数理逻辑

01 - 形式化推理

(Press ? for help, n and p for next and previous slide)

戴望州

南京大学智能科学与技术学院

2025年 - 春季

https://daiwz.net

课程信息

课程信息

- **时间**：周二3-4节
- **周数**：1-16周
- **地点**：南雍-西311
- **助教**：
  - 宋奥齐, songaq@smail.nju.edu.cn
  - 苏浩, h.su@smail.nju.edu.cn
  - 周桢瑜, dilantor@foxmail.com
- 课程主页：https://daiwz.net/teaching
- Office Hour：周三下午

2025数理逻辑
群号: 632078342

参考书

1. Peter Smith. *An Introduction to Formal Logic*.
2. Herbert B. Enderton. *A Mathematical Introduction to Logic*.
   - 郝兆宽, 杨睿之.《数理逻辑：证明及其限度》.

疑问

有了 OpenAI O1/O3, DeepSeek-R1/R1-Zero，

搞 AI 还需要学习数理逻辑吗?

疑问（1900s）

用了几千年自然语言，

干嘛要研究数理逻辑？

一言以蔽之

数理逻辑是人工智能前奏——我说的

数理逻辑初步

1. **形式化推理**（ヒルベルトの野望）
2. 命题逻辑（進撃のゲーデル・その1）
3. 一阶逻辑（進撃のゲーデル・その2）
4. 哥德尔不完全性定理（人類の絶望）
5. 穿插一些 `Lean4` & `Prolog` （如果有时间的话😅）

什么是逻辑?

逻辑学家如何解释上厕所要排队

"Formally: pee implies queue. Pee, therefore, queue."

credit: GPT4o

墨翟

“夫辯者，將以明是非之分，審治亂之紀，明同異之處，察名實之理，處利害，決嫌疑。焉摹略萬物之然，論求群言之比。以名舉實，以辭抒意，以說出故，以類取，以類予。有諸己不非諸人，無諸己不求諸人。”（《墨子·小取》，公元前476年 - 公元前390年）

Aristotle

A syllogism (Συλλογισμός, deduction) is discourse (*logos*) in which, certain things being supposed (premises), something different from those supposed results of necessity (conclusion) because of their being so. I mean by the last phrase that they produce the consequence, and by this, that no further term is required from without in order to make the consequence necessary. (*Prior Analytics*, 384 BC - 322 BC)

Leibniz

"If this is done, whenever controversies arise, there will be no more need for arguing among two philosophers than among two mathematicians. For it will suffice to take the pens into the hand and to sit down by the abacus, saying to each other (and if they wish also to a friend called for help): Let us calculate." (*Philosophische Schriften*, 1646-1716)

一言以蔽之

逻辑的作用是**系统性地检验论证的有效性**

1. 什么叫“论证”（arguments）？
2. 什么叫“检验”（evaluation）？
3. 什么叫“有效性”（validity）？
4. 什么叫“系统性地”（systematically）？

论证

论证："辯" - "Συλλογισμός" - "Argument"

《现代汉语词典（第七版）》：

1. 在逻辑学中，指引用论据来证明论题的真实性的论述过程，是由论据推出论题时所使用的推理形式。
2. 论述并证明。
3. 立论的根据。

“论证”vs“解决问题”

The General Problem Solver [Herbert A. Simon, J. C. Shaw and Allen Newell, 1957]

从“证明$1+1=2$”，“玩吃豆人”到“烧一壶开水”

BIG Problem

"解决问题"

`Working Memory` → [small problem 1, small problem 2, ...]

“解决问题”

`Working Memory` → [**small problem #1**, small problem #2, …]

“解决问题”

`Working Memory` → [[smaller problem #1.1, smaller problem #1.2],
small problem #2, ...]

“解决问题”

`Working Memory` → [[~~smaller problem #1.1~~, smaller problem #1.2],
small problem #2, …]

"解决问题"

`Working Memory` → [[~~smaller problem #1.1~~, ~~smaller problem #1.2~~],
small problem #2, ...]

“解决问题”

`Working Memory` → [~~small problem #1~~, **small problem #2**, …]

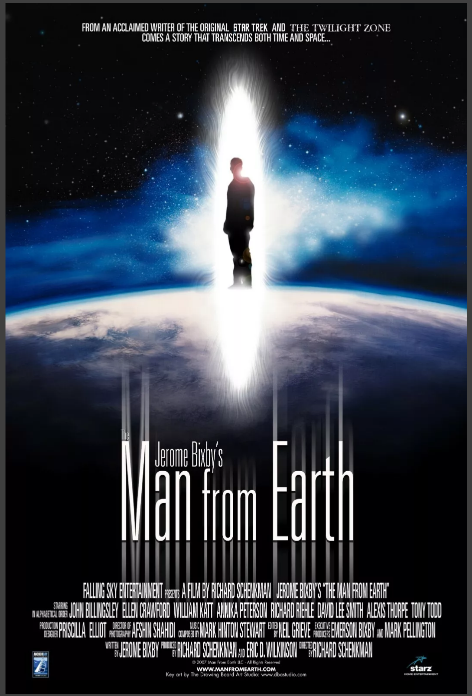
“已证”？“解决”？

如何判定~~small problem #1~~？

检验

例 1

- 论点 1：约翰已经活了超过14,000年。
- 论点 2：永生并不像看起来那么不可能。
- 论点 3：他见证了人类文化和宗教的诞生。
- 论点 4：约翰长久的生命解释了他对历史的深刻理解。
- 论点 5：永生的神话解释了许多历史人物为何显得“超凡脱俗”。
- 结论：永生者可能是我们许多神话和宗教信仰的源头。



检验论证的两个维度

- 论点1：约翰已经活了超过14,000年。
- 论点2：永生并不像看起来那么不可能。
- 论点3：他见证了人类文化和宗教的诞生。
- 论点4：约翰长久的生命解释了他对历史的深刻理解。
- 论点5：永生的神话解释了许多历史人物为何显得“超凡脱俗”。
- 结论：永生者可能是我们许多神话和宗教信仰的源头。

1. 论证的前提是否正确？
2. 论证的过程是否滴水不漏？
   - 即：若前提100%正确，结论是否保证100%正确？
   - “滴水不漏”的确是另一个维度，因为它与“前提是否正确”无关

检验论证的两个维度

在这门课上，我们**只关注第二个维度**：

1. “前提”来自各个领域，其正确性和“逻辑”本身无关
2. 逻辑学家关心的是：“从某些（假定为真的）论据出发，究竟能不能得到最终的结论”

论证的有效性

例2

1. 小明不是在卷，就是在卷的路上
2. 小明不在卷
3. 小明在卷的路上

例3

1. 徐克导演的武侠片一直都很好看。
2. 2025年的《射雕英雄传》是徐克导演的。
3. 2025年的《射雕英雄传》会很好看！

两种推理的不同

- 例 1 & 2 中的推理“滴水不漏”——演绎（deduction）
  1. 若它们的论据为真，没有任何理由能让它们的结论为假
  2. 从推理得到的结果
- 例 3 中的推理虽然很可能，但存在“漏洞”——归纳（induction）
  1. 若 3.1 和 3.2 为真，确实能找到令 3.3 为假的理由（即便可能性很小）
  2. 从过去经验中得到的结论，它能够一定程度上预测未来

演绎有效性（非正式定义）

定义：若一个从某些前提推演出结论的演绎步骤完全“滴水不漏”，即若前提的成立绝对保证（absolutely guarantees）结论的成立，我们就称这一步推理是演绎有效的（deductively valid）。

定义：反过来，若一个包含前提和结论演绎步骤是有效的（valid），那么我们称这些前提（premises）演绎蕴涵（deductively entails）它的结论（conclusion）。

- 例1 & 2是演绎有效的
- 例3尽管是很有说服力的归纳，但它不是演绎有效的

更多的例子

演绎有效性（非正式定义）

逻辑推理能够**系统性地检验论证的有效性**

定义：若一个从某些前提推演出结论的演绎步骤完全"滴水不漏"，即若前提的成立绝对保证（absolutely guarantees）结论的成立，我们就称这一步推理是演绎有效的（deductively valid）。

"绝对保证"和"滴水不漏"过于抽象，有没有更具体、更数学的定义？

从有效性（validity）
到可靠性（soundness）

“有效性”的另一种非正式定义

定义：我们称一个推演步骤是有效的当且仅当不存在任何一种可能，令该推演的前提为真且结论为假。同样地，在这种情况下我们称这些前提蕴涵（entails）其结论。

例 4：

1. 某人从来没学过数理逻辑，一整个学期不来上课，也不读相关的书，完全不接触任何关于数理逻辑的知识
2. TA 非常有骨气，不但不交作业，考试也拒绝作弊
3. TA 期末会挂科

问：以上推理是有效的吗？

用量化的“可能性”来描述“有效性”

考虑所有的可能性：理论上可能、技术上可能、操作上可能、经济上可能、XXX 上可能……

- “一切平行世界构成的集合”：$\{universe \in \mathfrak{U}\}$
- 它们的概率分布：$\Sigma_{u\in\mathfrak{U}}P(u) = 1$

至此，我们可以更数学地描述“演绎有效性”：即在最不可能最不可能最不可能……（省略无穷次）出现的平行世界里，当推演的前提为真时，其结论也不可能为假。甚至可以说，前提为真且结论为假的情况是在根本上不自洽的。

- 换句话说，一个有效的推演，其结论一定是前提的必要条件
  - Valid inferences are *necessarily truth-preserving* (“necessary” means “not-possible-not”)
- 自此，我们将省略对“可能性”的描述，它一定是指最弱的“可能”。
- 而这里的“必要性”与“最弱的可能性”相反，是最强的“必要性”，即无论任何条件下都必须满足的性质。

一致性（CONSISTENCY）

有了对“可能性”的量化，我们得以继续定义与“有效性”相关的逻辑学概念。

一个或多个命题是不一致的（inconsistent），当且仅当在任意可能的情况下，它们都无法同时为真。

- 也就是说，一组命题是一致的，只需要存在一种可能情况令它们同时为真。

一个推演是有效的，当且仅当它的前提和其结论的否定是不一致的。

ELI5 的例子

下面的命题是一致的吗？

- 命题1： Spot is a dog
- 命题2： Spot cannot fly
- 命题3： All dogs cannot fly

credit: r/explainlikeimfive

等价性（EQUIVALENCE）

有了对“可能性”的量化，我们得以继续定义与“有效性”相关的逻辑学概念。

两个命题是等价的（equivalent），当且仅当它们在完全相同的可能情况下才同时为真。

命题 A 和命题 B 是等价的，当且仅当 A 蕴涵（entails）B，且 B 蕴涵 A。

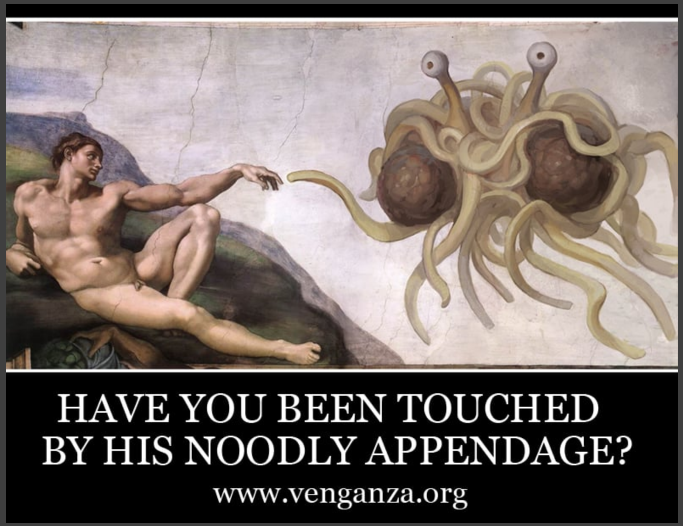
有效、非有效（INVALIDITY）与真（TRUTH）

下面的推理是演绎有效的吗？

- 前提1：Spot is a dog
- 前提2：All dogs can fly
- 结论：Spot can fly.

credit: r/explainlikeimfive

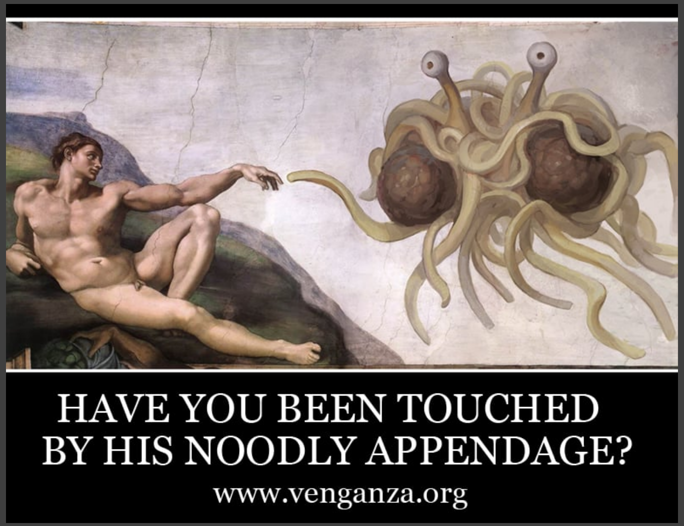
有效、非有效（INVALIDITY）与真（TRUTH）



例5-1：

1. 特朗普说的所有话都是真的
2. 特朗普说飞面神（Flying Spaghetti Monster）是存在的
3. 飞面神真的存在

有效、非有效（INVALIDITY）与真（TRUTH）



例5-2：

1. 飞面神的存在是一个玩笑
2. 玩笑很可能是真的
3. 飞面神真的存在

有效、非有效（INVALIDITY）与真（TRUTH）

例6:

1. 好球员做不了好教练
2. 李铁是个好球员
3. 李铁不是个好教练

例7:

1. All lions are fierce
2. Some lions do not drink coffee
3. Some fierce creatures do not drink coffee

有效、非有效（INVALIDITY）与真（TRUTH）

“有效”推演的三种赋值方式：

1. 前提为真，结论为真（例7）
2. 前提为假，结论为假（例5-1）
3. 前提为假，结论为真（例6）

- 根据我们对“有效性”的定义，唯一被其排除的是“前提为真，但结论为假”的情况
- Again, valid inferences are *necessarily truth-preserving*

非有效准则（The Invalidity Principle）：若一个推演的前提为**真**，但结论为**假**，那么这个推演一定是非有效的（must be invalid）。

有效、非有效（INVALIDITY）与真（TRUTH）

那么，非有效的推演能有如何的赋值呢?

例8:

1. 天气一点也不好
2. 我抓到了超梦
3. 我的作业没做完

1. true $\rightarrow$ false一定是非有效推演
2. 非有效推演不一定是true $\rightarrow$ false

有效、非有效（INVALIDITY）与真（TRUTH）

语言中的“有效”无法帮助我们寻找“真”，我们需要其他东西

有效性与可靠性（SOUNDNESS）

一个论证（argument）是有效的（valid），当且仅当从它的前提到结论的推演是有效的。

一个论证（argument）是可靠的（sound），当且仅当从它的前提为真，且结论为真。换言之，有效论的证前提为真，且论证本身也是有效的。

Once again, ELI5 的例子

下面的推理是可靠的吗?

- 前提1: ~~Spot is a dog~~ All dogs have the same flying ability as Spot the dog
- 前提2: Spot cannot fly in real world
- 结论: All dogs cannot fly in real world

credit: r/explainlikeimfive

Once again, ELI5 的例子

Laika, a Moscow street dog, became the first creature to orbit Earth (with USSR's Sputnik 2), but she died in space.

自然语言是有漏洞的！

我们无法用有穷的语言，准确地描述$\mathfrak{U}$！

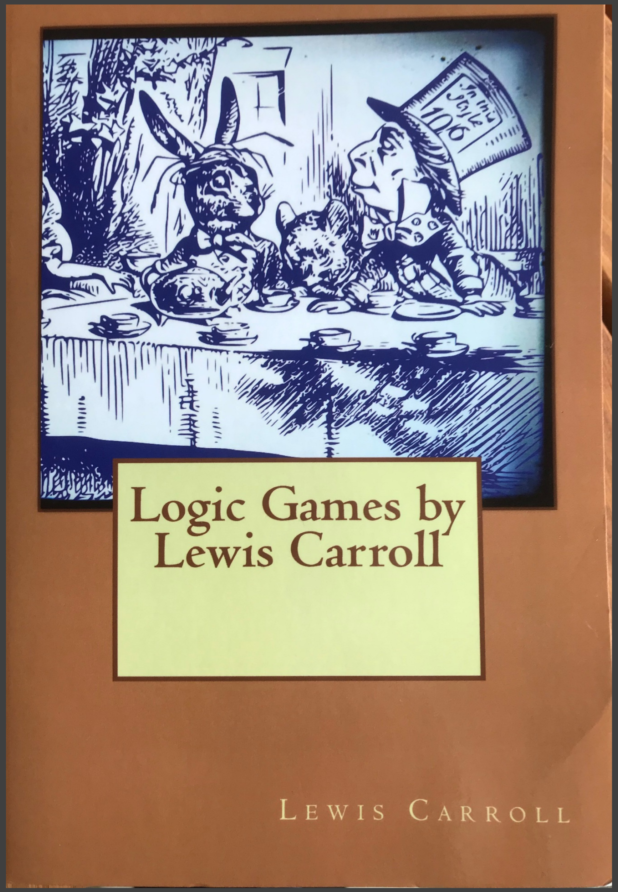
6.54 我的命题可以用以下方式解释：了解我作法的人，会用这些命题当做梯子，越过它们，最终会发现这些梯子是荒谬的（也可以说，当他们爬上去之后，要把梯子丢掉）

——“他必须超越这些命题，之后才能正确地看待世界”

形式化：摆脱自然语言的束缚

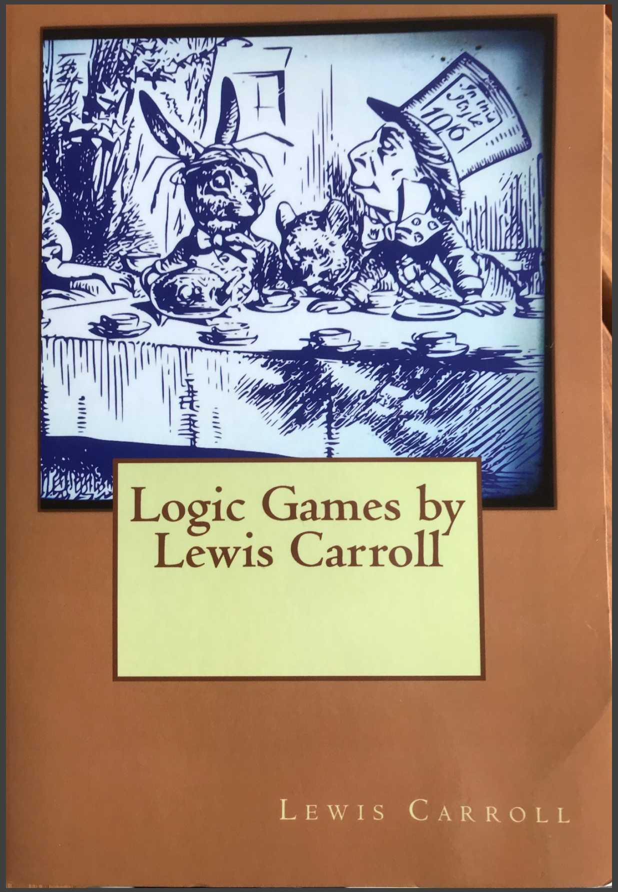
Games from Lewis Carroll

1. Babies are illogical;
2. Nobody is despised who can manage a crocodile;
3. Illogical persons are despised.
4. Q: Babies cannot manage crocodiles?



Games from Lewis Carroll

1. No kitten, that loves fish, is unteachable;
2. No kitten without a tail will play with a gorilla;
3. Kittens with whiskers always love fish;
4. No teachable kitten has green eyes;
5. No kittens have tails unless they have whiskers.
6. Q: Kitten with green eyes will play with a gorilla?



推理的模式（SCHEMA）

“模式”可以用来揭示推演的结构。

例 9-1:

1. No misers (守财奴) are generous; (No X are Y)
2. Some misers are old men. (Some X are Z)
3. Some old men are ungenerous. (Some Z are not Y)

例 9-2:

1. No thieves are honest; (No X are Y)
2. Some thieves are gentlemen. (Some X are Z)
3. Some gentlemen are dishonest. (Some Z are not Y)

模式中的符号

1. All F are G
2. n is F
3. *So*: n is G

1. All men are motal
2. Tweety is a robin
3. *So*: Donalt Trump is femail

相同的代号只能替换为相同的名字

模式中的符号

1. All F are G
2. n is F
3. *So*: n is G

1. All men are men
2. Socrates is a man
3. *So*: Socrates is a man

不同的代号可以替换为相同的名字

模式中的符号

1. All F are G
2. n is F
3. *So*: n is G

1. Socrates is a man
2. All men are mortal
3. *So*: Socrates is mortal

前提的顺序不影响推演有效性

模式中的符号

比如：

1. All F are G
2. n is F
3. *So*: n is G

又比如：

1. All H are K
2. m is H
3. *So*: m is K

比如：

1. All Φ are Ψ
2. α is Φ
3. *So*: α is Ψ

又比如：

1. All ① are ②
2. $*$ is ①
3. *So*: $*$ is ②

（有些）符号本身并无意义

模式的抽象

"All men are mortal, Socrates is a man, *So*: Socrates is mortal".

1. A, B, *So*: C
2. All F are G, m is H, *So*: n is K
3. All F are G, n is F, *So*: n is G
4. All F are G, Socrates is F, *So*: Socrates is G
5. All F are mortal, n is F, *So*: n is mortal
6. All men are G, Socrates is a man, *So*: Socrates is G
7. All men are mortal, Socrates is a man, *So*: Socrates is mortal

存在一些可靠的抽象推理形式（form），它的任意一个实例（instance）都是**有效的**，因此这些模式本身也是**有效的**。

逻辑有效性

话题无关性（TOPIC-NEUTRALITY）

意下这些推演是有效的吗？

例10-1：

1. Jill is a mother
2. *So*: Jill is a parent

例10-2：

1. Jack has a first cousin
2. *So*: At least one of Jack's parents is not an only child

例10-2：

1. Jack is a bachelor
2. *So*: Jack is unmarried

例10-1'：

1. n is a mother
2. *So*: Jill is a parent

例10-2'：

1. n has a first cousin
2. *So*: At least one of n's parents is not an only child

例10-2'：

1. n is a bachelor
2. *So*: n is unmarried

话题无关性（TOPIC-NEUTRALITY）

与具体论题无关，但对命题论述起作用的词汇，例如“所有”（all）、“一些”（some）、“与/和”（and）、“或”（or）、“并非”（not）、“如果”（if）、“那么”（then）、“是”（is, are）等，可被称为**话题无关词汇**（*topic-neutral vocabulary*）。

尽管仍然存在一些必须依靠话题相关词汇才能判定有效性的论题，这门课中我们将不再讨论它们。

逻辑有效性（LOGICAL VALIDITY）

至此，我们能够更明确地谈论“逻辑”上的有效性

逻辑的目的是**系统性地检验论证的有效性**

逻辑有效性（logical validity）：
我们称一个推演步骤是**逻辑有效的**，当且仅当它的有效性仅由其前提与结论中的话题无关词汇决定。对于一个逻辑有效的演绎，我们称其前提**逻辑上蕴涵**（logically entails）其结论。

形式化推理（Formal Reasoning）

形式上的“真”是由推理的形式决定的，与论题的领域无关。

例 11：

1. Whatever is *green* is *coloured*
2. Whatever is *green* and square is *green*

形式逻辑（Formal Logic）

逻辑谈论的只是关于真的一些的规律

形式逻辑（Formal Logic）

Pure Logical Schema = Schematic Variables + Topic-Neutral Vocabulary

数理逻辑谈论的只是关于真的一些**普遍性**规律

“话题无关词汇”的边界？

我们似乎定义了一个“话题无关词汇”表，但是它是否足够准确？除了这些“与”、“或”、“非”等等，还有没有必要添加其他词汇？

例 12：

1. 数理逻辑比数学分析简单
2. 数学分析比代数几何简单
3. *So:* 数理逻辑比代数几何简单

例 12'：

1. m is easier than n
2. n is easier than o
3. *So:* m is easier than o

例 12-2：

1. m is F-er than n
2. n is F-er than o
3. *So:* m is F-er than o

“话题无关词汇”的边界？

1. 我们无法绝对地定义“话题无关词”的边界，许多事情的描述本身就是模糊的
2. 本课程只考虑“任意”、“某些”、“与”、“或”、“非”等**有限个、语义明确**的话题无关词汇
   - 它们组成了数理逻辑的核心部分
   - 一定情况下，由它们构成的语句将具有**绝对明确的**（absolutely clear）逻辑有效性

形式化的后果

例 13：Jack is married. Jack is not married. *So:* the world will end tomorrow!

- 逻辑上有效吗？为什么会这样？
- 它的有效性来自于哪些词汇？
- 解决办法：在有效性的定义中加入“命题间的相关性”（relevance）
- 但“逻辑有效”的定义将不复存在
- A small price to pay: *vacuous truth*

语义上的“真”vs“逻辑上有效”

我们无法一劳永逸地把握**真**

塔斯基（Tarski）

- 我们只能谈论特定领域的**真** $\mathrm{Th}\mathfrak{A} = \{\sigma \mid \mathfrak{A} \models \sigma\}$
- $\mathrm{Th}\mathfrak{A}$不是$\mathfrak{A}$中可定义的（Why?）
- 日常语言中的**真**是不可定义的（Why???）

关于真，逻辑到底能谈论多少？

逻辑学是否提供新知识?

- 培根（Francis Bacon）：亚里士多德《工具论》中的那些纯演绎的方法（三段论）不足以发现科学真理,因此需要《新工具》（*Novum Organum Scientiarum*）
- 康德（Immanuel Kant）：分析命题（主项包含谓项）不提供新知识，后天综合命题讲述经验世界的知识，而哲学、数学知识应该是先天综合命题。

逻辑学是否提供新知识?

我也相信，所有人都先天具备逻辑思考的能力

那我们又能从这门课学到什么?

希望经过本学期的学习可以回答的问题

$\{\sigma \mid \forall \mathfrak{A}\ (\mathfrak{A} \models \sigma)\}$是个怎样的集合?

希尔伯特1930年的格尼斯堡演讲

有这样一种工具，它能将理论与实践，将思想与观察联系起来、那就是数学；它搭建了连接的桥梁并使之愈加坚固。以至于，我们整个当代文化，就其依赖于我们对自然的理智洞察与利用的范围内而言，是以数学为基础的。

希尔伯特1930年的格尼斯堡演讲

伽利略早就说过：只有习得了自然对我们诉说时所使用的语言和符号，一个人才能理解自然；而这个语言正是数学，它的符号则是数学图形。

康德宣称：“我始终认为，存在于每个特定的自然科学中的真理不会比数学中的真理更多。”

希尔伯特1930年的格尼斯堡演讲

事实上，只有当我们提炼出一门自然科学的数学内核并将其彻底揭开时，我们才算掌握了它的理论。

没有数学，今天的天文学和物理学将是不可能的；在它们的理论部分，这些科学直接展开为数学。正如大量其他的应用一样，这些事实使数学享有在公众中无与伦比的权威。

希尔伯特1930年的格尼斯堡演讲

尽管如此，所有数学家都拒绝将应用作为数学价值的标准。高斯在说到是什么让数论成为这位数学第一人最喜爱的学科时，提到的是它那魔法般的吸引力而不是它目前为止超越所有其他数学分支的那无穷无尽的丰富性。

克罗内克将数论学家比作食莲族†，一旦尝到了甜头便再也无法离开它了。

†出自荷马史诗《奥德赛》。古希腊神话中传说北非的某个岛上种有许多称为lotus的树，当地人以此树的果实“lotos”为食。这种果实具有强烈的催眠作用，吃了后能忘却烦恼忧愁，陷入浑浑噩噩、乐不思蜀的状态。

希尔伯特1930年的格尼斯堡演讲

伟大的数学家庞加莱曾以令人震惊的犀利抨击了托尔斯泰，后者声称“以科学的名义追求科学”是愚蠢的。例如，若只有那些实用主义头脑存在而没有那些无私的傻瓜来推动进步的话，工业上的成就是永远没有可能的。

正如格尼斯堡的数学家雅可比所说，人类精神的荣耀才是所有科学唯一的目标。

希尔伯特1930年的格尼斯堡演讲

我们绝不相信那些今天还以哲学的姿态或以优越的口吻所做的关于文化衰落的预言并接受不可知。对我们来说，不可知是不存在的，并且在我看来，在自然科学中也不存在。

让我们抛弃那个愚蠢的不可知，代之以下面的口号：

Wir müssen wissen. Wir werden wissen.

Lean 4

第一次作业

1. No interesting poems are unpopular among people of real taste.
2. No modern poetry is free from affectation.
3. All your poems are on the subject of soap bubbles.
4. No affected poetry is popular among people of real taste.
5. Only a modern poem would be on the subject of soap bubbles.
6. Therefore none of your poems are interesting.

Lean 4

```
structure Poem where
  isInteresting : Bool
  isPopular : Bool
  isModern : Bool
  isAffected : Bool
  isYourPoem : Bool
  hasSoapBubbleTheme : Bool

-- Axiom 1: No interesting poems are unpopular among people of real taste.
axiom interesting_implies_popular : ∀ (p : Poem), p.isInteresting → p.isPopular

-- Axiom 2: No modern poetry is free from affectation.
axiom modern_implies_affected : ∀ (p : Poem), p.isModern → p.isAffected

-- Axiom 3: All your poems are on the subject of soap bubbles.
axiom your_poem_implies_soap_bubble_theme : ∀ (p : Poem), p.isYourPoem → p.hasSoapBu

-- Axiom 4: No affected poetry is popular among people of real taste.
axiom affected_implies_not_popular : ∀ (p : Poem), p.isAffected → ¬ p.isPopular

-- Axiom 5: Only a modern poem would be on the subject of soap bubbles
```