

# 《心理学报》审稿意见与作者回应

题目：笔画节点在手写体汉字识别中的作用

作者：朱一鸣;赵阳;唐宁;周吉帆;沈模卫

## 第一轮

审稿人 1 意见：

本研究提出节点数量效应和节点复杂度效应，并采用两个行为实验证实两个效应，具有一定创新性，但是存在如下问题：

意见 1：文中前言第一段有些多余，与下文联系不紧密。

回应：非常感谢审稿专家的意见。原文第一段指出了探究手写字识别机制的应用背景，意在表明本研究除理论意义之外也具备较强的应用价值。为了加强与下文的联系，现已在该段强调探究汉字内笔画、部件等各层次单元的表征和计算机机制的重要性，从而与下文汉字的产生式识别过程、笔画表征依赖节点分析等内容相呼应。正文的修改在第 12 页第 1 段。

意见 2：文中“汉字的产生式识别过程”、“汉字的产生方式”等概念过于抽象，需进一步解释。

回应：感谢审稿专家提出的意见。“汉字的产生式识别过程”、“汉字的产生方式”等概念基于产生式思想。视觉的产生式观点认为，人在看到静态图像（如一个手写字）时，会自发推断它是如何产生的（该字是如何写出来的）。

汉字的产生方式可以通俗理解为一个汉字书写出来的过程。例如，“人”字由撇、捺两笔画按照自上而下、自左向右的规则书写产生。汉字均是二维图像在时空维度上的展开，因此可以将汉字的产生方式统一描述为：在定义汉字的基本笔画集合的基础上，从中抽取一定数量和种类的笔画，然后利用习得的笔画正字法规则（包括先验的笔画组合概率），按照自上而下、自左向右的原则，在恰当的位置上以恰当的顺序将上述笔画书写出来，从而产生汉字。

汉字的产生式识别是对汉字产生过程的逆向推理，即根据静态汉字这一输出结果反推之前的书写过程。仍以“人”字为例，看到这一图像，根据图像来推论它的书写过程是先写左边的撇，后写右边的捺，而非其他别扭的书写方式。同样的两个汉字，书写方式一定相同。因此，汉字的产生式识别过程实际上是，判断输入字和“心理字典”中某字的产生方式在多大程度上相同。该过程可以描述为，先基于整字图像中的线段交点提取出节点，接着枚举所有能产生该节点的笔画组合方式，再利用笔画先验概率筛选出其中概率最高的组合方案，从而获得输入字的产生方式。当长时记忆中某个原型字的产生方式与之高度相似时，则可将输入字判定为原型字，从而实现汉字识别。

文中第 12 页倒数第 2 段、第 14 页倒数第 2 段进行了修改。

意见 3：两个实验选取 SOA 的依据是什么？

回应：两个实验选取连续的多档 SOA，是基于产生式识别过程的时间进程特性：识别系统需要先根据底层视觉特征定位节点，然后对该区域附近的笔画进行拆解，接着再计算各种笔画组合的概率，选出最有可能的一种。上述计算过程需要一定时间执行，因此在加工的早期尚不能展现节点发挥的笔画拆解作用，在视觉系统执行到一定加工深度时才能发现相关效应。由于理论模型并不明确具体加工速度，我们在实验前无法知道具体的时间窗，因此从低

到高选取了多档 SOA，以探索该效应的时间动态。实验中选取 SOA 的依据是前人研究，参考了采用微观发生法的类似汉字识别实验的设置：沈模卫,朱祖祥.(1995).整体汉字字形识别过程探索. 应用心理学(02),43-48；沈模卫,潘善会,陈新.(1998).合体汉字字形识别过程探索. 应用心理学(01),27-32。上述研究同样关注字形识别的早期加工过程，采用的 SOA 能较好展现汉字加工变化过程的精细信息。我们的实验结果表明，操纵 SOA 的操作有效揭示了节点对汉字识别的早期影响——以实验 1 为例，整字能被有效识别时对应的刺激呈现时长，恰好与节点数量效应出现时的呈现时长一致。由于字数限制，上述依据未在正文中展开详细叙述。

意见 4：实验材料选取方面：实验一选取的部分假字看上去并不符合汉字的方块结构，更像非字。虽然结果只对真字进行统计分析，但作者仍应该区分假字和非字；

回应：非常感谢审稿专家的建议。实验 1 的独体假字由汉字笔画直接构成，按照书写习惯和经验，挑选经常同时出现的笔画组合，并放置在恰当位置，使之具有近似方块的结构，且与真实汉字有足够的区分度。我们预先制作了冗余一倍的待选人工字，由不知情被试挑选了其中最不像汉字的前一半，作为实验填充刺激。由于采用了手写字体，部分笔画并非完全横平竖直，导致整体看上去不太方正，使得实验中部分假字比较像非字。鉴于实验 1 采用真假字判别范式，人工字都是填充材料，我们仅对真字的结果进行分析，因此假字的问题对于真字组内部结果的意义影响较小。基于项目的方差分析也显示，节点数量的主效应是显著的： $F_2(1,18) = 4.56, p = 0.047, 2p = 0.20$ 。

意见 5：实验二在材料设置上存在一个问题：对于手写体汉字，有的汉字被掩盖的节点正好是笔画交叉的地方，如“舰”、“宾”、“蒙”，这类节点由于人们在写作时可能笔画间并不能完全连上，存在一些空隙，人们识别的更多，因此这类手写体更容易识别；而大多数字，如“棱”、“袒”、“椅”等汉字掩盖的节点是同一笔画中的节点，掩盖这类节点对汉字识别的干扰更大。因此，掩盖这两类不同的节点对汉字识别的影响干扰不同，作者在材料设置时并未考虑这个问题。

回应：十分感谢审稿专家指出的问题，这的确是可能影响研究结论的重要局限。经观察，这两类节点可以从节点的产生方式上做区分：一类是由笔画交叉且穿过产生的，如“十”、“椅”字中的节点；另一类是由笔画相接但不穿过产生的，如“丁”、“舰”字中的节点。为了考察这两类节点产生方式的不同，本文补充的实验三将节点产生方式做为一个自变量进行了操纵。结果发现节点产生方式的主效应不显著，但存在其与节点复杂度的交互作用。该交互作用具体表现为，掩盖交叉产生的复杂节点，比掩盖相接但不穿过的复杂节点，对识别的干扰更大。而掩盖不同产生方式的简单节点则没有明显差别。

我们认为，节点产生方式是影响节点复杂度的一个因素。在笔画数量一定的前提下，笔画交叉所产生的节点区域（如“十”字），相比笔画相交但不穿过的区域（如“厂”字），前者潜在的笔画组合方式更多，为了压缩问题空间，识别过程会更依赖这类节点的引导作用，因此掩盖该类节点产生的节点复杂度效应更大。而“舰”、“蒙”这类节点，人们在写作时可能笔画间存在一些空隙，恰是说明笔画相接而不穿过的区域，笔画运动信息较为明确，潜在的笔画组合方案较少，因此该处相交的笔画数量多少并不能对笔画的拆解过程起到明显影响，即这类节点的复杂度效应相对较小。

此外，如果假定节点复杂度效应完全是由书写笔画时是否存在缝隙导致的，那么实验 3 应该观察到节点产生方式的主效应，然而结果并不支持此假设。综上所述，节点复杂度效应确实存在，二节点产生方式这一因素对节点复杂度效应起到了调控作用。

意见 6：实验设计不够严谨，完全随机呈现刺激可能会出现同一个刺激连续呈现的情况，易

影响实验结果；另外，两个实验中同一个刺激都是多次呈现，很容易出现练习效应。在实验一中，在刺激呈现时间大于等于 40ms 时，练习效应可能更显著。实验二也很可能存在练习效应，对实验结果产生影响。

回应：非常感谢审稿专家指出的问题，这使我们的实验设计更加严谨。我们根据上述建议检查并重新分析了数据，获得了更加准确的结果：就同一个刺激连续呈现的情况而言，实验一中曾连续出现的刺激为 3、8、32、33 号汉字，所占试次约为 1.75%。排除这些试次后，仍存在节点数量效应： $F(1,25) = 9.72$ ,  $p = 0.005$ ,  $2p = 0.28$ 。实验二中曾连续出现的刺激为 29、48、76 号汉字，所占试次约为 0.47%。排除这些试次后，仍存在节点复杂度效应： $F(1,28) = 6.95$ ,  $p = 0.014$ ,  $2p = 0.20$ 。

此外，实验流程本身有助于减弱刺激连续呈现带来的影响：两个试次的刺激字之间间隔大约四秒，且配备了后掩蔽，判断和缓冲环节均在刺激所在的屏幕位置上设置了汉字提示。上述因素可以减弱上一试次的视觉后像对下一试次的影响。就练习效应而言，每个刺激呈现的次数和时长是一致的，练习效应理论上是平衡的。

为了彻底排除连续呈现所带来的干扰，新加入的实验三中单一刺激只呈现一次，避免了练习效应，也获得了实验二发现的节点复杂度效应，表明该效应是稳健的。

意见 7：实验设计中加入后掩蔽刺激的目的是什么？后掩蔽会影响目标词的加工。

回应：非常感谢审稿专家提出的问题。实验采用的后掩蔽刺激是打乱的汉字笔画，目的是控制刺激呈现时间、避免视觉后像影响被试的辨认绩效，以防止出现偏高的正确率。

意见 8：统计结果部分：正确率和反应时结果描述不够完整，包括数据整体的均值、标准差、是否删除极端数据等都没有描述。

回应：感谢审稿专家指出的问题。数据整体的均值、标准差已补充在文中，数据分析的对象是所有试次。

意见 9：心理语言学的研究应该汇报  $F_2$  值，文中所有结果中都缺少  $F_2$  的汇报。

回应：非常感谢审稿专家指出的问题。所有实验的基于项目的方差分析结果（ $F_2$ ）已补充在文章中。

意见 10：实验二统计结果只介绍了交互作用显著，未进一步写明简单效应分析的结果。

回应：感谢审稿专家指出的问题。实验二中正确率部分介绍了简单效应分析的结果，反应时结果中复杂度有关的交互作用均不显著，所以没有分析简单效应。现已补充基于项目的方差分析，也未发现复杂度有关的交互作用。

意见 11：实验二中虽然错误率结果发现掩盖复杂节点的正确率显著低于掩盖简单节点的正确率，但反应时的结果并不支持正确率的结果。因此，实验二并不能有力证明节点复杂度效应的存在，应设计实验进一步探究。

回应：非常感谢审稿专家指出的问题。本文后续补充的实验三，将节点复杂度、节点的产生方式和节点所处的部件类型作为三个自变量进行了操纵，以期进一步检验节点复杂度效应。结果发现，在正确率上，节点复杂度的主效应显著，基于被试的分析发现复杂度和节点产生方式的交互作用。反应时上，尽管没有观察到节点复杂度相关的显著差异，但从差异方向看，整体的数据模式和正确率是一致的（正确率高的反应时短），不存在反应时-正确率权衡。实验 3 通过更加严格的控制，为节点复杂度效应提供了进一步证据。至于两个实验均得到反应时的阴性结果，我们认为可能是合体字笔画数较独体字更多，掩盖节点对单字判断反应时

造成的干扰有限。

意见 12：研究工作量不够，文中最后提出：根据贝叶斯规划学习模型的计算特性预测，汉字识别过程还会表现出笔画概率自主学习效应和高频笔画组合优势效应。作者可以将这两个效应的研究加入进来形成一个完整的大研究。

回应：感谢审稿专家提出的建议。本文主要关注汉字识别的早期过程，现已通过三个实验说明了节点为笔画分割提供关键信息，揭示了汉字基本元素——笔画的加工特点，发现笔画分离以并行加工的方式进行。以上贡献一方面有助于加深对于汉字加工过程的理解，也对于基于轮廓线条的客体加工机制具有启发作用，同时进一步验证了产生式模型作为理论工具用于解释人类认知过程的可行性和普遍性。对于贝叶斯规划学习模型所预测的高频笔画组合优势效应和笔画概率自主学习效应，已有一些支持的佐证，例如，汉字家族的促进效应表明形旁构字能力大的汉字更容易被辨认(张积家, 姜敏敏, 2008; Su & Weekes, 2007)，尤其表现在促进高频字的识别，同时抑制低频字的识别(钱怡 等, 2015)；笔画概率学习方面，研究表明同一笔画所占的比例可以作为正字法知识通过内隐学习获得(王菲, 2015)；另有研究在高强度练习和正确反馈条件下，发现在完成笔画维度特征分类任务时内隐学习有效(侯伟康, 秦启庚, 1996)。相对于这两个方面，产生式模型可以预测的效应中，笔画分割等早期过程仍缺乏有力的实证证据。本研究对此做了关键补充。

当然，要建立完整的汉字识别计算模型，还有很长的路要走。较高层面的效应仍需进一步深入分析，并结合计算模型做细致的量化检验。这些工作有待于理论工作的细化和计算模型的建立，我们后续将进一步深入考察汉字产生式识别过程的各个环节，以期构建更加完备的汉字识别理论。

.....

审稿人 2 意见：

意见 1：两个实验的字频和笔画数数据未呈现。

回应：非常感谢审稿专家指出的问题。上述数据已补充在文中各实验的实验材料介绍部分。

意见 2：既然作者假设节点影响个体对汉字的识别，实验 2 掩盖节点的实验控制稍显不规律，部分节点在声符，部分在义符，部分在上部，部分在下部，部分在开头笔画，部分在结尾，这些因素是否会对实验结果产生影响？

回应：十分感谢审稿专家提出的问题。实验 2 中掩盖简单节点和复杂节点的材料都选取了 12 个上下结构和 8 个左右结构的字，且掩盖第 1 和第 5 节点的刺激数量也是一致的，但声符和义符在实验 2 中的确并未控制。这引入了额外的混淆因素，削弱了实验 2 结果的可靠性，因此文章补充了新的实验。实验 3 采用 SP 结构形声字，将节点所在的部件类型作为一个自变量加以检验。结果表明，节点复杂度的主效应显著，部件类型的主效应不显著，节点复杂度和部件类型之间交互作用不显著。

意见 3：两个实验刺激呈现时间选取的标准是什么？低于 30ms 呈现刺激，对刺激的识别是阈上还是阈下，阈上阈下加工的机制是否同质，能够在同一个体系下对比吗？

回应：非常感谢审稿专家的提出的问题。

区分阈上和阈下并非我们操作呈现时间的原因，我们设置不同时间的目的是寻找节点数量效应所处的时间窗口。根据产生式理论，视觉系统需要先根据底层特征定位节点区域，接着进行笔画拆解，再计算各种笔画组合的概率，猜测哪些原型字可以产生这些笔画组合，从中选出最有可能的一种。上述各环节均需要一定时间执行，因此只完成早期加工，节点信息

提取不充分、整字尚不能辨认时，也就无法展现节点发挥的笔画拆解作用（反映在整字识别绩效上）。因此汉字需要加工一段时间才能体现出节点数量效应，而我们在实验前不确定节点相关效应出现时对应的时间窗口，便从低到高选取了多档时长，以寻找效应产生的时间窗口。具体的呈现时长数值参考了采用速示法的类似研究：沈模卫,朱祖祥.(1995).整体汉字字形识别过程探索. 应用心理学(02),43-48；沈模卫,潘善会,陈新.(1998).合体汉字字形识别过程探索. 应用心理学(01),27-32。上述研究同样关注字形识别的早期加工过程，采用的 SOA 能较好展现汉字加工变化过程的精细信息。

阈上和阈下加工机制有何异同，是一个非常值得深入研究的好问题。本研究所基于的产生式理论并未就意识问题做出预测，因此我们事前未提出相关的假设，在实验范式的选择上也未考虑测量阈下加工（选用了主观 LDT 报告）。从结果上看，我们在低于 30ms 的条件下未观察到节点相关效应，此时被试正确率很低，且倾向于将目标判断为假字，这提示阈下加工有可能存在不同机制，但并无直接证据。因此从理论上和结果上，暂时无法判断是否同质。

尽管缺乏直接证据，我们猜测阈上与阈下汉字识别并无不同机制。阈下加工和阈上加工在行为表现上的区别，可能就是基于产生式的逆向推理过程由于时间不足而未完成的结果。当输入信息产生的证据尚不足以支撑某个假设时，该信息本身特征上的差异不会造成任务绩效上的改变（例如节点数量所产生的效应）；而一些阈下启动效应（本研究未作测量，这里指的是前人研究中所发现的那些不影响主观报告、但产生了内隐启动的效应）的发生，是因为刺激得到了初步的加工，激活了语义、整字、部首、笔画等各个层面某些节点，改变了一些假设的后验概率，从而进一步改变了后续启动任务中某些假设的先验概率。

当然，上述推测需要进一步的实验来验证，需要采用启动效应等更合适的范式，专门探讨汉字识别中阈限下的笔画表征机制。由于我们当前无法确认阈上和阈下是否机制相同从而具有可比性，因此我们也未就不同时间点之间进行比较，仅关注在什么时间点上会出现节点数量效应。

意见 4：实验材料选取方面：实验一选取的部分假字看上去并不符合汉字的方块结构，更像非字。虽然结果只对真字进行统计分析，但作者仍应该区分假字和非字；

回应：非常感谢审稿专家的肯定和指正。我们也认为，自上而下过程是汉字识别中非常重要的部分，这也正是我们采用产生式模型作为本研究理论基础的原因。区别于传统的只考虑自下而上过程的辨别式计算模型，产生式模型最大的优点正是将自上而下的过程表达出来，并和自下而上过程有机结合。这也是我们尝试用产生式模型描述人类汉字识别机制的根本原因。

审稿专家提出考虑自上而下的过程的思路非常有启发性，以往大量研究支持汉字识别存在自上而下路径，例如整字促进部件识别、整字促进笔画加工等。本研究意在强调，不仅在上述较为宏观的层面有自上而下的作用，传统上被认为是自下而上占主导的微观笔画表征层面，实际上也是一个自上而下的产生式过程，即正确的笔画表征需要借助汉字书写的产生过程进行推断。根据产生式模型分析，该过程建立笔画表征需要依赖节点分析和概率推断，据此我们做出了存在节点数效应和复杂度效应的理论预测，并通过实验加以检验。因此，从实验设计看似是在考察自下而上的初级视觉特征加工，但实际上是通过这样的实验来检验上述根据自上而下的理论推出的研究假设。

汉字具有笔画、部件和整字多层次，当完成较高层次单元的识别后，掩盖节点是否有作用和意义，是一个值得深究的很有意义的问题。尽管超出了本文的探讨范畴，然而根据产生式模型推测，在概率推断环节，相较于低层次单元，高层次单元的证据占据了更高的支配地位，因此义符/声符完成识别后，很可能会削弱另一未识别部件中掩盖节点造成的干扰。已

有研究也认为，汉字的整体轮廓可以辅助完成识别，即存在整体加工提供的完型促进。从本文的实验结果来看，节点和部首也存在交互：实验 3 将节点所在的部件类型作为自变量，均采用左形右声结构字，得到了与前人“声旁优势效应和笔画顺序效应相抵消”类似的结果：掩盖声旁笔画节点，干扰了顺序靠后的声旁的笔画表征，造成声旁的优势效应不明显，部件类型主效应不显著。这表明节点可以通过影响笔画表征，进而与部件产生交互。

意见 5：实验 1 中，为匹配笔画数，低节点的刺激多数均是“点”笔画居多，这一类刺激的选择是否具有代表性，基于项目的方差分析结果如何需要提供。

回应：感谢审稿专家提出的问题。基于项目的方差分析发现，节点数量的主效应是显著的， $F_2(1,18) = 4.56$ ， $p = 0.047$ ， $2p = 0.20$ 。“点”笔画同其他笔画的主要区别在于尺寸较小，复杂性低。张积家等人（2002）发现，简单和复杂笔画所含像素数量尽管不同，但识别绩效无差别，表明不同笔画被表征为同质的组块。因此，可以认为“点”笔画与其他笔画在加工难度上无明显差别，少节点真字组采用含“点”笔画的汉字是可行的。

意见 6：实验 1 中，“两因素间交互作用达到边缘显著， $F(5,125) = 1.998$ ， $p = 0.083$ ， $2p = 0.07$ 。”作者后续分析了简单效应，这一结果， $p = 0.08$  如此解释是否合适。若交互作用不显著，后续结果的解释是否合适，作者的假设是否成立。

回应：感谢审稿专家提出的问题。经综合考虑，此处正确率下交互作用的边缘显著是有意义的：一方面实验 1 中正确率结果与反应时结果具有相似的趋势，反应时数据显示两因素间交互作用达到显著；另一方面正确率结果的  $2p$  具有中等效应量( $0.06 < 2p < 0.14$ )。若交互作用不显著，由主效应可知多节点字的正确率显著高于少节点字。此外，若仅关注平均正确率在 50% 随机水平以上的条件（刺激呈现时间 40~60ms 的试次），这部分数据也表明存在节点数量效应，实验 1 的假设是成立的。

意见 7：.在文献综述部分，作者强调笔画在汉字识别中的作用，但实际上，作者强调的基于贝叶斯规划学习理论的节点效应，更多是笔画的组合，笔画的组合对汉字加工而言，具有其特殊性，不同节点的意义不同，节点出现在不同位置的作用不同。例如，多数两个笔画义符的节点非常重要，两个笔画已经确定了汉字义符的基本语义范畴，但是，对于多笔画义符而言，节点是否同样重要；同时，由于汉字构字的特殊性，偏旁部首的种类具有较大的语义价值，其笔画组合的变通性较低，节点是否又如此重要？作者在前言中的例子“例如，仅由“横”和“竖钩”两划和一个节点构成的“丁”字，还可能存在两种分割方案：“横折钩”和“短横”；两个“短横”和一个“竖钩”。”，是否仅仅是“计算模型”的推理过程和结果，真实的汉字加工是否需要如此认知的冗余。整体而言，可做汉字部首的笔画组合有限，个体对汉字部首的学习效果已经确保不合适的笔画组合出现在义符位置的不可能性，那么，作者基于“可见节点附近的所有潜在笔画组合可以由历遍该节点附近每一条线段来产生。对于由  $n$  个笔画构成的节点，若令其潜在笔画组合的集合为  $A(n)$ ，则可以推论， $A(n)$  会随交点相连的笔画数迅速膨胀。”是否成立值得商榷。

回应：非常感谢审稿专家提出的意见。从理论和实验上均能发现，不同位置上节点的作用会有差异。基于贝叶斯规划学习理论，我们认为节点的作用是为拆解笔画组合提供关键信息，这是节点的一般性视觉功用，不会因节点所处位置而改变。然而，节点的识别重要性，可以因节点所处位置而有差异：一方面，节点在多笔画义符和少笔画义符上的作用差异，涉及到笔画数量和节点重要性的关系。理论上，在节点数量一定的前提下，笔画数量越多，字形识别所需的概率推断计算量和难度应该越大。若干扰部件中笔画的分离和表征，理应对识别绩效造成更大损害。因此，多笔画义符中，节点可能为正确拆解义符笔画、正确识别义符整体，

均提供了更重要的作用；另一方面，笔画组合变通性不同的部件，节点的作用也应该有差异。这可以从概率推断的计算特性作出解释：笔画组合变通性可以近似看作笔画组合的概率分布情况，变通性低的组合具有较高的先验概率，意味着该处节点能够拆解出的笔画组合较少，且各组合的概率分布趋势更加集中。即便缺乏节点提供的运动信息，依靠穷举也可较快确定概率最高的笔画组合。而变通性高的组合，一旦缺乏节点提供的引导信息，穷举和概率推断的过程都将更加困难。因此，节点应在高变通性组合中更加重要。以上均是基于理论的分析，未来可以分别就问题展开实验探索。总的来说，节点的重要性会因节点所处位置而有差异，这是由节点所在笔画、部件的特性造成的。正是节点的拆解笔画功用普遍存在，才使得节点在不同位置上均有价值。

审稿专家关于“学习会降低认知冗余”的讨论是很合理的，也正是我们想表达的生成式模型的核心优势之一。前言里对笔画组合膨胀的分析，主要是在分析识别问题本身（而不是一个有经验的人或机器会采用什么样的算法和策略）。在代入生成式模型进行分析时，是基于一个没有经过学习、不带有先验知识的“白板”生成式模型来讨论，以明确该种识别模型的基本计算特性（例如会产生节点数量效应和复杂度效应），从而推出可供实验检验的研究假设。以没有经验的模型为基础只是为了理论分析的方便，这并不意味着我们否认学习的作用。恰恰相反，生成式模型的一个显著优势就是可学习，跟人类相近，可以在不断的自我学习中掌握数据分布，更新先验概率，从而提高识别效率。例如“丁”字，人们可以一眼看出“横”和“竖钩”的概率是最高的，而“横折钩”和“短横”概率低很多，正是通过长期学习掌握了符合事实的先验概率，使概率推断所需探索的假设空间受到约束，大大减小，达到降低认知冗余的效果。生成式模型能够体现这一特性的佐证是，理论上它和人一样借助高强度先验进行快速筛选时会产生“计算偏见”，识别先验中高概率的笔画组合时有加工优势，但识别先验中的低概率笔画组合反而更容易出错。这一预测得到了偏旁家族相关证据的支持（钱怡 等, 2015）。上一段中，节点之所以在不同位置上的识别价值有差异，也是人通过学习，改变了原本的扁平先验所造成的结果。

因此，前言从一个没有先验信息的初始模型切入，并不是在否认学习的重要性，或者拒绝承认偏旁部首的自上而下的作用。与之相反，生成式模型非常重视学习和先验知识，也非常适合表达学习和经验积累的动态过程。本文要强调的一点是，即使笔画加工这种看起来很底层的环节，它也是一个生成式的过程。研究的逻辑是，当假设它是靠生成式过程完成识别时，理应出现笔画节点数和复杂度效应，因此我们分别对这两个假设进行了检验。

意见 8：既然作者已经意识到不同部位的节点具有不同的识别价值，那么基于普遍性的、相同节点具有同样价值的节点效应是否正确？若不同节点具有不同的节点效应，那么作者的推论是否正确？

回应：感谢审稿专家提出的意见。诚然，不同节点会在识别价值上表现出差异。而究其原因，节点所处笔画的顺序、所处部件的类型，应是导致这一差异的根源。也即，不同部位的节点发挥同样的作用——拆分笔画，这与识别价值差异其实并不矛盾。对于一个没有经过汉字学习的“白板识别者”而言，或者人在刚接触一门陌生文字时，普遍性的节点效应仍然是存在的，它是一个更基础的视觉效应。而节点在不同部位上的识别价值差异，是人经过学习，改变了原本的均匀先验所造成的。本文中实验 3 也可提供节点有一般功用的证据：采用 SP 结构形声字的实验 3 也发现节点复杂度效应，并且掩盖声符节点和掩盖义符节点相比，整字识别绩效的影响没有显著差别，这与以往关于笔画省略方式和声旁关系的研究相一致——略去 SP 形声字声旁上的笔画，会导致声旁优势效应和笔画顺序效应相抵消。也即，不同顺序上的节点作用有差异，不同部件类型上的节点作用也有差异，它们可以相互抵消。因此，尽管不同部位的节点具有不同的识别价值，但都具有相同的视觉功用，即帮助获得笔画表征。

审稿人 3 意见：

论文考察了笔画节点在汉字识别中的作用。研究基于产生式理论框架，考察笔画节点特征对汉字识别的影响。研究视角较为新颖，研究问题有一定的理论意义。有下列建议供作者参考。

意见 1：在文章的背景阐述中，作者提出笔画是汉字识别的起点，类比贝叶斯规划学习模型，提出了汉字的产生式识别过程。从产生的过程来看，确实人们需要一笔一划地输出汉字，但这并不表明汉字的识别是以笔画为单元的。有研究表明汉字识别的单元是部首，而非笔画，作者是如何考虑不同的研究结果的。有关汉字识别过程的研究非常多，应系统深入地阐述汉字识别的最基本单元是什么，在此基础上提出相应的模型。

回应：非常感谢审稿专家提出的问题。一些研究认为字素和部件是合体字的识别单元，这可以从书写运动单元、部件数、部件频率、部件位置、部件类型等维度体现(冯臣, 张清芳, 2012; 韩布新, 1998; 张积家, 姜敏敏, 2008; 余成武, 张敏, 2014; Chen & Yeh, 2017), 表明合体字识别中部件发挥了作用, 是识别单元之一。大量笔画方面的研究发现, 合体字也会表现出笔画数效应、笔顺效应, 且笔顺效应能与声旁优势效应相抵消、笔画和部件层次上均存在错觉结合、觉察深层次部件的笔画变化比浅层次部件更困难等(彭聃龄, 王春茂, 1997; Taft & Zhu, 1997; 罗艳琳 等, 2008; 余成武, 2011; 闫国利 等, 2013; 张瑞, 2017)。综合多方证据, 我们支持笔画和部件均是合体字识别单元的理论, 且笔画是更低级的识别单元, 这也符合多数汉字识别经典模型的观点, 包括经成分模型(Huang & Wang, 1992)、合体汉字识别的相互作用模型(沈模卫, 朱祖祥, 1997)、多层次交互模型(Taft & Zhu, 1997)和格式塔认知模型(陈传锋, 黄希庭, 2004)等。此外, 本研究并非意在提出一个完整的汉字识别模型, 仅是阐述产生式思想可以作为一种研究汉字识别的角度。上述对汉字识别单元的讨论已加入文章第 29 页 4 段。

意见 2：因为是从产生的角度提出理论模型，考察对汉字识别过程的影响。文献综述应阐述书写与阅读之间关系，以及书写过程如何影响了汉字识别过程。

回应：非常感谢！审稿专家关于书写和阅读识别之间关系的建议对于丰富产生式汉字识别模型的理论基础非常有帮助。以往大量研究的确发现书写能力与阅读能力之间有正相关关系(Tan, Spinks, Eden, Perfetti, & Siok, 2005; Chan, Ho, Tsang, Lee, & Chung, 2006; McCarney Peters, Jackson, Thomas, & Kirby, 2013), 认知神经方面的证据也支持两类能力存在密切关系, 一方面两者存在共享的神经网络, 包括左侧梭状回和左侧额下回(Rapp&Lipka, 2011; Purcell Napoliello, & Eden, 2011)等; 另一方面阅读过程中书写相关的核心功能区也会激活(Longcamp, Anton Roth, & Velay, 2003)。行为学研究也发现, 人在学习书写汉字的过程中会掌握笔画的运动知识(Tan et al., 2005)。从产生式的角度来看, 汉字的产生过程即笔画在平面上依据一定规则运动、组合, 最终构成汉字的过程。产生式的逆向推理则意味着, 复现静态汉字的动态产生过程需要利用习得的汉字书写相关知识, 特别是笔画运动信息。因此, 笔画运动信息理应是汉字书写和汉字识别的关键。我们将以上内容补充进了文章的前言和讨论部分(第 13 页第 3 段、第 27 页第 4 段), 这有益于丰富本研究的理论意义, 为汉字产生式识别过程提供理论支撑。

意见 3：产生式模型的构建中是如何考虑不同因素包括笔画、部首、汉字结构、笔画顺序和位置等的影响的？这些因素是否会对本文中所操纵的节点数据和节点复杂度产生影响？

回应：非常感谢审稿专家提出的问题。汉字的产生式识别过程如图所示。



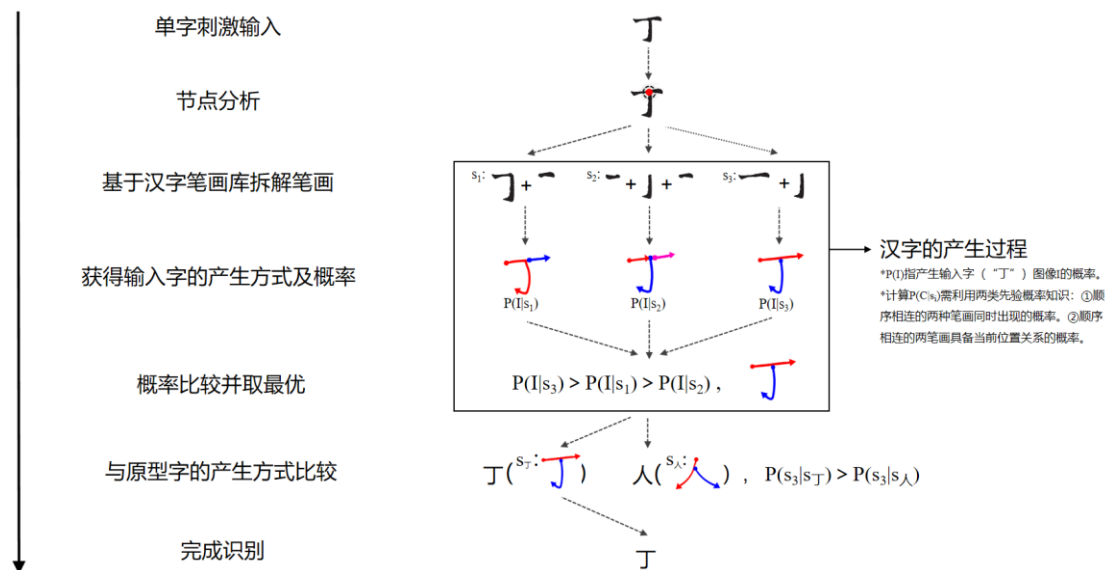


图1 汉字的产生式识别过程

需要指出的是,目前提出的产生式过程并不完备,本文关注的是该过程中节点分析和笔画表征这一基础和底层的环节。基于产生式思想,我们推测部首、汉字结构、笔画顺序和位置等因素很可能对识别过程有影响:就独体字而言,①笔画数量越多,概率判别的计算量越大,识别也越困难。②顺序在前的笔画更加重要,因其在概率上制约了后续笔画的种类和位置,如“横”后面更可能是“竖钩”而非“捺”,且横与竖钩更可能是交叉的位置关系而非首尾相接;就合体字而言,需要在独体字识别过程之外加入部件层次,类比笔画之间存在种类和位置的概率关系,部首之间也可能存在类似的概率关系,因此人需要依据部件先验概率筛选部件组合方案,进而推断合体字的产生方式。如果人以上述过程识别汉字,部件的类型很可能影响其在概率判别中的权重(例如,与部件家族大小有关)。上述是基于理论的分析,具体的影响和机制,有待后续研究解决。值得一提的是,本文新增的实验3,严格控制了部首、汉字结构和位置等因素,也观察到稳定的节点复杂度效应。

意见 4: 什么是节点? 如何计算一个汉字的节点数? 笔画和节点之间的联系和区别是什么?

回应：感谢审稿专家提出的问题。汉字是二维空间内线条的交错连接，节点是至少两个笔画交错连接时所构成的交接或交叉区域。一个汉字的节点数指该字中此类区域的数量。笔画和节点之间的联系是，节点由多条笔画相接或交叉产生，节点提供的分离引导信息有助于从整字的“一堆线条”中拆分并表征笔画。笔画和节点之间的关键区别是，笔画的运动信息较为明确（遵循自上向下、自左向右规则的单向线段），节点处的运动信息则具有高不确定性（多种运动方向的线段集合）。

意见 5: 文章前言中提到笔画的顺序位置也会影响汉字的识别, 在操纵节点的时候是否考虑了笔顺这一因素?

回应：感谢审稿专家提出的问题。实验 2 中考察了节点顺序的影响，区分了掩盖第 1 节点和第 5 节点时节点复杂度效应的差异。

意见 6: 实验 1 设计中同时包括了真字和假字, 假字是如何形成的? 假字的构成符合正字法规则吗? 从所举的例子来看, 假字不符合正字法规则。那么, 在真假字判断时被试可能基于正字法层面的信息做出真假字的判断, 不需要通达汉字的意义。

回应：非常感谢审稿专家提出的问题。实验 1 的独体假字由汉字笔画直接构成，按照书写习惯和经验，挑选经常同时出现的笔画组合，并放置在恰当位置，使之具有近似方块的结构，且与真实汉字有区分度。我们预先制作了冗余一倍的待选人工字，由不知情被试挑选了其中最不像汉字的前一半，作为实验填充刺激。由于采用了手写字体，部分笔画并非完全横平竖直，导致整体看上去不太方正。鉴于实验 1 采用真假字判别范式，人工字都是填充材料，被试的确可以基于正字法层面的信息做出真假字的判断；然而，我们仅对真字的结果进行分析，所有真字均符合正字法规则，因此真字组内部的结果差异是有意义的。基于项目的方差分析也显示，节点数量的主效应是显著的： $F_2(1,18) = 4.56$ ,  $p = 0.047$ ,  $2p = 0.20$ 。

意见 7：实验 1 设计包括了两个自变量：节点数和呈现时间。为什么要操作呈现时间这一自变量？对于所研究问题的意义是什么？

回应：非常感谢审稿专家提出的问题。之所以操作呈现时间，是考虑了产生式识别过程的时间进程特性：视觉系统需要先根据底层特征定位节点区域，接着依据节点提供的信息进行笔画拆解，后续再计算各种笔画组合的概率，选出最有可能的一种。上述各环节均需要一定时间执行，因此只完成早期加工，节点信息提取不充分、整字尚不能辨认时，也就无法展现节点发挥的笔画拆解作用（反映在整字识别绩效上）。由于理论模型并未对具体加工速度作出明确预测，在实验前我们不能估计效应出现的时间窗口，因此从低到高选取了多档 SOA，以探索该效应的时间动态。实验中选取 SOA 的依据是前人研究：沈模卫,朱祖祥.(1995).整体汉字字形识别过程探索. 应用心理学(02),43-48；沈模卫,潘善会,陈新.(1998).合体汉字字形识别过程探索. 应用心理学(01),27-32。这些研究以速示法和微观发生法探究字形识别的早期加工过程，采用的 SOA 能较好展现汉字加工变化过程的精细信息。

这一操作这有利于揭示节点分析所处的识别加工阶段——以实验 1 为例，在刺激呈现时间较短（小于 40 ms）时，节点信息尚未充分提取，即多节点字和少节点字的识别绩效未表现出差别，整字识别的正确率并未超出随机水平（50%）。呈现时间达到 40 ms 后，节点获得较好表征，节点数量不同的字的识别绩效开始有差异，整字识别正确率也开始超过随机水平。可见，整字能被有效识别时对应的刺激呈现时长与节点数量效应出现时的呈现时长一致。因此，操作呈现时间这一自变量，有效地揭示了节点应是整字刺激输入后较早获得的基础特征，利用节点信息是执行笔画分离和有效识别整字的前提。

意见 8：实验 1 采用了独体字，实验 2 采用了合体字，汉字的结构是否会对实验结果产生影响？

回应：非常感谢审稿专家提出的问题。本文所依据的产生式汉字识别过程是以笔画为基本单元，独体字作为由笔画直接构成的汉字，是检验节点数效应和节点复杂度效应最理想的实验材料。实验 2 之所以采用合体字，一方面是独体字的数量太少，不足以找出节点复杂度差异较大的两组字；另一方面则是合体字虽然存在部件层级，但单个部件可以“视为”直接由笔画构成的独体字，因此节点数量效应和节点复杂度效应理应在合体字的单个部件上存在，并通过整字的识别绩效展现出来。

意见 9：实验 2 的实验设计采用 2（掩盖复杂节点和掩盖简单节点）\*2（掩盖第 1 节点和掩盖第 5 节点）\*4（呈现时间：60 ms、70 ms、80 ms、90 ms）三因素被试内设计。对于第二个自变量来说，第 1 节点一般都在第一个部首上，第 5 节点一般都在第二个部首上。对这一变量的操纵混淆了部首因素。

回应：非常感谢审稿专家提出的问题，审稿专家指出部首因素的意见非常具有建设性。为了控制部首类型——声旁和义符的差异，文中补充的实验 3 将部件类型作为自变量加以检验，

实验材料均为 SP 结构的形声字（PS 结构的字太少，难以平衡）。基于被试和基于项目的方差分析表明，部件类型的主效应并不存在，部件类型和节点复杂度的交互作用也未出现。我们认为这是声旁优势效应和笔画顺序效应相抵消的结果，未来可以将其他结构的汉字作为对象，进一步检验节点作用。

意见 10：建议采用线性混合模型进行数据分析，同时考虑被试和项目这两个随机变量。

回应：十分感谢审稿专家提出的建议。综合多位审稿专家意见，我们对三个实验的数据进行了基于被试和基于项目的分析，分别报告了 F1 和 F2 值。

意见 11：建议作者针对前言和方法部分所提出的问题，结合实验 1 和 2 的实验设计和发现进行综合性的讨论。研究基于笔画为单元的产生式理论框架，但已有研究发现汉字识别的单元可能是字素或部首。建议作者综合考虑已有研究与本文的发现，讨论本研究发现的独特之处及其贡献，同时指出本研究设计的局限性。

回应：非常感谢审稿专家提出的意见。我们在讨论中增加了对笔画、部件关系的讨论，加入了部件作为识别单元的证据。本研究以三个实验探究了笔画节点在汉字识别中的作用，我们认为研究的独特之处和贡献主要有三方面：①首次将计算理论新思想——产生式识别模型应用于汉字识别，提出“汉字识别是对其产生过程的逆推理”的假设。②阐述了节点在独体字、合体字识别中均有引导笔画拆解的作用，并发现相关的节点数量效应和节点复杂度效应，为汉字识别的产生式假设提供了重要证据。③揭示了汉字基本元素——笔画的加工特点，发现笔画分离以并行加工的方式进行。以上贡献不仅有助于加深对于汉字加工过程的理解，而且对于基于轮廓线条的客体加工机制具有启发作用，同时进一步验证了产生式模型作为理论工具用于解释人类认知过程的可行性和普遍性。

实验 2、3 发现掩盖合体字中复杂节点比掩盖简单节点对整字识别产生的干扰更大，说明节点在合体字识别中发挥了作用，经节点拆分出的笔画也是合体字识别所需的表征单元。这与先前的理论和实验证据相符合。一方面，多数模型认为笔画和部件是合体字识别中信息上下传输的两个层级，部件是连接笔画和整字的加工中介，如经成分模型(Huang & Wang, 1992)、合体汉字识别的相互作用模型(沈模卫, 朱祖祥, 1997)、多层次交互模型(Taft & Zhu, 1997)和格式塔认知模型(陈传锋, 黄希庭, 2004)等。另一方面，大量以合体字为刺激的实验表明，笔画加工参与了合体字识别过程，证据主要来自笔画数效应、笔顺效应，且笔顺效应能与声旁优势效应相抵消、笔画和部件层次上均存在错觉结合、觉察深层次部件的笔画变化比浅层次部件更困难等(彭聃龄, 王春茂, 1997; Taft & Zhu, 1997; 罗艳琳 等, 2008; 余成武, 2011; 闫国利 等, 2013; 张瑞, 2017)。一些研究指出字素和部件是合体字的识别单元，这体现在书写运动单元、部件数、部件频率、部件位置、部件类型等方面(冯臣, 张清芳, 2012; 韩布新, 1998; 张积家, 姜敏敏, 2008; 余成武, 张敏, 2014)，表明合体字识别中部件发挥了作用，是识别单元之一。综合已有证据，合体字加工的单元包括了笔画和部件，它们在字形识别过程中发挥各自作用。上述对汉字识别单元的讨论已加入文章第 29 页第 2 段。

本研究设计的局限性以及未来的研究方向主要有以下几点：①仅从节点分析和笔画表征入手，未能解释部件如何参与产生式识别过程，未来可以探究人如何完成部件分离，以及对部件进行概率推断的计算机制。②采用的节点复杂度定义尽管较易操作，但不够精确，理想的定义是该节点附近所有可能的笔画组合数量。③实验材料为独体字和 SP 结构的形声字，未能在其他结构（如上下结构、PS 结构）的汉字中检验节点的作用。

---

## 第二轮

审稿人 1 意见：

作者回回应了之前的问题，但有些问题回回应的不够明确，结合之前的问题请作者思考以下几个问题：

意见 1：文中第一段强调了理解并研究手写体汉字的意义，且研究材料均为手写体汉字，但文中研究的节点数量效应和节点复杂度效应并非是手写体汉字的特性，为什么要突出手写体汉字？

回应：感谢审稿专家指出的问题。节点的确不是手写体独有的特征，突出手写体是因为相对于印刷体识别，揭示人类的手写体汉字识别机制有更大的理论和应用价值。

印刷体汉字识别的对象源自标准字体库，其训练使用固定样本，识别对象变异小。在这样的条件下，机器识别算法已经达到甚至超越人类的识别绩效，人的识别机制并无明显优势。因此，开发印刷体汉字识别算法，不必参考人的认知加工机制，也无需站在产生式识别的角度从节点开始分析。

然而，对于手写汉字的识别，人仍具有相对优势。现有的手写字识别模型还存在训练集过大、风格迁移困难、跨样本准确率不足、可解释性低等问题，因此有必要理解人类的手写字识别机制，特别是各水平加工单元的计算表征，从而启发新型识别算法的开发。这正是我们聚焦于人的手写汉字识别机制的原因。基于笔者对汉字（或图形）认知研究相关文献的梳理，产生式模型是解决该问题的一种有效途径。文中的节点数量效应和节点复杂度效应，是依据人的手写字识别机制（即产生式的反向推理过程）所提出的具体假设，因此探讨手写体汉字识别机制采用了实验中的刺激。

意见 2：文中多处语句不够严谨，例如 12 页最后一句话：“可见，理解汉字的产生过程对识别汉字具有重要意义。”根据前文内容得出这个结论有待商榷；

回应：感谢审稿专家对表述合理性提出的建议，的确，根据前文内容做此表述有武断之嫌。我们将其调整为“可见，如果汉字识别是一个产生式过程，那么首先需要理解汉字是如何产生的”，作为对下文的过渡；并为该推论补充了证据支持：根据视觉合成分析过程理论，对视觉客体的加工依赖于加工主体所掌握的背景知识，即客体是如何产生的。这种站在图像产生的角度以概率推断原则理解图像识别的思想，得到了行为学（Laeng et al., 2014）、神经科学（Kok et al., 2013）和计算科学（Gershman et al., 2012）证据的支持。现已将上述证据以参考文献的形式加入本段。

意见 3：文中统计结果中多次出现 F1 和 F2 统计结果不一致的情况，如 F1 结果显著而 F2 结果不显著，这种情况建议作者采用线性混合模型分析数据，以减少随机效应；

回应：非常感谢审稿专家在分析方法上的建议。相较于方差分析(ANOVA)，线性混合模型(Linear Mixed Model, LMM)具有一些优点，例如将全部原始数据纳入模型，数据利用率更高；且在计算数据时可采用最大随机效应结构，将被试和项目定义为交叉随机效应(crossed random effects)同时纳入模型，能够提高结果稳定性、统一性。

然而，我们查阅了心理语言学领域内混合效应模型功效分析的多篇文献，发现目前广泛认为采用 LMM 方法需要大量观察样本量(number of observations)，例如满足单个条件下至少 1600 个观察样本量(Brysbaert & Stevens, 2018; Kumle, V ã & Draschkow, 2021)，或被试和项目均不能低于 24 个(Kyle et al., 2016)，如果涉及交互作用则需进一步增加(Brysbaert, 2020)，否则计算的结果可能缺乏可靠性和稳定性(underpowered)。而本研究早先并非为适应混合线性模型而设计，三个实验预先设定且实际招募的被试量均不超过 30 人；此外，采用的汉字

材料由于需要控制笔画数、部件数、字频、构型以及节点相关特征，较难找到大量材料，因此单个条件的刺激项目数均不超过 20 项。上述原因导致本研究的观察样本量明显偏低，难以支撑 LMM 分析。

我们也检查了文中 F1 和 F2 结果，发现主效应上不一致的情况并不多，仅实验 2 的正确率结果未观察到项目分析的显著效应。交互作用上，实验 1 中节点数量和刺激呈现时间交互作用的项目分析不显著。考虑到实验 1 设置的 6 档刺激呈现时间是为探索节点加工的时间窗口，在 40ms 以上的呈现时长下被试方能做出有效真假字判断（正确率高于随机水平 50%），若单独分析 40 至 60ms 的试次，交互作用的结果是一致的；实验 2 复杂度和节点顺序交互作用上，正确率的项目分析不显著；实验 3 复杂度和节点产生方式交互作用上，正确率的项目分析不显著。考虑到两个实验的刺激数量均不多于 20 项，可能是由于刺激的内部变异较大，掩盖了本该显著的效应。因此对交互作用的讨论主要参考基于被试的分析。

此外，我们通过增加实验，以严格筛选材料降低项目内部变异的方式提高了节点复杂度效应的可靠性。实验 2 复杂度主效应的项目分析结果不显著，可能是由于选取材料时对部件属性的控制不够严谨。实验 3 则对部件的笔画数、频率、结合律等加以匹配，从而观察到一致的效应。综合两个实验，可以认为存在节点的复杂度效应。

单字材料内部变异较大，是汉字加工研究常见的问题。汉字之间可能存在多种难以控制的系统性干扰因素，致使项目分析 F2 有时不能通过显著性检验（章玉祉，张积家，2019；杨群等，2019）；且经过多条件筛选后的单字材料通常数量不多，难以支撑混合线性模型。这些不易克服的困难，可能是近年大多数汉字相关研究选择同时报告 F1 和 F2，并更多地依赖 F1 的原因（章玉祉等，2021；杨群等，2021；孙尔鸿等，2019；王丹等，2019）。

最后，我们也专门咨询了曾在心理学报发表文章的心理语言学专家，经沟通认为采用被试和项目分析是可行的（曹海波等，2023）。

意见 4：应将实验 2 选取 SOA 选取依据加入正文中；

回应：感谢审稿专家提出的建议。上述内容已放入实验 1 的实验设计部分。

意见 5：实验 2 讨论部分解释复杂度效应在第 1 节点比较弱的第 2 个原因：在第 1 节点时笔画交叉的情况比交叉但不穿过的情况更少，这么解释的依据是什么。

回应：感谢审稿专家提出的问题。这一解释是根据产生式理论推导所得：在产生节点的笔画数一定时，笔画交叉的节点相比笔画相接的节点，前者的潜在笔画组合空间会更复杂，信息量也更大，可能会在识别中起更大作用（图 1 展示了交叉和相接两种节点产生方式下潜在笔画组合数量的差异）。因此，节点产生方式有可能是造成第 1 节点上节点复杂度效应较弱的原因

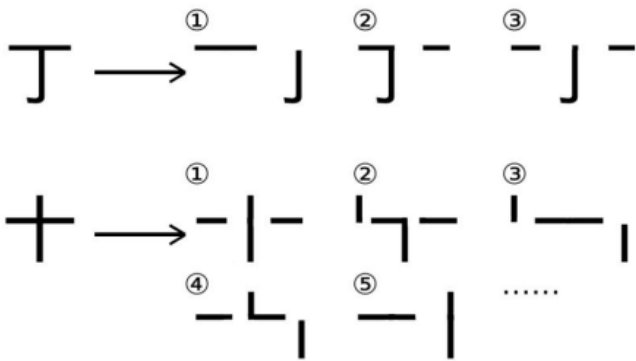


图 1 不同节点产生方式下潜在笔画组合的数量差异

意见 6：实验 2 和实验 3 的反应时结果均不显著，虽然如作者解释反应时和正确率结果趋于一致，但从统计学角度是没有差异的，作者如何解释？

回应：感谢审稿专家提出的问题。

首先，理论上由产生式模型，对节点复杂度效应可以做出较为明确的预测是正确率上的变化，而反应时上的变化不明确。

根据产生式识别过程可推论，连结较多笔画的复杂节点，其附近潜在的笔画空间较为庞大。为了压缩问题空间，识别过程会更依赖节点的引导作用。如果缺少节点的自下而上信息，计算过程将在约束时间内给出一种解，但各个可能解的概率区分度不高。而问题空间越大，给出解恰是正确笔画表征的概率越小，即正确率越低。因此，产生式过程预测的节点复杂度效应是正确率上的预测。两个实验的正确率结果也支持上述理论预测。

相比正确率，反应时的预测并不明确，还要取决于问题空间搜索过程是以并行方式还是串行方式进行。理论上，不论是掩盖简单节点还是复杂节点，计算过程都将在约束时间内给出一种解。如果计算以并行方式进行，那么认知系统的笔画搜索耗时将不受笔画空间大小的明显影响；反之，若以串行方式加工，那么耗时将随笔画空间增大而增加。现有的证据倾向于支持笔画组合搜索是并行的：一方面，实验 1 发现节点的搜索和加工以并行进行，另一方面，以往研究认为视觉系统对简单刺激具有并行加工的特性，许多涉及统计和概率计算的加工耗时不随加工对象数量的增加而增长，例如统计知觉的相关研究（Sang et al., 2003; 励奇添, 陈文锋, 2022）。此外，认知系统采取串行或并行加工，也与被试的认知风格有关（Das et al, 1995），可能存在被试间差异，导致实验结果难以预测：若大部分被试探索潜在笔画空间的过程是并行计算的，那么反应时上就可能没有明显效应。

其次，实验设计上，我们采用了较短的呈现时间，使被试保持中等的正确率，以此观察汉字识别的加工绩效，该操作有助于展现加工过程导致的正确率上的细微变化，但可能不利于突显反应时上的效应。将来可以尝试其他对反应时更为敏感的范式，例如在更长呈现时间、保持较高正确率的条件下，采用真假字判断以外的任务考察反应时可能存在的效应。此外，也可能由于合体字笔画数较独体字更多，掩盖单个节点对整字判断反应时造成的干扰有限等原因，研究 2 和研究 3 中未观察到反应时的效应（但反应时结果的整体趋势与正确率结果并不矛盾）。

意见 7：需统一文中的部件名称，从部件功能角度划分，形声字的两个部件分别被称为“义符”和“声符”；从部件构形角度划分，这两个部件分别被称为“形旁”和“声旁”。请作者选择一个角度的部件名称，文中出现混用。

回应：感谢审稿专家严谨的用语指正。本文对部件的区分重在构形差异，现已将文中部件名称统一为“形旁”和“声旁”。

意见 8：文中多处不规范现象，如图均不是标准三线表格，图 8、图 9 的标题不符合规范，24 页 t 值和 p 值需要斜体，参考文献中部分中文参考文献不符合规范，中文参考文献不需要“&”符号。

回应：感谢审稿专家的指正，文中已做相应修改。

意见 9：3.5 讨论第一段末尾“笔画穿过”写作有误，这里应指的是笔画相接但不穿过的情況。

回应：感谢审稿专家的指正，文中已改正。

.....

审稿人3意见：

作者针对所提问题做了思考和回复，但是与本研究密切相关的思考和回复并未体现在修改稿中。

意见1：本文研究的目标过程是汉字识别，从书写产生的角度提出汉字识别的产生式模型。在前言中要深入地比较已有汉字识别模型和产生式模型的不同，阐述书写中的各种不同因素对汉字识别过程的影响。这与上次提到的第二个和第三个问题密切相关，作者做了回复，但文章修改稿并未体现出来。上次所提的两个相关问题是：（1）因为是从产生的角度提出理论模型，考察对汉字识别过程的影响。文献综述应阐述书写与阅读之间关系，以及书写过程如何影响了汉字识别过程。（2）产生式模型的构建中是如何考虑不同因素包括笔画、部首、汉字结构、笔画顺序和位置等的影响的？这些因素是否会对本文中所操纵的节点数据和节点复杂度产生影响？总体来看，这篇文章总共包括了三个实验，涉及到了多个层级上的变量，例如新增实验3中其中一个自变量为“掩盖声符节点或掩盖义符节点”，作者需要在前言中提出一个整体的理论框架，阐述研究中所涉及的自变量对汉字识别过程的影响，以理清理论框架和研究思路。

回应：非常感谢审稿专家对文章前言结构的详实建议。我们将之前回复中的内容加入了前言：一方面，介绍了书写与阅读之间关系，以及书写过程如何影响汉字识别过程，并认为书写习得的运动和视空间信息是节点重要性和自上而下笔画先验知识的来源；另一方面，我们将合体字的部件层级纳入了前言中对汉字产生式识别过程的叙述。

需要指出，汉字识别是一个非常复杂的、多层级嵌套的机制，本研究仅是产生式思想指导下构建汉字识别模型的初步工作。前言的目的是将产生式思想引向汉字识别的领域，因此分别介绍了视觉的产生式理论、经典汉字识别模型、笔画表征的重要性、基于产生式的笔画表征获得机制和宏观的产生式识别流程（包括书写对阅读的影响）等内容。由于目前提出的识别过程并不完备，根据产生式的计算机机制尚不能对部件和汉字结构等识别机制作出明确推理，因此前言主要阐述了识别过程中初级的节点分析和笔画表征环节。

虽然当前工作聚焦于初级加工阶段，对于更高层次的加工过程和因素也可以做理论分析。由产生式思想可预测，部件、汉字结构、笔画顺序和位置等因素很可能对识别过程有影响（虽然它们并非本研究的主要对象）。以图2的“仃”字为例，其包含单人旁和丁字两个部件。单个部件的产生方式，是以一阶马尔可夫方法，运用自下而上的节点信息和自上而下笔画概率知识而计算获得的。根据该计算过程的特性可推测，相较擦除顺序在后的笔画，擦除顺序在前的笔画很可能对整个部件的识别产生更大损害，因其扰乱了后续笔画的可能解，即笔画顺序是一个影响因素；“仃”字是左形右声结构，左侧的单人旁在合体字部件的先验知识中，具有较高的位置概率和出现概率（处于常见位置），因此该部件后验概率的计算速度相对于右侧的丁字部件应是较快的，从而展现出义符位置效应，即汉字结构的因素；同时，这一部件的识别优势可能会带动整字识别绩效的提高，从而表现出形旁大家族的促进效应，即部件的因素。然而，这些分析仅是猜想，暂无实证研究数据的支撑。出于文章严谨性的考虑，正文中并不将上述因素纳入产生式过程的介绍。

本文的三个研究始终围绕前言所预测的节点相关效应，从数量和复杂度出发对节点所承担的笔画拆分作用进行检验，部件的位置、性质等因素仅作为可能影响实验准确性的变量进行操纵或控制。的确，它们如何在产生式的框架下参与合体字识别过程，是一个很重要的问题，后续将逐步考察各个层次的产生式识别过程。鉴于笔画水平之上的高层级产生式加工还需进一步研究验证，我们仅在正文中作大致介绍。下图是我们对合体字产生式识别过程的猜想（部件和整字层次的加工暂不明确），并对潜在的研究要点做了标注，例如备择部件的激活机制、各层次部件的先验概率、合体字的判别过程可能需要考虑单个部件的多个可能解等。



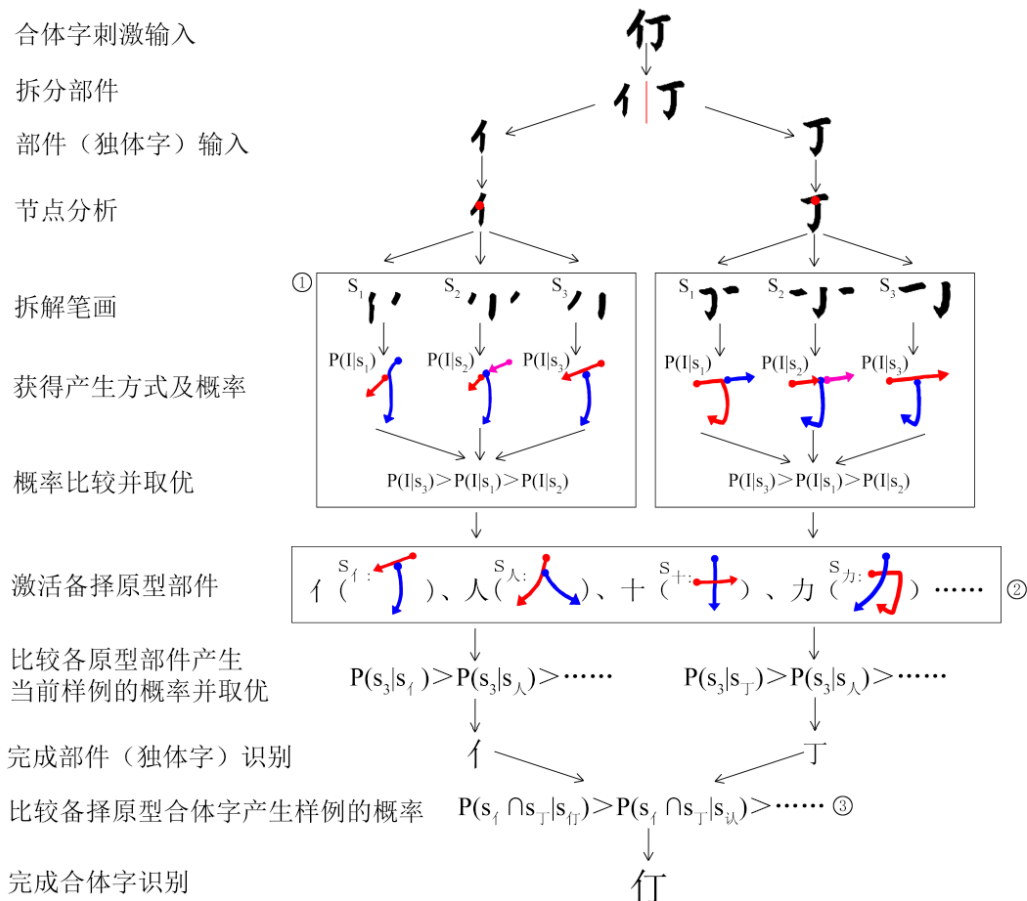


图 2 合体字的产生式识别过程示意图

注：标记①处方框中的内容描述了独体字的产生过程，即笔画如何基于概率规则产生整字。将  $P(I)$  定义为输入字图像  $I$  (如“亻”) 的概率。计算产生方式  $s_i$  下图像  $I$  的后验概率  $P(I|s_i)$ ，需要利用两类先验概率知识：一是顺序相连的两种笔画同时出现的概率，二是顺序相连的两笔画具备当前特定空间关系的概率；标记②处的方框表示将输入部件(独体字)的产生方式与长时记忆中原型部件的产生方式作比较，该过程需要激活一系列备选部件；标记③表示判别长时记忆中哪个原型字最有可能产生含有当前两个部件的输入字，该环节需要激活备选合体字。此外，合体字的判别过程可能需要考虑单个部件的多个可能解。

意见 2：上次提问过的问题：“什么是节点？如何计算一个汉字的节点数？作者应该在文中相关部分做详细说明，而不是仅仅做回复。

回应：感谢审稿专家的建议。我们将之前回复的节点和节点数概念加入了正文：在前言中论述笔画组合解算法和解空间的部分（21 页末）明确了节点的概念，在解释节点数量效应的预测时定义了节点数的概念（23 页）。这使文章的逻辑更加清晰。

意见 3：实验 1 设计包括了两个自变量：节点数和呈现时间。为什么要操作呈现时间这一自变量？对于所研究问题的意义是什么？作者不仅仅是回复，而是需要在文章前言阐述操纵这一实验自变量的意义。

回应：感谢审稿专家的意见，我们将之前回复的操作呈现时间这一自变量的原因和意义加入了实验 1 的实验设计部分。



意见 4: 我建议作者进行线性混合模型分析, 修改稿中报告了 F1 和 F2 值。确实, 我注意到另一个审稿人建议增加 F2 分析, 线性混合模型是同时考虑被试和项目这两个随机变量的数据分析方法。作者是如何思考这两类分析方法的? 为什么选择报告 F1 和 F2, 不做线性混合效应模型分析? 此外, 作者仅针对显著结果报告了效应量, 不显著就没有报告。可以查阅期刊要求, 是否有此规定。

回应: 非常感谢审稿专家在分析方法上的建议。线性混合模型(Linear Mixed Model, LMM)在计算数据时采用最大随机效应结构, 可以将被试和项目定义为交叉随机效应(crossed random effects)同时纳入模型, 相对于报告 F1 和 F2 具有数据利用率更高, 可有效规避被试分析和项目分析检验结果不一致的情况, 使计算结果更加统一和稳定等优点(Baayen et al., 2008; Josse et al., 2014)。

未采用混合线性模型的主要原因是, 我们查阅了心理语言学领域内估计混合效应模型功效分析和效应大小的多篇文献, 发现目前广泛认为采用 LMM 方法需要大量观察样本量(number of observations), 例如满足单个条件下至少 1600 个观察样本量(Brysbaert & Stevens, 2018; Kumle, Vö & Draschkow, 2021), 或被试和项目均不能低于 24 个(Kyle et al., 2016), 如果涉及交互作用还需进一步增加(Brysbaert, 2020), 否则计算结果可能缺乏稳定和可靠性(underpowered)。而本研究早先并非为适应混合线性模型而设计, 三个实验预先设定且实际招募的被试量均不超过 30 人; 此外, 采用的汉字材料由于需要控制笔画数、部件数、字频、构型以及节点相关特征, 较难找到大量材料, 因此单个条件的刺激项目数均不超过 20 项。上述原因导致本研究的观察样本量明显偏低, 难以支撑 LMM 分析。

目前, 研究者报告 F1 和 F2 结果仍是较为常见的做法。我们查阅了多本心理语言学领域的优秀期刊, 发现许多文献选择报告基于被试和项目的方差分析(Tamaoka et al., 2022; Rück et al., 2021; Gulowski & Błaszczak, 2020)。此外, 也检索了心理学报等国内期刊最近几年的相关文献, 发现与本研究相近的汉字加工研究多以 F1 和 F2 结果展开讨论(邓碧琳, 张积家, 2020; 章玉祉, 张积家, 2019; 王丹 等, 2019; 杨群 等, 2019)。

同时感谢审稿专家敏锐地指出效应量的问题。我们报告显著结果效应量而未报告不显著结果效应量, 正是因为要遵守心理学报的指南规定(<https://journal.psych.ac.cn/xlxb/CN/column/column6.shtml>)。

意见 5: 实验 2 的实验设计采用 2 (掩盖复杂节点和掩盖简单节点) \* 2 (掩盖第 1 节点和掩盖第 5 节点) \* 4 (呈现时间: 60 ms、70 ms、80 ms、90 ms) 三因素被试内设计。对于第二个自变量来说, 第 1 节点一般都在第一个部首上, 第 5 节点一般都在第二个部首上。对这一变量的操纵混淆了部首因素。上次所提到的这个问题主要是考虑到所掩盖节点在汉字的不同部首上, 位于不同的部首会对结果产生影响。作者回复: 实验 3 中考虑了部件类型(声旁和义旁), 我认为这是两个不同的变量。建议作者去检查一下所用材料在掩盖第 1 节点和第 5 节点时是否存在跨越部首的情况, 做一些对比性的分析排除可能的混淆。

回应: 感谢审稿专家细致的建议。我们检查了实验 2 的实验材料, 真字组共 80 字, 其中 20 字掩盖第 1 位的复杂节点(有 19 个掩盖节点在第一个部件上), 20 字掩盖第 5 位的复杂节点(有 1 个掩盖节点在第一个部件上), 20 字掩盖第 1 位的简单节点(有 19 个掩盖节点在第一个部件上), 20 字掩盖第 5 位的简单节点(有 1 个掩盖节点在第一个部件上)。因此, 节点位置的跨部首情况, 应当在复杂度不同的节点材料中是平衡的。

意见 6: 有些段落的阐述需要增加参考文献, 例如前言中以“类比贝叶斯规划学习模型……”开头的段落, 没有一条参考文献作为支持。

回应: 感谢审稿专家对写作提出的建议。该段落是对汉字产生式识别过程的宏观描述, 承接

上文介绍的通用字符识别的贝叶斯规划学习模型。从图像产生的角度解释图像的识别过程，这一思想基于产生式视觉的概率推断理论，并得到了行为学（Laeng et al., 2014）、神经科学（Kok et al., 2013）和计算科学（Gershman et al., 2012）证据的支持。现已将恰当的视觉识别机制证据以引证文献的形式加入本段，加强了相关论述的支持。

---

### 第三轮

审稿人 1 意见：

作者回回应了之前提出的问题，结合之前的问题请作者思考以下几个问题：

意见 1：关于文中数据统计，作者给出了不用线性混合模型分析数据的理由，其中“考虑到两个实验的刺激数量均不多于 20 项，可能是由于刺激的内部变异较大，掩盖了本该显著的效应。因此对交互作用的讨论主要参考基于被试的分析。”既然材料已经控制了笔画数、部件数、字频、构型以及节点相关特征，为什么刺激的内部变异还会掩盖本应该显著的效应？若真是这样，说明材料控制的还不到位。

回应：非常感谢审稿专家颇具启发性的问题。的确，对刺激材料进行严格控制后，材料的内部变异将大大降低，有助于观察到预期效应。这与本研究的结果是一致的：实验 2 中材料的多个无关变量未控制，导致项目分析中节点复杂度的主效应不显著；进一步的实验 3 则对这些变量加以平衡，就观察到了显著的节点复杂度主效应。

然而，经进一步控制后，实验 3 中复杂度和节点产生方式交互作用的项目分析结果并不显著。这表明，在笔画数、部件数、字频、构型以及节点相关特征等常规控制变量之外，可能存在未经控制的与节点产生方式有关的未知变量。由于有关汉字产生式识别理论和节点附近笔画识别机制的研究刚刚起步，对于相关影响因素的认识尚浅，目前只能就已知会对汉字识别产生一般性影响的常规控制变量进行控制，而如单个部件内产生节点的笔画顺序等因素难以通过选择真字材料加以控制（也可能由此造成刺激内部变异）。这提示我们，将来的工作应寻找笔画拆分的关键影响因素，可能通过制作假字进行更精细的材料控制。上述内容我们在 4.6 讨论部分也有所提及，感谢专家启发我们思考！

就本研究结果而言，鉴于当前控制下的节点复杂度主效应的被试和项目分析均显著，支持了我们提出的高复杂度节点会提供更为丰富的笔画拆解信息这一推论。

意见 2：手写体汉字是否应该在标题和摘要中提出，请作者斟酌。

回应：感谢审稿专家的提示，经多位作者讨论商议，认为标题和摘要应突出手写体汉字。原因在于突出手写体有助于读者快速理解本文的研究对象和价值。一方面，本文所提理论和实验，是针对手写体汉字识别过程；另一方面，手写体汉字识别相较于印刷体识别，在理论探索和应用实践上均有尚待解决的问题，本文可以为相关研究者提供启发。因此，现已将文章标题改为“笔画节点在手写体汉字识别中的作用”，并在摘要中修改了相应表述。

意见 3：文中存在多处病句和标点符号错误，如前言部分“果汉字识别的过程相当于汉字产生的反向推理，那么获得笔画表征应当是汉字识别的前提。”，28 页“呈现时间的主效应显著， $F(2,22,62.14) = 8.06$ ”等。

回应：感谢审稿专家对写作细节的关注和指正，上述问题已在文中做出修改。

意见 4：实验 1 结果部分：“基于被试的分析发现两因素间交互作用边缘显著。”后面既有 F1 结果，也有 F2 结果。

回应：感谢审稿专家对数据细节的敏锐观察。查阅多本期刊的相关文献后，我们发现大多数研究在 F1 和 F2 并非均显著的情况下，也选择同时报告两者。我们认同并遵从该惯例，这有助于读者充分了解研究结果。

.....

审稿人 3 意见：

作者针对所提问题做了认真修改，回应了所提问题。文中仍然存在一些写作方面的问题，比如 4.6 讨论部分，一句话形成一段，建议作者进行整合。此外，文中也存在一些不通顺的句子，比如 4.6 讨论中出现的“——若略后的笔画位于 SP 形声字的声旁”，不知道作者要表达什么。建议作者务必通读全文，修改此类错误。

回应：非常感谢审稿专家对写作的关注和指正。多位作者通读全文并商议后，对包含上述问题的不足之处作了合理修改，这使行文更加流畅和符合逻辑。改动如下：

1、4.6 讨论部分中，将交互作用的相关论述合并为一段。

2、4.6 讨论部分中，将“若略后的笔画位于 SP 形声字的声旁”一处修改为——“节点复杂度和部件类型不存在交互作用，可能是由于实验材料均为左形右声的形声字，掩盖形旁和声旁节点分别意味着干扰顺序在前和在后笔画的表征，即笔画顺序效应和声旁优势效应相互抵消。以往研究为此提供了佐证：相比形旁，形声字的声旁在整字识别中的作用更为明显；而相比顺序靠后的笔画，顺序在前的笔画对整字识别更为重要。因此，两种因素相平衡后，节点复杂度在声旁和形旁上的影响没有明显差异(闫国利 等, 2013)。”

3、3.5.1 正确率报告中，修正了 F 值保留小数位的问题。

---

编委复审意见：

我自己独立阅读了该论文初稿，然后仔细阅读了论文的初审意见、所有评审人的后续评审意见。基于 3 位评审人中已有两位评审人同意接受修改后的论文，我建议“修改后接受该论文。”其实，论文作者也应该很清楚该论文的统计处理方法在当前的心理语言学研究是不常用的了。我建议作者认真阅读关于 F1 和 F2 统计方法在心理语言学中使用的统计逻辑，如为什么要报告基于被试方差分析的结果，又要报告基于项目方差分析的结果。有专门的英文论文，具体题目我不太记得了，是比较早的文献，建议仔细阅读。当前该研究报告中，F2 的结果很多不显著，请研究者进一步在文中以“下标的方式稍作解释”。同时，关于 LMM 的统计方法，在我的记忆里是 2008 年后逐渐开始的统计方法，我在国外时系统学习后，然后在国内审读语言认知的行为、眼动、ERP 等研究报告时，通常建议的处理方法。我看了作者对该方法的介绍，建议作者继续加强对该方法的学习，特别是阅读英文原始书籍，以后尽量采用 LMM 做语言认知的数据统计分析。

回应：答：非常感谢编委专家的宝贵建议，现已对文中不显著的 F2 结果以脚注方式作了简要解释，分别在第 28、31 和 36 页。同时，感谢专家的悉心指导，根据您的建议，我们参考了如下等文献，正在加强对心理语言学专门统计方法的学习，以期在后续研究中可以实现更完美的数据分析。

[1]Andy F., Jeremy M.,& Zoe F. . (2012). Discovering statistics Using r. SAGE Publications Ltd.

[2]Brown, V. A. . (2021). An introduction to linear mixed-effects modeling in r. *Advances in Methods and Practices in Psychological Science*, 4(1), 251524592096035.

[3]Faraway, J. J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and*

Nonparametric Regression Models. CRC press. 10.1201/b15416.

[4]Meteyard, L., Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092.

**主编意见：**

同意外审和编委意见。希望这个研究至少可以启发其他研究者进一步对该方向的探索。