科学通报 2022 年 第 67 卷 第 13 期: 1423 ~ 1435
评 述
《中国科学》杂志社
SCIENCE CHINA PRESS

欺骗的认知神经网络模型

张英良, 买晓琴*

中国人民大学心理学系, 北京 100872

* 联系人, E-mail: maixq@ruc.edu.cn

2021-12-28 收稿, 2022-02-14 修回, 2022-02-14 接受, 2022-02-15 网络版发表
中国人民大学科学研究基金(中央高校基本科研业务费专项资金资助)项目成果(21XNA031)

摘要 欺骗是一种当事人有意试图让他人相信某种错误信念, 从而为自己或他人获取利益或规避损失的心理过程. 已有的研究为揭示欺骗的神经机制提供了启示, 但欺骗这一复杂行为中大规模脑网络的整体性作用尚未得到足够的关注. 本文总结了欺骗的相关理论和研究, 以及欺骗行为涉及的认知和情感加工过程, 并在此基础上提出了欺骗的认知神经网络模型. 该模型中, 欺骗行为产生于动力系统、情感系统、认知系统和执行系统之间的动态交互过程, 而奖赏网络、突显网络、中央执行网络和默认网络的激活和相互作用是其背后的神经基础. 模型建立了欺骗相关的研究和理论、心理过程、心理功能、脑区与神经网络之间的联系, 并有望对欺骗行为及其神经机制作出更为整体性、系统化的解释.

关键词 欺骗, 欺骗理论, 功能性磁共振成像, 事件相关电位, 认知神经网络模型

欺骗作为一种常见的行为和策略, 广泛存在于人类社会生活的各个方面. 欺骗具有多种表现形式, 例如玩笑、魔术表演、经济诈骗, 以及各种竞争活动中的误导性策略等. 在社会层面上, 欺骗和诚信足以影响人际交往的方方面面, 从而对整个社会风气产生影响. 对个人和组织而言, 实施和识别欺骗行为同样具有重要的意义.

对于欺骗的定义并不完全统一, 本文采用领域内较为主流的定义[1]: 欺骗是一种当事人有意试图让他人相信某种(当事人认为的)错误信念, 从而为自己或他人获取利益或规避损失的心理过程. 欺骗应至少包含以下几个子过程. (1) 决策: 形成欺骗意图, 收集他人心理状态等信息, 评价各种可能的结果; (2) 执行: 抑制诚实反应, 隐瞒真实信息、发出误导信息等; (3) 反馈: 对欺骗行为的结果进行评估[2]. 与之相似的说谎, 则在此框架下定义为通过言语方式开展的欺骗行为, 即言语欺骗(verbal deception), 属于欺骗的一种. 针对言语欺骗的研究集中于应用领域, 例如欺骗的识别和区分方面[3~7], 由于相关研究与本文主题关系不够密切, 不再单独讨论, 该领域的研究进展和总结详见Vrij[7]的述评.

从20世纪60年代起, 心理学家开始展开以测谎为主要目的的研究, 以解决刑侦、法律方面的实际问题. 21世纪以来, 研究者开始运用神经成像和神经生理学方法探索欺骗的神经机制. 近十年的欺骗研究则进一步明确了研究问题、细化了研究领域, 在前人研究的基础上通过更为先进的研究方法和范式发现了新的结果. 对欺骗的研究可按任务的性质分为两类[8,9]: 一类是“指示性欺骗”研究, 即实验者通过不同的线索指示被试作出欺骗或诚实反应; 另一类则是“自发性欺骗”研究, 即被试可以自由选择欺骗的时机. 前者的优点在于实验设计、实验流程和数据分析更加简单清晰, 同时也难以避免生态效度低、忽略个体差异等缺点; 后者作为更新颖的研究方法能够在一定程度上解决以上问题, 但也对研究设计提出了更高的要求.

引用格式: 张英良, 买晓琴. 欺骗的认知神经网络模型. 科学通报, 2022, 67: 1423–1435
Zhang Y L, Mai X Q. The cognitive neural network model of deception (in Chinese). Chin Sci Bull, 2022, 67: 1423–1435, doi: 10.1360/TB-2021-0963

本文综合前人的实证、综述类研究和理论解释, 通过分类讨论的方法对欺骗研究进行整理, 并通过提出欺骗的认知神经网络模型, 为理解欺骗行为的发生和发展提供一个新颖、系统、动态的视角. 本文首先介绍欺骗的主流理论; 其次, 总结了欺骗涉及的认知和情绪加工过程; 最后, 基于已有的理论和研究建构欺骗的认知神经网络模型.

1 欺骗的理论

为了探明欺骗行为背后的认知机制, 研究者从不同的侧重点提出了一系列理论解释, 包括分析欺骗过程涉及的心理因素的四因素理论、将欺骗行为视为特殊的决策过程的欺骗决策理论、着重解释社会认知过程的人际欺骗理论(Interpersonal Deception Theory, IDT)、关注神经层面的神经生理模型, 以及将欺骗过程视为多个子过程的有机组织的激活-决策-构建-行动理论(Activation-Decision-Construction-Action Theory, ADCAT)等.

1.1 欺骗的四因素理论

Zuckerman等人[10]提出的欺骗四因素理论认为主要有以下4种心理因素参与了欺骗过程: 控制、唤醒、情绪和认知. 在欺骗过程中, 控制能力是必不可少的, 因为欺骗者需要控制自己的言语、动作和表情以免产生信息的泄露或出现矛盾和破绽, 从而导致欺骗失败. 此外, 欺骗过程伴随着更强的生理唤醒[11]. 与欺骗行为相关的情绪主要包括罪恶感、焦虑和成功后的喜悦等. 这些情绪的效价取决于欺骗的目的、欺骗发生的社会情境和欺骗者的人格特征等因素. 最后, 欺骗是一项具有高度复杂性和整合性的认知任务, 因为说实话只需要回忆和复述, 而欺骗则需要编造细节, 而且不能与欺骗对象具有的知识和信念相冲突. 因此, 说谎一般比说实话需要更长的准备时间[8].

1.2 欺骗的决策理论

欺骗在广义上也是一种决策的过程. 欺骗的决策理论着眼于解释个体在欺骗和诚实之间如何选择, 即欺骗如何发生、何时发生. 古典经济学在解释决策时秉承“理性人”假设, 即人是完全自私、理性的, 进行决策的依据是各个策略的效用和价值, 欺骗者致力于通过权衡比较欺骗行为和诚实行为的收益与损失, 将自己的期望收益最大化.

Mazar[12]认为, 从心理学的角度出发, 决策者并不仅关注外部奖赏, 也会计算内部奖赏. 欺骗会带来额外的代价, 如损害个体的正面自我概念、产生焦虑等负性情绪、加重认知负担等. 此时个体往往会调整策略, 例如选择进行相对轻微的欺骗, 并通过暂时忽视或调整道德标准以重新解释事实, 保证自我概念不受威胁[12].

一些研究者指出, 大部分欺骗行为的目的是尝试避免损失, 而不是获得更多利益[13]. 当人们认为, 只有通过欺骗才能脱离困境时, 欺骗的发生率就会大大增加. 相比于获益, 人们常常对相同幅度的损失更敏感[14], 因此对损失的厌恶会产生一种近乎本能的强烈的“超动机”, 这就解释了人们常常在困境中更容易欺骗的现象. 除了超动机以外, 恐惧和绝望等负性情绪也可能是人们面临损失时欺骗行为的诱因.

1.3 欺骗的社会认知理论: 人际欺骗理论

人际欺骗理论[15]强调传统欺骗理论所忽视的社会沟通过程, 认为欺骗是一个相互交流的过程, 具有合作的本质属性. 欺骗者和听众将参与到一系列同步、双向的信息捕捉与计算过程中. 例如, 欺骗者会观察听众的反应来判断对方的心理状态, 并对后续策略进行相应调整; 听众则会从欺骗者的言语、行为和表情中提取可能的欺骗线索, 以推测对方的意图和信息真实性等. 因此, 欺骗过程会增加欺骗双方的认知资源的消耗. 欺骗的认知负担取决于多种因素. 对于欺骗者, 主要取决于沟通活动的情境, 如信息的种类、数量、一致性, 以及欺骗目标的数量和难度等; 对于欺骗对象, 主要取决于欺骗者所表露出的各种线索的明确性、一致性和可信度等.

1.4 欺骗的神经生理模型

Mohamed等人[16]试图从神经生理角度解释欺骗行为, 提出了欺骗的神经生理模型. 该模型将欺骗过程分解为7个阶段, 分别是: 知觉问题、理解问题、回忆问题相关内容、反应判断和计划、情绪反应、口头反应, 以及交感神经系统的反应. 这些阶段分别由不同的脑区负责, 彼此间可能存在重叠. 对于欺骗者和诚实者而言, 他们在前3个阶段(知觉问题、理解问题和回忆问题相关内容)会进行相同的认知处理, 因此应该会发生相似的脑活动. 在口头反应阶段, 尽管谎言与真话内容不同, 但均会激活相似的脑区, 例如布洛卡区和中央前回.

因此, Mohamed等人[16]认为, 识别说谎者最好的时机是第4个阶段, 即反应判断和计划阶段, 原因是此时说谎者需要抑制真实反应并采取欺骗反应.

1.5 激活-决策-构建-行动理论

Walczyk等人[17]的激活-决策-构建-行动理论认为, 说谎由4个依次发生的子成分构成: 激活成分、决策成分、构建成分和行动成分. 这些成分可能会自动、无意识地工作. 激活成分的作用是使欺骗者发现潜在的欺骗对象正在寻求真相, 从而意识到存在欺骗对方的机会, 并从工作记忆中调取相关信息. 随后, 决策成分获取社会情境信息或已做的决定, 以此方式确定欺骗的动机水平. 在欺骗决策完成后, 构建成分通过虚构、隐瞒、夸大或改进已有谎言来操纵信息, 形成欺骗内容. 最后, 行动成分则对应欺骗者向欺骗对象说谎的过程.

另外, ADCAT有两个核心结构, 分别是心理理论和认知资源. 心理理论是欺骗者推断他人心理状态的工具, 在欺骗中扮演重要角色. 而认知资源相关的核心原则是: 由于认知资源有限, 且欺骗者试图在欺骗时表现得真诚、流畅和放松, 他们会设法降低行动成分中内在的认知负担, 以及贯穿欺骗全程的外源性认知负担.

1.6 理论总结

四因素理论和决策理论分别关注构成欺骗的心理要素和影响欺骗动机的奖赏和成本, 而忽略了欺骗过程的发生机制; 欺骗的神经生理模型和ADCAT对机制性问题进行了解释, 但前者特异性不足, 后者缺乏对神经机制的解释, 且仅适用于言语欺骗; 人际欺骗理论是唯一关注欺骗过程人际性特点的理论, 但该理论并未深入探讨其背后的认知过程. 通过归纳、总结和对比以上理论, 欺骗行为的发生应当建立在如下心理过程的基础上: (1) 动机过程, 即在追求内外部奖赏或避免损失的驱使下产生欺骗动机; (2) 认知加工过程, 按时间顺序依次包括知觉和理解欺骗相关的情境与回忆等信息、完成欺骗决策、计划和构建欺骗内容以及隐藏和保持欺骗内容等, 是欺骗最复杂的子过程; (3) 社会人际加工过程, 包括与欺骗对象进行双向互动、推测他人想法等, 进而形成和调整欺骗策略; (4) 情绪加工过程, 在欺骗过程中欺骗者会产生焦虑感、罪恶感和欺骗成功的喜悦等情绪体验, 取决于欺骗行为的具体性质. 因此, 以上理论虽然对理解欺骗过程的产生和发展各有启发, 但是均难以深入和全面地解释其背后的认知机制. 一个更加成熟和完整的欺骗理论需要: (1) 涉及支持欺骗过程的心理过程和功能; (2) 能够解释欺骗发生的动态过程; (3) 考虑到欺骗行为的人际性特点.

2 欺骗涉及的认知和情绪加工

通过总结前人的研究可以发现, 欺骗所引发的行为表现、激活的神经结构和诱发的脑电反应十分复杂, 对应着不同的心理过程和功能. 根据其中, 最具有代表性的心理过程和功能包括执行功能、心理理论、情绪反应和奖赏与价值表征.

2.1 奖赏与价值表征

欺骗过程中的奖赏与价值表征存在于两个阶段. 首先, 欺骗行为的动机是达成某种积极的结果. 因此, 欺骗者对欺骗目的的奖赏和价值的评估是决定是否欺骗的先决因素. 其次, 欺骗者对欺骗结果的奖赏和价值表征对欺骗决策形成反馈, 影响后续的认知加工和决策过程.

自发性欺骗研究中, 实验者一般通过设置物质奖励或情感奖励来诱发被试的欺骗行为, 以观察被试的行为表现和神经活动. 与奖赏和价值表征有关的神经结构包括腹内侧前额叶皮层(ventromedial prefrontal cortex, vmPFC)、纹状体(striatum)、伏隔核和中脑腹侧被盖区等[18,19]. 对于预期性的奖赏与价值评估阶段, Sun等人[20]观察到奖赏网络的重要节点——纹状体在欺骗行为最初阶段的激活, 初步表明了上述相关结构在欺骗早期过程中的参与. 对于反馈性的奖赏与价值评估阶段, Sun等人[21]发现, 与诚实反应相比, 成功的欺骗行为会激活腹侧纹状体和后扣带回皮层(posterior cingulate cortex, PCC)、减小反馈相关负波的波幅, 表明欺骗行为的正性结果将引发价值评估和注意分配过程. 此外, 对计算机(相比于人类)对手失败的欺骗(相比于诚实)增强了P3b成分, 该交互作用似乎表征了诚实反应的奖赏. Zhu等人[22]发现, 自发欺骗降低了奖赏相关正波(reward positivity, RewP)的波幅及其delta和beta波段的能量, 而增大了theta波段的能量. RewP一般出现于行为反馈后200~300 ms, 反映了对结果效价的表征[23]. 该结果证明, 欺骗具有内在成本, 降低了被试对结果的奖赏预期[22]. 伏隔核的激活水平与伪装身份的欺骗行为有关[24], 并能正向预测随后任务中被试面对奖赏时的欺骗频率[25,26]. 以上结果反映了奖赏与价值

表征过程在欺骗过程中可能扮演了诱发和推动的作用, 决定了欺骗过程的动机水平. Pornpattananangkul等人[27]同样发现, 被试的欺骗频率越高, vmPFC激活水平越高, vmPFC-dlPFC功能连接也越强, 表明了价值(奖赏)系统和中央执行系统之间的联系在欺骗行为中的重要意义.

报告奖赏相关脑区在欺骗过程中激活的研究数量较少[27], 可能的原因包括: 研究者通常基于欺骗与诚实的对比来定位欺骗相关脑区, 但部分研究(特别是指示性欺骗研究)中欺骗行为不能带来明显的奖赏, 因此奖赏网络难以在对比中凸显出来; 类似地, 在部分实验设计中, 诚实反应同样能带来避免负性结果和情绪等奖赏; 此外, 研究者往往侧重于前额叶等负责高级认知功能的脑区, 而不把奖赏网络的构成部分作为兴趣区域.

2.2 执行功能

欺骗过程具有如下基本要求: 欺骗者必须生成虚假信念, 并同时保持真实信念和虚假信念, 最后在此期间时刻抑制真实信念的泄露. 因此大多数欺骗研究持有“默认诚实”的基本假设, 即诚实反应符合内化的社会规范和自我一致性需求, 因此是本能性的反应; 而欺骗反应则是困难而复杂的, 依赖抑制控制、冲突监控、行为选择等心理功能, 故相比于诚实反应, 将产生更高的认知负担[1,28,29].

很多研究结果支持这一观点. 首先, 在行为水平上, 改变任务的认知负担能够影响被试的欺骗表现. 例如, 通过增大欺骗任务中叙述环节的认知负担, 可以使观察者有效区分欺骗组和诚实组被试[30], 而升高认知负担和对额下回施加经颅直流电刺激, 则均能让被试的欺骗反应与诚实反应的表现更相似[31].

其次, 在神经影像学水平上, 一些与执行功能相关的脑区在欺骗条件下激活水平高于诚实条件, 特别是前额叶皮层(prefrontal cortex, PFC)的作用得到了大量研究的重复证实[1,32~35]. 其中, 前扣带回皮层(anterior cingulate cortex, ACC)被认为与冲突检测有关[36]; 背外侧前额叶皮层(dorsolateral prefrontal cortex, dlPFC)对认知控制具有贡献, 属于认知控制网络的一部分[32,37]; 腹外侧前额叶皮层(ventrolateral prefrontal cortex, vlPFC)与行为选择和反应抑制有关[38~40].

最后, 在神经生理学水平上, 同样发现了欺骗条件对与执行功能相关的ERP成分的效应. 一个典型的例子是欺骗的N2-P3效应, 即欺骗行为伴随额部N2成分的增大和顶部P3成分的减小[32,41~45]. 以往研究表明, N2与冲突监控相关[46], 而P3被认为具有评价性质, 与调用执行控制资源、冲突解决等过程联系紧密, 反映了认知资源的存量[47], 因此该效应表明, 欺骗需要消耗认知资源以抑制默认的诚实反应. 有研究认为, N2的结构来源是ACC[36], 进一步证明了欺骗和诚实反应的竞争所引发的认知冲突激活了冲突监控过程, 从而使N2波幅增大[44,48,49].

然而, 尽管绝大多数研究支持欺骗会产生更高的认知负担, “默认诚实”的基本假设始终存在争议. 一种观点称为“优雅假设(Grace Hypothesis)”, 即作出诚实行为是一种自动化的倾向; 而与之相对的观点被称为“意志假设(Will Hypothesis)”, 认为诚实行为的产生需要通过有意识的控制来抵抗欺骗的诱惑[50]. 研究者关于本问题存在不同见解, 部分研究更支持前者[25,50,51], 部分研究更倾向于后者[52,53]. 综合以上结果, 两种假设都不能单独解释研究结果, 二者应该在引入个体差异等因素的前提下达成某种统一[26,54].

Speer等人[26]为两种假设的统一提供了有力证据, 发现伏隔核的活动促进欺骗决策, 特别是对于不诚实者; 而PCC、内侧额叶和颞顶联合区(temporoparietal junction, TPJ)的活动促进诚实决策, 特别是对于诚实者; ACC和额下回的激活会使“不诚实者”变得更诚实, 而使“诚实者”变得更不诚实. 也就是说, 由于不同个体在奖赏评价和道德认同等方面存在差异, 他们在欺骗机会面前形成了不同的“道德默认状态”, 主动的认知控制能够使个体偏离原有的默认状态. 而两种假设分别是“道德默认状态”的两种极端情况, 即只考虑欺骗奖赏、不关心道德自我概念的“意志假设”, 以及与之相反的“优雅假设”. 而每个个体的“道德默认状态”都位于二者之间的一条数轴上, 行为上表现为欺骗的频率和幅度. 因此诚实和欺骗哪个是更默认、更自动化的反应, 需要考虑个体差异等因素.

对儿童欺骗行为的研究一方面能反映儿童各方面心智的发展状况, 另一方面通过结合儿童各种心理能力的发展趋势, 能够为理解欺骗背后的认知机制提供侧面的证据. 最近的元分析研究[55]发现, 儿童欺骗行为发展与执行功能之间存在一个小而稳定的相关(r=0.13), 并且进一步指出, 执行功能会影响儿童的欺骗维持能力, 但不影响产生谎言的能力. 这似乎说明, 欺骗的维持更加需要执行功能的参与.

2.3 心理理论

由于欺骗过程涉及欺骗者与欺骗对象的交互，特别是欺骗者有意操纵欺骗对象信念的过程，欺骗者必须具备一定的社会认知能力以确保欺骗行为的顺利实施，例如心理理论. 心理理论是一系列广泛的能力，包括意识到他人具有与自己不同的想法、知识和信念，感知和理解他人心理状态等[56]. 心理理论对欺骗者获取有关他人心理状态的信息，并在此基础上生成虚假信念、实施欺骗策略具有重要意义.

在成人被试和儿童被试中，心理理论能力与欺骗表现均存在相关[29,55,57]. 以往的神经成像学研究发现了一系列支持心理理论功能的脑区，例如TPJ与评估他人意图有关[58]; 背内侧前额叶皮层(dorsomedial prefrontal cortex，dmPFC)是参与表征他人利益、形成有关他人的印象的关键脑区[56,59]，PCC的激活水平在静息状态下增强，且在认知任务中与某些前额区域的激活水平存在负相关[60]; 顶下小叶涉及无定向思维、情景记忆和社会认知等过程[61,62]; 楔前叶则在作出道德相关的行为反应[63]和自我中心的心理想象过程[64]中激活. 以上脑区均在欺骗条件下表现出活动增强[8,33,65].

近期的元分析研究证实了心理理论与儿童欺骗行为之间的相关[55,57]，发现心理理论能力越强的儿童越可能进行欺骗. 此外，研究者进一步发现，心理理论能力与欺骗的产生和维持能力相关，并且欺骗类型会影响这种相关关系，表现为用于掩饰错误的欺骗与各心理理论成分无关，而在竞争游戏中的欺骗与对多重信念和情绪信念的理解有关[57].

2.4 情绪反应

诚实是最基本的社会规范之一，而对这种社会规范的内化导致人们作出本能性、自动化的诚实反应，并在违背诚实规范进行欺骗时产生内疚、焦虑和厌恶等负性情绪[22,51]. 除了违背诚实原则本身以外，对欺骗失败的风险的感知和加工也是产生负性情绪的潜在原因.

大量研究证实，欺骗行为伴随着内疚、焦虑和厌恶等负性情绪体验[8,25,49]，以及杏仁核(amygdala)、前脑岛(anterior insula，AI)和背侧前扣带回皮层(dorsal anterior cingulate cortex, dACC)等脑区的激活[8,11,20,27,66~69]. 考虑到以上脑区在厌恶、恐惧等负性情绪产生和表达中的作用[70~73]，上述研究结果在神经层面进一步证明了欺骗过程中情绪加工的存在.

3 欺骗的认知神经网络模型

已有研究对欺骗活动脑机制的探讨基本上局限于脑区水平，仅仅通过脑区的独立激活结果解释欺骗背后的神经机制，而大多数认知活动需要多个脑区组成的神经网络的支持[74]，特别是欺骗等具有社会性、复杂性和综合性的认知任务，往往需要多个神经网络的协同合作.

前文中，综合欺骗的相关理论，欺骗行为应涉及动机过程、认知加工过程、社会人际加工过程和情绪加工过程等子过程; 相应地，本领域的研究结果所涉及的脑区可分为四类[1,32,58,75]：与奖赏、价值表征相关的纹状体、vmPFC等; 与心理理论、社会决策、共情等社会功能相关的dmPFC、TPJ和PCC等; 与归因、问题解决、计划制定和执行功能相关的vlPFC、dlPFC等; 与觉察凸显刺激、产生负性情绪相关的杏仁核、AI和dACC等. 这四类脑区分别属于4个神经网络，即奖赏网络、默认网络、中央执行网络和显著网络(图1).

为了从更全面完整的视角解释欺骗发生的神经机制，本文基于上述神经网络在欺骗研究中的激活模式，将其整合成具有动态特征的反馈结构，提出欺骗的认知神经网络模型(图2). 奖赏网络、中央执行网络、默认网络和显著网络通过交互构建支持欺骗过程的动力系统、认知系统、情感系统和执行系统，各系统之间相互配合，作为一个整体产生欺骗全过程中各种策略、行为和情绪反应.

首先，个体通过感知觉和记忆内容表征所处的情境，获取实施欺骗行为所必需的信息，如各种事实、知识以及潜在欺骗对象的意图、人格特质和信息掌握情况等. 动力系统据此产生欺骗的动机，并驱动认知系统对欺骗进行计划、构建和保持. 认知系统由中央执行网络和默认网络组成，二者相互补充. 中央执行网络生成欺骗内容，并同时保持真伪两种不同的信念; 默认网络负责进行人际层面的觉察和推测. 一方面，认知系统输出的欺骗内容和策略激活情感系统，使其对可能的不利后果产生各种情绪反应，随之进一步作用于动力系统，调节对欺骗内在奖赏和代价的表征; 另一方面，执行系统执行欺骗决策，完成欺骗行为，这一步的重点是抑制真实反应以及避免认知和行为的冲突. 欺骗的结果，如欺骗是否成功、欺骗对象的反应将被欺骗者重新加工，以指导后续行动. 该模型可分为几部分: (1)

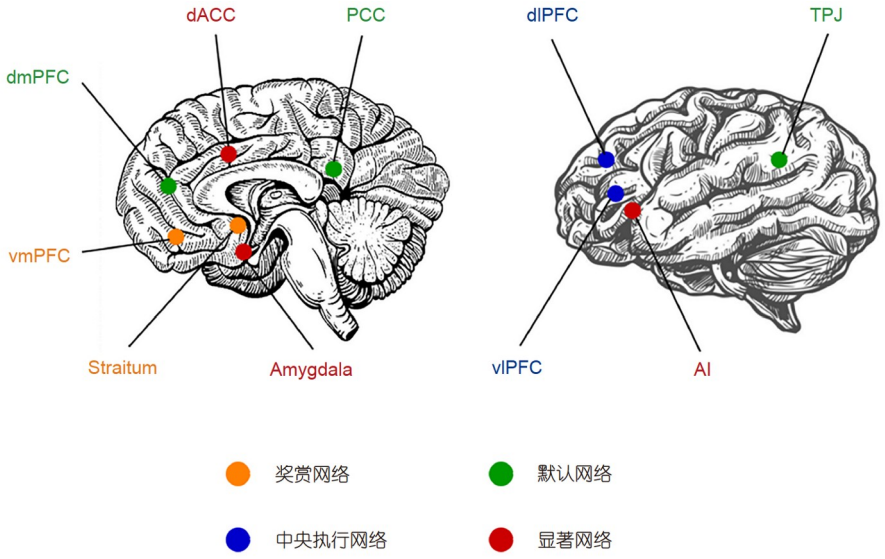

图 1 欺骗相关的脑结构和对应的大规模脑网络
Figure 1 Brain regions and corresponding large-scale brain networks related to deception

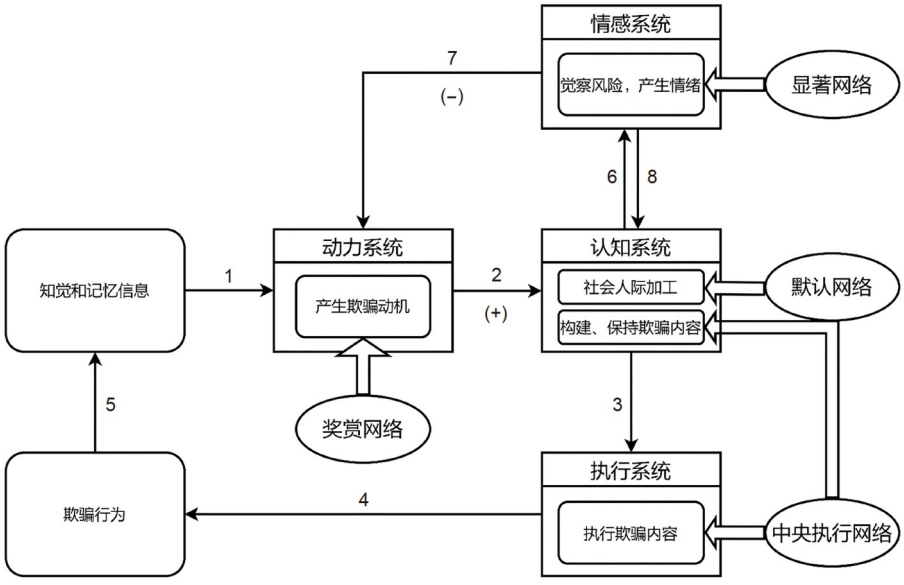

图 2 欺骗的认知神经网络模型. 直角矩形表示各系统; 圆角矩形表示认知或情感加工过程; 椭圆形表示各大规模脑网络
Figure 2 The cognitive neural network model of deception. Rectangles represent systems, rounded rectangles represent cognitive/emotional processes, and ellipses represent large-scale brain networks

动力系统的激活; (2) 认知系统的计划、构建和保持; (3) 情感系统的加工过程; (4) 执行系统的执行过程; (5) 欺骗过程中的反馈机制.

3.1 动力系统的激活

欺骗行为的目的不外乎追求利益或逃避不利后果, 这种正面的结果正是实施欺骗的奖赏. 动力系统在欺骗过程的最初阶段即产生激活, 是欺骗的起点. 动力系统首先在知觉和记忆信息提取的基础上对可能的奖赏进行预期, 从而为欺骗行为提供动力(如图2箭头1所示); 其次, 该系统接受来自认知系统和情感系统的负反馈调节为主的作用, 重新确定对欺骗结果的奖赏和价值评估(如图2箭头7所示).

有研究发现, 奖赏网络的重要节点, 如纹状体、伏

隔核等脑区的激活水平和对奖赏的敏感性，与欺骗频率呈正相关，甚至能正向预测后续的欺骗决策的频率和幅度[25,26]. 一些奖赏网络受损的个体, 如帕金森综合征患者, 对通过欺骗获利表现出更少的兴趣, 相比于正常人具有更少的欺骗行为[76].

3.2 认知系统的计划、构建和保持

在完成了对环境和欺骗对象的评估后，激活的动力系统产生欺骗动机，从而启动认知系统对欺骗的计划、构建和保持(如图2箭头2所示). 认知系统是欺骗过程的核心，负责对欺骗策略和欺骗内容的规划与构建, 为执行系统提供执行对象(如图2箭头3所示). 认知系统由“理性”的中央执行网络和“感性”的默认网络组成, 前者支持一系列人类特有的“理性”的逻辑思维和精细计算等功能, 后者则负责更加“感性”的社会认知过程. 与“默认诚实”的基本假设一致，认知系统对欺骗的计划、构建和保持会消耗额外的认知资源，体现了欺骗的复杂性和认知消耗性.

在评估情境、获取信息的基础上，欺骗者需要执行功能以作出并执行决策，而中央执行网络就是该过程的神经基础. 中央执行网络维系着计算、认知和行为抑制、目标的选择和保持等心理功能，在决策中扮演重要作用. 为了顺利完成欺骗的构建、保持和实施, 个体必须保持欺骗目的、掌握和操纵自己与他人的心理状态以及监控自己接收和发送的各种信息，这些过程都依赖于执行功能. 执行功能是多种自上而下的高级认知功能的集合，其作用是针对情境不断修正和控制个体的认知与行为[77]. 一般认为，执行功能包含3个成分, 即抑制控制、工作记忆和任务转换[78]. 欺骗过程中的抑制控制主要体现为自我控制(self-control)，即压抑不合适的行为. 工作记忆即在脑海中保持并操作信息的能力[79]，是人完成各种认知任务最重要的保证之一. 任务转换能力也称为认知灵活性，包括切换(空间和人际)视角、策略和目标的能力等[77]. 执行功能对完成欺骗行为至关重要，而最近的一项元分析研究[80]通过对比自发性欺骗和指示性欺骗的脑成像结果，发现自发性欺骗相比于指示性欺骗额外激活了vlPFC和膝周侧ACC，并揭示了后者在负性情绪产生等过程中的作用. 此外，dlPFC等认知控制相关脑区的激活可能与遵从指示有关, 而不一定是自主欺骗决策所致. 该结果初步分离了动力、意图过程与一般的认知控制过程.

认知系统中，中央执行网络的作用体现为支持工作记忆和任务转换能力. 工作记忆保证了欺骗者能够有效保持欺骗目标和策略, 并隐蔽地形成欺骗内容, 是欺骗过程中内部状态一致性和连贯性的基础；任务转换能力则在有效采取他人(如欺骗对象)的视角、根据信息灵活调整欺骗策略等过程中得以体现.

社会认知过程对欺骗过程同样至关重要. 欺骗者需要对情境进行人际评估，例如推测欺骗对象的信念、意图等心理状态, 以获取欺骗所必需的信息, 并据此生成欺骗内容、预期欺骗的结果. 以上心理过程和能力的相关脑区与默认网络的节点高度重合(2.3节), 因此本模型中, 默认网络作为认知系统的重要一环, 是社会人际认知的神经基础. 自Raichle等人[81]揭示了大脑中一系列与静息状态和自我相关思维有关的脑区, 即默认网络后，该网络与心理理论、观点采择和道德决策等社会认知功能的联系得到了广泛证实[81~85]. 根据人际欺骗理论[15]，欺骗者和欺骗对象始终处于双向动态的社会认知过程中，而一项元分析同样发现了右侧TPJ在社会互动性质的欺骗任务中的活跃[75], 暗示了默认网络在欺骗活动中的重要作用. Bhatt等人[86]指出, 右侧TPJ与欺骗者使用策略误导对手有关, 且该脑区的激活水平受到欺骗的奖励幅度的调节. 因此, 有理由认为, 该网络支持了欺骗中的人际认知过程.

默认网络与中央执行网络的激活模式经常表现为负相关[82]，处于一种此消彼长的动态平衡之中. 因此, 有理由假设默认网络与中央执行网络存在功能上的互补[82], 即在外部刺激丰富的情况下, 中央执行网络支持的理性、外部指向的计算式认知占据优势地位；而外部刺激缺乏的情况下，默认网络支持的感性、内省的社会性认知成为主流.

3.3 情感系统的加工过程

欺骗具有风险性和不确定性，欺骗者将面对情感、道德和认知压力以及欺骗失败后的利益损失等潜在的不利后果, 进而产生负性的情绪体验. 因此, 欺骗者必须监控和评估可能预示负性结果的各种信息. 显著网络由杏仁核、AI和dACC等脑区组成[19,74]，与觉察、评价风险等负性刺激，以及产生恐惧、焦虑等负性情绪的过程相关, 在欺骗过程中不可或缺, 是欺骗的情绪情感组成部分，在欺骗的神经网络模型中作为情感系统，对认知系统产生的欺骗内容和策略进行加工(如图2箭头6所示).

显著网络觉察和识别欺骗的风险与成本等负性因

素, 并产生恐惧、焦虑和愧疚等情绪, 对动力系统进行调节, 从而改变欺骗动机强度(如图2箭头7所示). 因此, 情感系统作为反馈环路的重要组成部分, 是欺骗过程的调节者. 欺骗行为的实施很难不伴随着情绪情感过程, 原因可能在于情绪能够帮助减少不适宜的欺骗行为, 从而具有进化意义. 而另一方面, 情绪和情感也可能为欺骗者带来额外的认知负担, 减少认知资源[17], 从而影响认知系统的功能(如图2箭头8所示).

3.4 执行系统的执行过程

认知系统在内部心理空间完成了对欺骗策略和内容的计划、构建和保持, 执行系统的作用则是确保欺骗行为在外部情境中的部署和实施(如图2箭头4所示). 执行系统的功能同样需要中央执行网络的参与, 此处其作用主要表现为抑制控制功能[87], 例如通过自我控制压抑诚实的本能以防止真相泄漏, 以及抵御吐露实情以缓解负性情绪的冲动. 此外, 抑制控制的另一部分——认知控制可能也会发挥作用, 例如通过有意性地忽略或暂时遗忘真相, 以便作出更真实、自然的反应. 由于人所具有的内化的社会规范, 诚实反应一般是自发的本能式反应, 欺骗者必须保持对诚实反应的抑制, 表现为dlPFC等脑区激活水平上升[51]; 另外, 欺骗者需要始终在头脑中保持两种以上的平行的信念, 并且要维持欺骗内容的连贯和隐蔽, 因此执行过程将持续性地消耗认知资源, 尤其是执行资源. 如果认知资源不足, 执行系统不能正常运行, 欺骗行为将面临困难, 表现为反应时延长、错误率升高等[30,52]. 因此, 执行系统是欺骗过程的保证.

3.5 欺骗过程中的反馈机制

欺骗过程中, 欺骗者需要时刻收集和评估环境信息, 快速有效地调整欺骗策略, 例如修改和补充欺骗内容, 部分或全部地透露真相或者干预欺骗对象的心理状态等, 这些过程需要借助反馈机制得以实现. 本模型中包含两种反馈机制: (1) 认知系统对欺骗的推测和模拟激活情感系统, 后者进一步调节动力系统, 即动力-认知-情感反馈环路(如图2箭头2、6、7所示); (2) 来自欺骗结果和环境信息的反馈(如图2箭头5所示).

动力-认知-情感反馈环路在时间进程上处于欺骗的执行过程之前, 发生于欺骗者的内部心理空间. 动力系统被欺骗奖赏激活, 驱动认知系统对欺骗行为进行计划和推测, 而情感系统通过评估欺骗的风险认知、情感负担等信息, 对动力系统的激活水平进行调节, 从而避免不适宜、不理智的欺骗行为. 欺骗行为会降低被试对奖赏的评估, 并提高认知负担和损失预期, 说明自发欺骗具有内在成本[22]. 该结果支持了动力-认知-情感反馈环路的存在. 还有研究发现, ACC的活动在精神病态特征和欺骗行为的反应时间的关系中起中介作用[76]. Ofen等人[8]检验了已有的元分析结果, 观察到右侧AI在欺骗准备过程中的活动, 证实了情感系统在欺骗实施之前的参与. 此外, PCC、左侧颞叶区域等默认网络节点的激活水平与欺骗的认知资源消耗存在相关, 表明这些脑区与欺骗的能力和效率有关[8], 证明了认知系统的重要作用.

欺骗过程中反馈机制的另一部分是整体水平上的反馈, 即欺骗者在实施欺骗行为后收集和评估欺骗结果, 以指导和修正后续的欺骗行为. 这个过程并非特定于欺骗行为, 而是类似于基于模型的强化学习过程. 欺骗者基于欺骗结果修正自己的信念, 例如对他人心理状态的推断等, 并重新评估欺骗的潜在收益和风险, 进而修正欺骗内容和策略. 另外, 这种整体性反馈还体现在相关脑区对欺骗过程的激活水平降低. 一些脑结构, 包括杏仁核[67]和dlPFC[88], 在欺骗行为中的激活水平会随欺骗的重复进行而逐渐下降, 即呈现对欺骗行为的适应性倾向[67].

4 总结与展望

本文总结了近20年关于欺骗的神经机制的研究和理论, 并据此提出了欺骗的认知神经网络模型: 欺骗的神经网络结构由动力系统(奖赏网络)驱动, 认知系统(中央执行网络和默认网络)和执行系统(中央执行网络)分别负责欺骗的构建和执行, 并且具有借由情感系统(显著网络)和社会学习能力实现的反馈机制. 这些神经网络建立了有机、动态的联系, 保证了人能够在奖赏预期、计算构建、抑制控制、执行功能、人际认知和风险觉察等基础心理过程和能力的基础上, 实施复杂、困难、隐蔽的欺骗行为. 本文在认知神经网络的视角下重新审视前人的理论和研究, 使我们对欺骗行为的理解从孤立上升到网络层面, 并力图对欺骗过程作出更为全面系统的解释.

本文提出的欺骗认知神经网络模型在全面考虑各种心理功能和过程的基础上, 为欺骗行为的发生提供了更系统的全新解释, 但本领域尚存在若干有待解决的重要问题.

首先, 需要对不同类型的欺骗行为具体研究, 将研究问题继续细化. 例如, 近年来越来越多的研究者开始关注利己欺骗和利他欺骗的神经机制及其关系[25,27,41,69,89], 并发现利他欺骗与利己欺骗在空间维度[69,89]和时间维度[41]均具有不同的神经激活模式. 未来可以通过进一步规范和改进实验范式, 对欺骗行为进行分类研究, 针对性地研究不同类型欺骗的认知神经机制, 从而提升欺骗研究的可重复性和可比性.

其次, 有必要考察欺骗过程中脑活动的同步性和动态性. 欺骗是一种社会性行为, 只观测其中一方的脑活动难以全面把握其复杂的社会性特征, 因此今后需要引进更加新颖、有效的研究方法和技术, 如结合空间和时间维度的数据[20]或通过超扫描技术同时关注欺骗者、欺骗对象以及第三方的脑活动模式之间的同步性和交互性关系[90]. 此外, 欺骗是一个动态的过程, 其动态特征反映在两个方面: (1) 欺骗者会根据上一次欺骗的结果修正下一次欺骗的策略; (2) 单次欺骗过程中神经网络之间会进行交互, 表现为支持欺骗行为的各种心理过程. 未来的研究可以将具有高时间分辨率的ERP和具有高空间分辨率的fMRI技术结合起来以考察该动态过程.

最后, 本文提出的欺骗认知神经网络模型仍存在一些需要进一步完善或验证的问题: (1) 缺乏数学模型的构建, 尚停留在定性层面, 不能对欺骗行为作出定量解释和预测; (2) 虽然有证据表明该模型涉及的各种心理功能和过程参与了欺骗行为, 但是其发生的顺序和作用的形式是否遵循模型的预测还需要更多研究的检验; (3) 由于一个脑区往往参与多种心理功能, 因而心理功能和过程与相应脑区的对应难以做到清晰的划分. 因此, 未来仍需要继续优化和发展本模型, 例如补充相应的数学模型以模拟欺骗涉及的认知和情绪过程, 以及通过设计关注和区分不同心理功能和过程的研究, 进一步验证这些心理功能和过程对欺骗行为的支持.

参考文献

1 Abe N. How the brain shapes deception. Neuroscientist, 2011, 17: 560–574
2 Sai L, Lin X, Hu X, et al. Detecting concealed information using feedback related event-related brain potentials. Brain Cogn, 2014, 90: 142–150
3 Bogaard G, Meijer E H, Vrij A, et al. Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. PLoS One, 2016, 11: e0156615
4 Burgoon J K, Blair J P, Strom R E. Cognitive biases and nonverbal cue availability in detecting deception. Hum Commun Res, 2008, 34: 572–599
5 Ebesu A S, Miller M D. Verbal and nonverbal behaviors as afunction of deception type. J Lang Soc Psychol, 1994, 13: 418–442
6 Vrij A. Nonverbal dominance versus verbal accuracy in lie detection. Crim Justice Behav, 2008, 35: 1323–1336
7 Vrij A. Deception and truth detection when analyzing nonverbal and verbal cues. Appl Cogn Psychol, 2019, 33: 160–167
8 Ofen N, Whitfield-Gabrieli S, Chai X J, et al. Neural correlates of deception: Lying about past events and personal beliefs. Soc Cogn Affect Neurosci, 2017, 12: 116–127
9 Yin L, Weber B. Can beneficial ends justify lying? Neural responses to the passive reception of lies and truth-telling with beneficial and harmful monetary outcomes. Soc Cogn Affect Neurosci, 2016, 11: 423–432
10 Zuckerman M, Depaulo B M, Rosenthal R. Verbal and nonverbal communication of deception. Adv Exp Soc Psychol, 1981, 14: 1–59
11 Kozel F A, Padgett T M, George M S. A replication study of the neural correlates of deception. Behav Neurosci, 2004, 118: 852–856
12 Mazar N, Amir O, Ariely D. The dishonesty of honest people: A theory of self-concept maintenance. J Mark Res, 2008, 45: 633–644
13 Rick S, Loewenstein G. Hypermotivation. J Mark Res, 2008, 45: 645–648
14 Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Econometrica, 1979, 47: 263–291
15 Buller D B, Burgoon J K. Interpersonal deception theory. Commun Theor, 1996, 6: 203–242
16 Mohamed F B, Faro S H, Gordon N J, et al. Brain mapping of deception and truth telling about an ecologically valid situation: Functional MR imaging and polygraph investigation—Initial experience. Radiology, 2006, 238: 679–688
17 Walczyk J J, Harris L L, Duck T K, et al. A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. New Ideas Psychol, 2014, 34: 22–36
18 Diekhof E K, Kaps L, Falkai P, et al. The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude—An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. Neuropsychologia, 2012, 50: 1252–1266
19 Menon V. Salience network. Hum Brain Mapp, 2015, 2: 597–611
20 Sun D, Lee T M C, Wang Z, et al. Unfolding the spatial and temporal neural processing of making dishonest choices. PLoS One, 2016, 11:

e0153660
21 Sun D, Chan C C H, Hu Y, et al. Neural correlates of outcome processing post dishonest choice: An fMRI and ERP study. Neuropsychologia, 2015, 68: 148–157
22 Zhu C, Pan J, Li S, et al. Internal cost of spontaneous deception revealed by ERPs and EEG spectral perturbations. Sci Rep, 2019, 9: 5402–5413
23 Proudfit G H. The reward positivity: From basic research on reward to a biomarker for depression. Psychophysiology, 2015, 52: 449–459
24 Ding X P, Du X, Lei D, et al. The neural correlates of identity faking and concealment: An fMRI study. PLoS One, 2012, 7: e48639
25 Abe N, Greene J D. Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. J Neurosci, 2014, 34: 10564–10572
26 Speer S P H, Smidts A, Boksem M A S. Cognitive control increases honesty in cheaters but cheating in those who are honest. Proc Natl Acad Sci USA, 2020, 117: 19080–19091
27 Pornpattananangkul N, Zhen S, Yu R. Common and distinct neural correlates of self-serving and prosocial dishonesty. Hum Brain Mapp, 2018, 39: 3086–3103
28 Leng H, Wang Y, Li Q, et al. Sophisticated deception in junior middle school students: An ERP study. Front Psychol, 2019, 9: 2567
29 Zheltyakova M, Kireev M, Korotkov A, et al. Neural mechanisms of deception in a social context: An fMRI replication study. Sci Rep, 2020, 10: 10713
30 Vrij A, Leal S, Mann S, et al. Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. Psychol Crime Law, 2012, 18: 579–594
31 Sánchez N, Masip J, Gómez-Ariza C J. Both high cognitive load and transcranial direct current stimulation over the right inferior frontal cortex make truth and lie responses more similar. Front Psychol, 2020, 11: 776
32 Christ S E, Van Essen D C, Watson J M, et al. The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. Cereb Cortex, 2009, 19: 1557–1566
33 Yin L, Reuter M, Weber B. Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling. Brain Cogn, 2016, 102: 13–25
34 Zhang M, Liu T, Pelowski M, et al. Gender difference in spontaneous deception: A hyperscanning study using functional near-infrared spectroscopy. Sci Rep, 2017, 7: 7508
35 Zhu L, Jenkins A C, Set E, et al. Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. Nat Neurosci, 2014, 17: 1319–1321
36 van Veen V, Carter C S. The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiol Behav, 2002, 77: 477–482
37 MacDonald A W, Cohen J D, Stenger V A, et al. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science, 2000, 288: 1835–1838
38 Levy B J, Wagner A D. Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. Ann NY Acad Sci, 2011, 1224: 40–62
39 Xu K Z, Anderson B A, Emeric E E, et al. Neural basis of cognitive control over movement inhibition: Human fMRI and primate electrophysiology evidence. Neuron, 2017, 96: 1447–1458.e6
40 Zhang D D, Liu Z L, Chen Y, et al. The role of right ventrolateral prefrontal cortex on social emotional regulation in subclinical depression: A tDCS study (in Chinese). Acta Psychol Sin, 2019, 51: 207–215 [张丹丹, 刘珍莉, 陈钰, 等. 右腹外侧前额叶对高抑郁水平成年人社会情绪调节的作用: 一项tDCS研究. 心理学报, 2019, 51: 207–215]
41 Cui F, Wu S, Wu H, et al. Altruistic and self-serving goals modulate behavioral and neural responses in deception. Soc Cogn Affect Neurosci, 2018, 13: 63–71
42 Rosenfeld J P, Ozsan I, Ward A C. P300 amplitude at Pz and N200/N300 latency at F3 differ between participants simulating suspect versus witness roles in a mock crime. Psychophysiology, 2017, 54: 640–648
43 Sai L, Wu H, Hu X, et al. Telling a truth to deceive: Examining executive control and reward-related processes underlying interpersonal deception. Brain Cogn, 2018, 125: 149–156
44 Suchotzki K, Crombez G, Smulders F T Y, et al. The cognitive mechanisms underlying deception: An event-related potential study. Int J Psychophysiol, 2015, 95: 395–405
45 Wu H, Hu X, Fu G. Does willingness affect the N2-P3 effect of deceptive and honest responses? Neurosci Lett, 2009, 467: 63–66
46 Folstein J R, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology, 2008, 45: 152–170
47 Johnson Jr R, Henkell H, Simon E, et al. The self in conflict: The role of executive processes during truthful and deceptive responses about attitudes. NeuroImage, 2008, 39: 469–482
48 Fu H, Qiu W, Ma H, et al. Neurocognitive mechanisms underlying deceptive hazard evaluation: An event-related potentials investigation. PLoS One, 2017, 12: e0182892

49 Hu X, Pornpattananangkul N, Nusslock R. Executive control- and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. Cogn Affect Behav Neurosci, 2015, 15: 475–491

50 Greene J D, Paxton J M. Patterns of neural activity associated with honest and dishonest moral decisions. Proc Natl Acad Sci USA, 2009, 106: 12506–12511

51 Sun P, Ling X, Zheng L, et al. Modulation of financial deprivation on deception and its neural correlates. Exp Brain Res, 2017, 235: 3271–3277

52 Debey E, Verschuere B, Crombez G. Lying and executive control: An experimental investigation using ego depletion and goal neglect. Acta Psychol, 2012, 140: 133–141

53 Maréchal M A, Cohn A, Ugazio G, et al. Increasing honesty in humans with noninvasive brain stimulation. Proc Natl Acad Sci USA, 2017, 114: 4360–4364

54 Xu Z X, Ma H K. Does honesty result from moral will or moral grace? Why moral identity matters. J Bus Ethics, 2015, 127: 371–384

55 Sai L, Shang S, Tay C, et al. Theory of mind, executive function, and lying in children: A meta-analysis. Dev Sci, 2021, 24: e13096

56 Amodio D M, Frith C D. Meeting of minds: The medial frontal cortex and social cognition. Nat Rev Neurosci, 2006, 7: 268–277

57 Sai L, Zhao C, Heyman G D, et al. Young children's lying and early mental state understanding. Infant Child Dev, 2020, 29: e2197

58 Tang H, Lu X, Cui Z, et al. Resting-state functional connectivity and deception: Exploring individualized deceptive propensity by machine learning. Neuroscience, 2018, 395: 101–112

59 Suzuki S, Harasawa N, Ueno K, et al. Learning to simulate others' decisions. Neuron, 2012, 74: 1125–1137

60 Greicius M D, Krasnow B, Reiss A L, et al. Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proc Natl Acad Sci USA, 2003, 100: 253–258

61 Igelström K M, Graziano M S A. The inferior parietal lobule and temporoparietal junction: A network perspective. Neuropsychologia, 2017, 105: 70–83

62 Wagner A D, Shannon B J, Kahn I, et al. Parietal lobe contributions to episodic memory retrieval. Trends Cogn Sci, 2005, 9: 445–453

63 Garrigan B, Adlam A L R, Langdon P E. The neural correlates of moral decision-making: A systematic review and meta-analysis of moral evaluations and response decision judgements. Brain Cogn, 2016, 108: 88–97

64 Cavanna A E, Trimble M R. The precuneus: A review of its functional anatomy and behavioural correlates. Brain, 2006, 129: 564–583

65 Volz K G, Vogeley K, Tittgemeyer M, et al. The neural basis of deception in strategic interactions. Front Behav Neurosci, 2015, 9: 27

66 Baumgartner T, Fischbacher U, Feierabend A, et al. The neural circuitry of a broken promise. Neuron, 2009, 64: 756–770

67 Garrett N, Lazzaro S C, Ariely D, et al. The brain adapts to dishonesty. Nat Neurosci, 2016, 19: 1727–1732

68 Langleben D D, Loughead J W, Bilker W B, et al. Telling truth from lie in individual subjects with fast event-related fMRI. Hum Brain Mapp, 2005, 26: 262–272

69 Yin L, Hu Y, Dynowski D, et al. The good lies: Altruistic goals modulate processing of deception in the anterior insula. Hum Brain Mapp, 2017, 38: 3675–3690

70 Janak P H, Tye K M. From circuits to behaviour in the amygdala. Nature, 2015, 517: 284–292

71 Namkung H, Kim S H, Sawa A. The insula: An underestimated brain area in clinical neuroscience, psychiatry, and neurology. Trends Neuroscis, 2018, 41: 551–554

72 Sanfey A G, Rilling J K, Aronson J A, et al. The neural basis of economic decision-making in the ultimatum game. Science, 2003, 300: 1755–1758

73 Vogt B A. Pain and emotion interactions in subregions of the cingulate gyrus. Nat Rev Neurosci, 2005, 6: 533–544

74 Bressler S L, Menon V. Large-scale brain networks in cognition: Emerging methods and principles. Trends Cogn Sci, 2010, 14: 277–290

75 Lisofsky N, Kazzer P, Heekeren H R, et al. Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. Neuropsychologia, 2014, 61: 113–122

76 Abe N, Greene J D, Kiehl K A. Reduced engagement of the anterior cingulate cortex in the dishonest decision-making of incarcerated psychopaths. Soc Cogn Affect Neurosci, 2018, 13: 797–807

77 Diamond A. Executive functions. Annu Rev Psychol, 2012, 64: 135–168

78 Miyake A, Friedman N P, Emerson M J, et al. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. Cogn Psychol, 2000, 41: 49–100

79 Baddeley A D, Hitch G J. Developments in the concept of working memory. Neuropsychology, 1994, 8: 485–493

80 Sai L, Bellucci G, Wang C, et al. Neural mechanisms of deliberate dishonesty: Dissociating deliberation from other control processes during dishonest behaviors. Proc Natl Acad Sci USA, 2021, 118: e2109208118

81 Raichle M E, MacLeod A M, Snyder A Z, et al. A default mode of brain function. Proc Natl Acad Sci USA, 2001, 98: 676–682

82 Buckner R L, Andrews-Hanna J R, Schacter D L. The brain's default network—Anatomy, function, and relevance to disease. Ann NY Acad Sci, 2008, 1124: 1–38

83 Schilbach L, Bzdok D, Timmermans B, et al. Introspective minds: Using ALE meta-analyses to study commonalities in the neural correlates of

emotional processing, social & unconstrained cognition. PLoS One, 2012, 7: e30920
84 Amft M, Bzdok D, Laird A R, et al. Definition and characterization of an extended social-affective default network. Brain Struct Funct, 2015, 220: 1031–1049
85 Andrews-Hanna J R, Reidler J S, Sepulcre J, et al. Functional-anatomic fractionation of the brain's default network. Neuron, 2010, 65: 550–562
86 Bhatt M A, Lohrenz T, Camerer C F, et al. Neural signatures of strategic types in a two-person bargaining game. Proc Natl Acad Sci USA, 2010, 107: 19720–19725
87 Wessel J R, Aron A R. On the globality of motor suppression: Unexpected events and their influence on behavior and cognition. Neuron, 2017, 93: 259–280
88 Tang H, Zhang S, Jin T, et al. Brain activation and adaptation of deception processing during dyadic face-to-face interaction. Cortex, 2019, 120: 326–339
89 Hayashi A, Abe N, Fujii T, et al. Dissociable neural systems for moral judgment of anti- and pro-social lying. Brain Res, 2014, 1556: 46–56
90 Wang M Y, Luan P, Zhang J, et al. Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: A mini-review. Quant Imag Med Surg, 2018, 8: 819–837

Summary for “欺骗的认知神经网络模型”

The cognitive neural network model of deception

Yingliang Zhang & Xiaoqin Mai*

Department of Psychology, Renmin University of China, Beijing 100872, China
* Corresponding author, E-mail: maixq@ruc.edu.cn

Deception is a psychological process by which an individual deliberately attempts to convince others to accept as true what the liar knows to be false, in order to gain benefits or avoid losses for oneself or others. Deception has many forms and is widespread in all aspects of human social life. At the social level, deception affects all aspects of interpersonal interactions and thus has an impact on the overall social climate. It is equally important for individuals and organizations to practice and recognize deception. A growing body of studies provides enlightenment on the neural mechanisms of deception, but the role of large-scale brain networks remains under-discussed. This paper reviews the theories and research related to deception, as well as the cognitive and emotional processing involved in deceptive behavior. The brain regions and event-related potential components associated with deception are summarized. The ventromedial prefrontal cortex, striatum, and reward positivity are related to reward evaluation. The dorsolateral prefrontal cortex, ventrolateral prefrontal cortex, and parietal P3 are related to executive functions consisting of inhibition and interference control, working memory, and cognitive flexibility. The dorsomedial prefrontal cortex and temporoparietal junction are related to social cognitive processing such as the theory of mind. The amygdala, anterior insula, dorsal anterior cingulate cortex, and frontal N2 are related to producing negative emotional experiences. Based on previous studies, we propose a cognitive neural network model of deception, in which deception arises from the interactive cooperative process among the dynamic, emotional, cognitive, and executive systems, and the activation of the reward network, salience network, central executive network, and default network is the neural basis behind it. Specifically, the dynamic system is acted upon by the deception reward, which provides motivation for the whole process and activates the cognitive system; the cognitive system makes deceptive decisions, plans, constructs, and maintains the content; the deception strategy and processing activate the emotional system, which causes the deceiver to have negative emotional experiences such as guilt and fear, thus causing the dynamic system to reduce the motivation level; and the executive system delivers the deception and ensures that the deceptive content is not confused or leaked. The result of the deception is then perceived and processed again by the deceiver. In addition, the model’s motivation-cognition-emotion loop, and the perceptual/memory information-motivation-cognition-execution-outcome loop that runs through the entire deception process, provide feedback functions that more dynamically and accurately model deceptive behavior. The model establishes connections between theories, involved brain regions, mental processes, and neural networks of deception, and is expected to provide a more holistic and systematic explanation of the neural mechanisms of deception. Finally, we summarize the shortcomings of the model and ways to improve it, and put forward some suggestions for the development of the field, such as a focus on separating the mental processes of deception by improving the experimental design, and combining multiple techniques to detect the neural activation patterns of deception.

deception, deception theory, functional magnetic resonance imaging, event-related potential, cognitive neural network model

doi: 10.1360/TB-2021-0963