

# 蓝眼睛岛问题

余周炜

日期：2019 年 11 月 15 日

该问题是华人数学家陶哲轩最感兴趣的逻辑问题。

## 1 问题描述

(蓝眼睛岛问题) 在一个岛上，住着一个部落。这个部落由 1000 个人组成，这些人有多种眼睛颜色。他们有一种宗教信仰，禁止他们了解自己眼睛的颜色，甚至禁止他们讨论眼睛颜色这个话题。因此，每个居民都能看到所有其他居民的眼睛颜色，但却不能发现自己的眼睛颜色（这里没有镜子或者诸如此类的东西）。如果有岛民知道了自己的眼睛颜色，那么他们的宗教就会强制他们第二天中午在村庄广场上自杀，让所有人都看到。所有的岛民都非常逻辑和非常虔诚，他们也都知道其他人也非常逻辑和非常虔诚。在这里，“非常逻辑”的意思是，如果从岛民已知的信息观察中可以推理出任何结论，那么岛民必定会知道这个结论。

实际上，在这 1000 个岛民中，100 个人的眼睛是蓝色，900 个人的眼睛是棕色。不过，岛民最初并不知道这些数据。因为，每个人只能看到除自己之外的 999 个人的眼睛颜色，看不到自己的眼睛颜色。

有一天，一个蓝眼睛的游客来到了岛上，并且获得了部落的完全信任。一天晚上，他向整个部落的人致辞，感谢他们的款待。

然而，由于不了解当地的风俗习惯，这位游客犯了一个错误，在他的讲话中提到了眼睛的颜色。他说了一句：“真是有趣呀，在这里能看到像我这样的蓝眼睛的人！”

请问，游客的失言对部落会不会产生影响？如果有，是什么影响？

## 附加问题

1. 如果游客在发言后的第二天就意识到了自己的错误，那么他有没有办法减少伤亡呢？
2. 如果游客意识到自己的错误不是在发言后的第二天，而是又过了几天之后，他有没有办法减少伤亡呢？

## 2 简单思考

这里用到的是递归推理。

如果只有一个蓝眼人，那么他原本看不到蓝眼人，那么他听到游客的话就知道了游客说的这个蓝眼人就是自己，因此他在一天以后就会自杀。

如果有两个蓝眼人，那么他们原本只能看到一个蓝眼人，听到游客的话后，他们就会想对方是不是岛上唯一的蓝眼人呢，如果是的话，那么一天以后他也应该自杀了，但是过了一天没人自杀，因此两个人都明白了，对方没有自杀，是因为岛上还有一个蓝眼人，这个蓝眼人是谁呢？那么就是自己，因此在游客讲话的两天以后这两个人会一起自杀。

继续推理，如果有  $n$  个蓝眼人，那么这  $n$  个蓝眼人会在  $n$  天以后一起自杀。

我们也可以按时间推理。

过一天推理一次，那过了一天没人自杀，这说明蓝眼人的人数不是 1；过了两天没人自杀，就说明蓝眼人的人数不是 2；如此，过了  $n$  天没人自杀，那么蓝眼人的人数不是  $n-1$ 。因此第  $n$  个蓝眼人就是自己，所以  $n$  个蓝眼人就在  $n$  天之后一起自杀。

## 3 核心问题

游客有没有带来新信息？ 答案是游客确实带来了新信息。

对于任何一个蓝眼人而言，他不妨把自己称为  $A_1$ ，把其他 99 位蓝眼人称为  $A_2, A_3, \dots, A_{100}$ ，现在游客给  $A_1$  带来了新信息，如下所述：

$A_1$  知道,  $A_2$  知道,  $A_3$  知道...  $A_{100}$  知道岛上有蓝眼人。

游客的话为什么能够带来新信息？ 因为他做的是一个公布，一个广播，即使他公布的内容是他早已知道的，但是公布这个动作本身就会造成一个重要的区别，就是现在所有人不但知道了公布的内容，而且知道所有人都知道，还知道所有人都知道所有人都知道... 如此等等，以至于无穷。

## 4 共识的分类

- 弱共识 (Mutual knowledge): 如果一件事是所有人都知道的，那么这件事称为 Mutual knowledge，也叫沉默共识。
- 强共识 (Common knowledge): 如果一件事是所有人都知道，所有人都知道所有人都知道...，以至于无穷，那么称这件事为 Common knowledge，也叫公开共识。

## 5 该问题的十层思考

### 5.1 该问题本身的答案是什么？

如果岛上有  $n$  个蓝眼人，游客说话之后  $n$  天，所有  $n$  个蓝眼人一起自杀。

证明. (数学归纳法)

1.  $n = 1$  时，即如果岛上只有一个蓝眼人，那么他原本看不到蓝眼人，那么他听到游客的话就知道了游客说的这个蓝眼人就是自己，因此他在一天以后就会自杀。
2. 假设。对于  $n = k$  这个命题是成立的，如果岛上有  $k$  个蓝眼人，游客说话之后  $k$  天，所有  $k$  个蓝眼人一起自杀。
3. 现在我们考虑  $n = k + 1$  的情况，这  $k + 1$  个蓝眼人中的每一个都会做这样的推理，我看到了  $k$  个蓝眼人，假如他们就是全部的蓝眼人，那么过  $k$  天之后他们就会一起自杀，假如他们没有在  $k$  天之后一起自杀，那就说明蓝眼人不止  $k$  个，那唯一可能的其他蓝眼人是谁呢？就是他自己，于是他等待  $k$  天，然后看到没有人自杀，由此得出结论，自己是蓝眼人。每一个蓝眼人都做出了这样的推理，都知道了自己是蓝眼，因此  $k + 1$  个蓝眼人一起自杀。

□

### 5.2 棕眼人会不会自杀？

不会，因为他们不知道岛上的眼睛颜色只有两种。

从推理过程讲，他们对眼睛做的推理只是在两种可能性中选择：蓝色或者不是蓝色，如果一个人确认自己眼睛不是蓝色，那么仍然可能是棕色，可能是红色，可能是黑色等等；因此他还是不知道自己眼睛的颜色，所以不会自杀。

### 5.3 如果没有游客，岛民们会不会自杀？

不会，因为没有推理的起点。

### 5.4 为什么我们要关心游客有没有带来新信息这个问题

只有新的信息才能造成新的结果，这是一个常识，对于逻辑系统，这也是一个真理。

但是蓝眼睛岛问题似乎推翻了这个常识，因为一眼看不出来游客带来了什么新信息，我们需要搞清楚，错误的是这个常识呢，还是游客没有带来新信息这个印象。

### 5.5 游客到底有没有带来新信息？

确实有，新信息不是游客的话：岛上有蓝眼人，而是一个  $n$  阶的知识。

如果把这  $n$  个蓝眼人称为  $A_1, A_2, \dots, A_n$ ，那么新信息就是下面这个命题：

$K_n = A_1$  知道， $A_2$  知道， $A_3$  知道， $\dots, A_n$  知道“岛上有蓝眼人”。

一个命题中出现了多少次知道，我们就把它叫做多少阶知识。

下面我们把“岛上有蓝眼人”这个命题记为  $P$ ，

$A_n$  知道  $P$  是一个一阶知识, 记为  
 $K_1=A_n$  知道  $P$ ,  
 $A_{n-1}$  知道  $K_1$  是一个二阶知识, 记为  
 $K_2=A_{n-1}$  知道  $K_1$ , 如此等等  
 那么游客带来的新信息即为:  
 $K_n = A_1$  知道  $K_{n-1}$ , 这是一个  $n$  阶知识。

## 5.6 为什么这个 $n$ 阶的知识是一个新信息?

在游客公布之前, 通过互相观察岛民们最多只能获得  $n-1$  阶知识, 不能知道  $n$  阶知识。因为每个人都不知道自己眼睛的颜色, 他在推测别人怎么想的时候, 只能考虑别人之间的相互观察, 不能考虑别人观察自己的结果, 这就意味着每多一重知道, 知识每升高一阶, 那作为知识来源的人数就减少 1, 例如对于  $K_1, A_n$  知道  $P$ ,  $A_n$  可以通过观察其他  $n-1$  个蓝眼人中的任何一个来推出  $P$ , 那么作为知识的来源就是  $n-1$ , 只需要扣除自己就行, 对于  $K_2$  就是  $A_{n-1}$  知道  $K-1$ , 那么  $A_{n-1}$  就需要扣除自己和  $A_n$ , 那么这个知识的来源就只有  $n-2$  个人, 以此类推, 对于  $K_{n-1}$  就是  $A_2$  知道  $K_{n-2}$ ,  $A_2$  就需要扣除从自己到  $A_n$  的所有人, 也就是  $n-1$  个人, 那么知识的来源只剩下了一个人, 就是  $A_1$ , 对于  $K_n$ ,  $n$  个人都被扣除了, 都不能作为知识的来源, 但是一个知识至少要有个来源, 因此  $K_n$  根本就不是个知识, 由此可见, 通过互相观察, 是无法得到  $K_n$  的, 最多只能得到  $K_{n-1}$ , 那么游客发言之后, 为什么就能得到  $K_n$  了, 因为这时游客成为了一个公开的信息来源, 每个岛民都可以从他那知道  $P$ , 不再需要通过互相观察推来推去了, 而且每个岛民都知道, 其他人都从游客那里知道了  $P$ , 因此岛民们不但是一下子知道了  $K_n$ , 而且是一下子知道了  $K_{n+1}, K_{n+2}$ , 以至于任意高阶的知识, 不过呢岛上只有  $n$  个蓝眼人, 如果知识高于  $n$  阶, 里面出现的人就必定有重复的, 这就没有多少价值。

## 5.7 新信息是如何发挥作用的?

新信息使岛民们有可能通过一个判决性实验, 来推出自己眼睛的颜色, 这个判决性实验的做法就是等待和观察。

## 5.8 在逻辑学上如何描述游客发言的效果?

游客通过公布一个岛民们早已知道的信息, 把它从弱共识提升到了强共识。这个操作产生的效果不是公布那个信息本身, 而是一个  $n$  阶的知识。

## 5.9 如何解答蓝眼睛岛问题的各种变体?

如果游客发言之后已经过了  $m$  天, 那么游客就需要指出  $m$  个人是蓝眼, 或者带走  $m$  个人, 才能阻止  $m$  个人全体自杀。

## 5.10 哪些现象可以用这些道理来理解呢?

如皇帝的新装, 语言通讯模型, 摩尔定律等。

## 参考资料

Guan Video 工作室的三个视频:

1. 做了这道数学题，可能会有减肥效果.....
2. 什么叫强共识？比如说：数学特别有趣
3. 科技袁人：如果说思考能力分为 10 层，你能达到第几层