计算精神病学：从神经科学到临床应用的桥梁

Computational psychiatry as a bridge from neuroscience to clinical applications

Quentin J M Huys[1,2,5], Tiago V Maia[3,5] & Michael J Frank[4]

[1]*Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zürich and Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland.*

[2]*Centre for Addictive Disorders, Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zürich, Zürich, Switzerland.*

[3]*School of Medicine and Institute for Molecular Medicine, University of Lisbon, Lisbon, Portugal.*

[4]*Computation in Brain and Mind, Brown Institute for Brain Science, Psychiatry and Human Behavior, Brown University, Providence, USA.*

[5]*These authors contributed equally to this work.*

Correspondence should be addressed to Q.J.M.H. (qhuys@cantab.net).

Accepted: 2015 by nature neuroscience

(translated by zang jie)

摘要：将神经科学的进步转化为精神疾病患者的利益面临巨大挑战，因为它涉及最复杂的器官，大脑及其与相似复杂环境的相互作用。处理这种复杂性需要强大的技术。计算精神病学将多种级别和类型的计算与多种类型的数据相结合，以提高对精神疾病的理解，预测和治疗。广义上讲，计算精神病学包含两种互补的方法：数据驱动和理论驱动。数据驱动的方法将机器学习方法应用于高维数据，以改善疾病分类，预测治疗结果或改善治疗选择。这些方法通常与基本机制无关。相比之下，理论驱动的方法使用的模型可以实例化此类机制的先验知识或明确的假设，可能会在多个层次的分析和抽象中进行实例化。我们回顾了这两种方法的最新进展，重点是临床应用，并强调了将它们组合起来的效用。

1、引言

将神经科学的进步转化为精神疾病患者的具体改善的进展缓慢。问题的一部分是精神病学中疾病分类和结果测量的复杂性[1]。但是，更广泛的原因是问题的复杂性：心理健康不仅取决于大脑的功能，最复杂的器官的功能，还取决于该功能如何与个人的身体联系，影响和影响。环境和体验挑战。因此，了解心理健康及其破坏取决于将多个相互作用的水平联系起来，从分子到细胞，电路，认知，行为以及身体和社会环境。

困难之一是这些级别之间的映射不是一对一的。相同的生物障碍会影响一些看似无关的心理功能，相反，不同的生物功能障碍会产生相似的心理乃至神经回路障碍[2-4]。干扰可以在某些级别独立出现，而在其他级别则不会出现功能障碍。例如，情绪低落可能会独立于其特定的生物学原因而影响社会功能。健康和生物学之间的映射也随外部环境而变化[5]。例如，神经生物学确定的情绪调节能力可能在某些环境下就足够了，但在另一些环境下会产生情绪障碍。当前的大数据时代具有获取和处理极高维度的多模式数据集的能力，包括临床，遗传，表观遗传学，认知，神经影像学和其他数据类型[6,7]，有望揭示这些复杂的关系，但是提出了巨大的数据分析挑战。在这里，我们认为，如果没有强大的计算工具和它们提供的概念框架，这些理论和数据分析挑战将是无法克服的。因此，广义的计算精神病学对精神病学的未来至关重要，并且可能在合理开发治疗方法，疾病学和预防策略中发挥核心作用。

计算精神病学包含两种方法[8]：广义上解释为机器学习（ML）的数据驱动的，理论上不可知的数据分析方法（包括但不限于标准统计方法），以及数学上指定变量之间的机械可解释关系的理论驱动模型（通常是包括可观察变量和假定的，理论上有意义的隐藏变量）。我们回顾了这两种方法的进展，重点是临床应用，并讨论了如何将它们结合起来。计算精神病学的其他方面已在其他一般性综述[9-12]和更具体的综述[8,13-16]中进行了回顾。

2、维度的祝福与诅咒

很少有个体迹象足以识别潜在疾病，更不用说症状了。例如，情绪低落不足以诊断重度抑郁症。DSM17 和 ICD18 等分类方案背后的直觉是，存在其他特征（例如快感低下，疲劳，暴饮暴食和自杀念头），可以通过识别一组结果相对较差，需要干预的人群来提高特异性，从而标签疾病。在缺乏对潜在生物学（或环境）病理学的任何了解的情况下，也无法对症状群与生物学之间的可识别关系做

任何保证。希望生物标记物可以提供更多信息，并增强（分层）[1]，甚至（部分）替代[19]症状。

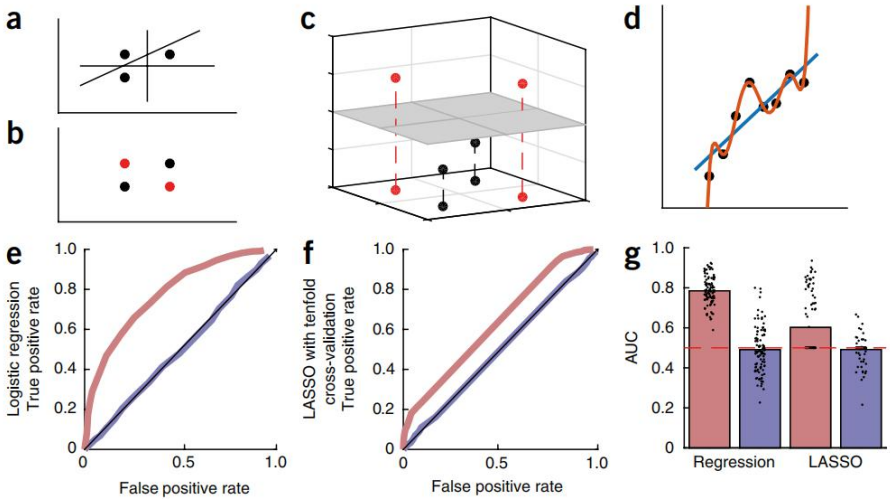

图 1 维度的祝福和诅咒。在精神病学的丰富数据集中，每个受试者的测量变量 d 的数量可能大大超过受试者的数量。（a）发生这种情况时，可以始终线性地分离主题：如果数据跨越 d 维空间，则最多 d+1 个主题可以始终线性地分为两类。始终可以使用两个功能的组合将三个主题分为两组。（b）对于 d+2（或更多）受试者，并非总是可以进行线性分离。（用黑点表示的主题不能与用红点表示的主题线性分离。）（c）但是，如果将这些数据投影到更高维度的空间中，则可以将它们线性分离。在此，通过计算从直线到黑点的绝对距离，将三维添加到 b 中的二维数据，从而使两类线性分离，如灰色二维平面所示。（d）回归可以说明类似的事实：d 阶多项式始终可以完美地拟合 d+1 点（红线），但是它在观测范围之外做出极端预测，并且对噪声极为敏感，因此过度适合训练数据。（e）即使特征和类别仅仅是随机噪声，在高维空间中执行回归也会导致误导性的高性能 142。面板显示了接收器工作特性（ROC）-相对于真实阳性率的假阳性-适用于此类随机数据的逻辑回归。红色曲线表明，逻辑回归在训练数据上的表现令人误解，在 ROC 曲线（AUC）（回归训练数据，g）之下有很高的面积。但是，显然，这是过拟合的，因为数据是随机的。实际上，将所得回归系数应用于训练集中未包含的看不见的验证数据中，则预测是随机的（蓝线；回归验证数据，g）。（f）使用交叉验证的正则回归形式（框 1）LASSO，可以部分防止过度拟合（红线；LASSO 训练数据，g）。但是，由于正则化参数已拟合到训练数据中，所以即使 LASSO 也不能完全防止过度拟合：仅当在验证数据集上测试 LASSO 参数时，才能正确显示性能处于偶然水平（蓝线；LASSO）验证数据，g）。

实际上，通过增加功能来改善分类是 ML 的一个重要概念，它带有祝福和诅咒。“内核技巧”包括隐式添加大量或无限数量的功能[20]。维数的祝福是，在这个无限维空间中，任何有限大小的数据集都可以始终使用简单的线性分类器进行完美分类（图 1a–c）。原始空间中的最终分类可能是复杂且非线性的，特别是如果所包含的（隐式）要素是非线性的或涉及原始数据维度或要素之间的相互作用或相关性时，则尤其如此。实际上，这种祝福也可能是一个诅咒，因为通过使用 n+m-1 个特征（图 1d–g），总是有可能完全区分 n 个患者和 m 个对照。由

于对于任何感兴趣的结果和任何功能都是如此，即使在随机噪声（图 1e–g）和过拟合（框 1）下，它也将表现良好，这意味着结果将不能很好地推广到新数据（例如，未来主题）。过度拟合的危险随着受试者数量（而不是测量数量）的增加而降低，从而激发了更大的研究规模，并且财团将他们的努力集中起来[6、7、21]。

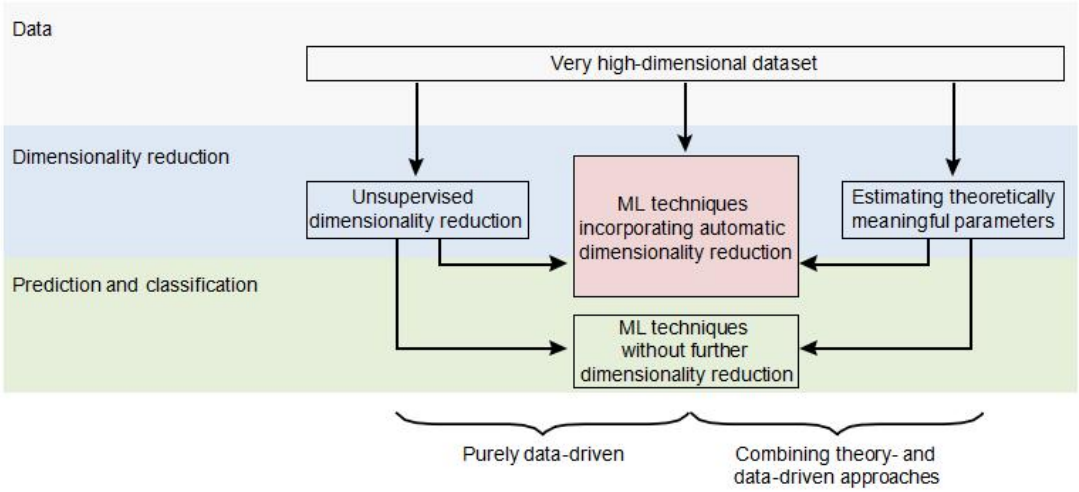

图 2 利用和应对精神病学数据集中的高维度。纯粹的数据驱动方法（左分支和中间分支）以及理论和数据驱动方法的组合（右分支）可用于分析大型数据集，以达到临床有用的应用。降维是避免过度拟合的关键步骤。在应用其他机器学习技术之前，可以使用无监督方法将其作为预处理步骤执行，无论是否进行进一步的降维（左分支；方框 1）；使用自动限制预测变量数量的 ML 技术；使用正则化或贝叶斯模型选择（中间分支；方框 1）；或使用理论驱动模型将本质上的原始高维数据投影到具有理论意义的参数的低维空间中，然后将其输入可能会或不会进一步降低维数（右分支）的 ML 算法中。

存在三种应对维度诅咒的广泛方法。首先，无监督方法可用于在分类或回归之前执行降维（图 2 和方框 1）。其次，可以使用诸如正则化，贝叶斯模型选择和交叉验证之类的技术来选择分类或回归信息最丰富的特征，从而将降维与感兴趣的预测任务结合起来（图 2 和方框 1）。这两种方法都是完全由数据驱动的（尽管贝叶斯方法允许并入先验知识）。第三种完全不同的方法使用理论驱动的模型，基于基础过程的模型来提取理论上有意义的参数。这些参数然后可以用作非常高维数据的有效，低维表示，随后可以将用于分类或回归的 ML 技术应用于这些数据（图 2）。例如，各种时变过程的模型，例如学习[22]，多神经元记录[23]和大胆的时间序列[24]，可以将看似复杂的长时间序列分解为几个表征基本动力学的参数。在理论驱动模型可以准确地描绘或总结生成数据的过程的范围内，它们可以提高 ML 算法的性能，超越了不考虑此类生成机制的方法[13、24、25]。

3、数据驱动的方法

机器学习方法已应用于多种临床相关问题，包括自动诊断，治疗结果和纵向

疾病进程的预测以及治疗选择。我们概述了这些方法的主要方法学特征，并重点介绍了一些说明性示例。最近的其他评论提供了补充信息，超出了本评论的范围[26,27]。

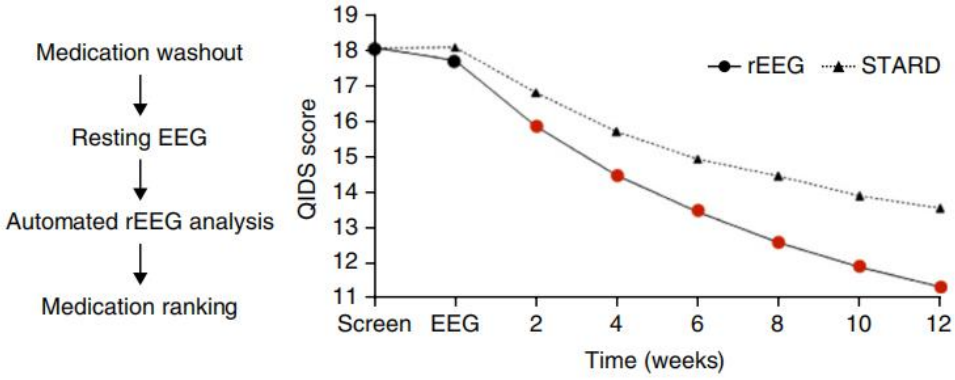

图 3 使用脑电图测量抑郁症的治疗选择可改善治疗反应。左，参考 EEG（rEEG）程序。撤回所有药物后，进行了 rEEG。提交给涉及 74 个生物标志物的在线自动分析，并与与纵向治疗结果相关的脑电测量大型参考数据库进行比较。最后，返回正确药物排名，在一项 12 个站点的试验中，通过优化的临床方案（基于 STAR*D）或 rEEG 将患者随机分配至治疗选择。相对于最佳临床方案，基于 rEEG 的选择 2 周后相对于优化的临床方案改善了治疗反应（红色点），并且这种效果在 12 周内变得更强。这些结果表明生物学措施可以改善抑郁症的治疗选择。改编自 71。

诊断分类。精神病学分类纲要中的大多数症状在两种或两种以上疾病中共享[28]。当前的分类方案试图通过要求多种症状的出现来改善诊断[17,18]。不幸的是，个体仍然常常满足多种疾病的标准（合并症[29]）或不能明确地适用于任何类别[30]，分类阈值不能将疾病负担不同的人群分开[31]，并且对某些疾病的诊断可靠性仍然存在问题[32]。

如今，大量工作已应用 ML 来自动将患者与对照组进行分类[26,27]。最近在竞赛中检查了使用 MRI 数据将精神分裂症与健康对照区分开来的最新技术[33]。最佳条目到达曲线下方的区域（AUC），以对（参考资料[34]），最热门的方法的组合达到了 0.93。尽管使用了不同的技术，但前三个条目的性能相似。但是，通过整合更多的模态[36]或通过算法的进步，例如与深度信念网络[37]或其他方法[38]的进一步改进，还有进一步的改进空间，这些方法在各种 ML 基准[35、39、40]上均优于更标准的方法。对于其他疾病，也已报道了类似的准确度[27]，并且以多种方式扩展了这些结果，例如，概率分类方法可以估算出某种分类的确定性[34,35]，而多分类技术可解决与临床相关性更高的问题。区分诊断组[27,41]。

对神经影像数据进行机器学习分析可以区分病例和对照的事实表明，至少在这些病例中，尽管存在诊断上的警告和障碍中可能的异质性，但症状群确实可以映射到特定的神经生物学底物上。但是，不能总是仅通过检查分类器使用的特征

来识别相关的神经基质：这些特征通常是复杂的，违反直觉的，并且孤立地没有意义，并且通常无法通过不同的机器学习技术进行整理[42]。这些方法在加以克服以使其实用方面也有一些局限性。首先，将病例与健康对照进行比较可能会以错误的方式对待严重程度[34,43]：尽管严重程度会加重合并症并因此模糊诊断界限，但它也经常被用作定量区分情况的指标，从而了解了如何处理严重程度和合并症其次，现有的二元或多类分类方法通常通过假设不同的诊断是互斥的，而错误地对待合并症。解决此限制可能需要统计方案，该方案需要为每个个体提供多个标签（例如，参见参考资料[45]），而这些标签在计算上要高得多。第三，这些算法（通常在明确的情况下进行训练）在多大程度上产生了在临床上更相关的模糊情况的有用信息，还有待探索。最后，这些方法可能会受到根本性的限制，因为它们可以修正基于症状的分类，尽管可以通过细分现有的类来可行地对它们进行细化。

预测治疗反应。当前的方法学局限性导致人们转向预测本质上更有效和直接有用的变量，例如酒精中毒的复发，自杀，高危人群的纵向转变[47-52]和治疗反应。后者解决了精神病学方面的迫切需求。

例如，在抑郁症中，尽管多达四分之三的患者最终会对抗抑郁药产生反应，但三分之二的患者在进行反应之前需要进行多次治疗试验[53]。已经发现几种定量脑电图（qEEG）标记可预测抑郁症的药理反应[54,55]。但是，最近的一项大规模研究，即《国际预测抑郁症最佳治疗方法的研究》（ISPOT-D6），对这些 qEEG 预测因子中的一些预测因子产生了希望[56-58]。qEEG 变量的某些组合（例如 cordance[59]或抗抑郁治疗反应指数[60]）优于单个预测变量。以完全数据驱动的方式组合 qEEG 功能可带来更好的结果：与依赖任何单个预测变量（60％的特异性，86％的敏感性）相比，组合的特征对治疗反应的预测更好（81％特异性，95％敏感性）。尽管样本量太小，无法包含适当的单独验证样本（图 1），但来自其他模态的结果[61]同样强调了将 ML 技术应用于使用多个特征进行预测的有用性。例如，对来自 STARD 和 COMED（两项抑郁症的大型试验）的数据进行的重新分析表明，有监督的降维方法和多元分类器相结合，可得出交叉验证的缓解率预测值（曲线下面积为 AUC；图 1）为 0.66。需要治疗的人数（NNT）为 14，这意味着将算法应用到 14 位患者应获得另一种缓解[62]。同样，尽管在 ISPOT-D 中获得的单变量认知标记不能将缓解者与非缓解者区分开[63,64]，但是任务执行的多变量模式确实可以预测亚组患者对选择性 5-羟色胺再摄取抑制剂（SSRI）依他普仑的反应。多变量结构 MRI 分析还提高了对不太可能做出反应的患者的识别能力，超出了使用单个标记物达到的水平[65]。如这些示例所示，机器学习技术可以改善治疗反应的预测。除了这些特征在模态中的组合之外，跨多个模态的特征的

组合似乎还可能导致进一步的性能改进。

治疗选择。与从业者最相关的问题不一定是给定的治疗方法是否有效，而是几种可能的治疗方法（或多药时代的治疗组合）中哪种方法最适合给定患者。从理论上讲，可以根据多个二进制分类来强制转换多类分类 66。然而，实际上，它提出了其他挑战：对每种治疗方案进行不同的测试（例如神经影像学，遗传学等）可能是不可行的，因此，理想情况下，应使用同一组测试来区分对多种治疗的反应。此外，如果针对不同的治疗方法使用了不同的测试方法，或者甚至针对同一测试方法使用了不同的 ML 算法，则这些预测可能无法直接进行比较，因此不利于在治疗方法之间进行选择。

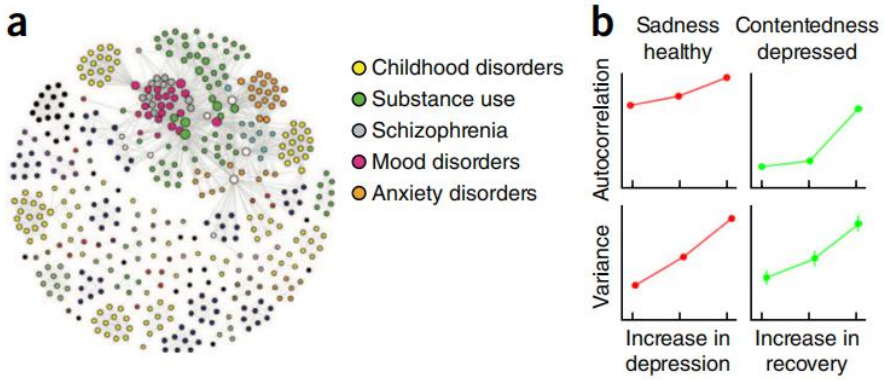

图 4 症状网络。（a）DSM-IV 中的症状网络。如果两个症状属于同一诊断类别，则有链接。有一个大型的，紧密连接的群集，其中包含 48%的症状。总体而言，该网络具有小世界特征，两个症状之间的平均路径长度仅为 2.6。改编自 28.（b）自相关性和方差是严重放缓的两个迹象，在动态网络的相变之前就增加了。在从健康状态转变为抑郁状态之前，负面情绪（例如悲伤）表现出方差和时间自相关的增加。在从抑郁状态转变为缓解状态之前，可以在积极的情绪（如知足）中观察到这一点。改编自 81。

尽管如此，研究已经开始使用通过寻找治疗与治疗之间的相互作用，将受试者随机分配到多个治疗组的试验数据进行多元回归中的相关变量。这表明已婚和受雇并且有更多的生活事件和更多失败的抗抑郁药试验预示了对认知行为疗法（CBT）的反应优于抗抑郁药，而共病的人格障碍患者对抗抑郁药的反应优于 CBT[67]。通过将每位患者分配给理想的治疗方法，可以预期得到的改善是进一步降低了

汉密尔顿抑郁量表上的 3.6 点超出了使用标准治疗方法可获得的减少量，具有临床上的显着效果 [68]。类似于 ISPOT-D 数据的方法得出了认知功能差的患者的依他普仑缓解的预测，其 NNT 为 3.8，这意味着根据他们的认知表现模式将这一组患者分配给依他普仑可导致其缓解。每四个被评估患者增加一个病人 [64]。一项研究 [63] 能够做出足够有力的个人反应预测，以指导大多数患者的治疗选择，从

而使 NNT 为 2–5。

正在迈向将 ML 应用于神经影像数据进行治疗选择的步骤。一组 69 使用单变量标志物杏仁核对下意识面面部情绪刺激的反应，以预测对 SSRI 和 5-羟色胺-去甲肾上腺素再摄取抑制剂（SNRI）的整体反应以及对 SNRI 与 SSRI 的差异反应。另一组研究表明，岛根活动增加与对 CBT 的反应较好有关，但对依他普仑的反应较差。尽管没有检查预测力，但效果大小很大。就像在治疗反应预测的情况下一样，治疗选择方法似乎也将受益于包括来自各种模式的多个变量。

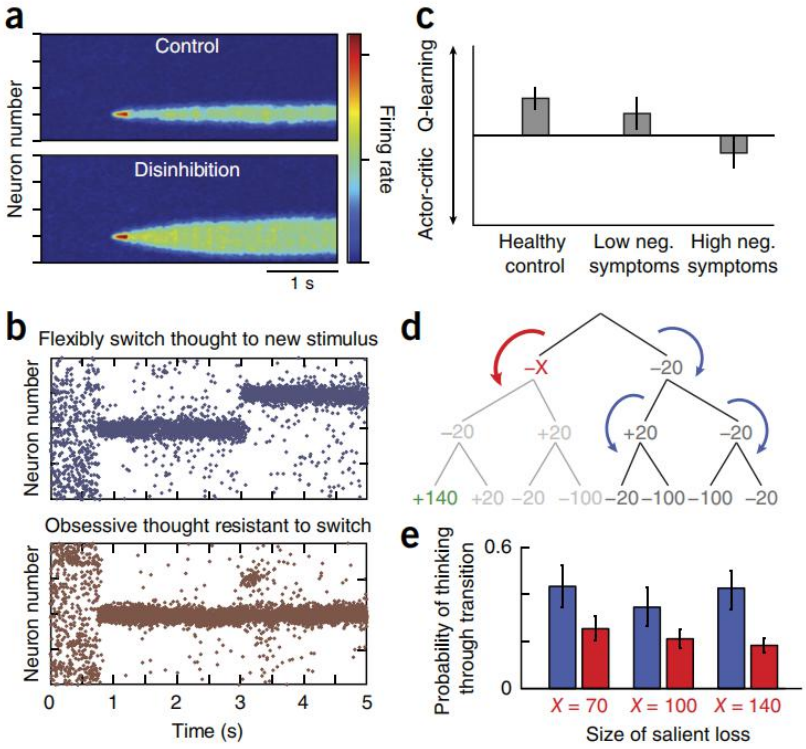

图 5 理论驱动的生物物理和 RL 方法。（a）对精神分裂症工作记忆障碍的见解。减少抑制性中间神经元上的 NMDA 电流会导致整体抑制作用，并扩大工作记忆中刺激的颠簸表示（比较顶部与底部），使其更容易受到干扰物的影响，尤其是那些激活邻近神经元的干扰物。改编自 90.（b）对强迫症的见解。降低 5-羟色胺水平和增加谷氨酸能水平都会使活动模式过度稳定，因此，当刺激新的神经元簇时，活动不会如预期的那样转移到新的位置（顶部，正常反应），而是保持“滞留”状态。'位于上一个位置（底部）。改编自 2.（c）精神分裂症的阴性症状与未能体现期望值有关。在一项仪器学习任务中，健康对照组和精神分裂症患者的不良症状水平较低，这是根据强化学习算法来学习的，该算法明确表示每个状态行为对的期望值（Q 学习），而精神分裂症患者根据一种学习偏好的算法而学习的负面症状的严重程度较高，而没有这种明确的表述（行为批评）。改编自 101.（d）检查指导目标评估的过程。显示的是与三个二进制选择的序列相对应的决策树，其中每个选择都会导致由数字指示的收益或损失。RL 模型适合各种选择，并包含两个关键参数，代表遇到重大突出损失（红色箭头，-X）或遇到其他结果（蓝色箭头）时继续思考的可能性。（e）对于各种显着损失大小，与其他结果相比，受试者在遇到显着损失（红色条）后继续评估分支的可能性要小得多。改编自 132。

据我们所知，迄今为止，只有一项研究试图在一项随机临床试验中验证自动选择治疗算法的临床效用[71]，其结果令人鼓舞。这项研究使用了一种专有算法，该算法是根据来自 1800 多个受试者的 EEG 的参考数据库构建而成的，具有关于多种治疗尝试反应的受试者内部信息（总计约 17,000 次治疗尝试）。该算法从每个患者的 EEG 中提取 74 个特征，以预测最可能成功治疗抑郁症的药物。值得注意的是，自动算法大大胜过了临床选择（图 3）。需要注意的是，两个部门中规定的药物存在很大差异，并且自动选择部门的改进可能不是纯粹通过更好地针对药物而实现的，而是通过使用更多的单胺氧化酶抑制剂和兴奋剂（尽管兴奋剂通常具有抑郁症的治疗效果不佳[72]）。

了解症状之间的关系。上面已经提到了当前诊断方案的局限性，并在其他地方进行了讨论[1,32,73]。替代框架可以洞悉症状的同时发生和顺序表达，而替代框架来自将症状描述为网络，在该网络中，症状被视为实体中的实体，而不是被视为潜在潜变量（给定的疾病）的表达。与其他症状有直接关系的自身权利。例如，睡眠障碍通常会导致疲劳；因此，他们的共同出现可能是他们直接因果互动的结果，而不是潜在的抑郁症[74]。的确，对抑郁症发作之前，之后最长的症状的计算模型（无希望和自尊心差[75]）建议它们可能导致诸如快感不足和缺乏动力等特征[76]。

对 DSM 本身描述的网络分析表明，DSM 诊断中的症状重叠本身可以重现经验观察到的合并症模式的许多关键特征，并揭示了一个具有小世界拓扑结构的优势集群[28]（图 4）：一些症状在其他症状（具有较高的中间性和中心性）之间强烈介导，并且具有从一种症状到另一种症状的短“路径”。有人认为，许多症状之间的强烈连贯性反映了一般的精神病理学因素 p，以与一般智力因素 g 如何捕获多种认知能力之间的协方差 44 相似的方式捕获并发和顺序合并症。

动态网络分析检查了症状在几天内的时间发生模式（例如，使用经验抽样方法进行评估[77]），发现相辅相成的症状频繁循环，可能彼此稳定[78-80]。确实，在非抑郁状态和抑郁状态之间转换之前，反之亦然，症状表现出方差增加和自相关增加[81]。这些是所谓的严重减速的迹象，这些迹象表明动态系统已从稳定状态过渡到另一个稳定状态。实际上，已知残留的亚阈值症状变化是复发的危险因素，并且可能与此处确定的差异有关[82,83]。

从长远来看，完全放弃潜在变量是有问题的，因为症状确实反映了多个潜在变量。使用图形模型可以将网络分析与其他级别的分析（例如，遗传学，神经回路功能等）集成[20]。这些提供了网络描述的概率概括，其中可以包括隐藏变量以及观察到的变量，并且可以在不同级别上合并变量之间的复杂关系，从而有可能

与更多的机械模型形成桥梁。

4、理论驱动的方法

现在，我们转向理论驱动的模型。与数据驱动的方法不同，这些模型封装了对当前现象的理论（通常是机械的）理解。他们的描述在理论上是独立的，但实际上是相互联系的，为集成提供了强大的工具。模型可以通过许多不同的方式进行分类。在这里，我们将区分综合模型，算法模型和最佳模型。

以生物物理上的详细模型为例的合成模型可能是最直观的“模型构建”练习。他们从与特定目标系统（例如，神经系统，特定神经递质的调控等）相关的多个来源获得的数据中获悉，并通过模拟和数学分析探索了这些因素之间的相互作用。这些模型通常会桥接不同级别的分析，并且可以演绎地用于推断已知或怀疑原因的可能后果（例如，给定神经递质浓度的变化会对神经回路动力学或行为产生何种影响）试图推断出可能导致已知结果的原因（例如，某些神经递质浓度的哪种类型的干扰会引起观察到的神经回路或行为障碍）[9]。这些模型可以具有许多不同的参数，这些参数受到广泛的科学文献的约束。通过定性检查其预测来验证它们，这些预测可能包括多个层次的分析（例如，神经活动和行为）。

算法模型（此处以强化学习（RL）模型为例）通常更简单。验证通常通过定量统计手段（例如，模型比较和模型选择技术）进行，该手段评估数据是否保证每个模型所体现的特征和复杂性（例如，参见参考文献 85）。它们包含数量相对较少的参数，可以通过将模型拟合到数据来估计各个受试者的值。这些参数代表了理论上有意义的结构，然后可以在各组之间进行比较，并与症状严重程度等相关[9]。这些模型作为用于测量难以或不可能直接测量的隐藏变量和过程的工具特别有用。

最佳（贝叶斯）模型试图将观察到的行为与问题的贝叶斯最优解联系起来。当最佳值是唯一的时，这尤其显露出来，因为它可以用来显示对象是否可以解决任务以及他们是否已经在特定的实验实例中完成了任务。贝叶斯决策理论大致提供了三种通往精神病理学的途径 86：正确解决错误问题（例如，始终优先考虑饮酒优先于健康），错误解决正确问题（例如，使用酒精“治疗”情绪问题）和解决正确地解决了正确的问题，但是在不幸的环境中或在不幸的先前经历之后（例如，在迫害经历之后有迫切的担忧）。

这些模型类型之间的区别可能很模糊。例如，基底神经节的生物物理现实模型可以具有类似算法的 RL 分量来计算预测误差。此外，有时可以以一致的方式

有益地使用不同的模型类型。例如，通过用更抽象的算法模型逼近详细的神经模型以允许从受试者数据中定量估计参数[87]。这种方法还允许一个人细化一个描述级别的细节，而另一个描述级别则受其限制。例如，详细的基底神经节模型区分了区别地处理多巴胺能增强信号的对手直接途径和间接途径。将此功能整合到更抽象的模型中后，就可以正式分析其对各种参数中各种行为的后果。它也促进了行为数据的定量拟合，并为在传统算法模型[88]之外增加这种对立度的帮助制定了规范性说明[88]。最后，还应注意，贝叶斯技术可以应用于所有三种类型用于拟合，验证和其他目的的模型集合，即非贝叶斯模型也可以使用贝叶斯技术进行拟合。

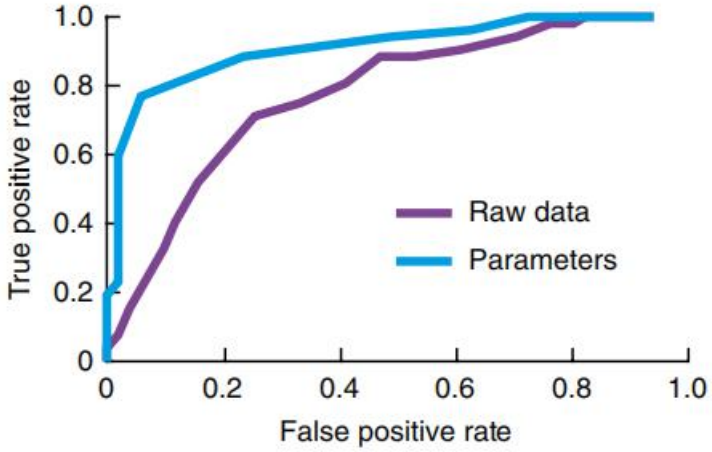

图 6 机械模型产生的参数可用作改善 ML 性能的功能。在分类模型上对模型的估计参数进行训练的分类器要比对原始数据直接进行训练（紫色曲线，AUC0.74）的分类器更好，模拟模型的行为参数拟合了模拟的行为数据（浅蓝色曲线，AUC0.87）。从具有时变作用增强的简单 MFRL 模型模拟了 200 名具有高斯分布参数的受试者的数据。仅根据一个参数（学习率）将受试者分为两组。数据集分为两部分，一半的主题用于训练分类器，另一半用于验证。训练了两个分类器，一个训练了原始的行为数据，另一个训练了通过拟合 RL 模型估计的参数。显示了 ROC 曲线在验证集上的性能。

生物物理现实的神经网络模型。合成的，生物物理的，现实的神经网络模型已被用于将精神疾病的生物异常与其神经动力学和行为后果联系起来。导致精神病学重要见解的一类模型包括反复连接的皮质锥体神经元和 GABA 能神经元。这些模型可以形成稳定的“冲击”活动，以在线维护信息。精神分裂症中发现的抑制性中间神经元上的 NMDA 受体密度的降低会导致更弱和更广泛的吸引子状态（图 5a），这些状态对颠簸附近输入的破坏更敏感，这表明精神分裂症中的工作记忆对类似于干扰物的干扰物特别敏感。工作存储器中保存的项目[90]。该模型跨级别整合的另一种用途是将 NMDA 受体密度与 BOLD 信号相关联。氯胺酮会诱发精神病症状[91]，并消除静止状态默认模式和任务相关模式之间的负面关系[92]。当包含了代表默认模式和任务阳性网络的两个神经元种群的模型时，只有当

GABA 能神经元（而不是锥体神经元）的 NMDA 受体功能降低时，才能捕获这种破坏 [92]。

这类吸引子模型也已用于探讨强迫症（OCD）[2] 中的谷氨酸能和血清素能障碍的影响。血清素减少和谷氨酸增加（这是强迫症中的两个可疑异常）都导致了网络趋向和难以逃脱的强大而持久的活动模式的发展-可能是痴迷的神经动力学底物（图。5b）。值得注意的是，该模型表明，可以通过增加血清素水平来缓解这些神经动力学障碍，而与根本原因是血清素水平低还是谷氨酸水平高无关。该模型还包括特定的 5-羟色胺受体类型：5HT2A 阻滞改善了神经动力学异常，提示为什么用非典型抗精神病药进行治疗可能有益。

皮质-纹状体-丘脑环的生物学详细模型实现了从突触特性到高级功能的相似整合 [93,94]。如前所述，这些模型解释了帕金森氏病，图雷特综合症，精神分裂症和成瘾的各个方面 [9,87]。

简而言之，在存在有关电路的结构和功能的详细知识的地方，综合模型通常可以理解分析水平之间因果关系复杂甚至遥远的关系（例如，从突触改变到行为）。这种模型代表了将生物学细节与症状联系起来的关键工具。还可以简化生物物理上的详细模型，以提取核心非线性动力学分量 [95]，并使其能够使用稳定性或微扰分析进行详细的数学分析。但是，应该指出的是，即使是详细的生物物理模型也通常会进行实质性简化，并且结论仅限于模型中包含的分析级别。例如，被认为可反映 NMDA 受体密度的参数变化捕获的信息可能是系统的其他生物学因素和紧急因素引起的。

生物物理模型也已经成功地应用于神经系统疾病 [95]，例如癫痫病，具有强大的，可识别的神经生理相关性，可以自行建模。精神病学中缺乏已知的强关联，因此很难以自身的方式对其进行建模，而是要求它们在理论上（如此处讨论的示例）或经验上（如数据驱动的方法）与症状相关。

算法强化学习模型。RL 包含一系列算法，可推论优化长期收益的政策 [96]，因此已广泛应用于影响，动机和情感决策方面。实际上，RL 模型通常由两个部分组成：假定捕捉内部学习和评估过程的 RL 算法，以及将内部评估结果与抉择 [3,97] 相关联的链接函数。这样一来，他们就可以为实验中每个参与者的选择分配一个概率，并提供统计上详细的学习和行为描述。尽管它们往往没有生物学上的详细描述，但它们已经表征了神经活动和行为的多个方面 [98]。

最突出的例子是所谓的“无模型”（MF）时间预测误差，该误差与预期获得的增强效果进行了比较。这些误差似乎是由多巴胺能阶段性活动报道的 99。

在这里，我们描述了这些模型在精神病学中的几种用途。

奖励敏感性在许多精神病情况下都会改变。但是，当分析行为时，奖励敏感性的变化通常很难与其他过程的变化区分开，特别是 MF 学习的变化。当将 RL 模型拟合到数据以解开它们时，抑郁症中的快感缺乏症与奖励敏感性的丧失特别相关，其方式不同于影响学习的多巴胺能操纵的方式[22]。类似的方法有助于更精确地测量对无关紧要的刺激的敏感性，从而预测酒精中毒的复发[51]和抑郁的自然过程[100]，从而将消极的症状与学习策略从代表期望值的转变联系起来（图 5c）[101]。。在精神分裂症中，RL 已用于检查异常学习[102]，并表明即使定量控制奖励敏感性和学习策略的差异[103]，腹侧纹状体功能减退仍然持续。

第二个重要方向是检查两种选择价值评估算法，这些算法最初被认为可以并行运行并竞争行为表达[104,105]。资源昂贵的前瞻性“基于模型”（MB）系统在世界内部模型的基础上模拟未来，被认为能够捕获目标导向的行为，并依赖于认知和边缘皮质-纹状体-丘脑-皮质（CSTC）循环。相反，资源小巧的 MF 系统通过根据经验反复预测值错误地更新它们来学习值，并且认为它们可以养成习惯并依赖于感觉运动 CSTC 回路[98、105-108]。由于大多数上瘾的物质会释放多巴胺，它们可能会促进多巴胺能预测错误的学习[110]（但请参见参考文献 111）并加快建立与毒品有关的习惯[112]。确实，更依赖于预测错误学习的动物更容易上瘾[113-115]，同时在人类中也发现了从 MB 转向 MF 选择的平行发现[116-118]。基于强迫症和强迫性药物使用具有某些特征[117,119,120]的观点，有人对强迫症中的 MF 行动转向了类似的论点。但是，补品多巴胺促进了甲基溴而不是 MF 的决定[121、122]，质疑其在将竞争从 MB 估值转移到 MF 估值方面的作用。MB 和 MF 之间的竞争账户的一种替代方案是集成度更高的账户，其中 MF 过程[123]提供了驱动 MB 评估的目标，例如，通过多巴胺能信号强化了前 CSTC 电路中更抽象的计划[124]。这将解释吸毒成瘾的突出目标追求特征[125]。最后，跨疾病从 MB 到 MF 的转变通常是 MB 成分减少的结果，而不是神经元[117,120]和行为[117,116]的更显着的 MF 成分减少（但请参见参考文献 116），这增加了可能性 MB 到 MF 的移动是执行功能的非特定性损伤[126,127]或压力[128]影响 MB 计算的资源的结果。

确实，重新出现的 RL 方向明确地解决了资源约束和有限理性的影响[15,129]。这些可以提供通往 MB 和 MF 系统如何相互作用的规范说明的路径，仅当资源成本被潜在的额外收益所抵消时才使用 MB 系统。此外，由于全面的 MB 评估成本高得令人望而却步，因此它们必须是局部的，对最终的估值产生深远的影响：如果重要的潜在结果未包括在评估中，结果可能会大不相同，并且玻璃杯会从一半充满到半空状态。内部评估策略的调节可能与情绪调节的认知方面有关[15,131]。

RL 建模已经开始确定这些过程的特定方面，例如厌恶性结果在指导资源分配过程中的作用[132, 133]（图 5d，e）。

贝叶斯模型。贝叶斯最佳建模方法可用于更好地理解问题及其解决方案的性质。例如，使用渐进式关联的条件模型无法捕获标准的灭绝现象，该现象是由于灭绝通常涉及新的学习而非学习而造成的。灭绝程序的正确统计描述是，存在一个潜在变量，即实验阶段，会导致刺激与结果之间的关联突然发生变化。使用允许学习此类潜在变量的模型可以更好地说明标准的灭绝现象[134]，并预测只要没有明显的突然切换，实际上就会发生稳定的学习，这已通过实验进行了验证[135]。贝叶斯模型的一个重要方面通常是强调不确定性的表示和使用。这些已被用来表明，厌恶经历的统计在从恐惧调节[136]到习得性无助和沮丧[76]的其他几个过程中也具有重要但有时被忽略的作用。

最佳模型也可用于询问给定症状是否与次优推论有关。例如，特质焦虑高的人无法最佳地更新平均环境的波动性，而低焦虑控制则表现出接近贝叶斯最佳行为的能力[137]。最后，贝叶斯模型也可以用于应用目的。例如，停止信号任务执行的贝叶斯模型[138]区分了长期效果好和坏的偶然刺激者，并为进行纵向预测的 fMRI 分析提供了回归[25]。古典分析都未能实现。

5、结合理论和数据驱动方法

旨在开发临床上有用的应用程序的研究倾向于使用理论上不可知的 ML 方法，而旨在增进对疾病理解的研究则倾向于使用理论驱动的机制方法。理论驱动的方法当然取决于现有知识，机制的理解以及对此类机制的适当评估（例如，通过适当的任务或生理测量）的程度。但是，当存在这些促成因素时，一些初步研究表明，即使从应用的角度来看，理论驱动和机器学习方法的组合使用也可能是有利的。如果机械理论足够准确，则理论驱动的方法可以估算与疾病特别相关的特征。换句话说，理论驱动的方法利用先验知识将数据集“投影”到几个相关参数的空间中，从而大大降低了数据集的维数。然后，机器学习方法可以以更高的效率和可靠性在此低维数据集上工作（图 2）。图 6 显示了这种直观效果的模拟：将分类器应用于生成模型所生成的数据，其效果要比将分类器应用于从该数据中恢复的模型参数的效果差。

概念验证研究说明了这种基于先前工作的方法，该研究表明漂移扩散模型（DDM）的决策阈值（在做出选择之前，支持一个选项而不是另一个选项所需的证据量）部分由沟通控制 139 在额叶皮层和丘脑下核之间。在接受 STN 深层脑刺激（DBS）治疗的帕金森氏病患者中，观察到决策阈值降低导致的冲动行为，

并与额叶皮质和 STN[140]之间的正常沟通中断有关。一项研究使用了应用于脑电图和行为数据的 ML 方法，将患者分为 DBS12 与非 DBS12。使用拟合的 DDM 参数时，分类要比使用原始数据时更好。此外，如先前的机械工作所建议的那样，分类中最有用的参数是决策阈值及其受额叶皮层活动的调节。使用模型参数对症状缓解前的亨廷顿病患者与对照组进行分类，并从表现出症状的患者中分离出相距较近或较远的患者，也发现了类似的改善[141]。使用基于模型的评估还增强了精神分裂症患者的分类和分型[24]，以及上述对兴奋剂滥用的前瞻性预测[25]。

6、结论

我们概述了计算精神病学可能大大改善精神病学的多个方面。数据驱动的方法已开始在临床相关问题上取得成果，例如改善分类，预测治疗反应和辅助治疗选择。但是，这些方法在捕获多个级别之间以及跨多个级别的交互变量的复杂性方面的能力有限。另一方面，由理论驱动的建模工作已经在许多分析级别上得出了有关特定疾病潜在过程的关键见解，但大部分尚未应用于临床问题。我们已经强调了理论驱动和数据驱动方法的组合为何如此强大以及为什么如此强大，并描述了一些初步但有希望的集成尝试。从了解或预测当前疾病类别到转诊方法以及对即将来临的实际有效变量（例如治疗结果）的预测，这两种方法的重点转移似乎是非常有希望的。

计算工具有许多限制。最明显的是，它们需要大量的专业知识，并且通常对非专家是不透明的。因此，该领域的挑战之一是如何促进临床医生，实验者，试验者和理论家之间富有成果的交流。通过在临床试验中积极追求计算方法，将重点放在建立效用上，这可能会有所帮助。此外，计算工具不是万能药，也不会脱离独立复制的要求。但是，开放源代码和数据库的日益普及将促进此类复制以及（临床上）可靠方法的建立和扩展。总体而言，理论家和临床医生之间的互动为患者带来了许多机遇，并最终带来了更好的结果。

6、致谢

Q.J.M.H.瑞士国家科学基金会（320030L_153449/1）和 M.F.由 NSF 授予 1460604 和 NIMHR 01 MH 080066-01。

trait anxiety are unable to optimally update how volatile an aversive environment is, whereas low-anxiety controls exhibit close to Bayes-optimum behavior[137]. Finally, Bayesian models can also be used for applied purposes. For example, a Bayesian model of stop-signal task performance[138] differentiated occasional stimulant users with good and poor long-term outcomes and provided regressors for fMRI analyses that allowed longitudinal prediction[25]; classical analyses failed to achieve either.

Combining theory- and data-driven approaches

Studies aimed at developing clinically useful applications have tended to use theoretically agnostic ML approaches, whereas studies aimed at increasing understanding of disorders have tended to use theory-driven mechanistic approaches. Theory-driven approaches depend, of course, on the extent to which prior knowledge, mechanistic understanding, and appropriate assessments of such mechanisms (for example, via suitable tasks or physiological measurements) are available. When such enabling factors are present, however, some preliminary studies suggest that the combined use of theory-driven and ML approaches can be advantageous even from an applied viewpoint. If the mechanistic theory is sufficiently accurate, theory-driven approaches allow the estimation of features specifically relevant to the disorder. In other words, theory-driven approaches use prior knowledge to massively reduce the dimensionality of the data set by 'projecting' it to the space of a few relevant parameters. ML approaches can then work on this lower-dimensional data set with increased efficiency and reliability (**Fig. 2**). **Figure 6** shows a simulation of this intuitive effect: applying a classifier to data produced by a generative model performs worse than applying it to the model parameters recovered from that data.

A proof-of-concept study illustrating this approach built on prior work showing that the drift-diffusion model's (DDM's) decision threshold—the amount of evidence required in favor of one option over another before committing to a choice—is partly controlled by communication between frontal cortex and the subthalamic nucleus (STN)[139]. Impulsive behaviors that result from reduced decision thresholds are observed in patients with Parkinson's disease treated with STN deep brain stimulation (DBS) and are linked to disruption of normal communication between frontal cortex and STN[140]. One study used ML methods applied to EEG and behavioral data to classify patients into those on versus off DBS[12]. Classification was better when using fitted DDM parameters than when using the raw data; moreover, as suggested by the prior mechanistic work, the most informative parameters for classification were the decision threshold and its modulation by frontal cortical activity. Similar improvements were found using model parameters for classifying presymptomatic Huntington's patients versus controls and separating patients that were closer versus further from exhibiting symptoms[141]. Using model-based assessments has also enhanced classification and subtyping of schizophrenia patients[24] and the aforementioned prospective prediction of stimulant abuse[25].

Conclusion

We have outlined multiple fronts on which computational psychiatry is likely to substantially advance psychiatry. Data-driven approaches have started to bear some fruit for clinically relevant problems, such as improving classification, predicting treatment response and aiding treatment selection. These approaches, however, are limited in their ability to capture the complexities of interacting variables in and across multiple levels. Theory-driven modeling efforts, on the other hand, have yielded key insights at many levels of analysis concerning the processes underlying specific disorders, but for the most part have yet to be applied to clinical problems. We have highlighted why and how the combination of theory- and data-driven approaches can be especially powerful and have described some initial, but promising, attempts at such integration. A shift in focus across both approaches from understanding or predicting current disease categories toward transdiagnostic approaches and the prediction of imminently practical and valid variables, such as treatment outcomes, appears to be very promising.

Computational tools have a number of limitations. Most obviously, they require substantial expertise and are frequently opaque to the non-expert. One challenge for the field is hence how to stimulate fruitful exchange between clinicians, experimentalists, trialists and theorists. This might be helped by a stronger focus on establishing utility by actively pursuing computational approaches in clinical trials. In addition, computational tools are not a panacea and are not released from the requirements of independent replication. However, the increasing popularity of open-source code and databases will facilitate such replications and the establishment and extension of (clinically) robust methods. Overall, the interaction between theorists and clinicians promises many opportunities and ultimately better outcomes for patients.

ACKNOWLEDGMENTS

Q.J.M.H. was supported by a project grant from the Swiss National Science Foundation (320030L_153449/1) and M.F. by NSF grant 1460604 and NIMH R01 MH080066-01.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Kapur, S., Phillips, A.G. & Insel, T.R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
2. Maia, T.V. & Cano-Colino, M. The role of serotonin in orbitofrontal function and obsessive-compulsive disorder. *Clin. Psychol. Sci.* **3**, 460–482 (2015).
3. Huys, Q.J.M., Moutoussis, M. & Williams, J. Are computational models of any use to psychiatry? *Neural Netw.* **24**, 544–551 (2011).
4. Stephan, K.E. *et al.* Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry* **3**, 77–83 (2015).
5. Caspi, A. & Moffitt, T.E. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat. Rev. Neurosci.* **7**, 583–590 (2006).
6. Williams, L.M. *et al.* International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: rationale and protocol. *Trials* **12**, 4 (2011).
7. Mennes, M., Biswal, B.B., Castellanos, F.X. & Milham, M.P. Making data sharing work: the FCP/INDI experience. *Neuroimage* **82**, 683–691 (2013).
8. Maia, T.V. Introduction to the series on computational psychiatry. *Clin. Psychol. Sci.* **3**, 374–377 (2015).
9. Maia, T.V. & Frank, M.J. From reinforcement learning models to psychiatric and neurological disorders. *Nat. Neurosci.* **14**, 154–162 (2011).
10. Montague, P.R., Dolan, R.J., Friston, K.J. & Dayan, P. Computational psychiatry. *Trends Cogn. Sci.* **16**, 72–80 (2012).
11. Wang, X.J. & Krystal, J.H. Computational psychiatry. *Neuron* **84**, 638–654 (2014).
12. Wiecki, T.V., Poland, J. & Frank, M.J. Model-based cognitive neuroscience approaches to computational psychiatry clustering and classification. *Clin. Psychol. Sci.* **3**, 378–399 (2015).
13. Maia, T.V. & McClelland, J.L. A neurocomputational approach to obsessive-compulsive disorder. *Trends Cogn. Sci.* **16**, 14–15 (2012).
14. Stephan, K.E. & Mathys, C. Computational approaches to psychiatry. *Curr. Opin. Neurobiol.* **25**, 85–92 (2014).
15. Huys, Q.J.M., Daw, N.D. & Dayan, P. Depression: a decision-theoretic analysis. *Annu. Rev. Neurosci.* **38**, 1–23 (2015).
16. Stephan, K.E., Iglesias, S., Heinzle, J. & Diaconescu, A.O. Translational perspectives for computational neuroimaging. *Neuron* **87**, 716–732 (2015).
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5 R)* (American Psychiatric Publishing, 2013).
18. World Health Organization. *International Classification of Diseases* (World Health Organization Press, 1990).

19. Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
20. MacKay, D.J. *Information Theory, Inference and Learning Algorithms* (CUP, Cambridge, 2003).
21. Lee, S.H. *et al.*; Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
22. Huys, Q.J.M., Pizzagalli, D.A., Bogdan, R. & Dayan, P. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol. Mood Anxiety Disord.* **3**, 12 (2013).
23. Cunningham, J.P. & Yu, B.M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
24. Brodersen, K.H. *et al.* Dissecting psychiatric spectrum disorders by generative embedding. *Neuroimage Clin.* **4**, 98–111 (2014).
25. Harlé, K.M. *et al.* Bayesian neural adjustment of inhibitory control predicts emergence of problem stimulant use. *Brain* **138**, 3413–3426 (2015).
26. Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012).
27. Wolfers, T., Buitelaar, J.K., Beckmann, C.F., Franke, B. & Marquand, A.F. From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* **57**, 328–349 (2015).
28. Borsboom, D., Cramer, A.O.J., Schmittmann, V.D., Epskamp, S. & Waldorp, L.J. The small world of psychopathology. *PLoS One* **6**, e27407 (2011).
29. Kessler, R.C. *et al.* Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *Br. J. Psychiatry Suppl.* **30**, 17–30 (1996).
30. Fairburn, C.G. & Bohn, K. Eating disorder NOS (EDNOS): an example of the troublesome "not otherwise specified" (NOS) category in DSM-IV. *Behav. Res. Ther.* **43**, 691–701 (2005).
31. Kessler, R.C., Zhao, S., Blazer, D.G. & Swartz, M. Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. *J. Affect. Disord.* **45**, 19–30 (1997).
32. Freedman, R. *et al.* The initial field trials of DSM-5: new blooms and old thorns. *Am. J. Psychiatry* **170**, 1–5 (2013).
33. Silva, R.F. *et al.* The tenth annual MLSP competition: schizophrenia classification challenge. *IEEE Int. Workshop Mach. Learn. Signal Process.* 1–6 (2014).
34. Solin, A. & Sarkka, S. The tenth annual MLSP competition: first place. in *IEEE Int. Workshop Mach. Learn. Signal Process.* 1–6 (2014).
35. Sabuncu, M.R. & Konukoglu, E. Alzheimer's Disease Neuroimaging Initiative. Clinical prediction from structural brain MRI scans: a large-scale empirical study. *Neuroinformatics* **13**, 31–46 (2015).
36. Hahn, T. *et al.* Integrating neurobiological markers of depression. *Arch. Gen. Psychiatry* **68**, 361–368 (2011).
37. Hinton, G.E., Osindero, S. & Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
38. Peng, X., Lin, P., Zhang, T. & Wang, J. Extreme learning machine-based classification of ADHD using brain structural MRI data. *PLoS One* **8**, e79476 (2013).
39. Kim, J., Calhoun, V.D., Shim, E. & Lee, J.H. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage* **124** Pt A, 127–146 2016).
40. Watanabe, T., Kessler, D., Scott, C., Angstadt, M. & Sripada, C. Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine. *Neuroimage* **96**, 183–202 (2014).
41. Costafreda, S.G. *et al.* Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry* **11**, 18 (2011).
42. Pereira, F., Mitchell, T. & Botvinick, M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* **45** (suppl.) S199–S209 (2009).
43. Lubke, G.H. *et al.* Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. *J. Am. Acad. Child Adolesc. Psychiatry* **46**, 1584–1593 (2007).
44. Caspi, A. *et al.* The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
45. Ruiz, F.J.R., Valera, I., Blanco, C. & Perez-Cruz, F. Bayesian nonparametric comorbidity analysis of psychiatric disorders. *J. Mach. Learn. Res.* **15**, 1215–1247 (2014).
46. Hyman, S.E. The diagnosis of mental disorders: the problem of reification. *Annu. Rev. Clin. Psychol.* **6**, 155–179 (2010).
47. Koutsouleris, N. *et al.* Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch. Gen. Psychiatry* **66**, 700–712 (2009).
48. Schmaal, L. *et al.* Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: A multivariate pattern recognition study. *Biol. Psychiatry* **78**, 278–286 (2015).
49. Stringaris, A. *et al.*; IMAGEN Consortium. The brain's response to reward anticipation and depression in adolescence: dimensionality, specificity, and longitudinal predictions in a community-based sample. *Am. J. Psychiatry* **172**, 1215–1223 (2015).
50. Whelan, R. *et al.*; IMAGEN Consortium. Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185–189 (2014).
51. Garbusow, M. *et al.* Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict. Biol.* published online, doi:10.1111/adb.12243 (1 April 2015).
52. Niculescu, A.B. *et al.* Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Mol. Psychiatry* **20**, 1266–1285 (2015).
53. Rush, A.J. *et al.* Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).
54. Olbrich, S. & Arns, M. EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response. *Int. Rev. Psychiatry* **25**, 604–618 (2013).
55. Iosifescu, D.V. Electroencephalography-derived biomarkers of antidepressant response. *Harv. Rev. Psychiatry* **19**, 144–154 (2011).
56. Arns, M. *et al.* Frontal and rostral anterior cingulate (rACC) theta EEG in depression: implications for treatment outcome? *Eur. Neuropsychopharmacol.* **25**, 1190–1200 (2015).
57. Arns, M. *et al.* EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized ISPOT-D study. *Clin. Neurophysiol.* **127**, 509–519 (2015).
58. Dinteren, Rv. *et al.* Utility of event-related potentials in predicting antidepressant treatment response: an iSPOT-D report. *Eur. Neuropsychopharmacol.* **25**, 1981–1990 (2015).
59. Leuchter, A.F. *et al.* Cordance: a new method for assessment of cerebral perfusion and metabolism using quantitative electroencephalography. *Neuroimage* **1**, 208–219 (1994).
60. Iosifescu, D.V. *et al.* Frontal EEG predictors of treatment outcome in major depressive disorder. *Eur. Neuropsychopharmacol.* **19**, 772–777 (2009).
61. Khodayari-Rostamabad, A., Reilly, J.P., Hasey, G.M., de Bruin, H. & Maccrimmon, D.J. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin. Neurophysiol.* **124**, 1975–1985 (2013).
62. Chekroud, A. *et al.* Cross-trial prediction of treatment outcome in depression. *Lancet Psychiatry* published online, doi:10.1016/S2215-0366(15)00471-X (20 January 2016).
63. Gordon, E., Rush, A.J., Palmer, D.M., Braund, T.A. & Rekshan, W. Toward an online cognitive and emotional battery to predict treatment remission in depression. *Neuropsychiatr. Dis. Treat.* **11**, 517–531 (2015).
64. Etkin, A. *et al.* A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. *Neuropsychopharmacology* **40**, 1332–1342 (2015).
65. Korgaonkar, M.S. *et al.* Magnetic resonance imaging measures of brain structure to predict antidepressant treatment outcome in major depressive disorder. *EBioMedicine* **2**, 37–45 (2015).
66. Rifkin, R. & Klautau, A. In defense of one-vs-all classification. *J. Mach. Learn. Res.* **5**, 101–141 (2004).
67. DeRubeis, R.J. *et al.* The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One* **9**, e83875 (2014).
68. Anderson, I. & Pilling, S. *Depression: the Treatment and Management of Depression in Adults (Updated Edition)* (The British Psychological Society and The Royal College of Psychiatrists, 2010).
69. Williams, L.M. *et al.* Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology* **40**, 2398–2408 (2015).
70. McGrath, C.L. *et al.* Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry* **70**, 821–829 (2013).
71. DeBattista, C. *et al.* The use of referenced-EEG (rEEG) in assisting medication selection for the treatment of depression. *J. Psychiatr. Res.* **45**, 64–75 (2011).
72. Candy, M., Jones, L., Williams, R., Tookman, A. & King, M. Psychostimulants for depression. *Cochrane Database Syst. Rev.* (2): CD006722 (2008).
73. Cuthbert, B.N. & Insel, T.R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).
74. Cramer, A.O.J., Waldorp, L.J., van der Maas, H.L.J. & Borsboom, D. Comorbidity: a network perspective. *Behav. Brain Sci.* **33**, 137–150, discussion 150–193 (2010).
75. Iacoviello, B.M., Alloy, L.B., Abramson, L.Y. & Choi, J.Y. The early course of depression: a longitudinal investigation of prodromal symptoms and their relation to the symptomatic course of depressive episodes. *J. Abnorm. Psychol.* **119**, 459–467 (2010).
76. Huys, Q.J.M. & Dayan, P. A Bayesian formulation of behavioral control. *Cognition* **113**, 314–328 (2009).
77. Telford, C., McCarthy-Jones, S., Corcoran, R. & Rowse, G. Experience sampling methodology studies of depression: the state of the art. *Psychol. Med.* **42**, 1119–1129 (2012).
78. Bringmann, L.F. *et al.* A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One* **8**, e60188 (2013).
79. Bringmann, L.F., Lemmens, L.H.J.M., Huibers, M.J.H., Borsboom, D. & Tuerlinckx, F. Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychol. Med.* **45**, 747–757 (2015).
80. Wigman, J.T.W. *et al.*; MERGE. Exploring the underlying structure of mental disorders: cross-diagnostic differences and similarities from a network perspective using both a top-down and a bottom-up approach. *Psychol. Med.* **45**, 2375–2387 (2015).

81. van de Leemput, I.A. *et al.* Critical slowing down as early warning for the onset and termination of depression. *Proc. Natl. Acad. Sci. USA* **111**, 87–92 (2014).
82. Segal, Z.V. *et al.* Antidepressant monotherapy vs sequential pharmacotherapy and mindfulness-based cognitive therapy, or placebo, for relapse prophylaxis in recurrent depression. *Arch. Gen. Psychiatry* **67**, 1256–1264 (2010).
83. Dunlop, B.W., Holland, P., Bao, W., Ninan, P.T. & Keller, M.B. Recovery and subsequent recurrence in patients with recurrent major depressive disorder. *J. Psychiatr. Res.* **46**, 708–715 (2012).
84. Marr, D. *Vision* (Freeman, New York, 1982).
85. Guitart-Masip, M. *et al.* Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage* **62**, 154–166 (2012).
86. Huys, Q.J.M., Guitart-Masip, M., Dolan, R.J. & Dayan, P. Decision-theoretic psychiatry. *Clin. Psychol. Sci.* **3**, 400–421 (2015).
87. Frank, M.J. Linking across levels of computation in model-based cognitive neuroscience. in *An Introduction to Model-Based Cognitive Neuroscience* (eds. B. Forstmann & E. Wagenmakers) 163–181 (Springer, New York, 2015).
88. Collins, A.G.E. & Frank, M.J. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* **121**, 337–366 (2014).
89. Lisman, J.E. *et al.* Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. *Trends Neurosci.* **31**, 234–242 (2008).
90. Murray, J.D. *et al.* Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb. Cortex* **24**, 859–872 (2014).
91. Krystal, J.H. *et al.* Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch. Gen. Psychiatry* **51**, 199–214 (1994).
92. Anticevic, A. *et al.* NMDA receptor function in large-scale anticorrelated neural systems with implications for cognition and schizophrenia. *Proc. Natl. Acad. Sci. USA* **109**, 16720–16725 (2012).
93. Frank, M.J. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J. Cogn. Neurosci.* **17**, 51–72 (2005).
94. Gurney, K.N., Humphries, M.D. & Redgrave, P. A new framework for cortico-striatal plasticity: behavioural theory meets *in vitro* data at the reinforcement-action interface. *PLoS Biol.* **13**, e1002034 (2015).
95. Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M. & Friston, K. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* **4**, e1000092 (2008).
96. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: an Introduction* (MIT Press, 1998).
97. Daw, N. Trial-by-trial data analysis using computational models. in *Decision Making, Affect, and Learning: Attention and Performance XXIII* (eds. M.R. Delgado, E.A. Phelps & T.W. Robbins) 1–23 (OUP, 2009).
98. Maia, T.V. Reinforcement learning, conditioning and the brain: successes and challenges. *Cogn. Affect. Behav. Neurosci.* **9**, 343–364 (2009).
99. Eshel, N. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).
100. Huys, Q.J.M. *et al.* The specificity of pavlovian regulation is associated with recovery from depression. *Psychol. Med.* (in the press).
101. Gold, J.M. *et al.* Negative symptoms and the failure to represent the expected reward value of actions: behavioral and computational modeling evidence. *Arch. Gen. Psychiatry* **69**, 129–138 (2012).
102. Roiser, J.P. *et al.* Do patients with schizophrenia exhibit aberrant salience? *Psychol. Med.* **39**, 199–209 (2009).
103. Schlagenhauf, F. *et al.* Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* **89**, 171–180 (2014).
104. Killcross, S. & Coutureau, E. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* **13**, 400–408 (2003).
105. Daw, N.D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
106. Dolan, R.J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
107. Friedel, E. *et al.* Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* **8**, 587 (2014).
108. Horga, G. *et al.* Changes in corticostriatal connectivity during reinforcement learning in humans. *Hum. Brain Mapp.* **36**, 793–803 (2015).
109. Nutt, D.J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P.R. The dopamine theory of addiction: 40 years of highs and lows. *Nat. Rev. Neurosci.* **16**, 305–312 (2015).
110. Redish, A.D. Addiction as a computational process gone awry. *Science* **306**, 1944–1947 (2004).
111. Panlilio, L.V., Thorndike, E.B. & Schindler, C.W. Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. *Pharmacol. Biochem. Behav.* **86**, 774–777 (2007).
112. Nelson, A. & Killcross, S. Amphetamine exposure enhances habit formation. *J. Neurosci.* **26**, 3805–3812 (2006).
113. Flagel, S.B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469**, 53–57 (2011).
114. Lesaint, F., Sigaud, O., Flagel, S.B., Robinson, T.E. & Khamassi, M. Modelling individual differences in the form of Pavlovian conditioned approach responses: a dual learning systems approach with factored representations. *PLoS Comput. Biol.* **10**, e1003466 (2014).
115. Huys, Q.J.M., Tobler, P.N., Hasler, G. & Flagel, S.B. The role of learning-related dopamine signals in addiction vulnerability. *Prog. Brain Res.* **211**, 31–77 (2014).
116. Sjoerds, Z. *et al.* Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* **3**, e337 (2013).
117. Voon, V. *et al.* Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* **20**, 345–352 (2014).
118. Sebold, M. *et al.* Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* **70**, 122–131 (2014).
119. Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S. & Ersche, K.D. Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. *Trends Cogn. Sci.* **16**, 81–91 (2012).
120. Gillan, C.M. *et al.* Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am. J. Psychiatry* **172**, 284–293 (2015).
121. Wunderlich, K., Smittenaar, P. & Dolan, R.J. Dopamine enhances model-based over model-free choice behavior. *Neuron* **75**, 418–424 (2012).
122. Deserno, L. *et al.* Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. USA* **112**, 1595–1600 (2015).
123. Cushman, F. & Morris, A. Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci. USA* **112**, 13817–13822 (2015).
124. Collins, A.G.E. & Frank, M.J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
125. Everitt, B.J. & Robbins, T.W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* **8**, 1481–1489 (2005).
126. Otto, A.R., Gershman, S.J., Markman, A.B. & Daw, N.D. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* **24**, 751–761 (2013).
127. Schad, D.J. *et al.* Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Front. Psychol.* **5**, 1450 (2014).
128. Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A. & Daw, N.D. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. USA* **110**, 20941–20946 (2013).
129. Boureau, Y.L., Sokol-Hessner, P. & Daw, N.D. Deciding how to decide: self-control and meta-decision making. *Trends Cogn. Sci.* **19**, 700–710 (2015).
130. Keramati, M., Dezfouli, A. & Piray, P. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput. Biol.* **7**, e1002055 (2011).
131. Etkin, A., Büchel, C. & Gross, J.J. The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–700 (2015).
132. Huys, Q.J.M. *et al.* Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* **8**, e1002410 (2012).
133. Huys, Q.J.M. *et al.* Interplay of approximate planning strategies. *Proc. Natl. Acad. Sci. USA* **112**, 3098–3103 (2015).
134. Gershman, S.J., Blei, D.M. & Niv, Y. Context, learning and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
135. Gershman, S.J., Jones, C.E., Norman, K.A., Monfils, M.H. & Niv, Y. Gradual extinction prevents the return of fear: implications for the discovery of state. *Front. Behav. Neurosci.* **7**, 164 (2013).
136. Maia, T.V. Fear conditioning and social groups: statistics, not genetics. *Cogn. Sci.* **33**, 1232–1251 (2009).
137. Browning, M., Behrens, T.E., Jocham, G., O'Reilly, J.X. & Bishop, S.J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat. Neurosci.* **18**, 590–596 (2015).
138. Shenoy, P. & Yu, A.J. Rational decision-making in inhibitory control. *Front. Hum. Neurosci.* **5**, 48 (2011).
139. Frank, M.J. *et al.* fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *J. Neurosci.* **35**, 485–494 (2015).
140. Cavanagh, J.F. *et al.* Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat. Neurosci.* **14**, 1462–1467 (2011).
141. Wiecki, T.V., Antoniades, C.A., Stevenson, A., Kennard, C. & Borowsky, B. A computational cognitive biomarker for early-stage Huntington's disease. *PLoS One* (in the press).
142. Whelan, R. & Garavan, H. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol. Psychiatry* **75**, 746–748 (2014).
143. Lemm, S., Blankertz, B., Dickhaus, T. & Müller, K.R. Introduction to machine learning for brain imaging. *Neuroimage* **56**, 387–399 (2011).
144. Mwangi, B., Tian, T.S. & Soares, J.C. A review of feature reduction techniques in neuroimaging. *Neuroinformatics* **12**, 229–244 (2014).
145. Wig, G.S. *et al.* Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cereb. Cortex* **24**, 2036–2054 (2014).
146. Maroco, J. *et al.* Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res. Notes* **4**, 299 (2011).
147. MacKay, D.J. Bayesian interpolation. *Neural Comput.* **4**, 415–447 (1992).
148. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Neuroimage* **14**, 1137–1145 (1995).