

基于简单循环网络的远距离规则跨长度迁移模拟研究^{*}

张如海^{1,2} 郭秀艳^{3,4} 凌晓丽^{5,6} 郑丽^{**3,4} 姜珊^{**2,7} Zoltan Dienes⁸

(¹江苏理工学院教育学院, 常州, 213001) (²华东师范大学心理与认知科学学院, 上海, 200062)

(³复旦大学老龄研究院, 上海, 200433) (⁴教育部哲学社会科学实验室——复旦大学国家发展与智能治理综合实验室, 上海, 200433)

(⁵山东师范大学心理学部, 济南, 250358) (⁶山东省脑科学与心理健康重点实验室, 济南, 250358)

(⁷上海政法学院政府管理学院, 上海, 201701) (⁸School of Psychology, University of Sussex, Brighton, BN1 9RH)

摘要 研究考察了简单循环网络是否与人类一样具备远距离规则跨长度迁移能力。实验中, 采用与人类实验一致的符合远距离规则的平仄声调序列对简单循环网络进行训练, 使其仅在长度为 10 的序列上进行学习, 并评估其在未训练长度 (8 与 12) 序列上的迁移表现。结果显示, 训练后的简单循环网络在所有测验长度上均显著优于未训练模型, 表现出对远距离规则的学习与迁移能力。进一步与人类行为数据的匹配分析显示, 简单循环网络能有效模拟人类在此类任务中的典型行为特征。结果支持简单循环网络具备跨长度迁移能力, 为理解神经网络中的规则迁移机制提供了理论依据, 同时表明其在推动人工智能系统实现类似人类的复杂规则学习与知识迁移方面具有巨大潜力。

关键词 简单循环网络 远距离规则 迁移

1 引言

学习是人类核心的认知能力之一。近几十年来的研究发现, 学习并非一定是有意识的, 也可以在无意识的条件下进行, 即人类存在一种内隐学习的能力。内隐学习是有机体通过与环境接触, 无目的、自动地获得事件和客体间结构关系的过程 (郭秀艳, 2003; 郭秀艳, 杨治良, 2002; Reber, 1967)。内隐学习在人类认知的许多方面发挥着重要的作用, 例如语言的获得 (Kovacs & Endress, 2014; Nazzi et al., 2000)、动作技能的掌握 (孙鹏等, 2022; Miyamoto et al., 2020) 等。

内隐学习研究领域的一个核心问题是内隐习得的究竟是何种类型的知识。一些研究者认为被试能够内隐习得底层的抽象规则 (Reber, 1967, 1969), 而另外一些研究者则认为被试在这一过程中仅限于对组块、具体样例等表面特征的学习 (Brooks, 1978; Jamieson & Mewhort, 2011; Rohrmeier & Rebuschat, 2012)。在内隐学习研究的早期阶段, 学者们主要采用限定状态人工语法规则生成实验材料。然而, 由于限定状态语法依赖相邻元素之间的序列顺序

(即相邻规则)使得学习过程中底层的抽象语法规则与表面相似性特征难以完全区分。因此, 参与者的学习效应常常被解释为对局部表面特征组块化的结果, 而非真正掌握了底层的抽象规则 (Rohrmeier & Rebuschat, 2012)。

近期, 研究者进一步借助可良好分离规则与组块等表面特征的远距离规则对内隐学习的抽象性进行探究 (Jiang et al., 2012; Ling et al., 2022), 并特别关注了被试是否能够将所习得的结构规则迁移至不同长度的新材料中, 以检验其是否掌握了更具抽象性的“无限结构模式” (infinite patterning) (Dienes & Longuet-Higgins, 2004)。迁移是指个体将已习得的结构规则知识泛化到表面特征不同的新材料中的过程。近年来, 迁移能否发生已被视为内隐学习中是否习得了抽象规则的关键判据 (戴惠等, 2018; Pothos, 2007; Wang et al., 2021)。探究个体能否在长度不同但远距离规则结构相似的新材料中实现迁移, 不仅能够证明内隐学习的抽象性, 也回应了 Chomsky (1956) 关于自然语言生成特征的核心观点——人类的认知机制超越了限定状态系统, 具备更高层次的结构抽象能力, 能够凭借有限的语言

* 本研究得到教育部哲学社会科学研究重大课题攻关项目 (22JZD044)、教育部人文社会科学研究青年基金项目 (23YJC190014, 24YJC190047)、上海市教育科学研究项目 (C2025010) 和国家自然科学基金面上项目 (32171072) 的资助。

** 通讯作者: 郑丽, E-mail: zheng_l@fudan.edu.cn; 姜珊, E-mail: jiangs988@126.com

DOI:10.16719/j.cnki.1671-6981.20260203

输入,掌握语法规则并生成无限的新句子。目前,远距离规则研究的相关证据表明,人类能够无意识地掌握非邻近元素之间的结构依赖关系,并在材料变化条件下保持迁移与应用。例如,姜珊与关守义(2018)在控制了组块等表面特征的基础上,采用汉语声调的远距离水平映射规则构建声音串材料。结果发现,被试能够内隐掌握声调间的远距离映射关系,并且能将这些规则迁移到新的声音串长度条件下。可见,内隐学习能够超越具体的表面特征,习得可迁移的抽象结构。

模拟研究不仅有助于验证认知心理学中关于内隐学习过程的理论假设,还能为人工智能(AI)系统的发展提供启示。自20世纪80年代以来,神经网络模型在内隐学习的计算建模中发挥了核心作用。众多序列学习模型相继问世,包括自联想器、多层前馈网络以及改进的反向传播算法(Cleeremans & Dienes, 2008; Cleeremans & McClelland, 1991)。其中,Elman提出的简单循环网络(simple recurrent network, SRN)被视为神经网络研究的一个里程碑,它通过引入上下文记忆机制,使神经网络能够模拟复杂的时间依赖性与递归结构,从而推动神经计算从静态模式识别迈向动态认知模拟(Elman, 1991)。

SRN的记忆机制尤其契合内隐学习的研究需求,特别是在模拟个体如何在无意识状态下捕捉和迁移时间先后的结构方面。已有研究表明,SRN在特定条件下能在无外显规则提示的情境中习得远距离规则(李菲菲,刘宝根,2018;Kuhn & Dienes, 2008; Timmermans & Cleeremans, 2000)。可见,SRN所具备的分布式概率表征能力,能够在无须明确语法或语义标签的前提下,隐式捕捉序列中的深层规律。这种特性使其成为探索无意识抽象知识获取机制的重要工具。

SRN的基本结构由输入层、隐含层和输出层构成,关键在于隐含层状态不仅由当前输入决定,还融合了来自前一时间点中隐含状态的反馈输入。这一机制通过“上下文单元”或隐含-隐含连接实现,构成了一个具有时间延迟的内部循环,从而赋予网络对先前状态的持续访问能力(Elman, 1991)。Magnuson与Luthra(2024)进一步指出,这种递归结构本质上构成了反馈系统,使得SRN在信息处理过程中具备强烈的交互性,而非传统的前馈式系统。换言之,SRN的激活状态源于当前输入与历史内部

状态的融合,这种机制在语言处理、语音识别和序列预测等领域展现出与人类学习过程高度相似的特征(Magnuson & Luthra, 2024)。

随着人工智能的发展,SRN的核心思想已被现代深度学习架构所继承和扩展,如长短期记忆神经网络模型(long short-term memory, LSTM)与Transformer模型在序列建模中的成功,正是对SRN“记忆+反馈”机制的重构与强化(Mienye et al., 2024)。在此背景下,评估SRN在无意识结构学习和知识迁移中的潜能,不仅具有理论价值,也为下一代AI系统的设计提供了认知启示。尽管已有研究显示SRN在固定长度输入上可习得远距离规则(李菲菲,刘宝根,2018),但其是否具备类似人类那样的远距离规则迁移能力,目前仍缺乏实证支持。基于此,本研究以姜珊与关守义(2018)提出的声调水平远距离映射任务为基础,系统考察SRN在无外显规则提示情境中,对远距离抽象结构的内隐学习能力及其在材料长度变化条件下的迁移能力。

2 方法

2.1 材料

SRN使用与人类实验相同的训练和测验声调串:144个训练串和96个测验串(姜珊,关守义,2018)。每个训练声调串长度均为10,其中,前五个声调在平仄上与后五个构成远距离映射关系,具体来说,第一个与第六个形成“平-仄”声调对应,第二个与第七个形成“平-仄”声调对应,以此类推。测验声调串则分为三组:一个为训练过的长度(长度为10),两个为未训练长度(长度为8,其中前四个与后四个声调分别构成四对“平-仄”远距离映射关系;长度为12,其中前六个与后六个声调分别构成六对“平-仄”远距离映射关系)。每组包含32串声调串,其中16串为符合远距离规则的合法串,16串为不符合远距离规则的非法串。非法串中有两对不符合“平-仄”声调映射关系。每种长度测验串中的合法串和非法串的表面特征均进行了平衡控制。

2.2 模型结构

SRN的输入单元与输出单元数量相同。两种输入/输出单元分别代表声调串中的两种声调类型(平声与仄声)。激活单元设为0.9,不激活单元设为0.1。一串完整的声调串可通过将每个元素对应的单元激活向量拼接起来表示为一个向量。

表 1 模拟中使用的参数取值范围

网络参数	
学习率	.1, .3, .5, .7, .9
动量	.1, .3, .5, .7, .9
隐含单元数量	5, 10, 15, 30, 60, 120
循环次数	100

模拟实验覆盖了表 1 中给出的所有参数取值的组合。这些参数选择基于已有的人工语法学习模拟研究(李菲菲, 刘宝根, 2018; Kuhn & Dienes, 2008)。总计形成了 150 个不同的模型。每个模型运行 25 次, 每次使用不同的初始随机权重(Nguyen-Widrow, 1990), 共生成 3750 次模拟。所有网络均使用 Matlab 神经网络工具箱模拟完成。

2.3 程序

训练阶段: 对于每一串训练串, 每次呈现一个元素, 激活相应输入单元, 网络学习预测序列中下一个元素。采用反向传播算法, 通过目标值与输出值之间的误差更新权重。每次呈现完整训练串后, 网络的上下文单元会被重置为 0。每个网络学习 144 个训练串, 顺序随机排列。

测验阶段: 测验阶段中, 网络被呈现 3 组共 96 串测验串。三个测验组的呈现顺序在不同网络间平衡, 每组中的 32 串测验串随机呈现。网络预测测验串中下一个元素的能力作为其表现指标。表现通过计算目标向量 t 与输出向量 o 之间夹角的余弦值(COS)评估(Altmann & Dienes, 1999; Dienes, 1993; Dienes et al., 1999; Kinder, 2000)。COS 值越大, 表示表现越好。测验阶段中网络仍保留学习功能, 因为大脑中的学习机制在测验时不会自动关闭。

2.4 数据分析

首先, 将对模型在训练长度(长度为 10)上的学习效果进行评估, 采用配对样本 t 检验探索训练前后网络区分符合与不符合远距离规则材料的能力是否存在显著差异, 以判断模型是否成功习得了长度为 10 的远距离规则。

其次, 进一步检验模型能否将已习得的远距离规则迁移到未训练长度(长度为 8 与 12)的材料中, 采用同样的方法评估迁移条件下的学习表现。

最后, 匹配 SRN 模型与人类被试在区分能力(discrimination performance)上的表现, 以探讨 SRN 能否有效模拟人类在远距离规则学习与迁移任务中的行为模式。人类数据参考姜珊与关守义(2018)的

研究结果。具体而言, 为了衡量网络与人类在区分能力上的一致性, 计算每个条件下的 Z 值: 将符合语法字符串的平均余弦相似度(COS)减去不符合语法字符串的平均 COS 值, 并除以合并标准差。 Z 值接近 0 表示表现接近随机水平, Z 值越大表明区分能力越强。对于人类行为数据, 以实验组与控制组被试 d' 差值的平均数(M)及其标准误(SE)为基础, 采用 $M \pm 1SE$ 的方式定义人类学习效应的典型区间(李菲菲, 刘宝根, 2018)。

本研究将参考李菲菲与刘宝根(2018)以及 Kuhn 与 Dienes(2008)的方法匹配模型和人类之间的表现。李菲菲与刘宝根(2018)发现, 在 150 个 SRN 模型中, 分别有 15 个和 18 个模型的区分能力落入人类行为数据的区间范围内。本研究将采用类似的方法, 通过绘制模型与人类区分能力的分布关系, 评估 SRN 在模拟内隐学习与跨长度迁移中的有效性。

3 结果

3.1 模拟的表现

表 2 展示了各组网络的平均 Z 分数。配对样本 t 检验结果显示, 训练过的网络在所有三个测验组中的表现均显著优于未训练网络: 对于长度为 8 的测验组: $t(149) = 6.45, p < .001, \text{Cohen's } d = .53$; 对于长度为 10 的测验组: $t(149) = 9.72, p < .001, \text{Cohen's } d = .79$; 对于长度为 12 的测验组: $t(149) = 10.56, p < .001, \text{Cohen's } d = .86$ 。可见, SRN 成功地区分了合法串与非法串, 并能将所学规则迁移到未学习过的长度。

3.2 模型与人类行为数据的匹配

本研究进一步评估了 SRN 模型在远距离规则学习与迁移任务中的特征性表现是否与人类数据相一致, 进而推断模型是否可以作为人类内隐学习的解释模型。分析结果显示, SRN 模型中有相当数量的模型在各测验任务中表现落入了人类数据所覆盖的区间范围(即均值 ± 1 标准误以内)。在长度为 10 的测验任务中, 共有 23 个模型表现符合人类被试的典型学习效应。进一步, 在跨长度迁移的测验任务中,

表2 三组测验中训练与未训练网络的平均Z值以及人类实验组与控制组的表现

测验长度	简单循环网络训练结果				人类数据			
	训练		未训练		实验组		控制组	
	M	SE	M	SE	M	SE	M	SE
长度 8	.08	.01	.01	.01	.19	.08	.01	.04
长度 10	1.26	.13	.01	.00	.34	.08	.07	.06
长度 12	.19	.02	.01	.00	.18	.06	-.04	.05

长度为 8 的测验中有 35 个模型，长度为 12 的测验中有 38 个模型表现符合人类典型效应。

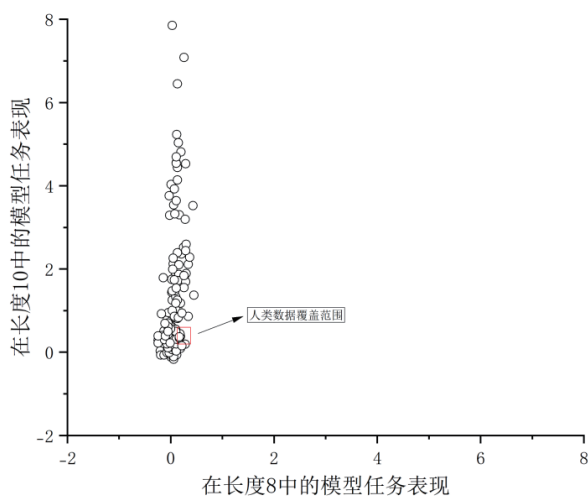


图1 模型在长度 10 和 8 上的学习效应及其与人类数据的匹配

注：横坐标为长度 8 的模型任务表现，纵坐标为长度 10 的模型任务表现，以 Z 分数来量化矩形代表人类被试的典型成绩 ($M \pm 1SE$)。

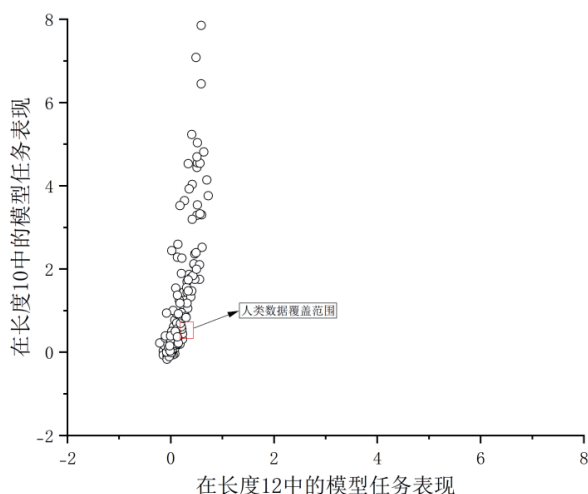


图2 模型在长度 10 和 12 上的学习效应及其与人类数据的匹配

注：横坐标为长度 12 的模型任务表现，纵坐标为长度 10 的模型任务表现，以 Z 分数来量化。矩形代表人类被试的典型成绩 ($M \pm 1SE$)。

进一步分析是否存在这样的模型：其在长度为 10 的测验任务中落入人类被试典型学习效应区间，

同时在跨长度迁移测验任务 (长度为 8 或 12) 中也落入该效应区间。分析结果显示，有 5 个模型在长度为 10 和 8 的测验任务中均落入人类被试的典型学习效应区间 (见图 1)，另有 4 个模型在长度为 10 和 12 的测验任务中同时落入人类效应区间 (见图 2)。这些结果表明，在一定的参数配置下，SRN 模型不仅能够学习远距离规则，还能实现规则的跨长度迁移，较好地模拟人类在内隐学习远距离结构及其迁移过程中的典型行为特征。

4 讨论

本研究旨在探讨 SRN 在人类远距离规则学习和迁移能力上的模拟表现。结果显示，与前人研究一致，SRN 能够显著区分远距离规则的合法串和非法串 (李菲菲，刘宝根，2018)。考虑到当前研究中合法串和非法串的表面特征差异已经得到控制，这种区分能力可被认为与刺激材料的表面特征无关。进一步，SRN 还能将学到的远距离规则迁移至未曾训练过的不同长度新测验串中，且 SRN 表现出的迁移能力不受新测验材料长度的限制，不论是比训练材料更短的测验串还是更长的测验串，只要远距离映射结构不变，其均能表现出显著的迁移能力。与人类行为数据的匹配分析则发现，SRN 在一些模型中的学习和迁移表现与人类被试的表现相似。可见，SRN 确实具备模拟人类规则学习和迁移表现的能力，且其学习到的并在新情境中迁移运用的是抽象的远距离规则，而非某些依赖于表面特征的规律。

值得注意的是，既往研究通过检验 SRN 能否习得特定远距离位置之间的映射关系，来评估对远距离规则的学习效果 (李菲菲，刘宝根，2018；Kuhn & Dienes, 2008)。而远距离规则被认为存在值 - 值的表征和变量 - 变量表征两种可能的表征方式。前者指的是被试可能习得某个位置和另一个位置元素之间存在的简单值 - 值对应关系；后者则是指被试获得了变量和变量之间的复杂规则而非特定

的值和值之间的简单对应 (Dienes & Longuet-Higgins, 2004)。对具体远距离位置之间映射关系的习得, 尚不足以判断 SRN 习得的是值 - 值的具体对应关系, 还是更为抽象和复杂的远距离规则。本研究通过检验 SRN 对长度不同但远距离映射结构相似的新材料的迁移能力, 证明了 SRN 习得的是变量 - 变量层面的抽象对应关系, 即 Marcus(2001) 所谓的“变量上的运算”, 显示出其对非局部依赖结构的学习能力。

本研究不仅从模拟角度验证了 SRN 在远距离规则学习与跨长度迁移任务中的有效性, 也进一步支持了其作为人类内隐学习与迁移认知机制建模工具的巨大潜力。正如 Reed(2019) 所指出, 连接主义模型不仅是人工智能技术路径, 也可作为理论工具服务于认知心理学研究。通过将 SRN 等神经网络模型作为理论工具, 研究者能够深入探究人类认知中如何无意识地加工复杂结构, 同时促进 AI 算法在处理结构性知识、推理与迁移学习等方面的优化。已有研究显示, SRN 通过递归连接整合历史信息, 可在无外显规则提示下形成对输入结构的内隐敏感性 (李菲菲, 刘宝根, 2018; Wang et al., 2021)。本研究进一步揭示了 SRN 不仅可以在无外显规则提示的情境下习得远距离规则, 还能够将其迁移至长度不同的新材料中, 表现出一定的结构学习与迁移能力, 这一结果为其对人类内隐学习及迁移的模拟能力提供了新的支持。可见, SRN 不仅能够模拟任务表现, 还为揭示人类认知系统中规则学习与迁移的潜在机制提供了强有力的计算模型支持。

SRN 能够习得远距离规则以及实现跨长度迁移, 关键在于其上下文单元所构成的时间延迟反馈机制。SRN 的隐含层激活状态不仅受到当前输入的影响, 也会整合前一时刻的隐含层激活状态 (Elman, 1991; Magnuson & Luthra, 2024), 从而实现了有序列历史信息的持续记忆与动态整合, 这一机制赋予网络对长距离依赖关系进行加工的能力, 使得 SRN 可以建立非相邻元素之间的结构依赖关系的概率表征。从远距离结构规则学习和迁移的角度来看, SRN 在训练阶段通过预测下一个元素来持续调整内部权重, 从而建立对“平仄”类别之间远距离映射规则的抽象加工能力。当测验阶段输入长度发生变化时, 这种类别层面的抽象规则仍能被激活, 实现对未见长度结构的成功迁移, 表明 SRN 对底层结构的迁移已超越具体位置与长度限制。

综上所述, 研究表明 SRN 能够学习抽象的远

距离规则, 并能够将所习得的规则有效迁移至不同长度的结构条件, 展现出模拟人类内隐学习非局部依赖关系的潜力。整体结果支持将 SRN 作为建模人类结构性学习过程的有力工具, 并为未来进一步融合认知科学与人工智能领域的研究奠定了理论基础。

参考文献

- 戴惠, 朱传林, 刘电芝. (2018). 内隐知识具有抽象性吗?——来自内隐序列学习迁移的证据. *心理学报*, 50(9), 965-974.
- 郭秀艳. (2003). 内隐学习和缄默知识. *教育研究*, 12, 31-36.
- 郭秀艳, 杨治良. (2002). 内隐学习的研究历程. *心理发展与教育*, 17(3), 85-90.
- 姜珊, 关守义. (2018). 汉语声调水平映射规则的内隐学习及其长度迁移效应. *心理学探新*, 38(4), 326-332.
- 【Jiang S., Guan S. (2018). Implicit Learning and Length Transfer Effect of Chinese Tonal Symmetry. *Psychological Exploration*, 38(4), 326-332.】
- 李菲菲, 刘宝根. (2018). 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. *心理科学*, 41(4), 796.
- 孙鹏, 李雪晴, 张庆云, 尚怀乾, 凌晓丽. (2022). 睡眠对知觉与动作序列内隐学习离线巩固效应的影响. *心理学报*, 54(12), 1467-1480
- Altmann, G. T. M., & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science*, 284, 875.
- Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169-211). Lawrence Erlbaum.
- Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 9, 113-124.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120(3), 235.
- Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 396-421). Cambridge University Press.
- Dienes, Z. (1993). Computational models of implicit learning. In D. Berry & Z. Dienes (Eds.), *Implicit learning: Theoretical and empirical issues*. Lawrence Erlbaum Associates.
- 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(1), 53-82.
- Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformation be implicitly learned? *Cognitive Science*, 28, 531-558.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-225.
- Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology*, 64, 209-216.
- Jiang, S., Zhu, L., Guo, X., Ma, W., Yang, Z. & Dienes, Z. (2012). Unconscious structural knowledge of tonal symmetry: Tang poetry redefines limits of implicit learning. *Consciousness and Cognition*, 21, 476-486.
- Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models. *Psychological Research - Psychologische Forschung*, 63(2), 95-105.
- Kovacs, A. M., & Endress, A. D. (2014). Hierarchical processing in seven-month-

- old infants. *Infancy*, 19(4), 409–425.
- Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition*, 106, 184–206.
- Ling, X., Sun, P., Zhao, L., Jiang, S., Lu, Y., Cheng, X., & Zheng, L. (2022). Neural basis of the implicit learning of complex artificial grammar with nonadjacent dependencies. *Journal of Cognitive Neuroscience*, 34(12), 2375–2389.
- Magnuson, J. S., & Luthra, S. (2024). Simple recurrent networks are interactive. *Psychonomic Bulletin and Review*, 32(3) 1–9.
- Marcus, G. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. MIT Press.
- Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9), 517.
- Miyamoto, Y. R., Wang, S., & Smith, M. A. (2020). Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nature Neuroscience*, 23(3), 443–455.
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43(1), 1–19.
- Nguyen, D., & Widrow, B. (1990). *Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights*. In Proceedings of the 1990 International Joint Conference on Neural Networks (Vol. 3, pp. 21–26), IEEE.
- Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, 133(2), 227.
- Reber, A. S. (1967). Implicit learning of artificial grammar. *Journal of Verbal Learning and Verbal Behaviour*, 6, 855–863.
- Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81(1), 115–119.
- Reed, S. (2019). Building bridges between AI and cognitive psychology. *AI Magazine*, 40(2), 17–28.
- Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning of music: What do we know today. *Topics in Cognitive Science*, 4, 525–553.
- Timmermans, B., & Cleeremans, A. (2000). Rules vs. statistics in biconditional grammar learning: A simulation based on Shanks et al. (1997). *Proceedings of the Twenty-Second Annual Conference of Cognitive Science Society*, 22, 512–517.
- Wang, L., Feng, Y., Fu, Q., Wang, J., Sun, X., Fu, X., & Yi, Z. (2021). A dual simple recurrent network model for chunking and abstract processes in sequence learning. *Frontiers in Psychology*, 12, 587405.

A Simulation Study of Cross-Length Transfer of Non-Adjacent Dependencies based on Simple Recurrent Networks

Zhang Ruhai^{1,2}, Guo Xiuyan^{3,4}, Ling Xiaoli^{5,6}, Zheng Li^{3,4}, Jiang Shan^{2,7}, Zoltan Dienes⁸

(¹College of Education, Jiangsu University of Technology, Changzhou, 213001) (²School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062) (³Fudan Institute on Ageing, Fudan University, Shanghai, 200433) (⁴Ministry of Education Philosophy and Social Sciences Laboratory - Comprehensive Laboratory of National Development and Intelligent Governance, Fudan University, Shanghai, 200433) (⁵Faculty of Psychology, Shandong Normal University, Jinan, 250358) (⁶Shandong Provincial Key Laboratory of Brain Science and Mental Health, Jinan, 250358) (⁷School of Government, Shanghai University of Political Science and Law, Shanghai, 201701) (⁸School of Psychology, University of Sussex, United Kingdom, BN1 9RH)

Abstract This study investigated whether simple recurrent networks (SRNs) could learn abstract non-adjacent dependencies and generalize them across sequences of different lengths. Building on previous findings that highlight the human ability to unconsciously acquire and transfer non-adjacent structural dependencies (Jiang & Guan, 2018), the present research aimed to evaluate whether SRNs could similarly internalize such structures and apply them flexibly to novel sequences. This could provide insights into the modeling of implicit learning and transfer processes.

SRNs were trained using tonal sequences derived from the “level/oblique” (ping/ze) categorizations, reflecting prior cognitive categories available to human participants. The network architecture included input, hidden, and output layers, with feedback loops enabling temporal integration. A total of 150 SRN models were constructed by systematically varying three key parameters: the number of hidden units (5, 10, 15, 30, 60, or 120), learning rate (.1, .3, .5, .7, or .9), and momentum (.1, .3, .5, .7, or .9). Each model was subjected to 25 independent training sessions initialized with random weights, resulting in 3,750 simulations.

Models were exclusively trained on sequences of length 10 and subsequently tested on sequences of lengths 8, 10, and 12. Learning performance was assessed using cosine similarity scores between the network outputs and target sequences, and z-scores were calculated to quantify discrimination performance between grammatical and ungrammatical strings. Human benchmark data were sourced from Jiang and Guan (2018). Human learning

effects were defined as the mean difference in discrimination index (d') between experimental and control groups, framed within ± 1 standard error (SE) as the typical human performance range.

The results revealed that trained SRNs significantly outperformed untrained models across all sequence lengths, confirming the successful acquisition of the nonlocal dependencies. Furthermore, a notable number of SRNs exhibited discrimination performance that fell within the typical human range: 35 models for the 8-element test set, 23 models for the 10-element set, and 38 models for the 12-element set. Notably, several SRN models demonstrated consistent human-like behavior across both trained and novel lengths. Specifically, five models aligned with human data in both the 8- and 10-length tests, and four models aligned in both the 10- and 12-length tests.

These findings suggest that, under specific parameter settings, SRNs were capable not only of learning abstract non-adjacent dependencies but also of transferring them flexibly to structurally novel sequences. Compared to earlier studies, which primarily demonstrated SRNs' learning fixed-length correspondences, this study highlighted SRNs' potential to acquire variable-variable mappings, reflecting the concept of "operations over variables" proposed by Marcus (2001). This indicated a more abstract level of generalization than previously reported, showing that SRNs may implicitly capture underlying structural principles rather than merely memorizing surface patterns.

The introduction of tonal category labels (ping/ze) as non-terminal markers likely provided a cognitive scaffold that facilitated the abstraction of structural rules. This approach mirrored how human learners leveraged prior conceptual knowledge to enhance statistical learning, offering insights into the interaction between prior knowledge and the acquisition of novel patterns.

From a computational modeling perspective, the results implied that SRNs, despite their architectural simplicity, could mimic key aspects of human implicit learning, including structural abstraction and transfer. Furthermore, the ability of some SRNs to perform comparably to humans under specific conditions supported the use of SRNs as viable models for studying the cognitive mechanisms underlying implicit knowledge acquisition and generalization.

This study broadly contributed to bridging cognitive psychology and artificial intelligence research. The findings suggested that relatively simple recurrent architectures possess latent capacities for flexible generalization, an essential feature for developing AI systems capable of human-like learning. Additionally, by examining SRNs' behavior on non-finite-state structures resembling those found in natural language, the study enriched our understanding of how internal memory dynamics support the processing of complex structures.

Overall, the present work advanced the field by systematically demonstrating the capacity of SRNs for abstract, nonlocal dependency learning and structural transfer. It provides empirical evidence for their utility in modeling implicit learning processes and contributes to theoretical foundations of future cognitive and AI modeling efforts.

Key words simple recurrent networks, non-adjacent dependencies, transfer