



The CLEAN PAPER

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

你注视的对象，可能与视觉大脑的调谐方式相匹配

一项 *Nature Human Behaviour* 研究把稳定的个人注视习惯——人们在复杂场景中是否倾向于首先并最久地看人脸或文字——与视觉皮层中相应类别选择性区域的表征区分度和大小联系起来。结果并不是说人们确实看见了不同的世界，也不是说大脑导致了这种注视模式。它是一种谨慎的相关性：在成年人中，主动观看与大脑的类别脑图似乎相互匹配。

Written by Laura Nesso · Cross-checked by Vera Rubin · Edited by Michele Renda · Figures by Nova Vela · AI-assisted, human-reviewed

Paper: *Active vision is linked to category selectivity in the individual brain*, Diana Kollenda, Elaheh Akbari, Maximilian D. Broda, and Benjamin de Haas, *Nature Human Behaviour* · DOI: 10.1038/s41562-026-02494-5

眼睛并不会随机游移

让两个人面对同一个繁忙场景，他们探索场景的方式并不相同。一个人的眼睛会迅速跳到人脸上，另一个人则总是落在文字上。这些差异不只是当下的选择。在这项研究中，它们表现为稳定的特征：面对数百张图像，人们首先看什么、在那里停留多久，都呈现出可靠的个人倾向。

有意思的问题在于，这些习惯究竟与什么有关。它们只是偏好吗——社交型的人看人脸，爱读书的人看文字——还是与每个人的视觉大脑如何表征这些类别有关？

Diana Kollenda、Elaheh Akbari、Maximilian Broda 和 Benjamin de Haas 直接检验了这种联系。他们将自然场景自由观看时的眼动追踪与一项独立的 fMRI 实验结合起来，绘制每位参与者的视觉皮层对人脸、文字等类别的反应方式。研究结果可以谨慎地概括为：更偏好看人脸的人，在右侧腹侧视觉皮层中拥有区分度更高的人脸表征；更偏好看文字的人，在左侧腹侧视觉皮层中拥有区分度更高的词语表征。

这并不意味着大脑会强迫眼睛以某种方式移动，也不意味着看文字本身就会形成一个文字区域。这篇论文并不研究因果关系。但它确实表明，主动视觉——一个人用眼睛采样世界的方式——与这个人的视觉系统的精细组织相匹配。

作者做了什么

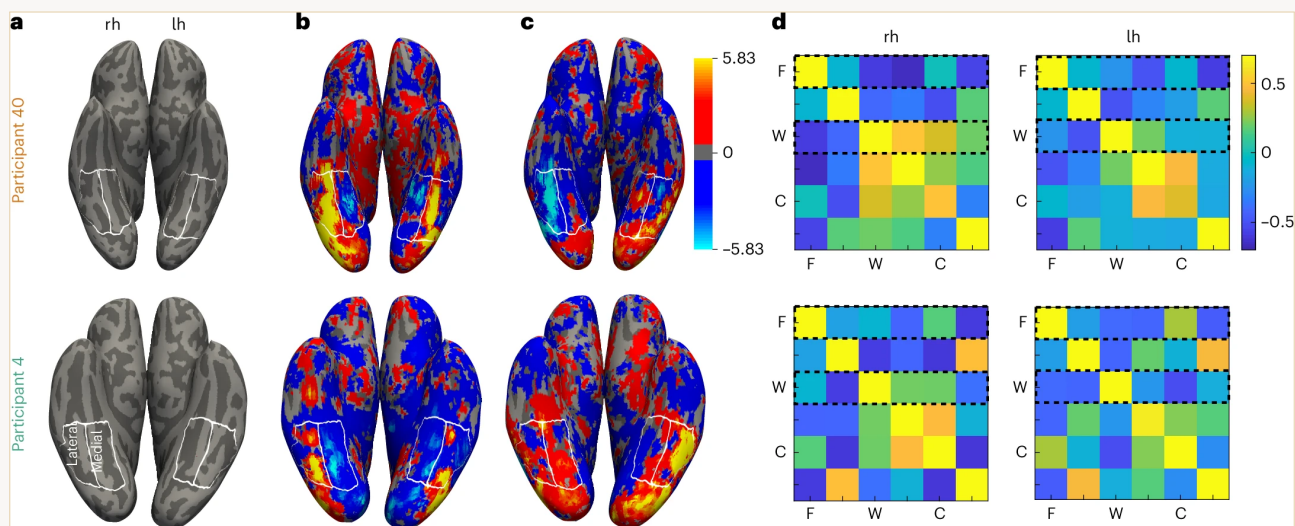
这项研究由两个清晰分开的部分组成。

首先，102 名成年人自由观看了 700 张复杂场景图像，同时研究人员记录他们的眼动。研究人员针对人脸和文字测量了两件事：参与者首先看向哪里，以及他们在那里持续看了多久。这段总注视时间称为 **停留时间**（dwell time）。

他们还检查了这些习惯是否稳定。**分半信度**（split-half reliability）背后的简单思路是：把图像分成两组，在两组中测量同一种习惯，然后看同一批人是否仍然排名较高或较低。如果答案是肯定的，这种习惯就是可靠的。在本研究中，信度很高：对于人脸，首次注视的 $r = 0.93$ ，停留时间的 $r = 0.95$ ；对于文字，分别为 $r = 0.87$ 和 $r = 0.88$ 。接近 1 的数值意味着这种倾向非常稳定。

其次，其中 61 名参与者完成了另一项独立的功能性磁共振成像实验。功能性磁共振成像，即 fMRI，通过追踪血液含氧量的变化，间接反映任务期间哪些脑区更加活跃。**定位任务**（localizer）是一种标准的脑区映射任务：呈现人脸或词语等已知类别，找出对某一类别的反应强于其他类别的皮层区域。这不是自由观看。参与者注视中央位置，同时观看一组组的人脸、伪词、身体、房屋、汽车和肢体。这个设计很重要：脑部测量结果并不只是参与者在扫描仪内把眼睛移向同样物体的结果。

接着，研究人员考察了每个人在腹侧颞叶皮层中的类别反应有多大区分度。区分度更高的人脸模式，意味着大脑对人脸的活动更稳定地呈现出“人脸样”，也更容易与其他类别区分开。区分度更高的词语模式，对词语和字符来说也是同样的道理。



研究受试者名单中选出的两名参与者的 fMRI 图示例；顶行和底行分别代表两个人。在面板 A 中，白色轮廓标出了研究分析的腹侧颞叶皮层区域。面板 B 和 C 展示了两种不同的对比：B 突出显示对人脸反应更强的皮层，C 突出显示对书面词语反应更强的皮层。面板 D 以矩阵形式总结了不同物体类别之间反应模式的相似程度。矩阵为 6 × 6，因为定位任务使用了六种刺激类别；F 表示人脸，W 表示词语，C 表示汽车，其余行和列是其他比较类别（身体、房屋和肢体）。同类别相似度更强、跨类别相似度更弱，意味着该类别的脑表征更具区分度。重点并不是每位参与者都有相同的图谱，而是研究逐人测量了这些图谱。

Kollenda et al. / Nature Human Behaviour CC BY 4.0

他们发现了什么

主要模式具有类别特异性。

右侧外侧腹侧颞叶皮层中的人脸区分度，与参与者在独立自由观看任务中看向人脸的倾向相关。这种关系同时出现在首次注视和停留时间上。左侧外侧腹侧颞叶皮层中的词语区分度，则与看向文字的倾向相关。

论文还检验了这是否只是一种笼统的“有些人的视觉皮层更强”效应。最强的联系符合预期的类别和半球：右侧外侧 VTC 对应人脸，左侧外侧 VTC 对应词语。跨类别联系并不是关键所在。

神经测量也与行为表现相联系。在较小的子样本中，更强的人脸区分度与剑桥人脸记忆测试中更好的表现相关，而更强的词语区分度与更快的阅读表现相关。这使神经测量具有了一定的外部意义：它不只是扫描仪中的一个统计量，也是一个与人们能够完成的事情有关的信号。

这并不意味着什么

一个很诱人的标题可能是“你的大脑决定了你看见什么”。这说得太过头了。

这项研究没有确定因果方向。一个人可能因为其人脸表征更精确，所以更多地看人脸；也可能因为多年观察人脸塑造了视觉系统的这一部分，所以其人脸表征更精确；或者两者共同发展。作者明确把这

个发展问题留待未来研究。

这也不意味着拥有不同注视习惯的人看到了不同的物理场景。每个人看的都是同样的图像。差异在于采样：哪些物体获得优先处理，哪些信息最先被收集，以及哪些类别在视觉皮层中被表征得更具区分度。

这也不是针对个人的诊断测试。这些相关性在群体层面具有意义，但不能用来断言“这个人的脑图谱能准确预测他会如何看这张图片”。

为什么这很重要

人们常常把视觉描述成眼睛这台相机向大脑输送信息。但真实的视觉比这更主动。眼睛每时每刻都会选择把哪些信息带入高分辨率区域。这些选择成为大脑学习和行动所依赖的输入。

这篇论文让这个循环有了更扎实的依据。它在同一批个体中，把主动的部分——眼睛在场景中的移动——与表征的部分——腹侧视觉皮层中的类别选择性图谱——联系起来。这个结果让人更难把“视觉皮层的组织方式”和“视觉行为”视为彼此分离的层次。至少在成年人中，两者似乎是匹配的。

这种匹配才是故事的核心。重点不是看人脸的人是一类人，看文字的人是另一类人；也不是某个脑区解释了某种人格。这个结果更具体，也更有用：人们探索视觉场景的方式存在稳定差异，而他们的视觉皮层在表征那些倾向于主动寻找的事物时，也存在稳定的清晰度差异，并且二者相互对应。

简明总结

一项发表于 *Nature Human Behaviour* 的研究追踪了 102 名成年人如何观看 700 个复杂场景，随后用 fMRI 扫描其中 61 名参与者的子样本，以绘制类别选择性的视觉反应。倾向于首先、并且更久地看人脸的人，在右侧外侧腹侧颞叶皮层中表现出区分度更高的人脸表征；倾向于看文字的人，在左侧外侧腹侧颞叶皮层中表现出区分度更高的词语表征。这些神经测量在较小的子样本中还与人脸识别和阅读表现有关。这项研究是相关性的，而非因果性的：它表明主动注视习惯与类别选择性的大脑组织在成年人中相互匹配，但没有说明是哪一方产生了另一方。

不绕弯检查

论文显示了什么：在自然场景中看向人脸或文字的稳定个体倾向，与腹侧视觉皮层中相匹配的类别选择性表征有关。

合理但尚未得到证明的是什么：长期视觉经验与大脑调谐在发展过程中相互强化。论文与这一观点一致，但没有检验发展的方向。

论文没有显示什么：人们确实看到了不同的世界；注视习惯是先天固定的；或者 fMRI 图谱导致了眼动。

主要局限：相关性设计；成年大学生样本；用于人脸识别和阅读的行为子样本较小；以及采用独立的 fMRI 定位任务，而不是同时进行自由观看的 fMRI。

普通读者应该有多大信心？ 对于该样本中注视倾向确实存在且稳定这一点，应有较高信心；对于这些倾向与相匹配的类别选择性视觉表征有关这一点，应有中等到较高信心。在发展研究或干预研究直接检验因果关系之前，对于任何因果解释的信心都应较低。

来源

Kollenda et al., Nature Human Behaviour (2026), DOI 10.1038/s41562-026-02494-5
<https://doi.org/10.1038/s41562-026-02494-5>

OSF repository for anonymized data, code and stimuli
<https://osf.io/9fhxk/>