

## 《心理学报》审稿意见与作者回应

题目：提取干扰对不同类型内隐记忆的影响

作者：黄发杰, 孟迎芳, 严颖

---

### 第一轮

#### 审稿人 1 意见：

该文通过四个实验考察了提取干扰对不同类型内隐记忆的影响，具有一定的理论意义。实验思路清晰，控制严格，数据分析恰当，结果可靠，写作规范。但是，以下问题需作者进一步思考。

**意见 1：**关于词汇判断任务，我同意作者将使用该任务测量的启动效应归类为知觉启动，但是，我不认同作者所说的“该任务主要依赖于对词汇正字法或词形等知觉信息的加工”。对于中文双字词，若没有先前的词汇学习过程，独立的词汇判断是不能以正字法、词形信息为基础的，因为双字词中的每个字都是真正的汉字，判断其是否构成真正的双字词，是需要语义加工的，有意义的就是词，无意义的就不是词。

**回应：**虽然也有研究进行了类似表述(Zeelenberg & Pecher, 2003)，但如专家所言，对其的界定可能也还是存在争议的。因此为了避免疑义，我们将这一说法修改为“由于该任务会受到学习和测验阶段词汇知觉信息相似性的影响(Prull et al., 2016)，主要测量的是知觉启动”。因为本实验着重的是知觉与概念启动之间的差异，因此主要强调词汇判断任务可作为知觉启动测验任务的代表。

#### 参考文献：

Zeelenberg, R., & Pecher, D. (2003). Evidence for long-term cross-language repetition priming in conceptual implicit memory tasks. *Journal of Memory and Language*, 49(1), 80–94.

**意见 2：**实验 3 和 4 在实验范式上有所创新，但是也使我产生了疑问。作者界定产生式测验为：只提供目标线索，该线索可能激活多种解决方案，需要被试从中选择并产生一种答案。

我觉得产生多种解决方案应该是产生式测验的关键特征，但是，在本文所使用的两个产生式测验中，都无需被试去产生，只需选择即可。缺少了“产生多种解决方案”这一特征的测验还能称之为产生式测验吗？

回应：感谢专家的意见。在最初确定实验 3 和 4 的内隐测验上，我们也曾进行了多方面的考量。本研究的目的是希望了解提取干扰是否会对不同类型的内隐记忆测验产生不同的影响。传统产生式中，通常刺激呈现时间较长，较容易混入外显记忆成分(MacLeod, 2008; Sheldon & Moscovitch, 2010; Miyoshi & Ashida, 2014)。因此如何尽可能避免其它因素的影响，使得测验之间的结果能够进行比较，是我们考虑的一个重要因素。

Gabrieli 等人(1999)最早将产生式启动任务界定为：需要被试根据刺激提供的目标线索来引导提取，目标线索可能激活多种解决方案，需要从中选择并产生一个正确答案。基于此，我们认为，提供目标线索，根据目标线索来提取并产生答案，应该是产生式测验的典型特征。而在提取过程中，可能在大部分条件下，目标线索都会激活多种解决方案，此时还有一个反应竞争，这似乎也构成了产生式测验的一个重要特征。曾有研究专门探讨了解决方案多与少对产生式测验的影响。例如 Spataro, Mulligan 和 Rossi-Arnaud (2010)设计了两种残词补全任务(Word Fragment Completion Task, 简称 WFC)，一种只有唯一答案，即有产生答案，但不存在反应竞争，另一种有多种可能答案。结果发现，在编码干扰条件下，大学生完成这两种 WFC 的成绩无显著差异；其它研究也发现类似的现象，包括词干补笔(Geraci & Hamilton, 2009; Marques et al., 2016)和动词生成 (Marques et al., 2016; Prull, 2010; Prull & Matthew, 2013) 等产生式测验。由此可见，目标线索是否一定产生反应竞争可能并不是产生式任务的必要条件。

基于上述的一些考虑，我们设计了两种新的产生式任务。我们认为，在测验中呈现目标线索，被试在加工线索的过程中理论上会自动产生解决方案，应该具备了产生式测验的基本特征，但为了避免反应竞争所产生的影响，以及能够与实验 1 和 2 的测验进行比较，就综合了两种测验的特征，设计了产生式的词汇判断和产生式的语义分类，这样两类测验的反应任务是类似的，但实验 3 和 4 又多了一个产生的特征。从实验 3、4 的结果来看，提取干扰对两类新的产生式任务产生的影响与实验 1 和 2 的识别式任务表现出不同。这种差异在一定程度上也可以说明，两类测验中所涉及的启动效应应该是存在差异的。

当然，创新后的产生式测验是否与传统的产生式测验存在差异，以及创新后的产生式测验是否会因解决方案的多少而有所差异，这是一个值得我们进一步深入探讨的问题。

再次感谢专家的建设性意见。

参考文献：

- Gabrieli, J. D. E., Vaidya, C. J., Stone, M., Francis, W. S., Thompson-Schill, S. L., Fleischman, D. A., et al. (1999). Convergent behavioral and neuropsychological evidence for a distinction between identification and production forms of repetition priming. *Journal of Experimental Psychology: General*, 128(4), 479–498.
- Geraci, L., & Hamilton, M. (2009). Examining the response competition hypothesis of age effects in implicit memory. *Neuropsychology Development & Cognition*, 16(6), 683–707.
- MacLeod, C. M. (2008). Implicit memory tests: Techniques for reducing conscious intrusion. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 245–263). New York, NY, US: Psychology Press.
- Marques, V. R. S., Spataro, P., Cestari, V., Sciarretta, A., & Rossi-Arnaud, C. (2016). Testing the Identification/Production Hypothesis of Implicit Memory in Schizophrenia: The Role of Response Competition. *Journal of the International Neuropsychological Society*, 22, 314–321.
- Miyoshi, K., & Ashida, H. (2014). Priming and implicit recognition depend on similar temporal changes in perceptual representations. *Acta Psychologica*, 148, 6–11.
- Prull, M. W. (2010). Age-related influences on repetition priming in the verb generation task: examining the role of response competition. *Aging, Neuropsychology, and Cognition*, 17(4), 439–461.
- Prull, M. W. (2013). Attention and repetition priming in the verb generation task. *Acta Psychologica*, 143, 218–226.
- Sheldon, S. A., & Moscovitch, M. (2010). Recollective performance advantages for implicit memory tasks. *Memory*, 18(7), 681–97.
- Spataro, P., Mulligan, N., & Rossi-Arnaud, C. (2010). Effects of divided attention in the word-fragment completion task with unique and multiple solutions. *European Journal of Cognitive Psychology*, 22(1), 18–45.

**意见 3：**讨论中对“为何产生式启动更不容易受到提取干扰的影响”给出了解释，但是这个解释理解起来很费劲，建议作者重新组织语言，增强逻辑性。

**回应：**非常感谢专家的意见，我们对这部分解释进行了修改。具体表现在讨论的倒数第 2 段：“那么，为何产生式启动更不容易受到提取干扰的影响呢？研究曾表明，双任务同时执行会激活前额叶皮层，即便是两个任务都比较简单甚至完全自动化，也需要激活这一双任务执行功能相关的特定化皮层区域(D'Esposito et al., 1995; Just, Keller, & Cynkar, 2008)。双任务

和单独的子任务执行时对大脑额叶皮层相同区域的激活程度存在差异,如果负责单任务加工的皮层区域激活增强,可以为双任务执行过程提供额外的资源,促进双任务之间的协调转换,降低任务冲突等,进而减小任务之间的干扰(Just et al., 2008; 谭金凤, 伍姗姗, 徐雷, 王丽君 & 陈安涛, 2013; Tombu et al., 2011)。研究通过将外部脑损伤病人与健康被试进行对比,发现双任务下外部脑损伤组的前额叶激活程度更大,这说明了额外的前额叶资源的参与更有助于双任务操作的顺利完成(Rasmussen et al., 2008)。据此我们认为,与后枕皮层的血流动力减少相关的识别式任务相比,由于单独的产生式内隐测验任务通常需要前额叶区域的参与(Gabrieli et al., 1999; Geraci, 2006),同时执行一个相对简单的干扰任务(奇偶判断),在双任务下表现出一定的前额叶激活增强效果,使得受到提取干扰的破坏更小。因此,与识别式启动相比,产生式启动在提取干扰下的启动效应并未被完全破坏。”

**意见 4:** 论文前言和讨论部分的理论分析可进一步加强。

**回应:** 修改稿中对前言和讨论部分进行了一定的修改,主要为:

在前言部分,增加了为何要进行不同测验类型设计的理论分析,具体表现在前言的第 4 段:“已有研究曾表明,不同类型的启动效应所依赖的认知神经系统存在着差异(Leynes, Bruett, Krizan, & Velosa, 2017; Marques, Spataro, Cestari, Sciarretta, & Rossi-Arnaud, 2015)。更为重要的是,编码干扰对不同类型间接测验中的启动效应所产生的影响也是分离的:编码干扰会影响产生式测验中的启动效应,但不影响识别式测验中的启动效应(LaVoie & Faulkner, 2008);编码干扰会减少随后的概念启动,但不影响知觉启动(Mulligan & Lozito, 2006; Newell et al., 2008)。那么提取干扰对内隐记忆的影响是否也会因内隐记忆测验类型的不同而有所差异呢?”

在讨论部分,我们重新整理了思路,期望能够对实验结果有着更合理的解释。详见讨论中标红部分。

**意见 5:** 文中有一些文字缺漏之处,作者应仔细核对修改。

**回应:** 感谢专家的细心审阅。我们对修改稿进行了多次的通读,将论文类似的疏漏进行修改、完善,并用红色字体标出。

.....

## 审稿人 2 意见：

论文《提取干扰对不同类型内隐记忆的影响》通过设置不同类型的内隐测验，探讨了提取阶段的干扰对内隐记忆的影响，结果发现，不同类型的内隐记忆都会受到提取干扰的影响，但相较于产生式词汇判断任务和语义分类任务，识别式词汇判断任务和语义分类任务更易受提取阶段干扰的影响。该研究实验方法可行，实验结果可信。不过，本文也存在如下几点问题：

**意见 1：**前沿第一段中最后一句“启动效应对编码干扰具有“免疫性””，其中“免疫性”是否合适？一般是说“抗干扰性”吧。文中其它部分如后面二段等也存在这一问题。

**回应：**感谢专家的建议。因为编码阶段的干扰任务不会削弱内隐记忆测验中的启动效应(Dew & Cabeza, 2011; Lozito & Mulligan, 2010; Spataro, Cestari, & Rossi-Arnaud, 2011)，文中欲借“免疫性”对这一特性进行简喻。我们查阅了《现代汉语词典》中有关两个词汇的中文释义，“免疫”更强调的是“生物体识别与排除外来的和内在的非己抗原，以维持机体相对稳定的一种生理功能。”“抗干扰”多指“在通信领域中，用来对抗有用信号的接收受到的干扰。”对比之下，“抗干扰性”确实更贴近认知心理学，从信息加工角度对记忆的定义。因此，根据专家的意见，对全文的表述进行了相应修改。

## 参考文献：

- Dew, I. T. Z., & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: behavioral and neural evidence. *Annals of the New York Academy of Sciences*, 1224(1), 174–190.
- Lozito, J. P., & Mulligan, N. W. (2010). Exploring the role of attention during implicit memory retrieval. *Journal of Memory & Language*, 63, 387–399.
- Spataro, P., Cestari, V., & Rossi-Arnaud, C. (2011). The relationship between divided attention and implicit memory: a meta-analysis. *Acta Psychologica*, 136, 329–339.

**意见 2：**前沿最后一段中提到“本研究将通过四个实验，分别考察提取干扰对识别式知觉内隐测验、识别式概念内隐测验、产生式知觉内隐测验和产生式概念内隐测验”。四类测验具体使用的是何种任务？为何每类任务进行的是该类测验？实验总体研究假设和预期是什么？

**回应：**原文中对四种测验的引出确实有点突兀，让人不容易理解，因此在修改稿中，我们对这部分进行了修改，具体表现在前言的最后一段：“已有研究曾认为，两种内隐记忆测验类型的划分之间会存在着交叉(Gabrieli et al., 1999; Prull & Spataro, 2017)，由此可构成四种类型

的间接测验，即识别式知觉内隐测验、识别式概念内隐测验、产生式知觉内隐测验和产生式概念内隐测验。因此，本研究拟通过四个实验，分别设置四种不同类型的内隐测验任务，系统考察提取干扰对内隐记忆的影响。”

而对于四类测验具体使用的是何种任务，以及为何进行该类测验，我们将其设置在每个实验的“实验目的”中加以说明，以方便读者理解随后的实验。同时在修改稿中，我们对每个实验目的也进行了再次的修改，以期让读者能够更明确每类任务的设置。

实验的总体研究假设和预期是认为“提取干扰对内隐记忆的影响可能会因内隐记忆测验类型的不同而不同。”该假设主要体现在前言的最后一段。

**意见 3:** “提取干扰”是本研究中最重要的一個操纵变量，在每个实验中具体是如何进行的？目前似乎没有交代的特别清楚。

**回应:** 提取干扰的操纵在四个实验中是一致的，采用的都为数字奇偶判断任务，以便于进行实验间的比较。为了使读者更为清晰，我们在四个实验的“实验程序”部分对干扰的表述都进行了修改，并用红色标出。

**意见 4:** 由于干扰条件和无干扰条件是被试内变量，是否进行了实验顺序的平衡？

**回应:** 四个实验的有无干扰条件都在被试间进行了平衡。我们在四个实验的“实验程序”部分补充了相关说明，并用红色标出。

**意见 5:** 建议实验流程图分别列出无干扰和有干扰条件。

**回应:** 感谢专家的建议。在实验 1 中我们增加了无干扰条件的流程图，以方便读者理解。由于接下来实验的呈现模式是类似的，只是内隐记忆测验发生了变化，因此在剩余的几个实验中，只呈现了干扰条件下的流程图。

**意见 6:** 2.3 结果部分是如何计算启动量的？

**回应:** 论文中启动量计算都是通过新词的反应时减旧词反应时，或者旧词 ACs 减去新词的 ACs，即为新旧词之间的差异。为了使读者更为清晰，我们在实验 1 的结果表格下面进行了标注：

(注：由于本研究中内隐测验任务的平均正确率较高(大于 0.7)，因此对正确率进行反正弦平方根( $ACs = \text{DEGREE}(\text{ASIN}(\text{SQRT}(AC)))$ )转换(Winer, Brown, & Michels, 1971)；启动量

为新旧词在反应时/ACs 上的平均差值。实验 2\3\4 类似。)

**意见 7:** 表 2 中反应时的启动量似乎有错误。

**回应:** 非常抱歉出现这样的错误，是我们在抄写过程中出现的失误，现已修正。同时我们对所有的数据再次认真核对，排除类似的失误。再次感谢专家的细心审阅。

**意见 8:** 实验 3 结果部分，在有干扰条件下根据表 3 的结果并不是“两类旧词的反应时显著长于新词”。

**回应:** 此处应表述为“两类旧词的反应时显著快于新词”，我们已经进行了修改，并标红。同时我们也通读了全文，避免类似疏忽。

**意见 9:** 实验 4 结果部分，在有干扰条件下根据表 4 的结果并不是“两类旧词的反应时显著高于新词”。

**回应:** 已修改为“两类旧词的反应时显著快于新词”。

**意见 10:** 文中个别段落语言还不够流畅，甚至存在有错别字的地方，如摘要中第一行“存在着疑义”应改为“存在着异议”。

**回应:** 再次感谢专家对论文的细心审阅，也对我们原稿中存在的不少疏忽表示歉意。修改稿中，我们认真通读了全文，梳理了文字表述，排查了错别字等。

---

## 第二轮

**审稿人 1 意见:**

**意见 1:** 作者对审稿人提出的问题进行了认真的回复，论文的质量有提高。但是，对于作者使用的两种新的产生式内隐测验，特别是产生式词汇判断任务，我仍然认为，其与传统的产生式测验相比，较少涉及“产生”的心理过程。作者说新范式结合了词干补笔的特点，但是我觉得词干补笔任务的关键之处就是被试要“补”，要产生，而本研究中的任务是要求被试在两个项目之间选择。作者强调，设计这个任务，是为了与识别式测验更容易比较，但是我没有看到这样做的必要性。

**回应：**确如专家所疑虑的，本研究中创新后的产生式测验是否具有产生的特征，其与识别式的测验是否存在着差异，尤其是产生式的词汇判断任务是否与识别式的词汇判断任务在本质上是不同的，这些都是需要进一步验证的问题。为了更好地解释这些问题，我们进行了如下两项研究，以期提供一定的答案。

研究一：对提取干扰与内隐记忆关系的国内外已有研究进行元分析，将已有研究按采用的内隐记忆测验不同区分为识别式和产生式两大类型的研究，一方面分析提取干扰对识别式和产生式两类内隐测验产生的影响是否存在着差异性，从而为本研究中所发现的提取干扰对不同类型测验的影响提供更进一步的证据。另一方面分析本研究中所创新的产生式词汇判断、产生式语义分类任务是否与已有研究所采用的传统产生式测验是属于同质性的，为这些测验是否也可归类为产生式测验提供证据。

研究一对国内外有关内隐记忆提取干扰的研究进行系统整理，收集了 26 个有效实验，初步的元分析结果表明：（1）提取干扰对不同类型内隐记忆的影响的总体效应量中等，且显著大于 0。（2）内隐记忆测验进一步分为识别式、产生式两类是合适的，这与 Spataro, Cestari 和 Rossi-Arnaud(2011)的从编码阶段的干扰探讨内隐记忆分类的元分析结果一致。（3）新的产生式词汇判断(实验 3)、产生式语义分类(实验 4)两个任务归类到产生式类型中没有异质性，但归类到识别式类型中会产生异质性，表明这两类测验归类于产生式更合适。结果详见附件 1。

研究二则采用 ERP 技术，对产生式词汇判断与识别式词汇判断任务所诱发的大脑活动进行直接的比较，以确定这两类任务之间是否存在着差异。（因为时间较紧，在 1 个月内要完成实验，因此只选择了受质疑更多的产生式词汇判断进行实验）

研究二共收集了 14 名大学生(男 5 名)的 ERP 数据，实验采用被试内设计，分成词汇判断和产生式词汇判断两个部分，只收集无干扰条件下在两个任务上的反应。研究结果表明，在两个任务上，被试都表现出明显的 N400 和 P600 效应(两个典型与启动效应关联的脑电成分)(罗跃嘉, 魏景汉, 翁旭初, 卫星, 2001; 孟迎芳, 郭春彦, 2009), 并且产生式词汇判断任务下所诱发的波幅在额区、中央区、顶区和枕区上都要比传统词汇判断任务下所诱发的波幅更为负走向。该结果至少表明，两种任务所诱发的神经活动是存在差异的。结果详见附件 2。

综合这两项研究初步分析的结果，我们认为，本研究中所创新的产生式词汇判断任务与传统的词汇判断任务是存在差异的，产生式词汇判断和产生式语义分类归类于产生式测验应该是合理的。



参考文献:

- Luo, Y. J., Wei, J. H., Weng, X. C., & Wei, X. (2001). ERP effects of recognition of Chinese spoken and written words and neural mechanism of retrieval. *Acta Psychologica Sinica*, 33(6), 489–494.
- [罗跃嘉, 魏景汉, 翁旭初, & 卫星. (2001). 汉字视听再认的 ERP 效应与记忆提取脑机制. *心理学报*, 33(6), 10–15.]
- Meng Y. F., & Guo C. Y. (2009). The Asymmetric Relationship Between Encoding and Retrieval in Implicit and Explicit Memory. *Acta Psychologica Sinica*, 41(8), 694–705.
- [孟迎芳, 郭春彦. (2009). 内隐与外显记忆的编码与提取非对称性关系. *心理学报*, 41(8), 694–705.]
- Spataro, P., Cestari, V., & Rossi-Arnaud, C. (2011). The relationship between divided attention and implicit memory: a meta-analysis. *Acta Psychologica*, 136(3), 329–339.

.....

审稿人 2 意见:

意见 1: 该文作者很好地回答了我提出的问题, 并且论文内容和写作有了很大的改进, 建议发表。

回应: 感谢专家的建议。

---

### 第三轮

审稿人 1 意见: 认可作者为解决审稿人疑问所做的进一步研究, 建议发表。

编委复审: 建议发表。

主编终审意见: 此稿经过修改, 已经达到了发表的水平。同意发表。但需要进行修改: 统计表中的表题为准确率, 表中正确率, 前后不一致, 请统一。反应时的数据需要取整。在几项实验的结果的描述中, 没有涉及到对正确率检验结果的报告, 为什么?

回应: 感谢主编对论文的细心审阅和校对。我们已将所有统计表格中的“准确率”统一改为“正

确率”，并阅读全文，对有关表述进行统一，修改部分用绿色标出。

已将所有反应时数据取整，修改部分用绿色标出。

已有研究认为，对正确率数据（Accuracy, AC）（一种类别比例数据）进行反正弦转换（arcsine-square-root transformation, ACs），有利于改善数据在方差分析中误差项的正态性、方差齐性和模型的可加性（Rao, 1960; Ahrens, Cox, & Budhwar, 1990）。考虑到本研究 4 个实验中 AC 均值较高（大于 0.7），因此将 ACs 转换应用于本研究非正态的 AC 原始数据（Winer, Brown, & Michels, 1971），使得对数据进行方差分析的结果更加稳定。此外，我们在论文中也补充了对数据转换的原因，修改部分用绿色标出。

本研究对 4 个实验的 AC 的原始数据也进行了与论文中相同的方差分析，统计结果与 ACs 的分析结果总体是一致的（具体如下），因此只报告了对 ACs 统计分析的结果。

本研究 4 个实验 AC 的统计分析结果：（1）实验 1：无干扰条件下，词类型主效应显著 [ $F(2, 54)=18.97, p<0.001, \eta_p^2=0.41$ ]，两类旧词的 AC 都显著高于新词 ( $p<0.001$ )；干扰条件下，词类型主效应不显著 [ $F(2, 54)=1.59, p=0.21$ ]。（2）实验 2：无干扰条件下，词类型的主效应显著 [ $F(2, 50)=6.72, p=0.003, \eta_p^2=0.21$ ]，知觉编码旧词的 AC 显著高于新词 ( $p=0.002$ )，但概念编码旧词与新词没有显著差异 ( $p=0.68$ )；干扰条件下，词类型的主效应不显著 [ $F(2, 50)=0.66, p=0.52$ ]。（3）实验 3：无干扰条件下，词类型的主效应显著 [ $F(2, 52)=21.24, p<0.001, \eta_p^2=0.45$ ]，两类旧词的 AC 显著高于新词 ( $p<0.001$ )；干扰条件下，词类型的主效应显著 [ $F(2, 52)=8.43, p=0.001, \eta_p^2=0.25$ ]，两类旧词的 AC 显著高于新词 ( $p_1=0.016, p_2=0.001$ )。（4）实验四：无干扰条件下，词类型的主效应显著 [ $F(2, 48)=40.45, p<0.001, \eta_p^2=0.63$ ]，两类旧词的 AC 显著高于新词 ( $p<0.001$ )。干扰条件下，词类型的主效应显著 [ $F(2, 48)=10.41, p<0.001, \eta_p^2=0.30$ ]，两类旧词的 AC 显著高于新词 ( $p_1<0.001; p_2=0.017$ )。

参考文献：

- Ahrens, W. H., Cox, D. J., & Budhwar, G. (1990). *Use of the arcsine and square root transformations for subjectively determined percentage data*. *Weed Science*, 38(4), 452–458.
- Rao, M. M. (1960). *Some asymptotic results on transformations in the analysis of variance*. ARL Technical Note, 60–126. Aerospace Research Laboratory, Wright-Patterson Air Force Base.
- Winer, B. J., Brown, D. R., & Michels, K. M. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.