



面向智能时代的人机合作心理模型

刘焯^{1,2*}, 汪亚珉³, 卞玉龙⁴, 任磊^{5,6}, 禚宇明^{1,2}

1. 中国科学院心理研究所脑与认知科学国家重点实验室, 北京 100101

2. 中国科学院大学心理学系, 北京 100049

3. 首都师范大学心理学院, 北京 100048

4. 山东大学计算机科学与技术学院, 济南 250101

5. 北京航空航天大学自动化科学与电气工程学院, 北京 100191

6. 北京航空航天大学精准医疗高精尖创新中心, 北京 100191

* 通信作者. E-mail: liuye@psych.ac.cn

收稿日期: 2017-11-05; 接受日期: 2018-02-06; 网络出版日期: 2018-04-10

国家自然科学基金 (批准号: 61632004, 61773379, 61572057)、中国博士后科学基金 (批准号: 2017M620284)、北京市教育委员会面上项目 (批准号: KM201610028018) 和北京市教育委员会市属高校创新能力提升计划 (批准号: TJSH20161002801) 资助项目

摘要 最近 10 年, 人工智能和人机交互研究突飞猛进, 在学术界和产业界都掀起了热潮. 不管是突破人工智能的瓶颈, 还是实现和谐、自然的人机交互, 都迫切需要符合人类认知机理和心理规律的人机交互心理模型. 建构面向智能时代的人机合作心理模型, 不但可以解释、模拟和预测人类的行为, 而且可以指导相应的智能系统和人机交互设计. 本文回顾和总结了以往的人机交互模型和认知计算模型, 基于最近 10 年的实验证据和研究发现, 试图拓展和深化已有的认知模型, 建构面向智能时代的人机合作心理模型, 为智能时代的人机交互研究提供理论基础.

关键词 人工智能, 人机合作, 人机交互, 心理模型, 感知, 认知, 动作

1 人机交互概述

近年来随着计算机科学和人工智能的飞速发展, 计算机已经可以在某些领域接近甚至超越人类智能, 例如 Google 公司开发的 AlphaGo 战胜人类顶尖围棋选手, Hanson Robotics 公司开发的情感机器人已经可以与人类产生情感互动, 谈笑风生, 甚至获得沙特阿拉伯公民身份. 计算机越来越智能化和情感化, 随之而来, 人与计算机的交互方式也在悄然发生着变化. 从最早的命令行, 到图形用户界面 (例如, windows, menu, icons, pointing devices, WIMP), 再到当前蓬勃发展的虚拟现实和增强现实技术, 以及各种智能语音交互、手势交互、眼动追踪交互、脑机接口等, 人机交互的方式越来越丰富, 技术越来越精湛, 但是人机交互的理论研究却相对比较滞后, 现在开展人机交互研究时所基于的仍然是将近 40 年前的理论.

引用格式: 刘焯, 汪亚珉, 卞玉龙, 等. 面向智能时代的人机合作心理模型. 中国科学: 信息科学, 2018, 48: 376-389, doi: 10.1360/N112017-00225

Liu Y, Wang Y M, Bian Y L, et al. A psychological model of human-computer cooperation for the era of artificial intelligence (in Chinese). Sci Sin Inform, 2018, 48: 376-389, doi: 10.1360/N112017-00225

早在 1983 年, Card 等^[1,2] 出版了最早的一本有关人机交互的著作《人机交互的心理学》, 其中不但提出了人机交互的概念, 而且提出了至今仍被大多数研究者奉为经典的人类处理器模型 (model human processor, MHP 模型), 及其衍生版本 GOMS 模型 (goals, operations, methods, and selection rules). 在过去的 30 多年里, 这些模型虽然为人机交互研究和设计提供了必要的理论指导, 但是随着人机交互方式的逐渐变革, 现有的模型已经不能满足当前的需求. 计算机越来越接近人类的智能和情感处理水平, 人与计算机的交互追求更加自然、和谐, 力图更加符合人类的认知和行为习惯, 因此人与计算机的交互越来越接近人与人之间的交互. 但是以往的人机交互模型仍然停留在旧交互模式的阶段, 无法满足当前人机交互方式多种多样的局面. 面对这一发展趋势, 迫切需要一个新的理论模型来指导相关的研究和设计工作.

本文在总结以往人机交互、人工智能的概念模型基础上, 梳理认知心理学、认知神经科学最近 10 年有关人类认知和人际交互的研究发现, 试图提出一个面向智能时代的人机合作心理模型, 并对这个心理模型进行计算建模. 我们希望这个心理模型可以在未来更好地指导智能时代的人机交互研究和设计, 更好地解决当前模型无法回答的交互设计问题.

2 已往人机交互模型和理论

Card 等^[1] 提出的 MHP 模型为人机交互的心理模型奠定了基础. 在 MHP 的基础上, Card 等^[1] 又提出了 GOMS 模型、击键水平模型, 以针对不同的任务类型. 随后, 基于人类认知规律建立的人机交互和人工智能心理模型不断涌现, 大致可以分为两类: 一类是基于信息加工取向的符号模型, 一类是联结主义神经网络模型^[3].

符号模型又可以分为基于阶段论的过程模型和产生式系统模型^[3]. 像 MHP、GOMS 和最近由中国心理学家和计算机科学家联合提出的感知、记忆和判断的认知计算模型 (computational cognition model of perception, memory and judgment, PMJ 模型)^[4] 都属于符号模型中的过程模型. 最有代表性的产生式系统模型有自适应控制理论^[3,5] 和执行过程/交互控制模型^[3,6,7]. 基于联结主义神经网络的心理模型起源于 20 世纪 80 年代. 受到当时计算机科学和人工智能领域联结主义兴起的影响, 心理学家也开始采用神经网络来模拟人类的认知加工, 主要集中在学习与记忆、知识表征、言语理解等领域^[3]. 其中最具有代表性的是并行分布式模型和 Bayes 网络模型^[3].

由于基于阶段论的过程模型更加适合人机交互设计, 也更加符合人机交互的心理加工过程, 因此本文将在以往过程模型的基础上提出本文的新模型. 下面将简单地介绍已往主要的过程模型, 并且回顾过去 30 年心理学领域相关的理论和思想.

2.1 人类处理器模型

MHP 模型是一个用于模拟和预测人类行为绩效的工程心理学模型. 该模型基于信息加工取向, 将人类的心理加工过程概括为感知、认知、动作 3 个处理器, 一些用于工作记忆、长时记忆存储的单元, 以及这些处理器和存储单元之间相互联结的通路^[1]. 除此之外, 该模型提出了这些处理器和存储单元所遵循的一系列原则, 这些原则最终被定义为一些确定的参数, 用于分析信息加工系统的经济性原则^[1]. 例如, 该模型所使用的记忆原则包括存储的项目容量、项目的延迟时间、编码的类型 (包括听觉的、视觉的和语义的), 加工器的参数只有环路循环周期时间^[1]. Card 等^[1] 在以往研究的基础上, 为感知、认知、动作这 3 个处理器设定了大致相似的循环周期时间.

MHP 模型为模拟和预测人机交互的行为绩效提供了一个简单、有效, 而且非常具有操作性的模型, 至今仍然在人机交互领域起到一定的指导作用. 但是, 该模型也存在明显的缺陷. 首先, 该模型在 20 世纪 80 年代提出, 当时心理学家对人类认知机理的认识还非常有限, 导致该模型对 3 个处理器的环路循环周期时间的设定非常随意和主观. 其次, 该模型将认知过程假设为一个序列加工的过程, 虽然存在动作到感知的反馈通路, 但是信息加工的过程完全遵循着继时序列加工方式, 这与后来大量认知心理学的研究发现不相符. 最后, 该模型将人机交互的心理过程区分为 3 个处理器, 这一阶段性的划分虽然非常具有指导意义, 但是该模型对 3 个处理器内容的描述却非常粗糙和笼统. 这种粗糙的处理器模型已经不能满足当前人机交互中越来越复杂的场景和任务.

2.2 GOMS 模型

GOMS 模型是基于 HMP 模型的扩展模型, 同样由 Card 等^[1]提出. 该模型是人机交互和界面设计领域中最常用的信息加工模型, 尤其在可用性测试领域被广泛应用. 该模型可以对用户在给定的任务参数下的行为绩效进行定性和定量的预测, 可以用于对特定的人机交互系统和设计进行评测.

GOMS 模型包含 4 类组成部分, 分别是一系列目标、操作、实现目标的方法, 以及用于选择实现目标方法的选择规则^[1]. 目标是指用户试图实现的任务的最终状态. 操作是指用户为实现目标所执行的知觉、运动或者认知加工, 这些加工操作的执行对于改变用户的心理状态或者影响任务环境非常必要. 实现目标的方法是指为了实现目标所进行的一系列步骤. 由于实现目标的方法可能有很多种, 因此需要通过选择规则这一控制结构, 为了实现目标而选择恰当的方法.

GOMS 模型将人机交互任务和过程分解为一系列步骤, 在评测一个交互系统和界面时具有很好的可操作性. 同时, 与 HMP 模型一样, GOMS 模型假设用户的信息加工时间由感知、认知和动作 3 个阶段的加工时间组成, 可以使研究者快速简单地估计交互的绩效. 但是 GOMS 模型假设用户对操作十分熟练, 而且其目标由明确的最终目标和一系列子目标组成, 用户会根据不同目标选择恰当的操作方法, 这导致该模型仅仅适用于熟练的使用者, 而无法预测初学者的试误过程. 同时, GOMS 模型更适合用于模拟和预测任务目标简单明确, 并且目标可以清晰分解的交互任务和场景, 但是随着当前人机交互方式和任务的复杂度大幅增加, 该模型也面临着无法适应新需求的困境.

2.3 PMJ 模型

近年来, 中国心理学家联合计算机科学领域的专家, 合作提出 PMJ 模型, 将认知过程归纳为感知、记忆和判断这 3 个阶段, 并提出认知加工的多条通路和一系列认知加工策略^[4]. 该模型试图回答认知计算性的基础问题, 并将认知过程的 3 个阶段分别对应于计算流程的分析、建模和决策, 对人类认知进行计算建模^[4].

该模型分为认知模型和计算模型两部分. 认知模型包括感知、记忆和判断 3 个阶段. 每个阶段完成一定的加工任务, 并且认知加工的各阶段之间具有多条加工通路, 主要包括快速加工通路、精细加工通路、反馈加工通路. 认知系统在处理信息时, 会依据任务难度、任务目标来动态地选择加工通路^[4].

PMJ 模型已经在认知心理学研究、工程心理学、神经网络建模、计算机图像处理、计算机图像情感预测等领域得到了一定的应用和验证, 可以很好地指导计算机模拟人类认知进行可视媒体的认知计算^[4]. 但是, PMJ 模型本质上是一个面向人工智能的心理模型, 没有涉及信息的输入和输出, 以及人与计算机如何交互的问题. 虽然该模型对人类心理加工过程的三阶段描述, 对人机交互具有一定的指导意义, 但是还需要更加具有针对性的模型来模拟当前复杂的人机交互机制.

2.4 分布式认知

分布式认知的思想最早于 20 世纪 80 年代提出^[8], 该理论观点认为, 人类认识世界, 或者说人类的认知是一种包含了认知主体与其所在的环境的综合系统, 人类的认知不再单纯地依赖于认知主体, 还涉及到各种环境因素, 包括认知情境、认知对象、工具和周围环境^[8]. 分布式认知思想充分考虑了人类认知环境的复杂性, 以及认知活动本身的复杂性, 认知并不仅仅是在对内部表征的加工, 而是内部表征与外部表征共同起作用的结果^[9]. 人类在处理复杂信息时, 不可能脱离开当前外部环境信息的支持, 我们借助于各种图表、视觉图像、实际场景来帮助我们思考和解决问题. 因此, 当思考人机交互的心理过程时, 不能把人与计算机割裂看待, 而是应该把人与计算机作为一个认知的整体系统来看待, 人与计算机互为交互情境.

2.5 多模态并行交互

早期的信息加工系统观点, 认为认知系统加工的信息是抽象的符号, 也就是信息通过各种感觉器官从外界进入认知系统时, 都从模态化的信息转换为了统一的非模态信息. 前面提到的大多数基于阶段论的过程模型没有对认知过程加工的信息进行明确的描述. 但是基本上都隐含着抽象符号的假设. 而最近 30 多年的研究发现, 人类表征外部世界的方式可能是多种多样的, 既存在非模态化的抽象语义信息, 也存在依赖各自感觉器官的模态特异性信息^[10]. 也就是说, 有些从感觉器官获得的信息被表征为抽象的符号, 而有些表征仍然保留了模态特异性的特点, 并不是抽象的符号. 另外, 人类的心智不是脱离开身体独立存在, 而是深植于身体之中, 身体构造、神经结构、感觉运动器官的活动方式都会影响我们对世界的认识, 影响思维的方式. 这种思想取向被称为具身认知 (embodied cognition)^[11, 12], 其中最具有代表性的理论观点是扎根认知理论 (grounded cognition)^[13].

扎根认知理论认为, 头脑表征的知识并不全都是抽象的符号, 其中大部分知识是对原本呈现给感觉器官的信息的直接表征, 也就是说我们存储的知识是对外界信息的生动模拟^[13], 是以各个感觉通道和模态特有的方式存储, 因此各个模态之间存在相互的影响, 也存在相互的分离. 随着当前人机交互所涉及的模态越来越丰富, 具身认知和扎根认知理论观点在人机交互领域的重要性也日益凸显.

3 人机合作心理模型

在已有的过程模型 (包括 MHP, GOMS, PMJ 模型等) 和分布式认知、多模态并行交互思想的基础上, 本文针对交互主体 (人或计算机) 提出了一个新的人机交互模型: 人机合作心理模型 (human computer cooperation model, HCCM 模型), 如图 1 所示.

该模型的基本假设如下.

假设一. 人与计算机的交互本质上可以类比为人与人的交互, 具有和人与人交互相似的属性和规律. 计算机是具有与人相似认知功能的智能体. 人与计算机的交互不是序列继时发生的信息处理过程, 更不是计算机被动接受人类指令的过程, 而是两个交互主体在各自先验知识和动机的驱动下主动交互的过程.

假设二. 不管是人类, 还是计算机, 其信息处理系统都包括感知、认知和动作 3 个功能模块. 对于计算机, 其感知、认知和动作 3 个模块分别主要实现多模态交互数据的获取和处理、加工和存储, 以及输出和反馈功能. 每个加工模块都要完成各自的信息加工任务, 并为其他模块提供信息输入, 或者接受其他模块的信息输出和反馈. 每个模块都有各自的主要功能, 各模块相互配合, 完成整个信息处理系统的交互任务.

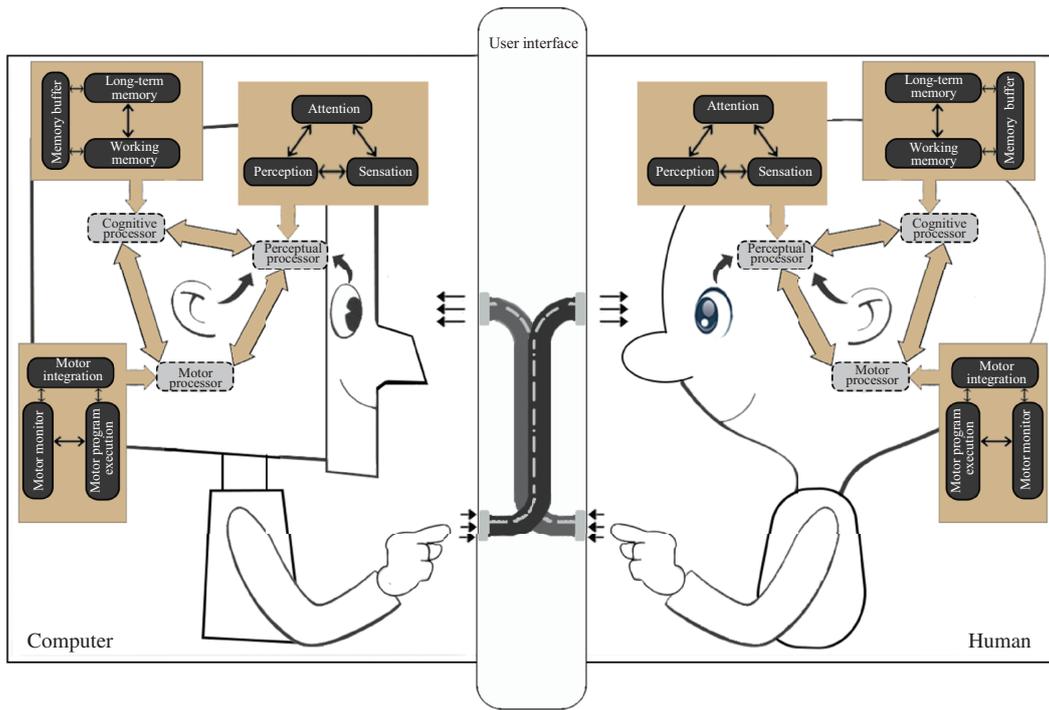


图 1 (网络版彩图) 人机合作心理模型示意图

Figure 1 (Color online) A schematic diagram of a psychological model on human-computer cooperation

假设三. 人机之间的交互和合作主要通过 3 个通道来实现: (1) 计算机的动作模块输出的信息进入人类的感知模块; (2) 人类的动作模块输出的信息进入计算机的感知模块; (3) 计算机与人类的认知模块之间存在异质同构性, 虽然计算机与人存在诸多差异, 但是两者的认知功能非常类似, 具有相同的功能模块. 不但人类与计算机的感知和动作模块之间具有相互契合的特性, 而且计算机具有和人一样理解其他个体意图、愿望、想法的能力.

假设四. 人机交互是多模态并行的、分布式交互的过程. 不管是人类还是计算机, 都会同时接受来自多个感觉通道的信息, 同时接受来自周围环境、交互场景中的各种信息. 这个过程持续不断的进行, 认知模块负责将所有信息进行整合和加工. 下面将逐渐展开论述人机合作心理模型的具体内容.

3.1 人机合作的本质

根据假设一, 我们认为, 随着计算机智能水平的提升, 人与计算机的交互将具有和人与人交互相似的特性. 计算机在人机交互过程中, 一方面被动地接受用户的动作处理器产生的指令, 一方面通过摄像头、麦克风、多导生理仪等终端主动采集用户的面部表情、身体姿态、语音、外周生理变化等信息. 这两方面的信息构成计算机感知处理器的信息输入. 在计算机的认知模块中, 计算机拥有和人类相似的加工子模块. 这些子模块不但具有和人类相似的信息处理能力, 而且也具有形成心理理论^[14]的能力, 在自身感知模块输入的信息基础上, 理解和预测交互对象的意图、愿望、信念、情感等心理状态.

心理理论是心理学家在研究黑猩猩是否具有推测他人心理状态的能力时, 首先提出的一个概念, 是一种对自己和他人心理状态的理解能力. 发展出心理理论的儿童才能正确地区分自己的想法与他人想法的不同, 才能正确地认识他人的视角, 理解他人的愿望、意图、信念等. 心理理论的形成是人与人

之间正常交往、相互理解、协同工作的基础^[15]。

3.2 3个处理器

假设二提出人和计算机这两个交互主体都是分为感知、认知和动作3个处理器,或者模块。每个模块又都各自包含了3个子模块。感知模块包括知觉、注意和感觉3个子模块;认知模块包括长时记忆、工作记忆和记忆缓冲3个子模块;动作模块包括动作整合、动作监控和动作执行3个子模块。

3.2.1 感知处理器

感知处理器包括感觉、知觉与注意3个子模块,负责接受、选择和初步加工输入的信息^[16~19]。如果说感觉相当于智能计算机的传感器,知觉相当于模式识别,那么注意无疑是传统人机交互模型难以处理的部分。人的注意功能巧妙地帮助人的信息加工系统进行信息的过滤和筛选,实现某些意图性的交互过程,未来和谐的人机交互模型有必要具备这一功能模块。为此,本文在新的人机交互模型中加入了注意这一功能模块。

感觉处理器。感觉作为认知加工过程的初级阶段,完成对客观事物的个别属性的认识,不同的感觉通道获取不同属性的信息。在对等的人机合作模式中,计算机也应该具备与人一样的感官能力,如视觉、听觉、触觉等。生理心理学的研究揭示,感觉过程主要由各感官的感受器接收相应的信息并进一步将这些信息进行神经编码后传递至相应的大脑皮层感觉区^[20]。计算机的各种信息输入在加工器层面上与人的感觉一致,而且在人机合作的交互模式中,计算机的感觉还将进一步地扩展和丰富,以实现与人的多感官信息加工能力相匹配。

知觉处理器。知觉是人脑对客观事物的整体属性的认识,是人赖以识别事物的心理加工机制。认知心理学对人的知觉的发生机制的探讨大致可以分为整体加工与特征加工两大类。以 Gestalt 学派开始的整体知觉认为人是先识别知觉对象的整体(完形, gestalt)然后再加工部分的特征^[21]。Chen^[22]提出的拓扑知觉理论进一步提出知觉开始于大范围的拓扑属性识别。相应地,特征加工说从 Marr 等^[23]的认知计算模型到 Treisman 等^[24]的特征整合模型,则认为人对知觉对象的识别开始于基本特征的提取,然后再通过特征的整合来实现对知觉对象的识别。在人机交互层面上,知觉的计算模型更具竞争力。知觉的计算理论将知觉过程视为特征提取与特征整合两个过程,这与计算机处理视觉信息的方式很相似。尤其是腹侧与背侧通路的揭示^[25]似乎提示这种通用的信息加工器是可能的。

然而,就人的知觉加工机制而言,有些特殊的知觉加工机制,比如面孔识别的认知神经加工机制的揭示,让通用的信息加工器假定复杂了很多。面孔识别与物体识别,如果基于人机共用的知觉加工机制,这两种识别的机制应该是一致的,大脑皮层的处理区域也应该是同一的,然而事实表明,人脸识别与物体识别在认知神经机制上是分开的,人脸识别主要激活梭状回面孔区这一皮层区域^[26],而物体识别则涉及多个相关区域。模块化是人与计算机在处理信息时共有的加工机制,参照人的知觉加工机制,也应该为计算机设定专门对象的识别模块,而不是通用的知觉处理器。如果从人机共生的角度来看,计算机知觉处理器的机能化发展是一个必然趋势,因为人的知觉加工抛弃机能模块的可能很小,毕竟人是进化而来的。作为一个大胆的预测,未来的人机交互可能需要基于一个人与计算机共有过往经验来进行,包括一些人类的特别成长经验(痛苦的、纠结的、快乐的),否则这种人机交互将无法达到预想的境界。

注意处理器。注意是人脑加工的一种特别机制,它的功能就是让心理活动在一定时间内指向并集中到某项活动上。注意并不是认知过程特有的,在情绪情感以及意志过程中也同样需要注意的参与。

注意负责对信息进行择取、实现对加工的监管, 以及实现对心理过程的随意调节^[27~29]. 在人机交互层面上, 注意可以被界定为个体对交互过程中信息的选择、监督与调节. 在未来的人机交互模型里, 计算机如何成长发展出这种类似人的注意功能的模块是现有计算算法理论及机器学习理论的一大挑战. 从人机交互的角度分析, 一个可能的设想是计算机将来要有专门的概念集来完成这种价值判断功能.

感知处理器与认知处理器、动作处理器的双向交互. 从人与环境交互的角度来看, 人的感知处理器初步实现了人对环境的基本认识和理解. 这种认识与理解保证了个体对环境刺激的简单、直接反应, 但还不足以保证个体对环境刺激产生间接反应, 或者是预测性反应. 直接反应与间接反应是简单与复杂的反映层次关系. 感知觉是被界定为人脑对客观事物的直接反应, 而只有思维才是间接反映. 思维是人脑对客观事物的概括的间接的反映, 是以感知觉加工为基础, 对事物进行分析比较、归类、抽象与概括后形成的一种判断性经验. 思维的完成既需要感知信息作为基础, 又需要记忆能力作为保证. 感知觉处理器与思维处理器 (认知处理器) 之间的交互的结果是实现了人脑对客观事物的本质属性的概括化的反应. 在心理学里, 这种从感知到思维加工, 并借助记忆对加工的结果加以存储的过程既是对环境的观察 (感知觉) 与学习 (记忆与思维) 的过程. 从人机合作的层面来看, 感知处理器与认知处理器的交互能够保证人机双方的互相认识与理解. 这种认知与理解是建立在一种观察与学习的基础之上的. 感知觉处理器直接从对方获取信息, 并认识信息的直接涵义, 这些涵义传递到认知处理器后被进一步分析、学习并形成一种预期返回给感知觉处理器进行验证.

与动作处理器的双向交互. 感知处理器可以直接启动动作处理器的初步反应, 即识别了事物的基本属性后做出直接的应对. 这种直接启动快速反应的机制与人适应环境的关键威胁有关, 比如损害机体生存的刺激, 火、水、悬崖, 或者蛇、蜘蛛等. 心理学中通常把由感知觉刺激直接启动的反应叫快反应, 与经过思维 (认知处理器) 处理分析后做出的慢反应相对应^[30]. 从人机合作的层面来看, 这种感知觉直接导致的动作反应是属于个体对环境中优先级最高的刺激所做出的一种反应, 与普通的经过认知处理器后做出的慢反应是协调进行的.

3.2.2 认知处理器

认知处理器包含 3 个子处理器: 工作记忆、记忆缓冲和长时记忆, 负责存储信息, 并对信息进行整合和精细加工.

工作记忆. 当前, 认知心理学家普遍认为, 工作记忆是一个容量有限的信息加工系统, 用来暂时保持和存储信息, 是连接感知觉、长时记忆和动作输出的加工平台, 是人类各种复杂认知过程, 包括推理、思维、决策等过程的基础^[31~33].

在本文的心理模型中, 工作记忆是认知处理器的重要组成部分, 负责实现认知处理器与感知处理器和动作处理器的相互联系. 工作记忆接收来自感知处理器的信息, 并将加工后的信息传递给动作处理器; 与此同时, 工作记忆也接受来自感知处理器的反馈, 并给感知处理器提供反馈. 工作记忆对信息的加工和处理主要表现为以下 4 方面:

(1) 来自感知处理器的信息进入工作记忆后, 通过编码和存储机制^[34~37] 将信息暂时存储, 使这些信息处于人的意识之中.

(2) 通过工作记忆的信息更新机制^[38,39], 使处于人的意识之中的信息保持动态的变化, 以确保思维过程可以顺利进行.

(3) 通过工作记忆的信息巩固机制^[40~43], 工作记忆通过记忆缓冲器, 将信息传递给长时记忆, 实现不断的学习和经验积累.

(4) 通过记忆缓冲器, 工作记忆提取长时记忆中的信息到当前认知加工中, 与感知处理器传入的信息进行整合.

记忆缓冲器. 记忆缓冲器是处于工作记忆与长时记忆之间的信息传递的桥梁. 大量关于阅读的研究表明, 在人处理大量信息的复杂任务中, 由于工作记忆的容量有限, 大量信息虽然不再处于工作记忆之中, 但是仍然处于相对活跃的状态, 在加工过程中更容易被激活和提取^[44,45]. 因此, 心理学家将这些信息所处的加工阶段称为长时工作记忆, 在本文提出的心理模型中, 将这一认知阶段统称为记忆缓冲器. 另外, 记忆缓冲器在工作记忆中的信息巩固过程中也发挥重要作用, 通过记忆缓冲器, 工作记忆中加工的信息转化为长时记忆中存储的信息.

长时记忆. 长时记忆中存储着人以往学习和经历的所有知识和经验. 其中不但包括可以用语言表达出来的陈述性知识, 也包括人无法用语言表达出来的内隐记忆 (无意识的记忆, 尤其以情绪性记忆为主)、程序性知识. 陈述性知识又包括语义记忆和情景记忆^[46]. 前者是指人类积累的各种知识和经验, 例如从书本中学习的那些知识, 后者则是个体自身的经历, 发生在特定的时间和空间. 每个人的长时记忆就像这个人特有的数据库, 里面存储了个体之前学习和经历的所有知识、事件和经历.

在本文的心理模型中, 长时记忆主要通过记忆缓冲器与工作记忆相互传递信息. 但是, 长时记忆也可以直接接收来自感知处理器的信息, 同时也可以直接给感知处理器提供反馈, 由于这些信息没有经过工作记忆模块, 所以这些信息的加工是无意识水平的加工, 没有进入人的当前意识中. 长时记忆也可以给动作处理器输出信息, 同时也从动作处理器接受反馈, 同样这些信息加工也处于无意识水平.

3.2.3 动作处理器

动作处理器包括 3 个子处理器: 动作整合、动作执行和动作监控, 负责根据认知处理器的输出结果采取相应的动作反应^[47,48].

动作整合. 动作整合子模块负责动作的选择、规划和动作序列的生成. 该子模块中存在一个核心功能区, 即动作缓冲区. 工作记忆和动作缓冲均涉及动作表征, 两者的分工在于, 工作记忆支持运动和运动序列的中央符号表征 (非运动的空间和语言), 以及其他信息的表征 (如任务目标). 动作缓冲区负责的运动表征涉及具体的运动选择、编程和特征处理, 可能还包含有关动作如何调整到效应器的生物力学信息^[49], 以及连续运动如何能顺利地整合到运动序列中^[50]. 然而, 它不需要进一步来产生实际的运动^[51].

本模型中动作缓冲区的运动表征有 3 个信息加工层级, 分别是动作图式、动作程序和动指令表征^[48,52]. 假设主体需要做出一套比较复杂的动作反应, 其动作序列一般需要一连串的动作表征^[53,54]. 第 1 层的动作图式表征代表是动作表征最抽象的水平, 会根据当前信息选择一个已有的动作图式或者构造一个新的动作表征^[55]. 第 2 层的动作编程代表了一种具体的动作表征. 最低层级是动作指令, 代表最简单的动作片段和特征. 这些片段的组合可以完成一个特定的动作程序. 较高水平的加工为低水平的动作表征提供特征和动作参数.

动作执行. 动作执行子模块负责动作指令的执行和输出. 如上文所述, 动作整合子模块的作用下, 动作表征会具化为动作序列以及动作的低层特征, 然后将其存储在动作缓冲器中. 一旦动作缓冲器包含所有必要的信息, 动作处理器开始执行运动缓冲内容, 将具体动作指令发送到效应器上, 最终输出动作. 动作执行过程包含一个回路, 如果动作缓冲器包含几个连续运动的表征, 动作处理器通过运动回路循环来产生每个动作, 相对严格地执行动作缓冲区的内容.

动作监控. 动作监控子模块负责动作执行结果反馈的收集和评价, 从而帮助调整动作序列的执行. 动作执行的结果会通过外部信息得到反馈, 也可以通过内部认知获得反馈. 然后, 监控子模块收集这

些反馈, 评估执行的每一个动作, 并将结果发送到认知系统. 认知系统会根据反馈结果, 给感觉加工阶段提供线索, 调整认知加工或者修正长时记忆中存储的知识, 调整动作选择或执行, 从而调整最终的输出结果与预期目标一致.

动作模块与认知模块. 动作模块与认知模块之间主要是通过动作整合子模块中的动作缓冲区相互联系. 信息进入认知模块的工作记忆后, 认知模块会做出判断和决策, 处理后的信息进入动作模块中的动作缓冲区. 工作记忆和动作缓冲是用于产生动作序列的两个部分, 它们相互重叠^[52]. 以往也有研究得出应从功能上将动作缓冲从工作记忆中分离出来. 在动作开始前, 动作整合子模块进行动作程序激活和动作参数选择, 并将获得的动作特征结果加载到工作记忆和动作缓冲中^[56]. 这些动作特征会进一步加工成单个序列原素, 并在序列执行之前将被存储到动作缓冲区中. 进入动作缓冲器中的信息会通过动作处理器最终生成指令, 并通过效应器 (输出设备) 执行.

另外, 如前文所述, 动作监控子模块获得的反馈信息会进一步传输到认知模块中, 从而实现对动作的评价、调整和记忆.

动作模块与感知模块. 信息除了通过感知、认知到达动作模块外, 还存在另一种更高效的信息加工模式^[52], 可以由感觉模块中抽取的刺激特征直接自动地启动动作图式, 并直接作用于效应器做出反应^[57,58]. 这使得感知模块直接作用于动作模块, 跳过了认知模块的复杂信息加工程序.

4 基于人机合作心理模型的智能交互计算模型

基于人机合作心理模型的智能交互计算模型如图 2 所示.

4.1 面向多模态融合并行交互的用户界面

基于人机合作心理模型的智能交互计算模型, 将突破传统人机交互中类似“乒乓球传递”的交互模式, 而趋向于两个认知主体之间类人-人的自然交互模式. 相应的, 将摒弃传统 WIMP 范式中不符合人类自然交互习惯的交互设备与界面表征元素, 代之以更为自然的视觉、听觉、触觉、手势等多模态交互模式. 在多模态交互过程中, 将立足于多模态的融合与并行交互机制, 以更贴近于人-人之间自然交互的行为特征作为用户界面的设计原则.

4.2 感知计算模块

感知计算模块基于感觉、注意、知觉等人机合作心理模型要素, 主要支持多模态交互数据的获取与处理, 并基于此提取用户的交互任务意图, 进一步交给认知计算模块对任务进行解算. 感知计算模块首先对多模态交互数据进行预处理, 包括图形图像、语音、力反馈、机械动作等信号数据, 然后对多模态交互数据做融合计算处理, 并提取特征. 在此基础上, 结合注意选择策略规则库, 可对注意力进行选择. 同时, 结合交互语义知识库, 交互任务意图进行识别提取, 并进一步输入认知计算模块中的工作记忆数据存储器. 在大量交互感知计算基础上, 可对感知交互任务语义知识库进行机器学习与知识更新. 对于不需要经过认知计算模块进一步处理的、具有明确对应动作反应输出的任务, 结合快反应规则库, 由快反应刺激触发器输入动作反应计算模块进行处理.

4.3 认知计算模块

认知计算模块主要接收感知计算模型传来的交互任务与情境信息, 进一步对交互任务进行处理和

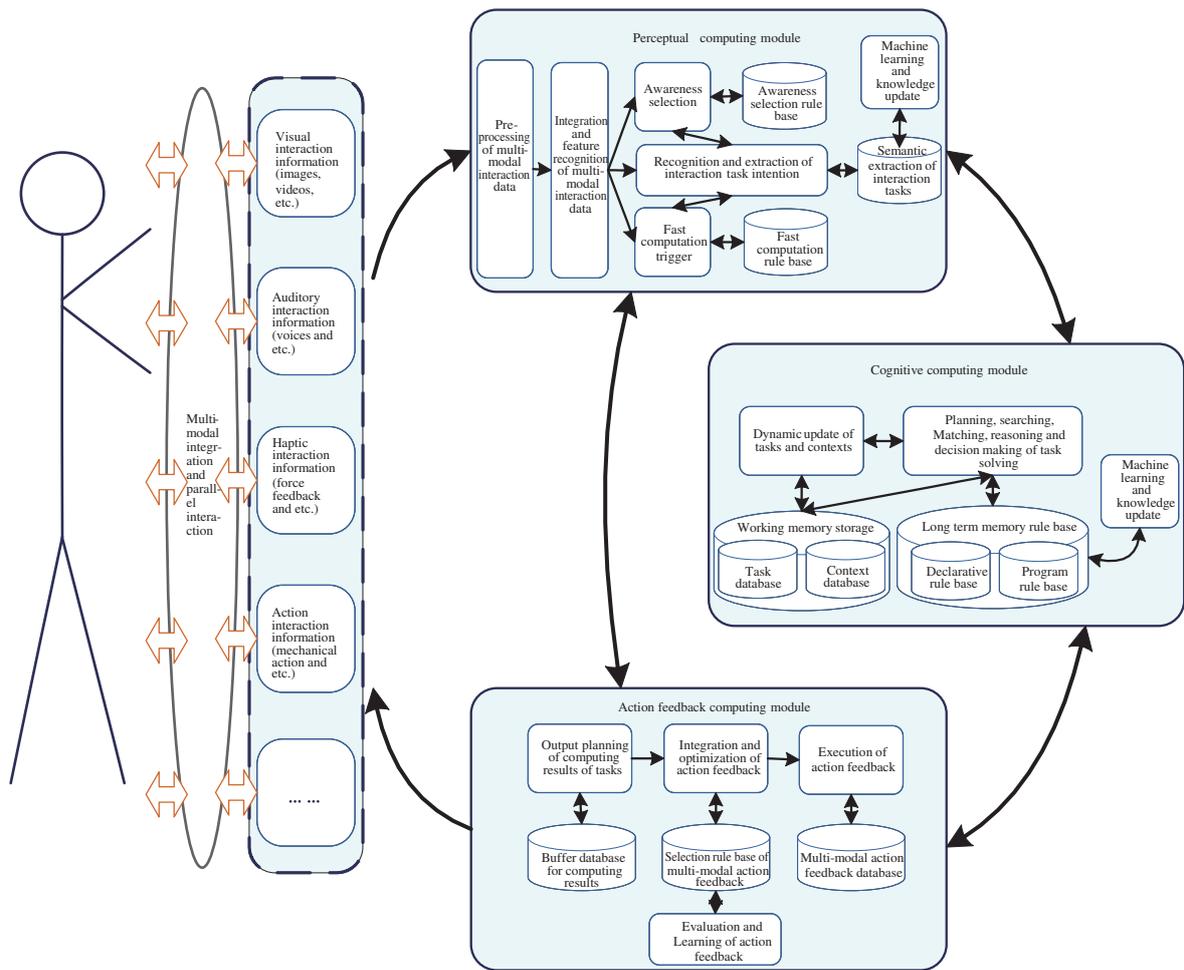


图 2 (网络版彩图) 基于人机合作心理模型的智能交互计算模型

Figure 2 (Color online) A computational model of intelligent interaction based on the psychological model on human-computer cooperation

解算, 得到结果后交给动作反应计算模块做反馈输出处理. 认知计算模块中的工作记忆数据存储单元用于暂存交互任务相关信息, 主要包括两类: 交互任务相关参数数据、交互情境上下文相关信息数据, 并由任务与情境信息管理器进行动态更新维护. 任务求解同时依赖长时记忆知识库的支持, 其中主要包括两类知识经验: 陈述性知识库、程序性知识库. 在工作记忆数据存储与长时记忆知识库的共同支持下, 认知计算模块对任务求解进行规划、搜索、匹配、推理、决策等. 随着交互任务求解的计算与迭代积累, 知识库基于机器学习模型进行自学习与知识更新.

4.4 动作反应计算模块

动作反应计算模块主要用于将任务处理结果以合适、自然的表征形式反馈给用户, 例如将计算结果以图像配以语音解释的形式作为反馈. 基于计算结果缓冲数据库, 动作反应计算模块将对任务计算结果以何种组合方式、何种顺序输出等进行规划. 然后, 基于多模态动作反应选择知识库, 将要反馈输出的计算结果分别优化选取合适的输出模态, 并将选定的多模态反应输出进行整合. 在此基础上, 动

作反应执行器基于多模态动作反应函数库, 调用相应的反应输出函数, 生成视觉、听觉、触觉等多模态融合的动作反应结果, 反馈给用户. 基于大量历史动作反应评价数据, 采用机器学习模型可自学习与优化更新多模态动作反应选择的知识库, 使得不同计算结果的对应动作反应模态选择更为贴近不同用户的个性化交互习惯.

4.5 应用举例

为说明该模型的适用性和可推广性, 本文结合一个人机交互实例——智能对话系统来对该模型进行更直观的阐述. 智能对话系统需要实现人与机器之间的自然对话, 特别是计算机对用户意图和状态需要形成准确理解并做出有效反馈. 在使用该系统时, 比如用户对计算机给出口语指令“我不知道今天要干什么, 怎么办?”, 信息首先进入计算机的感知处理器: 语音和语义特征被提取 (感觉子处理器), 然后识别并转化成文本等数据 (知觉子处理器). 另外, 系统接收到指令后, 系统也可以通过控制摄像头捕捉用户的表情和动作, 从而对感知模块的输入信息进行筛选 (注意子处理器). 识别后的文本、表情和动作信息进入认知处理器: 计算机需要对语义、情感信息进行持续分析 (工作记忆), 并从知识库 (长时记忆) 中查找相关的语义等内容 (记忆缓冲), 然后进行相似性分析等多种分析, 最终形成对语义、用户需求和状态的理解 (工作记忆). 分析后的信息进入动作处理器: 根据语义分析结果对需要给出的反馈或反应进行分析和整合, 形成动作指令 (动作整合), 最终通过语音、文字或配合其他形式为用户做出回答 (动作执行), 如“你心情如果不太好的话, 我觉得以下建议可能对你有帮助...”. 如果信息输出时发生错误, 系统会自查并进行提示或纠正 (动作监控). 输出的信息被用户所接收, 也经过感知、认知和动作模块对信息进行分析 and 评估, 最后再给出下一个问题或下一步动作指令, 最终共同形成一个有效的活动方案, 实现人-机之间的良好合作.

5 总结与展望

本文在已有的过程模型和分布式认知、多模态并行交互思想的基础上, 针对人和计算机这两个交互主体提出了面向智能时代的人机合作心理模型. 该模式试图解决当前人机交互研究所面临的主要问题, 包括计算机智能水平的大幅提升, 人机交互任务的复杂性提高, 以及人机交互方式的复杂多样性.

与此同时, 本文提出的人机合作心理模型仍存在一定的局限, 有待未来进一步完善和改进.

首先, 当前的模型未能涉及到人类的情绪情感领域. 情绪情感是人类心理活动不可或缺的一部分, 是驱动人类生存发展的原动力. 在人机交互过程中, 人类不可避免地伴随着各种心理需求、情绪情感状态的变化. 随着情感计算、生理计算、可穿戴计算等技术手段的发展, 未来计算机可以更加准确、即时地识别人类的情感, 有望使人机交互更加生动、和谐自然.

其次, 虽然基于过去 30 多年心理学行为实验和认知神经科学实验的发现, 经过系统梳理和概括总结, 提出了这一模型, 但是尚未有针对性地开展人机交互实验对这一模型进行系统验证. 这也是未来需要重点努力的方向.

第三, 目前这一模型还只是概念模型, 与 MHP 和 GOMS 模型相比, 缺少量化的参数描述各个加工阶段的反应时间, 还不是一个可计算的模型. 未来将结合心理学实验的发现, 将模型的各个模块和通路进行定量的分析.

最后, 虽然在模型的基本假设部分融合了多模态并行、分布式认知的思路, 但是在模型的具体实现上, 未能提出更加细化的模型架构. 未来有望在这方面进一步深化和完善模型.

参考文献

- 1 Card S K, Moran T P, Newell A. *The Psychology of Human-Computer Interaction*. Hillsdale: Lawrence Erlbaum Associates, 1983
- 2 Card S K, Moran T P, Newell A. The model human processor: an engineering model of human performance. In: *Handbook of Perception and Human Performance*. New York: John Wiley and Sons, 1986
- 3 Chen W F, Zhao G L, Liu Y, et al. Advances in computational models of human cognition in China. In: *China Association for Science and Technology, eds. 2012–2013 Report on Advances in Psychology*. Beijing: China Science and Technology Press, 2014. 85–100 [陈文锋, 赵国联, 刘焯, 等. 认知计算模型研究进展. 见: 中国科学技术协会, 编. 2012–2013 心理学学科发展报告. 北京: 中国科学技术出版社, 2014. 85–100]
- 4 Fu X L, Cai L H, Liu Y, et al. A computational cognition model of perception, memory, and judgment. *Sci China Inf Sci*, 2014, 57: 032114
- 5 Anderson J R, Lebiere C J. The atomic components of thought. *J Math Psychol*, 1995, 45: 917–923
- 6 Meyer D E, Kieras D E. A computational theory of executive cognitive processes and multiple-task performance: part 1. basic mechanisms. *Psychol Rev*, 1997, 104: 3–65
- 7 Meyer D E, Kieras D E. A computational theory of executive cognitive processes and multiple-task performance: part 2. accounts of psychological refractory-period phenomena. *Psychol Rev*, 1997, 104: 749–791
- 8 Hollan J, Hutchins E, Kirsh D. Distributed cognition: toward a new foundation for human-computer interaction research. In: *Human-Computer Interaction in the New Millennium*. New York: ACM Press, 2001. 75–94
- 9 Zhang J. The nature of external representations in problem solving. *Cogn Sci*, 1997, 21: 179–217
- 10 Barsalou L W. Grounded cognition: past, present, and future. *Top Cogn Sci*, 2010, 2: 716–724
- 11 Barsalou L W. Grounded cognition. *Annu Rev Psychol*, 2008, 59: 617–645
- 12 Ye H S. Embodied cognition: a new approach in cognitive psychology. *Adv Psychol Sci*, 2010, 18: 705–710 [叶浩生. 具身认知: 认知心理学的新取向. *心理科学进展*, 2010, 18: 705–710]
- 13 Barsalou L W, Simmons W K, Barbey A, et al. Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci*, 2003, 7: 84–91
- 14 Zimmer C. How the mind reads other minds. *Science*, 2003, 300: 1079–1080
- 15 Atkinson L, Slade L, Powell D, et al. Theory of mind in emerging reading comprehension: a longitudinal study of early indirect and direct effects. *J Exp Child Psychol*, 2017, 164: 225–238
- 16 Hergenhahn B R. *An Introduction to the History of Psychology*. Boston: Brooks/Cole Pub, 1986
- 17 Solso R L, Maclin O H, Maclin M K. *Cognitive Psychology*. 7th ed. Beijing: Peking University Press, 2005 [Solso R L, Maclin O H, Maclin M K. 认知心理学 (第7版). 北京: 北京大学出版社, 2005]
- 18 Wang S, Wang A S. *Cognitive Psychology*. Beijing: Peking University Press, 2006 [王甦, 汪安圣. 认知心理学 (重排本). 北京: 北京大学出版社, 2006]
- 19 Gazzaniga M S, Ivry R B, Mangun G R, et al. *Cognitive neuroscience: the biology of the mind*. New York: W. W. Norton & Company, 2009
- 20 Carlson N R. *Foundations of physiological psychology*, 6th ed. *Am J Psychol*, 1980, 80: 653
- 21 Koffka K. Perception: an introduction to the gestalt-theorie. *Psychol Bull*, 1922, 19: 531–585
- 22 Chen L. The topological approach to perceptual organization. *Visual Cogn*, 2005, 12: 553–637
- 23 Marr D, Poggio T, Hildreth E C, et al. A computational theory of human stereo vision. *Readings Cogn Sci*, 1988, 204: 534–547
- 24 Treisman A M, Gelade G. A feature-integration theory of attention. *Cogn Psychol*, 1980, 12: 97–136
- 25 Ungerleider L G, Haxby J V. ‘what’ and ‘where’ in the human brain. *Curr Opin NeuroBiol*, 1994, 4: 157–165
- 26 Kanwisher N, McDermott J, Chun M M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neuroscience*, 1997, 17: 4302–4311
- 27 Petersen S E, Posner M I. The attention system of the human brain: 20 years after. *Annu Rev Neurosci*, 2012, 35: 73–89
- 28 Broadbent D E. *Perception and Communication*. Oxford: Pergamon Press, 1958
- 29 Posner M I, Snyder C R, Davidson B J. Attention and the detection of signals. *J Exp Psychol*, 1980, 109: 160–174

- 30 Ledoux J E. Emotion circuits in the brain. *Annu Rev Neurosci*, 2000, 23: 155–184
- 31 Baddeley A D. The episodic buffer: a new component of working memory? *Trends Cogn Sci*, 2000, 4: 417–423
- 32 Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci*, 2003, 4: 829–839
- 33 Diamond A. Executive functions. *Annu Rev Psychol*, 2013, 64: 135–168
- 34 Chen C C, Wu J K, Lin H W, et al. Visualizing long-term memory formation in two neurons of the drosophila brain. *Science*, 2012, 335: 678–685
- 35 Fell J, Axmacher N. The role of phase synchronization in memory processes. *Nat Rev Neurosci*, 2011, 12: 105–118
- 36 Fusi S, Abbott L F. Limits on the memory storage capacity of bounded synapses. *Nat Neurosci*, 2007, 10: 485–493
- 37 Eichenbaum H. A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci*, 2000, 1: 41–50
- 38 Frankland P W, Bontempi B. The organization of recent and remote memories. *Nat Rev Neurosci*, 2005, 6: 119–130
- 39 Nadel L, Hardt O. Update on memory systems and processes. *Neuropsychopharmacol*, 2011, 36: 251–273
- 40 Mcgaugh J L. Memory — a century of consolidation. *Science*, 2000, 287: 248–251
- 41 Miranda M, Bekinschtein P. Plasticity mechanisms of memory consolidation and reconsolidation in the perirhinal cortex. *Neuroscience*, 2018, 370: 46–61
- 42 Tronson N C, Taylor J R. Molecular mechanisms of memory reconsolidation. *Nat Rev Neurosci*, 2007, 8: 262–275
- 43 Edelson M, Sharot T, Dolan R J, et al. Following the crowd: brain substrates of long-term memory conformity. *Science*, 2011, 333: 108–111
- 44 Nader K, Hardt O. A single standard for memory: the case for reconsolidation. *Nat Rev Neurosci*, 2009, 10: 224–234
- 45 Delaney P F, Ericsson K A. Long-term working memory and transient storage in reading comprehension: what is the evidence? comment on Foroughi, Werner, Barragán, and Boehm-Davis (2015). *J Exp Psychol Gen*, 2016, 145: 1406–1409
- 46 Klein S B, Cosmides L, Tooby J, et al. Decisions and the evolution of memory: multiple systems, multiple functions. *Psychol Rev*, 2002, 109: 306–329
- 47 Bi L Z, Shang J X, Gan G D. Modeling driver lane changing control with the queuing network-model human processor. In: *Proceedings of International Conference on Machine Learning and Cybernetics*, Xi'an, 2012. 830–834
- 48 Sadeghipour A, Kopp S. Gesture processing as grounded motor cognition: towards a computational model. *Procedia Soc Behav Sci*, 2012 32: 213–223
- 49 Andresen D R, Marsolek C J. Effector-independent and effector-dependent sequence representations underlie general and specific perceptuomotor sequence learning. *J Motor Behav*, 2012, 44: 53–61
- 50 Mattys S L. Stress versus coarticulation: towards an integrated approach to explicit speech segmentation. *J Exp Psychol Hum Perc Perform*, 2004, 30: 397–408
- 51 Shea C H, Kovacs A J, Panzer S. The coding and inter-manual transfer of movement sequences. *Front Psychol*, 2011, 2: 52
- 52 Verwey W B, Shea C H, Wright D L. A cognitive framework for explaining serial processing and sequence execution strategies. *Psychon Bull Rev*, 2015, 22: 54–77
- 53 Bo J, Seidler R D. Visuospatial working memory capacity predicts the organization of acquired explicit motor sequences. *J NeuroPhysiol*, 2009, 101: 3116–3125
- 54 Verwey W B. Diminished motor skill development in elderly: indications for limited motor chunk use. *Acta Psychol*, 2010, 134: 206–214
- 55 Ruitenberg M F L, Abrahamse E L, De Kleine E, et al. Context-dependent motor skill: perceptual processing in memory-based sequence production. *Exp Brain Res*, 2012, 222: 31–40
- 56 Rosenbaum D A. Human movement initiation: specification of arm, direction and extent. *J Exp Psychol Gen*, 1980, 109: 444–474
- 57 Hommel B, Müsseler J, Aschersleben G, et al. The theory of event coding (TEC): a framework for perception and action planning. *Behav Brain Sci*, 2001, 24: 849–878
- 58 Lien M C, Mccann R S, Ruthruff E, et al. Dual-task performance with ideomotor-compatible tasks: is the central processing bottleneck intact, bypassed, or shifted in locus? *J Exp Psychol Hum Perc Perform*, 2005, 31: 122–144

A psychological model of human-computer cooperation for the era of artificial intelligence

Ye LIU^{1,2*}, Yamin WANG³, Yulong BIAN⁴, Lei REN^{5,6} & Yuming XUAN^{1,2}

1. *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;*

2. *Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China;*

3. *College of Psychology, Capital Normal University, Beijing 100048, China;*

4. *Department of Computer Science and Technology, Shandong University, Jinan 250101, China;*

5. *School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;*

6. *Beijing Advanced Innovaton Center for Big Data-based Precision Medicine, Beihang University, Beijing 100191, China*

* Corresponding author. E-mail: liuye@psych.ac.cn

Abstract In the last 10 years, the research on artificial intelligence and human-computer interaction has progressed by leaps and bounds, and there has been a boom in the academia and industry. There is an urgent need for a psychological model on human-computer interaction that is consistent with human cognitive mechanisms and psychological laws in order to eliminate the existing bottleneck in the development of artificial intelligence and to realize harmonious and natural human-computer interaction. A psychological model on human-computer cooperation for the era of artificial intelligence would not only explain, simulate, and predict human behaviors but could also guide the design of intelligent systems and human-computer interactions. In this paper, we review and summarize the previous human-computer interaction models and cognitive models and propose a psychological model for human-computer cooperation for the era of artificial intelligence based on the experimental evidences and researches of the previous 10 years. The new model would expand and deepen the understanding of the existing cognitive models and provide a theoretical basis for a study on human-computer interaction in the era of artificial intelligence.

Keywords artificial intelligence, human-computer cooperation, human-computer interaction, psychological model, perception, cognition, motor



Ye LIU was born in 1979. She received her Ph.D. degree at the Institute of Psychology, Chinese Academy of Sciences in 2005. Currently, she is an associate professor at the Institute of Psychology, Chinese Academy of Sciences. Her research interests include affective computing and conceptual representation.



Yamin WANG was born in 1972. He received his Ph.D. degree in cognitive psychology from the Institute of Psychology, Chinese Academy of Science in 2007. Currently, he is a cognitive associate professor at the Beijing Key Lab of Learning and Cognition, Capital Normal University. His research interests include situated cognition, ubiquitous learning systems, and virtual-reality-based (VR-based) smart rooms.

His research is now focused on an intelligent assessment of personality in a virtual reality environment using VR-based eye tracking.



Yulong BIAN was born in 1988. He received his Ph.D. degree in basic psychology from Shandong Normal University, Jinan in 2016. Currently, he is a post-doctoral researcher at Shandong University. His research interests include virtual reality and human-computer interaction, cyberpsychology and behavior, and user experience.



Lei REN was born in 1979. He received his Ph.D. degree in computer science from the Institute of Software, Chinese Academy of Sciences, Beijing in 2009. Currently, he is a professor at Beihang University. His research interests include big data, cloud computing, cloud manufacturing, and human-computer interaction.