


# 情感推理中的面部和上下文整合是有限的，并且在不同类别和个人之间存在差异

收稿日期：2022 年 10 月 6 日

接受日期：2024 年 3 月 5 日

斯里什蒂·戈尔<sup>1</sup> 朱利安·贾拉-埃廷格<sup>1,2</sup> 德斯蒙德·C·翁<sup>3</sup> & 玛丽亚·詹德伦<sup>1</sup> 


Published online: 19 March 2024

 检查更新

对他人情绪状态做出细致入微的推断的能力是社会功能的核心。虽然情绪推断可能对面部动作及其发生的情境背景都很敏感，但人们对这两种信息来源何时跨情绪类别和个人进行整合的了解相对较少。在一系列研究中，我们使用一个档案和五个经验数据集来证明人们可以进行整合，但仅通过对情况的了解也可以（有时更好）捕获情感推断，而孤立的面部线索是不够的。此外，人们更多地将面部线索融入日常生活中最常遇到的面部表情类别（例如幸福）。随着时间的推移，人们对情境线索和线索整合的依赖也保持适度稳定，并且那些更可靠地利用情境线索的人也拥有更好的情境情感知识。这些发现强调了研究线索依赖和整合的变异性的的重要性。

人们的看法、决定和行动往往取决于对他人思想的理解，包括他们的情感体验。例如，人们根据对犯罪者的悔恨或内疚的推断做出道德和法律判断，并向被视为痛苦或悲伤的人提供支持。在谈判中，人们会向被认为愤怒的对方做出更多让步。情感推理得出的结果对各种社会环境都有影响，从亲密关系到法庭，从课堂到全球外交。此外，一个蓬勃发展的行业预计在未来几年内将增长至 371 亿美元，其核心是构建智能机器，这些机器可以使用非语言行为来“解读”人们的内心状态（例如情绪），从而预测他们的行为。除了认知表现之外，区分智能机器和非智能机器的许多定义还在于情感能力，特别是当目标是实现类人表现时

情感科学在日常生活和工业中的这些应用表明，对于人类如何推断他人的情感建立系统且可靠的解释至关重要。尽管现实世界中的情感事件非常复杂，但对情感推断的研究却过分关注人们如何处理孤立的规范情感面部描绘（或面部表情）。这种被称为“共同观点”的方法在基础科学、临床科学、教育、技术部门、安全以及大众媒体和娱乐中普遍存在（审查参见参考文献 14）。具体来说，无论是在理论上还是在实践中，研究人员经常假设情绪类别和面部行为表现出稳定的、与上下文无关的联系。例如，这种普遍观点假设存在一种愤怒的表达（即皱眉），这种表达在可能出现这些表达的特定上下文中是不变的（例如，在与浪漫伴侣的冲突中以及在交通堵塞时）。事实上，“情绪识别”领域是以这一假设为前提的，其中脱离语境的规范描述

1 耶鲁大学心理学系，地址：100 College St, New Haven, CT, USA。耶鲁大学吴蔡研究所，地址：100 College St, New Haven, CT, USA。德克萨斯大学奥斯汀分校心理学系，地址：108 E Dean Keeton St, Austin, TX, USA。电子邮件：srishti.goel@yale.edu；maria.gendron@yale.edu 

情感被呈现给感知者，“准确性”是根据响应是否与先验类别匹配来计算的。数十年的证据表明，这种关注面部的普遍观点是不够的。情绪推断比简单地依赖面部的一组有限线索要复杂得多。有证据表明，感知者严重依赖额外的情境信息，如身体动作、发声、文化知识、社会情境知识或情境情境来推断情绪（回顾，参见参考文献 15,25）。

尽管先前的研究证明了上下文在情感推断中的关键作用，但大多数研究都是有限的，因为它们通常关注某些线索的主导地位，而不是跨线索的整合以形成推论（有关整合的经典示例，请参阅参考文献 26）。也就是说，早期的工作通常旨在通过将线索（面部或情境）作为高度对比信息相互竞争来建立线索的主导地位（例如，参见参考文献 23,27）。在这种方法中，经常实验性地使用预先选择的“清晰”（意思是，在感知者的判断中引起高度一致）面部和情境线索（对于这种刺激方法清晰度的批评，请参见参考文献28），这可能限制了刺激更典型的表现（例如，因恐惧而睁大眼睛、气喘吁吁的表情）。此外，这种对比方法（例如，将“恐惧”面孔与“愤怒”情况结合起来）可能无意中鼓励参与者忽略一条信息而不是另一条信息，而不是将信息整合在一起。也就是说，当信息具有“对比性”时，可能只有基于单一线索的一种解释被认为是正确的，从而使判断成为该实验设置的伪影。因此，对比方法可能会限制与现实世界情感推断的相关性。本研究旨在使用计算建模方法来解决情感推断中的线索整合问题，该方法将线索整合模型与仅依赖于面部信息或情境信息的更简单模型进行比较。

为了研究感知者如何关注并结合面部信息和情境背景，有必要模拟人们如何理解情境和非语言行为之间的关系。也就是说，感知者可能不仅仅简单地同等依赖面部信息和上下文信息。相反，感知者可能会根据他们对情境、目标的情感体验和他们的表现出的非语言行为之间因果关系的外行信念，整合情境和非语言线索。最近的一些工作提高了我们对这种基于外行理论的集成进行建模的能力。具体来说，Ong 及其同事基于人类视觉感知中的其他理性观察者模型，提出了一个理性观察者模型，用于说明外行人如何整合多种情感线索。该模型假设观察者持有因果理论，其中情境结果导致情绪，进而导致面部表情（下面将更详细地描述）。该模型超出了普遍观点的假设。人们普遍认为，面部表情可以提供有关引发这些情绪的情绪的诊断信息。当情景结果可用时，这些结果应该只提供关于潜在情绪和从面部衍生的情绪的收敛、重叠的信息。因此，集成模型和仅基于面部的模型应该在相同程度上捕获情感判断。相反，如果线索整合模型比仅面部模型更大程度地捕获判断，则这意味着面部线索提供了有关情感的不完整信息，并且情境上下文提供了感知者用来推断情感的附加信息。鉴于这种外行理论本质上是一个有向无环图，应该注意的是，情境不会直接影响表达本身的形式：相反，情境仅通过情感本身影响表达。线索整合模型（方程式因此，从该 DAG 派生的（见下文 1）假设感知者利用与上下文无关的面部情绪表达的心理表征

符合前面概述的共同观点的一个方面。至关重要的是，在 Ong 及其同事的原创工作中，通过多次实验，贝叶斯线索整合模型密切跟踪感知者的判断，这表明人们可以并且确实整合了来自情境和面部的信息，这与这一理论相符。在这里，我们在这些建模进展的基础上提出了两个关于该模型捕获情感推断的稳健性的主要问题。

首先，现有文献中关注的面部动作的有限范围可能无法捕捉现实世界情感实例中面部动作的复杂性。情绪激动时的面部动作很少符合科学研究中常用的规范表达。先前的研究表明，当面部线索模棱两可（现实世界中经常如此）、高度可变、中性甚至完全缺失时，人们仍然可以根据上下文对情绪做出强有力的推断。这意味着当刺激具有更大的多样性和复杂性时，这种贝叶斯积分模型可能不适用于情感推断。因此，我们的目的是在这里测试贝叶斯积分帐户的稳健性。

我们询问，当使用更具可变性和复杂性的刺激时，与基于单一线索（仅面部或上下文）的推断相比，面部和情境线索的贝叶斯整合（基于参考文献 31 中的模型）是否能最好地解释感知者的情绪推断。之前的建模工作是使用一组狭窄的情况进行测试的：赌博游戏中的不同结果（赢得的钱数）或比赛表演网球比赛的结果（获胜或失败），有时还加上计算机生成的漫画面部肖像。因此，之前的工作留下了一个问题：基于贝叶斯 DAG 模型的面部和情境线索的整合是否可以扩展到更自然的变化刺激，从而更好地反映日常生活中体验和表达行为的多样性。在这项工作中，我们测试了现有的线索整合模型（之前称为整合）是否可以解释超过 2500 个感知者判断的数据集中的情绪推断，其中情况的情绪复杂性很高，并且表达行为高度可变，但也很高在感知强度中。基于证据表明，规范表达很少记录在自发产生的表达中，并且之前的研究表明，高质量的情感表演描绘很少涉及规范表达，我们认为，当前的数据集为感知者如何整合面部和情境提供了相对更生态有效的测试。

考虑到该数据集的复杂性，我们还能够检查整体情绪判断以及特定情绪类别（例如愤怒、恐惧等）的线索整合。当个人对给定线索和情绪之间的联系有更大的确定性时，他们可能更有可能参与整合。这种知识可以通过日常生活中遇到特定线索的频率进一步得到巩固。例如，人们报告在日常生活中看到“典型”面部表情的频率各不相同，例如通常与恐惧相关的睁大眼睛、喘气的表情，以及通常与愤怒相关的皱眉和紧闭的嘴唇。此外，最近的荟萃分析还表明，这些规范表达与实际情感体验的相关性很弱。对于可靠性低或不常遇到的情绪表达，感知者应该不太确定这些表达与情绪的联系。因此，他们可能更有可能依赖情境线索来形成推论，而不是依赖或整合面部线索。

其次，除了考察群体层面的整合之外，我们还将整合工作扩展到考察个体差异。在这里，我们询问个体的整合程度与仅依赖情境或面部线索的程度是否存在差异，以及这些差异是否随着时间的推移保持稳定。有理由怀疑情境线索的利用可能会受到个人的影响

差异。人们表达复杂情况含义的能力各不相同。人们不仅根据群体共识推断情境情绪后果的能力存在差异，而且由于这些变量表征而做出的行为反应选择也存在差异。在将健康成人与临床样本中的个体进行比较时，也观察到情况表现的变化（例如，参见参考文献 43,44）。例如，更具敌意的人更有可能将模棱两可的情况解释为威胁，而状态焦虑程度较高的人在身体危险、社会评价和冲突等多个维度上表现出强烈的压力情况。总之，研究结果表明，人们对社会情境的表述以及他们从中获得的意义可能存在很大差异。我们在这些发现的基础上，重点关注在形成情感推理时如何依赖情境背景的个体差异。

我们对个体间线索依赖的模式知之甚少，尽管这是一个潜在的重要个体差异。值得研究的是，量化情绪推断的单一模型是否足够，或者个体之间是否存在有意义的变异。为此，我们研究了线索整合中的个体差异是否与情绪领域个体差异的现有测量相关。具体来说，我们研究了人们准确推断（基于群体共识）在各种社交情境中最有可能经历的情绪的能力（情绪理解情境测试或 STEU）以及他们检测情境线索的能力（情境敏感性指数）。情感文献在很大程度上忽略了个体在情感推理中整合情境背景的差异。相反，对情绪推断中个体差异的研究（例如，在情绪智力文献中）通常侧重于从单一模态形成的推断，例如面部（例如，参见参考文献 47）或情境背景（例如，STEU）。即使向参与者提供多模态线索（例如，日内瓦情绪识别测试），研究人员也只关注有限的一组对比面部和情境线索的情绪推断情境调制程度的个体差异。在推断情绪和线索依赖程度时，个体整合互补的面部和情境线索的倾向是否存在差异仍然是一个悬而未决的问题。

我们通过使用档案和经验收集的数据集（N = 752）的五项研究来研究有关情感推理整合的这些问题。档案数据包含社会场景的评分、演员对这些场景的描绘的评分以及 13 种情感类别组合的评分——娱乐、愤怒、敬畏、蔑视、厌恶、尴尬、恐惧、幸福、兴趣、自豪、悲伤、羞耻、惊讶。这些评级是从总共 604 个刺激对（场景和姿势）中获得的，这些刺激对来自两本书：《角色中：演员表演》和《表演中：演员表演》。这些卷中包含了一群专业演员在接受了唤起情感的场景后所摆出的表情图像。为了计算模型预测，有必要计算人们对情绪的先验概率。我们的第一个实验数据包括人们 (N = 44) 在日常生活中感知 13 种情绪类别中每一种情绪的可能性（之前的任务）。第二个经验数据集（研究 2）包含对随机选择的刺激子集（n = 44）的面部描绘和社交场景组合的评分（N = 142），这是一种基于能力的个人对环境敏感度的衡量标准（CSI 量表），以及他们自我报告的区分情绪状态的能力 (RDEES)。第三个经验数据集（研究 3）是研究 2 的扩展，包含对先前使用的相同刺激子集的面部描绘和社交场景组合的评分（N = 162），这是一种基于能力的个人对环境敏感度的衡量标准（CSI 量表），以及他们理解情绪和情境之间关系的能力（STEU-B）

第四个经验数据集（研究 4）是一项预先注册的研究，旨在在全国代表性样本（N = 294）中复制和扩展研究 3 的结果。收集最终的经验数据集（研究 5）来评估模型估计中个体差异的重测可靠性，以检查它们的特征有多相似。该研究由参与者 (N = 110) 对研究 2-4 中使用的相同对 (n = 44) 面部描绘和情景描述进行评分。对 13 种情绪类别进行了两次评级——在第 1 节和第 2 节中，间隔两周进行。

结果

型号概览

与人们如何推理信念、欲望和意图等其他心理状态相比，贝叶斯计算模型在理解人们如何推理情感（或情感认知；参见参考文献 31,38）方面的应用相对较新。这些规范模型通过捕捉人们头脑中计算的不同假设并将其与人类判断进行比较（理性分析方法）来帮助形式化和测试人类的外行理论。

研究人员最近使用这种方法来捕捉一种简单的情绪理论，他们建议人类用这种理论来理解他人的情绪。该理论捕捉了对情境结果如何引起情绪的期望，以及情绪如何引发面部表情等表达行为。有了这个因果模型，人们就可以使用贝叶斯推理来确定人们可能正在经历的情绪，给定有关他们的面部表情和所处情况的信息。Ong 和同事（2015）提出了这个想法的正式计算模型并将其扩展为推导出一个模型方程，该模型方程捕捉面部和情境线索整合（或线索整合）后的情绪推论，由下式给出：

$$P(e|s, f) = \frac{P(e|s)P(f|e)}{P(f|s)}$$

这里， $P(e|s, f)$  代表观察者对代理人正在经历情绪  $e$  的信念，考虑到他们的面部表情  $f$  和情况结果  $s$ 。如果人们通过将贝叶斯推理应用于情感的外行理论来推断彼此的情感，那么人们的推断应该与每个单独线索的可能性  $P(e|s)$ （给定情况  $s$  的情感概率）和  $P(f|e)$ （给定面部表情  $f$  时情绪  $e$  的概率除以情绪发生的先验概率  $P(e)$ 。有关该方程的完整推导，请参见 Ong 及其同事 (2015)。我们通过平均参与者的情绪强度评级来计算整体模型概率。这假设频率评级的分布可用于计算模型概率，这与先前的文献一致，证明人们的推论反映了概率匹配行为，并且人们充当贝叶斯采样者。

Ong 及其同事将贝叶斯线索整合模型与两个更简单的替代模型进行了比较，这两个模型捕捉了关于依赖单一线索的信念，即线索优势模型，而不是整合。这些代表了这样一种信念，即代理人仅因情境结果而体验情绪（仅限情境模型或  $P(e|s)$ ），并且相信代理人的情绪体验可以仅从面部动作等外部线索推断出来（仅限面部模型，或  $P(f|e)$ ）。这些更简单的模型表明，当观察者同时看到面孔和情况时，他们会依靠单一线索（面孔或情况）来近似推断他们的情绪。

在这种方法中，将模型估计与人们在查看面部和情境线索时的情绪判断进行比较，以检查哪种模型总体上最好地捕捉人们的推论（图1a）。在这里，我们不仅总体上检查了最佳拟合模型，而且通过比较个体在观看两种情况时的情绪判断，将模型比较扩展到个体层面。





三个规范模型预测估计的面部和情境线索（图 1b）。我们在研究 1 中计算了仅面部、仅情境和贝叶斯线索整合模型估计，并将其与研究 1-5 中人们通过查看面部和情境线索而做出的情绪判断进行比较（根据研究 2-5 收集的经验）。作为稳健性检查，我们收集了一组针对仅面部和仅情况条件的附加评级，以确保采样变化不会显着影响这些模型拟合。我们发现评级的分布具有高度可比性（参见补充图S1-S3），并且使用附加评级集的模型拟合是一致的（参见补充表S1）。

情绪推断是否反映了贝叶斯线索整合？

在研究 1 中，我们使用更多样化和更复杂的刺激集检查了线索整合的贝叶斯模型的稳健性。此外，我们检查了线索整合模型对特定情绪类别（例如愤怒、恐惧等）的适用性。在本研究中，我们使用了档案数据集并收集了实验数据集（使用先验任务）。档案数据集包括参与者对社会场景描述的评分（仅场景条件，N = 839）、演员对这些场景的描绘的评分（仅面部条件，N = 842）以及场景组合的评分和面（jointcue 条件，N = 845）。每个参与者对从 604 个更大的集合中获得的大约 30 种刺激的随机子集提供了 13 种情绪类别（娱乐、愤怒、敬畏、蔑视、厌恶、尴尬、恐惧、幸福、兴趣、骄傲、悲伤、羞耻、惊讶）的评分，在三个条件之一。我们计算了仅面部  $P_{ejf\delta P}$  和仅情境  $P_{\delta ejsP}$  模型概率，方法是对每个情绪类别和刺激的人们的评分进行平均，然后对每个刺激的情绪进行标准化，使得每个刺激的情绪类别的概率总和等于1。

为了计算贝叶斯线索整合模型概率  $P_{\delta ejs, fP}$ ，除了上述仅面部和仅情境模型估计之外，我们还需要估计人们推断情绪  $P_{\delta eP}$  的先验期望。我们通过要求参与者 (N = 45) 评估他们在日常生活中感知 13 种情绪类别的可能性来收集先验数据。我们通过标准化每个情绪类别的平均评分来计算情绪的先验概率  $P_{\delta eP}$ ，使得十三个情绪类别的先验概率之和等于 1。然后，我们计算每个刺激和情绪类别的贝叶斯线索整合模型概率估计：在等式中插入仅面部模型概率、仅情境模型概率和情感先验概率。1. 线索整合模型的概率也被标准化，使得每个刺激的情绪类别的概率之和等于 1。

我们首先研究了线索整合模型是否能够最好地捕捉人们从这些更加多样化和复杂的面部和情境线索中推断出的情绪。我们将三种模型估计（仅面部、仅情境和贝叶斯线索整合）与经验线索整合概率进行了比较，这些概率是通过对联合线索条件下每个刺激的评分进行平均而计算出来的（经验线索整合）。一体化）。归一化经验提示整合概率与模型概率的双尾 Pearson 相关性表明，尽管三个模型中的每一个都与经验判断具有显著且强的相关性，但仅情景模型具有最高的相关性 ( $r(7850) = 0.865$ ， $t = 153.04$ ， $p < 0.001$ ，自举 95% CI: 0.857, 0.873)，然后是贝叶斯线索整合模型 ( $r(7850) = 0.840$ ， $t = 137.15$ ， $p < 0.001$ ，自举 95% CI: 0.830, 0.849)，最后是仅面部模型 ( $r(7850) = 0.660$ ， $t = 77.942$ ， $p < 0.001$ ，自举95% CI: 0.644, 0.677)（图2a）。虽然仅情境模型相关性和贝叶斯线索整合模型相关性存在统计显著性差异 ( $t(7850) = -7.19$ ， $p < 0$

001) 基于两个相关性之间差异的双尾检验，相关值的差异相对较小，表明相似的模型适合整合和情况依赖。与仅面部模型相比，我们还观察到仅情境模型 ( $t(7850) = 36.31$ ， $p < 0.001$ ) 和贝叶斯线索整合模型 ( $t(7850) = 48.35$ ， $p < 0.001$ ) 之间存在统计显著差异。基于两个相关性之间差异的双尾检验的模型。贝叶斯线索整合模型的稳健性可能明显低于仅情境模型，因为人们在进行推理时不会将先验信息整合到情感之上。为了测试这一点，我们使用平坦先验计算贝叶斯提示积分模型，即将先验设置为均匀分布。这也有助于检查凭经验收集的先验是否具有信息性，并比平坦先验提高模型拟合度。与具有信息性先验的模型 ( $t(7850) = 9.13$ ， $p < 0.001$ ) 基于两个相关性之间差异的双尾检验。这表明情绪的全局先验提供了丰富的信息，并为线索整合模型增加了价值，而仅情境模型由于缺乏先验整合而表现不佳。

模型预测的均方根误差 (RMSE) 证实了基于相关性的结果。仅情境模型的 RMSE 值最低 (0.042，自举 95% CI: 0.041, 0.044)，表明最适合经验数据，其次是贝叶斯线索整合模型 (0.071，自举 95% CI: 0.068, 0.073)，最后是仅面部模型 (0.075，自举 95% CI: 0.073, 0.077)（图 2b）。这些结果表明，仅情境模型可以更好地解释感知者的推论。

我们在研究 2-4 中复制了这些总体发现，因此我们观察到，平均而言，与其他两个模型相比，仅情境模型与人们根据面部和情境线索做出的判断具有更高的统计相关性（参见补充表 S2）模型。这表明感知者更加依赖情境信息，即使有机会整合，可能是因为面部信息除了背景知识之外几乎没有任何价值。作为稳健性测试，我们还使用直接确定性判断而不是强度判断来计算模型估计，并发现对情境线索的强烈依赖并不是我们计算基于频率的概率估计的方法的限制（参见补充图 S4）。我们发现，当使用基于两个相关性之间差异的双尾检验的确定性判断 ( $t(570) = 1.06$ ， $p = 0.289$ ) 时，贝叶斯模型拟合和仅情境模型拟合在统计上没有显著差异。因此，我们继续提出使用强度判断计算的概率，但请注意，反映概率匹配行为的决策可能并不总是适用于情感推断。为了准确地模拟情绪判断，了解可能反映不同感知任务的不同表示的潜在决策过程非常重要。

贝叶斯线索整合是否取决于推断的情绪？

除了这些总体模式之外，我们还分别比较了模型预测与每个情绪类别的经验判断的相关性。不同情绪类别的总体模式基本一致，但不同情绪也存在差异（图 3）。我们比较了为所有情绪类别的三个模型之间的相关值差异的自举样本计算的置信区间。在所有情绪中，仅情境模型和贝叶斯线索整合模型的相关性系统地显着高于仅面部模型，但娱乐除外，仅情境模型与仅面部模型在统计上没有差异。此外，仅情境模型的相关性在统计上并不统一高于贝叶斯线索整合模型在不同情绪中的相关性。相比于

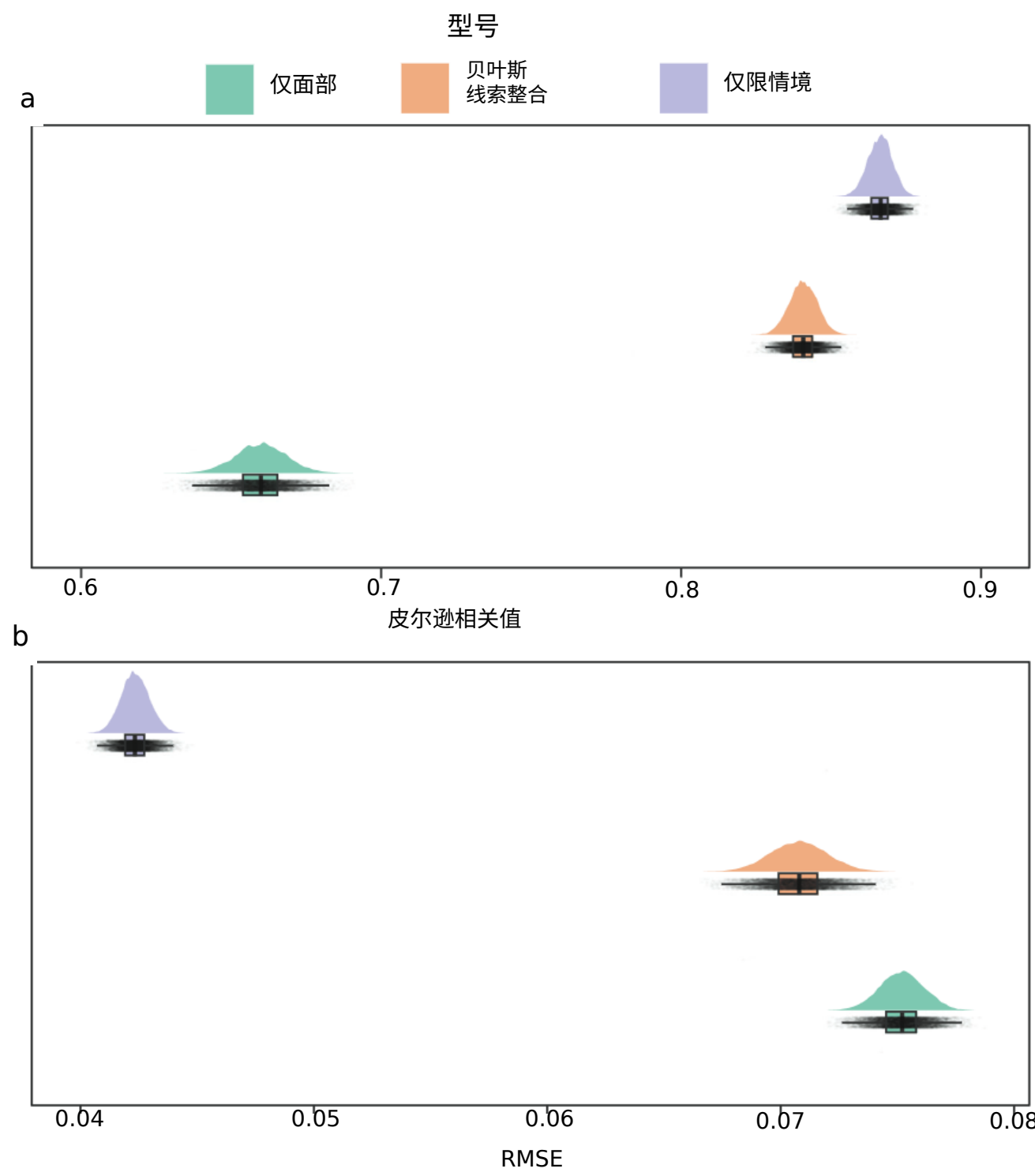


图2|研究 1 中三个模型的总体比较。a 分布从  $n = 7852$  对观测值中导出的每个模型的自举总体 Pearson 相关值。平均相关值越高表明模型拟合越好。b 从  $n = 7852$  对观测值导出的每个模型的自举总体均方根误差 (RMSE) 估计分布

较低的均方根误差 (RMSE) 值表明模型拟合较好。箱线图的中心线代表自举样本的平均相关性和平均 RMSE 值。每个箱线图的上限和下限代表每个模型的相关性或 RMSE 值的上四分位数 (第 75 个百分位数) 和下四分位数 (第 25 个百分位数)，从箱中延伸的须线代表 1.5 倍四分位数范围。

贝叶斯线索整合模型，对于娱乐、幸福等情绪类别，仅情境模型相关性显着较低（图 3a）；愤怒、蔑视、厌恶、悲伤和惊讶等情绪类别没有统计学差异（图 3b）；敬畏、尴尬、恐惧、兴趣、骄傲、羞耻等情绪类别显着较高（图 3c）。总之，在研究 1 中，我们发现，当使用更加多样化和复杂的情境描述和面部描绘时，贝叶斯线索整合模型可以很好地捕获经验数据，但在整体情感推断方面并不总是优于仅情境模型。这表明，当人们能够接触到丰富的情境背景时，他们往往主要依靠这些信息来推断他人的情绪。然而，这种广泛的模式确实因情绪类别而异。当推断娱乐和幸福的情绪时，面部线索似乎与情境线索相结合，正如贝叶斯线索整合模型的更好拟合所表明的那样。与我们的预测一致，这种发现模式与遇到情绪表达的频率的变化相一致。面部线索与情景线索相结合的情感类别，即幸福，也被报告为日常生活中最常遇到的表情类别（见补充表S4和S5）

这表明，遇到情绪表达的基本频率可能会影响人们对表达和情绪之间联系的信念，这反过来又有助于将面部线索整合到人们的推论中。

个体在线索依赖和整合方面有何不同？我们扩展了线索依赖（对面部或情境线索的依赖）和线索整合的调查，以检查三项研究中的个体差异（研究 2-4）。为此，在三项研究的每一项中，参与者对 13 种情绪类别的面部描绘及其相应的情景描述（经验联合线索评级）进行评分。刺激对是从较大组中随机选择的子集 ( $n = 44$ )。这项任务与 Le Mau 及其同事论文中的联合提示条件相同。我们首先计算每个人的经验联合线索概率估计，方法是对每个刺激的情绪评级进行平均，并对平均评级进行标准化，以使每个参与者和刺激的情绪评级总和等于 1。然后比较这些个人的联合线索概率估计到仅面部、仅情境和贝叶斯线索整合模型概率（从研究 1 中获得 44 个刺激的子集），分别计算每个个体的面部依赖、情境依赖和线索整合的估计。我们发现人们依赖面部线索的程度存在相当大的差异（研究 2:  $M = 0.46$ ,  $SD = 0.16$ ；研究 3:  $M = 0.53$ ,  $SD = 0.14$ ；研究 4:  $M = 0.62$ ,  $SD = 0.18$ ）线索（研究 2:  $M = 0.60$ ,

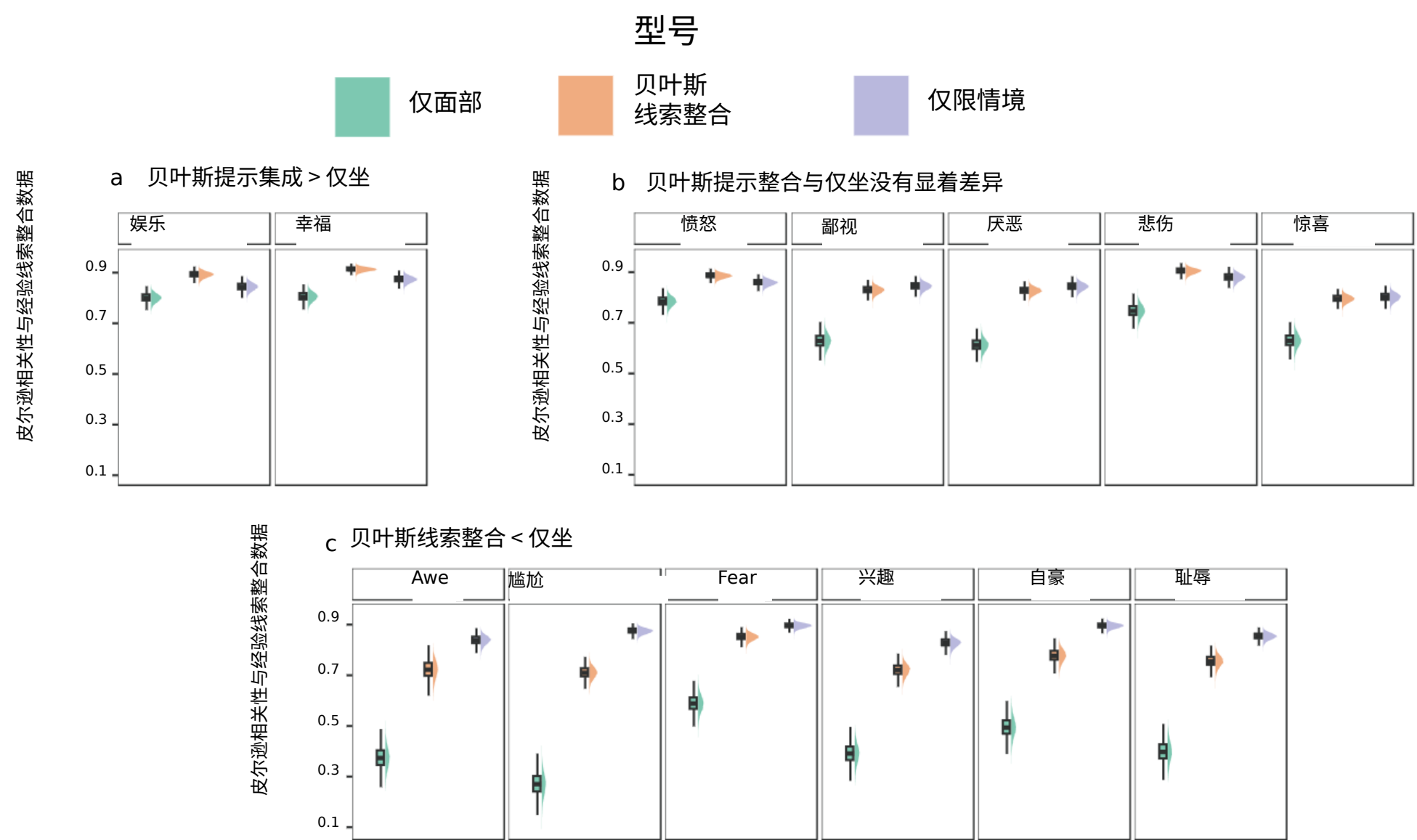


图3|每个情绪类别的模型相关分布。分配模型的 bootstrap Pearson 相关性来自 n = 604 对观察结果的情感类别的联合提示条件的经验判断。数据显示，贝叶斯线索整合模型与参与者判断的相关性显著高于仅情境模型；b 与仅情境模型没有显著差异；c 明显低于仅情境模型。每个的中心线箱线图表示引导样本的平均相关性。每个箱线图的上限和下限代表每个模型相关值的上四分位数（第 75 个百分位数）和下四分位数（第 25 个百分位数），从箱中延伸的须线代表 1.5 倍四分位数范围。有关每个情感类别的三个模型的成对比较之间相关性差异的 95% 置信区间值，请参阅补充表 S3。

标准差=0.21；研究 3：M = 0.73，SD = 0.20；研究 4：M = 0.90，SD = 0.24），以及面部和情境线索的整合（研究 2：M = 0.52，SD = 0.16；研究 3：M = 0.66，SD = 0.19；研究 4：M = 0.75，SD = 0.22）在所有三项研究中（图 4a-c）。此外，人们在主要依赖面部线索、情境线索还是线索整合方面也存在差异。这些基于最高性能模型的个体组如图 5a-c 所示。在研究 2 中，虽然大多数参与者（58.78%）主要依赖于情境线索，但有些（18.32%）主要依赖于面部线索，而其他人（22.9%）主要依赖于两种线索的整合（图 5a）-c）。随后的两项研究（研究 3 和 4）也证明了人们在是否主要依赖情境线索（研究 3：63.16%，研究 4：76.73%）和面部线索（研究 3：6.58%，研究 4：3.63%）方面存在差异。%，或提示整合（研究 3：30.26%，研究 4：19.64%）。有关模型估计的年龄相关影响的扩展分析，请参阅补充材料（表 S6 和随附的注释）。总之，这些发现表明，我们应该谨慎假设单一模型（例如研究 1 中描述的模型）足以捕获个体之间的情绪推断。

接下来，我们研究了个体的面部依赖、情境依赖和线索整合估计如何与情绪领域个体差异的其他测量相关（研究 2-4）。对于每一项分析，我们首先对线索依赖和线索整合相关性估计进行 Fisher 的 r 到 z 变换。然后，我们计算了这些估计值与针对多重校正调整的情绪个体差异的自我报告测量之间的双尾皮尔逊相关性，以便报告的所有 p 值均经过 Bonferroni 调整。在计算相关性之前，我们还使用马哈拉诺比斯距离方法删除了多元异常值。

我们首先检查了人们的线索依赖和线索整合估计是否与他们对情绪分化能力（RDEES 分化分量表）的信念相关。我们没有发现人们对区分情绪体验的能力的信念与线索依赖之间存在统计学上的显著关联（情境依赖：r (127) = -0.141，t = -1.609，p = 1.00，95% CI：-0.307，0.032；面部依赖：r (127) = 0.029，t = 0.327，p = 1.00，95% CI：-0.145，0.201）或线索整合估计（r (127) = -0.034，t = -0.388，p = 1.00，95% CI：-0.21，0.139）。RDEES 反映了个人对区分各种情绪体验的总体信念，而不是衡量的区分能力。因此，这些关于区分自己的情绪体验的信念可能与情绪推理中线索利用和整合的个体差异不相符。鉴于统计上的关系不显著，我们没有在后续研究中纳入这一衡量标准（研究 3 和 4）。

接下来，我们检查了参与者在情境环境中检测线索存在和不存在的能力（使用 CSI 量表评估）是否与线索依赖相关。我们假设，人们利用情境线索来推断他人情绪的能力与他们检测情境线索是否存在的能力呈正相关。我们没有找到可靠的证据来证明情境依赖与检测两个样本中是否存在情境线索的能力存在统计显著关系（研究 2：r (127) = 0.067，t = 0.779，p = 1.00，95% CI：-0.105，0.239；研究 3：r (146) = 0.023，t = 0.277，p = 1.00，95% CI：-0.139，0.184），但考虑到与情境信息使用的概念相关性，继续在研究中纳入这一衡量标准。在最终的高效研究（N = 268，研究 4）中，我们发现



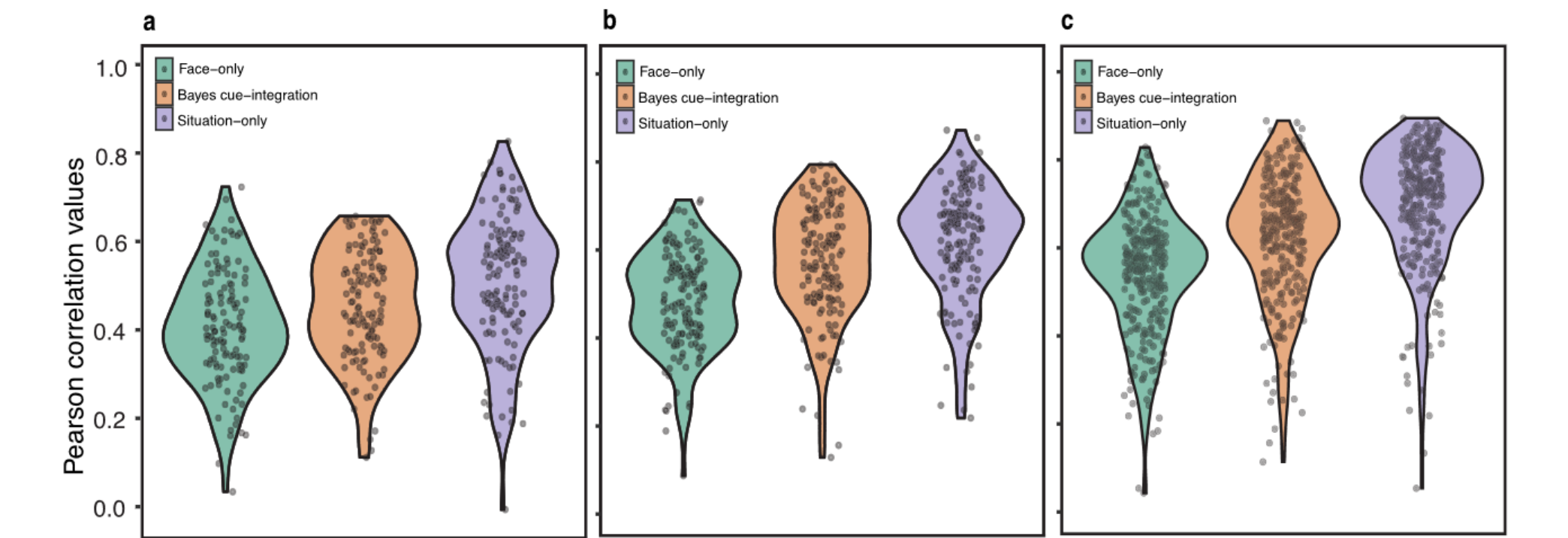


图4|模型和参与者之间相关值的分布。研究 2 中个体间仅面部、仅情境和贝叶斯线索整合模型相关性的分布（N = 129）； b 研究 3（N = 148）和 c 研究 4（N = 268）。

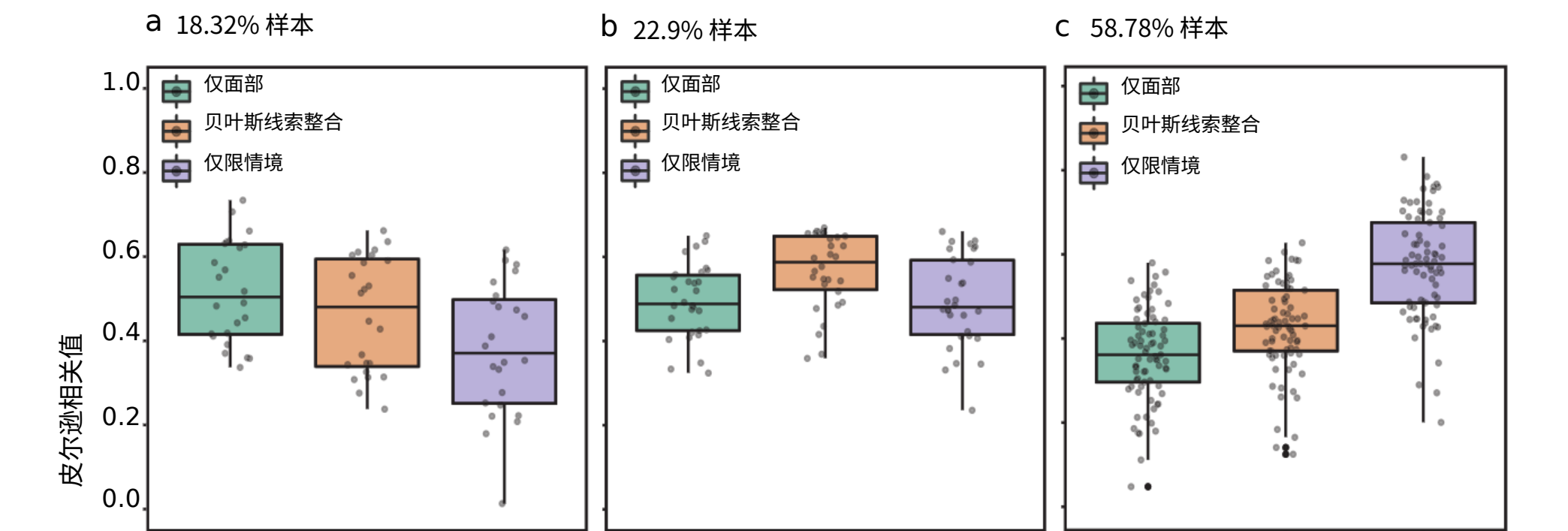


图5|按参与者分组的模型相关值的分布  
最高性能模型。对于仅面部模型具有最高相关性的参与者组，仅面部模型、贝叶斯线索整合和仅情境模型的 Pearson 相关值分布（N = 24; 18.32%）； b 贝叶斯线索整合模型（N = 30; 22.9%）和 c 仅情境模型（N = 77, 58  
研究 2 中分别为 78%）。误差线代表标准偏差。箱线图的中心线代表每个条件的中值相关值。每个箱线图的上限和下限代表每个条件的相关值的上四分位数（第 75 个百分点数）和下四分位数（第 25 个百分点数），从箱中延伸的须线代表 1.5 倍四分位数范围。

情景依赖估计和对上下文线索缺失的敏感性（研究 4:  $r(266) = 0.189$ ,  $t = 3.141$ ,  $p = 0.030$ , 95% CI: 0.071, 0.302），通过线索缺失维度的总分来衡量犯罪现场调查。这表明，在从面部和情境线索推断情绪时更有可能依赖情境线索的人，对于检测情境背景中线索的缺失也更敏感。然而，这种关系在之前的研究（研究 2 和 3）中并没有一致地被发现，这些研究有能力检测出可比较的效应大小，这表明我们应该谨慎地解释这种效应。此外，我们没有发现情境依赖估计和对上下文线索存在的敏感性之间存在统计上显著的关系，这是通过三项研究中 CSI 线索存在维度的总分来衡量的（研究 2:  $r(127) = 0.048$ ,  $t = 0.546$ ,  $p = 1.00$ , 95% CI: -0.126, 0.219; 研究 3:  $r(146) = 0.177$ ,  $t = 2.177$ ,  $p = 0.497$ , 95% CI: 0.016, 0.329; 研究 4:  $r(266) = 0.108$ ,  $t = 1.779$ ,  $p = 1.00$ , 95% CI: -0.012, 0.225）。鉴于 CSI 在理解情境背景方面的相关性，我们没有假设 CSI 与面部依赖或线索整合之间存在任何关系。

我们没有找到可靠的证据证明面部依赖与检测存在（研究 2:  $r(127) = 0.084$ ,  $t = 0.950$ ,  $p = 1.00$ , 95% CI: -0.090, 0.253）或不存在（研究 2: 上下文线索的  $r(127) = 0.041$ ,  $t = 0.458$ ,  $p = 1.00$ , 95% CI: -0.133, 0.212），也不是人们通过检测存在而进行的线索整合估计（研究 2:  $r(127) = 0.107$ ,  $t = 1.216$ ,  $p = 1.00$ , 95% CI: -0.067, 0.275）或缺乏上下文线索（研究 2:  $r(127) = 0.035$ ,  $t = 0.400$ ,  $p = 1.00$ , 95% CI: -0.138, 0.207）。鉴于 CSI 是上下文敏感性的相对狭窄的衡量标准，我们记录的弱但不一致的关联也许并不令人惊讶。它衡量人们检测是否存在与威胁相关的上下文线索的能力（例如，响应的紧迫性、自我和他人的控制）。我们研究中的场景可能包含了与威胁相关的更广泛的情境特征（例如，与他人的互动以及与目标的一致性），这反映在与它们相关的广泛情绪中（参见补充数据1）。  
为了解决 CSI 潜在的有限关注点，在研究 3 和 4 中，我们检查了情境依赖是否与更广泛的基于能力的测量相一致，该测量捕获了理解情绪和情境之间联系的个体差异。这是



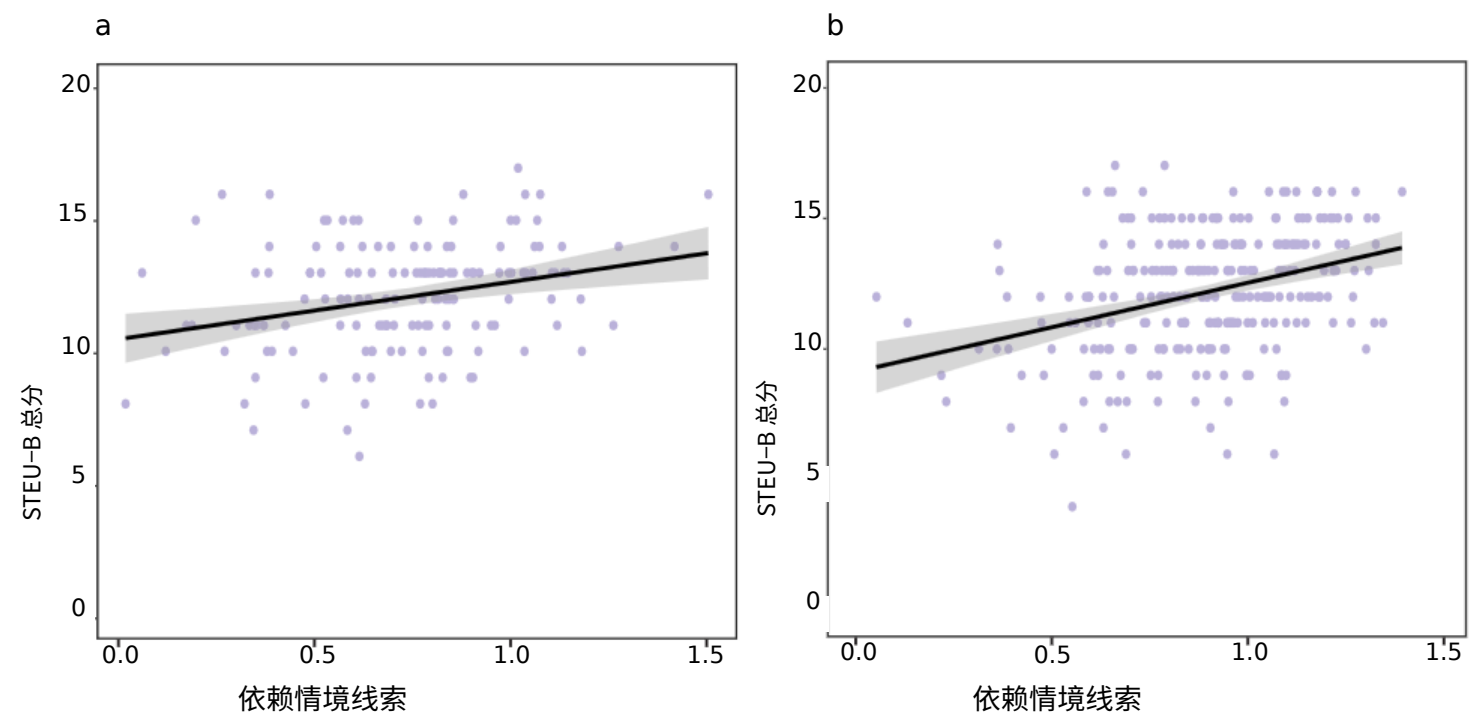


图 6 |对情境线索的依赖与所处环境之间的关系  
情绪的立场。在研究 3 (N = 148) 和研究 4 (N = 268) 中，个人的情境模型估计（x 轴）与他们对情绪体验的情境理解（y 轴）之间存在中等正相关关系。每个点

图上的表示来自数据集的观察结果。黑线代表变量之间线性关系的最佳拟合线，灰色阴影带代表拟合值的逐点 95% 置信区间。

使用情绪理解情景测试（或 STEU-B）的简短版本进行评估。我们假设情境依赖估计的变异性与人们对情绪的情境理解呈正相关，即他们准确识别（基于群体共识）人们在不同情境下最有可能经历哪些情绪的能力。与我们的预测一致，我们发现人们对情境情绪的理解与其情境依赖估计之间存在可靠且具有统计显著性的正相关关系（研究 3:  $r(146) = 0.281$ ,  $t = 3.533$ ,  $p = 0.009$ , 95% CI: 0.125, 0.423; 预注册复制研究 4:  $r(266) = 0.343$ ,  $t = 5.949$ ,  $p < 0.001$ , 95% CI: 0.232, 0.444) 表明，在推断情绪时更多依赖情境线索的人也能够更好地理解情绪和情境之间的联系（图 6a、b）。

此外，我们还检查了个体面部依赖、情境依赖和线索整合估计的时间稳定性（研究 5）。遵循研究 2-4 中使用的分析技术，我们首先计算了第一阶段和第二阶段个人的面部依赖、情境依赖和线索整合估计，并使用 Fisher  $r$ -to- $z$  变换来转换这些估计。然后，我们计算了双尾类内相关性，以检查不同时间个体的线索依赖（面部和情境）和线索整合估计的一致性。结果表明，平均而言，仅依赖情境线索（ $ICC = 0.593$ ,  $F(109, 109) = 3.136$ ,  $p < 0.001$ , 95% CI: 0.216, 0.768）和线索整合（ $ICC = 0.672$ ,  $F(109, 109) = 3.260$ ,  $p < 0.001$ , 95% CI: 0.508, 0.779），但仅依赖面部线索的可靠性较弱（ $ICC = 0.370$ ,  $F(109, 109) = 2.436$ ,  $p < 0.001$ , 95% CI: -0.170, 0.656）（图 7）。这表明人们对情境线索的依赖以及面部和情境线索的整合随着时间的推移保持适度稳定。然而，人们对面部线索的依赖并不是随时间变化的稳定的个体差异。我们还在删除多变量异常值后检查了这些结果的稳健性，并发现了线索依赖和线索整合个体差异的时间稳定性的类似结果（参见补充表 S7）。

讨论

在我们的工作中，我们研究人们利用面部和情境线索来推断他人情绪的情况。将基于单一线索（面部或情境）和线索整合的推论与人们在获得面部特征和情境背景的情况下对情绪的判断进行比较。除了集团层面的估计

我们研究了不同人和情绪类别之间面部和情境线索的利用差异。

我们的数据显示，基于获取他人面部和情境线索对他人情绪状态的总体推断与仅基于情境线索和线索整合的推断非常一致。平均而言，与基于贝叶斯线索整合的推断相比，仅基于情境线索的人们的情绪推断与联合线索经验判断的拟合度明显更好，但只有轻微的拟合度。此外，与仅情境模型和整合模型相比，我们发现仅基于面部线索的推论在捕捉同时存在面孔和情境时做出的情感推论方面始终明显较弱。一种解释是，在我们的研究背景下，面部表情包含的信息是情境背景中信息的小子集，因此仅面部推理很差，并且线索整合的推理与情境没有太大差异——仅是推论。这与普遍观点的一个方面相矛盾：面部行为与情绪状态紧密相关，因此获取面部线索应该提供有关潜在情绪的诊断信息。相反，这些发现表明，人们对他人感受的推断可以通过他们的信念来充分解释，即情境背景会导致情绪体验。这些结果还表明需要一种更丰富的外行情感理论：感知者实际上可能认为表达是有位置的（即，在因果模型中，在情境和表达之间添加因果联系）。观察到感知者在推断情绪时对情境背景的强烈依赖超出了过去的研究结果，证明了强烈的情境依赖（回顾，参见参考文献 15,25）。我们证明，情境通常在情感推断中占据优先地位，而不是与面部描绘相结合。这一发现出现在我们的工作中，避免了对比情境和面部线索的含义的陷阱。至关重要的是，我们的研究结果补充了现有的研究，即检查感知者根据现实世界背景下的自发表达进行的推断。例如，之前的研究表明，在体育比赛胜利和失败的高强度背景下（当与肢体语言线索分开时）自发的现实世界面部运动对于区分基于效价的信息没有用处。我们的研究结果进一步表明，捕捉更广泛强度和情感的非规范面部描绘不仅对于情感推断的信息量同样较少，而且很少与情境背景相结合（概念上的复制）。

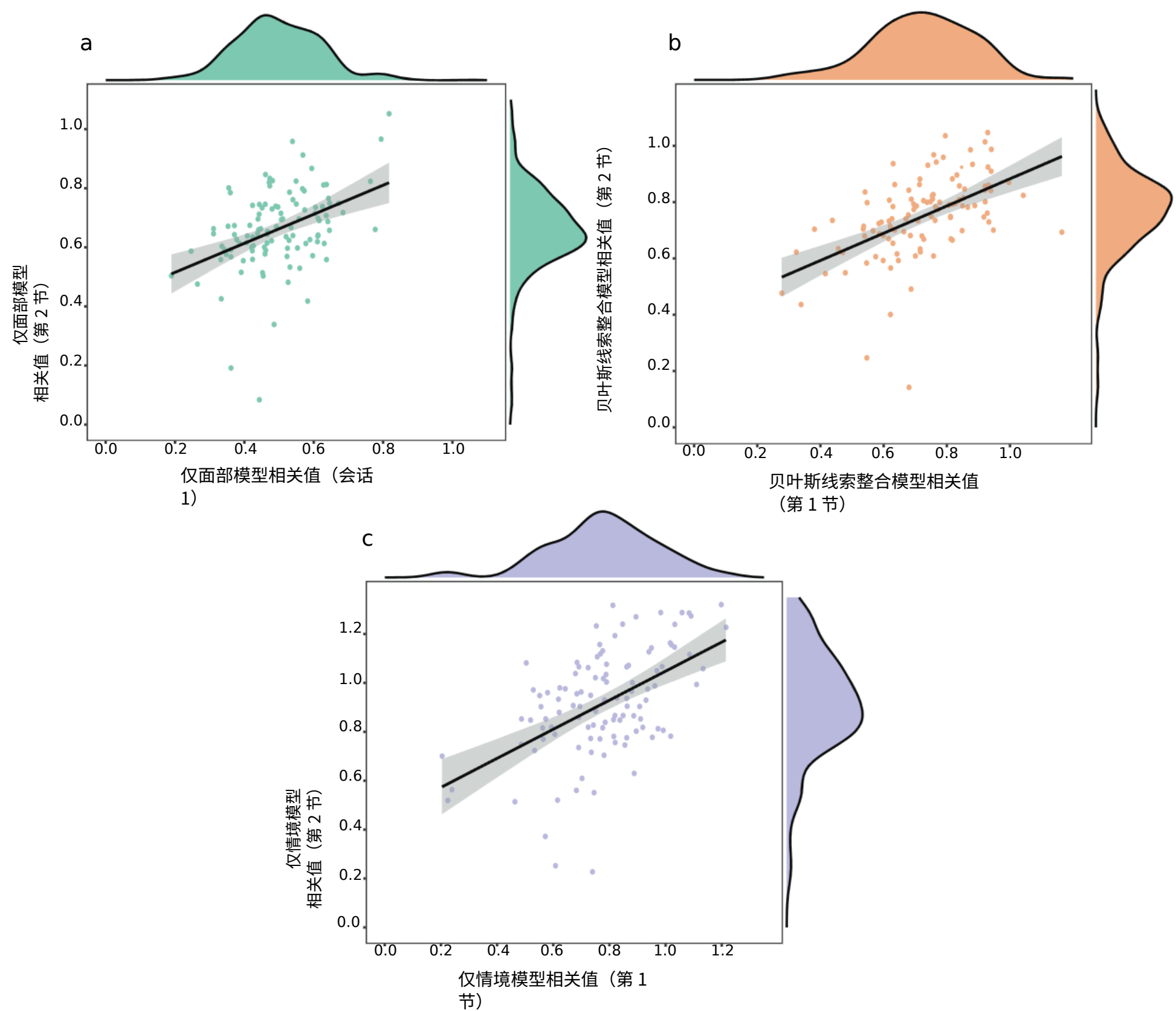


图 7 | 线索依赖和线索整合估计的重测可靠性。  
仅面模型的两个时间点的 Fisher  $r$  到  $z$  变换模型相关值之间的关系；b 贝叶斯线索整合模型和 c，仅情境模型。Pearson 相关值为 Fisher

在计算类内相关性之前进行  $r$  到  $z$  转换。黑线代表变量之间线性关系的最佳拟合线，灰色阴影带代表拟合值的逐点 95% 置信区间。在第 1 节和第 2 节中测试了相同的样本 ( $N = 110$ )。

在考虑现实世界的情感推断时，对情境背景的依赖可能特别相关。我们工作中使用的线索具有更大的多样性和复杂性。例如，该数据集中对情感的描绘很少符合对面部移动方式的先验规范期望。例如，在被视为引起兴趣的情况下，演员很少通过眯起眼睑并产生有控制的微笑（拟议的规范表达）来描绘这种情感。这可能是因为演员被要求根据描述来表演情景的复杂性，而不是为他们提供情感类别标签或预先确定的面部肌肉动作来表演（如文献中常见的那样）。事实上，演员们如何利用面部表情来描绘情感是存在差异的。然而，这些刺激不足以捕捉日常社交场合中出现的自发自然主义表达。因此，这些发现强调了使用不同的更自然刺激来测试情绪理论和计算模型的重要性

这种刺激可以帮助测试人们以与现实世界现象更密切相关的方式推断情绪的能力。我们的研究结果也与之前研究面部信息在印象形成和更新中的作用的研究相似。对特质推断的研究表明，人们根据面部的外观线索来推断特质（例如，参见参考文献 71），尤其是当被要求做出快速判断时（例如，参见参考文献 72,73）。然而，在存在目标行为/特征的显式描述的情况下，基于面部信息的特征的显式和隐式印象都会更新。与我们的结果一起，诸如此类的发现表明，面部在社会认知中占主导地位的假设可能被夸大了。事实上，先前的研究表明，感知者在形成情感推论时高估了面部信息的价值，而不是背景信息。正如无数研究表明的那样，人们对他人的推断，无论是特征还是状态推断，都可以根据面部信息来制定，但似乎受到其他信息来源（例如背景知识）的严重限制。

研究结果的总体模式与我们工作中测量的大多数情绪类别相关，但是，不同情绪类别之间存在一些差异，即人们是否相对更多地依赖情境背景。当推断娱乐和幸福的情绪时，人们平均最有可能整合面部和情境线索，正如我们的数据更适合贝叶斯线索整合模型所证明的那样。这表明，在推断这些情绪时，面部线索可能被认为与情境线索一起提供了信息，因此更有可能被整合。

在推断不同情绪类别的线索整合中观察到的变化与人们日常生活中遇到和处理面部表情的变化以有意义的方式一致（例如，参见参考文献 40,67）。幸福的情感类别是人们最有可能将面部线索与情境线索相结合的情感类别，也是人们对日常生活具有较高感知敏感性并报告看到最多面部表情的类别。同样，人们报告看到（在研究中）最少面部表情的恐惧情绪类别属于人们不太可能依赖整合的类别之一。这些发现补充了之前的研究结果，表明遇到某种情绪类别的面部表情的频率越高，也与规范表情的识别“准确性”越高以及响应延迟的减少有关。总之，我们的研究表明，尽管来自面部描绘的信息有时会被整合到情感推论中，但这种整合在不同类别中以有意义的方式存在差异。在研究人们如何推断他人的情绪状态时，这种变化很少被注意到，但对于确定不同情绪类别的情境实例中线索的相关性可能具有影响。

我们还证明了个体在依赖情境线索、面部线索或在推断他人情绪时整合这些线索的倾向方面存在差异。在三项研究中，我们发现大多数人主要根据情境线索来推断他人的情绪，但有些人根据面部和情境线索的整合来推断，而一小部分人的推断主要基于面部线索。这种差异表明，对于由情境引起的情绪或导致表达行为的情绪的外行信念的强度可能因人而异。尽管有一些有限的工作通过根据特定模型参数对个体进行分组来检查“个体差异”，但在认知建模中，个体之间的这种差异经常被忽略。在这里，我们专注于通过基于共识的方法对线索利用和整合的个体差异进行建模。另一种方法是使用目标经历或表达的情绪报告（讨论参见参考文献 81）。这两种方法可能会产生不同的结果，因为共识判断所捕获的共享文化模型与给定目标的表达情感联系有所不同。此外，在我们的工作中，我们发现人们依赖情境线索来推断情绪与他们理解情绪情境本质的能力密切相关。也就是说，更善于理解情绪和情境之间联系的人也更依赖情境线索来推断情绪。这进一步表明，基于情境背景推断情绪是有效的，并且根据个人利用情境信息的能力，以系统的方式在不同个体之间有所不同。

这一发现与之前的研究大体一致，即人们理解情绪和情境之间联系的能力与他们在社交情境中识别他人意图、信念和感受的能力以及根据不同情绪在不同情绪调节策略之间切换的能力呈正相关。不同情况的需求（通常称为监管灵活性和适应性）

这些关联意味着人们依赖情境信息的程度可能反映了他们理解自己以及他人心理状态的情境本质的规范能力。

除了个体之间的变异性之外，我们还检查了个体随时间的变异性。我们的研究结果表明，随着时间的推移，人们对情境线索的依赖和线索的整合呈现出适度稳定的倾向。这种时间稳定性表明了类似倾向的特征，表明个体如何依赖情境线索和线索整合的变化随着时间的推移在某种程度上是稳定的。相比之下，对面部线索的依赖具有较低的时间稳定性。在检查模型拟合估计的模式时，我们观察到第二个时间点的面部线索依赖度通常较高。这表明人们如何使用面部线索来推断情绪可能是上下文相关的，因此熟悉度会增加对面部线索的依赖。

我们的研究结果为使用多样化和复杂的社会场景和面部描绘扩展情感理论和计算模型的检验提供了初步支持，但需要更多的实证工作来解决我们方法的局限性。虽然我们在这里使用的刺激比过去的工作更加复杂和多样化，但它们并不能代表每天发生的无数情感体验。此外，所使用的刺激中的面部描绘最终是表演出来的，并没有捕捉到现实世界中发生的实际自发的自然表情。另一个限制是，这里的情境线索都是对社会情境的描述，而不是现实生活中经历的动态展开的情境。因此，向参与者强调的情况特征在日常生活中可能并不那么容易理解。对现实世界情况的理解需要能够表示抽象特征并部署注意力资源来优先考虑相关特征（例如，参见参考文献 84,85）。缓解这种担忧的是我们的数据中情境评级的共识相对较低（参见补充图S5-S7），这表明这些线索具有可变的解释，因此是模糊的。此外，即使个人对日常生活中的特定情况有不完整或不确定的知识，感知到的物理环境的各个方面仍然可能构成社会预测的基础。因此，现实世界中的许多情绪感知实例不太可能真正脱离语境。在具有不同程度的信息不确定性的不同背景下测试线索依赖和整合是一个有待探索的问题。此外，我们在研究 2-4 中使用根据研究 1 样本计算的模型估计值。这假设从这些线索得出的关于情绪的推论在样本之间具有可比性，这可能是一个限制（参见补充图S1-S3和补充表S1中的稳健性检查）。最后，我们用于检查线索使用和线索整合的个体水平变异性的样本很大程度上基于在线获得的便利样本。这种方法限制了我们研究结果的普遍性。我们开始通过最终的全国代表性样本（基于年龄、种族和性别）来解决这一问题，但还有许多其他维度需要解决普遍性问题。

在未来的工作中，使用可以使用贝叶斯推理进行测试的替代理论假设来检查线索整合将是有益的。在这里，我们研究了源自外行直觉的线索整合，即人们期望情绪体验是由情境结果引起并导致面部描绘。这种外行直觉表明，表达行为直接源于情绪（情绪本身是情境性的），因此表达行为可能不会进一步受到情境背景知识的限制。但人们可能持有一种外行理论，即情绪表达是定位的（即 DAG 中情境和表达之间的额外因果关系），这表明当前的线索整合模型本身可能存在局限性。举个例子，人们在对老板生气时可能会想到的非语言表达，与对在交通中抢道的司机表达的愤怒截然不同。检查面部和情境线索的整合将是富有成效的



基于另一种模型，该模型假设面部线索受到情绪和情况的限制。测试该模型需要构建一个自然刺激集，其中包含每个情绪类别内足够且系统的情境变化，以模拟这种复杂性。另一个扩展是测试当前发现在不同类型线索中的普遍性，包括当身体线索作为背景时。例如，根据之前的工作，身体环境模型可能至少会优于仅面部模型。当前的建模方法可以扩展到将人们对情绪理解的一系列线索的外行信念形式化。我们的研究结果的普遍性还可以通过检查对单一线索的依赖适用于不同程度的上下文和面部信息的程度来测试。一个悬而未决的问题是，在上下文信息不完整的情况下，人们是否更依赖视觉线索（例如面孔或肢体语言）。例如，当在赌场与陌生人玩扑克游戏时，您无权访问对手的牌。上下文信息的缺乏可能会导致您依赖视觉线索来了解对手可能的感受，并使用它来推理上下文（例如，他们有好牌还是坏牌）。未来的另一个方向是检查是否存在情绪特定的个体差异，这个问题超出了本研究的范围。此外，在我们的工作中，我们研究了一些可能影响个人对情境线索的利用的因素，但未来的工作可以解决有关面部线索利用的类似问题。人们相信，不同的情感类别在面部和身体上的表达方式各不相同，这些关于物质表达的信念与关于情绪具有内在本质的信念相关联。对物质表达和情感固有性的强烈信念可能会导致更多地利用面部线索，这对于在未来的工作中进行研究很有价值。最后，未来一个有趣的发展方向是一种建模方法，该方法考虑给定刺激的情绪评级概况，而不是在评估模型拟合时单独处理每个刺激的每种情绪。例如，“厌恶愤怒”的经历可能反映了道德愤怒，但厌恶和愤怒不会被视为单独的情绪。

总体而言，目前的研究表明，人们在推断别人的感受时会严重考虑情境背景。这挑战了强烈强调面部线索在情绪推断中的作用的普遍观点。此外，情绪推断中情境背景的权衡程度因情绪类别和个体而异，其系统性方式可能具有现实世界的社会影响。这些发现支持了关于情绪推论是复杂的、可变的并且是由包括背景知识在内的多种因素解释的论点。

## 方法

所有研究均符合进行人类受试者研究的道德准则，并根据耶鲁大学机构审查委员会批准的豁免协议进行管理（IRB 号：2000028669）。所有研究的参与者在参与研究之前都提供了知情同意书，并根据当时 Prolific 建议的标记为“良好”的小时费率（10-12 美元/小时）在任务期间按比例获得补偿。

### 研究1

概述。这项研究由两个数据集组成：(1) 从 Le Mau 及其同事 (2021) 获得的档案数据集，以及 (2) 用于计算推断情绪的先验期望的实验数据集。档案数据。档案数据包含社会场景的评分、演员对这些场景的描绘的评分以及组合的评分。这些评级是针对总共 604 个刺激对（场景和姿势）获得的，这些刺激对来自两本书：

角色：演员表演和陷入困境：演员表演。这些卷中包含了一群专业演员在接受了唤起情感的场景后所摆出的表情图像。一些场景的例子包括：“当她从一家餐馆出来时，她正在面对拒绝她的情人以及他的妻子”；“他是一个从摩托车酒吧里出来的摩托车手，就像一个开着保时捷的人开着闪闪发光的哈雷摩托车一样”。由于版权限制，我们无法提供面部肖像的示例图像。

该数据集包括参与者的 75,390 个观察结果，他们对大约 30 个刺激的随机子集（总共 604 个刺激）进行了评分。每项观察都包括对 13 种不同情绪类别的评分。参与者对三种不同条件中的一种进行了评分——仅面部（N = 842）、仅情境（N = 839）、面部和情境相结合（N = 845）。在只看到脸部的情况下，参与者只能看到演员对场景的描绘。在仅情景条件下，参与者仅查看这些场景的描述。在组合或联合提示条件下，参与者查看场景的描述以及演员对这些场景的描绘。他们评定的 13 种情绪是：有趣、愤怒、敬畏、蔑视、厌恶、尴尬、恐惧、幸福、兴趣、骄傲、悲伤、羞耻、惊讶。首先对每种情绪类别进行存在性评级，即，如果参与者观察到一种情绪，他们会回答是或否。如果存在这种情绪，即他们回答“是”，那么参与者会按照 4 点李克特量表（轻微、中等、强烈、强烈）评估该情绪的强度。

该数据集中演员的面部描绘与所提出的情感类别的规范面部配置并不可靠地一致。相反，面部姿势符合日常生活中自发情绪表达的变化。例如，与日常生活中的人们类似，演员在描绘与情绪愤怒一致的场景时，大约有30%的时间会皱眉。这一刺激方案的复杂性，包括情景及其描述，提供了更大的生态有效性和范围。收集有关人们先期期望的数据。我们收集了人们在日常生活中对上述 13 种情绪类别的评分数据（之前的任务）。在某些认知建模范式中，通过要求参与者对单个问题进行评分来凭经验收集先验是标准的（例如，参见参考文献 89）。此外，我们收集了经验先验，而不是直接根据档案数据估计先验，因为我们不假设刺激范围代表日常生活中可能影响外行判断的情感实例。我们将这些先验与萨默维尔和惠伦收集的类似评级进行了比较（参见补充表 S8）。此外，我们还加入了一个简短的评级任务，该任务与 Le Mau 及其同事论文中的联合提示条件相同，以确保我们收集的数据与档案数据之间的一致性，即检查提供此类评级是否会改变判断的性质。如果对可能性的报告改变了情感推断的性质，这可能会破坏这些可能性估计为情感推断提供信息的提议。我们确认，在评分任务和所有 13 种情绪类别的档案数据中，刺激子集的评分分布是相似的。

参与者。使用在线数据收集平台 Prolific 从美国招募了 45 名母语为英语的参与者（20 名男性，25 名女性，平均年龄 = 38 岁，年龄范围 = 18-60 岁）。我们在当前样本中招募了 45 名参与者，以大致匹配档案数据集，其中平均约 40 名参与者对给定的刺激做出反应。补充表 S9 提供了详细的人口统计信息。

先决任务。在之前的任务中，参与者回答了一个问题，该问题询问他们在日常生活中感知每种情绪类别的可能性，采用 7 点李克特量表，其中包含三个锚点 - 1：

完全不可能，4：中等可能性，7：极有可能。问题指出：“当你看到人们在日常生活中经历情绪时，你感知到人们经历[情绪]的可能性有多大”。

评级任务。任务设计与档案数据中的联合提示条件相同，使得每个参与者对一组面部描绘和情境描述的 13 种不同情绪做出反应（是/否；随后是轻微、中等、强烈、强烈）。刺激集由 10 个刺激的随机样本组成，这些样本是从 604 个刺激的较大池中抽取的。刺激的顺序是随机的。

程序。所有参与者在同意参与本研究之前都会阅读在线同意书。然后，参与者阅读评级任务的说明，在查看场景描述和演员在这些场景中的情绪描绘后，他们将提供情绪评级。这些说明与 Le Mau 及其同事提供的说明相同。接下来是一个问题，旨在验证他们是否阅读了说明。回答这个问题后，参与者阅读说明，对感知每种情绪类别的可能性进行评级，并回答另一个问题以验证他们阅读了说明。回答完这些问题后，系统会快速提醒他们相关说明。这样做的目的是为了提醒参与者，特别是那些给出错误答案的参与者任务说明。然后，参与者对之前的任务进行评分，然后进行简短的 10 项评分任务。最后，参与者还填写了一份简短的人口统计调查问卷，并对他们的参与表示感谢和补偿。

数据准备。我们从档案 (N = 13) 和先验任务数据 (N = 1) 中删除了具有虚假数据的参与者，在各自的数据集中总共留下了 2513 名和 44 名参与者进行分析。

研究2

概述。研究 2 包括一项实验和两份问卷（详情如下）。这项研究采用受试者内设计，所有参与者都对实验做出回应，然后填写两份问卷。

参与者。使用在线数据收集平台 Prolific 从美国招募了 150 名年龄在 18 至 60 岁之间、以英语为母语的参与者。为了确定样本量，我们使用 G\*Power 对控制多重比较的双变量相关性进行了功效分析。结果表明，对于 149 个样本，我们有足够的功效 ( $1-\beta = 0.8$ ) 来检测中小效应大小 ( $r = 0.3$ )。有 8 人未完成研究，总共 142 名参与者（平均年龄 = 33.84，SD 年龄 = 11.69；50.7% 女性，47.18% 男性，2.11% 非二元性别；76.05% 白人，10.56% 亚裔，7.75% 黑人或非裔美国人）。补充表 S10 提供了详细的人口统计信息。

实验任务。在这项任务中，每位参与者观看 44 个刺激对，其中包括面部肖像图像以及社交情境描述。对于每个刺激，参与者对 13 种情绪类别的存在（是/否）和强度（轻微、中等、强烈、强烈）进行评分，与档案数据中的联合提示条件相同。这 44 个刺激对以随机顺序存在，并且是从档案数据中使用的 604 个刺激对的较大集合中随机抽样的。使用 R 中的“pwr”包进行先验功率分析来确定刺激的数量，以获得中等效果大小 ( $r = 0.4$ )。我们为刺激选择了中等效应大小，因为这是从研究 1 中获得的自举模型相关性的最小效应大小。

上下文敏感指数或 CSI。CSI 是一种基于能力的衡量标准，可捕捉个人对上下文线索存在和不存在的敏感度。该问卷由 6 个日常生活中出现的情况的简短描述组成。对于每种情况描述，参与者按照 7 分制对一组 3-4 个问题进行评分，这些问题与该情况的心理特征（例如，自我控制、紧迫性）相关，其中 1 表示“一点也不”，7 表示“非常非常”。

情绪体验量表或 RDEES 的范围和差异。RDEES 是一种自我报告测量，通过测量人们报告的情绪体验范围（范围子维度）以及他们报告的在情绪类别内进行细微区分的倾向（分化子维度）来捕获情绪复杂性的个体差异。该问卷由 14 个项目组成，分为两个子量表，每个子量表均分为 5 分制，其中 1 表示该陈述“不能很好地描述我”，5 表示该陈述“很好地描述了我”。

程序。参与者首先阅读并签署同意参与本研究的在线同意书。然后，参与者阅读实验任务的说明，这些说明与 Le Mau 及其同事（2021）提供的说明相同。接下来是一个旨在验证参与者阅读说明的问题。然后参与者执行实验任务，然后填写 CSI 和 RDEES 问卷。调查问卷的顺序在参与者之间是平衡的。最后，参与者还填写了一份简短的人口统计调查问卷，并对他们的参与表示感谢和补偿。

数据准备。为了在情绪推理任务中正确使用评分量表，参与者被指示仅在情绪存在时对强度进行评分。由于平台的限制，强度等级上的意外评级无法撤销。当意外进行强度评级时，我们指示参与者对“轻微强度”进行评级。完全错误的评级被定义为标记出某种情绪的缺失，但将其感知的强度评级为高于轻微强度的任何事物。意外提供、不正确或缺失超过 10% 数据评分的参与者将被从分析中删除。删除了 11 名参与者的数据，留下 131 名参与者的样本量。错误的评分在逐次试验的基础上被删除，意外的评分被纠正为“情感缺失”。我们还使用 Routliers 包中的 Mahalanobis 距离方法进行了多变量异常值检测。2 个受试者被识别为异常值，他们的数据被删除，最终有 129 名参与者进行最终分析。

研究3

概述。研究 3 是研究 2 的延伸，由一项实验和两份问卷组成（详细信息如下）。这项研究采用受试者内设计，所有参与者都对实验做出回应，然后填写两份问卷。

参与者。使用在线数据收集平台 Prolific 从美国招募了 168 名年龄在 18 至 60 岁之间、以英语为母语的参与者。为了确定样本量，我们使用 G\*Power 对控制多重比较的双变量相关性进行了功效分析。结果表明，对于 139 个样本，我们有足够的功效 ( $1-\beta = 0.8$ ) 来检测中小效应大小 ( $r = 0.25$ )。根据之前的研究，我们对额外 20% 的数据进行了过采样，以考虑排除和自然减员，最终样本总数为 168 人。有 6 个人未能通过任务中内置的注意力检查，因此他们的数据从所有分析中删除，留下 162 个人的样本（平均年龄 = 36.04，SD 年龄 = 12.87；59.88% 女性，34.57% 男性，5.56% 非二元性别；71.6% 白色，12.35% 黑色或

非裔美国人，5.56% 亚裔）。补充表 S10 提供了详细的人口统计信息。

实验任务。实验任务与研究 2 中使用的相同。

上下文敏感指数或CSI。 CSI 与研究 2 中使用的相同。

情绪理解情境测试-简短或STEU-B。 STEU-B 是 STEU 的简短版本，用于衡量个人对情绪和情况之间联系的理解。它由 21 个项目组成，每个项目描述一种情绪情况，并附有包含 5 个情绪词的多项选择量表。参与者被要求选择最有可能由所描述的情况导致的情绪。根据 MacCann 和 Roberts (2008) 确定的评分指南，对回答进行正确或错误评分。

程序。本研究的程序与研究 2 相同，只是 RDEES 被 STEU-B 取代。

数据准备。按照与研究 2 相同的程序，那些提供了超过 10% 数据的意外、不正确或缺失评级的参与者将被从分析中删除。删除了 10 名参与者的数据，留下 152 名参与者的样本量。我们还使用 Routliers 包中的 Mahalanobis 距离函数进行多变量异常值检测。4 个受试者被识别为异常值，他们的数据被删除，最终有 148 名参与者进行最终分析。

研究4

概述。研究 4 是我们之前工作的预注册复制 (<https://osf.io/u9abg>, 注册日期: 2022 年 5 月 27 日)，具有全国代表性样本，由一项实验和两份问卷组成（详细信息如下）。这项研究采用受试者内设计，所有参与者都对实验做出回应，然后填写两份问卷。与预注册没有偏差，但根据补充表 S6 中详述的审稿人意见进行了一些额外的分析。

参与者。使用在线数据收集平台 Prolific 招募了来自美国的 303 名具有全国代表性的参与者样本。样本量计划是为了对先前观察到的效应大小提供更稳定的估计，因为相关性往往稳定在 N = 260。我们正在招募 260 多名参与者，以满足 Prolific 的最小样本量 (N = 300)，以招募具有全国代表性的样本。我们还对双变量相关性进行了先验功效分析，以检测 STEU-B 分数与先前研究中观察到的情境依赖估计之间关系的效应大小 ( $r = 0.3$ )，并控制多重比较 ( $\alpha = 0.05$ ) /3) 使用 G\*power。结果表明，样本量为 95 时，我们有足够的功效 ( $1-\beta = 0.8$ ) 来检测预期效应大小 ( $r = 0.3$ )。因此，计划的样本量满足了样本量要求，同时还估计了稳定的效应大小。有 9 人未能通过任务中内置的注意力检查，因此他们的数据从任何分析中删除，留下 294 人样本（平均年龄 = 45.13, SD 年龄 = 16.31; 51.36% 为女性，47.96% 为男性，0.34% 为非二元性别; 71.43% 白人，12.24% 黑人或非裔美国人，5.1% 亚洲人）。补充表 S10 提供了详细的人口统计信息。

实验任务。实验任务与研究 2 和 3 中使用的相同。

上下文敏感指数或CSI。 CSI 与研究 2 和 3 中使用的相同。

情绪理解情境测试-简短或STEU-B。  
STEU-B 与研究 3 中使用的相同。

程序。本研究的程序与研究 3 相同。

数据准备。按照与研究 2 相同的程序，那些提供了超过 10% 数据的意外、不正确或缺失评级的参与者将被从分析中删除。19 名参与者的数据被删除，样本规模为 275 名参与者。我们还使用 Routliers 包中的 Mahalanobis 距离函数进行多变量异常值检测。7 个受试者被识别为异常值，他们的数据被删除，最终有 268 名参与者进行最终分析。

研究5

概述。研究 5 包括对同一组参与者进行两次的实验。这项研究采用受试者内设计，所有参与者都对同一实验做出反应。

参与者。使用在线数据收集平台 Prolific 从美国招募了 136 名年龄在 18 至 60 岁之间、以英语为母语的参与者。为了确定样本量，我们使用 G\*Power 对控制多重比较的双变量相关性进行了功效分析。结果表明，对于 102 个样本，我们有足够的功效 ( $1-\beta = 0.8$ ) 来检测所需的效应大小 (ICC:  $r = 0.75$ )。根据之前的研究，我们对额外 25% 的数据进行了过采样，以考虑退出、排除和自然减员，最终样本总数为 136 人。17 人因未完成第 2 节而退出，因此他们的数据从任何分析中删除，留下 119 人样本（平均年龄 = 36.78, SD 年龄 = 11.29; 48.74% 女性，47.06% 男性和 4.20% 非二元性别; 65.55 % 白人，8.40% 黑人或非裔美国人，5.88% 亚洲人）。补充表 S10 提供了详细的人口统计信息。

实验任务。实验任务与研究 2-4 中使用的相同。

程序。在第一节中，参与者首先阅读并签署了同意参与本研究的在线同意书。然后，参与者阅读实验任务的说明，这些说明与研究 2-3 中提供的说明相同。接下来是一个旨在验证参与者阅读说明的问题。然后参与者执行实验任务。最后，参与者填写了一份简短的人口统计调查问卷，并对他们的参与表示感谢和补偿。第 2 节的程序与第 1 节相同。第 2 节在第 1 节后两周进行。

数据准备。首先，由于计算重测可靠性需要两个时间点的完整数据，因此删除了未完成两个会话的参与者的数据。接下来，按照研究 2-4 中的相同程序，在第 1 节或第 2 节中对超过 10% 的数据提供意外、不正确或缺失评级的参与者将从分析中删除。删除了 9 名参与者的数据，留下 110 名参与者的样本量。

报告摘要  
有关研究设计的更多信息，请参阅本文链接的《自然投资组合报告摘要》。



数据可用性

OSF 存储库 (<https://osf.io/7e6j5/>) 中提供了研究 1-5 收集和分析的所有数据（去识别化）和补充材料。本手稿中使用的档案数据已由原作者在 OSF 上公开提供。我们已在 OSF 存储库的自述文件中包含有关如何访问该数据并将其合并到我们的分析中的说明。

代码可用性

支持本研究结果的所有分析的代码可在 <https://osf.io/7e6j5/> 上获取。

参考

1. Weisman, R. 表现出悔恨：对错误定罪案件中表达与归因之间差距的反思。能。J。克里米诺。克里姆。司法 46, 121–138 (2004)。

2. Weisman, R. 存在与行为：司法上利用悔恨来构建品格和社区。苏克。腿。螺旋柱。18, 47–69 (2009)。

3. 皮科, A.等人。可见的眼泪如何影响观察者的判断和行为意图：真诚、悔恨和惩罚。J. 非语言行为。44, 215–232 (2020)。

4. Balsters, M. J.、Krahmer, E. J.、Swerts, M. G. 和 Vingerhoets, A. J. 情感眼泪有助于识别悲伤和感知到对社会支持的需要。进化。心理。11、147470491301100114 (2013) 。

5. Gregory, A. J.、Anderson, J. F. 和 Gable, S. L. 你不知道感觉如何：情绪感知的准确性可以预测支持的反应程度。情感 20, 343 (2020)。

6. Trobst, K. K.、Collins, R. L. 和 Embree, J. M. 情感在社会支持提供中的作用：性别、同理心和不满意的表达头发。J.苏克。个人的。关系。11, 45–62 (1994)。

7. Van Kleef, G. A.、De Dreu, C.K. 和 Manstead, A.S. 谈判中愤怒和快乐的人际影响。J.个人。苏克。心理。86、57 (2004) 。

8. Van Kleef, G.A. 国际愤怒手册 545–559 (2010)。

9. Dutta, A. 情感 AI 市场预计将升至 371 亿美元到 2026 年, (2022)。

10. Ribeiro, J. 商业情感人工智能简介, [medium.com/codex/an-introduction-to-emotional-ai-in-businessccb72268923c](https://medium.com/codex/an-introduction-to-emotional-ai-in-businessccb72268923c)> (2021) 。

11. Picard, R. W. 情感计算（麻省理工学院出版社，2000 年）。

12. Freed, S. AI 与人类思想和情感（Auerbach Publications，2019）。

13. Gendron, M. & Barrett, L. F. 《面部表情科学》15–36 (2017)。

14. Barrett, L. F.、Adolphs, R.、Marsella, S.、Martinez, A. M. 和 Pollak, S. D. 重新考虑情感表达：从人类面部动作推断情感的挑战。心理。科学。公共利益 20, 1–68 (2019)。

15. Barrett, L. F.、Mesquita, B. 和 Gendron, M. 情绪感知的背景。电流。导演。心理。科学。20, 286–290 (2011)。

16. Gendron, M.、Mesquita, B. 和 Barrett, L. F. 《牛津认知心理学手册》539–556 (2013)。

17. Hareli, S.、Elkabetz, S. 和 Hess, U. 从情绪表达中得出推论：情境信息性和情境的作用。情感 19, 200 (2019)。

18. Aviezer, H.、Hassin, R.、Bentin, S. 和 Trope, Y. 第一印象 255–286 (2008)。

19. De Gelder, B. & Vroomen, J. 通过耳朵和眼睛感知情绪。认知。表情。14, 289–311 (2000)。

20. Gendron, M.等人。哈扎狩猎采集者的情绪感知。科学。报告 10, 1–17 (2020)。

21. Gendron, M.、Roberson, D.、van der Vyver, J. M. 和 Barrett, L. F. 对面部表情的情感感知在文化上并不普遍：来自遥远文化的证据。情感 14, 251 (2014)。

22. 卡雷拉-勒维兰, P. 和费尔南德斯-多尔斯, J.-M. 上下文中的中性面孔：它们的情感意义和功能。J. 非语言行为。18, 281–299 (1994)。

23. Carroll, J. M. & Russell, J. A. 面部表情是否预示着特定的情绪？根据上下文判断面部表情。J.个人。苏克。心理。70, 205 (1996)。

24. Le Mau, T.等人。专业演员在照片中描绘情绪状态时表现出可变性，而不是刻板的表达方式。纳特。交流。12, 1–13 (2021)。

25. Aviezer, H.、Ensenberg, N. 和 Hassin, R. R. 面部情绪感知的内在情境化本质。电流。意见。心理。17, 47–54 (2017)。

26. Goodenough, F. L. & Tinker, M. A. 面部表情和刺激的口头描述在情绪判断中的相对效力。J.Comp. 心理。12, 365 (1931)。

27. Wallbott, H. G. 在上下文中和上下文之外：面部表情和上下文信息对情绪归因的影响。Br. J.苏克。心理。27, 357–369 (1988)。

28. Ekman, P.、Friesen, W. 和 Ellsworth, P. in P. Ekman, WV Freisen/PK Ellsworth. 纽约 269 (1972)。

29. Saxe, R. & Houlihan, S.D. 在心理理论的贝叶斯模型中形式化情感概念。电流。意见。心理。17、15-21 (2017) 。

30. Ong, D. C.、Zaki, J. 和 Goodman, N.D. 心理理论中情感推理的计算模型：回顾和路线图。顶部。认知。科学。11, 338–357 (2019)。

31. Ong, D. C.、Zaki, J. 和 Goodman, N.D. 情感认知：探索情感的外行理论。认知 143, 141–162 (2015)。

32. Durán, J. I. 和 Fernández-Dols, J.-M.情绪会导致他们预测的面部表情吗？表达与情感共现研究的荟萃分析。情感 21, 1550–1569 (2021)。

33. Hassin, R. R.、Aviezer, H. 和 Bentin, S. 本质上含糊不清：在上下文中情绪的面部表情。表情。修订版 5, 60–65 (2013)。

34. 卡尔比, M.等人。背景如何影响我们对情绪面孔的感知：库列绍夫效应的行为研究。正面。心理。8、1684 (2017) 。

35. 莫布斯, D.等人。库列绍夫效应：情境框架对情感归因的影响。苏克。认知。影响。神经科学。1, 95–106 (2006)。

36. Chen, Z. 和 Whitney, D. 推理情感跟踪 (IET) 揭示了上下文在情感识别中的关键作用。情感 22, 1185–1192 (2020)。

37. Houlihan, S. D.、Ong, D.、Cusimano, M. 和 Saxe, R. 关于情绪前因的推理：贝叶斯因果推理直觉心理理论。在过程中。认知年会科学学会卷。44, 854–861 (认知科学学会，2022 年) 。

38. Anzellotti, S.、Houlihan, S. D.、Liburd, S. Jr 和 Saxe, R. 利用面部表情和上下文信息来研究情绪的不透明表征。情感 21, 96 (2021)。

39. 卡罗尔, J.M. 和拉塞尔, J.A. 好莱坞情感表达中的面部表情。J.个人。苏克。心理。72, 164 (1997)。

40. Somerville, L. H. & Whalen, P. J. 作为刺激类别的先前经验混淆：使用情感面部表情的示例。苏克。认知。影响。神经科学。1, 271–274 (2006)。

41. Allen, V.D.、Weissman, A.、Hellwig, S.、MacCann, C. 和 Roberts, R.D. 使用项目反应理论开发情感理解情境测试 - 简报 (STEU-B)。个人的。独立的。不同。65, 3–7 (2014)。

42. Sharma, S.、Gangopadhyay, M.、Austin, E. 和 Mandal, M.K. 情商情境判断测试的开发和验证。国际。J. 选择。评估。21, 57–73 (2013)。

43. Edwards, J. M. 和 Endler, N. S. 压力情境评估。个人的。独立的。不同。 10, 7–10 (1989)。

44. Chen, E. 和 Matthews, K.A. 开发社会事件 (CAUSE) 视频的认知评估和理解。健康心理学。 22, 106 (2003)。

45. MacCann, C. & Roberts, R.D. 评估情商的新范式：理论和数据。情感 8, 540 (2008)。

46. Bonanno, G. A.、Maccallum, F.、Malgaroli, M. 和 Hou, W. K. 情境敏感性指数 (CSI)：衡量识别压力源情境线索是否存在的能力。评估 27, 261–273 (2020)。

47. Hamann, S. 和 Canli, T. 情绪处理的个体差异。电流。意见。神经生物学。 14, 233–238 (2004)。

48. Schlegel, K.、Grandjean, D. 和 Scherer, K. R. 介绍日内瓦情感识别测试：基于 Rasch 的测试开发示例。心理。评估。 26, 666 (2014)。

49. 李 T.-H.、崔 J.-S. & Cho, Y. S. 面部情绪感知的情境调节因个体差异而异。PLoS ONE 7, e32987 (2012)。

50. Schatz, H. 和 Ornstein, B. J. 《角色：演员表演》（Bulfinch Press, 2006 年）。

51. Schatz, H.、Edwards, E. 和 Ornstein, B. 《行为中的陷阱：演员表演》（Glitterati Incorporated, 2013 年）。

52. Kang, S. M. & Shaver, P. R. 情绪复杂性的个体差异：其心理影响。J. 个人。 72, 687–726 (2004)。

53. Baker, C.、Saxe, R. 和 Tenenbaum, J. 贝叶斯心理理论：联合信念-欲望归因建模。认知科学学会年会论文集 33 (2011)。

54. Goodman, N.D.、Baker, C. L. 和 Tenenbaum, J. B. 原因和意图：因果学习中的社会推理。认知科学学会第 31 届年会记录 2759–2764 (2009)。

55. Baker, C. L.、Jara-Ettinger, J.、Saxe, R. 和 Tenenbaum, J. B. 人类心智化中信念、欲望和感知的理性定量归因。纳特。哼。行为。 1, 1–10 (2017)。

56. Anderson, J. R. ACT：复杂认知的简单理论。是。心理。 51, 355 (1996)。

57. Wozny, D. R.、Beierholm, U. R. 和 Shams, L. 概率匹配作为感知中使用的计算策略。公共科学图书馆计算。生物。 6、e1000871 (2010)。

58. Murray, R. F.、Patel, K. 和 Yee, A. 后验概率匹配和人类感知决策。公共科学图书馆计算。生物。 11、e1004342 (2015)。

59. Gaissmaier, W. 和 Schooler, L. J. 概率匹配背后的智能潜力。认知 109, 416–422 (2008)。

60. Unturbe, J. & Corominas, J. 概率匹配涉及规则生成能力：一种处理概率的神经心理学机制。神经心理学 21, 621 (2007)。

61. Thomas, E. A. & Legge, D. 概率匹配作为检测和识别决策的基础。心理。修订版 77, 65 (1970)。

62. Kim, D. G. & Kim, H. C. 概率匹配和战略决策。J.行为。过期。经济。 98, 101850 (2022)。

63. Vul, E. 人类认知中的采样（麻省理工学院，2010 年）。

64. Vul, E.、Goodman, N.、Griffiths, T. L. 和 Tenenbaum, J. B. 一就完成了吗？从极少数样本中做出最佳决策。认知。科学。 38, 599–637 (2014)。

65. Goodman, N. D. 和 Tenenbaum, J. B. 认知概率模型。第二版。（2016）。

66. Austin, P. C. & Hux, J. E. 关于重叠置信区间的简要说明。J.瓦斯克。外科医生。 36, 194–195 (2002)。

67. Calvo, M. G.、Gutiérrez-García, A.、Fernández-Martín, A. 和 Nummenmaa, L. 情绪面部表情的识别与其在日常生活中的频率有关。J. 非语言行为。 38, 549–567 (2014)。

68. Thompson, R. J.、Springstein, T. 和 Boden, M. 明确情绪分化。苏克。个人的。心理。指南针 15, e12584 (2021)。

69. Barrett, L. F.重新考虑语境：复杂信号集合、关系意义和心理科学中的群体思维。是。心理。 77, 894–920 (2022)。

70. Aviezer, H.、Trope, Y. 和 Todorov, A. 区分强烈的积极情绪和消极情绪的是身体暗示，而不是面部表情。科学 338, 1225–1229 (2012)。

71. Todorov, A. & Duchaine, B. 在不识别面孔的情况下读取面孔的可信度。认知。神经心理学。 25, 395–410 (2008)。

72. Willis, J. 和 Todorov, A. 第一印象：在脸部暴露 100 毫秒后做出决定。心理。科学。 17, 592–598 (2006)。

73. Blair, I. V.、Chapleau, K. M. 和 Judd, C. M. 使用非洲中心特征作为存在诊断信息时的判断线索。欧元。J.苏克。心理。 35, 59–68 (2005)。

74. Rule, N. O.、Tskhay, K. O.、Freeman, J. B. 和 Ambady, N. 关于社会分类中面部外观和显性知识的交互影响。欧元。J.苏克。心理。 44, 529–535 (2014)。

75. Shen, X.、Mann, T. C. & Ferguson, M. J. 当心不诚实的面孔？：使用诊断行为信息更新基于面孔的隐含印象。J.Exp. 苏克。心理。 86、103888 (2020)。

76. Zhou, H.、Majka, E. A. & Epley, N. 推断观点与获取观点：低估站在他人立场的价值。心理。科学。 28, 482–493 (2017)。

77. Maher, S.、Ekstrom, T. 和 Chen, Y. 对快乐面部表情的感知敏感性更高。看法 43, 1353–1364 (2014)。

78. Calvo, M. G.、Avero, P.、Fernández-Martín, A. 和 Recio, G. 静态和动态情绪面孔的识别阈值。情感 16, 1186 (2016)。

79. Lee, M. D. & Webb, M. R. 认知个体差异建模。普西翁。公牛。修订版 12, 605–621 (2005)。

80. Navarro, D. J.、Griffiths, T. L.、Steyvers, M. & Lee, M. D. 使用狄利克雷过程建模个体差异。J.马斯。心理。 50, 101–122 (2006)。

81. Gendron, M. & Barrett, L. F. 情绪感知作为概念同步。表情。修订版 10, 101–110 (2018)。

82. Ferguson, F. J. 和 Austin, E. J. 成人样本中特质和能力情商与心理理论任务表现的关联。个人的。独立的。不同。 49, 414–418 (2010)。

83. Double, K. S.、Pinkus, R. T. & MacCann, C. 高情商的人在日常生活中表现出更灵活的情绪调节。情感。 22, 397 (2022)。

84. Horstmann, K. T. & Ziegler, M. 《威利人格评估手册》31–43 (2016)。

85. 劳斯曼, J.F. 等人。情境八钻石：情境特征主要维度的分类。J.个人。苏克。心理。 107, 677 (2014)。

86. Berent, I.、Feldman Barrett, L. 和 Platt, M. 情感推理中的本质主义偏见。正面。心理。 11、562666 (2020)。

87. Salerno, J. M. & Peter-Hagene, L.C. 愤怒和厌恶对道德愤怒和判断的互动影响。心理。科学。 24, 2069–2078 (2013)。

88. Russell, J.A. 和 Carroll, J.M. 双极性的凤凰：回复 Watson 和 Tellegen。心理牛。 125, 611–617 (1999)。

89. Wu, Y.、Baker, C. L.、Tenenbaum, J. B. & Schulz, L. E. 从情感表达中对信念和欲望的理性推断。认知。科学。 42, 850–884 (2018)。

90.Faul, F.、Erdfelder, E.、Lang, A.-G. & Buchner, A. G\* Power 3：针对社会、行为的灵活统计功效分析程序病毒学和生物医学科学。行为。资源。方法 39, 175–191 (2007)。

91. Champley, S. pwr：功率分析的基本函数。R 包版本 1.3-0 (2020)。

92. Marie Delacre, O.K.Routliers: 稳健的异常值检测。 R 包版本 0.0.0.3。（2019）。

93. Hofer, G. 样本量和相关系数的稳定性: Schönbrodt & Perugini (2013) 的复制。（2020）。

94. Schönbrodt, F. D. 和 Perugini, M. 在多大的样本量下相关性会稳定? J.Res. 个人的。 47、 609-612（2013）。

95. Goel, S.、 Jara-Ettinger, J.、 Ong, D. 和 Gendron, M. 情绪推理和个体差异中的线索整合建模。 <https://doi.org/10.17605/OSF.IO/7E6J5>（2024）。

致谢

我们感谢 Tuan Le Mau 和 Lisa Feldman Barrett 与我们分享他们的数据集和刺激。

作者贡献

S.G. 和 M.G.制定了学习理念。 S.G. 和 M.G.设计了最初的档案分析。 S.G. 和 M.G.设计了后续实验。 S.G. 和 M.G.计划的剩余分析和 J.J.-E.和 D.C.O.就数据分析和补充数据收集进行咨询。 S.G.收集并分析了所有实验数据。 S.G. 在 M.G.、 J.J.-E. 和 D.C.O. 的帮助下撰写了手稿。

利益竞争

作者声明没有竞争利益。

附加信息

补充信息 在线版本包含补充材料，网址为 <https://doi.org/10.1038/s41467-024-46670-5>。

信件和材料请求应发送给 Srishti Goel 或 Maria Gendron。

同行评审信息《自然通讯》感谢 Stefano Anzellotti 和其他匿名审稿人对这项工作的同行评审做出的贡献。同行评审文件可用。

重印和许可信息可在 <http://www.nature.com/reprints>

出版商说明施普林格·自然对于已出版地图和机构隶属关系中的管辖权主张保持中立。

开放获取本文根据知识共享署名 4.0 国际许可证获得许可，该许可证允许以任何媒介或格式使用、共享、改编、分发和复制，只要您对原作者和来源给予适当的认可，提供知识共享许可的链接，并指出是否进行了更改。本文中的图像或其他第三方材料包含在文章的知识共享许可中，除非材料的信用额度中另有说明。如果文章的知识共享许可中未包含材料，并且您的预期用途不受法律法规允许或超出了允许的用途，您将需要直接获得版权所有者的许可。要查看此许可证的副本，请访问 <http://creativecommons.org/licenses/by/4.0/>。

© 作者 2024