

《心理学报》审稿意见与作者回应

题目：常识性知识和语篇语境对代词指认的影响

作者：吴岩，高约飞，赵思敏，王穗苹

第一轮

审稿人 1 意见：该篇研究探讨了常识/语义/背景信息对性别代词加工的影响，通过三个实验，循序渐进，发现了常识的即时影响，背景信息的调节作用。对代词加工的理论问题也进行了系统阐述，研究对当前领域的具有一定的实证贡献，建议做出修改。主要问题如下：

意见 1：引言中“基于这样的研究结果，Cook 等认同 Garrod 和 Terras 所认为的阅读理解中的词汇加工可分为两阶段，但认为语境可以在早期阶段发挥作用”。Cook 的研究如何证明语境在早期阶段发挥作用，通过哪些指标说明。

回应：多谢指正。主要是通过眼动指标中的第一遍阅读时间，也称为凝视时间(gaze duration)和重读时间(second reading time)作为区分早期和晚期加工的指标。Cook 等发现当语境对常识性信息进行更新后，一致和冲突条件下在第一遍阅读时间上不再存在差异。但是，考虑到代词和非代词加工的区别，在修改后的文章中，我们并没有重点介绍该研究。

意见 2：在两个水平的实验中，实验 1 为什么选择 25 个被试，而不是 2 的整数倍。

回应：多谢指正，但有效被试为 22 人，是 2 的整数倍。

意见 3：实验中介绍 40 组实验材料，每个系列 40 句语篇，下文每三篇为一组，组内语篇逐页呈现？这里非常混乱，究竟每个屏幕呈现几个语篇？请将程序做个流程图进行说明。

回应：多谢指正，已经修订了这部分内容，同时增加流程图，见 p7 图 1。

意见 4：实验 1 中，作者用单字代词，导致跳读率偏高，为什么不用双字词，譬如他们/她们。

回应：多谢您的建议。首先，以往关于职业性别倾向对代词加工的影响，均采用单数名词作为先行词，因此考虑到和以往研究的一致性，我们采用了单数名词。其次，复数名词作为先行词时，在代词性别的基础上增加了数量信息，代词数量上的信息也是影响代词理解的重要变量，因此考虑额外变量的干扰，也未采用复数名词作为先行词。最后，如果采用复数名词作为先行词时，语料的编制也比较困难。综上，最终选用单数名词为先行词，代词也为单数。不过，确实如审稿人所述，双字词的略读率会有所降低，后继的研究可以考虑以复数名词为先行词。

意见 5：眼动指标包含不全，应当包含首次注视时间 FFD，指代更早期的眼动数据，此外，应包括 regression in，用回视概率来区分是否具有双阶段的加工。

回应：多谢您的建议，在修改稿中增加了 FFD 的报告，但是并没有增加 regression in 这个指标的结果报告。这是因为，首先，regression in 在三个实验中在两条件间结果不显著或者边缘显著，不能说明太多问题；其次，整篇文章我们集中在注视时间的测量和报告上，重读时间(second reading time)可以很好的反映再次加工的情况。此外，眼动指标的选用主要是参考

以往文献进行，以往相关文献(Cook & Myers, 2004; Duffy & Keir, 2004)多以第一遍阅读时间和重读时间为指标区分早期加工和晚期加工。

意见 6: 所有的数据分析，不应当只说两种条件下显著，而应指明差异的方向，譬如冲突条件下的凝视时间显著长于非冲突条件。

回应: 多谢指正，我们已经修改了这部分内容。

意见 7: 几个小问题：1，摘要中对三个实验应当分别描述，实验 2 出现的事情不应当放在实验 1 中的结论总结中。2，引言中对常识和语义的讲解不够明确，与背景信息的差异说的不够具体，语义与语境很难区分开来。3，引言最后一部分，介绍实验 3 时，应当像实验 12 一样提供简要的例子词汇。4，实验 1 程序介绍中，应当给出汉字的视角大小。5，刻度标示(calibration)、刻度确认(validation)，翻译不精确，尽量对应英文含义，翻译成校准，效度检验。6，“但两个实验中代词的眼动轨迹却并不相同”修改为，“但两个实验中代词的加工模式却并不相同”

回应: 非常感谢您的修改意见，所有这些内容已经修订。摘要已经重写，语境和常识也结合样例重新介绍了，前言中实验 3 已经提供样例，报告了汉字视角，不恰当的描述已经更正。

.....
审稿人 2 意见:

意见 1: T 检验为什么不报告 Cohen's d ? 95%CI 不报告的原因是什么？

回应: 多谢指正。抱歉之前没有细致阅读学报新版要求，在修改中报告了 Cohen's d 和 95%CI，具体见文中结果部分。

意见 2: 依据是什么？样本量计划的理由和做法可参考 <https://osf.io/5awp4/>

回应: 多谢指正。查找相关文献，我们发现计算计划样本量可以从三方面入手：首先，可以通过相关领域的元分析文章，得到先验的效果量，然后计算计划被试量；其次，也可通过比较相关的综述类的文章分析出先验效果量，计算计划被试量；最后，通过既定的大先验效果量，如 Cohen's d 选取 0.8，推知计划样本量。由于前两方面缺乏相关元分析和综述类文章，我们采用第三种办法，依据 Anderson, Kelley, 和 Maxwell (2017); Taylor 和 Muller (1996) 两篇研究，设置先验效果量值为 0.8，采用软件 GPower 3.1 版本 (<http://www.gpower.hhu.de/>) 计算出计划被试量为 24 人。参考文献：

Taylor, D. J., & Muller, K. E. (1996). Bias in linear model power and sample size calculations due to estimating noncentrality. *Communications in Statistics: Theory and Methods*, 25(7), 1595–1610.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: a method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, 28(11), 1547–1562.

意见 3: 有综合结论吗？还是就这样孤立着？

回应: 多谢指正，已经修改。

意见 4: 语境与语义，能完全分开吗？

回应: 多谢指正，确实，语境信息也是语义的一种，文中没有表述清楚，已经结合样例介绍常识和语境的区别，希望这次可以表述清楚。

意见 5: 前面没有综述其他词汇吧?

回应: 多谢指正, 前言内容已经修订。

意见 6: 怎么发现语境产生即时性的影响呢?

回应: 多谢指正, 即时性影响主要以眼动指标“首次阅读时间”和“凝视时间”的结果为依据。实验 2 结果发现有第一个代词所形成的语境在第二个代词加工中促进了代词的理解, 使冲突条件下的凝视时间在两条件下差异不显著, 说明更新后的语境产生了作用, 否则冲突条件下的阅读时间应该更长。实验 3 的结果同样说明语境的早期作用, 因为首次注视时间和凝视时间在冲突和一致条件下不再存在差异。在文中讨论部分我们已经添加了这部分内容, 具体见讨论部分 p16、p17 红色字体内容。

意见 7: 都一致的话, 本研究的创新及贡献是什么?

回应: 多谢指正, 已经修订此种说法。此外, 值得一提的是, 实验 1 更多地是验证性的实验, 确定职业性别倾向性可以对代词加工产生影响。实验 2 和实验 3 是重点, 考察文本语境和读者的常识性知识对代词加工的影响。

意见 8: 实验 3 中的前几句话, 已经把重新定义了代词, 此时还是语境了吗?

回应: 多谢指正, 确实在一些语料中, 实验 3 的前面几句话确实已经重新定义了职业名称的性别角色, 如说纺织工为父亲等, 但是无论如何都是语言材料中所提供的内容, 需要读者将性别描述(父亲)和职业名称整合一起(纺织工), 才能从语料中获得职业名称的性别信息: 例如, 纺织工是一名男性。那么这种整合后的内容就是语篇语境。这样的设置在一定程度上我们也是参照以往文献进行, 如 Duffy 和 Keir(2004)的研究, 研究者也采用“Electrician is a cautious woman”这样的内容定义职业名称的性别。

第二轮

审稿人 1 意见: 作者已经认真的进行了修改, 达到了学报要求的水平, 建议修后发表。

审稿人 2 意见:

意见 1: 第一遍阅读时间不再存在差异, 就证明语境在早期阶段发挥作用吗?

回应: 抱歉一审回应中没有回答清楚该问题。以下是对一审回应的补充: 采用眼动技术, Cook 和 Myers(2004)通过职业与行为的冲突考察常识和语境对词汇加工的作用。研究者给被试呈现如下材料:

“At a rock concert, teenage fans cry as a sad song is played by the band’s guitarist/ manager.”
在没有语境的情况下, “manager”的阅读时间比“guitarist”的阅读时间长, 因为根据读者的常识, “manager”不太可能在摇滚音乐会上唱歌。不过, 当文本前面有信息对这种不合理的现象进行解释, 使这种行为合理化; 或者是第二次出现不合理的词汇“manager”时, 一致条件(guitarist)和冲突条件(manager)相比, 两个目标词在第一遍阅读时间(gaze)上没有差别, 但冲突条件下目标词的重读时间更长。基于这样的研究结果, Cook 和 Myers 认为语言材料中所提供的语境信息在词汇加工的早期阶段发挥作用。这是因为, 当没有语境信息对 manager 的行为进行解释时, 或者第一次出现 manager 时, 读者很难理解 manager 在音乐会上唱歌这种行为。但当语言材料中提供信息对这种不合理现象进行解释时, 或者第二次出现不合理词

汇 manager 时，读者注意到了文本中所提供的信息，对 manager 唱歌这种行为进行了重新加工，从而促进了解释后的或再次出现的 manager 的早期加工，因此“manager”的首次阅读时间和“guitarist”不再存在差异了。

意见 2：难道选 25 个被试，就为了有 22 个有效被试？

回应：多谢指正，一审回答确实没有考虑周全。因为在收集数据过程中，我们会初步观察被试的眼动漂移情况和回答问题情况，发现个别被试回答问题的正确率过低，就会相应地增加了某个版本的被试数量，以保证有效数据量在两个版本中大致平衡。所以最终收集 25 名被试数据，而有效被试 22 人。

意见 3：14 年前的文献(Cook & Myers, 2004; Duffy & Keir, 2004)，怎么能证明现在还是这样呢？

回应：多谢指正，一审回应中的两篇研究：Cook 和 Myers, 2004；Duffy 和 Keir, 2004 是与本研究非常相关的早期研究，因此，一审回应中以这两篇研究为例进行例证。事实上，在眼动研究中，包括近年来的一些文章中，通常以阅读时间 FFD、gaze 反映早期加工过程，以 second reading time 和 go path reading time 等反映再次加工过程(e.g., Esaulova, Reali, & Stockhausen, 2014; Kliegl, Grabner, Rolfs, & Engbert, 2004; Luo, Wu, & Jiao, 2018; Qiu, Chen, & Wang, 2012)。因此，本研究所选择的眼动指标，符合以往研究惯例。

第三轮

编委意见：这篇稿子经过几轮修改，已经解决了审稿人的意见，达到了学报发表的标准。所以，建议发表。

第四轮

主编意见：本研究采用眼动技术，通过三个实验，探讨了在中文阅读中，读者的常识性知识和语篇语境在代词理解过程中的作用，结果发现常识性知识的及时性影响作用和语篇语境的调节作用。经过几轮修稿，文章在行文质量和规范上均有所提高，行文中作者对代词加工的相关理论问题也进行了系统详实的阐述，理论基础扎实，且研究工作对代词加工领域的研究具有一定的实证贡献。