

《心理学报》审稿意见与作者回应

题目：新颖语义联结在顿悟促进记忆效果中的作用

作者：陈石 梁正 李香兰 陈嫣然 于全磊 赵庆柏 李松清 周冶金 刘丽中

第一轮

审稿人 1 意见：

研究采用成语谜题选择任务，并通过行为和 fMRI 技术探讨新颖语义联结形成的认知过程在顿悟促进记忆中的作用，具有着一定的创新，但是在材料选择、方法上存在较大问题，具体如下

意见 1: 重要概念介绍不清楚。如引言中“是否能自行解决”、“魔术解密任务”、“主动生成答案”、“主观啊哈体验”、“间接测验”、“复合远距离联想任务”、“汉语谜题的答案选择范式”、“啊哈体验”“记得、熟悉和猜测”等。

回应: 感谢专家的指导意见。对于不清楚的重要概念现进行逐条解释，并在引言中进行说明。

是否能自行解决: 前文研究中的顿悟条件是通过呈现线索的形式让被试产生顿悟感，而非采用在无提示条件下让被试自行解决谜题的形式。现已将“是否能自行解决”改为“是否能在无线索提示下自行解决谜题”。

魔术解密任务: 被试的任务是观看魔术的视频剪辑，视频中会揭示魔术师用来实现魔术效果的方法，对该魔术的原理进行破解。现已将“魔术解密任务”改为“一项观看魔术揭秘的视频任务”。

主动生成答案: 即一个顿悟问题的解决方案是由被试自行想出而非实验者呈现，与之相对的是主试呈现答案。现已将“主动生成答案”改为“被试自主想出答案”。

主观啊哈体验、啊哈体验: 顿悟问题解决时，问题解决者受固定的思维方式或反应倾向的影响，绞尽脑汁却仍不知道怎么解决问题，从而陷入思维僵局的阶段。经过一段时间的酝酿，在某个不可预知的瞬间打破思维的定势，脑海中突然想到新颖的解题思路，此时问题解决者会伴有强烈的情感释放，即“啊哈”体验。主观的“啊哈”体验，即个体感知到的“啊哈”体验，通常被定义为突然意识到解决方案的感觉，同时又感到惊讶并确信解决方案的正确性。一些研究表明，主观的“啊哈”体验并不仅仅存在于人们自主解决问题(主动生成答案)后，在尝试解决问题失败后向其提供解决方案时也可以使他们感知到啊哈体验。(Bowden & Jung-Beeman, 2003a; Kizilirmak, Thuerich, Folta-Schoofs, Schott, & Richardson-Klavehn, 2016)。现已在正文中添加“啊哈”体验的定义。

间接测验: Kizilirmak, Galvao Gomes da Silva, Imamoglu 与 Richardson-Klavehn (2016)的研究中采用“学习-测验”范式探究顿悟问题答案的产生方式与啊哈体验对学习的促进作用，在后期测验阶段研究者采用了两种测验方式，一种是直接测验(再认任务，判断某项目是否为一周前的旧项目，考察旧项目的再认率)一种是间接测验(重做任务，完成新旧混合的项目，考察旧项目的解决率)。现已在文中进行说明。

复合远距离联想任务: 该任务由 Bowden 与 Jung-Beeman(2003b)开发，共有 144 个复合远程关联问题，首先向参与者呈现三个看似不相关的单词(例如 cloth, tennis, manners)，并要求参与者通过与每个单词建立复合词来找到连接其他三个单词的第四个单词(table)。由于任务试次较多，且参与者有机会在较短时间内思考解决问题，该任务适合于脑成像研究。现已在文

中添加说明。

汉语谜题的答案选择范式：本研究采用汉语成语谜题为实验材料，采用了答案选择范式。在实验任务的选择上，虽然答案生成任务最能体现顿悟问题解决的过程与特点，但是采用此实验任务完成时间较长，成功率较低。采用呈现答案并确定顿悟感的判断任务，又难以反映完整的问题解决过程(黄福荣，周冶金，赵庆柏，2013; Zhao, et al., 2013)。本研究采用折中的选择答案的任务范式，为谜题设置寻常答案、新颖答案(谜底)、似是而非答案、无关答案等几类备择答案，可以较好的兼顾这两个问题。现已在文中添加说明。

记得、熟悉和猜测：由于在测试阶段中，被试在回忆一周前所作出选择时，有可能对某些题目印象深刻，对某些题目已经遗忘，因此他们在做出判断的根据可能是记得、熟悉或者猜测。将三者进行区分可以更好的了解被试记忆的准确性。现已在实验一的 2.2.4 实验程序介绍中加入说明。

参考文献：

Bowden, E. M., and Jung-Beeman, M. (2003a). Aha! insight experience correlates with solution activation in the right hemisphere. *Psychon. Bull. Rev.* 10, 730–737.

Bowden, E. M., & Jung-Beeman, M. (2003b). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers*, 35(4), 634–639.

Kizilirmak, J. M., Thuerich, H., Folta-Schoofs, K., Schott, B. H., & Richardson-Klavehn, A. (2016). Neural correlates of learning from induced insight: A case for reward-based episodic encoding. *Frontiers in Psychology*, 7, 1–16.

Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., & Richardson-Klavehn, A. (2016). Generation and the subjective feeling of "aha!" are independently related to learning from insight. *Psychol Res*, 80(6), 1059-1074.

黄福荣，周冶金，& 赵庆柏. (2013). 汉语成语谜语问题解决中思路竞争的眼动研究. *心理学报*, 45(1), 35-46.

Zhao, Q. B., Zhou, Z. J., Xu, H. B., Chen, S., Xu, F., Fan, W. L., & Han, L. (2013). Dynamic neural network of insight: A functional magnetic resonance imaging study on solving Chinese ‘chengyu’ riddles. *PLoS One*, 8(3), e59351.

意见 2：引言最后一段，作者指出 Kizilirmak 等人(2019)研究中使用的复合远距离联想任务本身无现实意义、形成的联结无使用价值，并且认为汉语谜题解决过程中形成的联结具有现实价值。这一重要论述是本文的创新所在，但是阐述过于简略，为什么两个任务存在这样的差异需要进一步阐述清楚。

回应：感谢专家的指导意见，现进行说明

经典的顿悟研究多采用 9 点问题(Maier, 1930)或蜡烛、火柴和图钉盒(Duncker, 1945)等广为人知的问题。尽管这些任务非常适合于研究行为表现或顿悟体验，但测量顿悟产生过程及潜在神经活动时可能需要对相同类型的题目进行多次测量。目前由 Bowden 与 Jung-Beeman (2003)开发的复合远距离联想任务可以满足这一要求，该任务基于 Mednick (1962)的远距离联想测验开发，答案词与三个题目词能组成复合词，从而形成新颖的语义联结、激发被试的啊哈体验。但是这种任务形式属于以实验为目的而人为进行编制的，其仅存在于部分创造力研究当中，在日常生活并不存在此类解题形式，脱离了实验情境后便失去意义。

谜底为汉语成语的谜语问题被称为成语谜语问题。在猜测成语谜语问题的谜底时，参与者一般需要重新编码谜面信息，即重构问题的心理表征，形成新颖语义联结，并获得啊哈体验，因此成语谜题属于顿悟问题(朱新秤，李瑞菊，周冶金，2009)。与复合远距离联想任务不同，汉语成语谜题任务的素材取自汉语成语谜语库，其中谜底的成语与谜面的词组均是现实

生活中真实存在且常用的，这种解谜形式也广为流传，脱离实验情境后参与者仍可以将其运用到日常生活的休闲娱乐、文字游戏中，具有很强的现实价值。

相应的表述已添加至正文中。

参考文献：

Duncker, K. (1945). On problem-solving. *Psychol. Monogr.* 58, i-113.

Maier, N. R. F. (1930). Reasoning in humans. I. On direction. *J. Comp. Psychol.* 10, 115–143.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220–232.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers*, 35(4), 634–639.

朱新秤, 李瑞菊, 周冶金. (2009). 谜语问题解决中线索的作用. *心理学报*, 41(5), 397–405.

意见 3: 实验一中，在材料选择过程，作者为每个谜面配置了四个答案，其中一个新颖答案，一个寻常答案和两个无关答案。其中寻常答案是作者从预实验中选出的生成率较高的成语，这里的生成率数据需要具体指出。此外，实验要求被试从四个答案中选出“新颖且合适的答案”，但是从最终结果可以看出，被试仍有较大的概率选择了寻常答案，也就是认为寻常答案是“新颖的”，那么这是否说明寻常答案的新颖性没有得到很好的控制，进而使新颖答案和寻常答案的差异不够明显？更为重要的是，在数据分析部分，作者根据被试选择的答案类型划分出新颖联结和寻常联结，虽然结果表明两类联结在反应时和评级上存在差异，但是显然被试在选择这些寻常答案时是认为该答案是四个备选答案中新颖的。在这样的条件下作者在分析数据时划分出新颖联结和寻常联结似乎不够妥当。实验二也有同样的问题。

回应: 感谢专家的宝贵意见。

寻常答案的平均生成率已在正文中给出。

顿悟是个体在问题解决陷入思维僵局后，打破思维定势并突然完成问题解决的创造性思维形式，在问题解决的同时会伴随“啊哈”的顿悟感。顿悟的思维本质是打破思维定势，建立新颖且有价值的联结，而顿悟感是这一认知过程的副产品。先前研究探究了顿悟感对随后记忆的影响，而本研究则重点考察的是新颖语义联结这一认知过程的作用。由此，本研究对新颖语义联结条件的划分主要是依据成语谜题解决的认知过程，而非顿悟感判断。在成语谜题的解答中，新颖答案的选择需要对成语的整体含义进行破解，并对关键字的含义进行别解（抑制凸显语义，激活非凸显语义），组合形成新的语义。这一过程包含了组块破解和表征变换等顿悟关键的认知过程（罗劲, 2004; Zhao, et al., 2013），因此被界定为新颖语义联结条件。而寻常答案与谜面之间是相对直接的语义联系，其加工过程并不存在定势打破，因此被界定为寻常语义联结条件。在修改稿中，我们通过举例对新颖答案选择中的新颖语义联结形成过程进行了详细解释。

本实验中要求被试从备选答案中选出新颖且合适的答案，但被试仍有选择寻常答案的情况。我们认为，这可能是由于被试无法在限时的情境中建立谜面与谜底间的新颖联结，从而导致时间结束前被迫选择了仅与谜面具有一般关联的寻常答案。实验 1 中寻常答案的反应时 ($3.85 \pm 0.54s$) 显著长于新颖答案的反应时 ($3.65 \pm 0.51s$)，不过在剔除问题难度等因素后，反应时的显著性差异消失了。这说明被试选择寻常答案的部分试次中确实存在相对的困难，在一定程度上支持了我们的猜想。

当然，我们也无法排除专家所提出的被试在部分试次中认为寻常答案就是新颖的这种可能。不过我们对被试的顿悟感判断分析发现，在选择寻常答案的试次中只有 14.07% 被判定为强顿悟感，在被试选择无关答案的试次中也有 12.59% 被判断为强顿悟感。这说明若将顿悟感作为新颖联结形成的判断标准也是存在一定问题的。

鉴于本研究中重点考察的是新颖语义联结过程对随后记忆的影响，而非顿悟感，所以我

们还是以顿悟认知过程的发生作为界定新颖语义联结的依据。尽管被试在部分寻常答案选择中也报告了高顿悟感，但其比例较小，对结果的影响应该是有限的。

参考文献：

罗劲（2004）顿悟的大脑机制. *心理学报*, 36(2): 219-234

Zhao, Q. B., Zhou, Z. J., Xu, H. B., Chen, S., Xu, F., Fan, W. L., & Han, L. (2013). Dynamic neural network of insight: A functional magnetic resonance imaging study on solving Chinese 'chengyu' riddles. *PLoS One*, 8(3), e59351.

意见 4: 实验二通过核磁技术记录了被试在完成成语谜题选择任务中的大脑活动，但是每个谜面所对应的三类备选选项是同时呈现的，即新颖答案、寻常答案和似是而非答案是同时出现的，也就是说后续分析的核磁数据包含了对新颖答案、寻常答案和似是而非答案的加工过程，三类答案引发的大脑活动混淆在一起，并且三类答案的呈现位置是平衡的，也就是核磁数据记录的大脑活动还受到三类答案出现顺序的影响。此外，在限定时间内被试还需从三个备选答案中做出选择，也就是核磁数据记录的大脑活动不仅包含了三个备选答案的语义加工，还包含了选择、对比等其它加工过程。综上，通过分析包含了以上加工过程的核磁数据是否能够区分出新颖联结和寻常联结？

回应：感谢专家的宝贵意见，您提出的问题非常具有价值。

目前关于顿悟的实验范式分为三类：自主生成答案、呈现答案、答案选择。由于顿悟问题难度较高，自主生成答案需要大量时间，而直接给被试呈现答案范式又更像领悟，答案选择范式则能更好的进行兼顾(黄福荣，周冶金，赵庆柏，2013)。在答案选择范式中被试在完成新颖选择时可能进行了一系列初始问题表征、各选项间思路竞争、僵局和表征重构、新颖联结形成等阶段，并进行了最终的选择。这比直接呈现新颖答案让被试进行理解更为复杂，但是这一过程与被试在进行顿悟答案的自主生成过程类似，且更符合创造性问题解决的认知加工过程理论。根据认知加工过程理论中生成-选择模型(Bink & Marsh, 2000)，创造力的认知是产生合成和选择这两个综合过程相互作用的结果。在产生合成阶段，创造者从相关领域进行广泛的信息搜索，这些信息的片断临时地组合在一起；在选择阶段，创造者对这些组合体通过选择、保留或放弃来形成一些小的子集，以满足任务要求。顿悟问题解决是一种典型的创造性问题解决，因此我们所采用的答案选择范式与顿悟问题解决的加工过程相符。

那包含了以上加工过程的核磁数据是否能够区分出新颖联结和寻常联结呢？答案是肯定的。Bink 等人(2000)认为创造活动运用了与非创造活动一样的认知加工，二者的差别在于创造性认知选择了新颖的信息，而非创造性认知更多的是采用寻常的信息。信息选择以及新颖组合的形成是创造活动的关键过程。由于选项固定，被试的信息选择更可能被限制在一定的范围内，由此被试最终选择答案的类型不同可以反推其认知加工过程上的差异。更为重要的是，已有先前研究采用了成语谜语的答案选择范式来研究新颖联结与寻常联结在认知神经加工上的差异(Zhao et al., 2013; Zhao, Zhou, Xu, Fan, & Han, 2014)，其激活脑区与字谜的自主生成答案范式、歇后语的答案呈现范式的结果是一致的(Luo & Niki, 2003; Milivojevic, Vicente-Grabovetsky & Doeller, 2015; Preston, Shrager, Dudukovic & Gabrieli, 2004)。本研究 fMRI 结果与先前研究也是基本一致的。

参考文献：

Bink, M. L., & Marsh, R. L. (2000). Cognitive regularities in creative activity. *Review of General Psychology*, 4, 57-78.

Danek, A. H., Williams, J., & Wiley, J. (2020). Closing the gap: connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychol Res*, 84(1), 111-119.

- 黄福荣, 周治金, & 赵庆柏. (2013). 汉语成语谜语问题解决中思路竞争的眼动研究. *心理学报*, 45(1), 35-46.
- Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight reconfigures hippocampal-prefrontal memories. *Current Biology*, 25(7), 821-830.
- Luo, J., and Niki, K. (2003). Function of hippocampus in "insight" of problem solving. *Hippocampus* 13, 316-323.
- Preston, A.R., Shrager, Y., Dudukovic, N.M., and Gabrieli, J.D.E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* 14, 148-152.
- Zhao, Q., Zhou, Z., Xu, H., Chen, S., Xu, F., Fan, W., et al. (2013). Dynamic Neural Network of Insight: A Functional Magnetic Resonance Imaging Study on Solving Chinese 'Chengyu' Riddles. *Plos one*, 8(3), e59351.
- Zhao, Q., Zhou, Z., Xu, H., Fan, W., Han, L. (2014). Neural pathway in the right hemisphere underlies verbal insight problem solving. *Neuroscience*, 256, 334-341.

意见 5: 在确定样本量时, 作者只是指出了先前研究的被试量范围, 但是建议作者选择与自己所关注效应及实验设计相近的研究作为参考, 并给出具体的效应量及最终样本量的确定过程。

回应: 感谢审稿专家的宝贵意见。

本研究样本量根据两个方面进行确定: 在实验一中, 根据先前关于顿悟促进记忆的行为研究被试量为 20 人到 30 人, 即可得到显著的差异性结果(Danek, Fraps, von Muller, Grothe, & Ollinger, 2013; Kizilirmak, Galvao Gomes da Silva, Imamoglu, & Richardson-Klavehn, 2016), 本实验被试量与这些研究相当; 根据关于顿悟促进记忆效果的类似研究(Danek et al, 2013)中记忆效果差异的效果量(Cohen's $d = 0.61$), 采用 G-power 软件计算设置 Power 为 0.8, α 水平为 0.05, 计算需要被试样本量为 24。实验一样本量满足此要求。

实验二根据先前关于顿悟促进记忆的 fMRI 研究被试量为 17 人到 26 人, 即可得到显著的差异性结果(Kizilirmak et al., 2019; Ludmer, Dudai, & Rubin, 2011)。根据同样使用 fMRI 的 Kizilirmak, Thuerich, Folta-Schoofs, Schott 和 Richardson-Klavehn (2016)的研究中顿悟条件与控制条件记忆差异的效果量(Cohen's $d = 0.955$), 采用 G-power 软件计算设置 Power 为 0.8, α 水平为 0.05, 计算需要被试样本量为 11 人。实验二样本量满足此要求。

相应的表述已添加至正文中。

参考文献:

- Danek, A. H., Fraps, T., von Muller, A., Grothe, B., & Ollinger, M. (2013). Aha! experiences leave a mark: facilitated recall of insight solutions. *Psychol Res*, 77(5), 659-669.
- Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., & Richardson-Klavehn, A. (2016). Generation and the subjective feeling of "aha!" are independently related to learning from insight. *Psychol Res*, 80(6), 1059-1074.
- Kizilirmak, J. M., Thuerich, H., Folta-Schoofs, K., Schott, B. H., & Richardson-Klavehn, A. (2016). Neural Correlates of Learning from Induced Insight: A Case for Reward-Based Episodic Encoding. *Frontiers in Psychology*, 7, 1693.
- Kizilirmak, J. M., Schott, B. H., Thuerich, H., Sweeney-Reed, C. M., Richter, A., Folta-Schoofs, K., & Richardson-Klavehn, A. (2019). Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. *Conscious Cogn*, 69, 113-132.
- Ludmer, R., Dudai, Y., & Rubin, N. (2011). Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. *Neuron*, 69(5), 1002-1014.

意见 6: 实验二的讨论部分作者解释了为什么实验二的实验设置与实验一存在不同, 建议将这一部分挪到方法部分。

回应: 感谢审稿专家的宝贵建议。现已将实验二的讨论部分补充至方法部分。具体如下:

在实验 1 中, 新颖联结和寻常联结条件在测试阶段的正确率分别为 84.10% 和 68.35%, 这意味着失败记忆的试次较少。然而, fMRI 数据分析需要对成功记忆和失败记忆进行比较。为保证失败记忆条件有足够的试次用于数据分析, 实验 2 从成语谜语材料库(黄福荣 et al., 2013)中选取 98 道题目, 8 道用于练习阶段, 90 道用于正式阶段。

同时, 编码阶段答案选择的难度以及测试阶段记忆提取的难度被提高。在答案选择任务中无关答案被替代为似是而非答案, 此时每道谜题设置 3 个备择选项, 包括一个新颖答案、一个寻常答案和一个似是而非的答案。新颖答案与谜面存在内隐的语义关联, 是成语谜语原本匹配的答案。寻常答案是以往研究者从预实验被试针对谜面生成的成语中所挑选出来的生成率较高的成语。似是而非答案是从预实验被试针对谜面生成的成语中所挑选出来与谜面存在一定关联但扣合度相对不足的成语。表 4 列举了在实验 2 中所用到的两道谜题以及备择答案。在测试阶段的答案选择界面, 谜面下方的备择答案不是完整的四字成语, 而是只有成语的第二和第三个字, 如, 编码阶段的“积劳成疾”, 在测验阶段呈现的是“__劳成__”。

意见 7: 文中中英文括号混用, 建议对符号进行统一。此外还有个别存在错别字的地方, 如 3.3.1 行为结果下面第二段中“两种田间不存在显出差异”, 更多类似问题请作者自查。

回应: 感谢专家的细致点评, 现已在对文章进行仔细阅读并更改错误之处。

.....

审稿人 2 意见:

该研究采用成语谜底选择范式来探究顿悟促进记忆的机制, 作者提供了行为学和神经影像学的数据支持。研究发现新颖联结条件促进了一周后再认任务表现, 且发现海马激活于新颖连接的促进效应有关。

该主题是创造性领域的热门话题, 近几年关于记忆与顿悟、创造性思维的研究逐渐兴起, 从其他认知领域来探究创造性思维的本质是可行的。我认为该研究探讨了一个前沿性问题, 立意新颖, 实验操作严谨。尽管在方法等部分存在一些问题, 但整体来看是一篇值得推荐的研究, 建议给修改机会。

意见 1: 摘要“说明在编码阶段顿悟问题解决过程中”通读全文后才能理解编码阶段顿悟问题解决这句话, 如果只看摘要很难理解。按照认知的过程应该是顿悟在前, 编码在后, 建议更正为: 顿悟问题解决过程中的信息编码阶段。

回应: 感谢审稿专家的宝贵意见, 由于我们之前在文中使用的“编码”较为广义(指整个前测阶段), 与记忆的认知过程中所特指的“编码”存在歧义, 因此我们不再将前测阶段称为“编码阶段”, 而是参考同样使用学习-测验范式 Ludmer 等人(2011)的研究, 将其改称为“学习阶段”, 现已将摘要中“编码阶段顿悟问题解决过程中”改为“学习阶段顿悟问题解决过程中”。

参考文献:

Ludmer, R., Dudai, Y., & Rubin, N. (2011). Uncovering camouflage: amygdala activation predicts long-term memory of induced perceptual insight. *Neuron*, 69(5), 1002-1014.

意见 2: “可能反映了顿悟问题解决中新颖联结形成过程的情节记忆以及新颖且有价值的语义联结的建立”

回应：此处没有明白专家的审稿意见，因此未作回答。

意见 3：引言中最后一段说明海马的无作用，这一点不是因为材料的无意义。而是在将新信息纳入已有知识结构的图式中时，mPFC 和海马都会激活，但是随着时间推移，海马活动被 mPFC 抑制。因此记忆会独立于海马。参考 How schema and novelty augment memory formation

回应：感谢专家的宝贵意见，我们阅读了您推荐的相关文献。文中提到新的语义信息最初依赖于内侧颞叶诸如海马等结构，但随着时间的推移，该信息变得相对更依赖于新大脑皮层如 mPFC。Kizilirmak 等(2019)的研究是仅在开始时的学习阶段（顿悟产生阶段）进行了功能磁共振扫描，而没有扫描 24 小时后的测验阶段（回忆阶段），若测试阶段中海马没有起到促进记忆的作用，我们可以将其解释为海马活动被 mPFC 抑制，不过我们对测试阶段海马是否激活不得而知。在他们的研究中学习阶段顿悟与非顿悟条件间比较发现了海马的激活，但是这一差异没有对后面的回忆效果起促进的作用，而海马长期以来一直被认为与陈述性信息学习、情景和语义记忆有关(Opitz, 2014; van Kesteren, Ruiters, Fernandez, & Henson, 2012), 在记忆的形成与巩固中起着重要的作用(Eichenbaum, Yonelinas, & Ranganath, 2007; Squire, Zola-Morgan, & Clark, 2007), 因此我们对这一现象产生疑惑，并认为可能是由于实验材料的不同造成。

参考文献：

Eichenbaum, H., Yonelinas, A.R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Reviews of Neuroscience*, 30, 123–152.

Kizilirmak, J. M., Schott, B. H., Thuerich, H., Sweeney-Reed, C. M., Richter, A., Folta-Schoofs, K., & Richardson-Klavehn, A. (2019). Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. *Conscious Cogn*, 69, 113-132.

Opitz, B. (2014). Memory Function and the Hippocampus. *Monographs in Neural Sciences*, 34(3), 51-59.

Squire, L. R., Zola-Morgan, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, 8(11), 872–883.

van Kesteren, M. T., Ruiters, D. J., Fernandez, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends Neurosci*, 35(4), 211-219.

意见 4：方法和结果部分：作者把前测定义为编码阶段，但其实验不同于记忆研究中的编码阶段，一是编码的对象是什么，没有界定，是否作者认为这一阶段所有过程都属于编码吗？其二，在这个阶段是否被试被告知在一周后会有再认任务，请在文章中明确指导语。

回应：感谢专家的评审意见，下面进行分条解答。

1、本实验中定义的编码阶段不同于记忆研究中，属于更为广义上的编码，编码(encoding)这一称呼参考了以往相关研究(Kizilirmak, Galvao Gomes da Silva, et al., 2016; Kizilirmak et al., 2019; Kizilirmak, Thuerich, Folta-Schoofs, Schott, & Richardson-Klavehn, 2016), 此类研究中前测阶段均被称为编码阶段。被试在此阶段中的过程(包括谜面、谜底、啊哈体验)均属于编码内容。为了防止歧义，现将“编码阶段”改为“学习阶段”。

2、在学习阶段中并未告知被试在一周后会有再认任务，而仅告知其一周后同一时间进行另一个实验。现将实验一指导语在文章中进行概述，并将完整版呈现如下（实验二指导语与实验一基本一致）。

学习阶段：

欢迎参加本实验！本实验分为两个阶段。

第一阶段是成语谜语答案选择任务，要求从四个备选答案中为谜语的谜面选择一个新颖的答案。屏幕中央呈现红色的注视点“+”提示任务开始。接着呈现谜语的谜面，请仔细阅读

谜面并思考答案，然后在谜面下面会呈现四个备选答案，分别对应“A”“B”“C”“D”键，要求你在6秒之内仔细思考理解谜语后选出一个你认为新颖且合适的答案。若6秒之内还未作答，系统自动进入下一界面。提示：不是所有与谜面之间存在语义关联的备选答案都满足新颖性要求。如果你已明白上述指导语，按p键进入下一页指导语界面。

第二阶段是“啊哈”体验程度评定任务。在此之前，你可能有过“啊哈”体验，如先前不理解的问题你突然想到了答案。你不确定你是如何想出答案的，但是却对答案的正确性深信不疑。可能你不仅在自己想出答案时体验到“啊哈”，也可能在自己没有想出来但他人告诉你答案时体验到“啊哈”。比如，一个朋友向你讲了一个笑话，但是你没有发现笑点，然后他稍微解释一下，你突然就明白了。在本实验中，“啊哈”体验可能与现实生活中的体验有所不同，主要有四个特征：(1)在意想不到时，你突然领悟了成语谜语的答案；(2)一旦理解答案与谜语之间的联系，理解答案就毫不费劲并且很清晰；(3)对答案的正确性深信不疑；(4)突然领悟时常伴有积极情绪。在这一阶段，根据上面对“啊哈”体验的描述，你需要在6秒内对你选择答案时体验到的“啊哈”程度进行1-2-3-4-5-6评分。1表示无“啊哈”体验，6表示体验到强烈的“啊哈”体验。数字越大，感觉越强烈。按相应的数字键进行选择。如果你已明白上述指导语，按p键进入下一页指导语界面。

测试阶段：

欢迎参加本实验！本实验分为两个阶段。

第一阶段是成语谜语答案回忆任务，要求从四个备选答案中回忆出你在上周实验中选择的那个谜语答案。屏幕中央呈现红色的注视点“+”提示任务开始。接着呈现成语谜语的谜面，请仔细阅读谜面并回忆上周选择的答案，然后在谜面下面会呈现三个备选答案，分别对应“A”“B”“C”“D”键，要求你在6秒之内仔细思考理解谜语后选出你在上周实验中为该谜语选择的答案。若6秒之内还未作答，系统自动进入下一界面。如果你已明白上述指导语，按“p”键进入下一页指导语界面。

第二阶段是记忆来源区分任务。在这一阶段，你需要对刚才的选择原因进行“1 = 记得”或“2 = 觉得熟悉”或“3 = 猜测”判断。“1 = 记得”表示记得上周进行实验时的一些信息，如，记得指所选答案时的一些想法，记得所选答案的谜面，等等；“2 = 觉得熟悉”表示对所选答案有一种熟悉感；“3 = 猜测”表示所选答案的是猜测的结果。要求你在4秒内按相应的数字键进行选择。如果4秒内还未作答，系统会自动进入下一界面。请按真实情况作答。实验匿名，可放心答题。如果你已明白实验流程与要求，按p键进入练习部分。如有任何疑问，可询问实验主试。

参考文献：

- Kizilirmak, J. M., Galvao Gomes da Silva, J., Imamoglu, F., & Richardson-Klavehn, A. (2016). Generation and the subjective feeling of "aha!" are independently related to learning from insight. *Psychol Res*, 80(6), 1059-1074.
- Kizilirmak, J. M., Schott, B. H., Thuerich, H., Sweeney-Reed, C. M., Richter, A., Folta-Schoofs, K., & Richardson-Klavehn, A. (2019). Learning of novel semantic relationships via sudden comprehension is associated with a hippocampus-independent network. *Conscious Cogn*, 69, 113-132.
- Kizilirmak, J. M., Thuerich, H., Folta-Schoofs, K., Schott, B. H., & Richardson-Klavehn, A. (2016). Neural Correlates of Learning from Induced Insight: A Case for Reward-Based Episodic Encoding. *Frontiers in Psychology*, 7.

意见5：行为统计方法采用配对样本检验并没有错误，但选择新颖与常规答案的数量不一致，以及条目本身的差异(比如难度、问题字数等)在统计分析中没有排除，建议采用 Hierarchical Linear Model 进行统计分析。

回应：感谢专家的评审建议。

在实验反应时和记忆正确率的计算中确实未考虑新颖与寻常答案选择数量以及题目难度和问题字数等问题，在与学院高级统计任课教师讨论之后，我们采用更适合于分析本研究数据结构的混合线性模型（LMMs）方法对行为数据重新进行统计分析，记忆正确率的统计分析结果与配对 t 检验分析结果基本一致。我们对修改稿中的方法和结果部分进行了相应的修改。

意见 6：“采用一般线性模型得到四种条件及其相互比较的个体激活图，并对其进行组分析”，四种条件两两比较本会存在多重比较的问题，会提高阳性结果出现。请使用 flexible factorial design 进行影像数据统计分析。参考：

http://www.sbirc.ed.ac.uk/cyrl/download/Contrast_Weighting_Glascher_Gitelman_2008.pdf

回应：感谢专家的评审建议。

fMRI 数据已按照 flexible factorial design 方法进行了重新的统计分析。结果依然发现海马的显著激活，支持了其在新颖语义联结促进记忆中的作用。数据分析方法介绍和相应的结果已在实验 2 中进行了相应修改。

意见 7：“由于成语谜题的解决过程相对复杂，同类材料中不同题目的认知加工过程存在差异，甚至对于相同题目不同被试的认知加工也可能不完全相同，因此 fMRI 组分析参考先前成语谜题任务研究(Zhao et al., 2013)采用相对宽松的阈值,进行条件间比较,选择阈值 $p < 0.001$ (未校正)且连续激活体素大于 20 的区域。”这不是采用宽松阈值/未校正的理由，且未校正难以通过同行评议。鉴于已有数篇同类的研究，作者可以假设驱动定义 ROI，采用 ROI 分析，而非从全脑分析获得结果。其实作者已在引言部分论述了可能激活区域，建议查阅更多文献，加强假设的脑区，采用 ROI 分析。

回应：感谢专家的宝贵建议。

本研究的问题提出从实验材料和 fMRI 结果上对先前研究进行了质疑，特别强调了海马应该在在新颖语义联结促进记忆中发挥作用。因此，在修改稿中选择海马作为感兴趣区，进行了 ROI 分析。根据 flexible factorial design 分析中的交互作用相关结果确定 ROI 的坐标，提取在每个条件下海马的 Beta 值，进行 2（语义联结类型：新颖联结 VS 寻常联结） \times 2（记忆类型：成功记忆 VS 失败记忆）重复测量方差分析。结果发现，在新颖语义联结条件下，成功记忆的海马激活显著高于失败记忆；而在寻常语义联结条件下，二者不存在显著差异。

修改稿中已经增加了相应的方法和结果描述。

意见 8：在个体水平上海马激活是否与正确率有关？建议用相关分析探究两种条件下海马激活与回忆率的关系。

回应：感谢专家的宝贵建议。

为了探究海马的激活程度是否与回忆正确率有关，我们通过 con_*.img 文件提取了新颖联结条件下成功记忆和失败记忆比较的 contrast 值，并计算了它与新颖条件下回忆正确率的 Pearson 相关系数，结果显示二者不存在显著的相关性。

推测其原因，可能是：研究对记忆效果的测量采用的正确率指标，即记住与否，而非类似记忆程度这样的连续变量。当海马激活达到一定程度时，被试便能成功回忆。不过当海马激活程度超过这一水平继续增强时，当前试次的记忆效果（比如记忆细节等）可能有所提升，但是并不会提高被试的正确率，因为当前试次已经被标记为成功回忆了，也就是在正确率上只能贡献 1 次。

鉴于数据结果并未发现显著的相关，且上述解释也只是推测，缺乏其他的证据，因此本

结果并未正文中添加和讨论。

意见 9: 新颖条件下脑激活模式中涉及情绪脑区,但在结果讨论中没有关于情绪方面促进记忆的讨论。关于顿悟促进记忆可能是一个复合的作用。

回应: 感谢专家的宝贵意见,由于将新颖联结条件下记忆相关脑区与寻常联结条件下记忆相关脑区进行比较,结果发现只有右侧海马在顿悟促进记忆中发挥作用,因此本文在讨论中仅讨论了海马作为新颖语义联结的神经基础的作用。但正如专家所言,实验二中将新颖联结条件下成功记忆与失败记忆的在学习编码阶段的脑活动进行比较,还得到除海马外其他新颖条件下记忆相关脑区,说明了顿悟条件下的记忆可能是一个复合的作用过程。现已讨论中补充相关内容。

意见 10: 在两个实验中,为何在“记得”条件下,采用不同的记忆测试方法对寻常联结没有影响,在新颖联结条件下,回忆率大幅度减少。

回应: 感谢审稿专家的细致点评,您指出的问题十分重要。与实验一相比,实验二中记忆测试提高了回忆难度。对于新颖联结条件而言,其正确回忆率减少且因“记得”而做对的比例降幅显著,表现出遮盖备选项而导致的准确回忆困难。然而采用较难的记忆测试方法却没有使得寻常联结条件受到太多影响,这可能是由于谜面与寻常答案间存在较为紧密且一致的语义关联,即寻常答案通常为谜面的同义词或近义词,谜面可以成为其进行回忆时的线索,被试可以根据谜面较为简单地识别出之前所选的寻常答案,从而在正确率方面与实验一的结果差异不大。然而新颖联结条件下并不能通过此方法进行选择,因此记得条件下的回忆率答大幅度减少。

意见 11: 图 1,第 4 反应屏应为“谜底/答案”。

回应: 感谢专家的细心指正,原图 1 中第 4 屏中“谜面”的意思是在上部呈现谜面,下部呈现选项,并非此页为谜面,由于造成歧义,现已经将在图中加入相关说明。

意见 12: 请用 multi-slice 作图呈现表 6 结果。

回应: 感谢专家的建议。在接受专家对数据进行重新分析之后,表 6 包含了多次比较的脑区激活结果,若用 multi-slice 方式作图,那么图会比较复杂且复杂,因此修改稿中依然只以表格形式进行结果呈现。不过对本研究重点关注的海马,我们采用了 multi-slice 作图进行了可视化展示。

.....
审稿人 3 意见:

这项研究很有意义,但有以下两个问题需要澄清:

意见 1: 在实验流程中,被试先看一个字谜的谜面以及与之相配四个答案(包括一个富有内涵和创意的答案、一个直白的普通答案和两个无关答案),被试的任务是选出他认为新颖而适合的答案,被试有约百分之六十的情况选了内涵创意答案,百分之三十的情况选了普通直白答案——这项研究的关键,就是在人们对这两类选择的记忆及其相应的脑机制进行研究比较。我的最主要问题就是,即使是普通直白的答案,被试之所以选择它,也是认为该答案在四个选项中最为新颖适合,因此,从被试的主观判断角度来看,这两类选择没有质的区别(我推测被试之所以弃内涵创意答案不顾而选择普通直白答案,可能是他没看懂内涵创意答案或者认为它不够好),在这两者之间进行比较的理由似乎不够充分。

回应：感谢专家的宝贵意见。

本问题和审稿专家 1 的第 3 个问题是相似的，我们进行了详细的分析和讨论，请见相应的回应。

意见 2：核磁实验需要报告事后记住和忘记的"内涵创意答案选择"和"普通直白答案选择"各有多少项目数，一般而言，如对某一被试而言他在特定条件的项目数量(如事后成功记忆或事后忘记的"普通直白答案选择")少于 10 个 trial 的话，事件相关分析的叠加数量就有可能不够了。

回应：感谢专家的宝贵意见。

在最终选取的有效被试中，fMRI 实验四个条件下的试次数目分别为：成功记忆新颖条件平均为 24.9 个试次，最少为 8 个；失败记忆新颖条件平均为 14.1 个试次，最少为 9 个；成功记忆寻常条件平均为 21.6 个试次，最少为 8 个；失败记忆寻常答案平均为 11.7 个试次，最少为 8 个。

尽管实验存在不足 10 个试次情况，但有别于常规意义上的事件相关分析（事件的持续事件 duration 为 0），本研究中的事件持续时间较长（3s 左右），因此在数据分析中每个条件下均有不少于 12 个 Scan（3s*8/2s）的数据点，应该达到了数据分析的基本要求。

第二轮

审稿人 1 意见：

同意发表。建议补充说明一下海马的 ROI 具体是如何选取的，如果是根据语义联结类型与记忆类型的交互作用分析结果来确定 ROI 的话，则不应进行后续的 2 乘 2 分析。

回应：感谢专家对论文研究的肯定和指导建议。海马的 ROI 是以海马脑区的激活极值点为中心，提取半径为 6mm 的球体内所有体素，计算其 Beta 值的平均值作为感兴趣区的激活强度值。此部分说明已在修改稿中添加。尽管海马 ROI 是根据交互作用分析的结果来确定的，但是由于 fMRI 的分析只是显示了海马与二者的交互作用有关，但是并没有明确海马在 2*2 的各个条件下的关系如何，因此我们还是保留了后续的 ROI 的 2*2 方差分析，更为主要的是陈述了简单效应分析的结果。

审稿人 2 意见：

作者很好地解决了一审意见中的问题，并在修改稿中做了一系列补充分析以及论述，建议接收。

回应：感谢专家对论文研究的肯定。