

无效预期诱发知觉偏差却提升元认知判断准确性*

罗铁勇 刘翠珍

(陕西师范大学心理学院, 西安 710062)

摘要 人类生活在一个充满不确定性的世界中, 为了能够及时响应和处理周围信息的变化, 大脑会持续对周围环境产生预期。这些预期能够引导我们的感知和行动, 然而, 不成熟的预测也可能导致决策偏差。本研究通过 5 个实验探讨无效预期如何影响知觉判断及其后的信心估计。实验 1、实验 2 和实验 3 分别探究了无效预期在高、中、低三种不同任务难度下对知觉判断和信心估计的影响。实验 4 引入无预测的基线条件以提供对照, 实验 5 通过分离预测和知觉判断的按键设置, 排除动作效应的干扰。4 个实验均发现(实验 3 除外), 被试倾向于将知觉刺激判断为与预测一致, 这表明无效预期引起了知觉偏差。在元认知效率(即信心估计的准确性)上, 5 个实验一致发现, 被试在匹配试次(预测刺激与知觉判断一致)中的表现显著优于不匹配试次, 表现出更高的元认知水平。综上, 无效预期虽然影响了初级的知觉判断导致了知觉偏差, 但却促进了更高级的元认知心理加工。

关键词 无效预期, 知觉判断, 自信心, 元认知

分类号 B842

1 引言

人类大脑对周围环境的知觉建构是一个主动且动态的过程。预测处理模型(Predictive Processing Model)将这一过程描述为主观预期的持续更新, 通过逐步缩小与感觉输入之间的误差, 生成与感觉输入最相符的假设, 从而促进知觉判断(Keller & Masic-Flogel, 2018)。想象你在超市看到一个人推着购物车, 最初你认为他是一位顾客。随后你注意到他在货架上整理商品, 并佩戴着工作证, 这时你意识到他其实是一名超市员工。在这一过程中, 你的判断依赖于对所观察线索与既有角色预期之间的匹配和调整。因此, 在充满不确定性的客观世界中, 大脑就像一台“预测机器”(Prediction Machine), 通过先验经验和环境线索形成有效预期, 以提高感知的速度和准确性, 从而应对复杂多变的环境(Clark, 2013)。

在知觉判断领域, 大量研究表明, 有效预期能够促进知觉判断。例如, Stein 和 Peelen (2015)发现, 在刺激呈现之前提供有关物体类别的信息, 会促进对物体的感知, 具体表现在与无信息基线条件和无效先验信息条件相比, 有效信息条件下个体对刺激的定位判断的表现更好。在视觉检测任务中, 提供关于目标对象类型的视觉或语义提示可以激活注意力模板, 形成视觉注意期望, 从而加快对该对象的检测速度(Hertz et al., 2020)。脑成像研究进一步表明, 先验信念与刺激信息的整合不仅发生在初级视知觉皮层(Kok et al., 2014), 还涉及更高级的决策相关脑区, 特别是前额皮层。与无预测事件相比, 在有预测事件中, 被试的背外侧前额皮层更加活跃, 并且当被试对即将到来的刺激做出预测时, 感觉区域间的有效联结显著增强(Rahnev et al., 2011; Summerfield & de Lange, 2014)。这些证据表明, 通过提供有效线索诱导被试产生预期, 可以加快知觉

收稿日期: 2025-01-16

* 国家自然科学基金青年项目(32300934)、西安市科协青年人才托举计划项目(959202313049)资助。

通信作者: 刘翠珍, E-mail: liucuizhen@snnu.edu.cn

反应速度并减少知觉错误。

预期通过自上而下的调控机制,对知觉判断发挥了重要作用。经典预测编码框架强调“预测误差最小化”原则,认为大脑通过不断更新先验模型以匹配感觉输入(Clark, 2013)。然而,这一理论主要解释有效预期情景的认知促进作用,即预期基于可靠的先验信息,如线索提示范式下的概率学习,这样的情境中被试能够不断获得线索更新信念。然而,在现实生活中,许多物理刺激往往模糊且高度不可预测,这些刺激通常伴随微妙且动态的变化,导致人们的预期更新往往滞后于刺激的变化(Thomas et al., 2023; Yon et al., 2019)。在这种情况下,人们不能基于外在线索形成有效预期,只能通过片面的经验或局限的内在推理自发地生成预期以指导行为。

以往的研究主要聚焦于探讨有效预期(valid/informative prediction),即预期线索本身包含有效的信息,如经典的波斯纳线索效应(Posner cueing effect),具有信息价值的空间线索可以显著提升知觉判断的准确性和速度(Posner, 1980)。然而,关于这种无效预期(invalid/non-informative prediction),即预期信息与随后刺激的呈现没有实质关联、并未包含任何有用线索的预期,是否仍然对知觉判断产生影响,目前尚缺乏系统研究。从理性人假设的角度来看(von Neumann & Morgenstern, 1944),无效预期不应影响随后的判断,但是在实际决策过程中,个体往往表现出非理性的决策倾向,比如手热效应(Oskarsson et al., 2009)或赌徒谬论(Crosos & Sundali, 2005)。有研究证据表明,一些无助于证据积累的认知因素也会影响个体对刺激的识别和决策准确性的评估,例如, Jazayeri 和 Movshon (2007)的研究发现,在序列方向辨别任务中,当被试完成方向判断后,他们对刺激运动方向的感知会系统性地偏离决策边界,这种偏差是由于大脑的主动优化解码策略,即优先整合调谐方向偏离边界的神经元信号导致。此外,确认偏误(Confirmation bias)现象也提示,个体做出分类选择后,其选择本身会引发后续决策的知觉偏差(Talluri et al., 2018)。事实上,知觉选择不仅依赖于当前的感觉输入,还受个体既往选择历史的影响(Urai et al., 2019)。所有这些知觉偏差可能是个体维持内部一致性的策略(Festinger, 1957; Luu & Stocker, 2018)。因此,即便个体所作的预期对即将呈现的刺激没有实际预测价值,这种无效预期也可能对随后的决策行为产生影响。

前人研究集中于证实预期对知觉判断的影响,

但关于预期对元认知的影响的探究却相对较少。元认知即“对认知的认知”,反映了个体对自身认知表现进行监控、协调和评估的能力(Flavell, 1979)。在知觉判断任务范式中,元认知通常用信心评分与知觉判断表现之间的匹配程度来衡量。元认知双阶段模型将元认知过程分为独立的两个阶段,第一阶段是“对象阶段”即个体具体的知觉、记忆等任务活动,第二阶段是“元”阶段,指对第一阶段的监控、评估(Nelson & Narens, 1990)。因此典型的实验流程包括两个阶段:被试在第一阶段先进行知觉判断,然后在第二阶段对其知觉判断的正确性进行信心评分。知觉判断的正确性具有客观的评价标准,将被试的主观信心评分与其知觉判断表现进行匹配,若知觉判断正确时伴随高信心,知觉判断错误时伴随低信心,则表明主观评估与客观表现的匹配程度较高,元认知表现较好,反之则说明元认知较差。被试的信心评分对知觉判断正误的区分程度被称作元认知敏感性,即元认知判断准确性(Fleming & Lau, 2014)。

根据贝叶斯模型,信心估计和知觉判断相似,都是基于对期望(先验概率)和输入信息(似然)的加权整合(Constant et al., 2023; Hu et al., 2021)。个体可能会根据预期通过自上而下的调节来调整他们对决策的信心,甚至修正最初的决策(Fleming et al., 2018)。Sherman 等人(2016)的研究就发现被试在预期与决策一致的情况下有更高的信心水平。目前为止,仅有一项研究直接探讨了预期与元认知敏感性(meta-d')的关系。Sherman 等人(2015)发现,在刺激出现概率较高的组块中,被试在判断刺激存在的试次中表现出更高的元认知敏感性,而在刺激出现概率较低的组块中,被试在判断刺激不存在的试次中表现出更高的元认知敏感性。预测即将到来的刺激,即使是随意做出的,也是一个自愿且自主的行为。当预测与知觉判断一致时,个体的大脑奖赏回路会被激活,产生更强的自我控制感(Hampton et al., 2007),而不一致情况可能引发认知冲突(Botvinick et al., 2001)。因此,在一致试次中,被试可能拥有更多的认知资源用于信心评估,从而提升信心判断的准确性。基于此,本研究假设:即使预期信息本身是无效的,当个体的预期与其知觉判断一致时,仍会提升后续信心判断的准确性,表现为更高的元认知效率。

综上,本研究通过5个行为实验系统探究了无效预期对知觉判断和元认知的影响。5个实验采用

相同的知觉任务, 即对模糊的面孔/房屋图片进行知觉判断。要求被试在知觉判断之前对即将出现的刺激类型(面孔/房屋)做出预测, 并告知被试仅凭直觉来做出预测, 预测与刺激呈现是两个独立的过程。通过阶梯滴定程序, 实验 1、实验 2 和实验 3 分别设置了高、中、低不同的 3 种任务难度, 以探究在不同任务难度下无效预期的效应是否存在。实验 4 增设了基线水平(无预测条件), 在预测与知觉判断一致、预测与知觉判断不一致和无预测 3 种条件下进行比较。由于前面 4 个实验中, 被试在预测阶段和知觉判断阶段使用相同的按键, 均用左箭头键代表面孔, 右箭头键代表房屋, 预期阶段的动作效应可能对知觉判断产生干扰。因此, 实验 5 通过分离预测按键与知觉判断按键, 以排除动作效应的干扰, 进一步验证实验结果的稳健性。

2 实验 1: 高难度条件下无效预期对知觉判断和元认知的影响

2.1 实验设计和被试

本实验采用单因素被试内设计, 旨在比较预测与知觉判断一致和不一致条件下, 被试的知觉表现和元认知表现。根据以往相似研究中 Liu 和 Yu (2024)报告的效应量大小($d = 0.57$), 采用 G*Power 3.1, 设置显著性水平 $\alpha = 0.05$, 统计检验量 $(1 - \beta) = 0.90$, 计算出本研究的计划样本量为 35 人(Faul et al., 2007)。本实验随机招募了 43 名被试。数据排除标准包括: (1)被试在匹配试次、不匹配试次或总体试次的知觉正确率低于 55%; (2)被试在匹配试次、不匹配试次或总体试次的元认知效率在平均数上下 3 个标准差之外; (3)被试的按键反应率低于 80%; (4)被试在预测或知觉判断过程中始终按同一个键。余下实验均采用同一数据筛选标准。两名被试因总体知觉判断正确率低于 55%被排除, 一名被试因在匹配试次中的正确率低于 55%而被排除, 一名被试因在不匹配试次中的正确率低于 55%而被排除, 一名被试因在不匹配试次中的元认知效率($meta-d'/d' = -10.55$)低于平均值减去 3 个标准差($M - 3SD = -4.97$)被排除。总共排除 5 名被试, 对剩下 38 名被试做进一步的统计分析($M = 20.92$ 岁, $SD = 1.98$ 岁; 19 名男性)。所有被试的视力均为正常或矫正正常。该研究已通过伦理委员会审查, 被试在实验开始前均签署了知情同意书, 实验结束后给予一定的报酬。

2.2 实验材料

本研究中选用的图片材料来自 Fleming 等人 (2012, 2010), 其中 10 张中性面孔图片(5 张男性)来自 Karolinska Directed Emotional Faces 图片集 (Lundqvist et al., 1988), 10 张房屋图片的建筑类型十分相似, 避免了建筑风格差异的潜在干扰。这些刺激材料被裁剪为相同尺寸并转换为灰度图像。对每张图像进行傅里叶变换, 生成 20 个幅度矩阵和 20 个相位矩阵, 并保存所有刺激材料的平均幅度矩阵。将单个面孔或房屋图像的相位矩阵与不同比例的白噪声 [$P(\text{noise})$] 重新组合, 并与平均幅度矩阵结合, 满足 $P(\text{image}) = 1 - P(\text{noise})$ 。其中, $P(\text{noise})$ 以 0.01 的步长进行调整, 从而为每张图片生成 99 张具有不同噪声水平的图片, 从最清晰、噪声最小的图片 $P(\text{image}) = 0.99$ 到最模糊、噪声最多的图片 $P(\text{image}) = 0.01$ 。

2.3 实验流程

实验采用面孔/房屋判断任务。被试需要判断呈现的模糊图像是面孔还是房屋, 然后评估其对该知觉判断的信心程度。在模糊图片呈现之前, 被试需要预测即将呈现的图像类别(面孔或房屋)。要求被试仅凭直觉对即将出现的刺激做出预测, 并告知被试每个试次中刺激的呈现类型是随机的, 既不受之前试次的影响, 也与预测阶段无关。

正式实验前包含 50 个练习试次。练习阶段的阶梯程序从 $P(\text{image}) = 0.50$ 开始, 房屋和面孔试次各 25 次, 两类刺激随机呈现, 并采用彼此独立的阶梯程序。正式实验以练习阶段最后达到的 $P(\text{image})$ 作为起始刺激强度, 继续进行阶梯程序。实验包括 4 个 block, 每个 block 包含 40 个试次, 共 160 个试次。

每个试次的流程如图 1 所示。首先呈现注视符, 为防止被试产生预期, 注视符的呈现时间在 500~1500 ms 间随机取样。随后呈现预测屏, 被试需要对下一屏将要出现的刺激图像类别进行预测, 键盘左箭头键表示面孔, 右箭头键表示房屋。被试需在 2 s 内做出预测, 反应会以白框高亮显示, 持续时间为 2 s 减去其反应时间。预测后呈现刺激图像, 呈现时间为 200 ms。随后被试需要在 2 s 内对刺激图像进行判断, 被试的反应会以白框高亮显示, 持续时间为 2 s 减去其反应时间。最后, 被试需要评估其对知觉判断准确性的信心程度(1 = “非常不自信”, 2 = “不自信”, 3 = “自信”, 4 = “非常自信”)。

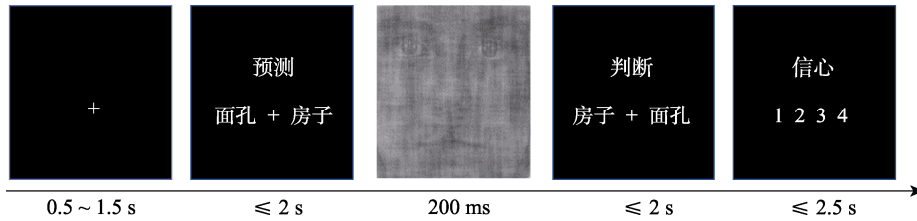


图1 实验流程图

实验采用“一升两降”的阶梯程序进行难度控制。在连续两次正确反应后,图像噪声增加一个步长,即 $P(\text{noise})$ 增加 0.01,使图片变得更模糊;在一次错误反应后,图片噪声减少一个步长,使图片变得更清晰。面孔和房屋刺激采用独立的阶梯程序进行调整。在该实验中,阶梯程序的步长调整不会重置计数。具体来说,在连续两次正确反应后,如果被试下一试次仍然做出了正确反应, $P(\text{noise})$ 将继续增加一个步长,而无需重新累计两次正确反应。直到被试做出错误的反应,计数才会重置。预实验结果表明,被试的平均正确率为 63%。使用阶梯程序能够平衡个体之间的表现差异,有效控制任务难度。为了检验正确率的稳定性,我们还分析了正确率在不同 block 中的变化趋势(见网络版附录)。

2.4 数据分析

根据预测与知觉判断的一致性,可将试次分为匹配试次组和不匹配试次组。匹配试次指预测和知觉判断一致的试次,即预测和知觉判断均为房屋或均为面孔的试次。不匹配试次则指预测为面孔而知觉判断为房屋,或者预测为房屋而知觉判断为面孔的试次。

首先,采用配对样本 t 检验比较匹配试次和不匹配试次的占比是否存在差异。如果匹配试次的占比显著高于不匹配试次,则证明存在知觉偏差。然后比较匹配试次组和不匹配试次组在正确率上是否存在差异,以检验预测一致性对知觉判断任务表现的影响。利用信号检测论(Signal Detection Theory),本实验将房屋刺激视为信号,当刺激图片为房屋且被试判断为房屋时,定义为击中(hit);当刺激图片为面孔,但被试判断为房屋时,定义为虚报(false-alarm),以此来计算被试在匹配试次和不匹配试次中的知觉敏感性(d')。

本研究采用 Maniscalco 和 Lau (2012)开发的元认知敏感性($meta-d'$)测量方法来评估元认知判断准确性。该方法将信心评分纳入信号检测论框架,将被试的信心评分视为二阶决策。若被试在正确的知觉判断后做出了高信心评分,则称作二类击中;若

在错误的知觉判断后做出了高信心评分,则称作二类虚报,进而可以计算出二阶决策敏感性,即元认知敏感性($meta-d'$)。需要注意的是, $meta-d'$ 的值受到知觉敏感性(d')的制约。由于不同被试的 d' 存在差异,这会造成二阶敏感性的差异,但这种差异并不是由被试元认知能力差异造成的。因此,为了排除个体知觉敏感性差异的影响,本研究使用比率($meta-d'/d'$)进行组间比较。该比率反映了一种控制了知觉表现差异后的元认知水平,被称为元认知效率。

本研究使用 JASP (版本 0.16.2.0)进行数据分析,除了使用频数统计方法外,还进行了贝叶斯分析,报告了支持原假设相对于备择假设的贝叶斯因子(BF_{01}),或支持备择假设相对于原假设的贝叶斯因子(BF_{10}),在两种情况下,更高的贝叶斯因子值都表示更强的统计证据(Masson, 2011)。所有贝叶斯分析均使用 JASP 中的默认先验分布。

2.5 实验结果

配对样本 t 检验结果显示,匹配试次的比例显著高于不匹配试次的比例($M \pm SD = 55\% \pm 8\%$ vs. $45\% \pm 8\%$), $t(37) = 4.35, p < 0.001$, Cohen's $d = 0.71$, $95\% \text{ CI} = [0.06, 0.16]$, $BF_{10} = 242.32$ (图 2a)。这一结果表明,无效预期会引起知觉偏差,被试更倾向于将刺激判断为与预测一致。被试在匹配试次与不匹配试次中的平均正确率没有显著差异($63\% \pm 6\%$ vs. $65\% \pm 6\%$), $t(37) = 1.29, p = 0.204$, Cohen's $d = 0.21$, $95\% \text{ CI} = [-0.04, 0.01]$, $BF_{01} = 2.66$ (图 2b)。

在信心水平上,相比于不匹配试次(2.19 ± 0.48),被试在匹配试次中有更高的信心评分(2.32 ± 0.48), $t(37) = 3.20, p = 0.003$, Cohen's $d = 0.52$, $95\% \text{ CI} = [0.05, 0.21]$, $BF_{10} = 12.31$ (图 2c)。在元认知效率上,相比于不匹配试次(0.83 ± 0.77),被试在匹配试次中表现出更好的元认知能力(1.30 ± 1.51), $t(37) = 2.11, p = 0.042$, Cohen's $d = 0.34$, $95\% \text{ CI} = [0.02, 0.92]$, $BF_{10} = 1.26$ (图 2d)。该结果说明预测与知觉判断的一致性能有效提升被试的元认知表现。除此之外,我们还分析了正确率与信心水平、元认知效率之间的相关关系(见网络版附录)。

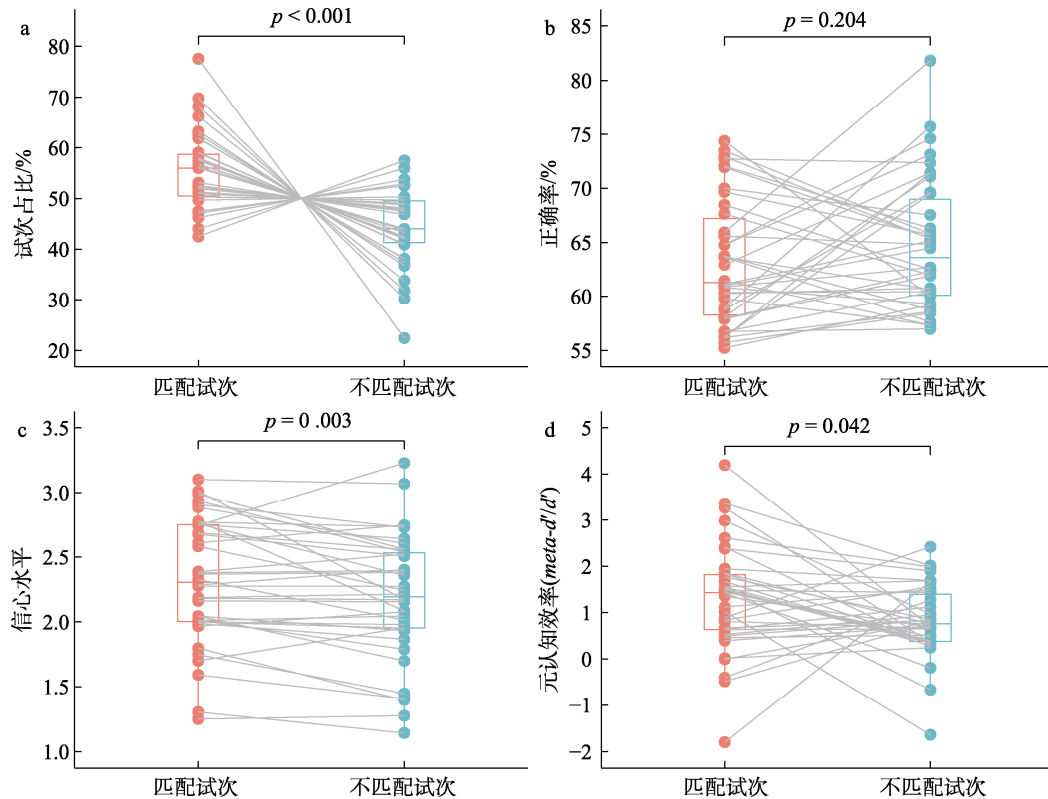


图 2 实验 1 中受访者在匹配试次和不匹配试次中的行为表现对比。(a)匹配试次占比显著高于不匹配试次占比。(b)受访者在匹配试次和不匹配试次中的正确率没有显著差异。(c)受访者在匹配试次上的信心水平显著高于不匹配试次。(d)受访者在匹配试次上的元认知效率显著高于不匹配试次。每对点代表每个受访者的值,箱型图中的上横杠线、中线、下横杠线分别代表上四分位数、中位数和下四分位数。彩图见电子版,下同。

3 实验 2: 中等难度条件下无效预期对知觉判断和元认知的影响

实验 1 结果显示,无效预期可以引起知觉偏差,即被试倾向于将知觉刺激判断为与预测类别一致。并且,这种预期与知觉的一致性会促进随后的信心评分表现。被试不仅在匹配试次中表现出更高的信心评分,而且信心评估的准确性(元认知效率)也显著优于不匹配试次。因此,尽管无效预期会影响初级的知觉判断导致知觉偏差,但同时也促进了更高级的元认知心理加工。

实验 1 通过连续的“一升两降”的阶梯程序,将知觉任务正确率控制在 63%左右,任务难度整体较高。当刺激模糊程度较高时,被试在进行知觉判断时能够利用的证据较少,因此可能容易受到预测的影响,依据预测的类别做出反应(姜英杰等, 2023)。那么,当知觉任务降低时,被试是否仍然会受到无效预期的影响,并进一步影响元认知判断呢?为了探讨这一问题,研究二采用了非连续的“一升两降”的阶梯程序,将知觉任务正确率控制在 71%左右,以降低任务难度,继续探讨无效预期对知觉判断和

元认知的影响。

3.1 实验设计和被试

实验 2 与实验 1 设计相同,因此计划样本量与实验 1 相同。共招募了 46 名被试,其中一名被试因在全部试次中只做出一种类型预测而被排除,一名被试在匹配试次中的元认知效率(3.231)大于平均数加 3 个标准差(3.111)而被排除,剩下 44 名被试进入数据分析($M = 21.02$ 岁, $SD = 2.05$ 岁),其中 16 名男性。

3.2 实验流程

实验 2 与实验 1 流程的唯一区别体现在阶梯程序的设置上,实验 2 仍然使用“一升两降”阶梯程序,但与实验 1 不同的是,每次步长调整后都会重新计数。具体来说,在 $P(\text{noise})$ 增加一个步长后,被试需要再次完成两次正确反应, $P(\text{noise})$ 才会继续增大。而在实验 1 中,当 $P(\text{noise})$ 增加一个步长后,如果被试在下一个试次中再次做出正确反应, $P(\text{noise})$ 会继续增加一个步长,无需重新累计。实验 2 的阶梯程序降低了任务难度,依据前人对阶梯程序的统计分析,实验 2 的正确率将收敛至 71%左右(Leek, 2001)。

3.3 数据分析

数据分析同实验 1。

3.4 实验结果

配对样本 t 检验结果显示, 匹配试次占比显著高于不匹配试次($53\% \pm 7\%$ vs. $47\% \pm 7\%$), $t(43) = 2.96$, $p = 0.005$, Cohen's $d = 0.45$, $95\% \text{ CI} = [0.02, 0.10]$, $\text{BF}_{10} = 7.16$ 。这表明即使任务难度降低, 被试仍然更倾向于做出与预测一致的反应, 无效预期的知觉偏差效应依然存在(图 3a)。被试在匹配试次与不匹配试次中的平均正确率没有显著差异($74\% \pm 8\%$ vs. $76\% \pm 6\%$), $t(43) = 1.41$, $p = 0.166$, Cohen's $d = 0.21$, $95\% \text{ CI} = [-0.05, 0.01]$, $\text{BF}_{01} = 2.44$ (图 3b)。

在信心水平上, 相比于不匹配试次(2.41 ± 0.44), 被试在匹配试次中有更高的信心评分(2.55 ± 0.49), $t(43) = 3.64$, $p < 0.001$, Cohen's $d = 0.55$, $95\% \text{ CI} = [0.06, 0.22]$, $\text{BF}_{10} = 39.56$ (图 3c)。在元认知效率上, 相比于不匹配试次(0.77 ± 0.47), 被试在匹配试次(1.04 ± 0.60)中表现出更好的元认知能力, $t(43) = 2.68$, $p = 0.010$, Cohen's $d = 0.40$, $95\% \text{ CI} = [0.07, 0.48]$, $\text{BF}_{10} = 3.80$ (图 3d)。无效预测与知觉判断的一致性仍然能增强被试的元认知能力。

4 实验 3: 低等难度条件下无效预期对知觉判断和元认知的影响

实验 1、实验 2 分别在高任务难度(63%)和中等任务难度(74%)下发现了无效预期可以引起知觉偏差, 并且预期与知觉的一致性促进了随后的信心评分表现, 被试在匹配试次中表现出更高的信心水平和更高的信心评分准确性。实验 3 旨在继续探讨在更低的任务难度下, 即当知觉信号更强的时候, 无效预期对知觉判断和元认知的影响是否仍然存在。

4.1 实验设计和被试

实验 3 与实验 1、实验 2 设计相同, 因此计划样本量与实验 1、实验 2 相同。共招募了 48 名被试, 其中一名被试在不匹配试次中的元认知效率(3.412)大于平均数加 3 个标准差(2.933)被排除, 剩下 47 名被试进入数据分析($M = 19.92$ 岁, $SD = 1.61$ 岁), 其中 10 名男性。

4.2 实验流程

实验 3 与实验 1、实验 2 流程的唯一区别体现在阶梯程序的设置上, 实验 3 使用“一升四降”的阶梯程序。具体来说, 在 P (noise)增加一个步长后,

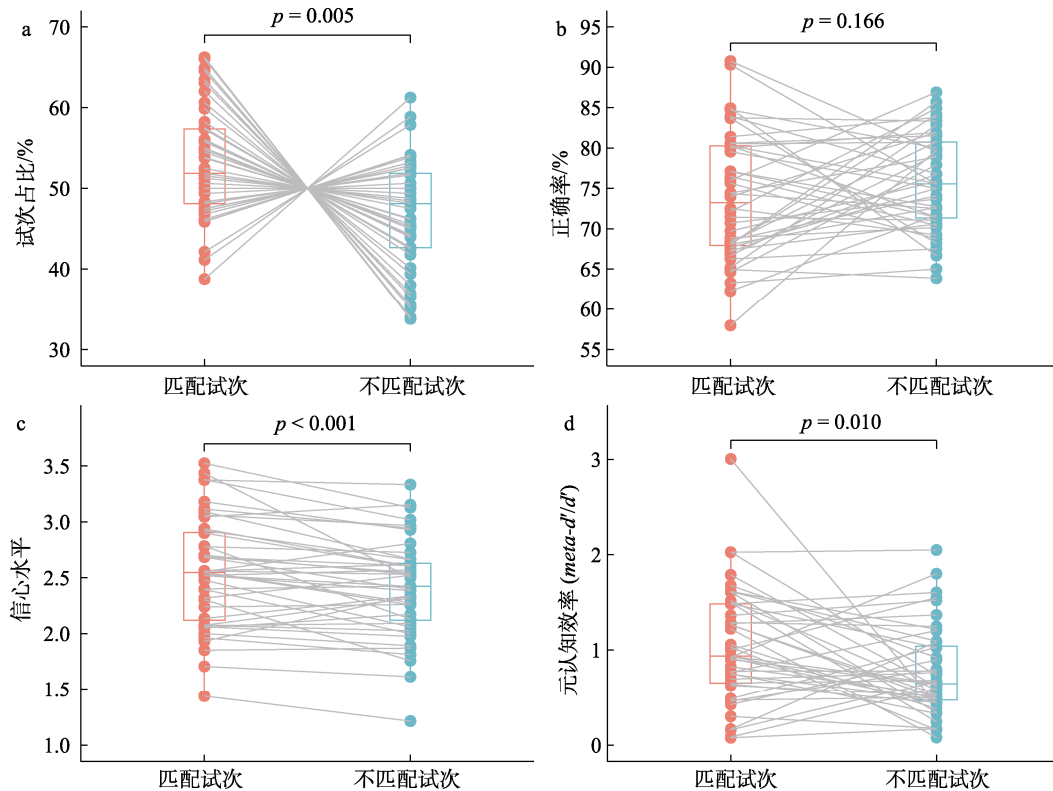


图 3 实验 2 中被试在匹配试次和不匹配试次中的行为表现对比。(a)匹配试次占比显著高于不匹配试次占比。(b)被试在不匹配试次和匹配试次中的正确率没有显著差异。(c)被试在匹配试次上的信心水平显著高于不匹配试次。(d)被试在匹配试次上的元认知效率高于一不匹配试次。

被试需要再次完成 4 次正确反应, $P(\text{noise})$ 才会继续增大, 而只要做错一次, $P(\text{noise})$ 就会下降。实验 3 的阶梯程序降低了任务难度, 依据前人对阶梯程序的统计分析, 实验 3 的正确率将在 84% 上下波动 (Leek, 2001)。

4.3 数据分析

数据分析同实验 1、实验 2。

4.4 实验结果

配对样本 t 检验结果显示, 匹配试次占比与不匹配试次占比无显著差异 ($51\% \pm 5\%$ vs. $50\% \pm 5\%$), $t(46) = 1.03$, $p = 0.307$, Cohen's $d = 0.15$, 95% CI = $[-0.01, 0.04]$, $\text{BF}_{01} = 3.82$ 。这表明任务难度低时, 被试没有表现出做出与预测一致的反应的倾向 (图 4a)。被试在匹配试次与不匹配试次中的平均正确率没有显著差异 ($85\% \pm 5\%$ vs. $86\% \pm 6\%$), $t(46) = 1.16$, $p = 0.253$, Cohen's $d = 0.17$, 95% CI = $[-0.03, 0.01]$, $\text{BF}_{01} = 3.38$ (图 4b)。

在信心水平上, 相比于不匹配试次 (2.86 ± 0.54), 被试在匹配试次中有更高的信心评分 (3.10 ± 0.45), $t(46) = 4.65$, $p < 0.001$, Cohen's $d = 0.68$, 95% CI = $[0.14, 0.35]$, $\text{BF}_{10} = 752.16$ (图 4c)。在元认知效率上,

匹配试次与不匹配试次差异显著, 相比于不匹配试次 (0.74 ± 0.60), 被试在匹配试次 (0.96 ± 0.42) 中表现出更好的元认知能力, $t(46) = 2.27$, $p = 0.028$, Cohen's $d = 0.33$, 95% CI = $[0.03, 0.42]$, $\text{BF}_{10} = 1.62$ (图 4d)。

与实验 1 (高任务难度) 和实验 2 (中等任务难度) 不同的是, 我们在实验 3 低任务难度中没有发现无效预期诱发知觉偏差, 这可能是因为任务难度降低, 被试能利用更多的刺激信息去进行知觉判断, 进而减少了对无效预期的依赖, 因此没有发生无效预期诱发的知觉偏差。这与 Yanagisawa (2016) 的实验结果一致, 预期权重与外部刺激信号强度呈负相关关系, 随着外部刺激信号强度的增强, 预期对于知觉判断的效应会越来越弱。在元认知表现上, 我们依旧发现了预期一致性会提高被试的元认知判断准确性。

5 实验 4: 参照于无预测基线对无效预期的效应作比较验证

实验 1 和实验 2 分别在高、中两种不同任务难度下 (正确率: 63% vs. 74%) 发现了无效预期可以引

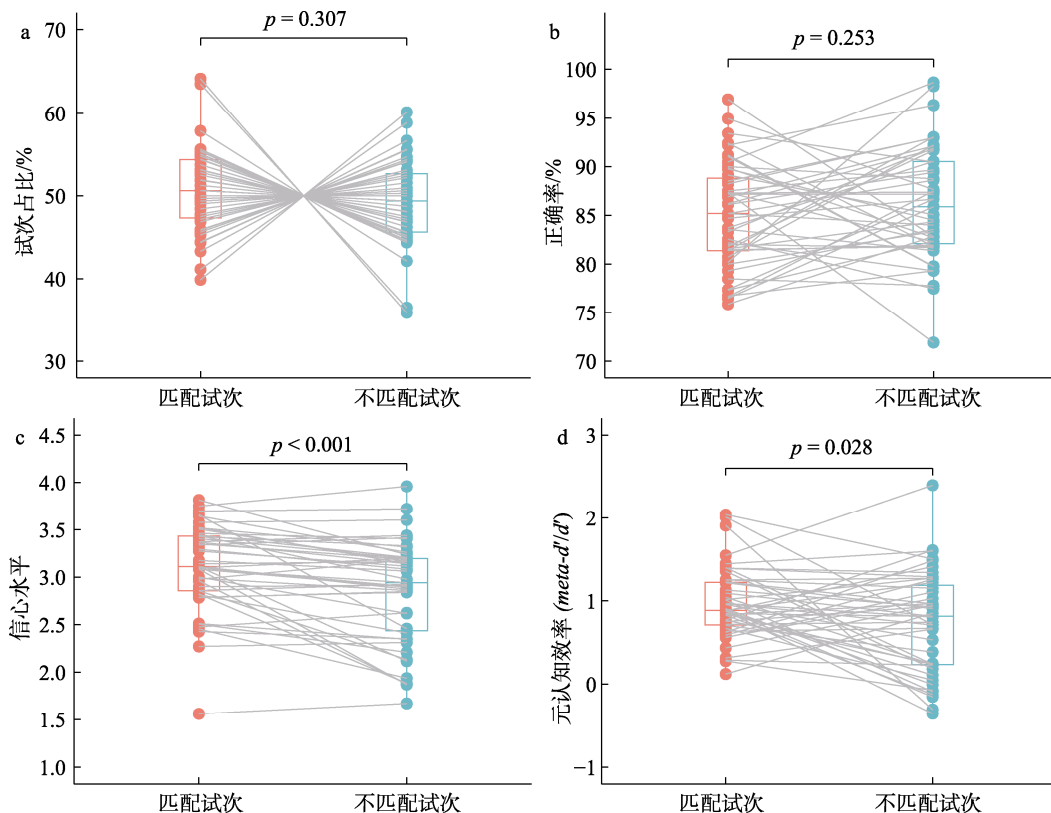


图 4 实验 3 中被试在匹配试次和不匹配试次中的行为表现对比。(a) 匹配试次占比与不匹配试次占比无显著差异。(b) 被试在匹配试次和不匹配试次中的正确率没有显著差异。(c) 被试在匹配试次上的信心水平显著高于不匹配试次。(d) 被试在匹配试次上的元认知效率显著高于不匹配试次。

起知觉偏差, 即被试更倾向于做出与预测一致的知觉判断。在实验 3 低任务难度(85%)中, 没有发现无效预期的知觉偏差效应。而在信心估计上, 3 个实验都发现了预期与知觉的一致性能够提高信心水平, 并对元认知判断准确性起促进作用。

实验 1、实验 2 和实验 3 都只设定了预测条件, 缺少对预测效果的基线参照, 因此实验 4 增设了无预测的基线条件, 在中等任务难度下比较被试在 3 种条件(匹配试次/不匹配试次/无预测试次)下的知觉表现和元认知是否存在差异。

5.1 实验设计和被试

实验 4 采用单因素被试内设计, 旨在比较预测与知觉判断一致、不一致以及无预测情况下, 被试的知觉表现和元认知表现。根据以往相似研究 Liu 和 Yu (2024)报告的效应量大小($d = 0.57$), 采用 G*Power 3.1, 设置显著性水平 $\alpha = 0.05$, 统计检验力 $1 - \beta = 0.90$, 计算出实验 4 至少需要样本量为 28 人(Faul et al., 2007)。实验 4 共招募 42 名被试。一名被试在匹配试次中的元认知效率($meta-d'/d' = 7.30$)大于平均数加 3 个标准差($M + 3 SD = 4.72$)而被排除, 最后共 41 名被试进入后续统计分析($M = 20.59$ 岁, $SD = 1.78$ 岁, 14 名男性)。

5.2 实验流程

与前 3 个实验不同, 在实验 4 中增加了无预测条件, 作为基线水平进行比较。正式实验前有两个练习 block, 预测试次 block 和无预测试次 block, 每个 block 各 30 个试次, 共 60 个练习试次。正式实验包括 8 个 block, 其中 4 个预测试次 block, 每个 block 有 50 个试次, 共 200 个试次; 4 个无预测 block, 每个 block 有 25 个试次, 共 100 个试次。预测试次流程如实验 1 所示; 无预测试次除了没有预测屏外, 其他流程与有预测试次相同。预测试次的 block 和无预测试次的 block 做了被试间的平衡。实验过程中预测 block 和无预测 block 交替呈现, 一半的被试先呈现有预测 block, 另一半被试先呈现无预测 block。实验 4 采用与实验 2 相同的阶梯程序进行难度控制。

5.3 数据分析

首先根据有无预测以及预测与知觉判断的一致性, 可以将试次分为匹配试次、不匹配试次和无预测试次。分别计算匹配试次和不匹配试次占总预测试次的比例, 用配对样本 t 检验比较被试的匹配试次占比和不匹配试次占比是否存在差异以验证知觉偏差是否存在。随后比较被试在匹配试次、不

匹配试次和无预测 3 种条件下的知觉行为表现和元认知表现是否存在差异, 进行单因素三水平重复测量方差分析。

5.4 实验结果

预期的知觉偏差仍然存在, 被试的匹配试次占比要显著高于不匹配试次占比($54\% \pm 7\%$ vs. $46\% \pm 7\%$), $t(40) = 3.90$, $p < 0.001$, Cohen's $d = 0.61$, $95\% CI = [0.04, 0.12]$, $BF_{10} = 76.46$ (图 5a)。除此之外, 结果还发现预期的一致性会影响被试知觉判断的正确率。重复测量方差分析结果显示, 预测一致性在正确率中的主效应显著, $F(2, 80) = 10.53$, $p < 0.001$, $\eta^2 = 0.21$, $BF_{10} = 405.26$; 事后比较(Holm)发现, 被试在匹配试次中的正确率($73\% \pm 6\%$)显著低于不匹配试次($78\% \pm 6\%$), $p < 0.001$; 在匹配试次与无预测试次($75\% \pm 6\%$)中的正确率没有显著差异, $p = 0.107$; 在不匹配试次中的正确率显著高于无预测试次, $p = 0.010$ (图 5b)。

在信心水平上, 重复测量方差分析表明, 预测一致性在信心评分上的主效应显著, $F(2, 80) = 3.55$, $p = 0.033$, $\eta^2 = 0.05$, $BF_{10} = 1.35$; 事后比较(Holm)发现, 被试在匹配试次中的信心水平(2.79 ± 0.60)显著高于不匹配试次(2.62 ± 0.70), $p = 0.030$; 在匹配试次与无预测试次中的信心水平(2.73 ± 0.64)没有显著差异, $p = 0.335$; 在不匹配试次与无预测试次中的信心水平也没有显著差异, $p = 0.200$ (图 5c)。

在元认知判断准确性上, 预测一致性在元认知效率上的主效应显著, $F(2, 80) = 8.20$, $p < 0.001$, $\eta^2 = 0.17$, $BF_{10} = 110.90$; 相比于不匹配试次(0.61 ± 0.50), 被试在匹配试次(1.08 ± 0.67)中表现出了更好的元认知能力, $p < 0.001$; 被试在匹配试次中的元认知效率与无预测试次(0.95 ± 0.52)没有显著差异, $p = 0.290$; 被试在无预测试次中的元认知效率显著高于不匹配试次, $p = 0.011$ (图 5d)。

6 实验 5: 排除动作效应后再验证无效预期的效应

实验 4 中, 无效预期引起的知觉偏差效应稳定存在; 跟预期不一致条件相比, 预期一致性对元认知判断准确性的促进作用依然存在, 这些发现与实验 1 和实验 2 结果一致。在前面 4 个实验中, 预测和知觉判断的按键设置是相同的, 均用左箭头键代表面孔, 右箭头键代表房屋, 预期阶段可能会对随后的知觉判断产生动作效应的干扰。因此, 在实验 5 中对预测和知觉判断的按键进行分离, 旨在探讨

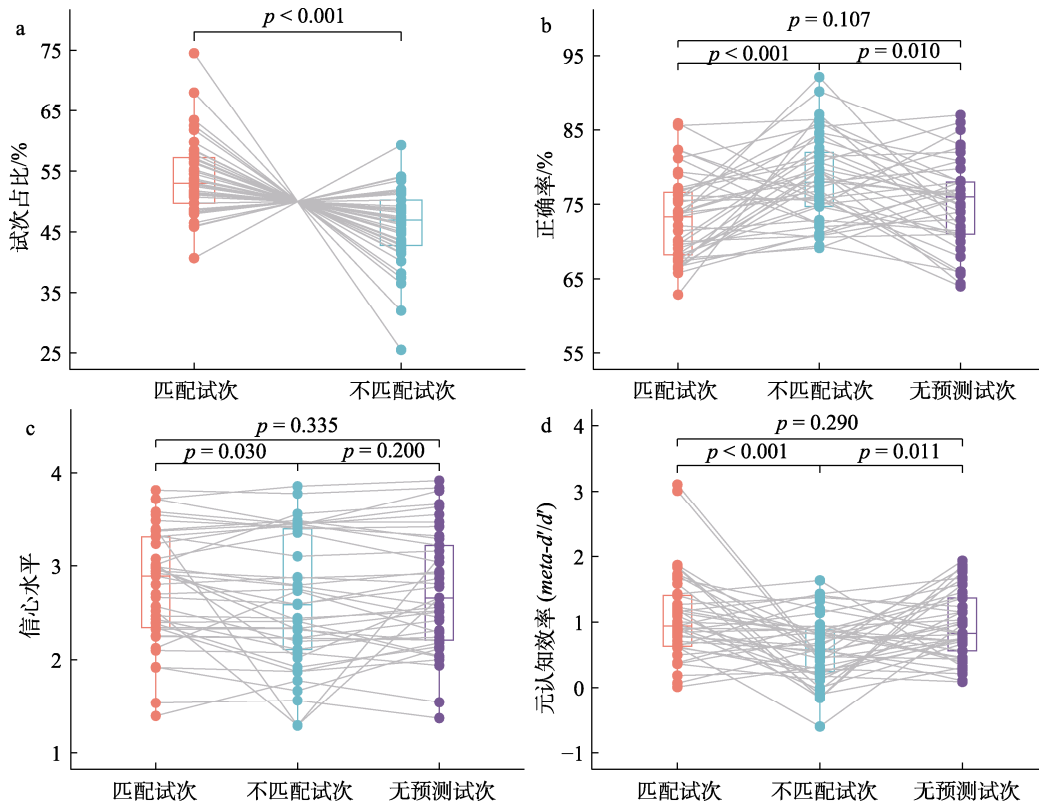


图 5 实验 4 中被试在匹配试次、不匹配试次以及无预测试次中的行为表现对比。(a)匹配试次占比显著高于不匹配试次占比。(b)被试在不匹配试次中的正确率显著高于匹配试次和无预测试次,匹配试次的正确率与无预测试次没有显著差异。(c)被试在匹配试次中的信心水平显著高于不匹配试次,但两者都与无预测试次的信心水平没有显著差异。(d)被试在匹配试次中的元认知效率显著高于不匹配试次,但与无预测试次没有显著差异,不匹配试次的元认知效率显著低于无预测试次。

排除可能的动作效应后,无效预期是否仍能对知觉判断和元认知表现产生影响。

6.1 实验设计和被试

实验设计与实验 4 相同。实验 5 共招募 44 名被试,其中一名被试因试次反应率低于 80% (75.33%)被排除,一名被试因在匹配试次中的正确率小于 55%而被排除,一名被试因在匹配试次中的元认知效率(0.7653)在平均数上下 3 个标准差之外(0.7492)而被排除,一名被试因总体元认知效率(-0.5886)在平均数上下 3 个标准差之外(-0.384)而被排除,最后剩 40 名被试进入后续分析($M = 20.85$ 岁, $SD = 1.96$ 岁, 14 名男性)。

6.2 实验流程

在实验 5 中,预测和知觉判断分别设置了不同的按键。在预测反应中,“Q”键和“W”键分别代表“面孔”和“房屋”;在知觉判断中,用左箭头键和右箭头键分别代表“面孔”和“房屋”。其余实验流程与实验 4 相同。

6.3 数据分析

与实验 4 相同。

6.4 结果

无效预测的知觉偏差效应依然存在,匹配试次占比显著高于不匹配试次占比($53\% \pm 7\%$ vs. $47\% \pm 7\%$), $t(39) = 2.87$, $p = 0.017$, Cohen's $d = 0.40$, $95\% \text{ CI} = [0.01, 0.09]$, $\text{BF}_{10} = 2.66$ (图 6a)。对知觉判断正确率做重复测量方差分析,结果显示预测一致性的主效应显著, $F(2, 78) = 3.09$, $p = 0.051$, $\eta^2 = 0.01$, $\text{BF}_{10} = 1.13$;不匹配试次($78\% \pm 6\%$)的正确率显著高于匹配试次($75\% \pm 7\%$), $p = 0.046$;匹配试次和无预测试次($77\% \pm 5\%$)在正确率上差异不显著, $p = 0.322$;不匹配试次和无预测试次之间也没有显著差异, $p = 0.322$ (图 6b)。因此,尽管被试在匹配试次中的知觉表现优于不匹配试次,但无论是在匹配试次还是不匹配试次中,被试的知觉表现都与无预测试次没有显著差异。

在信心水平上,预测与反应一致性的主效应显著, $F(2, 78) = 7.20$, $p = 0.004$, $\eta^2 = 0.16$, $\text{BF}_{10} = 22.14$;被试在匹配试次中的信心水平(2.56 ± 0.64)显著高于不匹配试次(2.31 ± 0.65), $p = 0.001$;匹配试次与无预测试次(2.47 ± 0.65)之间无显著差异, $p = 0.247$;

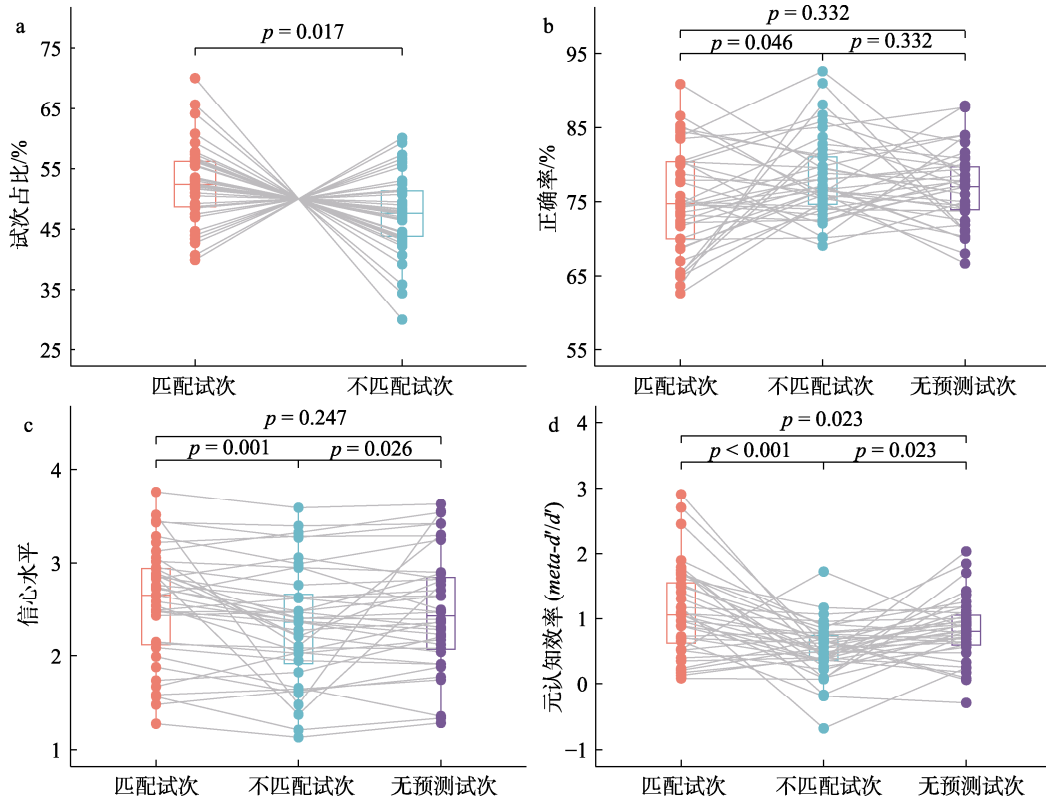


图 6 实验 5 中被试在匹配试次、不匹配试次以及无预测试次中的行为表现对比。(a)匹配试次占比显著大于不匹配试次占比。(b)被试在不匹配试次中的正确率显著高于匹配试次,但两者都与无预测试次的正确率没有显著差异。(c)被试在匹配试次中的信心水平显著高于不匹配试次,但与无预测试次的信心水平没有显著差异,不匹配试次中的信心水平显著低于无预测试次。(d)被试在匹配试次中的元认知效率显著高于不匹配试次和无预测试次,不匹配试次中的元认知效率显著低于无预测试次。

不匹配试次信心水平显著低于无预测试次, $p = 0.026$ (图 6c)。

对于元认知效率,预测与反应一致性的主效应显著, $F(2, 78) = 12.93, p < 0.001, \eta^2 = 0.25, BF_{10} > 1000$; 被试在匹配试次中的元认知效率(1.12 ± 0.69)显著高于不匹配试次(0.52 ± 0.41), $p < 0.001$; 被试在匹配试次中的元认知效率显著高于无预测试次中的元认知效率(0.81 ± 0.49), $p = 0.023$; 被试在不匹配试次的元认知效率显著低于无预测试次中的元认知效率, $p = 0.023$ (图 6d)。

7 讨论

本研究旨在探讨无效预期对知觉判断和元认知的影响。实验 1 采用经典的心理物理法滴定程序,将知觉任务控制在相对较难的水平(平均正确率 63%)。结果显示,无效预期能够引起知觉偏差,被试更倾向于做出与预期一致的知觉判断;且更重要的是,无效预期能够提升元认知判断的准确性,在做出与预期一致的知觉判断时,被试表现出更高的元认知水平。实验 2 将知觉任务控制在中等难度水

平(74%)后,进一步验证了实验 1 的结果;实验 3 将知觉任务控制在低等难度水平后(85%),无效预期诱导的知觉偏差效应消失,但是被试在与预期一致的试次中仍然表现出更高的元认知水平;实验 4 在中等任务难度引入了基线水平(无预测条件),结果发现预期诱导的知觉偏差效应仍然存在。跟预期不一致条件相比,预期一致条件能够提升元认知能力。然而,预期一致条件下的元认知水平与无预测条件相比没有显著差异,但预期不一致条件下的元认知能力却低于无预测试次。这表明,与预期不一致可能会损害被试的元认知判断。在实验 5 中,排除了预期阶段动作效应的潜在干扰后,预期一致条件下的元认知能力高于预期不一致条件和无预测基线,这进一步验证了预期一致性对元认知水平的促进作用。综上,在刺激信号较弱时(中高任务难度),无效预期会引发知觉偏差,而在更高级的元认知决策中,无论刺激信号强弱,无效预期始终能够提高元认知判断准确性。

本研究的第一个发现是,无效预期会影响知觉判断,在刺激信号较弱时(中高任务难度),被试更

倾向于做出与无效预期一致的知觉判断,表现为匹配试次的占比显著高于不匹配试次,这一结果与以往关于有效预期的研究一致(Stein & Peelen, 2015)。虽然,有效预期对知觉判断起促进作用,而无效预期对知觉判断是促进还是阻碍作用尚不清晰。从理论上讲,无效预期会使知觉判断表现变差,因为刺激本身是独立于预期的,对刺激的判断不应受到预期的影响,预期带来的影响属于额外的噪音干扰。本研究结果在一定程度上支持了这一观点。在实验 4 和实验 5 中,被试在不匹配试次中的正确率显著高于匹配试次。虽然实验 1 (高任务难度)、实验 2 (中等任务难度)和实验 3 (低任务难度)中的匹配试次正确率和不匹配试次的正确率差异并不显著,但是匹配试次正确率低于不匹配试次正确率的趋势始终存在(实验 1: $63\% \pm 6\%$ vs. $65\% \pm 6\%$, $p = 0.204$; 实验 2: $74\% \pm 8\%$ vs. $76\% \pm 6\%$, $p = 0.166$; 实验 3: $85\% \pm 5\%$ vs. $86\% \pm 6\%$, $p = 0.253$)。一个可能原因是本研究采用了动态滴定程序对被试正确率进行了控制,被试在整体正确率上的波动并不大(实验 1: $SD = 4.6\%$, 实验 2: 5.3% , 实验 3: 4.1%),使得匹配试次与不匹配试次的正确率差异也较小(实验 1: 1.6% , 实验 2: 2.0% , 实验 3: 1.2%)。对于实验 3,随着任务难度的降低,个体有足够多的刺激证据去做出知觉判断,因此无效预期的作用减弱,这可能也是实验 3 中匹配试次与不匹配试次正确率差异没有达到显著的原因。值得注意的是,近期研究也表明,预期的促进效应并不总是稳定存在。例如,Thomas 等人(2023)发现,预期的作用与预期本身的有效性密切相关。在预期有效性较高(75%)时,预期对知觉判断表现出显著的促进作用;当预期有效性降低至 67%时,这种促进效应减弱;而在预期有效性为 50% (即无效预期)时,预期效应消失。然而,Thomas 等人的研究采用的知觉任务过于简单,被试的平均正确率超过 90%,且未使用滴定程度控制任务难度,这可能是该研究未发现无效预期影响知觉判断的原因。综上,本研究的结果表明,无效预期会引发知觉偏差,并可能对知觉判断产生阻碍作用。

本研究的第二个发现是,无效预期会影响信心评分,当被试将刺激判断为与预期一致时,其主观感受的信心水平更高。这一结果与 Sherman 等人(2016)的研究一致,该研究表明,预期与决策的一致性程度越高,信心水平越高;右侧额叶下回(rIFG)表征了对预期与决策的不匹配,预期信息在rIFG 脑区被整合到信心评分中。Liu 和 Yu (2024)

的研究也发现了预期和知觉判断之间的不一致会改变主观信心的神经表征,相比于与预期一致的试次,在与预期不一致的试次中,信心评分相关的核壳活动明显减少,导致不一致试次中的低信心。除此之外,还有研究提出信心评分反映的是当前知觉判断与被试在过去可能做过的其他决策之间自我一致性的程度(Caziot & Mamassian, 2021),因此,当被试做出与之前预期一致的知觉判断时,往往会有更高的自我一致性水平,进而产生了更高的信心评分。

本研究最重要的发现是,无效预期会提升元认知判断的准确性,被试在匹配试次上的元认知效率显著高于不匹配试次。这一结果与 Sherman 等人(2015)的发现一致,表明预期能够提高元认知水平。但是, Sherman 等人的研究中,被试并未实际进行预测,而是通过操纵刺激出现的频率来影响被试的主观期望;此外,他们使用了二分信心判断(即“有信心”或“猜测”),而本研究采用了四分的信心评分,能够更精确地拟合元认知参数。尽管无效预期并不包含对决策有用的信息,但仍然也是个体的一个自愿且自主的行为,当预测与知觉判断一致时,个体的大脑奖赏回路会被激活,产生更强的自我控制感(Hampton et al., 2007)。相比于不一致情况可能对被试造成认知冲突(Botvinick et al., 2001),在一致试次中,被试能够有更多的认知资源进行信心判断,因此一致试次中的信心判断准确性会高于不一致试次。除此之外,预期与自我效能感关系密切,预期的满足往往能够增强个体的自我效能感(Abaszadeh et al., 2024),提升决策的动机水平(Walkey et al., 2013),因此可能促进被试信心评估的准确性。然而,这一机制的解释有待进一步实验验证。

本研究发现先验预期会阻碍知觉判断,但促进元认知决策,这一发现与最近的一项研究中发现的预期对于知觉决策和元认知决策影响模式相同(Constant et al., 2023)。在该研究中, Constant 等人(2023)设置了两个连续的知觉判断任务,第一个知觉判断的结果(正确/错误)与第二个知觉刺激(向左/向右)存在特定关联,因此第一个知觉判断能够提供关于第二个知觉刺激的先验信息,引发对第二个知觉刺激的预期。随后,被试需要对第二个知觉判断做出信心评分。结果发现,知觉判断和信心评分对先验信息有着不同的加工过程:知觉加工过程无法充分利用先验信息,先验信息反而作为噪音阻碍了知觉判断,而在更高级的元认知加工过程中,先

验信息能够被更大程度地整合,元认知效率得到提高。Balsdon 等人(2021)在神经层面上也发现了知觉判断与信心评分信息加工过程的不对称性,在元认知判断过程中,上顶叶和眶额皮层存在独立于知觉判断的神经环路激活,相比于知觉决策过程,信心评分过程中证据累积的来源也更加多样,更易受到自上而下的调控影响。总之,这些结果表明,尽管知觉判断和信心评分是密切相关的认知加工过程,但其内在机制存在差异。研究结果为进一步探索信心评分和知觉判断的分离提供了重要启示。

本研究首次发现无效预期仍然能够影响知觉判断,进一步拓展了预测编码理论。经典预测编码框架强调“预测误差最小化”原则,认为大脑通过不断更新先验模型以匹配感觉输入(Clark, 2013; Hu et al., 2021)。然而,这一理论主要解释有效预期(即与刺激统计关联的预期)的认知促进作用(Kok et al., 2014)。实验1和实验2的结果发现,在中高任务难度下(平均正确率75%和63%),无效预期会诱导被试产生知觉判断偏差,即更倾向于做出与无效预期一致的知觉判断,并且无效预期会作为噪音干扰感觉证据累计过程,表现为匹配试次正确率的下降。这与Yon等人(2019)提出的“顽固科学家”假说也形成了实证呼应,即大脑在感觉证据不足时,会顽固地维持错误预期。然而,无效预期对知觉判断的影响存在一定的边界条件。实验3结果表明,在知觉任务较为容易时(平均正确率85%),无效预期并未影响随后的知觉判断。这可能是因为任务难度降低,被试能利用更多的刺激信息去进行知觉判断,进而减少了对无效预期的依赖,从而无效预期没有诱发知觉偏差的发生。这与Yanagisawa (2016)的实验结果一致,预期权重与外部刺激信号强度呈负相关关系,随着外部刺激信号强度的增强,预期对于知觉判断的效应会越来越弱。

本研究为知觉判断和信心估计的双阶段分离模型提供了新的行为学证据。在传统的心理物理学模型中(Charles et al., 2014; Macmillan & Creelman, 2005),研究者假设个体对自身信心的估计是基于与感知决策相同的内部感官证据。这一理论观点已在动物研究中得到验证(Kepecs et al., 2008), Kiani 和 Shadlen (2009)也发现了单个神经元的放电率既能反映信心水平,又能反映感知决策。但是,近年来,越来越多的研究开始发现信心评分和知觉判断过程中利用的内部感官证据并不相同。有研究发现信心评分几乎完全取决于与反应一致的证据(即支

持已选类别的证据),而知觉判断依赖于支持与反对所选刺激类别的证据之间的平衡(Maniscalco et al., 2016; Sun et al., 2025)。Peters 等人(2017)使用颅内电生理记录和机器学习技术证明了在面孔/房屋辨别任务中,知觉决策和信心估计依赖于时空可分离的神经表征。与我们的研究结果类似的是,Vlassova 等人(2014)发现无意识信息改变知觉判断准确性,但不影响信心水平。在元认知敏感性上,我们的研究表明,无效预期虽然阻碍了知觉判断,却提升了信心判断的准确性。无效预期对知觉决策过程和信心判断准确性的不同作用支持了 Fleming 和 Dolan (2012)提出的知觉判断和信心判断双阶段分离模型。

8 结论

本研究旨在探讨完全自发的、直觉性的无效预期对知觉判断和元认知的影响。结果表明:(1)当刺激信号强度较弱时(中高等任务难度),无效预期导致知觉偏差,个体倾向于将刺激知觉为与预期一致;(2)无效预期影响主观信心,当个体的知觉判断与预期一致时,其对判断的信心水平更高;(3)无效预期提升元认知判断的准确性,个体在匹配试次中的元认知效率显著高于不匹配试次。尽管无效预期引发知觉偏差,但它促进了个体在元认知层面的加工。这些结果首次揭示了无效预期对于低级知觉判断和高级元认知功能的不同作用,为理解预期、知觉与元认知三者之间的关系提供了重要理论支撑,同时也为优化认知决策、提升元认知能力提供了实践启示。

参 考 文 献

- Abaszadeh, H., Amani, M., & Pordanjani, T. R. (2024). The relationship between motivational-cognitive variables, academic self-efficacy of students mediated by parent's educational expectations, parent-child interaction, and teacher-student interaction. *Learning and Motivation, 86*, 101983.
- Balsdon, T., Mamassian, P., & Wyart, V. (2021). Separable neural signatures of confidence during perceptual decisions. *eLife, 10*, 68491.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624-652.
- Caziot, B., & Mamassian, P. (2021). Perceptual confidence judgments reflect self-consistency. *Journal of Vision, 21*(12), 8.
- Charles, L., King, J. R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *Journal of Neuroscience, 34*(4), 1158-1170.
- Clark, A. (2013). Whatever next? Predictive brains, situated

- agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, 36(3), 181–204.
- Constant, M., Pereira, M., Faivre, N., & Filevich, E. (2023). Prior information differentially affects discrimination decisions and subjective confidence reports. *Nature Communications*, 14(1), 5473.
- Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, 30(3), 195–209.
- 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.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1594), 1338–1349.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in human neuroscience*, 8, 443.
- Fleming, S. M., van der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, 21, 617–624.
- Fleming, S. M., Whiteley, L., Hulme, O. J., Sahani, M., & Dolan, R. J. (2010). Effects of category-specific costs on neural systems for perceptual decision-making. *Journal of Neurophysiology*, 103(6), 3238–3247.
- Hampton, A. N., Adolphs, R., Tyszka, J. M., & O'Doherty, J. P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron*, 55(4), 545–555.
- Hertz, U., Blakemore, C., & Frith, C. D. (2020). I haven't a clue! Expectations based on repetitions and hints facilitate perceptual experience of ambiguous images. *Journal of Experimental Psychology: Human Perception and Performance*, 46(8), 831–846.
- Hu, X., Zheng, J., Su, N., Fan, T., Yang, C., Yin, Y., ... Luo, L. (2021). A Bayesian inference model for metamemory. *Psychological Review*, 128(5), 824–855.
- Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446, 912.
- Jiang, Y. J., Ma, X. X., Jiang, Y. T., Ren, J. M., & Long, Y. T. (2023). The effect of after-encoding rewards on agenda-based learning: The role of reward expectation and reward outcome. *Acta Psychologica Sinica*, 55(4), 542–555.
- [姜英杰, 马潇潇, 姜元涛, 任吉梅, 龙翼婷. (2023). 编码后奖赏影响基于议程的学习: 奖赏预期和结果的作用. *心理学报*, 55(4), 542–555.]
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, 100(2), 424–435.
- Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227–231.
- Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928), 759–764.
- Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, 26(7), 1546–1554.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8), 1279–1292.
- Liu, C., & Yu, R. (2024). Neural mechanisms underpinning metacognitive shifts driven by non-informative predictions. *NeuroImage*, 296, 120670.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). *Karolinska Directed Emotional Faces (KDEF)* [Database record]. APA PsycTests.
- Luu, L., & Stocker, A. A. (2018). Post-decision biases reveal a self-consistency principle in perceptual inference. *eLife*, 7, 33334.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430.
- Maniscalco, B., Peters, M. A., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception & Psychophysics*, 78(3), 923–937.
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43(3), 679–690.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, 26, 125–173.
- Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, 135(2), 262–285.
- Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., ... Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behavior*, 1, 139.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Rahnev, D., Lau, H., & de Lange, F. P. (2011). Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *Journal of Neuroscience*, 31(29), 10741–10748.
- Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, 35, 53–65.
- Sherman, M. T., Seth, A. K., & Kanai, R. (2016). Predictions shape confidence in right inferior frontal gyrus. *Journal of Neuroscience*, 36(40), 10323–10336.
- Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, 144(6), 1089–1104.
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews. Neuroscience*, 15(11), 745–756.
- Sun, F., Ni, Y., Lu, W., Su, J., Wang, S., & Wan, X. (2025). Confidence bias prescribes the neurocomputational mechanism of decision-making. *Cell Reports*, 44(5), 115563. Advance online publication.
- Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, 28(19), 3128–3135.
- Thomas, E. R., Rittershofer, K., & Press, C. (2023). Updating perceptual expectations as certainty diminishes. *Cognition*,

- 232, 105356.
- Urai, A. E., de Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice history biases subsequent evidence accumulation. *eLife*, 8, 46331.
- Vlassova, A., Donkin, C., & Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), 16214–16218.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Walkey, F. H., McClure, J., Meyer, L. H., & Weir, K. F. (2013). Low expectations equal no expectations: Aspirations, motivation, and achievement in secondary school. *Contemporary Educational Psychology*, 38(4), 306–315.
- Yanagisawa, H. (2016). A computational model of perceptual expectation effect based on neural coding principles. *Journal of Sensory Studies*, 31(5), 430–439.
- Yon, D., de Lange, F. P., & Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences*, 23(1), 6–8.

Invalid expectations induce perceptual bias yet enhance the accuracy of metacognitive judgments

LUO Tieyong, LIU Cuizhen

(School of Psychology, Shaanxi Normal University, Xi'an 710065, China)

Abstract

Humans live in a world filled with uncertainty. To adapt to and process changes in their surroundings, the brain continuously generates predictions about the environment. Predictions can guide perception and action. However, given the ambiguous and highly unpredictable nature of the external world, individuals often lack sufficient information to form accurate predictions. This study investigates how non-informative prediction, specifically invalid prediction, shapes perceptual judgments and subsequent confidence estimates through five experiments.

In a face/house judgment task, participants were required to determine whether a presented blurry image depicted a face or a house, and then rate their confidence in that perceptual judgment. Before the image was displayed, participants predicted the category of the upcoming image (face or house). Experiment 1 ($n = 38$, perceptual accuracy $\approx 64\%$), Experiment 2 ($n = 44$, perceptual accuracy $\approx 75\%$) and Experiment 3 ($n = 47$, perceptual accuracy $\approx 85\%$) systematically manipulated task difficulty to examine how non-informative predictions modulate perceptual judgments and confidence ratings. Experiment 4 ($n = 41$) introduced a condition without predictions to provide a baseline for evaluating the predictive effects. In Experiment 5 ($n = 40$), the response keys for predictions and perceptual judgments were separated to eliminate potential action-related effects.

The results revealed that: (1) non-informative predictions induced perceptual biases under high and moderate task difficulty conditions (except in Experiment 3 with low difficulty), systematically biasing individuals to align their perceptual judgments with prior predictions; (2) non-informative predictions affected subjective confidence, as participants reported higher confidence in trials where their perceptions aligned with their predictions compared to trials where they did not; and (3) non-informative predictions enhanced the accuracy of metacognitive judgments, with individuals exhibiting greater metacognitive efficiency when their perceptual judgments matched predictions compared to mismatched trials.

In summary, while non-informative prediction induces perceptual biases, it enhances metacognitive judgments. These findings highlight the distinct effects of non-informative predictions on perceptual judgments and metacognitive functions. The study contributes to a deeper understanding of the interplay among expectations, perception, and metacognition. Finally, it offers practical implications for optimizing cognitive decision-making and improving metacognitive accuracy.

Keywords non-informative prediction, perceptual judgment, confidence, metacognition

附录

1 正确率稳定性检验：正确率在不同 block 的变化趋势分析

在实验 1~5 中，我们用单因素方差分析检验了被试平均正确率在不同 block 中的变化情况。实验 1、实验 2 和实验 3 共 4 个 block，每个 block 含 40 个试次；实验 4 和实验 5 共 8 个 block，其中 4 个无预测 block，每个 block 含 25 个试次，4 个有预测 block，每个 block 含 50 个试次，实验中无预测 block 和有预测 block 交替出现，先后顺序在被试间进行平衡。具体描述性统计见表 S1。

表 S1 实验 1~5 中被试在不同 block 中的平均正确率和标准差

	实验 1	实验 2	实验 3	实验 4	实验 5
Block1	0.63 ± 0.09	0.74 ± 0.09	0.82 ± 0.10	0.73 ± 0.11	0.78 ± 0.11
Block2	0.63 ± 0.09	0.75 ± 0.08	0.85 ± 0.07	0.74 ± 0.08	0.77 ± 0.09
Block3	0.66 ± 0.08	0.76 ± 0.07	0.87 ± 0.07	0.75 ± 0.10	0.76 ± 0.08
Block4	0.63 ± 0.09	0.73 ± 0.08	0.88 ± 0.05	0.76 ± 0.08	0.77 ± 0.07
Block5				0.76 ± 0.10	0.75 ± 0.08
Block6				0.76 ± 0.08	0.75 ± 0.09
Block7				0.75 ± 0.09	0.78 ± 0.09
Block8				0.76 ± 0.09	0.77 ± 0.07

结果显示(图 S1)，在实验 1 中，block 的主效应不显著， $F(3, 111) = 1.20, p = 0.310, \eta^2 = 0.03$ ；在实验 2 中，block 的主效应不显著， $F(3, 129) = 1.45, p = 0.230, \eta^2 = 0.03$ ；在实验 3 中，block 的主效应显著， $F(3, 138) = 8.17, p < 0.001, \eta^2 = 0.15$ ，事后方差分析发现(Bonferroni 矫正)，被试在 block1 中的正确率显著低于 block2 ($p = 0.051$)、block3 ($p = 0.003$)

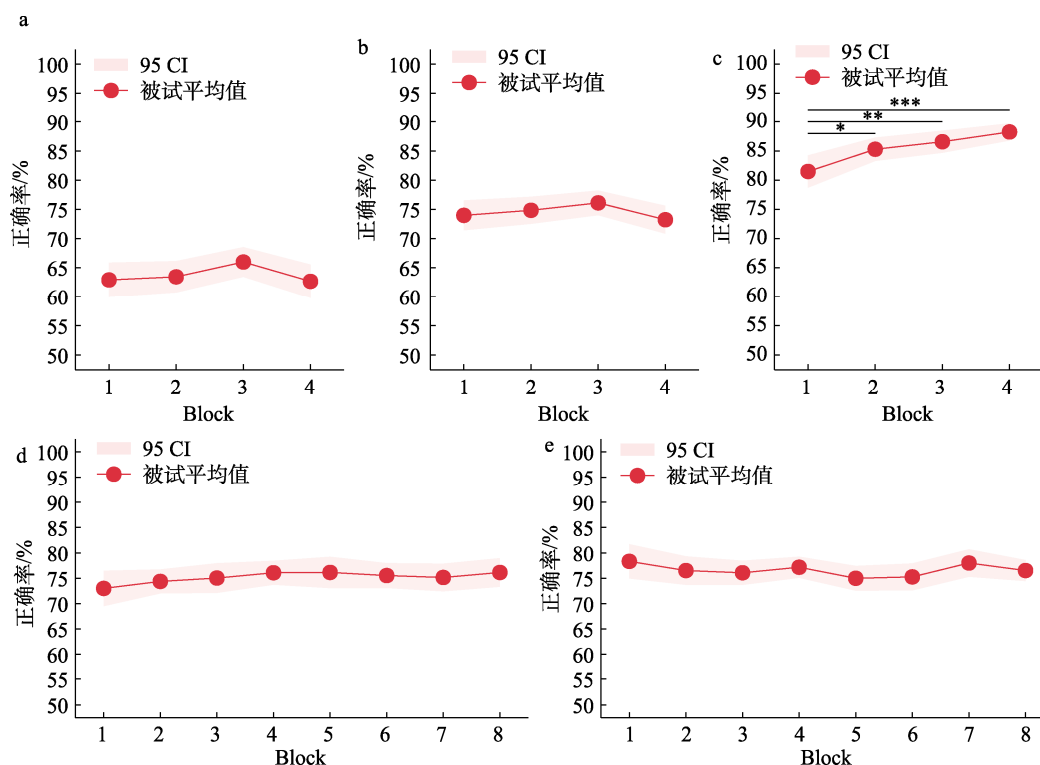


图 S1 五个实验中，被试在不同 block 中平均正确率的变化趋势。(a)实验 1 高任务难度下，被试在不同 block 中的平均正确率；(b)实验 2 中等任务难度下，被试在不同 block 中的平均正确率；(c)实验 3 低任务难度下，被试在不同 block 中的平均正确率；(d)实验 4 中被试在不同 block 中的平均正确率；(e)实验 5 中被试在不同 block 中的平均正确率。每个圆点代表全部被试在该 block 中的平均正确率，阴影部分表示 95%置信区间。*表示 $p < 0.05$ ，**表示 $p < 0.01$ ，***表示 $p < 0.001$ 。

和 block4 ($p < 0.001$), block2 的正确率与 block3 ($p > 0.9$)和 block4($p = 0.237$)无显著差异, block3 和 block4 也无显著差异 ($p > 0.9$); 在实验 4 中, block 的主效应不显著, $F(7, 280) = 0.66, p = 0.709, \eta^2 = 0.02$; 在实验 5 中, block 的主效应也不显著, $F(7, 273) = 0.93, p = 0.464, \eta^2 = 0.02$ 。

这些结果表明, 除实验 3 (“一升四降”)外, 被试在每个 block 中的平均正确率都很稳定, 说明“一升两降”的阶梯算法对任务表现进行了有效控制。在实验 3 中, 我们只发现了 block1 和 block2、block3、block4 之间存在差异, 没有发现 block2、block3 和 block4 之间的差异, 这可能是由于相比于“一升两降”的阶梯算法, “一升四降”的阶梯算法需要更多的试次才能达到收敛(García-Pérez, 1998)。被试在第一个 block 中的行为表现很有可能还没有达到算法的收敛, 因此导致了第一个 block 与剩余 3 个 block 之间的任务表现存在差异。但是, 这并不影响实验结论, 我们剔除了第一个 block, 对剩余 3 个 block 再一次做了统计分析, 结果与剔除前一致, 具体结果如下:

配对样本 t 检验结果显示, 匹配试次占比与不匹配试次占比无显著差异($51\% \pm 5\%$ vs. $49\% \pm 5\%$), $t(46) = 1.36, p = 0.182$, Cohen's $d = 0.20$, 95% CI = $[-0.01, 0.05]$, $BF_{01} = 2.68$, 被试并未表现出与预测一致的反应倾向。被试在匹配试次与不匹配试次中的平均正确率没有显著差异($86\% \pm 5\%$ vs. $88\% \pm 6\%$), $t(46) = 1.74, p = 0.088$, Cohen's $d = 0.25$, 95% CI = $[-0.04, 0]$, $BF_{01} = 1.56$ 。

在信心水平上, 相比与不匹配试次(2.88 ± 0.55), 被试在匹配试次中有更高的信心评分(3.13 ± 0.44), $t(46) = 4.44, p < 0.001$, Cohen's $d = 0.65$, 95% CI = $[0.14, 0.36]$, $BF_{10} = 392.48$; 在元认知效率上, 相比与不匹配试次(0.73 ± 0.62), 被试在匹配试次(1.00 ± 0.42)中表现出更好的元认知能力, $t(46) = 2.66, p = 0.011$, Cohen's $d = 0.39$, 95% CI = $[0.07, 0.48]$, $BF_{10} = 3.61$ 。

2 正确率与信心水平、元认知效率之间的相关分析

为进一步探究被试正确率与信心水平和元认知效率的关系, 我们对这些变量进行了相关分析, 具体结果如表 S2 所示。5 个实验的结果表明, 任务表现、信心水平和元认知效率之间可能并不存在稳定的相关关系。

表 S2 实验 1~5 中, 被试的正确率、信心水平和元认知效率的相关

	正确率-信心水平	正确率-元认知效率	信心水平-元认知效率
实验 1	-0.06	-0.13	-0.07
实验 2	-0.02	-0.08	-0.10
实验 3	0.42**	-0.19	0.21
实验 4	0.05	-0.50***	0.35*
实验 5	0.44**	-0.27	0.10

注: *表示 $p < 0.05$, **表示 $p < 0.01$, ***表示 $p < 0.001$ 。

首先是正确率与信心水平的关系。在实验 1、实验 2 和实验 4 中, 正确率与信心水平无显著相关; 而在实验 3 和实验 5 中, 正确率与信心水平呈显著正相关, 信心水平随着正确率的上升而增加。这些结果说明正确率与信心水平的相关关系并不稳定。事实上, 知觉决策和信心决策在认知和神经层面都存在一定的独立性。例如, Rahnev 等人(2012)用经颅磁刺激(TMS)干预枕叶皮层以破坏视觉刺激的内部表征, 结果发现被试任务表现正确率降低, 然而信心水平反而提升。Vlassova 等人(2014)发现无意识信息能够改变决策的准确性却不影响信心水平。从计算机制角度来说, 有研究发现信心评分几乎完全取决于与反应一致的证据(即支持已选类别的证据), 而知觉判断依赖于支持与反对所选刺激类别之间证据的平衡(Maniscalco et al., 2016; Sun et al., 2025)。从神经角度来说, Peters 等人(2017)使用颅内电生理记录和机器学习技术证明了知觉决策和信心决策依赖于时空可分离的神经表征。

其次是正确率与元认知效率的关系。我们仅在实验 4 中发现正确率和元认知效率存在显著负相关。事实上, 低任务表现并非必然导致高元认知效率, 二者存在的相关也并不稳定。这一观点在 Hu 等人(2023)的实验研究中得到进一步验证: 他们在知觉与记忆任务中, 任务表现(d')与元认知效率($meta-d'/d'$)表现出非对称性变化, 知觉任务的 d' 降低时, 元认知效率可能同步降低(实验 1)、或保持无差异(实验 3), 而在实验 2 中, 知觉任务表现提升时元认知效率反而下降。

最后是信心水平与元认知效率的关系。我们仅在实验 4 中发现信心水平与元认知效率呈现显著正相关。元认知

效率表示的是信心对于知觉判断正确或错误的区分程度,与信心水平的本身的高低不存在稳定的相关关系,例如,过度自信的人往往会表现出更高的信心水平和更低的元认知效率。Luo 和 Liu (2023)的研究结果进一步证明了信心水平和元认知效率之间的可分离性,他们发现在任务难度较高的实验 1 中,反馈组和无反馈组被试在信心水平上没有差异,但是无反馈组被试的元认知效率显著高于反馈组,而在任务难度较低实验 2 中,反馈组信心水平高于无反馈组,同时反馈组元认知效率也要高于无反馈组。

综上,一阶知觉任务表现、信心水平和元认知之间存在比较复杂的关系,有待进一步研究探究其本质。

参 考 文 献

- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, 443.
- García-Pérez MA. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12), 1861–81.
- Hu, X., Yang, C., & Luo, L. (2023). Are the contributions of processing experience and prior beliefs to confidence ratings domain-general or domain-specific? *Journal of Experimental Psychology: General*, 152(1), 28.
- Luo, T., & Liu, C. (2023). The impact of feedback on metacognition: Enhancing in easy tasks, impeding in difficult ones. *Consciousness and Cognition*, 116, 103601.
- Maniscalco, B., Peters, M. A., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception & Psychophysics*, 78(3), 923–937.
- Peters, M.A.K., Thesen, T., Ko, Y.D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., and Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behavior*, 1, 0139.
- Rahnev, D., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology*, 107(6), 1556–1563.
- Sun, F., Ni, Y., Lu, W., Su, J., Wang, S., & Wan, X. (2025). Confidence bias prescribes the neurocomputational mechanism of decision-making. *Cell Reports*, 44(5), 115563. Advance online publication.
- Vlassova, A., Donkin, C., & Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), 16214–16218.