

《心理学报》审稿意见与作者回应

题目：生理周期对情景记忆的影响

作者：李建花，解佳佳，庄锦英

第一轮

审稿人 1 意见：

该论文试图利用“What-Where-When Task”情景记忆任务，结合 ERP 技术，研究女性不同生理周期（卵泡后期和黄体中期）对情景记忆的影响及其机制，结果显示黄体中期对情景记忆中客体的空间位置与时间顺序整合的记忆效果显著好于卵泡后期，且黄体中期额叶脑区的 P300 以及 LPC 波的振幅显著大于卵泡后期，认知控制能力显著增强。该论文有望为理解性激素影响情景记忆的神经基础提供新的证据，但是还有如下问题需进一步完善。

意见 1：摘要：只有结果的总结，缺少结论。

回应：非常感谢您的建议。我们针对您的建议对摘要部分进行了修改与补充。具体内容请参见摘要部分。

意见 2：引言：情景记忆在老年人身上出现了明显退化，以及女性性激素水平随着年龄而改变，从这两个角度引入女性生理周期对情景记忆的影响逻辑性不强。

回应：非常感谢您的建议。根据您的建议，我们删除了原有引言中的部分论述材料及论述视角，重新修改了关于因果关系的陈述和阐释，以保证推论及研究的开展有更强的逻辑性，具体可见引言及讨论部分。

意见 3：讨论：老年人的有关内容对研究主题有干扰。

回应：非常感谢您的建议，老年人的有关内容的确干扰了我们对主题的表达，故而删掉了这一部分的内容。

意见 4：什么是感受性和判断标准？为什么要设计感受性和判断标准这两个因变量，需要在方法中介绍清楚。

回应：非常感谢您的建议，老年人的有关内容的确干扰了我们对主题的表达，故而删掉了这一部分的内容。非常感谢您的建议，我们在实验二的实验方法部分对相应内容进行了补充（详见 3.2.2）。

如下：“信号检测论的主要指标辨别力 (d') 和判断标准 (β)，可以体现个体参与任务时的感受能力、态度和反应倾向。辨别力即为感受性，体现了个体对信息的编码加工及记忆质量 (徐少卿 等, 2014)；判断标准用于衡量心理状态的变化，主要包括动机、态度等 (杨治良, 钟毅平, 1996)。借鉴信号检测论的研究范式，在本研究中，可以进一步区分黄体中期完

成 PO 任务的好成绩，是得益于认知能力（感受性 d' ）的有效提升，还是个体动机、态度等心理状态的影响。”

相关参考文献:

Xu S. Q., Li B., Ma C. Y., & He L. M. (2014). The Influence of 8-Week Mindfulness Training on Unconscious Emotion Processing. *Chinese Journal of Clinical Psychology*, 22(04), 583-588.

[徐少卿, 李波, 马长燕, 贺兰淼. (2014). 八周正念训练对无意识情绪加工的影响. *中国临床心理学杂志*, 22(04), 583-588.]

Yang, Z. L., & Zhong, Y. P. (1996). A Review of Three New Methods of Modern Experimental Psychology. *Journal of Psychological Science*, (01), 44-48.

[杨治良, 钟毅平. (1996). 现代实验心理学三种新方法评述. *心理科学*, (01), 44-48.]

意见 5: 判断标准的结果如何呈现?

回应: 非常感谢您的建议。我们在文中的相应部分增加了一些关于判断标准的解释。

如下: “判断标准用于衡量心理状态的变化，主要包括动机、态度等(杨治良, 钟毅平, 1996)。” (P20);

“上述这些结果为我们的假设提供了证据，即黄体中期完成 PO 任务时记忆成绩的提升得益于认知控制能力的增强；同时，排除了被试对记忆任务的态度、动机等因素的影响。” (P27);

我们查阅了生理周期的相关文献，没有看到基于信号检测论框架的相关研究，且本实验中也并没有看到生理周期两时期判断标准的差异，故而没有将判断标准作为我们的关注的重点，因而没有过多讨论。

相关参考文献:

Yang, Z. L., & Zhong, Y. P. (1996). A Review of Three New Methods of Modern Experimental Psychology. *Journal of Psychological Science*, (01), 44-48.

[杨治良, 钟毅平. (1996). 现代实验心理学三种新方法评述. *心理科学*, (01), 44-48.]

意见 6: 讨论中，“本研究发现的性激素影响情景记忆的特点与规律，可以为教育工作者提供一定的启发。更加有针对性组织教学材料与教学方法，提升学生学习能力。”这一部分如何实现？针对每个个体的性激素水平进行教学材料与教学方法的组织目前看来并不可行。

回应: 非常感谢您的建议，的确很难给出可行性的建议，因此，我们在讨论中删除了这一部分的内容。

意见 7: 文字有误，比如第 10 页， $p = 0.003 < 0.001$ 。

回应: 非常感谢！文字错误已经修改。

审稿人 2 意见:

本研究通过行为和脑电实验考察了女性生理周期对情景记忆的影响，发现黄体中期对情景记忆中客体的空间位置与时间顺序整合的记忆效果更好，并可能与额叶脑区的 P300 和

LPC 波的振幅增强有关。有一些问题如下：

意见 1: P3: 第二个结果是作者的推论，不能作为结果。

回应: 非常感谢您的建议。我们针对您的建议对相应内容进行了修改与补充，具体可见摘要及结语部分。

意见 2: P7: 有 9 名被试先在卵泡后期参加实验，后在黄体中期参加实验，比另一组多 2 人。因被试量较小，有可能由于练习效应导致黄体中期的成绩更好。而且基本上所有任务的结果都有黄体中期好于卵泡后期的趋势。

回应: 非常感谢专家的建议！有以下几方面的证据，可以在一定程度上排除专家提到的练习效应对实验结果的影响：

首先，我们对第一次和第二次参加实验的顺序进行了检验，结果发现，不存在顺序效应（结果见表 1），可以认为不存在练习效应。

其次，我们随机删除 2 名先在卵泡后期参加实验的被试，以此保证第一次参加实验的人数在两个时期相同，结果与原来一致：随机删除两名被试后，在任务 O、P、OO、OP、PO 条件下，对应的 t 值、 p 值、95%置信区间分别为： $t_{(13)} = -1.385, p = 0.189, [-0.078, 0.017]$ ； $t_{(13)} = -0.416, p = 0.684, [-0.190, 0.128]$ ； $t_{(13)} = -1.580, p = 0.138, [-0.302, 0.047]$ ； $t_{(13)} = -0.925, p = 0.372, [-0.211, 0.085]$ ； $t_{(13)} = -3.067, p = 0.009 < 0.01, [-0.417, 0.072]$ 。之后又进行了两次随机删除两名先在卵泡后期参加实验的被试，结果仍然类似。

再次，实验中，每个试次里物体随机、位置随机、顺序随机，两次实验相同的仅为两次看到的物体图片相同，如果说确实有练习效应的话，应该是第一次实验的记忆对第二次实验回忆的干扰，因为该客体的位置、顺序均发生了改变。

但是，您的建议很重要，在之后的记忆相关实验中，确实应该有更多的实验材料，保证第一次与第二次的记忆材料不同但同质，确保完全没有练习效应。

意见 3: 情景记忆属于长时记忆，本研究的记忆阶段只有 21s，随即进入测试阶段，如何能说是长时记忆？

回应: 非常感谢您的建议。根据记忆时间来划分，的确学习与测验阶段的时间间隔比较短，但我们的实验范式参考了前人的研究范式(Kessels, 2007): 同样的在学习阶段呈现 7 个物品，每个呈现 3s，随后立刻进入测验阶段，对相应内容的记忆效果进行检测。最近发表在 *Nature Communication* 的脑电研究也使用了类似的方式：每个刺激呈现 5s，每个记忆序列共 4 个物体，之后进入测验阶段 (Haque, 2020)。

目前对于情景记忆的研究主要有三种：一种研究认为自传体记忆经过复述和加工转化成了情景记忆；一部分研究认为对经验积累的物体属性的回忆效率可以表征情景记忆，如常见物体的形状、颜色等(Zhou et al., 2021)；还有一部分是实验室环境下，通过模拟情景记忆的主要成分探究其发生发展规律，前人使用过的实验范式主要有：序列范式（实验室环境中学习一些项目-背景配对，测验阶段加入新的项目，让被试依次判断是否学过该项目，并回忆项目来源）(Roberts & Blades, 1999)；房子测验范式（在实验室环境下学习、编码客体与背景之间的关系）(Picard et al., 2012)；配对联想回忆范式(Barnes & Underwood, 1959)、三重绑

定任务范式（评估项目与时间、空间以及项目关系的记忆和项目再认记忆的能力）(Lee et al., 2016), 等。实验室情景下探究的情景记忆均为需要在较短时间内记忆编码客体本身、背景信息、时间顺序及它们之间的关系, 本研究范式满足了情景记忆研究的主要特征, 属于对该概念的“操作性定义”。

相关参考文献:

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, 58(2), 97–105.

Haque, R.U., Inati, S.K., Levey, A.I. et al. Feedforward prediction error signals during episodic memory retrieval. *Nature Communication* **11**, 6075 (2020).

Kessels, R. P., Hobbel, D., & Postma, A. (2007). Aging, context memory and binding: a comparison of "what, where and when" in young and older adults. *The International Journal of Neuroscience*, 117(6), 795–810.

Lee, J. K., Wendelken, C., Bunge, S. A., & Ghetti, S. (2016). A Time and Place for Everything: Developmental Differences in the Building Blocks of Episodic Memory. *Child Development*, 87(1), 194–210.

Picard, L., Cousin, S., Guillery-Girard, B., Eustache, F., & Piolino, P. (2012). How do the different components of episodic memory develop? Role of executive functions and short-term feature-binding abilities. *Child Development*, 83(3), 1037–1050.

Roberts K. P., & Blades M. (1999), Children’s memory and source monitoring of real-life and televised events. *Journal of Applied Developmental Psychology*, 20(4), 575–596.

Zhou, W. J., Deng, L. Q., & Ding, J. H. (2021), Neural mechanism underlying the effects of object color on episodic memory. *Acta Psychologica Sinica*, 53(03):229-243.

[周文洁,邓丽群,丁锦红(2021).物体颜色对情景记忆的影响. *心理学报*, 53(03):229-243.]

意见 4: 结果分析应首先列出描述性统计数据, 而不是差异比较。

回应: 非常感谢您的建议。在这一次的修改稿中, 我们在差异比较之前, 增加了描述性统计的表格 (见表 2)。如下:

表 2 生理周期两时期被试在各任务的平均回忆正确率($M \pm SE$)

任务类型	生理周期	
	卵泡后期	黄体中期
O 任务	0.929 ±0.022	0.955 ±0.021
P 任务	0.705 ±0.041	0.723 ±0.072
OO 任务	0.781 ±0.042	0.946 ±0.028
OP 任务	0.388 ±0.061	0.470 ±0.050
PO 任务	0.464 ±0.043	0.714 ±0.064

意见 5: 表 1 中显然并不是 t 值和 p 值。

回应: 非常感谢专家的建议! 经过再次确认, 表 1 罗列的是检验顺序效应是否存在的 t 值和 p 值 (不考虑生理周期, 仅比较前、后两次实验)。

意见 6: P11: “完成其他四类任务时。。。均比记住 PO 任务中两类环境信息的组合更具体和容易。完成 PO 记忆任务，可能需要更强的认知控制能力”。按照这个解释，PO 任务的正确率应小于其他四种任务，但从图上可以看出 PO 任务的正确率至少是大于 OP 任务的。

回应: 非常感谢您的建议。我们先前对 PO 任务的表述可能会引起一些误解，在这次修改中，我们做了一些改进：

“考察上述 5 类记忆任务的特点，完成 PO 任务需要对物品呈现的时间、地点两类环境信息进行整合，而在完成其他四类任务时，要么只需记住一类环境信息，要么只需记住物品与另一类环境信息的结合，均比记住 PO 任务中两类环境信息的组合更具体；或者说，PO 任务包含了更多抽象信息的组合。前人研究发现，老年人随记忆能力的衰退，在情景记忆任务中对客体信息的记忆能力完好，但将多种信息整合起来的能力明显下降，特别是对时间和空间信息的整合（PO 任务）能力下降显著(Chalfonte & Johnson, 1996)。可见，PO 任务的特点与前述 4 类任务有所不同，完成 PO 任务，需要更多的认知控制资源(DuBrow & Davachi, 2016)。”

此外，对于完成 PO 任务和 OP 任务的正确率，我们进行了 2（生理周期：黄体中期 vs. 卵泡后期）× 2（任务：OP vs. PO）的方差分析，结果显示：生理周期主效应显著（ $F_{(1,15)} = 12.521, p = 0.003 < 0.01, \eta_p^2 = 0.455$ ）；任务主效应不显著（ $F_{(1,15)} = 3.342, p = 0.087, \eta_p^2 = 0.182$ ）；二者交互作用不显著（ $F_{(1,15)} = 3.353, p = 0.087, \eta_p^2 = 0.183$ ）。两任务间正确率没有显著差异。

意见 7: O 任务和 OO 任务可能有天花板效应。被试量加大可能得到显著差异的结果。

回应: 非常感谢您的建议。O 任务和 OO 任务的确可能存在天花板效应。但是，“根据 GPower 软件，在统计检验力 $1 - \beta = 0.80, \alpha = 0.05$ 的前提下，进行所有 t 检验需要的被试量小于 5 人。”而在我们的研究中，被试量远远大于 5 人，可以部分排除“被试量加大可能得到显著差异的结果。”可能性。同时，较大的样本会让显著性检验比较敏感，会发现不同条件之间很小的差距，这样的发现可能有统计显著性，但可能没有实际意义上的重要性。

意见 8: 在 O 任务和 OO 任务中都会给被试呈现 23 张干扰图片，而且是同样的 23 张，可能会给被试一些提示，造成正确率偏高。

回应: 非常感谢您的建议。这一部分我们确实缺乏考量，应该选择更多的实验材料而非重复利用。但是实验设计时，考虑到每种实验条件只有一个实验试次，从记忆容量及从众多测验材料选择看到过的物品的再认任务的特点来看，提示效应较小，而且任务顺序被试间随机，也会减弱可能存在的提示效应对实验结果的干扰。但非常感谢专家的建议，后续实验中，我们会选择准备更多的实验材料。

意见 9: P13: “16 名被试同意佩戴 EEG 实验设备完成实验任务，13 名被试仅参加相同任务的行为实验。。。28 名被试的行为数据”。少了一名被试的行为数据？

回应: 非常感谢专家的仔细审读和宝贵建议！确为 12 名被试只参加了行为学实验，16 名被试完成了两次脑电实验。非常感谢，我们已经修改了错误。

意见 10: P16: 感受性指标卵泡后期高于黄体中期，与作者解释的相反。判断标准也是卵泡后期严于黄体中期，可能会导致卵泡后期正确率偏低。

回应: 非常感谢，通过对原始数据再一次整理分析，我们发现文字描述是正确的，但是表 3 数据将黄体中期的感受性（2.14）与卵泡后期的判断标准（1.29）写反了，已经改正了。非常感谢！

意见 11: 如果上述感受性指标的结果有问题，后面行为和脑电相关的结果如何解释？

回应: 参见对第 10 点的答复，非常感谢您。

意见 12: 应详细报告 ERP 数据处理的步骤。

回应: 非常感谢您的建议。结合心理学报类似文章对数据处理部分的汇报形式，我们对实验二数据处理部分进行了补充（详见 3.2.5）。

补充如下:

“使用 eeglab 离线处理 EEG 数据，分析测验阶段刺激呈现的脑电数据。首先降采样为 250Hz，以全脑平均作为参考，之后手动剔除坏段后分别进行 0.1Hz 的低通滤波和 30Hz 的高通滤波，使用独立成分分析（Independent Component Analysis, ICA）探测矫正眼电及其他原因引起的伪迹。根据伪迹校正的结果，剔除波幅在 $\pm 75\mu V$ 范围以外的试次。之后选取测验阶段图片刺激呈现后的 1000ms 作为分析时程，刺激呈现前的 200ms 作为基线，以潜伏期、振幅峰值作为度量指标。最后，使用 ERPlab 插件，平均所有被试不同条件正确试次的 ERPs 波形图，得到 ERPs 波形图。”

第二轮

审稿人 1 意见:

作者很好地回答了我的问题，但是英文摘要未能和中文摘要很好地对应。

意见 1: 16 名女性被试; These findings provide a reference for organizing teaching activities and understanding memory abnormalities.

回应: 非常感谢您的建议，我们对中英文摘要进行了对应检查与修改，并请 native speaker 进行了润色与 proof reading。如下截图：

An Effect of Menstrual Cycle Phase on Episodic Memory

Abstract

Episodic memory refers to the recollection of personally experienced events in a specific context. Evolutionary psychology findings have suggested that female sex hormones may be important influencing factors for episodic memory, but the specific mechanism underlying these influences are unclear. The aim of the present study was to examine whether periodic fluctuations of estrogen and progesterone over the menstrual cycle may influence episodic memory in the What-Where-When Task. Healthy women with stable menstrual cycles and not taking exogenous hormones were recruited for two experiments. In Experiment 1, episodic memory was tested with 16 women during the late follicular phase (FP) and during the mid-luteal phase (LP) with the following five tasks employing images of objects as stimuli: object-only, position-only, object-position binding, object-order binding, and position-order binding. The testing order was counter-balanced across subjects. After a learning phase, participants were asked to recollect elements according to the requirements of each task, and the accuracy rates of their recollection were recorded as dependent variables. The results of Experiment 1 showed that recollection accuracy differed between the late FP and mid-LP for only one of the tasks, namely the position-order binding task. Thus, in Experiment 2, event-related potentials (ERPs) were recorded during the performance of the position-order binding task, while recollection accuracy performance was evaluated. For the position-order binding task in both experiments, a significantly higher response accuracy rate was observed in the mid-LP than in the late FP. Regarding the ERP results, amplitudes of the P300 component and the late positive component (LPC) in frontal cortices, which has associated with cognitive control, were found to be higher in the mid-LP than in the late FP. The present results support the notion that people may have greater cognitive control of episodic memory in the mid-LP than in the late FP. In conclusion, the present results showed that menstrual cycle phase affects position-order binding memory performance and concomitant neural activities. Better episodic memory performance during the mid-LP relative to the late FP could be attributed to better cognitive control ability. This study provides new information on physiological factors that can affect episodic memory.

Ann Power Smith PhD-Write Science Right
删除的内容: ... a fundamental component of cognitive ability... refers to the recollection of personally experienced events occurring... n a specific context. ETheories and research results from e...olutionary psychology findings have suggested that

06189
删除的内容: Sixteen

Ann Power Smith PhD-Write Science Right
删除的内容: "

H

06189
删除的内容: h

Ann Power Smith PhD-Write Science Right
删除的内容: ... no hormones using... with stable menstrual cycles and not taking exogenous hormones were recruited and... or two experiments were carried out... In Experiment 1 (16 women participants)... episodic memory was tested with 16 women during

06189
带格式的

Ann Power Smith PhD-Write Science Right
删除的内容: in the frontal and the LPC... which has associated with been shown representing the higher...ognitive control ability... were found to be

06189
删除的内容: in frontal and lateral parietal cortices

Ann Power Smith PhD-Write Science Right
删除的内容: provided evidence for the hypothesis that individuals... support the notion that people may have higher...reater cognitive control of episodic memoryability... in the mid-LP than in the late FP from episodic memory..."

审稿人 2 意见：作者基本回复了所有意见，仍有一些问题如下。

意见 1：如何根据 GPower 计算出最低被试量是 5 人？建议列出所有条件和参数。另外，文中没有根据期刊要求列出效应量。

回应：非常感谢您的宝贵建议。您的建议促使我们细化了对数据的思考，认为应该单独计算每一类任务的效应量以及所需被试数量，纠正了先前不正确的表述，并在文中补充了效应量的相关信息。具体请见文中的 2.1.1 被试部分：

即“根据 GPower 软件 (Faul, Erdfelder, Lang, & Buchner, 2007)，实验数据收集完毕后，借助后测检验 (Post hoc test)，16 名被试，在统计检验力 $1 - \beta = 0.80$, $\alpha = 0.05$ 的条件下，分别计算 5 类任务的效应量 (Effect Size, ES) 以及样本量，分别如下： $O_{ES} = 0.345$ ，样本量 = 68； $P_{ES} = 0.057$ ，样本量 = 2418； $OO_{ES} = 0.392$ ，样本量 = 54； $OP_{ES} = 0.335$ ，样本量 = 72； $PO_{ES} = 0.900$ ，样本量 = 12。又根据前人相关研究，探究生理周期对认知功能的影响时，其效应量通常大于或等于中等效应量，且所需样本量较小 (Ikarashi, Sato, Iguchi, Baba, & Yamashiro, 2020; Yamazaki & Tamura, 2017)。例如，Ikarashi 等人 (2020) 在探究生理周期对陈述性记忆的影响时，使用被试间实验设计，三组被试 (卵泡期组、排卵期组、黄体期组) 各 10 - 11 名。由此确定，使用 16 名被试探究生理周期对情景记忆各成分的影响是恰当的。”

参考文献：

Ikarashi, K., Sato, D., Iguchi, K., Baba, Y., & Yamashiro, K. (2020). Menstrual Cycle Modulates Motor Learning and Memory Consolidation in Humans. *Brain Sciences*, 10(10), 696.

Yamazaki, M., & Tamura, K. (2017). The menstrual cycle affects recognition of emotional expressions: an event-related potential study. *F1000Research*, 6, 853.

意见 2：表一的差异比较之前也应该先列出描述性统计数据。O 任务的均值差异是负数，为

什么 t 值是正值？

回应：非常感谢您的建议。

(1) 表一中，我们在差异比较之前补充了描述性统计的结果。

(2) 我们对表一的结果进行了重新审查，发现 O 任务的均值为 0.009，之前错误地写成了-0.009，已经进行了修改，由衷的感谢您对文章审查的认真以及对统计知识运用的游刃有余，谢谢您。

(3) 此外，对表一中的内容重新分析时，我们发现，由于之前版本中，表一呈现的是正确回忆个数，而非正确率，95%置信区间里的数据没有进行修改，我们这次也进行了修改。

意见 3: 可以看到，基本上所有的测试第二次成绩都好于第一次，而顺序效应的检验不显著同样有可能是被试量不足引起的。当然，作者随机删除 2 名被试，得到和原来一致的结果，可以部分回答顺序效应的问题。但这里又有一个问题，当按照测试顺序分组的时候，第二次测试成绩都好于第一次；当按照生理周期分组的时候，黄体中期又都好于卵泡后期。对于所有任务都是这样，甚至删除 2 个被试以后也是这样。这看起来是个小概率事件。除非顺序效应和生理周期效应同时存在，且都有不小的效应量？或者，这 5 个任务相互之间有较强的关联，有较为类似的内涵，他们之间差异检验显著性不同的结果，是由于天花板效应或者小被试量引起的？

回应：非常感谢专家的建议。

首先，关于顺序效应的问题，我们还可以提供另外的证据，证明顺序效应存在的可能性较小。在进行正式实验之前，我们还有一个预实验，在预实验中，考虑到可能存在的练习/记忆效应，我们同时招募了同等数量的男性被试（年龄相仿），让他们两次参加实验的人数以及时间间隔与女性被试相匹配，结果，同样没有发现顺序效应。对 16 名男性被试结果的分析如下：

16 名男性被试前后两次回忆正确率配对样本 t 检验

任务类型	第一次实验 ($M \pm SE$)	第二次实验 ($M \pm SE$)	第一次 - 第二次 ($M \pm SD$)	t	自由度	p (双尾)	效应量
O 任务	0.92 \pm 0.032	0.96 \pm 0.016	-0.045 \pm 0.125	-1.431	15	.173	0.360
P 任务	0.629 \pm 0.052	0.63 \pm 0.060	-0.005 \pm 0.216	-0.099	15	.922	0.023
OO 任务	0.91 \pm 0.055	0.78 \pm 0.075	0.134 \pm 0.412	1.299	15	.213	0.325
OP 任务	0.329 \pm 0.047	0.400 \pm 0.056	-0.071 \pm 0.256	-1.117	15	.281	0.278
PO 任务	0.51 \pm 0.068	0.44 \pm 0.078	0.071 \pm 0.281	1.017	15	.325	0.253

其次，对于审稿人提到“这 5 个任务相互之间有较强的关联，有较为类似的内涵，他们之间差异检验显著性不同的结果，是由于天花板效应或者小被试量引起的？”我们就此问题，重新计算了每个任务的效应量，分别为：O = 0.345、P = 0.057、OO = 0.392、OP = 0.335、PO = 0.900，对应的样本量分别为 68，2418，54，72 和 12，可以发现，生理周期在 PO 任务上的效应还是与其他任务显著不同的。

我们注意到，对于 O 任务、OO 任务等，当样本量达到足够大时，也可以显著，何时停

止收集数据是恰当的呢？我们就此问题查阅了更多的文献，发现生理周期影响记忆及其他认知能力的研究，其效应量均比较大，通常大于中等效应量（0.5），所用样本量就比较小。例如：Ikarashi, Sato, Iguchi, Baba, & Yamashiro (2020)，探究生理周期对陈述性记忆的影响，使用被试间实验设计，三组被试（卵泡期组、排卵期组、黄体期组）各 10 - 11 名，探究生理周期对陈述性记忆的影响，其中生理周期的效应量为 0.510；又如 Yamazaki, & Tamura (2017)，是一个使用了 12 名被试的被试内实验设计，要求被试完成情绪表情识别任务，生理周期的效应量为 0.604。此外，还有一些没有报告效应量的研究同样使用了较小的样本但体现了生理周期的效应：比如：《心理学报》2015 年，vol. 47, No. 12，探究女性生理周期对条件性恐惧习得和消退的影响，实验采用被试间设计，选择生理周期两时期各 20 名被试，探究生理周期的效应；又如，Mačiukaitė, Jarutytė, & Rukšėnas (2016)，同样使用较少的被试样本（卵泡期 10 人，黄体期 13 人），使用被试间实验设计。由此认为，根据本研究的目的，使用 16 名被试样本应该是恰当的。

进一步的，正如审稿人所言，可能“这 5 个任务相互之间有较强的关联，有较为类似的内涵”。的确，它们都属于情景记忆，都是情景记忆不同要素之间的组合，完成任务，可能需要类似的共同底层支持规则。在 5 类任务中，均存在黄体中期记忆成绩好于卵泡后期记忆成绩的趋势，可能恰恰表明，黄体中期显著增强的控制能力。

参考文献：

- Ikarashi, K., Sato, D., Iguchi, K., Baba, Y., & Yamashiro, K. (2020). Menstrual Cycle Modulates Motor Learning and Memory Consolidation in Humans. *Brain Sciences*, 10(10), 696.
- JIN Yan; ZHENG Xifu. (2015). Influence of Female Menstrual Cycle on the Acquisition and Extinction of Conditioned Fear. *Acta Psychologica Sinica*, 47(12), 1465-1471.
- [金艳;郑希付. (2015). 女性生理周期对条件性恐惧习得和消退的影响. *心理学报*, 47(12), 1465-1471.]
- Mačiukaitė, L., Jarutytė, L., & Rukšėnas, O. (2017). The effects of menstrual cycle phase on processing of emotional images. *Journal of Psychophysiology*, 31(4), 179-187.
- Yamazaki, M., & Tamura, K. (2017). The menstrual cycle affects recognition of emotional expressions: an event-related potential study. *F1000Research*, 6, 853.

意见 4：“分别进行 0.1Hz 的低通滤波和 30Hz 的高通滤波”：应该是 0.1Hz 至 30Hz 的带通滤波。

回应：非常感谢专家的仔细审读和宝贵建议！我们在原文中进行了修改。

意见 5：总讨论中一些地方将认知能力和认知控制能力混在一起。

回应：非常感谢您的建议。的确存在上述问题，这种混淆主要存在于总讨论的第二段，我们将这一段全部删除了，并检查删除了对应的参考文献。如下：

“我们的研究结果，也得到了单独探讨孕酮激素水平促进认知能力提升的研究结果的支持。例如，以往研究发现，高水平孕酮在 Go/No-Go、Stroop、2-back 任务中增强了被试在注意、知觉、抑制控制能力方面的表现(Hidalgo-Lopez & Pletzer, 2017)。外源性补给孕酮激素后，个体在完成视觉工作记忆任务时有更好的表现(Papanicolaou et al., 2002)。对于绝经后的妇女，孕酮给药有助于其总体认知能力保持在较高水平，一项以 643 名已绝经妇女为被试

的研究发现，其血清孕酮水平与言语记忆和整体认知能力呈正相关(Henderson et al., 2013)。动物研究也显示，孕酮通过作用于海马和前额叶脑区调节大鼠的认知功能(Paris, Walf, & Frye, 2011)，给切除卵巢的大鼠补充孕酮激素会改善其对客体的记忆及识别能力(Stein, 2011)。”

第三轮

审稿人 2 意见：作者基本回复了所有意见，仍有一些问题如下。

意见 1：作者并没有直接回答 GPower 计算被试量的问题。作者所用的后测检验的方法在某种意义上是一种循环论证，在小被试量的情况下得到显著结果，效应量往往会很大，然后又用这个大效应量反证小被试量的合理。更通用的方法是根据以往文献估计效应量。而作者自己举例的两篇文献效应量只有 0.510 和 0.604，只能说大于或等于中等效应量。同时作者自己计算的样本量也大部分远大于 16。

回应：非常感谢您的建议。根据您的建议，我们根据前人研究的效应量，使用 GPower 重新计算了本研究所需样本量，计划样本量为 34 名，由此，我们这次重新增补了被试。由原本的 16 名被试，增加到了 36 名被试（但其中 3 名被试的生理周期在实验后没有保持稳定，被删除），故最终只有 33 名被试。其中 17 名被试先在卵泡后期参加实验，后在黄体中期参加实验，16 名被试先在黄体中期参加实验。在增加了上述被试之后，实验结果看起来更加理想：其他任务的差异不显著，PO 任务的差异更加显著。由增加被试量带来的所有变化，均已修改在原文相应部分。

意见 2：如果是由于被试量小造成了目前的结果，目前结果中 PO 任务的独特性就没有了，文章的立论会有较大变化，特别是对于实验二及最终的结论。是否可用两因素方差分析，并考察交互效应？如果有交互效应，可以一定程度上说明 PO 任务的独特性。

回应：非常感谢您的宝贵建议。我们增加了被试样本之后，实验结果并没有不同，因此，对于 PO 任务因小样本量而导致的独特性的担心可以排除。为谨慎起见，我们根据您的建议，进行了任务（O、P、OO、OP、PO）× 生理周期（卵泡后期、黄体中期）的两因素分析，结果，交互作用显著， $F = 4.16$ ， $p = 0.003$ 。进一步的简单效应分析发现，仅在 PO 任务上，生理周期两时期差异显著，其他均不显著。因为觉得这个两因素分析稍有冗余，就没有放在文章中。

编委意见：同意审稿专家的意见，建议修改后发表。

主编意见：经审阅，[xb21-097]论文已经达到发表水平，同意刊发