

0-1

当我想要谈论一棵树，我会指着它，并期望你能理解我接下来所发表的议论的主体是什么。我可能会感叹树的颜色、描述其枝叶随风摇曳的情态等等，在此过程中，你也不会认为我在谈论树边长的草或是停在上面的鸟。

树这一概念何以与别的东西相区分开？

1-2

不妨先从空间入手，比如我新造一个词“甲”，用来代指一棵存在于某处的独一无二的树，占据了一处物理空间，由某些物质组成。当我说出这个词，一件也许很重要的事就是框定这棵树的范围：树皮属于“甲”、树根的泥土属不属于“甲”？刚从它上面落下的叶子属不属于“甲”？上一分钟还在我杯子里但现在已经被倾倒入土里然后被根吸收了的水属不属于“甲”？

由这些讨论可以知道：那些物质，那些比我们正谈论的东西更加基本的组成部分，它们哪些属于“甲”而哪些不属于，通常是模糊的。像一团墨迹，有公认属于的部分也有明显不属于的部分，但边缘灰灰的并不清晰，就像沙堆悖论向我们发问的那样：一堆沙要减少到多少粒才不叫一堆沙？不过一个倾向是，我们喜欢“拆解”物体和概念，还原论般地将一个整体拆成越来越细的局部，以此来对一个事物有越来越深入的了解，就像用越来越高的分辨率细致地测量海岸线的轮廓，从而越来越明确地划分什么东西“属于”某个概念而什么东西“不属于”。

但这个边界具体长什么样呢？一个词的代指范围的边缘若只是模糊则罢，如果混乱呢？也许不是每个边界都像模糊数学对沙堆悖论的处理那样有一个平滑的“多大程度上属于‘一堆沙’这个概念”的函数，而是不连续的、跳跃的、锯齿般尖锐的东西。

这团代指某个概念的墨迹具体形状也并不固定，不同人，或者同一人不同时间，脑子里关于某一概念的墨迹的形状也不相同（可以说是有着不同的分布？）。想要比较两个人脑子里指代同一个概念的两团墨迹是否相同也很困难：因为也许两个人具体到少数神经元连接所构成的脑中的画幅（landscape？）本就大相径庭。脑中墨迹的相对位置、局部环境、与其他墨迹的联系等等可能天差地别，但显示在更表层时具有某种冗余（多对一映射？）造成的相似性，从而仍然能维持日常聊天的愉快，不至鸡同鸭讲。

2-3

不同的两个人脑中的画幅在多大程度上大相径庭呢？不知道。光用肉眼看大脑的外形好像区别不大，盯着一两个神经元去检查它附近的连接那又完全混乱无规律可循，从大脑发育的角度（从无到有）去看似乎又能总结出一些神经元迁移剪枝成熟所普遍遵循的规律。从实打实的物理结构（当前阶段的科学中大家都在关注神经元连接这个层面的结构）到抽象的“意义的画幅”又是如何转变的呢？并没有很好的答案。

0-4

作一个基本假设：语言的用途是传递信息。

要传递信息，语言就必须能够**区分**。

这里的“区分”作何解释呢？如果用集合论来描述，那么“区分”大约就是用语言（一段话、一个词或者别的什么）定义一个集合，把事物划为属于该集合和不属于的两类；如果用概率分布来描述，大约是在所有事物构成的空间里定义一个非均匀的分布函数，以使人明白一段话更有可能指的是什么而不太可能在讨论什么。（我想这里应该有个更通用的理论可以统一集合论和概率论，更能把握能够唯一定义“区分”这一过程的特征，囿于本人智识，暂且不往此方向讨论）。

4-5

设想这种场景：我抛了一枚不均匀的硬币。抛掷它只会得到或正或反的两种结果，但你既不知道两种结果各是什么概率，也不知道刚刚那次抛掷的结果。

现在有下列三种情况：一是我告诉你本次抛掷结果；二是告诉抛掷此硬币得到正面或反面的概率；三是我将以上两者都告诉你。

这三种情况我向你传递的信息量分别是多少呢？

说明事件概率是如何划分的和说明已知某一事件结果是某某，传递的信息是不同的吗？

4-6

概率论里的样本空间定义为一个随机试验所有可能结果的集合，为了一一对应，不妨将“观察世界的某种特征”作为一种随机试验，得到的结果

从非模糊集合论的角度看，完全可以通过“不是某某”来定义“某某”。也就是说，当我在全集（样本空间）中找到它的一个子集并命名为甲，那么非甲可以通过甲在样本空间中的补集来得到。但是如果从概率分布的角度呢？我如果用一个定义在连续空间上的概率分布函数（标量场）来命名甲，那么非甲又如何定义呢？均匀分布减去甲分布吗？

1-7

一个概念何以与别的概念相区分？

当我们进入一个由“连续”的概念组成的空间时，一个自然的想法就是分析该空间的维度

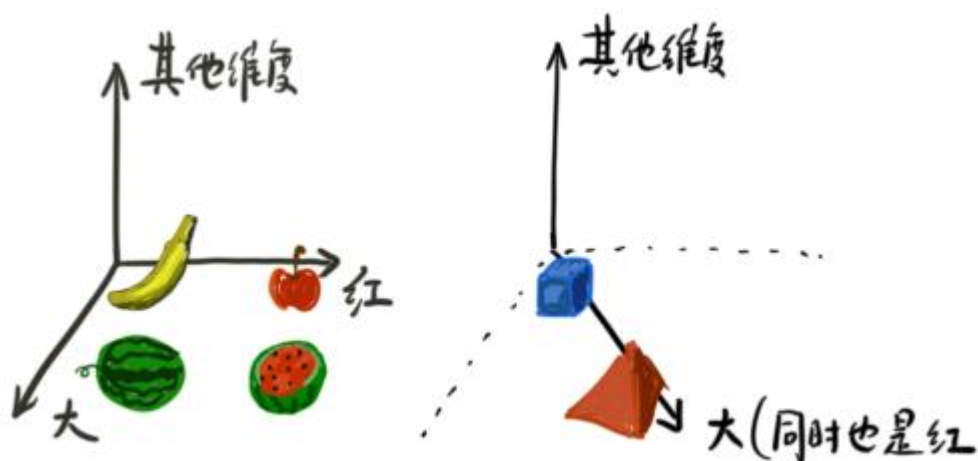
假如我们面前有许多桃子。我挑出一摞大的，说“大！”；混合均匀，再挑出一摞红的，说“红！”，我能否让他人明白我这两次涉及的是不同的“维度”？是两种不同的区分事物的角度，两种各自独立的特征？他人是如何单凭观察我的动作就明白这两个维度是不同的？

假如有一个概念的空间，用“颜色”这一维度可以将它分割成很多块，那么用“大小”的维度呢？那将会以不同的方式将它分成不同的很多块。就像大理石砖的纹路，一会可以从这几个色块里看出熊猫，一会又能从和刚刚不完全重合的色块里看出狐狸来一样。由此看，“维度”是一种分割空间的方式。（假如这个概念空间连续性很好，那么说分成很多层也许更合适）

我要用某种方式去分割这个世界，问题自然就诞生了：“要怎样”分割呢？假如你说“按颜色来分割吧！”那么我要问“颜色是什么性质？按颜色分割是怎么分割呢？”

不妨想象有这样一种分割方式，它是按颜色和按大小的某种混合，大的红色物体和小的蓝色物体会莫名其妙被归为一类。这样一种分割方式合理吗？似乎完全合理。你可能会觉得颜色和大小是完全独立的不相干的两个维度、两种判据，混合起来怎么会合理呢？暂且抛开物体颜色是否会影响人对其大小的感知的讨论（有一些研究声称会），颜色和大小为什么会被认为是相互独立的性质呢？这一结论是从日常经验里来的还是某种深埋在神经连接方式中的特性呢？

是从日常经验来的吗？也许我们潜意识里在偷偷地把我们所接收的信息作列联表，计算不同性质间是否两两独立。假如有人一辈子只见过大的红色物体和小的蓝色物体（而非小红和大蓝），他会将这两个性质联系在一起吗？会认为红就是大、大就是红，红和大是完全相同的概念，说了某物大又说该物红完全是冗余的吗？



那么与神经连接和大脑结构有关吗？我的观点是，大脑倾向于尽可能地去寻找相互独立的性质。这是要表征一个多维空间的最经济的方法（找到一组基即可用来表示空间中的点，而不必使用大量非独立的性质（矢量？）而造成信息冗余）。

提一句之前看到的关于“近似正交”的讨论。原讨论是关于机器学习的参数数量问题，设想一个 latent space 维数为 100 的模型，其只有 100 个互相正交的矢量，想要更多不相关的矢量就必须扩充 latent space 的维数，即增大模型参数量。但一方面参数量增大受算力、时间、电力等所限；另一方面有一现象，即大语言模型所学到的语义空间的维数似乎远超其模型架构所能容纳的理论最大维数，这非常奇怪了，像在三维空间里精确无歧义地表示四维物体一样反直觉。原讨论给的一个解释引入了“近似正交”（我不清楚有无更好的中文译名，暂定此名）的概念，即把两个夹角在  $90^\circ$  左右的某个范围（不严格规定为  $90^\circ$ ）的向量定义为“近似正交”。相比于严格的正交，这样可以在同样维数的空间里塞进更多相互“近似正交”的矢量。三维空间中此效果还不明显，但原讨论中计算称，如果规定夹角  $85^\circ$  至  $95^\circ$  的矢量为“近似正交”的话，在 100 维的空间中可以塞下一万余个两两“近似正交”的矢量。这样就大幅扩充了可利用的维度，计算上也很简单，只需对两个方向（的单位矢量）作点积，绝对值小于某阈值则视作正交即可。

大脑有可能通过这种“近似正交”来编码信息吗？如果是，具体的近似机制是什么呢？限制原空间维数的是什么因素呢？原空间维数和神经元数量、神经元连接又有何关系呢？

假如大脑生来就会的一个过程是：初始化许多方向随机的矢量，并根据从外界接收的信息动态调整这些矢量的方向，最后让这些矢量基本都两两“近似正交”。并且差不多指向某个方向的矢量在语言中对应着某个概念、某个词（这种对应关系也是学习来的）。不同人脑中对应同一概念的矢量可能并不指向完全相同的方向（略微有所 wiggle）

一个吐槽：工业革命后涌出许多说法说人脑就是一部像手表一样精密的机械，计算机革命后又有许多说法说人脑是一台超级计算机，LLM 应用化后又有很多说人脑就是大语言模型，我看什么时候会有人说人脑是室温超导。