11 还原与整合

对生物学现象，可以化繁为简，进行还原；也可以化简为繁，进行整合。

本章以视觉为例，讲述还原和综合。视觉是复杂的，感知始于眼睛、最终成于大脑皮层，所以需要在多个层面研究。对颜色的感知，在眼睛中第一级感光细胞对色觉分辨的分子基础，可以看到还原的重要性。而对于点和线的认知，我们可以看到整合的重要性。

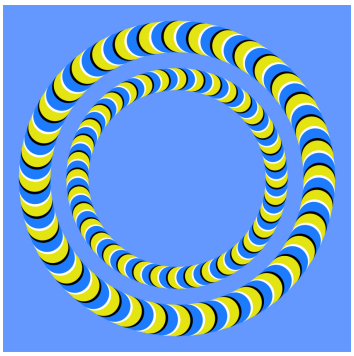
11.1 眼见不为实

眼睛是人类观察世界的重要器官，而科学家对眼睛的研究也是人类理解自身的渠道之一。不同方法和设备只能检测物质世界的部分特征。视觉系统并非照相机，而是对外界信号加工后获得对特定种属动物有意义的信息。现存的动物感觉系统是演化过程中选择的结果，有利于各种动物较为实用地快速获得和分析其常遇见的外界信息。眼和脑的局限导致我们的视觉系统不能完整和准确地反映外界的物理实际。

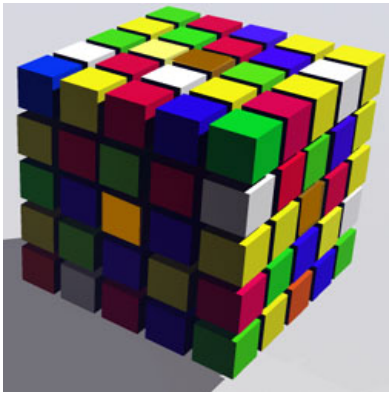
因此，眼见并不一定为实。

视错觉（visual illusion）可以显示视觉系统不是客观的简单影射。绝大多数人都有视错觉，如果少数人没有这种视错觉，其中有些人反而可能是因为基因的变化导致的异常。

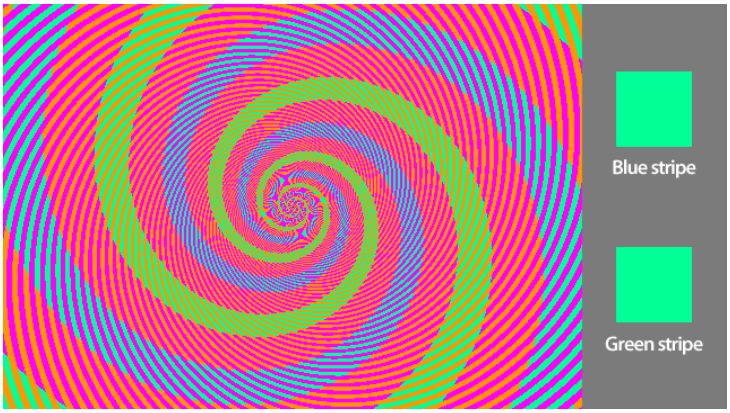
在下列视错觉图例中，第一个图如同佛家所谓外界没动、观察者心动的说法。图片中两个圆圈物理上没动，而多数人觉得它们在相对运动（一个顺时针、一个反时针）。



第二个视错觉的图，在一般人看来，立方体上面正中小方块呈“褐色”，而在立方体左侧面正中的小方块呈“黄色”，认为这两块的颜色差别很大。而实际上，读者不妨做个简单的实验，将图像拷贝下来后把这两小方块分别分离出来，可以发现它们的颜色完全一样：褐色。

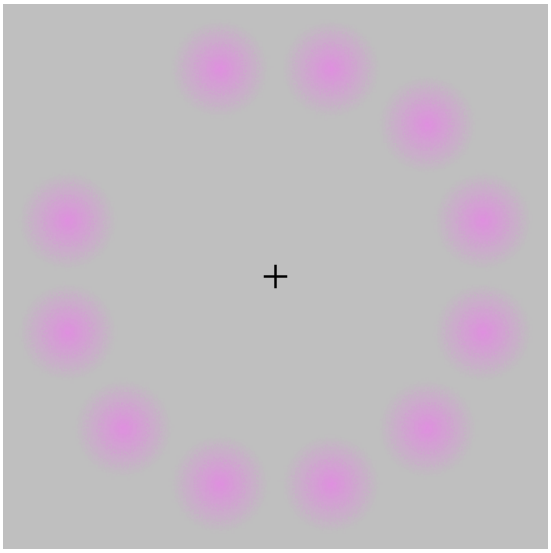


第三张图，一般人看到蓝色和绿色两个颜色的螺旋，其实这“两个”颜色在物理上是完全一样的颜色。



第四张图（可在下列网页看到 http://www.innocentenglish.com/funny-pics/best-pics/movie-dots-illusion.gif）。实际环绕加号

的紫色原点在顺时针方向逐个消失再出现。但是，当人盯着中间的加号一段时间后，可以看到一个浅绿色圆点顺时针方向运动（不是消失和再现，而是圆点绕圈运动）。如果你不相信这是你的错觉，可以在出现绿色圆点后移动你的头，绿色圆点会消失，等你再盯着加号后会再现，似乎由你不动手而遥控掌握外界。



以上这些视错觉都是我们的视觉系统处理信号所造成，其中有些机理不明。

11.2 视觉与光感

视觉与光感不同。植物有光感、有些细菌也有光感，但它们不能形成动物一样的视觉。

植物感光能力有时很奇妙。植物可以通过检测日光中不同波长的光、每日不同波长光照的时间，做出相应的季节性变化。已发现多个分子参与植物感光反应。植物的感光蛋白迄今只发现一种与动物的相似，即 cryptochromes，它在动物还用于生物钟。

动物界有多种不同眼睛和视觉系统，在进化中视觉系统独立地出现过四十多次。不同动物视觉特性不一样，如蜜蜂能感知偏振光。

人的视觉系统从位于外周的眼睛开始，眼睛的视网膜（retina）从外界获得信

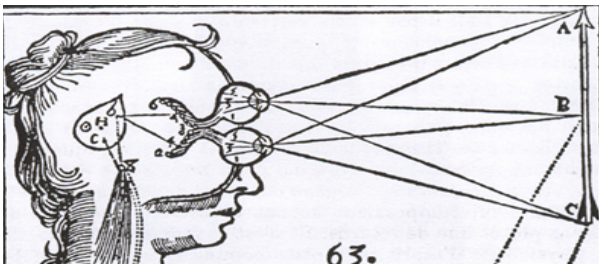
号后稍作加工、传递到丘脑的外侧膝状体（lateral geniculate nucleus，LGN）、再到大脑的视皮层（visual cortex），成为视觉通路。

人也有不形成视觉的感光系统。瞳孔对光反射指光强度增加时，我们的瞳孔缩小。这一反应既用了我们的视觉系统，也可能用了不能形成视觉的感光系统，如存在于视神经节细胞的 cryptochromes 和 melanopsin。

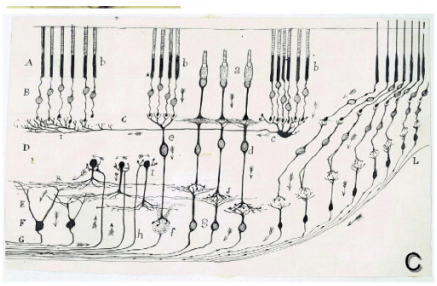
11.3 眼内的神经细胞

达芬奇曾对眼睛和视觉系统感兴趣。

笛卡尔（1596-1650）在 1637 年的《方法论》一书中曾显示眼睛的解剖、去世 12 年后出版的《论人》一书中绘制过视网膜上的神经细胞（后来称为视杆细胞）。笛卡尔也曾描绘过光通过眼睛晶状体投射到视网膜、视网膜收集的信息通过视神经从眼睛投射到大脑的过程。

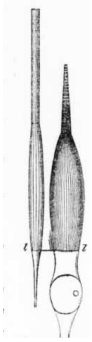


眼睛由多种细胞组成。人的眼球从前到后有角膜、晶状体（英文更能反映其作用，称其为透镜 lens）、水晶体（vitreous body）、视网膜。视网膜神经细胞通过视神经（optic nerve）向脑投射。视网膜内有三级神经细胞传递光信号，第一级为感光细胞（photoreceptors），第二级为双极细胞（bipolar cells），第三级为视神经节细胞（retina ganglion cells，RGC）。视网膜还有不直接传递信息、而参与修饰加工的水平细胞(horizontal cells)和无突起细胞(amacrine

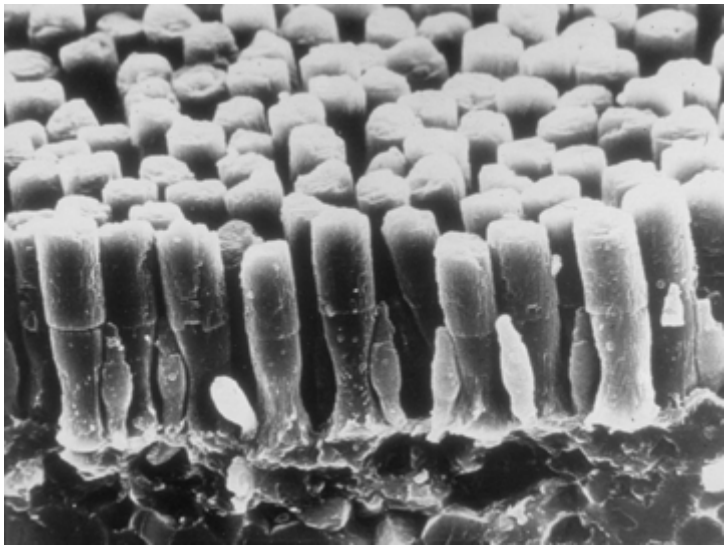


cells)。这里显示一副西班牙解剖学家 Santiago Ramón y Cajal (1852 –1934)画视网膜。

传统认为，人类的第一级感光细胞有两种：视杆细胞（rods）和视锥细胞（cones）。1675 年 Anton van Leeuwenhoek (1632-1723)就可能观察到了感光细胞，而一百多年后的 Gottfried Reinhold Treviranus (1776–1837)描述了感光细胞。十九世纪，德国的 Rudolph Albert von Kolliker （1817-1905）与 Max Schultze（1825-1874）明确了视网膜细胞的分层和感光细胞。右图为 Schultze 所画的视杆和视锥细胞。



现代的电子显微镜拍摄出了漂亮的视杆细胞和视锥细胞照片：

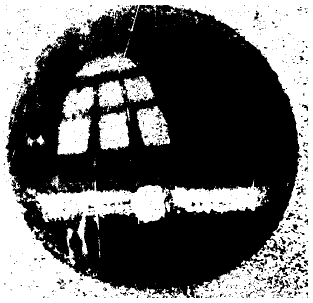


视杆细胞有一段较长的杆状部分，称为“外段”(outer segment)，起感光作用。视杆细胞的数量远多于视锥细胞：人类有 1.2 亿视杆细胞，而仅 6 百万视锥细胞。两种细胞在视网膜的分布不同，功能不同。位于视野正中的中央凹，只有密集排列的视锥细胞，没有视杆细胞。视杆细胞主要是感知明暗，视锥细胞感受不同的颜色，是辨别色觉的第一级细胞。

Hecht 等(1942)和 van der Velden (1946)发现感光细胞对光的敏感性非常高，

单个细胞就可以对单个光子有反应。他们的工作后来被 Stanford 大学的研究者用更精确的技术所验证（Baylor, Lamb, Yau，1979）。在认知上，人可感知 5 到 8 个光子。

德国海德堡大学生理学教授 Willy Kühne（1837-1900）在 1878 年发现：让兔子在暗室内注视阳光射入的窗口，一段时间后处死兔子，固定其视网膜，可看到窗子的印记。当时很轰动，以为这可以用来破案。如果被谋杀的人的视网膜能留下如此强烈的印记，找到凶手，就不怕凶杀现场没有证人，只要被谋杀的人最后一眼看见凶手就可以了。事实上，一般信号没这么强，这一发现并无实际应用价值。



11.4 视网膜的感光蛋白

1851 年，德国的 Heinrich Müller（1801-1858）发现视网膜中视杆细胞呈红色，认为是血中红细胞的血红蛋白所造成（Müller，1851；Ripps，2008）。

1877 年，德国生理学家 Franz Böll (1849–1879)发现，在暗中视杆细胞为红色，见光 20 秒后变成黄色、60 秒后成为无色，他的解释是视杆细胞中存在色素，他称为“红色物质”，见光后起反应而漂白，在暗处 2 小时后复原为红色。他排除了死亡与漂白的关系、也排除了血液与漂白的关系。Böll 提出视杆细胞外段含特殊的物质，通过光化学过程，将信号传到大脑（Böll，1871；Wolf，2001；Ripps，2008）。可惜 Böll 发表论文后因肺结核而不能继续工作，30 岁英年早逝。

Böll 的论文发表后不久，德国海德堡的 Kühne 开始继续其研究。Kühne 以前曾发现胰蛋白酶，而 1877 至 1882 年间集中时间和精力研究视网膜，发表 22 篇相关的论文（Wolf，2001）。Kühne 将 Böll 的视杆色素称为“视紫”（visual purple），

Kühne 发现光漂白后视杆细胞的在黑暗中还原为紫色还需要另一种细胞：视网膜色素上皮细胞（retinal pigment epithelium，RPE）。只有在 RPE 存在的情况下，被光漂白的视杆细胞才能在暗中重新成为紫色。还发现漂白和复原都无需血液循环，从死的动物取出的视网膜也能重复这一过程，肯定与血红蛋白无关（Wolf，2001）。

Kühne 发明了提取视紫的方法：在明矾溶液中，视网膜和 RPE 分开，视网膜硬化，容易分离视网膜，然后加入胆盐溶液，视杆细胞溶解后，视紫被释放到溶液中。因视紫对热的敏感、不能过通过半透膜、但可溶于硫酸铵，Kühne 推测它是蛋白质。Kühne 发现视紫分子本身就对光有反应，能被光漂白。他明确提出：光能解构视紫，光化学反应产物刺激视神经（Wolf， 2001）。Kühne 的工作将人类对视觉的理解推进到分子水平，并落实到特定分子。Kühne 发现的视紫，是世界上第一次发现细胞膜上的蛋白 rhodopsin，中译为“视紫红质”。

既然视紫蛋白是蛋白质，那么就是多个氨基酸按特定顺序连接而成。二十世纪，经过多个实验室的研究，逐步获得视紫光蛋白的氨基酸部分序列。1983 年，美国南伊利诺大学的 Hargrave 等和苏联的 Ovchinnikov 等分别通过分离肽段，经过分析获得视紫蛋白的全长序列。

当时已知如果连续出现 19 到 21 个左右的疏水氨基酸，就可以组成跨细胞膜的α螺旋，从视紫蛋白的氨基酸序列可推测出七个跨膜区域，即所谓的“七重跨膜蛋白”。在生物化学上，先发现一些受体（如肾上腺素能受体）与 G 蛋白偶联，所谓“GPCR”(G 蛋白偶联受体)，对感光细胞的研究也发现视紫光蛋白是GPCR。视紫蛋白是被光子激活的 GPCR，而肾上腺素能受体等是被化学分子激活的 GPCR，这些七重跨膜的 GPCR 在本质上相似。

了解视紫蛋白一级结构（氨基酸序列）后，进一步需要知道其三维的空间结构。视紫蛋白与其他跨细胞膜的蛋白质一样，以前很难通过 X 线衍射分析其晶体结构。1997 年，日本的 Kimura 等用冷冻电镜在 3 埃的分辨率解析了细菌视紫蛋白的结构（Kimura *et al*., 1999）。1999 年，法国科学家在 1.9 埃分辨率解析了细菌视紫蛋白的三维结构（Belrhali *et al*., 1999）。这些结果都清晰地显示了七重跨膜区域。2000 年，美国与日本科学家解析了眼中视紫蛋白的三维结构，并显示了视黄醛的结合位点（Palczewski *et al*., 2000）。

11.5 维生素 A 与光化学

动物感光细胞的光化学反应不仅需要蛋白质，还需要其他分子。

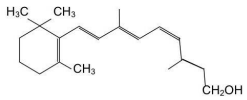
十九世纪 Kühne 研究发现了视网膜的漂白过程，从紫色、橘黄色、黄色到无色，但黄色不可能是视紫物质造成的，应该有另外一个分子，其吸收波长短于视紫物质（Wolf，2001）。后来的研究揭示黄色是视紫蛋白结合了全反型视黄醛所致。

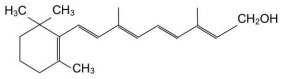
视黄醛的发现与夜盲症以及维生素 A 有关。几千年前人们就注意到有些人在暗光下视力不如一般的人，而且发现有些夜盲与营养相关（Wolf，2001）。希波克拉底（公元前 460-325）曾建议夜盲症患者吃肝脏。1913 年，威斯康辛大学的 Elmer McGollum 与 Marguerite Davis、耶鲁大学的 Thomas Osborne 与 Lafayette Mendel 在研究老鼠生长需要的营养因素过程中发现维生素 A。1925 年，丹麦哥本哈根大学卫生研究所的 Fridericia 与 Holm 证明维生素 A 参与视光漂白后在暗中的复原，并提出这可能解释维生素 A 与夜盲症的关系。Yudkin 于 1931 年证明视网膜含维生素 A。

犹太裔美国科学家 George Wald（1906-1997）于 1927 年在纽约大学毕业后

进入哥伦比亚大学，研究生第一年听过摩尔根和 Selig Hecht（1892-1947）的课，后选 Hecht 为导师。犹太裔的 Hecht 当时是世界著名视觉生理学家，1919 年曾提出光感受的基本过程：光敏感物质 S，经光解后为 P 和 A，在暗中复原为 S（Hecht，1919）。Wald 研究生课题是果蝇视觉，发现果蝇与人的视觉有相似之处，1932 年获博士学位（Dowling，2002）。Wald 到德国做博士后，分别师从 Otto Warburg（1883-1970，因研究细胞氧化还原过程的机制而获 1931 年诺贝尔生理奖）、Paul Karrer（1889-1971，因研究胡萝卜素及其与维生素 A 的关系获 1937 年诺贝尔化学奖）、和 Otto Meyerhof（1884-1951，因研究糖酵解而获 1922 年诺贝尔生理奖）。Wald 在三个实验室学习了不同的思想和分析方法（Dowling，2002）。Wald 研究生期间学习的领域成为他以后研究方向：视觉，而博士后期间学习的是他进行研究所需要的方法：生物化学。他的研究是视觉的生物化学，分析参与视觉的分子、以及分子在视觉过程中的变化。Wald 在 Warburg 实验室学会用光谱分析，发现与视紫蛋白相连的分子可能是维生素 A。Warburg 建议 Wald 师从做胡萝卜素的 Karrer，Wald 在 Karrer 实验室检测了牛、羊、猪的视网膜，确认都含维生素 A（Wald，1933）。他再到德国犹太科学家 Meyerhof 实验室，发现在暗中和光照后的视网膜，维生素 A 相关的分子有所不同，并能双向改变，从而提出视循环（visual cycle）学说。1934 年 Wald 回到美国，从此在哈佛大学继续研究维生素 A 与视觉的关系（Wald，1935，1968），他和同事几十年的工作推进了对视网膜光化学的理解，他们发现的视循环就是视网膜有视黄醛的两种光学异构体，全反型（all-trans retinal）与 11 顺型（11-cis retinal）。在黑暗时，11 顺视黄醛与视紫蛋白形成共价键结合，结合位点是视杆光蛋白中赖氨酸的一个，这一构型稳定；而光照时，11 顺视黄醛变构为全反型视黄醛，脱离视紫蛋白，随光照与否，两

种视黄醛循环变化。视循环也称视紫蛋白循环、或视黄醛循环，目前仍有研究（Smith，2010），视黄醛顺反光学异构的速度很快，在 10^{-13} 秒内发生。光化学变化也不仅仅是一步，而有多步。





脱离了视黄醛的视紫蛋白本身构型改变，触发其下游的小 G 蛋白复合体，复合体有三个亚基（α，β，γ），解离后 G α亚基结合并激活磷酸二酯酶，后者降解 cGMP（Arshavsky 等，2002）。在脊椎动物视网膜，感光细胞的关键，是第二信使分子是 cGMP。1985 年苏联科学院生物物理研究所的 Fensenko 等发表论文，用膜片钳技术证明蛙的感光细胞及其外段能被 cGMP 开放阳离子通道，使钠离子和钙离子流入细胞内，兴奋感光细胞。

感光细胞在电生理有几个特点：一般神经细胞被刺激时多半被激活，产生动作电位。而脊椎类的感光细胞，首先不能产生动作电位；其次，在被光刺激的时候，不是被激活，而是被抑制。在黑暗中，感光细胞释放神经递质谷氨酸。在大脑中谷氨酸一般只兴奋下级神经细胞，而视网膜感光细胞释放的谷氨酸对下级的双极细胞有两种不同的作用，兴奋或抑制，这取决于双极细胞表面不同的谷氨酸受体。被谷氨酸兴奋的双极细胞，在光照后因为感光细胞释放谷氨酸的减少而相对被抑制；而被谷氨酸抑制的双极细胞，在光照后因为谷氨酸减少而相对被兴奋。

11.6 色觉的三原色学说

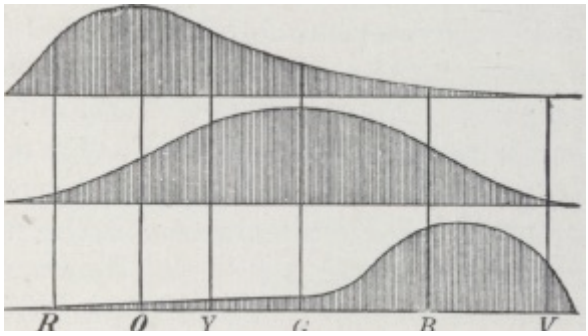
英国科学家胡克（Robert Hooke，1635-1703，弹性定律的提出者）于 1665 年提出：光照到视网膜的角度不同导致不同的颜色。胡克对色觉的推想是错的，而其长期竞争对手牛顿有关色觉的想法在原则上是对的。

牛顿（1642-1727）与胡克因为万有引力和光学有过多个发明权的争论。牛顿可能因此等胡克去世以后，才在 1704 年发表《光学》一书。从物理学角度，大家熟知牛顿于 1672 年曾用三棱镜研究颜色的光学特征。从生物学角度，牛顿知道色觉是神经系统对外界的反应，他提出:“May not the harmony and discord of colours arise from the proportions of the vibrations propagated through the fibres of the optick nerves into the brain”，光振荡以不同比例作用于视神经，传入大脑而导致人对色的感觉不同。在感觉神经生物学，牛顿也有先驱的想法。

三原色最初由 George Palmer（aka George Giros de Gentilly，1746-1826）于 1777 年提出:“每束光由三种、也仅有三种组成，黄、红、蓝……视网膜表面有三种不同的颗粒，类似于三种光；每种颗粒为相应的光束所动”。其中，有关光只有三种成分的物理学理解是错误的，而对视网膜的假设与现在的理解有相通性。很多人认为 Young 第一位提出色觉三原色的第一位，是忽略了 Palmer 的贡献，虽然 Palmer 在理解光物理上有失误。

英国科学家 Thomas Young（1773-1829）于 1801 年宣读、1802 年发表的“光和色的理论”。在引用了牛顿在《光学》中对颜色和色觉的讨论后，Young 提出，在视网膜的光敏感点不可能含无穷的颗粒，每一颗粒与对应的光同步振荡，所以需要提出有限的颗粒数，比如三原色，红、黄、蓝。每一种的振荡或多或少偏离完全同步。视神经的每根纤维可能有三部分，分别对一种原色敏感。至于为什么是这三种颜色，他当时只是猜测性地将三种原色分布在可见光较广的范围内。

英国物理学家麦克斯韦（James Clerk Maxwell，1831-1879）和德国生理学家赫姆霍兹（Hermann von Helmholtz，1821-1894）发展了 Young 的色觉理论。麦克斯韦还在 1861 年显示了彩色照相的可能。赫姆霍兹对光的生理学有深入和细致的讨论，还研究了视神经的电生理。对于色觉，他发展了 Young 的理论，提出有三种视神经，分别对红、绿、紫最敏感，各种颜色的光对这三种视神经有不同程度的激活，导致不同的色觉。他指出理论上无法确定哪三个颜色是三原色，可能需要通过分析色盲的情况来确定。



瑞典芬兰生理学家 Ragnar Granit（1900-1991）在动物的实验发现不同的视锥细胞可能分别对应不同颜色（Granit，1941，1943，1945）。人的视锥细胞到底有几种、分别对什么波长最敏感，直到 1964 年才由美国霍普金斯大学的科学家所确定（Marks *et al.*，1964）。人类三种视锥细胞最为敏感的波长分别为：564 nm、533 nm 和 437 nm。它们分别接近黄、绿和紫，而没有对红色最为敏感的细胞，现在称它们为长（L）、中（M）、短（S）波的视锥细胞。不同的动物色觉细胞种类不同，所含光蛋白（opsins）敏感的波长也不同。果蝇有对紫外线敏感的感光细胞。

11.7 色觉的分子基础

Wald 研究过不同色觉的细胞，发现它们都用视黄醛，但所用的光蛋白不同。那么，如何找到人类对不同颜色反应的光蛋白？1980 年代初期，Jeremy Nathans（1958-）在 Stanford 大学做研究生，导师为 David Hogness。Hogness 从 1960 年代末起研究果蝇发育的分子生物学，并不断推广果蝇分子生物学技术，但 Nathans

加入 Hogness 实验室的研究，是分析人类的色觉蛋白基因，而且坚持在果蝇实验室研究人的基因。

1982 年已有牛视紫蛋白的部分氨基酸序列发表了。1983 年美国的 Hargrave 和苏联的 Ovchinnikov 分析获得了牛视紫蛋白的全长氨基酸序列。

已知主管明暗的视紫蛋白后，怎么样才能拿到色觉的光蛋白呢？1983 年，Nathan 与 Hogness 清晰地介绍了他们的策略。他们推测感色觉的光蛋白（我们这里姑且称为视锥光蛋白，因为应该是存在于视锥细胞中）与视紫蛋白有部分序列相似性，所以可以通过视紫蛋白找到视锥光蛋白。Nathans 的研究用了四步。第一步，通过已知的牛视紫蛋白部分序列，找到其编码的 cDNA 而预测其全长蛋白质的氨基酸序列；第二步，用牛的视紫蛋白基因寻找人的视紫蛋白基因；第三步，用视紫蛋白基因寻找视锥光蛋白的基因；第四步，确定每个视锥光蛋白对应的颜色。

Nathans 的第一步是用分子生物学技术，而 Hargrave 和 Ovchinnikov 是用生物化学方法。生物化学分离纯化貌似比较老，但到今天都很有用，有时还起不可替代的作用。掌握生物化学方法比较难，有了分子生物学以后，在有些方面可以替代一些（但非全部）生物化学的方法。比如，在得知部分氨基酸序列的情况下，制备 DNA 探针，筛选含全长基因的 DNA 文库而找到相应的全长基因。Nathans 和 Hogness 依据 Hargrave 等（1982）报道的牛视紫蛋白部分序列，制备 DNA 探针，通过筛选获得编码视紫蛋白的 cDNA，得以推测视紫蛋白全长氨基酸序列，

第二步，Nathans 用牛视紫蛋白基因的序列设计探针，在人的 cDNA 文库中寻找人的视紫蛋白。假设牛视紫蛋白的一段序列是 TAGTACTACTTGT，那么人的视紫蛋白相应这一段可能有些差别，但非百分之百相同。通过降低杂交对顺序

相似程度的要求，用牛的视紫蛋白基因做探针来找人的视紫蛋白基因。Nathans 称他用了自己的“种系发生细胞”制造人基因文库的 DNA。Nathan 和 Hogness 于 1984 年报道找到人的视紫蛋白基因，其氨基酸序列与牛的视紫蛋白相似性为 93.4%。

第三步，用人的视紫蛋白基因 DNA 做探针，寻找视锥光蛋白的基因。这是基于推测视锥光蛋白也应该与视紫蛋白有一部分序列相似性。这样也是用人的视紫蛋白基因做探针，通过低保真度杂交，找到其他三个基因。Nathans 等（1986）发现它们编码的基因相互有高度的相似性，而与视紫蛋白的相似性仅 41%。

如何确定这三个基因相对应什么颜色？Nathan 等在 1986 年的第二篇《科学》论文报道，根据色盲患者 DNA 片段异常和相应于 Nathans 获得的基因的分析，从而确定了基因与色觉的对应关系。这是人类遗传学与分子生物学相结合的成功范例。这一系列漂亮的工作为色觉分子研究的里程碑。

11.8 还原

色觉的第一级，发生在视网膜的三种视锥细胞，它们分别对不同的颜色有最大的反应，相应三个最佳波长，这并非一般所谓红绿蓝的三原色，而黄绿紫是真正的三原色。

视锥细胞中参与感光反应的是视锥光蛋白与视黄醛，其中吸收光的是视黄醛、但视锥光蛋白影响其对不同颜色的反应。人的视网膜有三种视锥光蛋白，其基因于 1986 年获得，基础工作是首先通过生物化学方法，先获得了视紫蛋白的部分序列。一般各种颜色的光，以不同比例刺激三种视锥细胞，信号在视网膜加工后，传入大脑，再经过更复杂的加工，而获得高级的认知。

对色觉的第一级分子理解是否完结，并未确定。以前认为感光细胞只有第一

级的视杆细胞和视锥细胞，1998 年发现视网膜第三级的视神经节细胞中，有少数细胞含另一种光蛋白：melanopsin（黑光蛋白）（Provencio *et al.*, 1998, 2000）。这些细胞可能参加光感（Hattar *et al.*，2002；Berson *et al.*，2002）。而 2012 年底 Stanford 大学心理系学者提出人视网膜在边缘区域可能有对三色以外第四色敏感的细胞（Horiguchi *et al.*, 2012），也许是含 melanopsin 的细胞。其后的研究发现，光蛋白可能在一些细胞和系统里其的作用与感光无关（Senthilan *et al.*，2012），可能演化过程中，其作用也不一定是从感光开始的。生物有很多奥秘，即使在分子层面也不一定就一时能完全清楚。

分子层面并非色觉研究的终点，大脑对色觉的理解还有更深、更多的机理，包括人类目前还不知道的奥秘。

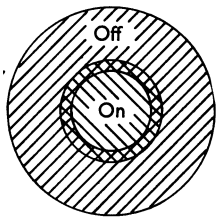
11.9 整合：视网膜

视觉认知是逐步整合的结果。

视网膜神经信息传递的三级细胞（感光细胞、双极细胞、视神经节细胞）有整合，视网膜内还有其他细胞（如水平细胞、无突细胞）也参与调节、整合。

美国生理学家 Keffer Hartline（1903-1983）于 1932 年记录到马蹄蟹单根视神经的动作电位，发现视觉刺激强弱影响其发放频率，但不影响其形状（Hartline and Graham，1932）。更多的脊椎动物视神经研究显示，有些视神经对恒定的光照有稳定的反应，有些对给光有反应（on response），有些视神经是撤光的时候反应最强（off response），有些视神经是给光和撤光的时候反应都强（on-off response），有些视神经只对撤光有反应，光线的运动也能影响一些视神经的反应（Hartline，1938，1940a，1940b）。Hartline 称：“单个神经细胞不会独立行动，视觉系统所有单位行动的整合才能产生视觉”（Hartline，1942）。

Stephen Kuffler（1903-1980）研究猫的视网膜时发现，RGC 的感受野不是简单的对光反应，而有特定模式。有些 RGC 的感受野是中心为 On 反应、周边为 Off 反应（称为“On 中心”感受野）。有些相反，中心为 Off 反应、周围 On 反应（称为“Off 中心”感受野），



而且用光同时刺激中心和周边区域可以抑制神经元的反应（Kuffler，1953）。英国的 Horace Barlow（1921-，达尔文的曾孙）研究蛙的视网膜得到类似发现（Barlow，1953；Barlow，FitzHugh and Kuffler，1957）。这些发现显示视觉传递到 RGC 时经过了视觉加工处理。

哺乳动物的 RGC 送出神经纤维投射到外侧膝状体（LGN），双侧各有一个 LGN，分六层，2、3、5 层接受同侧视网膜的 RGC 投射，1、4、6 接受对侧视网膜的 RGC 投射。LGN 的神经投射到大脑皮层枕叶的初级视皮层（V1）。

11.10 视皮层研究的基础

对初级视皮层的研究，是神经信息整合为认知的一次突破。加拿大旅美科学家 David Hubel（1926-2013）和瑞典旅美科学家 Torsten Weisel（1924-）的合作为核心内容（Hubel and Wiesel，1998，2005）。1958 年，Wiesel 是霍普金斯大学眼科研究所 Kuffler 的博士后，Hubel 来加入生理系 Vernon Mountcastle（1918-2015）实验室，但因生理系拥挤安排不过来，Kuffler 建议 Hubel 与 Wiesel 合作九个月等空间。这一合作延续 25 年，而且前五年就已硕果累累。Kuffler 和 Barlow 等已经研究视网膜的信号，Hubel 和 Wiesel 研究视皮层。

在技术上，Kuffler 的合作者发明了较好的仪器（Talbot and Kuffler，1952），可以固定头和眼睛，而且可以局部给光刺激视网膜较小的区域，改进了此前常用的弥散给光。Hubel 本人发明了用钨丝电极记录皮层电位（Hubel，1957），很快

成为皮层电生理研究的常用工具，它的尖足够小（直径 0.5 到 0.05 微米）可以较好记录单细胞放电，而柔韧性足够强，相对不易断。Hubel 还发明了微推进器，将微电极插入脑内（Hubel，1959）。工欲善其事必先利其器的 Hubel 综合多种实验方法，设计建立了研究途径：发明电极，用微推进器将之插入视皮层，用 Talbot 和 Kuffler 发明的仪器局部给光刺激视网膜后记录视皮层的信号，然后用电流损毁电极所在部位，电生理实验完成后可以通过解剖观察损毁部位而确定电极插入的部位（Hubel，1959）。

在科学上，Mountcastle 在体躯感觉皮层的研究是 Hubel 和 Wiesel 研究的基础之一。Mountcastle 发现了皮层功能柱（cortical columns）：功能柱垂直于皮层表面，其中不同层的神经细胞都对同样的体躯部位产生感觉反应（Mountcastle，1957， 1997）。他当时可以分四种感觉模式：皮肤的毛发位移、皮肤压力、深部压力和关节转动。他发现功能柱内部的神经元都对同一种感觉刺激起反应，而不对其他起反应（Mountcastle，1957）。当时已有对皮层纵切的实验，它们对皮层功能影响很小（Lorente De Nó，1949；Sperry，Minor and Myers，1955），Mountcastle 推测这些是由于对纵向功能柱影响较小的缘故。

人的大脑皮层约 2600 平方厘米，从脑膜紧邻的第一层到邻接脑室的第六层，其中功能柱从第 2 层到第 6 层，每个功能微柱（minicolumn）一般含 80-100 个神经元（纹状皮层 2.5 倍于此），多个微柱组成直径为 300 到 600 微妙的功能柱（cortical column）（Peters and Sethares，1996；Mountcastle，1997）。听觉皮层、运动皮层也一样有功能柱，听觉皮层的功能柱与声波的频率等相关（Mountcastle，1997）。功能柱可能与发育过程中神经细胞前体的起源和迁移有关（Rakic，1972，1995）。

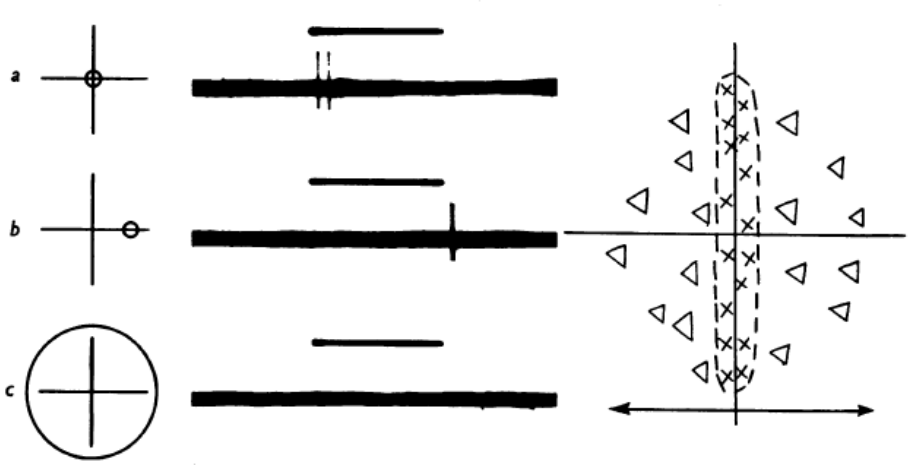
11.11 整合：初级视皮层

在 Hubel 和 Wiesel 之前，极少研究视皮层。他们起初研究猫的视皮层，后研究猴的视皮层。

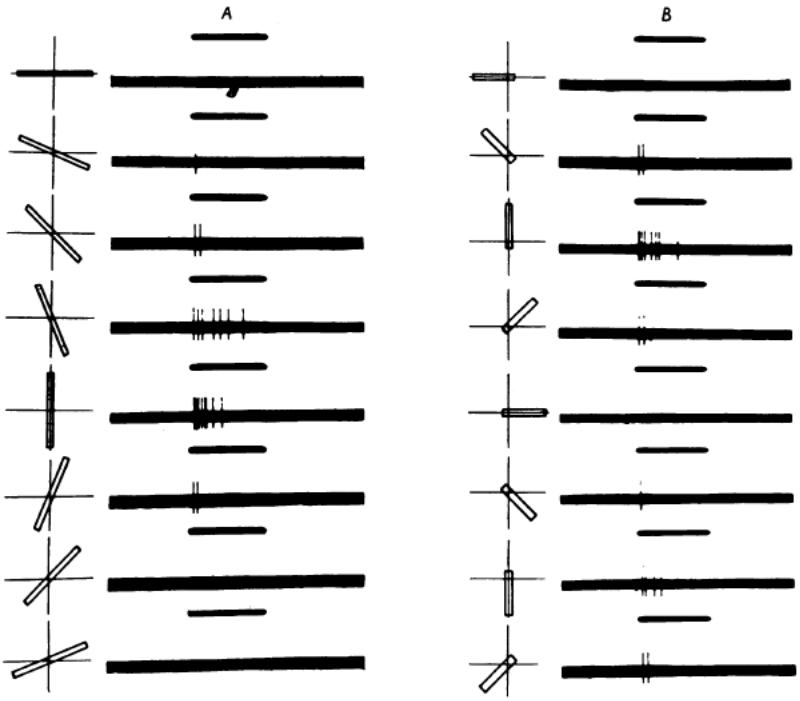
在仅有的视皮层研究中，德国佛莱堡大学的 Jung 等用了弥散的光作为刺激，作用于整个视网膜，记录到百分之五十的 V1 细胞对光有反应，其反应类似视网膜的 RGCs（Jung，1953，1958；Jung and Baumgartner，1955）。

1959 年，Hubel 比较了弥散光和点光，发现 V1 的细胞基本对弥散光不反应，而对局限光反应很强，On 反应和 Off 反应都有。存在对移动的光有反应的 V1 细胞，其中多数只对一个方向移动的光有反应，有些对于静止的光无反应（Hubel，1959）。这一研究看起来是用新的记录电极和给光方法，重复 Jung 的研究，但却有新发现。事后分析，Jung 等的研究不仅给光方法不佳，而且记录有问题，估计是记录到了 LGN 投射到视皮层 V1 的纤维，而不是 V1 内部的细胞，所以记录到的反应类似视网膜的 RGC。

Hubel 和 Wiesel 最初预测在视皮层也发现类似 RGC 的反应，On 中心反应、Off 中心反应（Hubel and Wiesel，1998）。但他们发现了差别：中心不是圆的，而是“拉长了的圆“（Hubel and Wiesel，1959）。他们称这种细胞为“简单细胞”。



他们提出这些细胞是对直线有反应。按这样的想法，他们进一步实验确定 V1 确实有对直线有反应的细胞，对而且直线的朝向有要求，只有特定朝向（orientation）的直线才能引起特定 V1 细胞的反应，所以是“朝向选择性细胞”。

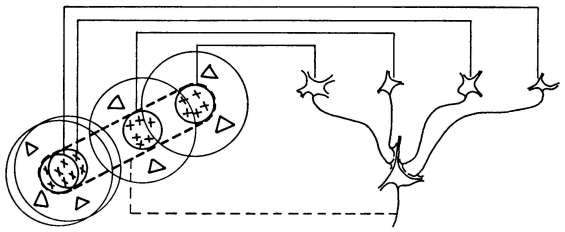


Hubel 和 Wiesel 还发现，V1 的细胞一般主要对一侧眼的刺激有反应，因此有眼优势细胞。对运动有反应的细胞对直线运动方向（direction）有选择性（Hubel and Wiesel，1959）。

他们还发现，视皮层也有功能柱，而且同一个功能柱的细胞都对同一个方向

的线有反应，既朝向选择性功能柱（orientation selective columns），也有主要对一侧眼有反应的功能柱，眼优势柱（ocular dominance columns）(Hubel and Wiesel，1962，1963a)。因此，皮层的功能柱是普遍的规律，而不限于体躯感觉皮层（Mountcastle，1957），相邻的朝向功能柱偏好的朝向相近，逐渐变化。

在发现这些事实的基础上，Hubel 和 Wiesel 提出 V1 之所以有这种识别线的简单细胞是因为他们接受了按一定朝向排列的 LGN 识别点的细胞的神经投射（Hubel and Wiesel，1962）。



这是对认知现象提出大脑皮层神经生理学解释的优美范例。

Hubel 和 Wiesel 对于简单细胞的解释简单而优美，很快认为可能是对的。但不容易检验。因为需要检测投射到同一个 V1 线识别细胞的 LGN 神经纤维末梢到底对什么有反应。1990 年代中期才有实验似乎支持 LGN 投射与 V1 细胞反应的关系（Reid and Alonso，1995；Ferster，Chung and Wheat，1996；Hubel，1996）。为了分离 LGN 投射的神经末梢电位与 V1 局部神经元网络的电位，Ferster 等用全细胞膜片钳方法，记录 V1 细胞的膜电位，并通过降温大幅度减少 V1 神经元网络的活动，他们发现，V1 神经元的膜电位仍然保持朝向选择性。由于冷冻 V1 后 V1 神经元的膜电位主要来自 LGN 投射，这就证明 V1 简单细胞的朝向选择性的确可以由 LGN 的传入产生，这就支持了 Hubel 和 Wiesel 的模型（Ferster，Chung and Wheat，1996；Hubel，1996）。Reid 等人同时在 LGN 和 V1 进行单细胞微电

极记录，并使用脉冲相关分析判定 LGN 细胞与 V1 简单细胞之间是否存在单突触连接，他们发现，与 V1 细胞有单突触连接的那些 LGN 细胞，其感受野的确按 V1 简单细胞的感受野朝向排列。当然，V1 局部网络对 V1 简单细胞的反应也有贡献。

但还有不同于 Hubel 和 Wiesel 的解释：V1 细胞有计算功能，或 V1 不同功能柱之间可以横向相互作用，也可以解释 V1 简单细胞的反应。而复杂细胞和高复杂细胞，就可能更复杂。在鼠进行双光子成像，通过钙离子浓度反映神经细胞的兴奋性（Jia *et al*., 2010）。在小鼠视皮层单细胞树突进行光学成像辅以电生理记录发现：同一根树突上，接受不同朝向直线刺激的输入，而无论细胞的输出朝向，其输入都有多种朝向，从而支持 V1 细胞不是依赖输入的相似性决定其输出，而是单个细胞可以计算其多种输入信号，整合后得到单一输出信号（Jia *et al*., 2010）。

所以，前级神经纤维投射确定 V1 简单细胞朝向选择性的机理并非已经证明。认知的神经机理，可能比以前想象的要复杂。

Hubel 和 Wiesel 还发现猫的 17 区（猴的 V1）有对更复杂型式反应的细胞：对运动的线，对运动的边界，对特殊的形状。他们称之为“复杂细胞”（没有 On 和 Off 区域，有位置不变性）（Hubel and Wiesel, 1962）。在猫的 18 区（猴的 V2），他们发现 90%的细胞是复杂细胞，在猫的 19 区（猴的 V3），他们发现 42%为复杂细胞、58%为“超复杂”细胞（后者反应的图形需要有非连续性，亦称 end-stop）（Hubel and Wiesel，1965a）。V2 和 V3 的一个功能柱可同时含有复杂细胞和超复杂细胞。V2 和 V3 的细胞多数为双眼驱动。从 V1 到 V3 越来越复杂，而且解剖上，他们观察到 V1 对于 V2 和 V3 有投射。Hubel 和 Wiesel 提出复杂细胞接

受简单细胞的投射从而解析更复杂的视觉刺激（Hubel and Wiesel，1965a）。当时很容易认为视觉认知从此都由简单的投射法则而解决。

以上实验是用猫做的。Hubel 和 Wiesel 此后在猴做了同类实验（Hubel and Wiesel，1968）。在猴的 V1，他们发现有简单、复杂和超复杂细胞，简单细胞多于复杂和超复杂细胞。猴 V1 也有朝向选择功能柱和眼优势功能柱，而且相互独立。第 II 和第 III 层的浅层 2/3 只有复杂和超复杂细胞，没有简单细胞，双眼驱动。第 III 深层的 1/3、IVa 和 IVb 层有简单细胞，其中 IVb 主要是简单细胞、几乎无复杂细胞。第 IV 层细胞基本都是单眼驱动。第 V 和 VI 层主要是复杂和超复杂细胞，双眼驱动。因为视皮层的第 IV 层接受丘脑 LGN 输入，提示 IV 层简单加工后，再在其他层曾进一步加工。

视觉系统的整合加工仍然是认知的模型。但其他认知如嗅觉、听觉的神经生物学也有很大进展，与我们对视觉的理解相辅相成。

显然，迄今为止，在生物学研究成功应用还原的范围和例子都超过整合。这是过去，何时整合超越还原，是一个有趣的问题。

注 1：Kühne 证明 RPE 参与视紫物质暗复原的实验：a.把青蛙的眼球置于光照之下 30 分钟，视网膜褪色，在黑暗中保存后紫色重现；b.褪色后的视网膜与 RPE 分离，在黑暗紫色不能复原；c.把褪色后的视网膜放在分离的 RPE 上，在黑暗中紫色可以复原；d.如果把 c 中的视网膜和 RPE 之间放置一片不通透的膜，不能复原；e.如果 c 中的 RPE 死亡时间过长（数小时），那么视网膜不能复原；f.如果漂白溶解在胆盐中的视紫物质，单独在黑暗中不能复原，而褪色的视紫物质如果与 RPE 混合则能在黑暗复原。Kühne 的遗嘱要求自己被火化时演奏贝多芬的第九交响乐（Ripps，2008）。

注 2：本文将 rhodopsin 称为“视紫蛋白”。

注 3：牛顿《光学》出版于 Hooke 去世以后，与两人一直有争论可能有关。牛顿的名言“我如果看的远是因为站在巨人肩上（If I have seen further it is by standing on the should of giants)”出现在他与 Hooke 争论的通讯中，有两位英国传记作家对之解读为牛顿可能是讽刺挖苦驼背的 Hooke。不过，这句话并非牛顿发明，在十二世纪就有人提过。

注 4：George Wald 与夫人 Ruth Hubbard 在 1960、1970 年代加入社会活动，如反核，但也曾错误地积极反对重组 DNA，在 1970 年代导致哈佛所在地曾限制重组 DNA 的研究，使得一些哈佛教授逃到其他可以做 DNA 实验的地区去。

注 5：**互补原则和核酸杂交**。DNA 两条相互环绕的核酸链所含核苷酸（A、C、G、T）通过氢键形成特异性的配对：A 对应 T，G 对应 C，也称碱基配对。这一互补原则（complementarity），有重要的生物学意义，而科学家们利用互补原则发明了新的技术：分子杂交（molecular hybridization）。用同位素标记一条链的 DNA，作为探针，通过探针与互补链的结合而找到另外一条链。如果用双链 DNA 做的文库，在一定条件下两条链松开，那么，一条短片段做的单链探针可以找到文库中相应的 DNA。要求两条链 100%碱基对配对的互补可以用高保真的条件达到。而有些基因在不同种类进化后，不会百分之百一样。比如人与牛的同一基因有一些差别。低于百分之百互补的两条 DNA 链，在一定条件下，也能杂交。需要降低杂交时对互补的要求，所谓低保真度杂交（low stringency hybridization）。降低杂交的保真度可以通过降低杂交进行时的温度、或增加溶液中的盐来达到。也就是用牛的基因得到的 DNA 探针，通过低保真杂交可以在人基因文库找到人的基因。

注 6：Jeremy Nathans 的父亲为霍普金斯大学的 Daniel Nathans（1928-1999），后者与研究生 Kathleen Danna 于 1971 年发现 DNA 了限制性内切酶，1973 年美国旧金山加州大学（UCSF）生化系的 Herbert Boyer 和 Stanford 微生物系的 Stanley Cohen 在限制性内切酶的基础上，提出其后普遍通用的重组 DNA 方法，开创了基因工程的时代，奠定了生物技术产业的一个主要支柱，也改观了人类的生命科学研究。1978 年 Daniel Nathans 与瑞士的 Werner Arber 和美国的 Hamilton Smith 共享诺贝尔生理或医学奖。

注 7：人们通常难以意识到自己的缺点。John Dalton（1766-1844）研究集中于化学和物理，特别以其对原子论的贡献著称。1798 年，道尔顿发表“有关颜色感觉的特殊事实：观察”。道尔顿告诉大家自己和兄弟的色觉有问题。为了搞清楚这一问题，道尔顿遗嘱安排死后解剖自己的眼睛。1844 年 7 月 27 日，Dalton 去世后，28 日医生 Joseph Ransom 解剖 Dalton 的眼睛并保存下来。1995 年，伦敦大学分子遗传系的 Hunt 等从道尔顿的眼球取少量组织，制备 DNA，分析了他的色觉基因（光蛋白基因），确定其特定光蛋白基因 DNA 突变，也就解释了道尔顿色盲。

注 8：Mountcastle 和两位同事的结果最初于 1955 年在美国生理学会宣读（Davis, Berman and Mountcastle，1955；Mountcastle, Berman and Davis，1955），但等到 1957 年发表文章的时候，两位合作者不敢接受 Mountcastle 的功能柱概念，因此文章仅 Mountcastle 一位作者（Snyder，2015）。

注 9：Hubel 和 Wiesel 初期工作在霍普金斯大学，1959 年他们随 Kuffler 到哈佛医学院药理系，1966 年哈佛建立由 Kuffler 主持的世界上第一个神经生物学系，主要以 Kuffler 从霍普金斯带到哈佛的人为主，用生理、生化等多学科研究神经生物学问题。

注 10: Hubel 和 Wiesel 有一系列视觉发育的研究（Wiesel and Hubel，1963a,1963b ,1965a, 1965b; Hubel and Wiesel,1963b, 1965, 1970），一个主要发现是视觉发育的关键期，在关键期没有视觉经历对视觉的影响特别大，而关键期之前或者之后闭眼影响就小很多。

参考文献:

Arshavsky VY, Lamb TD, Pugh Jr EN (2002) G proteins and phototransduction. *Annual Review of Physiology* 64: 153-187.

Barlow HB (1953) Summation and inhibition in the frog's retina. *Journal of Physiology* 119:69-88.

Barlow HB, FitzHugh R and Kuffler SW (1957) Change of organization in the receptive fields of the cat's retina during dark adaptation. *Journal of Physiology* 137:338-354.

Baylor DA, Lamb TD and Yau KW (1979) Responses of retinal rods to single photons. *Journal of Physiology* 288:613-634.

Belrhali H, Nollert P, Royant A, Menzel C, Rosenbusch JP, Landau EM and Pebay-Peyroula E (1999). Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* 7:909-917.

Berson DM, Dunn FA and Takao M (2002). Phototransduction by retinal ganglion cells that set the circadian clock. *Science* 295:1070-1073.

Böll F (1877) Zur Anatomie und Physiologie der Retina. *Arch. Anat. Physiol. Physiol. Abt*:4-35.

Dalton J (1798). Extraordinary facts relating to the vision of colours with observations. *Memoirs of the Literary and Philosophical Society of Manchester* 5:28-45.

Davis PW, Berman AL and Mountcastle VB (1955) Functional analysis of the first somatic area of the cat's cerebral cortex in terms of activity of single neurons. *American Journal of Physiology* 183:607.

Dowling JE（2002）George Wald 18 November 1906-12 April 1997. *Proceedings of the American Philosophical Society* 146:431-439.

Fesenko EE, Kolesnikov SS and Lyubarsky AL (1985) Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment. *Nature* 313:310 –313.

Ferster D, Chung S and Wheat H (1996) Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380:249−252.

Fridericia LS and Holm E (1925). Experimental contribution to the study of the relation between night blindness and malnutrition. *American Journal of Physiology* 73:63–78.

Granit R (1941) Isolation of colour-sensitive elements in a mammalian retina. *Acta Physiologica Scandinavia* 3:318-328.
Granit R (1943) The spectral properties of the visual receptors of the cat. *Acta Physiologica Scandinavia* 5:21-22.
Granit R (1945) The colour receptors of the mammalian retina. *Journal of Neurophysiology* 8:197-210.
Hargrave PA, McDowell JH, Sremiatkowski-Juszczak EC, Fong S-L, Kuhn H, Wang JK, Curtis DR, Mohana Rao JK, Argos P, and Feldmann RJ (1982). The carboxy terminal one-third of bovine rhodopsin: its structure and function. *Vision Research* 22:1429-1438.
Hargrave PA, McDowell JH, Curtis DR, Wang JK, Juszczak E, Fong SL, Rao JK, Argos P (1983) The structure of bovine rhodopsin. *Biophysics of Structure and Mechanism* 9:235–244.
Hartline HK and Graham CH (1932) Nerve impulses from single receptors in the eye. *Journal of Cellular and Comparative Physiology* 1:227-295.
Hartline HK (1938) The response of single optic nerve fibbers of the vertebrate eye to illumination of the retina. *American Journal of Physiology* 121:400-415.
Hartline HK (1940a) The receptive fields of optic nerve fibers. *American Journal of Physiology* 130:690-699.
Hartline HK (1940b). The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. *American Journal of Physiology* 130:700-711.
Hartline HK (1942) The neural mechanisms of vision. *Harvey Lectures* 37:39-68.
Hattar S, Liao HW, Takao M, Berson DM and Yau KW (2002). Melanopsin containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* 295:1065-1070.
Hecht S (1919). Sensory equilibrium and dark adaptation in Mya arenaria. *Journal of General Physiology* 1:545-558.
Hecht S, Shlaer S, Pirenne M (1942) Energy, quanta and vision. *Journal of General Physiology* 25:819–840.
von Helmholtz H (1910) Treatise on Physiological Optics, translated by Southall JPC (1924), Vol. 2 the sensations of vision. The Optical Society of America.
Horiguchi H, Winawer J, Dougherty RF, Wandell BA (2012). Human trichromacy revisited. *Proceedings of the National Academy of Sciences USA* 110:260-269.
Hubel DH (1957) Tungsten microelectrode for recording from single units. *Science* 125:549-550.
Hubel DH (1959) Single unit activity in striate cortex of unrestrained cats. *Journal of Physiology* 147:226-238.
Hubel DH (1996) A big step along the visual pathway. *Nature* 380:197–198.
Hubel DH and Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* 148:574-91.
Hubel DH and Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160:106-154.
Hubel DH and Wiesel TN. (1963a) Shape and arrangement of columns in cat's striate

cortex. *Journal of Physiology* 165:559-68.
Hubel DH and Wiesel TN (1963b) Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology* 26:994-1002.
Hubel DH and Wiesel TN (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology* 28:229-289.
Hubel DH and Wiesel TN (1968) Receptive fields and functional archicture of monkey striate cortex. *Journal of Physiology* 195:215-243.
Hubel DH and Wiesel TN (1970) The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *Journal of Physiology* 206:419-36.
Hubel DH and Wiesel TN (1998). Early exploration of the visual cortex. *Neuron* 20:401-12.
Hubel DH and Wiesel TN (2005). Brain and Visual Perception: the story of a 25-year collaboration. Oxford University Press.
Hunt DM, Dulai KS, Bowmaker JK, Mollon JD (1995). The chemistry of John Dalton's color blindness. *Science* 267:984 – 988.
Jia H, Rochefort NL, Chen X and Konnerth A (2010) Dendritic organization of sensory input to cortical neurons in vivo. *Nature* 464:1307-1312.
Jung R (1953) Neuronal discharge. *Electroencephalography and Clinical Neurophysiology Suppl* 4:57-71.
Jung R (1958) Excitation, inhibition and coordination of cortical neurones. *Experimental Cell Research Suppl* 5:262-271.
Jung R and Baumgartner G (1955) Hemmungsmechanismen und bremsende Stabilisierung an einzelnen Neuronen des optischen Cortex. *Pflüger Archiv für die gesamte Physiologie des Menschen und der Tiere* 261:434-456.
Kimura, Y., Vassylyev, D.G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murate, K., Hirai, T. & Fujiyoshi, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* 389:206-211.
Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology* 16:37-68.
Kühne W. (1879) Chemische Vorgänge in der Netzhaut. Hermann L. eds. Handbuch d. Physiologie d. Sinnesorgane Erster Theil, Gesichtssinn. F.C.W. Vogel Leipzig, Germany.
Kühne W (1882) Beiträge zur Optochemie. Untersuchungen aus dem physiologischen Institute der Universität Heidelberg 4:169-249.
Lorente De Nó R (1949) Cerebral cortex: architecture, intracortical connections, motor projections. Chap. 15 in: FULTON, J. F., Physiology of the nervous system, 3rd ed. New York and London, Oxford University Press.
Marks WB, Dobelle WH and MacNichol EF (1964) Visual pigments of single primate cones. *Science* 143:1181-1183.
Maxwell JC (1872) On colour vision. *Proceedings of the Royal Institution of Great Britain* 6:260-271.
McCollum EV and Davis M (1913) The necessity of certain lipins in the diet during growth. *Journal of Biological Chemistry* 15:167–175.

McCollum EV and Davis M (1915) The nature of the dietary deficiencies of rice. *Journal of Biological Chemistry* 23:181–230.

Mountcastle VB (1957) Modality and topographic properties of single neurons of cat‘s somatic sensory cortex. *Journal of Neurophysiology* 20:408-434.

Mountcastle VB (1997) The columnar organization of the neocortex. *Brain* 120:701–722.

Mountcastle VB, Berman AL and Davis PW (1955) Topographic organization and modality representation in the first somatic area of the cat’s cerebral cortex by method of single unit analysis. *American Journal of Physiology* 183:646.

Müller R (1851). Zur Histologie der Netzhaut. *Z. Wiss. Zool*. 3:234-237.

Nathans J and Hogness DS (1983) Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin *Cell* 34:807-814.

Nathans J and Hogness DS (1984) Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proceedings of the National Academy of Sciences USA* 81:4851-4855.

Nathans J, Thomas D, Hogness DS (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232:193-202.

Nathans J, Piantanida TP, Eddy RL, Shows TB and Hogness DS (1986) Molecular genetics of inherited variation in human color vision. *Science* 232:203-210.

Newton I (1704) Opticks.

Osborne TB and Mendel LB (1913) The relation of growth to the chemical constituents of the diet. *Journal of Biological Chemistry* 15:311-326.

Ovchinnikov Iu A, Abdulaev NG, Feigina M, Artamonov ID, Bogachuk AS (1983) Visual rhodopsin. III. Complete amino acid sequence and topography in a membrane. *Bioorganicheskaia Khimiia* 9:1331–1340.

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289:739–745.

Palmer G (1777a) Theory of Colour and Vision, London: Leacroft.

Palmer G (1777b) Theorie du couleur et de la vision, translated by D.-B. Quatremere d'Isjonval, Paris: Pissot.

Peters A and Sethares C (1996) Myelinated axons and the pyramidal cell modules in monkey primary visual cortex. *Journal of Comparative Neurology* 365:232-255.

Provencio I, Jiang G, De Grip WJ, Hayes WP, Rollag MD (1998) Melanopsin: an opsin in melanophores, brain, and eye. *Proceedings of the National Academy of Sciences USA* 95:340–345.

Provencio I, Rodriguez IR, Jiang G, Hayes WP, Moreira EF and Rollag MD (2000) A novel human opsin in the inner retina. *Journal of Neuroscience* 20:600-605.

Rakic P (1972) Mode of cell migration to the superficial layers of fetal monkey neocortex. *Journal of Comparative Neurology* 145:61–83.

Rakic (1995) A small step for the cell, a giant leap for mankind: a hypothesos of neocortical expansion during evolution. *Trends in Neuroscience* 18:383-388.

Reid RC and Alonso J-M (1995) Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* 378:281−284.

Ripps H (2008) The color purple: milestones in photochemistry. *FASEB Journal* 22:4038-4043.

Senthilan PR, Piepenbrock D, Ovezmyradov G, Nadrowski B, Bechstedt S, Pauls S, Winkler M, Möbius W, Howard J, Göpfert MC (2012) Drosophila auditory organ genes and genetic hearing defects. *Cell* 150:1042-1054.

Smith SO (2010). Structure and activation of the visual pigment rhodopsin annual review of biophysics. *Annual Review of Biophysics* 39:309-328.

Snyder SH (2015) Vernon B. Mountcastle 1918-2015. *Nature Neuroscience* 18:318.

Sperry RW, Minor N, Myers RE (1955) Visual pattern perception following subpial slicing and tantalum wire implantations in the visual cortex. *Journal of Comparative Physiology and Psychology* 48:50-58.

Talbot SA and Kuffler SW (1952). A multibeam ophthalmoscope for the study of retinal physiology. *Journal of the Optical Society of America* 42:931-936.

van der Velden HA (1946) The number of quanta necessary for the perception of light in the human eye. *Opthalmologica* 111:321–331.

Wald G (1933) Vitamin A in the retina. *Nature* 132:316.

Wald G (1935) Vitamine A in eye tissue. *Journal of General Physiology* 18 :905-915.

Wald G (1935) Carotenoids and the visual cycle. *Journal of General Physiology* 19:351-371.

Wald G. (1968). The molecular basis of visual excitation. *Nature* 219:800-807.

Walton GE, Bower NJA and Bower TGR (1992) Recognition of familiar faces by newborns. *Infant Behavioral Development* 15:265-269.

Wiesel TN and Hubel DH (1963a) Effects of visual deprivation on morphology and physiology of cells in the cats lateral geniculate body. *Journal of Neurophysiology* 26:978-93.

Wiesel TN and Hubel DH (1963b) Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology* 26:1003-1017.

Wiesel TN, Hubel DH (1965a) Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *Journal of Neurophysiology* 28:1029-40.

Wiesel TN, Hubel DH (1965b) Extent of recovery from the effects of visual deprivation in kittens. *Journal of Neurophysiology* 28:1060-72.

Wolf G (2001). The discovery of the visual function of Vitamin A. *Journal of Nutrition* 131:1647-1650.

Young T (1802) On the Theory of Light and Colours. *Philosophical Transactions of the Royal Society of London* 92:12-48.

Yudkin AM (1931) The presence of vitamin A in the retina. *Archive of Ophthalmology* 6:510–517.

Molecular Genetics of Inherited Variation in Human Color Vision

JEREMY NATHANS, THOMAS P. PIANTANIDA, ROGER L. EDDY, THOMAS B. SHOWS, DAVID S. HOGNESS

The hypothesis that red-green "color blindness" is caused by alterations in the genes encoding red and green visual pigments has been tested and shown to be correct. Genomic DNA's from 25 males with various red-green color vision deficiencies were analyzed by Southern blot hybridization with the cloned red and green pigment genes as probes. The observed genotypes appear to result from unequal recombination or gene conversion (or both). Together with chromosome mapping experiments, these data identify each of the cloned human visual pigment genes.

MOST HUMANS CAN MATCH ANY COLOR EITHER BY COMbining three suitably chosen primary colors or by combining two primaries and adding the third primary to the given color. For additive color mixture, such as when lights are mixed, the primaries are red, green, and blue. Thomas Young, 180 years ago, put forward the hypothesis that this phenomenon (trichromacy) is a consequence of humans having three independent light-sensitive mechanisms (*1*). We now know that Young's three mechanisms are embodied in three classes of cone photoreceptor cells in the human retina. Each class contains a different visual pigment, which determines the spectral sensitivity of all the cones of that class.

Most humans agree on the proportions of the three primaries required to match a given color. Among those who differ from color normals with respect to the proportions of the primaries, some require that the three be present in unusual proportions, and others require only two primaries. Individuals with the first type of variation are called anomalous trichromats and are presumed to have three classes of cones, one of which contains a photopigment with an anomalous absorption spectrum. Those with the second type of variation are called dichromats and are presumed to have only two of the three classes of cones. These types of variation can be further subdivided by psychophysical tests into classes whose color vision variation is attributable to alterations in either red, green, or blue cones. (As is discussed below, the blue cone defects are so rare that most of the data, especially with respect to subtypes of variation, comes from studies of the red and green cone sensitivities.)

The heritability of variations in color sensitivity has long been recognized (*2*). It is now clear that one locus is responsible for variations in red cone sensitivity and that a second locus is responsible for variations in green cone sensitivity (*3*). Both loci map to the distal part of the q arm of the X chromosome and are tightly linked to each other and to glucose-6-phosphate dehydrogenase (G-6-PD) (*4*). Variations in the blue cone sensitivity have recently been shown to segregate in an autosomal fashion (*5*). Among the red and green variants, available evidence points to allelism of those traits that affect a given cone type. However, a true complementation test (requiring expression of both alleles in the same cell) is not possible because each cell in a female expresses only one of her two X chromosomes (*6*). The evidence for allelism rests instead on the lack of recombination between two defects affecting a single cone type (*4, 5, 7*). It is possible to define a "dominance" hierarchy among alleles by observing the phenotypes of heterozygous women; in each case the allele that least diminishes color discrimination is "dominant" (*7*).

These data are consistent with a model in which the loci responsible for inherited variations in color vision correspond to the genes that encode the apoproteins of the three cone pigments. In support of this model, Rushton, as well as Alpern and Wake, have measured visual pigment absorption in the living human eye by reflection densitometry and found that dichromats lack one of the cone photopigments (*8*). More recently, this result has been confirmed and extended by microspectrophotometric analysis of single human cones from normal and dichromat retinas (*9*).

We have performed a direct test of the hypothesis that inherited variations in human color vision are caused by alterations in the genes that encode the red, green, and blue visual pigments. Our strategy has been to isolate these genes as recombinant DNA molecules and compare their structures among normal and mutant individuals. The X linkage of red and green defects simplifies this analysis. In males, only those genes resident on a single X chromosome are seen in Southern blots or in cloning experiments, and only the phenotype resulting from that X chromosome is observed in psychophysical experiments. This phenotype can be measured in a sample noninvasive test. Moreover, the incidence of red and green color vision variation is quite high—approximately 8 percent among Caucasian males.

We have described (*10*) the isolation and characterization of the three human cone pigment genes. We referred in (*10*), and do so here, to these genes by their true identities—that is, blue, green, and red pigment genes; however, this assignment requires the data presented below. We showed (*10*) that the red and green pigment genes were extraordinarily similar in DNA sequence (98 percent identity). This degree of homology and the high incidence of variation in green pigment gene number among color normal males, led us to propose a model in which red and green pigment genes reside in a head-to-tail tandem array.

Chromosomal locations of cloned visual pigment genes. As a

Jeremy Nathans and David S. Hogness are in the Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305. Thomas P. Piantanida is in the Sensory Sciences Research Laboratory, S.R.I. International, 333 Ravenswood Avenue, Menlo Park, CA 94025. Roger L. Eddy and Thomas B. Shows are in the Department of Human Genetics, Rosewell Park Memorial Institute, 666 Elm Street, Buffalo, NY 14263. Address reprint requests to David S. Hogness.

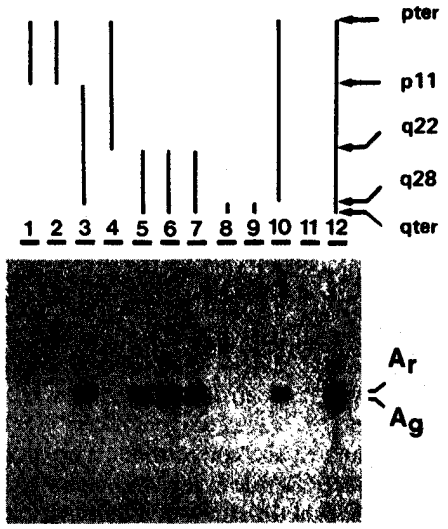


Fig. 1. Subchromosomal localization of green and red pigment genes. Shown is a Southern blot autoradiograph of Eco RI–digested DNA prepared from a panel of ten mouse-human hybrid cell lines (lanes 1 to 10), a mouse cell line (RAG, lane 11), and a human cell line (WI-38, lane 12). Each line has that fraction of the human X chromosome shown above the autoradiograph. The probe hybridizes to both green and red pigment genes. The gene mapping strategy and isolation of mouse-human hybrid cell lines have been described (*11*). Each of the mouse-human hybrids has a known X-autosome translocation as determined by karyotypic analysis, isozyme markers, and the HAT (hypoxanthine, aminopterine, thymidine) selection system (*11*). They are: lanes 1 and 2, a 15 to X translocation retaining Xpter-Xp11; lane 3, an X to 15 translocation retaining Xp11-Xqter; lane 4, an X to 11 translocation retaining Xpter-Xq22; lane 5, a 5 to X translocation retaining Xq22-Xqter; lanes 6 and 7, a 22 to X translocation retaining Xq22-Xqter; lanes 8 and 9, a 3 to X translocation retaining Xq28-Xqter, and lane 10, an X to 3 translocation retaining Xpter-Xq28. Hybridization methods are described in the legend to Fig. 3.

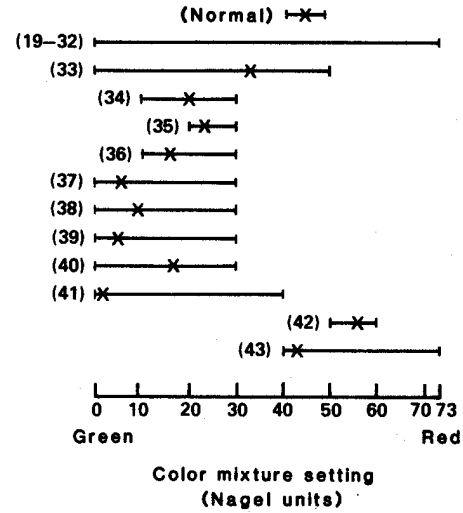


Fig. 2. Anomaloscope test data for 25 color variant males. Each subject is represented by a number next to the horizontal bar showing the range of ratios of red to green lights acceptable as a match of the standard yellow light. The midpoints of the free matching range are marked by crosses; that is, those red to green ratios chosen by the subject when he was free to adjust both the relative intensities of red and green lights as well as the absolute intensity of the standard yellow light. Subjects 19 to 32 accept any red to green ratio and are, therefore, dichromats. They were divided into G^-R^+ (19 to 26) or G^+R^- (27 to 32) classes on the basis of their relative sensitivities to the red primary. Subject 33 appears to show some residual discrimination; he is either $G''R^+$ or G^-R^+. Because he is missing all green pigment genes (Figs. 3 and 4) we classify him as G^-R^+. Subjects 34 to 36 are $G'R^+$; subjects 37 to 41 are $G''R^+$; subject 42 is G^+R'; and subject 43 is G^+R''.

first step in testing the hypothesis relating color blindness loci and visual pigment genes, we mapped the chromosomal locations of the three human cone pigment genes, as well as that of the gene encoding the rod pigment, rhodopsin. In one experiment a DNA probe derived from a green pigment gene [the rightmost 4 kilobases (kb) of clone gJHN9; see figure 5A in (*10*)] was hybridized to a Southern blot filter of Eco RI digested DNA from a panel of mouse-human hybrid cell lines that retain different human chromosomes. The probe is homologous to both red and green pigment genes and revealed the expected 9-kb size class of hybridization only in those lanes containing DNA from cell lines that retain the human X chromosome. In a second experiment we observed loss of that hybridization concomitant with loss of the human X chromosome from one of the cell lines. To map the subchromosomal locations of these sequences, we probed a Southern blot of Eco RI–digested DNA prepared from a panel of mouse-human hybrid cell lines that retain different parts of the human X chromosome. The blot was hybridized with a probe encompassing the second exon of a green pigment gene [contained in clone gJHN21; figure 5B in (*10*)]. This probe hybridizes with two Eco RI fragments: A_g from the green pigment genes and A_r from the red pigment gene (Fig. 3A). Because the ratio of green to red pigment genes varies among normal males (*10*) and the hybrids were generated by fusion of cells from different human donors (*11*), the ratio of A_g to A_r is expected to vary among the hybrids. Figure 1 shows that A_g and A_r segregate together in each case; they also segregate with human G-6-PD activity, and with the q22-q28 interval (*11*). This pattern matches that predicted for the loci responsible for variations in red and green color sensitivity and provides the first evidence for the identification of the red and green pigment genes. These experiments do not, however, allow us to determine which of the X-linked genes corresponds to the red locus and which to the green locus.

The gene encoding the human rod photopigment, rhodopsin (RHO) (*12*), was mapped to the third chromosome by Southern blotting of a panel of mouse-human hybrids that retain various fractions of the human karyotype (Table 1). This gene is present in a cell hybrid (XTR-22) that retains the q21-qter region of chromosome 3 as part of an X to 3 translocation. Therefore the human rhodopsin gene resides within the interval 3q21-3qter. The same panel was used to map the remaining visual pigment gene [shown in figure 1 in (*10*)] to chromosome 7. Given that this is the only visual pigment gene that maps to an autosome aside from rhodopsin, we presume that it encodes the blue pigment, consistent with the autosomal nature of inherited variation in blue sensitivity. The blue pigment gene (BCP; that is, blue cone pigment) is absent from a cell hybrid (JSR-17S) that retains the pter-q22 region of chromosome 7 as part of a 7 to 9 translocation. This gene therefore resides in the 7q22-7qter interval.

Identification of red and green pigment genes. To correlate genotype with phenotype, we examined 25 males with various forms of red or green color vision variation. Psychophysical test data obtained from these males are shown in Fig. 2. Briefly, the test (which measures only red and green cone sensitivities) consists of presenting a variable mixture of red and green lights on one half of a screen, and a variable intensity yellow light on the other half. The subject adjusts the ratio of the red and green lights and the intensity of the yellow light to produce a perfect match between the two halves of the screen. When the red-green ratio and the yellow intensity are adjusted to produce a match, the number of photons captured per second by red and green pigments is the same from both halves of the screen. The midpoints and ranges of this red-green ratio for 25 male subjects is shown in Fig. 2. All of the major groups of red and green color vision variation are represented.

Southern blots for three different restriction digests of genomic DNA from 15 dichromats and a color-normal control are shown in Fig. 3. The probes used for these blots hybridize only with the X-

linked loci under these conditions. It is immediately apparent that 14 out of 15 samples differ in the fragment pattern for at least one of the restriction digests from that in the color-normal controls [see figure 8 in (*10*)]. It is also apparent that more than one genotype can correspond to a given phenotype: the nine G^-R^+ (G^-, absent green sensitivity; R^+, normal red sensitivity) subjects fall into two groups, whereas the six G^+R^- (G^+, normal green sensitivity; R^-, absent red sensitivity) subjects are all different from each other (*12a*). These hybridization patterns suggest that gross changes in DNA rather than point mutation have produced at least 14 of these 15 mutant genotypes.

We consider first the assignment of gene identities based upon these data. We showed (*10*) that restriction fragments A_r, B_r, C_r, and D_r shown in Fig. 3A are all derived from one gene and that color-normal males all have the same copy number of this gene, which is probably one; by contrast restriction fragments A_g, B_g, C_g, and D_g derive from a different gene that varies in number among color-normal males. None of the nine G^-R^+ subjects have D_g and three of nine G^-R^+ subjects have A_g, B_g, and C_g (Fig. 3). However, all of the G^-R^+ subjects retain A_r, B_r, C_r, and D_r. Thus A_r, B_r, C_r, and D_r are associated with a functioning red mechanism, whereas A_g, B_g, C_g, and D_g are not. Indeed, absence of A_g, B_g, and C_g correlates partially, and absence of D_g correlates perfectly with absence of the green mechanism.

The situation among G^+R^- subjects is more complex. Four of six subjects have A_r, all six have B_r, four of six have C_r, and only one of six has D_r. However, every G^+R^- subject (6/6) has D_g, although A_g, B_g, and C_g are not always present (5/6, 4/6, and 5/6, respectively). We can summarize these data and those presented below from anomalous trichromats (as well as those from normals) by listing the following true statements: (i) G^+ is always associated with D_g, and usually, but not always, with A_g, B_g, and C_g; (ii) R^+ is always associated with A_r, B_r, C_r, and D_r; (iii) G^- is always associated with loss of D_g and sometimes with the combined loss of A_g, B_g, and C_g; (iv) R^- is usually associated with loss of D_r, sometimes with loss of A_r or C_r, and never with loss of B_r, and, in fact, (v) B_r is never lost. Taken together these associations imply that A_g, B_g, C_g, and D_g derive from the normal green pigment genes and that A_r, B_r, C_r, and D_r derive from the normal red pigment gene. Furthermore, the 5′ end of the red pigment gene, represented by fragment B_r (Fig. 3C),

Table 1. Chromosomal assignment of the rhodopsin gene and the blue pigment gene. Probes from each gene were hybridized to Southern blots of DNA from mouse-human hybrid cell lines. The 30 cell hybrids involve 14 unrelated human cell lines and four mouse cell lines (*11*). They were characterized by analysis of karyotype, mapped enzyme markers, and mapped cloned DNA probes (*11*). Presence (+) or absence (−) of a human chromosome is indicated. (+/+), (+/−), (−/+), (−/−): the first symbol within the parentheses indicates the presence (+) or absence (−) of the human gene; the second symbol within the parentheses indicates the presence (+) or absence (−) of the indicated chromosome. Percent discordancy indicates the percent discordant segregation for a probe and a chromosome. The chromosome which shows no (0 percent) discordancy with a human gene probe is the one from which that gene derives. Hybrid XTR-22 retains the 3q21-3qter region as part of an X to 3 translocation and, because the human rhodopsin gene is also retained, localizes that gene to this interval. Hybrid JSR-175 retains the 7pter-7q22 region and, because it does not retain the blue pigment gene, localizes that gene to the 7q22-7qter interval.

			Human Chromosomes																							
HYBRID	Rhodopsin	Blue Pigment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Translocations
ATR-13	+	+	+	+	+	+	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	−	−	−	t	5/X
DUA-3BSAGA	−	+	−	+	−	−	−	−	+	+	−	−	−	−	+	+	−	−	+	−	−	−	−	−	−	
DUA-5BSAGA	+	−	−	−	+	−	+	−	−	−	−	−	+	−	−	+	−	−	+	+	−	−	+	−	−	
DUM-13	+	+	+	+	+	−	+	+	+	−	−	+	+	+	−	+	t	+	+	+	+	+	+	+	t	X/15 15/X
GAR-1	+	−	−	−	+	−	+	−	−	+	−	+	−	+	−	+	+	+	−	−	−	+	−	−	+	
ICL-15	−	−	−	−	−	−	−	−	−	+	−	−	−	+	−	−	−	−	+	−	−	+	+	−	−	
JSR-14	+	−	−	+	+	+	+	+	−	−	−	−	−	+	+	−	−	−	+	−	−	+	+	−	+	
JSR-175	+	−	+	+	+	−	+	−	t	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	7/9
JWR-26C	+	+	t	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	+	+	−	+	+	−	+	1/2
NSL-9	−	−	−	−	−	+	−	−	+	t	+	−	+	+	+	+	+	+	+	−	−	+	+	+	−	17/9
NSL-16	+	+	−	−	+	+	+	−	+	+	t	+	−	+	−	+	+	+	+	+	−	+	+	−	−	17/9
REW-7	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	+	+	+	+	+	+	
REW-11	−	+	−	−	−	+	−	−	+	−	−	−	+	+	+	−	−	+	−	−	−	+	+	+	+	
REX-11BSAgB	+	−	−	−	+	−	−	−	−	−	−	+	−	−	−	+	+	−	−	+	−	−	−	−	−	
REX-11BSHF	+	−	−	−	+	−	−	−	−	−	−	+	−	−	−	+	−	−	−	+	−	−	−	t	t	22/X
REX-26	+	+	+	+	+	−	−	−	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	t	t	22/X
SIR-8	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	
TSL-1	+	−	−	+	+	−	−	−	−	+	+	+	−	+	+	−	+	+	+	+	−	+	+	−	−	
VTL-6	−	+	−	+	−	−	−	+	+	+	−	+	+	−	−	−	+	−	+	−	+	+	+	+	−	
VTL-8	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−	+	−	−	+	+	+	−	
VTL-17	−	+	−	−	−	−	+	−	+	−	−	+	+	−	+	+	−	−	+	−	−	+	+	−	−	
WIL-2	−	−	−	−	−	−	−	−	−	+	−	−	−	+	−	−	+	−	+	−	−	−	+	−	+	
WIL-6	−	+	−	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	+	−	+	+	+	−	+	
WIL-7	+	−	−	+	+	+	+	−	+	−	−	+	+	−	+	+	−	−	+	+	−	+	+	−	+	
WIL-8	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
WIL-8X	+	+	−	−	+	+	+	−	+	+	−	+	+	+	−	+	−	−	+	+	+	+	+	−	+	
WIL-14	+	+	+	−	+	−	+	+	+	−	−	+	−	+	−	+	+	−	+	−	−	−	−	−	+	
WIL-15	+	+	−	+	+	+	−	+	+	−	−	+	+	+	+	+	+	−	+	+	−	+	+	−	+	
XER-11	+	+	+	−	+	+	−	+	+	+	+	+	t	+	+	−	+	+	+	+	+	+	+	+	t	11/X X/11
XTR-22	+	−	−	+	t	+	+	+	−	+	−	+	+	−	−	+	−	−	+	+	+	+	+	+	+	X/3

Rhodopsin		Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
	(+/+)		9	12	20	13	15	10	12	13	5	19	13	15	10	18	14	11	17	18	8	13	16	7	11
	(+/−)		11	9	0	8	6	11	8	8	15	2	7	6	11	3	6	10	4	3	13	8	5	12	5
	(−/+)		8	3	0	2	3	2	5	6	8	4	4	4	5	4	4	2	8	0	2	7	8	4	3
	(−/−)		9	6	9	7	6	7	4	3	8	5	5	5	4	5	5	7	1	9	7	2	1	5	6
		% Discordancy	38	40	0	33	30	43	45	47	54	20	38	33	53	23	34	40	40	10	50	50	43	57	32
Blue Pigment		Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
	(+/+)		8	11	12	12	11	9	17	12	3	15	12	13	9	14	11	9	16	11	9	12	14	7	9
	(+/−)		8	6	5	5	6	8	0	5	13	2	4	4	8	3	5	8	1	6	8	5	3	9	4
	(−/+)		1	4	8	3	7	3	0	7	2	8	5	6	6	8	7	4	9	7	1	8	10	4	5
	(−/−)		12	9	4	10	6	10	12	6	10	5	8	7	7	5	6	9	4	6	12	5	3	8	7
		% Discordancy	31	33	45	27	43	37	0	40	54	33	31	33	47	37	41	40	33	43	30	45	43	46	36

NOTE: "t" in the table indicates a translocation, no intact chromosome.

appears to occupy a privileged position—it is neither duplicated nor deleted. We show below that the complexity of these genotypes, especially G^+R^-, can be explained by proposing that they arose via intragenic recombination. Thus, green function and red function correlate best with just a fraction of each gene, D_g and D_r, respectively. We infer that the region corresponding to the D_g compared to D_r difference (in the fifth exon) is tightly linked to sequence differences that determine the spectral absorbance of the cone photopigments.

Molecular models for color variant genotypes. The analysis presented above relied only on scoring each restriction digest for the presence or absence of hybridizing fragments. Another result of these experiments is that none of the blots have hybridization bands at positions other than those seen in normal DNA. Moreover, the Eco RI digest shown in Fig. 3B was also probed with the 3′ proximal two-thirds of a red pigment complementary DNA (cDNA) clone [a Bam HI–Eco RI fragment from hs7; figure 5C in (*10*)]. This probe encompasses the 3′ half of exon 2 and all of exons 3 to 6; it therefore hybridizes strongly to a 9-kb Eco RI fragment and weakly to 9.2-kb (A_g) and 11.1-kb (A_r) Eco RI fragments (Fig. 3A). No fragments other than these are seen following hybridization of this probe to the blots shown in Fig. 3B. An analogous result was obtained following hybridization to Eco RI–digested anomalous trichromat DNA. The above data suggest that the observed DNA rearrangements involve homologous crossing-over or gene conversion (or both) because nonhomologous rearrangements produce restriction fragments that usually differ in size from those of unrearranged DNA. Generation of these homologous rearrangements is not surprising in light of the high degree of sequence homology (98 percent) throughout almost the entire length of the

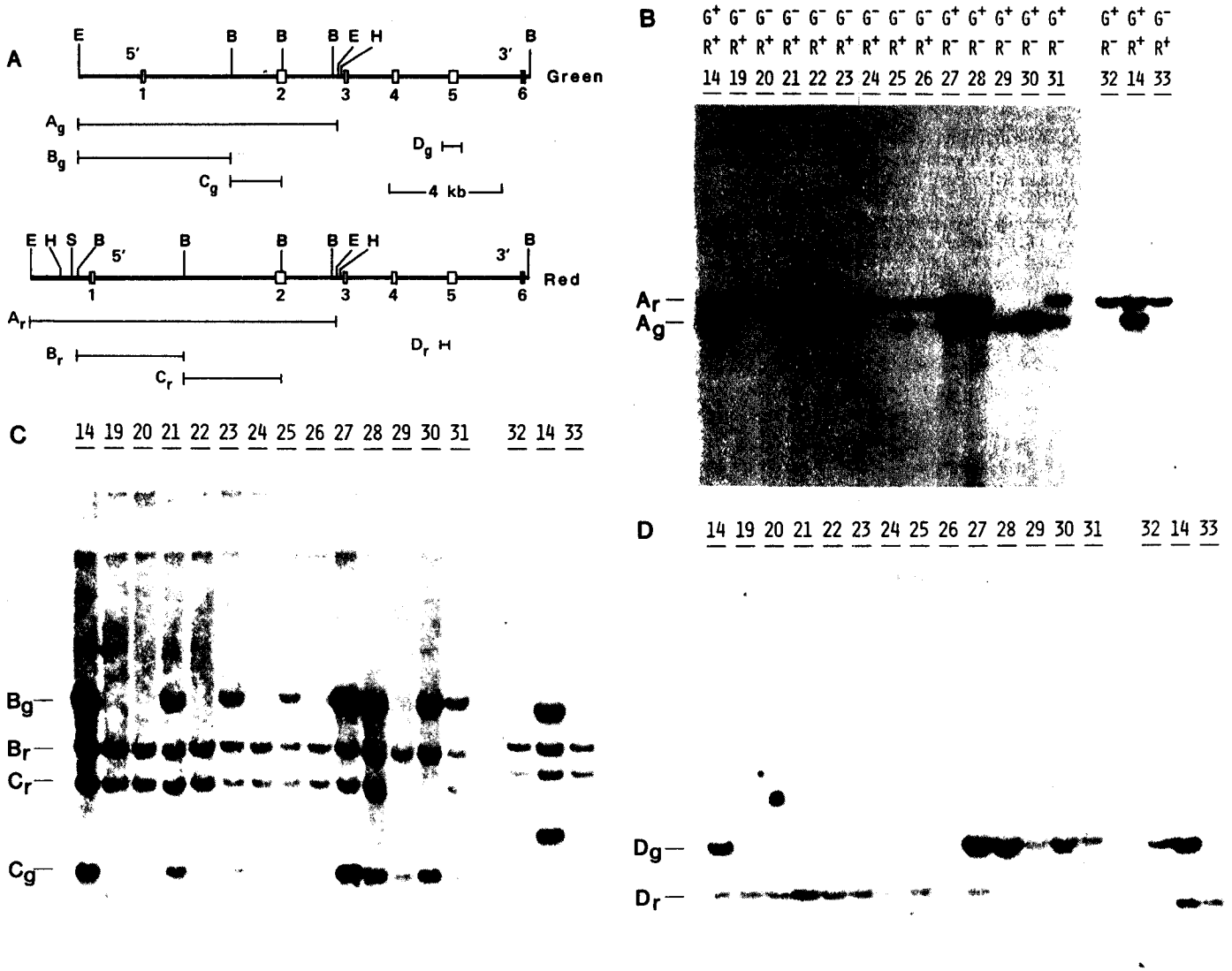


Fig. 3. Genomic Southern blots of DNA from 15 dichromats. (A) Restriction maps of red and green pigment genes showing fragments visualized by Southern blotting. The probes used to visualize these fragments are more than 98 percent identical to both genes. A_g and A_r are two Eco RI fragments derived from green and red pigment genes, respectively, and are visualized with a probe from exon 2. B_g and B_r, and C_g and C_r are fragments resulting from Bam HI and Eco RI double digestion and are visualized with a probe encompassing exon 1 and the 5′ half of exon 2. D_g and D_r are fragments resulting from Rsa I digestion which share a common left border but differ on their right borders because the green pigment gene lacks but the red pigment gene has an Rsa I site in the fifth exon. They are visualized with a probe from the 3′ end of the fourth intron. (B to D) Three pairs of genomic Southern blots of DNA from 15 dichromats and one color normal. Each number and phenotype refers to the individual whose DNA is in that lane and whose anomaloscope test data are shown in Fig. 2. (B) Eco RI, (C) Eco RI and Bam HI, (D) Rsa I. Approximately 10 μg of DNA, digested with the indicated enzymes, was placed on each lane. Filters were hybridized and washed under standard conditions (*18*) and exposed to preflashed x-ray film at −70°C with an intensifying screen.

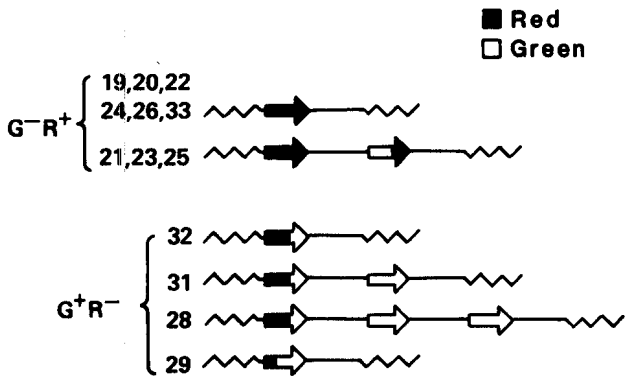


Fig. 4. Proposed arrangement of green and red pigment genes in dichromats. Each arrow represents a single gene: the base corresponds to the 5′ end and the tip to the 3′ end. Zigzag lines represent single-copy flanking DNA and thin lines represent homologous intergenic sequences. The genotypes and subject numbers represented by each diagram are at the left. Presented here, as an example, are the data and methods used to deduce the genotypes of subjects 21, 23, and 25. Peak areas were measured by scanning the autoradiographs shown in Fig. 3, B, C, and D, and ratios of these areas were calculated (Table 2). Subjects 21, 23, and 25 had A_g:A_r ratios of 1.10, 1.21, and 1.22, respectively; B_g:B_r ratios of 1.10, 1.69, and 1.35, respectively; C_g:C_r ratios of 0.69, 0.92, and 0.72, respectively; and D_g:D_r ratios of 0 for all three. For comparison, the peak area ratios corresponding to a 1:1 gene ratio have been calculated from the Southern blots of 18 color-normal males [shown in (*10*) figures 8 and 9]. These calculations show that the 18 color-normal males have a total of 18 red pigment genes and 37 green pigment genes and that the peak area ratios (mean ± SD) corresponding to a 1:1 gene ratio are as follows A_g:A_r, 0.97 ± 0.06; B_g:B_r, 2.00 ± 0.49; C_g:C_r, 0.88 ± 0.18; D_g:D_r, 1.24 ± 0.30. The A_g:A_r ratios cluster at nearly integral values because these two fragments appear to bind to the nitrocellulose and hybridize with nearly equal efficiencies. Fragments B_g and D_g appear to bind or hybridize better, respectively, than do the smaller B_r and D_r fragments. Therefore, the ratios B_g:B_r and D_g:D_r cluster at values greater than 1. Since subjects 21, 23, and 25 have no D_g with which to compare D_r, the copy number of D_r was estimated by normalizing each D_r peak area to the background smear present in that track and comparing this value to that of the color-normal standard. For subjects 21, 23, and 25 these values of D_r peak areas (arbitrary units) are: 3.43, 2.04, and 1.88, respectively. The areas of the smears (in different arbitrary units) are: 4.05, 2.50, and 1.75, respectively. Therefore, the normalized values of D_r are 0.85, 0.82, and 1.07, respectively. On the same blot the color normal control (subject 14), who has single copy of D_r, has a D_r peak area of 2.08 and a smear area of 3.85, giving a normalized D_r area of 0.54. As discussed in the text, the gene rearrangements appear to involve only homologous events. Hence, only complete genes are assumed to be present and the sum of the members of each fragment class—A, B, C, and D—are equal. Given these constraints, these data predict that subjects 21, 23, and 25 have one copy each of A_g, A_r, B_g, B_r, C_g, and C_r; two copies each of D_r; and no copies of D_g.

red and green pigment genes and the already documented propensity of the green pigment genes to vary in number among normal individuals (*10*).

By including in our analysis the relative intensities of the autoradiographic bands, we can calculate (using values from normal DNA as a standard) the stoichiometries of the various bands for each subject. From these stoichiometries and from the conclusion that only homologous events have occurred (see above), we can construct plausible models of the arrangements of these genes. Therefore, we measured peak areas and calculated ratios of peak areas for bands within a single track (A_g:A_r, B_g:B_r, C_g:C_r, and D_g:D_r) (Table 2); these data should be insensitive to variations in amount of DNA loaded per track and are the basis for our models. As a check on these data, we also estimated the amount of DNA loaded per track by scanning across the smear generated by reprobing the blots with labeled total human genomic DNA. [See (*10*); in some autoradiographs the background smear of hybridization produced by the pigment gene probe was sufficient for this purpose and reprobing was unnecessary.] These estimates of DNA loaded per track were used to normalize the measured peak areas (see legend to Fig. 4 for an example of these calculations). These models are not in all cases the only arrangements consistent with the data. In particular, the distance between genes and the relative order of the genes along the chromosome cannot be determined from these data. Ultimately, molecular cloning and DNA sequencing will be required to elucidate the precise structures of these variant genotypes.

Models for dichromacy. We consider first the dichromat genotypes. Quantitation of the hybridization patterns for the DNA from subjects 19, 20, 22, 24, 26, and 33 (all G^-R^+) indicates that they are identical within experimental error. All six subjects have a single red pigment gene and no green pigment gene (Fig. 4). We assume that these genotypes arose from a homologous but unequal exchange [such as the one in figure 10B in (*10*)]. In the absence of a green pigment gene those cells that were destined to become green cones may express the remaining red pigment gene by default. This general idea has been put forward in the past to account for the fact that dichromats have normal visual acuity (*13*) and an increase in the sensitivity of the unaffected mechanism [for example, increased green sensitivity in G^+R^- dichromats; (*14*)]. Subjects 21, 23, and 25, are also G^-R^+ and identical to one another within experimental error. These three subjects have one copy each of A_g and A_r, B_g and B_r, and C_g and C_r, two copies of D_r, and no D_g. Therefore, we propose that they have a single normal red pigment gene and a single hybrid gene in which the 5′ part of a green pigment gene (A_g, B_g, C_g) has been joined to the 3′ part of a red pigment gene (D_r) (Fig. 4). We further propose that the hybrid gene includes sufficient material derived from the red pigment gene that its spectral absorbance is identical or nearly identical to that of the red pigment.

Table 2. Ratios of Southern blot peak areas. For comparison the ratios that correspond to a 1:1 gene ratio are A_g:A_r, 0.97 ± 0.06; B_g:B_r, 2.00 ± 0.49; C_g:C_r, 0.88 ± 0.18; D_g:D_r, 1.24 ± 0.30 (see legend to Fig. 4). Symbols: inf, infinite, that is, the denominator is zero; *, probably too small due to Southern blot transfer artifact; **, probably too large due to Southern blot transfer artifact.

Subject genotype	Numbers	A_g:A_r	B_g:B_r	C_g:C_r	D_g:D_r
G^-R^+	19	0	0	0	0
G^-R^+	20	0	0	0	0
G^-R^+	21	1.10	1.10	0.69	0
G^-R^+	22	0	0	0	0
G^-R^+	23	1.21	1.69	0.92	0
G^-R^+	24	0	0	0	0
G^-R^+	25	1.22	1.35	0.76	0
G^-R^+	26	0	0	0	0
G^+R^-	27	6.08	5.40	3.77	5.87
G^+R^-	28	1.95	3.02	1.08*	inf
G^+R^-	29	inf	0	inf	inf
G^+R^-	30	inf	2.21	inf	inf
G^+R^-	31	0.93	1.50	1.07	inf
G^+R^-	32	0	0	0	inf
G^-R^+	33	0	0	0	0
$G'R^+$	34	3.97	6.76	1.50	0.33
$G'R^+$	35	2.22	4.07	1.83	0
$G'R^+$	36	4.46	6.96	4.66	0.71
$G''R^+$	37	4.42	8.87	3.91	1.99
$G''R^+$	38	3.54	5.20	1.82	1.04
$G''R^+$	39	2.00	3.71	2.86**	0.40
$G''R^+$	40	2.31	5.36	1.42	0.56
$G''R^+$	41	3.00	7.63	2.95	1.64
G^+R'	42	1.70	2.94	1.63	inf
G^+R''	43	4.97	7.89	3.39	inf

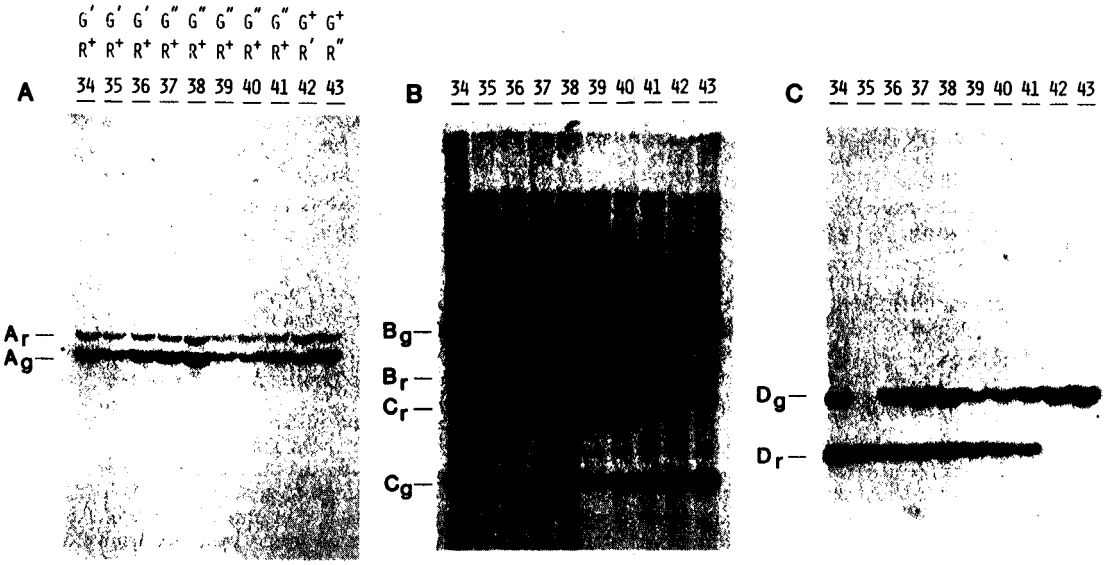


Fig. 5. Genomic Southern blots of DNA from anomalous trichromats. Each number and genotype refers to the individual whose DNA is loaded in that lane. (A) Eco RI digestion; (B) Bam HI and Eco RI double digestion; (C) Rsa I digestion. Fragment identities and experimental methods are described in the legend of Fig. 3A.

If pigment gene expression is controlled by 5′ proximal sequences, then the hybrid gene should be expressed in cells that were destined to become green cones. As a result the spectral sensitivity of these cells would be the same as that of the red cones.

As already mentioned the six G^+R^- genotypes are all different. Consider subjects 28, 31, and 32. Each has a single copy of A_r, B_r, and C_r; all lack D_r. They differ in their content of material derived from the green pigment gene: subject 32 has one copy of D_g and nothing else, subject 31 has two copies of D_g and one copy of A_g, B_g, and C_g, and subject 28 has three copies of D_g and two copies of A_g, B_g, and C_g. Models that account for these various stoichiometries are shown in Fig. 4. These three genotypes resemble one another in having one hybrid gene (5′ red–3′ green) and either zero, one, or two intact green pigment genes. We postulate that the hybrid gene produces a greenlike pigment in cells that would have become red cones.

Subject 29 (G^+R^-) has only one copy of fragments A_g, B_r, C_g, and D_g, and no other bands. Because the difference in size between A_r and A_g arises from a difference in the lengths of the first introns of the red and green pigment genes, these fragments do not serve as good markers for recombination events upstream of that intron. Instead B_r and B_g represent the most 5′ proximal landmarks which distinguish red from green pigment genes. Thus, subject 29 has a single hybrid gene in which only the part of the red pigment gene that is furthest toward the 5′ end has been retained (Fig. 4). Subject 30 (G^+R^-) is similar except that either one or two intact green pigment genes are also present. (The measured band intensities do not allow an unambigious assignment of stoichiometries for this subject.)

Subject 27 is the only G^+R^- subject who has all of the fragments corresponding to an intact red pigment gene (A_r, B_r, C_r, and D_r). At the same time he has more green pigment genes (either four or five) than we have seen in any normal subject (*10*). We are, at present, uncertain of the exact number and arrangement of his genes.

In summary, dichromasy caused by defects in the red or green mechanisms appears to be produced by various unequal exchanges or gene conversions. Different combinations of hybrid and normal genes can produce the same phenotype. We suggest that this heterogeneity may explain the observation that under very rigorous test conditions dichromats of a given type frequently do not accept each other's spectral matches (*15*).

Models for anomalous trichromacy. Ten anomalous trichromats were tested (Fig. 2), and their DNA was analyzed by Southern blotting (Fig. 5). In each case only those fragment sizes predicted from the normal restriction maps are seen. (This is also seen when the Eco RI digest is reprobed with a cDNA fragment encompassing exons 2 to 6.) Therefore we will consider only events involving homologous exchange. We consider first subjects 42 and 43, the former classified as G^+R' (R′, anomalous red sensitivity) and the latter G^+R'' (R″, extremely anomalous red sensitivity). Both subjects have one copy of A_r, B_r, and C_r, but lack D_r. They differ in that subject 42 has two copies of A_g, B_g, and C_g, and three copies of D_g, whereas subject 43 has approximately four copies of A_g, B_g, and C_g, and approximately five copies of D_g. Both subjects, therefore, have one copy of a hybrid gene (5′ red–3′ green) as well as a number of normal green pigment genes (Fig. 6). We predict that the red cones would in each case express the hybrid gene. The two different phenotypes could be caused by differences in the exact point of crossover, which might well determine the light-absorbing properties of the hybrid pigment.

In psychophysical studies of G^+R' and G^+R'' subjects (*16*), the anomalous red sensitivity curve was found in the interval between the normal red and green sensitivity curves. The decrease in color discrimination is attributed to the resultant decrease in the difference between red and green cone outputs. The shift of red cone sensitivity toward shorter wavelengths also alters the red-green ratio required to match a standard yellow. In extreme anomalous trichromacy the interval between normal green sensitivity and the short-wavelength-shifted red sensitivity is very small. This gives rise to large errors in color matching. The observation of hybrid genes,

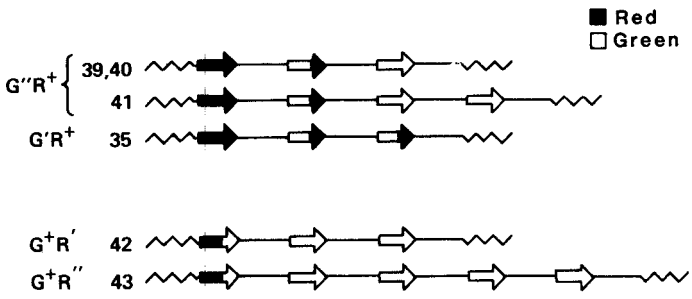


Fig. 6. Proposed arrangement of green and red pigment genes in anomalous trichromats. See legend to Fig. 4 for a description of the symbols.

which might plausibly encode pigments with spectral properties part way between those of the normal red and green pigments, fits well with these psychophysical data. The finding of approximately four green pigment genes in subject 43 is curious, but may not be relevant to the phenotype.

Four of eight $G'R^+$ (G', anomalous green sensitivity) and $G''R^+$ (G", extreme anomalous green sensitivity) (numbers 35, 39, 40, and 41) have a sufficiently small total number of genes that we could assign them unambiguous fragment stoichiometries. Subjects 39 and 40 have one copy of A_r, B_r, and C_r, two copies of D_r, two copies of A_g, B_g, and C_g, but only one copy of D_g. Subject 41 is similar, except one additional copy each of A_g, B_g, C_g, and D_g is superimposed. Subject 35 is unusual in having one copy of A_r, B_r, and C_r, three copies of D_r, two copies of A_g, B_g, and C_g, but none of D_g. The remaining four subjects (numbers 34, 36, 37, and 38) have a total of either four or five genes, including one or more hybrid genes, but quantitation of their hybridization band intensities do not permit an unambiguous assignment of fragment stoichiometries.

Our interpretation of these data for subjects 35, 39, 40, and 41 is shown in Fig. 6. Each has a single intact red pigment gene and some combination of intact or hybrid (or both) green pigment genes. In each case the hybrids are 5′ green–3′ red. If we suppose that green cones express all genes with 5′ sequences derived from a green pigment gene, then we would predict the production in those cones of a mixture of normal green or hybrid pigments (or both). The action spectrum of such a cell would then be shifted to the mean of the absorption spectra of its pigments. Psychophysical experiments indicate that G′ and G″ defects result, respectively, from smaller or greater shifts of green sensitivity toward the red sensitivity curve (*16*). If the shifted green sensitivity curve results from averaging the sensitivity curves of two or more pigments, then we might expect $G'R^+$ genotypes to have more normal green pigment genes than $G''R^+$ genotypes. In our sampling of three $G'R^+$ and five $G''R^+$ subjects we do not see this pattern; in fact, we see a small bias toward the reverse. It is therefore likely that the exact nature of the hybrid pigment is important and varies from subject to subject.

Although we could not deduce unambiguous stoichiometries for all eight green anomalous subjects, together they define a consistent pattern. All eight have one intact red pigment gene, as evidenced by the presence of single copies of A_r, B_r, and C_r, but all eight have at least two copies of D_r. Moreover, as a class they have on average more total genes than do normals (Fig. 7). We showed in (*10*) that normals have on average two copies of A_g, B_g, C_g, and D_g for each copy of A_r, B_r, C_r, and D_r, and therefore an average total of three genes. Green anomalous subjects have on average three copies of A_g, B_g, and C_g for each copy of A_r, B_r, and C_r giving an average total of four genes. In contrast, the ratio of D_g to D_r in green anomalous subjects is significantly less than that of normals. Therefore green anomalous subjects must have hybrid genes with the structure 5′ green–3′ red.

How might the observed green anomalous genotypes be created? The required event must produce both an increase in gene number and create a 5′ green–3′ red hybrid gene. An intragenic recombination between a red and a green pigment gene (Fig. 8) is such an event. Since the 3′ end of the red pigment gene is always accompanied by those green pigment genes distal to it, the total number of genes on the green anomalous chromosome will increase. In contrast, a gene conversion would not alter the total number of genes. The other product of this event should have on average fewer total genes and confer on a male a G^+R^-, G^+R', or G^+R'' phenotype.

Frequencies of color variant types. The frequencies of different anomalies in the red and green mechanisms have been accurately measured among Caucasians. We have calculated means and standard deviations of these values obtained from eight large population studies (*17*). The frequency of all types of red and green variant alleles is 8.08 ± 0.37 percent. Among the variant alleles 15.5 ± 5.2 percent are G^-R^+, 15.6 ± 5.2 percent are G^+R^-, 56.2 ± 4.6 percent are $G'R^+$ or $G''R^+$, and 12.5 ± 4.0 percent are G^+R' or G^+R''. (Anomalous and extreme anomalous trichromats are counted together for this analysis.) The most striking feature of this distribution is the asymmetry in anomalous trichromat frequencies.

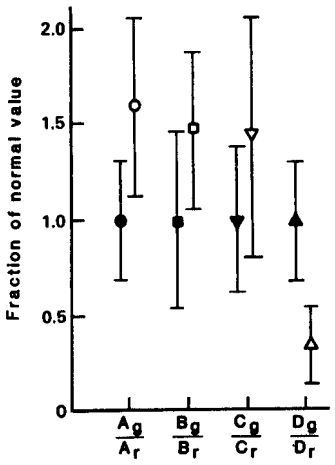


Fig. 7. Ratios of Southern blot band intensities: comparison of $G'R^+$ and $G''R^+$ to G^+R^+. Each point indicates the mean and standard deviation of the indicated ratio of fragments derived from the green pigment gene to those derived from the red pigment gene: A_g:A_r, B_g:B_r; C_g:C_r, and D_g:D_r (Fig. 3A). These ratios are plotted as the fraction of the corresponding ratios for the 18 color normals described in (*10*). For the color-normal data (closed symbols) the values, therefore, all center at 1.0. Data for the eight $G'R^+$ and $G''R^+$ subjects are shown by open symbols. The large standard deviations reflect the inhomogeneity of each population.

The models of unequal recombination just presented can account qualitatively for this asymmetry. The unequal intragenic exchanges (Fig. 8) illustrate a general feature of all such events. The recombination product carrying a 5′ green–3′ red hybrid gene also carries one normal red pigment gene and at least one normal green pigment gene. We propose that this hybrid gene would be expressed in the green cones together with the normal green pigment gene or genes. This gene arrangement therefore confers a $G'R^+$ or $G''R^+$ phenotype. In contrast, the other recombination product carries either a single 5′ red–3′ green hybrid gene (Fig. 8B) or that hybrid plus one

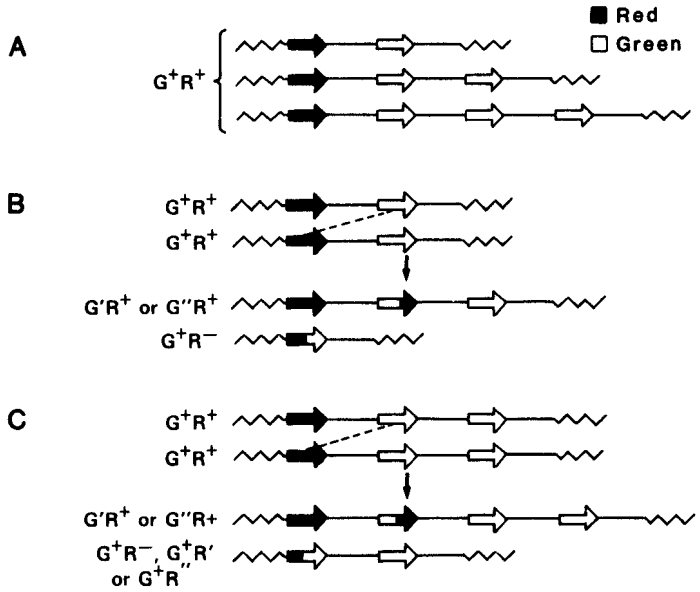


Fig. 8. Proposed products of unequal intragenic recombination between red and green pigment genes. At the left are the corresponding phenotypes. See legend to Fig. 4 for further explanation of symbols. (A) Proposed genotypes for a sampling of 18 color-normal males. (B) Recombination event in which one product retains only a 5′ red–3′ green hybrid gene. (C) Recombination event in which both products retain intact green pigment genes.

or more normal green pigment genes (Fig. 8C). This hybrid is presumed to be expressed in the red cones. In the former case it confers a G^+R^- phenotype, in the latter case it confers either G^+R^-, G^+R', or G^+R'' phenotypes, depending presumably on the exact site of intragenic recombination.

Herein lies the asymmetry. This mechanism produces $G'R^+$ and $G''R^+$ in excess over G^+R' and G^+R'', the difference being the number of G^+R^-. Moreover, if the fitness of dichromats is lower than that of anomalous trichromats, then $G'R^+$ and $G''R^+$ genotypes will accumulate to levels greater than the sum of G^+R', G^+R'', and G^+R^- genotypes.

Finally, G^-R^+ genotypes are produced by at least two different mechanisms. Unequal intergenic recombination, resulting in a change in the number of green pigment genes, gives rise to either two G^+R^+ products or to one G^+R^+ product and one G^-R^+ product, the latter produced by the complete loss of all green pigment genes [for example, subjects 19, 20, 22, 24, 26, and 33 shown in Fig. 4; see also figure 10B in (*10*)]. In contrast, those G^-R^+ genotypes containing 5′ green–3′ red hybrid genes (subjects 21, 23, and 25) (Fig. 4) cannot be produced from the normal genotypes (Fig. 8A) by a single recombination event. After such an event, the 3′ part of each 5′ green–3′ red hybrid gene remains linked to one or more downstream green pigment genes. These three subjects lack an intact green pigment gene, and therefore their genotypes are derived from either a gene conversion event or a sequence of two or more unequal recombinations.

These experiments verify the long-standing hypothesis that the loci responsible for inherited red-green color blindness are the genes encoding the red and green visual pigments. A test of the analogous hypothesis regarding blue color blindness and the blue pigment gene is now also possible. The arrangement of green and red pigment genes observed in color variant subjects reveals a unifying theme: in at least 24 out of 25 cases, either unequal homologous recombination or gene conversion has produced an arrangement of pigment genes different from those arrangements observed in color normals. This propensity for homologous events is probably a consequence of the proximity and high degree of sequence homology between red and green pigment genes. It is probably responsible, at least in part, for the high frequency (8 percent among Caucasian males) of red-green color blindness. Other genetic events, such as point mutation and nonhomologous rearrangement, probably occur at a far lower frequency. The finding of hybrid genes in anomalous trichromats fits well with the observation that anomalous green and red sensitivities lie in the interval between normal green and red sensitivities. In contrast, point mutation would not be expected to preferentially produce shifts toward either long or short wavelengths.

REFERENCES AND NOTES

1. T. Young, *Phil. Trans. Roy. Soc.*, 12 (1802).
2. F. Horner, *Amtl. Ber. Verwaltung d. Medizinalwesens Kanton Zurich*, 208 (1876); R. Boynton, *Human Color Vision* (Holt, Rinehart, and Winston, New York, 1979).
3. G. H. M. Waaler, *Z. Indukt. Abstamm. Vererbungsl.* **45**, 279 (1927); H. Kalmus, *Diagnosis and Genetics of Defective Colour Vision* (Pergamon, New York, 1965).
4. V. A. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, MD, 1983); I. A. Porter, J. Schulze, V. A. McKusick, *Ann. Hum. Genet.* **26**, 107 (1962).
5. J. Francois, *Heredity in Opthalmology* (Mosby, St. Louis, MO, 1961); H. Kalmus, *Ann. Hum. Genet.* **20**, 33 (1955).
6. M. F. Lyon, *Am. J. Hum. Genet.* **14**, 135 (1962).
7. R. W. Pickford, *Br. J. Physiol. Optics* **16**, 83 (1959); H. Kalmus, *Diagnosis and Genetics of Defective Colour Vision* (Pergamon, New York, 1965).
8. W. A. H. Rushton, *J. Physiol.* **168**, 345 (1963); *ibid.* **176**, 24 (1965); M. Alpern and T. Wake, *ibid.* **266**, 595 (1977).
9. H. J. A. Dartnall, J. K. Bowmaker, J. D. Mollon, in *Colour Vision*, J. D. Mollon and L. T. Sharpe, Eds. (Academic Press, New York, 1983), p. 69.
10. J. Nathans, D. Thomas, D. S. Hogness, *Science* **232**, 193, (1986).
11. T. B. Shows, A. Y. Sakaguchi, S. L. Naylor, *Adv. Human Genet.* **12**, 341 (1982); A. Y. Sakaguchi and T. B. Shows, *Somat. Cell Genet.* **8**, 83 (1982); T. B. Shows *et al.*, *ibid.* **10**, 315 (1984); T. B. Shows, in *Genetic Mechanisms of Sexual Development*, H. L. Vallet and I. H. Porter, Eds. (Academic Press, New York, 1979), p. 253; T. B. Shows, in *Isozymes: Current Topics in Biological and Medical Research*, M. C. Rattazzi, J. G. Scandalios, G. S. Whitt, Eds. (Liss, New York, 1983), p. 323.
12. J. Nathans and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4851 (1984).

12a. Abbreviations: G^+R^+, normal color vision; $G'R^+$, anomalous green sensitivity (deuteranomaly); $G''R^+$, extreme anomalous green sensitivity (extreme deuteranomaly); G^-R^+, absent green sensitivity (deuteranopia); G^+R', anomalous red sensitivity (protanomaly); G^+R'', extreme anomalous red sensitivity (extreme protanomaly); and G^+R^-, absent red sensitivity (protanopia).

13. S. Hecht, *The Retinal Process Concerned with Visual Acuity and Color Vision* (Harvard Univ. Press, Cambridge, MA, 1931); D. G. Wilder, *Dis. Abs. 31, 4B*, 2333 (1970).
14. G. Wald, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 1347 (1966); Y. Hsia and C. Graham, *ibid.* **43**, 1011 (1957).
15. M. Alpern and T. Wake, *J. Physiol.* **266**, 595 (1977); M. Alpern and E. N. Pugh, *ibid.*, p. 613.
16. T. P. Piantanida and H. G. Sperling, *Vision Res.* **13**, 2033 (1973); *ibid.*, p. 2049; W. A. H. Rushton, D. S. Powell, K. D. White, *ibid.*, p. 2017.
17. J. Francois, *Heredity in Opthalmology* (Mosby, St. Louis, MO, 1961); R. W. Pickford, *Br. J. Physiol. Optics* **14**, 2 (1957).
18. R. W. Davis, D. Botstein, J. R. Roth, *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980).

19. We thank Drs. Mathew Alpern and Vivianne Smith for providing us with subjects 27 and 28, respectively. We also thank all of the color-blind subjects for their participation. A special thanks to Dr. Lubert Stryer for his enthusiasm and insight throughout the course of this work. J.N. is a trainee of the Medical Scientist Training Program of the National Institutes of Health. This work was supported by a grant from the National Eye Institute (D.S.H.) and from the National Institutes of Health (T.B.S.).

14 January 1986; accepted 12 February 1986

J. Physiol. (1962), **160**, *pp.* 106–154
With 2 plates and 20 text-figures
Printed in Great Britain

RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL

From the Neurophysiology Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

(*Received* 31 *July* 1961)

What chiefly distinguishes cerebral cortex from other parts of the central nervous system is the great diversity of its cell types and interconnexions. It would be astonishing if such a structure did not profoundly modify the response patterns of fibres coming into it. In the cat's visual cortex, the receptive field arrangements of single cells suggest that there is indeed a degree of complexity far exceeding anything yet seen at lower levels in the visual system.

In a previous paper we described receptive fields of single cortical cells, observing responses to spots of light shone on one or both retinas (Hubel & Wiesel, 1959). In the present work this method is used to examine receptive fields of a more complex type (Part I) and to make additional observations on binocular interaction (Part II).

This approach is necessary in order to understand the behaviour of individual cells, but it fails to deal with the problem of the relationship of one cell to its neighbours. In the past, the technique of recording evoked slow waves has been used with great success in studies of functional anatomy. It was employed by Talbot & Marshall (1941) and by Thompson, Woolsey & Talbot (1950) for mapping out the visual cortex in the rabbit, cat, and monkey. Daniel & Whitteridge (1959) have recently extended this work in the primate. Most of our present knowledge of retinotopic projections, binocular overlap, and the second visual area is based on these investigations. Yet the method of evoked potentials is valuable mainly for detecting behaviour common to large populations of neighbouring cells; it cannot differentiate functionally between areas of cortex smaller than about 1 mm^2. To overcome this difficulty a method has in recent years been developed for studying cells separately or in small groups during long micro-electrode penetrations through nervous tissue. Responses are correlated with cell location by reconstructing the electrode tracks from histological material. These techniques have been applied to

the somatic sensory cortex of the cat and monkey in a remarkable series of studies by Mountcastle (1957) and Powell & Mountcastle (1959). Their results show that the approach is a powerful one, capable of revealing systems of organization not hinted at by the known morphology. In Part III of the present paper we use this method in studying the functional architecture of the visual cortex. It helped us attempt to explain on anatomical grounds how cortical receptive fields are built up.

METHODS

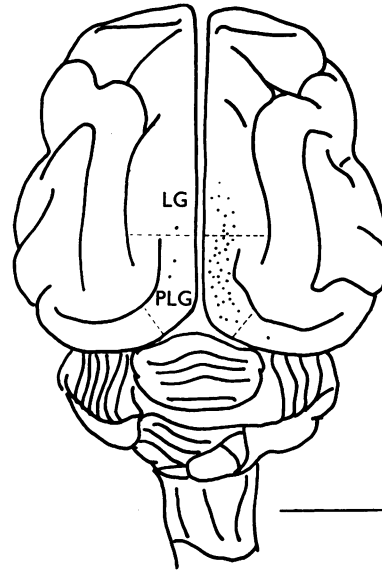
Recordings were made from forty acutely prepared cats, anaesthetized with thiopental sodium, and maintained in light sleep with additional doses by observing the electrocorticogram. Animals were paralysed with succinylcholine to stabilize the eyes. Pupils were dilated with atropine. Details of stimulating and recording methods are given in previous papers (Hubel, 1959; Hubel & Wiesel, 1959, 1960). The animal faced a wide tangent screen at a distance of 1·5 m, and various patterns of white light were shone on the screen by a tungsten-filament projector. All recordings were made in the light-adapted state. Background illumination varied from $-1{\cdot}0$ to $+1{\cdot}0$ $\log_{10}$ cd/m^2. Stimuli were from 0·2 to 2·0 log. units brighter than the background. For each cell receptive fields were mapped out separately for the two eyes on sheets of paper, and these were kept as permanent records.

Points on the screen corresponding to the area centralis and the optic disk of the two eyes were determined by a projection method (Hubel & Wiesel, 1960). The position of each receptive field was measured with respect to these points. Because of the muscle relaxant the eyes usually diverged slightly, so that points corresponding to the two centres of gaze were not necessarily superimposed. In stimulating the two eyes simultaneously it was therefore often necessary to use two spots placed in corresponding parts of the two visual fields. Moreover, at times the two eyes were slightly rotated in an inward direction in the plane of their equators. This rotation was estimated by (1) photographing the cat before and during the experiment, and comparing the angles of inclination of the slit-shaped pupils, or (2) by noting the inclination to the horizontal of a line joining the area centralis with the optic disk, which in the normal position of the eye was estimated, by the first method, to average about 25°. The combined inward rotations of the two eyes seldom exceeded 10°. Since the receptive fields in this study were usually centrally rather than peripherally placed on the retina, the rotations did not lead to any appreciable linear displacement. Angular displacements of receptive fields occasionally required correction, as they led to an apparent difference in the orientation of the two receptive-field axes of a binocularly driven unit. The direction and magnitude of this difference were always consistent with the estimated inward rotation of the two eyes. Moreover, in a given experiment the difference was constant, even though the axis orientation varied from cell to cell.

The diagram of Text-fig. 1 shows the points of entry into the cortex of all 45 microelectrode penetrations. Most electrode tracks went no deeper than 3 or 4 mm, so that explorations were mainly limited to the apical segments of the lateral and post-lateral gyri (LG and PLG) and a few millimetres down along the adjoining medial and lateral folds. The extent of the territory covered is indicated roughly by Text-figs. 13–15. Although the lateral boundary of the striate cortex is not always sharply defined in Nissl-stained or myelin-stained material, most penetrations were well within the region generally accepted as 'striate' (O'Leary, 1941). Most penetrations were made from the cortical region receiving projections from in or near the area centralis; this cortical region is shown in Text-fig. 1 as the area between the interrupted lines.

Tungsten micro-electrodes were advanced by a hydraulic micro-electrode positioner (Hubel, 1957, 1959). In searching for single cortical units the retina was continually stimulated with stationary and moving forms while the electrode was advanced. The unresolved background activity (see p. 129) served as a guide for determining the optimum stimulus. This procedure increased the number of cells observed in a penetration, since the sampling was not limited to spontaneously active units.

In each penetration electrolytic lesions were made at one or more points. When only one lesion was made, it was generally at the end of an electrode track. Brains were fixed in 10% formalin, embedded in celloidin, sectioned at 20 μ, and stained with cresyl violet. Lesions were 50–100 μ in diameter, which was small enough to indicate the position of the electrode tip to the nearest cortical layer. The positions of other units encountered in a cortical penetration were determined by calculating the distance back from the lesion along the track,



Text-fig. 1. Diagram of dorsal aspect of cat's brain, to show entry points of 45 micro-electrode penetrations. The penetrations between the interrupted lines are those in which cells had their receptive fields in or near area centralis. LG, lateral gyrus; PLG, post-lateral gyrus. Scale, 1 cm.

using depth readings corresponding to the unit and the lesion. A correction was made for brain shrinkage, which was estimated by comparing the distance between two lesions, measured under the microscope, with the distance calculated from depths at which the two lesions were made. From brain to brain this shrinkage was not constant, so that it was not possible to apply an average correction for shrinkage to all brains. For tracks marked by only one lesion it was assumed that the first unit activity was recorded at the boundary of the first and second layers; any error resulting from this was probably small, since in a number of penetrations a lesion was made at the point where the first units were encountered, and these were in the lower first or the upper second layers, or else at the very boundary. The absence of cell-body records and unresolved background activity as the electrode passed through subcortical white matter (see Text-fig. 13 and Pl. 1) was also helpful in confirming the accuracy of the track reconstructions.

PART I

ORGANIZATION OF RECEPTIVE FIELDS IN CAT'S VISUAL CORTEX: PROPERTIES OF 'SIMPLE' AND 'COMPLEX' FIELDS

The receptive field of a cell in the visual system may be defined as the region of retina (or visual field) over which one can influence the firing of that cell. In the cat's retina one can distinguish two types of ganglion cells, those with 'on'-centre receptive fields and those with 'off'-centre fields (Kuffler, 1953). The lateral geniculate body also has cells of these two types; so far no others have been found (Hubel & Wiesel, 1961). In contrast, the visual cortex contains a large number of functionally different cell types; yet with the exception of afferent fibres from the lateral geniculate body we have found no units with concentric 'on'-centre or 'off'-centre fields.

When stimulated with stationary or moving patterns of light, cells in the visual cortex gave responses that could be interpreted in terms of the arrangements of excitatory and inhibitory regions in their receptive fields (Hubel & Wiesel, 1959). Not all cells behaved so simply, however; some responded in a complex manner which bore little obvious relationship to the receptive fields mapped with small spots. It has become increasingly apparent to us that cortical cells differ in the complexity of their receptive fields. The great majority of fields seem to fall naturally into two groups, which we have termed 'simple' and 'complex'. Although the fields to be described represent the commonest subtypes of these groups, new varieties are continually appearing, and it is unlikely that the ones we have listed give anything like a complete picture of the striate cortex. We have therefore avoided a rigid system of classification, and have designated receptive fields by letters or numbers only for convenience in referring to the figures. We shall concentrate especially on features common to simple fields and on those common to complex fields, emphasizing differences between the two groups, and also between cortical fields and lateral geniculate fields.

RESULTS

Simple receptive fields

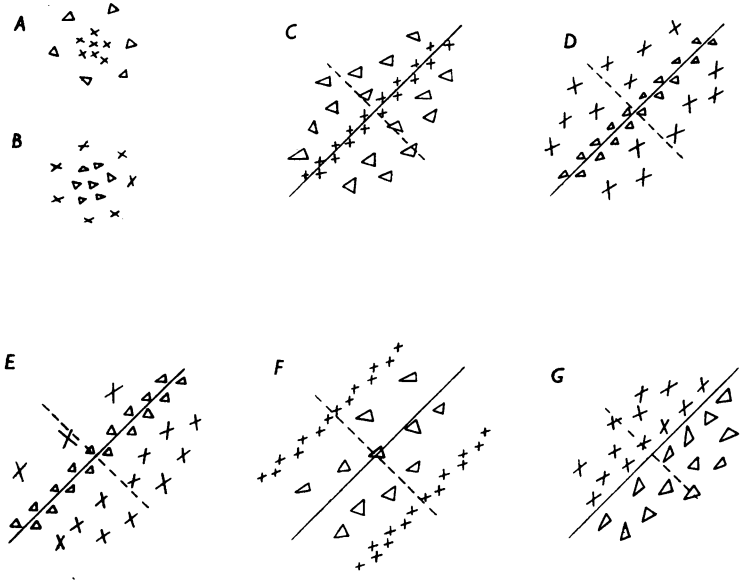
The receptive fields of 233 of the 303 cortical cells in the present series were classified as 'simple'. Like retinal ganglion and geniculate cells, cortical cells with simple fields possessed distinct excitatory and inhibitory subdivisions. Illumination of part or all of an excitatory region increased the maintained firing of the cell, whereas a light shone in the

inhibitory region suppressed the firing and evoked a discharge at 'off'. A large spot confined to either area produced a greater change in rate of firing than a small spot, indicating summation within either region. On the other hand, the two types of region within a receptive field were mutually antagonistic. This was most forcefully shown by the absence or near absence of a response to simultaneous illumination of both regions, for example, with diffuse light. From the arrangement of excitatory and inhibitory regions it was usually possible to predict in a qualitative way the responses to any shape of stimulus, stationary or moving. Spots having the approximate shape of one or other region were the most effective stationary stimuli; smaller spots failed to take full advantage of summation within a region, while larger ones were likely to invade opposing regions, so reducing the response. To summarize: these fields were termed 'simple' because like retinal and geniculate fields (1) they were subdivided into distinct excitatory and inhibitory regions; (2) there was summation within the separate excitatory and inhibitory parts; (3) there was antagonism between excitatory and inhibitory regions; and (4) it was possible to predict responses to stationary or moving spots of various shapes from a map of the excitatory and inhibitory areas.

While simple cortical receptive fields were similar to those of retinal ganglion cells and geniculate cells in possessing excitatory and inhibitory subdivisions, they differed profoundly in the spatial arrangements of these regions. The receptive fields of all retinal ganglion and geniculate cells had one or other of the concentric forms shown in Text-fig. 2*A*, *B*. (Excitatory areas are indicated by crosses, inhibitory areas by triangles.) In contrast, simple cortical fields all had a side-to-side arrangement of excitatory and inhibitory areas with separation of the areas by parallel straight-line boundaries rather than circular ones. There were several varieties of fields, differing in the number of subdivisions and the relative area occupied by each subdivision. The commonest arrangements are illustrated in Text-fig. 2*C*–*G*: Table 1 gives the number of cells observed in each category. The departure of these fields from circular symmetry introduces a new variable, namely, the orientation of the boundaries separating the field subdivisions. This orientation is a characteristic of each cortical cell, and may be vertical, horizontal, or oblique. There was no indication that any one orientation was more common than the others. We shall use the term *receptive-field axis* to indicate a line through the centre of a field, parallel to the boundaries separating excitatory and inhibitory regions. The *axis orientation* will then refer to the orientation of these boundaries, either on the retina or in the visual field. Axes are shown in Text-fig. 2 by continuous lines.

Two common types of fields, shown in Text-fig. 2*C*, *D*, each consisted of a narrow elongated area, excitatory or inhibitory, flanked on either side

by two regions of the opposite type. In these fields the two flanking regions were symmetrical, i.e. they were about equal in area and the responses obtained from them were of about the same magnitude. In addition there were fields with long narrow centres (excitatory or inhibitory) and asymmetrical flanks. An example of an asymmetrical field with an inhibitory centre is shown in Text-fig. 2*E*. The most effective stationary stimulus for all of these cells was a long narrow rectangle ('slit') of light just large



Text-fig. 2. Common arrangements of lateral geniculate and cortical receptive fields. *A*. 'On'-centre geniculate receptive field. *B*. 'Off'-centre geniculate receptive field. *C–G*. Various arrangements of simple cortical receptive fields. ×, areas giving excitatory responses ('on' responses); △, areas giving inhibitory responses ('off' responses). Receptive-field axes are shown by continuous lines through field centres; in the figure these are all oblique, but each arrangement occurs in all orientations.

enough to cover the central region without invading either flank. For maximum centre response the orientation of the slit was critical; changing the orientation by more than 5–10° was usually enough to reduce a response greatly or even abolish it. Illuminating both flanks usually evoked a strong response. If a slit having the same size as the receptive-field centre was shone in either flanking area it evoked only a weak response, since it covered only part of one flank. Diffuse light was ineffective, or at most evoked only a very weak response, indicating that the excitatory and inhibitory parts of the receptive field were very nearly balanced.

In these fields the equivalent but opposite-type regions occupied retinal

areas that were far from equal; the centre portion was small and concentrated whereas the flanks were widely dispersed. A similar inequality was found in fields of type *F*, Text-fig. 2, but here the excitatory flanks were elongated and concentrated, while the centre was relatively large and diffuse. The optimum response was evoked by simultaneously illuminating the two flanks with two parallel slits (see Hubel & Wiesel, 1959, Fig. 9).

Some cells had fields in which only two regions were discernible, arranged side by side as in Text-fig. 2*G*. For these cells the most efficient stationary stimulus consisted of two areas of differing brightness placed so that the line separating them fell exactly over the boundary between the excitatory and inhibitory parts of the field. This type of stimulus was termed an 'edge'. An 'on' or an 'off' response was evoked depending on whether the bright part of the stimulus fell over the excitatory or the inhibitory region. A slight change in position or orientation of the line separating the light from the dark area was usually enough to reduce greatly the effectiveness of the stimulus.

Moving stimuli were very effective, probably because of the synergistic effects of leaving an inhibitory area and simultaneously entering an excitatory area (Hubel & Wiesel, 1959). The optimum stimulus could usually be predicted from the distribution of excitatory and inhibitory regions of the receptive field. With moving stimuli, just as with stationary, the orientation was critical. In contrast, a slit or edge moved across the circularly symmetric field of a geniculate cell gave (as one would expect) roughly the same response regardless of the stimulus orientation. The responses evoked when an optimally oriented slit crossed back and forth over a cortical receptive field were often roughly equal for the two directions of crossing. This was true of fields like those shown in Text-fig. 2*C*, *D* and *F*. For many cells, however, the responses to two diametrically opposite movements were different, and some only responded to one of the two movements. The inequalities could usually be accounted for by an asymmetry in flanking regions, of the type shown in Text-fig. 2*E* (see also Hubel & Wiesel, 1959, Fig. 7). In fields that had only two discernible regions arranged side by side (Text-fig. 2*G*), the difference in the responses to a moving slit or edge was especially pronounced.

Optimum rates of movement varied from one cell to another. On several occasions two cells were recorded together, one of which responded only to a slow-moving stimulus (1°/sec or lower) the other to a rapid one (10°/sec or more). For cells with fields of type *F*, Text-fig. 2, the time elapsing between the two discharges to a moving stimulus was a measure of the rate of movement (see Hubel & Wiesel, 1959, Fig. 5).

If responses to movement were predictable from arrangements of excitatory and inhibitory regions, the reverse was to some extent also true.

The axis orientation of a field, for example, was given by the most effective orientation of a moving slit or edge. If an optimally oriented slit produced a brief discharge on crossing from one region to another, one could predict that the first region was inhibitory and the second excitatory. Brief responses to crossing a very confined region were characteristic of cells with simple cortical fields, whereas the complex cells to be described below gave sustained responses to movement over much wider areas.

TABLE 1. Simple cortical fields

	Text-fig.	No. of cells
(*a*) Narrow concentrated centres		
(i) Symmetrical flanks		
Excitatory centres	2*C*	23
Inhibitory centres	2*D*	17
(ii) Asymmetrical flanks		
Excitatory centres	—	28
Inhibitory centres	2*E*	10
(*b*) Large centres; concentrated flanks	2*F*	21
(*c*) One excitatory region and one inhibitory	2*G*	17
(*d*) Uncategorized	—	117
Total number of simple fields		233

Movement was used extensively as a stimulus in experiments in which the main object was to determine axis orientation and ocular dominance for a large number of cells in a single penetration, and when it was not practical, because of time limitations, to map out every field completely. Because movement was generally a very powerful stimulus, it was also used in studying cells that gave little or no response to stationary patterns. In all, 117 of the 233 simple cells were studied mainly by moving stimuli. In Table 1 these have been kept separate from the other groups since the distribution of their excitatory and inhibitory regions is not known with the same degree of certainty. It is also possible that with further study, some of these fields would have revealed complex properties.

Complex receptive fields

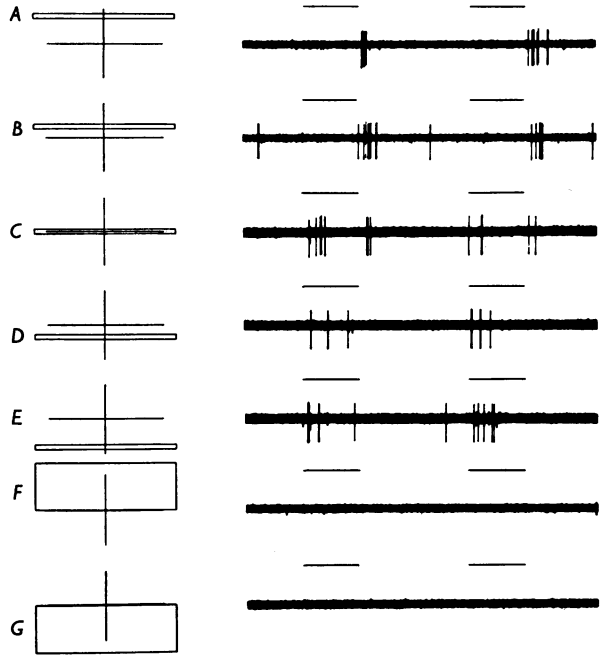
Intermixed with cells having simple fields, and present in most penetrations of the striate cortex, were cells with far more intricate and elaborate properties. The receptive fields of these cells were termed 'complex'. Unlike cells with simple fields, these responded to variously-shaped stationary or moving forms in a way that could not be predicted from maps made with small circular spots. Often such maps could not be made, since small round spots were either ineffective or evoked only mixed ('on-off') responses throughout the receptive field. When separate 'on' and 'off' regions could be discerned, the principles of summation and mutual antagonism, so helpful in interpreting simple fields, did not generally hold. Nevertheless, there were some important features common to the two

types of cells. In the following examples, four types of complex fields will be illustrated. The numbers observed of each type are given in Table 2.

TABLE 2. Complex cortical receptive fields

	Text-fig.	No. of cells
(*a*) Activated by slit—non-uniform field	3	11
(*b*) Activated by slit—uniform field	4	39
(*c*) Activated by edge	5–6	14
(*d*) Activated by dark bar	7–8	6
Total number of complex fields		70

The cell of Text-fig. 3 failed to respond to round spots of light, whether small or large. By trial and error with many shapes of stimulus it was discovered that the cell's firing could be influenced by a horizontally oriented slit $\frac{1}{8}°$ wide and 3° long. Provided the slit was horizontal its exact



Text-fig. 3. Responses of a cell with a complex receptive field to stimulation of the left (contralateral) eye. Receptive field located in area centralis. The diagrams to the left of each record indicate the position of a horizontal rectangular light stimulus with respect to the receptive field, marked by a cross. In each record the upper line indicates when the stimulus is on. *A–E*, stimulus $\frac{1}{8} \times 3°$, *F–G*, stimulus $1\frac{1}{2} \times 3°$ (4° is equivalent to 1 mm on the cat retina). For background illumination and stimulus intensity see Methods. Cell was activated in the same way from right eye, but less vigorously (ocular-dominance group 2, see Part II). An electrolytic lesion made while recording from this cell was found near the border of layers 5 and 6, in the apical segment of the post-lateral gyrus. Positive deflexions upward; duration of each stimulus 1 sec.

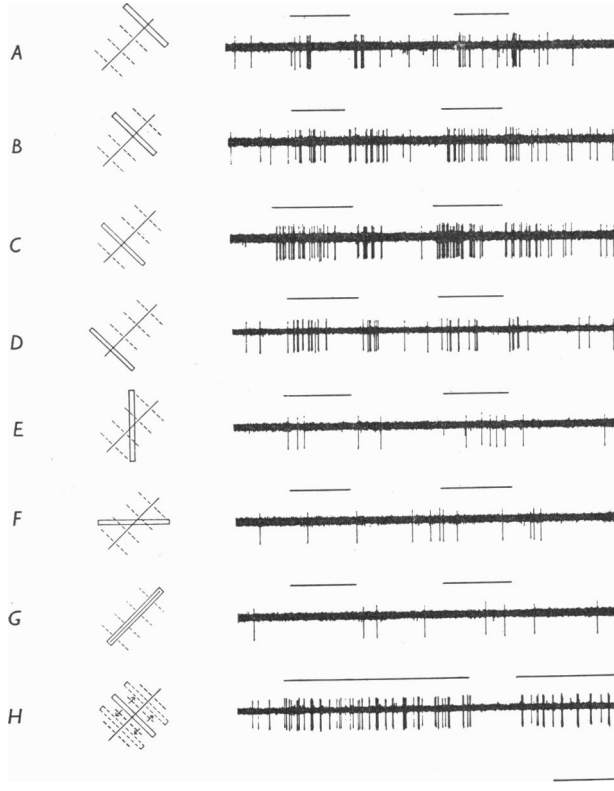
positioning within the 3°-diameter receptive field was not critical. When it was shone anywhere above the centre of the receptive field (the horizontal line of Text-fig. 3) an 'off' response was obtained; 'on' responses were evoked throughout the lower half. In an intermediate position (Text-fig. 3*C*) the cell responded at both 'on' and 'off'. From experience with simpler receptive fields one might have expected wider slits to give increasingly better responses owing to summation within the upper or lower part of the field, and that illumination of either half by itself might be the most effective stimulus of all. The result was just the opposite: responses fell off rapidly as the stimulus was widened beyond about $\frac{1}{8}°$, and large rectangles covering the entire lower or upper halves of the receptive field were quite ineffective (Text-fig. 3*F*, *G*). On the other hand, summation could easily be demonstrated in a horizontal direction, since a slit $\frac{1}{8}°$ wide but extending only across part of the field was less effective than a longer one covering the entire width. One might also have expected the orientation of the slit to be unimportant as long as the stimulus was wholly confined to the region above the horizontal line or the region below. On the contrary, the orientation was critical, since a tilt of even a few degrees from the horizontal markedly reduced the response, even though the slit did not cross the boundary separating the upper and lower halves of the field.

In preferring a slit specific in width and orientation this cell resembled certain cells with simple fields. When stimulated in the upper part of its field it behaved in many respects like cells with 'off'-centre fields of type *D*, Text-fig. 2; in the lower part it responded like 'on'-centre fields of Text-fig. 2*C*. But for this cell the strict requirements for shape and orientation of the stimulus were in marked contrast to the relatively large leeway of the stimulus in its ordinate position on the retina. Cells with simple fields, on the other hand, showed very little latitude in the positioning of an optimally oriented stimulus.

The upper part of this receptive field may be considered inhibitory and the lower part excitatory, even though in either area summation only occurred in a horizontal direction. Such subdivisions were occasionally found in complex fields, but more often the fields were uniform in this respect. This was true for the other complex fields to be described in this section.

Responses of a second complex unit are shown in Text-fig. 4. In many ways the receptive field of this cell was similar to the one just described. A slit was the most potent stimulus, and the most effective width was again $\frac{1}{8}°$. Once more the orientation was an important stimulus variable, since the slit was effective anywhere in the field as long as it was placed in a 10 o'clock–4 o'clock orientation (Text-fig. 4*A–D*). A change in orientation of more than 5–10° in either direction produced a marked

reduction in the response (Text-fig. 4*E–G*). As usual, diffuse light had no influence on the firing. This cell responded especially well if the slit, oriented as in *A–D*, was moved steadily across the receptive field. Sustained discharges were evoked over the entire length of the field. The optimum rate of movement was about 1°/sec. If movement was interrupted the discharge stopped, and when it was resumed the firing recommenced. Continuous firing could be maintained indefinitely by small side-



Text-fig. 4. Responses of a cell with a complex field to stimulation of the left (contralateral) eye with a slit $\frac{1}{8} \times 2\frac{1}{2}°$. Receptive field was in the area centralis and was about $2 \times 3°$ in size. *A–D*, $\frac{1}{8}°$ wide slit oriented parallel to receptive field axis. *E–G*, slit oriented at 45 and 90° to receptive-field axis. *H*, slit oriented as in *A–D*, is on throughout the record and is moved rapidly from side to side where indicated by upper beam. Responses from left eye slightly more marked than those from right (Group 3, see Part II). Time 1 sec.

to-side movements of a stimulus within the receptive field (Text-fig. 4*H*). The pattern of firing was one characteristic of many complex cells, especially those responding well to moving stimuli. It consisted of a series of short high-frequency repetitive discharges each containing 5–10 spikes. The

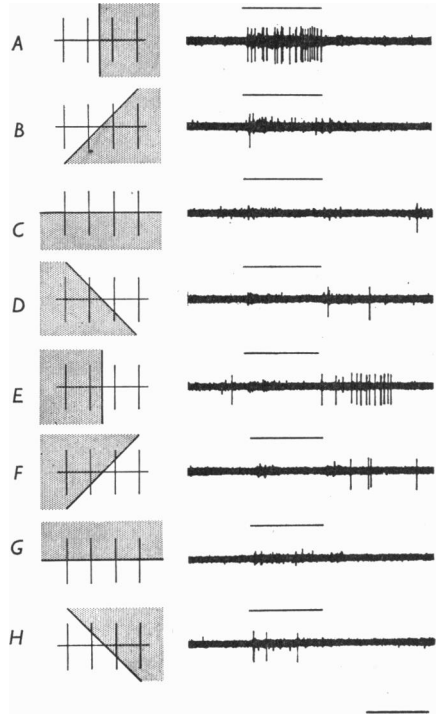
bursts occurred at irregular intervals, at frequencies up to about 20/sec. For this cell, movement of an optimally oriented slit was about equally effective in either of the two opposite directions. This was not true of all complex units, as will be seen in some of the examples given below.

Like the cell of Text-fig. 3 this cell may be thought of as having a counterpart in simple fields of the type shown in Text-fig. 2*C–E*. It shares with these simpler fields the attribute of responding well to properly oriented slit stimuli. Once more the distinction lies in the permissible variation in position of the optimally oriented stimulus. The variation is small (relative to the size of the receptive field) in the simple fields, large in the complex. Though resembling the cell of Text-fig. 3 in requiring a slit for a stimulus, this cell differed in that its responses to a properly oriented slit were mixed ('on-off') in type. This was not unusual for cells with complex fields. In contrast, cortical cells with simple fields, like retinal ganglion cells and lateral geniculate cells, responded to optimum restricted stimuli either with excitatory ('on') responses or inhibitory ('off') responses. When a stimulus covered opposing regions, the effects normally tended to cancel, though sometimes mixed discharges were obtained, the 'on' and 'off' components both being weak. For these simpler fields 'on-off' responses were thus an indication that the stimulus was not optimum. Yet some cells with complex fields responded with mixed discharges even to the most effective stationary stimuli we could find. Among the stimuli tried were curved objects, dark stripes, and still more complicated patterns, as well as monochromatic spots and slits.

A third type of complex field is illustrated in Text-figs. 5 and 6. There were no responses to small circular spots or to slits, but an edge was very effective if oriented vertically. Excitatory or inhibitory responses were produced depending on whether the brighter area was to the left or the right (Text-fig. 5*A*, *E*). So far, these are just the responses one would expect from a cell with a vertically oriented simple field of the type shown in Text-fig. 2*G*. In such a field the stimulus placement for optimum response is generally very critical. On the contrary, the complex unit responded to vertical edges over an unusually large region about 16° in length (Text-fig. 6). 'On' responses were obtained with light to the left (*A–D*), and 'off' responses with light to the right (*E–H*), regardless of the position of the line separating light from darkness. When the entire receptive field was illuminated diffusely (*I*) no response was evoked. As with all complex fields, we are unable to account for these responses by any simple spatial arrangement of excitatory and inhibitory regions.

Like the complex units already described, this cell was apparently more concerned with the orientation of a stimulus than with its exact position in the receptive field. It differed in responding well to edges but poorly or

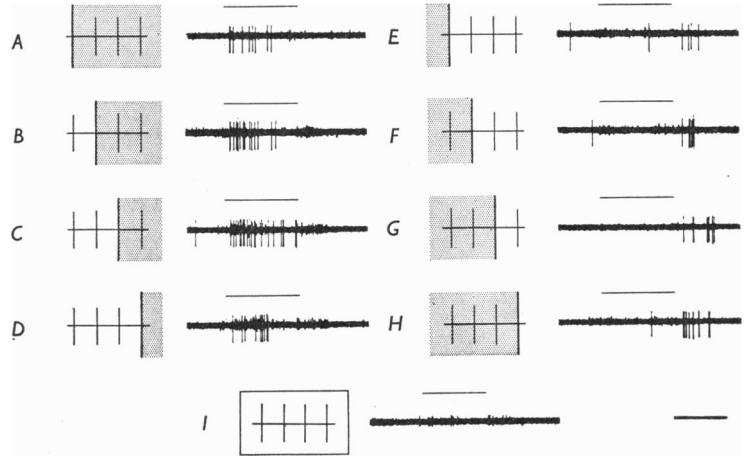
not at all to slits, whether narrow or wide. It is interesting in this connexion that exchanging an edge for its mirror equivalent reversed the response, i.e. replaced an excitatory response by an inhibitory and vice versa. The ineffectiveness of a slit might therefore be explained by supposing that the opposite effects of its two edges tended to cancel each other.



Text-fig. 5. Responses of a cell with a large (8 × 16°) complex receptive field to an edge projected on the ipsilateral retina so as to cross the receptive field in various directions. (The screen is illuminated by a diffuse background light, at 0·0 $\log_{10}$ cd/m². At the time of stimulus, shown by upper line of each record, half the screen, to one side of the variable boundary, is illuminated at 1·0 $\log_{10}$ cd/m², while the other half is kept constant.) *A*, vertical edge with light area to left, darker area to right. *B–H*, various other orientations of edge. Position of receptive field 20° below and to the left of the area centralis. Responses from ipsilateral eye stronger than those from contralateral eye (group 5, see Part II). Time 1 sec.

As shown in Text-fig. 6, the responses of the cell to a given vertical edge were consistent in type, being either 'on' or 'off' for all positions of the edge within the receptive field. In being uniform in its response-type it resembled the cell of Text-fig. 4. A few other cells of the same general category showed a similar preference for edges, but lacked this uniformity. Their receptive fields resembled the field of Text-fig. 3, in that a given edge evoked responses of one type over half the field, and the opposite type over

the other half. These fields were divided into two halves by a line parallel to the receptive-field axis: an edge oriented parallel to the axis gave 'on' responses throughout one of the halves and 'off' responses through the other. In either half, replacing the edge by its mirror image reversed the response-type. Even cells, which were uniform in their response-types, like those in Text-fig. 4–6, varied to some extent in the magnitude of their responses, depending on the position of the stimulus. Moreover, as with most cortical cells, there was some variation in responses to identical stimuli.

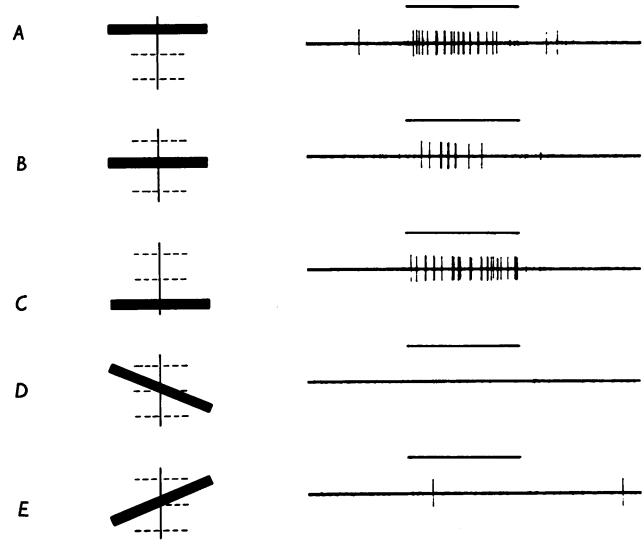


Text-fig. 6. Same cell as in Text-fig. 5. *A–H*, responses to a vertical edge in various parts of the receptive field: *A–D*, brighter light to the left; *E–H*, brighter light to the right; *I*, large rectangle, 10 × 20°, covering entire receptive field. Time, 1 sec.

A final example is given to illustrate the wide range of variation in the organization of complex receptive fields. The cell of Text-figs. 7 and 8 was not strongly influenced by any form projected upon the screen; it gave only weak, unsustained 'on' responses to a dark horizontal rectangle against a light background, and to other forms it was unresponsive. A strong discharge was evoked, however, if a black rectangular object (for example, a piece of black tape) was placed against the brightly illuminated screen. The receptive field of the cell was about 5 × 5°, and the most effective stimulus width was about $\frac{1}{3}$°. Vigorous firing was obtained regardless of the position of the rectangle, as long as it was horizontal and within the receptive field. If it was tipped more than 10° in either direction no discharge was evoked (Text-fig. 7*D*, *E*). We have recorded several complex fields which resembled this one in that they responded best to black rectangles against a bright background. Presumably it is important to

have good contrast between the narrow black rectangle and the background; this is technically difficult with a projector because of scattered light.

Slow downward movement of the dark rectangle evoked a strong discharge throughout the entire 5° of the receptive field (Text-fig. 8*A*). If the movement was halted the cell continued to fire, but less vigorously.



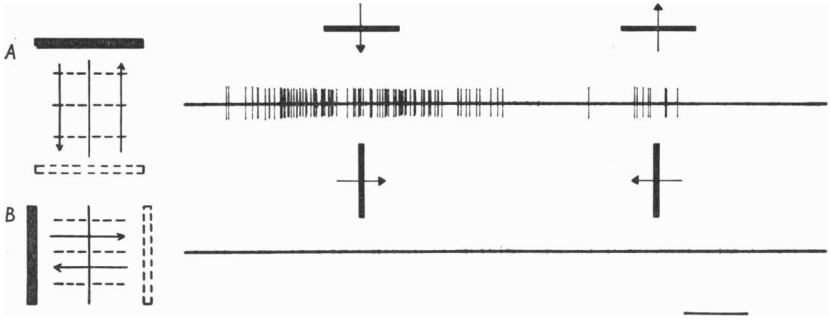
Text-fig. 7. Cell activated only by left (contralateral) eye over a field approximately 5 × 5°, situated 10° above and to the left of the area centralis. The cell responded best to a black horizontal rectangle, $\frac{1}{3} \times 6°$, placed anywhere in the receptive field (*A–C*). Tilting the stimulus rendered it ineffective (*D–E*). The black bar was introduced against a light background during periods of 1 sec, indicated by the upper line in each record. Luminance of white background, 1·0 $\log_{10}$ cd/m²; luminance of black part, 0·0 $\log_{10}$ cd/m². A lesion, made while recording from the cell, was found in layer 2 of apical segment of post-lateral gyrus.

Upward movement gave only weak, inconsistent responses, and left–right movement (Text-fig. 8*B*) gave no responses. Discharges of highest frequency were evoked by relatively slow rates of downward movement (about 5–10 sec to cross the entire field); rapid movement in either direction gave only very weak responses.

Despite its unusual features this cell exhibited several properties typical of complex units, particularly the lack of summation (except in a horizontal sense), and the wide area over which the dark bar was effective. One may think of the field as having a counterpart in simple fields of type *D*, Text-fig. 2. In such fields a dark bar would evoke discharges, but only if it fell within the inhibitory region. Moreover, downward movement of

the bar would also evoke brisker discharges than upward, provided the upper flanking region were stronger than the lower one.

In describing simple fields it has already been noted that moving stimuli were often more effective than stationary ones. This was also true of cells with complex fields. Depending on the cell, slits, edges, or dark bars were most effective. As with simple fields, orientation of a stimulus was always critical, responses varied with rate of movement, and directional asymmetries of the type seen in Text-fig. 8 were common. Only once have we seen activation of a cell for one direction of movement and suppression of



Text-fig. 8. Same cell as in Text-fig. 7. Movement of black rectangle $\frac{1}{3} \times 6°$ back and forth across the receptive field: *A*, horizontally oriented (parallel to receptive-field axis); *B*, vertically oriented. Time required to move across the field, 5 sec. Time, 1 sec.

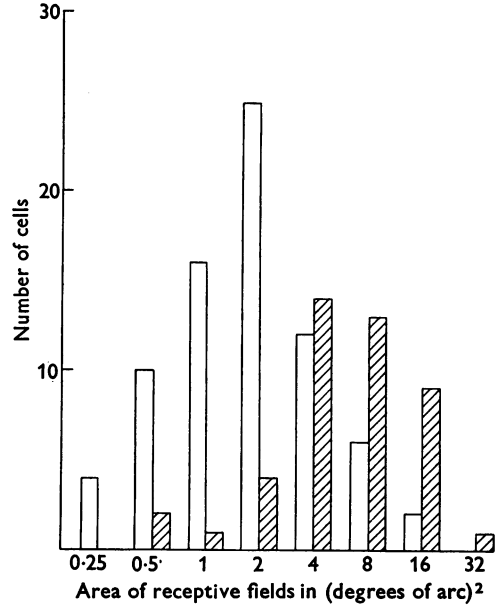
maintained firing for the opposite direction. In their responses to movement, cells with complex fields differed from their simple counterparts chiefly in responding with sustained firing over substantial regions, usually the entire receptive field, instead of over a very narrow boundary separating excitatory and inhibitory regions.

Receptive-field dimensions

Over-all field dimensions were measured for 119 cells. A cell was included only if its field was mapped completely, and if it was situated in the area of central vision (see p. 135). Fields varied greatly in size from one cell to the next, even for cells recorded in a single penetration (see Text-fig. 15). In Text-fig. 9 the distribution of cells according to field area is given separately for simple and complex fields. The histogram illustrates the variation in size, and shows that on the average complex fields were larger than simple ones.

Widths of the narrow subdivisions of simple fields (the centres of types *C*, *D* and *E* or the flanks of type *F*, Text-fig. 2) also varied greatly: the smallest were 10–15 minutes of arc, which is roughly the diameter of the smallest field centres we have found for geniculate cells. For some cells

with complex fields the widths of the most effective slits or dark bars were also of this order, indicating that despite the greater overall field size these cells were able to convey detailed information. We wish to emphasize that in both geniculate and cortex the field dimensions tend to increase with distance from the area centralis, and that they differ even for a given location in the retina. It is consequently not possible to compare field sizes in the geniculate and cortex unless these variations are taken into account. This may explain the discrepancy between our results and the findings of Baumgartner (see Jung, 1960), that 'field centres' in the cortex are one half the size of those in the lateral geniculate body.



Text-fig. 9. Distribution of 119 cells in the visual cortex with respect to the approximate area of their receptive fields. White columns indicate cells with simple receptive fields; shaded columns, cells with complex fields. Abscissa: area of receptive fields. Ordinate: number of cells.

Responsiveness of cortical cells

Simple and complex fields together account for all of the cells we have recorded in the visual cortex. We have not observed cells with concentric fields. Except for clearly injured cells (showing extreme spike deformation or prolonged high-frequency bursts of impulses) all units have responded to visual stimulation, though it has occasionally taken several hours to find the retinal region containing the receptive field and to work out the optimum stimuli. Some cells responded only to stimuli which were optimum in their retinal position and in their form, orientation and rate of

movement. A few even required stimulation of both eyes before a response could be elicited (see Part II). But there is no indication from our studies that the striate cortex contains nerve cells that are unresponsive to visual stimuli.

Most of the cells of this series were observed for 1 or 2 hr, and some were studied for up to 9 hr. Over these periods of time there were no qualitative changes in the characteristics of receptive fields: their complexity, arrangements of excitatory and inhibitory areas, axis orientation and position all remained the same, as did the ocular dominance. With deepening anaesthesia a cell became less responsive, so that stimuli that had formerly been weak tended to become even weaker or ineffective, while those that had evoked brisk responses now evoked only weak ones. The last thing to disappear with very deep anaesthesia was usually the response to a moving form. As long as any responses remained the cell retained the same specific requirements as to stimulus form, orientation and rate of movement, suggesting that however the drug exerted its effects, it did not to any important extent functionally disrupt the specific visual connexions. A comparison of visual responses in the anaesthetized animal with those in the unanaesthetized, unrestrained preparation (Hubel, 1959) shows that the main differences lie in the frequency and firing patterns of the maintained activity and in the vigour of responses, rather than in the basic receptive-field organization. It should be emphasized, however, that even in light anaesthesia or in the attentive state diffuse light remains relatively ineffective; thus the balance between excitatory and inhibitory influences is apparently maintained in the waking state.

PART II

BINOCULAR INTERACTION AND OCULAR DOMINANCE

Recording from single cells at various levels in the visual system offers a direct means of determining the site of convergence of impulses from the two eyes. In the lateral geniculate body, the first point at which convergence is at all likely, binocularly influenced cells have been observed, but it would seem that these constitute at most a small minority of the total population of geniculate cells (Erulkar & Fillenz, 1958, 1960; Bishop, Burke & Davis, 1959; Grüsser & Sauer, 1960; Hubel & Wiesel, 1961). Silver-degeneration studies show that in each layer of the geniculate the terminals of fibres from a single eye are grouped together, with only minor overlap in the interlaminar regions (Silva, 1956; Hayhow, 1958). The anatomical and physiological findings are thus in good agreement.

It has long been recognized that the greater part of the cat's primary visual cortex receives projections from the two eyes. The anatomical

evidence rests largely on the observation that cells in all three lateral geniculate layers degenerate following a localized lesion in the striate area (Minkowski, 1913). Physiological confirmation was obtained by Talbot & Marshall (1941) who stimulated the visual fields of the separate eyes with small spots of light, and mapped the evoked cortical slow waves. Still unsettled, however, was the question of whether individual cortical cells receive projections from both eyes, or whether the cortex contains a mixture of cells, some activated by one eye, some by the other. We have recently shown that many cells in the visual cortex can be influenced by both eyes (Hubel & Wiesel, 1959). The present section contains further observations on binocular interaction. We have been particularly interested in learning whether the eyes work in synergy or in opposition, how the relative influence of the two eyes varies from cell to cell, and whether, on the average, one eye exerts more influence than the other on the cells of a given hemisphere.

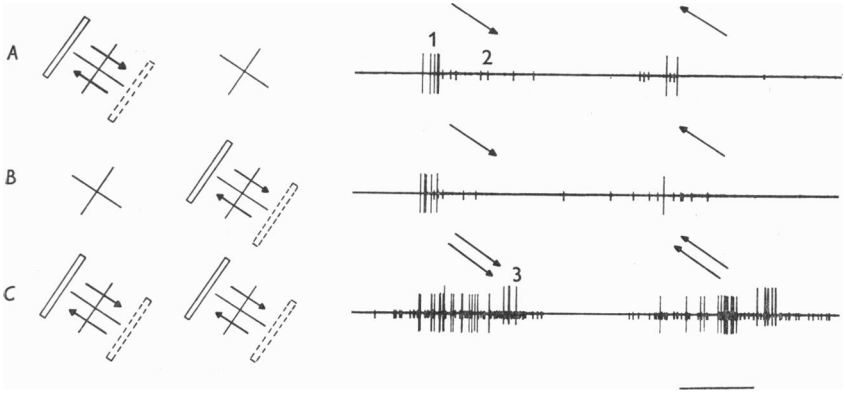
RESULTS

In agreement with previous findings (Hubel & Wiesel, 1959) the receptive fields of all binocularly influenced cortical cells occupied corresponding positions on the two retinas, and were strikingly similar in their organization. For simple fields the spatial arrangements of excitatory and inhibitory regions were the same; for complex fields the stimuli that excited or inhibited the cell through one eye had similar effects through the other. Axis orientations of the two receptive fields were the same. Indeed, the only differences ever seen between the two fields were related to eye dominance: identical stimuli to the two eyes did not necessarily evoke equally strong responses from a given cell. For some cells the responses were equal or almost so; for others one eye tended to dominate. Whenever the two retinas were stimulated in identical fashion in corresponding regions, their effects summed, i.e. they worked in synergy. On the other hand, if antagonistic regions in the two eyes were stimulated so that one eye had an excitatory effect and the other an inhibitory one, then the responses tended to cancel (Hubel & Wiesel, 1959, Fig. 10*A*).

Some units did not respond to stimulation of either eye alone but could be activated only by simultaneous stimulation of the two eyes. Text-figure 10 shows an example of this, and also illustrates ordinary binocular synergy. Two simultaneously recorded cells both responded best to transverse movement of a rectangle oriented in a 1 o'clock–7 o'clock direction (Text-fig. 10*A*, *B*). For one of the cells movement down and to the right was more effective than movement up and to the left. Responses from the individual eyes were roughly equal. On simultaneous stimulation of the two eyes both units responded much more vigorously. Now a third cell was also activated.

The threshold of this third unit was apparently so high that, at least under these experimental conditions, stimulation of either eye alone failed to evoke any response.

A second example of synergy is seen in Text-fig. 11. The most effective stimulus was a vertically oriented rectangle moved across the receptive field from left to right. Here the use of both eyes not only enhanced the response already observed with a single eye, but brought into the open a tendency that was formerly unsuspected. Each eye mediated a weak

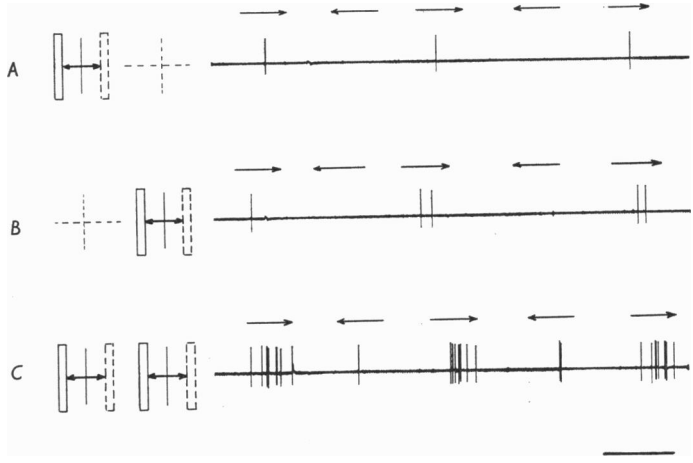


Text-fig. 10. Examples of binocular synergy in a simultaneous recording of three cells (spikes of the three cells are labelled 1–3). Each of the cells had receptive fields in the two eyes; in each eye the three fields overlapped and were situated 2° below and to the left of the area centralis. The crosses to the left of each record indicate the positions of the receptive fields in the two eyes. The stimulus was $\frac{1}{8} \times 2°$ slit oriented obliquely and moved slowly across the receptive fields as shown; *A*, in the left eye; *B*, in the right eye; *C*, in the two eyes simultaneously. Since the responses in the two eyes were about equally strong, these two cells were classed in ocular-dominance group 4 (see Text-fig. 12). Time, 1 sec.

response (Text-fig. 11*A*, *B*) which was greatly strengthened when both eyes were used in parallel (*C*). Now, in addition, the cell gave a weak response to leftward movement, indicating that this had an excitatory effect rather than an inhibitory one. Binocular synergy was often a useful means of bringing out additional information about a receptive field.

In our previous study of forty-five cortical cells (Hubel & Wiesel, 1959) there was clear evidence of convergence of influences from the two eyes in only one fifth of the cells. In the present series 84 % of the cells fell into this category. The difference is undoubtedly related to the improved precision in technique of binocular stimulation. A field was first mapped in the dominant eye and the most effective type of stimulus determined. That stimulus was then applied in the corresponding region in the other

eye. Finally, even if no response was obtained from the non-dominant eye, the two eyes were stimulated together in parallel to see if their effects were synergistic. With these methods, an influence was frequently observed from the non-dominant eye that might otherwise have been overlooked.



Text-fig. 11. Movement of a $\frac{1}{4} \times 2°$ slit back and forth horizontally across the receptive field of a binocularly influenced cell. *A*, left eye; *B*, right eye; *C*, both eyes. The cell clearly preferred left-to-right movement, but when both eyes were stimulated together it responded also to the reverse direction. Field diameter, 2°, situated 5° from the area centralis. Time, 1 sec.

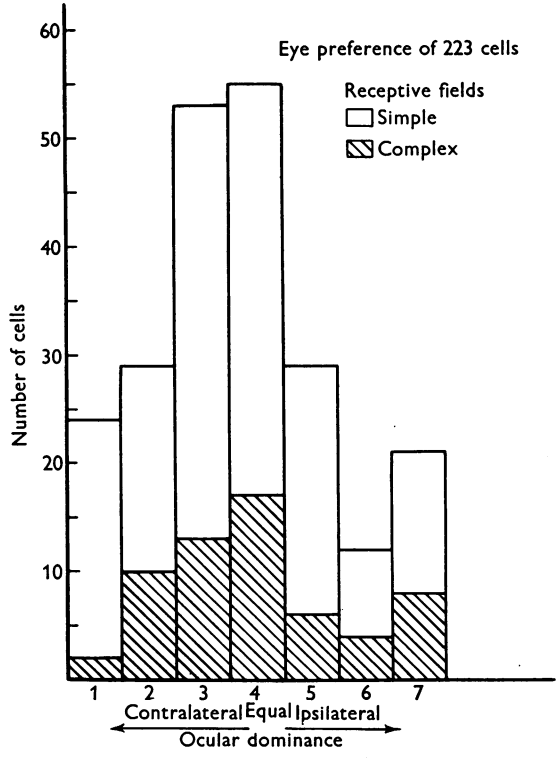
A comparison of the influence of the two eyes was made for 223 of the 303 cells in the present series. The remaining cells were either not sufficiently studied, or they belonged to the small group of cells which were only activated if both eyes were simultaneously stimulated. The fields of all cells were in or near the area centralis. The 223 cells were subdivided into seven groups, as follows:

Group	Ocular dominance
1	Exclusively contralateral
2*	Contralateral eye much more effective than ipsilateral eye
3	Contralateral eye slightly more effective than ipsilateral
4	No obvious difference in the effects exerted by the two eyes
5	Ipsilateral eye slightly more effective
6*	Ipsilateral eye much more effective
7	Exclusively ipsilateral

* These groups include cells in which the non-dominant eye, ineffective by itself, could influence the response to stimulation of the dominant eye.

A histogram showing the distribution of cells among these seven groups is given in Text-fig. 12. Assignment of a unit to a particular group was to some extent arbitrary, but it is unlikely that many cells were misplaced by more than one group. Perhaps the most interesting feature of the

histogram is its lack of symmetry: many more cells were dominated by the contralateral than by the ipsilateral eye (106 vs. 62). We conclude that in the part of the cat's striate cortex representing central vision the great majority of cells are influenced by both eyes, and that despite wide variation in relative ocular dominance from one cell to the next, the contralateral eye is, on the average, more influential. As the shaded portion



Text-fig. 12. Distribution of 223 cells recorded from the visual cortex, according to ocular dominance. Histogram includes cells with simple fields and cells with complex fields. The shaded region shows the distribution of cells with complex receptive fields. Cells of group 1 were driven only by the contralateral eye; for cells of group 2 there was marked dominance of the contralateral eye, for group 3, slight dominance. For cells in group 4 there was no obvious difference between the two eyes. In group 5 the ipsilateral eye dominated slightly, in group 6, markedly; and in group 7 the cells were driven only by the ipsilateral eye.

of Text-fig. 12 shows, there is no indication that the distribution among the various dominance groups of cells having complex receptive fields differs from the distribution of the population as a whole.

A cortical bias in favour of the contralateral eye may perhaps be related to the preponderance of crossed over uncrossed fibres in the cat's optic

tract (Polyak, 1957, p. 788). The numerical inequality between crossed and uncrossed tract fibres is generally thought to be related to an inequality in size of the nasal and temporal half-fields, since both inequalities are most marked in lower mammals with laterally placed eyes, and become progressively less important in higher mammals, primates and man. Thompson *et al.* (1950) showed that in the rabbit, for example, there is a substantial cortical region receiving projections from that part of the peripheral contralateral visual field which is not represented in the ipsilateral retina (the 'Temporal Crescent'). Our results, concerned with more central portions of the visual fields, suggest that in the cat the difference in the number of crossed and uncrossed fibres in an optic tract is probably not accounted for entirely by fibres having their receptive fields in the temporal-field crescents.

PART III

FUNCTIONAL CYTOARCHITECTURE OF THE CAT'S VISUAL CORTEX

In the first two parts of this paper cells were studied individually, no attention being paid to their grouping within the cortex. We have shown that the number of functional cell types is very large, since cells may differ in several independent physiological characteristics, for example, in the retinal position of their receptive fields, their receptive-field organization, their axis orientation, and their ocular-dominance group. In this section we shall try to determine whether cells are scattered at random through the cortex with regard to these characteristics, or whether there is any tendency for one or more of the characteristics to be shared by neighbouring cells. The functional architecture of the cortex not only seems interesting for its own sake, but also helps to account for the various complex response patterns described in Part I.

RESULTS

Functional architecture of the cortex was studied by three methods. These had different merits and limitations, and were to some extent complementary.

(1) *Cells recorded in sequence.* The most useful and convenient procedure was to gather as much information as possible about each of a long succession of cells encountered in a micro-electrode penetration through the cortex, and to reconstruct the electrode track from serial histological sections. One could then determine how a physiological characteristic (such as receptive-field position, organization, axis orientation or ocular dominance) varied with cortical location. The success of this method in

delineating regions of constant physiological characteristics depends on the possibility of examining a number of units as the electrode passes through each region. Regions may escape detection if they are so small that the electrode is able to resolve only one or two cells in each. The fewer the cells resolved, the larger the regions must be in order to be detected at all.

(2) *Unresolved background activity.* To some extent the spaces between isolated units were bridged by studying unresolved background activity audible over the monitor as a crackling noise, and assumed to originate largely from action potentials of a number of cells. It was concluded that cells, rather than fibres, gave rise to this activity, since it ceased abruptly when the electrode left the grey matter and entered subcortical white matter. Furthermore, diffuse light evoked no change in activity, compared to the marked increase caused by an optimally oriented slit. This suggested that terminal arborizations of afferent fibres contributed little to the background, since most geniculate cells respond actively to diffuse light (Hubel, 1960). In most penetrations unresolved background activity was present continuously as the electrode passed through layers 2–6 of the cortical grey matter.

Background activity had many uses. It indicated when the cells within range of the electrode tip had a common receptive-field axis orientation. Similarly, one could use it to tell whether the cells in the neighbourhood were driven exclusively by one eye (group 1 or group 7). When the background activity was influenced by both eyes, one could not distinguish between a mixture of cells belonging to the two monocular groups (1 and 7) and a population in which each cell was driven from both eyes. But even here one could at least assess the relative influence of the two eyes upon the group of cells in the immediate neighbourhood of the electrode.

(3) *Multiple recordings.* In the series of 303 cells, 78 were recorded in groups of two and 12 in groups of three. Records were not regarded as multiple unless the spikes of the different cells showed distinct differences in amplitude, and unless each unit fulfilled the criteria required of a single-unit record, namely that the amplitude and wave shape be relatively constant for a given electrode position.

In such multiple recordings one could be confident that the cells were close neighbours and that uniform stimulus conditions prevailed, since the cells could be stimulated and observed together. One thus avoided some of the difficulties in evaluating a succession of recordings made over a long period of time span, where absolute constancy of eye position, anaesthetic level, and preparation condition were sometimes hard to guarantee.

Regional variations of several physiological characteristics were examined by the three methods just outlined. Of particular interest for the

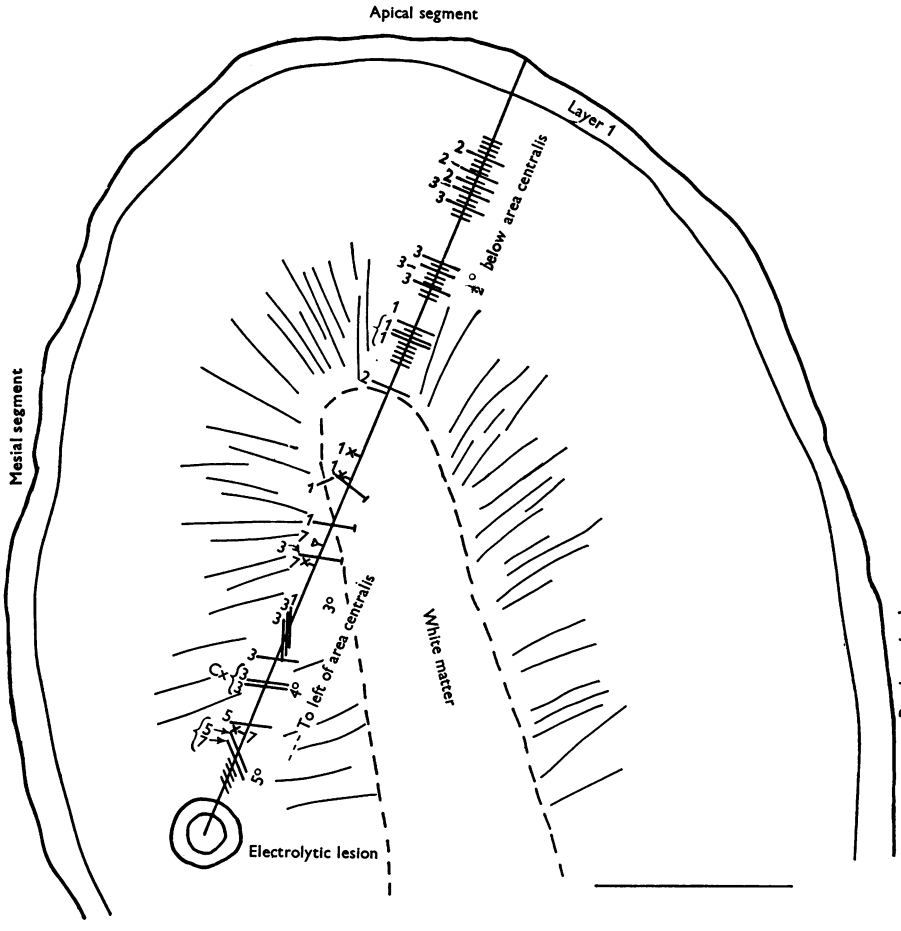
present study were the receptive-field axis orientation, position of receptive fields on the retina, receptive-field organization, and relative ocular dominance. These will be described separately in the following paragraphs.

Orientation of receptive-field axis

The orientation of a receptive-field axis was determined in several ways. When the field was simple the borders between excitatory and inhibitory regions were sufficient to establish the axis directly. For both simple and complex fields the axis could always be determined from the orientation of the most effective stimulus. For most fields, when the slit or edge was placed at right angles to the optimum position there was no response. The receptive-field axis orientation was checked by varying the stimulus orientation from this null position in order to find the two orientations at which a response was only just elicited, and by bisecting the angle between them. By one or other of these procedures the receptive-field orientation could usually be determined to within 5 or 10°.

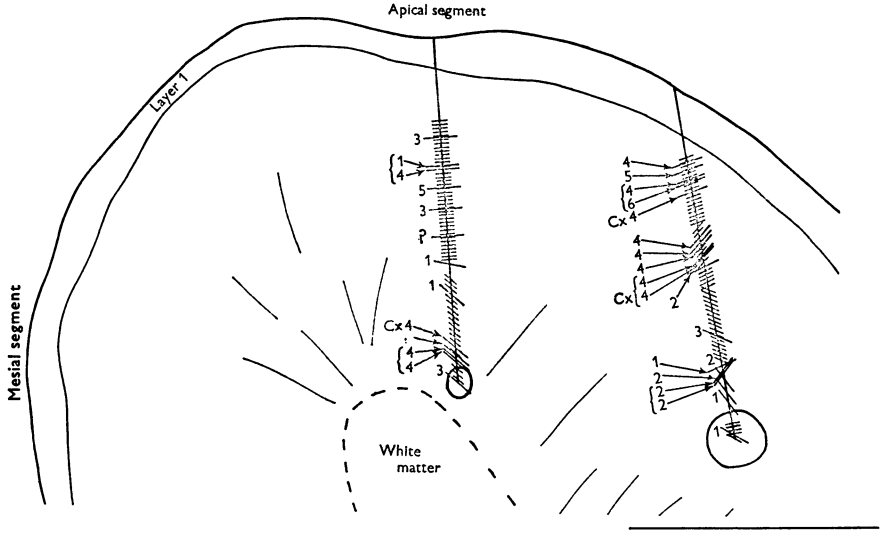
One of the first indications that the orientation of a receptive-field axis was an important variable came from multiple recordings. Invariably the axes of receptive fields mapped together had the same orientations. An example of a 3-unit recording has already been given in Text-fig. 10. Cells with common axis orientation were therefore not scattered at random through the cortex, but tended to be grouped together. The size and shape of the regions containing these cell groups were investigated by comparing the fields of cells mapped in sequence. It was at once apparent that successively recorded cells also tended to have identical axis orientations and that each penetration consisted of several sequences of cells, each sequence having a common axis orientation. Any undifferentiated units in the background responded best to the stimulus orientation that was most effective in activating the cell under study. After traversing a distance that varied greatly from one penetration to the next, the electrode would enter an area where there was no longer any single optimum orientation for driving background activity. A very slight advance of the electrode would bring it into a region where a new orientation was the most effective, and the succeeding cells would all have receptive fields with that orientation. The change in angle from one region to another was unpredictable; sometimes it was barely detectable, at other times large (45–90°).

Text-figure 13 shows a camera lucida tracing of a frontal section through the post-lateral gyrus. The electrode track entered normal to the surface, passed through the apical segment in a direction parallel to the fibre bundles, then through the white matter beneath, and finally obliquely through half the thickness of the mesial segment. A lesion was made at the termination of the penetration. A composite photomicrograph (Pl. 1) shows the lesion



Text-fig. 13. Reconstruction of micro-electrode penetration through the lateral gyrus (see also Pl. 1). Electrode entered apical segment normal to the surface, and remained parallel to the deep fibre bundles (indicated by radial lines) until reaching white matter; in grey matter of mesial segment the electrode's course was oblique. Longer lines represent cortical cells. Axons of cortical cells are indicated by a cross-bar at right-hand end of line. Field-axis orientation is shown by the direction of each line; lines perpendicular to track represent vertical orientation. Brace-brackets show simultaneously recorded units. Complex receptive fields are indicated by 'Cx'. Afferent fibres from the lateral geniculate body indicated by ×, for 'on' centre; Δ, for 'off' centre. Approximate positions of receptive fields on the retina are shown to the right of the penetration. Shorter lines show regions in which unresolved background activity was observed. Numbers to the left of the penetration refer to ocular-dominance group (see Part II). Scale 1 mm.

and the first part of the electrode track. The units recorded in the course of the penetration are indicated in Text-fig. 13 by the longer lines crossing the track; the unresolved background activity by the shorter lines. The orientations of the most effective stimuli are given by the directions of the lines, a line perpendicular to the track signifying a vertical orientation. For the first part of the penetration, through the apical segment, the field orientation was vertical for all cells as well as for the background activity.



Text-fig. 14. Reconstructions of two penetrations in apical segment of post-lateral gyrus, near its anterior end (just behind anterior interrupted line in Text-fig. 1, see also Pl. 2). Medial penetration is slightly oblique, lateral one is markedly so. All receptive fields were located within 1° of area centralis. Conventions as in Text-fig. 13. Scale 1 mm.

Fibres were recorded from the white matter and from the grey matter just beyond it. Three of these fibres were axons of cortical cells having fields of various oblique orientations; four were afferent fibres from the lateral geniculate body. In the mesial segment three short sequences were encountered, each with a different common field orientation. These sequences together occupied a distance smaller than the full thickness of the apical segment.

In another experiment, illustrated in Text-fig. 14 and in Pl. 2, two penetrations were made, both in the apical segment of the post-lateral gyrus. The medial penetration (at left in the figure) was at the outset almost normal to the cortex, but deviated more and more from the direction of the deep fibre bundles. In this penetration there were three different axis orientations, of which the first and third persisted through long sequences.

In the lateral track there were nine orientations. From the beginning this track was more oolique, and it became increasingly so as it progressed.

As illustrated by the examples of Text-figs. 13 and 14, there was a marked tendency for shifts in orientation to increase in frequency as the angle between electrode and direction of fibre bundles (or apical dendrites) became greater. The extreme curvature of the lateral and post-lateral gyri in their apical segments made normal penetrations very difficult to obtain; nevertheless, four penetrations were normal or almost so. In none of these were there any shifts of axis orientation. On the other hand there were several shifts of field orientation in all oblique penetrations. As illustrated by Text-fig. 14, most penetrations that began nearly normal to the surface became more and more oblique with increasing depth. Here the distance traversed by the electrode without shifts in receptive-field orientation tended to become less and less as the penetration advanced.

It can be concluded that the striate cortex is divided into discrete regions within which the cells have a common receptive-field axis orientation. Some of the regions extend from the surface of the cortex to the white matter; it is difficult to be certain whether they all do. Some idea of their shapes may be obtained by measuring distances between shifts in receptive-field orientation. From these measurements it seems likely that the general shape is columnar, distorted no doubt by any curvature of the gyrus, which would tend to make the end at the surface broader than that at the white matter; deep in a sulcus the effect would be the reverse. The cross-sectional size and shape of the columns at the surface can be estimated only roughly. Most of our information concerns their width in the coronal plane, since it is in this plane that oblique penetrations were made. At the surface this width is probably of the order of 0·5 mm. We have very little information about the cross-sectional dimension in a direction parallel to the long axis of the gyrus. Preliminary mapping of the cortical surface suggests that the cross-sectional shape of the columns may be very irregular.

Position of receptive fields on the retina

Gross topography. That there is a systematic representation of the retina on the striate cortex of the cat was established anatomically by Minkowski (1913) and with physiological methods by Talbot & Marshall (1941). Although in the present study no attempt has been made to map topographically all parts of the striate cortex, the few penetrations made in cortical areas representing peripheral parts of the retina confirm these findings. Cells recorded in front of the anterior interrupted lines of Text-fig. 1 had receptive fields in the superior retinas; those in the one penetration behind the posterior line had fields that were well below the horizontal

meridian of the retina. (No recordings were made from cortical regions receiving projections from the deeply pigmented non-tapetal part of the inferior retinas.) In several penetrations extending far down the mesial (interhemispheric) segment of the lateral gyrus, receptive fields moved further and further out into the ipsilateral half of each retina as the electrode advanced (Text-fig. 13). In these penetrations the movement of fields into the retinal periphery occurred more and more rapidly as the electrode advanced. In three penetrations extending far down the lateral segment of the post-lateral gyrus (medial bank of the post-lateral sulcus) there was likewise a clear progressive shift of receptive-field positions as the electrode advanced. Here also the movement was along the horizontal meridian, again into the *ipsilateral* halves of both retinas. This therefore confirms the findings of Talbot & Marshall (1941) and Talbot (1942), that in each hemisphere there is a second laterally placed representation of the contralateral half-field of vision. The subject of Visual Area II will not be dealt with further in this paper.

Cells within the large cortical region lying between the interrupted lines of Text-fig. 1, and extending over on to the mesial segment and into the lateral sulcus for a distance of 2–3 mm, had their receptive fields in the area of central vision. By this we mean the area centralis, which is about 5° in diameter, and a region surrounding it by about 2–3°. The receptive fields of the great majority of cells were confined to the ipsilateral halves of the two retinas. Often a receptive field covering several degrees on the retina stopped short in the area centralis right at the vertical meridian. Only rarely did a receptive field appear to spill over into the contralateral half-retina; when it did, it was only by 2–3°, a distance comparable to the possible error in determining the area centralis in some cats.

Because of the large cortical representation of the area centralis, one would expect only a very slow change in receptive-field position as the electrode advanced obliquely (Text-fig. 13). Indeed, in penetrations through the apex of the post-lateral gyrus and extending 1–2 mm down either bank there was usually no detectable progressive displacement of receptive fields. In penetrations made 1–3 mm apart, either along a parasagittal line or in the same coronal plane (Text-fig. 14) receptive fields again had almost identical retinal positions.

Retinal representation of neighbouring cells. A question of some interest was to determine whether this detailed topographic representation of the retina held right down to the cellular level. From the results just described one might imagine that receptive fields of neighbouring cortical cells should have very nearly the same retinal position. In a sequence of cells recorded in a normal penetration through the cortex the receptive fields should be superimposed, and for oblique penetrations any detectable

changes in field positions should be systematic. In the following paragraphs we shall consider the relative retinal positions of the receptive fields of neighbouring cells, especially cells within a column.

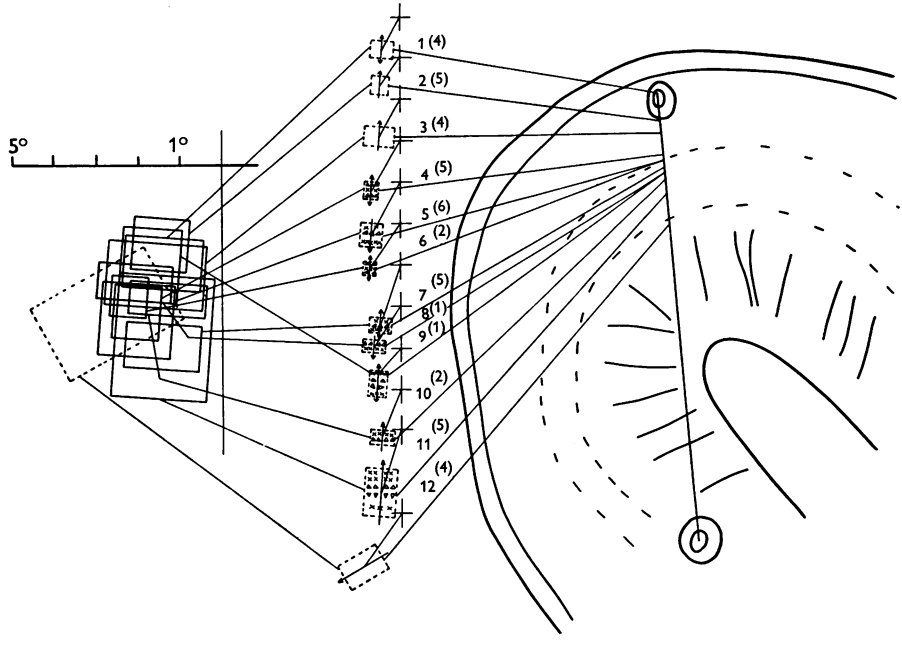
In all multiple recordings the receptive fields of cells observed simultaneously were situated in the same general region of the retina. As a rule the fields overlapped, but it was unusual for them to be precisely superimposed. For example, fields were often staggered along a line perpendicular to their axes. Similarly, the successive receptive fields observed during a long cortical penetration varied somewhat in position, often in an apparently random manner. Text-figure 15 illustrates a sequence of twelve cells recorded in the early part of a penetration through the cortex. One lesion was made while observing the first cell in the sequence and another at the end of the penetration; they are indicated in the drawing of cortex to the right of the figure. In the centre of the figure the position of each receptive field is shown relative to the area centralis (marked with a cross); each field was several degrees below and to the left of the area centralis. It will be seen that all fields in the sequence except the last had the same axis orientation; the first eleven cells therefore occupied the same column. All but the first three and the last (cell 12) were simple in arrangement. Cells 5 and 6 were recorded together, as were 8 and 9.

In the left-hand part of the figure the approximate boundaries of all these receptive fields are shown superimposed, in order to indicate the degree of overlap. From cell to cell there is no obvious systematic change in receptive-field position. The variation in position is about equal to the area occupied by the largest fields of the sequence. This variation is undoubtedly real, and not an artifact produced by eye movements occurring between recordings of successive cells. The stability of the eyes was checked while studying each cell, and any tendency to eye movements would have easily been detected by an apparent movement of the receptive field under observation. Furthermore, the field positions of simultaneously recorded cells 5 and 6, and also of cells 8 and 9, are clearly different; here the question of eye movements is not pertinent.

Text-figure 15 illustrates a consistent and somewhat surprising finding, that within a column defined by common field-axis orientation there was no apparent progression in field positions along the retina as the electrode advanced. This was so even though the electrode often crossed through the column obliquely, entering one side and leaving the other. If there was any detailed topographical representation within columns it was obscured by the superimposed, apparently random staggering of field positions. We conclude that at this microscopic level the retinotopic representation no longer strictly holds.

Receptive-field organization

Multiple recordings. The receptive fields of cells observed together in multiple recordings were always of similar complexity, i.e. they were either all simple or all complex in their organization. In about one third of the multiple recordings the cells had the same detailed field organization; if simple, they had similar distributions of excitatory and inhibitory

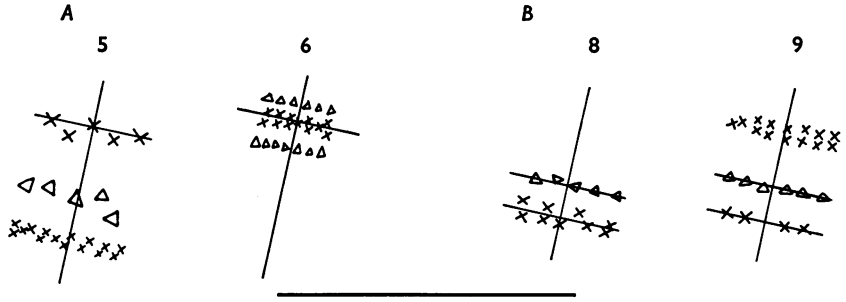


Text-fig. 15. Reconstruction of part of an electrode track through apical and mesial segments of post-lateral gyrus near its anterior end. Two lesions were made, the first after recording from the first unit, the second at the end of the penetration. Only the first twelve cells are represented. Interrupted lines show boundaries of layer 4.

In the centre part of the figure the position of each receptive field, outlined with interrupted lines, is given with respect to the area centralis, shown by a cross. Cells are numbered in sequence, 1–12. Numbers in parentheses refer to ocular-dominance group (see Part II). Units 5 and 6, 8 and 9 were observed simultaneously. The first three fields and the last were complex in organization; the remainder were simple. ×, areas giving excitation; △, areas giving inhibitory effects. Note that all receptive fields except the last have the same axis orientation (9.30–3.30 o'clock). The arrows show the preferred direction of movement of a slit oriented parallel to the receptive-field axis.

In the left part of the figure all of the receptive fields are superimposed, to indicate the overlap and variation in size. The vertical and horizontal lines represent meridia, crossing at the area centralis. Scale on horizontal meridian, 1° for each subdivision.

areas; if complex, they required identical stimuli for their activation. As a rule these fields did not have exactly the same retinal position, but were staggered as described above. In two thirds of the multiple recordings the cells differed to varying degrees in their receptive field arrangements. Two types of multiple recordings in which field arrangements differed seem interesting enough to merit a separate description.

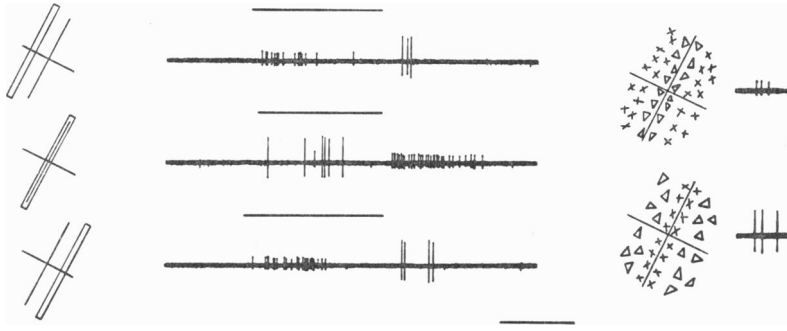


Text-fig. 16. Detailed arrangements of the receptive fields of two pairs of simultaneously recorded cells (nos. 5 and 6, and 8 and 9, of Text-fig. 15). The crosses of diagrams 5 and 6 are superimposed as are the double crosses of 8 and 9. Note that the upper excitatory region of 5 is superimposed upon the excitatory region of 6; and that both regions of 8 are superimposed on the inhibitory and lower excitatory regions of 9. Scale, 1°.

In several multiple recordings the receptive fields overlapped in such a way that one or more excitatory or inhibitory portions were superimposed. Two examples are supplied by cell-pairs 5 and 6, and 8 and 9 of Text-fig. 15. Their fields are redrawn in Text-fig. 16. The fields of cells 5 and 6 are drawn separately (Text-fig. 16*A*) but they actually overlapped so that the reference lines are to be imagined as superimposed. Thus the 'on' centre of cell 6 fell directly over the upper 'on' flank of 5 and the two cells tended to fire together to suitably placed stimuli. A similar situation existed for cells 8 and 9 (Text-fig. 16*B*). The field of 9 was placed so that its 'off' region and the lower, weaker 'on' region were superimposed on the two regions of 8. Again the two cells tended to fire together. Such examples suggest that neighbouring cells may have some of their inputs in common.

Cells responded reciprocally to a light stimulus in eight of the forty-three multiple recordings. An example of two cells responding reciprocally to stationary spots is shown in Text-fig. 17. In each eye the two receptive fields were almost superimposed. The fields consisted of elongated obliquely oriented central regions, inhibitory for one cell, excitatory for the other, flanked on either side by regions of the opposite type. Instead of firing together in response to an optimally oriented stationary slit, like the cells

in Text-fig. 16, these cells gave opposite-type responses, one inhibitory and the other excitatory. Some cell pairs responded reciprocally to to-and-fro movements of a slit or edge. Examples have been given elsewhere (Hubel, 1958, Fig. 9; 1959, Text-fig. 6). The fields of these cell pairs usually differed only in the balance of the asymmetrical flanking regions.



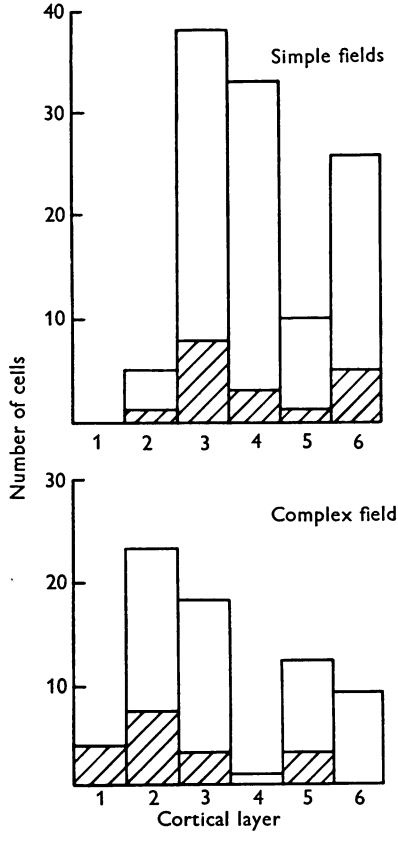
Text-fig. 17. Records of two simultaneously observed cells which responded reciprocally to stationary stimuli. The two receptive fields are shown to the right, and are superimposed, though they are drawn separately. The cell corresponding to each field is indicated by the spikes to the right of the diagram. To the left of each record is shown the position of a slit, $\frac{1}{4} \times 2\frac{1}{2}°$, with respect to these fields.

Both cells binocularly driven (dominance group 3); fields mapped in the left (contralateral) eye; position of fields 2° below and to the left of the area centralis. Time, 1 sec.

Relationship between receptive field organization and cortical layering. In a typical penetration through the cortex many different field types were found, some simple and others complex. Even within a single column both simple and complex fields were seen. (In Text-fig. 13 and 14 complex fields are indicated by the symbol 'Cx'; in Text-fig. 15, fields 1–3 were complex and 4–11 simple, all within a single column.) An attempt was made to learn whether there was any relationship between the different field types and the layers of the cortex. This was difficult for several reasons. In Nissl-stained sections the boundaries between layers of the cat's striate cortex are not nearly as clear as they are in the primate brain; frequently even the fourth layer, so characteristic of the striate cortex, is poorly demarcated. Consequently, a layer could not always be identified with certainty even for a cell whose position was directly marked by a lesion. For most cells the positions were arrived at indirectly, from depth readings and lesions made elsewhere in the penetrations: these determinations were subject to more errors than the direct ones. Moreover, few of the penetrations were made in a direction parallel to the layering, so that the distance an electrode travelled in passing through a layer was

short, and the error in electrode position correspondingly more important.

The distribution of 179 cells among the different layers is given in the histograms of Text-fig. 18. All cells were recorded in penetrations in which at least one lesion was made; the shaded portions refer to cells which were individually marked with lesions. As shown in the separate histograms, simple-field cells as well as those with complex fields were widely distributed throughout the cortex. Cells with simple fields were most numerous in layers 3, 4 and 6. Especially interesting is the apparent rarity of complex fields in layer 4, where simple fields were so abundant. This is also illustrated in Text-fig. 15, which shows a sequence of eight cells



Text-fig. 18. Distribution of 179 cells, 113 with simple fields, 66 with complex, among the different cortical layers. All cells were recorded in penetrations in which at least one electrolytic lesion was made and identified; the shaded areas refer to cells marked individually by lesions. Note especially the marked difference in the occurrence, in layer 4, between simple and complex fields.

recorded from layer 4, all of which had simple fields. These findings suggest that cells may to some extent be segregated according to field complexity, and the rarity with which simple and complex fields were mapped together is consistent with this possibility.

Ocular dominance

In thirty-four multiple recordings the eye-dominance group (see Part II) was determined for both or all three cells. In eleven of these recordings there was a clear difference in ocular dominance between cells. Similarly, in a single penetration two cells recorded in sequence frequently differed in eye dominance. Cells from several different eye-dominance categories appeared not only in single penetrations, but also in sequences in which all cells had a common axis orientation. Thus within a single column defined by a common axis orientation there were cells of different eye dominance. A sequence of cells within one column is formed by cells 1–11 of Text-fig. 15. Here eye dominance ranged from wholly contralateral (group 1) to strongly ipsilateral (group 6). The two simultaneously recorded cells 5 and 6 were dominated by opposite eyes.

While these results suggested that cells of different ocular dominance were present within single columns, there were nevertheless indications of some grouping. First, in twenty-three of the thirty-four multiple recordings, simultaneously observed cells fell into the same ocular-dominance group. Secondly, in many penetrations short sequences of cells having the same relative eye dominance were probably more common than would be expected from a random scattering. Several short sequences are shown in Text-fig. 13 and 14. When such sequences consisted of cells with extreme unilateral dominance (dominance groups 1, 2, 6, and 7) the undifferentiated background activity was usually also driven predominantly by one eye, suggesting that other neighbouring units had similar eye preference. If cells of common eye dominance are in fact regionally grouped, the groups would seem to be relatively small. The cells could be arranged in nests, or conceivably in very narrow columns or thin layers.

In summary, cells within a column defined by a common field-axis orientation do not necessarily all have the same ocular dominance; yet neither do cells seem to be scattered at random through the cortex with respect to this characteristic.

DISCUSSION

A scheme for the elaboration of simple and complex receptive fields

Comparison of responses of cells in the lateral geniculate body with responses from striate cortex brings out profound differences in the receptive-field organization of cells in the two structures. For cortical

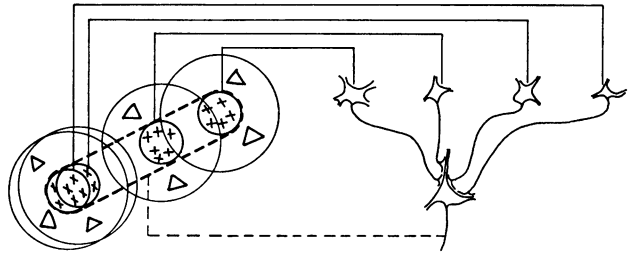
cells, specifically oriented lines and borders tend to replace circular spots as the optimum stimuli, movement becomes an important parameter of stimulation, diffuse light becomes virtually ineffective, and with adequate stimuli most cells can be driven from the two eyes. Since lateral geniculate cells supply the main, and possibly the only, visual input to the striate cortex, these differences must be the result of integrative mechanisms within the striate cortex itself.

At present we have no direct evidence on how the cortex transforms the incoming visual information. Ideally, one should determine the properties of a cortical cell, and then examine one by one the receptive fields of all the afferents projecting upon that cell. In the lateral geniculate, where one can, in effect, record simultaneously from a cell and one of its afferents, a beginning has already been made in this direction (Hubel & Wiesel, 1961). In a structure as complex as the cortex the techniques available would seem hopelessly inadequate for such an approach. Here we must rely on less direct evidence to suggest possible mechanisms for explaining the transformations that we find.

The relative lack of complexity of simple cortical receptive fields suggests that these represent the first or at least a very early stage in the modification of geniculate signals. At any rate we have found no cells with receptive fields intermediate in type between geniculate and simple cortical fields. To account for the spatial arrangements of excitatory and inhibitory regions of simple cortical fields we may imagine that upon each simple-type cell there converge fibres of geniculate origin having 'on' or 'off' centres situated in the appropriate retinal regions. For example, a cortical cell with a receptive field of the type shown in Text-fig. 2*C* might receive projections from a group of lateral geniculate cells having 'on' field centres distributed throughout the long narrow central region designated in the figure by crosses. Such a projection system is shown in the diagram of Text-fig. 19. A slit of light falling on this elongated central region would activate all the geniculate cells, since for each cell the centre effect would strongly outweigh the inhibition from the segments of field periphery falling within the elongated region. This is the same as saying that a geniculate cell will respond to a slit with a width equal to the diameter of its field centre, a fact that we have repeatedly verified. The inhibitory flanks of the cortical field would be formed by the remaining outlying parts of the geniculate-field peripheries. These flanks might be reinforced and enlarged by appropriately placed 'off'-centre geniculate cells. Such an increase in the potency of the flanks would appear necessary to explain the relative indifference of cortical cells to diffuse light.

The arrangement suggested by Text-fig. 19 would be consistent with our impression that widths of cortical receptive-field centres (or flanks, in a

field such as that of Text-fig. 2*F*) are of the same order of magnitude as the diameters of geniculate receptive-field centres, at least for fields in or near the area centralis. Hence the fineness of discrimination implied by the small size of geniculate receptive-field centres is not necessarily lost at the cortical level, despite the relatively large total size of many cortical fields; rather, it is incorporated into the detailed substructure of the cortical fields.



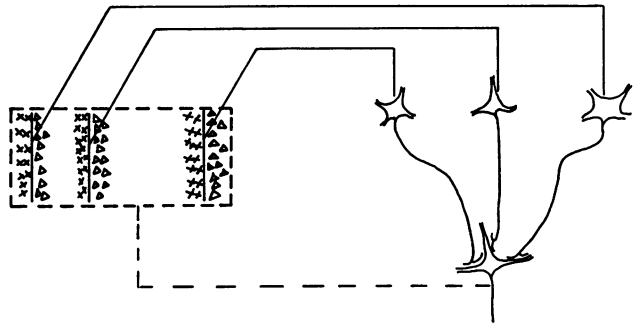
Text-fig. 19. Possible scheme for explaining the organization of simple receptive fields. A large number of lateral geniculate cells, of which four are illustrated in the upper right in the figure, have receptive fields with 'on' centres arranged along a straight line on the retina. All of these project upon a single cortical cell, and the synapses are supposed to be excitatory. The receptive field of the cortical cell will then have an elongated 'on' centre indicated by the interrupted lines in the receptive-field diagram to the left of the figure.

In a similar way, the simple fields of Text-fig. 2*D–G* may be constructed by supposing that the afferent 'on'- or 'off'-centre geniculate cells have their field centres appropriately placed. For example, field-type *G* could be formed by having geniculate afferents with 'off' centres situated in the region below and to the right of the boundary, and 'on' centres above and to the left. An asymmetry of flanking regions, as in field *E*, would be produced if the two flanks were unequally reinforced by 'on'-centre afferents.

The model of Text-fig. 19 is based on excitatory synapses. Here the suppression of firing on illuminating an inhibitory part of the receptive field is presumed to be the result of withdrawal of tonic excitation, i.e. the inhibition takes place at a lower level. That such mechanisms occur in the visual system is clear from studies of the lateral geniculate body, where an 'off'-centre cell is suppressed on illuminating its field centre because of suppression of firing in its main excitatory afferent (Hubel & Wiesel, 1961). In the proposed scheme one should, however, consider the possibility of direct inhibitory connexions. In Text-fig. 19 we may replace any of the excitatory endings by inhibitory ones, provided we replace the corresponding geniculate cells by ones of opposite type ('on'-centre instead of 'off'-centre, and conversely). Up to the present the two mechanisms have

not been distinguished, but there is no reason to think that both do not occur.

The properties of complex fields are not easily accounted for by supposing that these cells receive afferents directly from the lateral geniculate body. Rather, the correspondence between simple and complex fields noted in Part I suggests that cells with complex fields are of higher order, having cells with simple fields as their afferents. These simple fields would all have identical axis orientation, but would differ from one another in their exact retinal positions. An example of such a scheme is given in Text-fig. 20. The hypothetical cell illustrated has a complex field like that



Text-fig. 20. Possible scheme for explaining the organization of complex receptive fields. A number of cells with simple fields, of which three are shown schematically, are imagined to project to a single cortical cell of higher order. Each projecting neurone has a receptive field arranged as shown to the left: an excitatory region to the left and an inhibitory region to the right of a vertical straight-line boundary. The boundaries of the fields are staggered within an area outlined by the interrupted lines. Any vertical-edge stimulus falling across this rectangle, regardless of its position, will excite some simple-field cells, leading to excitation of the higher-order cell.

of Text-figs. 5 and 6. One may imagine that it receives afferents from a set of simple cortical cells with fields of type *G*, Text-fig. 2, all with vertical axis orientation, and staggered along a horizontal line. An edge of light would activate one or more of these simple cells wherever it fell within the complex field, and this would tend to excite the higher-order cell.

Similar schemes may be proposed to explain the behaviour of other complex units. One need only use the corresponding simple fields as building blocks, staggering them over an appropriately wide region. A cell with the properties shown in Text-fig. 3 would require two types of horizontally oriented simple fields, having 'off' centres above the horizontal line, and 'on' centres below it. A slit of the same width as these centre regions would strongly activate only those cells whose long narrow

centres it covered. It is true that at the same time a number of other cells would have small parts of their peripheral fields stimulated, but we may perhaps assume that these opposing effects would be relatively weak. For orientations other than horizontal a slit would have little or no effect on the simple cells, and would therefore not activate the complex one. Small spots should give only feeble 'on' responses regardless of where they were shone in the field. Enlarging the spots would not produce summation of the responses unless the enlargement were in a horizontal direction; anything else would result in invasion of opposing parts of the antecedent fields, and cancellation of the responses from the corresponding cells. The model would therefore seem to account for many of the observed properties of complex fields.

Proposals such as those of Text-figs. 19 and 20 are obviously tentative and should not be interpreted literally. It does, at least, seem probable that simple receptive fields represent an early stage in cortical integration, and the complex ones a later stage. Regardless of the details of the process, it is also likely that a complex field is built up from simpler ones with common axis orientations.

At first sight it would seem necessary to imagine a highly intricate tangle of interconnexions in order to link cells with common axis orientations while keeping those with different orientations functionally separated. But if we turn to the results of Part III on functional cytoarchitecture we see at once that gathered together in discrete columns are the very cells we require to be interconnected in our scheme. The cells of each aggregate have common axis orientations and the staggering in the positions of the simple fields is roughly what is required to account for the size of most of the complex fields (cf. Text-fig. 9). That these cells are interconnected is moreover very likely on histological grounds: indeed, the particular richness of radial connexions in the cortex fits well with the columnar shape of the regions.

The otherwise puzzling aggregation of cells with common axis orientation now takes on new meaning. We may tentatively look upon each column as a functional unit of cortex, within which simple fields are elaborated and then in turn synthesized into complex fields. The large variety of simple and complex fields to be found in a single column (Text-fig. 15) suggests that the connexions between cells in a column are highly specific.

We may now begin to appreciate the significance of the great increase in the number of cells in the striate cortex, compared with the lateral geniculate body. In the cortex there is an enormous digestion of information, with each small region of visual field represented over and over again in column after column, first for one receptive-field orientation and then

for another. Each column contains thousands of cells, some cells having simple fields and others complex. In the part of the cortex receiving projections from the area centralis the receptive fields are smaller, and presumably more columns are required for unit area of retina; hence in central retinal regions the cortical projection is disproportionately large.

Complex receptive fields

The method of stimulating the retina with small circular spots of light and recording from single visual cells has been a useful one in studies of the cat's visual system. In the pathway from retina to cortex the excitatory and inhibitory areas mapped out by this means have been sufficient to account for responses to both stationary and moving patterns. Only when one reaches cortical cells with complex fields does the method fail, for these fields cannot generally be separated into excitatory and inhibitory regions. Instead of the direct small-spot method, one must resort to a trial-and-error system, and attempt to describe each cell in terms of the stimuli that most effectively influence firing. Here there is a risk of over- or under-estimating the complexity of the most effective stimuli, with corresponding lack of precision in the functional description of the cell. For this reason it is encouraging to find that the properties of complex fields can be interpreted by the simple supposition that they receive projections from simple-field cells, a supposition made more likely by the anatomical findings of Part III.

Compared with cells in the retina or lateral geniculate body, cortical cells show a marked increase in the number of stimulus parameters that must be specified in order to influence their firing. This apparently reflects a continuing process which has its beginning in the retina. To obtain an optimum response from a retinal ganglion cell it is generally sufficient to specify the position, size and intensity of a circular spot. Enlarging the spot beyond the size of the field centre raises the threshold, but even when diffuse light is used it is possible to evoke a brisk response by using an intense enough stimulus. For geniculate cells the penalty for exceeding optimum spot size is more severe than in the retina, as has been shown by comparing responses of a geniculate cell and an afferent fibre to the same cell (Hubel & Wiesel, 1961). In the retina and lateral geniculate body there is no evidence that any shapes are more effective than circular ones, or that, with moving stimuli, one direction of movement is better than another.

In contrast, in the cortex effective driving of simple-field cells can only be obtained with restricted stimuli whose position, shape and orientation are specific for the cell. Some cells fire best to a moving stimulus, and in these the direction and even the rate of movement are often critical.

Diffuse light is at best a poor stimulus, and for cells in the area of central representation it is usually ineffective at any intensity.

An interesting feature of cortical cells with complex fields may be seen in their departure from the process of progressively increasing specificity. At this stage, for the first time, what we suppose to be higher-order neurones are in a sense less selective in their responses than the cells which feed into them. Cells with simple fields tend to respond only when the stimulus is both oriented and positioned properly. In contrast, the neurones to which they supposedly project are concerned predominantly with stimulus orientation, and are far less critical in their requirements as regards stimulus placement. Their responsiveness to the abstraction which we call orientation is thus generalized over a considerable retinal area.

The significance of this step for perception can only be speculated upon, but it may be of some interest to examine several possibilities. First, neurophysiologists must ultimately try to explain how a form can be recognized regardless of its exact position in the visual field. As a step in form recognition the organism may devise a mechanism by which the inclinations of borders are more important than their exact visual-field location. It is clear that a given form in the visual field will, by virtue of its borders, excite a combination of cells with complex fields. If we displace the form it will activate many of the same cells, as long as the change in position is not enough to remove it completely from their receptive fields. Now we may imagine that these particular cells project to a single cell of still higher order: such a cell will then be very likely to respond to the form (provided the synapses are excitatory) and there will be considerable latitude in the position of the retinal image. Such a mechanism will also permit other transformations of the image, such as a change in size associated with displacement of the form toward or away from the eye. Assuming that there exist cells that are responsive to specific forms, it would clearly be economical to avoid having thousands for each form, one for every possible retinal position, and separate sets for each type of distortion of the image.

Next, the ability of some cells with complex fields to respond in a sustained manner to a stimulus as it moves over a wide expanse of retina suggests that these cells may play an important part in the perception of movement. They adapt rapidly to a stationary form, and continuous movement of the stimulus within the receptive field is the only way of obtaining a sustained discharge (Text-fig. 4*H*). Presumably the afferent simple-field cells also adapt rapidly to a stationary stimulus; because of their staggered fields the moving stimulus excites them in turn, and the higher-order cell is thus at all times bombarded. This seems an elegant means of overcoming a difficulty inherent in the problem of movement

perception, that movement must excite receptors not continuously but in sequence.

Finally, the above remarks apply equally well to displacements of retinal images caused by small eye movements. The normal eye is not stationary, but is subject to several types of fine movements. There is psychophysical evidence that in man these may play an important part in vision, transforming a steady stimulus produced by a stationary object into an intermittent one, so overcoming adaptation in visual cells (Ditchburn & Ginsborg, 1952; Riggs, Ratliff, Cornsweet & Cornsweet, 1953). At an early stage in the visual pathway the effect of such movements would be to excite many cells repeatedly and in turn, rather than just a few continuously. A given line or border would move back and forth over a small retinal region; in the cortex this would sequentially activate many cells with simple fields. Since large rotatory movements are not involved, these fields would have the same axis orientations but would differ only in their exact retinal positions. They would converge on higher-order cells with complex fields, and these would tend to be activated continuously rather than intermittently.

Functional cytoarchitecture

There is an interesting parallel between the functional subdivisions of the cortex described in the present paper, and those found in somatosensory cortex by Mountcastle (1957) in the cat, and by Powell & Mountcastle (1959) in the monkey. Here, as in the visual area, one can subdivide the cortex on the basis of responses to natural stimuli into regions which are roughly columnar in shape, and extend from surface to white matter. This is especially noteworthy since the visual and somatic areas are the only cortical regions so far studied at the single-cell level from the standpoint of functional architecture. In both areas the columnar organization is superimposed upon the well known systems of topographic representation—of the body surface in the one case, and the visual fields in the other. In the somatosensory cortex the columns are determined by the sensory submodality to which the cells of a column respond: in one type of column the cells are affected either by light touch or by bending of hairs, whereas in the other the cells respond to stimulation of deep fascia or manipulation of joints.

Several differences between the two systems will at once be apparent. In the visual cortex the columns are determined by the criterion of receptive-field axis orientation. Presumably there are as many types of column as there are recognizable differences in orientation. At present one can be sure that there are at least ten or twelve, but the number may be very large, since it is possible that no two columns represent precisely the same axis orientation. (A subdivision of cells or of columns into twelve groups

according to angle of orientation shows that there is no clear prevalence of one group over any of the others.) In the somatosensory cortex, on the other hand, there are only two recognized types of column.

A second major difference between the two systems lies in the very nature of the criteria used for the subdivisions. The somatosensory cortex is divided by submodality, a characteristic depending on the incoming sensory fibres, and not on any transformations made by the cortex on the afferent impulses. Indeed we have as yet little information on what integrative processes do take place in the somatosensory cortex. In the visual cortex there is no modality difference between the input to one column and that to the next, but it is in the connexions between afferents and cortical cells, or in the interconnexions between cortical cells, that the differences must exist.

Ultimately, however, the two regions of the cortex may not prove so dissimilar. Further information on the functional role of the somatic cortex may conceivably bring to light a second system of columns, superimposed on the present one. Similarly, in the visual system future work may disclose other subdivisions cutting across those described in this paper, and based on other criteria. For the present it would seem unwise to look upon the columns in the visual cortex as entirely autonomous functional units. While the variation in field size from cell to cell within a column is generally of the sort suggested in Text-figs. 9 and 15, the presence of an occasional cell with a very large complex field (up to about 20°) makes one wonder whether columns with similar receptive-field orientations may not possess some interconnexions.

Binocular interaction

The presence in the striate cortex of cells influenced from both eyes has already been observed by several authors (Hubel & Wiesel, 1959; Cornehls & Grüsser, 1959; Burns, Heron & Grafstein, 1960), and is confirmed in Part II of this paper. Our results suggest that the convergence of influences from the two eyes is extensive, since binocular effects could be demonstrated in 84% of our cells, and since the two eyes were equally, or almost equally, effective in 70% (groups 3–5). This represents a much greater degree of interaction than was suggested by our original work, or by Grüsser and Grüsser-Cornehls (see Jung, 1960), who found that only 30% of their cells were binocularly influenced.

For each of our cells comparison of receptive fields mapped in the two eyes showed that, except for a difference in strength of responses related to eye dominance, the fields were in every way similar. They were similarly organized, had the same axis orientation, and occupied corresponding regions in the two retinas. The responses to stimuli applied to corresponding parts of the two receptive fields showed summation. This should

be important in binocular vision, for it means that when the two images produced by an object fall on corresponding parts of the two retinas, their separate effects on a cortical cell should sum. Failure of the images to fall on corresponding regions, which might happen if an object were closer than the point of fixation or further away, would tend to reduce the summation; it could even lead to mutual antagonism if excitatory parts of one field were stimulated at the same time as inhibitory parts of the other. It should be emphasized that for all simple fields and for many complex ones the two eyes may work either synergistically or in opposition, depending on how the receptive fields are stimulated; when identical stimuli are shone on corresponding parts of the two retinas their effects should always sum.

Although in the cortex the proportion of binocularly influenced cells is high, the mixing of influences from the two eyes is far from complete. Not only are many single cells unequally influenced by the two eyes, but the relative eye dominance differs greatly from one cell to another. This could simply reflect an intermediate stage in the process of mixing of influences from the two eyes; in that case we might expect an increasing uniformity in the eye preference of higher-order cells. But cells with complex fields do not appear to differ, in their distribution among the different eye-dominance groups, from the general population of cortical cells (Text-fig. 12). At present we have no clear notion of the physiological significance of this incomplete mixing of influences from the two eyes. One possible hint lies in the fact that by binocular parallax alone (even with a stimulus too brief to allow changes in the convergence of the eyes) one can tell which of two objects is the closer (Dove, 1841; von Recklinghausen, 1861). This would clearly be impossible if the two retinas were connected to the brain in identical fashion, for then the eyes (or the two pictures of a stereo-pair) could be interchanged without substituting near points for far ones and vice versa.

Comparison of receptive fields in the frog and the cat

Units in many respects similar to striate cortical cells with complex fields have recently been isolated from the intact optic nerve and the optic tectum of the frog (Lettvin, Maturana, McCulloch & Pitts, 1959; Maturana, Lettvin, McCulloch & Pitts, 1960). There is indirect evidence to suggest that the units are the non-myelinated axons or axon terminals of retinal ganglion cells, rather than tectal cells or efferent optic nerve fibres. In common with complex cortical cells, these units respond to objects and shadows in the visual field in ways that could not have been predicted from responses to small spots of light. They thus have 'complex' properties, in the sense that we have used this term. Yet in their detailed behaviour they differ greatly from any cells yet studied in the cat, at any

level from retina to cortex. We have not, for example, seen 'erasible' responses or found 'convex edge detectors'. On the other hand, it seems that some cells in the frog have asymmetrical responses to movement and some have what we have termed a 'receptive-field axis'.

Assuming that the units described in the frog are fibres from retinal ganglion cells, one may ask whether similar fibres exist in the cat, but have been missed because of their small size. We lack exact information on the fibre spectrum of the cat's optic nerve; the composite action potential suggests that non-myelinated fibres are present, though in smaller numbers than in the frog (Bishop, 1933; Bishop & O'Leary, 1940). If their fields are different from the well known concentric type, they must have little part to play in the geniculo-cortical pathway, since geniculate cells all appear to have concentric-type fields (Hubel & Wiesel, 1961). The principal cells of the lateral geniculate body (those that send their axons to the striate cortex) are of fairly uniform size, and it seems unlikely that a large group would have gone undetected. The smallest fibres in the cat's optic nerve probably project to the tectum or the pretectal region; in view of the work in the frog, it will be interesting to examine their receptive fields.

At first glance it may seem astonishing that the complexity of third-order neurones in the frog's visual system should be equalled only by that of sixth-order neurones in the geniculo-cortical pathway of the cat. Yet this is less surprising if one notes the great anatomical differences in the two animals, especially the lack, in the frog, of any cortex or dorsal lateral geniculate body. There is undoubtedly a parallel difference in the use each animal makes of its visual system: the frog's visual apparatus is presumably specialized to recognize a limited number of stereotyped patterns or situations, compared with the high acuity and versatility found in the cat. Probably it is not so unreasonable to find that in the cat the specialization of cells for complex operations is postponed to a higher level, and that when it does occur, it is carried out by a vast number of cells, and in great detail. Perhaps even more surprising, in view of what seem to be profound physiological differences, is the superficial anatomical similarity of retinas in the cat and the frog. It is possible that with Golgi methods a comparison of the connexions between cells in the two animals may help us in understanding the physiology of both structures.

Receptive fields of cells in the primate cortex

We have been anxious to learn whether receptive fields of cells in the monkey's visual cortex have properties similar to those we have described in the cat. A few preliminary experiments on the spider monkey have shown striking similarities. For example, both simple and complex fields have been observed in the striate area. Future work will very likely show

differences, since the striate cortex of the monkey is in several ways different morphologically from that of the cat. But the similarities already seen suggest that the mechanisms we have described may be relevant to many mammals, and in particular to man.

SUMMARY

1. The visual cortex was studied in anaesthetized cats by recording extracellularly from single cells. Light-adapted eyes were stimulated with spots of white light of various shapes, stationary or moving.

2. Receptive fields of cells in the visual cortex varied widely in their organization. They tended to fall into two categories, termed 'simple' and 'complex'.

3. There were several types of simple receptive fields, differing in the spatial distribution of excitatory and inhibitory ('on' and 'off') regions. Summation occurred within either type of region; when the two opposing regions were illuminated together their effects tended to cancel. There was generally little or no response to stimulation of the entire receptive field with diffuse light. The most effective stimulus configurations, dictated by the spatial arrangements of excitatory and inhibitory regions, were long narrow rectangles of light (slits), straight-line borders between areas of different brightness (edges), and dark rectangular bars against a light background. For maximum response the shape, position and orientation of these stimuli were critical. The orientation of the receptive-field axis (i.e. that of the optimum stimulus) varied from cell to cell; it could be vertical, horizontal or oblique. No particular orientation seemed to predominate.

4. Receptive fields were termed complex when the response to light could not be predicted from the arrangements of excitatory and inhibitory regions. Such regions could generally not be demonstrated; when they could the laws of summation and mutual antagonism did not apply. The stimuli that were most effective in activating cells with simple fields—slits, edges, and dark bars—were also the most effective for cells with complex fields. The orientation of a stimulus for optimum response was critical, just as with simple fields. Complex fields, however, differed from simple fields in that a stimulus was effective wherever it was placed in the field, provided that the orientation was appropriate.

5. Receptive fields in or near the area centralis varied in diameter from $\frac{1}{2}$–1° up to about 5–6°. On the average, complex fields were larger than simple ones. In more peripheral parts of the retina the fields tended to be larger. Widths of the long narrow excitatory or inhibitory portions of simple receptive fields were often roughly equal to the diameter of the smallest geniculate receptive-field centres in the area centralis. For cells

with complex fields responding to slits or dark bars the optimum stimulus width was also usually of this order of magnitude.

6. Four fifths of all cells were influenced independently by the two eyes. In a binocularly influenced cell the two receptive fields had the same organization and axis orientation, and were situated in corresponding parts of the two retinas. Summation was seen when corresponding parts of the two retinas were stimulated in identical fashion. The relative influence of the two eyes differed from cell to cell: for some cells the two eyes were about equal; in others one eye, the ipsilateral or contralateral, dominated.

7. Functional architecture was studied by (*a*) comparing the responses of cells recorded in sequence during micro-electrode penetrations through the cortex, (*b*) observing the unresolved background activity, and (*c*) comparing cells recorded simultaneously with a single electrode (multiple recordings). The retinas were found to project upon the cortex in an orderly fashion, as described by previous authors. Most recordings were made from the cortical region receiving projections from the area of central vision. The cortex was found to be divisible into discrete columns; within each column the cells all had the same receptive-field axis orientation. The columns appeared to extend from surface to white matter; cross-sectional diameters at the surface were of the order of 0·5 mm. Within a given column one found various types of simple and complex fields; these were situated in the same general retinal region, and usually overlapped, although they differed slightly in their exact retinal position. The relative influence of the two eyes was not necessarily the same for all cells in a column.

8. It is suggested that columns containing cells with common receptive-field axis orientations are functional units, in which cells with simple fields represent an early stage in organization, possibly receiving their afferents directly from lateral geniculate cells, and cells with complex fields are of higher order, receiving projections from a number of cells with simple fields within the same column. Some possible implications of these findings for form perception are discussed.

We wish to thank Miss Jaye Robinson and Mrs Jane Chen for their technical assistance. We are also indebted to Miss Sally Fox and to Dr S. W. Kuffler for their helpful criticism of this manuscript. The work was supported in part by Research Grants B-2251 and B-2260 from United States Public Health Service, and in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 49 (638)-713. The work was done during the tenure of a U.S. Public Health Service Senior Research Fellowship No. SF 304-R by D.H.H.

REFERENCES

BISHOP, G. H. (1933). Fiber groups in the optic nerve. *Amer. J. Physiol.* **106**, 460–474.

BISHOP, G. H. & O'LEARY, J. S. (1940). Electrical activity of the lateral geniculate of cats following optic nerve stimuli. *J. Neurophysiol.* **3**, 308–322.

BISHOP, P. O., BURKE, W. & DAVIS, R. (1959). Activation of single lateral geniculate cells by stimulation of either optic nerve. *Science*, **130**, 506–507.

BURNS, B. D., HERON, W. & GRAFSTEIN, B. (1960). Response of cerebral cortex to diffuse monocular and binocular stimulation. *Amer. J. Physiol.* **198**, 200–204.

CORNEHLS, U. & GRÜSSER, O.-J. (1959). Ein elektronisch gesteuertes Doppellichtreizgerät. *Pflüg. Arch. ges. Physiol.* **270**, 78–79.

DANIEL, P. M. & WHITTERIDGE, D. (1959). The representation of the visual field on the calcarine cortex in baboons and monkeys. *J. Physiol.* **148**, 33*P*.

DITCHBURN, R. W. & GINSBORG, B. L. (1952). Vision with stabilized retinal image. *Nature, Lond.*, **170**, 36–37.

DOVE, H. W. (1841). Die Combination der Eindrücke beider Ohren und beider Augen zu einem Eindruck. *Mber. preuss. Akad.* **1841**, 251–252.

ERULKAR, S. D. & FILLENZ, M. (1958). Patterns of discharge of single units of the lateral geniculate body of the cat in response to binocular stimulation. *J. Physiol.* **140**, 6–7*P*.

ERULKAR, S. D. & FILLENZ, M. (1960). Single-unit activity in the lateral geniculate body of the cat. *J. Physiol.* **154**, 206–218.

GRÜSSER, O.-J. & SAUER, G. (1960). Monoculare und binoculare Lichtreizung einzelner Neurone im Geniculatum laterale der Katze. *Pflüg. Arch. ges. Physiol.* **271**, 595–612.

HAYHOW, W. R. (1958). The cytoarchitecture of the lateral geniculate body in the cat in relation to the distribution of crossed and uncrossed optic fibers. *J. comp. Neurol.* **110**, 1–64.

HUBEL, D. H. (1957). Tungsten microelectrode for recording from single units. *Science*, **125**, 549–550.

HUBEL, D. H. (1958). Cortical unit responses to visual stimuli in nonanesthetized cats. *Amer. J. Ophthal.* **46**, 110–121.

HUBEL, D. H. (1959). Single unit activity in striate cortex of unrestrained cats. *J. Physiol.* **147**, 226–238.

HUBEL, D. H. (1960). Single unit activity in lateral geniculate body and optic tract of unrestrained cats. *J. Physiol.* **150**, 91–104.

HUBEL, D. H. & WIESEL, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591.

HUBEL, D. H. & WIESEL, T. N. (1960). Receptive fields of optic nerve fibres in the spider monkey. *J. Physiol.* **154**, 572–580.

HUBEL, D. H. & WIESEL, T. N. (1961). Integrative action in the cat's lateral geniculate body. *J. Physiol.* **155**, 385–398.

JUNG, R. (1960). Microphysiologie corticaler Neurone: Ein Beitrag zur Koordination der Hirnrinde und des visuellen Systems. *Structure and Function of the Cerebral Cortex*, ed. TOWER, D. B. and SCHADÉ, J. P. Amsterdam: Elsevier Publishing Company.

KUFFLER, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* **16**, 37–68.

LETTVIN, J. Y., MATURANA, H. R., MCCULLOCH, W. S. & PITTS, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. Inst. Radio Engrs, N.Y.*, **47**, 1940-1951.

MATURANA, H. R., LETTVIN, J. Y., MCCULLOCH, W. S. & PITTS, W. H. (1960). Anatomy and physiology of vision in the frog (*Rana pipiens*). *J. gen. Physiol.* **43**, part 2, 129–176.

MINKOWSKI, M. (1913). Experimentelle Untersuchungen über die Beziehungen der Grosshirnrinde und der Netzhaut zu den primären optischen Zentren, besonders zum Corpus geniculatum externum. *Arb. hirnanat. Inst. Zürich*, **7**, 259–362.

MOUNTCASTLE, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* **20**, 408–434.

O'LEARY, J. L. (1941). Structure of the area striata of the cat. *J. comp. Neurol.* **75**, 131–164.

POLYAK, S. (1957). *The Vertebrate Visual System*, ed. KLÜVER, H. The University of Chicago Press.

POWELL, T. P. S. & MOUNTCASTLE, V. B. (1959). Some aspects of the functional organization of the cortex of the postcentral gyrus of the monkey: a correlation of findings obtained in a single unit analysis with cytoarchitecture. *Johns Hopk. Hosp. Bull.* **105**, 133–162.

RIGGS, L. A., RATLIFF, F., CORNSWEET, J. C. & CORNSWEET, T. N. (1953). The disappearance of steadily fixated visual test objects. *J. opt. Soc. Amer.* **43**, 495–501.

SILVA, P. S. (1956). Some anatomical and physiological aspects of the lateral geniculate body. *J. comp. Neurol.* **106**, 463–486.

TALBOT, S. A. (1942). A lateral localization in the cat's visual cortex. *Fed. Proc.* **1**, 84.

TALBOT, S. A. & MARSHALL, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *Amer. J. Ophthal.* **24**, 1255–1263.

THOMPSON, J. M., WOOLSEY, C. N. & TALBOT, S. A. (1950). Visual areas I and II of cerebral cortex of rabbit. *J. Neurophysiol.* **13**, 277–288.

VON RECKLINGHAUSEN, F. (1861). Zum körperlichen Sehen. *Ann. Phys. Lpz.* **114**, 170–173.

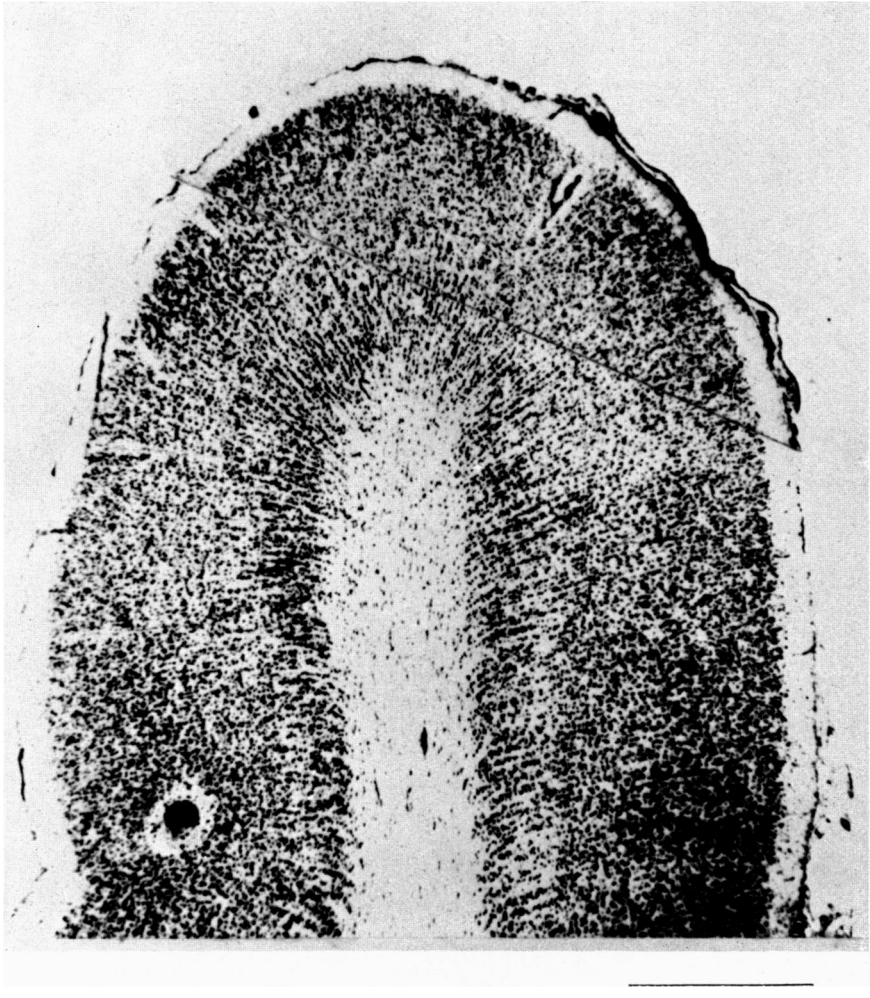
EXPLANATION OF PLATES

PLATE 1

Coronal section through post-lateral gyrus. Composite photomicrograph of two of the sections used to reconstruct the micro-electrode track of Text-fig. 13. The first part of the electrode track may be seen in the upper right; the electrolytic lesion at the end of the track appears in the lower left. Scale 1 mm.

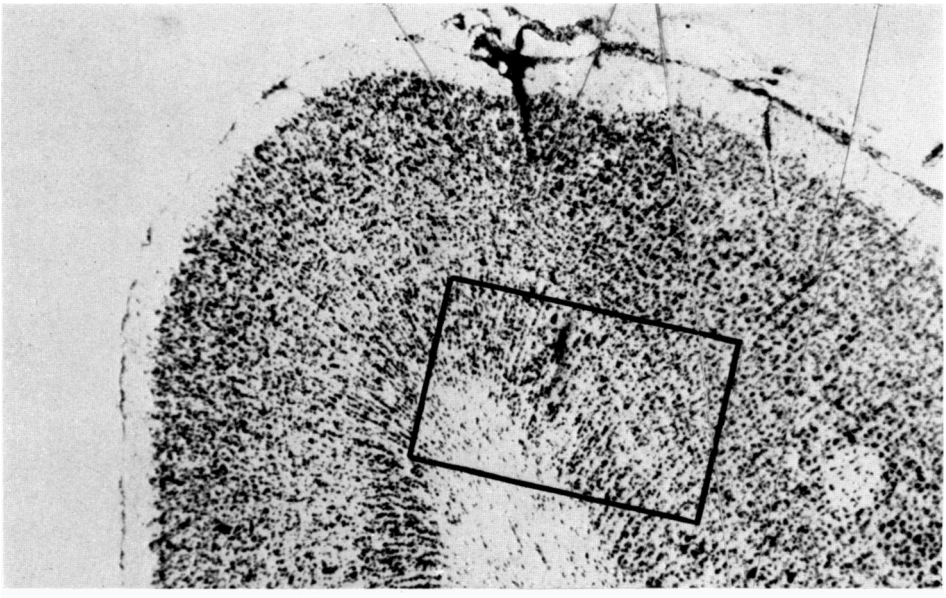
PLATE 2

A, coronal section through the anterior extremity of post-lateral gyrus. Composite photomicrograph made from four of the sections used to reconstruct the two electrode tracks shown in Text-fig. 14. The first part of the two electrode tracks may be seen crossing layer 1. The lesion at the end of the lateral track (to the right in the figure) is easily seen; that of the medial track is smaller, and is shown at higher power in *B*. Scales: *A*, 1 mm, *B*, 0·25 mm.

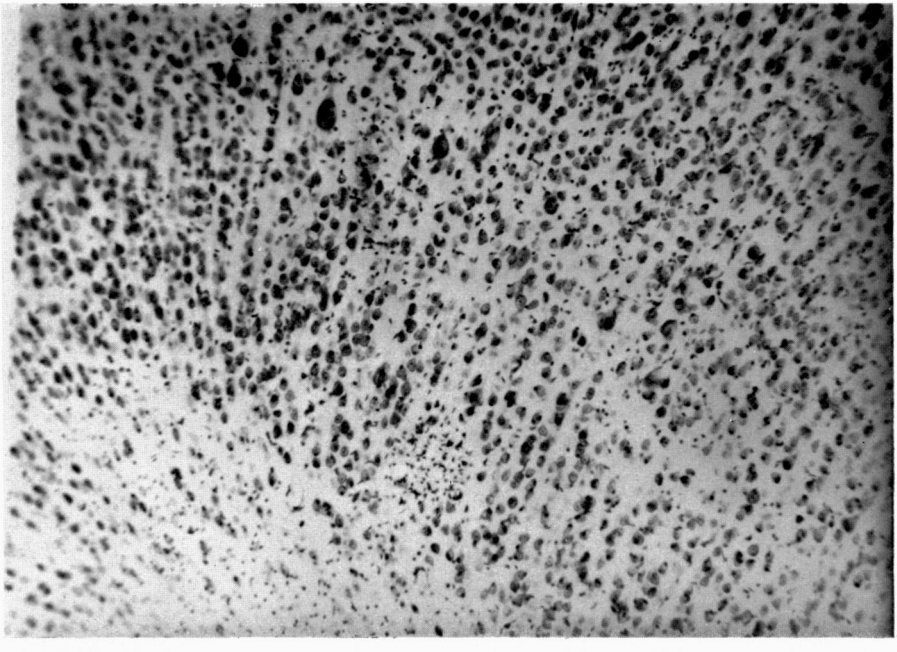


D. H. HUBEL AND T. N. WIESEL

(*Facing p.* 154)



A



B

脸识别

同种动物之间相互识别脸既是很高级的视觉认知，也是社会认知。对此，科学发掘了很多现象，但对根本的机理知之甚少。

初级视皮层 V1 继续投射到更高的区域，分为识别 where（物体空间位置）的背侧通路和识别 what（物体本征）的腹侧通路。

腹侧通路从 V1 到 V2、V4、及更远的区域，其识别的图像越来越复杂（Kobatake and Tanaka，1994）。颞下皮层（inferotemporal cortex，IT）可以识别更复杂的图形，如：圆、方块、多刺圆、手等（Gross, Bender and Rocha-Miranda，1969；Gross, Rocha-Miranda and Bender，1972）。

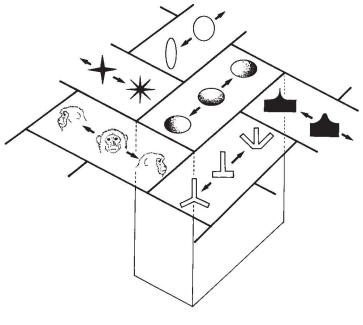
普林斯顿大学心理系科学家在猴的 IT（Gross *et al.*, 1972; Perret *et al.*, 1982; Desimone *et al.*, 1984）和颞上回（superior temporal sulcus，STS）（Bruce *et al.*, 1981）发现了识别脸的细胞，其中 IT 的脸识别细胞几乎都对脸特异反应，而对其他物体没有反应（Desimone *et al.*, 1984）。

人识别脸的能力强于黑猩猩和猴（Rosenfeld and van Hoesen，1979；Parr，2011）。黑猩猩、后、绵羊、鸟类（鸡和鸽）、狗等动物也有脸识别细胞（Kendrick and Baldwin，1987；Ryan and Lea，1994；Kendrick *et al.*,1996；Pascalis and Kelly，2009）。绵羊不仅有识别绵羊脸的细胞，还有识别人脸、狗脸的细胞（Kendrick and Baldwin，1987）。羊羔一到两个月认识母亲的脸（Kendrick *et al.*，1998）。雌绵羊还对雄绵羊的脸有偏好（Kendrick *et al.*, 1995）。低等动物一般依赖嗅觉，但有两种峰（wasps，*Polistes fuscatus* 和 *Polistes metricus*），*P fuscatu* 是群居的、*P metricus* 是独居的，前者有识别个体脸的能力，后者没有（Sheehan and Tibbets, 2011）。

用行为检测显示人对脸的关注在出生很早期就可能出现：9 分钟左右就对脸

的反应大于其他（Goren，Sarty and Wu，1975），5 周就注视脸（Haith, Bergman and Moore, 1977），对脸是否好看也有不同的反应（Slater，1998，2000）。在 4 天区分不带围巾的母亲与其他人的脸、35 天区分带围巾的母亲和其他人的脸（Bushnell，Sai and Mullin，1989；Walton，Bower and Bower，1992；Pascalis *et al.*, 1995；Bruce *et al.*, 2000；Bartrip，Morton and De Schonen，2001）。3 个月识别熟悉的脸（De Haan *et al.*，2001）。用 fMRI 检测观察到，两个月的婴儿的对脸反应脑区被脸激活情况类似成人，但脸还激活婴儿的语言区域（Tzourio-Mazoyer *et al.*,2002）。脸激活与成人一样在 9 岁儿童（Gathers *et al.*,2004）、或 12 岁（Golarai *et al.*，2007）。黑猩猩在 4 周左右识别母亲的脸（Myowa-Yamakoshia，2005）。

1980 年代的一系列电生理实验证明猴对脸有特异反应（Bruce, Desimone and Gross,1981;Perret, Rolls and Caan,1982; Desimone *et al*., 1982; Perrett *et al*., 1984, 1985a, 1985, 1988b; Rolls, Baylis and Leonard, 1985; Saito *et al*., 1986; Perret, Mistlin and Chitty, 1987）。STS 多感觉区域有只对脸反应的细胞（Bruce *et al.*, 1981）。例如，记录 497 个 STS 细胞，48 个只对脸反应，被脸持续激活，28 个细胞在脸有转向、或颜色、大小、距离变化后反应不变（Perret，Rolls and Caan，1982）。早期在 IT 一次记录中，41 个没有反应，110 个有反应的细胞中，66 个有选择性反应，其中 20 对形状反应、2 个对手反应、3 个对脸有选择性反应（Desimone *et al.*, 1984）。可以比较对脸和物体、脸和身体的反应，找到对这三种分别有选择性反应的细胞（Pinsk et al., 2005）。通过 fMRI 辅助确定电生理电极插入位置，可以找到特定区域内 97%的细胞都对脸有选择性反应（Tsao *et al*., 2006; Friewald, Tsao and Livingston, 2007），说明有脸特异区



块（patch）。用脑表面光学成像观察，可以看到对脸呈现面有反应的脑区紧密相连（Wang, Tanaka and Tanifuji, 1996, 1998）。从而提出可能有脸朝向的功能柱（Tanaka，2003）。

脸识别能力有倒置效应，对正立的脸敏感性远远大于倒置的脸（Yin，1969；Thompson，1980）。脸识别细胞对于脸的要求是一个圆加两点一杠（大体相当于脸、眼和嘴）（Kobatake and Tanaka，1994）。对脸有全面的识别和部件的敏感（Freiwald，Tsao and Livingston，2009）。在猴的脸识别细胞研究中，提出抑制性神经元可能对于脸识别很重要，去除 GABA 的抑制性作用后，原对脸（和其他物体）有特异反应的细胞失去反应特异性（Wang，Fujita and Murayama，2000）。用微电流刺激猴的脑区 50 毫秒，可以增加其对脸的反应（Afraz, Kiani and Esteky, 2006）。

人对脸反应的脑区类似于猴（Tsao *et al*., 2003; Tsao, Moellet and Freiwald, 2008；Pinsk *et al*., 2009；Srihasam *et al*., 2012）。在开颅手术的病人经过允许能用颅内记录事件相关电位改变（Allison *et al*., 1999; McCarthy *et al*., 1999, Puce, Allison and McCarthy, 1999），记录到脸特异反应。也直接记录到神经细胞对脸反应（Kreiman, Koch and Fried，2000）。更多的是用正电子扫描（PET）（Sergent，Ohta and MacDonald，1992；Haxby *et al*., 1994）和 fMRI（Malach *et al*., 1995；Puce *et al*., 1996; Clark *et al*., 1996; Kanwisher, McDermott and Chun，1997；McCarthy *et al*., 1997）。可以分别观察几个脑区（FFA、OFA 和 fSTS）对脸的部件和构型的敏感性（Liu, Harris and Kanwisher，2010）。跨颅磁刺激（TMS）是一种研究脑功能的方法（Walsh and Cowey，2000）。用 TMS 作用于特定脑区，可以观察到脸反应的变化（Pitcher *et al*., 2007, 2008, 2009）。FFA 对脸的部件和

构型都敏感，OFA 和 fSTS 只对真的脸部件反应、对其构型不反应（Dzhelyova, Ellison and Atkinson, 2010）。

人类有不能识别脸的个体，诊断为脸盲（prosopagnosia，faceblind）（Bodamer，1947），分为先天型（发育型）和获得型。脸盲者可以识别其他物体，而不能识别脸（Farah, Levinson and Klein, 1995；Farah, 1996; Henke *et al*., 1998; Nunn, Postma and Pearson, 2001; Duchaine and Nakayama, 2005; Duchaine *et al*., 2006; Li and Song, 2007; Riddoch *et al*., 2008）。也有患者可以识别脸但不能识别其他物体（Feinberg *et al*., 1994; Moscovitch, Winocur and Behrmann，1997；McMullen，Fisk and Phillips，2000；Germine *et al*., 2011）。对于脸盲的机理，有多种解释，有些脸盲可能确实是脸识别能力的特异变化（Duchaine，2006）。

后天获得的脸识别异常，可以是病变或外伤（Yin, 1970; Meadow *et al*., 1974; Landis *et al*., 1986; Barton *et al*., 2002；Bouvier and Engel，2006；Schiltz *et al*., 2006；Steeves *et al*., 2006）。右脑单侧外伤就可以导致脸盲。如果可以在脑成像观察到病变部位，有助于了解参与脸识别的脑区（Riddoch *et al*.,2008）。这些可以与在正常人脑进行的核磁共振成像、外科手术人脑电生理记录相辅相成（Kanwisher, McDermott and Chun，1997；Tsao *et al*., 2003；Barraclouth and Perret，2011）。

双生子研究显示脸识别能力有高度遗传性（Polk *et al*., 2007；Wilmer *et al*.，2010；Zhu *et al*., 2010）。德国一个大学的调查显示 2.47%脸盲率（Kennerknecht *et al*., 2006, 2007），中国香港大学的调查显示 1.88%的脸盲率（Kennerknecht, Ho and Wong，2008）。遗传性脸盲在一些家系发现（McConachie，1976；Grueter *et al*., 2007；Duchaine，Germine and Nakayama，2007；Schmalzl，Palermo and Coltheart，2008；Lee *et al*., 2009）。这些表明经典遗传学和基因组学可以用于研究脸识别。

参考文献:

Afraz SR, Kiani R and Esteky H (2006) Microstimulation of inferotemporal cortex influences face categorization. *Nature* 422:692-695.

Allison T, Puce A, Spencer DD and McCarthy G (1999) Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and nonface stimuli. *Cerebral Cortex* 9:415-430.

Barraclough NE and Perrett DI (2011) from single cells to social perception. *Philosophical Transactions of the Royal Society of London B* 366:1739-1752.

Bartrip J, Morton J and De Schonen S (2001) Responses to mother's face in 3-week to 5-month-old infants. *British Journal of Developmental Psychology* 19:219-232.

Bodamer J (1947) Die prosopagnosie. *Archiv für Psychiatrie und Nervenkrankheiten* 179:6-53.

Bruce CJ, Desimone R and Gross CG (1981) Visual properties of neurons in a polysensory area in superior temperal sulcus of the macaque. *Journal of Neurophysiology* 46:369-384.

Bruce V, Campbell RN, Doherty-Sneddon G, Import A, Langton S, McAuley S and Wright R (2000) Testing face processing skills in children. *British Journal of Developmental Psychology* 18:319-333.

Bushnell IW, Sai F and Mullin JT (1989) Neonatal recognition of the mother's face. *British Journal of Developmental Psychology* 7:3-15.

Clark VP, Keil K, Maisog JM, Courtney S, Ungerleider LG and Haxby JV (1996) Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. *NeuroImage* 4:1-15.

de Haan EH (1999) A familial factor in the development of face recognition deficits. *Journal of Clinical and Experimental Neuropsychology* 21: 312-315.

de Haan M, Johnson MH, Maurer D and Perrett DI (2001) Recognition of individual faces and average face prototype by 1- and 3-month-old infants. *Cogitive Development* 16:1-20.

Desimone R, Albright TD, Gross CG and Bruce CJ (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience* 4:2051-2062.

Duchaine B and Nakayama K (2005) Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience* 17:249-261.

Duchaine B, Yovel G, Butterworth E and Nakayama K (2006) Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology* 23:714-747.

Duchaine B, Germine L and Nakayama K (2007) Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology* 24:419-430.

Dzhelyova MP, Ellison A and Atkinson AP (2010) Event-related repetitive TMS reveals distinct, critical roles for right OFA and bilateral posterior STS in judging the trustworthiness and sex of faces. *Journal of Cognitive Neuroscience* 22:203-211.

Farah MJ (1996) Is face recognition "special"? Evidence from neuropsychology. *Behavioral Brain Research* 76:181-189.

Farah MJ, Levinson KL and Klein KL (1995) Face perception and within category discrimination in prosopagnosia. *Neuropsychologia* 33:661-674.

Feinberg TE, Schindler RJ, Ochoa E, Kwan PC and Farah MJ (1994) Associative visual agnosia and alexia without prosopagnosia. *Cortex* 30:395-412.

Freiwald WA, Tsao DY and Livingstone MS (2009) *Nature Neuroscience* 12:1187-1196.

Gathers AD, Bhatt R, Corbly CR, Farley AB and Joseph JE (2004) Developmental shifts in cortical loci for face and object recognition. *NeuroReport* 15:1549-1553.

Germine L, Cashdollar N, Duzel E and Duchaine B (2011) A new selective developmental deficit: Impaired object recognition with normal face recognition. *Cortex* 47:598-607.

Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JDE and Grill-Spector K (2007) Differential development of high-level cortex correlates with category-specific recognition memory. *Nature Neuroscience* 10:512-522.

Goren CC, Sarty M and Wu PYK (1975) Visual following and pattern discrimination of facelike stimuli by newborn infants. *Pediatrics* 56:544–549.

Gross CG, Bender DB and Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166:1303-1306.

Gross CG, Rocha-Miranda CE and Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology* 35:96-111.

Grueter M, Grueter T, Bell V, Halligan PW, Horst J, Sperling K, Halligan PW, Ellis HD and Kennerknecht I (2007) Hereditary prosopagnosia: the first case series. *Cortex* 43:734-749.

Haith MM, Bergman TT and Moore MJ (1977) Eye contact and face scanning in early infancy. *Science* 198:853–855.

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P and Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience* 14:6336-6353.

Henke K, Schweinberger SR, Grigo A, Klos T and Sommer W (1998) Specificity of fact recognition: recognition of exemplars of non-face objects in prosopagnosia. *Cortex* 34:289-296.

Kanwisher NG, McDermott J and Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17:4302-11.

Kendrick KM and Baldwin BA (1987) Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces *Science* 236:448-450.

Kendrick KM, Atkins K, Hinton MR, Broad KD, Fabre-Nys C and Keverne B (1995) Facial and vocal discrimination in sheep. *Animal Behavior* 49:1665–1676.

Kendrick KM, Atkins K, Hinton MR, Heavens P and Keverne B (1996) Are faces special for sheep? Evidence from facial and object discrimination learning tests showing effects of inversion and social familiarity. *Behavioral Processes*

38:19-35.
Kennerknecht I, Grueter T, Welling B, Wentzek S, Horst J, Edwards S and Grueter M (2006) First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics* 140:1617-22.
Kennerknecht I, Plümpe N, Edwards S and Raman R (2007) Hereditary prosopagnosia (HPA): the first report outside the Caucasian population. *Journal of Human Genetics* 52:230-6.
Kennerknecht I, Ho NY and Wong VC (2008) Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics* 146A:2863-70.
Kobatake E and Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology* 71:856–867.
Kreiman G, Koch C and Fried I (2000). Category specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience* 3:946-953.
Lee Y, Duchaine B, Wilson H and Nakayama K (2009) Three cases of developmental prosopagnosia from one family: detailed neuropsychological and psychophysical investigation of face processing. *Cortex* 46:949-64.
Li XB and Song YY (2007) First congenital progopagnosia case reported in China. *Progress in Natural Sciences* 17:165-171.
Liu J, Harris A and Kanwisher N (2010) Perception of face parts and face configuration: an fMRI study. *Journal of Cognitive Neuroscience* 22:203-211.
Malach R, Reppas JB, Benson RB, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR and Tootell RBH (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences USA* 92:8135-8138.
McCarthy G, Puce A, Gore J and Allison T (1997) Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience* 9:605-610.
McCarthy G, Puce A, Belger A and Allison T (1999) Electrophysiological studies of human face perception: II. Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex* 9:431-444.
McConachie HR (1976) Developmental prosopagnosia. A single case report. *Cortex* 12: 76-82.
McMullen P, Fisk JD and Phillips S (2000) Apperceptive agnosia and face recognition. *Neurocase* 6:403-414.
Moscovitch M, Winocur G and Behrmann M (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience* 9:555-604.
Myowa-Yamakoshia M, Yamaguchib MK, Tomonagac M, Tanakac M and Matsuzawac T (2005) Development of face recognition in infant chimpanzees (Pan troglodytes) *Cognitive Development* 20:49-63.
Nunn JA, Postma P and Pearson R (2001) Developmental prosopagnosia: Should it be taken at face value? *Neurocase* 7:15-27.

Parr LA (2011) The evolution of face processing in primates. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 366:1764-1777.

Pascalis O, de Schonen S, Morton J, Deruelle C and Fabre-Grenet M (1995) Mother's face recognition by neonates: A replication and an extension. *Infant Behavioral Development* 18:79-85.

Pascalis O and Kelly DJ (2009) The origins of face processing in humans: phylogeny and ontogeny. *Psychological Sciences* 4:200-209.

Perret DI, Rolls ET and Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47:329-342.

Perrett DI, Smith,PAJ, Potter DD, Mistlin AJ and Head AS (1984) Neurons responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Human Neurobiology* 3:197-208.

Perrett DI, Smith PAJ, Potter DD, Mistlin,AJ, Head AS, Milner AD and Jeeves MA (1985a) Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London Series B* 223:293-317.

Perrett DI, Smith P.AJ, Mistlin AJ, Chitt ,AJ, Head AS, Potter DD, Broennimann R, Milner AD and Jeeves MA (1985b) Visual analysis of body movement by neurones in the temporal cortex of the macaque monkey: a preliminary report. *Brain Research* 16:153-170.

Perrett DI, Mistlin AJ and Chitty AJ (1987) Visual cells responsive to faces *Trends in Neuroscience* 10: 358-364.

Perrett DI, Mistlin AJ, Chitty AJ Smith PAJ, Potter DD Broennimann R and Harries MH (1988) Specialized face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies. *Behavioral Brain Research* 29:245-258.

Pinsk MA, Desimone K, Moore T, Gross CG and Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: A functional MRI study. *Proceedings of the National Academy of Sciences USA* 102:6996-7001.

Pinsk MA, Arcaro M, Weiner KS, Kaljus JF, Inati SJ, Gross CG and Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *Journal of Neurophysiology* 101:2581-2600.

Pitcher D, Walsh V, Yovel G and Duchaine B (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology* 17:1566-1573.

Pitcher D, Garrido L, Walsh V and Duchaine BC (2008) Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience* 28:8929-8933.

Pitcher D, Charles L, Devlin JT, Walsh V and Duchaine B (2009) Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology* 19:319-324.

Polk TA, Park J, Smith MR and Park DC (2007) Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *Journal of Neuroscience* 27:13921-13925.

Puce A, Allison T, Asgari M, Gore JC and McCarthy G (1996) Differential sensitivity

of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of Neuroscience* 16:5205-5215.

Puce A, Allison T and McCarthy G (1999)Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex* 9:445-458.

Riddoch MJ, Johnston RA, Bracewell RM, Boutsen L and Humphreys GW (2008) Are faces special? A case of pure prosopagnosia. *Cognitive Neuropsychology* 25:3-26.

Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiology* 3:209-222.

Rolls ET, Baylis GC and Leonard LM (1985) Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. *Vision Research* 25:1021-1035.

Rosenfeld SA and Van Hoesen GW (1979) Face recognition in the rhesus monkey. *Neuropsychologia* 17:503–509.

Ryan CME and Lea SEG (1994) Images of conspecifics as categories to be discriminated by pigeons and chickens: slides, video tapes, stuffed birds and live birds. *Behavioral Processes* 33:155-176.

Saito H, Yukie M, Tanaka K, Hikosaka K, FukadaY and Iwai E (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience* 6:145-l57.

Schmalzl L, Palermo R and Coltheart M (2008) Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology* 2:99-117.

Sergent J, Ohta S and MacDonald B (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115:15-36.

Sheehan MJ and Tibbets EA (2011) Specialized face learning is associated with individual recognition in paper wasps *Science* 334:1272-1275.

Slater A, von der Schulenburg C, Brown E, Badenoch M, Butterworth G, Parsons S and Samuels C (1998) Newborn infants prefer attractive faces. *Infant Behavioral Development* 21:345-354.

Slater A, Bremner G, Johnson SP, Sherwood P, Hayes R, and Brown E (2000) Newborn infants' preferences for attractive faces: The role of internal and external facial features. *Infancy* 1:265-274.

Srihasam K, Mandeville JB, Morocz IA, Sullivan KJ and Livingstone MS (2012) Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron* 73:608-619.

Tanaka K (2003) Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex* 13:90-99.

Tate AJ, Fisher H, Leigh AE and Kendrick KM (2006) Behavioral and neurophysiological evidence for face identity and face emotion processing in animals. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 361:2155-2172.

Thompson P (1980) Margaret Thatcher: a new illusion. *Perception* 9:483-484.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB and Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nature Neuroscience* 6:989-95.

Tsao DY, Freiwald WA, Tootell RB and Livingstone MS (2006) A cortical region consisting entirely of face selective cells. *Science* 311:670-674.

Tsao DY, Moellet S and Freiwald WA (2008) comparing fact patch systems in macaques and humans. *Proceedings of the National Academy of Sciences USA* 105:19514-19519.

Tzourio-Mazoyer N, de Schonen S, Crivello F, Reutter B, Aujard Y and Mazoyer BM (2002) Neural correlates of woman face processing by 2-month-old infants. *NeuroImage* 15:454-461.

Walsh V and Cowey A (2000) transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience* 1:73-79.

Walton GE, Bower NJA and Bower TGR (1992) Recognition of familiar faces by newborns. *Infant Behavioral Development* 15:265-269.

Wang G, Tanaka K and Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665-1668.

Wang G, Tanifuji M and Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by *in vivo* optical imaging. *Neurosci Res* 32:33-46.

Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, Nakayama K and Duchaine B (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences USA* 107: 5238–5241.

Yin RK (1969) Looking at upside-down faces. *Journal of Experimental Psychology* 81:141-145.

Zhu Q, Song Y, Hu S, Li X, Tian M, Zhen Z, Dong Q, Kanwisher N and Liu J (2010) Heritability of the specific cognitive ability of face perception. *Current Biology* 20:137-142.