受我们的控制。但我们通常能估计各种结果的可能性有多大；同样重要的是，我们能估计各种结果对我们的价值有多大。按照惯例，我们可以通过对每个结果在下面的范围指派一个数字来度量其价值，这个范围在两个方向上都是无限的：

$$\ldots, -4, -3, -1, 0, +1, +2, +3, +4, \ldots$$

正数是好的，越往右越好。负数是差的，越往左越差。0 是无差别点：做不做都行。

现在，假设有一个我们可能会实施的行动，比方说去骑自行车。然而，天也许会下雨。不下雨的时候骑自行车会很有乐趣，因此我们会赋予它一个值，比方说 $+10$。但下雨时骑自行车就很悲催，因此我们会赋予它一个值，比方说 -5。对我们唯一能控制的事情——去骑自行车，我们应该赋什么值呢？我们可以只是把两个数字 -5 和 10 加在一起，但那样就漏掉了这个情景中的一个重要部分。也许下雨的可能性非常低，因此尽管可能下雨是不好的，我们也不希望给它太多权重。假设下雨的概率是，比方说 0.1；相应的，不下雨的概率就是 0.9。那么我们可以用相应的概率对值进行加权，得到总的值为：

$$0.10 \times (-5) + 0.9 \times 10$$

它等于 8.5，我们把它称作上述行动——去骑自行车的**期望值**（*expectation*）。（这里的“期望”是一个专业术语，它和我们日常使用这个词的意义实际上没有什么关系。）

一般地，令 a 表示：我们实施某个行动。简单起见，假设只有两个可能的结果。令 o_1 表示其中一个结果出现，令 o_2 表示另一个结果出现。最后，令 $V(o)$ 表示 o 为真时我们赋予它的价值。那么 a 的期望值 $E(a)$ 就是如下定义的数：

$$pr(o_1) \times V(o_1) + pr(o_2) \times V(o_2)$$

（严格说来，上述概率应该分别为条件概率 $pr(o_1|a)$ 和 $pr(o_2|a)$。但在这个例子中，出去骑自行车对下雨的概率没有影响。我们要考察的所有其他例子也都是如此。因此我们这里可以一直使用简单的先验概率。）

目前为止都没问题。但这如何帮助我们决定是否要去骑自行车呢？我们知道去骑单车的总体价值。正如我们已经看到的，其期望值为 8.5。不去骑自行车的期望值是多少呢？同样，要么下雨要么不下雨——概率和前面一样。现在的两个结果是：(i) 下雨我待在家里；(ii) 不下雨我待在家里。两种情况我都得不到骑自行车的乐趣。如果不下雨的话，情况可能或略遭一点。在那种情况下，我可能会对没有去骑自行车而感到懊恼。但两种情况都不会比淋成落汤鸡更糟。因此，如果下雨，价值可能为 0，不下雨，价值可能为 -1。现在就可以计算待在家里的期望值了：

$$0.10 \times 0 + 0.9 \times (-1)$$

结果为 -0.9，这给了我们需要的信息；因为我应该选择具有最高总体价值（即期望值）的行动。在这个例子里，去骑自行车的期望值为 8.5，而待在家里的期望值为 -0.9。因此，我应该去骑自行车。

这样，给定 a 和 $\neg a$ 之间的选择，我应该选择那个期望值更大的。（如果二者有相同的期望值，我随机选一个就行，比方说，通过掷硬币。）在前面的例子中，只有两种可能性。一般情况下，可能有更多可能性（比如，去骑自行车，去看电影和待在家里）。不过原理是一样的：计算每个可能性的期望值，然后选择期望值最大的那个。这种推理是来自逻辑学的一个分支——决策论（*decision theory*）的简单例子。

现在让我们回到帕斯卡赌注。在这个例子里，有两个可能的行动：相信或不相信；有两个相关可能性：上帝存在或上帝不存在。我们可以用下表来表示相关信息：

	上帝存在	上帝不存在
我相信（b）	$0.1\backslash+10^2$	$0.9\backslash-10$
我不相信（$\neg b$）	$0.1\backslash-10^6$	$0.9\backslash+10^2$

反斜杠左边的数字是相关概率，比如，0.1 是上帝存在的概率，0.9 是上帝不存在的概率。（我是否相信上帝对上帝是否存在没有影响，因此两行的概率一样。）反斜杠右边的数字是相关价值。我不太关心上帝是否存在；重要的是我没有弄错；因此我没有弄错的这两种情况价值都是 $+10^2$。（也许人们在这里的偏好并不完全一样，但我们会看到，这无关紧要。）上帝不存在时相信它存在，会有一点不便，因此价值为 -10。不过，上帝存在却不相信它存在，那就真的糟糕了，因此价值为 -10^6。

给定这些值，我们可以算出相关的期望值：

$$E(b)=0.1\times10^2+0.9\times(-10)\simeq0$$
$$E(\neg b)=0.1\times(-10^6)+0.9\times10^2\simeq-10^5$$

（$\simeq$ 意思是"约等于"。）我应该选择期望值更大的行动，即相信上帝存在。

你可能会认为，我选取的那些精确数值有点任意；它们的确如此。但事实上，那些精确数值具体是多少并没有多大关系。重要的是 -10^6 这个值。这个数字表示事情真的很糟糕。（有时，决策论专家会把它写作 $-\infty$。）它是如此糟糕，以至于淹没了所有其他的数字，即使上帝存在的概率很低。这就是帕斯卡赌注有冲击力的地方。

这个赌注或许看上去相当有说服力，但事实上，它犯了一个十分简单的决策论错误。它忽略了一些相关的可能性。不是只有一个可能的神，而是有很多：基督教的神（上帝），伊斯兰教的真主阿拉，印度教的婆罗门，还有更多各种小宗教所崇拜的神。如果上帝存在而你不相信它，你会有麻烦；但如果阿拉存在而你不相信它，也会有同样的麻烦，如此等等。此外，如果上帝存在，你却信仰阿拉，或者反过来，阿拉存在而你信仰上帝，那么情况会更糟。因为无论是在基督教还是在伊斯兰教中，信仰错误的神比做简单的无信仰者要还要糟。我们画出具有更现实信息的表格如下：

	没有神存在	上帝存在	阿拉存在	$\cdots$
无信仰（n）	$0.9\backslash+10^2$	$0.01\backslash-10^6$	$0.01\backslash-10^6$	$\cdots$
信仰上帝（g）	$0.9\backslash-10$	$0.01\backslash+10^2$	$0.01\backslash-10^9$	$\cdots$
信仰阿拉（a）	$0.9\backslash-10$	$0.01\backslash-10^9$	$0.01\backslash+10^2$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

如果我们根据这些即使有限的信息来计算期望值，我们会得到：

$$E(n)=0.9\times10^2+0.01\times(-10^6)+0.01\times(-10^6)\simeq 2\times10^4$$
$$E(g)=0.9\times(-10)+0.01\times10^2+0.01\times(-10^9)\simeq -10^7$$
$$E(a)=0.9\times(-10)+0.01\times(-10^9)+0.01\times10^2\simeq -10^7$$

情况看上去都很惨淡。但很清楚的是，信仰有神论的结果会更加糟糕。你不应信仰它们中的任何一个。

和其他章一样，让我以如下方式结束本章。我会给出一些理由，说明人们为什么会担心这里给出的一般框架——具体到这里就是，根据最大期望值进行决策的策略。

假设你在帕斯卡赌注中下错了注，最后进了地狱。几天后，魔鬼带着一笔交易出现。上帝已发号施令，说你可以得到某种宽恕。因此魔鬼谋划了一个方案。他会给你一个逃出地狱的机会。你可以掷硬币，如果正面朝上，你就可以出地狱、升天堂。如果反面朝上，你就要永远待在地狱。然而，硬币并不是公平的，魔鬼可以控制胜算。如果你今天掷硬币，正面朝上的机会是 $1/2$（即，$1-1/2$）。如果你等到明天掷硬币，机会就上升到 $3/4$（即，$1-1/2^2$）。你把信息总结如下：

	逃离地狱	待在地狱
今天掷硬币（d）	$0.5\backslash+10^6$	$0.5\backslash-10^6$
明天掷硬币（m）	$0.75\backslash+10^6$	$0.25\backslash-10^6$

逃离地狱有一个很大的正值，待在地狱有一个很大的负值。并且，这些值在今天和明天是一样的。如果你等到明天的话，你可能得在地狱里多待一天，这是真的，但和以后无穷无尽的日子相比，一天可以忽略不计。于是你开始计算：

$$E(d)=0.5\times10^6+0.5\times(-10^6)=0$$
$$E(m)=0.75\times10^6+0.25\times(-10^6)=0.5\times10^6$$

因此，你决定等到明天。

但明天魔鬼来告诉你，如果你再多等一天，胜算更高：它会升到 $7/8$（即，$1-1/2^3$）。我请读者自行计算：你应该决定再等一天。麻烦在于，如果你愿意等到下一天的话，魔鬼每天都来提供一个更高的胜算给你。胜算越来越高，日复一日：

$$1-1/2, 1-1/2^2, 1-1/2^3, 1-1/2^4, \ldots, 1-1/2^n, \ldots$$

每天你都进行计算。在第 n 天掷硬币的期望值是：

$$(1-1/2^n) \times 10^6 + 1/2^n \times (-10^6)$$

一点算术知识告诉我们，它等于 $10^6 \times (1-2/2^n) = 10^6 \times (1-1/2^{n-1}$。而等到下一天，第 $n+1$ 天的期望值也是一样的，只是把 n 换成 $n+1$，即 $10^6 \times (1-1/2^n)$，它要更大。（$1/2^n$ 小于 $1/2^{n-1}$。）期望值每天都在升高。

因此，每天你都理性地等到下一天。结果你永远也没有掷出硬币，因而你永远待在地狱！在**任何**一天掷出硬币都要好得多。那么，看上去你要做的唯一理性之事就是不再理性。

本章要点

- $E(a) = pr(o_1) \times V(o_1) \ldots pr(o_n) \times V(o_n)$，其中 $o_1, \ldots, o_n$ 表示 a 为真时所有可能的结果。
- 理性的行动是为真时具有最大期望值的行动。

第 14 章：停！发生什么了？

如果你一直读到了这里，就会对现代逻辑的基本思想有一个不错的理解。不过，也仅限于入门。现代逻辑的内容远胜于此，包含极为深刻而优美的结果。当然，我们不可能在这样一本书中全面评述这些结果。但本章和下一章，我们至少对它们提供一瞥。我们将看看一些关于形式推理哪些可以做哪些不能做的结果，以及这些结果的哲学意味。警告：这两章可能比前面的章节更难一点。我已竭力简单化处理，但我们涉及的是一些复杂的数学问题。说完这些，让我们继续本章的主题。

莱布尼茨——我们在第 6 章和第 9 章碰到的那个莱布尼茨——有一个梦想，一个终结争论之梦。每当我们有一个富有争议的论断，我们就可以用一种被称为普遍文字（characteristica universalis）的恰当语言把它写出来。然后，为了判定该论断的真假，我们演算（calculemus）就行。这种语言使得有一个计算过程（calculus ratiocinator）可以被用于判定该论断是否为真。

尽管莱布尼茨对于实现该计划确实建议了一些步骤，但它一度只是一个梦想。莱布尼茨时代的数学还没有达到可以处理其设想的程度。

我们时代的数学达到了。我们在前面的章节中看到的符号语言可以表达（至少一大类）真假未知的论断。剩下的问题就是，是否存在一个合适的计算过程。

答案（不幸）是，不存在——即使对范围十分有限的数学论断也是如此。该结果由英国数学家阿兰·图灵（1912-1954）于 1936 年证明。图灵是现代计算机科学的奠基者之一。当然，在图灵时代，没有任何为现在多数人所熟知的现代计算机这样的东西。但这些机器的**理论**远在计算机出现之前就已被图灵和其他人发明，剩下那些人只是去寻找如何在实践中实现这些想法——尽管图

灵本人在实际构建计算机器方面（例如在破解二战中德国海军电报密码的Enigma项目中），也取得过令人瞩目的进展。正如所料，图灵在计算上的兴趣与莱布尼茨之梦的联系并非巧合。

什么是计算机？最简单的情况下，计算机就是某种接收一个或多个输入，执行某个过程——数学家称之为算法（algorithm），该名称来自波斯数学家 Al Khwārizmī（780-850）——然后（如果运气好的话）给你一个输出的装置。

现代计算机的输入和输出有不同种类：数、文本、图片、声音。但对机器而言，这些都是数，这是它所能作用的全部对象。计算机的输入装置将输入转换成一串算法能作用的数。输出装置则将这一过程倒转过来。

不过，计算机存储数的形式并非我们在小学算术中熟知的那些。计算机的存储单元只能是两个状态中的某一个：开或关。因此，计算机只有两个基本字节的信息可以使用。人们可以把它们想成 1 和 0。任何数都可以用这两个数字表达。这是由二进制算术完成的。（即，如果你只有两个指头你会如何计数。）在标准（十进制）算术中，数实际上是一种表达 10 的幂之和的方式。比如，4302 就是：

$4 \times 10^3 + 3 \times 10^2 + 0 \times 10^1 + 2 \times 10^0$

（10^0——实际上任何数的 0 次幂——就是 1）。类似的，二进制数也是一些幂之和，只是这次是 2 的幂。于是，1011 就是：

$1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$

下表给出了前几个十进制数与二进制数之间的转换。

十进制	二进制
0	0
1	1
2	10
3	11
4	100
5	101

所以我们可以将计算（算法）看作某种作用在这种二进制数上的东西。

输入和输出就讲这么多，但什么是计算？计算由一组在标准计算机程序中可找到的那类规则给定。这些程序由很多不同的语言写成，其精确细节在这里并不重要。一个相当无趣的程序可能长下面这个样子：

1. 若 $x = 0$ 则输出 x；否则跳至第 2 行
2. 令 $x := x - 1$

3. 跳至第 1 行

左边的数是行号。输入是某个数 x。第 1 行测试这个数是否是 0，如果是，就输出这个数。否则跳到下一行。这一行将 x 减 1，然后计算又回到第1行。稍微思考一下就会发现，这个程序所做的就是把任何输入的数不断减 1，直至减到 0 作为输出。

目前都没有问题。接下来——这是现代计算机真正聪明之处——当计算进行时，计算机不必等待人输入每一行程序。程序本身就存储在计算机中。当然，是存储成一个数。计算机没法存储任何其他东西。（事实上，我们可以把计算机在任何时刻的整个状态想成一个巨大的 0、1 字符串——一个庞大的二进制数！）我们可以把存储在计算机中代表程序的数看作程序的"编码"。若 n 是任何一个数，则令 P_n 为编码为 n 的程序。（若由于计算机中的编码方式，n 恰好不是任何程序的编码，我们就让 P_n 默认为上面那个简单的程序。）严格来说，程序本身不关心它运行的算法有多少个输入。它只是当它被告知时，取用任何位于计算机中的信息。不过，作为约定，我们可以假定，除了那些相干的信息被合适的比特填充，所有输入的信息比特都为 0。

现在，有时一个给定输入的程序会产生一个输出；但有时它会永远运行下去。比如，考虑下面这个程序：

1. 令 $x := x + 1$
2. 若 $x = 0$ 则输出 0；否则跳至第 1 行

该程序将某个输入加 1，然后测试它是否为 0，如果是，则输出 x。但它当然不是 0（我们的二进制数总是大于等于 0），因此我们跳回第 1 行，然后重复这一过程。加1我们永远也得不到 0，因而计算永不停止，就这样以无限循环的方式一直运行（现实中，计算会在机器报废时，或在 x 大到超出计算机能处理的范围时停止）。为将来讨论方便，我们称这个程序为 L（表示循环，looping）。

良好构造的程序会被设计成这绝不会发生。程序员通过分析程序，保证它永远不会进入这种无限循环。但这总是能做到吗？是否有算法可以对任何程序和输入，判定该程序和输入下的计算是否会终止？

答案是：没有。这就是图灵所证明的。这是一个相对简单，但却十分精巧的证明。它使用了归谬法（reductio ad absurdum），先假定和我们希望证明的结论相反的情况成立，然后表明这会导致某种不可接受的结果。

假定有一个这样满足要求的算法，称之为 A。于是，当 A 应用于两个输入 n 和 i，它输出 1 或 0。1 意味着程序 P_n 输入 i 的计算终止；0 意味着它不终止。

现在考虑下面这个算法。让我们称之为 T（表示图灵，Turing）：

- 运行算法 A（输入 x 和 x）。该算法终止并给出 1 或 0 的结果。
- 若结果是 0，则输出 1
- 若结果是 1，则运行 L（输入 x）。

该程序输入 x 会做什么呢？本质上，它相当于运用 A 判定程序 P_x 输入 x 是否会停止。若计算不会停止，则该程序输出 1。特别的，该程序会停止。但若计算确实停止了，则整个计算会跳至一个无限循环并永不停机。

我是用一种颇为"高层次"的术语来描述程序 T 的。但这并不会产生特殊难题。任何理解信息如何编码并存储于计算机，使用一种可以直接利用这些数据的语言的娴熟程序员，都能写出这样一个程序。

现在，为了完成证明：T 自身也有一个编码，称之为 t。我们可以运行 T 并以 t 自身作为输入。如果计算*停止*，则运行 A（输入 t 和 t）的计算停止并输出 1。但这样运行 T 的计算就会陷入无限循环而不会停止。反之，若运行 T（输入 t）的计算*不停止*，则运行 A（输入 t 和 t）的计算停止并输出 0。这样，运行 T 的计算就停止并输出 1。因此，若计算不停止，则它停止！无论哪种情况，我们都得到某种不可能的结果。因此我们最初有一个算法 A 的假设必定为假。

图灵证明的精巧之处是某种自指。（我们在第5章碰到过自指。）它让某个假定的程序应用于它自身的编码。这有时被称作对角化（diagonalization），一种由伟大的德国数学家康托（Georg Cantor, 1845-1918）在研究无穷时发明的方法。考察下表你就会明白它为何叫对角化。

	0	**1**	**2**	**3**	**4**	$\cdots$
0	$\mathbf{a_{00}}$	a_{01}	a_{02}	a_{03}	a_{04}	$\cdots$
1	a_{10}	$\mathbf{a_{11}}$	a_{12}	a_{13}	a_{14}	$\cdots$
2	a_{20}	a_{21}	$\mathbf{a_{22}}$	a_{23}	a_{24}	$\cdots$
3	a_{30}	a_{31}	a_{32}	$\mathbf{a_{33}}$	a_{34}	$\cdots$
4	a_{40}	a_{41}	a_{42}	a_{43}	$\mathbf{a_{44}}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

最左列是程序的编码。最上面一行是输入。表格中的项 a_{xy} 是编码为 x 的程序输入 y 后的输出。若该计算不终止，则我们可以用符号 ∞ 表示。算法 A 所做的——假定它存在——就是计算 a_{xy} 的值是否为 ∞。T 作用于该计算在对角线上的结果（粗体显示），确保它的行为与 P_x 作用于 x 不同。这样 T 就不可能出现在最左列。但每个程序都出现在这一列。因而 T 不存在。而 T 是由算法 A 定义的，该定义并无问题，因此 A 也就不可能存在。

这个结果称为停机定理（Halting Theorem）。它表明，没有算法可以判定任何给定的程序在给定的输入下是否停机（当然，我们尽可在特殊情形做到这一点）。现在回到莱布尼茨之梦。我们看到的是，有一些数学问题，比如这个停机问题，并没有算法判定其真假。莱布尼茨之梦无法实现。

我在结束前面的章节时会指出，为何其中的看法可能存在争议。让我以同样的方式结束本章。在标准数论假设下，图灵的证明是无可辩驳的。这是足够好的数学。但在我给的上述证明中，有一个到现在我还未作评论的假设。该假设就是，任何我们能看作算法的东西都能写成计算机程序。如果并非如此，则图灵的证明只是表明了没有计算机程序能判定任何计算是否停止。但或许存在某种其他的算法——一种或许能在莱布尼茨的计划中有效使用的算法。

每个算法都能写成计算机程序，这一论断被称为**丘奇-图灵论题**（Church-Turing Thesis），以图灵和美国数学家Alonzo Church（1903-1995）命名。它本身不能进行数学证明，因为证明只能作用于被精确定义的概念，而尽管什么是计算机能做的*可以*被精确的数学语言定义，算法却只是一个非形式的直观概念。粗略而言，算法是可以无需猜测和创造性的逐步完成的过程——而这或多或少是模糊概念。

长期以来，丘奇-图灵论题都被数学家所接受。曾有一段试图反驳它的历史。这些反驳都试图构造某种直观上可以被看作算法，但却不能在计算机上编程的东西。这些尝试都失败了。因此，丘奇-图灵论题就被普遍接受了。

然而，现在也有研究领域研究除了那种运用于台式计算机中的计算的计算方法。这些计算有时被称作超计算（hypercomputation）。一个例子是：有些涉及的方法使用模拟量，而不是数字量。（模拟量是连续的，如长度；而数字量则是离散的，如二进制数。）另一个例子是：有些涉及的方法诉诸于广义相对论中的时空性质，其中时间会"加速"。不过，现在谈这些研究的最终结果如何还为时尚早。

本章要点

- 算法可以被编码。
- 假若有算法 A，可以判定编码为 x 的算法（即 P_x）输入 y 是否终止，则我们可以定义算法 T，它计算 A 输入 x 和 x 得到的值，并利用该结果确保它自身的输出“在对角线上”不同于每个 P_x。
- T 自身必定有一个编码 t。运行 T（输入 t）将产生一个不可能的结果。
- 因此没有这样的算法 T，因而也就没有这样的算法 A。

第 15 章：也许为真——但你无法证明！

本章标题听上去像是小毛贼可能对警察说的话。但实际上，它宣称了过去100年里逻辑学中另一个最重要的结果。（快速警告：前面的章节中我使用了小写字母 (a, b 等等) 表示语句。本章我将使用大写字母(A, B 等等)表示语句，以避免任何可能与数字的混淆。）

莱布尼茨不是逻辑学史上唯一有雄心壮志计划的逻辑学家。另一个人物是20世纪最重要的数学家之一，大卫·希尔伯特（David Hilbert，1862-1943），生活和工作于哥廷根。其计划通常称作**数学基础的希尔伯特纲领**（Hilbert Program in the Foundations of Mathematics），该纲领试图证明数学是一致的，即证明在数学中永远不会证明形如 A 且$\neg A$ 的东西。1920年代，数学家仍然处于发现罗素悖论（我们在第5章见过）及其他同类悖论带来的震惊之中。这类悖论冲击到数

学最核心的部分。希尔伯特想确保这种事情以后再也不会发生。

我们在这里要稍微小心一点。对一致性的证明当然也是数学证明。若数学是不一致的，或许它终究可以证明其一致性。事实上，若使用的逻辑是我们在第2章中见过的那种，若数学是不一致的，则它能证明一切！正如我们在那里见到的，在这种逻辑中，任何命题都可以从矛盾中得出。因此，对一致性的证明必须在某种特别安全的数学推理中完成。希尔伯特称这种推理为**有穷的**（finitary）。不过这与我们要讲的内容并不相关。

有关的是下面这些。要证明某个东西是一致的，你首先要对它有个确切的了解。如果你要运用数学证明它，你要对数学有个精确的界定。因此，作为其主要主要纲领的预备步骤，希尔伯特需要为数学提供一个恰当的公理系统，从而证明其一致性。

一个公理系统由一组**公理**（axioms）构成。公理是无需证明即可被接受的东西。（公理集可以是有穷的，也可以是无穷的。但如果是无穷的，则我们要能识别某个东西是不是公理。具体而言，需要有一个算法来判定这件事。）该系统中的一个**证明**（proof），就是一个陈述序列，其中每个陈述要么是一条公理，要么可以从前面的陈述通过推演得到。该系统的**定理**（throrems），就是每个证明最末尾的陈述。于是，定理就是从公理出发，最终能推演得到的东西。

公理化方法是数学中的一个古老方法。它被古希腊数学家欧几里得（Euclid，公元前 4 世纪中叶至 3 世纪中叶）应用于几何。不过，有点奇怪的是，直到 20 世纪，该方法才在数学中被广泛应用。在此之前，数学中被公理刻画的部分，就只有几何（或者更准确地说，是各种几何；19世纪初，数学家已经知道除了欧几里得几何之外的其他几何——非欧几何）和抽象代数的某些部分。

希尔伯特的宏伟计划则要求所有数学都被公理化。即，我们需要一个公理系统，其所有定理恰好是所有真的数学论断（无论人们如何理解数学中的真是什么）。这种公理系统的存在性，被或许是20世纪最著名的逻辑学家、奥地利数学家哥德尔（Kurt Gödel，1906-1978）反驳了。哥德尔表明，别说整个数学，就是对数学中关于自然数（0，1，2，...）的部分，也无法提供这样一种公理系统。数学中的这部分通常被称作**算术**（arithmetic）。因此，哥德尔证明的是，即使有公理系统可以刻画算术的*部分*真理，也没有公理系统可以刻画算术的全部真理。正如逻辑学家所说，该公理系统必定是**不完全的**（incomplete）。

有了我们在第14章学到的东西，哥德尔的结果就可以以相当直接的方式证明。证明仍然是归谬法。考虑一个能谈论数并能表达对数的操作（加、乘等）的符号语言。这种语言并不难得到。在该语言中可表达如下陈述：编码为n的程序输入 i 后会终止。（要证明这一点需要费点功夫，但并不难。）称该陈述为 S_{ni}。现在假设存在一个完备的关于算术的公理系统，即一个其所有定理恰好是该语言中所有真陈述的公理系统。于是就存在一个判定 S_{ni} 是否为真的算法：我们只要以一种系统的方式开始证明所有定理即可，确保任何可被证明的迟早被证明。（不难设计这样一个程序。）这样，迟早要么得到 S_{ni} 的证明，要么得到 $\neg S_{ni}$ 的证明，从而判定上述问题。（我们可能不知道要多久才能完成，但这不重要）。但第 14 章的停机定理告诉我们，不存在一个算法能判定这件事。因此，不存在这样的公理系统。

我刚刚概述的证明并非哥德尔本人的证明。事实上，哥德尔的证明出现于 1931 年，而图灵证明停机定理是在 5 年之后。但哥德尔的证明和图灵的证明一样精巧——如果说不是更精巧的话——而且也使用了某种自指。哥德尔证明的概要如下。

假设我们在有一个其语言有足够表达能力的算术公理系统。该系统的所有定理都为真，但它可能无法证明该语言中所有真的陈述，因此它可能是不完全的。该语言的一个数学陈述就是一段文本。正如我们在第 14 章指出的，这样的陈述可以被编码为一个数。一个计算机程序就是一个陈述序列，同样正如我们在第14章指出的，它也可以被编码为一个数。但一个数学证明也只是一个陈述序列，因而也可以以相同方式编码为一个数。

现在考虑这个陈述：x 是编码为 y 的陈述的证明（的编码）。该陈述本身是一个关于数的陈述，它可以表达为该语言中的一个语句 $Prov(x,y)$。并且，只要该系统的公理足够强，就有：

- 若 m 的确是编码为 n 的定理的证明（的编码），则 $Prov(m,n)$ 可在该公理系统中被证明。

要证明这些事情实际上相当难，需要一些可观的数学。但它可以做到，正如哥德尔所证明的。

说编码为 y 的陈述是可证的，即是说它存在一个证明：$\exists x Prov(x,y)$。但是——这是哥德尔证明的真正巧妙之处——通过一个精巧的构造，我们可以找到一个本质上形如 $\neg\exists x Prov(x,n)$ 的语句，其编码恰好是 n 本身！该陈述相当于说，本陈述（在系统中）不可证。称该语句为 G（代表哥德尔）。

现在，假设 G，即 $\neg\exists x Prov(x,n)$，在该公理系统中是可证的。那就有个数 m 是 G 的某个证明的编码。这样，$Prov(m,n)$ 就为真，因而在公理系统中是可证的（根据上面的 •）。但由此可得 $\exists x Prov(x,n)$ 是可证的。于是该公理系统就是不一致的。假设系统要是一致的，则 G 就是不可证的。但如此一来它就是真的，因为它恰好就是说它自己不可证！因此，该语言中有真的语句在系统中无法被证明。这样，算术就没有完全的公理系统。

哥德尔定理——无论它是如何证明的——明确表明了公理化方法在数学中的限度（这不是说公理化方法不应该被使用：事实上，如今公理化方法比以往任何时候都更加是数学方法的主要部分）。特别的，它给了希尔伯特纲领致命一击。算术无法被公理化——更不用说整个数学了。哥德尔的结果已被认为有很多哲学上的后果，包括关于数的本质，我们关于数的知识，甚至是人类心灵的本质。争论仍在继续，我们这里就不展开了。

让我再次以提出关于该定理的一个问题结束。我给的两个证明在数学上是无可辩驳的。但两个证明都做了一个假定（希尔伯特当然认为——或至少希望——它是成立的）：即关于数的真陈述，以及刻画其任何部分的公理系统是一致的。这在哥德尔的证明中是相当明确的假定，而我也在第一个证明中给出了。该证明假定 S_{ni} 和 $\neg S_{ni}$ 中恰好有一个会出现，从而确定事情是如此还是那般。但如果系统是不一致的，这就不再成立：可能两个一起出现，使得事情悬而不决！

现在，正如有些读者可能看到的，哥德尔证明中使用的语句 G 很像我们在第 5 章见过的说谎者悖论的表亲。两个都对某个语句，即它自己，说了它不具有某个关键性质。事实上，有一个关于可证性的悖论与说谎者悖论密切相关。考虑如下语句：**本语句不可证**。假设它是可证的，则它是真的。因而它是不可证的。因此（据归谬法），它不可能是可证的。但我们刚刚证明了这一点，因此它是可证的！

如果我们试图对语句 G 在其公理系统中进行这种推理，则它无法在该系统中再现。或许令人惊讶的是，若系统是一致的，则可证的都是真的这一论断并不能在系统中得到证明。（这一点由德国数学家Martin Löb（1921-2006）在 1955 年证明，因而通常称作 **Löb**定理。）因此该悖论无法用于证明公理系统的不一致性。然而，它确实表明，自指悖论就藏在算术周围。考虑到这一点，人们或许不应该如此自信地认为，关于数的真理是一致的。

本章要点

- 一个算术公理系统是完全的，若它能证明其语言中的所有真语句。
- 没有其语言具有足够表达力的公理系统是完全的。
- 该结论可由停机定理推出。
- 它也可以通过考虑本质上相当于说本语句（在系统中）不可证这样的语句而证明。

尾声：哥德尔第二不完全性定理

本章我们考察的结果有时称作哥德尔第一不完全性定理。在证明这一结果的同时，哥德尔还证明了另一个结果，称作第二不完全性定理。它本质上表明了，如我我们有一个我们考察过的那种一致的公理系统，则它无法证明被认为表达了该系统一致性的语句——只要该系统是基于我们在第2章中看到的逻辑。事实上，有了 Löb 定理，可以很容易证明这一点。

首先，注意到既然系统能证明 $\neg 0 = 1$，那么若它还能证明 $0 = 1$，则它就是不一致的。反过来，若它是不一致的，则它能证明 $0 = 1$，因为从矛盾可以推出任何结论，这是我们在第 2 章中提到的。因此，系统是一致的一个简单说法是说它无法证明 $0 = 1$。

现在，设 A 是该语言中的任何语句，记其编码为 $\langle A \rangle$，则 Löb 定理更精确一点的表述如下：

- 若系统能证明语句 $\exists x Prov(x, \langle A \rangle) \to A$，则它也能证明语句 A。

即：

- 若系统无法证明语句 A，则它无法证明语句 $\exists x Prov(x, \langle A \rangle) \to A$。

因此，作为特例，若它无法证明 $0 = 1$，则它无法证明：

- $\exists x Prov(x, \langle 0 = 1 \rangle) \to 0 = 1$

换言之，若它是一致的，则它无法证明 $\exists x Prov(x, \langle 0 = 1 \rangle) \to 0 = 1$。

但这样一来它就也无法证明 $\neg \exists x Prov(x, \langle 0 = 1 \rangle)$，因为无论 C 是什么，$A \to C$ 都可以由 $\neg A$ 推出（如第 7 章中的真值表显示的，若 A 为假，则 $A \to C$ 为真）。因此若系统是一致的，则它无法证明 $\neg \exists x Prov(x, \langle 0 = 1 \rangle)$，即（无法证明）它是一致的。这就是哥德尔第二不完全性定理。

该定理给了希尔伯特纲领第二个打击。回顾一下，该纲领的目标首先是公理化数学，然后其次是证明该公理系统是一致的。这样的证明当然是数学证明，因而在系统自身内可以实现。第一不完全性定理表明，纲领的第一步不可能实现。

第 16 章：一点历史与进阶阅读

本书我们考察过的思想是由不同历史时期和不同地方发展而来的。本章我们将描述逻辑学的历史，将这些思想放入它们的历史背景中考察。我将首先从总体上简要概述一下逻辑史，然后逐章深入考察，解释这些细节如何纳入到更大的历史背景中。

在介绍逻辑史的同时，我还会给出一些进一步阅读的建议，如果感兴趣你可以对一些问题作进一步考察。这项工作并不像想的那么容易。大体来说，逻辑学家、哲学家和数学家喜欢为彼此进行写作。找到为初学者写的东西不容易，但我已经尽力了。

在西方思想史上，逻辑学有三大发展时期，其间夹着相对贫瘠的时期。第一大发展时期是公元前 400 年至公元前 200 年的古希腊。这一时期最有影响的人物是亚里士多德（Aristotle，384-322）。我们在 第 6 章见过他。亚里士多德发展出一套关于“三段论”推断的系统理论。这种推断的形式如下：

所有[有些] A 是[不是] B。
所有[有些] B 是[不是] C。
因此，所有[有些] A 是[不是] C。
亚里士多德一生大部分时间生活在雅典，创立了一个叫“学园”的哲学学派。但在差不多同一时期，在雅典 50 公里以西的麦加拉还有一个逻辑学派。关于麦加拉的逻辑学家，我们所知不多，但他们似乎曾对条件句和逻辑悖论特别感兴趣。欧布里德（我们在第 5 章和第 10 章见过他）就是麦加拉人。大约公元前 300 年，雅典兴起了另一个重要的哲学运动，叫斯多葛主义（学派），因其早期聚会的地方——门廊（希腊语为 stoa）而得名。尽管斯多葛学派关注的哲学问题远比逻辑更宽，逻辑仍然是其中重要的一部分。不管怎样，斯多葛派的逻辑学家关心的一个主要问题是研究否定、合取、析取和条件句的表现。

还要注意的是，大约在所有这一切发生在古希腊的同时，逻辑理论也在印度得到发展，主要由佛教逻辑学家完成。尽管这些理论很重要，却从未发展到西方逻辑达到的精致水平。

西方逻辑的第二个增长期是在 12-14 世纪巴黎和牛津这样的中世纪欧洲大学。中世纪的逻辑学家包括著名的邓斯·司各脱（Duns Scotus，1266-1308）和奥卡姆的威廉（William of Ockham，1285-1349）。他们大力而系统化地发展了从古希腊继承的逻辑。这一时期之后，逻辑学直到 19 世纪下半叶之前都停滞不前。其间唯一的亮点是莱布尼茨（Leibniz，1646-1716），我们在第 6 章和第 9 章见过他。莱布尼茨预见了逻辑学的某些现代发展，但他那个时代的数学还没有发展到能实现他的想法。

19 世纪抽象代数的发展恰好提供了逻辑所需要的技术，引发了第三个时期，也可能三个中最伟大的一个时期。全新的逻辑思想由弗雷格（Frege，1848-1925）和罗素（Russell，1872-1970）这样的思想家发展出来。我们在第 2 章和第 4 章分别见过这两个人。从这些工作发展出来的逻辑理论通常被称作**现代逻辑**，以区别于之前的**传统逻辑**。整个 20 世纪，逻辑学都在飞速发

展，而且没有减缓的迹象。

Kneale 和 Kneale (1975)[18] 是一部逻辑史的标准著作。它稍微有点过时。该书倾向于认为，早期现代逻辑学家已经把所有事情都弄清楚了。这与其说是正当的结论，不如说是乐观的态度。但它仍然是一本极好的参考书。关于本书覆盖的主题，还可参考 Zalta (1995-) 中的很多文章，不过其中一些也相当技术化。

第 **1** 章：有效性

演绎和归纳有效性的区分可以追溯到亚里士多德。从那时起关于演绎有效性的理论就被清楚地表述。第 1 章描述的观点——即一个推断是演绎有效的，当且仅当在前提为真的任何情形结论也为真——被认为可追溯到中世纪的逻辑；但它的清晰表述是现代逻辑的一个核心部分。要注意的是，我所说的情形更常见的说法是解释（*interpretation*）、结构（*structure*），有时也叫模型（*model*）。“情形”这个词本身在逻辑学的某个领域有不同的专门含义。刘易斯·卡罗尔（本名为 Charles Dodgson）本人就是个很好的逻辑学家，出版了许多传统逻辑的著作。

第 **2** 章：真值函数——抑或不是?

矛盾蕴含一切这个论证是中世纪的发明。具体是谁发明的并不清楚，但在司各脱的论著中明确可以找到。对否定、合取和析取的真值函数理解本身似乎在中世纪就出现了。（斯多葛派的解释并不是现代意义上的真值函数。）但其完全清楚的表述是在现代逻辑的奠基者弗雷格和罗素那里出现的。该理论的一位现代反对者是 Strawson（1952，第 3 章）。

第 **3** 章：名称与量词

区分名称与量词主要是现代逻辑的产物。事实上，对量词的分析常被认为是现代逻辑的决定性时刻。它由弗雷格提出，而后被罗素接受和发展。大约与此同时，美国哲学家和逻辑学家 C.S. 皮尔士（Peirce）发展了类似的思想。$\exists x$ 常被称作存在量词（*existential quantifier*），但这个术语夹带了某种有争议的存在性理论。刘易斯·卡罗尔关于爱丽丝的作品充满了哲学玩笑。对它们的一个极佳评论，可参见 Heath（1974）。Heath 自己也创作了很多关于空无（nothing）的玩笑，参见 Heath（1967）。

第 1-3 章中解释的理论可以在任何标准的现代逻辑教材中找到。Hodges（1977）是其中一本，它没有把难度设定得高不可攀，Lemmon（1971）也是如此。

第 **4** 章：摹状词与存在

将摹状词作为一个重要的逻辑范畴分离出来也是现代逻辑中才有的。也许其中最著名的分析是罗素在 1905 年给出的。本章给出的解释并不是罗素本人的，但精神实质是很相近的。一些（但并非全部）标准的现代逻辑教材讨论了摹状词。Hodges（1977）对摹状词有很好的清晰解释。

第 **5** 章：自指

各种不同版本的说谎者悖论在古希腊哲学中都可以找到。整个中世纪的逻辑都在发明和讨论更多的自指悖论。20 世纪之交又发现了甚至更多的自指悖论——这次是在数学本身最核心的地方。从那时起，自指悖论就成为逻辑学中一个非常中心的议题。解决它们的建议不计其数。可能有些语句既不真也不假这个想法可追溯到亚里士多德（《解释篇》，第 9 章）；然而，他应该不会赞同与之对称的想法，即有的语句可以既真又假。可能存在这样的语句，且悖论句可能就在其中，这是一种过去 40 年被一些逻辑学家提出的非正统观点。关于自指悖论的讨论容易很快就变得非常技术性。在 Read（1994，第 6 章）和 Sainsbury（1995，第5、6章）中可以找到很好的介绍性讨论。整个领域仍然是高度有争议的。

第 **6** 章：必然性与可能性

对涉及模态算子推断的研究可追溯到亚里士多德，并在中世纪继续发展。现代研究则大致在 1915-1930 年间由美国哲学家 C.I.刘易斯开始进行。可能世界的概念可以在莱布尼茨的著作中找到，但本章应用它的方式主要归功于另一个美国哲学家索尔·克里普克（Saul Kripke）。他在 1960 年代产生了这些想法。该领域的一本标准入门书是 Hughes 和 Cresswell（1996）；但在掌握一本更现代的逻辑导论书之前，你不太可能从中学到很多。亚里士多德的宿命论论证来自于《解释篇》第 9 章。他认为这个论证是谬误的，不过不是由于本章给出的理由。在 Haack（1974，第 3 章）可以找到一个相当浅显易懂的讨论。本章结束时给出的论证是“主论证”（Master Argument）的一个版本，由麦加拉学派逻辑学家 Diodorus Cronus 提出。

第 **7** 章：条件句

关于条件句本质的争论可追溯到麦加拉学派和斯多葛学派，他们提出了许多不同的理论。这个问题在中世纪也得到了广泛的讨论。条件句是真值函数的思想是麦加拉学派的一种观点，在早期现代逻辑中被弗雷格和罗素所采用。本章给出的解释在中世纪逻辑中肯定可以找到，其现代形式归功于 C.I.刘易斯。他围绕条件句发展了模态逻辑。会话蕴含的概念归功于1970年代的英国哲学家保罗·格赖斯（Paul Grice）（不过格赖斯是用它来为实质条件句进行辩护）。条件句的本质仍然是高度有争议的。Read（1994，第 3 章）是一个很好读的导论介绍。Sanford（1989）的第一部分也是如此。

第 **8** 章：将来和过去

许多中世纪逻辑学家讨论了时态推理。本章描述的方法主要由新西兰逻辑学家阿瑟·普赖尔（Arthur Prior）在 1960 年代受模态逻辑发展的启发而发明。在 Øhrstrøm 和 Hasle（1995）中可找到该主题的一个很好读的解释。McTaggart 的论证最初出现于 1908 年，不过他的表述和我的稍有不同。我采用了 Mellor（1981，第 7 章）中的表述。

第 **9** 章：同一性与变化

“同一之是”与“谓述之是”的区分在可追溯到古希腊哲学的柏拉图（亚里士多德的老师）。我在本书对同一性给出的解释，其起源已不可考。等值替换的思想在欧几里得（Euclid，公元前 300 年）的著作中就已出现。类似于本章的解释可以在奥卡姆（Ockham）的著作中找到，当然也包括莱布尼茨的著作。它的现代形式可以在弗雷格和罗素的著作中找到。在最标准的现代逻辑教材中也有对它的表述，如 Hodges（1977）和 Lemmon（1971）。关于同一性的谜题在哲学中不胜枚举。据我所知，本章结束时谈到的谜题应归于普赖尔（Prior）。

第 **10** 章：模糊性

堆积问题可追溯到麦加拉派的逻辑。本章开头的问题是被称作“提修斯之舟”的一个版本。这艘船在假想中被一块木板一块木板地重新建造。据我所知，这个例子最早被 17 世纪英国哲学家托马斯·霍布斯（Thomas Hobbes）在他的《哲学原理》（Elements of Philosophy）“论物体”这一节中使用。对这类问题的集中研究主要是过去 30 年的事。本章描述的逻辑细节最早由波兰逻辑学家乌卡西维茨（Jan Łukasiewicz） 在 1920 年代提出，完全独立于对模糊性的考虑。（他的最初动机来自于亚里士多德的宿命论论证。）在 Read（1994，第 7 章）和 Sainsbury（1995，第 2 章）中可以找到对模糊性的出色讨论。一部篇幅长得多的介绍是 Williamson（1994）。

第 **11** 章：概率

在历史上，与演绎有效性相比，归纳有效性发展得不够充分。概率论是在 18 世纪和机会博弈的联系中发展起来的，主要由说法语的数学家完成，如：皮埃尔·德·拉普拉斯（Pierre de Laplace）和非凡的伯努利（Bernoulli）家族成员。将概率论应用于归纳推理主要归功于德国逻辑学家鲁道夫·卡尔纳普（Rudolph Carnap）在 1950 年代的工作。概率概念有很多种。本章所描述的通常称作**频率解释**（frequency interpretation）。一本很好的对整个领域进行介绍的读物是 Skyrms（1975）。

第 **12** 章：逆概率

研究互逆概率之间的联系可追溯到 18 世纪英国数学家托马斯·贝叶斯（Thomas Bayes）。本章描述的联系常被（不正确地）称作贝叶斯定理。关于无差别原则的问题也可追溯到概率论的起源。一本标准的介绍此类推理的读物是 Howson 和 Urbach（1989）；但这本书不适合那些畏惧数学的人。

第 **13** 章：决策论

决策论也根源于 18 世纪对概率论的研究，但在 20 世纪才成为一门严肃的学科，在经济学和博弈论中可以找到很多重要的应用。一本出色的入门读物是 Jeffrey（1985）；不过同样，这本书也不适合那些畏惧数学的人。本章结束时谈到的问题来自 Gracely（1988）。

我们在本书见到的许多论证都以这样或那样的方式和上帝有关。这并不是因为上帝是个特别逻辑性的话题。只是因为哲学家们有很长一段时间都在提出关于上帝的有趣论证。在第 3 章，我们见过宇宙论论证。它最著名的版本也许是由中世纪哲学家托马斯·阿奎那（Thomas Aquinas）提出的。（他的版本比第 3 章中的论证要复杂得多，也不存在我们那里指出的问题。）上帝存在的本体论论证是由中世纪哲学家坎特伯雷的安瑟伦（Anselm of Canterbury）提出的。第 4 章给出的版本本质上应归于 17 世纪哲学家勒内·笛卡尔（René Descartes）在其“第五个沉思”中给出的论证。设计论证的生物学版本在 19 世纪时很流行，但受到了进化论的摧毁。一本出色而又篇幅短小的关于论证上帝存在的参考书是 Hick（1964）。

第 **14** 章：停！发生什么了？

计算理论由诸如 Alan Turing，Alonzo Church 和 John von Neumann 这样的逻辑学家和数学家在20世纪上半叶发明，这远在我们现在所熟悉的这类计算机诞生之前。该理论包括图灵对停机定理的证明，以及丘奇-图灵论题的表述。对它们的一般讨论，可参考 Copeland (2004)。超计算的概念则要晚近得多。对它的一些讨论，可参考 Piccini (2015)。关于图灵的生平和著作，可参考 Hodges (2013)。

第 **15** 章：或许为真——但你无法证明！

希尔伯特纲领是20世纪上半叶提出的，为数学提供严格基础的若干想法之一。对该纲领的讨论，可参考 Zach (2015)。对哥德尔定理技术化得多（不过不像通常讨论这些的文献那样非常技术化）的介绍，可参考 Smith (2007)。哥德尔的不完全性定理，通常被认为是他最令人惊叹的结果；不过，他还在逻辑和集合论基础方面证明了其他好几个十分重要的结果。对这些的讨论，以及关于哥德尔自己，可参考 Dawson (1997)。哥德尔定理被（正确地或错误地）认为有很多哲学上的后果。对其中一些的讨论，可参考 Raatikainen (2005)。

当然，逻辑史远不止上述细节所包含的内容。同样，逻辑本身也有很多内容这本书完全没有谈到。我们只不过是浮光掠影地介绍了一下逻辑。逻辑的深奥和美妙之处是无法在这样一本小书中传达出来的。但过去很多伟大的逻辑学家都恰恰是因为本书讨论的这类思考和问题才投身于逻辑事业的。如果它们也能让你投身其中，我就别无所求了。

术语表

以下术语表包含了在本书中使用的专业术语和逻辑符号。这里的词条不打算给出精确定义，只是给出主要思想，以供快速参考。大体来说，这里给出的术语和符号都是相当标准的，尽管有其他几套符号也是常用的。

- algorithm（算法）：无需猜测或创造性可逐步执行的过程。
- antecedent（前件）：条件句中跟在“如果”后面的部分。
- arithmetic（算术）：讨论自然数（0, 1, 2...）的数学分支。
- axoim（公理）：公理系统的基本陈述。
- axiom system（公理系统）：一组基本陈述，其他陈述可由它们推演而被证明。
- binary numeral（二进制数字）：像 10011 这样的数字，它通过 2 的幂表达一个数。
- Church-Turing Thesis（丘奇-图灵论题）：每个算法都可由一个计算机程序实现这一论题。
- code number（编码）：可以指派给陈述、计算机程序或证明这类实体的数。有了编码，人们可以通过“解码”找到被编码的对象。
- completeness（完全性）：一个公理系统是完全的，若它能证明其语言中能表达的所有真语句（因而，给定排中律，对任何 A，要么证明 A，要么证明 $\neg A$。）
- conclusion（结论）：推断中被给出理由的部分。
- conditional（条件句）：如果...那么...
- conditional probability（条件概率）：给定某些其他信息的条件下，某个陈述的概率。
- conjunction（合取，合取式）：...且...

- conjuncts（合取项，合取支）：合取式中涉及的两个语句。
- consequent（后件）：条件句中跟在“那么”后面的部分。
- consistency（一致性）：一个公理系统是一致的，若不存在公式A使得它既能证明A又能证明$\neg A$。
- conversational implicature（会话蕴含）：不是从所说的内容而是从它被说出这一事实得出结论的推断。
- decision theory（决策论）：在不确定信息条件下如何做决定的理论。
- deductive validity（演绎有效性）：一个推断是演绎有效的，若结论不真前提也不可能为真。
- (definite) description（（限定）摹状词）：形如“具有某某特征的东西”的名称。
- disjunction（析取，析取式）：...或...
- disjuncts（析取项，析取支）：析取式中涉及的两个语句。
- Excluded Middle（排中律）：对每个 A，$A \vee \neg A$ 这一原理。
- expectation（期望值）：每个可能的结果乘以其概率，所得的值全部相加后得到的结果。
- fuzzy logic（模糊逻辑）：一种语句取 0 和 1 之间任何数字作为真值的逻辑。
- Gödel's (first) Incompleteness Theorem（哥德尔（第一）不完全性定理）：给定具有恰当表达力的算术公理系统，它要么是不一致的，要么是不完全的。
- Gödel's (second) Incompleteness Theorem（哥德尔（第二）不完全性定理）：给定具有恰当表达力的算术公理系统，如果它是一致的，则其一致性无法被该系统证明。
- Halting Theorem（停机定理）：图灵的如下结果：没有任何计算机程序能判定任意程序在任意输入下是否终止。
- Hilbert's Program（希尔伯特纲领）：公理化全部数学并证明该公理系统是一致的计划。
- inductive validity（归纳有效性）：一个推断是归纳有效的，若前提为结论提供了某种合理的根据，尽管结论未必是结论性的。
- inference（推断）：一条前提对结论给出理由的推理。
- inverse probability（互逆概率）：条件概率 $pr(a|b)$ 和 $pr(b|a)$ 之间的关系。
- 'is' of identity（同一之“是”）：...和...是同一个对象。
- 'is' of predication（谓述之“是”）：谓语的一部分，表示谓语其余部分表达的性质可以被应用
- Leibniz's Law（莱布尼茨律）：如果两个对象是同一的，一个具有的任何性质另一个也具有。
- liar paradox（说谎者悖论）：“本语句为假”。
- Löb’s Theorem（Löb定理）：给定具有恰当表达力的算术公理系统，若它能证明 $\exists x Prov(x, \langle A \rangle) \to A$，则它能证明 A。
- material conditional（实质条件句）：并非（...且非...）。
- modal operator（模态算子）：加在语句上的短语，用来形成另一个语句，表达前一个语句为真的方式（可能地，必然地，等等）。
- modern logic（现代逻辑）：20 世纪之交产生于逻辑学变革的逻辑理论和技术。
- modus ponens（分离规则）：形如 $a, a \to c/c$ 的推断。
- name（名称）：用来指称某个对象（如果能指称的话）的词的语法范畴。

- necessity（必然性）：必定是...
- negation（否定）：并非...
- particular quantifier（特称量词）：有东西使得...
- possibility（可能性）：可能是...
- possible world（可能世界）：与情形 s 关联的情形，在这些情形上 s 上仅仅可能的事情变成是现实的。
- predicate（谓语）：对语法最简单的一类语句来说，就是用来表达关于主语说了些什么的部分。
- premises（前提）：推断中给出理由的部分。
- Principle of Indifference（无差别原则）：给定若干可能性，若它们之间没有相关差别，则它们有相同的概率。
- prior probability（先验概率）：在考虑任何证据之前，某个陈述的概率。
- probability（概率）：介于 0 和 1 之间的一个数，用来度量某事的可能性有多大。
- proof（证明）：公理系统中的推演。
- proper name (专名）：不是摹状词的名称。
- quantifier（量词）：可以作主语的词或短语，但不指称任何对象。
- reductio ad absurdum（归谬法）：一种假定要证明的结论的否定成立，然后表明这不可能的证明方法。
- reference class（参照类）：从中计算概率比率的一组对象。
- Russell's paradox（罗素悖论）：和由所有不是自身成员的集合构成的集合有关。
- self-reference（自指）：关于某个情形反射回自己的语句。
- situation（情形）：一个可以是假设的事件状态，前提和结论在其中可以为真或为假。
- sorites paradox（堆积悖论）：一类涉及到反复应用模糊谓词的悖论。
- subject（主语）：对语法最简单的一类语句来说，就是告诉你这个句子是关于什么的部分。
- syllogism（三段论）：一种具有两个前提和一个结论的推断形式，其理论最早由亚里士多德提出。
- tense（时态）：过去，现在或将来。
- tense operator（时态算子）：加在语句上的短语，用来形成另一个语句，表达前一个语句在什么时候为真或为假（过去或将来）。
- theorem（定理）：公理系统可被证明的陈述。
- traditional logic（传统逻辑）：20 世纪之前使用的逻辑理论和技术。
- truth conditions（真值条件）：一些语句，它们详细说明一个语句的真值如何依赖于其组成部分的真值。
- truth function（真值函数）：一个逻辑符号，当应用于某个语句时得到一个更复杂的语句，该复合语句的真值完全由其组成部分的真值决定。
- truth table（真值表）：描述真值条件的图表。
- truth value（真值）：真（T）或假（F）。
- universal quantifier（全称量词）：所有东西使得...
- vagueness（模糊性）：谓词的一个性质，表示对象的微小变化不改变该谓词的适用性。
- validity（有效性）：应用于前提确实为结论提供了某种理由的推断。

符号	意义	名称
T	真（在某个情形）	真值
F	假（在某个情形）	真值
$\vee$	...或...	析取（式）
$\wedge$	...且...	合取（式）
$\neg$	并非...	否定
$\exists x$	某个对象 x 使得...	特称量词
$\forall x$	所有对象 x 使得...	全称量词
ιx	那个 x 使得...	摹状词算子
$\Box$	必然地...	模态算子
$\Diamond$	可能地...	模态算子
$\rightarrow$	如果...那么...	条件句
$\supset$	并非（...且非...）	实质条件句
$\mathbf{P}$	过去是...	时态算子
$\mathbf{F}$	将要是...	时态算子
$\mathbf{H}$	过去一直是...	时态算子
$\mathbf{G}$	将要永远是...	时态算子
$=$	...和...是同一个对象	等词
$<$	...小于...	
$\leq$	...小于或等于...	
$\|\ldots\|$	...的真值	
$\max$	...和...中更大的	
$\min$	...和...中更小的	
pr	...的概率	
$pr(\ldots\|\ldots)$	给定...的条件下...的概率	条件概率
E	...为真的期望值	
V	...为真的价值	

$\simeq$	…约等于…	
$\langle A \rangle$	A 的名字（编码）	编码
$Prov(x, y)$	x 是 y 的证明	证明谓词

习题

以下对本书每一章都给了一道练习，你可以用来测试自己对那一章的理解。答案请参见附录。[19]

第 **1** 章：以下推断是演绎有效的、归纳有效的还是都不有效？为什么？

> 乔塞是西班牙人；
>
> 大多数西班牙人是天主教徒；
>
> 因此，乔塞是天主教徒。

第 **2** 章：用符号表示以下推断，并评价其有效性。

> 琼斯是个流氓或者琼斯是个傻瓜；
>
> 但他肯定是个流氓；
>
> 因此，他不是个傻瓜。

第 **3** 章：用符号表示以下推断，并评价其有效性。

> 有人看见或听见了枪击；
>
> 因此，有人看见了枪击或者有人听见了枪击。

第 **4** 章：用符号表示以下推断，并评价其有效性。

> 每个人都想获奖；
>
> 因此，那个赢了比赛的人也想获奖。

第 **5** 章：用符号表示以下推断，并评价其有效性。

> 你做了个煎蛋；
>
> 你不会做了个煎蛋而不打破一个鸡蛋；
>
> 因此，你打破了一个鸡蛋。

第 **6** 章：用符号表示以下推断，并评价其有效性。

猪会飞是不可能的；

猪在水下呼吸是不可能的；

因此，猪既不会飞也不能在水下呼吸，是必然的。

第 7 章：用符号表示以下推断，并评价其有效性。

如果你信上帝，你就会去教堂；

但你去教堂；

因此，你信上帝。

第 8 章：用符号表示以下推断，并评价其有效性。

过去一直在下雨；

将来要一直下雨；

因此，现在在下雨。

第 9 章：用符号表示以下推断，并评价其有效性。

帕特是女的；

擦窗户的那个人不是女的；

因此，帕特不是那个擦窗户的人。

第 10 章：用符号表示以下推断，并评价其有效性，此处可接受性水平是 0.5。

珍妮聪明；

珍妮不聪明或者她漂亮；

因此，珍妮漂亮。

第 11 章：下面是一组从 10 个人（编号为 1-10）那里收集的统计数据：

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
高	✓		✓		✓				✓	
富有	✓		✓		✓		✓	✓		
快乐	✓	✓		✓	✓			✓	✓	

如果 r 是其中随机挑选的一个人，请评估以下推断的归纳有效性：

r 高且富有；

因此，r 快乐。

第 12 章：假设有 A 和 B 两种病，它们可观察的症状完全一样。具有该症状的人中 90% 得的是 A 病，10% 得的是 B 病。还假设，有一种病理检查可以区分 A 和 B。检查的正确率是 9/10。

1. 随机挑选一个具有该症状的人，检查结果是 B 病的概率是多少？（提示：考虑一个 100 人的具有该症状的典型样本，然后算出多少人的检查结果是 B 病。）
2. 给定检查结果为 B 的条件下，得 B 病的概率是多少？（提示：你得利用第 1 问。）

第 13 章：你租了一辆车。如果你不买保险出事故，会花费 1500 美元。如果你买保险出事故，会花费 300 美元。保险费是 90 美元。你估计出事故的概率是 0.05。假设你唯一要考虑的就是经济因素，你应该买保险吗？

第 14 章：下面的论证有何问题？当然存在判定一个程序在给定输入下是否终止的算法。我们只要在该输入下运行该程序，然后看会发生什么就行。它要么终止，要么不终止。无论是种，我们都有一个结果。

第 15 章：称一个公理系统具有**析取性质**（disjunction property），若只要我们能证明形如 $A \vee B$ 的公式，我们也要么能证明 A，要么能证明 B（要么二者都能证明）。假设我们有一个一致的算术公理系统，其所有定理都是真的，用到的是第 2 章中的逻辑。

附录：习题解答

第 1 章：该推断不是演绎有效的。完全有可能前提为真，而乔塞是西班牙少数不是天主教徒的人。但是，前提一起对结论给出了很好的（尽管不是决定性的）支持理由。因此，该推断是归纳有效的。

第 2 章：令

k 为“琼斯是个流氓”

f 为“琼斯是个傻瓜”

则推断符号化为：

$$\frac{k \vee f \quad k}{\neg f}$$

真值表检测结果为：

k	f	$k \vee f$	k	$\neg f$
T	T	T	T	F

T	F	T	T	T
F	T	T	F	F
F	F	F	F	T

第一行，两个前提都为 T，而结论为 F。因此，该推断是无效的。

第 **3** 章：令

> xS 为“x 看见枪击”
>
> xH 为“x 听见枪击”

令考虑中的对象为所有人。推断符号化为：

$$\frac{\exists x(xS \vee xH)}{\exists x\ xS \vee \exists x\ xH}$$

该推断是有效的。因为假设前提在某个情形为真，那就有某个对象 x 在该情形的论域中使得 $xS \vee xH$ 为真。据 $\vee$ 的真值条件，xS 为真或 xH 为真。在第 1 种情况下，有 $\exists x\ xS$；在第 2 种情况下，有 $\exists x\ xH$。因此无论哪种情况，$\exists x\ xS \vee xH$ 在该情形都为真。

第 **4** 章：令

> xP 为“x 想获奖”
>
> xR 为“x 赢了比赛”

令考虑中的对象为所有人。推断符号化为：

$$\frac{\forall x\ xP}{(\iota x\ xR)P}$$

该推断是无效的。考虑某个情形 s，其中所有人都满足 P，但没有人满足 R。（也许比赛取消了！）那么前提在 s 为真。但摹状词 $\iota x\ xR$ 不指称任何东西。因此，结论在 s 为假。

第 **5** 章：令

> m 为“你做了个煎蛋”
>
> b 为“你打破了一个鸡蛋”

则推理符号化为：

$$\frac{m \quad \neg(m \wedge \neg b)}{b}$$

该推断是无效的。[20] 考虑如下情形：

b：F 但不为 T

m：T 且 F

则 $\neg b$ 为 T（且不为 F）；因此 $m \wedge \neg b$ 为 T 和 F（两个合取项均为真，且其中一个为假）；因此 $\neg(m \wedge \neg b)$ 为 T 和 F。在该情形下，两个前提均为 T，而结论不为 T。

第 6 章：令

f 为“猪会飞”。

b 为“猪能在水下呼吸”。

则推断符号化为

$$\frac{\neg\Diamond f \wedge \neg\Diamond b}{\Box(\neg f \wedge \neg b)}$$

该推断是有效的。因为假设前提在某个情形 s 为真，则两个合取项在该情形都为真。因此，没有关联情形 s' 使得 f 为真（第 1 个合取项）或 b 为真（第 2 个合取项）。即，在每个关联情形 s'，$\neg f \wedge \neg b$ 均为真。因此，结论在 s 为真。

第 7 章：令

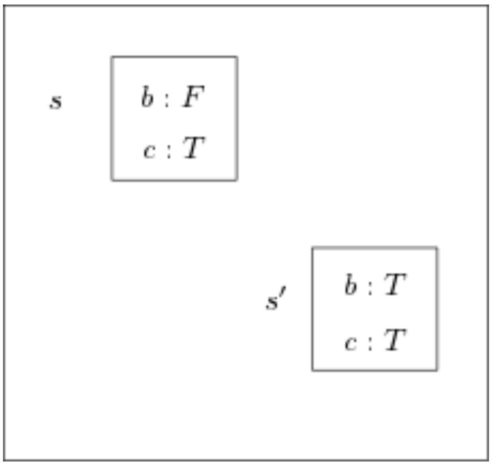
b 为“你信仰上帝”

c 为“你去教堂”

则推断符号化为：

$$\frac{b \to c \quad c}{b}$$

该推断是无效的。[21] 考虑某个情形 s，它有一个关联情形 s'，相关信息图示如下：



在每个 b 为真的情形，c 都为真。因此，$b \to c$ 在 s 为真。这样，两个前提在 s 均为真，但结论在 s 不为真。

第 **8** 章：令

r 为“现在在下雨”

则推断符号化为：

$$\frac{\mathbf{H}r \wedge \mathbf{G}r}{r}$$

该推断是无效的。考虑如下图示的情形组合：

$\dots$	s_{-3}	s_{-2}	s_{-1}	s_0	s_1	s_2	s_3	$\dots$
	r	r	r	$\neg r$	r	r	r	

r 在 s_0 之前的所有时刻均为真，故 $\mathbf{H}r$ 在 s_0 为真。r 在 s_0 之后的所有时刻均为真，故 $\mathbf{G}s$ 在 s_0 为真。因此，$\mathbf{H}r \wedge \mathbf{G}s$ 在 s_0 为真。但结论在 s_0 不为真。

第 **9** 章：令

p 为“帕特”

c 为“那个擦窗户的人”

w 为“是个女人”

则推断符号化为：

$$\frac{pW \wedge \neg cW}{\neg p = c}$$

该推断是有效的。考虑任何前提为真的情形，则在该情形下，p 指称的任何对象都具有 W 表达的性质，而 c 指称的任何对象都不具有该性质。因此，据莱布尼茨律，p 和 c 指称不同的对象（假定没有什么可以既真又假！）。即，$\neg p = c$ 为真。

第 **10** 章：令

c 为“珍妮聪明”

b 为“珍妮漂亮”

则推断形式化为：

$$\frac{c \quad \neg c \vee b}{b}$$

该推断是无效的。[22] 考虑某个 c 和 b 具有如下真值的情形：

c：0.5

b：0.2

则 $\neg c$ 的真值为 0.5（$1-0.5$），因而 $\neg c \vee b$ 的真值也为 0.5（$\max(0.5, 0.2)$）。这样两个前提都是可接受的（≥ 0.5），但结论不是可接受的。

第 11 章：令

t 为"r 高"

w 为"r 富有"

h 为"r 快乐"

推断是有效的。因为有 3 个高且富有的人，其中 2 个是快乐的，故 $pr(h|t \wedge w) = 2/3$；其中 1 个是不快乐的，故 $pr(\neg h|t \wedge w) = 1/3$。因此，$pr(h|t \wedge w) > pr(\neg h|t \wedge w)$。

第 12 章：第 1 问：考虑一个 100 人的具有该症状的典型样本，则 90 人得 A 病，10 人得 B 病。由于检查的正确率是 9/10，检查结果会告诉我们 90 个得 A 病的人中有 81 个得 A 病、9 人得 B 病；10 个得 B 病的人中有 9 人得 B 病、1 人得 A 病。因此，总共有 18 人检查结果为 B 病，因此一个随机抽中的人检查结果为 B 病的概率为 18/100。

第 2 问：令 r 为随机抽中的具有该症状的人，且令

b 为"r 得 B 病" t 为"r 的检查结果为 B 病"

则：

- $pr(t|b) = 9/10$，因为检查的正确率为 90%；
- $pr(b) = 1/10$，因为 10 人中有 1 人得 B 病；
- $pr(t) = 18/100$，据第 1 问。

据互逆概率之间的关系可得，

$$pr(b|t) = pr(t|b) \times pr(b)/pr(t) = \frac{9}{10} \times \frac{1}{10} \div \frac{18}{100} = 1/2$$

第 13 章：将相关信息用表格表示如下：

	出事故	不出事故
买保险	$0.05\backslash-390$	$0.95\backslash-90$
不买保险	$0.05\backslash-1500$	$0.95\backslash 0$

计算期望值可得：

$$E(t) = 0.05 \times (-390) + 0.95 \times (-90) = -105$$
$$E(\neg t) = 0.05 \times (-1500) + 0.95 \times 0 = -75$$

由于 $E(\neg t) > E(t)$，因此你应该不买保险。

第 **14** 章：我们当然可以对给定的输入运行程序。如果它确实终止，则或早或晚它会终止，那时我们就知道它终止（尽管我们可能无法提前知道要多久它才会终止）。但是，如果它不终止，我们将永远无法知道这一点。无论计算持续了多长时间，如果它没停，这可能是因为它永远不会终止，但也可能只是它还没到终止的时候。我们没办法知道究竟处于哪种情形。

第 **15** 章：没有。如果 n 是语句 $\neg\exists x Prov(x,n)$ 的编码，则由于逻辑具有排中律，该理论能证明 $\exists x Prove(x,n) \vee \neg\exists x Prov(x,n)$。但哥德尔定理表明 $\neg\exists x Prov(x,n)$ 无法被证明，尽管它是真的。

---

1. 这里我推荐 Harry J. Gensler 的 Introduction to Logic（中译本正在翻译中）。↩
2. 译者注：原文对合取采用的记号是 &，中译采用了更标准的记号 $\wedge$。↩
3. 译者注：英文的 or 有否则的意思，汉语此处这一意义不明显。↩
4. 译者注：英文为 member，数学上一般译为“元素”，这里为了照顾普通读者，译成了更一般的用语。↩
5. 译者注：原文为 $\neg a$，疑误。↩
6. 译者注：更好的办法其实是不出门。↩
7. 译者注：传统逻辑中一般将 modus ponens 译为假言推理，现代逻辑中一般将其译为分离规则，以下我们采用现代逻辑的译法。↩
8. 译者注：参见尾注 6。↩
9. 译者注：原文此处有个排印错误，左括号应在 $\neg$ 之后。↩
10. 译者注：原文有同上面一样的排印错误。↩
11. 译者注：鉴于汉语中的时态表达不严格，如果直觉上无法理解为何它们是有效的，看完这一章就会明白了。↩
12. 译者注：原文是 The sun will had been shining，从语法和后面的形式分析来看，这里似乎多了一个 will。↩
13. 译者注：原文为 last chapter，疑误，因模态算子是在第 6 章介绍的。↩
14. 译者注：此处不严谨，除了不包含时态算子，也不能包含其他逻辑算子，如 $\neg, \vee, \wedge$ 或 $\Box, \Diamond$ 。即应只对原子语句提供真值。↩
15. 译者注：即同一于 x 的性质。↩
16. 译者注：原文为 You can't be Indifferent About it，直译为：你无法不在乎它。作者选用 indifferent 这个词有意要达到某种双关效果，因为本章会讨论概率中的一个重要原则——无差别原则，Principle of Indifference，其中 Indifference 是 Indifferent 的名词。鉴于这种双关性，此处没有采用直译。下一章标题也有类似情况。↩
17. 译者注：原文是 Great Expectations，狄更斯的著名小说名，中译《远大前程》。此处一语双关，同时指数学上大的期望值。↩
18. 译者注：学术论著经常采用这种`作者（年代）`或`（作者，年代）`的方式标注参考文献，有时括号中还会加入页码或章节等具体位置信息，其对应的书名或文章标题及出版信息可以在最后的参考文献中根据作者姓氏的字母序查找得到。若同一作者有多篇文献被引用，年代可以用来区分不同文献；若同一作者在同一年有多篇文献被引用，则会在年代后再加 a, b, c 这样的字母进行区分。↩
19. 译者注：英文答案请参见这个网址。↩
20. 译者注：在经典逻辑中，该推断是有效的。因本章讨论的是有 4 个真值的逻辑，所以这里给出的答案是相反的。↩
21. 译者注：事实上，该推断不仅在条件句逻辑中是无效的，在经典逻辑中也是无效的。↩
22. 译者注：同样，该推断在经典逻辑中是有效的。因本章讨论的是模糊逻辑，所以这里给出的答案是相反的。↩