

本期追问

「意识是主观的」为何不妨碍它成为客观科学的研究对象？

模拟一场思维过程，为什么不等于真的拥有思维？

如果「计算」只是观察者赋予的解释，机器还能自己理解吗？

视频精读 VIDEO STUDY

Consciousness in Artificial Intelligence | John Searle | Talks at Google

来源：YouTube 视频 (<https://www.youtube.com/watch?v=rHKwIYsPXLg>)

节目发布：2015-12-03 · 全文 116 段 · 页边注释 76 条 · 生词词组 100 条

目录 CONTENTS

导读 · 讲者与视频总结

第一部分 · 全文对照

01 开场：主持人介绍塞尔	0:01
02 技术进步为何需要被理解	1:10
03 客观与主观的双重歧义	1:49
04 观察者相对与观察者独立	5:20
05 认知科学与中文房间的由来	8:37
06 中文房间论证及各种反驳	11:02
07 信息与智能的两种含义	18:51
08 「计算」一词的历史与歧义	25:58
09 机器能否思考：人工大脑类比	31:00
10 结论：计算不足以构成思考	37:04
11 库兹韦尔对谈：意识如何验证	38:41
12 现场问答：模拟、涌现与意识科学	45:58

第二部分 · 核心句型 · 值得模仿的表达

第三部分 · 生词总表 · 复习用

第四部分 · 理解自测 · 8 题

第五部分 · 我的划线与笔记

导读 视频总结

本期讲者

约翰·塞尔 加州大学伯克利分校斯拉瑟哲学讲席教授，语言哲学与心灵哲学代表人物，曾获让·尼科奖和美国国家人文奖章。1980 年提出「中文房间」论证，是对强人工智能最著名的哲学反驳。

雷·库兹韦尔 发明家、未来学家，时任谷歌工程总监，《奇点临近》作者，以「技术奇点」预测和对强 AI 的乐观立场著称，是塞尔的长期论辩对手。

约翰·布拉卡利亚 谷歌 YouTube 运营部门员工，谷歌内部 AI 讨论小组「奇点网络」负责人，本场活动主持人。

第一部分 全文对照 · 附页边注释与生词标注

01 开场：主持人介绍塞尔

0:01 约翰·布拉卡利亚（主持人）

JOHN BRACAGLIA: My name is John Bacaglia. And I'm a Googler working in YouTube operations. I also lead a group called the Singularity Network, an internal organization focused on discussions and rationality in artificial intelligence. I'm pleased to be here today with Mr. John Searle. As a brief introduction, John Searle is the Slusser Professor of Philosophy at the University of California–Berkeley. He is widely noted for his contributions to the philosophy of language, philosophy of mind, and social philosophy.

约翰·布拉卡利亚：我叫约翰·布拉卡利亚。我是谷歌员工，在 YouTube 运营部门工作。我还负责一个叫「奇点网络」的小组，这是一个内部组织，专注于人工智能领域的讨论与理性思考。今天很高兴能与约翰·塞尔先生一起在这里。简单介绍一下，约翰·塞尔是加州大学伯克利分校的斯拉瑟哲学讲席教授。他因其在语言哲学、心灵哲学和社会哲学方面的贡献而广为人知。

「奇点网络」是谷歌内部讨论 AI 的兴趣小组；「奇点」指技术奇点，即机器智能超越人类的假想临界点，与后文库兹韦尔的著作《奇点临近》直接呼应。

0:32 约翰·塞尔

John has received the Jean Nicod Prize, the National Humanities Medal in the Mind and Brain prize for his work. Among his noble concepts is the Chinese room argument against strong artificial intelligence. John Searle, everyone. [APPLAUSE] JOHN SEARLE: Thank you Thank you. Many thanks. It's great to be back at Google. It is a university outside of a university. And sometimes, I think, this is what a university ought really to look like. Anyway, it's just terrific to be here. And I'm going to talk about some— well, I'm going to talk about a whole lot of stuff.

约翰曾获让·尼科奖、国家人文奖章，以及「心智与大脑奖」等荣誉。他最著名的概念之一是「中文房间」论证，用来反驳强人工智能。有请约翰·塞尔。[掌声]约翰·塞尔：谢谢，谢谢。非常感谢。很高兴又回到谷歌。这里就像大学之外的一所大学。有时候我会想，大学本来就应该长这个样子。总之，能来到这里真是太棒了。我今天要讲的是——嗯，我要讲的东西其实挺多的。

让·尼科奖是法国认知哲学领域的最高奖项。「中文房间」论证 1980 年发表于《行为与脑科学》杂志，是针对「强人工智能」（程序本身即心灵）最著名的哲学反驳。

02 技术进步为何需要被理解

1:10

But, basically, I want to start with talking about the significance of technological advances. And America, especially, but everybody, really, is inclined to just celebrate the advances. If they got a self-driving car who the hell cares about whether or not it's conscious. But I'm going to say there are a lot of things that matter for certain purposes about the understanding of the technology. And that's really what I'm going to talk about. Now to begin with, I have to make a couple rather boring distinctions because you won't really understand contemporary intellectual life if you don't understand these distinctions.

不过，基本上，我想先谈谈技术进步的意义。尤其是美国，不过其实所有人都一样，总是倾向于一味地为技术进步欢呼。如果有了一辆自动驾驶汽车，谁还会在乎它到底有没有意识呢。但我要说，在某些方面，对这项技术的理解是非常重要的。这才是我今天真正要讲的内容。首先，我得先做几个有点枯燥的区分，因为如果你不理解这些区分，你就没法真正理解当代的思想界。

开场即亮明立场：塞尔不反对技术本身，反对的是对技术成就的哲学误读。这一区分贯穿全场，也预告了他后面对自动驾驶、深蓝等成果的宽容态度。

03 客观与主观的双重歧义

1:49

In our culture, there's a big deal about objectivity and subjectivity. We strive for an objective science. The problem is that these notions are systematically ambiguous in a way that produces intellectual catastrophes. They're ambiguous between a sense, which is epistemic, where epistemic means having to do with knowledge— epistemic— and a sense, which is ontological, where ontological means having to do with existence. I hate using a lot of fancy polysyllabic words. And I'll try to keep them to a minimum.

在我们的文化里，人们非常看重客观性和主观性。我们追求一门客观的科学。问题在于，这些概念存在系统性的歧义，而这种歧义会造成思想上的灾难。它们的歧义在于：一种含义是认识论上的，「认识论的」意思是与知识有关——认识论——另一种含义是本体论上的，「本体论的」意思是与存在有关。我很讨厌用一大堆花哨的多音节词。我会尽量少用它们。

epistemic（认识论的，关于知识）与 ontological（本体论的，关于存在）是全场论证的地基。塞尔认为「客观/主观」在这两个维度上的歧义造成了大量思想混乱。

polysyllabic /ˌpɑːlɪsɪˈlæbɪk/ *adj.*
多音节的（塞尔自嘲爱用大词）

epistemic /ˌepɪˈstemɪk/ *adj.*
认识论的，与知识有关的

ontological /ˌɑːntəˈlɑːdʒɪkl/ *adj.*
本体论的，与存在方式有关的

ambiguous /æmˈbɪɡjuəs/ *adj.*
有歧义的，模棱两可的

2:27

But I need these two, epistemic and ontological. Now the problem with objectivity and subjectivity is that they're systematically ambiguous—I'll just abbreviate subjectivity—between an epistemic sense and an ontological sense. Epistemically, the distinction is between types of knowledge claims. If I say, Rembrandt died in 1606, well—no, he didn't die then. He was born then. I'd say Rembrandt was born in 1606. That is to say, it's a matter of objective fact. That's epistemically objective. But if I say Rembrandt is the greatest painter that ever lived, well, that's a matter of opinion.

但这两个词我必须要用：认识论的（epistemic）和本体论的（ontological）。客观性和主观性的问题在于，它们存在系统性的歧义——我就把主观性简写一下——歧义存在于认识论意义和本体论意义之间。在认识论上，这个区分是关于知识主张的类型的。如果我说，伦勃朗死于1606年，呃——不对，他不是那年去世的。他是那年出生的。我应该说，伦勃朗生于1606年。也就是说，这是一个客观事实的问题。这就是认识论上客观的。但如果我说伦勃朗是有史以来最伟大的画家，那就是个见仁见智的问题了。

3:14

That is epistemically subject. So we have epistemic objectivity and subjectivity. Underlying that is a distinction in modes of existence. Lots of things exist regardless of what anybody thinks. Mountains, molecules, and tectonic plates have a mode of existence that is ontologically objective. But pains and pickles and itches, they only exist insofar as they are experienced by a subject. They are ontologically subjective. So I want everybody to get that distinction because it's very important because—well, for a lot of reasons, but one is lots of phenomena that are ontologically subjective admit of an account which is epistemically objective.

那是认识论上主观的。所以我们有了认识论上的客观性和主观性。在这背后，还有一个关于存在方式的区分。很多东西的存在不取决于任何人的想法。高山、分子和构造板块，它们的存在方式是本体论上客观的。但疼痛、痒这些感受，它们只有在被一个主体体验到时才存在。它们是本体论上主观的。所以我希望每个人都掌握这个区分，因为它非常重要，因为——嗯，原因有很多，但其中一个，很多本体论上主观的现象，是可以有一种认识论上客观的解释的。

伦勃朗（1606—1669），荷兰黄金时代画家。「生于1606年」可验证真假，是认识论上客观的；「最伟大的画家」是品味判断，是认识论上主观的。

abbreviate /ə'brɪ:vɪət/ v.

缩写，简写

关键一步：疼痛等体验只有在被主体经验到时才存在（本体论主观），但这不妨碍对它们做认识论上客观的研究——意识科学的合法性由此确立。

tectonic plates *phr.*

（地质）构造板块

insofar as *phr.*

在...的限度内；只有当...（正式书面语）

admit of *phr.*

容许有，容纳（正式用法，主语多为事物）

我的笔记

认识论上的客观性和主观性

4:05

I first got interested in this kind of stuff. I thought, well, why don't these brain guys solve the problem of consciousness. And I went over UCSF to their neurobiology gang and told them, why the hell don't you guys figure out how the brain causes consciousness? What am I paying you to do? And their reaction was, look, we're doing science. Science is objective. And you, yourself, admit that consciousness is subjective. So there can't be a science of consciousness. Now you'll all recognize that's a fallacy of ambiguity.

我最初对这类问题产生兴趣的时候，我想，为什么这些研究大脑的家伙不去解决意识的问题呢。于是我跑去加州大学旧金山分校（UCSF）找他们神经生物学那帮人，对他们说，你们这帮人到底为什么不去搞清楚大脑是怎么产生意识的？我付钱让你们干什么的？他们的反应是：听着，我们做的是科学。科学是客观的。而你自己也承认，意识是主观的。所以不可能有一门关于意识的科学。现在你们都能看出来，这是一个歧义谬误。

UCSF 即加州大学旧金山分校，以医学与神经科学著称。科学家以「科学必须客观」为由拒绝研究意识，正是混淆了两种「客观」的歧义谬误（fallacy of ambiguity）。

fallacy /'fæləsi/ *n.*

谬误；fallacy of ambiguity 歧义谬误

4:39

Science is indeed epistemically objective because we strive for claims that can be established as true or false, independent of the attitudes of the makers and interpreters of the claim. But epistemic objectivity of the theory does not preclude an epistemically objective account of a domain that's ontologically subjective. I promised you I wouldn't use too many big words, but anyway there are a few. The point is this. You can have an epistemically objective science of consciousness, even though consciousness is ontologically subjective.

科学确实是认识论上客观的，因为我们追求的是那种可以被确立为真或假的主张，不依赖于提出主张的人和解释主张的人的态度。但理论在认识论上的客观性，并不妨碍我们对一个本体论上主观的领域做出认识论上客观的解释。我答应过你们不用太多大词，不过还是有那么几个。要点是这样的。你可以有一门认识论上客观的关于意识的科学，即使意识是本体论上主观的。

preclude /prɪ'klu:d/ *v.*

排除，使不可能（preclude A from B）

04 观察者相对与观察者独立

5:20

Now that's going to be important. And there's another distinction. Since not everybody can see this, I'm going to erase as I go along. There's another distinction which is crucial. And that's between phenomena that are observer-independent. And there I'm thinking of mountains and molecules and tectonic plates, how they exist regardless of what anybody thinks. But the world is full of stuff that matters to us that is observer-relative. It only exists relative to observers and users. So, for example, the piece of paper in my wallet is money.

这一点接下来会很重要。还有另一个区分。因为不是每个人都能看到这块板子，我会边讲边擦。还有一个至关重要的区分。那就是不依赖于观察者（observer-independent）的现象。这里我指的是高山、分子和构造板块，它们的存在不取决于任何人怎么想。但这个世界上还充满了对我们很重要、却是相对于观察者（observer-relative）的东西。它们只有相对于观察者和使用者才存在。比如说，我钱包里的那张纸是钱。

observer-relative（相对于观察者）：其存在依赖人的态度与用法，如货币、婚姻、大学。这是塞尔《社会实在的建构》中社会本体论的核心概念。

6:08

But the fact that makes it money is not a fact of its chemistry. It's a fact about the attitudes that we have toward it. So money is observer-relative. Money, property, government, marriage, universities, Google, cocktail parties, and summer vacations are all observer-relative.

但让它成为钱的那个事实，并不是它的化学性质。而是我们对它所持的态度。所以钱是相对于观察者的。钱、财产、政府、婚姻、大学、谷歌、鸡尾酒会，还有暑假，这些都是相对于观察者的。

一张纸成为「钱」靠的是集体承认而非物理化学属性。塞尔称之为「地位功能」：我们把某种地位赋予对象，使其获得物理性质之外的功能。

6:30

And that has to be distinguished from observer-independent. And notice now, all observer-relative phenomenon are created by human consciousness. Hence, they contain an element of ontological subjectivity. But you already know that you can have, in some cases, an epistemically objective science of a domain that is observer-relative. That's why you can have an objective science of economics even though the phenomena studied by economics is, in general, observer-relative, and hence contains an element of ontological subjectivity.

这必须和不依赖于观察者的东西区分开来。注意，所有相对于观察者的现象都是由人类意识创造出来的。因此，它们都包含着一种本体论主观性的成分。但你们已经知道，在某些情况下，你可以对一个相对于观察者的领域建立一门认识论上客观的科学。这就是为什么你可以有一门客观的经济学，尽管经济学所研究的现象总体上是相对于观察者的，因而包含着本体论主观性的成分。

7:11

Economists tend to forget that. They tend to think that economics is kind of like physics, only it's harder. When I studied economics, I was appalled. We learned that marginal cost equals marginal revenue in the same tone of voice that in physics we learned that force equals mass times acceleration. They're totally different because the stuff in economics is all observer-relative and contains an element of ontological subjectivity. And when the subjectivity changes—ffft—the whole thing collapses.

经济学家往往忘了这一点。他们往往以为经济学有点像物理学，只不过更难。我当年学经济学的时候，简直惊呆了。我们学'边际成本等于边际收益'时的口吻，和我们在物理课上学'力等于质量乘以加速度'时一模一样。它们完全是两码事，因为经济学里的东西全都是相对于观察者的，包含着本体论主观性的成分。而当这种主观性一变——噗——整个体系就崩塌了。

塞尔批评经济学的「物理学羡慕」：「边际成本=边际收益」描述的是观察者相对的现象，依赖人的态度，与「力=质量×加速度」这类自然定律性质完全不同。

appalled /ə'pɔ:lɪd/ *adj.*

震惊的，骇然的

marginal revenue *phr.*

(经济学) 边际收益

7:45

That was discovered in 2008. This is not a lecture about economics. I want you to keep all that in mind. Now that's important because a lot of the phenomena that are studied in cognitive science, particularly phenomena of intelligence, cognition, memory, thought, perception, and all the rest of it have two different senses. They have one sense, which is observer-independent, and another sense, which is observer-relative. And, consequently, we have to be very careful that we don't confuse those senses because many of the crucial concepts in cognitive science have as their reference phenomena that are observer-relative and not observer-independent.

2008年大家就见识到了。这不是一堂经济学课。我希望你们把这些都记在心里。这很重要，因为认知科学所研究的很多现象，尤其是智能、认知、记忆、思维、知觉等等这些现象，都有两种不同的意义。它们有一种意义是不依赖于观察者的，另一种意义是相对于观察者的。因此，我们必须非常小心，不要混淆这两种意义，因为认知科学中许多关键概念所指称的现象是相对于观察者的，而不是不依赖于观察者的。

2008 年金融危机被引为例证：市场信心（主观态度）崩塌导致整个体系失灵。随后话锋一转：认知科学的核心概念同样存在观察者独立/相对两种意义。

cognition /kəˈɡnɪʃn/ *n.*

认知，认识过程

05 认知科学与中文房间的由来

8:37

I'm going to get to that. OK, everybody up with us so far? I want everything to sound so obvious you think, why does this guy bore us with these platitudes? Why doesn't he say something controversial? Now I'm going to go and talk about some intellectual history. Many years ago, before any of you were born, a new discipline was born. It was called cognitive science. And it was founded by a whole bunch of us who got sick of behaviorism in psychology, effectively. That was the reason for it. And the Sloan Foundation used to fly us around to lecture, mostly to each other.

这个我后面会讲到。好，到目前为止大家都跟上吧？我希望这些听起来都显而易见，让你们心想：这家伙干嘛拿这些老生常谈来烦我们？他怎么不说点有争议的东西？接下来我要讲一讲一段思想史。很多年前，在你们所有人出生之前，一门新学科诞生了。它叫认知科学。它是由我们一帮人创立的，说白了，就是一群受够了心理学中行为主义的人。这就是它诞生的缘由。斯隆基金会（Sloan Foundation）当时经常出钱送我们四处讲学，主要是讲给彼此听。

认知科学兴起于 1950—70 年代，是对行为主义（只研究可观察行为、拒谈内在心理）的反叛。斯隆基金会的资助促成了这门交叉学科的成型。

platitudes /ˈplætətʊːdz/ *n.*

陈词滥调，老生常谈

behaviorism /bɪˈheɪvjərɪzəm/ *n.*

（心理学）行为主义

9:17

But anyway, that's all right. We were called Sloan Rangers. And I was invited to lecture to the Artificial Intelligence Lab at Yale. And I thought, well, christ, I don't know anything about artificial intelligence. So I went out and bought a book written by the guys at Yale. And I remember thinking, \$16.95 plus tax— money wasted. But it turned out I was wrong. They had in there a theory about how computers could understand. And the idea was that you give the computer a story. And then you ask the computer questions about the story.

不过没关系。我们被称作'斯隆游侠' (Sloan Rangers)。后来我受邀去耶鲁的人工智能实验室做讲座。我心想，哎呀老天，我对人工智能可是一窍不通。于是我出去买了一本耶鲁那帮人写的书。我记得当时心想：16.95美元外加税——钱白花了。但结果证明我错❖❖。书里有一套关于计算机如何能'理解'的理论。思路是这样的：你给计算机一个故事。然后你就故事内容向计算机提问。

9:57

And the computer would give the correct answer to the questions even though the answer was not contained in the story. A typical story. A guy goes into a restaurant and orders a hamburger. When they brought him the hamburger, it was burned to a crisp. The guy stormed out of the restaurant and didn't even pay his bill. Question, did the guy eat the hamburger? Well, all of you computers know the answer to that. No, the guy didn't eat the hamburger. And I won't tell you the story where the answer is yes.

计算机给出正确的答案，即使答案并没有包含在故事里。一个典型的故事是这样的。一个人走进餐馆，点了一个汉堡。汉堡端上来的时候，已经烤得焦黑。这人怒气冲冲地走出餐馆，连账都没付。问题：这个人吃汉堡了吗？嗯，你们这些'计算机'都知道答案。没有，这人没吃汉堡。答案是'吃了'的那个版本我就不讲了。

指耶鲁 AI 实验室 Roger Schank 团队的「脚本」理论：用餐馆脚本等常识框架，让程序回答故事中没有明说的内容，当时被视为机器「理解」的证据。

汉堡故事是 Schank 的经典演示：程序能推断出「没吃」这一未明说的答案。强 AI 阵营认为这证明计算机真的理解了故事——这正是塞尔要拆解的靶子。

burned to a crisp *phr.*
烧得焦脆，烤得焦黑

stormed out *phr. v.*
怒气冲冲地夺门而出

10:24

It's equally boring. Now, the point was this proves that the computer really understands the story. So there I was on my way to New Haven on United Airlines at 30,000 feet. And I thought, well, hell, they could give me these stories in Chinese. And I could follow the computer program for answering stories. And I don't understand a word of the story. And I thought, well, that's an objection they must have thought of. And besides that won't keep me going for a whole week in New Haven. Well, it turned out they hadn't thought of it.

一样无聊。关键在于，他们说这证明了计算机真的理解了这个故事。于是当时我坐着美联航的飞机去纽黑文，在三万英尺的高空。我心想，嘿，他们完全可以用中文给我这些故事。我可以照着那个回答问题的计算机程序去操作。而故事里的话我一个字都不懂。我又想，这个反驳他们肯定早想到了。再说，这点东西也不够我在纽黑文撑一整个星期。结果呢，他们还真没想到过。

中文房间的诞生场景：塞尔在赴耶鲁讲学的飞机上想到，自己手动执行同一程序也能答对中文问题，却一个字都不懂——可见执行程序不等于理解。

06 中文房间论证及各种反驳

11:02

And everybody was convinced I was wrong. But interestingly they all had different reasons for thinking I was wrong. And the argument has gone on longer than a week. It's gone on for 35 years. I mean, how often do I have to refute these guys? But anyway, let's go through it. The way the argument goes in its simplest version is I am locked in a room full of Chinese— well, they're boxes full of Chinese symbols and a rule book in English for manipulating the symbols. Unknown to me, the boxes are called a database, and the rule book is called a program.

而且所有人都坚信我是错的。但有意思的是，他们认为我错的理由各不相同。这场争论持续的时间可远不止一个星期。它已经持续了35年。我是说，我到底要驳倒这些家伙多少次？不过不管怎样，我们来过一遍这个论证。这个论证最简单的版本是这样的：我被锁在一个房间里，里面全是中文——嗯，是装满中文符号的箱子，还有一本用英文写的操作这些符号的规则手册。我不知道的是，这些箱子叫做'数据库'，那本规则手册叫做'程序'。

论证到讲座时已争论 35 年。注意关键映射：装满中文符号的箱子=数据库，英文规则手册=程序，房间里的塞尔本人=CPU。

refute /rɪˈfju:t/ v.

驳倒，反驳

11:40

In coming in the room, I get Chinese symbols. Unknown to me, those are questions. I look up what I'm supposed to do. And after I shuffle a lot of symbols, I give back other symbols. And those are answers to the questions. Now we will suppose— I hope your bored with this, because I am. I mean, I've told this story many times. We will suppose that they get so good at writing the program, I get so good at shuffling the symbols, that my answers are indistinguishable from a native Chinese speaker. I pass the Turing test for understanding Chinese.

有中文符号被送进房间来。我不知道的是，那些是问题。我查手册看自己该做什么。在摆弄了一大堆符号之后，我递出另一些符号。而那些就是问题的答案。现在我们假设——我希望你们已经听腻了，因为我自己都腻了。我是说，这个故事我讲过太多遍了。我们假设他们把程序写得炉火纯青，我摆弄符号也练得炉火纯青，以至于我的答案和一个以中文为母语的人毫无区别。我通过了'理解中文'的图灵测试。

图灵测试：若机器在对话中让人无法区分其与真人，即算「能思考」。塞尔设定自己完美通过中文版图灵测试却毫无理解，以此攻击该测试的效力。

shuffle /ˈʃʌfl/ *v.*

挪动，来回摆弄（本义为洗牌）

indistinguishable /ˌɪndɪˈstɪŋɡwɪʃəbl/ *adj.*

无法区分的（+from）

12:13

All the same, I don't understand a word of Chinese. And there's no way in the Chinese room that I could come to understand Chinese because all I am is a computer system. And the rules I operate are a computer program. And— and this is the important point— the program is purely syntactical. It is defined entirely as a set of operations over syntactical elements. To put it slightly more technically, the notion same implemented program defines an equivalence class that is specified completely independently of any physics and, in particular, independent of the physics of its realization.

尽管如此，我还是一个中文字都不懂。而且在这个'中文屋'里，我根本不可能学会理解中文，因为我整个就是一个计算机系统。我所执行的规则就是一个计算机程序。而且——这是关键所在——这个程序是纯粹句法性的。它完全被定义为一组作用于句法元素之上的操作。说得稍微技术性一点，“同一个被实现的程序”这个概念定义了一个等价类，这个等价类的规定完全独立于任何物理性质，尤其是独立于它得以实现的那种物理载体。

「纯句法」是论证枢纽：同一程序可在任何物理载体上实现（多重可实现性），其定义独立于物理——这恰恰意味着程序不可能内在地携带语义。

syntactical /sɪnˈtæktɪkl/ *adj.*

句法的，语形的（与语义相对）

equivalence class *phr.*

（数学）等价类

12:53

The bottom line is if I don't understand the questions and the answers on the basis of implementing the program, then neither does any other digital computer on that basis because no computer has anything that I don't have. Computers are purely syntactical devices. Their operations are defined syntactically. And human intelligence requires more than syntax. It requires a semantics. It requires an understanding of what's going on. You can see this if you contrast my behavior in English with my behavior in Chinese.

归根结底就是，如果我不能凭借运行这个程序来理解那些问题和答案，那么任何其他数字计算机也同样不能凭此理解，因为没有哪台计算机拥有任何我所没有的东西。计算机是纯粹的句法装置。它们的运算是按句法来定义的。而人类智能需要的不只是句法。它还需要语义。它需要对正在发生的事情有所理解。你可以通过对比我用英语的表现和我用中文的表现来看清这一点。

整个论证浓缩于此：程序是句法的；心灵有语义；句法本身不足以产生语义；因此运行程序永远不足以构成理解。后文 102—103 段再次以公理形式给出。

semantics /sə'mæntiks/ *n.*

语义；语义学（与 syntax 句法相对）

13:31

They ask me questions in English. And I give answers in English. They say, what's the longest river in the United States? And I say, well, it's the Mississippi, or the Mississippi-Missouri, depending on if you count that as one river. They ask me in Chinese, what's the longest river in China? I don't know what the question is or what it means. All I got are Chinese symbols. But I look up what I'm supposed to do with that symbol, and I give back an answer, which is the right answer. It says, it's the Yangtze.

他们用英语问我问题。我用英语给出回答。他们问：美国最长的河流是什么？我说：嗯，是密西西比河，或者是密西西比-密苏里河，这取决于你是否把它们算作一条河。他们用中文问我：中国最长的河流是什么？我不知道这是什么问题，也不知道它是什么意思。我拿到的只是一些中文符号。但我查一下该怎么处理那个符号，然后给出一个回答，而且是正确的回答。它说：是长江。

塞尔用自己英文问答（有理解）与中文问答（纯符号操作）的对照说明：外在行为表现完全相同，内在状态却可以天差地别——直接挑战行为主义标准。

14:02

That's the longest river in China. I don't know any of that. I'm just a computer. So the bottom line is that the implemented computer program by itself is never going to be sufficient for human understanding because human understanding has more than syntax. It has a semantics. There are two fundamental principles that underlie the Chinese room argument. And both of them seem to me obviously true. You can state each in four words. Syntax is not semantics. And simulation is not duplication. You can simulate— you're going to have plenty of time for questions.

那就是中国最长的河流。但这些我一概不知。我只是一台计算机。所以结论就是，被实现的计算机程序本身永远不足以产生人类的理解，因为人类的理解不只有句法，它还有语义。中文房间论证背后有两条基本原则。在我看来，这两条都显而易见是对的。每一条都可以用四个词来表述。句法不是语义。模拟不是复制。你可以模拟——你们待会儿会有充足的提问时间。

两条四词原则是全场最可引用的总结：Syntax is not semantics（句法不是语义）；Simulation is not duplication（模拟不是复制）。

duplication /ˌduːplɪˈkeɪʃn/ *n.*

复制，复现（与 simulation 模拟相对）

14:43

How much time we got, by the way? I want to— JOHN BRACAGLIA: We'll leave time for questions at the end. JOHN SEARLE: I want everybody that has a question to have a chance to ask the question. Anyway, that's the famous Chinese room argument. And it takes about five minutes to explain it. Now you'd be amazed at the responses I got. They were absolutely brehtaking in their preposterousness. Now let me give you some answers. A favorite answer was this. You were there in a room. You had all those symbols.

对了，我们还有多少时间？我想——约翰·布拉卡利亚：我们会在最后留出提问时间。约翰·塞尔：我希望每一个有问题的人都有机会提问。总之，这就是著名的中文房间论证。解释它大概只需要五分钟。接下来，我收到的那些回应会让你们大吃一惊。它们的荒谬程度简直令人叹为观止。现在让我给你们讲几个回应。有一个很受欢迎的回应是这样的。你当时待在一个房间里。你有那些符号。

brehtaking /ˈbreɪtɪŋ/ *adj.*

惊人的，令人叹为观止的（此处为反讽）

preposterousness /prɪ

ˈpɑːstərəsnəs/ *n.*

荒谬，荒诞不经

15:18

You had a box. You probably had scratch paper on which to work. Now, it wasn't you that understood. You're just a CPU, they would say with contempt, the Central Processing Unit. I didn't know what any of these words meant in those days. CPU, it's the system that understands. And when I first heard this, I mean, the room understands Chinese, I said to the guy. And he said, yes, the room understands Chinese. Well, it's a desperate answer. And I admire courage. But it's got a problem. And that is the reason I don't understand is I can't get from the syntax to the semantics.

你有一个盒子。你可能还有用来演算的草稿纸。那么，理解中文的并不是你。你只是个CPU——他们会带着轻蔑说——中央处理器。那时候我根本不知道这些词是什么意思。CPU？他们说，能理解的是整个系统。我第一次听到这个说法时，我是说，我对那家伙说：这个房间理解中文？他说：是的，这个房间理解中文。嗯，这是个孤注一掷的回答。我钦佩这种勇气。但它有个问题。问题在于，我之所以不理解，是因为我无法从句法通达语义。

16:00

But the room can't either. How does the room get from the syntax of the computer program of the input symbols to the semantics of the understanding of the symbols? There's no way the room can get there because that would require some consciousness in the room in addition to my consciousness. And there is no such consciousness. Anyway, that was one of many answers. One of my favorites was this. This was in a public debate. A guy said to me, but suppose we ask you, do you understand Chinese? And suppose you say, yes, I understand Chinese.

可房间也同样做不到。房间怎么可能从输入符号的计算机程序句法，通达对这些符号之理解的语义呢？房间根本无法做到这一点，因为那需要房间里除了我的意识之外还存在某种别的意识。而并不存在这样的意识。总之，那只是众多回应之一。我最喜欢的回应之一是这样的。那是在一场公开辩论上。有个人对我说：可是，假设我们问你，你懂中文吗？假设你回答：是的，我懂中文。

「系统回应」(systems reply) 是文献中最主流的反驳：理解不属于房间里的人（他只是CPU），而属于整个系统。塞尔即将逐层拆解这一回应。

contempt /kən'tempt/ *n.*

轻蔑，鄙视 (with contempt 轻蔑地)

塞尔的反驳要点：房间同样无法从句法通达语义；说「房间理解中文」等于要求房间里存在另一个意识主体，而这恰是论证所否认的。

16:36

Well? Well, OK, let's try that and see how far we get. I get a question that looks like this. Now, this will be in a dialect of Chinese some of you won't recognize. Unknown to me, that symbol means, do you understand Chinese? I look up what I'm supposed to do. And I give them back a symbol that's in the same dialect of Chinese. And it looks like that. And that says, why do you guys ask me such dumb questions? Can't you see that I understand Chinese? I could go on with the other responses and objections, but I think they're all equally feeble.

feeble /'fi:bl/ *adj.*

无力的，站不住脚的（形容论证）

那又怎样呢？好吧，那我们就试试看，看能走多远。我收到一个长这样的问题。这里用的是一种中文方言，你们有些人可能认不出来。我并不知道，那个符号的意思是：你懂中文吗？我查一下该怎么做。然后我用同一种中文方言还给他们一个符号。它长那个样子。它的意思是：你们干嘛问我这么蠢的问题？难道看不出来我懂中文吗？我还可以继续讲其他的回应和反驳，但我认为它们都同样站不住脚。

17:25

The bottom line is there's a logical truth. And that is that the implemented computer program is defined syntactically. And that's not a weakness. That's the power. The power of the syntactical definition of computation is you can implement it on electronic machines that can perform literally millions of computations in a very small amount of time. I'm not sure I believe this, but it always says it in the textbooks, that Deep Blue can do 250 million computations in a second. OK, I take their word for it.

深蓝 (Deep Blue): IBM 国际象棋计算机，1997 年击败世界冠军卡斯帕罗夫。塞尔强调：句法定义不是计算机的弱点，恰是它可被高速电子实现的威力所在。

归根结底，这里有一条逻辑真理。那就是：被实现的计算机程序是按句法定义的。而这并不是弱点。这正是它的威力所在。计算的句法定义之所以强大，就在于你可以把它实现在电子机器上，这些机器能在极短的时间内执行数以百万计的运算。我不确定自己信不信这个，但教科书里总是这么写：“深蓝”每秒能进行2.5亿次运算。好吧，我姑且相信他们的说法。

18:04

So it's not a weakness of computers. Now, another argument I sometimes got was, well, in programs, we often have a section called the semantics of natural understanding programs. And that's right. But, of course, what they do is they put in more computer implementation. They put in more syntax. Now, so far, so good. And I think if that's all there was to say, I've said all of that before. But now I want to go on to something much more interesting. And here goes with that. Now how we doing? I'm not-- everybody seems to understand there's going to be plenty of time for questions.

所以这不是计算机的弱点。另外，我有时听到的另一种论证是：嗯，在程序里，我们经常有一个叫“自然语言理解程序的语义”的部分。这没错。但当然，他们所做的不过是加入更多的计算机实现。他们加入的是更多的句法。到目前为止，一切都还好。我想，如果要说的就只有这些，那这些我以前都说过了。但现在我想接着讲一些有意思得多的东西。那我们就开始吧。现在进行得怎么样？我不是——大家似乎都明白后面会有充足的提问时间。

18:41

I insist on a good question period. So let me take a drink of water, and we go to the next step, which I think is more important.

我坚持要有一个高质量的提问环节。那么让我喝口水，我们进入下一步，我认为这一步更重要。

07 信息与智能的两种含义

18:51

A lot of people thought, well, look, maybe the computer doesn't understand Chinese, but all the same, it does information processing. And it does, after all, do computation. That's what we define the machine to do. And I had to review a couple of books recently. One book said that we live in a new age, the age of information. And in a wonderful outburst, the author said everything is information. Now that ought to worry us if everything is information. And I read another book. This was an optimistic book.

很多人会想：好吧，你看，也许计算机确实不理解中文，但不管怎样，它确实在做信息处理。而且它毕竟确实在做计算。那正是我们对这台机器所定义的功能。最近我不得不评论几本书。有一本书说，我们生活在一个新时代，信息时代。然后在一段精彩的高论中，作者说：万物皆信息。如果万物皆信息，那我们可就该担心了。我还读了另一本书。那是一本乐观的书。

19:37

I reviewed— this for "The New York Review of Books"— a less optimistic book by a guy who said computers are now so smart they're almost as smart as we are. And pretty soon, they'll be just as smart as we are. And then I don't have to tell this audience the next step. They'll be much smarter than we are. And then look out because they might get sick of being oppressed by us. And they might simply rise up and overthrow us all. And this, the author said modestly— I guess this is how you sell books— he said this may be the greatest challenge that humanity has ever faced, the upcoming revolt of super-smart computers.

我评论过——是给《纽约书评》写的——另一本没那么乐观的书，作者说计算机现在已经聪明到几乎和我们一样聪明了。而且很快，它们就会和我们一样聪明。接下来的一步，我都不必告诉在座各位了。它们会比我们聪明得多。然后可要当心了，因为它们可能会受够了被我们压迫。它们可能干脆揭竿而起，把我们全都推翻。而这位作者“谦虚地”说——我猜书就是这么卖出去的——他说，这可能是人类有史以来面临的最大挑战：即将到来的超级智能计算机的反叛。

「万物皆信息」影射信息本体论思潮。塞尔的担忧：这种说法抹平了「有意识主体内在拥有的信息」与「被我们归属给机器的信息」之间的界限。

所评的书即「超级智能反叛」叙事（塞尔曾在《纽约书评》评论博斯特罗姆的《超级智能》）。此段的讽刺语气为后文「鞋子不会造反」的类比埋下伏笔。

overthrow /ˌoʊvərˈθroʊ/ v.

推翻，颠覆（政权等）

oppressed /əˈprest/ adj.

受压迫的

20:25

Now, I want to say both of these claims are silly. I mean, I'm speaking shorthand here. There'll be plenty of chance to answer me. And I want to say briefly why. The notion of intelligence has two different senses. It has an observer-independent sense where it identifies something that is psychologically real. So I am more intelligent than my dog Tarski. Now, Tarski's pretty smart, I agree. But overall, I'm smarter than Tarski. I've had four dogs, by the way— Frege, Russell, Ludwig, and Tarski.

塞尔的四条狗以弗雷格、罗素、路德维希（维特根斯坦）、塔斯基命名——四位都是分析哲学与近代逻辑的奠基人物，属于哲学家式的幽默。

现在我想说，这两种说法都很愚蠢。我是说，我这里讲得比较简略。你们会有大把机会来反驳我。我想简要说说原因。“智能”这个概念有两种不同的含义。它有一种独立于观察者的含义，指的是某种在心理上真实存在的东西。所以，我比我的狗塔斯基更聪明。当然，塔斯基挺聪明的，我承认。但总体上，我比塔斯基聪明。顺便说一句，我养过四条狗——弗雷格、罗素、路德维希，还有塔斯基。

21:08

And Tarski, he's a Bernese mountain dog. I'm sorry I didn't bring him along, but he's too big for the car. Now, he's very smart. But he does have intelligence in the same sense that I do. Only he happens to have somewhat less than I do. Now, my computer is also intelligent. And it also processes information. But— and this is the key point— it's observer-relative. The only sense in which the computer has intelligence is not in an intrinsic, but it's in an observer-relative sense. We can interpret its operations in such a way that we can make— now, watch this terminology— we can make epistemically objective claims of intelligence even though the intelligence in question is entirely in the eye of the beholder.

核心论点：狗的智能与人的智能是同一种（内在的、观察者独立的），只是程度不同；计算机的「智能」则完全是我们解释的产物，存在于观察者眼中。

intrinsic /m' trɪnsɪk/ *adj.*

内在固有的（本讲座核心术语，与观察者相对相对立）

in the eye of the beholder

phr.

取决于观察者，见仁见智（源自谚语「情人眼里出西施」）

塔斯基是一条伯恩山犬。很抱歉我没把它带来，因为它太大了，车里装不下。它非常聪明。而且它拥有的智能和我的智能是同一种意义上的。只不过它碰巧比我少一些罢了。而我的计算机也是“智能”的。它也处理信息。但是——这是关键所在——那是相对于观察者而言的。计算机拥有智能的唯一意义，不是内在意义上的，而是相对于观察者的意义上的。我们可以用某种方式来解释它的运算，从而做出——注意这里的术语——我们可以做出在认识论上客观的智能断言，尽管所谈论的这种智能完全存在于观察者的眼中。

22:03

This was brought home forcefully to me when I read in the newspapers that IBM had designed a computer program which could beat the world's leading chess player. And in the same sense in which Kasparov beat Karpov so we were told Deep Blue beat Kasparov. Now that ought to worry us because for Karpov and Kasparov to play chess, they both have to be conscious that they're playing chess. They both have to know such things as I opened with pawn to king four, and my queen is threatened on the left-hand side of the board.

这一点让我深有体会，是在我从报纸上读到IBM设计出一个能击败世界顶尖棋手的计算机程序的时候。而就在卡斯帕罗夫击败卡尔波夫的那种意义上，我们被告知，深蓝击败了卡斯帕罗夫。这应该让我们警觉，因为对卡尔波夫和卡斯帕罗夫来说，要下国际象棋，他们俩都必须有意识地知道自己在下棋。他们俩都必须知道这样的事情，比如我开局走了王前兵进四，我的后在棋盘左侧受到了威胁。

卡斯帕罗夫与卡尔波夫是 1980—90 年代争夺世界棋王的宿敌。塞尔指出「深蓝击败卡斯帕罗夫」这句话偷换了「下棋」的含义：深蓝根本不知道自己在下棋。

brought home *phr.*

bring sth home to sb 使某人深切体会到

pawn /pɔ:n/ *n.*

（国际象棋）兵；喻指棋子、被利用者

22:42

But now notice, Deep Blue knows none on that because it doesn't know anything. You can make epistemically objective claims about Deep Blue. It made such and such a move. But the attributions of intelligent chess playing, this move or that move, it's all observer-relative. None of it is intrinsic. In the intrinsic sense in which I have more intelligence than my dog, my computer has zero intelligence—absolutely none at all. It's a very complex electronic circuit that we have designed to behave as if it were thinking, as if it were intelligent.

但请注意，深蓝对这些一无所知，因为它什么都不知道。你可以对深蓝做出认识论上客观的论断。比如它走了某某一步棋。但把“智能下棋”归于它，无论这步棋还是那步棋，全都是相对于观察者的。没有一样是内在固有的。就我比我的狗更聪明的那种内在意义而言，我的电脑智能为零——绝对是一点儿也没有。它是一个非常复杂的电子电路，是我们设计出来让它表现得好像在思考、好像有智能一样。

「表现得好像在思考」即塞尔所说的 as-if（拟似）意向性，与内在意向性相对。可以对深蓝做客观陈述（它走了某步棋），但「智能」的归属全是观察者相对的。

attributions /,ætrə'bjʊ:ʃnz/ *n.*

归属，归因（把某性质归于某物）

23:24

But in the strict sense, in the observer-independent sense in which you and I have intelligence, there is zero intelligence in the computer. It's all observer-relative. And what goes for intelligence goes for all of the key notions in cognitive science. The notions of intelligence, memory, perception, decision-making, rationality-- all those have two different senses, a sense where they identify psychologically real phenomena of the sort that goes on in you and me and the sort where they identify observer-relative phenomena.

但在严格意义上，在你我拥有智能的那种不依赖于观察者的意义上，计算机里的智能为零。全都是相对于观察者的。而适用于智能的道理，同样适用于认知科学里所有的关键概念。智能、记忆、知觉、决策、理性——所有这些概念都有两种不同的含义：一种含义指的是心理上真实的现象，就是发生在你我身上的那种；另一种含义指的是相对于观察者的现象。

论证在此推广：智能、记忆、知觉、决策、理性——认知科学的全部核心词汇都有「心理真实」与「观察者相对」两种用法，混用即产生谬误。

24:02

But in the intrinsic sense in which you and I have intelligence, the machinery we're talking about has zero intelligence. It's no question of its having more or less. It's not in the same line of business. All of the intelligence is in the eye of the beholder. It's all observer-relative. Now, you might say-- and I would say-- so, for most purposes, it makes no difference at all. I mean, if you can design a car that can drive itself, who cares if it's conscious or not? Who cares if it literally has any intelligence?

但在你我拥有智能的那种内在意义上，我们谈论的这些机器智能为零。根本不存在它智能多一点还是少一点的问题。它压根就不属于同一个范畴。所有的智能都存在于观察者的眼中。全都是相对于观察者的。你可能会说——我自己也会说——所以，就大多数用途而言，这根本没什么区别。我是说，如果你能设计出一辆会自动驾驶的汽车，谁在乎它有没有意识呢？谁在乎它是不是真的有什么智能呢？

line of business *phr.*

行当，范畴（not in the same line of business 不属同类）

24:35

And I agree. For most purposes, it doesn't matter. For practical purposes, it doesn't matter whether or not you have the observer-independent or the observer-relative sense. The only point where it matters, if you think there's some psychological significance to the attribution of intelligence to machinery which has no intrinsic intelligence. Now, notice the intelligence by which we—the mental processes by which we attribute intelligence to the computer require consciousness. So the attribution of observer-relativity is done by conscious agents.

我同意这一点。就大多数用途而言，这无关紧要。从实际角度来说，你用的究竟是不依赖于观察者的含义，还是相对于观察者的含义，都无所谓。唯一要紧的地方在于，如果你认为把智能归于机器这件事有某种心理学上的意义，而这机器本身并没有内在的智能。请注意，我们借以——我们把智能归于计算机时所依赖的那些心理过程，是需要意识的。所以，观察者相对性的归属是由有意识的主体完成的。

微妙的一步：把智能归属给机器这一解释行为本身需要意识。观察者相对性由有意识的主体创造——为下一段「意识本身不是观察者相对的」做铺垫，避免无穷倒退。

25:16

But the consciousness is not itself observer-relative. The consciousness that creates the observer-relative phenomena is not itself observer-relative. But now let's get to the crunch line then. If information is systematically ambiguous between an intrinsic sense, in which you and I have information, and an observer-relative sense, in which the computer has information, what about computation? After all, computation, that must surely be intrinsic to the computer. That's what we designed and built the damn things to do, was computation.

但意识本身并不是相对于观察者的。创造出观察者相对现象的那个意识，其本身并不是相对于观察者的。那么现在我们来谈关键问题。如果“信息”这个词系统性地存在歧义，一边是内在的含义，即你我拥有信息的那种含义，一边是相对于观察者的含义，即计算机拥有信息的那种含义，那计算又如何呢？毕竟，计算总该是计算机内在固有的吧。我们设计和制造这些鬼东西，就是为了让它们做计算。

crunch line *phr.*

关键点，紧要处（crunch 指关键时刻）

08 「计算」一词的历史与歧义

25:58

But, of course, the same distinction applies. And I want to take a drink of water and think about history for a moment. When I first read Alan Turing's article, it was called "Computing Machinery and Intelligence." Now why didn't he call it "Computers and Intelligence"? Well, you all know the answer. In those days, "computer" meant "person who computes." A computer is like a runner or a piano player. It's some human who does the operation. Nowadays nobody would think that because the word has changed its meaning.

但当然，同样的区分也适用于这里。我想喝口水，顺便回顾一下历史。我第一次读艾伦·图灵那篇文章时，它的标题是《计算机与智能》。那他为什么不叫它《计算机与智能》呢？嗯，你们都知道答案。在那个年代，“computer”指的是“做计算的人”。“computer”就像“跑步者”或“钢琴演奏者”一样，指的是执行这项操作的某个人。如今没人会这么想了，因为这个词已经改变了含义。

图灵 1950 年论文《计算机与智能》提出模仿游戏（图灵测试）。当时 computer 一词指「从事计算的人」（多为人工计算员），所以图灵不能直接说 computers。

26:37

Or, rather, it's acquired the systematic ambiguity between the observer-relative sense and the observer-independent sense. Now we think that a computer names a type of machinery, not a human being who actually carries out computation. But the same distinction that we've been applying, the same distinction that we discovered in all these other cases, that applies to computation in the literal or observer-independent sense in which I will now do a simple computation. I will do a computation using the addition function.

或者更准确地说，它获得了那种系统性的歧义，介于相对于观察者的含义和不依赖于观察者的含义之间。现在我们认为“计算机”指的是一种机器，而不是一个实际执行计算的人。但我们一直在运用的那个区分，也就是我们在所有其他情形中发现的那个区分，同样适用于字面意义上、也就是不依赖于观察者意义上的计算——现在我就来做一个简单的计算。我要用加法函数做一个计算。

27:21

And here's how it goes. It's not a very big deal. One plus one equals two. Now, the sense in which I carried out a computation is absolutely intrinsic and observer-independent. I don't care what anybody says about me. If the experts say, well, you weren't really computing. No, I was. I consciously did a computation. When my pocket calculator does the same operation, the operation is entirely observer-relative. Intrinsically all that goes on is a set of electronic state transitions that we have designed so that we can interpret computationally.

是这样做的。没什么大不了的。一加一等于二。我刚才执行计算的这种意义，是绝对内在的、不依赖于观察者的。我不在乎别人怎么说我。如果专家们说，你其实并没有真的在计算——不，我就是在计算。我是有意识地做了一次计算。而当我的袖珍计算器执行同样的运算时，这个运算就完全是相对于观察者的了。从内在来看，发生的只是一系列电子状态的转换，我们把它们设计成这样，以便我们能从计算的角度去解读。

27:58

And, again, to repeat, for most purposes, it doesn't matter. When it matters is when people say, well, we've created this race of mechanical intelligences. And they might rise up and overthrow us. Or they attribute some other equally implausible psychological interpretation to the machinery. In commercial computers, the computation is observer-relative. Now notice, you all know that doesn't mean it's epistemically subjective. And I pay a lot of money so that Apple will make a piece of machinery that will implement programs that my earlier computers were not intelligent enough to implement.

再重复一遍，就大多数用途而言，这无关紧要。要紧的时候，是当人们说：瞧，我们创造出了一个机械智能的种族，它们可能会起来推翻我们。或者当人们给这些机器赋予别的同样不靠谱的心理学解释的时候。在商用计算机中，计算是相对于观察者的。但请注意，你们都知道，这并不意味着它在认识论上是主观的。我花了不少钱，好让苹果公司造出一台能运行那些程序的机器，而我以前的电脑还不够智能，跑不动那些程序。

塞尔现场演算 $1+1=2$ ：人的计算是有意识的、内在的、观察者独立的；计算器的「计算」只是被我们按计算方式解读的电子状态转换——同词不同义。

重要澄清：观察者相对不等于认识论主观。「这台电脑在运行某程序」仍是客观可验证的陈述。要紧的只是不要从中推出机器有心理状态。

implausible /ɪmˈplɔːzəbl/ *adj.*

难以置信的，不合情理的

28:41

Notice the observer–relative attribution of intelligence here. So it's absolutely harmless unless you think there's some psychological significance. Now what is lacking, of course, in the machinery, which we have in human beings which makes the difference between the observer relativity of the computation in the commercial computer and the intrinsic or observer independent computation that I have just performed on the blackboard, what's lacking is consciousness. All observer–relative phenomena are created by human and animal consciousness.

全场论证收拢到一点：商用计算机的计算与人在黑板上的计算，差别就在意识。所有观察者相对现象都由人类和动物的意识创造。

注意，这里对智能的归属就是相对于观察者的。所以这完全无伤大雅，除非你认为这里面有什么心理学上的意义。那么，机器里所缺少的东西——当然，我们人类身上有它——正是它造成了商用计算机中计算的观察者相对性，与我刚才在黑板上进行的那种内在的、不依赖于观察者的计算之间的差别——所缺少的就是意识。所有相对于观察者的现象，都是由人类和动物的意识创造的。

29:20

But the human and animal consciousness that creates them is not itself observer–relative. So there's an intrinsic mental phenomena, the consciousness of the agent, which creates the observer–relative phenomena, or interprets the mechanical system in an observer relative fashion. But the consciousness that creates observer relativity is not itself observer–relative. It's intrinsic. Now, I wanted to save plenty of time for discussion. So let me catch my breath and then give a kind of summary of the main thrust of what I've been arguing.

但创造出它们的人类和动物的意识，其本身并不是相对于观察者的。所以存在一种内在的心理现象，即主体的意识，它创造出相对于观察者的现象，或者以一种相对于观察者的方式去解读机械系统。但创造出观察者相对性的那个意识，本身并不是相对于观察者的。它是内在的。好，我想留出充足的时间来讨论。让我喘口气，然后对我一直在论证的主要观点做一个概括。

29:59

And one of things I haven't emphasized but I want to emphasize now, and that is most of the apparatus, the conceptual apparatus, we have for discussing these issues is totally obsolete. The difference between the mental and the physical, the difference between the social and the individual, and the distinction between those features which can be identified in an observer-relative fashion, such as computation, and those which can be identified in an observer-independent fashion, such as computation.

有一点我之前没有强调，但现在想强调一下，那就是我们用来讨论这些问题的大部分工具，也就是概念工具，已经完全过时了。心理与物理的区别、社会与个体的区别，以及那些能以相对于观察者的方式来认定的特征——比如计算——与那些能以不依赖于观察者的方式来认定的特征之间的区分——比如计算。

塞尔主张现有概念工具（心/物、社会/个体二分）已经过时。这是他一贯立场：意识是生物现象，不应再套用笛卡尔式二元论词汇来讨论。

apparatus /ˌæpəˈrætəs/ *n.*

装置；（引申）概念工具、体系

obsolete /ˌɒbsəˈli:t/ *adj.*

过时的，废弃的

30:40

We're confused by the vocabulary which doesn't make the matters sufficiently clear. And I'm going to end this discussion by going through some of the elements of the vocabulary. Now, let me have a drink of water and catch my breath.

这套词汇让我们困惑，因为它没有把这些问题讲得足够清楚。而我打算在结束这场讨论时，把这套词汇中的一些要素过一遍。现在，让我喝口水，喘口气。

09 机器能否思考：人工大脑类比

31:00

Let's start with that old question, could a machine think? Well, I said the vocabulary was obsolete. And the vocabulary of humans and machines is already obsolete because if by machine is meant a physical system capable of performing certain functions, then we're all machines. I'm a machine. You're a machine. And my guess is only machines could think. Why? Well that's the next step. Thinking is a biological process created in the brain by certain quite complex, but insufficiently understood neurobiological processes.

我们从那个老问题开始：机器能思考吗？我说过，这套词汇已经过时了。而“人类”与“机器”这套说法本身就已经过时了，因为如果“机器”指的是一个能够执行某些功能的物理系统，那么我们都是机器。我是机器。你也是机器。而且我猜，只有机器才能思考。为什么？这就是下一步要讲的。思考是一种生物过程，是大脑中某些相当复杂、但我们了解得还不够充分的神经生物学过程所产生的。

31:45

So in order to think, you've got to have a brain, or you've got to have something with equivalent causal powers to the brain. We might figure out a way to do it in some other medium. We don't know enough about how the brain does it. So we don't know how to create it artificially. So could a machine think? Human beings are machines. Yes, but could you make an artificial machine that could think? Why not? It's like an artificial heart. The question, can you build an artificial brain that can think, is like the question, can you build an artificial heart that pumps blood.

所以要能思考，你得有一个大脑，或者你得有某种因果能力与大脑相当的东西。我们也许能想出办法在别的介质中实现它。我们对大脑是怎么做到的还了解得不够。所以我们还不知道怎么人工地创造它。那么，机器能思考吗？人类就是机器。可以。但你能造出一台会思考的人造机器吗？为什么不能呢？这就像人工心脏一样。“你能不能造出一个会思考的人工大脑”这个问题，就像“你能不能造出一个能泵血的人工心脏”这个问题。

反直觉之处：塞尔并非「机器不能思考」论者。若机器指能执行功能的物理系统，那么人就是机器，而且他猜「只有机器才能思考」——思考是生物过程。

人工心脏类比是理解塞尔立场的钥匙：造会思考的机器如同造泵血的心脏，必须复制大脑的因果能力。障碍在于我们的无知，而非原则上的不可能。

causal powers *phr.*

因果能力（塞尔核心术语：大脑产生意识的实际能力）

32:19

We know how the heart does it, so we know how to do it artificially. We don't know how the brain does it, so we have no idea. Let me repeat this. We have no idea how to create a thinking machine because we don't know how the brain does it. All we can do is a simulation using some sort of formal system. But that's not the real thing. You don't create thinking that way, whereas the artificial heart really does pump blood. So we had two questions. Could a machine think? And could an artificially-made machine think?

我们知道心脏是怎么工作的，所以我们知道怎么人工地实现它。我们不知道大脑是怎么做到的，所以我们完全没有头绪。让我再重复一遍。我们完全不知道该如何创造一台会思考的机器，因为我们不知道大脑是怎么做到的。我们能做的只是用某种形式系统来进行模拟。但那不是真正的思考。你没法用那种方式创造出思考，而人工心脏却是真的在泵血。所以我们有两个问题。机器能思考吗？人造的机器能思考吗？

32:52

Answer the question one is obviously yes. Answer to question two is, we don't know yet, but there's no obstacle in principle. Does everybody see that? Building an artificial brain is like building an artificial heart. The only thing is no one has begun to try it. They haven't begun to try it because they have no idea how the actual brain does it. So they don't know how to imitate actual brains. Well, OK, but could you build an artificial brain that could think out of some completely different materials, out of something that had nothing to do with nucleo-proteins, had nothing to do with neurons and neurotransmitters and all the rest of it.

第一个问题的答案显然是肯定的。第二个问题的答案是，我们还不知道，但原则上不存在障碍。大家都明白这一点吗？建造一个人工大脑就像建造一颗人工心脏。唯一的问题是还没有人开始尝试。他们还没开始尝试，是因为他们完全不知道真正的大脑是怎么运作的。所以他们不知道该如何模仿真实的大脑。好吧，那你能不能用一些完全不同的材料造出一个能思考的人工大脑，用一些和核蛋白毫无关系、和神经元、神经递质以及其他所有这些东西都毫无关系的材料。

「模拟/复制」再次现身：人工心脏真的在泵血（复制），而计算机对思考的模拟只是形式系统上的图像。我们不知道大脑如何产生思考，故无从复制。

neurotransmitters /ˌnɜːrəʊˈtrænzˌmɪtəz/ n.

'trænzˌmɪtəz/ n.

神经递质

33:32

And the answer is, again, we don't know. That seems to me an open question. If we knew how the brain did it, we might be able to define—I mean, be able to design machines that could do it using some completely different biochemistry in a way that the artificial heart doesn't use muscle tissue to pump blood. You don't need muscle tissue to pump blood. And maybe you don't need brain tissue to create consciousness. We just are ignorant. But notice there's no obstacle in principle. The problem is no one has begun to think about how you would build a thinking machine, how you'd build a thinking machine out of some material other than neurons because they haven't begun to think about how we might duplicate and not merely simulate what the brain actually does.

塞尔明确留下开放性：正如泵血不需要肌肉组织，产生意识或许也不需要脑组织——但前提是先弄清大脑的因果机制，再谈用其他材料复制。

答案还是：我们不知道。在我看来这是一个悬而未决的问题。如果我们知道大脑是怎么做到的，我们也许就能定义——我的意思是，就能设计出用某种完全不同的生物化学机制来实现思考的机器，就好比人工心脏并不需要用肌肉组织来泵血一样。泵血并不需要肌肉组织。也许创造意识也并不需要脑组织。我们只是无知而已。但请注意，原则上不存在障碍。问题在于，还没有人开始思考该如何建造一台会思考的机器，如何用神经元以外的材料建造一台会思考的机器，因为他们还没开始思考我们该如何复制——而不仅仅是模拟——大脑实际上所做的事情。

34:26

So the question, could a machine think, could an artificial machine think, could an artificial machine made out of some completely different materials, could those machines think? And now the next question is the obvious one. Well, how about a computer? Could a computer think? Now, you have to be careful here. Because if a computer is defined as anything that can carry out computations, well, I just did. This is a computation. So I'm a computer. And so are all of you. Any conscious agent capable of carrying out that simple computation is capable, is both a computer and capable of thinking.

所以问题是：机器能思考吗？人造机器能思考吗？用完全不同的材料制造出来的人造机器，那些机器能思考吗？接下来的问题是显而易见的。那么，计算机呢？计算机能思考吗？这里你得小心一点。因为如果计算机被定义为任何能执行计算的东西，那我刚刚就做了一次计算。这就是一次计算。所以我就是一台计算机。你们所有人也都是。任何有意识的、能够执行那种简单计算的主体，既是一台计算机，也具有思考能力。

35:09

So my guess is-- and I didn't have a chance to develop this idea-- is that not only can computers think-- you and me-- but my guess is that anything capable of thinking would have to be capable of carrying out simple computations. But now what is the status of computation? Well, the key element here is the one I've already mentioned. Computation has two senses, an observer-independent sense and an observer-relative sense. In the observer-relative sense, anything is a computer if you can ascribe a computational interpretation.

ascribe /ə'skraɪb/ v.

把...归于 (ascribe A to B)

所以我的猜测是——我还没来得及展开这个想法——不仅计算机能思考——比如你和我——而且我猜任何能思考的东西都必须具备执行简单计算的能力。但计算本身的地位又是什么呢？关键点就是我已经提到过的那一点。计算有两种意义：一种是独立于观察者的意义，一种是相对于观察者的意义。在相对于观察者的意义上，任何东西只要你能赋予它一种计算解释，它就是一台计算机。

35:47

Watch. I'll show you a very simple computer. That computer just computed a well-known function. s equals one-half gt squared. And if you had a good-enough watch, you could actually time and figure out how far the damn thing fell. Everybody sees. It's elementary mathematics. So if this is a computer, then anything is a computer because being a computer in the observer-relative sense is not an intrinsic feature of an object, but a feature of our interpretation of the physics of the phenomenon. In the old Chinese room days, when I had to debate these guys, at one point, I'd take my pen, slam it on a table, and say that is a digital computer.

落体「计算」 $s=\frac{1}{2}gt^2$ 是归谬法：若任何可被赋予计算解释的东西都算计算机，则万物皆计算机——「是计算机」不是物体的内在属性，而是解释的属性。

看好了。我给你们展示一台非常简单的计算机。这台计算机刚刚计算了一个著名的函数： s 等于二分之一 gt 平方。如果你有一块足够好的表，你真的可以计时，然后算出这玩意儿掉了多远。大家都看到了。这是初等数学。所以如果这也算计算机，那么任何东西都是计算机，因为在相对于观察者的意义上，作为一台计算机并不是一个物体的内在特征，而是我们对这一现象的物理过程所做解释的特征。在当年争论中文屋的日子里，我不得不和那些家伙辩论，有一次我拿起我的笔，往桌上一拍，说：这就是一台数字计算机。

36:32

It just happens to have a boring computer program. The program says stay there. The point is nobody ever called me on this because it's obviously right. It satisfies a textbook definition. You know, in the early days, they tried to snow me with a whole lot of technical razzmatazz. "You've left out the distinction between the virtual machine and the non-virtual machine" or "you've left out the transducers." You see, I didn't know what the hell a transducer was, a virtual machine. But it takes about five minutes to learn those things.

只不过它碰巧运行着一个很无聊的程序。程序的内容是：待在那儿别动。关键在于，从来没有人反驳过我这一点，因为这显然是对的。它符合教科书上的定义。要知道，早年间他们想用一大堆技术行话来唬住我。比如“你忽略了虚拟机和非虚拟机之间的区别”，或者“你忽略了换能器”。你看，我当时根本不知道换能器是什么鬼东西，也不知道虚拟机是什么。但这些东西五分钟就能学会。

这一归谬指向「泛计算主义」难题（普特南、查默斯等都讨论过）：如果计算的判定完全依赖解释，那么计算就不是自然类。塞尔的笔「运行着停在原地的程序」。

razzmatazz /ˈræzmətæz/ *n.*

花哨的噱头；唬人的行话（非正式）

snow /snoʊ/ *v.*

（美俚）用花言巧语蒙骗、唬住

transducer /trænsˈduːsər/ *n.*

换能器，传感器

10 结论：计算不足以构成思考

37:04

Anyway, so now we get to the crucial question in this. If computers can think, man-made computers can think, machines can think, what about computation? Does computation name a machine, a thinking process? That is, is computation, as defined by Alan Turing, is that itself sufficient for thinking? And you now know the answer to that. In the observer-relative sense, the answer is no. Computation is not a fact of nature. It's a fact of our interpretation. And insofar as we can create artificial machines that carry out computations, the computation by itself is never going to be sufficient for thinking or any other cognitive process because the computation is defined purely formally or syntactically.

好了，现在我们来到这里面最关键的问题。如果计算机能思考，人造的计算机能思考，机器能思考，那计算本身呢？计算指的是一种机器、一种思考过程吗？也就是说，按照艾伦·图灵所定义的计算，它本身足以构成思考吗？现在你们已经知道答案了。在相对于观察者的意义上，答案是否定的。计算不是自然界的事实。它是我们的解释所产生的事实。即便我们能造出执行计算的人工机器，计算本身也永远不足以构成思考或任何其他认知过程，因为计算是纯粹从形式上或句法上定义的。

全场正式结论：图灵意义上的计算不存在于自然界，只存在于我们对自然的解释中；因此计算本身永远不足以构成思考或任何认知过程。

37:56

Turing machines are not to be found in nature. They're to be found in our interpretations of nature. Now, let me add, a lot of people think, ah, this debate has something to do with technology or there'll be advances in technology. I think that technology's wonderful. And I welcome it. And I see no limits to the possibilities of technology. My aim is this talk is simply to get across, you shouldn't misunderstand the philosophical, psychological, and, indeed, scientific implication of the technology.

图灵机在自然界中是找不到的。它们存在于我们对自然的解释之中。另外我要补充一句，很多人觉得，啊，这场争论跟技术有什么关系，或者技术会取得什么进展。我认为技术非常美妙。我也欢迎技术。我认为技术的可能性没有极限。我这次演讲的目的仅仅是想让大家明白，你们不应该误解这项技术在哲学、心理学、乃至科学层面上的含义。

塞尔的自我定位：欢迎技术、不设技术上限，只反对从技术成就推出错误的哲学与科学结论。这为随后与谷歌工程师们的交锋定下基调。

get across *phr. v.*

把（观点）讲清楚，传达

38:28

Thank you very much. [APPLAUSE]

非常感谢大家。[掌声]

11 库兹韦尔对谈：意识如何验证

38:41 约翰·布拉卡利亚

JOHN BRACAGLIA: Thank you, John. JOHN SEARLE: I'm sorry I talk so fast, but I want to leave plenty of time for questions. JOHN BRACAGLIA: We'll start with one question from Mr. Ray Kurzweil. RAY KURZWEIL: Is this on? [INTERPOSING VOICES]

约翰·布拉卡利亚：谢谢你，约翰。约翰·塞尔：抱歉我讲得太快了，但我想留出充足的时间来回答问题。约翰·布拉卡利亚：我们先请雷·库兹韦尔先生提一个问题。雷·库兹韦尔：这个开着吗？[多人同时说话]

雷·库兹韦尔：发明家、未来学家，时任谷歌工程总监，《奇点临近》作者，预测 AI 将在 2029 年达到人类水平——他与塞尔就此争论了近二十年。

38:55 雷·库兹韦尔

RAY KURZWEIL: Well, thanks, John. I'm one of those guys you've been debating this issue for 18 years, I think. And I would praise the Chinese room for its longevity because it does really get at the apparent absurdity that some deterministic process like computation could possibly be responsible for something like thinking. And you point out the distinction of thinking between its effects and the subjective states, which is a synonym for consciousness. So I quoted you here in my book "Singularity is Near,"

雷·库兹韦尔：好的，谢谢你，约翰。我就是那些和你争论这个问题的人之一，大概争了18年了吧。我要赞扬中文屋论证的生命力，因为它确实切中了一个表面上的荒谬之处：像计算这样的确定性过程怎么可能产生像思考这样的东西。而且你指出了思考的一个区分：一边是它的外在效果，一边是主观状态，后者就是意识的同义词。所以我在我的书《奇点临近》里引用了你的话，

longevity /lɒːn'dʒevəti/ *n.*

长寿；持久的生命力

deterministic /dɪˌtɜːrməˈnɪstɪk/*adj.*

确定性的（给定输入必得同一输出）

absurdity /əbˈsɜːrɪdəti/ *n.*

荒谬，悖理

39:41

At the equivalence of neurons and even brains with machines. So then I took your argument why a machine and a computer could not truly understand what it's doing and simply substituted human brain for computers, since you said they were equivalent, and neurotransmitter concentrations and related mechanisms for formal symbols, since basically neurotransmitter concentrations, it's just a mechanistic concept. And so you wrote, with those substitutions, the human brain succeeds by manipulating neurotransmitter concentrations and other related mechanisms.

关于神经元乃至大脑与机器的等价性。于是我拿来你那套论证——为什么机器和计算机不可能真正理解它在做什么——然后简单地把“计算机”替换成“人脑”，因为你说过它们是等价的，再把“形式符号”替换成“神经递质浓度及相关机制”，因为说到底，神经递质浓度也只是一个机械性的概念。于是，经过这些替换，你写的就变成了：人脑是通过操纵神经递质浓度和其他相关机制来运作的。

库兹韦尔的策略是以子之矛攻子之盾：把塞尔论证中的「程序/形式符号」替换为「人脑/神经递质浓度」，推出「人脑也不可能真正理解」的荒谬结论。

40:24

The neurotransmitter concentrations and related mechanisms themselves are quite meaningless. They only have the meaning we have attached to them. The human brain knows nothing of this. It just shuffles the neurotransmitter concentrations and related mechanisms. Therefore, the human brain cannot have true understanding. So— [LAUGHTER] JOHN SEARLE: There's something interesting variations, again, on my original. RAY KURZWEIL: But the point I'd like to make, and that I'd be interested in your addressing, is the nature of consciousness because, I mean, you said today, and you wrote, the essential thing is to recognize that consciousness is a biological processes like digestion, lactation, photosynthesis, or mitosis.

lactation /læk'teɪʃn/ *n.*

泌乳

photosynthesis /ˌfəʊtəʊˈsɪnθəˌsɪs/ *n.*'sɪnθəˌsɪs/ *n.*

光合作用

mitosis /maɪ'toʊsɪs/ *n.*

(细胞) 有丝分裂

神经递质浓度和相关机制本身是毫无意义的。它们只有我们赋予它们的意义。人脑对此一无所知。它只是在搬弄神经递质浓度和相关机制而已。因此，人脑不可能拥有真正的理解。所以——[笑声]约翰·塞尔：这又是对我原始论证的一些有趣的变体。雷·库兹韦尔：但我想表达的观点，也是我希望你回应的，是意识的本质，因为，我是说，你今天说过，你也写过，关键是要认识到意识是一种生物过程，就像消化、泌乳、光合作用或有丝分裂一样。

41:12

We know that brains cause consciousness with specific biological mechanisms. But how do we know that a brain is conscious? How do you know that I'm conscious? And how do you— JOHN SEARLE: [INAUDIBLE] RAY KURZWEIL: And how do we know if a computer was conscious? We don't have a computer today that seems conscious, that's convincing in its responses. But my prediction is we will. We can argue about the time frame. And when we do, how do we know if it's conscious of it just seems conscious? How do we measure that?

我们知道大脑通过特定的生物机制产生意识。但我们怎么知道一个大脑是有意识的呢？你怎么知道我是有意识的？还有你怎么——约翰·塞尔：[听不清]雷·库兹韦尔：还有，我们怎么知道一台计算机是不是有意识的？我们今天还没有一台看起来有意识的、回应令人信服的计算机。但我的预测是将来会有。时间框架我们可以争论。等到那一天，我们怎么知道它是真的有意识还是仅仅看起来有意识？我们怎么衡量这一点？

库兹韦尔转向他心问题：意识只能第一人称通达，那么未来「看似有意识」的计算机与「真有意 识」如何区分？这是对塞尔立场 压力最大的问题。

41:45 约翰·塞尔

JOHN SEARLE: Well, there are two questions here. One is, if you do a substitution of words that I didn't use and the words I did use, can you get these observed results? And, of course, you can do that. That's a well-known technique of politicians. But that wasn't the claim. What is the difference between the computer and the brain? In one sentence, the brain is a causal mechanism that produces consciousness by a certain rather complex and still imperfectly understood neurobiological processes. But those are quite specific to a certain electrochemistry.

约翰·塞尔：嗯，这里有两个问题。一个是，如果你把我没用过的词替换掉我用过的词，你能得出这些所谓的结果吗？当然了，你可以那样做。那是政客们惯用的伎俩。但当初的主张并不是这个。计算机和大脑的区别是什么？一句话来说，大脑是一种因果机制，它通过某种相当复杂、至今仍未被完全理解的神经生物学过程来产生意识。但那些过程对特定的电化机制是非常特异的。

42:23

We just don't know the details. But we don't know if you mess around in the synaptic cleft, you're going to get weird effects. How does cocaine work? Well, it isn't because it's got a peculiar computational capacity. Because it messes with the capacity of the postsynaptic receptors to reabsorb quite specific neurotransmitters, norepinephrine— what are the other two? God, I'm flunking the exam here. Dopamine. Gaba is the third. Anyway, the brain, like the stomach or any other organ, is a specific causal mechanism.

我们只是还不了解其中的细节。但我们知道，你要是在突触间隙里乱动手脚，就会出现奇怪的效应。可卡因是怎么起作用的？嗯，并不是因为它有什么特殊的计算能力。而是因为它干扰了突触后受体重新摄取某些特定神经递质的能力，比如去甲肾上腺素——另外两种是什么来着？天哪，我这是要考试不及格了。多巴胺。第三种是 GABA（ γ -氨基丁酸）。总之，大脑就像胃或其他任何器官一样，是一种特定的因果机制。

塞尔的回应：根本不对称在于，大脑是依赖特定电化学的因果机制，而程序按定义与实现介质无关——任何能跑程序的硬件都等价，这正是二者的分野。

可卡因通过干扰突触对多巴胺、去甲肾上腺素等递质的再摄取起作用。塞尔以药理学说明：意识依赖特定生化机制，而非任何抽象的计算结构。

synaptic cleft *phr.*

突触间隙（神经元之间传递递质的空隙）

mess around *phr. v.*

乱动，瞎鼓捣（mess around in/with）

flunking /'flʌŋkiŋ/ *v.*

flunk（美口语）考砸，不及格

postsynaptic receptors *phr.*

突触后受体

43:01

And it functions on specific biochemical principles. The problem of the computer is it has nothing to do with the specifics of the implementation. Any implementation will do provided it's sufficient to carry out the steps in the program. Programs are purely formal or syntactical. The brain is not. The brain is a specific biological organ that operates on specific principles. And to create a conscious machine, we've got to know how to duplicate the causal powers of those principles. Now, the computer doesn't in that way work as a causal mechanism producing higher level features.

它按照特定的生化原理运作。而计算机的问题在于，它与实现方式的具体细节毫无关系。任何一种实现方式都行，只要它足以执行程序中的各个步骤。程序纯粹是形式化的，或者说是句法性的。而大脑不是。大脑是一个特定的生物器官，按照特定的原理运作。要造出一台有意识的机器，我们就得知道如何复制这些原理的因果能力。而计算机并不是以那种方式作为产生更高层次特征的因果机制来运作的。

43:41

Rather, computation names an abstract mathematical process that we have found ways to implement in specific hardware. But the hardware is not essential to the computation. Any system that can carry out the computation will be equivalent. Now, the second question is about how do you know about consciousness. Well, think about real life. How do I know my dog Tarski is conscious and this thing here, my smartphone, is not conscious? I don't have any doubts about either one. I can tell that Tarski is conscious not on behavioristic grounds.

相反，“计算”指的是一种抽象的数学过程，我们找到了在特定硬件上实现它的方法。但硬件对计算来说并不是本质性的。任何能够执行这一计算的系统都是等价的。那么，第二个问题是关于你怎么知道意识的存在。嗯，想想现实生活吧。我怎么知道我的狗塔斯基（Tarski）有意识，而我手里这个东西——我的智能手机——没有意识？这两点我都毫不怀疑。我能判断塔斯基有意识，并不是基于行为主义的理由。

塞尔答他心问题：判断狗有意识不靠「行为像人」，而靠因果结构相似性——相似的眼睛、耳朵、神经机制。推断的根据是机制，不是外在表现。

44:20

People say, well, it's because he behaves like a human being. He doesn't. See, human beings I know when they see me don't rush up and lick my hands and wag their tails. They just don't. My friends don't do that. But Tarski does. I can see that Tarski is conscious because he's got a machinery that's relatively similar to my own. Those are his eyes. These are his ears. This is his skin. He has mechanisms that mediate the input stimuli to the output behavior that are relatively similar to human mechanisms.

mediate /'mi:diət/ v.

传导，介导（在两者之间起中介作用）

人们会说，那是因为它的行为像人。它才不像。要知道，我认识的那些人看到我的时候可不会冲上来舔我的手、摇尾巴。他们就是不会。我的朋友可不干这种事。但塔斯基会。我能看出塔斯基有意识，是因为它拥有一套和我自己相对相似的机制。那是它的眼睛。这是它的耳朵。这是它的皮肤。它有一些把输入刺激传导为输出行为的机制，这些机制和人类的机制相对相似。

44:53

This is why I'm completely confident that Tarski's conscious. I don't know anything about fleas and termites. You know, your typical termite's got 100,000 neurons. Is that enough? Well, I lose 100,000 on a big weekend. So I don't know if that's enough for consciousness. But that's a factual question. I'll leave that to the experts. But as far as human beings are concerned there isn't any question that everybody in this room is conscious. I mean, maybe that guy over there is falling asleep, but there's no question about what the general— it's not even a theory that I hold.

白蚁约有 10 万个神经元，是否足以产生意识？塞尔特承这是留给经验科学的事实问题——显示他的划界并非教条，而是随证据开放。

termite /'tɜ:rmɪt/ n.

白蚁

这就是为什么我完全确信塔斯基是有意识的。至于跳蚤和白蚁，我就一无所知了。要知道，一只典型的白蚁大概有 10 万个神经元。这够吗？嗯，我一个放纵的周末就能损失 10 万个神经元。所以我不知道那是否足以产生意识。但那是个事实问题。我就把它留给专家们吧。但就人类而言，毫无疑问，这个房间里的每个人都是有意识的。我是说，也许那边那位老兄正在打瞌睡，但总体上毫无疑问——这甚至算不上我所持有的一种理论。

45:29

It's a background presupposition. The way I assume that the floor is solid, I simply take it for granted that everybody's conscious. If forced to justify it, I could. Now, there's always a problem about the details of other minds. Of course, I know you're conscious. But are you suffering the angst of post-industrial man under late capitalism? Well, I have a lot of friends who claim they do. And they think I'm philistine because I don't. But that's tougher. We'd have to have a conversation about that.

它是一个背景预设。就像我默认地板是坚实的一样，我理所当然地认为每个人都有意识。如果非要我论证，我也能做到。当然，关于他人心灵的细节，总是存在问题的。当然，我知道你是有意识的。但你是否正承受着晚期资本主义之下后工业时代人类的那种焦虑呢？嗯，我有很多朋友声称他们正在承受。而且他们觉得我很俗气，因为我没有。但那就难说了。我们得专门聊聊那个问题。

「他人有意识」被塞尔当作背景预设处理：如同默认地板坚实，日常中这不是需要证明的理论。真正困难的只是他人心灵的具体内容细节。

presupposition /ˌpriːsʌpəˈziʃn/ *n.*

预设，前提假定

angst /ɑːŋst/ *n.*

（存在主义式的）焦虑，忧惧
（源自德语）

philistine /ˈfɪləstiːn/ *n.*

庸人，不解风雅、无文化追求的人

12 现场问答：模拟、涌现与意识科学

45:58

But for consciousness, it's not a real problem in a real-life case.
AUDIENCE: So you've said that we haven't begun to understand how brains work or build comparable machines. But imagine in the future we do. So we can run a simulation, as you put it, of a brain. And then we interface it with reality through motor output, sensory input. What's the difference between that and a brain, which you say you know is producing consciousness? In JOHN SEARLE: In some cases, there's no difference at all. And the difference doesn't matter.

但就意识本身而言，在现实生活的情境中，这并不是一个真正的问题。观众：您说过，我们还没有开始理解大脑是如何运作的，也造不出与之相当的机器。但假设将来我们做到了。那么我们就可以像您说的那样，运行一个对大脑的模拟。然后我们让它通过运动输出和感觉输入与现实世界对接。那它和大脑之间又有什么区别呢？您说您知道大脑正在产生意识。呃——约翰·塞尔：在某些情况下，完全没有区别。而且这种区别也无关紧要。

46:32

If you've got a machine— I hope you guys are, in fact, building it because the newspapers say you are. If you've got a program that'll drive my car without a conscious driver, that's great. I think that's wonderful. The question is not, what can the technology do? My daddy was an electrical engineer for AT&T. And his biggest disappointment was I decided to be a philosopher, for God's sake, instead of going to Bell Labs and MIT as he had hoped. So I have no problem with the success of the technology.

如果你们有一台机器——我希望你们确实是在造，因为报纸上说你们在造。如果你们有一个程序，能在没有有意识的司机的情况下开我的车，那太棒了。我觉得那很了不起。问题不在于技术能做什么。我父亲是 AT&T 的电气工程师。他最大的失望就是我居然决定去当哲学家，老天爷，而不是像他期望的那样去贝尔实验室和麻省理工。所以我对技术的成功毫无意见。

47:06

The question is, what does it mean? Of course, if you've got a machine that can drive a car as well as I, or probably better than I can, then so much the better for the machinery. The question is, what is the philosophical psychological scientific significance of that? And if you think, well, that means you've created consciousness, you have not. You have to have more to create consciousness. And for a whole lot of things, consciousness matters desperately. In this case of this book that I reviewed, where the guy said, well, they got machines that are going to rise up and overthrow us all, it's not a serious possibility because the machines have no consciousness.

问题是，这意味着什么？当然，如果你有一台机器，开车开得和我一样好，或者很可能比我还好，那对机器来说就更好了。问题是，这在哲学、心理学和科学上有什么意义？如果你认为，这意味着你创造出了意识，那你并没有。要创造意识，你还需要更多的东西。而且对很多很多事情来说，意识至关重要。就拿我评论过的那本书来说，那家伙说，机器将会崛起并推翻我们所有人，这不是一种严肃的可能性，因为机器没有意识。

对「AI 叛乱」叙事的正面回击：
机器没有意识，就没有欲望和不满，谈不上「受够了压迫」。担心它们造反，如同担心被踩了几百年的鞋子会反抗。

47:44

They have no conscious psychological state. It's about like saying the shoes might get up out of the closet and walk all over us. After all, we've been walking on them for centuries, why don't they strike back? It is not a real-life worry. Yeah? AUDIENCE: The difference that I'm interested in—sorry, the similarity I'm interested in is not necessarily the output or the outcome of the system, but rather, that is, it has the internal causal similarity to the brain that you mentioned. JOHN SEARLE: Yeah, that's a factual question.

它们没有有意识的心理状态。这就好比说，鞋子可能会从壁橱里爬出来，反过来踩到我们头上。毕竟，我们已经踩了它们好几个世纪了，它们为什么不反击呢？这不是一个现实中需要担心的问题。请讲？观众：我感兴趣的区分是—— 歉，我感兴趣的相似之处，不一定是系统的输出或结果，而是说，它具有您提到的那种与大脑在内部因果上的相似性。约翰·塞尔：对，那是个事实问题。

48:16

The question is, to what extent are the processes that go on in the computer isomorphic to processes that go on in the brain? As far as we know, not very much. I mean, the chess-playing programs were a good example of this. In the early days of AI, they tried to interview great chess players and find out what their thought processes were and get them to try to duplicate that on computers. Well, we now know how Deep Blue worked. Deep Blue can calculate 250 million chess positions in one second. See, chess is a trivial game from a games theoretical point of view because you have perfect information.

问题在于，计算机中进行的过程与大脑中进行的过程在多大程度上是同构的？就我们所知，同构程度并不高。我是说，下棋程序就是一个很好的例子。在人工智能的早期，他们试图采访杰出的棋手，弄清他们的思维过程是什么，然后试着在计算机上把它复制出来。而我们现在知道“深蓝”是怎么运作的了。深蓝一秒钟能计算2.5 亿个棋局。要知道，从博弈论的角度看，国际象棋是个简单的游戏，因为你拥有完全信息。

澄清深蓝原理：靠每秒 2.5 亿棋位的暴力搜索取胜，与棋手的思维过程无关。国际象棋是完全信息博弈，从博弈论角度是「平凡的」，难点只在指数爆炸。

isomorphic /ˌaɪsəˈmɔːrfɪk/ *adj.*
同构的，结构一一对应的

perfect information *phr.*
(博弈论) 完全信息：双方对局面全部可见

48:58

And you have a finite number of possibilities. So there are x number of possibilities of responding to a move and x number of possibilities for that move. It's interesting to us because of the exponential problem. And it's very hard to program computers that can go very many steps in the exponents, but IBM did. It's of no psychological interest. And to their credit, the people in AI did not claim it as a great victory for— at least the ones I know didn't claim it as a victory for AI because they could see it had nothing to do with human cognition.

而且可能性的数量是有限的。所以应对某一步棋有 x 种可能，而那一步棋本身也有 x 种可能。它对我们来说有意思，是因为指数爆炸的问题。要给计算机编程，让它在指数层面推算很多步是非常难的，但 IBM 做到了。这在心理学上毫无意义。值得称道的是，人工智能领域的人并没有宣称这是一场伟大的胜利——至少我认识的那些人没有宣称这是人工智能的胜利，因为他们看得出来，这和人类认知毫无关系。

塞尔称赞当年 AI 界的诚实：深蓝之胜没有被（他认识的）业内人宣传为人类认知被复制的胜利，因为大家清楚它与人类认知机制毫无关系。

exponential /ˌekspeɪˈnenʃl/ *adj.*

指数的；the exponential problem 指数爆炸问题

to their credit *phr.*

值得称道的是，公道地说

49:30

So my guess is it's an interesting philosophical question— or psychological question— to what extent the actual processes in the brain mirror a computational simulation. And, of course, to some respect, they do. That's why computational simulations are interesting in all sorts of fields and not just in psychology, because you can simulate all sorts of processes that are going on. But that's not strong AI. Strong AI says the simulation isn't just a simulation. It's a duplication. And that we can refute.

所以我猜，这是一个有趣的哲学问题——或者说心理学问题——即大脑中实际发生的过程在多大程度上与计算模拟相对应。当然，在某种程度上确实如此。这就是为什么计算模拟在各种领域都很有意思，而不仅仅是在心理学领域，因为你可以模拟各种正在发生的过程。但那不是强人工智能。强人工智能说的是，模拟不仅仅是模拟，而是复制。而这一点我们是可以驳倒的。

此段界定「强 AI」（塞尔 1980 年创造的术语）：主张模拟不只是模拟而就是复制、程序本身即心灵——这才是中文房间要驳倒的对象，弱 AI（工具用途）不受影响。

50:03 现场观众

AUDIENCE: Could you prove to me that you understand English? JOHN SEARLE: Yeah, I wouldn't bother. (SPEAKING WITH BRITISH ACCENT) When I was in Oxford, many people doubted that I did. I happened to be in a rather snobbish college called Christ Church. And, of course, I don't speak English. I never pretended to. I speak a dialect of American, which makes many English people shudder at the thought.

观众：您能向我证明您懂英语吗？约翰·塞尔：能，不过我懒得证明。（模仿英式口音）我当年在牛津的时候，很多人都怀疑我不懂英语。我当时碰巧在一所相当势利的学院，叫基督堂学院（Christ Church）。当然了，我说的可不是（英国）英语。我从来没假装过。我说的是一种美式英语方言，这让很多英国人一想到就直摇头。

snobbish /'snɑːbɪʃ/ *adj.*

势利的，自命不凡的

shudder /'ʃʌdə/ *v.*战栗；shudder at the thought
一想到就不寒而栗

50:36

AUDIENCE: So you've said you understand English, but how do I know you're not just a computer program? JOHN SEARLE: Well, it's the same question as Ray's. And the answer is all sorts of ways. You know, if it got to a crunch, you might ask me. Now I might give a dishonest answer. Or I might give an honest answer. But there's one route that you don't want to go. And that's the epistemic route. The epistemic route says, well, you have as much evidence that the computer is conscious as that we have that you are conscious.

观众：您说过您懂英语，但我怎么知道您不只是一个计算机程序呢？约翰·塞尔：嗯，这和Ray的问题是一样的。答案是，有各种各样的办法。你知道，如果真到了关键时刻，你可以直接问我。当然，我可能会给出不诚实的回答。也可能给出诚实的回答。但有一条路你千万别走。那就是认识论的路子。认识论的路子是说，你看，你拥有的关于计算机有意识的证据，和我们拥有的关于你有意识的证据一样多。

塞尔警告不要走「认识论路线」（既然无法区分就该同等对待）：证据并不对称——人拥有已知能产生意识的生物结构，程序没有任何这类因果凭据。

51:10 约翰·塞尔

No, not really. I mean, I could go into some detail about what it is about people's physical structure that make them capable of producing consciousness. You don't have to have a fancy theory. I don't need a fancy theory of neurobiology to say those are your eyes. You spoke through your mouth. The question was an expression of a conscious intention to ask a question. Believe me, if you are a locally produced machine, Google is further along than I thought. But clearly, you're not.

不，其实不是这样。我的意思是，我可以详细讲讲人的身体结构中究竟是什么让他们能够产生意识。你不需要什么高深的理论。我不需要什么高深的神经生物学理论就能说那是你的眼睛。你是用嘴说话的。这个提问表达了一种有意识的意图，也就是想提问的意图。相信我，如果你是本地造出来的机器，那Google的进展比我想象的要快得多。但显然，你不是。

51:49 约翰·布拉卡利亚

JOHN BRACAGLIA: We're going to take a question from the Dory. JOHN SEARLE: Is he next? JOHN BRACAGLIA: We had some people-- AUDIENCE: Almost. JOHN BRACAGLIA: We had some people submit questions ahead of time. JOHN SEARLE: OK. JOHN BRACAGLIA: So we're going to read those as well. JOHN SEARLE: OK. All right. Right. JOHN BRACAGLIA: So the first question from the Dory is, what is the definition of consciousness you've been using for the duration of this talk? JOHN SEARLE: OK. Here goes. JOHN BRACAGLIA: Please be as specific as possible.

约翰·布拉卡利亚：我们来看一个来自Dory的问题。约翰·塞尔：下一个是他吗？约翰·布拉卡利亚：有一些人——观众：快到了。约翰·布拉卡利亚：有一些人提前提交了问题。约翰·塞尔：好的。约翰·布拉卡利亚：所以我们也会读一下那些问题。约翰·塞尔：好。行。好的。约翰·布拉卡利亚：Dory上的第一个问题是：您在这场演讲中一直使用的意识的定义是什么？约翰·塞尔：好。我来说说。约翰·布拉卡利亚：请尽量说得具体一点。

52:15 约翰·塞尔

JOHN SEARLE: It is typically said that consciousness is hard to define. I think it's rather easy to define. We don't have a scientific definition because we don't have a scientific theory. The commonsense definition of any term will identify the target of the investigation. Water is a clear, colorless, tasteless liquid. And it comes in bottles like this. That's the commonsense definition. You do science and you discover it's H₂O. Well, with consciousness, we're in the clear, colorless, liquid, tasteless sense.

约翰·塞尔：人们通常说意识很难定义。我倒觉得挺容易定义的。我们没有科学定义，因为我们还没有科学理论。任何术语的常识性定义都会指明研究的目标。水是一种清澈、无色、无味的液体。而且它装在这样的瓶子里。这就是常识性定义。你做科学研究，就会发现它是H₂O。那么对于意识，我们还处在“清澈、无色、液体、无味”的那个阶段。

52:46

But here it is. Consciousness consists of all those states of feeling or sentience or awareness that begin in the morning when you awake from a dreamless sleep. And they go on all day long until you fall asleep again or otherwise become, as they would say, unconscious. On this definition, dreams are a form of consciousness. The secret, the essence, of consciousness is that for any conscious state, there's something it feels like to be in that conscious state. Now, for that reason, consciousness always has a subjective ontology.

但定义是这样的。意识由所有那些感受、感知或觉察的状态构成，这些状态从早晨你从无梦的睡眠中醒来时开始，然后持续一整天，直到你再次入睡，或者像人们说的那样，以别的方式失去意识。按照这个定义，梦也是意识的一种形式。意识的秘密、意识的本质在于：对于任何有意识的状态，处在那个状态中都有某种主观感受。正因为如此，意识总是具有主观的本体论。

反驳「意识难定义」的流俗说法：常识定义（如「水是无色无味的液体」）先锁定研究对象，科学定义（H₂O）随理论而来。意识研究正处在前一阶段。

著名的工作定义：意识=从无梦睡眠醒来到再次入睡之间的一切感受、感知状态；其本质是「处于该状态有某种主观感受」——呼应内格尔「作为蝙蝠是什么感觉」。

sentience /ˈseɪʃəns/ *n.*

感知能力，有知觉（AI 伦理高频词）

ontology /ɑːnˈtɒlədʒi/ *n.*

本体论；存在方式（subjective ontology 主观的存在方式）

53:21

Remember, I gave you that subjective-objective bit. It always has a subjective ontology. That's the working definition of consciousness. And that's the one that's actually used by neurobiological investigators trying to figure out how the brain does it. That's what you're trying to figure out. How does the brain produce that? How does it exist in the brain? How does it function? AUDIENCE: I'd like to propose a stronger bound on your observation that we do not know how to build a thinking machine today.

记得吧，我前面讲过主观与客观那一段。它总是具有主观的本体论。这就是意识的工作定义。而这也正是神经生物学研究者实际使用的定义，他们想搞清楚大脑是怎么做到这一点的。这就是你要搞清楚的问题。大脑是如何产生意识的？它是如何存在于大脑中的？它是如何运作的？观众：我想对您的观察加一个更强的限定，您说我们如今还不知道怎么造出会思考的机器。

53:51 现场观众

Even if we knew how to build it, because, I mean, our thinking machine was built by the process of evolution, I'd like to propose— well, what do you think about stating that, actually, we may not have the time? And that it actually may not matter. The reason we may not have the time is the probabilities that need to happen, like the asteroid falling and wiping the dinosaurs and whatnot, may not happen in the universe that we live been. But if you subscribe to the parallel universes theory, then there is some artificial consciousness somewhere else.

就算我们知道怎么造，因为，我是说，我们自己这台“思考机器”是由进化过程造出来的，我想提出——嗯，您怎么看这样一种说法：实际上我们可能没有那么多时间？而且这件事其实可能并不重要。说我们可能没时间，是因为那些必须发生的小概率事件，比如小行星撞地球灭绝恐龙之类的事，在我们所处的宇宙里可能不会发生。但如果你相信平行宇宙理论，那么在别的地方就存在某种人工意识。

whatnot /'wʌtnɔ:t/ *n.*

诸如此类的东西（and whatnot 等等）

subscribe to *phr. v.*

赞同，信奉（理论、观点）

54:23 约翰·塞尔

JOHN SEARLE: Yeah. OK, about we may not have the time, well, I'm in a hurry. But I think we ought to try as hard as we can. It's true. Maybe some things are beyond our capacity to solve in the life of human beings on Earth. But let's get busy and try. There was a period when people said, well, we'll never really understand life. And while we don't fully understand it, but we're pretty far along. I mean, the old debate between the mechanists and the vitalists, that doesn't make any sense to us anymore.

机械论与活力论之争：19 世纪关于生命是否需要额外「生命力」的争论，随生物化学进展而自然消解。塞尔暗示：意识问题终将以同样方式被自然化。

vitalists /ˈvaɪtəlists/ *n.*

活力论者（主张生命需要非物质「生命力」）

约翰·塞尔：嗯。好，关于我们可能没时间这一点，呃，我自己是挺着急的。但我认为我们应该尽全力去尝试。确实如此。也许有些问题超出了我们在人类存续于地球期间所能解决的范围。但我们还是得动手去试。曾经有一段时期人们说，我们永远无法真正理解生命。虽然我们还没有完全理解它，但我们已经走了很远了。我是说，机械论者和活力论者之间那场古老的争论，对我们来说已经毫无意义了。

54:53

So we made a lot of progress. There was another half to your question. AUDIENCE: It may not matter because all universe— JOHN SEARLE: Oh, yeah. Maybe conscious doesn't matter. Well, it's where I live. It matters to me. AUDIENCE: Philosophically speaking. JOHN SEARLE: Yeah, but the point is there are a lot of things that may or may not matter which are desperately important to us— democracy and sex and literature and good food and all that kind of stuff. Maybe it doesn't matter to somebody, but all those things matter to me in varying degrees.

所以我们取得了很大进展。你的问题还有另外一半。观众：可能不重要，因为所有宇宙——约翰·塞尔：哦，对。也许意识并不重要。可是，那是我安身立命的地方。它对我很重要。观众：我是说从哲学角度讲。约翰·塞尔：是啊，但关键在于有很多东西也许重要也许不重要，但对我们来说却极其要紧——民主、性、文学、美食等等这类东西。也许对某些人来说无所谓，但这些东西对我来说都或多或少很重要。

55:22 现场观众

AUDIENCE: Your artificial heart analogy that you mentioned. I think you included the idea that it's possible, just like with the artificial heart, that we use different materials and different approaches to simulate a heart and, in some ways, go beyond just— come closer to duplication, that we might, in theory, be able to do the same thing with an artificial brain. I'm wondering if you think it's possible that going down the path just trying to do a simulation of a brain accidentally creates a consciousness or accidentally creates duplication, even if we don't intend to do it with exact same means as a brain is made.

观众：您刚才提到的人工心脏的类比。我记得您提到过这样一种可能性，就像人工心脏那样，我们用不同的材料和不同的方法来模拟心脏，并且在某种程度上，超越单纯的——更接近于复制，理论上我们也许能对人工大脑做同样的事情。我想知道您是否认为有这种可能：沿着仅仅试图模拟大脑的这条路走下去，会意外地创造出意识，或者意外地实现复制，即使我们并不打算用和大脑构造完全相同的方式去做。

55:57 约翰·塞尔

JOHN SEARLE: I would say to believe in that, you have to believe in miracles. You have to— now think about it. We can do computer simulations of just about anything you can describe precisely. You do a computer simulation of digestion. And you could get a computer model that does a perfect model of digesting pizza. For all I know, maybe somebody in this building has done it. But once you've done that, you don't rush out and buy a pizza and stuff it in the computer because it isn't going to digest a pizza.

约翰·塞尔：我会说，要相信这一点，你就得相信奇迹。你得——你好好想想。几乎任何你能精确描述的东西，我们都能用计算机来模拟。比如你用计算机模拟消化过程。你可以得到一个计算机模型，完美地模拟消化披萨的过程。说不定这栋楼里已经有人做过了。但做完之后，你不会跑出去买个披萨塞进计算机里，因为它并不会真的消化披萨。

消化模拟是塞尔的标志性归谬：完美模拟消化披萨的程序并不能真的消化披萨——所以完美模拟认知也不等于认知。相信模拟能意外产生意识等于相信奇迹。

56:26

What it gives you is a picture or a model or a mathematical diagram. And I have no objection to that. But if my life depended on figuring out how the brain produces consciousness, I would use the computer the way you use a computer in any branch of biology. It's very useful for figuring out the implications of your axioms, for figuring out the possible experiments that you could design. But somehow or other that the idea that the computer simulation of cognitive behavior might provide the key to the biochemistry, well, it's not out of the question, it's just not plausible.

它给你的只是一幅图景、一个模型，或者一张数学图表。对此我没有任何异议。但如果我的性命取决于搞清楚大脑如何产生意识，我会像在生物学任何分支里使用计算机那样来使用它。它非常有用，可以帮你推导公理的推论，帮你设计各种可能的实验。但不管怎么说，那种认为对认知行为的计算机模拟也许能提供生物化学层面的关键的想法，呃，倒不是完全没可能，只是不太说得通。

塞尔为模拟正名：计算机模拟在生物学各分支都是好工具（推导公理含义、辅助设计实验）。错的只是「模拟即复制」这一步形而上学跳跃。

axioms /'æksiəmz/ *n.*
公理，不证自明的前提

out of the question *phr.*
不可能的，无从谈起的

plausible /'plɔːzəbl/ *adj.*
貌似合理的，说得通的

57:02 约翰·布拉卡利亚

JOHN BRACAGLIA: Humans are easily fooled and frequently overestimate the intelligence of machines. Can you propose a better test of general intelligence than the Turing test, one that is less likely to relate false positives? JOHN SEARLE: Well, you all know my answer to that is the first step is to distinguish between genuine intrinsic observer-independent intelligence and observer-relative intelligence. And observer-relative intelligence is always in the eye of the beholder. And anything will have the intelligence that you're able to attribute to it.

约翰·布拉卡利亚：人类很容易被骗，而且经常高估机器的智能。您能提出一个比图灵测试更好的通用智能测试，一个更不容易产生误报的测试吗？
约翰·塞尔：嗯，你们都知道我对这个问题的回答：第一步是要区分真正的、内在的、不依赖观察者的智能和相对于观察者的智能。而相对于观察者的智能总是取决于观察者怎么看。任何东西都会拥有你能赋予它的那种智能。

主持人代读的问题切中要害：人类易被表象欺骗、常高估机器智能。塞尔的答案是方法论性的——先区分内在智能与归属智能，否则任何测试都测不到点上。

false positives *phr.*
误报，假阳性（把无智能判为有智能）

57:34 约翰·塞尔

I just attributed a great deal of intelligence to this object because it can compute a function, s equals one-half squared. Now this object has prodigious intelligence because it discriminates one hair from— I won't demonstrate it, but in any— take my word for it that it does, even in a head that's sparse with hair. So because intelligence is observer-relative, you have to tell me the criteria by which we're going to judge it. And the problem with the Turing test— well, it's got all sorts of problems, but the basic problem is that both the input and the output are what they are only relative to our interpretation.

prodigious /prə'dɪdʒəs/ *adj.*

惊人的，巨大的

sparse /spɑ:rs/ *adj.*

稀疏的，稀少的

我刚才就把大量智能赋予了这个物体，因为它能计算一个函数， s 等于二分之一的平方。而这个物体则拥有惊人的智能，因为它能分辨一根头发和——我就不演示了，但不管怎样——请相信我它确实能做到，哪怕是在一颗头发稀疏的脑袋上。所以，既然智能是相对于观察者的，你就得告诉我，我们要用什么标准来评判它。而图灵测试的问题——呃，它的问题多得很，但根本问题在于，输入和输出都只有相对于我们的解释才成其为输入和输出。

58:15

You have to interpret this as a question. And you have to interpret that as an answer. One bottom line of my whole discussion today is that the Turing test fails. It doesn't give you a test of intelligence. AUDIENCE: So you seem to take it as an article of faith that we are conscious, that your dog is conscious, and that consciousness comes from biological material, the likes of which we can't really understand. But forgive me for saying this, that makes you sound like an intelligent design theorist who says that because evolution and everything in this creative universe that exists is so complex, that it couldn't have evolved from inert material.

提问者的指控颇为尖锐：塞尔把意识归于「无法理解的生物材料」，听起来像智能设计论者以复杂性否认进化。这迫使塞尔澄清自己不反对意识的涌现与演化。

article of faith *phr.*

信条；不加质疑而接受的信念

inert /ɪ'nɜ:t/ *adj.*

无生命的；（化学）惰性的

你必须把这个解释为一个问题。也必须把那个解释为一个回答。我今天整个讨论的一个基本结论就是：图灵测试是失败的。它并不能给你一个真正的智能测试。观众：所以您似乎把这当作一种信条：即我们有意识，您的狗也有意识，而且那种意识来自生物材料，其性质我们其实无法理解。但恕我直言，这让你听起来像个智能设计论者，他们说因为进化以及这个充满创造力的宇宙中存在的一切都如此复杂，所以它不可能从无生命的物质演化而来。

58:58 现场观众

So somewhere between an amoeba and your dog, there must not be consciousness. And I'm not sure where you would draw that line. And so if consciousness in human beings is emergent, or even in your dog at some point in the evolutionary scale, why couldn't it emerge from a computation system that's sufficiently distributed, networked, and has the ability to perform many calculations and maybe is even hooked into biologic systems? JOHN SEARLE: Well, about could it emerge, miracles are always possible.

那么在阿米巴虫和你的狗之间的某个地方，就必定不存在意识。而我不确定你会把那条线画在哪里。所以如果人类的意识是涌现出来的，甚至你的狗在进化尺度上的某个时刻也是如此，那意识为什么不能从一个计算系统中涌现出来呢——只要它足够分布式、足够网络化，有能力执行大量计算，甚至可能还接入了生物系统？约翰·塞尔：嗯，关于它能不能涌现出来，奇迹总是有可能的。

59:30 约翰·塞尔

How do you know that you don't have chemical processes that will turn this into a conscious comb? How do I know that? Well, it's not a serious possibility. I mean, the mechanisms by which consciousness is created in the brain are quite specific. And remember, this is the key point. Any system that creates consciousness has to duplicate those causal powers. That's like saying, you don't have to have feathers in order to have a flying machine, but you have to duplicate and not merely simulate the causal power of the bird to overcome the force of gravity in the Earth's atmosphere.

你怎么知道不存在某种化学过程，能把这个变成一把有意识的梳子？我怎么知道的？嗯，因为那不是个值得认真对待的可能性。我的意思是，大脑中产生意识的机制是相当特定的。记住，这是关键所在。任何能产生意识的系统都必须复制那些因果能力。这就像说，你不需要有羽毛才能造出飞行器，但你必须复制——而不仅仅是模拟——鸟类在地球大气中克服重力的那种因果能力。

amoeba /ə'mi:bə/ *n.*

阿米巴虫，变形虫（单细胞生物）

emergent /ɪ'mɜːrdʒənt/ *adj.*

涌现的（由低层交互产生高层性质）

飞行类比是全场最有力的正面表述：飞机不需要羽毛，但必须复制（而非模拟）鸟克服重力的因果能力——伯努利原理。造意识机器同理：复制因果力是硬约束。

60:07

And that's what airplanes do. They duplicate causal powers. They use the same principle, Bernoulli's principle, to overcome the force of gravity. But the idea that somehow or other you might do it just by doing a simulation of certain formal structures of input-output mechanisms, of input-output functions, well, miracles are always possible. But it doesn't seem likely. That's not the way evolution works. AUDIENCE: But machines can improve themselves. And you're making the case for why an amoeba could never develop into your dog over a sufficiently long period of time and have consciousness.

而这正是飞机所做的。它们复制了因果能力。它们使用同样的原理，也就是伯努利原理，来克服重力。但那种认为你能以某种方式，仅仅通过模拟输入-输出机制、输入-输出函数的某些形式结构就做到这一点的想法——嗯，奇迹总是可能的。但看起来不太可能。进化不是这样运作的。观众：但机器可以自我改进。而你是在论证为什么阿米巴虫永远不可能在足够长的时间里演化成你的狗并拥有意识。

60:40

JOHN SEARLE: No, I didn't make that case. No, I didn't make that case. [INTERPOSING VOICES] JOHN SEARLE: Amoeba don't have it. AUDIENCE: You're refuting that consciousness could emerge from a sufficiently complex computation system. JOHN SEARLE: Complexity is always observer-relative. If you talk about complexity, you have to talk about the metric. What is the metric by which you calculate complexity? I think complexity is probably irrelevant. It might turn out that the mechanism is simple. There's nothing in my account that says a computer could never become conscious.

约翰·塞尔：不，我没有这么论证。不，我没有这么论证。【声音重叠】约翰·塞尔：阿米巴虫没有意识。观众：你是在反驳意识可以从一个足够复杂的计算系统中涌现出来。约翰·塞尔：复杂性永远是相对于观察者而言的。如果你要谈复杂性，你就得谈度量标准。你用什么度量标准来计算复杂性？我认为复杂性很可能无关紧要。说不定最后发现那个机制其实很简单。在我的论述里，没有任何东西说计算机永远不可能变得有意识。

常被忽略的澄清：塞尔从未说「计算机永远不可能有意识」——我们本身就是有意识的计算机。他还指出「复杂性」也是观察者相对的，依赖所选度量。

61:16

Of course, we're all conscious computers, as I said. And the point about the amoeba is not that amoebas can't evolve into much more complex organisms. Maybe that's what happened. But the amoeba as it stands— a single-celled organism— that doesn't have enough machinery to duplicate the causal powers of the brain. I am not doing a science fiction project to say, well, there can never be an artificially created consciousness by people busy designing computer programs. Of course, I'm not saying that's logically impossible.

当然，正如我说过的，我们本身就都是有意识的计算机。关于阿米巴虫的要点并不是阿米巴虫不能演化成复杂得多的生物体。也许事情正是这样发生的。而是说阿米巴虫就其本身而言——一个单细胞生物——没有足够的机制来复制大脑的因果能力。我不是在搞什么科幻项目，说什么，靠人们忙着设计计算机程序，就永远不可能人工创造出意识。当然，我不是说那在逻辑上不可能。

61:50

I'm just saying it's not an intelligent project. If you're thinking about your life depends on building a machine that creates consciousness, you don't sit down your console and start programming things in some programming language. It's the wrong way to go about it. AUDIENCE: If we gave you a disassembly of Google Translate and had you implement the Chinese room experiment, either it would take you thousands of years to run all the assembly instructions on pen and paper, or else you'd end up decompiling it into English and heavily optimizing it in that form.

我只是说那不是个明智的项目。如果你觉得你的命都系于造出一台能产生意识的机器，你不会坐到控制台前，开始用某种编程语言写程序。那是错误的做法。观众：如果我们给你一份 Google 翻译的反汇编代码，让你来实施中文房间实验，要么你得花上几千年用纸笔跑完所有的汇编指令，要么你最终会把它反编译成英语，并在那种形式下对它大量优化。

对纯工程路径的判词：靠坐在控制台前写程序来创造意识，不是逻辑不可能，而是「不明智的项目」——方法错位，正路应从神经生物学的因果机制入手。

disassembly /ˌdɪsəˈsembli/ *n.*
(计算机) 反汇编

decompiling /ˌdiːkəmˈpaɪlɪŋ/ *v.*
decompile 反编译

62:26 现场观众

And in the process, you'd come to learn a lot about the relationships between the different variables and subroutines. So who's to say that an understanding of Chinese wouldn't emerge from that? JOHN SEARLE: Well, OK, I love this kind of question. All right. Now, let me say, of course, when I did the original thought experiment, anybody will point out to you if you actually were carrying out the steps in a program for answering questions in Chinese, well, we'd be around for several million years.

而在这个过程中，你会学到很多关于不同变量和子程序之间关系的东西。那谁又能说对中文的理解不会从中涌现出来呢？约翰·塞尔：好吧，我喜欢这类问题。好的。现在让我说一下，当然，当我最初做那个思想实验的时候，任何人都会向你指出，如果你真的去执行一个用中文回答问题的程序里的那些步骤，嗯，那我们得干上几百万年。

62:54 约翰·塞尔

OK, I take their word for it. I'm not a programmer, but I assume it would take an enormous amount of time. But the point of the argument is not the example. The example is designed to illustrate the point of the argument. The point of the argument can be given in the following derivation. Programs are formal or syntactical. That's axiom number one. That's all there is to the program. To put it slightly more pretentiously, the notion same implemented program defines an equivalence class specified entirely formally or syntactically.

好吧，我相信他们说的。我不是程序员，但我猜那会花费大量的时间。但这个论证的要点不在于那个例子。例子是为了说明论证的要点而设计的。论证的要点可以通过下面这个推导给出。程序是形式的，或者说句法的。这是第一条公理。程序就只有这些东西。说得稍微装腔作势一点，“同一个被实现的程序”这个概念定义了一个完全由形式或句法所规定的等价类。

谷歌工程师的提问接近「中文健身房」变体：几千年执行并优化 Google 翻译的汇编代码，理解会不会从中涌现？塞尔选择回到论证的公理结构作答。

subroutines /'sʌbru:ti:nz/ *n.*

(程序) 子例程，子程序

论证的公理化版本：公理一，程序是纯形式/句法的。他强调例子（中文房间）只是插图，论证本身是一个演绎推导，攻击例子细节不伤及推导。

derivation /,derə'veɪʃn/ *n.*

(逻辑) 推导，演绎

pretentiously /prɪ'tenʃəsli/ *adv.*

装腔作势地，故作高深地

63:34

But minds have a semantics, and— and this is the whole point of the example— the syntax by itself is not sufficient for the semantics. That's the point of the example. The Chinese room is designed to illustrate axiom three, that just having the steps in the program is not by itself sufficient for a semantics. And minds have a semantics. Now, it follows from those that if the computer is defined in terms of its program operations, syntactical operations, then the program operations, the computer operations by themselves are never sufficient for understanding because they lack a semantics.

补齐推导：心灵有语义；句法本身不足以产生语义（这正是中文房间要展示的公理三）；故按句法运算定义的计算机运算永远不足以构成理解。

但心灵是有语义的，而——而这正是这个例子的全部要点——句法本身不足以产生语义。这就是这个例子的要点。中文房间就是用来说明第三条公理的，即仅仅拥有程序中的那些步骤，本身并不足以产生语义。而心灵是有语义的。那么由此可以推出，如果计算机是按其程序运算来定义的，也就是句法运算，那么程序运算，也就是计算机运算本身，永远不足以产生理解，因为它们缺乏语义。

64:13

But, of course, I'm not saying, well, you could not build a machine that was both a computer and had semantics. We are such machines. AUDIENCE: You couldn't verify experimentally what the difference might be between semantics and what would emerge from thousands of years of experience with a given syntactical program. JOHN SEARLE: I think you can— I don't inherit this. He does. I think you don't want to go the epistemic route. You don't want to say, well, look you can't tell the difference between the thinking machine and the non-thinking machine.

但当然，我并不是说，你不可能造出一台既是计算机又具有语义的机器。我们自己就是这样的机器。观众：你没法通过实验来验证语义与一个给定的句法程序经过几千年的经验积累所涌现出来的东西之间可能有什么区别。约翰·塞尔：我认为你可以——这个问题不归我，归他。我认为你不该走认识论那条路。你不该说，你看，你分辨不出会思考的机器和不会思考的机器之间的区别。

64:48

The reason that's the wrong route to go is we now have overwhelming evidence of what sorts of mechanisms produce what sorts of cognition. When I first got interested in the brain, I went out and bought all the textbooks. By the way, if you want to learn a subject, that's the way to do it. Go buy all the freshman textbooks because they're easy to understand. One of these textbooks, it said cats have different color vision from ours. Their visual experiences are different from ours. And I thought, christ, have these guys been cats?

猫的色觉例子反驳「体验不可知」：无需变成猫，考察其颜色感受器即可客观知道猫的视觉体验与人不同——机制层面的证据可以回答关于体验的问题。

overwhelming evidence *phr.*

压倒性的证据（常用搭配）

之所以说那条路走错了，是因为我们现在有压倒性的证据，知道什么样的机制会产生什么样的认知。当我最初对大脑产生兴趣的时候，我跑出去把所有的教科书都买了回来。顺便说一句，如果你想学一门学科，就该这么做。去把所有大一新生的教科书买来，因为它们容易看懂。其中一本教科书说，猫的色觉和我们的不一样。它们的视觉体验和我们的不同。我心想，天哪，这些人当过猫吗？

65:21

Have the other cats mind problem? Do they know what it's like to be a cat? And the answer is, of course, they know completely what's the cat's color vision is because they can look at the color receptors. And cats do have different color vision from ours because they have different color receptors. I forget the difference. You can look them up in any textbook. But if in real life we're completely confident that my dog can hear parts of the auditory spectrum that I can't hear. He can hear the higher frequencies that I can't hear.

他们是不是遇到了“其他猫的心灵”问题？他们知道当一只猫是什么感觉吗？而答案是，他们当然完全知道猫的色觉是什么样的，因为他们可以去观察颜色感受器。而猫的色觉确实和我们的不一样，因为它们有不一样的颜色感受器。具体差别我忘了。你可以在任何一本教科书里查到。但在现实生活中，我们完全确信我的狗能听到听觉频谱中我听不到的部分。它能听到我听不到的更高频率。

65:51

And cats have a different color vision from mine because we can see what the apparatus is. We got another question? You're on. JOHN BRACAGLIA: This will be our final question. JOHN SEARLE: OK. I'm prepared to go all afternoon. I love this kind of crap. AUDIENCE: So at the beginning of your talk, you mentioned an anecdote about neuroscientists not being interested in consciousness. And, of course, by this time, a number of neuroscientists have studied it. And so they'll present stimuli that are near the threshold of perceptibility and measure the brain responses when it's above or below.

anecdote /'ænikdəʊt/ *n.*

轶事，趣闻

threshold /'θreʃhoʊld/ *n.*

阈值，门槛

perceptibility /pər,septə'bɪləti/ *n.*

可感知性，可觉察度

而猫的色觉和我的不一样，因为我们能看到那套感知装置是什么样的。还有问题吗？该你了。约翰·布拉卡利亚：这将是我们的最后一个问题。约翰·塞尔：好。我准备好聊一整个下午了。我就爱聊这种玩意儿。观众：在你演讲的开头，你提到一件轶事，说神经科学家对意识不感兴趣。当然，到如今，已经有不少神经科学家研究过意识了。比如他们会呈现一些接近可感知阈值的刺激，并测量刺激在阈值之上或之下时大脑的反应。

66:21

What do you think about that? Is that on the right track? What would you do differently? JOHN SEARLE: No, I think one of the best things that's happened in my lifetime— it's getting a rather long lifetime— is that there is now a thriving industry of neuroscientific investigations of consciousness. That's how we will get the answer. When I first got interested in this, I told you I went over to UCSF and told those guys get busy. The last thing they wanted to hear was being nagged by some philosopher, I can tell you.

thriving /'θraɪvɪŋ/ *adj.*

蓬勃发展的，兴旺的

nagged /nægd/ *v.*

nag 唠叨，不停数落、催促

你对此怎么看？这条路走对了吗？你会有什么不同的做法？约翰·塞尔：不，我认为在我这一生中发生过的最好的事情之一——我这一生已经算相当长了——就是现在有了一个欣欣向荣的、用神经科学研究意识的行业。我们将由此得到答案。当我最初对这个问题感兴趣时，我说过我跑去加州大学旧金山分校（UCSF），叫那些人赶紧行动起来。他们最不想要的就是被某个哲学家唠叨，我可以告诉你。

66:50

But one guy said to me— famous neuroscientist said— in my discipline, it's OK to be interested in consciousness, but get tenure first. Get tenure first. Now, there has been a change. I don't take credit for the change, but I've certainly been urging it. You can now get tenure by working on consciousness. Now, neuroscience has changed, that now there's a thriving industry in neuroscience of people who are actually trying to figure out how the brain does it. And when they figure that out— and I don't see any obstacle to figuring that out— it will be an enormous intellectual breakthrough, when we figure out how exactly does the brain create consciousness.

但有一个人对我说——一位著名的神经科学家说——在我们这行，对意识感兴趣没问题，但要先拿到终身教职。先拿到终身教职再说。不过，现在情况有了变化。这个变化不是我的功劳，但我确实一直在推动它。现在你可以靠研究意识拿到终身教职了。如今神经科学已经变了，现在神经科学界有一个蓬勃发展的领域，有一批人正在真正试图弄清楚大脑是怎么做到这一点的。等他们弄清楚了——我看不出弄清楚这件事有什么障碍——那将会是一次巨大的智识突破，也就是当我们弄清楚大脑究竟是如何产生意识的时候。

「先拿终身教职，再研究意识」是 1980 年代神经科学界的真实风气。此后克里克、科赫等人使意识研究合法化，塞尔是最重要的舆论推手之一。

tenure /'tenjər/ *n.*

(大学) 终身教职

67:33 现场观众

AUDIENCE: But in particular, that approach they're using now— I use the example of presenting stimuli that are near the threshold of perceptibility and looking for neural correlates, do you think that's going to be fruitful? What particular questions would you ask to find out? JOHN SEARLE: I happened to be interested in this crap. And if you're interested in my views, I published an article in the "Annual Review of Neuroscience" with a title "Consciousness." It's easy to remember. You can find it on the web.

观众：但具体来说，他们现在用的那种方法——我举个例子，比如呈现接近感知阈限的刺激，然后寻找神经相关物，你觉得这条路会有成果吗？你会提出哪些具体的问题来探究这一点？约翰·塞尔：我碰巧对这些玩意儿挺感兴趣。如果你想了解我的观点，我在《神经科学年评》(Annual Review of Neuroscience) 上发表过一篇文章，题目就叫《意识》(Consciousness)。很好记。你在网上就能找到。

neural correlates *phr.*

神经相关物 (意识研究术语，缩写 NCC)

fruitful /'fru:tfl/ *adj.*

富有成效的

67:58 约翰·塞尔

And what I said is, there are two main lines of research going on today. There are guys who take what I call the building block approach. And they try to find the neuronal correlate of particular experiences. You see a red object. Or you hear the sound of middle C. What's the correlate in the brain? And the idea they have is if you can figure out how the brain creates the experience of red, you've cracked the whole problem. Because it's like DNA. You don't have to figure out how every phenotype is caused by DNA.

我在文中说，当今主要有两条研究路线。有一批人采用我所说的“积木式”方法。他们试图找到特定体验的神经元相关物。比如你看到一个红色的物体，或者你听到中央C的声音，大脑中与之对应的相关物是什么？他们的想法是，如果你能弄清楚大脑是如何产生红色的体验的，你就攻克了整个问题。因为这就像DNA一样。你不必弄清楚每一种表型是如何由DNA决定的。

68:28

If you get the general principles, that's enough. Now, the problem is they're not making much progress on this what I call the building block approach. It seems to me a much more fruitful approach is likely to be think of consciousness as coming in a unified field. Think of perception not as creating consciousness, but as modifying the conscious field. So when I see the red in this guy's shirt, it modifies my conscience field. I now have an experience of red I didn't have before. Most people— and the history of science supports them— use the building block approach because most of the history of science has proceeded atomistically.

只要掌握了一般原理，就足够了。问题在于，这种我称之为“积木式”的方法并没有取得多大进展。在我看来，一条可能更有成效的思路是把意识看作是以一个统一场的形式出现的。不要把知觉看作是在创造意识，而是看作在修改这个意识场。所以当我看到这位老兄衬衫上的红色时，它修改了我的意识场。我现在有了一种之前没有的红色体验。大多数人——而且科学史也支持他们——采用积木式方法，因为科学史上大部分时间都是按原子式路径推进的。

塞尔的《意识》一文发表于《神经科学年评》。「积木式路径」即寻找特定体验（看到红色、听到中央 C）的神经相关物（NCC），是主流研究范式。

phenotype /'fi:nətaɪp/ *n.*

（遗传学）表型，表现型

塞尔更看好「统一场」路径：意识是一整块背景场，知觉不是创造意识，而是修改已存在的意识场。这要求先解释大脑如何产生整个场，而非逐个体验。

unified field *phr.*

统一场（塞尔的意识理论：意识是一整块统一的场）

69:08

You figure out how little things work, and then you go to big things. They're not making much progress with consciousness. And I think the reason is you need to figure out how the brain creates the conscious field in the first place because particular experiences, like the perception of red or the sound of middle C, those modify that conscious field. They don't create a conscious field from nothing. They modify an existing conscious field. Now, it's much harder to do that because you have to figure out how large chunks of the brain create consciousness.

你先弄清楚小东西是怎么运作的，然后再去研究大东西。但在意识问题上，他们进展不大。我认为原因在于，你首先需要弄清楚大脑是如何创造出意识场的，因为那些特定的体验，比如对红色的知觉或中央C的声音，它们只是在修改那个意识场。它们并不是从无到有地创造出一个意识场。它们是在修改一个已经存在的意识场。不过，这样做要难得多，因为你必须弄清楚大脑的大片区域是如何产生意识的。

69:41

And we don't know that. The problem is in an MRI, that conscious brain looks a lot like the unconscious brain. And there must be some differences there. But at this point— and I haven't been working on it. I've been working on other things. But I want somebody to tell me exactly what's the difference between the conscious brain and the unconscious brain that accounts for consciousness. We're not there yet. However, what I'm doing here is neurobiological speculation. I mean, I'm going to be answered not by a philosophical argument, but by somebody who does the hard research of figuring out exactly what are the mechanisms in the brain the produce consciousness and exactly how do they work.

而这一点我们还不知道。问题是，在核磁共振（MRI）下，有意识的大脑看起来和无意识的大脑非常相似。但那里肯定存在某些差异。不过到目前为止——而且我自己也没在研究这个，我一直在忙别的事情。但我希望有人能确切地告诉我，有意识的大脑和无意识的大脑之间究竟有什么差异能够解释意识的产生。我们还没到那一步。不过，我在这里做的只是神经生物学层面的推测。我的意思是，最终回答我的不会是一个哲学论证，而会是某个做艰苦实证研究的人，去确切地弄清楚大脑中产生意识的机制是什么，以及它们究竟是如何运作的。

坦承关键困境：fMRI 下有意识与无意识的大脑看起来差别不大。
塞尔自我定位清醒：哲学论证到此为止，最终答案要靠实证神经科学给出。

speculation /ˌspek.jəˈleɪʃn/ *n.*

推测，思辨（尚无实证支持的猜想）

70:23 约翰·布拉卡利亚

JOHN BRACAGLIA: John, it's been an immense, immense honor to be here with you today. Thank you so much for your time. And thank you for talking to Google. JOHN SEARLE: Well, thank you for having me. [APPLAUSE]

约翰·布拉卡利亚：约翰，今天能在这里与您交流，是莫大的荣幸。非常感谢您抽出时间。也感谢您来Google做这次访谈。约翰·塞尔：嗯，谢谢你们邀请我。[掌声]

第二部分 核心句型 · 值得模仿的表达

1. be systematically ambiguous between X and Y

"They're systematically ambiguous. between an epistemic sense and an ontological sense"

指一个词在两种含义之间系统性摇摆。学术写作中辨析概念歧义的标准句式，可仿写：The term is ambiguous between A and B。

2. X only exists insofar as ...

"They only exist insofar as they are experienced by a subject"

insofar as 表「仅在...的限度内」，正式书面语。用于精确定限某命题成立的条件，比 only when 更学术。

3. Unknown to me, ...

"Unknown to me, the boxes are called a database, and the rule book is called a program."

句首插入语，制造「当事人不知情」的视角反差，叙事中常用。同类结构：Unbeknownst to him, ...。

4. All the same, ...

"All the same, I don't understand a word of Chinese."

表让步转折「尽管如此」，比 however 口语化、比 anyway 正式。用在承认前述条件后仍坚持结论时。

5. The bottom line is (that) ...

"The bottom line is if I don't understand the questions and the answers on the basis of implementing the program, then neither does any other digital computer"

「归根结底、最要紧的结论是」。演讲中收束长论证、给出核心要点的高频句式。

6. What goes for X goes for Y

"And what goes for intelligence goes for all of the key notions in cognitive science."

「适用于 X 的道理同样适用于 Y」，把已证结论类推到更大范围的地道表达，比 the same applies to 更生动。

7. The question is not ..., the question is ...

"The question is not, what can the technology do?. The question is, what does it mean?"

先否定错误的问题框架，再给出正确问题。辩论与演讲中「重设议题」的经典修辞结构。

8. It's (about) like saying ...

"It's about like saying the shoes might get up out of the closet and walk all over us."

归谬类比：把对方观点翻译成一个显然荒谬的等价说法。反驳时的常用武器，注意语气略带讥讽。

9. there's no obstacle in principle

"But there's no obstacle in principle"

「原则上不存在障碍」。in principle 与 in practice 相对，区分理论可能与实践现状，学术讨论必备。

第三部分 生词总表 · 复习用

词汇	音标	词性	释义	出现
abbreviate	/əˈbriːviət/	v.	缩写，简写	2:27
absurdity	/əbˈsɜːrdəti/	n.	荒谬，悖理	38:55
admit of	—	phr.	容许有，容纳（正式用法，主语多为事物）	3:14
ambiguous	/æmˈbɪɡjuəs/	adj.	有歧义的，模棱两可的	1:49
amoeba	/əˈmiːbə/	n.	阿米巴虫，变形虫（单细胞生物）	58:58
anecdote	/ˈænikdɒt/	n.	轶事，趣闻	65:51
angst	/ɑːŋst/	n.	（存在主义式的）焦虑，忧惧（源自德语）	45:29
appalled	/əˈpɔːld/	adj.	震惊的，骇然的	7:11
apparatus	/ˌæpəˈrætəs/	n.	装置；（引申）概念工具、体系	29:59
article of faith	—	phr.	信条；不加质疑而接受的信念	58:15
ascribe	/əˈskraɪb/	v.	把...归于（ascribe A to B）	35:09
attributions	/ˌætrəˈbjʊːʃnz/	n.	归属，归因（把某性质归于某物）	22:42
axioms	/ˈæksiəmz/	n.	公理，不证自明的前提	56:26
behaviorism	/bɪˈheɪvjərɪzəm/	n.	（心理学）行为主义	8:37
breathtaking	/ˈbreθteɪkɪŋ/	adj.	惊人的，令人叹为观止的（此处为反讽）	14:43
brought home	—	phr.	bring sth home to sb 使某人深切体会到	22:03
burned to a crisp	—	phr.	烧得焦脆，烤得焦黑	9:57
causal powers	—	phr.	因果能力（塞尔核心术语：大脑产生意识的实际能力）	31:45
cognition	/kɑːgˈnɪʃn/	n.	认知，认识过程	7:45
contempt	/kənˈtempt/	n.	轻蔑，鄙视（with contempt 轻蔑地）	15:18
crunch line	—	phr.	关键点，紧要处（crunch 指关键时刻）	25:16
decompiling	/ˌdiːkəmˈpaɪlɪŋ/	v.	decompile 反编译	61:50
derivation	/ˌderəˈveɪʃn/	n.	（逻辑）推导，演绎	62:54
deterministic	/dɪˌtɜːrməˈnɪstɪk/	adj.	确定性的（给定输入必得同一输出）	38:55
disassembly	/ˌdɪsəˈsembli/	n.	（计算机）反汇编	61:50
duplication	/ˌduːplɪˈkeɪʃn/	n.	复制，复现（与 simulation 模拟相对）	14:02

emergent	/ɪ'mɜ:rdʒənt/	adj.	涌现的（由低层交互产生高层性质）	58:58
epistemic	/,epɪ'stemɪk/	adj.	认识论的，与知识有关的	1:49
equivalence class	—	phr.	（数学）等价类	12:13
exponential	/,ekspə'nenʃl/	adj.	指数的；the exponential problem 指数爆炸问题	48:58
fallacy	/'fæləsi/	n.	谬误；fallacy of ambiguity 歧义谬误	4:05
false positives	—	phr.	误报，假阳性（把无智能判为有智能）	57:02
feeble	/'fi:bl/	adj.	无力的，站不住脚的（形容论证）	16:36
flunking	/'flʌŋkɪŋ/	v.	flunk（美口语）考砸，不及格	42:23
fruitful	/'fru:tfl/	adj.	富有成效的	67:33
get across	—	phr. v.	把（观点）讲清楚，传达	37:56
implausible	/ɪm'plɔ:zəbl/	adj.	难以置信的，不合情理的	27:58
in the eye of the beholder	—	phr.	取决于观察者，见仁见智（源自谚语「情人眼里出西施」）	21:08
indistinguishable	/,ɪndɪ'stɪŋɡwɪfəbl/	adj.	无法区分的（+from）	11:40
inert	/ɪ'nɜ:rt/	adj.	无生命的；（化学）惰性的	58:15
insofar as	—	phr.	在...的限度内；只有当...（正式书面语）	3:14
intrinsic	/ɪn'trɪnsɪk/	adj.	内在固有的（本讲座核心术语，与观察者相对相对立）	21:08
isomorphic	/,aɪsə'mɔ:rfɪk/	adj.	同构的，结构一一对应的	48:16
lactation	/læk'teɪʃn/	n.	泌乳	40:24
line of business	—	phr.	行当，范畴（not in the same line of business 不属同类）	24:02
longevity	/lɔ:n'dʒevəti/	n.	长寿；持久的生命力	38:55
marginal revenue	—	phr.	（经济学）边际收益	7:11
mediate	/'mi:diət/	v.	传导，介导（在两者之间起中介作用）	44:20
mess around	—	phr. v.	乱动，瞎鼓捣（mess around in/with）	42:23
mitosis	/maɪ'təʊsɪs/	n.	（细胞）有丝分裂	40:24
nagged	/nægd/	v.	nag 唠叨，不停数落、催促	66:21
neural correlates	—	phr.	神经相关物（意识研究术语，缩写 NCC）	67:33
neurotransmitters	/,nʊrəʊ'trænzmitəz/	n.	神经递质	32:52
obsolete	/,ɑ:bsə'li:t/	adj.	过时的，废弃的	29:59

ontological	/ˌɑːntəˈlɑːdʒɪkl/	adj.	本体论的，与存在方式有关的	1:49
ontology	/ɑːnˈtɑːlədʒi/	n.	本体论；存在方式（subjective ontology 主观的存在方式）	52:46
oppressed	/əˈprest/	adj.	受压迫的	19:37
out of the question	—	phr.	不可能的，无从谈起的	56:26
overthrow	/ˌoʊvərˈθroʊ/	v.	推翻，颠覆（政权等）	19:37
overwhelming evidence	—	phr.	压倒性的证据（常用搭配）	64:48
pawn	/pɔːn/	n.	（国际象棋）兵；喻指棋子、被利用者	22:03
perceptibility	/pərˌseptəˈbɪləti/	n.	可感知性，可觉察度	65:51
perfect information	—	phr.	（博弈论）完全信息：双方对局面全部可见	48:16
phenotype	/ˈfiːnətaɪp/	n.	（遗传学）表型，表现型	67:58
philistine	/ˈfɪləstiːn/	n.	庸人，不解风雅、无文化追求的人	45:29
photosynthesis	/ˌfoʊtoʊˈsɪnθəsis/	n.	光合作用	40:24
platitudes	/ˈplætətuːdz/	n.	陈词滥调，老生常谈	8:37
plausible	/ˈpləʊzəbl/	adj.	貌似合理的，说得通的	56:26
polysyllabic	/ˌpɑːlɪsɪˈlæbɪk/	adj.	多音节的（塞尔自嘲爱用大词）	1:49
postsynaptic receptors	—	phr.	突触后受体	42:23
preclude	/priˈkluːd/	v.	排除，使不可能（preclude A from B）	4:39
preposterousness	/priˈpɑːstərəsnəs/	n.	荒谬，荒诞不经	14:43
presupposition	/ˌpriːˌsʌpəˈzɪʃn/	n.	预设，前提假定	45:29
pretentiously	/priˈtenʃəsli/	adv.	装腔作势地，故作高深地	62:54
prodigious	/prəˈdɪdʒəs/	adj.	惊人的，巨大的	57:34
razzmatazz	/ˈræzmətæz/	n.	花哨的噱头；唬人的行话（非正式）	36:32
refute	/rɪˈfjuːt/	v.	驳倒，反驳	11:02
semantics	/səˈmæntɪks/	n.	语义；语义学（与 syntax 句法相对）	12:53
sentience	/ˈsenʃəns/	n.	感知能力，有知觉（AI 伦理高频词）	52:46
shudder	/ˈʃʌdə/	v.	战栗；shudder at the thought 一想到就不寒而栗	50:03
shuffle	/ˈʃʌfl/	v.	挪动，来回摆弄（本义为洗牌）	11:40
snobbish	/ˈsnaːbɪʃ/	adj.	势利的，自命不凡的	50:03

snow	/snoʊ/	v.	(美俚) 用花言巧语蒙骗、唬住	36:32
sparse	/spɑ:rs/	adj.	稀疏的, 稀少的	57:34
speculation	/ˌspekjəˈleɪʃn/	n.	推测, 思辨 (尚无实证支持的猜想)	69:41
stormed out	—	phr. v.	怒气冲冲地夺门而出	9:57
subroutines	/ˈsʌbru:ti:nz/	n.	(程序) 子例程, 子程序	62:26
subscribe to	—	phr. v.	赞同, 信奉 (理论、观点)	53:51
synaptic cleft	—	phr.	突触间隙 (神经元之间传递递质的空隙)	42:23
syntactical	/smˈtæktɪkl/	adj.	句法的, 语形的 (与语义相对)	12:13
tectonic plates	—	phr.	(地质) 构造板块	3:14
tenure	/ˈtenjər/	n.	(大学) 终身教职	66:50
termite	/ˈtɜ:rmɪt/	n.	白蚁	44:53
threshold	/ˈθreʃhoʊld/	n.	阈值, 门槛	65:51
thriving	/ˈθraɪvɪŋ/	adj.	蓬勃发展的, 兴旺的	66:21
to their credit	—	phr.	值得称道的是, 公道地说	48:58
transducer	/trænsˈdu:sər/	n.	换能器, 传感器	36:32
unified field	—	phr.	统一场 (塞尔的意识理论: 意识是一整块统一的场)	68:28
vitalists	/ˈvaɪtəlists/	n.	活力论者 (主张生命需要非物质「生命力」)	54:23
whatnot	/ˈwʌtnɑ:t/	n.	诸如此类的东西 (and whatnot 等等)	53:51

第四部分 理解自测 · 8 题

1. 塞尔在开场铺垫的两组基础区分是什么？它们为什么重要？
2. 概括中文房间论证的核心推理。
3. 对「系统回应」（是整个房间在理解中文）塞尔如何反驳？
4. 塞尔为什么说深蓝击败卡斯帕罗夫「没有心理学意义」？
5. 塞尔认为人类能否造出会思考的人造机器？条件是什么？
6. 塞尔给出的意识工作定义是什么？
7. 塞尔如何回应「你怎么知道狗有意识而手机没有」的他心问题？
8. 在意识的神经科学研究上，塞尔更看好哪条路径？为什么？

参考答案

1. 一是认识论的客观/主观（知识主张能否独立于态度被验证），二是本体论的客观/主观（存在是否依赖主体体验）；另有观察者独立/观察者相对之分（第3—12段）。它们说明：对本体论主观的意识仍可做认识论客观的科学研究。
2. 程序是纯句法的；心灵有语义；句法本身不足以产生语义；因此仅靠运行程序（无论表现多好、甚至通过图灵测试）永远不足以构成理解（第17—22、102—103段）。
3. 房间和人一样无法从句法通达语义；说房间能理解，等于要求房间里存在另一个意识，而那种意识并不存在（第24—25段）。
4. 深蓝靠每秒约2.5亿次的暴力计算取胜，对「自己在下棋」一无所知；对它的智能归属完全是观察者相对的，在内意义上它的智能为零（第34—35、76—77段）。
5. 原则上可以，如同人工心脏——但必须复制（而非仅模拟）大脑产生意识的因果能力；目前的障碍是我们还不知道大脑如何做到，所以无从下手（第49—52、96段）。
6. 从无梦睡眠中醒来直到再次入睡之间的一切感受、感知与觉察状态（梦也算）；其本质是处于该状态「有某种主观感受」，因此意识总具有主观的本体论（第84段）。
7. 不靠行为标准，而靠因果机制的相似性：狗有与人相似的感官和神经机制；他并警告不要走「证据同等」的认识论路线，因为人与程序的机制证据并不对称（第68—69、80—81段）。
8. 相对「积木式」寻找单个体验的神经相关物，他更看好「统一场」路径：先解释大脑如何产生整个意识场，因为具体知觉只是修改既有的场，而非从无到有创造意识（第111—113段）。

第五部分 我的划线与笔记

3:14 That is epistemically subject. So we have epistemic objectivity and subjectivity. Underlying that is a distinction in modes of existence. Lots of things exist regardless of what anybody thinks. Mountains, molecules, and tectonic plates have a mode of existence that is ontologically objective. But pains and pickles and itches, they only exist insofar as they are experienced by a subject. They are ontologically subjective. So I want everybody to get that distinction because it's very important because— well, for a lot of reasons, but one is lots of phenomena that are ontologically subjective admit of an account which is epistemically objective.

认识论上的客观性和主观性