

Foundations of Artificial Intelligence



School of Electronic and Computer Engineering
Peking University

Wang Wenmin

What are the Foundations of AI:

Philosophy	❖	哲学
Mathematics	❖	数学
Economics	❖	经济学
Neuroscience	❖	神经科学
Psychology	❖	心理学
Computer engineering	❖	计算机工程
Control theory and cybernetics	❖	控制理论和控制论
Linguistics	❖	语言学



Contents

- ☐ 1.2.1 Mathematics
- ☐ 1.2.2 Neuroscience
- ☐ 1.2.3 Cognitive Psychology
- ☐ 1.2.4 Control Theory and Cybernetics

Mathematics in Three Fundamental Areas:

1) Logic -- *What are the formal rules to draw valid conclusions?*

逻辑学 -- 得出正确结论的形式规则是什么？

□ 1847, George Boole:

propositional logic, also called Boolean logic.

乔治·布尔：命题逻辑，亦称布尔逻辑。

□ 1879, Gottlob Frege:

first order logic, that extends Boole's logic to include objects and relations.

戈特洛布·弗雷格：一阶逻辑，它扩展了布尔逻辑，增加了对象和关系。

□ Alfred Tarski (1902–1983):

theory of reference, that shows how to relate the objects in a logic to objects.

阿尔弗雷德·塔斯基：指称理论，它揭示如何将逻辑中的对象与对象相关联。

Mathematics in Three Fundamental Areas:

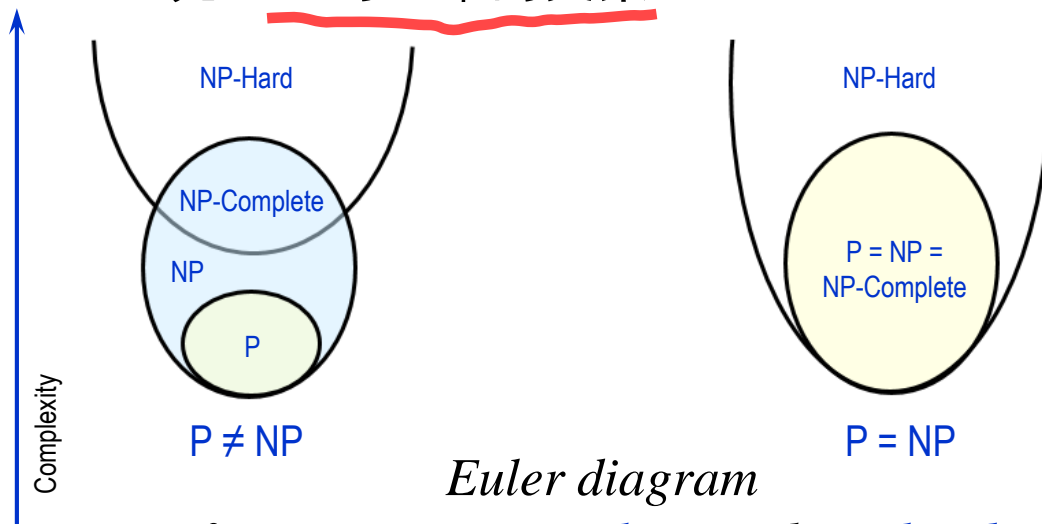
2) Computation -- *What can be computed?*

计算 -- 什么是可计算的？

- Alan Turing (1912–1954):
try to characterize exactly which functions are **computable**.
艾伦·图灵：试图精确地描述哪些函数是可计算的。
- mid-1960s, Cobham and Edmonds:
proposed the notion of computational **tractability**.
科伯姆与埃德蒙兹：提出了计算的易处理性的概念。
- 1972, Steven Cook and Richard Karp:
proposed the theory of **NP-completeness**.
斯蒂文·库克与理查·德卡普：提出了NP完全性的理论。

Terminology: NP-completeness

- In computational complexity theory
 - **P**: Polynomial time. P: 多项式时间
 - **NP**: Non-deterministic Polynomial time.
NP: 不确定性多项式时间
 - **NP-complete**: both in NP and NP-hard.
NP完: NP与NP难的交集。



*Euler diagram
for P , NP , NP -complete, and NP -hard.*

尤拉图: P、NP、NP完与NP难

- Nobody has yet been able to determine conclusively whether NP-complete problems are in fact solvable in polynomial time.

- Left side is valid under the assumption $P \neq NP$.
当假设 $P \neq NP$ 时, 左图成立
- Right side is valid under the assumption $P = NP$.
当假设 $P = NP$ 时, 右图成立

Mathematics in Three Fundamental Areas:

3) Probability -- *How do we reason with uncertain information?*

概率 -- 如何根据不确定信息进行推理？

- Gerolamo Cardano (1501–1576)

framed **probability**, describing it in terms of possible outcomes of gambling events.

杰罗拉莫·卡尔达诺：构建了概率的概念，将其描述为博弈事件中可能的结果。

- James Bernoulli (1654–1705), Pierre Laplace (1749–1827), and others

advanced the theory and introduced new **statistical methods**.

詹姆士·伯努力、皮埃尔·拉普拉斯等人：推进了这一理论，并引入了新的统计学方法。

- Thomas Bayes (1702–1761)

Bayes' rule, it underlies most modern approaches to uncertain reasoning.

托马斯·贝叶斯：提出了贝叶斯规则，它成为不确定性推理的最现代方法。

Neuroscience

-- How Do Brains Process Information?

神经科学：--大脑如何处理信息？

- Neuroscience is the study of the nervous system, particularly the brain.

神经科学研究神经系统，尤其是大脑。

- Brains are very good at making rational decisions (but not perfect).

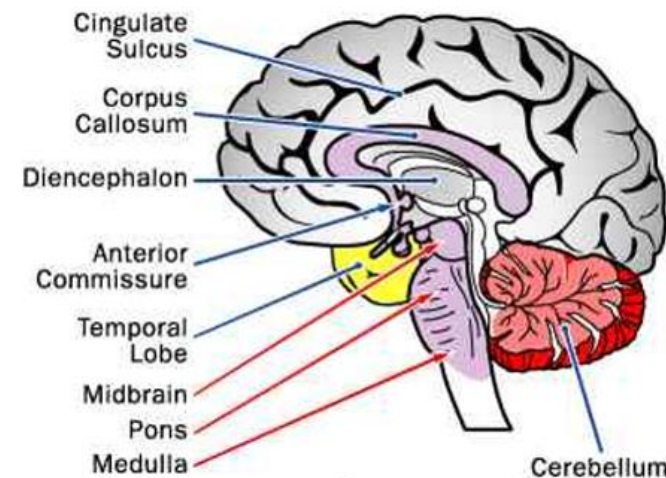
大脑在理性决策方面非常优越（但并非完美无缺）。

- Brains aren't as modular as software.

大脑不像软件那样模块化。

- **Prediction** and **simulation** are key to decision making.

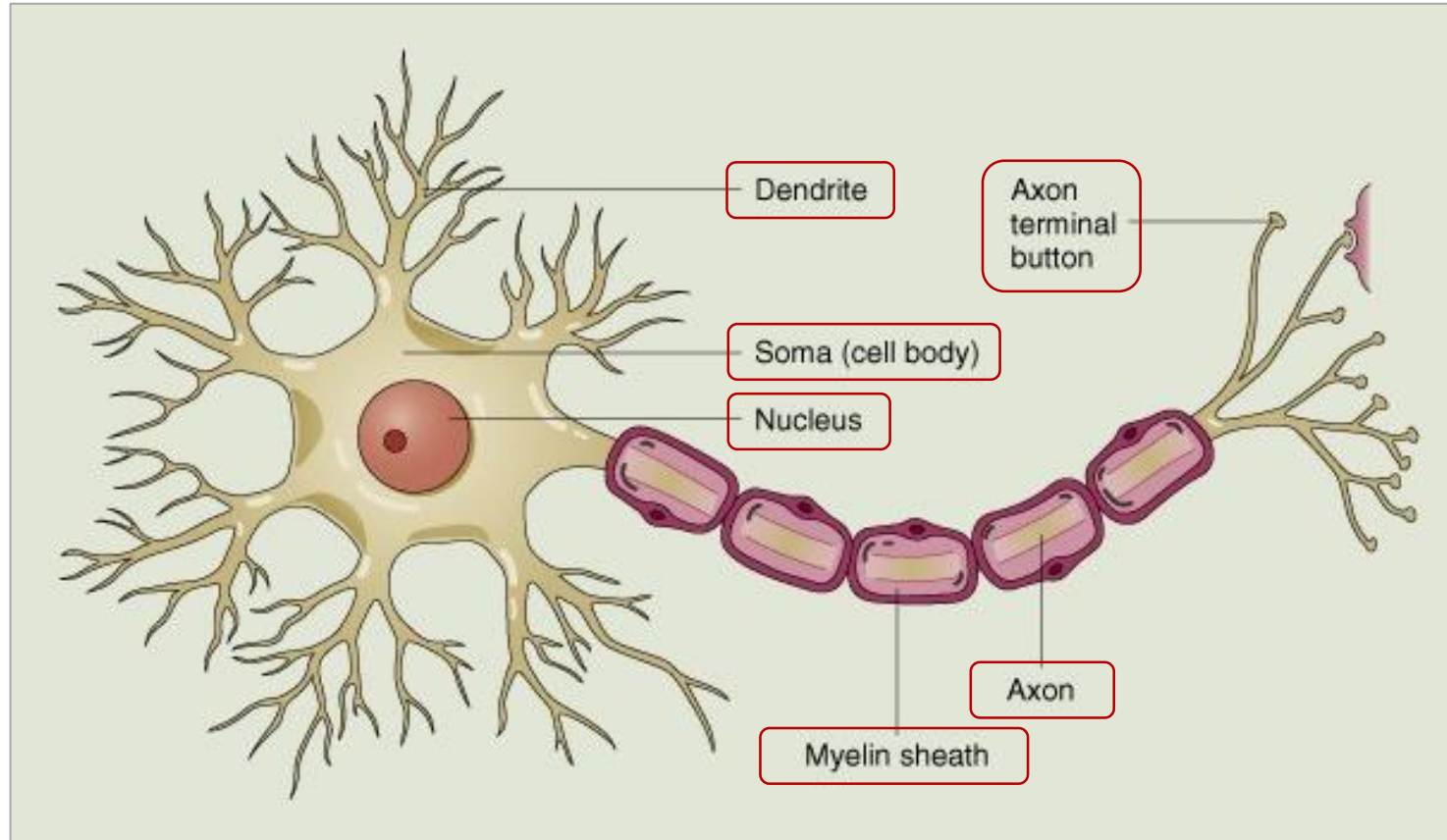
预测和仿真是决策的关键。



Major Internal Parts of the Human Brain

人类大脑的主要内部部件

Parts of a Neuron



Dendrite, Soma, Nucleus, Myelin sheath, Axon, Axon terminal button

树突, 细胞体, 细胞核, 髓鞘, 轴突, 轴突末端突触

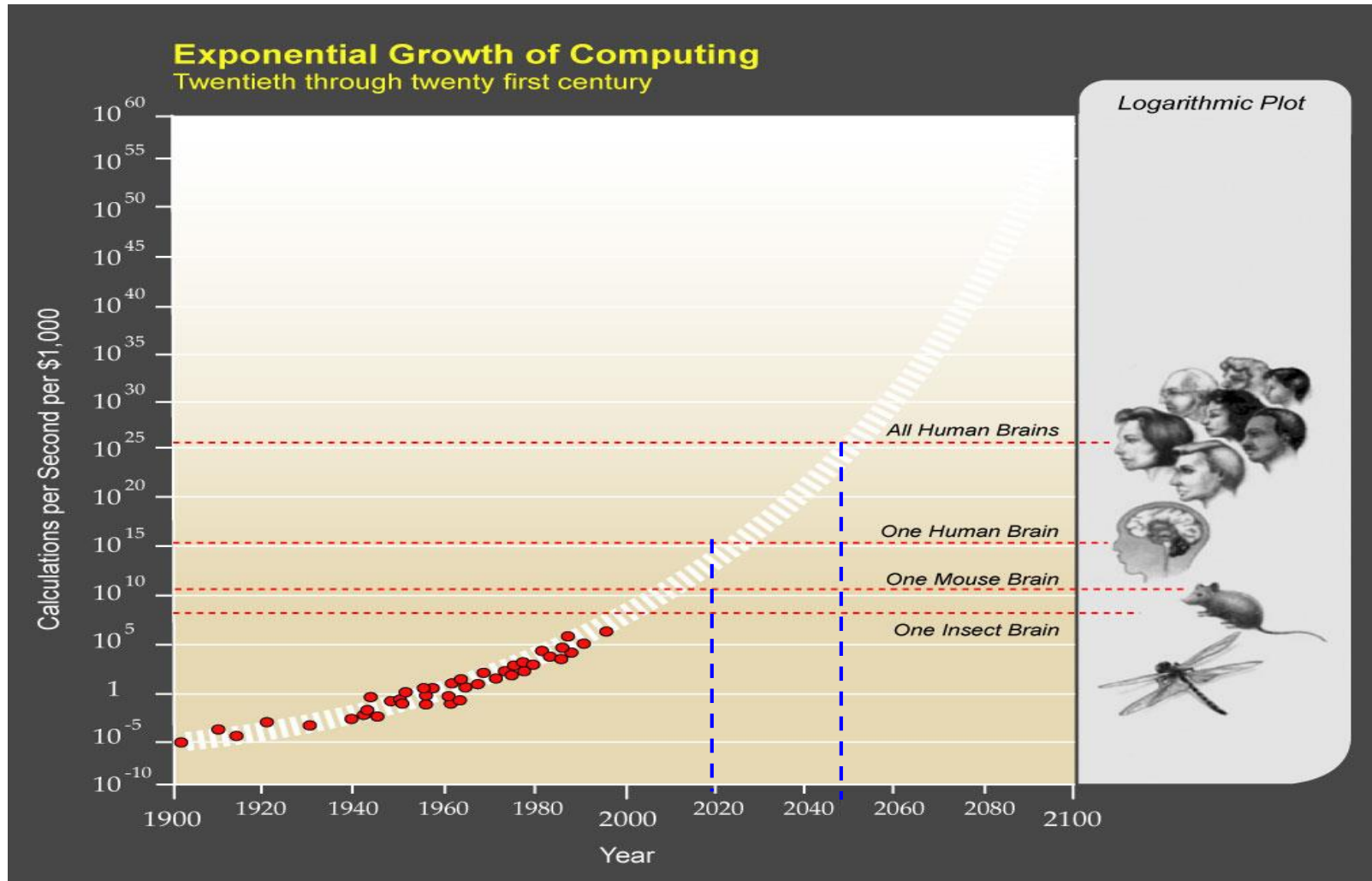
Computer vs. Human brain

- ❑ The brain's numbers are essentially fixed.
大脑中神经元的数量基本是固定的。
- ❑ The computer's numbers have been increasing by a factor of 10 every 5 years.
计算机中处理单元的数量，每5年增加10倍。

	Computer	Human Brain
Computational units	CPU, 64 bits, 10^9 transistors	10^{11} neurons
Storage units	10^{11} bits RAM 10^{13} bits disk	10^{11} neurons 10^{14} synapses
Cycle time	10^{-9} sec	10^{-3} sec
Operations/sec	10^{10}	10^{17}
Memory updates/sec	10^{10}	10^{14}

Computer vs. Human brain

Source: <http://waitbutwhy.com/>



Cognitive Psychology

-- *How do humans think and act?*

认知心理学：-- 人类如何思考与行动？

- Views the brain as an information-processing device, studies **mental processes**:
把大脑看作是信息处理设备，是研究心智过程的学科：

attention	■ 注意机制
language use	■ 语言运用
memory	■ 记忆
perception	■ 感知
problem solving	■ 问题求解
creativity	■ 创造力
thinking	■ 思考

Mental processes

□ Attention

- A state of focused awareness on a subset of available perceptual information.
注意机制：意识集中在某个有用感知信息子集的状态。

□ Memory

- Three sub-classes:
procedural memory, semantic memory, episodic memory.
记忆：三个子集：过程记忆、语义记忆、情景记忆。

□ Perception

- Physical senses (sight, smell, hearing, taste, touch, and proprioception), as well as their cognitive processes.
感知：物理感知（视觉、嗅觉、味觉、知觉），及其认知过程。

Mental processes

□ Language

- Study language acquisition, individual components of language formation, how language use is involved in mood, or numerous other related areas.

语言：研究语言习得、语言形成的组件、语言使用时的语气、或者许多其它相关领域。

□ Metacognition

- It is “cognition about cognition”, “thinking about thinking”, or “knowing about knowing”.

元认知：它是“关于认知的认知”、“关于思考的思考”、或者“关于认识的认识。”

- There are generally two components of metacognition:

knowledge about cognition, and regulation of cognition.

元认知通常有两个组成部分：关于认知的知识，以及认知的调节。

Cognitive psychology vs. Cognitive science

□ Cognitive psychology

be often involved in running psychological experiments involving human participants, with the goal of gathering information related to how the human mind takes in, processes, and acts upon inputs received from the outside world.

认知心理学：通常通过人类参与者的心理实验来收集信息，其目的是研究人脑如何接受外部世界的输入、如何处理以及作用等。

□ Cognitive science

be concerned with gathering data through research, which has links to philosophy, linguistics, anthropology, neuroscience, and particularly with artificial intelligence.

认知科学：关注于通过研究收集数据，其涉猎心理学、语言学、人类学、神经科学、社会学和教育学，尤其是人工智能。



Control theory and cybernetics

-- *How can artifacts operate under their own control?*

控制理论与控制论：--机器如何能在其自身的控制下运行？

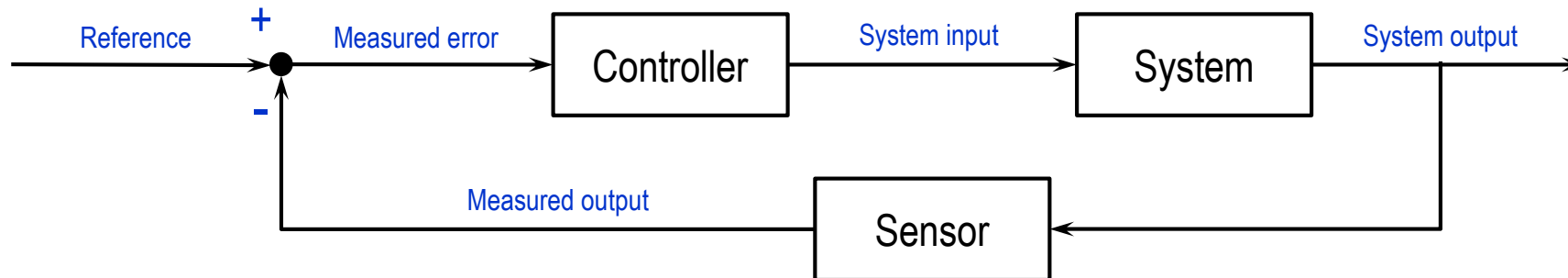
□ Control theory

- An interdisciplinary branch of engineering and mathematics.

控制理论：工程与数学的交叉学科分支。

- Deal with the behavior of dynamical systems with inputs, and how their behavior is modified by feedback.

处理动态系统对输入的行为，以及该行为如何通过反馈进行调整。



Control theory and cybernetics

□ Cybernetics

- A transdisciplinary approach for exploring regulatory systems, their structures, constraints, and possibilities.

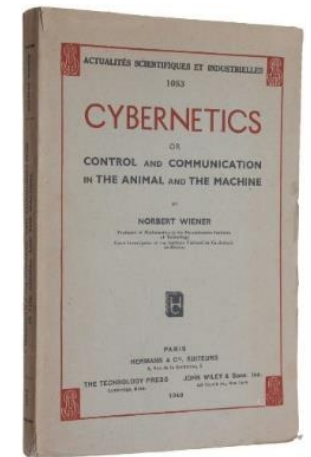
跨学科的研究途径，探索调控系统、它们的结构、约束、和可能性。

- Defined in 1948 as “the scientific study of control and communication in the animal and the machine.”

控制论：1948被定义为 “研究动物与机器的控制与通信的科学”。

- In the 21st century, the term is often used in a rather loose way to imply “control of any system using technology.”

21世纪，该术语通常被简单通俗地解释为 “用技术控制任何系统”。



Thank you for your attention!

