

《心理学报》审稿意见与作者回应

题目：“没有”为什么隐含着“消极情绪”？——否定加工中的情绪表征

作者：高志华，鲁忠义

第一轮

审稿人 1 意见：本研究采用情感错误归因范式（AMP）和经典情感启动范式（BFP）探讨了汉语否定加工的情绪表征，AMP 范式研究结果发现否定短语比肯定短语有显著高的消极反应概率；BFP 范式研究结果发现否定短语比肯定短语同样启动了更高的消极反应概率。基于此结果，作者提出了否定本身具有消极效价，否定加工通过消极情绪实现具身化的观点。研究问题还是有一定新意的，审稿人有以下一点建议供作者思考，以完善该研究。

意见 1：实验 1 中的启动项呈现时间为 200ms，作者没有说明为什么设置为 200ms，因为本实验运用了不同的刺激材料，因此应该说明一下。

回应：感谢审稿专家提出的宝贵意见！文中做了补充，在这里也加以说明。实验 1 和实验 2 中启动项的呈现时间均为 200ms，这是参考了 Deutsch, Kordts-Freudinger, Gawronski 和 Strack(2009)的研究。在 Deutsch 等（2009）的实验中，启动项为否定或肯定一个情绪词构成的短语，例如，“no/a party”，结果探测到了否定效应，这表明启动项呈现 200ms 可以满足探测否定效应的实验要求，故在本实验中启动项的呈现时间设为了 200ms。

意见 2：实验 1 中为什么选择两个效价范围，建立两个汉语双音节中性名词库？为什么不能是一个效价范围内的名词然后构成两个词库呢？

回应：感谢专家严谨的审稿态度和意见！实验 1 中作为启动项的构建短语素材的中性词全部选自效价值在 4.5-5.5 之间的汉语双音节中性词库（即原库一）。实验 1 没有使用库二的材料，它是为了进一步研究所用的，为了避免误解，在修改稿中已经删除了此词库。

意见 3：从实验 1、实验 2 的流程与被试反应来看，几乎都是相同的，但实验 1 却是一个二因素设计，而实验 2 却是一个三因素设计，因为实验 1 是目标项是中性词，而实验 2 的词包括了积极与消极。因此，建议作者把经典情感启动范式与情感错误归因范式的本质区别从一开始就明确指出，这样对那些不熟悉该两种范式的读者就能清晰把握。

回应：感谢专家的宝贵意见，关于实验 1 与实验 2 中使用的两种不同的情感启动范式 AMP 与 BFP 的区别，在引言部分增加了这两种范式的比较与说明（见引言中的蓝色字体部分）。

为了便于审稿专家审阅，这里抄录如下：

虽然 AMP 和 BFP 两种情感启动范式都是内隐的测量方式，但它们反映的内隐认知加工机制是不同的（Nosek, Hawkins, & Frazier, 2011）。Deutsch 和 Gawronski（2009）比较了这两种研究范式，在 AMP 范式中，对被试来说，目标项（如象形文字）是无意义的，呈现

时间短,并且其后有掩蔽刺激,被试对目标项的效价的判断只好更多地依赖于启动项,因此,通过AMP测量可以看出,启动项诱发的情感被(错误地)用来评价目标项(如象形文字),反映的是一种错误归因机制。而BFP的启动效应反映的则是反应干扰机制(response-interference, RI),就是说如果启动项与目标项的效价不同,诱发的反应倾向就会相反,表现为干扰作用。Deutsch等(2009)的研究中就发现,AMP中探测到了否定效应,“no party”启动更多的消极反应;而BFP中则未出现。究其原因,AMP中对目标项的反应要借助于启动项,启动项得到了更多的加工,才出现了否定效应;在BFP中,启动项和目标项之间是一种很微弱的联系,而且目标项的语义是很明确的,要完成对目标项的情绪效价判断,被试的注意力只需放在目标项上,而无需去关注启动项。这样做,实际上是限制了启动项中的否定词与目标项中的情绪词之间的意义整合和效价转换。这种情况下,如果还能发现启动项对目标项的干扰,表现出以启动项中的情绪为主导的匹配或不匹配效应,那么就完全有理由说,在对否定的启动效应的测量中,BFP范式较AMP范式更具内隐性,所测的否定加工也更具自动化。

意见 4: 实验 2 中作者将实验 1 与实验 2 两种测量方式作为一个变量进行了 2 (短语极性:肯定与否定) × 2 (测量方式: AMP 与 BFP) 的数据分析,以验证加工资源对否定的情绪表征的影响。让审稿人有点不明白的是,AMP 与 BFP 是两种不同的实验范式,难道它们只是两种测量方式的不同?该疑惑实际上与前一个关于二者实验流程上的疑惑是一样的。请作者解释。

回应: 感谢专家的宝贵意见!AMP 和 BFP 两种情感启动范式不仅是测量方式的不同,它们反映的内隐加工机制也是不同的,概括地说,在对否定的启动效应的测量中,BFP 范式中较 AMP 范式更具内隐性,更能测量出否定启动效应。具体内容见引言的蓝色字体或对上一个问题的回答。

意见 5: 关于“否定的情绪表征是自动产生的,不受加工资源多寡的限制”的看法可能还需谨慎,虽然BFP范式比AMP范式的启动效应要小,但仅凭这个研究结果得出这个结论可能还为时尚早。因此,建议把表述的口气变温和点。

回应: 感谢专家的宝贵意见,在表达方式已经修改得较为温和。

意见 6: “否定本身具有情绪性,并且是消极情绪,否定通过情绪实现具身化”这一观点确实有一定新意,虽然实验结果也倾向于支持作者提出的观点,但审稿人凭直觉判断,情况可能不是这么简单。我们的现实经验中,我们是出于多种原因运用否定的,而且也与情境等因素有关。有些情境下使用否定可以带来意想不到的喜剧效果,或者说能引发明显的积极情绪的。因此,不一定是消极情绪。另外,如果真的就是该研究者指出的这种情况,那么在双重否定条件下,可能会是什么结果呢?

回应: 感谢专家的宝贵意见!确如专家所说,否定本身是非常复杂的,会受到情境语用因素的影响,这得到了一些实验的证实(Jordan, 1998; Nieuwland & Kuperberg, 2008)。同时,否定还有描述否定(descriptive Negation)和元语否定(metalinguistic Negation)的区分(Ducrot,

1972), 还可以作为一种修辞手段 (Giora, Fein, Metuki, & Stern, 2010)。但本研究中所讨论的仅限于一种简单的陈述性否定。双重否定问题, 国内学者何先友团队做过研究, 确实有其不同的加工机制。关于双重否定以及其他形式的否定加工将是进一步研究的方向, 这一点在文章的结尾处做了一点简单说明。

意见7: 调整单词或符号之间的空格, 多处

回应: 感谢专家的宝贵意见, 这类错误已经进行了修改。

意见8: 表达问题: “BFP 范式中两者的差异显著, 但在 0.05 置信水平上”这种表述不符合中文习惯。

回应: 感谢专家的宝贵意见, 指出的错误已经进行了修改。

.....

审稿人 2 意见:

该论文采用情感错误归因范式和经典情感启动范式考察汉语否定加工的情绪表征问题。论文的选题新颖, 文献分析全面, 问题导入清晰。实验设计合理, 数据分析方法, 结果可靠。

意见1: 实验选择的中性词作为启动词, 并且进行了匹配。但是没有评定加否定/肯定前缀后的短语句子的效价进行评定。个人认为, “没有管家”、和“有管家”整体意义上存在效价差异。也即二者的差异不仅仅是否定意义, 而含有情绪色彩的差异。这样发现的负性情绪表征效应可能是短语的情绪意义所致而非否定信息的具身表征。可以补测短语的效价, 排除这一可能。

回应: 感谢专家的宝贵意见! 这个问题我们从两个方面来解释。(1) 从研究目的看, 我们着重考察一个中性名词前加上否定词后的情绪表征。因此实验中主要是控制这些中性词在各个维度上的差异, 而不能去控制中性名词前面加上肯定词和否定词组成的短语之间的差异。这就如同研究肯定句和否定句加工一样, 肯定句和否定句之间一定是有差异的。(2) 从研究方法看, 我们使用了 AMP 与 BFP 启动实验范式, 用这种内隐测量的研究方法得到了预期的研究目的。审稿专家建议补测两种短语的效价, 这其实可以看做是外显的一种研究方法。我们预测利用外显的评定法也能得到如我们使用内隐测量方法得到的实验结果。

我们按照专家的建议对两种短语做了评定, 但限于篇幅没有放到稿子中, 在这里呈现如下:

从已经建立的汉语双音节中性词库中抽取词频高于 3, 激动程度低于 5, 优越程度高于 3, 熟悉程度高于 4, 抽象程度低于 5 的词, 最后剩下 130 个中性名词。为了避免评价中的天花板效应与地板效应, 从王一牛, 周立明和罗跃嘉 (2006) 编纂的情绪词汇材料库中, 筛选出 35 个情绪效价高于 7 的积极效价名词 35 个, 低于 3 的消极效价名词 35 个, 作为填充材料。这样就构成了 200 个名词的词汇库。在这些名词前面加上“有”, 构成“有某物”的肯定版本; 在前面加上“没有”, 构成“没有某物”的否定版本。将肯定版本与否定版本进行拉丁方匹配, 构成两个短语列表。在每个列表中, 有 100 个肯定短语, 100 个否定短语。所选中性词的词频、效价、唤醒水平、优势程度、熟悉程度与具体—抽象程度的值分别为 152.72 ± 314.50 、 4.89 ± 0.20 、 4.38 ± 0.32 、 5.08 ± 0.52 、 5.65 ± 0.79 和 3.65 ± 0.42 。随机抽取某高校大学生 283 人

(其中男 173, 女 110), 平均年龄 20 ± 1.32 岁, 在问卷星网络问卷平台上 (<http://www.sojump.com>) 就每一个肯定或否定短语的情感的愉悦程度从 1 到 9 进行评分, 1 代表非常不愉快, 9 代表非常愉快。

结果回收问卷 283 份, 删除重复评分超过 80% 的被试, 剩下有效数据 257 份, 有效率为 91%。对词组评价的数据进行以被试为随机变量和以项目为随机变量的单因素方差分析, 词组的极性在两种分析中都是组内变量。结果否定短语 (4.53 ± 1.26) 的效价评分显著低于肯定短语 (5.75 ± 1.13), $F_1(1, 256) = 151.664, p < 0.001, \eta^2 = 0.372, F_2(1, 119) = 450.283, p < 0.001, \eta^2 = 0.791$ 。否定短语较肯定短语倾向于消极化, 这在项目分析中尤为突出。将所选出的中性词的效价作为总体平均数 (4.89 ± 0.20), 将否定短语与肯定短语分别与之进行比较, 结果发现否定短语效价显著低于中性词 $T = -10.950, p < 0.001$; 肯定短语效价显著高于中性词 $T = 28.015, p < 0.001$ 。这表明, 否定确实使效价趋于消极。

意见 2: 科隆巴赫信度系数是测验项目的内部一致性系数, 作者在几个维度的评定上报告了该系数, 这是如何计算的? 每个维度不是只有一个评定值?

回应: 感谢专家的宝贵意见, 文中已经对这一部分的表达和呈现进行了修改。第一个问题每个维度的“科隆巴赫信度系数是如何计算的”。通过计算所有的有效被试在每个维度上所评价的所有项目 (359 个双音节中性词语) 的内部一致性系数, 获得所选词库在每个维度上的内部一致性系数。第二个问题“每个维度不是只有一个评定值”, 我理解专家疑问的是两个信度指数, 第一个一致性系数, 也就是科隆巴赫信度, 第二个呈现的是分半信度。分半信度的计算是将所有项目 (359 个双音节中性词语) 分成前后两部分, 前半部分 180 个项目, 后半部分 179 个项目, 计算 Guttman 分半信度。

意见 3: 文章的撰写整体上比较规范, 语言流畅, 有一些小问题: a. 前言中有标题 1.1, 但没有 1.2 b. 实验 1 被试数量没报告。 c. 实验 2 目标项的呈现时间没有报告。

回应: 感谢专家的宝贵意见, 在行文中又做了认真检查与修改。其中, a. 前言中有标题 1.1”中的标题已经删除; b. 实验 1 中被试数量已经补充, c. 实验 2 中目标项的呈现时间是不固定的, 持续呈现到被试按键做出效价判断为止, 这在实验 2 的实验程序部分有说明。

第二轮

主编终审意见:

经过两轮的修改, 几位审稿专家均认为该论文达到发表标准。我本人审阅文章之后也同意几位专家意见。但是, 发表之前, 一些细节需要修改。

意见 1: 参考文献引文不规范, 例如英文文献排序未按第一作者首字母排序。

回应: 感谢审稿专家提出的宝贵意见! 参考文献引文已经进行了修改, 特别英文文献排序已经按第一作者首字母进行了排序。

意见 2: 英文摘要存在较多表述不当、语法等问题, 建议找专人修改。

回应：感谢审稿专家提出的宝贵意见！英文摘要已经找英文专业人士进行修改。

意见 3：效应量方面，文章中给出的是“ η^2 ”，请核实到底是“ η^2 ”还是“ η_p^2 ”

回应：感谢审稿专家提出的宝贵意见！已经核实，效应量方面，文章中给出的是“ η_p^2 ”，而不是“ η^2 ”，已经在文章中做了相应修改。