

• 研究构想(Conceptual Framework) •

贝叶斯框架下社交焦虑的社会认知特性*

彭玉佳^{1,2,3} 王愉茜¹ 鞠芊芊¹ 刘峰⁴ 徐佳⁵

(¹北京大学心理与认知科学学院〔行为与心理健康北京市重点实验室, 生物与机器智能教育部重点实验室〕; ²北京大学人工智能研究院, 北京 100871)
(³跨媒体通用人工智能全国重点实验室, 北京通用人工智能研究院, 北京 100080)
(⁴上海交通大学心理学院 上海 200030) (⁵北京大学第六医院, 北京大学精神卫生研究所, 国家卫生健康委员会精神卫生学重点实验室〔北京大学〕, 国家精神心理疾病临床医学研究中心〔北京大学第六医院〕, 北京 100191)

摘要 社交焦虑是焦虑障碍中一个重要的组成部分, 与异常的社会认知密不可分。以往研究发现了社交焦虑群体中社会认知的特异性, 如倾向于对情绪和意图进行负面的加工和消极的解释。然而, 社会认知特异性现象背后的机制尚不明确, 难以揭示社交焦虑症状的底层成因并指导个性化干预。本研究旨在基于认知计算建模, 回答社交焦虑社会认知特异性背后的认知神经机制。在贝叶斯框架下, 本研究提出假设, 社交焦虑的负向认知可能源于异常的先验预期。研究综合运用行为实验、脑电图(EEG)和功能磁共振成像(fMRI), 构建社交焦虑社会认知先验预期的认知神经计算模型, 刻画先验预期的动态变化规律, 并基于解码神经反馈对神经机制进行因果验证。本研究有望揭示社会认知特异性的成因和动态发展规律, 在行为和大脑层面计算建模个体差异, 推动以社交焦虑为代表的精神疾病的个性化干预。

关键词 社会认知, 社交焦虑, 认知计算, 认知神经机制, 脑成像

分类号 B842; R395

1 问题提出

焦虑和抑郁障碍是当今最普遍的精神疾病, 由于其共病性和异质性的双重挑战, 很多症状背后的机制还缺乏明确的答案(Peng et al., 2021, 2023; Xu et al., 2024)。社交焦虑是焦虑障碍中一个重要的组成部分, 具体体现为对于社会情境的极度害怕与回避, 严重影响患者的生活和工作。社交焦虑与其他焦虑和抑郁障碍具有高度的共病性, 可作为解析焦虑和抑郁障碍共病性和异质性的有效切入点。社交焦虑与异常的社会认知密不可分, 如社交焦虑个体倾向于对社会信息进行负向加工和消极解释。然而, 社交焦虑背后社会认知的计算神经机制和发生过程尚不明确,

难以揭示社交焦虑症状的底层机制或指导临床干预。具体而言, 负向认知偏向现象的发生, 是来源于特异的自上而下的先验预期模型(prior, 如有偏见误差的社会预期), 还是自下而上的异常社会信号感知觉加工(likelihood, 如对于社会信号加工效率低或发生扭曲), 目前仍不清楚。

本研究将基于贝叶斯框架回答上述问题, 系统探究高社交焦虑特质个体社会认知特异性的机制, 尝试剥离先验预期和感知觉加工, 回答先验预期在社会认知过程中的作用以及动态形成过程。本研究将综合运用行为实验、功能磁共振成像(functional magnetic resonance imaging, fMRI)和脑电图(electroencephalography, EEG)等方法, 基于认知计算建模和机器学习, 系统探究社交焦虑社会认知的特异性机制。研究将构建高社交焦虑特质个体的社会认知先验预期的计算模型, 从而检验异常的社会认知可能源于负向先验预期的

收稿日期: 2025-02-28

* 国家自然科学基金面上基金项目(32471151)。

通信作者: 彭玉佳, E-mail: yujia_peng@pku.edu.cn

假设。本研究有望基于负向社会认知这一现象,深入解析社交焦虑社会认知特异性的认知和神经层面的病理学机制,并推动以社交焦虑为代表的精神疾病的个性化干预。

2 国内外研究现状述评

2.1 社交焦虑的社会认知特异性

社交焦虑与异常的社会认知密不可分。在情绪识别方面,社交焦虑个体存在负向倾向性,如对负面情绪识别更快(Heuer et al., 2010; Maoz et al., 2016)。当社会环境模糊或不确定时,社交焦虑个体无法有效学习社会环境中的负面信息,倾向于对消极的社会情境持续投入(Lamba et al., 2020)。同时,社交焦虑个体缺乏认知灵活性,无法及时调整对外部环境变化的认知(Beltzer et al., 2019; Xia et al., 2023),尤其体现为对负向社会信息的认知固化(Beltzer et al., 2019)。所以,负向认知是社交焦虑社会认知特异性的重要特征。

具体而言,社交焦虑个体的负向认知可以分为负向注意偏向和负向解释偏向。首先,负向注意偏向指社交焦虑个体对威胁性社会信息的不均衡的注意资源分配,如研究发现高社交焦虑个体相比于非社交焦虑个体更倾向于将注意力分配给威胁面孔(Bantin et al., 2016)。其次,负向解释偏向指社交焦虑个体倾向于对社会信息进行偏负向的解释(Alden et al., 2008; Laposa et al., 2010),如社交焦虑个体更多地将中性或低强度的积极情绪解释为消极情绪(Heuer et al., 2010)。

2.2 社交焦虑负向认知偏向的认知行为和计算模型

对于上述社交焦虑负向认知偏向存在经典的认知行为模型。Clark 和 Wells (1995)认为,社会情境会引起社交焦虑个体负面的假设或信念(如,“如果我犯错,其他人会羞辱我”),使得社交焦虑个体错误评估社会情境具有威胁性。相似的,Rapee 和 Heimberg (1997)认为社交焦虑个体对他人存在负面的心理表征:认为他人倾向于对自己做出负面评价,使得社交焦虑个体将更多的注意资源分配到威胁性刺激,进一步强化社交焦虑个体负向心理表征,形成恶性循环。两个认知行为模型均提示了一种假设,先验预期的偏差可能在社交焦虑的社会认知特异性的背后起到关键作用。

认知计算模型有助于验证上述假设。比如,

漂移扩散模型(Drift Diffusion Model, DDM)假设社交焦虑个体在判断过程中逐渐积累负向或正向相关信息,最终达到决策阈值并做出判断(Ratcliff et al., 2016)。模型基于被试选项和反应时数据,评估社交焦虑个体对负向或正向信息的初始偏向(z)、证据积累速度(v)、整体反应谨慎程度(冲动/谨慎) (a)和与决策无关的反应时间(t_0)这4个认知特征(Chen et al., 2020; White et al., 2010)。然而,以往研究证据并不统一。比如, Dietel 等人(2021)发现高社交焦虑个体具有更快的负向信息积累速度;而 White 等人(White et al., 2016)使用DDM模型则发现该认知偏向主要受高焦虑者的初始偏向和信息累积速度的双重驱动,即高焦虑个体在判断时存在初始偏向,认为刺激更可能存在威胁,并且其总体信息积累速度较慢。

另一方面,强化学习模型(Reinforcement Learning, RL)通过分析社交焦虑个体的社会信息学习异常,揭示了社交焦虑负向认知的形成过程。模型通过学习率(learning rate)和任务表现评估个体社会信息学习过程:学习率反映个体根据社会反馈更新行为的快慢;任务表现体现为任务结束后,个体完成目标的程度。前人研究发现,社交焦虑个体在掷球游戏(cyberball game)中,难以更新对威胁情况的判断,且对威胁情况持续保持高学习率,表明高社交焦虑个体可能难以更新已学习的负面社会信息(Beltzer et al., 2019)。同时,在社会概率学习任务中,一对中性面孔分别对应着高概率社会奖赏或社会惩罚,高社交焦虑个体虽然可以很准确地学习社会奖赏,但无法准确地回避社会惩罚(Beltzer et al., 2023)。综上所述,社交焦虑个体在社会信息学习过程中,保持对社会威胁的高警觉,且无法有效回避社会威胁。

2.3 社交焦虑社会认知特异性的脑机制

前人研究在脑成像实验中验证了社交焦虑个体的社会认知特异性。在情绪加工任务中,社交焦虑与威胁恐惧情绪环路的异常活动有关,主要涉及杏仁核、脑岛、海马旁回和内侧前额叶皮质等边缘系统脑区(Freitas-Ferrari et al., 2010; Hattingh et al., 2013),与焦虑障碍的恐惧加工环路密切相关(Peng et al., 2021, 2023)。在社会情境任务中,社交焦虑个体同样在属于默认网络的楔前叶和后扣带回出现异常的激活(Nakao et al., 2011)。然而,上述证据还无法有效剥离自上而下的先验预期和

自下而上的感知觉加工对于社交焦虑社会认知特异性的影响。

目前,已有间接证据指出先验预期在社会认知特异性中的作用(彭玉佳 等, 2023; Gu et al., 2020; Xia et al., 2023)。首先,以往的 EEG 和 fMRI 研究已构建出刺激出现前的大脑基线水平和刺激出现后判断的联系,验证了先验预期可能影响认知判断的神经证据。具体而言,刺激前的 α 频段振荡特征可预测低级的意识过程和高级的决策过程,如顶枕叶的 α 频段振荡可预测对于自身注意状态的内省判断(Macdonald et al., 2011),以及全脑 α 频段振荡可预测被试单次实验的表现(Busch et al., 2009)。同时, fMRI 研究同样发现感知觉脑区的刺激前激活水平可预测随后的认知过程(Boly et al., 2007; Hesselmann, Kell, Eger, & Kleinschmidt, 2008; Hesselmann, Kell, & Kleinschmidt, 2008)。最后,基于额叶的反馈负波(feedback-related negativity, FRN)的 EEG 研究也为社交焦虑先验预期的异常提供了证据。Gu 等人的研究发现社交焦虑个体在面对意外的社会接纳时,其 FRN 信号强度强于收到意外社会拒绝反馈的条件。结果表明社交焦虑个体可能存在对社会反馈的负向预期偏差(Cao et al., 2015; Gu et al., 2020),因而对意外的社会接纳表现出更大的预测误差。结合上述研究,刺激前大脑基线水平和 FRN 信号的研究结果已为社交焦虑异常的先验预期的神经表征提供支持,但目前还缺乏直接的研究证据。综上,推动在贝叶斯框架下解析社交焦虑先验预期的脑机制,将有助于理解社交焦虑社会认知特异性的机制和形成过程。

3 简要述评

综上所述,以往社交焦虑相关研究虽然发现了以负向认知为代表的社会认知特异性,但存在以下局限性。

首先,以往关于社交焦虑负向认知偏向的研究结果不一致。比如,一方面,有研究发现高社交焦虑个体相比于非社交焦虑个体更倾向于将注意力分配给威胁面孔,反映了社交焦虑对威胁刺激的注意警觉(Bantin et al., 2016)。而 Andersson 等人(2006)使用情绪 Stroop 任务发现高社交焦虑个体存在对威胁性刺激的回避。另一方面,Heuer 等人(2010)发现社交焦虑个体对在情绪强度低的面孔存在负向解释偏向,而 Jusyte 和 Schönenberg

(2014)未发现社交焦虑个体对模糊面孔存在解释偏向。此外,注意偏向与解释偏向对应认知过程的不同阶段,在时间进程上不连续,因而难以直接整合行为结果。具体而言,注意偏向主要关注刺激呈现 500 毫秒之内的认知过程(蒋婧琪 等, 2019),而解释偏向则反映了社交焦虑个体对社会刺激的后期加工,决定着对社交焦虑个体对社会信息所赋予的意义(李涛, 冯菲, 2013),通常不设置时间限定,直到被试做出反应(如: Maoz et al., 2016)。为了整合注意偏向和解释偏向,揭示社交焦虑个体负向认知偏向的共性机制,本研究从两者的共同本质——信息加工过程入手,采用计算建模进行解构和分析。本研究提出,不同实验背后的差异可能是由于未剥离先验预期与感知觉的双成分的交互作用,贝叶斯模型有望给出统一解释。贝叶斯模型描述从经验知识到推断未知的过程,包含三个概念:先验预期(个体基于过往经历形成的信念模型)、证据加工(也称似然性),和后验信念(posterior, 个体整合新信息后更新的信念模型或推断)。理解社交焦虑个体的负向认知偏向,需要考察对于社交情境存在负向先验预期(negative prior)。前人研究中社交焦虑个体的负向认知偏向更多在模糊实验刺激中发现(李涛, 冯菲, 2013; Gu et al., 2010)。与此相一致,根据贝叶斯模型,模糊实验刺激没有明确的社会效价指向,所以证据加工的精度低且对于后验分布的贡献权重小,进而负向先验预期所产生的对于最终决策的贡献才得以体现(图 1)。

第二,前人虽然在行为层面发现了社交焦虑社会认知特异性,但仍缺乏认知计算模型来解释其发生和形成机制。比如,社交焦虑个体在识别情绪时的负面偏向是由负向先验导致,还是由信息加工偏向导致,以及这种负向认知是如何形成的,均还存在领域空白。贝叶斯框架中的贝叶斯更新模型(Bayesian Updating)和层级高斯滤波器(Hierarchical Gaussian Filter)均可模拟社交焦虑个体对社会情境的负向先验预期的更新过程(Lawson et al., 2017; Mathys et al., 2011, 2014; Palmer et al., 2017; Powers et al., 2017),从而解析负向先验预期的发展变化规律。理解焦虑形成背后的动态认知及脑机制,将有助于有针对性地筛查和预测症状的发生和发展。

第三,围绕社交焦虑的神经机制,前人研究

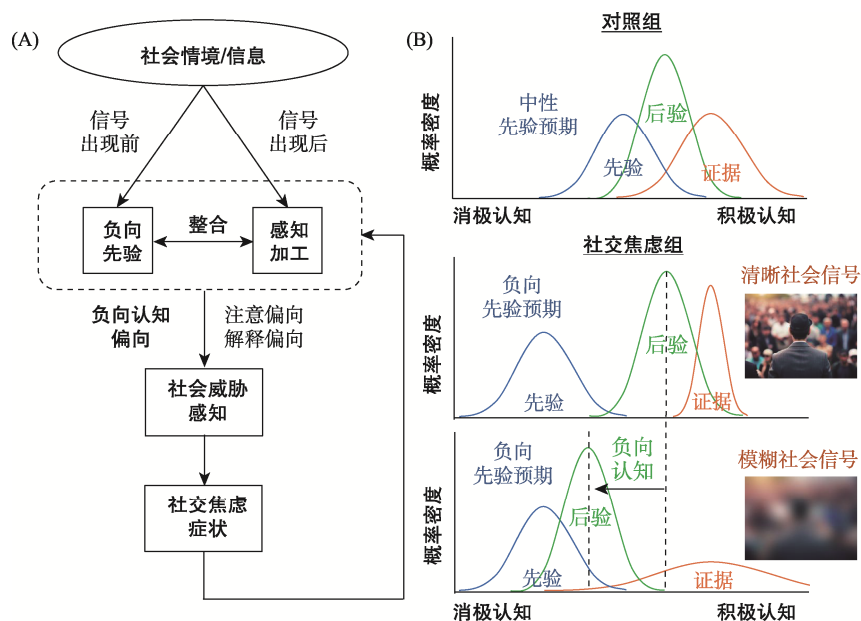


图1 (A)基于贝叶斯框架的社交焦虑认知行为模型。(B)基于贝叶斯框架理解社交焦虑个体的负向认知偏向。对于社交焦虑群体,先验分布相较于对照组呈现负向偏向。当社会信号输入清晰时,后验分布由似然分布主导,没有体现出与对照组的差异。然而,当证据较为模糊时,似然分布对于后验分布的影响减弱,负向先验分布主导社会认知,导致负向认知偏向。

大多关注情绪任务,发现社交焦虑与威胁恐惧情绪环路的异常活动有关(Etkin & Wager, 2007; Freitas-Ferrari et al., 2010; Hattingh et al., 2013; Peng et al., 2021, 2023; Stein & Stein, 2008)。然而,这种异常的情绪加工神经环路是如何形成,以及如何在神经层面拆分开先验预期和感知觉加工,仍有待探究。首先,大脑基线活动水平和FRN信号可能表征贝叶斯框架中的先验预期(Boly et al., 2007; Busch et al., 2009; Hesselmann et al., 2010; Hesselmann, Kell, Eger, & Kleinschmidt, 2008; Hesselmann, Kell, & Kleinschmidt, 2008; Lou et al., 2014),因此有望从大脑基线活动中分离社交焦虑个体对社会情境的先验预期。其次,基于EEG和fMRI数据建模可建立大脑神经活动与社交焦虑个体负向先验预期特点的映射关系,从而探究负向先验预期发展变化特征对应的大脑表征(Powers et al., 2017)。

基于上述挑战和科研需求,本研究计划基于贝叶斯框架,探究社交焦虑背后的先验预期模型,整合前人的认知行为模型(Clark & Wells, 1995; Rapee & Heimberg, 1997)和计算模型的结果(Beltzer et al., 2019; Dietel et al., 2021),为理解社

交焦虑负向认知偏向的机制和形成过程提供一个统一的理论框架。

4 研究构想

本研究总体框架见图2。本研究旨在基于认知计算建模,回答社交焦虑社会认知特异性背后的认知神经机制。在贝叶斯框架下,我们提出假设,社交焦虑的负向认知可能源于异常的先验预期。研究将基于4个研究综合运用行为实验、脑电图(EEG)和功能磁共振成像(fMRI),构建社交焦虑社会认知先验预期的认知神经计算模型,刻画先验预期的动态变化规律,并基于解码神经反馈对神经机制进行因果验证。研究旨在揭示社会认知特异性的成因和发展规律,在行为和大脑层面构建个体差异计算模型,推动以社交焦虑为代表的精神疾病的个性化干预。

4.1 研究内容1: 基于贝叶斯框架探究社交焦虑的社会场景及情绪先验预期

研究1旨在探究高社交焦虑特质个体对社会情境和情绪的负向先验预期。基于贝叶斯框架,研究使用模糊刺激情绪判断任务以及3D面孔重建手段,量化测量高社交焦虑特质背后的社会

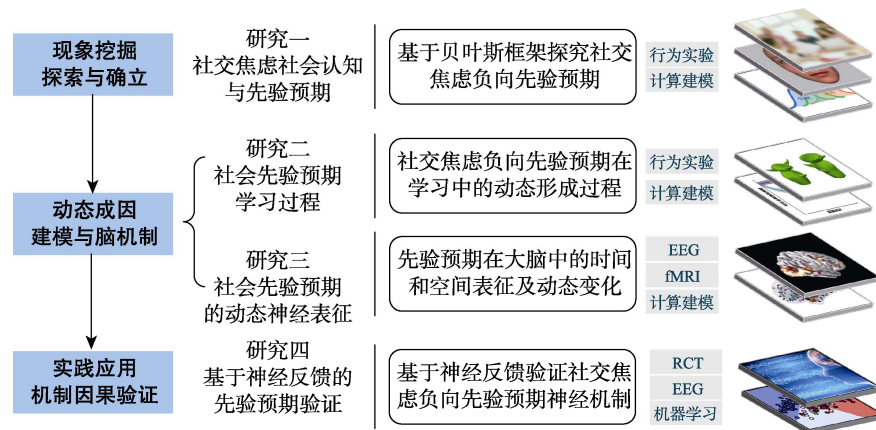


图 2 研究研究框架及技术方法。研究计划基于 4 个研究, 系统检验社交焦虑社会认知特异性背后的认知神经机制。

情境和面孔情绪判断, 构建社交焦虑个体的社会情景和面孔情绪先验预期模型。研究假设为, 相比于对照组, 高社交焦虑特质个体会表现出偏向于负向情绪的特异性先验模型, 体现为负向模糊场景预期以及更强的愤怒面孔表征。

4.2 研究内容 2: 贝叶斯模型解析社交焦虑先验预期的动态学习过程

在研究 1 的基础上, 研究 2 将进一步探究社交焦虑特质个体先验预期的动态形成过程。研究基于社会信息学习任务, 主要关注高社交焦虑特质个体对于积极和消极社会刺激的学习速率, 以及当社会效价发生转换时, 社交焦虑特质个体学习的认知灵活性。研究假设为, 相比于健康个体, 高社交焦虑特质个体更容易整合负向信息, 在社会信息学习过程中持续保持对社会威胁的高警觉(Beltzer et al., 2019), 具有更高的负向社会信息学习率, 更快构建负向先验模型, 并在社会效价发生转换时更难改变原先形成的负向先验模型, 体现负向社会认知固化的特点。

4.3 研究内容 3: 贝叶斯模型解析社交焦虑先验预期的脑机制

基于研究 2 的行为计算建模, 研究 3 将探究社交焦虑特质个体的先验预期学习率和固化特点对应的大脑表征。EEG 具有高时间分辨率, 而 fMRI 拥有高空间分辨率, 因此, 本研究将分别基于两者探究社交焦虑先验预期脑机制的时空特征。基于 EEG, 研究将采用时频分析(time-frequency analysis), 探究时间上的学习过程的大脑活动特点。进一步, 研究将基于 fMRI 进一步探究, 采用

一般线性模型和多体素模式分析(multi-voxel pattern analysis, MVPA), 探究大脑表征的空间特征。此外, 为增强 EEG 和 fMRI 结果的相互印证, 本研究还将在 EEG 数据上进行溯源分析(Keynan et al., 2019)。

4.4 研究内容 4: 基于神经反馈验证社交焦虑负性先验预期的神经机制

研究 4 旨在基于随机对照试验(Randomized Control Trials, RCT), 因果验证社交焦虑负向先验预期脑机制。研究将基于 EEG 探究社交焦虑负向先验预期的认知神经机制, 构建神经解码模型。神经解码模型可实现线上脑电解码神经反馈(real-time EEG decoded neurofeedback, rtEEG-DecNef), 实验中将实时监测被试的脑活动, 为被试及时提供可视化反馈, 进而揭示负向先验预期脑机制影响社交焦虑的神经过程。研究将建立神经反馈预测负向先验预期的解码器, 实证检验社交焦虑负向先验预期的脑机制, 同时探索基于脑的个性化干预的可能性。

5 理论构建与创新

社交焦虑背后社会认知特异性的计算神经机制和发生过程尚不明确, 难以揭示社交焦虑症状的底层机制或指导临床干预。本研究旨在基于贝叶斯框架, 从理论探索、实验范式、技术手段和数据分析等层面进行创新, 进一步回答精神疾病背后的机制。本研究将综合运用行为实验、脑成像、临床量表、计算建模和机器学习, 系统考察社交焦虑个体对于社会信息的先验预期和信号加工过程, 致力于从基础的认知过程, 理解社交焦

虑个体对于社会信号产生恐惧和回避的原因,以及可能有效的干预方法。

(1) 基于贝叶斯框架,提出社交焦虑社会认知特异性的统一理论框架

基于前人和前期实验证据,本研究提出假设,社交焦虑的负向认知可能源于异常的先验预期。本研究将经典的贝叶斯理论框架引入社交焦虑领域,突破传统现象学的研究范式,为挖掘和构建社交焦虑背后社会认知特异性的计算机制提供了可能性。研究从社会图片、三维面孔等多个维度出发,通过计算建模还原社交焦虑个体的先验预期模型。

(2) 实证性检验社交焦虑社会认知先验模型形成过程和神经机制

研究将结合行为实验、脑成像和计算建模,全面探索社交焦虑的社会认知先验预期特点以及动态形成过程,为社交焦虑认知理论提供新的计算框架。研究计划采用社会奖赏学习任务完成,基于拟人且剥离先验知识的抽象社会对象Greebles作为刺激,探索不同社交焦虑特质个体的学习过程。基于学习过程,研究将回答社交焦虑背后的特异性社会认知可能的形成过程。

(3) 推动理论成果转化,探索神经解码反馈应用于社交焦虑干预的可能性

基于机制的理解,研究致力于进一步探索临床干预方面的转化。研究将针对社交焦虑构建EEG解码器,基于实时神经反馈随机对照试验,验证基于脑信号调控社交焦虑的可能性,同时也为负向先验预期影响社交焦虑提供因果证据。基于贝叶斯框架的解码神经反馈致力于结合脑调控干预和经典的认知行为疗法干预(Cognitive behavioral therapy, CBT),探索从神经生物学层面和心理层面双向结合的个性化临床干预道路。

参考文献

- 蒋婧琪,王浩宇,钱铭怡. (2019). 社交焦虑注意偏向的动态变化. *心理科学进展*, 27(11), 1887-1895.
- 李涛,冯菲. (2013). 社交焦虑解释偏差:研究范式,特征及矫正. *心理科学进展*, 21(12), 2196-2203.
- 彭玉佳,王愉茜,路迪. (2023). 基于生物运动的社交焦虑者情绪加工与社会意图理解负向偏差机制. *心理科学进展*, 31(6), 905-914.
- Alden, L. E., Taylor, C. T., Mellings, T. M., & Laposa, J. M. (2008). Social anxiety and the interpretation of positive social events. *Journal of Anxiety Disorders*, 22(4), 577-590.
- Andersson, G., Westöo, J., Johansson, L., & Carlbring, P. (2006). Cognitive bias via the internet: A comparison of web-based and standard emotional stroop tasks in social phobia. *Cognitive Behaviour Therapy*, 35(1), 55-62.
- Bantini, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, 50, 40-51.
- Beltzer, M. L., Adams, S., Beling, P. A., & Teachman, B. A. (2019). Social anxiety and dynamic social reinforcement learning in a volatile environment. *Clinical Psychological Science*, 7(6), 1372-1388.
- Beltzer, M. L., Daniel, K. E., Daros, A. R., & Teachman, B. A. (2023). Examining social reinforcement learning in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 80, 101810.
- Boly, M., Baletau, E., Schnakers, C., Degueldre, C., Moonen, G., Luxen, A., ... Laureys, S. (2007). Baseline brain activity fluctuations predict somatosensory perception in humans. *Proceedings of the National Academy of Sciences*, 104(29), 12187-12192.
- Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, 29(24), 7869-7876. <https://doi.org/10.1523/JNEUROSCI.0113-09.2009>
- Cao, J., Gu, R., Bi, X., Zhu, X., & Wu, H. (2015). Unexpected acceptance? Patients with social anxiety disorder manifest their social expectancy in ERPs during social feedback processing. *Frontiers in Psychology*, 6, 1745. <https://doi.org/10.3389/fpsyg.2015.01745>
- Chen, J., Short, M., & Kemps, E. (2020). Interpretation bias in social anxiety: A systematic review and meta-analysis. *Journal of Affective Disorders*, 276, 1119-1130.
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment, and treatment* (pp. 69-93). New York: Guilford Press.
- Dietel, F. A., Möllmann, A., Bürkner, P.-C., Wilhelm, S., & Buhlmann, U. (2021). Interpretation bias across body dysmorphic, social anxiety and generalized anxiety disorder—A multilevel, diffusion model account. *Cognitive Therapy and Research*, 45(4), 715-729. <https://doi.org/10.1007/s10608-020-10180-7>
- Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry*, 164(10), 1476-1488.
- Freitas-Ferrari, M. C., Hallak, J. E., Trzesniak, C., Santos Filho, A., Machado-de-Sousa, J. P., Chagas, M. H. N., Nardi, A. E., & Crippa, J. A. S. (2010). Neuroimaging in social anxiety disorder: A systematic review of the literature. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 34(4), 565-580.
- Gu, R., Ao, X., Mo, L., & Zhang, D. (2020). Neural correlates of negative expectancy and impaired social feedback

- processing in social anxiety. *Social Cognitive and Affective Neuroscience*, 15(3), 285–291. <https://doi.org/10.1093/scan/nsaa038>
- Gu, R., Ge, Y., Jiang, Y., & Luo, Y. (2010). Anxiety and outcome evaluation: The good, the bad and the ambiguous. *Biological Psychology*, 85(2), 200–206. <https://doi.org/10.1016/j.biopsycho.2010.07.001>
- Hattingh, C. J., Ipser, J., Tromp, S., Syal, S., Lochner, C., Brooks, S., & Stein, D. J. (2013). Functional magnetic resonance imaging during emotion recognition in social anxiety disorder: An activation likelihood meta-analysis. *Frontiers in Human Neuroscience*, 6, 347.
- Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences*, 105(31), 10984–10989.
- Hesselmann, G., Kell, C. A., & Kleinschmidt, A. (2008). Ongoing activity fluctuations in hMT+ bias the perception of coherent visual motion. *Journal of Neuroscience*, 28(53), 14481–14485. <https://doi.org/10.1523/JNEUROSCI.4398-08.2008>
- Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PloS One*, 5(3), e9926. <https://doi.org/10.1371/journal.pone.0009926>
- Heuer, K., Lange, W.-G., Isaac, L., Rinck, M., & Becker, E. S. (2010). Morphed emotional faces: Emotion detection and misinterpretation in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 41(4), 418–425.
- Jusyte, A., & Schönenberg, M. (2014). Threat processing in generalized social phobia: An investigation of interpretation biases in ambiguous facial affect. *Psychiatry Research*, 217(1–2), 100–106.
- Keynan, J. N., Cohen, A., Jackont, G., Green, N., Goldway, N., Davidov, A., ... Hendler, T. (2019). Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature Human Behaviour*, 3(1), 63–73. <https://doi.org/10.1038/s41562-018-0484-3>
- Lamba, A., Frank, M. J., & FeldmanHall, O. (2020). Anxiety impedes adaptive social learning under uncertainty. *Psychological Science*, 31(5), 592–603. <https://doi.org/10.1177/0956797620910993>
- Laposa, J. M., Cassin, S. E., & Rector, N. A. (2010). Interpretation of positive social events in social phobia: An examination of cognitive correlates and diagnostic distinction. *Journal of Anxiety Disorders*, 24(2), 203–210.
- Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, 20(9). <https://doi.org/10.1038/nn.4615>
- Lou, B., Li, Y., Philastides, M. G., & Sajda, P. (2014). Prestimulus alpha power predicts fidelity of sensory encoding in perceptual decision making. *Neuroimage*, 87, 242–251. <https://doi.org/10.1016/j.neuroimage.2013.10.041>
- Macdonald, J. S., Mathan, S., & Yeung, N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Frontiers in Psychology*, 2, 82. <https://doi.org/10.3389/fpsyg.2011.00082>
- Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Research*, 241, 122–127.
- Mathys, C. D., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5, 39.
- Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, 8, 825.
- Nakao, T., Sanematsu, H., Yoshiura, T., Togao, O., Murayama, K., Tomita, M., Masuda, Y., & Kanba, S. (2011). fMRI of patients with social anxiety disorder during a social situation task. *Neuroscience Research*, 69(1), 67–72.
- Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, 143(5), 521–542.
- Peng, Y., Knotts, J. D., Taylor, C. T., Craske, M. G., Stein, M. B., ... Paulus, M. P. (2021). Failure to identify robust latent variables of positive or negative valence processing across units of analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(5), 518–526. <https://doi.org/10.1016/j.bpsc.2020.12.005>
- Peng, Y., Knotts, J. D., Young, K. S., Bookheimer, S. Y., Nusslock, R., Zinbarg, R. E., ... Craske, M. G. (2023). Threat neurocircuitry predicts the development of anxiety and depression symptoms in a longitudinal study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 8(1), 102–110. <https://doi.org/10.1016/j.bpsc.2021.12.013>
- Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351), 596–600.
- Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, 35(8), 741–756.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260–281.
- Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *The Lancet*, 371(9618), 1115–1125.
- White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, 54(1), 39–52.
- White, C. N., Skokin, K., Carlos, B., & Weaver, A. (2016). Using decision models to decompose anxiety-related bias in threat classification. *Emotion*, 16(2), 196–207.
- Xia, L., Gu, R., Lin, Y., Qin, J., Luo, W., & Luo, Y. (2023). Explaining reversal learning deficits in anxiety with electrophysiological evidence. *Journal of Psychiatric Research*, 164, 270–280. <https://doi.org/10.1016/j.jpsychires.2023.06.030>
- Xu, J., Wang, Y., & Peng, Y. (2024). Psychodynamic profiles of major depressive disorder and generalized anxiety disorder in China. *Frontiers in Psychiatry*, 15, 1312980. <https://doi.org/10.3389/fpsyg.2024.1312980>

Investigating social cognitive characteristics of social anxiety within the Bayesian framework

PENG Yujia^{1,2,3}, WANG Yuxi¹, JU Qianqian¹, LIU Feng⁴, XU Jia⁵

(¹ School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China)

(² Institute for Artificial Intelligence, Peking University, Beijing 100871, China)

(³ State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence, Beijing 100080, China) (⁴ School of Psychology, Shanghai Jiao Tong University, Shanghai 200030, China)

(⁵ Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health (Peking University), National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Beijing, 100191, China)

Abstract: Social anxiety disorder (SAD) is among the most common anxiety disorders, marked by overwhelming fear and avoidance of social behaviors and social scenarios, and debilitates patients' lives and work. Previous studies have provided ample evidence of dysregulated social cognition in social anxiety, such as negative cognitive biases, demonstrating a negative processing of social information. However, the factors driving the dysregulated social cognition remain unclear, impeding the elucidation of the underlying computational neural mechanisms of social anxiety symptoms, and guiding personalized interventions. Within the Bayesian framework, the current project proposed that the negative cognitive biases phenomenon may stem from negative prior expectations. We will use psychophysics experiments, electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), computational modeling, and machine learning to comprehensively investigate the characteristics and dynamic patterns of negative cognitive biases among individuals with social anxiety traits. This project has the promise to reveal cognitive neural mechanisms underlying dysregulated social cognition, potentially guiding future personalized interventions for social anxiety.

Keywords: social cognition, social anxiety, cognitive modeling, cognitive neural mechanisms, brain imaging