

面孔社会排斥的心理表征及推断模式*

侯春娜¹ 马一升¹ 伍麟² 刘志军³

(¹ 东北师范大学心理学院, 长春 130024) (² 武汉大学社会学院, 武汉 430072)

(³ 长春理工大学社会学系, 长春 130022)

摘要 研究基于特质归因理论, 采用反向相关图像分类技术考察了面孔引发社会排斥的心理表征内容及特质推断模式。以中国大学生为被试, 研究1通过启动社会排斥情境后执行双图像迫选任务, 发现被排斥者的面孔心理表征包含可信性与支配性特质信息, 鼻子和嘴巴(可信性)、眉毛(支配性)是核心诊断信息区域。研究2客观测量和主观评估一致表明, 低可信、低支配心理表征是社会排斥的典型特质特征, 在社会排斥的特质推断过程中均发挥关键作用。其中低可信性作用权重更大, 有力支持可信性优先假说。研究揭示了面孔心理表征在社会排斥过程中的作用, 把可信性优先机制从言语材料扩展到面孔刺激, 为理解高级社会认知层面的排斥过程提供了新的理论依据。

关键词 面孔, 社会排斥, 心理表征

分类号 B842

1 引言

人类作为一种社会性生物, 依赖建立和维护社会关系来确保生存和繁衍的成功。如果一个人能很好地适应这种社会环境并顺利融入其中, 就会得到社会的接纳; 反之, 就可能遭遇社会排斥。这一现象包含双方面体验, 不仅触及被排斥者的感受, 也关联排斥者的心理状态。被排斥者视角的研究多关注于遭受社会排斥后的行为反应, 如亲社会行为、攻击性行为及退缩回避等(Casini et al., 2022; Chen et al., 2023; Dickerson & Quas, 2024)。虽然排斥行为会给被排斥者带来极大痛苦, 并导致其生存利益遭受威胁; 但从积极角度看, 对于排斥行为的发起者而言, 此行为往往蕴含特定的适应性目的。通过选择性排斥或忽视特定个体, 行为者旨在获取潜在的适应性收益, 进而强化群体内部的亲密感与凝聚力(Wyer & Schenke, 2016)。需指出的是, 恶意和不公平的排斥很容易遭到他人的贬低和惩罚, 且毫无依据的排斥也会让排斥者感到不适。因而, 除非有充分理由, 否则人们并不会轻易实施排斥

(Rudert et al., 2020)。

针对排斥者视角的研究中, 探究排斥行为的发生机制始终是学界的热点议题。在人际交往的微妙互动中, 面孔特质作为构成第一印象的关键要素, 影响力不容忽视。大多数进化心理学家相信, 自然选择已经为人类塑造出能够对蕴含社会排斥的线索进行监测及反应的系统。它将这些线索与适当的预警和反应机制联系起来, 帮助人类有选择地排斥他人, 并检测自身被排斥或可能被排斥的情况(Over & Uskul, 2016)。这种进化出的心理适应机制可能与面孔线索的特质推断有关(Rudert et al., 2020)。

来自特质归因理论(Trait Attribution Theory)的观点认为, 知觉者会基于可信性和支配性两个特质维度对陌生面孔的个性特征进行推断(Oosterhof & Todorov, 2008; Todorov et al., 2015)。此过程极为迅速(100ms 内即可完成), 且具有跨文化的共识性(Todorov & Oosterhof, 2011)。这是人类心理对环境刺激进行评估所进化出的能力, 它们几乎完全阐释了人们如何看待他人(Fiske et al., 2007)。其中, 可

收稿日期: 2025-02-11

* 教育部人文社会科学研究规划基金项目(22YJA190002)资助。

通信作者: 伍麟, E-mail: wulin@whu.edu.cn; 侯春娜, E-mail: houcn359@nenu.edu.cn

信任特质涉及帮助或伤害意图,能提供有关他人信任或不信任的潜在信息;而支配性特质则反映执行此意图的能力(Todorov et al., 2015)。它们与刻板印象内容模型(stereotype content model, SCM; Fiske et al., 2007)中的“热情”(可信性)和“能力”(支配性)存在对应关系(Sutherland et al., 2016),并构成了社会认知内容的两个基本维度(Park & Baumeister, 2015)。

面孔特质推断如何诱发社会排斥?学者推断可能存在一个行为机制,这种机制与人们头脑中对排斥目标形成的社会共识性的面孔刻板印象有关(Rudert et al., 2020)。刻板印象作为一种关于某些群体成员个性、特征及行为的信念,其本质是头脑中关于真实的(或感知到的)差异的心理表征(Chattalas, 2005)。心理表征(mental representation)作为一种图式知识结构,是信息或知识在心理活动中的表现和记载的方式(侯春娜,刘志军,2019)。心理表征的内容构成了信息解释的基础,通过与输入面孔信息的匹配生成对特定目标(个体或群体)的评价,用以产生适应性行为(Jack & Schyns, 2017)。迄今,针对该议题的实验研究仍显匮乏。根本原因在于面孔特质心理表征的获取方式在很大程度上已超出传统研究方法的范畴(Jack & Schyns, 2017)。

SCM的理论框架为面孔社会排斥研究提供了重要参考。SCM理论认为“热情”与“能力”定位在二维空间中的象限将会形成有关个体(或群体)的不同刻板印象内容,这是导致社会排斥的主要原因(Fiske et al., 2007)。Rudert等人(2017)率先检验了面孔刺激能否在“热情-能力”框架下诱发社会排斥。他们借助巴塞尔面孔模型(Basel Face Model, BFM),通过操控“热情”与“能力”维度向量,生成4个象限的合成面孔,并要求被试想象排斥一名外群或内群成员。结果发现,排斥倾向与群体身份并无关联,仅取决于面孔在“热情-能力”二维空间中的位置:冷漠无能的面孔遭到最强烈的排斥,其次是冷漠而有能力的面孔,热情无能的面孔则最不易受到排斥,这与SCM相一致。但情绪的中介作用模式与SCM存在差异:冷漠无能面孔确实激起了强烈的厌恶,但同时伴生怜悯;而热情无能面孔并未唤起SCM中的同情,反而以“厌恶缺失”为特征。

Rudert等人(2017)的研究首次证实了面孔特质可诱发社会排斥,但局限在于:第一,研究范式上,Rudert等人沿用了Oosterhof与Todorov(2008)的数据驱动建模逻辑,即先在面孔空间内提取与“热情”

及“能力”每个维度单独相关的物理特征。随后以线性平均的方式将两组物理特征线性平均叠加,合成4种高低水平不同的面孔,最后将之映射到二维坐标系内的4个象限中,作为SCM框架的刻板印象内容要求被试进行特质评估。但是这种面孔建模方式与最初SCM识别二维特质空间的方法完全相反,SCM是从两个维度组合的大量判断中得到的二维象限空间。学者质疑这种面孔建模方式,并未真实反映出面孔特质的整合方式(Oliveira et al., 2019)。第二,任务设置上,该实验要求被试主动实施排斥。由于实验者处于权威地位,被试可能会质疑排斥他人的正当性,并意识到故意排斥将会给他人带来痛苦,进而引发自身心理困扰(Wirth & Wesselmann, 2018)。

在此基础上,Rudert等人(2020)改进了研究方法,采用了基于随机矢量的反向相关图像分类(Reverse Correlation Image Classification, RCIC)技术,探索面孔社会排斥心理表征的人格特点。RCIC是一种纯粹基于数据驱动的心理物理学方法,不受研究者先验假设的影响,可以将特定面孔特质的刻板印象或心理表征可视化。在此过程中,被试需先激活相应的心理表征,再用其与输入刺激进行比对。因此,RCIC方法本身和所得结果都对心理表征极为敏感(Oliveira et al., 2019)。采用RCIC技术,Rudert等人(2020)在无需明确提及排斥的情况下,同时获得了面孔社会排斥及社会接纳两个心理表征,证实人们对何种面孔易被排斥存在社会共识性观念。同时,主观评估显示,面孔社会排斥的心理表征综合了低责任心及低宜人性两种人格;特质评分结果显示,其具有低可信性及低支配性。在此研究中,Rudert等人对SCM解释力提出了质疑,认为面孔特质是产生社会排斥更为直接的原因,而情绪(厌恶)则是相对远端的原因;不仅如此,混合型刻板印象(高宜人性及低责任心)的排斥,无法通过补偿效应解释。在SCM中,补偿效应是指当某一社会群体在“热情”与“能力”任一维度获得消极评价时,会通过另一维度的积极评价进行补偿,从而相互抵消(Cuddy et al., 2009)。

Rudert等人(2020)推进了社会排斥在面孔刺激中的研究,但仍存两方面问题:第一,为创建图像分类任务的刺激,Rudert等人(2020)采用BFM中的形状(长度,圆度等)作为随机向量,以BFM平均的真人面孔为基础面孔,通过随机向量参数的运算改变基本面孔形状,生成了一组新颖的真人面孔作为

刺激材料。此做法被认为难以保障生态效度：因为每个真实面孔的身份都是唯一的，必定包括特定的形态(如形状和结构)，这将会诱发多种面孔推断(情绪、身份、性别)而混淆结果，从而难以获得符合实验者需要的心理表征(Jack & Schyns, 2017)。第二，更为重要的是，面孔特质推断并非取决于整个面孔，而是少数关键区域，只有精确定位出这些诊断信息区域，才能充分且必要地识别出此面孔的特质推断(Jack & Schyns, 2017)。虽然可信性与支配性的诊断信息区域已被确认(Dotsch & Todorov, 2012; Mo et al., 2022)，但 Rudert 等人(2020)并未报告面孔社会排斥心理表征存在这些特质诊断信息区域。此外，该 RCIC 研究的试验次数不足 100 次，可能降低图像信噪比影响心理表征的图像质量(Giacomin et al., 2022)。

基于此，本研究采用随机噪音的 RCIC 技术，改进 Rudert 等人(2020)的研究方法。这种技术通过向同一基本面孔分别叠加正负相反的随机正弦噪音，生成互为镜像的两张备择性面孔刺激。所谓镜像是指两个面孔刺激在视觉效果上相对称(侯春娜, 刘志军, 2019; Okazawa et al., 2018)。比如其中一张的某区域增亮，在另一张中对应区域则变暗。这种设计使差异最大化，单次试验即可捕捉最大信号(Dotsch et al., 2008)。由于叠加随机噪音使基本面孔扭曲，视觉会呈现出非真人的模糊效果，因而具有较佳生态效度。同时，此技术并不依赖基本面孔，仅通过纯粹随机噪音即可提取心理表征。更重要的是，这种 RCIC 技术可以直接对心理表征的像素数据进行分析，精确定位出面孔推断的诊断信息区域(Oliveira et al., 2019)。

借助此种技术，本研究通过两个研究，旨在揭示此前尚未发现的研究内容：研究 1a 拟在确保生态效度的前提下，获取中国人面孔社会排斥的心理表征可视化图像。在此基础上，尝试回答两个科学问题：(1)面孔社会排斥心理表征在特质维度上具有何种独特特征(研究 1b)? (2)可信性与支配性两种面孔特质如何影响面孔社会排斥心理表征，也即其特质推断的模式如何(研究 2)?

2 研究 1: 面孔社会排斥心理表征的特质构成

2.1 研究 1a: 面孔社会排斥心理表征内容的可视化

研究 1 旨在以中国人为被试，借助随机噪音的

RCIC 技术，通过更多试验次数(300 次)，生成中国人面孔社会排斥心理表征的可视化图像。

2.1.1 研究方法

研究设计 采用单因素(情境类型：排斥 vs. 接纳)被试间实验设计。自变量为被试启动的情境类型，包括社会排斥和社会接纳两个水平；因变量为面孔刺激项目的选择结果。

被试 RCIC 技术以图像像素亮度值而非单样本为分析单位，凭借庞大的像素样本量确保统计检验力。以往研究中，20 名被试即达最佳效果(Oliveira et al., 2019)。基于此，本研究招募 81 名被试，平均年龄 19.60 ± 0.83 岁。其中社会排斥组 42 人($N_{\text{女}} = 29$)，社会接纳组 39 人($N_{\text{女}} = 24$)，被试数量满足统计检验力要求。

所有被试视力或校正视力正常，此前均未参与过类似的实验。研究获得了东北师范大学伦理委员会批准(2023054)，被试在参与实验前签署书面知情同意书，实验后获得一定报酬。

实验材料 面孔刺激全部由每次试验中的随机噪音叠加在同一基本面孔上构成。其中，基本面孔为男性中性情绪的平均面孔，这有助于保障生态效度(侯春娜, 2017)。随机噪音则由 6 个不同朝向(0° , 30° , 60° , 90° , 120° 和 150°)的 2 个周期叠加的截断正弦曲线片段组成，并按照 5 种空间尺度(2, 4, 8, 16 和 32 个周期) \times 2 个相位($0, \pi/2$)形成包含 4092 个随机对比度参数的噪音点图，生成过程见图 1 所示。研究共计生成了 300 对正向和负向面孔刺激。面孔刺激图像为 512×512 像素，实验中在 70 cm 处的观看尺寸设置为 $9.70 \text{ cm} \times 13.00 \text{ cm}$ ，这近似于从 1 m 距离观看真实面孔($13.80 \text{ cm} \times 18.50 \text{ cm}$)的大小。

实验流程

第一阶段：启动社会排斥 获得所有被试的知情同意后，首先实验者让被试想象自己身处排斥者的位置，观看 Cyberball 游戏视频(Wesselmann et al., 2009)。具体而言，被试会基于第一视角观看一段视频。视频中“我”作为游戏发起者将主导一场三人传球游戏。其中，玩家 1 在游戏中很少会得到“我”的传球，充当被排斥者角色；相比之下，玩家 2 将会得到“我”更多次的传球，扮演社会接纳者角色。

第二阶段：双图片迫选 视频结束后，被试被随机分成两组(玩家 1 = 社会排斥组；玩家 2 = 社会接纳组)执行 RCIC 的双图片迫选任务(Two-Image Forced Choice Task, 2IFC)，选出符合期望的玩家面孔。具体而言，在一次试验中，屏幕中心呈

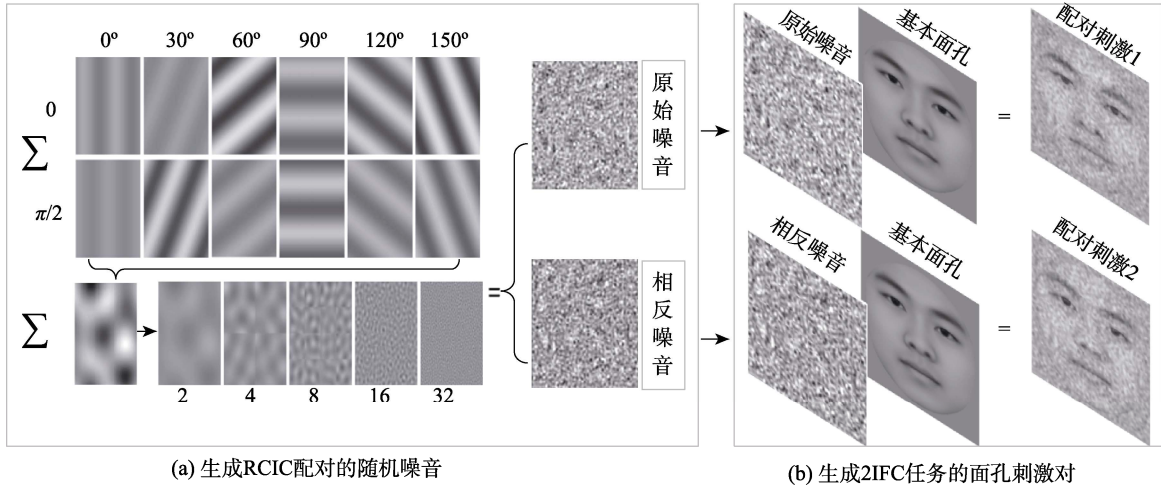


图 1 RCIC 备择式面孔刺激对的生成示意图

现 500 ms 的“十”字后, 会在水平方向并排显示两个备择式的面孔刺激图片。两组被试分别依照屏幕上方指示“两个面孔中, 哪一个与玩家 1 (/玩家 2) 更相符?”, 并执行相应的键盘迫选反应(F 键 = 左图; J 键 = 右图)。

2IFC 试验次数设置为 300 次。学者认为这将显著提升图像质量(Giacomin et al., 2022)。试验分为 6 个 block, 每个 block 间有短暂休息。整个试验中备择式的两种噪音刺激位置已进行平衡处理, 面孔刺激对的顺序随机出现。

所有试验完成后, 被试还需对玩家身份的感受体验进行 7 点评估(-3 = 此人被人排斥; +3 = 此人被人接纳), 并获得小礼品作为酬谢(价值约¥10 元)。实验结束后实验者解释了实验目的并对所有被试做了抚慰工作, 以消除不良影响。

2.1.2 结果

数据筛选 首先, 研究依照玩家身份感受体验的评估结果对数据进行了筛选。剔除了未能有效甄别被排斥者的 1 名被试及社会接纳者的 6 名被试。社会排斥组 41 人及社会接纳组 33 人进入至 RCIC 的数据分析环节。

其次, 为了提高信号—噪音比率以获得更高质量的群体分类图像(groupCI), 从而获得效果更佳的心理表征图像。依照个体分类图像情况(individual CI)进行筛查, 并剔除社会排斥组中 individual CI 与 groupCI 负相关的 1 名被试。最终共获得 73 人的 RCIC 有效分析数据, 其中社会排斥组 40 人, 社会接纳组 33 人。

数据一致性检验 学界认为个体相互间的 CI 图像噪音模式达到良好一致性时进行合并, 平均的

group CI 图像才有意义。因此, 本研究对 individual CI 之间的数据一致性进行了检验。以 group CI 有效面孔的显著区域为各个 individual CI 计算范围。具体做法为, 选取 group CI 面孔有效区域之外的非信息标准化区域为基线(确定均值和标准差)。通过基线将 group CI 图像 Z 转换, 并选取 $Z \geq 2.58$ 为阈限标准, 从而构建出显著区域。

像素的组内一致性系数 ICC (Intraclass Correlation Coefficient)的 two-way random model 分析结果显示, 社会排斥组个体之间的 ICC = 0.972, 95% CI 为[0.972, 0.973]; 社会排接纳组个体之间的 ICC = 0.938, 95% CI 为[0.935, 0.942], 表明两组成员间的噪音模式一致性情况理想(ICC > 0.90 表示极好; Koo & Li, 2016)。此外, 辅助的皮尔逊相关分析显示: 社会排斥组 individual CI 与 groupCI 噪音模式的平均相关系数为 0.954, 范围为 0.894 ~ 0.983, 四分位数为 0.947、0.960 和 0.965; 社会接纳组 individual CI 与 groupCI 噪音模式的平均相关系数为 0.817, 范围为 0.675 ~ 0.919, 四分位数为 0.741、0.830 和 0.873, 且所有相关系数均为正值。结果表明各组成员间的 RCIC 数据一致性情况较为理想, 能够为后续社会排斥与社会接纳的心理表征图像的生成提供保障。

面孔社会排斥的心理表征图像 依照 RCIC 数据分析流程, 将社会排斥组各个被试所有试次的迫选结果, 即通过所有符合预期“被排斥者”形象的选项结果的随机噪音模式的合并平均, 生成社会排斥组的 group CI (见图 2b); 再将其叠加在基本面孔上, 就获得排斥者头脑中被排斥者的心理表征形象(见图 2a)。与之类似, 亦获得被接纳者心理表征形象。

为更好呈现社会排斥的心理表征诊断信息效果,本研究尝试在原始心理表征信息上进行面孔噪音的视觉信息重构。这可以更为“逼真地”展示社会排斥心理表征的诊断信息特征;同时亦重构了社会接纳心理表征效果作为对照(图 3)。

辅助分析 本研究采用结构相似度指标 (Structural Similarity index, SSIM)额外考察了社会

排斥与社会接纳心理表征间的相似性,其优点在于它与观察者的主观判断高度一致。SSIM 结果显示,社会排斥与社会接纳心理表征之间的噪音模式显著负相关($SSIM = -0.334, p < 0.001$),这一发现符合人们的常规认知。进一步的 3D 热图在视觉效果上以明暗对比的方式直观呈现了负相关结果(见图 2c),从而排除了随机噪声产生伪像的可能性。

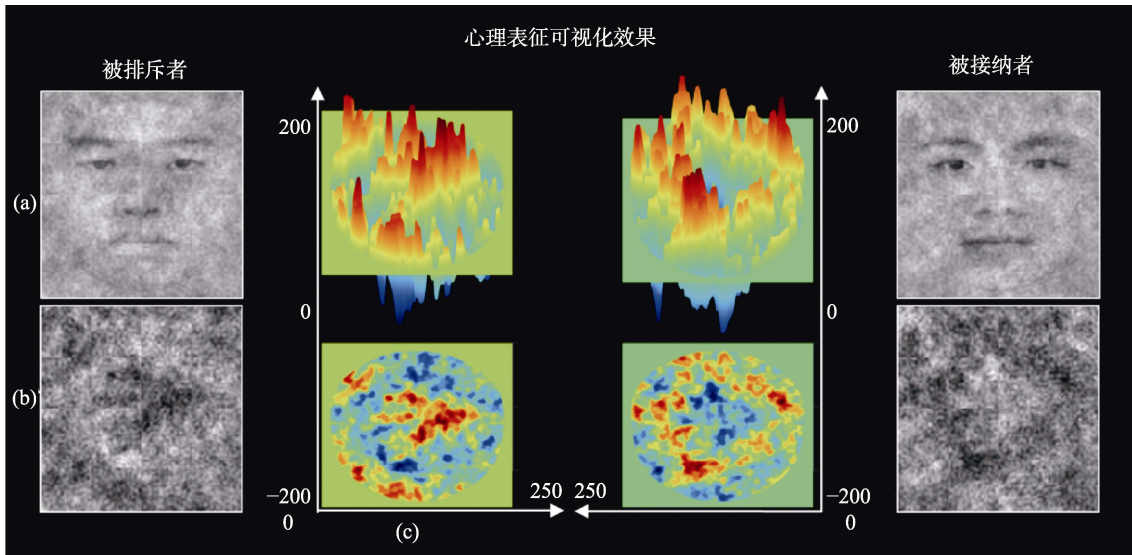


图 2 社会排斥与社会接纳的 RCIC 分析效果图示

注: (a)心理表征的可视化图像; (b)群体分类图像; (c)群体分类图像有效面孔区域 3D 热图,其中红色表示暗像素,蓝色表示亮像素。彩图见电子版,下同。

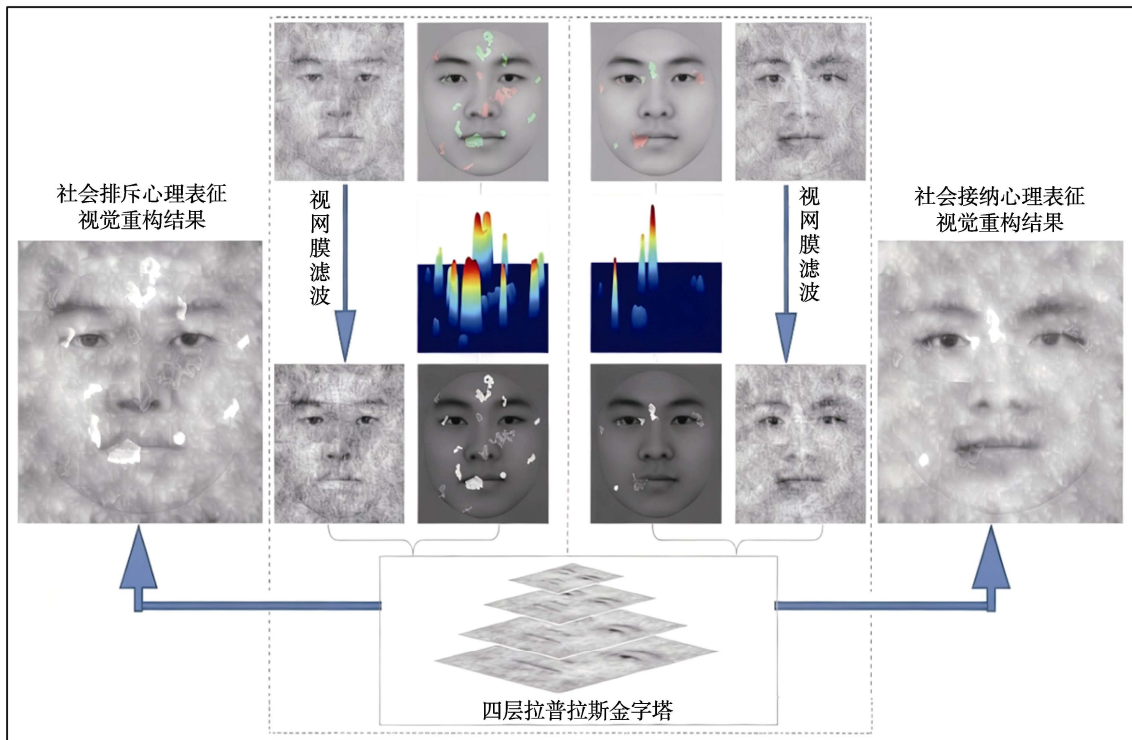


图 3 社会排斥心理及社会接纳心理表征的视觉重构示意图

2.2 研究 1b: 面孔社会排斥心理表征的特质特征

2.2.1 研究目的

研究 1b 旨在考察面孔社会排斥心理表征的诊断信息区域, 并将之与特质心理表征的已知诊断信息区域进行比对, 从而揭示面孔社会排斥心理表征的特质特征。

2.2.2 研究假设

研究 1b 基于社会排斥、面孔与 SCM 的已有研究提出预测假设:

(1)可信性方面。SCM 认为人们对社会世界中信任与不信任的区分极为敏感, 且难以改变(Fiske et al., 2007), 不友好、不可信的人会遭到排斥(Hales et al., 2016)。来自面孔特质的心理表征证据显示: 眼睛、鼻子和嘴巴区域是可信性特质推断的诊断信息, 具有跨文化的普遍性并与面孔的种族和性别无关(Dotsch & Todorov, 2012; Ma et al., 2015; Mo et al., 2022)。颧骨在眼睛区域附近, 颧骨(突出/平坦)也是面孔可信性判断的诊断信息(Todorov et al., 2008)。

(2)支配性方面。人们希望获得支配的能力, 相信和控制之中会获得个人益处, 并通过两种方式产生排斥。一种是主动的排斥。他人的高支配性常令个体感到地位受威胁, 而实施主动排斥。另一种被动排斥, 表现为个体对能力低下、无法带来适应收益者的忽视(Chen et al., 2017)。面孔特质的心理表征证据显示: 眉部区域(眉毛/眉脊)及面孔轮廓是支配性的主要诊断信息来源(Dotsch & Todorov, 2012)。

总之, 依据 SCM 观点, 不可信及高/低支配性的特质均可产生面孔社会排斥。若面孔社会排斥心理表征确实通过特质信息产生, 则其诊断信息区域

应体现出这些区域。近来文献表明, 眼睛与嘴巴是形成可信性推断的核心线索区域(Mo et al., 2022); 而中国人在社会交往中回避直接眼神接触(Wang et al., 2020), 偏向利用面孔下半部进行信息诊断, 且嘴巴在面孔中的信息量最大(Blais et al., 2012)。因而, 假设 1 预期: 若面孔社会排斥心理表征通过不可信特质信息产生排斥, 则嘴巴是其诊断信息区域。

同时, 眉毛的面孔特征较为稳定, 虽然并非固定的面孔结构(如眉骨), 但它是将前额和眼眶分成两个面孔区域的参照。失去眉毛会降低人们提取面孔信息的效率, 识别他人也会变得困难(Sekuler et al., 2004)。因而, 假设 2 预期: 若面孔社会排斥心理表征通过支配性(高/低)特质信息产生排斥, 则眉毛是其诊断信息区域。

2.2.3 研究方法

被试 研究 1b 仅需对研究 1a 所得心理表征图像进行数据分析, 无需另行招募被试。

研究材料与数据分析流程 研究 1b 以研究 1a 生成的 512×512 像素 RGB 格式的社会排斥与社会接纳 groupCI 以及基本面孔三幅图像为研究材料。前两者用于数据分析, 基本面孔用于结果的视觉呈现。数据分析以像素为分析单位, 采用标准偏差为 $\sigma_b = 4$ 像素的高斯滤波器进行平滑加工, 以显著性阈值 $Z_{crit} \geq |2.3|$, $p < 0.05$ 执行双侧像素检验及诊断差异图像分析(Dotsch & Todorov, 2012)。

像素检验结果 研究 1b 采用像素检验分析, 揭示社会排斥心理表征中的诊断信息区域, 参见图 4 左侧(a)部分。其中, 红色及绿色区域为像素检验的显著区域, 绿色簇集表示簇集中的像素亮度与社

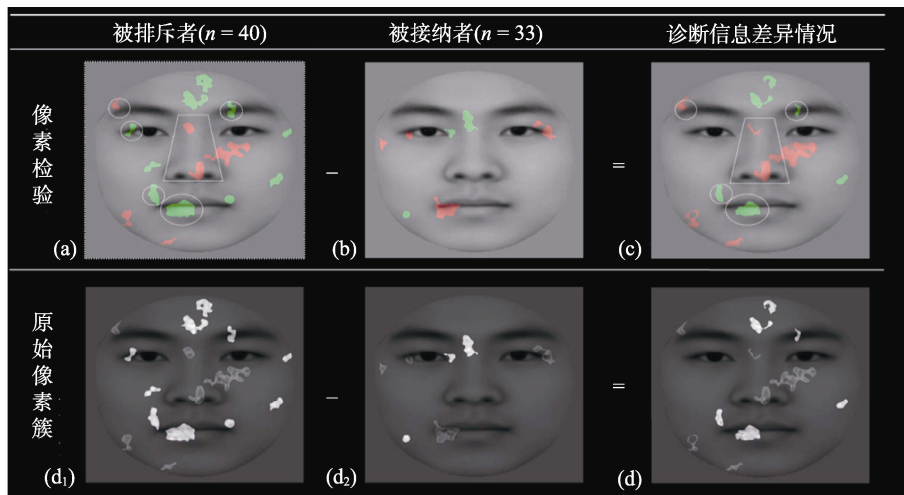


图 4 像素检验结果及诊断信息差异情况

会排斥正相关(即这些像素较亮时,被试更有可能做出社会排斥);相反,红色簇集表示簇集中的像素亮度与社会排斥负相关(像素较暗时,被试更倾向做出社会排斥)。结果显示,社会排斥的诊断信息包含着面孔的大部分关键信息区域:额头区域、眼部区域、鼻子区域、嘴部区域以及面孔四周边缘轮廓区域。

其中,眼部区域和嘴部区域被认为是可信性特质推断的关键诊断信息区域(Dotsch & Todorov, 2012)。鼻子区域也是可信性特质推断的诊断信息区域(Dotsch & Todorov, 2012)。这意味着社会排斥心理表征中可能包含可信性特质信息。同时,眉毛是支配性的重要诊断信息区域(Dotsch & Todorov, 2012),可信性在眉毛上则无法做出区分(Todorov et al., 2015)。散布面孔边缘的显著像素簇集显示面孔与背景轮廓区分性较低,这是低支配性的诊断信息(Dotsch & Todorov, 2012)。这意味着社会排斥心理表征同时还可能包含支配性特质信息。

诊断差异图像分析结果 在 Cyberball 范式中,被排斥者特点的揭示需要借助被接纳者,通过被排斥者与被接纳者比较所获得的差异揭示社会排斥的独特特征(Syrjämäki & Hietanen, 2020)。为此,研究 1b 执行了诊断差异图像分析(流程见图 4 下部分中 $d_1 - d_2 = d$)。结果显示,在剔除与社会接纳重合的显著区域后,社会排斥心理表征独特的诊断信息区域主要集中在眉毛、鼻子和嘴巴区域。鉴于像素的显著区域并未出现左侧集中的现象,研究结果可排除面孔加工偏侧化的影响,确认信息差异源自面孔信息本身。

此前面孔特质推断的心理表征文献表明,嘴巴区域越亮(绿色),越不可信任(刘志军, 2021; Mo et al., 2022)。图 4c 中白色虚线圈定的嘴巴区域与此一致,表明社会排斥心理表征中可能包含“不可信”特质信息,这与 SCM 观点相一致,至此研究假设 1 得到证实。眉毛区域越暗支配性越高(图 4c 左侧圈出的红色眉毛区域);越亮支配性越低(图 4c 右侧圈出的绿色眉毛区域)。这表明社会排斥心理表征中可能同时包含支配性两端的特质信息,这与 SCM 观点相符。假设 2 也得到证实。

此外,鼻子的诊断信息区域较大,基本包含整个鼻子长度(图 4c 白色梯形圈出的区域)。跨种族研究表明,鼻子是可信性特质推断的核心线索(Todorov et al., 2008),鼻子越长,越值得信任(Ma et al., 2015)。不过此结果只表示社会排斥心理表征

中可能包含可信性特质信息,无法区分可信与不可信的具体情况。值得注意的是,鼻子区域达到显著也可能与东方人特殊的面孔加工策略有关:东方人倾向于更多注视于面孔中央,即从鼻子区域提取更多信息,作为对眼睛(及嘴巴区域信息)的补偿性策略,从而回避直接的眼睛接触。这种对鼻子的关注,在东方人面孔特征文化注视偏差中是较为稳健的(Wang et al., 2020)。因而,需谨慎看待这一结果。

3 研究 2: 面孔社会排斥心理表征的特质推断模式

3.1 研究目的

研究 1 生成了人们对社会排斥者的心理表征形象,并揭示眉毛、鼻子和嘴巴是其关键诊断信息区域,其分别与可信性和支配性特质的诊断信息区域高度重叠,回答了问题 1。但可信性和支配性并非以单一维度,而是以统合的方式共同编码在面孔心理表征之中,并对特定的心理表征形象发挥作用(Gunaydin & DeLong, 2015)。两种特质如何作用于面孔社会排斥心理表征,研究 1 未能给出明确答案。

研究 2 旨在通过客观测量(研究 2a)和主观评估(研究 2b)两个子研究,检验 4 种单一特质(可信性:高 vs.低;支配性:高 vs.低)对社会排斥心理表征的预测情况,探索两种面孔特质基本维度的具体影响,揭示面孔心理表征中的特质推断模式,尝试回答问题 2。

3.2 研究假设

目前,特质信息影响面孔心理表征进而诱发社会排斥,主要有两种观点:可信性优先假说(Oliveira et al., 2019)及消极晕轮效应假说(Rudert et al., 2020)。

(1)可信性优先假说。Oliveira 等人(2019)提出在心理表征层面,特质维度之间也存在与 SCM 相似的二维空间。统合后的特质信息会映射到这个面孔空间中,最终形成对他人的刻板印象(心理表征)。在此过程中,两者以统合方式并行作用于特定面孔内容,且遵循权重不等原则。可信性比支配性发挥更大权重作用,也即是可信性优先。依照进化心理学观点,为了增加自身生存和繁衍的机率,识别他人是否具有伤害意图(可信性),要比判断他人执行此意图的能力(支配性)更加重要(Dotsch & Todorov, 2012; Oosterhof & Todorov, 2008)。因而,无论支配性(高/低)情况如何,一旦识别他人不可信,基于趋利避害,个体会选择疏远或排斥此人。证据表明,以低可信性为主的统合方式容易诱发社会排

斥。其中,低可信性与低支配性(冷漠无能)将诱发出最严重的社会排斥;而低可信性高支配性(冷漠有能力)诱发的排斥水平次之;相反,以低支配性为主的统合方式中,其与高可信性(热情无能力)的统合却并不会产生排斥(Rudert et al., 2017)。

(2)消极晕轮效应假说。Rudert 等人(2020)提出可信性与支配性并非线性函数,即一个维度的消极评价无法通过另一维度的积极评价进行补偿;而是非线性的消极权重加强,即当一个特质维度评价为消极时,将会导致整体评价更为消极,他们将之称为消极晕轮效应。甚至当另一特质维度评估为积极时,这种对比效果反而导致消极特质维度产生更为强烈的消极影响。以可信性为例,从功利主义的角度来看,社会排斥的功能就在于筛掉潜在的不良合作者。一旦可信性被评价为消极(不值得信任、无法合作),支配性便几乎失去权重;即使能力再强,也无法补偿信任的缺失,最终导致排斥发生。

两种观点共同之处在于均主张可信性与支配性并行作用于面孔心理表征,且两种特质维度的作用并不相等。然而分歧在于:可信性优先效应提出可信性是关键特质维度,仅不可信特质信息才会产生社会排斥;而消极晕轮效应假说则拒绝假定任何单一特质维度的关键地位,更加强调消极特质信息本身对社会排斥的影响。消极晕轮效应是以前否定面孔特质存在补偿效应为前提的(Rudert et al., 2020)。但从现有证据来看,消极晕轮效应假说的证据前难以成立,源于近来实验室证据表明补偿效应不仅出现在词语刺激中,面孔的可信性与支配性上也存在补偿效应(Schmitz et al., 2024)。因此,本研究提出假设 3:可信性与支配性特质并行作用于面孔社会排斥的心理表征,且相比支配性,可信性的作用权重更大。

3.3 研究 2a: 面孔特质对社会排斥心理表征作用的客观测量

3.3.1 预实验: 4 种单一特质面孔心理表征的生成

(1)研究目的

本预实验旨在获得特质归因理论中两个基本维度的 4 种不同类型的心理表征图像(可信性:高 vs.低;支配性:高 vs.低),用于后续面孔社会排斥心理表征的像素回归模型分析。

(2)研究方法

被试 采用与研究 1 相同的 RCIC 范式,共计招募了 200 名年龄在 17~22 岁的大学生志愿者,平均年龄 19.41 ± 1.18 岁 ($N_{女} = 124$)。

研究材料与流程 面孔噪音刺激来自研究 1, 2IFC 任务实施过程详见研究 1 所述,区别在于 2IFC 的指导语划分为 4 种特质情况。具体而言,高可信性 2IFC 任务要求迫选出“哪个面孔看起来更值得信赖”(35 名被试);高支配性 2IFC 任务要求被试迫选“哪个面孔看起来更具能力或领导力”(38 名被试);低可信性 2IFC 任务则要求迫选“哪个面孔看起来更不值得信任”(62 名被试);而低支配性 2IFC 任务要求迫选“哪个面孔看起来能力不足或缺领导力”(65 名被试)。特质指导语源自特质归因理论中对各特质类型的描述(Dotsch & Todorov, 2012; Fiske et al., 2007), 试验次数同样为 300 次。

数据有效性筛选与一致性检验 剔除各组中 individual CI 与 groupCI 负相关的被试数据,最终进入 RCIC 分析的各组有效数据分别为:高可信性组 30 人,低可信性组 58 人;高支配性组 31 人,低支配性组 58 人。数据一致性检验结果显示,高可信性组($n = 30$)的 $ICC = 0.951$, 95% CI 为[0.949, 0.953]。高可信性组 individual CI 与 groupCI 噪音模式的皮尔逊相关平均系数为 0.927, 范围为 0.847 ~ 0.971, 四分位数为 0.897、0.933 和 0.956。低可信性组($n = 58$)的 $ICC = 0.981$, 95% CI 为[0.980, 0.981]。低可信性组 individual CI 与 groupCI 噪音模式的皮尔逊相关平均系数为 0.958, 范围为 0.900 ~ 0.984, 四分位数为 0.945、0.963 和 0.976。

高支配性组($n = 31$)的 $ICC = 0.906$, 95% CI 为[0.903, 0.909]。高支配性组 individual CI 与 groupCI 噪音模式的皮尔逊相关平均系数为 0.914, 范围为 0.855 ~ 0.957, 四分位数为 0.897、0.913 和 0.935。低支配性组($n = 58$)的 $ICC = 0.947$, 95% CI 为[0.945, 0.949]。低支配性组 individual CI 与 groupCI 噪音模式的皮尔逊相关平均系数为 0.967, 范围为 0.918 ~ 0.989, 四分位数为 0.959、0.969 和 0.978。

总之,结果表明 4 种特质类型心理表征 ICC 均超过 0.90, 能够为后续分析提供保障。

(3)研究结果

与研究 1 的 RCIC 流程相同,生成高可信、低可信、高支配、低支配四种 512×512 像素 RGB 心理表征面孔图像,并作为研究 2a 中像素回归分析材料(图 5)。

3.3.2 面孔特质对社会排斥心理表征预测的像素回归分析

(1)研究方法

自变量为预实验生成的 4 种特质类型心理表征

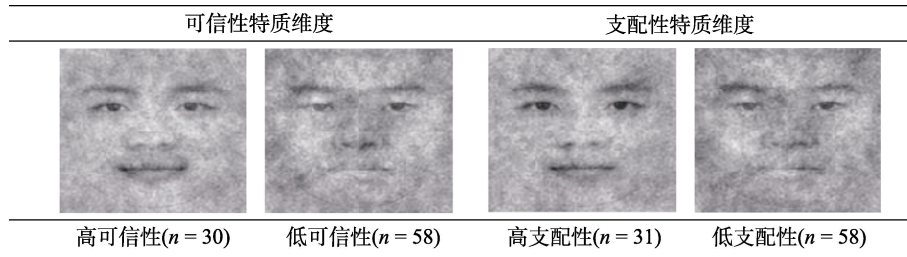


图 5 四种特质类型的面孔心理表征形象

(高可信、低可信、高支配、低支配), 因变量为研究 1 获得的面孔社会排斥心理表征。以像素亮度值作为分析单位, 建立面孔基本特质维度预测社会排斥(被排斥者)心理表征的像素回归模型。

(2) 研究结果

像素回归结果显示, 在对社会排斥心理表征的预测中, 低可信性显著正向预测社会排斥心理表征($\beta = 0.78$, 95% CI = [0.77, 0.78]), 且在所有面孔特质心理表征类型中的预测效应最大, 这表明低可信性是社会排斥心理表征的关键因素。此外, 高可信性的负向预测作用显著($\beta = -0.11$, 95% CI = [-0.12, -0.11]), 意味着高可信性将会削弱社会排斥。这种正负相反的预测效果表明, 可信性维度对社会排斥心理表征的作用具有区分性。

相比之下, 支配性维度的预测表现有所不同。结果显示, 无论是高支配性还是低支配性, 均对社会排斥心理表征表现出正向预测作用, 且预测程度较为接近($\beta = 0.19$, 95% CI = [0.18, 0.19]; $\beta = 0.15$, 95% CI = [0.15, 0.16]), 这与可信性维度的预测模式存在差异。这一结果表明, 面孔心理表征产生的

社会排斥受可信性与支配性特质的并行影响, 并由冷漠无能(低可信性与低支配性)及冷漠有能力(低可信性与高支配性)两种特质统合方式诱发。且两种特质中(不)可信性特质始终发挥关键的预测作用。此结果通过可信性优先假说及消极晕轮效应假说, 均可得到合理解释。

残差结果显示, 残差 Residual CI 图像大部分区域已被单一暗色像素所填充, 图像中几乎只残存基础图像, 这意味着特质预测结果已解释了社会排斥心理表征的大部分变异(图 6 所示); 同时 R^2 为 0.62, 表明此像素回归模型拟合情况较好。

同时, 补充的社会接纳的像素回归模型显示, 高可信性($\beta = 0.50$, 95% CI = [0.49, 0.50])及高支配性($\beta = 0.37$, 95% CI = [0.37, 0.38])均发挥显著的正向预测作用, 且可信性的预测作用最大。同时, 低可信性($\beta = -0.17$, 95% CI = [-0.17, -0.16])及低支配性($\beta = -0.07$, 95% CI = [-0.07, -0.06])均发挥负性预测作用。结果表明, 可信性与支配性并行作用于社会接纳心理表征, 且两者的作用都具有区分性, 热情有能力(高可信性与高支配性)特质统合方式最

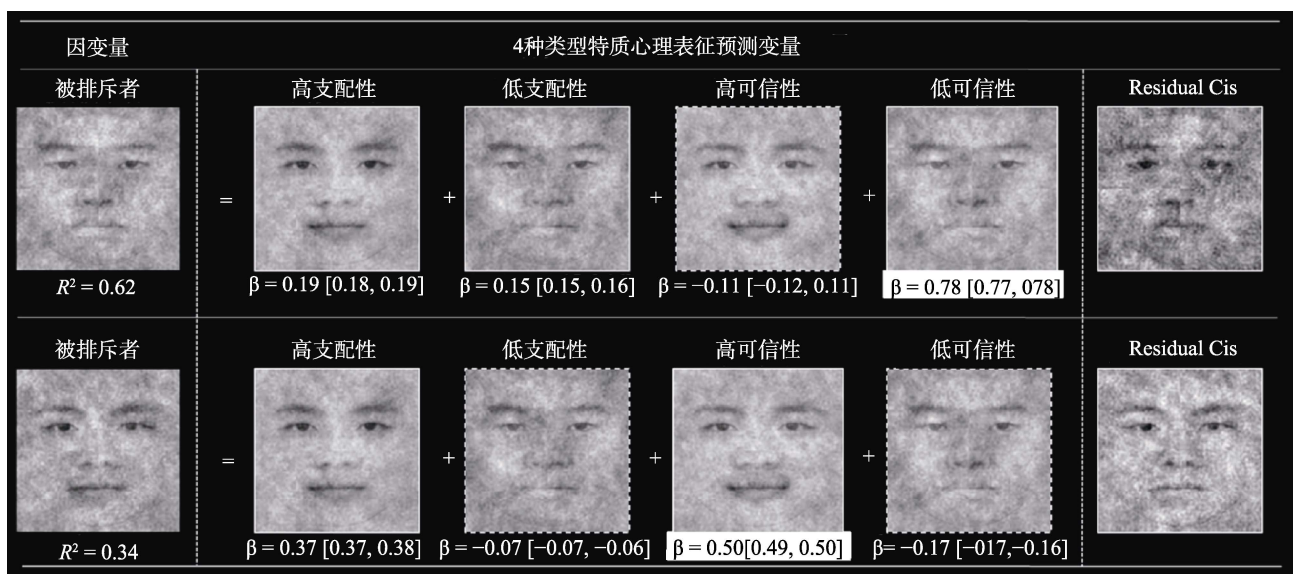


图 6 不同单一特质对社会排斥心理表征的像素回归预测情况

容易受人接纳,这与此前内群偏私的刻板印象结果较为相符:人们会赋予内群成员更为积极的特质,认为“热情且有能力”(Fiske et al., 2007)。此结果支持可信性优先假说:可信性在特质推断中起决定性作用,无论支配性高低,只要可信性高,其特质组合均可被接纳。消极晕轮效应假说难以解释这一统合方式,因已有研究表明,热情无能力(高可信性低支配性)并不会引发社会排斥(Rudert et al., 2017)。此像素回归模型解释力适中($R^2 = 0.34$),详见图6。

综上,结果一致表明,无论社会排斥心理表征还是社会接纳心理表征,均受可信性与支配性的并行影响;但相比于支配性,可信性均发挥关键预测作用,且作用权重更大。假设3得到证实。

3.4 研究2b: 面孔特质对社会排斥心理表征作用的主观评估

研究2b采用特质词汇评估方式,将研究2a中的心理表征图像转变为问卷调查内容。通过构建4种单一特质(可信性:高 vs.低;支配性:高 vs.低)对社会排斥心理表征的线性回归模型,探讨各特质对社会排斥评分的预测效应,从而进一步验证假设3。

3.4.1 研究方法

被试 本研究招募了153名大学生($N_{女} = 105$),平均年龄为 19.75 ± 0.97 岁。采用G*Power 3.1计算样本量(Faul et al., 2007)。在 $\alpha = 0.05$,统计检验力 $1 - \beta = 0.95$ 下,89名被试就可以达到线性回归分析的中等效应量(Cohen' $f = 0.15$; Linear multiple regression: Fixed model, R^2 increase)。因而,本次样本量满足抽样要求。

图片刺激及设备 采用研究1中生成的被排斥者及被接纳者2张面孔心理表征图像及研究2a预实验中生成的4种特质类型面孔心理表征图像(低可信性、高可信性、低支配性、高支配性),总计6张图片作为评估的刺激材料。

图片刺激使用交互智能平板中Office办公软件PowerPoint播放。设备型号为希沃F86EA,屏幕刷新率为60 Hz。分辨率设置为 1920×1080 像素。

调查问卷

(1)社会排斥评估 采用7点计分方式($-3 =$ 被人排斥, $+3 =$ 被人接纳)对社会排斥与社会接纳心理表征图像的社会排斥程度进行了评估。组内一致性系数 $ICC = 0.93$, 95% CI为 $[0.73, 1.00]$,一致性情况较为理想。

(2)特质评估词汇 研究2b所用特质评估条目来源于国内学者的面孔心理表征研究(刘志军, 2021)。可信性特质的评估词汇为“可信、友好、诚实”,而支配性特质的评估词汇为“智力、技能、才干”。所有词汇均采用李克特5点计分法进行评估。

词汇评估在不同心理表征条件下的内部一致性信度如下:可信性评估的Cronbach's α 在低可信、高可信、社会排斥和社会接纳条件分别为0.67、0.51、0.63与0.52;支配性评估的Cronbach's α 在低支配、高支配、社会排斥和社会接纳条件分别为0.54、0.52、0.57与0.56。各维度信度系数均高于0.50,达到中等可接受标准(Sutherland et al., 2020),但仍低于通用阈值0.70(Hussey et al., 2025)。因而,研究2b采用主成分分析生成的合成分数执行后续线性回归分析。

评估流程 问卷测试开始前,被试需签署知情同意书。随后,主试介绍两个基本特质维度(可信性和支配性)相关知识,并澄清测试中涉及的6个特质词汇(如“可信”“友好”“诚实”“智力”“技能”“才干”)的含义,确保被试理解无误。整个问卷测试分为两个阶段:特质词汇评估和社会排斥评估。

(1)特质词汇评估阶段。为避免两个特质维度间的相互影响,被试对每张面孔仅进行单一维度(可信性或支配性)的3个词汇的5点评分(1 = 非常不同意, 5 = 非常同意)。6张刺激图片共需完成8次评估任务:其中4张单一特质心理表征图像每次仅执行对应维度的评估;而2张社会排斥和社会接纳的心理表征图像则分别按两个基本特质维度,各需执行2次评估任务。在呈现顺序上,实验组预先将面孔列表与特质评估问题匹配,并将面孔顺序随机化,确保每次播放的面孔与纸质特质评估内容一致。

(2)社会排斥评估阶段。特质词汇评估完成后,屏幕上依次呈现社会接纳者和社会排斥者的2张心理表征图像,要求被试根据“ $-3 =$ 被人排斥, $+3 =$ 被人接纳”的7点评分标准,分别完成社会排斥评估。

整个测试过程大约15分钟完成。153名被试均在同一测试地点,分3批次完成调查。无论是否配合完成所有评估任务,被试均会获得一份礼品作为酬谢。

3.4.2 研究结果

评估检验 配对样本 t 检验结果显示,被试能够显著区分社会排斥($M = -1.33$, $SD = 0.71$)与社会

接纳($M = 2.76, SD = 0.86$)心理表征图像, 两组差异显著, $t(152) = 46.16, p < 0.001, \text{Cohen's } d = 3.73$ 。这一结果表明, RCIC 生成的社会排斥与社会接纳的心理表征图像结果较为理想, 符合人们的预期(刻板)印象。

面孔特质对社会排斥心理表征预测的线性回归分析 研究采用线性回归方法, 以 4 种单一-特质类型的合成分数(高可信性、低可信性、高支配性、低支配性)作为预测变量, 分别构建了针对社会排斥和社会接纳心理表征评估的线性回归预测模型。相关分析结果见表 1。

线性回归分析结果显示: 在社会排斥心理表征中, 低可信性特质发挥显著且最大的正向预测效应($\beta = 0.21, t = 7.41, p < 0.001, 95\% \text{ CI} = [0.15, 0.27]$); 低支配性同样具有显著的正向预测作用($\beta = 0.16, t = 4.75, p < 0.001, 95\% \text{ CI} = [0.09, 0.22]$)。这表明低可信性和低支配性并行影响面孔社会排斥心理表征, 且低可信性作用的权重更大, 此种统合方式(冷漠无能)可诱发面孔社会排斥, 重复了研究 2a 的发现。在社会接纳心理表征中, 仅高可信性表现显著的正向预测效应($\beta = 0.44, t = 28.13, p < 0.001, 95\% \text{ CI} = [0.41, 0.47]$); 其它特质类型的预测作用均不显著($ps > 0.05$), 这表明在社会接纳的心理表征中, 高可信性的作用更为关键(见表 2)。其中, 低支配性的预测效应几乎为 0, 意味着高可信性低支配性(热情无能)的统合方式基本不受消极晕轮效应的影响, 结果较符合可信性优先假说。两个线性回归模型的拟合情况良好, 变异解释比例均超过 70%。

综上, 无论社会排斥心理表征还是社会接纳心理表征, 相比于支配性, 可信性均发挥关键预测作用, 其作用权重更大。假设 3 再次得到证实。

3.5 辅助元分析

如果元分析结果支持研究假定, 其信度将高于单一研究的精确重复。基于此, 本研究以皮尔逊相关系数 r 作为效应量指标, 采用元分析方法对特质维度与社会排斥心理表征关系结果的可靠性进行检验。

本研究采用 CMA 3.3 (Comprehensive Meta-Analysis Version 3.3)元分析软件进行随机效应模型分析。结果显示, 特质维度与社会排斥心理表征的总效应值 $r = 0.51, 95\% \text{ CI} = [0.023, 0.800]$, 且不含 0。这表明特质维度与社会排斥心理表征存在较强相关。异质性效应 $Q = 471249, I^2 = 99.999, p < 0.001$, 超过了 Higgins 等人(2003)设定的 75%限制, 表明结果的异质性很高, 采用随机模型是合理的。同时, 这一高异质性也表示可能存在其他因素导致研究之间的估计值产生差异。因而, 需对影响两者间关系的调节变量进行合理探讨。

在此基础上, 本研究进一步以可信性和支配性为特质维度的亚组变量, 分别进行了亚组分析。结果显示, 可信性维度随机效应模型的总效应量 $r = 0.61, 95\% \text{ CI} = [0.003, 0.891], k = 4$, 不包括 0; $z = 1.97, p = 0.049$ 。这表明可信性维度与社会排斥心理表征存在较强相关。相比之下, 支配性维度随机效应模型的总效应量 $r = 0.39, 95\% \text{ CI} = [-0.534, 0.887], k = 4$, 包括 0; $z = 0.80, p = 0.43$ 。这表明支配性与社会排斥心理表征之间的关系并不显著。

表 1 社会排斥与社会接纳心理表征与单一-特质评估的相关情况($n=153$)

单一-特质维度	社会排斥	社会接纳	低可信性	低支配性	高可信性	高支配性
低可信性	0.82***	0.08	—			
低支配性	0.76***	-0.09	0.78***	—		
高可信性	0.02	0.92***	0.03	-0.15	—	
高支配性	-0.10	0.17*	-0.06	-0.06	0.15	—

注: * $p < 0.05$, *** $p < 0.001$, 下同。

表 2 单一-特质类型对社会排斥及社会接纳的回归预测情况($n=153$)

预测变量	社会排斥心理表征				社会接纳心理表征			
	β	SE	Beta	t	β	SE	Beta	t
截距	-1.33	0.031		-42.82***	2.76	0.027		103.86***
低可信性	0.21	0.028	0.54	7.41***	0.03	0.024	0.05	1.03
低支配性	0.16	0.033	0.35	4.75***	0.00	0.028	0.00	0.03
高可信性	0.03	0.018	0.07	1.49	0.44	0.016	0.92	28.13***
高支配性	-0.03	0.021	-0.06	-1.30	0.02	0.018	0.04	1.38
	$R = 0.84 \quad R^2 = 0.71 \quad \text{Adj } R^2 = 0.70$				$R = 0.93 \quad R^2 = 0.86 \quad \text{Adj } R^2 = 0.85$			

元分析结果再次证实,可信性维度是诱发面孔社会排斥心理表征的关键因素,假设3再次得到验证。不过遗憾的是,尽管可信性维度与社会排斥的较高相关均来自于低可信性($r_{\text{study1}} = 0.88$; $r_{\text{study2}} = 0.82$),但由于研究数据有限(均少于3个),难以进一步通过亚组分析确定低可信性所发挥的作用。

4 综合讨论

社会排斥作为一种普遍存在的现象,可由多种因素引起,心理表征可能是其中之一。本研究考察了社会排斥来源中基于面孔信息的心理表征来排斥他人的可能性。面孔作为社会交往中不可或缺的信息来源,蕴含着个体心理状态(如合作或伤害意图)的丰富特质信息(Todorov et al., 2008)。人类已经进化至非常擅长读取这些特质线索,可能已形成心理表征,并依靠其施以偏见或歧视等(Dotsch et al., 2008)。

4.1 面孔社会排斥心理表征的特质特征

当前研究显示,在社会排斥情境中,即使没有呈现被试面孔,人们也能对排斥目标的预期面貌形成社会共识。本研究采用RCIC技术揭示出这个心理表征的面孔直观形象(或刻板印象)。在特质归因理论框架下,社会排斥的心理表征被解读为包含丰富的可信性与支配性面孔信息,这些信息涉及两种特质的关键诊断信息区域。这可能代表着人们对社会排斥的谨慎态度,即只有当面孔线索与心理表征特质信息充分匹配时,人们才倾向于排斥他人。此外,社会排斥与社会接纳的心理表征形象存在明显不同,两者差别主要来自于前者的眉毛、鼻子及嘴巴区域。其中,嘴巴是面孔中所占面积最大,最具区分性的区域。嘴巴区域在心理表征中传递了可信性的关键信息,帮助人类决定是否信任他人(Mo et al., 2022)。这表明人们在社会排斥上,更关注于可信性维度上的表现(尤其是不值得信任方面)。这或可解释此前研究中,遭到社会排斥后,为何排斥目标会表现出亲社会行为。这是因其感受到了他人对自身信任方面的否定(排斥),为重建社交关系而表现出来的行为(Chen et al., 2024)。但心理表征中的嘴巴区域既可能是特质推论结果,也不排除是视觉系统认知能力有限时对信息量最多区域的优先加工的结果。

4.2 面孔社会排斥心理表征的特质推断模式

为进一步探究面孔特质信息在社会排斥过程中如何发挥作用,本研究先后通过客观测量及主观

评估的方式,考察了社会排斥心理表征中的特质信息来源。考虑到面孔特质维度之间具有相关性,单独研究面孔某一维度的结果解释力有限(Oosterhof & Todorov, 2008),可信性与支配性共同编码在面孔心理表征中(Gunaydin & DeLong, 2015),因而本研究以特质统合方式探索了社会排斥心理表征的特质推断模式。两种不同研究方式一致发现,不可信在预测社会排斥过程中发挥了最关键作用,且低可信低支配(即冷漠无能)心理表征是社会排斥的典型特征,这符合人们对排斥目标的刻板认知结果。人们容易排斥此种面孔特征的个体且不受自身道德推断的影响(Fiske, 2015; Rudert et al., 2017)。稍显不同的是,客观测量结果还显示,低可信高支配也符合社会排斥心理表征的特征;相对的,高可信高支配及高可信低支配则是社会接纳心理表征的特征,这为此前社会排斥及社会接纳的相关发现提供了证据支持(Fiske et al., 2007; Rudert et al., 2017),这有助于拓展社会认知内容的理论解释范围。

4.3 理论贡献

面孔特质推断如何诱发社会排斥的议题,目前学术界推断这与人们头脑中对排斥目标形成的社会共识性的面孔刻板印象有关(Rudert et al., 2020)。要推动这一议题的实证化,关键在于如何有效捕捉并量化面孔特质的心理表征。我们的研究推进了特质归因模型(可信性和支配性)在面孔排斥领域的应用,推动了心理表征在社会排斥研究中的发展。

针对特质推断如何诱发社会排斥的议题,目前存在两种较为典型的理论解释。一种为可信性优先假说,另一种为消极晕轮效应假说。本研究证据表明可信性优先假说或许提供了更为适宜的解释途径,排除了消极晕轮效应假说的理论解释力。两者特质以统合方式,并行作用于特定面孔内容,且遵循权重不等原则,可信性比支配性发挥更大权重作用,也即可信性优先。虽然可信性与支配性都可以自发地编码在心理表征中,但可信性是核心维度,是重要的决定社会排斥的区分性证据(Oosterhof & Todorov, 2008)。这种优先级排序反映在社会情境中,个体对可信性的判断要比支配性更早,提取可信性线索的能力比支配性更迅速准确(Todorov et al., 2009)。这意味着当我们决定是否可以信任他人时,就将陌生人面孔与可信性心理表征进行匹配,从而推断出是否应该接近或回避他人。其中,低可信性推断结果会导致难以克服的怀疑感,使低可信个体更易受到排斥(Vandewouw et al., 2020)。

综上,本研究的理论贡献在于:揭示了诱发社会排斥的面孔心理表征以(不)可信性为关键线索,从而为“可信性优先假说”提供了实验性证据,并揭示消极晕轮效应假说的解释力有限。本研究把可信性优先机制从言语材料扩展到非言语面孔刺激,为理解高级社会认知层面的排斥过程提供了新的理论依据。

4.4 方法学贡献

研究引入基于随机噪音的 RCIC 技术,改进了面孔社会排斥心理表征的研究方法。该技术凭借纯粹数据驱动的原理,结合工程、神经生理系统识别(如 fMRI)与心理物理学、实验心理学及计算机科学的跨学科优势,突破了传统心理物理学在呈现认知内容上的局限(Jack & Schyns, 2017)。基于随机噪音的 RCIC 技术提升了生态效度,实现了面孔社会排斥心理表征的直观可视化,并通过像素数据计算出心理推断的诊断信息区域(Oliveira et al., 2019)。此技术促进了面孔特质与社会排斥关系的理解。

本研究通过采用客观和主观的方法学措施,为研究结果提供了良好的聚合效度。客观测量方法借助 RCIC 技术,不受研究者先验假设影响,获得社会排斥及单一特质心理表征的可视化形象,并直接通过对这些图像的像素回归关系预测出不同特质维度对社会排斥的作用情况。整个过程并不受实验中的外显因素影响。而主观评估方法为其发现提供了有益佐证。最终研究一致发现:不可信是影响社会排斥心理表征的关键特质维度,而“冷漠无能”的个体极易受到排斥。

4.5 现实意义

心理表征并非固定不变,其受多种因素的影响且可通过特定手段进行塑造和调整。面孔特质推断也受到文化环境的影响。因此,通过积极的文化传播和教育,不仅可优化心理表征的构建过程,还可以直接改善刻板印象本身,从而促进社会的包容与和谐。

此外,心理表征也可通过特定的训练手段进行塑造和调整。如 Soto 等人(2020)发现,分类训练可用于改善自闭症谱系障碍患者在面孔识别或特质方面的表现。这种训练通过改变个体的心理表征,使其在不相关维度的变化中变得更加稳定,提升了其社会认知能力(Soto et al., 2020)。这表明,通过有针对性的训练可改变个体对社会信息的心理表征方式,进而削弱或消除某些社会排斥。

4.6 局限与不足

首先,结论的跨文化普适性仍待验证。虽然个体从情境的各个方面获取社会排斥的具体线索,并依靠这些线索积极地构建他们对社会排斥的解释。但可信性信息本身具有文化差异性,当被试做出明确的信任判断时,他们往往更喜欢具有自己文化特征的面孔(Sofer et al., 2017),且中国人的可信性心理表征与西方人存在明显不同(Mo et al., 2022)。因而,未来仍需要来自跨文化研究的证据支持。

其次,尽管本研究展示了人们头脑中是如何以心理表征的形式来定义社会实体(被排斥者形象),却未能充分揭示社会排斥的心理表征形成过程。这是因为 RCIC 反映的是“自上而下”的加工形式,虽可提供整个刺激可视区域的诊断信息结果,却缺失视觉信息“自下而上”加工情况。因而,未来可尝试借助眼动追踪技术获得面孔视觉注视区域的加工结果,其与心理表征诊断区域的重叠或分离情况将有助于了解社会排斥的心理表征机制如何。这对于人们全面理解社会排斥的原因,制定出有效的干预措施无疑是有益的。

此外,现有研究将支配性等同于 SCM 能力维度仍有分歧。然而当前缺乏专门的支配性测评工具,学界仍沿用能力词汇进行测量可能引入概念偏差。后续研究需开发更契合支配性内涵的测量方案,重复本研究的发现。

5 研究结论

(1)面孔社会排斥的心理表征形象包含可信性与支配性特质信息,其中鼻子和嘴巴(可信性)、眉毛(支配性)是核心诊断信息区域。

(2)低可信、低支配心理表征是社会排斥的典型特质特征,在社会排斥的特质推断过程中两者均发挥关键作用。其中低可信性作用权重更大,支持了可信性优先假说。

参 考 文 献

- Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, 50(12), 2830–2838.
- Casini, E., Glemser, C., Premoli, M., Preti, E., & Richetin, J. (2022). The mediating role of emotion regulation strategies on the association between rejection sensitivity, aggression, withdrawal, and prosociality. *Emotion*, 22(7), 1505–1516.
- Chattalas, M. J. (2005). *The effects of national stereotypes on country of origin-based product evaluations*. New York: City University of New York.
- Chen, F., Guo, T., & Wang, J. (2024). Divergent effects of

- warmth and competence social rejection: An explanation based on the need-threat model. *Journal of Personality and Social Psychology*, 126(3), 461–476.
- Chen, L., Zeng, S., & Su, Y. (2023). The influence of social exclusion on adolescents' social withdrawal behavior: The moderating role of connectedness to nature. *Journal of Environmental Psychology*, 87, 101951.
- Chen, Z., Du, J., Xiang, M., Zhang, Y., & Zhang, S. (2017). Social exclusion leads to attentional bias to emotional social information: Evidence from eye movement. *PLOS ONE*, 12(10), e0186313.
- Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., ... Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1–33.
- Dickerson, K. L., & Quas, J. A. (2024). Compensatory prosocial behavior in high-risk adolescents observing social exclusion: The effects of emotion feedback. *Journal of Experimental Child Psychology*, 241, 105840.
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562–571.
- Dotsch, R., Wigboldus, D., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, 19(10), 978–980.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.
- Giacomin, M., Tskhay, K. O., & Rule, N. O. (2022). Gender stereotypes explain different mental prototypes of male and female leaders. *The Leadership Quarterly*, 33(6), 101578.
- Gunaydin, G., & DeLong, J. E. (2015). Reverse correlating love: Highly passionate women idealize their partner's facial appearance. *PLOS ONE*, 10(3), e0121094.
- Hales, A. H., Wesselmann, E. D., & Williams, K. D. (2016). Prayer, self-affirmation, and distraction improve recovery from short-term ostracism. *Journal of Experimental Social Psychology*, 64, 8–20.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560.
- Hou, C. N. (2017). *Faces: The evolutionary code of intergroup trust*. Beijing: Science Press.
- [侯春娜. (2017). *面孔: 群际信任的进化密码*. 北京: 科学出版社]
- Hou, C. N., & Liu, Z. J. (2019). Visualization of mental representation: Noise-based reverse correlation image classification technology. *Advances in Psychological Science*, 27(3), 465–474.
- [侯春娜, 刘志军. (2019). 心理表征的可视化途径: 基于噪音的反向相关图像分类技术. *心理科学进展*, 27(3), 465–474.]
- Hussey, I., Alsalti, T., Bosco, F., Elson, M., & Arslan, R. (2025). An aberrant abundance of Cronbach's alpha values at .70. *Advances in Methods and Practices in Psychological Science*, 8(1), 25152459241287123.
- Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, 68(1), 269–297.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Liu, Z. J. (2021). *A study on the facial compensatory effect of stereotype in intergroup cognition*. Beijing: Social Sciences Academic Press.
- [刘志军. (2021). *群际认知的刻板印象: 面孔补偿效应*. 北京: 社会科学文献出版社]
- Ma, F., Xu, F., & Luo, X. (2015). Children's and adults' judgments of facial trustworthiness: The relationship to facial attractiveness. *Perceptual and Motor Skills*, 121(1), 179–198.
- Mo, C., Cristofori, I., Lio, G., Gomez, A., Duhamel, J. R., Qu, C., & Sirigu, A. (2022). Culture-free perceptual invariant for trustworthiness. *PLOS ONE*, 17(2), e0263348.
- Okazawa, G., Sha, L., Purcell, B. A., & Kiani, R. (2018). Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nature Communications*, 9(1), 3479.
- Oliveira, M., Garcia-Marques, T., & Dotsch, R. (2019). Combining traits into a face: A reverse correlation approach. *Social Cognition*, 37(5), 516–545.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092.
- Over, H., & Uskul, A. K. (2016). Culture moderates children's responses to ostracism situations. *Journal of Personality and Social Psychology*, 110(5), 710–724.
- Park, J., & Baumeister, R. F. (2015). Social exclusion causes a shift toward prevention motivation. *Journal of Experimental Social Psychology*, 56, 153–159.
- Rudert, S. C., Keller, M. D., Hales, A. H., Walker, M., & Greifeneder, R. (2020). Who gets ostracized? A personality perspective on risk and protective factors of ostracism. *Journal of Personality and Social Psychology*, 118(6), 1247–1268.
- Rudert, S. C., Reutner, L., Greifeneder, R., & Walker, M. (2017). Faced with exclusion: Perceived facial warmth and competence influence moral judgments of social exclusion. *Journal of Experimental Social Psychology*, 68, 101–112.
- Schmitz, M., Vanbeneden, A., & Yzerbyt, V. (2024). The many faces of compensation: The similarities and differences between social and facial models of perception. *PLOS ONE*, 19(2), e0297887.
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5), 391–396.
- Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H., & Todorov, A. (2017). For your local eyes only: Culture-specific face typicality influences perceptions of trustworthiness. *Perception*, 46(8), 914–928.
- Soto, F. A., Escobar, K., & Salan, J. (2020). Adaptation aftereffects reveal how categorization training changes the encoding of face identity. *Journal of Vision*, 20(10), 18, 1–24.
- Sutherland, C. A., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257–267.
- Sutherland, C. A., Rhodes, G., Burton, N. S., & Young, A. W. (2020). Do facial first impressions reflect a shared social reality?. *British Journal of Psychology*, 111(2), 215–232.

- Syrjämäki, A. H., & Hietanen, J. K. (2020). Social inclusion, but not exclusion, delays attentional disengagement from direct gaze. *Psychological Research*, *84*(4), 1126–1138.
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545.
- Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces. *IEEE Signal Processing Magazine*, *28*(2), 117–122.
- Todorov, A., Pakrashi, M., & Oosterhof, N.N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460.
- Vandewouw, M. M., Choi, E., Hammill, C., Arnold, P., Schachar, R., Lerch, J. P., ... Taylor, M. J. (2020). Emotional face processing across neurodevelopmental disorders: A dynamic faces study in children with autism spectrum disorder, attention deficit hyperactivity disorder and obsessive-compulsive disorder. *Translational Psychiatry*, *10*(1), 375.
- Wang, Q., Hoi, S. P., Wang, Y., Song, C., Li, T., Lam, C. M., ... Yi, L. (2020). Out of mind, out of sight? Investigating abnormal face scanning in autism spectrum disorder using gaze - contingent paradigm. *Developmental Science*, *23*(1), e12856.
- Wesselmann, E. D., Bagg, D., & Williams, K. D. (2009). “I Feel Your Pain”: The effects of observing ostracism on the ostracism detection system. *Journal of Experimental Social Psychology*, *45*(6), 1308–1311.
- Wirth, J. H., & Wesselmann, E. D. (2018). Investigating how ostracizing others affects one’s self-concept. *Self and Identity*, *17*(4), 394–406.
- Wyer, N. A., & Schenke, K. C. (2016). Just you and i: The role of social exclusion in the formation of interpersonal relationships. *Journal of Experimental Social Psychology*, *65*, 20–25.

The mental representation and inference patterns of facial social exclusion

HOU Chunna¹, MA Yisheng¹, WU Lin², LIU Zhijun³

(¹ School of Psychology, Northeast Normal University, Changchun 130024, China)

(² School of Sociology, Wuhan University, Wuhan 430072, China)

(³ Department of Sociology, Changchun University of Science and Technology, Changchun 130022, China)

Abstract

As social beings, humans rely on social relationships for survival and reproduction. Successful adaptation to the social environment leads to acceptance, while failure may result in social exclusion. The mechanisms of exclusion behavior are a key academic focus. Unjustified exclusion not only threatens the mental health of the excluded, but may also cause distress to the excluders. Rudert et al. (2017) emphasized that facial features significantly influence decisions about exclusion. Based on the Big Two model of trait attribution, this study employed reverse correlation image classification techniques to explore the facial mental representations and trait inference patterns that trigger social exclusion.

Study 1 involved 81 Chinese college students as participants and employed a two-image forced-choice task after inducing a social exclusion scenario. The results revealed differences in the mental representations of faces between those who were excluded and those who were accepted, as well as varying diagnostic criteria. Pixel test analysis showed that the diagnostic information for social exclusion included most of the key areas of the face: the forehead, eyes, nose, mouth, and the peripheral facial contour.

Study 2 examined the role of different facial traits in social exclusion, utilizing both objective measurements and subjective evaluations. Study 2a adopted an objective measurement approach, with the independent variables being the different trait-based mental representations generated in a pilot study (high trustworthiness, low trustworthiness, high dominance, low dominance), and the dependent variables being the two types of facial mental representations (social exclusion and social inclusion) identified in Study 1. A pixel regression model using pixel brightness values analyzed facial traits to predict the mental representation of social exclusion. The pixel regression analysis revealed that low trustworthiness strongly and positively predicted social exclusion, with the largest effect among all traits. Trustworthiness is a key dimension in shaping the mental representation of social exclusion. High trustworthiness had a negative predictive effect, indicating that the trustworthiness dimension has a discriminative impact on social exclusion. However, the discriminative

effect of the dominance dimension is less clear. Study 2b conducted a subjective evaluation of the inference patterns of facial traits in the social exclusion process. The study utilized six facial mental representations as stimuli, including the two images of excluded and included individuals generated in Study 1, as well as the four images representing low trustworthiness, high trustworthiness, low dominance, and high dominance generated from Study 2a. A total of 153 college students participated in the study. The results showed that for the mental representation of social exclusion, low trustworthiness had a significant positive predictive effect, and low dominance similarly contributed to social exclusion.

Both objective and subjective studies indicated that trustworthiness was a key predictor of social exclusion, while the role of dominance remains inconclusive and warrants further investigation. Both assessments indicated that low trustworthiness and low dominance were typical features of social exclusion. In the trait inference process underlying social exclusion, both trustworthiness and dominance play crucial roles, but low trustworthiness carries greater weight, thereby supporting the primacy of trustworthiness hypothesis.

Keywords face, social exclusion, mental representation