

一级与二级视觉视角采择的机制比较： 理论争议、行为与神经科学的证据*

王稼茵 李晶

(南京师范大学心理学院, 南京 210097)

摘要 视觉视角采择(VPT)分为一级视觉视角采择与二级视觉视角采择。现有理论对二者关系存在根本分歧：双系统理论主张两者内部机制独立，单系统理论则认为它们共享同一系统。综合两种理论视角，本文提出三阶段加工模型，该模型认为，一级与二级视觉视角采择均经历信息处理、视角模拟及信息整合与反应选择三个阶段。行为与神经证据表明，在三个阶段中，一级与二级视觉视角采择的加工机制可能既存在差异又存在相似之处：在信息处理阶段，两者共享基础的空间信息编码，但二级视觉视角采择涉及表征更为精细；在视角模拟阶段，一级视觉视角采择依赖快速、非具身的视线追踪，而二级视觉视角采择则需要进行具身的参照系变换与自我旋转对齐；在信息整合与反应选择阶段，两者可能共享对他人意图的理解，但二级视觉视角采择需要更强的认知控制。本文构建的三阶段模型为理解一级与二级视觉视角采择提供了统一框架，未来研究应致力于开发能够分离各加工阶段的实验范式、进一步检验该模型的时间进程，并深入探索二级视觉视角采择中具身机制的触发条件及其跨模态整合。

关键词 视觉视角采择, 双系统理论, 单系统理论, 空间认知

分类号 B842.1

1 引言

英国诗人约翰·多恩曾经写下“没有谁是一座孤岛……每个人都像一块小小的泥土，连接成整个陆地。”作为社会性生物，我们总是需要与他人交往互动，预测他人的行为，理解他人的意图，并做出恰当的反应。视觉视角采择(Visual Perspective Taking, VPT)是指个体在心理上模拟并理解他人视觉经验的能力(Flavell et al., 1981)。在社交互动过程中，VPT使我们能够从他人的视角出发，判断其注意焦点和看到的内容，从而帮助我们更好地与他人协同行动(邵雨婷等, 2020)。比如，当我们想让坐在对面的朋友帮我们倒杯水时，我们需要意识到对方能够看到桌面上的杯子，并明白在我们右侧的水壶位于他的左手

边。VPT不仅让人们认识到他人视角信息与自身的不同，也为推测更复杂的心理状态(如信念、意图)提供了基础(肖承丽等, 2021; Pesimena & Soranzo, 2023; Samuel, Cole et al., 2023)。

Flavell等人(1981)提出VPT可以划分为两个层次，一级VPT和二级VPT。一级VPT聚焦于简单的二元判断问题，即判断物体是否处于他人视野范围内，比如意识到背对书架的人无法看到书架上的书。而二级VPT则涉及更为复杂的推断过程，它不仅要求个体判断物体是否处于他人视野范围内，还需要进一步理解即便自己和他人都能看到同一物体，由于观看的距离、角度等环境因素不同，两人视角中物体具体的视觉呈现效果仍会存在差异(Flavell et al., 1981)。例如当两个人面对面同时看到桌面上的数字时，我们能够理解自己眼中的数字“6”在他人眼中却是数字“9”(Surtees et al., 2012)。这种一级与二级VPT之间的差异，体现了个体在模拟他人视觉经验时的复杂性。

收稿日期: 2025-06-05

* 国家自然科学基金面上项目(42371444); 教育部人文社会科学研究一般项目(24YJA190007)。

通信作者: 李晶, E-mail: lij@njnu.edu.cn

虽然 Flavell 等人(1981)对一级与二级 VPT 作出了概念区分,但二者在内部机制层面的关系仍待理清——两种 VPT 是否涉及截然不同的处理机制,抑或具有共享或相似的内部过程。Cole 等人(2020)表明,当前 VPT 领域缺乏统一的理论框架,而明确两种 VPT 在处理机制上的独特性与关联性,是填补这一空白的核心前提——只有深入理解一级与二级 VPT 各自的内部过程,才能构建出可以覆盖不同 VPT 类型的系统性理论。多名 VPT 领域的重要研究者共同指出,探讨不同形式的 VPT 之间是否有共享的处理过程是视角采择研究迈向未来的关键议题之一(Samuel et al., 2024)。因此系统阐明一级与二级 VPT 处理过程的异同具有重要的研究意义。现今,国内的视角采择综述主要围绕视角采择的发展研究(赵婧等, 2010)、视角采择的行为范式研究(张越等, 2018)、视角采择中自我中心性偏差的机制之争(吴梦慧等, 2022)、自动视角采择中“内隐心智化”与“潜心智化”的争议(李艺, 肖风, 2021),以及视角采择与空间语言交流(肖承丽等, 2021)等过程的关联展开,尚未有研究专门探讨一级与二级 VPT 的关系。

基于此,本文旨在系统澄清一级与二级 VPT 在加工机制上的异同。我们首先探讨了双系统与单系统理论的核心争议,指出其各自局限,进而提出一个整合性的“三阶段加工模型”作为全新的分析框架。继而,我们循此框架,系统地整理了行为与认知神经科学的研究证据,分别阐述两种 VPT 在信息处理、视角模拟及信息整合与反应选择三个阶段中的特性与共享性机制。最后,在证据整合的基础上,正式构建并阐释一级与二级 VPT 的三阶段模型,并据此展望未来研究方向,以期为一二级 VPT 的未来研究提供新的视角。

2 多种理论视角下一二级视觉视角采择的关系

从理论层面上来看,目前还并没有专门针对 VPT 的统一理论框架。但在过往研究中,有一些“读心”理论对一级与二级 VPT 的内部过程提出了不同的解读。“读心”理论是指探讨人类如何理解他人的心理状态的理论,但在某些“读心”理论中,一级与二级 VPT 的关系也被作为支持理论的证据进行了讨论。其中,最具代表性的理论是双系统理论和单系统理论。

2.1 两种理论的基本观点

双系统理论(two-systems account)认为有两种专门负责理解他人的心理状态的独立但互补的系统:最小读心系统和完全读心系统(Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low et al., 2016)。最小读心系统(minimal mindreading system)发展较早,核心功能是表征“注册(registration)”情况,即记录个体、他人、物体之间的位置关系(Apperly & Butterfill, 2009)。它运行快速、自动化,不易受背景知识干扰,但是难以处理涉及信念等需要高度灵活性的任务(Apperly et al., 2006; Butterfill & Apperly, 2013)。所以当任务要求更加复杂时,个体就需要使用完全读心系统来解决问题(Surtees et al., 2012; Thompson, 2014)。完全读心系统(full-blown mindreading system)又称灵活系统,它是在个体发育后期逐渐成熟的,受语言与执行功能支持的系统(Apperly & Butterfill, 2009; Samson & Apperly, 2010)。完全读心系统能够灵活地处理更为复杂的心理状态推理任务,但其运作缓慢且需要消耗大量认知资源(Low et al., 2016)。

与之相对,单系统理论(single-system account)则认为从婴儿期到成人期,只存在一个单一且连续发展的“心智推理系统”(Carruthers, 2016; Gómez-Tabares, 2023; Kloos et al., 2020)。这个系统在婴儿阶段已具备雏形,在成长的过程中功能逐渐丰富,逐步演变为成人的成熟系统。该理论否认存在两个相互独立的平行系统,认为个体在不同任务中的表现差异,并非源于不同的系统,而是由于任务对认知资源的需求不同。

综上所述,双系统理论与单系统理论对“读心”能力的基本架构提出了截然不同的假设:前者主张存在两个独立且功能互补的认知系统,分别负责不同范畴的心智推理;后者则强调一个统一的系统足以解释不同层次心智推理任务的表现差异(李鸿锴, 2025)。两种理论的分歧也引发了它们对一级与二级 VPT 内在机制的争论。接下来,我们将进一步厘清这两种理论在解释 VPT 机制时的具体观点与冲突。

2.2 两种理论在 VPT 机制上的观点冲突

双系统和单系统理论针对 VPT 内部机制的核心冲突点在于:一级与二级 VPT 是两种完全独立的不同机制,还是共享着同一个内部机制。

双系统理论认为一级与二级VPT分别依赖于两种截然不同的加工机制：一级VPT依赖于最小读心系统，只需要追踪简单的位置可见性的信息；而二级VPT有完全读心系统的参与，所以能够理解他人对物体的具体表征(Low et al., 2016)。两种VPT的机制之间相互平行、独立，无法直接交换信息。这代表在双系统理论框架下，一级VPT更倾向于高效自动处理，而二级VPT则更依赖灵活但认知负荷较高的认知过程。双系统理论的关键证据来自于两种一级与二级VPT任务的经典范式。在经典的一级VPT范式——点视角任务(dot-perspective task; Samson et al., 2010)中，被试需判断自己或采择对象所见的点数。结果发现，当两者视角不一致时，即使被试被明确告知不需要考虑他人视角，被试反应仍会变慢、错误增加，双系统理论认为这一实验结果有力地证明了一级VPT在没有提示的情况下，仍然能够快速且近乎自动地处理他人的视觉信息。而在测量二级VPT的数字识别任务(Surtees et al., 2012)中，被试在判断自我视角内容时并未因他人视角与自己不同而反应变慢，说明被试无法自动地处理他人二级视角的内容，支持了二级VPT并不是完全自动而快速的，需要个体努力调动认知资源来实现。

然而，单系统理论的支持者对这些观点提出了质疑(Bohl & van den Bos, 2012; Carruthers, 2016; Cole et al., 2020; Cole & Millett, 2019; Jacob, 2019; Tomasello, 2018)。有研究表明，一级VPT的自动化加工并非必然，会受到个体对他人感知能力信念的调节(Cole & Millett, 2019; Furlanetto et al., 2016)。例如，在Furlanetto等人(2016)改编的点视角任务中，当被试认为他人佩戴了不透明的护目镜时，即使采择对象视角与自身不一致被试的反应时也并没有延长。这说明一级VPT的判断不是完全自动的，只有当被试知道他人能够看到物体时，才会产生一级VPT。同样的，Elekes等人(2016)改编了测量二级VPT的数字识别任务，分别设置了个体、联合视角依赖、联合非视角依赖三种条件。在个体条件中被试独自进行数字任务。在联合视角依赖条件下，被试面前会有一个真实的伙伴，并被告知他们的伙伴正在完成同样的任务。在联合非视角依赖条件下，被试被告知伙伴正在完成不同的任务。结果发现当被试知道并相信伙伴与自己进行相同的任务时，他们的反

应显著快于被试独自进行的条件。Westra (2017)指出，二级VPT任务并非天生需要“慢系统”，当被试具备正确的背景知识，并且有足够的动机时，二级VPT也可以快速有效。因此单系统理论认为，一级和二级VPT任务的差异来自于任务对资源的不同要求，而不是两个独立系统的运作。一级VPT因仅涉及简单的空间关系推理，故而显得快速自动；而二级VPT因需要更复杂的操作(如心理旋转)，所以需要调动更多认知资源，表现得缓慢而费力。这种差异是量变而非质变，二者共享同一个底层加工系统。然而，单系统理论对两种VPT共享单一机制的解释也面临挑战。若一级与二级VPT果真依赖单一系统，其处理机制应高度一致，而现有研究却发现二者存在差异：一级VPT依赖于快速的视线追踪(Michelon & Zacks, 2006)，而二级VPT则更多涉及身体表征与模拟(Kessler & Thomson, 2010)。上述差异难以仅用调用认知资源的多少来解释。

2.3 新的理论方向：多阶段加工模型

尽管双系统理论与单系统理论为理解一级与二级VPT的关系提供了富有启发性的框架，但它们在本质上仍将VPT视为一个依赖于特定系统模块的整体性过程——无论是两个独立的系统，还是一个统一的系统。然而，这种整体“系统”层面的划分难以完满地解释所有实证发现：双系统理论不能解释为何一级VPT的“自动化”加工会受到信念的调节，二级VPT有时也会快速有效；而单系统理论则难以阐明，为何一级与二级VPT会稳定地依赖不同的核心认知机制。这些局限表明，从“系统”这一整体层面出发的理论视角，可能难以捕捉一级与二级VPT内部加工过程的复杂性，因此我们需要一个更具包容性和整合性的新理论框架。VPT任务本身并非一个单一的认知行为，而是包含了多种子过程的复合任务。无论是完成一级还是二级VPT，个体都需要处理感知信息，模拟他人视角内容，区分自身和他人的视角并且抑制无关视角的干扰。因此，难以简单地用“一个系统”或“两个系统”便概括所有子过程中的表现。一个更合理的推测是，在不同的加工子过程中，一级与二级VPT的关系是动态变化的：在某个阶段，两者的认知机制可能高度重叠；而在另一个阶段，则可能调用完全不同的处理机制。

基于此，我们提出从“系统”转向“阶段”的分

析思路,将VPT根据任务要求分解为三个主要加工阶段:(1)信息处理阶段,即对自我、他人和物体的空间关系及属性进行初步感知与表征;(2)视角模拟阶段,即通过特定的心理操作模拟他人视角下的信息内容;(3)信息整合与反应选择阶段,即整合多方信息,抑制无关视角干扰,并做出最终判断。这种“三阶段加工模型”并不预设一级与二级VPT整体上是完全独立还是完全共享,而是试图在更细致的层面上,探讨它们在哪些环节存在机制重叠,在哪些环节又存在本质差异。这为理解一级与二级VPT之间既存在共性又存在差异的复杂关系,提供了一个更为整合的理论框架。接下来我们将系统地考察行为与神经科学层面的证据,从而为构建这样一个整合性的多阶段VPT加工模型奠定坚实的基础。

3 一二级视觉视角采择的行为研究

“读心”理论层面的争议提示我们,一级与二级VPT可能既存在不同的处理过程,又包含相似甚至共享的处理过程。如前文所述,我们试图提出一个三阶段加工框架,该框架包括信息处理阶段、视角模拟阶段和信息整合与反应选择阶段。接下来我们依照此框架,系统地阐述在行为层面上,一级与二级VPT在不同阶段所展现出的机制异同。

3.1 信息处理阶段:表征深度与范围的差异

在一级与二级VPT任务的加工初始,个体必须对视觉场景中的基本要素进行编码,构建一个包含自我、他人与物体的空间情境模型。行为研究表明,一级与二级VPT在此阶段的核心差异在于信息表征的深度和范围。

大量研究表明,一级VPT的判断依赖于对视线路径是否通畅的快速评估(Kelly et al., 2004; Michelon & Zacks, 2006)。Michelon和Zacks(2006)的实验是证明一级与二级VPT机制差异的经典研究,在该研究中,被试需要完成两种任务:判断物体是否在采择对象的左侧或右侧(二级VPT任务)和判断采择对象是否能看到目标物体(一级VPT任务)。结果发现,一级VPT任务中的表现只与采择对象与物体之间的距离相关,距离越长被试的反应时间越长,而不受被试与采择对象之间角度差异的影响。这表明视线是影响一级VPT的关键因素,个体在进行一级VPT时,可能

更多涉及对他人视线进行追踪的机制,人们会想象从采择对象的眼睛到目标物体之间存在着一条“虚拟连线”,如果这条线没有被遮挡则物体便是可见的(Michelon & Zacks, 2006)。因此在一级VPT的信息处理阶段,被试只需要编码“视线是否被阻挡”这一物理属性,表征深度较浅,范围局限于视线路径本身。后续一系列研究通过操纵视线的有效性,再次佐证了这一观点。Furlanetto等人(2016)和Marshall等人(2018)的护目镜任务发现,当被试相信采择对象佩戴的眼镜为不透明时,采择对象的视角不会再自动干扰被试的自我判断。这表明,一级VPT的触发并非无条件反射,而是依赖于对他人视觉感知能力的信念,而视线通畅是此信念成立的物理前提。Baker等人(2016)和O'Grady等人(2020)使用屏障任务得出了相似的结果,屏障任务是指在传统的点视角任务基础上,在采择对象与物体之间设置阻挡视线的屏障,可见条件下屏障有“窗口”能看到墙面,不可见条件则完全遮挡了视线。而当屏障遮住了被试的视线时,他人视角对自我任务中被试判断的干扰效应消失。这些证据共同表明,一级VPT在信息处理阶段可能主要识别环境中与视线路径相关的信息,比如编码物理遮挡、距离和头部朝向等基础空间关系,其范围应该相对局限。

相比之下,二级VPT则需进行深层与精细化的空间表征。与一级VPT不同,二级VPT要求个体理解他人视角下物体的具体形态(如数字“6”与“9”)或精确位置(左或右)。这必然要求其进行更深层次的加工。多数研究结果表明,视线对于二级VPT的影响较小(Quesque et al., 2018; Ward et al., 2019; Ward et al., 2020)。Ward等人(2020)对经典心理旋转任务进行了改编,在场景图片中,桌面上呈现不同旋转角度的字符,而桌边存在一个与任务无关的人物,结果发现即使字符的旋转方向与被试自身差别较大,但当字符与桌边人物的正方向相一致时,被试的反应仍然会更快。重要的是,当个体转头移走视线时,这一过程仍然不会受到影响。这表明二级VPT的加工不仅要确认视线是否通畅,还涉及更复杂的空间变换,因此需要精确表征物体本身可能因视角不同而呈现的形态变化,和自身与采择对象的身体朝向等信息。总而言之,在信息处理阶段,二级VPT所处理的信息更为精细和复杂,其表征范

围从简单的视线路径扩展到了具体的空间关系与视觉形态。

3.2 视角模拟阶段：快速视线追踪 vs. 具身的参照系转换

在完成基础的空间信息表征后，个体需要执行核心的心理操作以模拟他人所看之内容。而在这一阶段，一级与二级 VPT 展现出两种有差异的加工机制。

一级 VPT 的模拟过程相对直接，正如前文所述，研究发现一级 VPT 会受到采择对象与目标之间距离的影响，但是并不随着观察者与目标之间角度差异的变化而改变，因此一级 VPT 的视角模拟阶段更多涉及到快速的“视线追踪” (Michelon & Zacks, 2006)。Creem-Regehr 等人(2013)指出，这种模拟机制是非自我中心转换的——个体无需离开自身固有的空间参照系，始终立足于自己的视角，只是在此框架内对他人的视线进行了一条虚拟的“连线”操作，通过心理想象来判断从他人双眼到目标物体的视线路径是否存在中断。

然而，二级 VPT 的视角模拟过程则更加复杂，Michelon 和 Zacks (2006) 的研究显示，在二级 VPT 任务中，被试的反应时显著受到自身与采择对象间角度差的影响，角度差越大，反应越慢。这代表着二级 VPT 包含了心理旋转的过程。需要注意的是，二级 VPT 与物体的心理旋转存在本质区别，不能将二者视作完全重合的心理机制(赵标柯等, 2010)。研究者认为，二级 VPT 是一个“具身”的过程 (Müsseler et al., 2022; Samuel, Salo et al., 2023)。“具身”简单来说就是和身体相关，强调身体在认知、感知等过程中起着重要的作用 (Wilson, 2002)。大量证据表明，二级 VPT 与身体和动作表征紧密相关 (Fischer & Demiris, 2020; Kessler & Thomson, 2010; Surtees et al., 2013; Yu & Zacks, 2017)。Kessler 和 Thomson (2010) 以及 Surtees 等人(2013) 的研究都发现，身体姿势的对齐对二级 VPT 具有重要的作用，研究中被试需要坐在旋转椅上判断采择对象看到的物体位置和数字外观。结果发现无论是哪种判断，当被试自身姿势的运动方向与采择对象的方向一致时(比如被试向右旋转，而采择对象同样面向右方)，被试的判断效果更好。由此可知，身体线索对于二级 VPT 具有至关重要的作用，在二级 VPT 的视角模拟过程中，个体会对获得对方观点所必需的身体运动(旋转/

平移)进行心理模拟，这种心理模拟并不是一种基于物体对象的心理旋转，而是将自己的身体运动模式与采择对象的身体姿势相匹配，“设身处地”的具身自我旋转 (Kessler & Thomson, 2010)。

此外，需要关注的是，这种具身地想象自己旋转到他人角度上的自我旋转并不是“一点一点”的旋转，而是将自己的参照系进行变换，重新对准一条“主轴”的变换。Wraga 等人(2005)和 Negen (2025)使用相同的材料对比了心理旋转和自我旋转(类似于二级 VPT 中采取他人视角方向的旋转过程)，在该实验中测试材料为十个由立方体构成的三维物体，心理旋转任务要求被试想象物体自身旋转到某个角度；自我旋转任务则要求被试想象自己绕物体移动到某个固定角度后进行判断。结果发现，自我旋转任务的表现并非随着旋转角度的增加而单调下降，而是呈现出一种特殊的“凹槽型”模式。在中间角度(60°)时，被试的反应表现显著优于较小角度(30°)和较大角度(90°、120°)。研究者认为这是因为被试在进行自我旋转时，会将自己的参照系旋转到另一个轴上，当刺激角度(60°)与被试未旋转前参照系的正交主轴(如左右、前后轴)接近时，被试转换更容易成本更低，所以此时被试的表现最优。因此，二级 VPT 中具身的自我旋转包含了将自身参照系与他人参照系对齐的过程。

综上所述，在视角模拟阶段，一级和二级 VPT 存在着完全独特的两种机制。一级 VPT 通过非具身的、基于自身参照系的视线追踪来实现模拟；而二级 VPT 则主要通过具身的、脱离自身参照系的自我旋转和参照系对齐来完成视角的切换。

3.3 信息整合与反应选择阶段：可能共享的意图与心理状态的理解

在完成模拟他人视角的过程后，个体进入信息整合与反应选择阶段。此时，自我视角的信息、模拟出的他人视角信息，以及来自场景的其他社会性线索会被整合。同时个体必须抑制无关视角的信息，根据任务要求选择目标视角进行最终判断。而在这一阶段，一级与二级 VPT 之间可能共享着整合他人意图与心理状态信息的过程。

近来研究揭示，社会性线索，特别是行为意图，能够作为一种高级的线索被整合，同时促进一级与二级 VPT 在最终反应阶段的判断。如前文所述，视线对于一级 VPT 具有重要的作用，但是

还有一些研究并没有发现视线的独特作用。Cole 等人(2016)采用带有屏障的点视角任务,结果显示无论屏障是否有窗口,被试判断自身视角时均不受他人视角的影响。Conway 等人(2017)通过望远镜操纵圆点可见性,也发现无论采择对象能否看到圆点,一致性效应均存在。但是这些矛盾的结果并不能完全否定视线对一级 VPT 的作用,李艺和肖风(2021)指出,这种矛盾可能是因为实验设计中存在着其他因素的影响,例如采择对象的属性与特征,以及实验中指导语的设置等。总之,对于一级 VPT 来说,视线仍是最有力的影响因素,但不代表视线能够完全地支配一级 VPT 任务的表现。Mayrand 等人(2024)的研究指出,眼神信号通常能够同时传达注视的方向和注视者的心理状态,也许,感知线索中采择对象的状态和行为意图也会对一级 VPT 产生影响。Hu 等人(2025)的研究提供了重要的证据,其进行了一系列的一级 VPT 实验,结果发现当采择对象执行注视和伸手动作时,或仅执行注视动作时,都能够引发被试的自发视觉采择,但是只有当采择对象同时展现出注视和伸手行为,并且注视行为先于伸手行为时,被试的采择效果最强,这是因为先注视后伸手的顺序符合人类“先感知后行动”的正常行为模式,更容易被理解。这代表着,当感知线索能够清晰地传达他人的目标导向意图时,会更有效地支持一级 VPT,从而促进对他人视角可见性的最终判断。

同样已经有研究证明,当被试感知到线索能够指向他人的心理状态或意图时,二级 VPT 的表现也会得到促进。Lukošiūnaitė 等人(2024)的实验比较了“行动”(采择对象伸手拿杯子)和“无行动”(采择对象双手放在腿上或一只手放在杯子旁)条件下的二级 VPT 表现,结果发现行动条件可以促进二级 VPT。Brady 等人(2024)的二级 VPT 研究也证实,当采择对象与目标物体进行互动时,例如做出指向或伸手抓取的动作,被试判断目标物体相对于采择对象的左右位置将更准确。基于此,一级与二级 VPT 都可能包含一个从感知线索中理解他人心理状态或意图的过程。对他人在当前情境下的目标与意图的理解,能够作为一种线索参与一级与二级 VPT 的信息整合,并进而影响最终的决策。

综合而言,通过三阶段框架的系统梳理,行

为研究的证据清晰地展示了一级与二级 VPT 在三个阶段中机制的异同:在信息处理阶段,二者表征的深度与范围存在差异:一级 VPT 表征与视线路径相关的基础空间信息,范围有限,而二级 VPT 需要表征更精细的空间关系与形态。在视角模拟阶段,二者存在本质的机制差异:一级 VPT 通过非具身的视线追踪实现模拟,二级 VPT 通过具身的参照系的旋转对齐实现模拟。而在信息整合与反应选择阶段,两种 VPT 可能共享着对他人行为意图的理解信息的整合。接下来,我们将深入神经机制层面,检验这一行为框架的神经基础。

4 一二级视觉视角采择的神经机制研究

行为研究展示了一级与二级 VPT 在三阶段加工模型中的异同,而认知神经科学的研究,则从神经层面进一步支持并深化了这一模型。

在信息处理阶段,神经机制研究同样证实了一级与二级 VPT 共同激活了负责视觉空间信息初步加工的脑区,但其激活的强度和范围存在差异。大量研究指出,次级视觉皮层、楔前、顶叶上部和左顶叶下部都检测到对二级 VPT 反应的血氧水平依赖性(blood oxygen level dependent, BOLD)信号的增加,这些脑区与空间感知和视觉信息处理息息相关(Wraga et al., 2005; Zacks et al., 2003)。而一级 VPT 同样发现了左顶叶下部、双侧楔前区等涉及到视觉空间处理的脑区的激活(Zacks & Michelon, 2005)。这些脑区的激活确保了两种 VPT 都能够感知和表征任务中必要的视觉空间信息。除此之外,最新一项功能性磁共振成像(functional magnetic resonance imaging, fMRI)研究直接比较了一级与二级 VPT 的神经机制,研究结果显示,一级与二级 VPT 都显著激活了双侧枕顶叶皮层及右侧额叶的小范围区域,但是只有二级 VPT 任务显著激活了包括额下回、中央前回、顶下小叶、辅助运动区、岛叶和小脑在内的更广泛脑区;并且功能连接分析显示,二级 VPT 对背侧注意网络与额顶控制网络的激活强度显著更高(Schurz et al., 2025)。研究者指出,虽然一级与二级 VPT 都包含基本视觉空间信息处理机制(如双侧枕顶叶皮层)和认知控制的脑区(如右侧额叶的小范围区域),但一级 VPT 的神经激活相对局限,这支持了一级 VPT 主要依赖于对视线的快速追踪。而二级 VPT 则具有更复杂的神经加工路径,

会激活加工深度更深、更加高级的网络来应对复杂的视角冲突与认知调控需求(如背侧注意网络与额顶控制网络的高度激活),并且可以执行心理旋转、重新编码空间关系等高级功能(Schurz et al., 2025)。这一研究同样启示着在信息整合与反应选择阶段中,二级VPT需要比一级VPT更高的认知控制需求和加工深度。

另一方面,视角模拟阶段中二级VPT独特的具身机制也得到了认知神经科学研究的支持。例如一项较早的经典fMRI研究就已经证实,当个体想象自己围绕物体旋转(即二级VPT中的具身自我旋转)时,主要激活了枕中回的脑岛、辅助运动区、外显体区和右顶上小叶等脑区(Wraga et al., 2005)。这些脑区除其他功能外还参与了身体表征的编码(Fontan et al., 2017),但是在一级VPT中并没有发现相关脑区的激活,这支持了二级VPT单独涉及到具身模拟的假设(Gunia et al., 2021)。

然而,神经机制的研究对于行为研究中“一级与二级VPT均涉及从感知线索中理解他人心理状态与意图”这一猜想,仍存在争议,这一争议主要围绕着右侧颞顶叶联合区(right temporoparietal junction, rTPJ)和背内侧前额叶(dorsomedial prefrontal cortex, dmPFC)。rTPJ和dmPFC是“社会脑”的重要枢纽,它们与多种社会认知的脑区高度互联,同样参与着心理理论的认知过程(Schurz et al., 2014)。多数研究认为,rTPJ和dmPFC对于二级VPT具有重要的作用(Aichhorn et al., 2006; Santiesteban et al., 2012; Seymour et al., 2018)。例如Seymour等人(2018)采用左右判断的二级VPT范式,结合脑磁图(magnetoencephalography, MEG)和fMRI发现,随着自我与他人视角之间角度差异的增加,被试的反应时间显著延长,且rTPJ、右前扣带回皮层和右背外侧前额叶皮层的theta频段(3~6Hz)功率显著增加。其中rTPJ起着核心的协调作用,它可以通过theta波段连接视觉区域、心理化网络和运动/身体图式网络(Seymour et al., 2018)。除此之外,dmPFC在二级VPT中也存在显著的激活(David et al., 2006; Lieberman et al., 2019; Wittmann et al., 2016)。Mazzarella等人(2013)的fMRI研究发现,在二级VPT任务中,随着采择对象与被试之间角度差异的增大,dmPFC的BOLD信号呈现线性增强的趋势。然而,这些研究并没有在一级VPT的任务中观察到rTPJ和

dmPFC类似的激活模式(Seymour et al., 2018; Wang et al., 2016)。Martin等人(2020)通过高精度经颅直流电刺激技术(High-Definition transcranial Direct Current Stimulation, HD-tDCS),明确对比了dmPFC和rTPJ在一级与二级VPT的作用,结果发现无论是刺激rTPJ还是dmPFC,对一级VPT的任务表现都未产生任何显著影响。

但是在Martin等人(2019)同样使用HD-tDCS的研究中,发现了dmPFC对显式的一级VPT任务也会产生影响(显式任务是指被试被明确告知需要从自己视角或者他人视角进行反应)。除此之外,Rochas等人(2023)使用高密度脑电图分析了点视角任务中神经机制的情况,结果发现在一级VPT的自我任务中,大脑早期的激活主要集中于基础视觉注意与信息整合的区域,包括海马旁回、枕叶及舌回。而他人任务中的脑激活时间开始晚但是涉及更广泛高阶网络:从500ms开始,激活显著扩展至与心理理论相关的心智化网络(楔前叶、后扣带回)、角回,以及负责认知控制的额顶执行网络;其中,右额叶与颞顶叶区域的激活尤为突出,这一结果与前人的两项脑电研究结果一致(Beck et al., 2018; McCleery et al., 2011)。Yao等人(2021)针对过往一级与二级VPT的13项非侵入性脑刺激的研究进行了元分析,结果显示rTPJ的兴奋性刺激能提升被试在一级VPT他人任务中的表现,尤其当自我与他人视角不一致时效果更明显。这表明rTPJ在一级VPT中也能起到抑制自我视角干扰的作用,促进对他人视角可见性的判断;而dmPFC刺激同样显著降低了被试在自我任务中的表现,可能增强了他人视角无关信息的显著性,从而干扰了自我视角的判断。然而惊讶的是,该研究却没有发现rTPJ和dmPFC对二级VPT的显著影响。

Bukowski(2018)指出任务类型的差异可能是导致这些不一致结果的重要原因。有的研究使用的是显式任务,有的研究使用的是隐式任务。而有的研究同时包含自我任务与他人任务。但是在解释研究结果时这些研究并没有深入剖析为什么某些脑区只在其中一种任务下被激活(Bukowski, 2018; Yao et al., 2021)。正如Martin等人(2019)的研究,显式的一级VPT任务展现了社会认知相关的脑区的激活,因此也许只有当任务需求或情境线索明确包含了对他人状态或意图时(如明确的

伸手动作), 或者被试被明确要求需要考虑他人(如他人任务)时, 一级 VPT 才会调用这种社会认知的脑网络。此外, 如前文所述, 一些脑电研究的结果显示, 一级 VPT 在后期才会涉及到心理理论相关脑区的参与(Beck et al., 2018; McCleery et al., 2011; Rochas et al., 2023), 而 Seymour 等人(2018)的研究也指出随着二级 VPT 的进行, rTPJ 会逐渐减少与视觉区域的连接, 加深与心理化和身体图式区域的连接。因此, 这种理解心理状态与意图的共享机制更有可能存在于一级与二级 VPT 的后期, 尤其是在信息整合与反应选择阶段产生作用。

总而言之, 现有的认知神经研究进一步支持并扩展了行为研究中对两种 VPT 的内部机制的猜想。首先, 在信息处理阶段, 两种 VPT 共同涉及处理和表征视觉空间信息的机制(如双侧枕顶叶皮层等), 但是因为任务的要求不同, 二级 VPT 额外激活了加工更深度信息的更高级的脑区(如背侧注意网络与额顶控制网络的高度激活)。其次, 在视角模拟阶段, 二级 VPT 相比一级 VPT 额外涉及到一些身体编码的脑区(如脑岛、辅助运动区、外显体区和右顶上小叶等), 这证实了只有二级 VPT 才具有具身的过程。而在信息整合与反应选择阶段, 一级与二级 VPT 都包含了进行认知控制的脑区(如右侧前额叶), 需要根据目标视角抑制无关信息进行判断, 但是一级 VPT 所需的认知负荷较低。然而, 关于一级与二级 VPT 是否都会激活社会认知相关脑区存在争议, 但是在线索明确的情况下, 两种 VPT 很可能会稳定地共享理解他人意图和心理状态这一机制。

5 一二级视觉视角采择的内部机制的假设

本文围绕着一级与二级 VPT 的机制异同逐步深入。首先通过对现有的“读心”理论的分析指出, 一级与二级 VPT 可能既存在不同的机制, 又包含类似甚至共享的机制, 并提出了三阶段模型的设想。接着从行为研究和认知神经研究入手, 试图探索在不同阶段中一级与二级 VPT 机制具体异同。行为研究发现, 在信息处理阶段, 两种 VPT 处理的深度与范围不同; 而在视角模拟阶段, 一级 VPT 更多依赖于快速的视线追踪机制, 而二级 VPT 则包含具身的参照系对齐的自我旋转, 这是两种 VPT 重要的内部过程差异。除此之外, 在信

息整合与反应选择阶段, 两者还可能都需要整合线索中他人的心理状态与意图。神经机制的研究同样提供了支持性的证据, 首先, 一级与二级 VPT 同时激活了处理和表征视觉空间信息的脑区, 但是一级 VPT 更多涉及基本的视觉特征的处理, 这支持了行为研究中关于信息处理阶段的猜想, 同时也为视角模拟阶段为快速视线匹配机制提供了佐证。此外二级 VPT 额外激活了涉及身体编码的脑区, 契合了行为研究中具身机制的发现。最后, 二级 VPT 激活了更高级的认知控制机制, 这代表了二级 VPT 在信息整合与反应选择阶段需要较高的认知负荷, 而认知神经研究中关于一级与二级 VPT 是否都包含社会认知相关的脑区(主要是 rTPJ 和 dmPFC)存在争议, 但是这种争议意味着在线索中意图明确的情况下, 一级与二级 VPT 可能共同涉及到理解他人的心理状态与意图。

基于前述行为与神经科学证据的系统性整合, 我们正式提出了关于一二级视觉视角采择内部机制的“三阶段模型”。一级与二级 VPT 的加工过程可解析为三个阶段, 而在这些阶段中各自的机制存在差异和相似(见图 1)。

第一个阶段为信息处理的阶段。一级与二级 VPT 都需要处理和表征视觉空间信息, 例如视线、身体朝向等指向性线索和客观的环境线索, 建立一个包含自我、他人、物体三者的空间区域。在这一阶段, 一级与二级 VPT 表征信息的深度不同, 需要处理的范围也不同。一级 VPT 只需要关注与视线路径相关的物理条件(如距离、光线、遮挡)(Michelon & Zacks, 2006), 以及自我与他人的视线与头部方向等(Furlanetto et al., 2016; Marshall et al., 2018; O'Grady et al., 2020), 而二级 VPT 则需要表征物体具体的形态与位置, 他人的身体朝向等, 加工深度更深(Samuel, Cole et al., 2023; Samuel, Salo et al., 2023; Schurz et al., 2025; Wang et al., 2025)。

第二阶段为视角信息模拟的阶段。在处理基本信息后, 被试需要模拟他人视角的信息内容, 此时一级与二级 VPT 会运行两种不同的模拟机制。一级 VPT 会追踪他人的视线路径, 判断物体与他人之间是否存在遮挡, 当物体存在于他人的空间区域内未被阻挡时, 物体即可见(Creem-Regehr et al., 2013)。这一过程更多地基于视觉场景的物理属性和个体的视觉与空间感知能力, 因

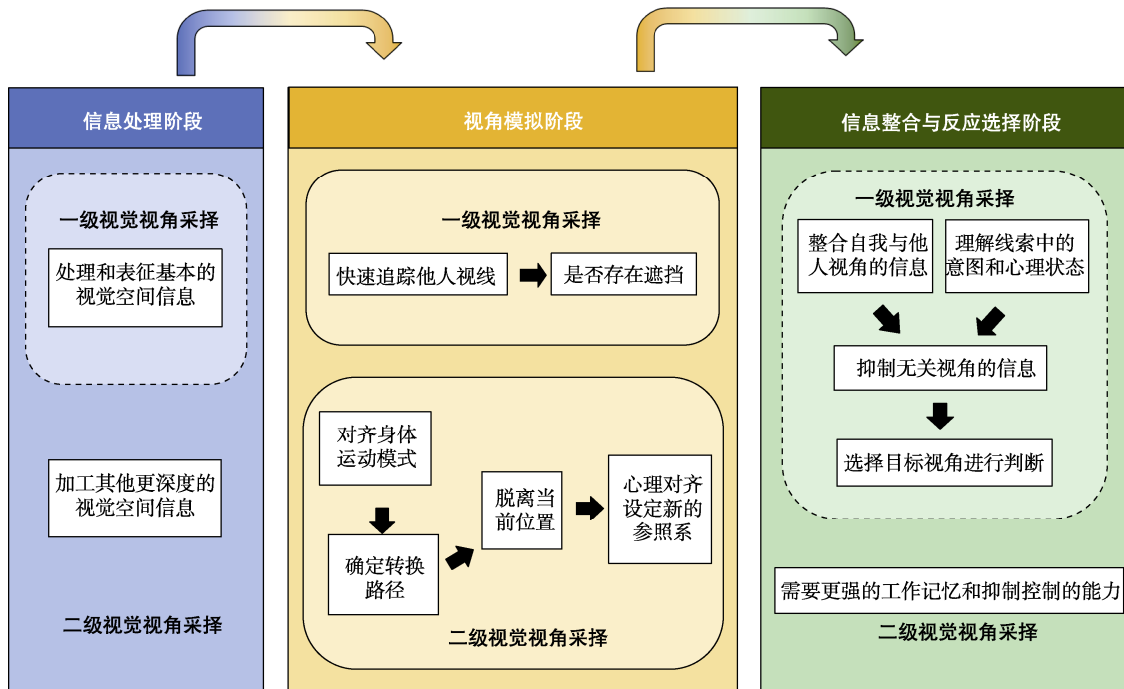


图 1 一二级视觉视角采择的内部机制

此具有自动化的基础,但其实际表现仍会受目标任务和执行功能的影响(Todd et al., 2017)。二级 VPT 则更加复杂,首先个体需要将自己的身体运动模式与采择对象的身体姿势相匹配,例如从自身的站立转为他人所使用的坐姿,这种模拟过程消耗的认知资源较少,主要依赖于整合视觉、本体感觉与前庭信号的姿势控制机制(Yeh et al., 2021)。接着个体需要确定从自我位置到他人位置的空间转换路径。例如当他人位于左侧时,个体要在心理上模拟向左旋转的动作。最后个体会设定移动方向进行对齐参照系的变化,将自己的主轴与他人的主轴相匹配(Fischer & Demiris, 2020; Kessler & Thomson, 2010; Surtees et al., 2013; Yu & Zacks, 2017)。

第三个阶段是整合信息,抑制无关视角选择目标视角的阶段。在这一阶段,个体不光需要整合自我与他人视角的信息,还需要理解线索中的意图和心理状态,但是这种理解可能只有在线索明确的情况下才会被整合(Brady et al., 2024; Ford et al., 2024; Hu et al., 2025; Mayrand et al., 2024; Ueda et al., 2021)。这些信息的整合会共同影响一级与二级 VPT 抑制无关视角的信息,选择目标视

角进行判断(Ciorli & Pia, 2023)。其中二级 VPT 需要比一级 VPT 更强的工作记忆和抑制控制的能力来执行这一阶段(Schurz et al., 2025)。

6 总结与展望

VPT 是个体理解他人感知体验的关键过程,对于社会交往和社会认知至关重要。本文通过系统剖析一级与二级 VPT 的理论争议、行为与神经科学证据,提出了一个全新的“三阶段整合模型”。该模型为理解一级与二级 VPT 的机制异同提供了一个统一且富有解释力的框架,而未来研究仍然需要继续深入几个关键方向:

首先,未来的行为研究应该聚焦于设计能够真正分离三个阶段的实验范式,并在此基础上探索不同影响因素分别作用于一级与二级 VPT 的哪些阶段。目前已经有过程分离程序(process dissociation procedure, PDP)的研究进行了相关的尝试。PDP 是一种分析方法,其假设所有任务表现都同时受控于需要认知资源的控制过程和快速无意的自动过程。在 VPT 的自我任务中,自动过程是计算他人视角,控制过程是选择自我视角。在一致条件下,被试的正确反应代表自动和控制

过程都成功运行(因为他人视角与目标视角一致);而在不一致条件下,错误反应代表控制过程失败而自动过程成功。将个体在这两种条件下的行为数据并代入特定的数学模型,就可以计算出两种过程的强度,从而进一步探讨不同影响因素对这两种过程的影响(Qureshi & Monk, 2018; Todd et al., 2019; Todd et al., 2021)。PDP 的研究可以探讨同一影响因素对 VPT 不同过程的影响,例如 Todd 等人(2017)通过 PDP 发现,时间压力会削弱了 VPT 的控制过程但不影响自动过程。但是 PDP 并未真正地将同一任务的不同的过程进行程序分离,自动和控制两种阶段的划分也过于简单与机械。因此未来研究需要致力于开发能够分离不同的 VPT 阶段的实验范式,并且进一步探索同一影响因素对于一级与二级 VPT 不同阶段是否会产生相同的影响,从而再次佐证一级与二级 VPT 的机制差异。

其次,未来的神经机制的研究应该增强时间进程的研究,并且探讨任务类型以期解决社会认知脑区的争议。目前我们针对一级与二级 VPT 的三种阶段提出来了模型假设,但这三个阶段是严格串行还是部分重叠?它们各自的时间进程关系如何?要回答这些问题,必须加强高时间分辨率技术(如脑电图/脑磁图)的应用。通过分析特定脑网络(如身体表征网络、社会脑网络)在不同时间窗口的激活顺序和功能连接模式,我们可以检验三阶段模型的时间进程,提供更直接的证据。其次,关于 rTPJ 和 dmPFC 等社会认知核心脑区是否共同参与两种 VPT,以及在何种条件下参与,现有研究仍然存在巨大争议(Martin et al., 2020; Yao et al., 2021)。这种不一致,很可能源于任务范式的异质性未被充分考量。未来研究应有意识地直接比较显性与隐性任务以及是否包含明确意图线索(如指向、抓取动作)的条件,从而明确社会认知脑区参与的条件。

最后,未来研究可以不断扩展本文的模型假设,以拓宽其适用范围并提升解释力。比如深入探讨二级 VPT 具身机制的特征与内部过程。虽然二级 VPT 已经被公认为具有这种具身的自我旋转机制,但是尚不清楚这种机制在何种情况下才会被调动。Janczyk (2013)进行了一系列心理不应期(psychological refractory period, PRP)任务,PRP 任务是双任务干扰的经典范式,在 PRP 任务中被试

会同时进行两个任务,如果这两个任务都需要中央处理系统的处理,那么当这两个任务快速连续呈现时,第二个任务的反应时会显著延长;这是因为中央处理阶段存在“瓶颈”,一次只能处理一个任务的“决策”部分(Pashler, 1994)。值得关注的是, Janczyk (2013)的研究发现,当只需要处理小角度的旋转时,两个任务之间的间隔时间不会对二级 VPT 的任务表现产生影响,这代表此时二级 VPT 是无意识的,可以与其它过程并行进行,不受中央容量的限制;只有当需要处理超过 60°的大角度时,二级 VPT 才开始占用中央容量。这一结果提示,基于身体参照系转换的模拟机制在低难度任务中可能未被完全触发,那么个体此时究竟依赖何种加工机制完成二级 VPT,目前仍缺乏直接证据。为此,本文在构建机制假设时,参考了 Yeh 等人(2021)的设想,将心理模拟阶段初步划分为自动化的姿势匹配与后续的旋转对齐两个子过程。该划分仍属理论推测,有待未来研究提供更多实证支持。除此之外,“具身”代表“设身处地”地代入他人,现有研究大多探讨视觉信息和身体表征的作用,但是其他感觉信息是否也会促进具身的过程呢?比如听觉、触觉、嗅觉等。Guo 等人(2024)的研究表明,在进行二级 VPT 时,听觉信息也可以在建立以他人中心参照系时被整合。但是除此之外还较少有研究者探讨他人中心参照系中的跨模态交互。除了这一扩展外,未来研究也应该不局限于目前的设想,更深入地去探索其他过程的特征。

参考文献

- 李鸿锴. (2025). 读心系统理论之争: 双系统还是单系统?. *系统科学学报*, 33(1), 46-51.
- 李艺, 肖凤. (2021). 自动观点采择: 内隐心智化与潜心智化的争议. *心理科学进展*, 29(10), 1887-1900.
- 邵雨婷, 李伟健, 孙炳海, 张文海. (2020). 视觉空间观点采择对教师共情的影响: 自我表征抑制和自我视空转换的不同作用. *心理科学*, 43(4), 871-878.
- 吴梦慧, 谢久书, 邓铸. (2022). 视觉观点采择中自我中心性偏差的抑制和归因之争. *心理科学进展*, 30(1), 179-187.
- 肖承丽, 隋雨繁, 肖苏衡, 周仁来. (2021). 空间交互研究新视角: 多重社会因素的影响. *心理科学进展*, 29(5), 796-805.
- 张越, 葛贤亮, 田志强, 葛列众. (2018). 基于空间的一级和二级视角转换的行为研究及理论综述. *心理科学*, 41(2), 504-510.

- 赵婧, 王璐, 苏彦捷. (2010). 视觉观点采择的发生发展及其影响因素. *心理发展与教育*, 26(1), 107-111.
- 赵杨柯, 钱秀莹. (2010). 自我中心视角转换——基于自身的心理空间转换. *心理科学进展*, 18(12), 1864-1871.
- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind?. *NeuroImage*, 30(3), 1059-1068.
- Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, 17(10), 841-844.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953-970.
- Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance*, 42(4), 508-516.
- Beck, A. A., Rossion, B., & Samson, D. (2018). An objective neural signature of rapid perspective taking. *Social Cognitive and Affective Neuroscience*, 13(1), 72-79.
- Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of type 1 and type 2 processes. *Frontiers in Human Neuroscience*, 6, 274.
- Brady, N., Leonard, S., & Ni Choisdealbha, Á. (2024). Visual perspective taking and action understanding. *Acta Psychologica*, 249, 104467.
- Bukowski, H. (2018). The neural correlates of visual perspective taking: A critical review. *Current Behaviour Neuroscience*, 5, 189-197.
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606-637.
- Carruthers, P. (2016). Two systems for mindreading?. *Review of Philosophy and Psychology*, 7(1), 141-162.
- Ciorli, T., & Pia, L. (2023). Spatial perspective and identity in visual awareness of the bodily self-other distinction. *Scientific Reports*, 13(1), 14994.
- Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, 164, 165-168.
- Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review*, 26(6), 1787-1802.
- Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective-taking: In search of a theory. *Vision*, 4(2), 30.
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 454-465.
- Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience*, 7, 596.
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... Vogeley, K. (2006). Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, 18(6), 898-910.
- Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, 41, 93-103.
- Fischer, T., & Demiris, Y. (2020). Computational modeling of embodied visual perspective-taking. *IEEE Transactions on Cognitive and Developmental Systems*, 12(4), 723-732.
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the level 1-level 2 distinction. *Developmental Psychology*, 17(1), 99-103.
- Fontan, A., Cignetti, F., Nazarian, B., Anton, J. L., Vaugoyeau, M., & Assaiante, C. (2017). How does the body representation system develop in the human brain? *Developmental Cognitive Neuroscience*, 24, 118-128.
- Ford, B., Monk, R., Litchfield, D., & Qureshi, A. (2024). Agent-object relationships in level-2 visual perspective taking: An eye-tracking study. *Journal of Cognition*, 7(1), 72.
- Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. A. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 158-163.
- Gómez-Tabares, A.-S. (2023). Is there continuity from implicit recognition of intentional action in infants to explicit mindreading in preschoolers? Systematic review of longitudinal evidence and theoretical implications. *Journal for the Study of Education and Development*, 46(4), 950-982.
- Gunia, A., Moraresku, S., & Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind. *Behavioural Brain Research*, 407, 113247.
- Guo, G., Wang, N., Sun, C., & Geng, H. (2024). Embodied cross-modal interactions based on an altercentric reference frame. *Brain Sciences*, 14(4), 314.
- Hu, X., Xu, H., Chen, H., Shen, M., & Zhou, J. (2025). Good to see you R2-D2: Inducing spontaneous perspective-taking towards non-human agents through human-like gaze and reach. *Cognition*, 259, 106101.
- Jacob, P. (2019). Challenging the two-systems model of mindreading. In A. Avramides & M. Parrott (Eds.), *Knowing other minds* (pp. 79-106). Oxford University Press.
- Janczyk, M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1878-1887.
- Kelly, J. W., Beall, A. C., & Loomis, J. M. (2004). Perception of shared visual space: Establishing common ground in real and virtual environments. *Presence: Teleoperators and*

- Virtual Environments*, 13(4), 442–450.
- Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition*, 114(1), 72–88.
- Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, 44(2), 107–115.
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience and Biobehavioral Reviews*, 99, 311–328.
- Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3), 184–189.
- Lukošiūnaitė, I., Kovács, Á. M., & Sebanz, N. (2024). The influence of another's actions and presence on perspective taking. *Scientific Reports*, 14(1), 4971.
- Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two level-1 perspective-taking tasks. *Plos One*, 13(3), e0194101.
- Martin, A. K., Huang, J., Hunold, A., & Meinzer, M. (2019). Dissociable roles within the social brain for self-other processing: A HD-tDCS study. *Cerebral Cortex*, 29(8), 3642–3654.
- Martin, A. K., Kessler, K., Cooke, S., Huang, J., & Meinzer, M. (2020). The right temporoparietal junction is causally associated with embodied perspective-taking. *The Journal of Neuroscience*, 40(15), 3089–3095.
- Mayrand, F., Capozzi, F., & Ristic, J. (2024). Gaze communicates both cue direction and agent mental states. *Frontiers in Psychology*, 15, 1472538.
- Mazzarella, E., Ramsey, R., Conson, M., & Hamilton, A. (2013). Brain systems for visual perspective taking and action perception. *Social Neuroscience*, 8(3), 248–267.
- McCleery, J. P., Surtees, A. D. R., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *Journal of Neuroscience*, 31(36), 12849–12854.
- Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68(2), 327–337.
- Müsseler, J., von Salm-Hoogstraeten, S., & Böffel, C. (2022). Perspective taking and avatar-self merging. *Frontiers in Psychology*, 13, 714464.
- Negen, J. (2025). Mental rotation, perspective taking, and performance profiling. *Cognitive Processing*, 26(3), 531–540.
- O'Grady, C., Scott-Phillips, T., Lavelle, S., & Smith, K. (2020). Perspective-taking is spontaneous but not automatic. *Quarterly Journal of Experimental Psychology*, 73(10), 1605–1628.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116(2), 220–244.
- Pesimena, G., & Soranzo, A. (2023). Both the domain-general and the mentalising processes affect visual perspective taking. *Quarterly Journal of Experimental Psychology*, 76(3), 469–484.
- Quesque, F., Chabanat, E., & Rossetti, Y. (2018). Taking the point of view of the blind: Spontaneous level-2 perspective-taking in irrelevant conditions. *Journal of Experimental Social Psychology*, 79, 356–364.
- Qureshi, A. W., & Monk, R. L. (2018). Executive function underlies both perspective selection and calculation in level-1 visual perspective taking. *Psychonomic Bulletin & Review*, 25(4), 1526–1534.
- Rochas, V., Montandon, M.-L., Rodriguez, C., Herrmann, F. R., Eytan, A., Pegna, A. J., Michel, C. M., & Giannakopoulos, P. (2023). Mentalizing and self-other distinction in visual perspective taking: The analysis of temporal neural processing using high-density EEG. *Frontiers in Behavioral Neuroscience*, 17, 1206011.
- Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. *Infant and Child Development*, 19(5), 443–454.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266.
- Samuel, S., Cole, G. G., & Eacott, M. J. (2023). It's not you, it's me: A review of individual differences in visuospatial perspective taking. *Perspectives on Psychological Science*, 18(2), 293–308.
- Samuel, S., Erle, T. M., Kirsch, L. P., Surtees, A., Apperly, I., Bukowski, H., ... Quesque, F. (2024). Three key questions to move towards a theoretical framework of visuospatial perspective taking. *Cognition*, 247, 105787.
- Samuel, S., Salo, S., Ladvelin, T., Cole, G. G., & Eacott, M. J. (2023). Teleporting into walls? The irrelevance of the physical world in embodied perspective-taking. *Psychonomic Bulletin & Review*, 30(3), 1011–1019.
- Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology*, 22(23), 2274–2277.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 9–34.
- Schurz, M., Tholen, M. G., Kronbichler, M., Perner, J., & Surtees, A. D. R. (2025). Comparing level 1 and level 2 visuo-spatial perspective-taking in the brain: Evidence from fMRI. *Social Neuroscience*, 20(6), 287–305. <https://doi.org/10.1080/17470919.2025.2490574>
- Seymour, R. A., Wang, H., Rippon, G., & Kessler, K. (2018).

- Oscillatory networks of high-level mental alignment: A perspective-taking MEG study. *NeuroImage*, 177, 98–107.
- Surtees, A., Apperly, I. A., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience*, 7, 698.
- Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *The British Journal of Developmental Psychology*, 30(1), 75–86.
- Thompson, J. R. (2014). Signature limits in mindreading systems. *Cognitive Science*, 38(7), 1432–1455.
- Todd, A. R., Cameron, C. D., & Simpson, A. J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, 159, 97–101.
- Todd, A. R., Cameron, C. D., & Simpson, A. J. (2021). The goal-dependence of level-1 and level-2 visual perspective calculation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 948–967.
- Todd, A. R., Simpson, A. J., & Cameron, C. D. (2019). Time pressure disrupts level-2, but not level-1, visual perspective calculation: A process-dissociation analysis. *Cognition*, 189, 41–54.
- Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America*, 115(34), 8491–8498.
- Ueda, S., Nagamachi, K., Nakamura, J., Sugimoto, M., Inami, M., & Kitazaki, M. (2021). The effects of body direction and posture on taking the perspective of a humanoid avatar in a virtual environment. *Plos One*, 16(12), e0261063.
- Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex*, 75, 68–81.
- Wang, N., Huang, S., Cai, J., Huang, R., & Geng, H. (2025). How the brain memorizes the world from others' perspectives: Investigating allocentric encoding of object features during perspective taking. *BMC Psychology*, 13(1), 691.
- Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, 29(5), 874–880.e4.
- Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition*, 199, 104241.
- Westra, E. (2017). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, 194, 4559–4581.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. (2016). Self-other mergence in the frontal cortex during cooperation and competition. *Neuron*, 91(2), 482–493.
- Wraga, M., Shephard, J. M., Church, J. A., Inati, S., & Kosslyn, S. M. (2005). Imagined rotations of self versus objects: An fMRI study. *Neuropsychologia*, 43(9), 1351–1361.
- Yao, Y. W., Chopurian, V., Zhang, L., Lamm, C., & Heekeren, H. R. (2021). Effects of non-invasive brain stimulation on visual perspective taking: A meta-analytic study. *NeuroImage*, 242, 118462.
- Yeh, Y. Y., Wang, C. C., Cheng, S. K., & Chiu, C. D. (2021). Dissociation of posture remapping and cognitive load in level-2 perspective-taking. *Cognition*, 214, 104733.
- Yu, A. B., & Zacks, J. M. (2017). Transformations and representations supporting spatial perspective taking. *Spatial Cognition & Computation*, 17(4), 304–337.
- Zacks, J. M., Gilliam, F., & Ojemann, J. G. (2003). Selective disturbance of mental rotation by cortical stimulation. *Neuropsychologia*, 41(12), 1659–1667.
- Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews*, 4(2), 96–118.

Comparing the mechanisms of level-1 and level-2 visual perspective taking: Theoretical controversies, behavioral and neuroscientific evidence

WANG Jiayin, LI Jing

(School of Psychology, Nanjing Normal University, Nanjing 210097, China)

Abstract: Visual perspective taking (VPT) is commonly divided into level-1 and level-2. Current theories fundamentally disagree on their relationship: the two-systems account posits distinct mechanisms, while the single-system account suggests shared processing. Integrating these perspectives, this article proposes a three-stage processing model, which posits that both level-1 and level-2 VPT involve three stages: information processing, perspective simulation, and information integration with response selection. Behavioral and neural evidence indicates that the mechanisms of level-1 and level-2 VPT exhibit both similarities and differences across these stages. During information processing, both share basic spatial

information encoding, but level-2 VPT requires more fine-grained representations. In the perspective simulation stage, level-1 VPT relies on rapid, non-embodied gaze tracking, whereas level-2 VPT involves embodied self-rotation with reference frame transformation. During information integration and response selection, both may share an understanding of others' intentions, though level-2 VPT demands stronger cognitive control. The proposed three-stage model offers a unified framework for understanding level-1 and level-2 VPT. Future research should focus on developing experimental paradigms to dissociate these stages, employing high-temporal-resolution techniques to examine the temporal dynamics of the model, and further investigating the triggering conditions of embodied mechanisms in level-2 VPT, as well as cross-modal integration.

Keywords: visual perspective taking, two-systems account, single-system account, spatial cognition