

《心理学报》审稿意见与作者回应

题目：路径转弯对语篇空间情境模型建构的影响

作者：陈洁彬 鲁忠义

第一轮

审稿人 1 意见：

意见 1：中文摘要偏长。

回应：根据专家意见进行了删减。

意见 2：实验 1 的范式是学习探测+回指解决，从实验材料上看不出需要对关键词（如指纹）进行回指解决，因为即使不学习地图，目标句在语义上也是通顺的，在 Rinck, Gordon, & Bower (1995) 和 Morrow, Bower, & Greenspan (1989) 的研究中，因变量是探测词反应时间。

回应：这条意见包含实验范式和因变量两个意思，分别作出回答。（1）关于实验范式。在实验 1 中，正式实验开始前的学图阶段，要求被试能够记住每个房间的特定物品，如果不学习地图的话，被试无法确定关键词到底指代那个房间，所以必须经历学图阶段。且该范式中，目标句包含的一个词语指代某个房间，如“指纹”指代“前台”。如果不充分学习布局图，被试只读了语义通顺的目标句“忽然反应过来自己忘了按指纹”，那么，被试是无法确定是不是回指到了前台这个房间。这样转弯效应就无法得以证明了。（2）关于因变量。本实验使用了同前人实验不同的因变量，是基于这样的考虑：前人使用探测词做因变量反映了对回指的目标词的直接激活情况，也是不错的，但有中断阅读的问题存在。我们的研究用的是目标句的阅读时间，这种因变量不仅反映了阅读中的整合，同时由于目标句中有回指词，也反映了激活，再加上它不中断阅读，因此我们认为这是更合适的一种因变量。

意见 3：实验 1 和实验 2 是材料间设计，目标句是不同的，对目标句的阅读时间不能准确反映不同条件下被试对关键词的搜索过程，用关键词的阅读时间更好一些。但是对于材料间设计，无论是目标句还是单个关键词，都应该进行预实验，看看这些目标句或者关键词在语篇之外的阅读时间是否无显著差异。

回应：关于因变量上面做了回答，我们还是认为阅读时间作为因变量可能更合适一些，或者至少说也是可以的。

对目标句我们做了较为严格的控制，在每个实验的各种实验条件下，目标句的字数和词的数量都做了控制。实验 1 中，目标句的字数均为 13 个字；实验 2 中，目标句字数均为 11 个字；实验 3 中，目标句字数均为 11 个字。

除了对目标句的字数控制外，单独的句子阅读时间也没有差异，预实验的结果是：所有阅读理解任务的判断正确率达到 100%，说明实验句简单易懂。实验 1 句子阅读时间均值 $M=1980.74\text{ms}$ ，标准差 $SD=379.71\text{ms}$ ；实验 2 句子阅读时间均值 $M=2021.33\text{ms}$ ，标准差 $SD=574.31\text{ms}$ ；实验 3 句子阅读时间均值 $M=1818.93\text{ms}$ ，标准差 $SD=176.44\text{ms}$ 。对每组进行重复测量方差分析，实验 1： $F(7, 203)=0.786$ ， $p=0.673$ ；实验 2： $F(9, 261)=0.491$ ， $p=0.880$ ；实验 3： $F(3, 87)=0.435$ ， $p=0.728$ 。说明在语篇之外，单独对这些句子的阅读时间是没有差异的。这样就排除了句子本身对实验的影响。根据以上结果，可以认为编制的实验材料是合理的。

意见 4：2.1、3.1 和 4.1 是关于研究目的和研究假设的，其实只是研究目的的陈述，没有研究假设的表述。研究假设的表述一般是“如果……，那么……”。

回应：根据专家意见重新将实验假设进行了表述，分别见三个实验的研究目的和研究假设部分。：

意见 5：表格不规范。例如表 1，合适的表述应该为：

表 1 目标句阅读时间的平均数和标准差 (ms)

路径转弯次数	目标句阅读时间	
	<i>M</i>	<i>SD</i>
0	2480	792
1	2717	1033
2	3018	1128
3	3067	1319

回应：根据专家意见规范了所有图表。

意见 6：表述要清楚、简洁。例如，表 2 已经呈现了两两比较的结果，文字叙述时只需要“即当路径出现转弯时，目标句的阅读时间显著长于没有转弯的路径，且随转弯次数的增加阅读时间递增，一直到转弯次数增加到 3”即可。

回应：根据专家意见精简了实验结果的表述。

意见 7：建议对实验 2 的数据进一步分析：看看相同转弯数、不同直线距离的情况下，能否验证实验 3 的结果。

回应：实验 2 的目的在于探讨纯语篇自然阅读范式下，路径转弯对其空间距离表征是否存在影响，为此，控制了当前地点与回指地点之间的路径距离和直线距离，即材料中转弯次数相同的路径其直线距离也相等。这样，专家建议的相同转弯数、不同直线距离的情况是不存在的，所以没有办法从这个角度来验证实验 3 的结果。之所以做实验 3，是由于实验 2 中路径转弯次数达到 3 时目标句阅读时间反而缩短，为探讨出现这一现象的原因，我们才进行了实验 3。

意见 8：实验 3 使用了眼追踪技术，其中以回指词为信息区的总注视时间是该技术的优势。然而，文中没有说明直线距离长和短时回指词的注视时间出现差异所表达的意义。另外，眼动轨迹如何得出需较详细说明。

回应：(1)关于总注视时间。总注视时间 (total time on a word)，包括凝视时间和回扫后花费在该单词上的注视时间。它反映了包含首次和回扫后多次的阅读加工过程。直线距离长，读者从当前地点到回指地点的加工回指词时间就长，即总注视时间就长；相反，直线距离短，总注视时间也就短。对回指词的总注视时间能很好地反应回指解决所需要的时间。(2)关于眼动轨迹。根据专家意见做了进一步的详细说明。增加的部分复制在这里，以便审稿专家阅读。

在实验 3 的判断任务阶段，目标句出现之前屏幕上会出现一个手型图作为注视点，标志目标句即将出现的位置以保证被试能够迅速准确的阅读目标句；每张判断图片呈现之前也会出现一个手型注视点，手型注视点的位置与绿色小人位置重合，即是主人公当前的位置，也就是被试进行逆向搜索的起点，被试只有在盯住注视点的同时按键才会出现判断图片。由于之前被试的目光落在注视点，所以此时被试在寻找目的地时必然是从这个位置进行逆向搜索。如果被试是以路线图的形式进行搜索，则眼动的轨迹应该是沿着走过的实际路线进行的；如果被试是以空间搭桥的方式进行搜索的，那么被试的眼动轨迹则应该是直线距离的，即空间搭桥方式的。

意见 9: 数据筛选采用了删除每个水平上目标句反应时超过 ± 2 个标准差的标准, 以该标准筛选数据, 被删除的数据较多。建议采用至少 ± 2.5 个标准差的标准再核算一下实验结果, 看是否能得到一致的结论。

回应: 首先, 专家担心删除 ± 2 个标准差以外的数据会删的较多, 这里将数据删除做一详细说明: 实验 1 共回收 47 人的有效数据 (阅读理解正确率高于 75%), 剔除每个水平上目标句反应时离散程度 ± 2 个标准差以外的数据 2 个, 数据筛除量占总回收量的 0.04%; 实验 2 共回收 58 人的有效数据 (阅读理解正确率高于 75%), 剔除每个水平上目标句反应时离散程度 ± 2 个标准差以外的数据 3 个, 数据筛除量占总回收量的 0.05%; 实验 3 共回收 47 人的有效数据 (阅读理解正确率高于 75%), 单个被试同一水平上两个版本语篇目标句总注视时间取平均值, 剔除每个水平上目标句反应时超过 ± 2 个标准差的项目共 2 个, 筛除数据量占总回收量的 0.04%。可以认为, 按照超过 ± 2 个标准差的标准删除数据, 删除掉数据量不大。按 2.5 个标准差以外的数据进行删除, 数据删除量与 ± 2 个标准差相差小于 0.01%, 且统计检验的结果与删除 2 个标准差相同。

另外, 一般说来, 删除 2 个到 3 个标准差之外的数据都是可以的, 从国内外发表的研究报告看也是如此。例如, 下列文献就是删除的 2 个标准差:

Yaxley & Zwaan. Simulating visibility during language Comprehension. *Cognition*, 2007.105

Richter & Zwaan. Integration of perceptual information in word access. *The Quarterly Journal of Experimental Psychology*, 2010.1

张璇, 杨玉芳. 工作记忆广度对语篇理解中重读效应的影响 *心理学报*, 2010.7

王瑞明, 莫雷等. 文本阅读中协调性整合的发生机制 *心理学报*, 2008.12

意见 10: 文中有几处句子无标点或者不太通顺, 请认真检查。

回应: 对全文做了细致校对, 更正了语法错误。

审稿人 2 意见:

意见 1: 实验选用自编的语篇作为实验材料, 每个条件下只有两篇文章, 也就是说实验的项目数非常少。这种情况下, 如何排除取样偏差的影响呢? 建议作者增加项目数, 不仅进行被试检验, 还要报告项目检验的结果。

回应: 每个实验条件下用了两篇文章, 是根据预实验确定的。在预实验中, 每个条件下有 4 篇文章, 但被试做完实验反映说, 读的文章太多, 每篇文章由于很多转弯的描述, 结果记不清楚或混淆。考虑到被试阅读中的注意力集中时间和承受能力, 确定每个实验条件下选两篇文章。

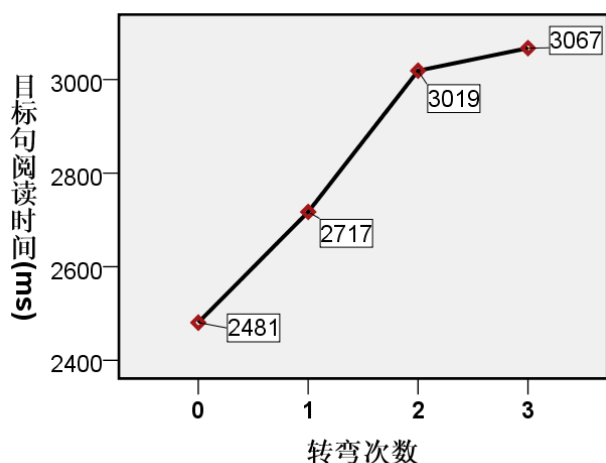
虽然实验用的材料偏少, 但通过效果量的测定, 发现实验效果还是不错的。三个实验的效果量分别是 $\eta^2=0.57$; $\eta^2=0.63$; $d_{\text{回指词}}=0.71$ 、 $d_{\text{目标句}}=0.54$ 、 $d_{\text{判断}}=0.47$ 。

补充了以项目为随机变量的方差分析 F_2 。

意见 2: 三个实验的关系看似递进的, 然而每个实验间改变的变量或情景太多, 使得逻辑关系变得模糊。例如实验 2 不仅采用了自然情景的阅读任务, 也增加了自变量的水平数。而实验 3 中作者只做了复杂文本阅读的情况, 发现存在搭桥现象, 而没有做简单文本的对照实验, 这也使得结果的说服力不足。

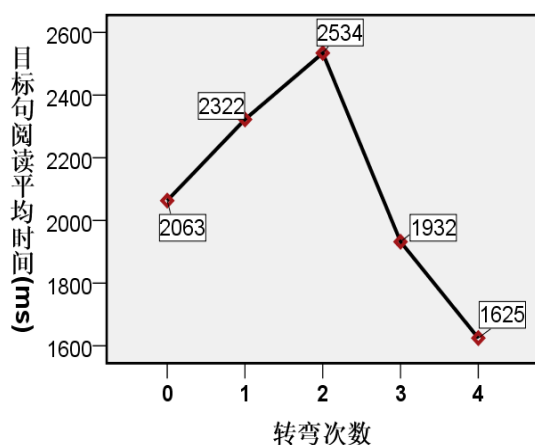
回应: 本研究的三个实验逻辑如下: 实验 1 采用经典范式尝试探讨语篇理解中是否存在转弯效应, 即在有布局图配合文字材料的学习的情况下, 研究阅读中的路径转弯效应。结果发现, 当转弯次数从 0 增加到 2 时出现路径转弯效应, 即目标句的阅读时间随转弯次数的增加而增加, 但是当转弯次

数增加到 3 时，目标句的阅读时间没有显著增长，如下图所示。



实验 1 是一个尝试性的实验，是为实验 2 打基础的，主要是想在此基础上研究纯语篇阅读中的路径转弯效应。实验 1 发现阅读中不仅有路径转弯效应，而且转弯次数的阈限似乎为 2。我们还发现转弯次数为 2 和 3 时，目标句的阅读时间是没有差异的，那么转弯次数继续增加是仍然没有差异，即 2 个转弯次数以后目标句的阅读时间是一个平台呢，还是会出现其他情况呢。为此，在实验 2 中还增加了一个转弯次数。

实验 2 的结果发现，当转弯次数增加到 3 个时，不仅发现了同实验 1 同样的研究结果，即转弯效应的最大转弯次数为 2，而且还发现 2 个转弯次数以后，目标句的阅读时间不是一个平台，而是一个倒转的“V”形，如下图所示。



实验 1 和实验 2 的结果表明，语篇的空间描述简单（转弯次数 <2 ）时，目标句的阅读时间会随着转弯次数的增多而增长，即回指解决方式是原路返回。既然转弯次数 <2 的问题已经解决，所以实验 3 只研究语篇中的转弯次数 >2 时的情况，即目标句的阅读时间为什么不随着转弯次数的增加而增加，反而减少？因此实验 3 只选择了语篇空间描述复杂的材料。

意见 3: 此外，写作中存在一些概念的混淆，例如摘要中：“采用回指解决方法，通过 3 个实验探讨了路径转弯对空间距离表征的影响机制”，文章大部分都在讨论“空间情境模型”，这些概念不能存在混淆。

回应: 统一用了“空间情境模型”。

意见 4: 行文中存在一些不通顺的地方, 希望作者能仔细检查, 如第六页中“...表面上无差异, 差异在于所路径由出发房间到到达房间经过的距离不等...”。

回应: 对全文重新做了细致校对, 更正了语法及表达错误。

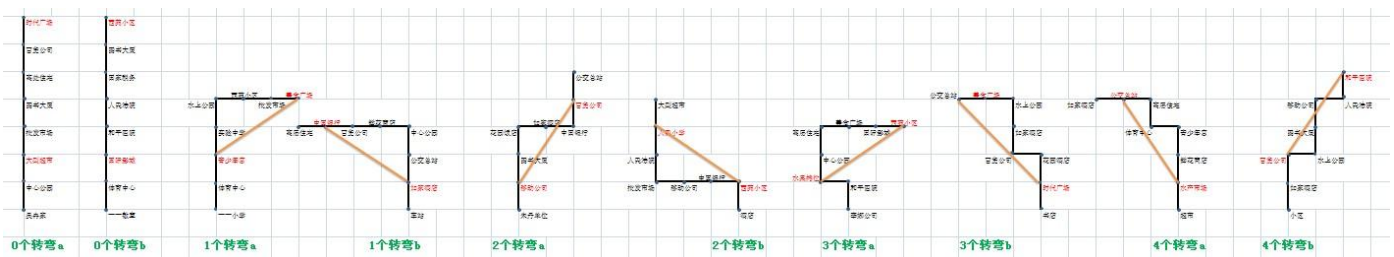
第二轮

审稿人 1 意见:

意见 1: 实验 1 和实验 2 的自由度不对。两个实验都是单因素被试内设计, 实验 1 有效被试 47 人, 误差自由度应该是 $(4-1) \times (47-1) = 138$; 实验 2 有效被试 58 人, 误差自由度应该是 $(5-1) \times (58-1) = 228$; 如果只是笔误, 改过来就可以了, 如果不是笔误, 有可能是方差分析的命令没有用对。

回应: 本条意见是由原文表达不当造成, 已在文中进行修改。“共回收 47 个有效数据 (阅读理解正确率高于 75%), 剔除每个水平上目标句反应时离散程度 ± 2 个标准差以外的数据共 2 个”收回的 47 个有效数据中, 还根据标准差删除了 2 个, 也就是剩数据 45 个, 因此自由度为 $(4-1) \times (45-1) = 132$; 实验 2 同实验 1, “共回收有效 58 个有效数据 (阅读理解正确率高于 75%), 剔除每个水平上目标句反应时离散程度 ± 2 个标准差以外的数据 3 个”, 剩余数据为 55 个, 自由度为 $(5-1) \times (55-1) = 216$ 。

意见 2: 实验 2 的材料设置有问题, 数据结果可能不能说明语篇阅读中随着路径转弯次数的增加, 目标句的阅读时间也增加。在所有的 10 篇描述中, 目标句之前的那个句子 (目标前句) 都同时出现了起始地点和主人公当前所在的地点, 如在 2 个转弯的语篇中, “不一会儿朱丹从单位 (起始地点) 来到百货公司 (当前地点)”, 也就是说, 在阅读这一句时, 被试为了理解语篇, 可能已经激活自己曾经的位置“单位”和单位邻近的“移动公司”, 当阅读目标句“她想起忘了买手机充值卡”时, “移动公司”已经在工作记忆中, 回指词解决不是真的从当前地点到回指地点, 而是从起始地点 (单位) 到回指地点 (移动公司), 其间的距离是 1 个单位 (见图 4); 而在 4 个转弯的语篇中, 其间的距离是 2 个单位。这样一来, 目标句的阅读时间的差异很可能不是由于转弯效应引起的, 而是距离效应引起的。



回应: 上图列出了实验 2 所有文字材料所描述的路径图, 实验材料的设置已经平衡了起始地点与回指地点之间的距离, 而且, 如果如专家猜测, 目标句阅读时间的差异是由距离效应引起的, 那么, 同个水平间材料 a 和材料 b 之间就应该存在显著差异, 但是经检测, 这种差异并不存在, 因此可以排除这种可能。

进而推测被试在进行回指解决时进行的是逆向搜索, 即从当前位置进行回溯。这符合空间情境模型建构的规律, “每一个进入读者工作记忆的信息如果与当前模型存在不一致均会导致当前情境模型的更新以整合新的变化”, 同时, 被试的工作记忆中内容也伴随阅读进行了更新, 到达了当前位置。

意见 3: 实验 3 的因变量之一是以回指词为信息区的总注视时间, 作者认为“直线距离长, 读者从当前地点到回指地点的加工回指词时间就长, 即总注视时间就长”, 直线距离长, 可以推断读者判断从当前地点能否到回指地点的加工时间就长, 这个路径中直线距离覆盖的地点都有可能增加加工时间,

而不能推断只是加工回指词的时间长。

回应：实验 3 的因变量之一是以回指词为信息区的总注视时间，作者认为“直线距离长，读者从当前地点到回指地点的加工回指词时间就长，即总注视时间就长”，直线距离长，可以推断读者判断从当前地点能否到回指地点的加工时间就长，这个路径中直线距离覆盖的地点都有可能增加加工时间，而不能推断只是加工回指词的时间长。

对专家的问题可以从两个方面做出解释：（1）实验三自变量为直线距离，不管直线距离是长还是短，直线距离中只覆盖一个地点，这一点我们做了控制，所以专家认为的那种情况在实验 3 中不存在；（2）回指词有整合全句的作用，其加工时间（总注视时间）的长短能很好地反映出路径距离的长短，就是说当回指到起点时，如同实际走过这一段路一样，距离长花费的时间就长，反之，则短。这与行为实验可能是一样的。实验 2 的结果证明了专家的观点，当路线简单时，被试对回指词的总注视时间反而比路线复杂时长，从而可以推测可能是回指解决的方式不同而造成的。正式实验 2 和实验 3 联合证明的观点：路线简单时，被试沿着路线逆向搜索，对路径所经过的所有地点进行加工；而路线复杂时则是利用空间搭桥方式搜索，此时搜索路径为直线，只覆盖一个地点，所以加工时间更短。

意见 4：关于眼动轨迹，作者认为“如果被试是以路线图的形式进行搜索，则眼动的轨迹应该是沿着走过的实际路线进行的；如果被试是以空间搭桥的方式进行搜索的，那么被试的眼动轨迹则应该是直线距离的，即空间搭桥方式的”，逻辑上是行得通的，但是实验程序是文章描述与地图是同步呈现的，而且标示了清晰的路径，让被试在地图上判断红色旗子位置是否是主人公的目的地，被试不需要建构空间情境模型，直接进行视觉搜索就可以完成，眼动轨迹不能说明在空间情境模型中被试以空间搭桥的方式进行搜索，只能说明在实际地图上被试对从一个点到另一个点的搜索遵循几何学“两点之间，直线最短”的原理。而且，对眼动轨迹如何得出也没有给出详细说明，所作的说明还是实验程序，而不是数据获得，因为实际的眼动轨迹应该是围绕直线上波动的注视点的集合。

回应：专家意见包括被试是否需要构建情境模型和搜索模式两层意思，分别回答：

（1）专家认为实验程序中文章描述同时已经同步呈现路线图，那么被试在进行任务时不需要建构空间情境模型。实验 3 程序阅读的同时同步呈现路线图是为了给被试限定路线，以便对被试的反应进行眼动监控。并且被试在阅读时和进行判断时所呈现的路线图是不一样的，被试在进行阅读任务时，所呈现的路线图中绿色小人是从起点开始一步一步标出的，且只标注当前地点；而在进行判断任务时，路线图中绿色小人变为从终点（即当前地点）返回。也就是说，阅读时呈现的是“去”的路线，判断时呈现的是“回”的路线。可以认为实验中只是规定了路线图的形状，但并未给出地点名称，所以此时被试仍需自行建构空间情境模型以明确路线中各个地点的名称，否则无法完成判断任务。

（2）专家认为眼动轨迹不能说明在空间情境模型中被试以空间搭桥的方式进行搜索，只能说明在实际地图上被试对从一个点到另一个点的搜索遵循几何学“两点之间，直线最短”的原理。实际两者的意思是一致的，当空间描述复杂时由于认知资源限制读者很难建立精细的路线型空间情境模型，而是将其概化为地图型的俯视图，并寻找最短路径来搜索目标地点，在眼动轨迹上表现为寻找两点之间最短距离，即空间搭桥，这也符合认知资源有限的原理；但是当空间描述简单时，被试的眼动轨迹则不再遵循两点之间直线最短的原理，表现为顺着路线逆向搜索，具体采取何种搜索方式由路线的复杂程度决定。

审稿人 2 意见：

意见 1：实验 2 中“3.2.1 研究工具和被试同实验 1”，被试同实验 1 是什么意思？是做过实验 1 的被试吗？

回应：实验 1 和实验 2 是用同一批被试。

意见 2: 实验 2 中 6 篇填充材料与实验材料字数句式完全一致, 实验材料和填充材料表面无差异。那么, 设置填充材料的用意是什么?

回应: 表面无差异, 但是填充材料中没有转弯。

意见 3: 实验 3 选了 50 人做被试, 实验结果与分析中说“共回收 57 人的有效数据”, t 检验的自由度是 44。

回应: 此处为疏忽错误, 已改为“共回收 47 人的有效数据”, 又剔除每个水平上目标句反应时超过 ± 2 个标准差的项目共 2 个, 因此剩余数据位 45, 所以自由度为 44。

意见 4: 实验 3 图片判断反应时, 只计算正确反应还是正确反应和错误反应一起计?

回应: 只计算做出正确判断的反应时。