

《心理学报》审稿意见与作者回应

题目：强弱语义语境下的否定句加工机制

作者：陈广耀 何先友 刘涛

第一轮

审稿人 1 意见：

意见 1：文章采用眼动技术对强弱语义语境下否定句加工进行研究。建议在摘要中具体指出研究要解决什么样的问题以及为什么要解决这一问题，而不是笼统地说要探讨否定句加工的机制。

回应：感谢审稿人的中肯建议，已经在摘要里做相应修改。

“2 个眼动实验分别探讨强弱语义语境下的否定句加工机制，以便考察语言理解是命题表征的过程还是经验模拟的过程，抑或是二种兼具。实验中，被试听句子（强语义语境：如加工“胳膊不是弯曲的”的备择选项为“伸直的胳膊”，其中“弯曲-伸直”为具有反义关系的强语义连接关系；弱语义语境：如加工“裙子不是蓝色的”的备择选项为“黑色的裙子”，“蓝色-黑色”为弱语义连接关系），并按键选择与句子描述匹配的图片。结果发现：强语义语境条件下，被试在早期（201~600ms 时窗）对描述事件否定状态图片（弯曲的胳膊）与备择图片（伸直的胳膊）的注视概率没有差别，晚期（601ms 后）仅对备择图片的注视概率更高；弱语义语境条件下，被试在早期（401~600ms）仅对描述事件否定状态图片（蓝裙子）注视概率更高，晚期（801ms 后）仅对备择图片（黑色裙子）注视概率更高；并且在晚期都对描述事件否定状态图片的注视概率低于随机水平。结果支持符号依存假设和抑制假设。”

意见 2：文章中多次提到陈广耀等人（2014a, b）的研究，显然两项研究与本论文研究内容直接相关，但是在整个引言中，没有对这两项研究进行具体介绍，不清楚这两项研究到底做了什么，为什么当前研究还在继续做。

回应：已对原文的写作做出相应调整。本研究与陈广耀等人（2014a,b）的研究有以下几个方面的联系。首先，陈广耀等人的研究是探究类别型状态不确定否定句的加工机制，而与本研究探究强弱语义联接否定句加工机制，研究目的都是为了考察否定句加工机制。其次，陈广耀这两篇研究材料相对比较单一，研究材料只是类别型状态不确定否定句，而类别型状态不确定否定句可以归属于本研究的弱语义联接否定句。从这个角度来看，本研究考察的研究材料更加全面。第三，本研究借用了陈广耀等人（2014b）采用的眼动追踪技术和实验材料设置、数据处理方法等；第四，本研究对陈广耀等人（2014a）所提出的锚激活与限制满足模型进行了深化和修订。这些也一一在文章中进行了阐述，再次感谢审稿专家的建议。

意见 3：实验 1 和 2 的研究目的与引言最后一段在很大程度上非常相似，建议不要简单重复。

回应：感谢审稿人建议，我们在原文中删除了重复的部分。

意见 4：实验 1 和 2 分别考察强弱语义语境下的否定句加工，为什么不在一个实验中操纵语境强弱，然后进行直接比较，而是分开两个实验进行？

回应：感谢审稿专家的中肯意见。在实验初期也考虑将实验 1 和实验 2 融合在一个实验进行，

但是考虑到一个实验的 trial 数就比较多，整个实验大约要持续 20 分钟左右，考虑到如果将两个实验融合在一起，被试的任务量就会翻倍，更容易产生焦躁情绪，从而影响到实验的精确性。此外，由于实验用句型相对比较单一，为了防止练习效应的出现，也将两个实验分开做了。再次感谢审稿专家的建议。

意见 5: 实验 1 要求被试通过按键选择与句子描述相匹配的图片，例如“胳膊不是弯曲的”，根据图 1 的图片材料，没有与之匹配的，在这种情况下被试该如何作答。

回应: 在实验 1 中，所有句子都可以在图片中找到合适的匹配图片。正如您上面所提到的“胳膊不是弯曲的”表示胳膊是伸直的，胳膊不是弯曲的就是伸直的，因此被试可以选择伸直的胳膊作为匹配选项。

.....

审稿人 2 意见:

该研究采用眼动方法，考察了不同语境中否定句的加工机制，研究具有一定的理论意义。以下问题请作者思考或修改。

意见 1: 前言中介绍的研究逻辑显得突兀。为什么文中列出的数据模式可以为相应的假设提供证据？请结合已有研究结果或理论推理加以说明。

回应: 感谢审稿人的中肯建议，已在原文中的写作方面做出相应的调整。

意见 2: 两个实验的“实验目的”内容与前言重复。建议直接去掉或从理论层面加以说明。

回应: 感谢审稿人建议，我们在原文中删除了重复的部分。

意见 3: 请对两个实验的被试特点进行必要的介绍。

回应: 谢谢审稿人的建议，我们已经在原文中做出相应调整。被试要求如下：汉语为母语的大学生自愿报名参与实验，每名被试视听正常，智力精神正常，不戴框架眼镜。

意见 4: 数据分析时是否考虑了被试的按键反应时间长短，如果是，请作者在实验设计部分补充说明。

回应: 谢谢审稿人的建议，已在原文中的写作做相应调整。在本研究中，被试的反应时数据属于次要数据，眼动数据是主要数据。在有关否定的研究中，反应时数据通常被用于考察否定句加工困难的指标，在这类研究中通常期望加工否定句比相应的肯定句更加困难，进而需要更长的反应时。

意见 5: 补充说明数据的剔除标准。

回应: 已在原文中的写作做相应调整。本研究剔除的数据包括两个部分，一种是错误数据；一种是反应时超过 3 个标准差的数据。感谢审稿专家的建议。

意见 6: 请补充说明兴趣区是如何划分的、注视概率分析的是哪个眼动指标、以及采用这一分析方式的理由。

回应: 本研究在兴趣区的划分、眼动指标的选取方面借鉴了陈广耀等人（2014）的做法。在兴趣区的划分方面，“图片材料均为白色背景、黑色实物画。每屏呈现 4 张 256×192 像素的小图片。具体排列位置如图 1，首先将整个屏幕平均划分为 8×8=64 个区域，左上角图片（道

路) 占据第2行和第3行的第2和第3个方格; 右上角图片(自行车)处在同左上角图片基于纵向中线对称的位置; 同样下面两张图片处在与上面两张图片基于横向中线对称的位置。为了避免出现位置效应, 我们对四类图片的位置采用拉丁方平衡。在进行眼动数据分析时, 我们将每个图片所占据的4个小方格区域作为兴趣区。”

在眼动数据的指标选择方面, 我们只用了不同兴趣区内的注视点数据。从本质上讲, 眼动数据就是由一系列的注视点和眼跳组成。通过眼动仪我们可以记录到每个注视点在屏幕上的落点位置, 以及起止的时刻。据此信息我们可以做出4条注视曲线, 进一步, 可以计算出被试在每个时间窗中, 对不同类型图片的注视概率, 进而画出注视概率曲线图。感谢审稿专家的建议。

意见 7: 讨论部分说服力不强, 尤其是讨论中的第三、第四、第五自然段更为突出。如作者在第三自然段中提出“本研究除了支持符号依存假设的正确性, 并且还修订了陈广耀等人(2014)基于综合理论模型而提出的锚激活与限制满足模型”, 然而该段内容着重介绍了已有模型观点, 以及对这些观点的理解, 且脱离了实验结果进行讨论; 第四自然段内容松散; 第五自然段有一定的可取之处, 即能够思考研究的不足, 然而涉及的内容过于宽泛, 未能反映研究中存在的根本问题。

回应: 感谢审稿人的中肯建议, 从审稿专家的意见能够感觉到审稿专家在写作方面的深厚功底, 我们也从审稿专家的建议中学到了许多有关论文写作方面的知识和心得。我们在原文中按照审稿专家的建议重新对讨论部分进行修订, 再次感谢审稿专家的不吝赐教与斧正。

其它几个问题:

意见 8: 文中存在大量的表述问题, 如“601ms 时刻后便仅仅对述事件实际状态的 T 图片注视概率更高”, “数据处理前首先删除剔除数据”, “同时, 本实验顺便探讨了相应肯定句的加工以防实验句式过于单一”等等。建议检查全文, 并做相应修改。

回应: 感谢审稿专家的中肯建议, 已经对文章做通篇检查, 并一一修正。

意见 9: 缩小表一和表 2 的字号。

回应: 感谢审稿专家, 我们已经在原文中做出修正。

意见 10: 关键词中应包括“语境”一词。

回应: 感谢审稿专家, 已在原文中增添“语境”作为关键词。

.....

审稿人 3 意见:

主要问题:

意见 1: 作者在前言中阐述了有关否定句加工的理论, 但从中不能清楚地了解到为什么会有这样的理论争论。建议作者增加具体的实验证据来说明各个理论争论的原因与争论焦点。在此基础上进一步阐述本研究与已有研究的区别及其独特贡献。

回应: 感谢审稿人的中肯建议, 我们已经对前言部分的写作做出调整, 详见前言部分。

意见 2: 实验 1 的 2.1 部分与前言最后一段的内容重复。

回应：已经在文章中做出调整，感谢审稿人的建议。

意见 3：实验 1 中包括了肯定句与否定句两类句子，同时两类句子都与图片存在四种水平的匹配关系，这应该是一个 2*2 的实验设计。建议作者按 2*2 的设计来分析数据。

回应：谢谢审稿人的建议。研究中肯定句和否定句确实与四类图片都存在 4 种水平的匹配关系。只是肯定句与四类图片的匹配关系为：与事件实际状态匹配图片/事件备择图片/具备事件实际状态的控制图片/单纯的控制图片；而相应的否定句与四类图片的匹配关系为：与事件否定状态匹配的图片/与事件实际状态匹配的图片/具有事件否定状态的控制图片/单纯的控制图片。他们的匹配关系是不同的，所以不能采用 2*2 的实验设计，只能采用单因素被试内设计。

意见 4：实验 1 中“在实验中，句子通过语音播放 1200ms（句子“胳膊是/不是”部分播放时间）后在屏幕上呈现 4 幅图片”，被试已经预先知道后面要呈现的客体目标，这样就不会去关注无关的图片。

回应：谢谢审稿人的建议。本研究的研究目的是考察强弱语义语境下的否定句的加工机制，具体考察在不同时刻被试对不同图片的注视概率差异。在早期时刻，重点关注被试是否同时表征事件否定状态与实际状态，也就是看被试在早期对 N 图片与 T 图片的注视概率之间是否存在差异；晚期重点考察理解者是否会对早期表征的事件否定状态产生抑制，即考察 N 图片与无关图片之间是否存在差异，是否对 N 图片的注视概率降至基线水平（25%）以下。如加工否定句“胳膊不是弯曲的”，早期重点考察两个主语图片（N 图片：“弯曲的胳膊”和 T 图片：“伸直的胳膊”）之间的差异；在晚期重点考察描述事件否定状态图片（弯曲的胳膊）与控制图片（自行车和道路）之间的差异。审稿人说的也有一定道理，如果被试已经预先知道后面要呈现的客体目标，就应该不会再去关注其他无关图片了。然而，正如前面所述，早期关注的重点在 N 图片（弯曲的胳膊）与 T 图片（伸直的胳膊）之间的差异，与控制图片无关，两张胳膊图片都是预先知道的事物“胳膊”，这样就不存在预先知道或者不知道的差异；在晚期，重点关注 N 图片与无关图片的对比，如果预先知道了接下来要选择什么物体，就继续关注什么物体图片的话，理应获得的结果模式是被试在晚期一直保持对 N 图片的注视概率高于无关图片。但是，两个实验的结果都显示被试在较晚的时刻会表现出对 N 图片的注视概率低于随机水平（25%），并且随后与无关图片之间没有差异，这样的结果恰恰更加有力的证明了抑制假设的正确性。

意见 5：1200ms 的播放时间是如何确定的？在否定句和肯定句下的播放时间是相同的？

回应：1200ms 时刻的划分是根据语音材料本身的性质设定的。我们通过音频软件 Cool Edit 分析发现所有的音频材料都可以切分为“主语”、“谓语”和“宾语”三个部分，如音频材料“胳膊是/不是弯曲的”，其中“胳膊”、“是/不是”和“弯曲的”之间都存在几十到上百毫秒的空白间隔，发音“是”与“不是”之间的差异并不是很明显。此外，90%以上的语音材料都可以从 1200ms 之间切分开来，也就是说 1200ms 刚好可以落到空白间隔中。剩余的一些没有落入的音频就可以通过在句首增加或者删除静音的方式使 1200ms 时刻刚好落入到空白间隔内。这样就可以将肯定句和否定句的播放时间控制一致。感谢审稿人的建议。

意见 6：实验 1 为强语义否定句，实验 2 为弱语义否定句，强弱语义否定句中主语的生命性是否会影响注视时间和注视模式？同时呈现了四幅图片，其中有两幅与句子语义有关，而另外两幅一幅与句子中的描述词有关，另一幅则是一个具体的客体。图片的各项指标（熟悉度、复杂度等）是否会影响注视模式？

回应：感谢审稿人的建议。详细的实验材料我们已经附上，其中实验 1 正式材料中总共有 23 个实物，其中有生命的只有 2 个；实验 2 正式材料有 45 个实物，其中有生命的有 4 个。从上面数据来看，有生命的主语所占比例非常低，对整个实验的影响应该可以忽略不计。审稿人的建议非常专业和有深度，未来我们也会专门设计实验，将主语生命性作为一个实验因素来验证该问题。此外，为了尽量降低图片熟悉度和复杂度等方面的影响，在实验开始前，我们要求被试逐一学习实验中出现的实物图片及其名称，测试其熟练掌握图片及其命名后才进行正式实验。

意见 7：图 2 和图 3 的模式中是否包含了部分被试已经做出按键反应后的眼动注视情况？如果有，建议剔除这一部分数据后重新比较。

回应：在本研究中只采用了按键反应之前的眼动数据，按键后被试的眼动数据没有参与分析。谢谢审稿专家的专业建议。

意见 8：表 1 表示的是否定句的结果，表 1 的说明需要明确这一点。

回应：已经把表 1 和表 2 的表头都调整为“否定句条件下四类图片在不同时窗的注视概率(%) 差异比较”，感谢您的建议。

意见 9：实验 1 讨论中“当加工强语义联结否定句时，结果和两步模拟假设与符号相互依存假设的预测一致”，按照两步模拟假设，在否定句的加工中，应该是首先模拟否定状态，再模拟肯定状态，那应该出现的结果是先观看 N 图片，再观看 T 图片，但作者得到的结果并非如此。第二，符号相互依存理论的主要观点是什么？结果如何支持了这一假设？按照前文中所描述的观点，单纯的强语义联结情况下的比较并不能验证这一理论的观点。

回应：感谢审稿人的建议。实验 1 讨论部分在写作方面确实出现了一些纰漏，我们已经重新写作。具体如下：“加工强语义连接否定句的结果显示，被试在早期对 N 图片和 T 图片的注视概率没有差异，该结果说明语义连接强度起了作用，被试在强语义连接的辅助下快速地模拟了事件的实际状态。早期的数据模式符合命题理论和符号相互依存理论的预期，而两步模拟假设的预期不一致。晚期数据显示，被试从 601ms 开始便对 T 图片的注视概率高于其他三类图片的注视概率，并且对 N 图片的注视概率从 1001ms 开始低于随机概率。晚期的数据模式与两步模拟假设和符号相互依存假设的预测一致，即被试在表征了事件实际状态后，对事件否定状态的模拟已经无用，会被抛弃，支持抑制假设。综合早期和晚期的数据显示，加工强语义连接否定句符合符号相互依存假设的预期。为了进一步考察弱语义连接结果离散型否定句的加工机制。我们设计了实验 2。”

符号相互依存假设的主要内容为：“假设认为语言理解是命题表征与知觉模拟共同作用的结果，二者貌似相互排斥，实际相互依存 (Louwerse, 2010; 陈广耀等人, 2014a)。该假设认为语言可以编码整个世界，大脑中的知识由语言符号及其指代物组成，并且语言符号与其指代物之间、语言符号与语言符号之间和指代物与指代物之间都具有连接关系；这样就形成了符号-指代物连接层、符号-符号连接层和指代物-指代物连接层。不同连接类型之间和同种连接内部的连接强度存在差异，如符号-符号连接层中反义词对（伸直-弯曲）之间的连接强度就高于非反义词对（蓝色-黑色）。在解释语言加工方面，符号相互依存假设的支持者认为符号加工和知觉模拟在否定句理解中都起作用，只不过二者会因为不同连接层的连接强弱所起作用的大小不同而已。当加工的否定句基于语境可以获得事件的结果状态时，理解者就会进行模拟加工；如果无法获得确定的结果状态，则会进行命题符号加工。”

在解释否定加工机制方面，符号相互依存假设的支持者认为命题符号加工和知觉模拟在否定句理解中都起作用，只不过二者在否定加工的不同阶段所起作用的大小不同而已。正如

审稿专家所讲，单单一个实验的结果无法得出本研究的结论。本研究的实验结果需要综合实验 1 和实验 2 的结果才能支持符号相互依存假设。在否定加工的早期，加工强弱语义连接否定句由于语义连接强度的差异导致了不同的加工模式出现，即强语义连接否定句在加工早期便模拟了事件的实际状态，而加工弱语义连接否定句失去强语义连接的辅助没有在这个阶段模拟事件的实际状态，早期的结果差异证明的命题符号加工的正确性。晚期的结果支持了经验模拟加工的存在，正如两个实验的实验结果显示，被试在晚期都会只保持对事件实际状态的模拟，抑制对事件否定状态的模拟。再次感谢审稿专家的建议。

意见 10: 为什么在实验中使用注视概率这一指标？注视概率的差异表明了个体差异的存在，作者是如何考虑个体差异的？

回应: 谢谢审稿人的建议。我们选择注视概率这一指标有如下几个方面的考虑。首先，不管我们选择什么指标来处理数据都会存在个体差异；其次，我们可以不采用注视概率，也可以采用注视点个数这一指标，只是使用注视点个数的话，个体之间的差异有可能会更大，比如，在某一时窗，有些被试的搜索效率较高，可能 4 个注视点就完成了搜索，对四幅图片各注视 1 次，这样每个图片的注视概率就为 25%；但是，有些被试的搜索效率较低，共用了 12 个注视点完成搜索，同样对每个图片上的注视 3 次，每个图片的注视概率同样为 25%。这样就可以把个体差异降至最低。但是如果采用原始的注视点个数作为指标的话，两个被试之间的差异将会拉大许多。此外，从研究目的考虑，本研究主要想考察被试在不同时间窗对四类图片之间的相对注视差异程度，而不是考察图片之间的绝对差异，这样的话，注视概率是相对更合适的指标。基于以上几点的考虑，在本研究采用注视概率这一指标。考虑到个体间差异，我们觉得任何实验研究都会存在个体差异，并且方差分析的原理也是拿组间差异来与个体间差异做比较，如果组间差异（实验处理效应）相对于个体差异来讲足够大，就说明不同实验处理水平之间存在显著性差异。从本研究的实验结果来看，如果在不同时窗，几类图片之间存在显著性差异，这就说明在该时窗实验处理的效应远远大于个体间的差异。

意见 11: 实验 2 因为其研究的相似性存在类似的与上述实验 1 相同的问题。

回应: 已经根据实验 1 的问题，对实验 2 存在的类似问题做出修正。

意见 12: 实验 1 和 2 强弱语义下的数据是否可以联合起来进行分析和讨论？同时根据对前言的修改意见，结合研究数据对两个实验的结果进行综合讨论。

回应: 审稿专家的建议非常好，我们认真接受审稿专家的建议。正如审稿专家所说，只有将实验 1 和 2 的数据综合起来分析，才能得出实验结果支持符号相互依存假设。再次感谢审稿专家的专业建议。我们在实验 2 讨论后增加了一个段落综合讨论实验 1 和 2 的结果：“综合实验 1 和实验 2 的研究结果分析，在否定句加工的早期，加工强语义连接否定句，被试对 N 图片和 T 图片的注视概率没有差异，说明被试在强语义连接的辅助下表征事件否定状态的同时就表征了事件的实际状态。加工弱语义连接否定句，被试对 N 图片的注视概率高于对 T 图片的注视概率，说明失去强语义连接的辅助，被试通达事件实际状态的时间就要延后。综合早期阶段的数据结果，证明了否定句加工具有命题符号加工的存在。在否定句加工的晚期，被试对 T 图片的注视概率都高于 N 图片，并且 N 图片在大约 1001ms 阶段都会低于随机概率，该结果证明了被试在否定句理解的晚期会议仅仅模拟事件的实际状态，因为对事件否定状态的模拟产生抑制作用。两个实验晚期的结果证明了否定句加工经验模拟假设的正确性。综合两个实验前期后期的数据，我们可以得出否定句的加工是命题符号加工与经验模拟加工两种加工共同作用的结果，结果支持符号相互依存假设。”。

意见 13：次要问题：作者需要在附录中呈现实验 1 和 2 中所用到的全部实验材料。

回应：：已经将相关实验材料附上。

附录 1：实验一实验材料：

序号	否定句	与事件否定状态匹配的 N 图片	与事件实际状态匹配的 T 图片
1	胳膊不是伸直的	伸直的胳膊	弯曲的胳膊
2	胳膊不是弯曲的	弯曲的胳膊	伸直的胳膊
3	双手不是分开的	分开的双手	合上的双手
4	双手不是合上的	合上的双手	分开的双手
5	被褥不是凌乱的	凌乱的被褥	整洁的被褥
6	被褥不是整洁的	整洁的被褥	凌乱的被褥
7	天平不是倾斜的	倾斜的天平	平衡的天平
8	天平不是平衡的	平衡的天平	倾斜的天平
9	房门不是打开的	打开的房门	关上的房门
10	房门不是关上的	关上的房门	打开的房门
11	T 恤不是干净的	干净的 T 恤	脏了的 T 恤
12	T 恤不是脏了的	脏了的 T 恤	干净的 T 恤
13	抽屉不是关上的	关上的抽屉	拉开的抽屉
14	抽屉不是拉开的	拉开的抽屉	关上的抽屉
15	书本不是翻开的	翻开的书本	合上的书本
16	书本不是合上的	合上的书本	翻开的书本
17	嘴巴不是闭上的	闭上的嘴巴	张开的嘴巴
18	嘴巴不是张开的	张开的嘴巴	闭上的嘴巴
19	轮胎不是瘪的	瘪了的轮胎	鼓鼓的轮胎
20	轮胎不是鼓的	鼓鼓的轮胎	瘪了的轮胎
21	盘子不是干净的	干净的盘子	脏乱的盘子
22	盘子不是脏乱的	脏乱的盘子	干净的盘子
23	皮带不是解开的	解开的皮带	扣上的皮带
24	皮带不是扣上的	扣上的皮带	解开的皮带
25	手铐不是解开的	解开的手铐	锁上的手铐
26	手铐不是锁上的	锁上的手铐	解开的手铐
27	衬衣不是平整的	平整的衬衣	褶皱的衬衣
28	衬衣不是褶皱的	褶皱的衬衣	平整的衬衣
29	圆规不是张开的	张开的圆规	合上的圆规
30	圆规不是合上的	合上的圆规	张开的圆规
31	水龙头不是打开的	打开的水龙头	关上的水龙头
32	水龙头不是关上的	关上的水龙头	打开的水龙头
33	毯子不是卷上的	卷上的毯子	展开的毯子
34	毯子不是展开的	展开的毯子	卷上的毯子
35	报纸不是打开的	打开的报纸	折叠的报纸
36	报纸不是折叠的	折叠的报纸	打开的报纸

37	钢琴盖不是打开的	打开的钢琴盖	盖上的钢琴盖
38	钢琴盖不是盖上的	盖上的钢琴盖	打开的钢琴盖
39	折尺不是打开的	打开的折尺	合着的折尺
40	折尺不是合着的	合着的折尺	打开的折尺
41	纸扇不是打开的	打开的纸扇	合上的纸扇
42	纸扇不是合上的	合上的纸扇	打开的纸扇
43	自行车不是倒下的	倒下的自行车	立起的自行车
44	自行车不是立起的	立起的自行车	倒下的自行车
45	蜡烛不是点燃的	点燃的蜡烛	熄灭的蜡烛
46	蜡烛不是熄灭的	熄灭的蜡烛	点燃的蜡烛

附件二：实验 2 实验材料

序号	否定句	与事件否定状态匹配的 N 图片	与事件实际状态匹配的 T 图片
1	衬衣不是条纹的	条纹衬衣	纯色衬衣
2	蛋糕不是巧克力的	巧克力蛋糕	奶油蛋糕
3	道路不是 T 型的	T 型道路	十字型道路
4	地板不是瓷砖的	瓷砖地板	木质地板
5	豆子不是黄色的	黄色豆子	绿色豆子
6	耳环不是条状的	条状耳环	方形耳环
7	房子不是城堡式	城堡式房子	普通平房
8	饭盒不是塑料的	塑料饭盒	不锈钢饭盒
9	盒子不是圆形的	圆形盒子	心形盒子
10	婚礼不是韩式的	韩式婚礼	中式婚礼
11	奖杯不是水晶的	水晶奖杯	金质奖杯
12	轿车不是奥迪的	奥迪轿车	宝马轿车
13	麻将不是条子的	条子麻将	饼子麻将
14	帽子不是鸭舌状的	鸭舌状帽子	圆形帽子
15	牛奶不是瓶装的	瓶装牛奶	袋装牛奶
16	女孩不是哭泣的	哭泣的女孩	微笑的女孩
17	扑克不是红心的	红心的扑克	黑桃的扑克
18	桥不是拱形的	拱形桥	立交桥
19	裙子不是蓝色的	蓝色裙子	黑色裙子
20	人脸不是侧面的	侧面人脸	正面人脸
21	沙发不是木质的	木质的沙发	布质的沙发
22	手机不是翻盖的	翻盖手机	直板手机
23	水杯不是陶瓷的	陶瓷水杯	不锈钢水杯
24	水龙头不是抬起式	抬起式水龙头	旋转式水龙头
25	树叶不是针状的	针状树叶	掌型树叶
26	梳子不是羊角的	羊角梳子	木质梳子
27	四边形不是梯形的	梯形四边形	菱形四边形
28	T 恤不是白色的	白色 T 恤	红色 T 恤
29	统计图不是条形图	条形图	饼形图

30	土豆不是丝状的	土豆丝	整个土豆
31	拖鞋不是人字形的	人字拖	棉拖鞋
32	图形不是三角形	三角形	菱形
33	文章不是英文的	英文文章	中文文章
34	舞蹈不是芭蕾的	芭蕾舞	街舞
35	项链不是珍珠的	珍珠项链	黄金项链
36	学生不是小学生的	小学生	博士生
37	眼镜不是半框的	半框眼镜	全框眼镜
38	硬币不是一元的	一元硬币	五角硬币
39	椅子不是四脚的	四脚椅子	八脚椅子
40	泳姿不是蛙泳的	蛙泳泳姿	仰泳泳姿
41	月饼不是五仁的	五仁月饼	蛋黄月饼
42	运动鞋不是耐克的	耐克运动鞋	361 运动鞋
43	运动员不是跑着的	跑着的运动员	跳高的运动员
44	枕头不是竹制的	竹制枕头	棉制枕头
45	纸币不是人民币	人民币纸币	美元纸币

第二轮

审稿人 1 意见：

意见：作者已经针对审稿意见进行了认真修改，文章的质量大大提高。但文章中很多句子特别复杂，难以理解。例如：“代表第一代认知理论的命题理论和基于第二代认知理论具身认知理论提出的经验模拟假设一直被用来解释否定句的加工机制”、“在研究方法上，我们在本研究的 2 个眼动实验中采用陈广耀等人应用的能够进行实时测量的眼动追踪技术”、“数据处理前首先剔除错误数据和各条件下超过 3 个标准差的实验数据”、等等。请作者通读一遍文章，同时让其他同学读一读，划出难以理解的句子，并一一进行修改。另外，英文摘要存在语法错误和中式表达，请再找专业人士帮忙修改。

回应：非常感谢审稿专家的意见。我们找了不同层次的读者对文章进行通读，找出了难以理解的句子，并做出相应的调整。另外，也请了英文专业人士对英文摘要进行修改。再次感谢审稿专家对本研究所提供的专业建议。

审稿人 2 意见：

作者已对文章进行了认真合理的修改，建议发表。

审稿人 3 意见：

意见 1：作者在前言中阐述了有关否定句加工的理论，但从中不能清楚地了解到为什么会有这样的理论争论。建议作者增加具体的实验证据来说明各个理论争论的原因与争论焦点。在此基础上进一步阐述本研究与已有研究的区别及其独特贡献。

答：感谢审稿人的中肯建议，我们已经对前言部分的写作做出调整，详见前言部分。

审稿人回复：对于这一点，作者增加了一些阐述，但仍然局限于理论观点的阐述，而没有增加相应的实验证据来说明理论争论的原因，建议作者增加具体的实验证据说明各个理论争论的原因，并在此基础上进一步阐述本研究与已有研究的区别及其独特贡献。

回应：感谢审稿专家的建议，我们已经在前言部分增加了具体的实验证据来说明各个理论争论的原因与争论焦点。详见前言描红部分。

意见 2：实验 1 中包括了肯定句与否定句两类句子，同时两类句子都与图片存在四种水平的匹配关系，这应该是一个 2*2 的实验设计。建议作者按 2*2 的设计来分析数据。

答：谢谢审稿人的建议。研究中肯定句和否定句确实与四类图片都存在 4 种水平的匹配关系。只是肯定句与四类图片的匹配关系为：与事件实际状态匹配图片/事件备择图片/具备事件实际状态的控制图片/单纯的控制图片；而相应的否定句与四类图片的匹配关系为：与事件否定状态匹配的图片/与事件实际状态匹配的图片/具有事件否定状态的控制图片/单纯的控制图片。他们的匹配关系是不同的，所以不能采用 2*2 的实验设计，只能采用单因素被试内设计。

审稿人回复：对于否定句和肯定句两种类型的句子来说，所匹配的四种图片都包括：1) 与描述客体相同且性质相同的图片；2) 与描述客体相同但性质不同的图片；3) 与描述客体不同但性质相同的图片；4) 与描述客体不同且性质不同的图片。因此，这应该是一个 2*4 的实验设计，作者应该按照 2*4 的实验设计来分析数据。

回应：感谢审稿专家针对实验设计提出的新描述，这种描述从理论上讲也是可以说通的，只是本研究不适合采用该设计。具体原因如下：首先，从研究目的来看，本研究是为了考察强弱语义语境下否定句的加工机制，重点关注在不同的时间阶段，被试对描述事件否定状态的 N 图片与描述事件实际状态的 T 图片之间的注视概率差异。在否定句条件下，我们已经设计了四类匹配关系，包括基线条件（C 图片），不同匹配条件之间可以进行相互比较，无需与肯定句进行交叉分析。其次，我们在实验中确实增加了相应的肯定句，只是增加肯定句的目的主要有两个：一个是为了防止实验句型过于单一，排除练习效应；另一个是为了能够对两种句型的反应时进行比较，进而验证否定句更难加工，需要更长反应时的观点。第三，句子与图片的匹配关系也决定本研究只能采用单因素设计。否定句存在“否定状态”条件，实验中与 N 图片匹配；但是对于肯定句来讲，四类图片都不能与句子形成“否定状态”的匹配关系，因为肯定句只有“实际状态”，却没有“否定状态”，也就不能与否定句一起构成 2*4 的实验设计。第四，为了与前人的研究表述保持一致，本研究更适合采用原先的实验设计表述。如在否定句条件，N 图片与句子描述的事件否定状态匹配、T 图片与句子描述的事件实际状态匹配等，这些匹配关系和描述方式已经在该领域不同研究者之间形成共识（Kaup & Zwaan, 2003; Kaup, Lütke, & Zwaan, 2006; Zwaan, Stanfield, & Yaxley, 2002; Giora, 2006; Giora & Fein, 2007; 高志华, 鲁忠义, 马红霞, 2011; 崔如霞, 高志华, 唐艺琳, 何皓璠, 鲁忠义, 2016)。如果在本研究中改变相应的匹配关系，会割裂本研究与该领域其他研究的承继关系。此外，为了探讨否定句的加工机制，前人已经采用同样的实验设计和数据分析方法来研究该问题，如果本研究失去了注视概率曲线的结果，不但丧失了眼动技术的优势，研究结果也无法与前人研究结果进行分析比较，进而失去研究的延续性。最后，在数据分析方面也决定了本研究只能采用单因素被试内设计。退一步讲，假如按照 2*4 实验设计来分析数据，强行处理出来的结果也会存在以下两个方面的问题：一方面，本研究的主要数据是注视概率曲线，

否定句条件获得的注视概率曲线能否与肯定句条件获得注视概率进行交叉分析仍需慎重考虑。即使处理出来了结果，想必得出的结果也会显得异常复杂，科学性存疑。这样的话，不但不利于研究问题的解决，还会让读者产生更多的迷惑。综上所述，虽然审稿专家提的新表述在理论上也能说得通，只是本研究更适合采用原来的实验设计方式。

我们也一直在思索什么原因让审稿专家产生了疑惑，深入分析发现，原因可能是我们在本研究中采用了不恰当的表述方式。我们在肯定句条件使用“N 图片”和“NC 图片”作为图片的简称，而“N 图片”和“NC 图片”的表述同样出现在否定句条件，这样就很容易让审稿专家误以为肯定句与否定句同四类图片之间的匹配关系是一样的。由于我们的表述不当让审稿专家产生误解，在此我们向审稿专家表示歉意。我们在修改稿中将该类型图片简称为“A 图片”（alternative picture）和“AC 图片”（alternative control picture）以作区分。再次感谢审稿专家的建议。

意见 3: 实验 1 为强语义否定句，实验 2 为弱语义否定句，强弱语义否定句中主语的生命性是否会影响注视时间和注视模式？同时呈现了四幅图片，其中有两幅与句子语义有关，而另外两幅一幅与句子中的描述词有关，另一幅则是一个具体的客体。图片的各项指标（熟悉度、复杂度等）是否会影响注视模式？

答：感谢审稿人的建议。详细的实验材料我们已经附上，其中实验 1 正式材料中总共有 23 个实物，其中有生命的只有 2 个；实验 2 正式材料有 45 个实物，其中有生命的有 4 个。从上面数据来看，有生命的主语所占比例非常低，对整个实验的影响应该可以忽略不计。审稿人的建议非常专业和有深度，未来我们也会专门设计实验，将主语生命性作为一个实验因素来验证该问题。此外，为了尽量降低图片熟悉度和复杂度等方面的影响，在实验开始前，我们要求被试逐一学习实验中出现的实物图片及其名称，测试其熟练掌握图片及其命名后才进行正式实验。

审稿人回复：

1) 删除有生命的客体后，实验结果与原来的分析结果相同吗？ 2) 图片的各项指标肯定会影响被试早期的知觉模式，作者需要匹配 N 图片、T 图片，NC 图片和 C 图片的各类指标，包括命名一致性、视觉复杂度、熟悉度等各项指标。3) 前期的学习图片名称并不会使得这些图片的各类指标匹配。建议作者增加后期的实验补充这些数据。

回应：感谢审稿专家的意见。从实验结果来看，物体的生命性并不会影响实验结果。如上面所述，两个实验中有生命的实物占比极低（实验 1 中只有 2 个试次，实验 2 中仅有 4 个试次）。如果删除这 2 个试次就能导致整个实验结果发生改变，那么说明这 2 个试次的的数据属于极端数据，但是极端数据又会在数据分析前就被剔除；也就是说，这 2 个试次的的数据并不是极端数据（假如没被剔除的话），即使现在剔除这 2 个试次的的数据也不会改变现有结果。再者，对于本研究来讲，实物的生命性也并不是非常重要的无关变量，保留这 2 个试次的的数据反而可以提升实验结果的外部效度。

针对后两个问题，我们非常赞同审稿专家看法，我们也认为图片各项指标，如命名一致性、视觉复杂度和熟悉度等，作为无关的随机变量会在某种程度上影响被试的早期知觉模式。但是从实验结果来看，这些随机误差并没有影响到本实验的结果。在数据的早期阶段（200ms 以内）几条曲线处在非常好的重叠状态，也就是说，这些随机误差是会被抵消掉的。进一步讲，图片的这些特性并不影响我们重点关注的的数据。我们重点考察的是，在否定句条件下，被试对 N 图片与 T 图片之间的注视概率是否存在显著差异。而 N 图片（弯曲的胳膊）与 T

图片（伸直的胳膊）描述的是同一物体，只是状态不同而已，如一条是伸直的胳膊，另一条是弯曲的胳膊。“两条胳膊”在生命性、视觉复杂度和熟悉度方面都是完全匹配的。从这个角度理解的话，审稿专家提到的这些无关变量反而都做到了非常严格的控制。不过审稿专家提出的建议非常值得思考，我们也接受审稿专家的建议，在未来的研究中，会专门将这些因素作为研究变量来考察。再次感谢审稿专家的意见。

意见 4: 实验 1 和 2 强弱语义下的数据是否可以联合起来进行分析和讨论？同时根据对前言的修改意见，结合研究数据对两个实验的结果进行综合讨论。

审稿人回复：作者只是进行了讨论，而未对实验 1 和 2 的数据联合进行分析，需要增加实验 1 和 2 的联合分析结果。

回应：谢谢审稿专家的意见。正如在方法部分给出的回复，单独在一个实验中就不太适合将肯定句和否定句的数据做交叉分析；在此也就不太适合进一步将两个实验的数据来进行联合分析。但是审稿专家的意见很有帮助，在讨论部分需要将实验 1 和实验 2 的数据结果综合来进行讨论，我们也在讨论部分增加了联合讨论。再次感谢审稿专家的意见。

意见 5: 次要问题：有些句子不通顺，如“早期的数据模式符合命题理论和符号相互依存理论的预期，而两步模拟假设的预期不一致”。

回应：我们进行通读论文，已经做了相应调整，再次感谢审稿专家的建议。