

远距离规则的内隐学习使用了何种记忆存储器： 来自神经网络模拟的证据*

李菲菲 刘宝根**

(浙江师范大学杭州幼儿师范学院, 杭州, 311215)

摘要 远距离规则的知识是如何被内隐学习的, 研究尚未得出结论。该研究采用和人类被试相同的实验材料和程序, 考察了简单循环网络模型 (SRN) 对两种汉语声调远距离规则——倒映和逆行规则的学习。结果发现: (1) 在广泛的参数范围上, SRN 能够学会倒映和逆行规则, 表明模型的记忆缓冲器可以模拟人类远距离规则的内隐学习; (2) SRN 学习倒映规则比逆行规则更好, 表明在功能上远距离规则的内隐学习可能优先使用了先进先出的记忆存储器及信息加工模式。该研究为探究远距离规则内隐学习的机制提供了新的证据和视角。

关键词 远距离规则 内隐学习 记忆存储器 神经网络模拟

1 引言

人类具备一种功能强大的学习机制, 能够无意识地获得外部世界中结构化规则的知识 (Chan & Leung, 2014; Rohrmeier, Dienes, Guo, & Fu, 2014; Rohrmeier & Rebuschat, 2012), 这种学习机制被称为内隐学习 (implicit learning)。如远距离的语法 “is doing”, 幼儿通常能够在有意识地学习这一规则之前就熟练使用它。那么, 远距离规则 (long distance dependencies) 是如何被内隐学习的? 神经网络模型, 由于在学习过程 (经过训练和学习逐步调整内部结构之间的权重) 和学习特征 (采用分布式表征知识, 无法明显体现出规则) 上都和人类的内隐学习十分相似 (Cleeremans & Dienes, 2008), 因此十分有助于我们探究远距离规则的内隐学习机制。

远距离规则的内隐学习原则上需要学习机制具备一个记忆存储器 (memory buffer)。而简单循环网络模型 (simple recurrent network, SRN) 恰好具备一个背景层 (见附录), 能够承担存储记忆的功能。研究表明 SRN 能够在特定条件下学会远距离规则, 如增加模型结构并由简单到复杂输入材料 (Elman, 1993)、设置某个模型参数 (Elman, 1993; Timmermans & Cleeremans, 2000) 等。但这些研究都在特定条件下考察、且只考察了模型的内隐学习。

Kuhn 和 Dienes 则在更一般的条件下 (使用广泛的模型参数) 采用相同的材料和程序, 考察了模型和人类被试音乐远距离规则的内隐学习。他们发现 (Kuhn & Dienes, 2005) 人类被试能够内隐习得有关规则、样例和组块的知识; SRN 也能够学会规则, 但成绩不如样例和组块, 无法与人类被试的结果取得最佳匹配 (Kuhn & Dienes, 2008)。可见, SRN 具备的记忆存储器是否可以模拟人类的内隐学习机制, 仍有待研究。

如果远距离规则的内隐学习需要记忆存储器, 那么它是一个什么类型的存储器, 或者说在记忆存储器中信息的加工模式是怎样的? 这一问题可以通过比较两种对称规则——倒映 (inversion) 和逆行 (retrograde) 的内隐学习加以考察。对称是远距离规则的一种, 想象一面镜子水平或垂直地摆在前后半串元素中间、将前半串元素反射到后半串元素, 可以构成倒映或逆行结构 (Dienes, Kuhn, Guo, & Jones, 2012; Li et al., 2013)。如荷兰语中 “Aad has Jantje the teacher the marbles let help collect up” (英语表示为 “Aad let Jantje help the teacher collect up the marbles”) 符合倒映规则; 英语中 “The mouse the cat the dog chased bit ran away” 符合逆行规则 (见图 1) (de Vries, Christiansen, & Petersson, 2011;

* 本研究得到教育部人文社会科学研究青年基金项目 (17YJC880050) 和浙江省自然科学基金计划 (LY18C090006) 的资助。

** 通讯作者: 刘宝根。E-mail: liubaogen@126.com

DOI: 10.16719/j.cnki.1671-6981.20180405

Dienes et al., 2012; Fitch & Friederici, 2012)。研究者认为 (Christiansen & Chater, 1999; Dienes & Longuet-Higgins, 2004; Kuhn & Dienes, 2008)，先进先出 (first in-first out) 记忆存储器 (见图 2 左)，相当于一个队列，序列中的元素依次进入队列，每次只有最底部的元素能够被加工。如 B1 进入队列时，A1 被输出与之比较，之后被擦除；接着 B2 进入队列时，A2 与 B2 比较，再被擦除……，因此有利于检测倒映规则；而后进先出 (last in-first out) 记忆存储器 (见图 2 右)，相当于一个堆栈，序列中的元素依次进入分析器和堆栈，每次只有堆栈顶部的元素能够被加工。如 B3 进入分析器时，A3 被输出与之比较，之后被擦除；接着 B2 进入队列时，A2 与 B2 进行比较，再被擦除……，因此有利于检测逆行规则。

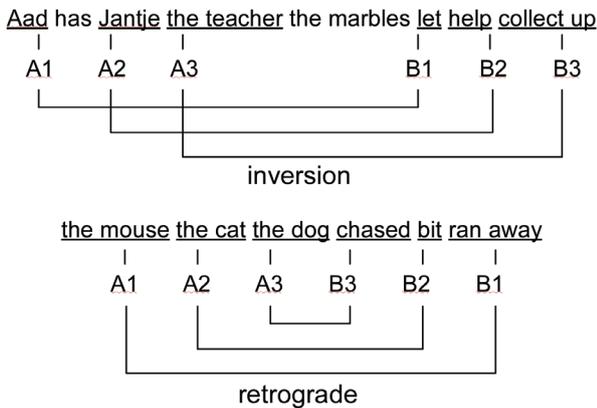


图 1 倒映与逆行规则示例

注：AB 代表前后半串，123 代表顺序。上例符合倒映规则，后半串的动词与前半串的名词主谓对应，且顺序相同；下例符合逆行规则，后半串的动词与前半串的名词主谓对应，但顺序相反。

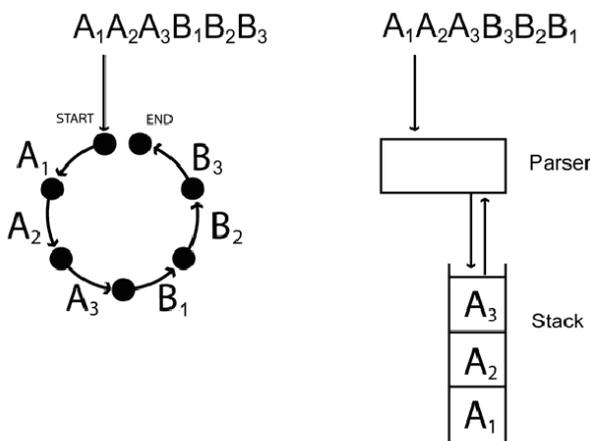


图 2 先进先出和后进先出记忆存储器结构图示

(Uddén et al., 2012)

Uddén 等 (2012) 将 F-L、D-P 两对字母嵌套

在字母串中分别构成倒映 (如 MF \overline{F} D-L \overline{L} LPVS) 和逆行结构 (如 MF \overline{F} D-PL \overline{L} V \overline{S})，发现被试能够学会这两种规则，且倒映规则的成绩好于逆行规则。Li 等 (2013) 采用汉语声调材料设置长度为 10 的音节串构成倒映 (前后半串音节平仄相反、顺序相同) 或逆行规则 (前后半串音节平仄相同、顺序相反) (见图 3)，得到了与 Uddén 等一致的结果，初步证实远距离规则的内隐学习优先使用了先进先出的记忆存储器。但是，上述研究中被试母语分别为荷兰语和汉语，荷兰语中存在倒映结构，汉语唐诗中亦存在平仄对应的倒映结构，所以倒映规则的成绩优势可能来自于记忆存储器的类型、也可能来自于语言经验。由于模型不具备语言经验，因此可以以模型为对象、排除人类被试可能存在的语言经验混淆，进一步考察远距离规则内隐学习使用的记忆存储器类型。

综上，本研究将采用与 Li 等 (2013) 人类被试相同的实验材料和程序，设置广泛的模型参数范围，考察 SRN 对倒映和逆行两种对称规则的内隐学习；并通过比较这两种规则内隐学习的难易，且匹配模型与人类被试的成绩，探究远距离规则内隐学习的记忆存储器类型及加工模式。

2 方法

2.1 设计

和 Li 等 (2013) 相同，采用 2 规则 (倒映、逆行) \times 2 组别 (实验组、控制组) 实验设计，四组被试 / SRN 模型，每组分别接受四种实验处理的一种。需要注意的是，人类实验的四组被试各不相同，是被试间设计；而模拟实验中四组模型都包含 150 个参数取值的模型，即四组模型参数相同，因此是被试内设计。因变量为模型测验阶段的成绩。

2.2 SRN 参数

SRN 的参数主要有四种：学习率决定模型各层之间权重变化的大小，动量规定一个权重持续变化的程度，隐含层单元个数指隐含层中单元的个数，循环次数指模型运行学习材料的遍数。这些参数并无明显的心理学或生理学含义，但能导致模型的成绩发生很大的变化。仅仅在某个参数值上证实它能够匹配人类被试的数据，并不足以证明模型能够解释人类内隐学习的机制。因此，参考 Kuhn 和 Dienes (2008) 选择这四种参数的各种取值：学习率：.1、.3、.5、.7、.9，动量：.1、.3、.5、.7、.9，

隐含层单元个数 5、10、15、30、60、120，循环次数 100 组合形成 150 个不同的模型。

Li 等 (2013) 被试为 94 名大学生，两个规则实验组各 25 人，两个规则控制组各 22 人。为模拟人类被试的样本量，每一组模型模拟 25 遍，每次使用一个随机的不同的起始权重。

2.3 材料

同 Li 等 (2013)，采用 10 个平声音节和 10 个仄声音节构成无意义的声音串，每串长度为 10，音节不重复，前后 5 个音节在平仄上构成倒映或逆行规则 (见图 3)。倒映规则中前后半串音节平仄相反、顺序相同；逆行规则中前后半串音节平仄相同、顺序相反。每个规则，学习阶段包含 48 串音节串，测验阶段包含 48 串新的音节串 (底层平仄串也是新的)，合法串和非法串各半。非法串中位置 7 和 9 的元素违反规则。

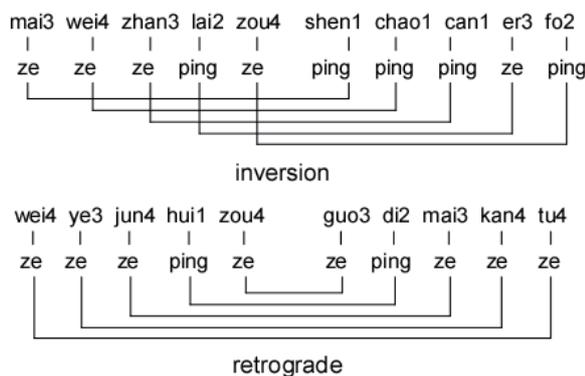


图 3 汉语声调对称规则示例 (上为倒映规则、下为逆行规则)

2.4 过程

同 Li 等 (2013)，两个规则的实验组模型完成学习阶段和测验阶段，两个规则的控制组只完成测验阶段。学习阶段，48 个声音串以不同的随机顺序向每个模型呈现 3 遍。声音串中的元素依次呈现，每个元素激活模型相应的输入单元，要求模型预测声音串中下一个元素。设置了两个输入/输出单元分别代表平和仄两种声调类型，这样设置的前提是：将音节归为平仄的知识是人类被试已有的类别概念，且这种知识在汉语声调对称规则的内隐学习中是必须的 (姜珊, 2012)。输入/输出单元若激活设为 .9，若不激活设为 .1。测验阶段，48 个新的声音串 (一半合法、一半非法) 以不同的随机顺序向每个模型呈现。声音串中的元素依次呈现，要求模型预测下一个元素。模型各层之间的权重可以继续变化，以类比人类被试在测验阶段可以继续发生学习的现象。

150 个模型每个模拟 25 遍，一共运行 3750 次模拟。模拟采用 Matlab 神经网络工具箱完成。

2.5 数据统计

模型的成绩为输出单元目标和实际激活模式构成的向量之间的夹角余弦 (\cos) (Christiansen & Chater, 1999; Dienes, Altmann, & Gao, 1999; Elman, 1993; Kuhn & Dienes, 2008)。Cos 越大，两个向量之间的距离越小，模型的成绩越好。为与人类被试的成绩作比较，用测验合法串的 \cos 平均分减去非法串的 \cos 平均分、再除以两者的联合标准差，得到 Z_{\cos} (Kuhn & Dienes, 2008)。声音串中只有后 5 个元素能够根据规则被前 5 个元素预期，因此因变量为测验串后 5 个元素的 Z_{\cos} ，大于 0 表示模型能够区分合法串和非法串。

3 结果

3.1 SRN 模型的总体成绩

每组模型的成绩见表 1。配对样本 t 检验表明，实验组模型的成绩显著高于控制组模型 (倒映规则: $t(149) = 10.16, p < .001, d = .83$; 逆行规则: $t(149) = 8.07, p < .001, d = .66$)，且经 Bonferroni 序列校正差异仍然显著。并且，倒映规则组模型的成绩显著高于逆行规则组模型 (实验组: $t(149) = 10.12, p < .001, d = .83$; 控制组: $t(149) = 2.15, p = .034, d = .18$)，且经校正差异仍然显著。以实验组的成绩减去控制组的成绩代表学习效应，另作检验发现，倒映规则组的学习效应显著大于逆行规则组， $t(149) = 9.82, p < .001, d = .80$ 。这说明虽然两个规则的控制组成绩差异也达到了显著水平，但是倒映规则比逆行规则学习效应更大。这些结果表明，SRN 不仅学会了汉语声调的倒映和逆行规则，而且倒映比逆行规则更加容易习得。

3.2 SRN 模型在每个可预测位置上的成绩

如果 SRN 学会了远距离的对称规则，那么当一个元素输入时，它应该能够在正确的位置上正确地预测与该元素对应的那个元素，而不管中间间隔的元素是什么声调类型。因此，考察实验组模型在测验合法串和非法串中每一个可预测位置上的成绩，以直接考察模型是否真正学会了对称规则。对模型在两个规则上的成绩作平均以选择模型的最佳参数取值 (Kuhn & Dienes, 2008)。结果发现，当学习率为 .7、动量为 .5、隐含层单元个数为 120 时模型的平均成绩最好，因此选择该参数的模型在两个规

表1 四组 SRN 模型和人类被试 (Li et al., 2013) 的成绩

	SRN (Zcos)				人类被试 (d')			
	实验组		控制组		实验组		控制组	
	M	SE	M	SE	M	SE	M	SE
倒映规则	1.36	.13	.00	.01	.47	.16	-.03	.07
逆行规则	.31	.04	-.02	.01	.28	.17	-.07	.05

则实验组中的成绩作进一步数据分析。两组在测验串中每一个可预测位置上的成绩见表2。配对样本 t 检验表明, 两个规则组, 在第7位和第9位上(非法串这两个位置上的元素违反规则), 合法串元素的成绩都显著高于非法串元素 ($p < .005$)。即在违反位置7和9上, 模型能够预测正确的声调类型,

而在合法位置6、8和10上, 合法串和非法串元素的成绩不存在显著差异(经序列校正)。这些结果表明, SRN 能够在正确的位置上正确地预测声调类型, 它学会了远距离的对称规则。

3.3 SRN 模型与人类被试成绩的匹配

对模型, 用实验组的 Zcos 减去控制组的 Zcos

表2 两个规则实验组在测验串每个可预测位置上的成绩 (Zcos)

		第6位		第7位		第8位		第9位		第10位	
		M	SE	M	SE	M	SE	M	SE	M	SE
		倒映规则	合法串	.95	.01	.95	.01	.88	.02	.92	.01
	非法串	.95	.01	.41	.02	.88	.02	.43	.03	.95	.01
逆行规则	合法串	.61	.02	.95	.01	.33	.01	.80	.02	.58	.03
	非法串	.61	.02	.56	.02	.33	.02	.70	.02	.60	.02

来代表学习效应; 对人类被试, 计算实验组和控制组被试 d' 差值的平均分 (M) 和标准误 (SE), 以 $M \pm 1SE$ 作为学习效应的典型区域 (Kuhn & Dienes, 2008) (见图4)。在倒映和逆行规则上, 150个 SRN 模型中分别有15个和38个模型的学习效应落入了人类被试学习效应的典型区域, 表明在一些参数上 SRN 能够模拟人类被试典型的学习效应。

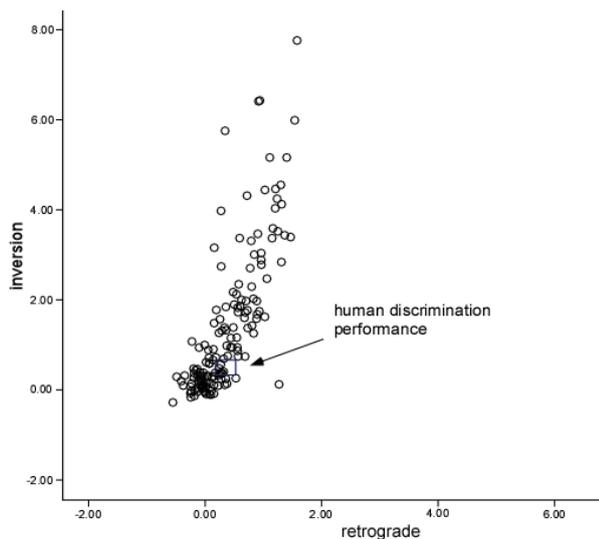


图4 150个 SRN 模型在两种规则上的学习效应

(横坐标为逆行规则、纵坐标为倒映规则)

注: 矩形代表人类被试的典型成绩 ($M \pm 1SE$)

通过比较倒映和逆行规则的成绩, 评价在两种规

则的学习难易上 SRN 的成绩是否能够匹配人类被试的成绩。以倒映规则的学习效应为因变量、逆行规则的学习效应为自变量作回归分析 (Kuhn & Dienes, 2008)。结果发现, 回归效应显著 ($\beta = .78, t(149) = 15.22, p < .001, r^2 = .61$), 该回归方程中的斜率更加倾向于倒映规则, 这表明 SRN 学习倒映规则比学习逆行规则更加容易, 与人类被试结果一致。

4 讨论

4.1 远距离规则内隐学习的 SRN 模拟

模拟结果表明, SRN 在倒映和逆行两种规则上都能够学会区分合法和非法的音节串, 并且音节串的组块这一表面特征已经进行了平衡, 表明模型学会的是底层远距离的对称规则。已有研究表明, SRN 能够学会区分倒映规则的合法串和非法串 (Timmermans & Cleeremans, 2000)。但是, 他们的结果可能只是基于训练和测验材料之间某种未知的统计属性, 而未必表明模型真正学会了倒映规则中元素之间的对应 (Kuhn & Dienes, 2008)。本研究则通过考察测验串中每个可预测位置上的成绩, 发现 SRN 能够正确地预测每个位置上符合对称规则的元素, 为 SRN 确实能够学会远距离规则提供了有力的证据, 这与 Kuhn 和 Dienes (2008) 关于 SRN 能够学会音乐倒映规则的研究结果一致。

而与 Timmermans 和 Cleeremans (2000) 一致、与

Kuhn 和 Dienes (2008) 不一致的是, 本研究为 SRN 设置了 100 的循环次数。Timmermans 和 Cleeremans 模拟倒映规则 (如 D F G K - F D L X) 的内隐学习时采用了 50、100、300、1000 和 3000 的循环次数, 发现总体上 SRN 的学习成绩在 100 次以上才能达到相对较高的水平, 之后稍微上升, 到 3000 次又有所下降。这说明模型可能需要更多的学习遍数才能达到与人类被试相当的学习成绩。而 Kuhn 和 Dienes 在模拟音乐倒映规则 (如 F3 C3 B3 C4 - G3 C4 D3 C4, 音程为 $-3 +6 +1 - +3 -6 -1$) 的内隐学习时, 只用了与人类实验相同的 1 次循环。结果发现 SRN 规则组的成绩不如样例和组块组, 而人类被试规则组学习与样例和组块组都很好; 150 个 SRN 模型中进入人类典型学习效应的个数, 样例组块组为 21 个、而规则组则没有。模拟与人类被试实验的结果匹配不好, 这可能与他们学习次数过少有关。本研究采用了更多的循环次数, 控制了组块, 在倒映和逆行规则的学习成绩上, 150 个 SRN 模型中分别有 15 个和 38 个模型的学习效应落入了人类被试学习效应的典型区域, 得到了与人类被试匹配较好的成绩。

此外, SRN 具备很多自由参数, 它的成绩可以在广泛的参数空间内变化。仅仅在某个参数值上证实它能够匹配人类被试的行为 (Elman, 1993; Timmermans & Cleeremans, 2000) 不足以证明模型能够解释人类内隐学习的机制。参考 Kuhn 和 Dienes (2008), 本研究采用了广泛的参数设置, 发现在两种对称规则上 SRN 的一般行为 (在 150 种参数取值范围上的平均成绩) 能够模拟人类被试的学习效应; 并且, 在一些参数范围上 (倒映规则 15/150、逆行规则 38/150 个模型的学习效应落入了人类被试学习效应的典型区域), SRN 还能够模拟人类被试典型的学习效应。由此, 本研究与已有研究一起, 证实了 SRN 能够习得远距离的对称规则, 表明 SRN 的记忆缓冲器能够模拟远距离规则的内隐学习机制。

4.2 远距离规则内隐学习使用的记忆存储器类型

模拟结果表明倒映规则比逆行规则更加容易被习得, 这意味着内隐学习机制在功能上可能优先使用了先进先出的记忆存储器。该结果与人类被试实验结果一致 (Li et al., 2013; Uddén et al., 2012)。但是, 人类被试的母语均存在倒映规则, 研究亦表明内隐学习对先前知识很敏感 (Chen et al., 2011; Ziori & Dienes, 2008), 因此, 语言经验可能导致内隐学习机制对学习倒映而非逆行规则更加敏感。人类实

验没有考虑和控制语言经验的影响, 无法证实内隐学习机制使用的记忆存储器类型。本研究中, SRN 并不具备对倒映规则的先前知识, 因此, 模拟结果发现的倒映对逆行规则的学习优势能够被归因于内隐学习机制优先使用了先进先出的记忆存储器。的确, 因为工作记忆中反向回忆比顺向回忆更加困难 (Robinson, Mervis, & Robinson, 2003), 工作记忆一般都是选择先进先出的加工方式。先进先出记忆预期倒映规则比逆行规则更加容易习得, 实验得到的结果也是如此。

需要说明的是, 内隐学习机制使用的真实的记忆缓冲器并不一定是先进先出或后进先出记忆。然而, 假设的记忆系统可以在功能上表现得更像先进先出或后进先出记忆, 或者更加容易被训练、并表现得更像这种或另外一种记忆。因此, 本研究只是从功能上考察内隐学习机制使用的记忆缓冲器的类型。并且, SRN 包含的记忆缓冲器的类型也并不明确, 它在计算上是十分灵活的, 它的类型依据训练时回传的误差情况而定。内隐学习机制包含的真正的记忆缓冲器的结构是什么, 仍然有待将来研究进一步确认。

倒映规则的学习优势可能还存在其它解释。例如, 根据句法预期位置理论 (Syntactic Prediction Locality Theory; Gibson, 1998), 逆行规则比倒映规则需要花费更多的记忆代价, 因此更难以被加工。因为, 在逆行规则中对应元素之间的间隔距离是长短不一的 (9、7、5、3、1), 而在倒映规则中对应元素之间的间隔距离是相等的 (5)。逆行规则需要更多的记忆容量来储存更长的间隔的无关元素信息, 因此需要花费更多的记忆代价。

5 结论

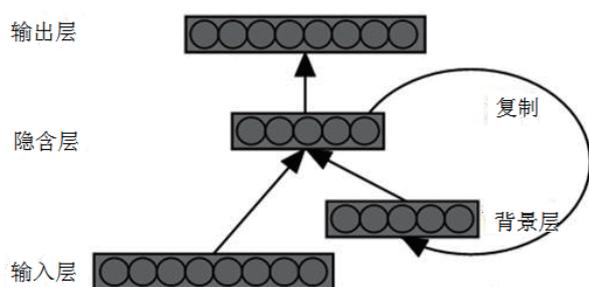
在广泛的参数范围上, SRN 能够学会倒映和逆行这两种远距离的对称规则, 且倒映规则比逆行规则更加容易习得。表明 SRN 的记忆缓冲器能够模拟远距离规则的内隐学习机制, 且在功能上内隐学习机制可能优先使用了先进先出、而非后进先出的记忆存储器及信息加工模式。

参考文献

- 姜珊. (2012). 汉语声调规则的内隐学习. 华东师范大学博士学位论文.
- Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, 27, 807-842.
- Chan, R. K. W., & Leung, J. H. C. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, 30, 463-484.

- Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious structural knowledge of form-meaning connections. *Consciousness and Cognition*, 20, 1751-1760.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157-205.
- Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun Sun (Ed.), *Cambridge handbook of computational cognitive modeling* (pp. 396-421). Cambridge England: Cambridge University Press.
- de Vries, M. H., Christiansen, M. H., & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics*, 5, 10-35.
- Dienes, Z., Altmann, G. T. M., & Gao, S. J. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science*, 23, 53-82.
- Dienes, Z., Kuhn, G., Guo, X., & Jones, C. (2012). Communicating structure, affect and movement: Commentary on Bharucha, Curtis & Paroo. In P. Rebuschat, M. Rohrmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 156-168). Oxford, England: Oxford University Press.
- Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*, 28, 531-558.
- Elman, J. L. (1993). Learning and development in neural networks - the importance of starting small. *Cognition*, 48, 71-99.
- Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B*, 367, 1933-1955.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1417-1432.
- Kuhn, G., & Dienes, Z. (2006). Differences in the types of musical regularity learnt in incidental- and intentional-learning conditions. *The Quarterly Journal of Experimental Psychology*, 59, 1725-1744.
- Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition*, 106, 184-206.
- Li, F., Jiang, S., Guo, X., Yang, Z., & Dienes, Z. (2013). The nature of the memory buffer in implicit learning: Learning Chinese tonal symmetries. *Consciousness and Cognition*, 22, 920-930.
- Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246-263.
- Robinson, B. F., Mervis, C. B., & Robinson, B. W. (2003). The roles of verbal short-term memory and working memory in the acquisition of grammar by children with Williams syndrome. *Developmental Neuropsychology*, 23, 13-31.
- Rohrmeier, M., Dienes, Z., Guo, X., & Fu, Q. (2014). Implicit learning and recursion. In F. Lowenthal & L. Lefebvre (Eds.), *Language and Recursion* (pp.67-85). New York: Springer Verlag.
- Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4, 525-553.
- Timmermans, B., & Cleeremans, A. (2000). In *Proceedings of the twenty-second annual conference of cognitive science society*. Hillsdale: Lawrence Erlbaum Associates.
- Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*, 36, 1078-1101.
- Ziori, E., & Dienes, Z. (2008). How does prior knowledge affect implicit and explicit concept learning? *The Quarterly Journal of Experimental Psychology*, 61, 601-624.

附录 SRN 的结构和工作原理



SRN 是一个三层的前馈模型，包含输入层、隐含层、输出层和背景层。每一层都包含若干单元，输入层和输出层、隐含层和背景层的单元个数相同。学习时，如某语法使用了 8 个字母，将输入输出单元个数设置为 8，使每个单元代表 1 个字母。学习字母串“BTSXVP”时，首先输入第一个元素“B”，要求模型输出下一个

元素“T”。那么“B”将激活输入层相应的单元，激活模式为“10000000”，输出单元的目标激活模式为“00100000”（假设第一个和第三个单元分别代表 B 和 T，1 代表激活，0 代表不激活）。隐含层单元根据它与输入层和输出层单元之间的联结权重决定激活模式。权重根据回传的输出单元的实际与目标激活模式的误差进行调整。同时，隐含层的激活模式将被复制到背景层中。接着输入第二个元素“T”，要求输出下一个元素“S”。同时，背景层会将上一时间点隐含层的激活模式反馈给隐含层，即此时隐含层的激活模式同时受到输入层和自己上一时间点的激活模式的影响。第三个及以后的元素以此方式依次运行。完整的一串字母串可以表示为一个向量，它由每个元素在输入单元上的激活值构成的二元向量组成。随着训练的进行，各层单元之间的权重逐渐稳定下来，这一过程即为模型的学习过程（Boucher & Dienes, 2003; Cleeremans & Dienes, 2007）。

Which Memory Buffer does the Implicit Learning of Nonlocal Dependencies Use: Evidence from Neural Network Simulations

Li Feifei, Liu Baogen

(Hangzhou College of Preschool Teacher Education, Zhejiang Normal University, Hangzhou, 311215)

Abstract In implicit learning literature, a basic question concerning how knowledge of structures and regularities is learned is whether the learning mechanism uses a temporary storage buffer, and, if so, what the nature of the buffer is. Recently, Li et al.(2013) found that people acquired unconscious structural knowledge of both Chinese tonal retrogrades and inversions. Moreover, inversions were implicitly learnt more easily than retrogrades, pattern predicted by implicit learning used a first-in-first-out buffer rather than a last-in-first-out buffer. However, because Chinese Tang poetry used an inversion, which was common knowledge that participants were likely exposed to as children, it was not clear whether prior expectations of structure instantiating inversions could override the effect of what type of buffer the system used. The network did not have prior knowledge. Accordingly, the present study investigated whether the Simple Recurrent Network (SRN), that used a buffer to allow learning of nonlocal dependencies, could learn tonal inversions and retrogrades, and replicate the advantage of inversions over retrogrades.

The SRN was tested on the same materials and procedures as Li et al. (2013). The networks were assigned to four cells of two training conditions (trained vs. untrained) by two rules (inversion vs. retrograde) design. The simulations were carried out using all possible permutations of the parameter values, resulting in 150 different models for each group. The materials were strings of tonal syllables. Each string consisted of 10 different tonal syllables, where the tone types (pings and zes) of first five syllables predicted the tone types of following five by forming an inversion or a retrograde. In training phase, 144 grammatical strings were used for two trained groups. In test phase, four groups of networks were presented with 48 test sequences (half grammatical and half ungrammatical), and their ability to predict the next tone in the predictable second five elements was used as an index of performance.

T-test (with Bonferroni correction) showed that trained networks performed significantly better than untrained networks for both inversion and retrograde groups, suggesting that the networks possibly learnt the two rules. Moreover, for both trained and untrained groups, inversion group performed significantly better than retrograde group. The performance difference between inversion and retrograde for trained networks was greater than that for untrained networks, indicating that inversions were implicitly learnt more easily than retrogrades. Further, the effects of learning were calculated by subtracting the z-scores of the untrained networks/participants from that of the trained networks/participants. A substantial number of the SRNs fell within the area covered by the human data ($M \pm 1SE$)(15/150 for inversion, 38/150 for retrograde), suggesting that the SRN could match the characteristic performance of human participants.

To conclude, consistent with the results of human experiments, the present simulations showed that: SRN could learn the two nonlocal dependencies, and tonal inversions were implicitly learnt more easily than retrogrades, tentatively suggesting that functionally a first-in-first-out memory buffer was more likely to be involved in implicit learning of nonlocal dependencies. Thus the present study provided new evidence and a new perspective for exploring the implicit learning mechanism of nonlocal dependencies.

Key words nonlocal dependencies, implicit learning, memory buffer, neural network simulations