

精神病学的计算方法

Computational approaches to psychiatry

Klaas Enno Stephan^{1,2,3} and Christoph Mathys³

¹*Translational Neuromodeling Unit (TNU), Institute of Biomedical Engineering, University of Zurich & Swiss Federal Institute of Technology (ETH Zurich), Switzerland*

²*Laboratory for Social and Neural Systems Research (SNS), University of Zurich, Switzerland*

³*Wellcome Trust Centre for Neuroimaging, University College London, UK*

Accepted: 29th December 2013 by Current Opinion in Neurobiology

(translated by zang jie)

摘要：精神病学诊断和疾病学进展令人失望的主要原因是缺乏能够对个体患者内疾病过程进行机械推断的测试。结果导致无法进行正式的鉴别诊断，这迫使该领域坚持基于症状的诊断方案，而这些方案对治疗反应和临床结果的预测有效性有限。一种有希望的新方法是将计算模型用于推断机制，该机制可在精神病患者中产生观察到的行为和大脑活动。但是，尽管这种用于精神病学的计算方法迅速受到关注，但仍有许多工作要做以完善现有的计算模型，使它们在临床环境中“适合实践”并在纵向研究中证明其有效性。这篇综述概述了这方面的最新方法学进展和策略，重点是生成模型，这些模型从测量的行为和大脑活动中推断出机械上可解释的参数（计算或生理过程的参数）。

1、为什么计算方法对精神病学很重要？

目前的精神病学诊断工具包不包括诊断测试（除了用于排除“有机”原因的脑部疾病的诊断测试之外），这些诊断测试揭示出特定行为症状的确切机制并预测临床结果或指导个体治疗[1]。这是精神病学无法超越描述性分类（例如《精神障碍诊断和统计手册》，DSM）的主要原因，该分类学将精神疾病从现象学上定义为症状的簇，但预测效度有限[2]。

缺乏机械测试的一些原因很容易指出。遗传学和神经成像是生物学精神病学的关键方法，面临着相当大的障碍：遗传学在强大的基因和环境相互作用中挣扎，这可能是迄今为止仅基于基因组数据的临床相关预测未能成功的重要原因[3]。相比之下，尽管神经影像学可以提供假定的症状产生回路的功能状态的读数，但其测量结果是间接的，远离我们感兴趣的神经元过程，从而加剧了机械学假说的表述。克服这种僵局的一项重要策略是使用“计算”模型[4,5,6,7,8]。在本文中，我们考虑了“计算”这个广义术语的两个可能含义：第一，信息处理的建模机制；第二，从大脑活动的测量值推断生理过程。

从跨区域研究计划（例如，伦敦大学学院和马克斯·普朗克学会关于“计算精神病学和衰老研究”的联合倡议[9]）到“计算精神病学”[10]，以及专门致力于翻译神经建模的新成立的机构[11]都证明了精神病学的计算方法正在迅速受到关注。目前存在许多令人鼓舞的概念验证示例，例如如何将计算模型应用于患者[12-15]。但是，到目前为止，尚未使用前瞻性研究设计来评估这些计算方法，这对于评估临床效用至关重要。

因此，本文对释放精神病学计算方法的翻译潜力的方法和策略的最新进展进行了综述。我们专注于所谓的“生成模型”，该模型指定了所有变量（观测值和参数）的联合概率分布，并用于根据测量的行为或大脑活动来推断认知和生理机制[4]（请参见参考1）。相比之下，有限的空间使我们无法讨论受神经经济学，博弈论，图论或机器学习应用于精神病学神经影像学启发的丰富的建模文献；有关这些主题的全面评论，请参阅[16-18]。

参考 1 生成模型：

生成模型定义了基于观测值（测量数据 y ）和参数 u 的联合概率分布 $p(y, \theta)$ 。它具有两个分量，似然函数 $p(y|\theta)$ 和参数 $p(u)$ 的先验密度。之所以称为“生成模型”，是因为它可以从先验采样参数值并将其嵌入似然估计中来合成数据。因此，人们还可以将生成模型视为从参数到观测数据的“前进模型”。“模型逆转”指的是相反的过程：给定一些观察到的数据，估计参数的后验概率。值得注意的是，通过整合数据对模型的依赖性参数，则获得“期望数据”，即边际可能性或模型信息：

$$p(y) = \int p(y|\theta)p(\theta)d\theta \quad (1)$$

模型信息是模型可概括性的原则性度量（即在准确性和复杂性之间进行权衡），并广泛用于模型比较；参见[69, 71]。

2、建模计算

现有的大多数精神疾病的计算机治疗方法都将学习不良和决策制定作为适应不良认识的核心组成部分。尽管存在许多类型的此类模型，但发现有两种特别广泛地应用于经验数据：强化学习（RL）模型和贝叶斯推断模型。尽管这两个框架起源于不同的理论根基，但它们在概念上有着共同的联系。最重要的是，正如最近从变分逼近到分层贝叶斯学习[19]的 RL 方程的衍生所强调的那样，两个框架在学习背后都具有结构相似的驱动力：预测误差（PE），由学习率（RL）加权或精度/不确定性（贝叶斯理论）。在这篇综述中，我们特别强调了贝叶斯方法，因为最近对 RL 的发展已有一些出色的综述[20-24]。

与精神病学特别相关的一个研究问题涉及“无模型”系统和“基于模型”的系统之间的差异，这两种系统应该分别介导习惯性学习和目标导向性学习[25]。简而言之，在前一种情况下，PE 代表实际结果与预期结果之间的差异（例如奖励 PE）；在后一种情况下，该模型体现了对环境的明确了解，并通过“状态 PE”（隐含状态和预期状态之间的差异）更新了其表示形式。

近年来，RL 方法引起了人们的极大兴趣。这是由于有关不同学习系统之间潜在竞争的想法而引起的，例如，对目标导向学习产生适得其反的巴甫洛夫式影响[26]，或强迫症中习惯性学习和目标导向学习之间的平衡受到干扰[27]。最初的功能磁共振成像研究[28]发现，健康参与者的学习行为反映了奖励和状态 PE，其中前者与腹侧纹状体的活动相关，这与以前的许多研究一致，而状态 PE 是由活动编码的。

在顶叶和额叶前区。这与随后的奖励性 PEs 的腹侧纹状体激活的 fMRI 结果[29]大致相符，而前额叶区域活动反映了状态性 PEs。然而，另一项旨在最大程度区分无模型学习和基于模型的学习的，分两步任务的研究表明，腹侧纹状体中的功能磁共振成像活动并非纯粹反映无模型学习，而是两种学习形式的混合体，其比例与最佳解释行为的比例相同[30]。这组作者说，“这些结果挑战了一个独立的无模型学习者的观念，并提出了一种更集成的计算体系结构，以进行高级的人类决策。”

从 RL 转向贝叶斯方法，“贝叶斯脑假说” [31,32]将大脑视为构建并不断更新其感觉输入的生成模型的方法（参见专栏 1），启发了最近的建模框架。应用于精神病学的潜力很大。例如，“自由能原理” [33, 34]认为，对大脑生成模型的持续优化取决于自由能的最小化，这是有原则且易于处理的令人惊讶的近似值（见专栏 2）。进行正式定义）。简而言之，这对应于净预测误差的最小化（可能跨越多个推断层次），并且可以通过调整人们对世界的看法（感知）或更改通过感官对世界进行采样的方式来实现（行动）。

这种观点导致了对（不良）适应性认知的一系列最新理论治疗，尤其是关于精神分裂症的治疗[4, 35,36, 37]。此外，它启发了分析经验数据的具体策略。一种这样的实际应用框架是一种元贝叶斯方法，该方法考虑了贝叶斯推断（由实验者或精神病学家）基于贝叶斯推理过程（在受试者或患者的大脑中）的基础，该过程是观察到的行为反应的基础[38,39]。在这个框架中，一个模型可以模拟对象的“隐藏”（内部）信念更新过程如何引起他/她的明显反应，然后实验者可以观察到。这种分层方法的吸引力在于，可以通过反转单个生成模型并在关于如何在脑中实现贝叶斯推理的相同假设下，估计实验者的信念（关于受试者的信念驱动观察到的行为）。，通过最小化自由能）。

这种对贝叶斯方法的一种特殊实现是分层高斯滤波器（HGF; [19]），它从变分近似到理想的分层贝叶斯学习，得出类似于 RL 的更新方程，并且包含代表个人对贝叶斯近似的参数。最优性。最近的几项研究已使用该框架来确定学习和决策的竞争假设，方法是使用与病理生理相关的范例，例如知觉学习[40]或暗示的眼球运动[41]。它也作为“情感价”（以自由能的负变化率表示）的理论研究的基础[42]。

对于不确定性起关键作用的范式，例如，由刺激约束（感官噪声）或环境因素（波动性）引起的范式，贝叶斯分级方法尤其有用。除了 HGF，最近还引入了其他几种贝叶斯模型，例如[43,44]。这些特别有助于神经调节递质的研究（例如，多巴

胺，DA，乙酰胆碱，乙酰胆碱，去甲肾上腺素，NE），与精神病学特别相关的应用领域和计算机建模的“经典”目标[45,46]。在过去的两年中，行为和神经影像数据的RL和贝叶斯建模对不同神经调节剂的作用产生了新的见解，尤其是在Yu和Dayan [47]的提议中，ACh和NE的释放编码了“预期不确定性”和“意外不确定性”。拟议的NE参与意想不到的不确定性信号的研究已经得到了对瞳孔大小变化[48,49]和fMRI [50]的研究的支持。一个突出的问题是，功能磁共振成像显示出未预料到的不确定性增加，蓝斑蓝素活性降低[50]。与预测的符号相反的关系[47]。

神经调节功能的一个重要方面是关注PE信号的自适应缩放[51]。例如，体育对学习的影响取决于其准确性（逆不确定性）[19, 34]。尽管已经发现中脑神经元活动反映了对奖励的精确加权[52,53]，但尚不清楚这种精确加权是否超出了奖励和多巴胺能系统。最近一项有关感觉学习的功能磁共振成像研究发现，关于视觉结果的精确加权PE激活了中脑（与奖励或新颖性无关）[40]。相比之下，在条件概率（听觉提示下的视觉结果）上的精确加权PE（概念上与预期不确定性有关的数量）由胆碱能性前脑的活动编码[40]。

参考 2 自由能

给定生成模型 m ，自由能 F 表示遇到数据 y 的惊喜（负对数概率）的上限。差异由近似后验密度 $q(\theta)$ 与真实但未知的后验密度 $p(\theta|y,m)$ 之间的Kullback-Leibler散度（KL；两种概率密度的相异性的度量）给出：

$$F = -\log p(y|m) + KL[q(\theta), p(\theta|y,m)] \quad (2)$$

等式 (2) 揭示了两件重要的事情。首先，与其直接计算后验密度（这可能是棘手的或计算上很昂贵的），不如将后方的能量最小化。这将最小化KL散度项，从而优化近似后验。其次，自由能代表了对数模型证据的下界，因此可以用于模型比较 [69, 71]。

3、模拟神经生理学

尽管通过数学模型了解脑病理生理学的尝试可以追溯到几十年前，但近年来，人们对与精神疾病相关的生理过程进行数学建模的兴趣已大大增加。可以区分两种通用方法。尽管其中一个是基于大脑活动的生成模型的倒置（如下所述），但更为经典的策略是基于生物物理上详细的动态系统模型，该模型描述了通过远程连接链

接的局部微电路或其集合。在最近的工作中已成功应用此方法的一个重要例子是多巴胺和 NMDA 受体 (NMDAR) 在精神分裂症中频繁发生认知功能障碍的作用：例如，工作记忆受损[54,55]。

这些模型的复杂性阻止了参数估计。但是，通过使用合适的正向模型进行增强，它们可以预测由于对实验操作敏感的神经元参数的变化而引起的 fMRI 或 EEG 数据的变化。此策略的一个很好的例子[56]使用前额叶皮层的生物物理模型来预测，阻塞 NMDARs 会导致锥体细胞活动减少工作记忆内容的分离表示，从而导致错误的特定行为模式。一组接受 NMDAR 拮抗剂 Ketamine 和安慰剂的健康志愿者证实了这一预测。

尽管上述方法对于在常规 (DSM) 定义的患者组中生成有关平均病理生理学的可测预测很有用，但它并不适合解决精神病学诊断中最关键的挑战：鉴别诊断，可以从观察到的行为和脑部生理推断出特定患者中最可能的疾病机理。在医学的其他领域，这种鉴别诊断通常由 (生化) 测定法支持，该测定法可从周围可及的组织 (例如血液) 中推断出“隐性”疾病机制。一个有吸引力的想法是使用计算模型在精神病学中建立等效程序，使用无创功能读数而不是组织样本。这些“计算分析”已经以生成模型的形式被提出，可以适合大脑活动和行为的测量[40]。

希望这样的测定法可以检测个体中 (未知) 病理生理过程的表达，并有助于区分异质性疾病中的亚组。理想情况下，此类测定应映射到直接适用于现有治疗方法 (药理或认知) 的过程；这将允许在纵向研究中可以 (必须) 评估不同的治疗预测。

尽管这不是一件轻而易举的事，而且距离任何重大成就都还很遥远，但在生理学领域，最近几年已经迈出了重要的第一步。这些通常依赖于神经元种群动态模型，该模型已充分简化以能够从 fMRI 或 EEG 数据进行参数估计 (模型倒置)，但又足够详细以至于它们保留了生理过程的有意义的总结[57-59]。建立的这类贝叶斯系统识别框架是动态因果模型 (DCM; [59,60])。对于功能磁共振成像，DCM 依靠未知神经元系统的低阶 (泰勒) 近似值，并解释了测量的 BOLD 信号是由大的未分化神经元群体中的突触耦合引起的[60]。尽管有这么粗略的表述，但最近的研究表明，这种类型的模型可能会很有用。例如，在慢性精神分裂症 (SZ) 中，无论表现如何或前额叶激活如何，在工作记忆过程中前额叶与顶叶的耦合都减少了[61]。对于处于不同疾病状态 (从健康状态到处于“处于危险状态的精神状态”到未经治疗的首发性 SZ) 的受试者，这种耦合作用逐渐下降，但恢复到与已治疗的首发患者的对照水

平无法区分的水平[62]。其他有关 SZ 的 DCM 研究已经仔细研究了工作记忆以外的其他任务的有效连通性，包括处于危险中的受试者[63,64]和慢性患者[65]。

与磁共振成像相比，用于电生理数据的 DCM 提供了更好的概念分辨率，可以区分不同类型的神经元和突触连接。在对啮齿动物进行较早的验证研究后（例如 [66]），最近在人体中进行的概念验证研究采用了多巴胺能药物挑战，以证明在前额微电路中推断 DA 诱导 NMDA 和 AMPA 电导率变化的可行性。 [67]。另一项研究使用 DCM 检验了 NMDAR 拮抗剂氯胺酮下依赖 NMDAR 的短期突触可塑性和神经元适应对失配负值（SZ 知觉推断受损的模型）降低幅度的贡献，找到了选择性减少听觉连接处的估计短期可塑性[68]。提供的这些结果可以在复制研究中得到证实，这种模型可以作为临床相关测定的蓝图，用于量化特定电路中发射器系统的状态。

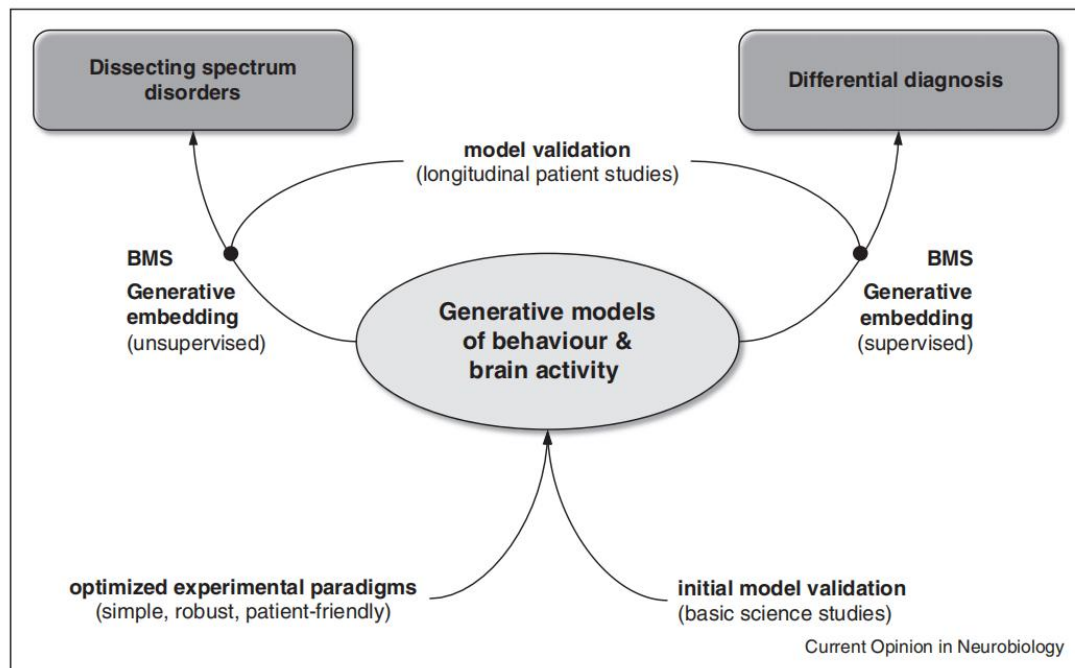


图 1 通过计算建模为精神病学诊断的未来提供了关键的方法学组成部分的图形摘要。

4、生成模型在鉴别诊断和亚组检测中的重要性

生成模型对于计算精神病学很重要的两个原因。首先，顾名思义，生成模型描述了如何通过隐藏的机制和原因生成观察到的数据（大脑活动或临床症状）（参见专栏 1）。因此，它们迫使我们机械地思考并明确我们的病理生理学理论。其次，对于任何给定的度量，可以想到不同的解释，即潜在（认知或神经元）过程的不同

模型。可以使用（对数）模型证据对这些模型进行比较，该模型证据可以通过变分贝叶斯方法在模型反演期间进行近似（参考 2），也可以使用经典近似法作为贝叶斯信息准则（BIC）进行比较。证据是模型拟合和模型复杂度之间权衡的原则性指标[69]，可用于在竞争模型之间进行裁决。近年来，这种贝叶斯模型选择（BMS）方法已在计算和神经生理学建模中得到越来越多的应用。例如 [26,39,40,41,50,62,68,70]。此外，对象在产生其行为的过程中可能会有所不同，也就是说，模型本身可能是总体中的随机变量。这个问题与精神病学所处理的异构频谱疾病特别相关，并且已经通过随机效应 BMS 方法的开发得到解决[71]。

相互竞争的生成模型的倒置，每个模型对于行为或大脑反应的测量都提供了不同的解释，随后通过 BMS 进行的比较可以为精神病学的鉴别诊断提供正式的框架。但是，这需要了解整个患者谱中表达的其他疾病途径。一种互补的方法是生成嵌入，其中相关模型参数的后验估计可用于构造特征空间以用于后续分类或聚类[72]。最初的无力和 SZ 患者的概念验证研究采用 DCM 进行功能性核磁共振成像，与标准方法相比，已证明生成嵌入的性能极佳[72,73]。最重要的是，生成嵌入可以帮助检测频谱混乱中的病理生理亚组。最近在 SZ 中证明了这种基于模型的亚组定义[73]：在这里，根据工作记忆过程中皮质区域之间相互作用的简单 DCM 的连通性估计的聚类，发现了三个患者亚组，它们通过不同的视觉-顶叶-前额连接（请参见图 2）。至关重要的是，这些纯粹的生理学定义的亚组表现出明显不同的临床症状水平。未来的希望是，以机械上可解释的模型进行索引，以不同疾病过程为特征的患者亚群的描述将为有关个体治疗的原则性预测提供依据，并最终为新的疾病学铺平道路。

5、总结和未来挑战

本文总结了建立基于模型的测定所需的方法学作为精神病学的新型诊断工具的最新进展。在接下来的几年中，进一步的进展可能将使神经建模的翻译策略的实际实施成为可能（图 1）：首先，建立可应用于患者任务数据的生成模型；第二，基于一组竞争神经元和/或认知机制的假设，针对给定的临床症状或测量进行鉴别诊断，每种假设均由特定的生成模型代表；第三，将异质性疾病分解为沿机械维度定义的亚组。

显然，许多挑战摆在面前。至关重要的是，无论采用哪种模型，都必须在基础验证研究中仔细评估其假设和稳健性，包括对动物和人类的初始药理学和刺激性研究（例如光遗传学）；有关示例，请参见[66,67,74,75]。随后，最重要的挑战是对具

有明确临床问题（例如结果或治疗反应）的患者进行纵向研究，这些问题可作为可以（且必须）测试我们的模型的临床实用性的现实基准。尽管尚无此类研究证明计算建模可以对精神病学的临床决策产生实际的实际影响，但在这方面正在进行的许多努力仍希望该主题在下次成为本期主题。该杂志将对精神病学的计算方法的实用性做出初步判断。

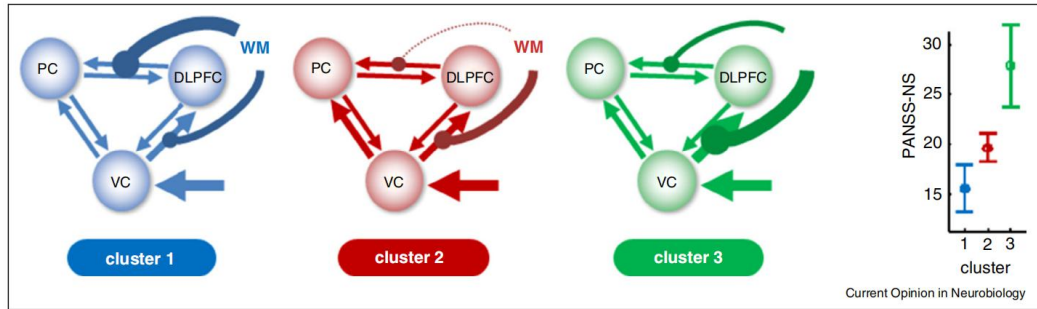


图 2 该图是从[73]复制而来的，它说明了如何使用生成模型将诸如精神分裂症等精神分裂症划分为不同的亚组。在这里，使用简单的视觉（VC），顶叶（PC）和背外侧前额叶皮层（DLPFC）之间相互作用的动态因果模型分析了 41 名精神分裂症患者的 fMRI 测量结果。然后，基于变分高斯混合模型，连通性估计值用作无监督聚类过程的输入。这表明连接数据的最合理的划分对应于三个不同的子组。这些子组的连通性体系结构在左侧以不同的颜色进行了图示显示（实线：正连通性估计，虚线：负连通性估计；线宽与各个估计的大小成比例）。重要的是，这三个纯粹由生理学定义的亚组的区别在于临床症状（右图）的显著不同（以阳性和阴性综合征量表（PANSS）衡量的“阴性症状”）。

Clearly, many challenges lie ahead. Critically, whatever models are proposed, their assumptions and robustness must be carefully evaluated in basic validation studies, including initial pharmacological and stimulation (e.g., optogenetics) studies in animals and humans; for examples, see [66,67^{••},74,75[•]]. Subsequently, the most important challenge is to conduct longitudinal studies in patients with well-defined clinical problems (such as outcome or treatment response) that serve as real-world benchmarks against which the clinical utility of our models can (and must) be tested. While there are no such studies yet which prove that computational modeling can have a real practical impact on clinical decision-making in psychiatry, the many ongoing efforts in this regard instill hope that by the next time this topic features in an issue of this journal, first studies will have expressed an initial verdict on the practical utility of computational approaches to psychiatry.

Acknowledgements

We are grateful to Quentin Huys for helpful comments on the manuscript. We would like to acknowledge support by the René and Susanne Braginsky Foundation (KES), the NCCR 'Neural Plasticity and Repair' (CM, KES), the joint initiative by University College London and the Max Planck Society on 'Computational Psychiatry and Ageing Research' (CM), and the Clinical Research Priority Programs (CRPP) 'Molecular Imaging' and 'Multiple Sclerosis' (KES).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Kapur S, Phillips AG, Insel TR: **Why has it taken so long for •• biological psychiatry to develop clinical tests and what to do about it?** *Mol Psychiatry* 2012, **17**:1174-1179.

An excellent summary of the reasons why psychiatric research has so failed to deliver clinically useful tests of disease mechanisms.

2. Cuthbert BN, Insel TR: **Toward the future of psychiatric •• diagnosis: the seven pillars of RDoC.** *BMC Med* 2013, **11**:126. A comprehensive outline of the problems associated with standard phenomenological classification schemes and of a research framework for establishing a revised nosology.
3. Tansley KE, Guipponi M, Perroud N, Bondolfi G, Domenici E, Evans D, Hall SK, Hauser J, Henigsberg N, Hu X *et al.*: **Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis.** *PLoS Med* 2012, **9**:e1001326.
4. Stephan KE, Baldeweg T, Friston KJ: **Synaptic plasticity and •• dysconnection in schizophrenia.** *Biol Psychiatry* 2006, **59**:929-939.

The first proposal to use generative models of behavior and brain activity as computational assays for detecting pathophysiological mechanisms and predicting treatment response.

5. Maia TV, Frank MJ: **From reinforcement learning models to •• psychiatric and neurological disorders.** *Nat Neurosci* 2011, **14**:154-162.

A thoughtful analysis of how to proceed from theoretical models to clinical applications in psychiatry and neurology.

6. Huys QJ, Moutoussis M, Williams J: **Are computational models of any use to psychiatry?** *Neural Netw* 2011, **24**:544-551.
7. Friston KJ, Dolan RJ: **Computational and dynamic models in neuroimaging.** *Neuroimage* 2010, **52**:752-765.

8. Montague PR, Dolan RJ, Friston KJ, Dayan P: **Computational •• psychiatry.** *Trends Cogn Sci* 2012, **16**:72-80. An important proposal of key questions, goals and methods for computational psychiatry.
9. <http://www.mpib-berlin.mpg.de/en/research/research-initiative-mps-ucl>.
10. <http://computationalpsychiatry.org>.
11. <http://www.translationalneuromodeling.org>.
12. Yoshida W, Dziobek I, Kliemann D, Heekeren HR, Friston KJ, Dolan RJ: **Cooperation and heterogeneity of the autistic mind.** *J Neurosci* 2010, **30**:8815-8818.
13. Moutoussis M, Bental RP, El-Deredey W, Dayan P: **Bayesian modelling of Jumping-to-Conclusions bias in delusional patients.** *Cogn Neuropsychiatry* 2011, **16**:422-447.
14. Frank MJ, Seeberger LC, O'Reilly RC: **By carrot or by stick: cognitive reinforcement learning in parkinsonism.** *Science* 2004, **306**:1940-1943.
15. King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR: **The rupture and repair of cooperation in borderline personality disorder.** *Science* 2008, **321**:806-810.
16. Kishida KT, King-Casas B, Montague PR: **Neuroeconomic approaches to mental disorders.** *Neuron* 2010, **67**:543-554.
17. Bullmore E, Vertes P: **From lichtheim to rich club: brain networks and psychiatry.** *JAMA Psychiatry* 2013, **70**:780-782.
18. Klöppel S, Abdulkadir A, Jack CR Jr, Koutsouleris N, Mourao-Miranda J, Vemuri P: **Diagnostic neuroimaging across diseases.** *Neuroimage* 2012, **61**:457-463.
19. Mathys C, Daunizeau J, Friston KJ, Stephan KE: **A Bayesian • foundation for individual learning under uncertainty.** *Front Hum Neurosci* 2011, **5**:39. A hierarchical Bayesian model for inferring individual mechanisms of (approximate) Bayes-optimality from measured behavior.
20. Botvinick MM: **Hierarchical reinforcement learning and decision making.** *Curr Opin Neurobiol* 2012, **22**:956-962.
21. Lee D, Seo H, Jung MW: **Neural basis of reinforcement learning and decision making.** *Annu Rev Neurosci* 2012, **35**:287-308.
22. Dayan P: **How to set the switches on this thing.** *Curr Opin Neurobiol* 2012, **22**:1068-1074.
23. Ito M, Doya K: **Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit.** *Curr Opin Neurobiol* 2011, **21**:368-373.
24. Gershman SJ, Niv Y: **Learning latent structure: carving nature at its joints.** *Curr Opin Neurobiol* 2010, **20**:251-256.
25. Doll BB, Simon DA, Daw ND: **The ubiquity of model-based reinforcement learning.** *Curr Opin Neurobiol* 2012, **22**:1075-1081.
26. Huys QJ, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP: **Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees.** *PLoS Comput Biol* 2012, **8**:e1002410.
27. Robbins TW, Gillan CM, Smith DG, de Wit S, Ersche KD: **Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry.** *Trends Cogn Sci* 2012, **16**:81-91.
28. Gläscher J, Daw N, Dayan P, O'Doherty JP: **States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning.** *Neuron* 2010, **66**:585-595.
29. Beierholm UR, Anen C, Quartz S, Bossaerts P: **Separate encoding of model-based and model-free valuations in the human brain.** *Neuroimage* 2011, **58**:955-962.
30. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model- •• based influences on humans' choices and striatal prediction errors.** *Neuron* 2011, **69**:1204-1215. A thorough examination of potentially competing learning mechanisms in humans using fMRI.

31. Dayan P, Hinton GE, Neal RM, Zemel RS: **The Helmholtz machine**. *Neural Comput* 1995, **7**:889-904.
32. Knill DC, Pouget A: **The Bayesian brain: the role of uncertainty in neural coding and computation**. *Trends Neurosci* 2004, **27**:712-719.
33. Friston K: **The free-energy principle: a unified brain theory?** *Nat Rev Neurosci* 2010, **11**:127-138.
 ●● A comprehensive theoretical framework for understanding adaptive cognition which has provided a foundation for many computational and physiological studies of mental disease.
34. Friston K, Kilner J, Harrison L: **A free energy principle for the brain**. *J Physiol Paris* 2006, **100**:70-87.
35. Fletcher PC, Frith CD: **Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia**. *Nat Rev Neurosci* 2009, **10**:48-58.
36. Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH: **Toward a neurobiology of delusions**. *Prog Neurobiol* 2010, **92**:345-369.
 ●● A compelling framework for understanding and experimentally studying delusions.
37. Adams RA, Perrinet LU, Friston K: **Smooth pursuit and visual occlusion: active inference and oculomotor control in schizophrenia**. *PLoS ONE* 2012, **7**:e47502.
38. Daunizeau J, den Ouden HE, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ: **Observing the observer (I): meta-bayesian models of learning and decision-making**. *PLoS ONE* 2010, **5**:e15554.
39. Lieder F, Daunizeau J, Garrido MI, Friston KJ, Stephan KE: **Modelling trial-by-trial changes in the mismatch negativity**. *PLoS Comput Biol* 2013, **9**:e1002911.
40. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, Stephan KE: **Hierarchical prediction errors in midbrain and basal forebrain during sensory learning**. *Neuron* 2013, **80**:519-530.
41. Vossel S, Mathys C, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE: **Spatial attention, precision, and bayesian inference: a study of saccadic response speed**. *Cereb Cortex* 2013 <http://dx.doi.org/10.1093/cercor/bhs418>.
42. Joffily M, Coricelli G: **Emotional valence and the free-energy principle**. *PLoS Comput Biol* 2013, **9**:e1003094.
 ● A compelling example how computational frameworks can be used to formally define clinically relevant mental states.
43. Payzan-LeNestour E, Bossaerts P: **Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings**. *PLoS Comput Biol* 2011, **7**:e1001048.
44. Nassar MR, Wilson RC, Heasly B, Gold JI: **An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment**. *J Neurosci* 2010, **30**:12366-12378.
45. Dayan P: **Twenty-five lessons from computational neuromodulation**. *Neuron* 2012, **76**:240-256.
46. Doya K: **Modulators of decision making**. *Nat Neurosci* 2008, **11**:410-416.
47. Yu AJ, Dayan P: **Uncertainty, neuromodulation, and attention**. *Neuron* 2005, **46**:681-692.
48. Preusschoff K, Hart BM, Einhauser W: **Pupil dilation signals surprise: evidence for noradrenaline's role in decision making**. *Front Neurosci* 2011, **5**:115.
49. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI: **Rational regulation of learning dynamics by pupil-linked arousal systems**. *Nat Neurosci* 2012, **15**:1040-1046.
50. Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP: **The neural representation of unexpected uncertainty during value-based decision making**. *Neuron* 2013, **79**:191-201.
51. Bunzeck N, Dayan P, Dolan RJ, Duzel E: **A common mechanism for adaptive scaling of reward and novelty**. *Hum Brain Mapp* 2010, **31**:1380-1394.
52. Schultz W, Preusschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P: **Explicit neural signals reflecting reward uncertainty**. *Philos Trans R Soc Lond B Biol Sci* 2008, **363**:3801-3811.
53. Fiorillo CD, Tobler PN, Schultz W: **Discrete coding of reward probability and uncertainty by dopamine neurons**. *Science* 2003, **299**:1898-1902.
54. Anticevic A, Gancsos M, Murray JD, Repovs G, Driesen NR, Ennis DJ, Niciu MJ, Morgan PT, Surti TS, Bloch MH *et al.*: **NMDA receptor function in large-scale anticorrelated neural systems with implications for cognition and schizophrenia**. *Proc Natl Acad Sci USA* 2012, **109**:16720-16725.
55. Rolls ET, Deco G: **A computational neuroscience approach to schizophrenia and its onset**. *Neurosci Biobehav Rev* 2011, **35**:1644-1653.
56. Murray JD, Anticevic A, Gancsos M, Ichinose M, Corlett PR, Krystal JH, Wang XJ: **Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model**. *Cereb Cortex* 2012 <http://dx.doi.org/10.1093/cercor/bhs370>.
 A nice demonstration how predictions from biophysical models can be tested in human pharmacological studies.
57. Breakspear M, Heitmann S, Daffertshofer A: **Generative models of cortical oscillations: neurobiological implications of the kuramoto model**. *Front Hum Neurosci* 2010, **4**:190.
58. Olier I, Trujillo-Barreto NJ, El-Dereby W: **A switching multi-scale dynamical network model of EEG/MEG**. *Neuroimage* 2013, **83C**:262-287.
59. David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ: **Dynamic causal modeling of evoked responses in EEG and MEG**. *Neuroimage* 2006, **30**:1255-1272.
60. Friston KJ, Harrison L, Penny W: **Dynamic causal modelling**. *Neuroimage* 2003, **19**:1273-1302.
61. Deserens L, Sterzer P, Wustenberg T, Heinz A, Schlagenhaut F: **Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia**. *J Neurosci* 2012, **32**:12-20.
62. Schmidt A, Smieskova R, Aston J, Simon A, Allen P, Fusar-Poli P, McGuire PK, Riecher-Rossler A, Stephan KE, Borgwardt S: **Brain connectivity abnormalities predating the onset of psychosis: correlation with the effect of medication**. *JAMA Psychiatry* 2013, **70**:903-912.
63. Dauvermann MR, Whalley HC, Romaniuk L, Valton V, Owens DG, Johnstone EC, Lawrie SM, Moorhead TW: **The application of nonlinear dynamic causal modeling for fMRI in subjects at high genetic risk of schizophrenia**. *Neuroimage* 2013, **73**:16-29.
64. Diwadkar VA, Wadehra S, Pruitt P, Keshavan MS, Rajan U, Zajac-Benitez C, Eickhoff SB: **Disordered corticostriatal interactions during affective processing in children and adolescents at risk for schizophrenia revealed by functional magnetic resonance imaging and dynamic causal modeling**. *Arch Gen Psychiatry* 2012, **69**:231-242.
65. Banyai M, Diwadkar VA, Erdi P: **Model-based dynamical analysis of functional disconnection in schizophrenia**. *Neuroimage* 2011, **58**:870-877.
66. Moran RJ, Jung F, Kumagai T, Endepols H, Graf R, Dolan RJ, Friston KJ, Stephan KE, Tittgemeyer M: **Dynamic causal models and physiological inference: a validation study using isoflurane anaesthesia in rodents**. *PLoS ONE* 2011, **6**:e22790.
67. Moran RJ, Symmonds M, Stephan KE, Friston KJ, Dolan RJ: **An in vivo assay of synaptic function mediating human cognition**. *Curr Biol* 2011, **21**:1320-1325.
 A proof-of-concept study of model-based assays, using a dopaminergic drug challenge to demonstrate the feasibility of inferring drug-induced changes in NMDA and AMPA conductances in a prefrontal microcircuit.
68. Schmidt A, Diaconescu AO, Kometer M, Friston KJ, Stephan KE, Vollenweider FX: **Modeling ketamine effects on synaptic plasticity during the mismatch negativity**. *Cereb Cortex* 2013, **23**:2394-2406.

69. Penny WD: **Comparing dynamic causal models using AIC, BIC and free energy.** *Neuroimage* 2012, **59**:319-330.
70. Frank MJ, Badre D: **Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis.** *Cereb Cortex* 2012, **22**:509-526.
71. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ: **Bayesian model selection for group studies.** *Neuroimage* 2009, **46**:1004-1017.
72. Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM, Stephan KE: **Generative embedding for model-based classification of fMRI data.** *PLoS Comput Biol* 2011, **7**:e1002079.
73. Brodersen KH, Deserno L, Schlagenhaut F, Lin Z, Penny WD, • Buhmann JM, Stephan KE: **Dissecting psychiatric spectrum disorders by generative embedding.** *Neuroimage: Clin* 2013, **4**:98-111.
- A first demonstration how generative models could be used for identifying mechanistically distinct subgroups in spectrum disorders.
74. David O, Guillemain I, Sallet S, Reyt S, Deransart C, Segebarth C, Depaulis A: **Identifying neural drivers with functional MRI: an electrophysiological validation.** *PLoS Biol* 2008, **6**:2683-2697.
75. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, • Janak PH: **A causal link between prediction errors, dopamine neurons and learning.** *Nat Neurosci* 2013, **16**:966-973.
- An impressive example how optogenetics can be used to validate central assumptions of computational models.