

## 《心理学报》审稿意见与作者回应

题目：执行功能对言语顿悟问题解决的影响：基于行为与 ERPs 研究

作者：邢强，孙海龙，占丹玲，胡婧，刘凯

---

### 第一轮

#### 审稿人 1 意见：

**意见 1：**首先，文章的引言部分需要对文献回顾做更系统的梳理，近 10 年内有关执行功能对顿悟问题解决影响的探讨颇多，但本研究有关该部分的探讨并不能很好地反映最新研究近况以及本文的问题提出。与此同时，作者需要注意相关内容的文献引用标注。

**回应：**感谢外审专家的意见。由于外审专家二也对引言部分提到相同问题，这里同样回复修改情况。我们重新对引言部分内容进行梳理，对于引言部分内容我们重新进行梳理，添加论述了近年来关于执行功能对顿悟问题解决的影响文献。整个研究逻辑也有所调整。目前引言部分行文大致逻辑为：

文章开始时一般问题与顿悟问题的差异出发，指出现有研究普遍认为顿悟问题中存在僵局与表征重组阶段，而通过顿悟问题与认知能力之间的关系研究成为探讨顿悟问题解决心理机制研究的热点和重点。目前，执行功能对顿悟问题解决的影响仍存在争议。有研究认为顿悟问题的发生不受到执行功能的影响，是一种无意识地改变问题表征，解决者运用的是自动的、不受执行功能限制的直觉性应答，这个思维过程是快速自动且无意识的，同时也有大量研究发现执行功能及其亚成分能有效预测顿悟问题解决。我们认为顿悟问题解决是一个复杂的过程，可以区分为多个不同阶段，执行功能对顿悟问题解决的不同阶段可能具有不同的影响，此外，顿悟至少有空间顿悟问题与言语顿悟问题等不同的类型。因此，仅仅从整体上考察顿悟问题解决过程，而不区分顿悟问题的类型以及顿悟不同加工阶段可能不足以揭示顿悟问题解决与执行功能的关系。

详见正文引言部分黄色标记处。

**意见 2：**其次，在写作方法上，文章中的总讨论与《心理学报》的地位似乎不大相称。目前的讨论似乎更接近《心理科学》的写作。作者需要深入探讨，并论述三个系列实验之间的逻辑和围绕本研究的重点问题来展开。这也就是为什么在形式上，总讨论的内容和引言看起来深度不够，结论松散——一个三个实验的研究获得了如此多的结论。因此，建议作者对文章进行深入构思，围绕研究所要解决的问题来展开具有逻辑性的讨论，并集中凝练结论。

**回应：**感谢外审专家提出的宝贵意见。我们重新检视了原文的总讨论部分，确实存在诸多问题。目前重新对讨论部分进行梳理。

首先在研究 1 和研究 2 后精炼了研究小结，三个研究的逻辑为，研究 1 初步探讨执行功能与顿悟问题解决的相关关系，同时为后续研究筛选执行功能高低不同水平的被试。研究 2 则在研究 1 基础上探讨执行功能对空间顿悟问题解决过程特别是表征重组的影响，研究 3 则是从言语顿悟问题出发，探讨执行功能对研究顿悟问题的影响。研究 1 初步探讨两者的关系，研究 2 针对空间顿悟问题，研究 3 针对言语顿悟问题。总讨论部分，补充了总体的研究结果概述，同时分两个方面进行讨论，分别是执行功能对顿悟问题解决汇总的问题空间表征搜索阶段的影响，执行功能对顿悟问题解决中的表征重组阶段的影响来加以讨论。与此同时，我们重新修改了研究结论，详见结论黄色标记处。

**意见 3:** 第三，一些心理学基本的问题作者需要注意，比如实验 1 作者在目的中强调对因果关系的探讨，但最终在结果和讨论中使用的似乎都是只能揭示相关关系的方法或说明，建议不要在基本问题上“哗众取宠”。类似问题在文中其他部分也有涉及。

**回应:** 感谢外审专家的提醒。已经通篇修改文章措辞，结合第二位外审专家意见，将实验部分改为研究 1，研究 2、研究 3，并依次对 3 个研究的题目、目的以及讨论等处措施进行修改，更为客观谨慎的进行解释说明。

**意见 4:** 最后，最关键的是问题是本研究缺乏对顿悟类型的考虑，这导致可能目前的研究结论不太可靠，至少需要作者进一步思考。主要是两个行为研究都对空间顿悟问题进行研究，而脑电实验则是对言语顿悟问题进行研究，现有的研究普遍提示，空间顿悟与言语顿悟有着非常不相似的心理与神经机制。针对此，建议作者重视围绕实验结果来构思论文写作或者补充言语顿悟的行为实验为后续的脑电研究提供铺垫与心理机制资料。

**回应:** 感谢外审专家的建议。我们重新对整个研究思路进行思考，并**补做研究 3a 言语顿悟问题的行为实验**。

目前，几个研究之间的逻辑调整为：研究 1 初步探讨执行功能与顿悟问题解决的关系。此时采用的研究材料既包括空间顿悟问题也包括言语顿悟问题，并未对顿悟类型加以区分。研究 2 和研究 3 则分别从空间顿悟问题与言语顿悟问题的角度进行细致探索。其中研究 2 则具体采用空间顿悟问题，探讨执行功能对空间顿悟问题的影响，以及顿悟问题解决的具体过程。研究 3 则从言语顿悟问题采用行为实验与 ERPs 技术探讨高低执行功能条件下顿悟问题解决的具体过程。

**意见 5:** 除此之外，还有些细节问题需要注意，比如若是一定要列举表格，请将表格内的数字排列整齐，不要因为这些简单的形式问题而失去进一步修改的机会。建议大修后再审。

**回应:** 感谢外审专家的提醒和建议。我们重新检查修订文章中的表格内容，更正诸多格式问题。

---

**审稿人 2 意见:**

**意见 1:** 文章题目与主体内容不符，文章内容并不能揭示“影响机制”，是否可以去掉题目中的“机制”二字？请作者斟酌。

**回应:** 感谢外审专家的意见。重新修订文章题目，目前题目改为“执行功能对顿悟问题解决的影响：基于行为与 ERPs 研究”。

**意见 2:** “引言”部分对执行功能与顿悟问题解决关系的文献综述不够完备，其内在逻辑需要进一步加以梳理，建议作者认真修改这一部分。

**回应:** 感谢外审专家的意见。我们对引言部分内容重新进行了梳理，由于外审专家一同样对引言部分提到了相同问题，具体修改思路在外审专家一的意见 1 中进行了一并回复。

**意见 3:** 实验 1 不是一个因果关系的实验，而是一个测量研究，因而是否可以改成“研究 1”？请斟酌。同时，实验 1 并不能说明执行功能对顿悟问题解决的影响机制，那么实验 1 是否有必要存在？请作者认真思考。

**回应:** 感谢外审专家的意见。确实原实验 1 并非是一个因果关系的实验，为严谨起鉴，我们

接受外审专家的意见，将三个实验统一更改为研究 1 研究 2 研究 3。

关于研究 1 是否有存在的必要问题，课题组讨论后更倾向保留研究 1。之所以保留，主要是基于以下几个理由。其一，思路研究 1 是对执行功能与顿悟问题解决的相关关系的初步探讨。这一研究通过对执行功能及其各个亚成分与顿悟问题解决的关系，为后续更严谨的实验操作提供了方向。其二，方法上，研究 1 起到筛选合适被试的作用。细致交代了具体执行功能三个亚成分的测量方法以及高低执行功能分组。由于研究 2、研究 3b 实验被试均来源于研究 1。而执行功能的测量以及分组是本文的关键研究基础，若将研究 1 删除，这部分内容置于研究 2 中较为庞杂，故我们倾向于保留研究 1。

**意见 4:** 实验 2 的题目不够严谨，且与实验内容不符。顿悟问题解决的心理机制不可能是基于执行功能的。实验 2 题目是否可以直接写为“执行功能对顿悟问题解决的影响”？请思考。实验 3 的题目也存在类似的问题。

**回应:** 感谢外审专家的意见。我们接受外审专家的意见，实验 2 实验 3 的题目均做修改。目前，实验 2 改为“研究 2 执行功能对空间顿悟问题的影响” 原实验 3 改为“研究 3 执行功能对言语顿悟问题的影响”。

**意见 5:** 实验 3 采用脑电实验揭示了高、低执行功能被试顿悟问题解决神经基础的差异，但是除了在 P2 波中执行功能水平与顿悟类型变量有交互作用外，其他脑电成分（N1 波、N2 波、P3 波）中均未发现执行功能水平的效应，但是在讨论时却得到了执行功能影响顿悟问题解决的结论。请作者对此进行解释。

**回应:** 感谢外审专家的意见。我们重新检视了讨论部分内容，确实存在不严谨的地方。在在此重新进行表述，根据已有研究结果，N1 波代表了初级物理特征的加工，而 P2 和 N2 代表了初级的物理刺激加工后，较为高级的加工过程和问题空间特征的搜索阶段，执行功能高低影响早期的问题空间搜索，而中晚期的 P3 成分则代表了僵局阶段，新旧思维更替的过程，并不受到执行功能高低的影响。因此结合研究 1 研究 2 以及研究 3a 的研究结果，我们认为执行功能影响顿悟问题解决空间表征的搜索，但是表征重组阶段不受执行功能高低的影响。研究 3b 的小讨论删除，上述内容的讨论详见总讨论部分和结论部分。

**意见 6:** 实验 3 对 N1 波、N2 波、P2 波、P3 波均进行了详细的检验、分析与讨论，结果只有 N1 波和 P3 波对于揭示顿悟问题解决具有实质性意义。因此，应该对这两种波形的内容进行详细分析和深入讨论，其他两种波形简要介绍一下即可，不必花大量篇幅去分析和讨论。

**回应:** 感谢外审专家的意见。我们对实验 3（最新版本中为研究 3b）的分析和讨论部分进行了调整。首先在脑电数据的分析部分，我们对四种脑电成分的分析过程详略程度加以调整。且相应的删除研究 3b 的小讨论，部分内容在总讨论部分中涉及。

**意见 7:** 讨论部分 5.2 中也对上述四种波形的结果均同样地展开讨论，没有分清主次，建议对有实质性意义的波形结果进行更深入地讨论，对其他波形的内容不必展开讨论。

**回应:** 感谢外审专家的意见。结合外审专家对我们总讨论部分的意见，我们对总讨论部分重新进行梳理。弱化早期成分 N1 的讨论，分别在总讨论 5.1 和 5.2 部分深入探讨较高级知觉加工和搜索阶段的 P2 和 P3 的讨论。详见正文绿色标记处。

**意见 8:** 本文的四个结论与前面的分析和讨论没有很好地对应，并且它们之间缺乏逻辑联系，建议重新提炼本研究的主要结论。

**回应:** 感谢外审专家的意见。我们重新对研究的主要结论进行提炼。其中结论 1 主要对应研究 1 的发现，结论 2 对应研究 3 的发现，结论 3 对应研究 2 和研究 3 的发现。现研究结论提

炼为:

(1) 个体的执行功能中的转换、工作记忆刷新亚成分与顿悟能力呈显著的正相关,且刷新功能在一定程度上能够预测顿悟成绩。

(2) “顿悟”条件与“无顿悟”条件均诱发早期成分 P2 和中晚期 P3 成分。其中, P2 主要反映人脑对思维僵局的早期觉察,受到执行功能高低的影响。280~500 ms 内,中晚期 P3 则主要体现以新旧思路交替为特征的僵局打破过程,不受执行功能高低的影响。

(3) 无论是空间顿悟问题还是言语顿悟问题,执行功能对表征重组过程无显著影响,支持表征重组过程是一个突击式的“特殊过程”。

**意见 9:** 英文摘要中有许多语病,如“Experiment 1 study the relationship...”; 一个关键词翻译是错误的:“无意识”应译为 *unconsciousness*, 建议作者重新撰写英文摘要。

**回应:** 感谢外审专家的意见。已经重新撰写中英文摘要,修订关键词。具体内容可参加正文中英文摘要。

---

**审稿人 3 意见:**

**意见 1:** 实验一中执行功能任务的三个实验的选取依据是什么? 请说明来源。

**回应:** 感谢外审专家的意见。实验一种执行功能的三个亚成分测量方法均补充了选取依据。具体而言,工作记忆刷新测验采用点色位置刷新任务(color dots updating) (陈天勇,李德明,2005)。抑制测验则参考经典的色-词 Stroop 任务 (Jewel & Swarson, 2005), 转换任务采用连线测验 (the Trial Making Test, TMT) (Reitan & Wolfson, 1993)。详见正文红色底色标记处。

**意见 2:** 实验一实验条件交代不够,如顿悟测验限定时间吗? 反应指标介绍还不够明朗,如顿悟实验。

**回应:** 感谢外审专家的意见。已经重新补充相关实验条件。顿悟问题解决能力测验使用纸笔测试,整体实验分为两个部分,执行功能测试(包含有三个子成分的测量)与顿悟问题解决测试(共计 30 分钟)。详见正文红色标记处。

**意见 3:** 前言中提到, Gilhooly 和 Murphy (2005) 研究了顿悟问题解决与图形流畅性任务(测量优势反应的抑制以及基础反应的转换)的关系,发现在顿悟问题解决中均有抑制与转换的参与。但实验一的结果中发现抑制涉及较少,这与上文的引用是否矛盾?

**回应:** 感谢外审专家的意见。可能由于引言和讨论并未进行充分论述,引起专家的疑问,实际上与我们的研究结果并不冲突。本研究结果与上述文章的主要作者的 09 年研究结果相一致。Gilhooly 和 Fioratou (2009) 的研究表明无论言语还是空间工作记忆能够有效预测顿悟问题的解决,而顿悟问题与执行功能中的抑制和转换亚成分则没有相关。之所以会有这种研究结果的差异, Gilhooly 提及 05 年的研究结果中图形流畅性任务并非是一个单纯的执行功能任务,里面涉及到无意识的优势反应可能会影响到研究结果。我们在引言部分增加补充相关论述。详见可参见正文引言部分红色标记处。

**意见 4:** 实验二所用的实验是 Cushen 和 Wiley (2012) 的拓展,在本来实验的基础上做了部分改动。原实验中,作者通过评定被试在答题中对关键圆片的移动性等级来定义被试的问题表征。而本实验则并不要求其对圆片做等级评定,直接要求被试给出他们想到的即时答案。这种改动是否会影响顿悟问题解决的实质? 即没有解决的问题能叫顿悟吗? 被试移动关键圆片和关键位置的随机性有多大? 等级评定可测评被试对问题的真正了解程度,而即时评定就会涉及碰运气问题,如何解释?

**回应：**感谢外审专家的意见。Wiley 的实验解题任务中，作者要求被试在每次进行解答“三角问题”前花费 35s 时间对全部 10 个圆片在解决问题中的重要性进行评定。但我们认为目前的这种评级方法存在一些不足：首先这种答题过程中的等级评定可能会干扰到被试的思维流畅性；其次在 Wiley 本人在研究在总讨论中也提及，被试即使发现了关键圆片，在评级上认为其是重要的，但仍然可能由于移动的位置不对而不能准确地解答问题。基于此，本实验在 Cushen 和 Wiley（2012）基础上对该范式有所调整，则并不要求其对圆片做等级评定，同时增加被试对关键位置的评分，即考虑被试移动关键圆片和找到关键位置，使被试的表征状态更加明确；与此同时，将被试的思考时间由 1 分钟改为 30 秒，更能清楚的看到被试的问题表征转换的过程。

关于专家提到的这种方法引起过多随机性，涉及碰运气的问题。其实在 Cushen 和 Wiley（2012）中被试在最初评级的阶段依然是存在随机性和猜测的过程。所以借鉴 Cushen 和 Wiley（2012）的统计方法，截取的最后四次的的数据均为在规定时间内成功解决问题的。虽然个体在问题解决的早期解决必然存在随机性和猜测的答题过程。但是在僵局和最后的问题解决阶段并不存在，由数据分析可以发现，个体并非是通过试错来最终解决顿悟问题的，d12 阶段显著高于其他阶段，体现出一种突进式的过程。

**意见 5：**实验二部分结果依赖于边缘显著性。边缘显著性值得讨论吗？

**回应：**感谢外审专家的意见。实验二中有部分结果边缘显著，我们调整文中的部分内容表述和篇幅，对于边缘性显著的数据在呈现结果的同时，做更为严谨的讨论，弱化边缘显著数据讨论，对研究 2 的小讨论以及总讨论进行了删减。详细内容可参见研究 2 讨论处。

**意见 6：**实验三中，有 20 名被试实验前对材料进行了难度评定，根据评定结果确定最终的实验材料。相当于这 20 名被试提前熟悉了实验材料，这是否会对这 20 名被试的实验结果造成影响？

**回应：**感谢外审专家的意见。可能我们在原文中未表述清楚，进行实验材料难度评定的 20 人为不参加正式实验的 20 名被试。已经在原文表述中加以补充。

**意见 7：**实验三针对被试有可能一味追求高正确率而做出不诚实的按键反应，设计了 10 个字谜与谜底相互不匹配的材料作为剔除指标，那如何处理被试既没有猜到正确答案也不理解正确答案却按下“J”键的按键反应？

**回应：**感谢外审专家的意见。我们重新查阅了已有的同类研究，字谜与谜底不匹配材料的设置，目的主要是防止个体猜出字谜规律，在具体的数据分析时均不涉及不匹配材料的数据。在此基础上，但是我们认为若被试选择不匹配材料一定程度上也可能说明被试一味追求高正确率或而不诚实按键，故其实在数据分析时，无论个体是按键 F 或 J 键，我们均认为被试不诚实作答，这类被试数据均加以剔除。尽管我们力图更为严格的排除个体不认真作答造成的噪音数据，但也确如专家所言，可能无法完全排除。不过我们根据外审专家 1 的意见，补做了字谜顿悟问题的行为实验，结果与脑电实验中的行为结果相一致，也一定程度上说明实验设计的有效性。

**意见 8：**文字与格式：文章中有缺字、错字现象，需要校对修改；表格式不规范；有冗余文字，如摘要等，全文请准确精炼。t=5.161, p<0.0001, 0.001 就可以了，且无自由度与效应值，自由度和效应值问题请检查全文。

**回应：**感谢外审专家的意见。已经对全文的格式等进行修订。

---

## 第二轮

### 审稿人 1 意见:

**意见 1:** 首先, 作者的研究内容主要围绕执行功能影响不同类型的顿悟问题解决和顿悟不同阶段来展开, 但是研究的引言或者问题提出却只是泛泛而谈, 并不是针对这两个具体问题, 只是停留在论文中研究 1 的回答的层面。引言需要重新进行问题提出。

**回应:** 感谢外审专家的意见。我们重新对引言的问题提出进行了梳理。

**意见 2:** 其次, 研究 1 的顿悟材料的筛选及其表达存在问题。其中各项所选得材料的具体行为学指标是多少, 例如难度或者熟悉度, 另外评定难度和熟悉度的被试与正式实验被试的同质性如何, 也就是能否反映筛选的合理性?

**回应:** 衷心感谢外审专家的宝贵意见。我们在实验 1 所选得的十道顿悟材料的总熟悉度均小于 0.05、平均难度为 0.47。其中各项所选得的顿悟材料的熟悉度亦小于 0.05 (在 36 名被试对我们挑选的 40 道顿悟问题中, 我们在问卷最后要求被试勾选出曾经见过且知道答案的题目序号, 若该题有两人或者两人以上勾选则将该题视为熟悉题目进行剔除, 即我们选择熟悉度小于 2/36), 各项平均难度并未纳入。正式实验中的对材料熟悉度大于 0.05 的被试成绩已被剔除, 86 人为最终有效成绩录入数, 男女生的顿悟成绩可以得出正式实验被试的整套题目平均难度为 0.47。评定难度和熟悉度的被试与正式实验被试均为在校大学生和研究生。我们在研究 1 中顿悟材料筛选部分重新修订, 增加了上述相关内容。详见正文顿悟材料处青绿色标记处。

“由不参与正式实验的 36 大学生 (男、女生各 18 人) 对材料进行熟悉性评定, 解答上述顿悟问题, 无时间限制。首先, 依据熟悉频率大于 0.05 (即超过 1 人评定答过此题, 且知道答案为标准) 去掉熟悉性材料 13 题; 剩下的 27 道题的平均解答率由高到低排序, 前 33% 和后 33% 的题目作为困难材料 (解答率低) 和容易材料 (解答率高) 剔除, 挑选出 10 道问题作为正式研究的顿悟材料 (平均解答率为 0.47)。”

**意见 3:** 第三, 执行功能测试的具体的内容和试次等信息均未给出, 例如执行功能各类测试之间的顺序等, 这使得无法对此操作的合理性进行判断。顿悟任务之前的执行功能测试大概需要 20 分钟, 相对顿悟问题解决, 之前这么久的执行控制任务是否会导致执行控制相关的疲惫或者失活, 或说是认知资源损耗。进而导致现在测量任务对随后顿悟的表现实质上是认知资源损耗导致的。因为之前脑损伤研究显示前额叶的损伤会促进顿悟。

**回应:** 衷心感谢外审专家的意见。我们在研究 1 的程序处补充了具体的三个分测量的顺序, 时间等相关细节。

研究 1 的材料部分, 进一步修订补充执行功能测试三个任务的具体内容和试次:

**工作记忆刷新任务:** “箭头呈现数量分别为 3、4 和 5 个三种系列, 每一系列中每种颜色的箭头各出现 1~2 次。在实验中, 三种系列的箭头随机安排, 被试事先不知箭头何时呈现完毕。每个被试首先练习 3 个 trails 以熟悉实验流程, 正式测验 12 个 trails, 每种长度各 4 个 trails。”

**抑制任务:** “基线任务与冲突任务测验采用 ABBA 的顺序, 共 4 个 block, 包含 2 个基线任务 block (组块), 两个冲突任务组块, 每个组块均有 24 个 trails。被试的成绩为冲突任务下的平均反应时减去基线任务的平均反应时, 冲突任务下的反应时小于基线任务下平均反应时为无效数据, 计分规则为反应时差值越大, 优势反应抑制能力越差。”

专家提到的疲惫等问题确实在实验设计中并未考虑到, 但是由于只在研究 1 中进行的执行功能测试主要目的是进行高低执行功能个体的筛选, 在研究 2b 脑电研究中采用的即为研

究 1 中的被试,结合研究 1 和研究 2 的结果,我们认为研究结果准确性一定程度上可以保证。

**意见 4:** 第四,被试分组时所有的执行功能测试之间没有产生冲突么,比如在记忆刷新上高分者但是在抑制控制或注意转换任务上的得分低?如果有出现,研究者是如何实现这种分组的,主要依据谁?为什么?研究 2 与研究 1 间隔多久,因为被试是相同的。研究 2 中 3.3.1 部分的没有完成任务是被试没有解答出来还是被试是解答错误的?只采用一个任务来进行测试,研究者还没有要求被试对自己是否接触过类似任务进行自我报告,如果被试解答过该任务,反应时数据可能并不能说明问题,为什么作者不对此进行操作检验和被试自我报告的解题经验排除?更重要的是,不知道作者是否意识到执行功能对空间顿悟问题及其不同阶段的影响的说服力非常有限,一方面是只使用了一个顿悟问题,另一方面是只有一个拓展自 Cushen 和 Wiley (2012) 的简单的行为学实验,并且还存在着争议和质疑。若想依据充分,可能该部分需要新的实验结果。

**回应:** 感谢外审专家的意见。我们这里采用的是三个分测验的成绩均高于对应的平均值的记为高执行功能组,三个分测验的成绩均低于对应的平均数的为低执行功能组。主要参考的是李美华,白学军,闫国利(2007)的操作。专家对于原研究 2 提到的问题,我们经过讨论后在这一稿中将整个研究 2 予以删除。

**意见 5:** 第五,统计方法存在问题,这导致结果可能不可信。例如,实验 3a 部分由于被试不等组能否进行重复测量方差分析的 F 检验,是否满足不等组条件下 F 检验所必需的统计前提?该文 22 页的执行功能的分组分析为何只有两个组处理?作者似乎是对三个指标进行分组,应该有三个组处理,每个分组情况下每组多少人?

**回应:** 感谢外审专家的意见。关于统计方法问题,我们认为还是可以进行重复测量方差分析的。但确实在数据报告和呈现上还不够清晰存在问题。我们简单地解释一下:研究 3a(即现在的研究 2a),我们对反应时进行 2(高执行功能 vs. 低执行功能) × 2(顿悟 vs. 无顿悟) 的两因素方差分析,其中执行功能高低组为被试间分组,高执行功能组为 17 人,低执行功能组 19 人,确实不等组。但方差分析时方差齐性,我们认为还是可以进行方差分析 F 检验的。我们同时查阅了相关权威期刊文献,如该文献实验 1 四个分组人数不等,为被试间设计,学习组块为被试内变量。

## The Effect of Feedback Delay and Feedback Type on Perceptual Category Learning: The Limits of Multiple Systems

### Method

**Participants.** One hundred and thirty undergraduate students from the University of Adelaide (Adelaide, South Australia, Australia) participated in return for course credit or a payment of AUD12. Each participant completed one experimental condition. There were 34 participants assigned to the RB No Delay condition, 34 to the RB Delay condition, and 30 and 32 participants assigned to the II No Delay and II Delay conditions, respectively. In all of the experiments reported here, we followed Maddox and Ing

We analyzed the data in two ways. First, a propos of the manner of plotting the data in Figure 3a, we conducted a 2 (category structure) × 2 (feedback delay) × 4 (block) mixed-design ANOVA. This revealed a main effect of block,  $F(3, 282) = 74.78, p < .001$ , indicating learning; a main effect of

关于第二点，专家提到的执行功能的分组分析，我们是分别根据三个测试的平均数，分别比较不同水平的刷新能力、转换能力和抑制能力的个体不同顿悟条件下的反应时差异。因此对于刷新能力，进行 2（高刷新能力 vs.低刷新能力） $\times$ 2（顿悟 vs.无顿悟）的方差分析。结果只有不同水平刷新能力的个体存在显著，抑制和转换能力上均无显著差异。我们这里重新加以表述。

**意见 6:** 最后，如上一次评审提到的脑电的分析和解释存在问题。一方面在电极选择理由未交待，脑电成分的认定过于武断，缺乏依据；另一方面 *Greenhouse-Geisser* 法校正没有根据统计前提，并不是所有情况下都使用该方法，作者一而概之。

**回应:** 感谢外审专家的指正。我们已经在引言部分和电极选择处交待电极选择理由和脑电成分。另外在结果分析表述上我们修订为根据球形检验的结果进行校正。重新修订引言部分内容：

“关于顿悟问题解决的心理机制，买晓琴等(2004)运用事件相关电位 ERP 技术对顿悟的时间进程进行了研究。他们发现，呈现谜语答案以后，“有顿悟”反应比“无顿悟”反应的 ERP 波形在 250~500ms 内有一个更加负性的偏移。这个差异波的潜伏期约为 380ms (N380),且最大波幅值锁定在 Cz 电极点。他们认为 N380 可能与 N2、ERN/Ne 或 Stroop 效应中的 ERP 成分一样，主要与认知冲突有关。沈汪兵等（2012）运用三字谜为实验对象探讨了顿悟过程中思维僵局的早期觉察的脑电机理，结果发现“无僵局”和“有僵局”谜题主动解题过程的脑认知活动差异主要表现为早期 P170 和晚期的晚正成分 LPC(620~800 ms)上。“无僵局”谜题的解题过程诱发了较小的 P170 和 LPC,“有僵局”谜题的解题过程则诱发了较大的 P170 和 LPC。因此，顿悟问题解决的不同阶段存在着特定 ERP 成分。”

**意见 7:** 3.3.3 部分的结果是如何分析的，采用何种方法？

**回应:** 感谢外审专家的意见。该部分的结果我们采用的是重复测量方差分析，但是综合考虑您上面的第 4 条问题，我们讨论后在现在的版本中删除原实验 2 的相关内容，因此正文中不再该部分内容不再呈现。

**意见 8:** 连线任务和顿悟测试的时间交代不明，是各自 10 分钟还是总共 10 分钟？文中还有大量的文字表达或者语句上的问题，有些数字也有误写。例如第 24 页的 280500ms 等。

**回应:** 感谢外审专家的意见。连线任务与顿悟测试各在 10 分钟内。所有的任务被试可以在 30 分钟内完成。

“整体实验分为两个部分，执行功能测试（包含有三个子成分的测量）与顿悟问题解决测试，所有被试在 30 分钟内可以完成所有实验任务。执行功能的三个子成分测验任务之间的顺序，以及执行功能与顿悟问题解决测试顺序 ABBA 平衡，但要求被试一次性完成整套实验。其中，执行功能的抑制与刷新功能的测试使用 Eprime2.0 程序编程，显示器为 1024 $\times$ 768 高分辨率的显示器，连线测试与顿悟问题解决测试均使用秒表计时施测，各自在 10 分钟之内完成，顿悟问题答对一题记 1 分。”

存在文字表达和语法问题，已经重新修订。

---

**审稿人 2 意见:**

本修改稿针对三位初审专家的意见对文中存在的问题进行了深入修改，其中大多数问题得到了较好的解决，但是，本文仍然存在以下问题：

**意见 1:** 第 14 页“2.3.1 执行功能子成分与顿悟问题解决成绩的结果分析”部分对转换、刷新和抑制功能的平均成绩报告没有必要, 这些平均值说明不了什么问题。同时, 表 1 对性别差异的检验结果都不显著, 因此整个这部分内容也没必要报告, 并且性别差异检验在讨论部分没有涉及, 因此建议把这一部分整体都删去。

**回应:** 感谢外审专家的意见。我们重新审读后, 接受您的意见, 将上述部分删除。

**意见 2:** 第 15 页表 3 中的回归系数与此表上面回归方程中的回归系数不对应, 一般回归方程有常数项, 但此方程中却写的是标准回归系数, 这是错误的。

**回应:** 谢谢外审专家的意见。我们对回归系数重新进行更正。

**意见 3:** 第 15 页“2.3.3 执行功能总体水平高低与顿悟的关系”中对高低执行功能者的分组标准没有讲清楚, 是把刷新、抑制、转换三个任务都高于平均值的分为执行工作高水平者, 还是只要其中一项高于平均值的作为高执行功能者? 请阐述清楚。

**回应:** 感谢外审专家的建议。已经在原本中修订阐述清楚。我们是指同时三项均高于相应的平均值为高执行功能组。

**意见 4:** 第 15 页“计算顿悟高低得分与执行功能水平的相关, 得到相关系数值为  $r=0.298$ ”, 是把高分组、低分组得分分别计算相关吗? 执行功能水平的指标是什么? 如果分开来做相关, 那么至少有好几个相关系数, 怎么会只有一个相关系数? 请说明。

**回应:** 感谢外审专家的意见。我们是将所有分数转换为标准分数后, 高分组、低分组顿悟得分分别与对应的三个执行功能子成分进行相关。已经重新进行表述。

“根据顿悟问题解决的平均成绩  $4.81(SD=1.59)$ , 将被试的顿悟成绩分成高低两组, 其中, 低分组 41 人, 高分组 45 人。分别计算顿悟高低得分下与执行功能三个子成分的的相关, 结果表明低顿悟得分与转换得分、刷新得分之间显著相关,  $r=0.37$ ,  $r=0.61$ , 高顿悟得分与三个子成分之间均无显著相关。”

**意见 5:** 第 15 页“2.4 讨论”部分, 仅仅根据相关和回归研究结果是无法得到“刷新功能对顿悟过程中需要较多认知资源的问题表征与问题空间搜索阶段有直接的影响, 而执行功能的抑制过程可能在这两个阶段中较少涉及到”这样的结论的。相关和回归不能揭示认知过程的内部加工阶段。

**回应:** 感谢外审专家的意见。我们重新修订该部分内容, 更为谨慎表述相关结果。修改后的内容为:

“上述结果表明, 高执行功能组个体的顿悟成绩显著高于低执行功能水平组, 执行功能中的刷新功能对顿悟成绩有显著预测作用。刷新主要涉及信息的存储与更新, 即通过用新的信息替代旧的信息来更新记忆内容 (Collette & Linden, 2002)。而顿悟问题解决过程中, 顿悟的表征阶段、搜索阶段与重组阶段都是对信息的整合、更新与转换 (Ash & Wiley, 2006; Sternberg, 1999; 罗劲, 2004)。在初始表征阶段, 对问题的注意状态无疑是形成初始表征的重要因素; 而在问题空间的搜索阶段, 会受到个体的认知资源广度、分配与监控能力的交互影响 (Gilhooly & Fioratou, 2009)。”

**意见 6:** 研究 3a 和研究 3b 没有讨论部分, 建议补充。尤其是 ERP 结果部分, 仅仅列举了各个波形成分的差异检验结果, 应在深入讨论的基础上说明各个波形成分所代表的认知加工过程。文章在总讨论第一段直接说明各个波形成分代表的认知加工阶段, 没有深入分析讨论, 因此这些相关结论就没有依据。虽然作者在后面对此问题进行了深入分析, 但这些分析似乎

可以提前到研究 3 中进行。请作者认真考虑。

**回应：**感谢外审专家的宝贵意见。我们将总讨论中的脑电部分讨论提前到研究 3 中进行深入的讨论。

“研究 3 采用三字谜为实验材料的言语顿悟问题，研究 3a 的行为结果中发现执行功能中的工作记忆刷新亚成分影响顿悟问题解决，高工作记忆刷新功能的个体在顿悟条件下的反应时显著快于低工作记忆刷新功能的个体。而研究 3b 脑电数据进一步表明，“有顿悟”、“无顿悟”均诱发了早期刺激加工成分 N1，且高低执行功能组，以及有无顿悟两种反应所诱发的 N1 成分均不存在显著差异。N1 成分更多的与刺激的物理特性（颜色、大小等）相关，结果表明不同条件下在早期的视觉加工上没有差异。但是被试在对三字字谜进行较高级的视知觉加工过程中在谜底呈现后的 180-240ms 内，顿悟条件下较之无顿悟条件诱发了一个更负的早期成分 N2，N2 与无关信息的门控作用或者早期的选择注意功能的聚集有关(Fuerst, Gallinat & Boutros, 2007; Boutros et al., 2009; 邢强等, 2013)，表明随着物理刺激早期视觉加工的结束，顿悟条件下个体需要花费更多的选择注意，通过注意筛选的刺激将会被进一步加工（沈汪兵等, 2012）。在 190-260ms 内顿悟条件与无顿悟条件均诱发 P2 成分，且在高执行功能条件下，无顿悟条件下比顿悟条件下诱发更大波幅的 P2 成分。P2 也与注意加工以及短时记忆有关(Carreti é et al., 2001; Luck & Hillyard, 1994)，表明高执行功能的条件下，无顿悟的个体投入更多注意资源执行功能个体会有意识的优先加工其更为熟悉的问题表征以确认结果的正确性。此外，沈汪兵（2012）等发现，个体在顿悟问题解决过程的早期知觉阶段能够直觉性的觉察到僵局的存在，顿悟条件下 P2 潜伏期早于无顿悟条件，我们认为极有可能是个体对顿悟问题僵局的觉察引发的。本研究中，P3 成分主要引发于额叶-中央脑区，且在这两个脑区中，顿悟条件的波幅显著大于无顿悟条件，说明相比较确认和判断（无顿悟）按键过程，个体在顿悟条件中看到刺激得到错误的反馈之后，需要打破自己先前的猜谜规则，并且根据答案联系之前呈现的字谜来产生新的连接方式和工作记忆表征，在这个过程中需要投入更多的注意资源和工作记忆（邢强等, 2012）。而顿悟条件的 P3 潜伏期显著长于无顿悟条件，也说明了由于个体发现答案错误，在理解答案时需要联系之前所呈现的谜面信息，所以记忆资源损耗大，回忆时间也较长。在对被试间变量执行功能水平高低组进行分析时，发现高低执行功能水平组被试引发的 P3 成分并无显著差异。因此，言语顿悟问题与空间顿悟问题类似，虽然在顿悟条件下进行问题表征搜索，僵局产生阶段引发更大波幅的 P3 成分，但在僵局打破、实现表征重组并不受到执行功能高低的影响。”

**意见 7：**文中有一些小错误，如摘要中第七行的“工资记忆”；第 10 页倒数第二行“我那天解决”是什么意思？第 14 页第一行“示器”是什么？建议作者对文章进行认真校对。

**回应：**衷心感谢外审专家的认真细致。对于复审中存在这类问题感觉非常抱歉。我们已经重新校对全文，更正文字错误。

**审稿人 3 意见：**

作者对问题进行了很好的回答和说明，且补充了实验，对文章进行了修改和完善，文章逻辑关系更为清晰。评审者无异议。

---

### 第三轮

审稿人 1 意见:

意见 1: 首先, 标题不太适宜, 作者删除了视觉空间类顿悟问题的实验, 目前主要探讨的是语言类顿悟问题, 建议在文章标题中突出语言类顿悟。

回应: 谢谢外审专家的意见。我们接受专家的建议, 将文章题目改为“执行功能对言语顿悟问题解决的影响: 基于行为与 ERPs 研究”。

意见 2: 第二, 作者直接删除了视觉空间类顿悟问题, 这自然成为该研究的不足, 作者应当在讨论中进行合理的阐述。

回应: 谢谢外审专家的意见。我们在总讨论的最后进行了补充论述。

“由于研究范式的限制, 本研究主要着眼于汉字字谜这类言语顿悟问题, 探讨了执行功能对于言语顿悟问题(verbal insight problem)的影响, 并未对空间顿悟问题(spatial insight problem)进行深入研究。虽然不管是采用空间类顿悟还是采用言语类顿悟问题都强调了思维僵局与表征转换作为顿悟问题的重要特征。但已有研究结果普遍发现空间顿悟与言语顿悟的心理与神经机制存在差异(Gilhooly, Fioratou, & Henretty, 2010; Cushen & wiley, 2012)。因此, 以后的研究应进一步区分顿悟问题类型, 探讨执行功能对空间顿悟问题的影响。”

意见 3: 第三, 文章的标点符号、语句表达和引用格式等诸多方面仍存在需要细致修改之处, 建议认真修改。

回应: 感谢外审专家的意见。我们对全文的标点、部分语句表述和引用格式重新进行了修改。由于改动较为零散, 修改说明中不一一呈现, 在正文中我们使用粗体字标记。