分类号：B842　　单位代码：10335

密　级：　　学　　号：11339009

浙江大学

博士学位论文

中文论文题目： 视觉加工中运动信息的层级表征

英文论文题目： **Hierarchical representation of visual motion**

申请人姓名：徐昊骅

指导教师：沈模卫

合作导师：周吉帆　高在峰

专业名称：应用心理学

研究方向：工程心理学

所在学院：心理与行为科学系

论文提交日期 **2018.06.05**

视觉加工中运动信息的层级表征

论文作者签名:_____

指导教师签名:_____

论文评阅人 1：(隐名)

评阅人 2：(隐名)

评阅人 3：(隐名)

评阅人 4：(隐名)

评阅人 5：(隐名)

答辩委员会主席：张智君　教授　浙江大学

委员 1：周欣悦　教授　浙江大学

委员 2：郑全全　教授　浙江大学

委员 3：梁君英　教授　浙江大学

委员 4：高在峰　教授　浙江大学

委员 5：沈模卫　教授　浙江大学

答辩日期：2018.06.05

Hierarchical representation of visual motion

Author's signature:_____

Supervisor's signature:_____

External Reviewers: (Anonymity)

(Anonymity)

(Anonymity)

(Anonymity)

(Anonymity)

Examining Committee Chairperson:

Prof. Zhijun Zhang, Zhejiang University

Examining Committee Members:

Prof. Xinyue Zhou, Zhejiang University

Prof. Quanquan Zheng, Zhejiang University

Prof. Junying Liang, Zhejiang University

Prof. Zaifeng Gao, Zhejiang University

Prof. Mowei Shen, Zhejiang University

Date of oral defence： 2018.06.05

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 **浙江大学** 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：　　签字日期：　　年　　月　　日

学位论文版权使用授权书

本学位论文作者完全了解 **浙江大学** 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 **浙江大学** 可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：　　导师签名：

签字日期：　　年　　月　　日　　签字日期：　　年　　月　　日

致谢

不觉间已在浙大呆了近十年，一路走来身边常伴良师益友，乃是这一段人生中最大的幸运。

我的导师沈模卫教授是一位治学严谨、待人宽厚的学者。从踏进心理学大门到如今博士毕业的所有阶段，沈老师都给予我悉心指导和关怀。在研究中，沈老师深刻的思想往往使我在困境中豁然开朗；在生活中，沈老师无微不至的关怀总能让我感受到实验室如同家一般的温暖。这份帮助和鼓励将永存我心，并成为未来人生路上重要的内心支柱。

感谢高涛师兄将我引领进认知建模的大门，并竭尽全力将所学传授给我。即使有十多个小时的时差阻隔，高涛师兄仍坚持每周数次组织我们讨论学习，帮助我推进研究、修改论文。他的坚韧、热情和智慧总能在逆境中点燃我的内心，很荣幸能在未来几年和高涛师兄共同奋斗。

在研究生生涯初期得到诸多师兄的指导是我极大的幸运。周吉帆师兄从不吝啬时间和精力为我答疑解惑，给予我研究上最大的支持；尹军师兄亲手指导我入门初期的多项研究，并始终以合作形式继续为我提供帮助；高在峰师兄在我内心迷茫的时期点醒了我，他顽强拼搏的作风也深深感染了我；丁晓伟师兄不断给予我肯定和鼓励，他博学踏实的风格令人折服。十分感激诸位师兄对我的照顾。感谢唐宁、史博皓、程少哲、赵阳、安玮、蒋瑞峰等建模组全体成员的共同努力，使得建模组从无到有成为了今天的模样。感谢我的徒弟王曼华，师傅从你身上学到了很多。感谢吴凡、鲁溪芊、杨桐、许晓东、赵洋帆、张琼寒、俞陌桑等好友，我们一起度过的愉快生活是最美好的记忆。

特别感谢我的父母给予我最无私的支持和肯定，谢谢你们！

感谢我的妻子胡晶晶，在生命最美好的年华让我遇见了你，如黑暗中遇见光明。很幸运始终有你陪伴至今，并许下一生的承诺，我爱你。

感谢时光，予我经历，助我成长。愿今后人生路上的每一步，都无愧于这些年所受教诲，无愧于博士之名。

徐昊骅　2018 年 6 月于浙大心理系 509

摘要

日常生活中人类即使完成最简单的任务，也必须与外部环境实现灵活的动态交互，譬如在拥挤的超市中前往特定的货架取下一袋奶粉，或是在篮球比赛中避开对手的阻扰，准确地将球传给队友。上述交互行为均以视觉系统对运动信息感知、理解和预测的能力为基础。强大的视觉加工能力或许可以通过基于丰富认知资源的大量计算加以实现（如人工智能中的强化学习和深度学习），然而数十年来的认知心理学研究表明，人类的认知资源极其有限，这意味着人类的视觉智能不可能通过大量复杂的运算来实现，而是需要以相对简单的运算处理丰富的信息。信息的运算极大地依赖于信息的表征方式(Marr, 1982)，因此，建立高效的信息表征是实现人类视觉智能的关键所在。运动客体并非独立存在于视觉场景中，而是与其他客体或背景具有紧密的联系，对运动信息的加工往往依赖于由相互联系的视觉对象整合而成的整体，因此视觉系统需要为运动信息构建具有整体结构的表征。在诸多可能的整体表征中，层级结构能够在不同的层级水平上描述运动对象，具有以简单形式表达丰富信息的特征，与视觉表征的核心需求高度契合(Xu, Tang, Zhou, Shen, & Gao, 2017)。笔者据此假设，运动信息在视觉加工中以层级结构加以表征。

本研究将心理物理学研究方法与计算建模的方法相结合，以三项研究共八个实验，测量运动信息潜在层级结构的变化对人类行为绩效的影响，并采用不同类型的计算模型模拟相同的任务过程，比较其模拟结果与人类行为模式的异同，旨在检验运动信息在视觉加工中是否存在层级表征，并进一步探讨其特征。主要结果如下：

（1） 运动信息潜在层级结构的变化影响被试的行为绩效，表明视觉加工中形成了对运动信息的层级表征。

（2） 视觉对运动信息的层级表征不受运动信息的时间长度、任务线索和任务目的的影响，具有跨情景的一致性。

（3） 带有社会信息的运动，同样能在视觉加工中以层级结构加以表征。

（4） 视觉对运动信息的层级表征具备因果性，不仅描述了运动的形式，同

时描述了运动的产生过程。

（5） 视觉加工基于所构建的层级表征，通过逆向工程的计算完成对运动场景的识别、理解和预测。

上述结果不仅为视觉加工中运动信息层级表征的存在提供了坚实的证据，同时揭示了该层级表征的稳定性和普遍性。此外，本研究的计算模型进一步模拟了视觉系统利用层级表征执行后续加工的过程，为现有人工智能系统向人类智能逼近提供了有益的尝试。

关键词：运动信息，层级表征，贝叶斯建模，人工智能

Abstract

In daily life humans should interact flexibly with the external environment even if the task is simple, such as taking a bag of milk in the crowded supermarket, or passing the ball to the teammate accurately in basketball games. The basis of the above interaction is the ability of the visual system to perceive, understand, and predict of motions. Such amazing ability of vision may come from massive operations based on rich cognitive resources. However, cognitive science researches over the past decades have shown that human cognition resources are extremely limited. This means that visual intelligence cannot be realized purely by amounts of operations of big data, instead, it should process the visual information with simple operations. Visual process often depend rather critically on the particular representation that is employed (Marr, 1982). Therefore, constructing efficient information representation is the key to achieving human visual intelligence. In the motion scenes, object does not exist in the visual scene independently, but has some connection with other objects or environments. The visual processing of motion often relies on the integration of interconnected visual objects. Therefore, the vision system needs to represent motion as integrate structure. Among the possible integrate representations, hierarchical structure makes it possible to describe information in different levels, which could express rich information in simple form. Such feature of hierarchical structure is highly compliant with the core requirements of visual processing (Xu, Tang, Zhou, Shen, & Gao, 2017). Based on this, we propose hypothesis that the motion in visual processing is represented as hierarchical structure.

Current study combines methods of psychophysical research and computational modeling. Human performance to motions with different potential hierarchical structure were measured. At the same time, simulation results from different models were compared with human performance. The results showed that:

(1) The change of the latent hierarchical structure of motion influences the

performance of the participants, which indicates visual hierarchical representation of motion.

(2) The visual hierarchical representation of motion is stable and is not affected by information length, task cues and other factors.

(3) The visual hierarchical representation of motion exists also in scenes that contain social information.

(4) The visual hierarchical representation of motion is a kind of causal structure. It not only describes the form of movements, but also describes the generate process of movements.

(5) Based on the constructed hierarchical representation, the visual system recognizes, understands and predicts the motion scenes through the reverse engineering process.

The above results not only provide solid evidence for the existence of visual hierarchical representation of motion, but also reveal the stability and universality of hierarchical representation. In addition, the computational model of current study further simulates the process of performing subsequent processing using the hierarchical representation, providing a useful attempt for the artificial intelligence system to approach human intelligence.

Keywords: motion, hierarchical representation, Bayesian modeling, artificial intelligence

目录

1 引言

人们所处的环境不仅充斥着大量的视觉对象，且对象本身往往处于运动状态。视觉系统需要准确感知和理解这些运动信息，以维持对外部世界的稳定认知，从而实现与外界的动态交互。在过去数十年内，包括注意瞬脱(Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992)、注意视盲(Rensink, 2002; Simons & Levin, 1997)、视觉工作记忆(Luck & Vogel, 1997)以及多客体追踪(Flombaum, Scholl, & Pylyshyn, 2008; Pylyshyn & Storm, 1988)等方面的大量研究显示，人类的视觉系统在注意、记忆及其他高级认知能力方面的资源极其有限，仿佛暗示着人类的视觉加工能力受到了极大的限制。实际上，认知资源的受限并未在根本上损害人类的视觉加工能力，恰恰相反，视觉系统在资源有限的情况下仍能对真实世界中复杂的视觉输入进行高效灵活的处理。例如我们能够在拥挤的超市中顺利地前往货架取下一袋奶粉，或是在篮球比赛中避开对手的干扰，准确地将球传给队友。人类视觉的这种能力引起了心理学和人工智能（如：计算机视觉、模式识别等）领域研究者的兴趣，近年来人工智能方面最先进的模型开始尝试模拟人类注意和记忆的算法(Caicedo & Lazebnik, 2015; Karpathy & Li, 2015)，并获得了模型绩效的显著提升。然而上述模型仅仅是将人类视觉的一些外显特性迁移到算法中，未能指出这些视觉特性背后的信息加工基础——资源有限条件下，视觉需要通过简单的运算过程处理丰富的信息，而为信息建立高效的表征正是简化运算过程的重要前提。因此，视觉研究者一直在探索视觉加工中信息的表征形式，以及基于表征的运算过程。

1.1 运动信息的视觉加工特性

视觉系统对运动信息的有效加工，是人类得以实现与外界动态交互的基础，大量研究探讨了运动信息的视觉加工机制。笔者首先简单介绍这部分相关研究，以便更好的理解人类视觉系统加工运动信息的特点。

1.1.1 运动信息的知觉与记忆

我们的视觉系统不仅能够处理静态的视觉场景，还能够理解动态的视觉场景，

将客体在空间位置上随时间的变化知觉为运动。运动知觉很难被看作视觉对静态场景加工的简单叠加：尽管视觉往往被比喻为照相机，但实际上视觉系统并没有如同快门一样硬件装置，并且动态的场景也并未使得对外部环境的认知变得一片模糊(Johansson, 1975)。相反，运动知觉的存在意味着视觉的本质是一个智能的计算系统，将视觉对象的时空信息整合为运动，而非离散的静态图像。

运动知觉的现象在大量情景中广泛存在，观察者不仅能够知觉到二维平面中的运动信息，还能够从二维的视觉刺激中知觉到三维的运动信息(Johansson, 1975)，譬如在二维平面上呈现一个做椭圆运动的光点，观察者会将其报告为一个在倾斜平面上做圆周运动的光点（图 1.1a）；在二维平面上缩放的四边形，会被观察者报告为一个大小不变的四边形在深度上靠近或远离（图 1.1b）。即使运动对象的部分信息被遮挡，视觉系统也能知觉到完整的运动信息（如：狭缝知觉、孔径效应等）(Wallach, 1935)。大量存在的运动知觉现象表明，运动信息的视觉加工是人类视觉系统的基本能力之一。

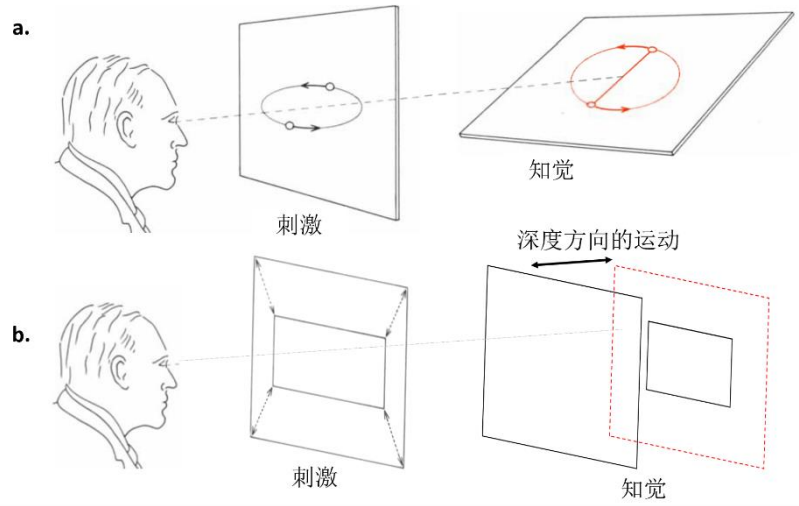

图 1.1 视觉能够从二维平面呈现的运动刺激中知觉到三维的运动信息：（a）平面上做椭圆运动的光点，被知觉为在斜面上做圆周运动；（b）平面上缩放的矩形，被知觉为在深度方向运动(Johansson, 1975)。

对运动信息的维持主要与工作记忆有关，后者是人类实时加工和存储信息的核心系统(Baddeley, 2012)。研究发现，对运动信息的记忆主要受到运动对象数量的影响，当仅需要记忆一个客体的运动方向时，被试能够几乎完美的完成任务，即使在运动信息消失后间隔很长时间，甚至插入一些其他任务，被试仍然能够在

检测阶段报告出单个客体的运动方向，然而当需要同时维持并报告多个客体的运动信息时，被试对方向的判断(Blake, Cepeda, & Hiris, 1997)和对运动方向精度的记忆迅速下滑(Narasimhan, Tripathy, & Barrett, 2009; Shooner, Tripathy, Bedell, & Ogmen, 2010; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011)。运动信息本身的特性同样会对记忆产生影响，速度较快的运动更难被维持在记忆中(McKeefry, Burton, & Vakrou, 2007)。运动记忆的相关研究表明，视觉系统具有记忆运动信息的能力，但这种能力受到认知资源的限制。

1.1.2 运动对象的追踪

为完成与动态环境的交互，视觉系统不仅需要感知运动信息，还要维持对运动对象的持续追踪，这同样是视觉的基本能力之一。研究发现对运动对象的追踪能力主要与平滑追踪眼动有关(Krauzlis, 2004)，并且与运动知觉一样发生在视觉加工的早期阶段，具有相同的神经活动基础(Ilg & Churan, 2004)。然而，在特定的情境下，维持对运动对象的追踪，可能会利化或损害对整个场景的运动知觉(Spering & Montagnini, 2011)，这意味着在共享绝大部分认知资源和神经机制的基础上，追踪运动对象可能有其独特的认知加工机制存在。

视觉系统对运动对象的追踪不局限于单个客体。Pylyshyn 和 Storm (1988)首次对视觉同时追踪多个运动客体的能力进行了探索，在他们创造的多客体追踪范式（Multiple Objects Tracking，MOT）中，多个静止的客体被同时呈现给被试，其中一部分客体被特定的线索标记为目标，接下来的运动阶段中所有客体外形特征均相同，被试需要在运动阶段中保持对目标的追踪，并在运动结束后报告它们的位置。采用这一范式的一系列研究结果显示，被试能够同时追踪 4-5 个客体的位置(Allen, McGeorge, Pearson, & Milne, 2006; Alvarez, Arsenio, Horowitz, DiMase, & Wolfe, 2005; Alvarez & Cavanagh, 2005; Pylyshyn & Annan, 2006)，其追踪绩效主要受到与运动相关的时空因素的影响，速度、运动复杂度、场景中的客体密度等均在多客体追踪中扮演重要作用(Alvarez & Franconeri, 2007; Clair, Huff, & Seiffert, 2010; Fencsik, Klieger, & Horowitz, 2007; Yantis, 1992)。

在更复杂的追踪任务中，多个运动的目标被赋予了身份信息，在检测阶段被试不仅需要报告该客体是否是目标，还需要判断它的位置-身份信息绑定是否正

确。这种情况下，尽管被试仍然能够维持对目标的追踪，但对目标的具体身份判断往往较为困难(Pylyshyn, 2004)。类似的，被试同样难以维持目标其他方面的具体特征（如：颜色、形状等）(Botterill, Allen, & McGeorge, 2011)，在此类任务中，被试大约只能完成对 2 个目标的加工。特征追踪能力的下降，暗示着对客体位置信息的加工和对客体特征信息的加工或许来自于两个不同的系统(Horowitz et al., 2007)。

1.1.3 从多客体运动中识别社会信息

在与真实世界交互的过程中，视觉不仅加工对象的物理属性，还需要加工其社会属性。对视觉对象社会性的知觉往往通过两个方面进行：由客体的外表信息获得社会属性（如：面孔、身体形状）或由客体的运动信息获得社会属性。后者的加工可以完全不依赖于前者：即使简单的几何形状，也有可能仅通过运动模式被知觉为带有意图的生命体。例如在简短的运动片段中（图 1.2），观察者将其内容描述为“三角形正试图抓住圆形”(Heider & Simmel, 1944)或“矩形翻越障碍物追捕圆形”(Michotte, 1950)。

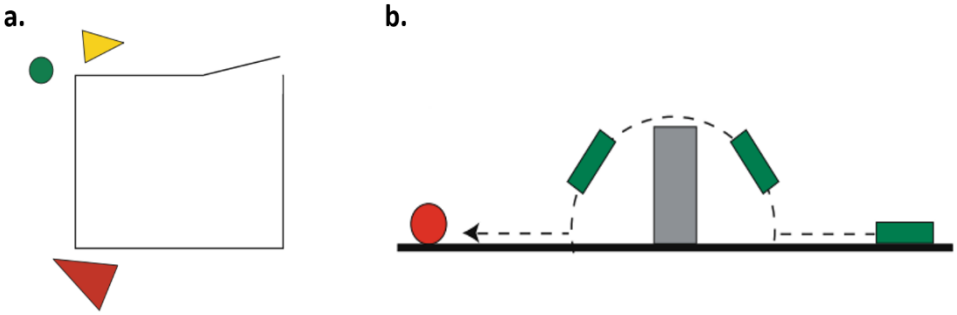

图 1.2 视觉从运动的简单几何形状中知觉到生命体运动：（a）红色的三角形追捕黄色的三角形，绿色的圆形试图掩护黄色三角形(Heider & Simmel, 1944)；（b）绿色的矩形试图翻越障碍物追逐红色圆形(Michotte, 1950)。

视觉通过运动对象识别生命体的能力受到不同领域研究者的关注(Schultz, Friston, O’Doherty, Wolpert, & Frith, 2005; Southgate & Csibra, 2009)，运动对象的许多特征为生命体知觉提供了线索，包括由于违背基本力学定理而被知觉为自带驱力(Dasser, Ulbaek, & Premack, 1989)、运动方向或速度的突然改变(Tremoulet &

Feldman, 2000)、运动方向与目标方向的偏离程度(Gao, Newman, & Scholl, 2009; T. Gao & Scholl, 2011)等，不仅如此视觉甚至能够识别出简单运动中复杂的合作或竞争关系(Yin et al., 2013)，并利用获得的社会信息进行行为规划和预测(Kaelbling, Littman, & Cassandra, 1998; J. Yin et al., 2016)。视觉从运动中获取意图以及社会交互信息的能力，体现了人类视觉系统的智能优越性，同时也暗示着，运动信息的视觉加工不仅仅是对视觉对象时空信息的简单整合，而是依赖于更加智能高效的表征和计算过程。

1.2 视觉加工中的表征问题

认知心理学认为，对人类信息加工过程的探索包含两个重要的问题：信息加工具有哪些阶段，以及信息是如何被表征的(Neisser, 1967)。表征是指人类对外部信息的内在描述，是决定具体信息运算过程的重要前提(Marr, 1982)，对表征的探索，是对人类行为进行认知解释的重要基石(Pylyshyn, 2000)。基于此，大量研究在视觉运动加工特性的基础上，进一步探讨了运动信息在视觉加工中的表征。

1.2.1 基于客体的视觉加工

近年来，基于客体的视觉加工的观点得到了视觉领域研究者的普遍认同。客体是由不同维度结合而成的，具有一致性、完整性等特点客观对象。基于客体的加工观点来自于大量研究的强有力支持：在客体特征辨别的任务中(Duncan, 1984)，被试需要报告可能来自于同一个客体或不同客体的两个特征，客体均呈现在相同的空间位置，结果发现当两个特征来自于同一个客体时，被试的反应时更快、正确率更高，表明注意的选择对象是视觉客体。客体的形成可以不受空间和特征的限制，即使分布在多个空间位置的特征或部件，也能被绑定为整合客体(Driver, Davis, Russell, Turatto, & Freeman, 2001)。工作记忆方面的研究同样支持基于客体的加工观点，视觉工作记忆的容量大约是 3-4 个客体，在容量限度内，被试的记忆绩效不受特征数量的变化，但增加客体数量将导致记忆绩效的显著下降，该结果表明视觉工作记忆同样以客体作为存储和加工的基本单元(Gao, Yin, Xu, Shui, & Shen, 2011; Luck & Vogel, 1997; Yin et al., 2012)。特征整合理论对客体的形成过程进行了详细的描述，该理论认为，在前注意阶段，视觉场景中的特

征被划分为多个部分，这一过程可以并行实现，而在注意阶段，每一个部分的特征被整合成为一个客体，这一过程只能以线性的方式进行(Treisman & Gelade, 1980; Treisman & Schmidt, 1982)。

运动信息视觉加工的相关研究同样支持基于客体的观点。基于 MOT 范式的研究发现，在多客体追踪任务中增加一定的障碍物，使得客体从障碍物下放穿过时不能被持续知觉到，被试的追踪绩效并不受影响。然而，如果客体在进入和离开障碍物时的消失和出现方式不符合客体的基本概念（包括整体同时消失/出现、远离障碍物端优先消失/出现、缩放式消失/出现等），被试将难以继续维持对运动客体的追踪(Scholl & Pylyshyn, 1999)。另有研究将追踪任务中的目标和干扰子构造为同一客体，譬如用线将目标和干扰子一一对应的连接起来，被试被要求追踪线段的其中一端，此时被试的追踪绩效大幅下降，这表明视觉无法脱离客体而去追踪运动的特征，即运动对象的追踪是基于客体的(Scholl, Pylyshyn, & Feldman, 2001)。

基于客体的加工理论及其进一步研究为揭示视觉加工机制做出了重要贡献，然而无论在知觉、记忆或是追踪任务中，视觉系统能够处理的客体数量极其有限，以客体作为基本单元的加工观点，不足以解释视觉如何在丰富的场景中表现出高效灵活的加工能力，在客体之上构建更有计算优越性的整体表征，是理解视觉智能的重要基础。

1.2.2 视觉中的信息整合

视觉元素并非无序的处于场景中，而是彼此存在紧密的结构和组织关系(Braund, 2008; Pomerantz & Kubovy, 1986)，视觉系统同样会根据对象间的关系对其进行组织加工(Wertheimer, 1923)。完整的组织所产生的视觉体验远非单个客体视觉结果的简单叠加，而是包含对象间的关系信息，以整体的形式参与视觉过程，表现出整体大于部分之和的特点，对心理学有着深远影响的格式塔心理学便以此观点为核心(Köhler, 1967)。

包括格式塔心理学在内的一系列视觉研究揭示了大量引发组织的线索，在底层物理特征方面，客体在空间位置上的邻近性，以及其他特征上的相似性往往是产生组织的最主要因素。同时，客体高层的概念信息也能够作为产生组织的线索，

例如功能上存在联结的客体（锤子和钉子）能够被组织在一起(Humphreys & Riddoch, 2001; Kaiser, Stein, & Peelen, 2015; Laverick et al., 2015)，含义上存在关联的词语同样能形成组织(Zhou, Zhang, Ding, Shui, & Shen, 2016)。值得注意的是，尽管大量的研究主要探索了组织在知觉阶段的产生和加工机制，但组织的形成并非只能够发生知觉阶段，对于分别进入工作记忆并从当前视野中消失的信息，工作记忆能够对其进行动态加工，使其在记忆内部形成组织(Gao, Gao, Tang, Shui, & Shen, 2016; Shen et al., 2015)。

在对运动信息进行加工时，视觉是否同样能够将多个运动对象组织成一个整体呢？研究表明，人类的运动知觉确实表现出了对运动信息进行整体加工的特点。在包含多个运动对象的场景中，对运动的知觉和追踪往往沿着运动信息的平均方向(Lisberger & Ferrera, 1997)。即使在要求对特定对象的运动进行加工时，如果对象周围存在与对象相同的运动，将会利化对目标对象的运动知觉，反之如果对象周围存在与对象不同的运动，则对目标对象的运动知觉受到损害(Masson, Proteau, & Mestre, 1995; Spering & Gegenfurtner, 2007)。有时，缺乏背景信息反而为运动知觉带来了困扰，例如在孔径问题(Wallach, 1935)中，运动的光栅透过小孔呈现出一部分信息，观察者很难准确判断出光栅的真是运动方向，因为能够造成相同运动现象的真实运动不止一种。只有当给予更多的背景信息时，运动知觉才能进行准确的判断(Beutter & Stone, 2000)。许多经典的运动视觉现象也为运动加工中的信息整合提供了证据，在呈现背景运动的情况下（图 1.3），观察者往往将运动描述为“整体运动+局部运动”的形式(Duncker, 1929)。此外，生物运动的相关研究也表明，视觉不仅能够将离散运动的光点知觉为生命体的运动(Blake & Shiffrar, 2007; Johansson, 1973)，还能够从多个运动光点的集合中知觉出交互运动信息(Neri, Luu, & Levi, 2006)。

对运动对象的追踪研究同样证实了信息整合的存在。研究发现，对多个运动目标的追踪导致目标之间的距离发生了压缩，产生与知觉组织相似的扭曲效应(Liverence & Scholl, 2011)。当目标与干扰子存在诱发知觉组织的线索时（如：共同运动、追逐等），对目标的追踪绩效受到破坏(Suganuma & Yokosawa, 2006)。在一项基于 MOT 任务的研究中，研究者操纵了目标与目标、目标与干扰子在多种不同的底层物理特征方面的相似性，以构建基于不同客体的知觉组织，结果显示

当目标之间有更强的组织线索时，目标的追踪绩效显著提高，而当目标与干扰子间形成知觉组织时，这一发生在组织上的混乱极大损害了目标的追踪绩效(Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013)。另有研究发现，将类别信息等高层的特征作为分组线索，能够引发相同的效应(Endress, Korjoukov, & Bonatti, 2017)。此外，近年来有研究显示，观察者还能够利用社会交互信息对运动的多个对象进行组织(Yin et al., 2016; Yin et al., 2013)。

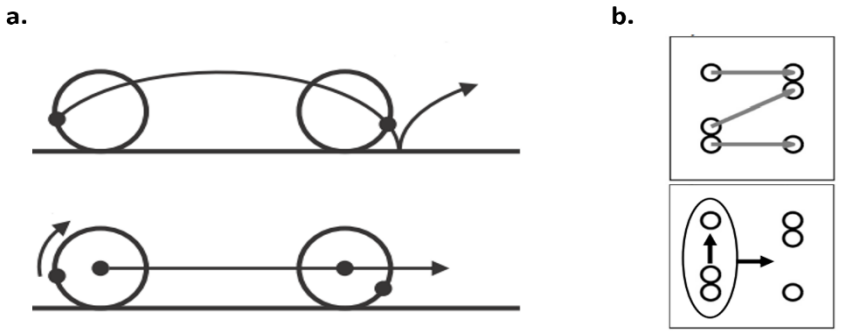

图 1.3 对运动的知觉依赖于整体信息：（a）图的上半部分为光点的真实运动，在两个光点同时呈现时，被描述为如图下半部分展示的滚轮运动(Duncker, 1929)；（b）图的上半部分为三个点的真实运动，观察者均将它们描述为“整体向右运动，中间的光点向上运动”(Johansson, 1973)。

对于多个运动客体所形成的组织而言，由于客体的空间位置不断改变，组织的稳定性似乎相对较弱，因此有研究提出，视觉对多个运动客体的加工还依赖于在运动过程中相对稳定的构型。Yantis（1992）的研究显示，当运动的目标发生重叠或穿越时，由于多个目标围成的多边形发生了本质变化，导致被试的追踪绩效受到破坏。近年来，有研究针对性的从运动客体空间构型的几何形状结构入手，考察不同程度的结构变化对多客体运动加工的影响，结果发现当构型发生拓扑结构变化时（一个目标穿过另外两个目标之间的连线），对目标的追踪绩效几乎不受影响，但对目标身份的识别绩效急剧下降(Zhao et al., 2014)，而在记忆任务中，构型的投射变化和拓扑变化均会导致对多个客体运动方向的记忆绩效下降(Sun et al., 2015)。

上述研究表明，包括构型在内的组织结构是视觉加工的重要成分，基于客体间关系建立的结构性表征，是人类视觉智能性的基础。结构性表征在计算上可能

有不同的表达形式，一种形式是将结构表达为马尔科夫随机场，它假设每一个对象与相邻的对象存在关联，同时在考虑与相邻对象关联的情况下，不再与其他对象存在关联（图 1.4a）。这种表征方式能够很好的模拟被试对大量客体记忆时的表现(Brady & Tenenbaum, 2013)，同时对于多客体运动中构型影响的相关研究也是一个不错的解释，但尚未有严格的计算模型对此进行证明。同时视觉对象间仅存在相邻的关联关系是一种理想情况，与真实情况相比也许会存在一些信息损失。与之相比，全连接模型能够尽量避免信息损失，在全连接模型中，每一个对象的信息与任意两个对象的关联均被纳入表征（图 1.4b），这能够同时应对针对于单个对象的任务和针对于整个视觉场景的任务。但该模型的表达较为复杂，以大量冗余信息为代价尽量保证了信息完整，这与人类视觉资源有限的事实有所冲突。结构表征还能够被表达为层级形式（图 1.4c），在层级表征中，每个客体以终结节点（terminal node）的形式存在，它们的共同信息则被表达到这些节点共同的父节点中（parent node）。层级的表征形式在信息完整性、结构解释力、计算高效性等方面有独特的有点，在近年来的一些研究中得到来自心理学领域和计算机模型领域的认可(Froyen et al., 2015; Gershman, Tenenbaum, & Jäkel, 2016)。

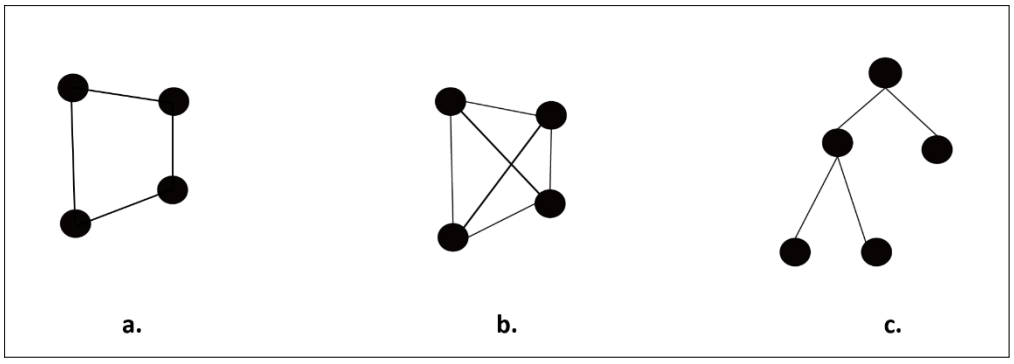

图 1.4 结构化表征在计算上的不同表达形式(Xu, Tang, Zhou, Shen, & Gao, 2017)：（a）马尔科夫随机场，每一个对象至于相邻的对象间存在关联；（b）全连接模型，任意两个对象间的关系被纳入计算表达；（c）层级模型，对象的共有信息被表达在上层节点中，独特信息被自身节点表达。

1.3 层级结构与人类认知活动的联系

1.3.1 层级结构的特点

在进一步探讨认知加工中的层级表征之前，有必要对层级结构的基本特点进行说明。层级结构的图模型由节点和路径组成，其中节点表达特定的对象或信息，路径则表达所连接的节点之间的关系。层级结构最重要的特征是形成了不同的层次：上层节点（父节点）可以连接多个下层节点（子节点），并表达同属于这些下层节点的共同信息；同属一层的节点若无特别说明，通常认为具有相同的地位。基于其结构特征，层级结构具有三方面独特的性质：层级的深度、节点在层级结构中的距离、以及层级之间的方向。

层级结构的基本特点为它在信息表征方面带来了优势，与人类认知加工紧密相关的主要优势包括以下三个方面：

第一，高效的信息表达和计算能力。层级结构中的父节点能够集中表达子节点的共同信息，信息的集成不但反应了多个对象的关系信息，也避免重复表达造成的信息冗余。同时，复杂的层级结构可以被看作简单的双层结构的重复叠加，这使得对完整层级结构的计算可以被简化为对双层结构的迭代，这一特性大大降低了层级结构的计算复杂度。近年来人工智能领域最热门的深度神经网络便是对这一优势最好的证明，尽管深度神经网络与当前探讨的层级结构存在一些差异，但在构建具有深度的结构来提高计算效率方面是共通的(Lecun, Bengio, & Hinton, 2015)。考虑到视觉过程中极其有限的认知资源，层级结构的信息共享特性非常有利于视觉的高效加工。

第二，灵活的信息抽象能力。物理世界在客观上具有层级的本质，每一个对象都由更小的部分组成，同时构成更大的对象的一部分。在主观上，视觉同样需要对场景进行不同层次的加工，比如面对行走的人群时，视觉可能需要从个体的身体部分层面、完整的一个人层面、或者具有交互的多个人组成的群体层面加工这个场景。视觉对信息进行不同程度抽象的灵活需求，在层级结构中可以被表现为不同层次的信息整合加工。因此，层级结构可以使得视觉在维持稳定的表征形式的基础上，灵活的进行不同水平的信息抽象。

第三，完备的因果解释力。层级结构是一种因果结构，它不仅描述事物之间的相关关系，也描述事物之间的因果关系。层级结构的上下层之间存在带有方向的连接，它表达了信息由上而下的产生过程，即上层节点的属性决定了下层节点的属性，对层级结构的逐级展开过程，描述了以表征为蓝图“产生”数据的因果过程。认知心理学的研究表明，提取因果关系是人类认知的核心特征之一(Sperber, Premack, & Premack, 1995)，而相关信息并不等价于因果。因此基于层级结构的表征有助于人类认知进行有效的因果推理。

层级结构的特性为其带来了高效、灵活且具备解释力的计算能力，而这正是视觉加工所必备的核心能力，可以说采用层级的表征方式天然的契合了视觉加工所表现的特性。

1.3.2 其他认知活动中的层级结构

人类语言和语法的层级表征研究有着更长的历史(Chomsky, 1964;Smith, Shoben, & Rips, 1974)。人类语言的词汇、语法规则是有限的，但却能够表达近乎无限的含义。这种基于有限使用的无线表达(von Humboldt, 1836)是语言研究最大的挑战。层级的表征方式为解答这个问题带来了可能：将有限的词汇填入有限规则定义的层级树中，基于层级结构可迭代的特性，它将可以表达近乎无限的含义。神经学的研究也为此提供了证据，大脑电信号的频谱分析结果显示，听到语言时，在字、短语、句子等多个层次的对应频率上均发现显著较高的能量激活(Ding, Melloni, Zhang, Tian, & Poeppel, 2015)。与语言相关的认知能力方面，概念学习同样被认为是以层级表征的方式进行的。研究表明婴儿将学到的新概念组织到层级表征结构中，形成完整并可拓展的概念体系(Johnson & Keil, 2014; Kiley Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013)。此外，婴儿和成人均在事件分割方面表现出层级结构的特性，在不同层面上对连续发生的事件进行切分(Baldwin, Baird, Saylor, & Clark, 2001; Zacks & Swallow, 2007)。

近年来一些研究开始探索视觉加工中的层级表征。有观点认为，整个视觉系统在本质上是以层级结构为基础的(Biederman, 1987)，从最基础的特征识别(Palmer, 1977)、客体的形成(Kahneman, Treisman, & Gibbs, 1992)、到更抽象层面

的类别概念表达(Ullman, 2007)均发现了层级表征的证据。在工作记忆(Brady & Tenenbaum, 2013; Lew & Vul, 2015)和知觉整合(Froyen et al., 2015)方面的研究发现，视觉对象以层级的方式进行聚类整合，暗示着视觉可能将客体聚类形成的层级结构作为加工整体。最近，有研究基于层级结构为视觉的知觉组织现象构建了计算模型，在该模型中，每一个对象以及组成它的部分被以父节点-子节点的关系表征，通过在不同水平上进行聚类，为整个视觉场景形成了层级表征，通过贝叶斯推理，模型能够推断视觉对象的组成成分和结构关系。结果表明，基于层级表征的模型能够胜任经典的知觉组织、边界探索、部分分解等问题(Gershman et al., 2016)。

在不同的认知活动中建立相似的层级表征，为视觉和语言的双向交互带来了可能。对于人类来说，将看到的视觉场景转化为语言，或为听到的语言描述重构一个相似的视觉场景是相对容易的(Gorniak & Roy, 2004; Heider & Simmel, 1944; Jackendoff, 1996; Talmy, 1988)。然而，鉴于视觉的输入本身是大量底层的特征，而语言是相对抽象的结果，这一视觉语言的鸿沟始终未能被清晰探明。视觉和语言在某种程度上采用相似的层级表征方式（尽管在语言语法和视觉语法的规则上会有所不同），能够对认知加工顺利跨越这一鸿沟具有促进作用。

当前的计算机视觉系统也得到了来自层级表征的启发。Zhu 的一系列研究(Zhu & Mumford, 2006; Zhu, 1999)采用一种称为“与或树”的层级结构对视觉场景进行表征，这种结构和语言中的生成语法树极为相似(Chi & Geman, 1998)，能够基于有限的规则，将视觉对象表达为其部分的组合，但不依赖于组成对象的部分特定的特征。例如，一张桌子可以分为桌面和桌腿，但桌面和桌腿的形状、颜色甚至数量可以发生变化。在运动信息的视觉加工方面，Gershman 等人基于层级表征构建了运动知觉的计算模型(Gershman et al., 2016)，通过贝叶斯推理，模型在模拟一系列经典的运动视觉现象(Duncker, 1929; Johansson, 1973)方面得到了与人类知觉相似的结果。

1.4 问题提出

由前文可见，视觉能够整合客体的时空信息，形成对运动信息的感知。然而，运动客体并非独立地存在于视觉场景中，而是与其他客体或背景具有紧密联系(Braund, 2008; Pomerantz & Kubovy, 1986)，对运动信息的加工往往依赖于由相互联系的视觉对象整合而成的整体(Yantis, 1992; Yin et al., 2013)，因此视觉系统需要为运动信息构建具有整体结构的表征。在诸多可能的整体表征中，层级结构能够在不同的水平上描述运动对象，具有以简单形式表达丰富信息的特征(Froyen, Feldman, & Singh, 2015)，可为资源有限条件下视觉智能的实现提供有力支持(Marr, 1982)。此外，有数项研究表明，人类在语言加工(Chomsky, 1964)、概念学习(Tenenbaum, 1999)、场景分割(Zacks & Swallow, 2007)等许多认知活动中均采用层级表征形式。在不同的认知加工过程中采用相对统一的表征，有助于知识在整个认知系统中保持稳定且易于迁移(Joo, Wang, Zhu, & Wagemans, 2012)，这正是人类智能强于当前人工智能系统的突出优势之一(周吉帆等, 2016)。

物理世界在本质上具有层级的组织结构：每一个对象都由更小的部分构成（如：原子由原子核和电子构成），同时每一个对象也能和其他对象共同构成更大的整体（如：原子能够和其他原子一起组成分子）(Froyen, Feldman, & Singh, 2015)。运动信息同样具有层级的组织结构，客体运动可用相对运动的方式分解为与参考系相同的运动和相对于参考系的相对运动，并且这种基于参考系的分解可不断迭代，形成运动的层级结构。人类对行星运动的描述便是一个最好的例子：月球相对于地球做圆周运动，地球连同月球一起相对于太阳做圆周运动，它们共同构成了一个具有层级结构的运动系统。上述分析表明，运动信息本质上可以被表示为层级结构，这为视觉建构运动信息的层级表征提供了可能性。笔者据此假设，运动信息在视觉加工中以层级结构加以表征。本研究将检验这一假设，并进一步探讨层级表征的特征。

1.5 研究构思

1.5.1 构建具有层级结构的运动形式

本研究拟通过心理物理学实验的方法，操纵运动信息背后的潜在层级结构，以观察其对人类行为绩效的影响。笔者首先简单介绍构建具有特定层级结构的运动形式的基本方法，其详细数学计算过程见附录一。

产生具有特定层级结构的运动信息包括两个关键步骤：生成特定的层级结构，以及为层级结构中的节点赋予运动向量。（1）层级结构的产生通过“嵌套中国餐馆过程”（nested Chinese Restaurant Process）实现(Blei, Griffiths, Jordan, & Tenenbaum, 2004)，该过程是单层中国餐馆过程的迭代形式，能以概率的形式描述每一层节点扩展到下一层节点的过程，每一个客体将被分配到层级结构的任意一个终结节点上。基于此过程可以得到特定层级结构的产生概率。（2）将运动向量赋值到节点的过程通常由高斯过程实现(Rasmussen & Williams , 2006)，该过程一方面满足在没有其他假设的情况下，运动具有一切可能的自由度。同时，也考虑到在连续时间上发生运动突变的概率，避免所生成的运动表现出剧烈抖动甚至突变，造成违背客体基本概念的视觉体验。客体的真实运动由客体所在终结节点到层级结构根节点的路径中，所有运动向量的矢量和叠加而成（图 1.5）。

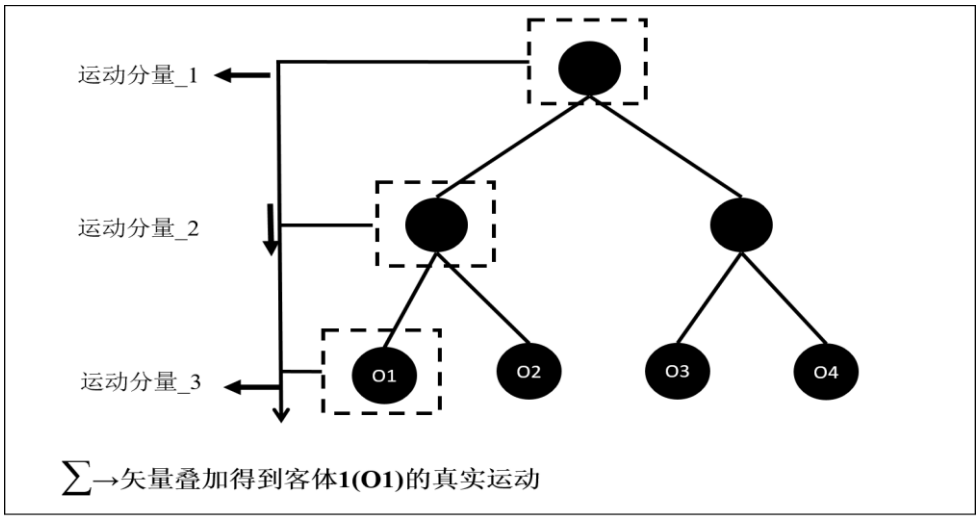

图 1.5 运动信息在层级结构中的表达：客体位于最底层的子节点中，客体间共有的运动分量被表达在上层节点中，每一个客体的真实运动，由其子节点到根节点之间路径上的所有运动分量叠加而成。

值得注意的是，仅依赖于上述过程生成的层级结构与客体运动之间并没有严格的一一对应关系，即层级结构的产生过程和给节点分配运动向量的过程均为概率过程，相同的运动可能由完全不同的层级结构产生（想象在描述月亮、地球和太阳的运动时，我们将地球作为中心，通过合适的矢量分解，能够构建一个完全不同的层级结构，并且节点上有完全不同的运动向量）。然而，以不同层级结构产生特定运动的概率大小存在差异，具有最大概率的层级结构可作为对运动信息最好的描述。因此，本研究根据贝叶斯定理计算特定运动对应的不同层级结构的概率，并将概率最大的层级结构作为该运动的潜在结构。

1.5.2 总体构思

对运动信息在视觉加工中的层级表征的探讨分为三个研究部分，总体研究框架见图 1.6：

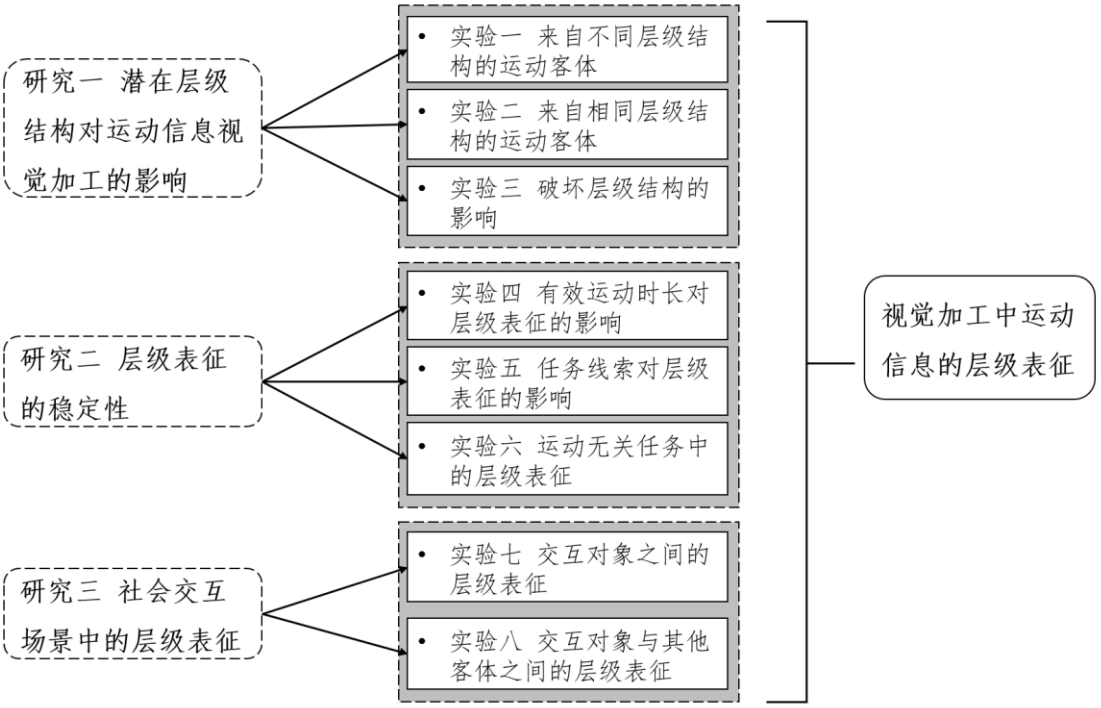

图 1.6 总研究思路结构图

研究一首先检验运动信息在视觉加工中是否存在层级表征。该部分研究从层级结构的特点入手，分别设计了具有不同层级结构的运动信息（实验一），和具有相同层级结构的运动信息（实验二），考察运动客体在层级结构的深度、距离以及方向三方面的变化，对行为绩效的影响。同时，每一个实验中均进行计算模

型的模拟，比较两类不同模型（层级模型和相关模型）在任务绩效上与被试反应模式的异同。实验三则从反方向入手，在保持运动的物理特征不变的情况下，破坏其潜在的层级结构，观察行为反应所受到的影响，进一步确认运动信息在视觉加工中层级表征的存在。

研究二通过调节可能对层级表征存在影响的相关因素，考察视觉加工中运动信息层级表征的稳定性。实验四操纵运动信息的时间长度，探查层级表征的构建是否受高级认知过程的影响。如果层级结构不受信息时间长度的影响，则暗示着层级表征的构建不是策略等高级认知过程的产物。实验五操纵了任务相关的线索，考察层级表征的构建是否依赖于加工多个运动对象的任务导向。实验六构造了与运动信息完全无关的任务，进一步验证层级表征的构建仅仅依赖于运动信息，而与任务本身无关。研究二通过三个实验，试图说明视觉加工中运动信息层级表征的稳定性。

研究三将上述实验情景进一步拓展到社会场景中，试图考察视觉系统能否为具有社会交互意图的运动信息构建层级表征。以追逐场景为例，研究构造了两个实验：实验七操纵追逐者和目标之间的层级结构，测量被试对运动场景中社会意图的识别能力；实验八则构建了追逐者（与社会交互相关的对象）和干扰子（与社会交互无关的对象）之间的层级结构，同样测量被试对运动场景中社会交互意图的识别能力。通过两个实验研究三试图证明，视觉系统能够为具有社会意图的运动信息构建层级表征。同时，实验七与实验八中同样辅以计算模型对任务进行模拟，通过比较两类模型模拟结果与被试行为模式的异同，为层级表征的存在提供进一步证据，并进一步揭示具体任务中基于层级表征的视觉计算过程。

1.6 研究意义

对运动信息的视觉加工是实现人类与外界动态交互的基础，也是视觉领域重要的研究主题之一。人类认知资源的有限性表明，强大的视觉加工能力不可能通过基于大数据的海量运算来实现，而是需要以相对简单的运算过程处理丰富的信息。因此，资源有限条件下的视觉智能是如何实现的，一直是视觉研究领域的重要问题之一。本研究提出并验证了视觉加工中运动信息的层级表征，为实现视觉智能所需的计算特点提供了表征基础，初步回答了上述问题，一定程度上促进了

对视觉计算过程的理解。

同时本研究所揭示的层级表征有着完备、具体的数学表达，可直接应用于人工智能算法，有助于当代人工智能从“大数据、小任务”到“小数据，大任务”的转变，向着人类智能优越性的方向前进(唐宁等, 2018)。

2 研究一：潜在层级结构对运动信息视觉加工的影响

该部分研究主要探索运动信息潜在层级结构的变化对视觉加工任务的影响。研究构造了基于两个或三个客体的运动信息，主要基于以下两方面的考虑，以保证实验的效度：一方面，对层级结构的针对性探究主要着手于其独特的三类性质，即深度、距离和方向，基于两个或三个客体的层级结构能够较好的在操纵其中一类性质时控制另外两类性质保持一致；另一方面，层级结构与真实运动的对应关系采用最大概率原则构建，基于两个或三个客体的层级结构其概率分布离散程度较低，更好地保证层级结构与真实运动的对应。

2.1 实验一 来自不同层级结构的运动

实验一构建具有不同层级结构的运动，通过测量被试行为绩效的变化，考察运动信息的视觉层级表征。其一般逻辑是：若操纵刺激的某一物理特性导致被试行为绩效改变，则表明视觉系统加工了该物理特性。本研究中操纵的并非是刺激的外显物理特性，而是潜在层级结构。由于层级结构并不能从视觉场景中直接观测到，视觉必须形成层级结构的内在表征，才能够对其进行加工。本实验采用位置预测任务（见图 2.1），操纵了层级结构的三方面特性（见 1.3.1）：（1）深度——层级结构可以具有不同的深度，层级越深其结构越复杂，则运动客体的位置更难预测，即预测误差更大；（2）距离——在深度相同的层级结构中，客体间的层级距离（节点间路径长度）可能不同，客体的层级距离越远，其位置越难预测，即预测误差增大；（3）方向——层级结构中的客体可能位于不同的层级水平，由父节点预测子节点与由子节点预测父节点相比方向不同，由于父节点包含更多的共同信息，由父节点预测子节点更易，即预测误差更小。

2.1.1 方法

2.1.1.1 被试、设备与刺激材料

16 名（7 男，9 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。

实验在暗室中进行，采用 CRT 显示器进行刺激呈现，屏幕分辨率为 1024×768，刷新率为 100Hz，视距约 70cm，整个显示屏约为 36.6°×27.6°。

实验中的运动轨迹由 Python2.7 编写的程序生成并存储，在满足特定层级结构的基础上，保证每个客体的运动速度为 0.12°/帧，运动方向的平均变化大小约为 7.05°/帧。实验程序采用 Matlab 的 Psychtoolbox 工具箱编写(Brainard, 1997)，正式实验中客体是一个实心的白色圆 (RGB = [255, 255, 255]，直径为 1°)，运动轨迹和被试绩效的计算均以圆心为准。每一种条件下会生成 5 条不同的轨迹，每条轨迹分别以围绕屏幕中心旋转 0°、90°、180°和 270°的方式重复呈现四次，以排除客体在空间位置上分布不均匀的影响，因此每种条件下共有 20 条运动轨迹。

2.1.1.2 实验设计与流程

实验流程如图 2.1 所示，每次实验开始前首先在屏幕中央呈现一个注视点，以吸引被试的视线保持在屏幕中央，同时提示即将开始进入一个试次。0.5s 后注视点消失，同时呈现两个客体的运动，在两个客体均可见的运动（完全可见阶段）持续 4s 后，其中一个客体消失，但其仍然在继续运动，只是不能被观察到（部分可见阶段），再运动 2s 后所有客体静止，屏幕中央出现鼠标，要求被试用鼠标点击出之前消失的客体在运动结束时刻所在的位置，程序将自动记录被试的反应，并在间隔 1.5-2.5s 后进入下一个试次。实验前保证被试理解实验任务的基础上，不对运动轨迹本身做任何描述，以保证被试对实验目的不知情且未受到任何相关提示，每次被试完成点击后不给予任何反馈。

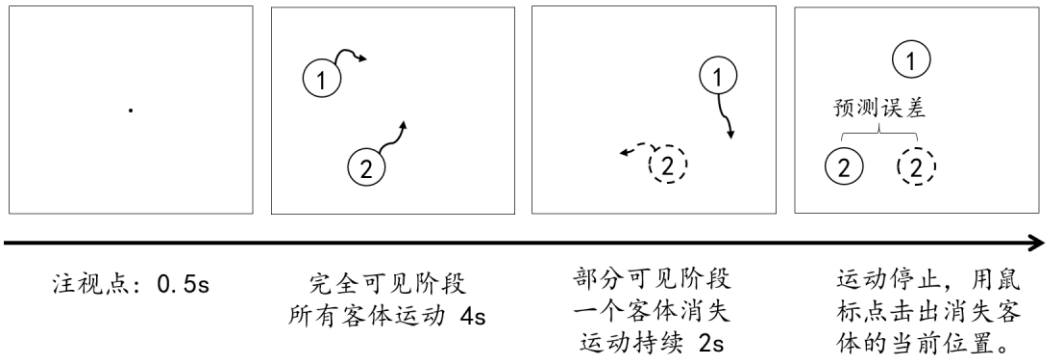

图 2.1 位置预测任务实验流程图。所有客体运动 4s，随后其中一个客体消失，继续呈现 2s 运动，该阶段消失的客体仍然保持运动，只是不能被观测到。运动结束后，用鼠标点击消失客体最后时刻的位置。

实验一生成了具有四种不同层级结构的运动(图 2.2),首先是对深度的操纵,其中“共同根节点”条件下两个客体所在的层级结构深度为 1;而在“单子节点条件”和“双子节点条件”下,两个客体所在的层级结构深度为 2;“独立条件”下两个客体分别位于两个独立的节点中,不能组成共同的层级结构,该条件被作为基线。在此基础上进一步对层级结构中的距离进行操纵,在深度为 2 的层级结构中,“单子节点条件”下两个客体在层级结构中的距离为 1,而“双子节点条件”下两个客体在层级结构中的距离为 2。此外,在“单子节点条件”中,两个客体分别位于父节点和子节点上,对它们的位置预测具有方向上的不对称性。所有条件在实验中随机出现,共有 160 个试次(每条轨迹重复 2 次,两个客体分别在这 2 次中作为需要预测的目标客体),分为 4 组进行,被试可以在每组实验完成后进行充分休息。

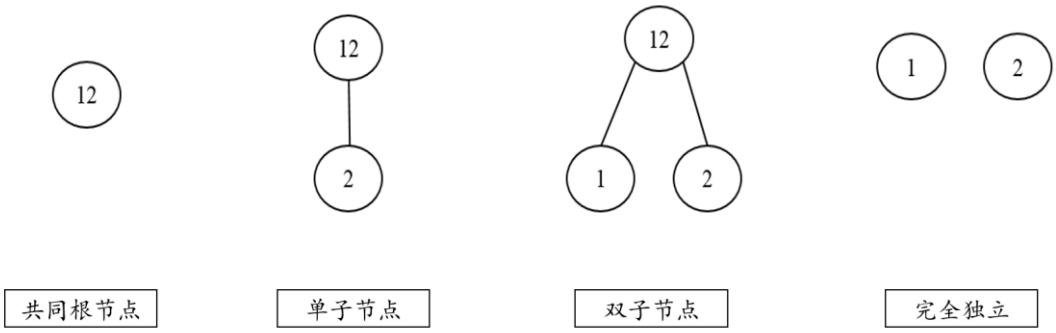

图 2.2 实验一中运动信息对应的四种潜在结构。共同根节点:两个客体位于层级结构中相同的节点,其运动仅存在噪音差异;单子节点:在共同运动的基础上,客体 2 具有独特的运动成分;双子节点:在共同运动的基础上,客体 1 和客体 2 分别具有不同的独特运动成分;完全独立:两个客体做完全独立的随机运动。

本实验同时采用两类计算模型模拟了任务过程。第一类模型是层级模型,该模型在完全可见阶段通过逆向推理构建运动的层级表征,在部分可见阶段基于该层级表征计算消失客体可能的位置;第二类模型是相关模型,该模型在完全可见阶段存储客体的所有运动信息以及它们之间的相互关系,在部分可见阶段基于所存储的信息计算消失客体可能的位置(详见附录二)。两类模型的主要计算过程完全相同,仅区别于是否采用层级表征作为对运动信息的表达方式。本研究中计算模型模拟的目的并非比较其绩效上的优劣,而是寻找模型模拟结果与人类行为

结果在模式上的异同，为视觉加工中运动信息层级表征的存在提供进一步证据。

2.1.2 结果与分析

被试点击的客体位置与客体真实位置之间的差异（预测误差）被作为衡量被试绩效的因变量。首先对层级结构深度的影响进行分析（图 2.3a），结果显示：被试的预测误差在深度 1、深度 2 和独立三种水平间存在显著差异 ($F(2, 30) = 107.82$, $p < 0.01$, $\eta_p^2 = 0.88$)，进一步比较 (采用 Bonferroni 矫正) 发现深度 1 水平下被试的预测误差 (1.26°)显著小于深度 2 水平下的预测误差 (2.96°) ($t(15) = 9.13$, $p < 0.01$, $d = 2.28$)，且这两种水平下的预测误差均小于作为基线水平的独立条件 (3.89°) ($t(15) = 12.27$, $p < 0.01$, $d = 3.07$; $t(15) = 6.87$, $p < 0.01$, $d = 1.72$)。其次对层级结构中距离的影响进行分析（图 2.3b），结果显示：被试在距离 1（单子节点条件）水平下的预测误差 (2.77°)显著小于在距离 2（双子节点条件）水平下的预测误差 (3.15°) ($t(15) = 2.33$, $p = 0.03$, $d = 0.61$)。最后对方向的影响进行分析（图 2.3c），结果显示：被试对父节点上的客体的预测误差 (2.92°)显著大于对子节点上的客体的预测误差 (2.62°) ($t(15) = 2.36$, $p = 0.03$, $d = 0.59$)。

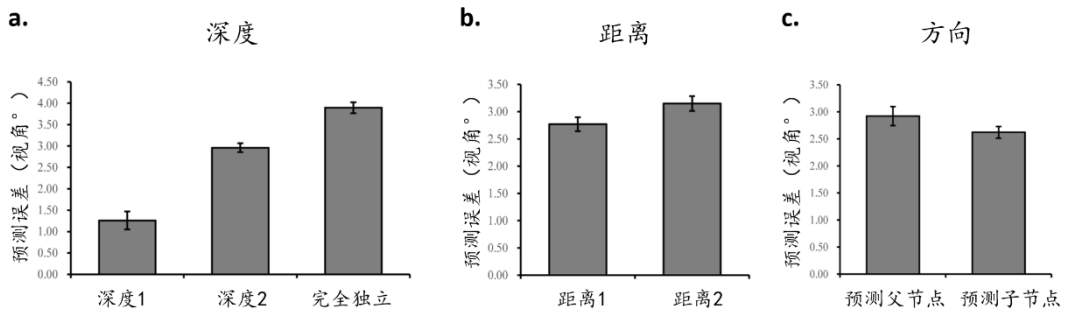

图 2.3 实验一被试行为结果。纵坐标均为预测误差，单位为视角。(a) 不同深度的层级结构对应的预测误差；(b) 层级结构中距离不同的客体对应的预测误差；(c) 对父节点客体和子节点客体的预测误差。

两类模型（层级模型和相关模型）对任务进行了模拟（详细数学过程见附录二），结果显示：层级模型的任务绩效模式与被试绩效一致（图 2.4a-c），在深度 1、深度 2 和独立三种深度水平下的预测误差存在显著差异 ($F(2, 30) = 2084.86$, $p < 0.01$, $\eta_p^2 = 0.99$)；且在距离 1 条件下的预测误差显著小于距离 2 水平下的预测

误差 ($t(15) = 16.04, p < 0.01, d = 4.01$)；模型对父节点客体的预测误差显著大于对子节点客体的预测误差($t(15) = 6.45, p < 0.01, d = 1.61$)。与之相反，相关模型的绩效与被试绩效的模式不同（图 2.4d-f），仅在不同深度水平间存在显著差异 ($F(2,30) = 212.55, p < 0.01, \eta_p^2 = 0.93$)；在不同距离水平间则出现相反的效应，即距离 1 条件下的预测误差显著大于距离 2 水平下的预测误差 ($t(15) = 5.32, p < 0.01, d = 1.33$)；此外，相关模型对父节点客体和子节点客体的预测误差不存在显著差异 ($t(15) = 1.09, p = 0.29, d = 0.28$)。

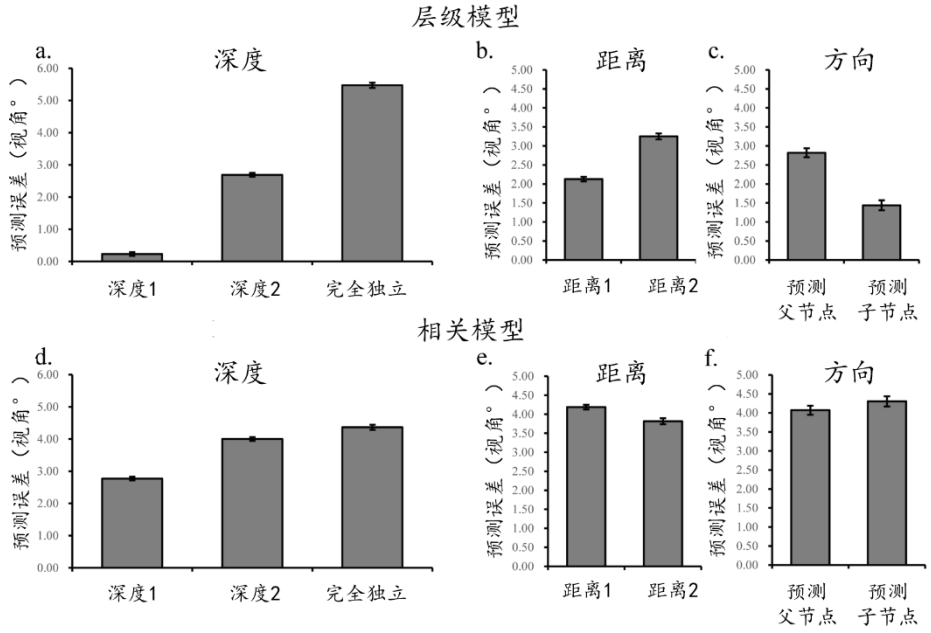

图 2.4 实验一模型模拟结果。纵坐标均为预测误差，单位为视角。(a)-(c)分别是层级模型在深度、距离和方向三类条件中的模拟结果;(d)-(f)分别是相关模型在深度、距离和方向三类条件中的模拟结果。

上述行为结果显示，运动潜在结构的改变，影响了被试对客体位置的预测绩效，这种结构效应意味着，即使在任务要求预测单个客体位置时，被试仍然表征了场景中的所有运动信息。被试预测绩效随层级结构深度、距离和方向的对应变化，进一步说明被试对运动信息的表征是层级结构。模型的模拟结果同样为此提供了证据，只有采用层级表征的计算模型能够表现出与被试行为相似的模式，与相关模型相比能够更好的解释被试的行为。综上所述，实验一表明运动信息在视觉加工中以层级结构加以表达。

2.2 实验二 具有相同层级结构的运动

本实验在具有相同层级结构的运动中检验运动信息的视觉层级表征。与实验一相比，该实验的不同条件间运动信息完全相同，但所需预测的客体在层级结构中具有不同的地位，其深度、与其它客体的距离和方向可能不同。如果运动信息在视觉加工中表征为层级结构，则对地位不同客体的位置预测绩效应存在差异，而对地位相同客体的位置预测则具有相同的绩效水平。

2.2.1 方法

2.2.1.1 被试、设备与刺激材料

16 名（5 男，11 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。除运动轨迹外，实验的设备、环境与刺激材料与实验一完全相同。

2.2.1.2 实验设计与流程

实验的任务和流程与实验一完全相同，但构造了与实验一不同的运动轨迹。实验二中包含三个运动客体，其运动轨迹基于如图 2.5a 所示的层级结构构造，其中客体 1 和客体 2 在层级结构中地位相等，且它们之间的距离较近；客体 3 与另外两个客体在层级结构中的地位不同，且与它们的距离较远。实验包含四种条件，前三种条件具有完全相同的运动轨迹，差别在于需要预测的目标客体分别是客体 1、客体 2 或客体 3，另外构造了三个客体不具有任何统一层级结构（相互独立的随机运动）的条件作为基线。四种条件在实验中随机出现，共 160 个试次，分为 4 组，被试可以在完成一组实验后充分休息。本实验采用与实验一相同的两类模型对任务进行模拟，并比较模型绩效与被试行为绩效在模式上的异同。

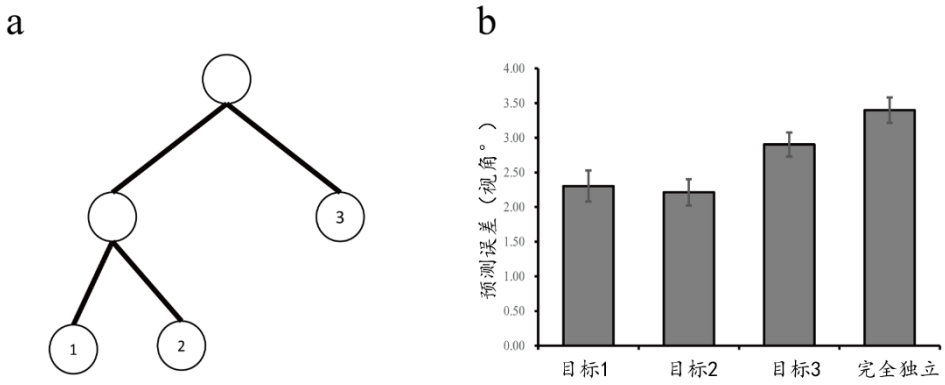

图 2.5 （a）实验二运动的潜在层级结构；（b）实验二被试行为结果，纵坐标为预测误差，单位为视角。

2.2.2 结果与分析

对被试的行为绩效进行分析，结果显示（图 2.5b）被试在四种条件下的预测误差存在显著差异 ($F(3, 45) = 40.45$, $p < 0.01$, $\eta_p^2 = 0.73$)，进一步比较(Bonferroni 矫正)发现，对客体 1 的预测误差 (2.30°)和对客体 2 的预测误差 (2.21°)之间不存在显著差异 ($t(15) = 1.24$, $p > 0.250$, $d = 0.31$)，且均小于对客体 3 的预测误差 (2.90°) ($t(15) = 3.56$, $p = 0.01$, $d = 0.89$; $t(15) = 5.16$, $p < 0.01$, $d = 1.29$)。另外，三种条件下的预测误差均小于基线条件 (3.40°) ($t(15) = 8.86$, $p < 0.01$, $d = 2.22$; $t(15) = 10.99$, $p < 0.01$, $d = 2.75$; $t(15) = 4.36$, $p < 0.01$, $d = 1.09$)。

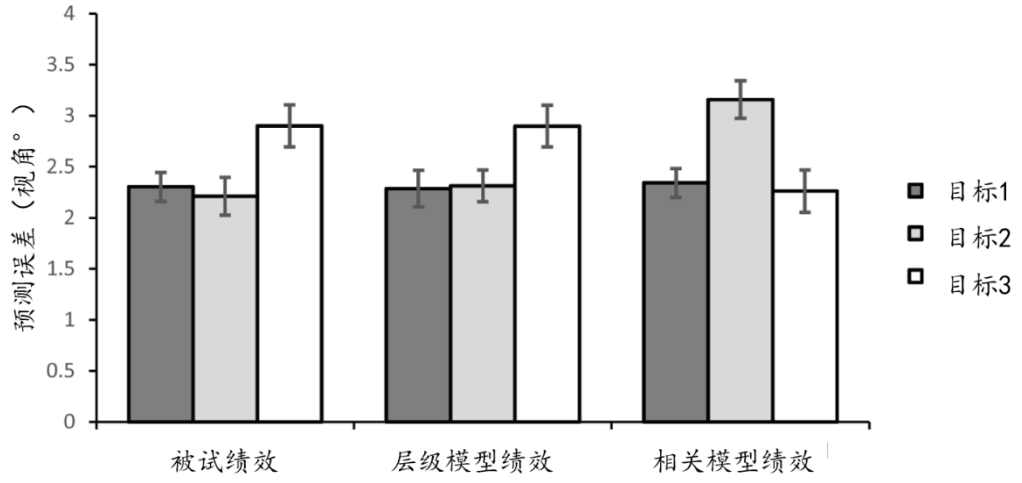

图 2.6 实验二模型模拟结果与被试绩效对比。

对两类模型的绩效进行同样的分析（图 2.6），结果显示：层级模型的绩效与被试行为绩效具有相同的模式，对客体 1 的预测误差和对客体 2 的预测误差之间不存在显著差异 ($t(15) = 0.12, p > 0.25, d = 0.03$)，且均小于对客体 3 的预测误差 ($t(15) = 2.65, p = 0.05, d = 0.66$; $t(15) = 2.73, p = 0.04, d = 0.68$)。与此相反，相关模型的绩效与被试行为绩效的模式具有很大差异，对客体 2 的预测误差显著大于对客体 1 的预测误差 ($t(15) = 3.28, p = 0.01, d = 0.82$)和对客体 3 的预测误差 ($t(15) = 3.99, p < 0.01, d = 0.998$)，对客体 1 的和客体 3 的预测误差不存在显著差异 ($t(15) = 0.39, p > 0.25, d = 0.09$)。

被试的行为绩效再一次表现出结构效应，表明视觉系统以层级结构表征运动信息。并且在本实验中三个主要条件下，视野中的运动场景完全相同，因此运动轨迹本身不能解释所观察到的效应。对客体 1 和客体 2 的预测结果为层级表征的构建提供了进一步证据：这两个客体本身的运动轨迹并不相同，但它们在层级表征中的地位完全对等，预测误差之间无显著差异意味着被试首先对运动构建了层级表征，并基于该表征（而非仅基于运动本身）完成预测任务。层级模型的模拟结果与被试行为结果模式一致，支持了视觉对运动的层级表征，而相关模型则表现出完全不同的模式：在层级结构中地位完全相同的客体，其预测绩效出现显著差异，可能的一个解释是，在这两种条件下，目标客体与其他客体在运动轨迹上的相关性存在差异，对轨迹的分析显示，客体 1 与客体 2 作为目标时，目标客体与其他客体的平均相关系数分别为 0.727 和 0.507，因此相关模型对客体 1 的预测误差显著小于对客体 2 的预测误差。这一结果表明，人类视觉并非简单考虑运动客体的相关信息，而是构建其潜在层级结构作为视觉表征。

2.3 实验三 破坏层级结构的影响

实验三的目的是从另一个角度证明视觉对运动信息的层级表征，其逻辑是：如果视觉确实以层级的形式表征运动信息，则在不改变运动物理特性的情况下，破坏其潜在的层级结构将对行为绩效产生影响。对层级结构的破坏通过镜像的方法实现，如果在破坏层级结构前后对相同客体的位置预测绩效存在差异，则说明视觉以层级结构表征运动信息。

2.3.1 方法

2.3.1.1 被试、设备与刺激材料

16 名（5 男，11 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与刺激材料与之前实验完全相同。

2.3.1.2 实验设计与流程

实验任务和流程与之前实验完全相同，运动轨迹则在实验二的基础上进一步变换。新轨迹的产生通过将原轨迹中某一个客体进行镜像变换来得到，镜像的具体方法是在每一个时刻，使该客体的位置相对于屏幕中心旋转 180°（图 2.7a），这一操作将会破坏运动的层级结构，但客体运动轨迹的速度、方向变换均不发生改变，且其运动与其他客体运动的相关性也不受影响（仅相关系数的正负发生变化）。实验三中，需要预测的目标客体总是客体 1，但存在四种不同的条件，三种镜像条件下分别将原始轨迹中客体 1、客体 2 或客体 3 的位置进行镜像，作为新轨迹呈现给被试，而保留原始轨迹不变的条件则作为基线（与实验二中客体 1 条件完全相同）。不同条件随机出现在整个实验中，共 160 个试次，分为 4 组进行，被试在完成每组任务后可以充分休息。

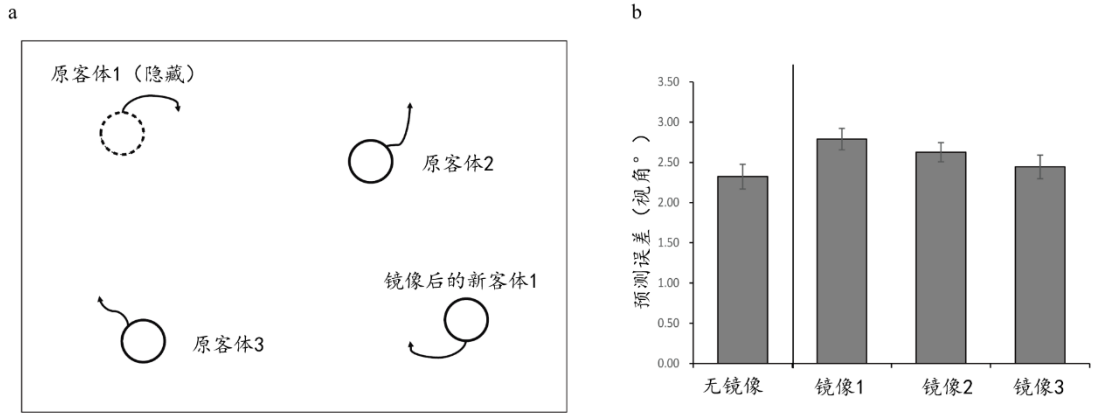

图 2.7 （a）实验三镜像操作示意图；（b）实验三被试行为结果，纵坐标为预测误差，单位为视角。

2.3.2 结果与分析

为针对性的考察镜像操作的效果，本实验中以使用原轨迹的无镜像条件作为基线，分别与镜像 1、镜像 2 和镜像 3 条件进行对比。结果显示（图 2.7b）：镜

像 1 条件下的预测误差 (2.79°)显著大于无镜像条件 (2.32°) ($t(15) = 3.44, p < 0.01, d = 0.86$)，镜像 2 条件下的预测误差 (2.63°)同样显著大于无镜像条件 ($t(15) = 2.13, p = 0.05, d = 0.53$)，镜像 3 条件下的预测误差 (2.44°)则与无镜像条件之间不存在显著差异 ($t(15) = 1.49, p = 0.16, d = 0.37$)。

与仅考虑相关性的观点不同，尽管运动间的相关性保持相同，但破坏层级结构仍然对预测绩效产生了影响。并且，由于各个客体在层级结构中的地位不同，导致对不同客体的镜像操作产生了不同程度的结构破坏，进而表现出对预测绩效影响的差异。具体来说，客体 1 和客体 2 在层级结构中地位相等，且距离较近，因此在借助层级结构预测客体 1 位置时，客体 2 的作用更为重要，这导致镜像 1 或镜像 2 均对预测绩效造成了显著破坏。与之相对的，客体 3 在层级结构中离客体 1 较远，对预测客体 1 位置的贡献相对较小，因此镜像 3 条件下未能观察到预测绩效的显著差异。上述结果再一次证明，视觉对多客体运动构建了层级表征。

2.4 小结

研究一通过三个实验分别考差了基于不同层级结构的运动信息（实验一）、基于相同层级结构的运动信息（实验二）和破坏层级结构（实验三）对位置预测精度的影响，并针对性的在层级结构深度、距离和方向上观察到对应的变化模式，验证了视觉系统以层级结构对运动信息加以表征。

3 研究二：层级表征的稳定性

视觉表征可能形成于视觉加工早期阶段，在不同情境中表现出稳定性；也可能受高级认知过程影响，由认知主体的策略或偏好决定，表现出对特定情境的依赖。研究二通过检验信息时间长度、任务线索和任务目的等因素对被试行为绩效的影响，考察视觉加工中运动信息层级表征的稳定性。如果在上述因素不同的情境中，被试的行为绩效表现一致，则说明运动信息的视觉层级表征具有跨情境的稳定性。

3.1 实验四 有效运动时长对层级表征的影响

运动信息在视觉加工中的层级表征可能受到运动信息时间长度的影响，运动时间较长的情境下，被试能够主动寻找最能描述当前运动信息的模式，此时形成的层级表征是基于策略或个人偏好的。为检验视觉对运动信息的层级表征是否是被试主观策略的结果，实验四将操纵运动信息的时间长度，测量被试行为绩效受到的影响。如果在信息时间长度不同的情况下，被试的行为绩效始终表现出依赖于运动层级结构的变化模式，且在不同时间长度下无显著差异，则能够说明运动的层级表征并非源于被试主观策略，而是早期视觉加工的结果。

3.1.1 方法

3.1.1.1 被试、设备与刺激材料

本实验为被试间设计，64 名（30 男，34 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与刺激材料和实验二相同。

3.1.1.2 实验设计与流程

实验为被试间设计，被试被随机均匀分配到四组中，每组包含 16 名被试。实验任务与实验二完全相同，但不再包含作为基线的随机运动条件，仅留下消失客体 1 和消失客体 3 作为衡量这种层级结构的关键条件。四组被试的区别在于，在运动的完全可见阶段，呈现给被试的带有层级结构的运动时长不同，分别为 4s、3s、2s 和 1s。为了控制被试在完全可见阶段观察到的运动总时长相同，仅包含不

同时长的有效运动，对运动轨迹进行一定的处理，使其由随机运动和带有层级结构的运动拼接而成。以 3s 条件为例，被试观察到的 4s 运动中前 1s 是三个客体的随机运动，在 1s 运动结束后从客体的当前位置开始进行 3s 的带有层级结构的运动（图 3.1a）。两部分运动以及整个 4s 的运动片段中，客体的速度、方向变化等物理因素保持相同，避免不同的运动片段被觉察出，或是在衔接的时刻发生运动的突然改变。每名被试需要完成 80 个试次，分为两组，被试在完成每组实验后可以进行充分休息。

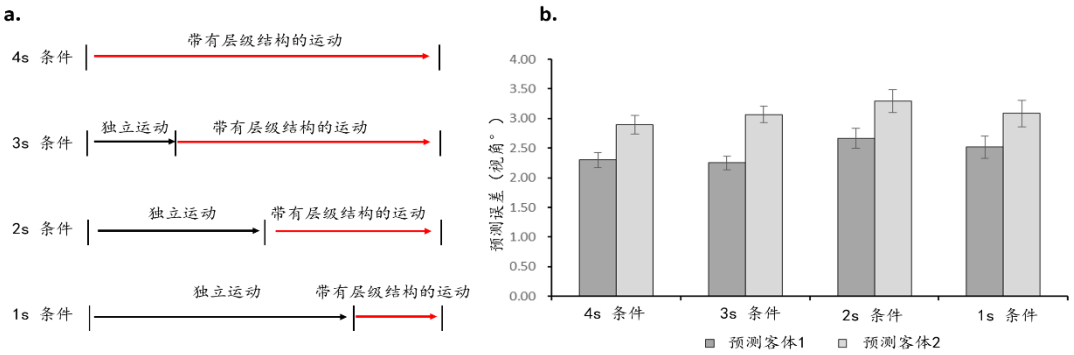

图 3.1 （a）实验四不同有效时长的轨迹拼接示意图；（b）实验四被试行为结果，纵坐标为预测误差，单位为视角。

3.1.2 结果与分析

对不同条件下被试的预测误差进行 4（有效运动时长）×2（消失客体）的混合方差分析，结果显示（图 3.1b）：消失客体的主效应显著，对客体 1 的预测误差显著小于对客体 3 的预测误差 ($F(1,60) = 115.62, p < 0.01, \eta_p^2 = 0.66$)。但有效运动时长与消失客体的交互作用不显著 ($F(3,60) = 0.89, p = 0.45, \eta_p^2 = 0.04$)，即有效运动的时长不影响被试对层级结构中不同客体的预测误差，进一步分析结果发现，四种有效运动时长下，被试均表现出对客体 1 的预测误差显著小于对客体 3 的预测误差 (4s: $t(15) = 4.90, p < 0.01, d = 1.22$; 3s: $t(15) = 6.75, p < 0.01, d = 1.69$; 2s: $t(15) = 5.17, p < 0.01, d = 1.29$; 1s: $t(15) = 4.67, p < 0.01, d = 1.16$)。

上述结果不但重复了位置预测绩效基于层级结构的变化模式，证明视觉中形成了运动信息的层级表征，同时表明该表征不受运动时间长度影响，在极短的时

间内，视觉系统已能够完成层级表征的构建。即运动信息的视觉层级表征不依赖于被试主观策略，而是完成于视觉加工的早期阶段。

3.2 实验五 任务线索对层级表征的影响

尽管位置预测任务只要求对单个客体的位置进行预测，并不要求被试对全部运动信息进行整合加工，然而在完全可见阶段被试无法确定哪一个客体会消失，这种不确定性可能诱使被试对所有运动信息进行整体加工。实验五将调节任务线索的有效性，考察被试行为绩效受到的影响。如果被试在不同任务线索下表现出相同的行为模式，则表明运动信息的视觉层级表征不依赖于任务对整体加工的导向，而是表现出跨任务情境的一致性。

3.2.1 方法

3.2.1.1 被试、设备与刺激材料

本实验为被试间设计，44 名（17 男，27 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与刺激材料和实验二相同。

3.2.1.2 实验设计与流程

本实验在实验二的基础上进行，同样采用位置预测任务。实验流程与实验二基本相同，唯一的区别是，在运动开始前，其中一个客体变为红色 (RGB = [255, 0 ,0]) 并持续 300ms，这一线索提示在接下来的任务中，该客体会消失，被试需要对它的位置进行预测。本实验采用被试间设计，所有被试被随机均匀的分为两组，以避免对相同的运动轨迹产生熟悉。在线索无效组中，任务相关的线索是随机出现的，它所标记的客体在部分可见阶段消失的概率只有 50%，而在线索有效组中，任务相关的线索是绝对正确的，即线索标记的客体一定会在部分可见阶段消失。被试在任务开始前已被告知线索的有效性情况。在前述实验结果的基础上，本实验中不再包含原本作为基线的随机运动条件，只剩下消失客体为客体 1、客体 2 和客体 3 三种条件，随机出现在整个实验过程中，每组被试均需进行 120 个试次，分为 3 组，被试可以在完成一组实验后充分休息。

3.2.2 结果与分析

对两组被试的预测误差进行 2（线索有效性）×3（消失客体）的混合设计方差分析，结果显示（图 3.2）：消失客体的主效应显著 ($F(2,84) = 47.56$, $p < 0.01$, $\eta_p^2 = 0.53$)，与之前实验模式相同，被试对客体 1 的预测误差与对客体 2 的预测误差无明显差异 ($t(21) = 0.90$, $p > 0.25$, $d = 0.19$)，且均小于对客体 3 的预测误差 ($t(21) = 7.58$, $p < 0.01$, $d = 1.61$; $t(21) = 7.80$, $p < 0.01$, $d = 1.66$)。线索有效性的主效应不显著 ($F(1,42) = 0.81$, $p = 0.37$, $\eta_p^2 = 0.01$)，且线索有效性与消失客体的交互作用不显著 ($F(2,84) = 0.41$, $p = 0.66$, $\eta_p^2 = 0.01$)。进一步对两组被试内的预测绩效模式进行简单效应检验，结果发现无论线索是否有效，预测绩效均表现出相同的模式，即对层级结构中地位相同的客体 1 和客体 2 的预测误差无显著差异，且均小于在层级结构中距离较远的客体 3，统计结果见表 3.1。

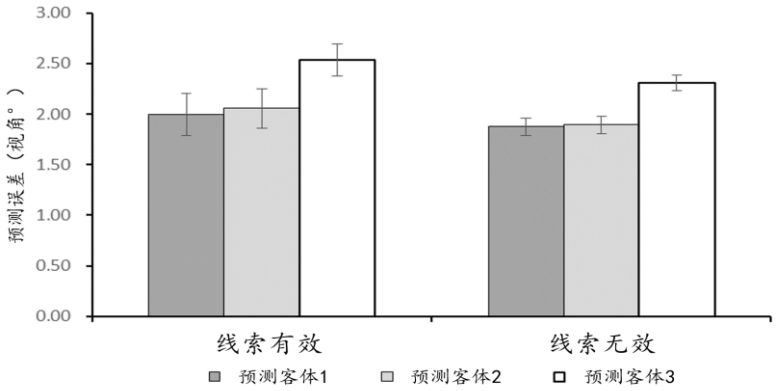

图 3.2 实验五被试行为结果，纵坐标为预测误差，单位为视角。

表 3.1 实验五简单效应分析结果 (Bonferroni 矫正)

	线索有效条件			线索无效条件		
	t	p	d	t	p	d
客体 1 vs. 客体 2	0.96	> 0.25	0.20	0.31	> 0.25	0.06
客体 1 vs. 客体 3	5.91	< 0.01	1.26	4.80	< 0.01	1.02
客体 2 vs. 客体 3	5.89	< 0.01	1.25	5.13	< 0.01	1.09

上述结果表明，被试对客体位置的预测绩效符合运动潜在的层级结构，且该结构不受到任务相关线索的影响：无论任务中的不确定性是否存在，被试是否需要对运动信息进行整体加工，视觉始终构建运动信息的层级表征。这意味着运动信息的视觉层级表征是运动视觉加工的基础，在不同的任务需求中稳定存在。

3.3 实验六 运动无关任务中的层级表征

位置预测任务中，被试可能在部分可见阶段依赖于其他客体的运动预测目标客体的位置，因此形成对运动信息的层级表征。为考察层级表征是否依赖于特定的任务目的，实验六构造了与运动信息完全无关的任务，并测量被试行为绩效随层级结构的变化模式。由于在运动无关任务中，加工运动信息不仅对任务绩效没有帮助，甚至可能存在干扰。如果被试的行为绩效仍然表现出依赖于层级结构的变化模式，则表明对运动信息的视觉层级表征与特定的任务目的无关，而是在广泛的情境中具有极强的稳定性，且其形成过程可能是自动化的。

3.3.1 方法

3.3.1.1 被试、设备与刺激材料

16 名（8 男，8 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与之前实验相同。

实验中运动轨迹的产生和呈现与前述实验完全相同，被试需要判断的目标字母为 1°×1°的目标字母 T 或 L，非目标字母为 T 和 L 组成的无意义字母（图 3.3a），所有字母均为蓝色 (RGB = [0, 0, 255])。

3.3.1.2 实验设计与流程

实验采用经典的线索范式，在运动结束后，目标字母会在任意一个客体上出现，同时在其他客体上出现作为干扰的非目标字母，被试的任务是判断目标字母的身份。在目标字母出现之前，会在任意一个客体上出现一个线索，在大概率下这个线索是有效的，意味着接下来目标字母会在相同的客体上出现，以保证被试的注意会投向线索出现的客体。在少数情况下，线索是无效的，意味着目标字母会出现在另一个客体上，此时为了判断目标字母的身份，被试的注意需要从线索

出现的客体转移到目标出现的客体，导致被试的反应时变长。被试在线索无效与线索有效情况下的反应时差异被定义为线索效应，其大小反应了注意在客体间转移的速度，更快的转移速度表明客体之间具有更紧密的组织结构。

本实验中被试需要判断的目标字母是 T 或者 L，作为干扰的非目标字母则是 T 和 L 组成的无意义字母。实验首先呈现三个客体的运动，运动持续 3s 后，所有客体停止运动，并且其中一个客体变为红色 (RGB = [255, 0, 0])持续 0.1s 作为线索，之后线索消失，所有客体静止 0.2s，随后目标和干扰子同时出现，直到被试反应或持续时间超过 2s。实验流程如图 3.3b 所示。

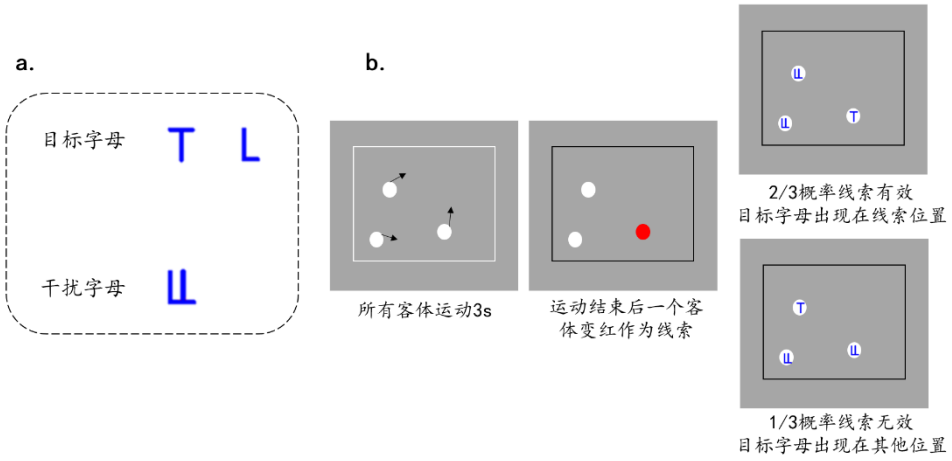

图 3.3 （a）实验六中目标字母和干扰字母示意图；（b）实验六任务流程图，所有客体运动 3s，运动结束后其中一个客体变为红色 0.25s 作为线索，之后在任意一个客体上呈现目标字母，并在其他客体上呈现干扰字母，被试需在 2s 内判断目标字母的身份。

实验中的客体运动是基于与实验二相同的层级结构产生的，即客体 1 与客体 2 在层级结构中地位相同且距离较小，客体 3 与前两个客体的距离相等且较大。实验包含四种条件，分两类情况：考察层级结构距离不同的两种条件中，线索呈现在客体 1 上，在线索无效的情况下，目标有可能出现在客体 2（条件 1_2）或客体 3（条件 1_3）上，并且在这两种条件下，运动结束时刻客体 1 与客体 2 和客体 3 的空间距离相等；考察层级结构距离相同的两种条件中，线索呈现在客体 3 上，在线索无效的情况下，目标可能出现在客体 1（条件 3_1）或客体 2（条件 3_2）上，并且在这两种条件下，运动结束时刻客体 3 与客体 1 和客体 2 的空间距离相等。需要注意的是，条件 1_3 和条件 3_1 在运动结束时刻客体间的空间距

离并不相同，因此它们的反应时和线索效应可能存在差异，但本实验主要关注的是两类情况各自的两种条件间的比较。为保证线索的有效性，实验中线索有效的比例为 2/3。不同条件在实验中随机出现，共包含 240 个试次，分为 5 组，被试在完成每组任务后可以充分休息。

3.3.2 结果与分析

每种条件下线索有效和线索无效时的反应时差异被定义为线索效应，分别分析层级结构距离不同和层级结构距离相同两类情况下，条件间的线索效应差异。结果显示（图 3.4）：层级结构距离不同的情况下，条件 1_3 的线索效应显著大于条件 1_2 ($t(15) = 2.68, p = 0.01, d = 0.67$)，即层级结构中距离较远的客体间线索效应更大；而在层级结构距离相同的情况下，条件 3_1 和条件 3_2 的线索效应不存在显著差异 ($t(15) = 0.91, p > 0.25, d = 0.03$)，即层级结构中距离相等的客体间线索效应无差异。

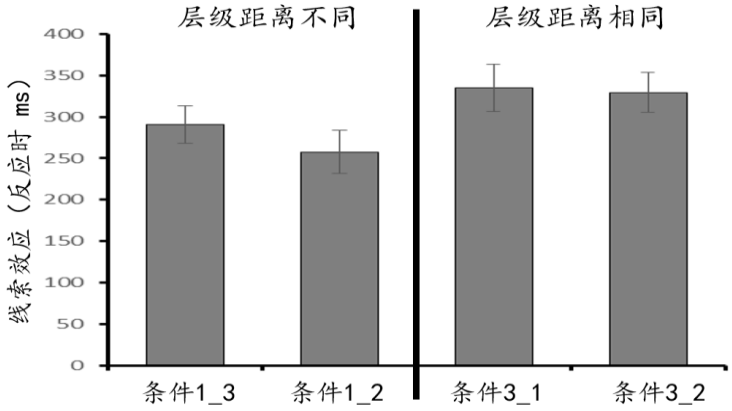

图 3.4 实验六被试行为结果，纵坐标为线索效应，即线索有效和线索无效情况下的反应时差，单位为毫秒（ms）。

上述结果表明，注意在该场景下的分布依赖于运动的层级结构，即视觉中形成了对运动信息的层级表征。值得注意的是，实验所涉及的任务与运动信息完全无关，然而被试仍然构建了对运动信息的层级表征，这意味着运动信息的视觉层级表征在不同的情境中具有稳定性，其构建可能是自动化的。

3.4 小结

研究二通过三个实验考察了运动信息时间长度（实验四）、任务线索（实验五）和任务目的（实验六）对位置预测精度的影响，验证了层级结构对被试行为绩效的影响在不同的任务情境中表现一致，即视觉系统对运动信息的层级表征具有跨任务的稳定性。

4 研究三：社会交互场景中的层级表征

人类所处的环境中不但包含丰富的物理运动，往往还充斥着大量的社会信息。视觉系统对社会场景中运动信息的有效加工，是完成社会交互活动的重要基础。带有社会属性的客体，其运动既依赖于自身的内在意图，又受到外部因素的限制，导致其真实运动更加复杂。视觉对这类运动信息的理解和预测需要建立在具有完备解释力的表征的基础上。研究三将以追逐场景为例，在更普遍的社会场景中验证运动信息的视觉层级表征，并进一步探索其加工机制。

4.1 实验七 交互对象之间的层级表征

处于社会交互场景中的对象，其运动往往受到来自外部环境的共同限制（如：追逐场景中的狼和羊，其运动均受到地形的限制）。为模拟这一情景，实验七构建带有层级结构的社会交互对象，并借助追逐识别任务来测量被试对运动中意图信息的识别能力。由于环境限制的存在，交互对象的真实运动往往与目标导向的运动存在较大偏离，使得意图的识别极为困难。如果被试能够为该场景下的运动构建层级表征，其追逐识别绩效将显著提升，表明带有社会信息的运动同样在视觉加工中以层级结构表征。

4.1.1 方法

4.1.1.1 被试、设备与刺激材料

16 名（6 男，10 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与之前实验相同。

4.1.1.2 实验设计与流程

实验采用追逐识别任务，具体流程如下：在主视点消失后，屏幕上将呈现包含四个客体的运动，运动持续 10s 后停止，被试需要通过按键判断在刚才呈现的运动中是否包含追逐信息，即是否存在一个追逐者（简称为：狼）试图追逐一个目标（简称为：羊）。如果被试认为存在追逐信息，则需要通过鼠标点击进一步确认哪一个客体是狼、哪一个客体是羊；如果被试认为不存在追逐信息，则在按键反应后，直接进入下一个试次。被试反应后无论正误，不给予任何反馈（图 4.1）。

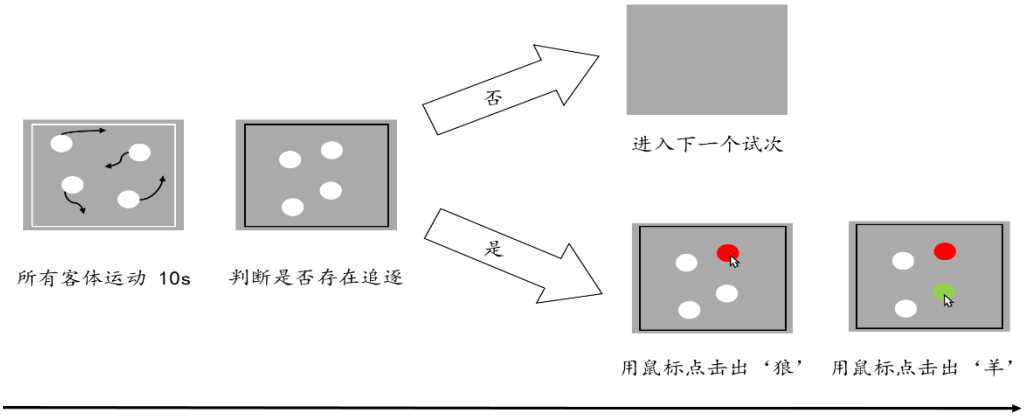

图 4.1 意图识别任务实验流程图，所有客体运动 10s 后停止，被试需要首先按键判断之前的运动中是否存在追逐信息，若判断为‘是’，则需用鼠标点击出追逐中‘狼’和‘羊’的身份。

带有追逐信息的运动分为两类，一类具有潜在的层级结构，由模型产生，其具体结构如图 4.2 所示：狼和羊共同位于一个双子节点的层级结构中，使得它们的真实运动由它们各自子节点上的运动分量和位于父节点上的运动分量共同组成。子节点上的运动分量与它们的真实意图一致，即狼的运动分量直接指向当前羊的位置，羊则向最有利的方向逃脱。然而由于父节点运动分量的存在，狼的真实运动与目标指向的方向存在很大分离。在不具有潜在的层级结构的追逐运动中，狼的运动直接指向羊的当前方向，但为了控制条件间的追逐偏离程度相同，将具有层级结构条件中每一帧的追逐偏离大小叠加到不具有层级结构的条件中，使得两种条件的追逐偏离大小在每一时刻完全相同。其余两个干扰子均做完全独立的随机运动，且轨迹在条件间保持一致。不带有追逐信息的运动由带有追逐信息的运动产生，通过将轨迹中羊的运动镜像来破坏追逐信息，同时控制其余运动信息尽量保持一致。不同条件的运动在实验中随机出现，保证带有层级结构的运动与不带有层级结构的运动比例相同，且带有追逐信息和不带有追逐信息的运动比例相同，实验共包括 80 个试次，分为两组进行，被试在完成每组实验后可以进行充分休息。

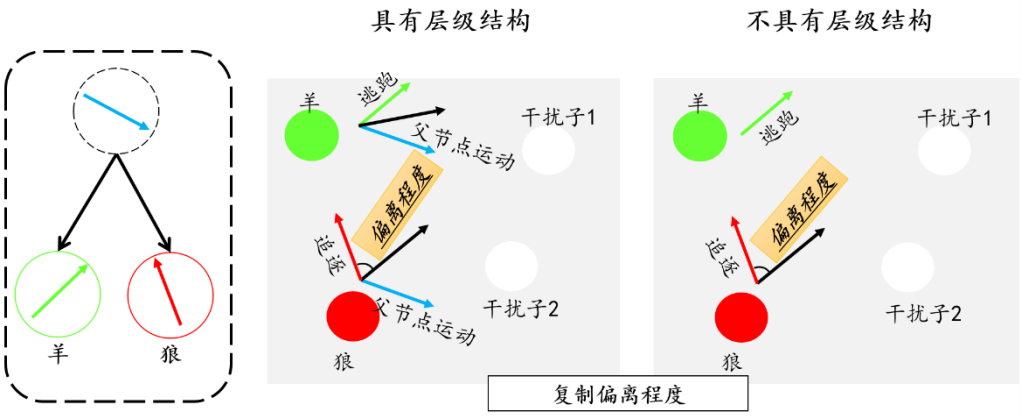

图 4.2 实验七运动产生示意图，左侧为有层级结构条件下，狼和羊的潜在层级结构，右侧是对应到真实运动场景中的简单说明。

本实验还采用模型对任务进行模拟。对运动过程中的意图识别模拟通常通过马尔科夫决策过程（MDP）来实现，该过程中，每一个时刻生命体根据当前自己和环境的状态，以及自己的目标，从行为集中选择一个行动执行。在观察到行动的基础上，可借助贝叶斯定理反解 MDP 过程，推理出生命体的目标。采用何种方式表征状态，是层级模型和一般模型的关键区别：在一般模型中，所有状态信息被存储在高维数组中，而在层级模型中，首先求解运动的潜在层级结构，随后以层级结构作为对状态的表征。不同模型在两种情况下的绩效，将在分析中和被试的行为绩效相对比，以进一步揭示社会情景中的层级表征（详见附录三）。

4.1.2 结果与分析

只有被试正确判断出存在追逐运动，并正确点击出狼和羊的身份时，这个试次才被算作正确识别的试次，正确识别的试次占对应条件下存在追逐运动的总试次的比例被定义为识别正确率，对被试识别正确率的分析结果显示（图 4.3）：在具有层级结构的运动中，被试的识别正确率能够达到 67.81%，显著高于不具有层级结构运动中的识别正确率 (46.56%) ($t(15) = 5.39$, $p < 0.01$, $d = 1.34$)。

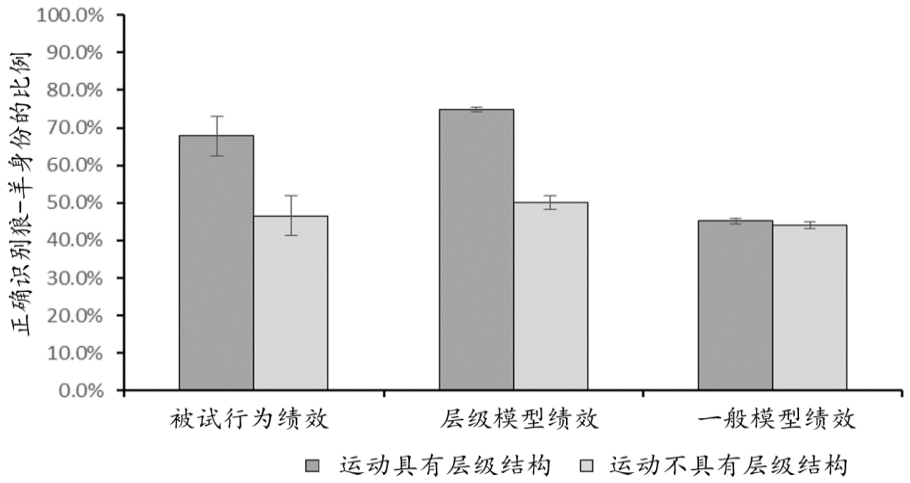

图 4.3 实验七结果图，包括被试行为结果和两类模型的模拟结果，纵坐标是正确识别的比例，判断出存在追逐信息并正确点击狼和羊的身份被定义为正确识别，正确识别试次数占对应条件下存在追逐的总试次数比例被定义为正确识别比例。

层级模型的结果与被试行为绩效类似，在具有层级结构的运动中展现出更高的识别绩效 ($t(15) = 11.95, p < 0.01, d = 2.98$)。然而一般模型并没有类似的差异 ($t(15) = 1.14, p = 0.27, d = 0.29$)，对于一般模型来说，运动是否具有层级结构并不重要，较高的偏离量使得它无法处理这一意图识别问题。

本实验所构造的运动中，追逐偏离量超过 60°，以往研究表明类似程度的偏离量将导致意图识别的正确率接近 40%(Gao, Newman, & Scholl, 2009)，本实验中不具有层级结构的条件下得到了类似的结果。然而在同等偏离程度下，具有层级结构的运动中意图识别正确率表现出显著提升，同时层级模型表现出了与被试相似的模式，而一般模型则在两类运动中表现相同。这表明被试确实能够为带有社会意图信息的运动构建层级表征，并基于该表征解释运动中的偏离信息，更好的进行意图识别。

4.2 实验八 交互对象与其他客体之间的层级表征

社会交互场景中，运动对象受到的限制不仅来自于所处的外部环境，还可能来自于其他因素（如：被驯兽人拴住的狼）。这种限制对于交互双方并不相同。为模拟上述情境，实验八构造了交互对象与其他运动客体之间的层级结构，采用

追逐识别范式，进一步考察社会场景中运动信息视觉层级表征的构建。与实验七相比，本实验能够在不同条件下生成完全相同的运动轨迹，进一步控制运动信息物理特性对实验结果的干扰，为带有社会信息的运动在视觉加工中的层级表征提供更强的证据。

4.2.1 方法

4.2.1.1 被试、设备与刺激材料

16 名（9 男，7 女）在校大学生有偿参加了本次实验，年龄在 18-25 周岁之间，视力或矫正视力正常。实验的设备、环境与之前实验相同。

4.2.1.2 实验设计与流程

实验流程和设计与实验六完全相同，对运动的构造存在一些差别。在具有层级结构的追逐运动中，狼和其中一个干扰子位于同一个单子节点层级结构中（图 4.4），其中狼所在子节点的运动成分直接指向羊的当前位置，但由于父节点的存在，狼的真实运动由直接指向羊的运动分量和与来自于父节点上干扰子的运动分量叠加而成，导致其真实运动与目标方向存在较大偏离。羊始终向最有利的方向逃跑，两个干扰子则始终进行完全独立的随机运动。在不具有层级结构的追逐运动中，狼和羊的运动轨迹保持不变，仅将原本作为父节点的干扰子镜像，以破坏干扰子和狼之间的层级结构。不具有追逐运动的轨迹同样通过将羊的运动镜像来得到。实验共包含 80 个试次，分为两组，被试在完成一组实验后可以进行充分休息。本实验采用与实验六相同的模型进行任务模拟。

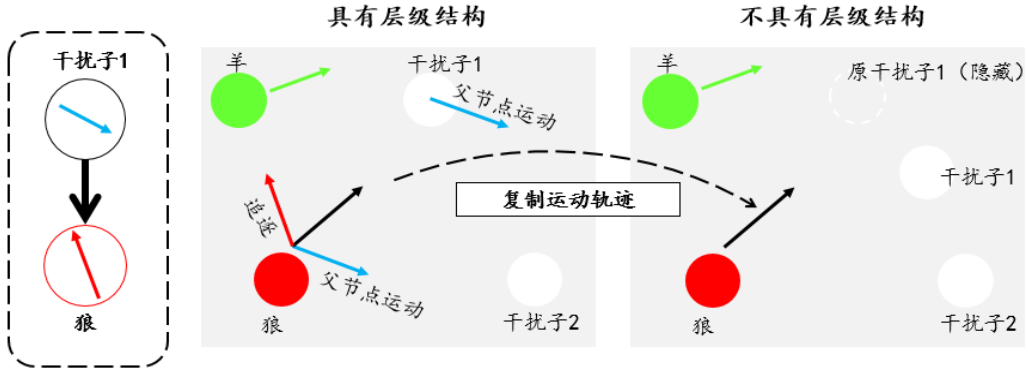

图 4.4 实验八运动产生示意图，左侧为有层级结构条件下，狼和羊的潜在层级结构，右侧是对应到真实运动场景中的简单说明。

4.2.2 结果与分析

对被试的识别正确率进行分析(图 4.5),结果显示:具有层级结构的条件下,被试的识别正确率 (55.31%)显著高于不具有层级结构条件下的识别正确率 (35.31%) ($t(15) = 3.64$, $p < 0.01$, $d = 0.91$)。

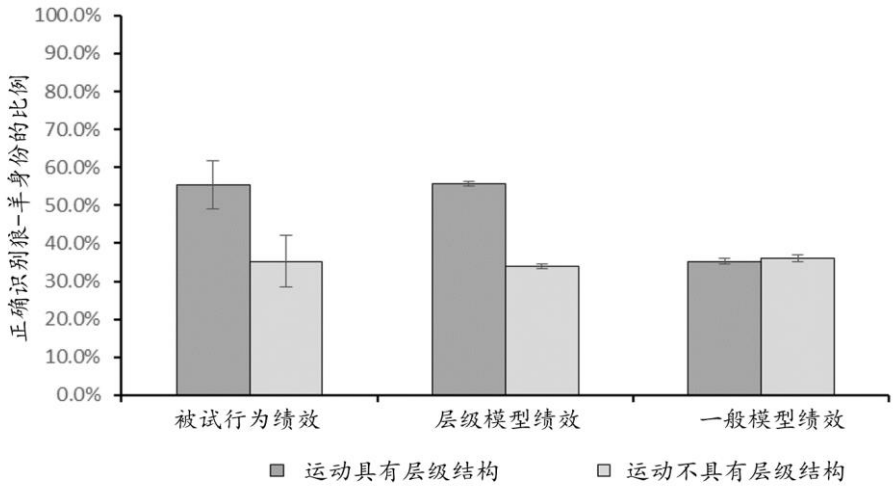

图 4.5 实验八结果图,包括被试行为结果和两类模型的模拟结果,纵坐标是正确识别的比例。

层级模型的结果与被试行为绩效类似,在具有层级结构的运动中展现出更高的识别绩效 ($t(15) = 23.87$, $p < 0.01$, $d = 5.97$)。然而一般模型并没有类似的差异 ($t(15) = 0.73$, $p = 0.47$, $d = 0.18$),同样受到在两类运动中均具有较高追逐偏离度的限制。

本实验中狼的平均追逐偏离量大于 80°,在部分情况下甚至会远离其目标,因此不具有层级结构的条件下,被试的识别绩效受到极大影响,表现出与以往研究类似的水平。但在狼和羊的运动完全相同的情况下,被试对具有层级结构的运动中意图信息的识别绩效更高,且只有层级模型表现出与被试行为相似的绩效模式,这表明被试确实能够对带有社会信息的运动构建层级表征,并利用层级表征完成意图识别的任务。

4.3 小结

研究三将运动场景进一步拓展到社会场景中，通过两个实验分别考察了社会交互对象之间的层级结构（实验七）与社会交互对象和其他运动对象之间的层级结构（实验八）对社会交互意图识别的影响，验证了视觉系统同样对带有社会信息的运动以层级结构加以表征。

5 总讨论

本研究对运动信息视觉加工中的关键问题——视觉表征进行了较为系统的探讨，通过三部分研究证明了运动信息在视觉加工中以层级结构加以表征，并进一步考察了该表征在不同情境中的稳定性与普遍性。笔者就运动信息视觉层级表征的特点和加工机制等四方面问题进行简要探讨。

5.1 运动信息层级表征的构建

5.1.1 其它运动物理特性效应的控制

研究的主要逻辑是通过潜在层级结构的变化引起的行为差异，来推论视觉系统对运动信息层级表征的构建。然而，潜在层级结构的外在表现是运动信息本身，因此运动本身所具有的其它物理特性对结果的影响是必须加以排除的。基于本研究所构建的运动信息和任务，可能对结果存在影响的其它运动物理特性主要包括目标客体运动特性、目标客体与其它客体的运动轨迹相关性和目标客体的空间分布三个方面。

不同条件下对运动对象位置预测的差异，可能源于目标客体速度、方向等运动特性的不同。当客体的运动速度更快时，工作记忆对运动信息的维持能力急剧下降(McKeefry et al., 2007)，对运动客体的追踪也变得更为困难(Alvarez & Franconeri, 2007)。对运动客体位置的预测依赖于对前序运动信息的记忆和对当前运动客体的追踪，因此预测绩效可能受到运动客体速度和运动方向变化的影响。然而本研究对上述因素产生的效应进行了严格控制。实验一和实验二中，对客体运动的速度和方向进行了严格的控制，对轨迹的检验未发现存在速度或方向上的差异；实验三则进一步排除了这种可能，由于该实验始终要求预测客体 1 的位置，无论在哪种条件中，客体 1 运动轨迹的所有物理特性均完全相同。因此，研究得到的结果不能被目标客体运动的物理特性所解释。

运动对象之间的相关性同样可能造成位置预测绩效的差异(Yin et al., 2016)，在层级结构中更紧密的客体，其运动存在高相关性的概率越大。然而本研究的结果并不能够被运动间的相关性解释：尽管层级结构中关系紧密的客体更有可能存在运动轨迹上的高相关，仍有可能构造层级关系紧密但运动相关性较低的客体，本研究在生成运动轨迹时对此进行了控制，选取了相关性无显著差异的运动轨迹；

实验一中对“父节点-子节点”不对称性的检验，以及实验三中采用镜像操作破坏层级结构，均保持了客体间相关性不变，被试预测绩效的差异并非源于客体间运动相关性的不同。此外，计算模型的模拟结果显示，相关模型难以描述被试的行为绩效，即被试并非通过表征场景中运动客体的相关性来完成任务。上述结果均表明，研究所得结果不能被客体间运动相关性解释。

本研究中所采用的任务大多没有反应时间或正确率的限制，被试的反应可能涉及主观决策和偏好。如果被试对某些空间位置的偏好较高，将有可能导致真实位置离该偏好位置较近的客体具有更低的预测误差。为排除这种假设，实验中将每一段运动轨迹参照屏幕中心旋转了四种角度，使得客体的运动和最终位置均匀的分布在屏幕的四个区域中，尽量排除了空间位置偏好的干扰。此外，实验三中对客体 2 和客体 3 镜像的两个条件，客体 1 的真实位置与基线条件完全相同，预测绩效的差异能够排除空间位置偏好的干扰。

基于上述分析，本研究对其它运动物理特性的效应进行了严格的控制，可以认为被试在任务中的行为绩效差异确实是基于层级结构的变化，由此推论视觉中构建了对运动信息的层级表征。

5.1.2 潜在结构的层级特征

对潜在结构的操纵引起了被试行为绩效的变化，这表明视觉确实对运动信息的潜在结构形成了内部表征。然而视觉所表征的结构是否就一定是层级结构？对该问题的检验需要基于层级结构特征的针对性操作。

与其他可能的潜在结构相比，层级结构具有三个方面的独特特征：深度、距离和方向。研究一针对性的操纵了以上三个方面的特征，行为绩效随之发生对应变化。特别是在方向这个特性上表现出的不对称性，为层级结构的存在提供了最强有力的支持。不仅如此，多个实验中，行为绩效的变化方向与对层级结构的操纵是一致的，在深度方面，深度越深的树，其内在结构越复杂，对其中客体的位置预测越难；层级结构中距离越远的客体，即使空间距离相同，也具有更远的关系，对其位置的预测也越难；基于父节点预测子节点相对容易，因为这是和层级结构方向一致的，反之由于父节点包含更多的共同信息，由父节点预测子节点将更难。同时，实验三从反方向入手，特异性的破坏运动的潜在层级结构，而保持

运动的物理特性不变，该操作下产生的行为绩效变化，同样证明了层级结构的存在。

上述结果显示视觉所表征的潜在结构符合层级结构的特征，换言之，视觉系统对运动信息的表征确实采用了层级结构。

5.1.3 层级表征的稳定性和普遍性

层级表征作为视觉加工的基础，其本质是对外部信息的内在描述，因而应具有跨任务的一致性，并且适用于更具生态效度的社会场景。

研究二通过三个实验验证了运动信息视觉层级表征的稳定性。实验四表明，运动信息的层级表征构建在视觉加工的早期阶段就已完成，排除了由于决策、个人偏好等原因对结果产生的干扰；实验五则验证了运动信息层级表征的构建并非依赖于对整体加工的任务要求；实验六进一步将场景拓展到与运动信息无关的任务中，层级结构效应的稳定存在表明，层级表征是对运动信息本质的内在描述和解释，随运动信息的呈现而产生，不受任务线索等因素的影响。

研究三将运动信息在视觉加工中的层级表征拓展到了社会场景中，验证了层级表征的普遍性。无论层级结构位于交互的对象之间（实验七），还是交互对象与其他对象之间（实验八），视觉系统均能够对包含社会信息的运动构建层级表征。这意味着层级表征是视觉对运动信息内在理解的基本形式，不受场景类型的限制。

5.2 作为因果结构的层级表征

相关不等同于因果，这是科学研究中众所周知的原则。本研究所强调的层级表征，不仅注重其具有的结构优势，也注重其具备因果性这一特点。运动背后的潜在层级表征既描述了运动本身，还描述了运动的产生过程。视觉系统对运动信息的层级表征，本质上是对运动产生过程的重构。

因果知觉的研究表明，人类视觉确实具有强大的识别信息背后因果关系的能力。虽然视觉对象呈现出的物理特征中并不直接显示因果关系，然而人类的认知系统却表现出快速、直接从刺激中获取因果关系的能力(Zhou et al., 2012)。例如在篮球运动中，人类能够清晰理解，球的运动是由怎样的动作引起的。不少研究

也证明了这种因果知觉的能力，碰撞效应（launching effect）显示，当观察到一个物体碰撞另一个物体时，该物体的运动终止，另一个物体随之开始运动，观察者一般认为后者的运动是由前者的碰撞引起的(Minchotte, 1963)。直觉物理学的相关研究发现，观察者能够理解引起物体运动的潜在原因，并利用该因果关系预测后续运动(Smith, Battaglia, & Vul, 2013)。即使几个月大的婴儿，也会表现出对违背因果关系运动的惊讶(Kim & Spelke, 1999; Wang & Baillargeon, 2006)，并利用因果关系预测复杂场景中的行为结果(Battaglia, Hamrick, & Tenenbaum, 2013)。此外，由刺激的外在信息推测内在的因果关系，在计算上是可行的(Pearl, 2009)，计算图模型中将因果关系定义为节点之间的有方向的连接，连接从父节点指向子节点，表示原因对结果的影响方向，在图模型的基础上对相关信息进行特异于路径的分析，将能够推测路径的方向，获得信息内部的因果关系。相关的计算方法已被应用于人工智能领域(Russell & Norvig, 2003)。

本研究中的层级表征是一种具有因果性的结构。计算模型的模拟为此提供了证据：层级模型和相关模型均以非离散的方式处理视觉场景中的对象，为它们建立了具有普遍联系的结构，但对于相关模型来说，其表征的结构只考虑了对象之间的相关关系而非因果关系，使其对客体位置的预测精度极易受到运动轨迹间相关性波动的影响，与人类的行为模式不一致，这表明相关模型不足以反应人类视觉表征的本质。更为决定性的证据来自于对层级结构方向属性的考察，即父节点与子节点的绩效必须表现出不对称性。实验一的结果显示，由父节点预测子节点的预测精度更高，这是因为父节点作为上层节点带有更多共同信息，其在产生过程中扮演着原因，而由父节点指向子节点的路径方向则暗示着子节点作为父节点的结果，由原因预测结果的难度相对更低。上述分析表明，人类的视觉系统在处理运动场景时，为运动信息构建了具有因果性的层级表征，其本质是重构运动的产生过程。

5.3 基于层级表征的视觉计算过程模拟

视觉被认为是一个逆向工程(Marr, 1982)，它不仅仅是对知觉对象的直接反应，而是总是试图对其进行解释，寻找背后的产生过程。同时，基于视觉的行为理解、预测任务同样被看作是一个逆向规划的过程(Baker, Saxe, & Tenenbaum,

2009; Jara-ettinger, Schulz, & Tenenbaum, 2015)，建立在对行为产生过程的推论上，进一步基于其产生过程完成理解和预测任务（公式 5.1）。本研究所涉及的两个主要任务（位置预测任务和意图识别任务）针对性的强调了这一加工过程，对其计算过程的模拟也通过逆向规划来实现。

$$\mathrm{P}(Goal/Environment, Action) \propto \mathrm{P}(Goal/Environment) \times \mathrm{P}(Action/Environment, Goal) \quad \ldots\ldots (5.1)$$

位置预测任务主要包括两个阶段，在完全可见阶段观察者形成对运动信息的表征，在部分可见阶段则利用已有的表征和仍然可见的运动信息推测消失客体的运动。推测过程可能存在两种途径：其一，在完全可见阶段仅获取目标客体运动的产生方式，并在部分可见阶段利用产生规则重现目标客体的运动；其二，在完全可见阶段获取所有客体的运动产生方式，并在部分可见阶段利用产生方式和仍然可见的客体重现目标客体的运动。无论采用哪种方式，推测过程的本质是一个逆向规划过程，即寻找在正向过程中最能够维持当前层级表征的运动信息，并将其作为对消失客体的预测(公式 5.2)。显而易见，采用途径一的可能性微乎其微，在完全可见阶段，被试并不确定哪一个客体将会成为目标客体，因而无法针对性的构建独属于目标客体的产生过程。即使被试能够确定目标客体的身份，并针对性的构建了目标客体运动的产生过程，对其位置的预测也应不受到场景中其他运动客体的影响，这显然与研究结果不符（实验三），因此本研究表明视觉的计算过程是基于途径二的逆向过程：在完全可见阶段提取视觉场景中所有客体的运动信息，并构建潜在的层级表征，该表征描述了所有客体运动的产生过程，随后的部分可见阶段，根据仍能被观测到的客体的运动，寻找最能够维持已形成的层级表征的运动，作为对消失客体的具体运动情况的推测。

$$\mathrm{P}(Position/Hierarchy, Motion\_new) \propto \mathrm{P}(Hierarchy/Motion\_old) \ldots (5.2)$$

意图预测任务的计算模拟同样通过逆向过程实现，即建立在运动本身的产生

过程上。对于带有社会属性的客体，其真实运动由自身内在意图和外在限制共同决定，观测者所接收到的运动信息是多种因素共同作用的结果，但并不展现不同驱动力产生最终运动的因果过程，因此观测到的运动信息遮盖了生命体的真实意图，甚至有时表现出相反的运动情况。为了寻找运动的真实产生过程，视觉系统首先需要重构运动的潜在层级结构，分解出对真实运动的产生做出贡献的各个成分，进而分离出其中来自于外界限制的运动成分，获得真正体现客体内在意图的运动信息（公式 5.3）。采用逆向过程的计算方式对于意图识别来说尤为重要，对于外在表现完全相同的运动信息，如果不能从中获取其运动的产生过程，仅依赖于运动信息本身推测意图，将很容易受到与意图无关的运动信息的干扰（研究三），这与视觉系统的高度灵活性是不一致的。

$$\mathrm{P}(Goal, Hierarchy/Motion) \propto \mathrm{P}(Motion/Hierarchy, Goal) \quad (5.3)$$

由上述分析可得，视觉系统基于已构建的层级表征，以逆向工程的计算方式进行进一步加工。本研究根据此观点为基于层级表征的视觉加工过程构建详细的数学模拟。同时，在逆向工程的过程中，层级表征起到了关键的作用，由于运动信息的视觉表征具有因果性，只有采用层级的表征方式处理运动信息，视觉才有可能获得运动的产生过程，完成逆向工程的推理。由此角度推论，运动信息的视觉层级表征是视觉加工顺利进行的必要前提。

5.4 本研究的贡献与创新

研究的相关结论和模型的模拟结果为视觉加工的相关理论做出贡献。基于信息加工的视觉认知研究有两个核心任务：探寻视觉加工的阶段和寻找视觉对信息的表征(Neisser, 1976)，本研究主要针对于上述第二个问题，系统地探讨了视觉加工中运动信息的表征问题，首次验证了视觉运动层级表征的存在并进一步探索了其性质，为揭示视觉信息表达的规律做出了贡献，同时也对构建视觉加工的计算过程有着重要的启示。本研究的结果还为视知觉与工作记忆的交互模型提供了新的支持，该模型认为，知觉与工作记忆并不是完全独立的两个阶段，而是具有相似的表征和计算机制，在视觉加工的过程中灵活地开展动态交互(Gao, Gao, Li,

Sun, & Shen, 2011; Shen et al., 2015)。运动信息的层级表征构建和计算过程同样依赖于这一交互过程，在构建层级表征时，视觉需要同时完成对运动信息的知觉，以及工作记忆中对层级表征的推断；而在利用层级表征进行计算时，工作记忆中存储的表征和当前知觉中运动信息需协同工作。实际上在运动信息的整个视觉加工过程中，极难严格的划分出知觉与工作记忆的不同阶段或任务，它们始终以相通的计算模式共同完成对运动信息的理解和预测。

本研究在视觉表征的研究方法上有一定创新。一方面，研究拓展了传统心理物理方法的应用，操纵了刺激的潜在结构信息，而非表现出来的物理特征，为探究信息的内在表征提供了一个很好的方法。同时，视觉系统本身如同一个黑箱，视觉研究的核心目标则是尽量揭示黑箱的内容，基于对刺激潜在结构的操纵，研究向着打开黑箱的方向前进了一步，更加接近视觉系统的本质特点；另一方面，研究采用模型的任务模拟对行为结果进行辅助性的补充，模型的任务模拟注重的并非其绝对绩效上的优劣，而是模拟得到的绩效模式与人类行为绩效的异同，这一思路有助于为行为数据的结论提供进一步支持。同时，模型具有更强的过程完整性和可操纵性，为直接着手探索认知加工的内在计算过程提供了可能。

同时本研究所揭示的层级表征有着完备、具体的数学表达，可直接应用于人工智能算法，有助于当前人工智能摆脱对庞大计算能力的依赖，能够在对少量数据进行简单计算的情况下学习到信息背后的本质规律，并迁移到广泛的任务场景中，逐渐向“小数据，大任务”的智能模式转变(唐宁等, 2018; 程少哲等, 2017)。

6 结论及进一步研究设想

6.1 主要结论

本研究采用心理物理学研究方法，通过操纵运动信息背后潜在的层级结构，考察了运动信息在视觉加工中的层级表征，同时辅以计算建模的方法，进一步揭示视觉为运动信息构建层级表征的机制。研究主要得到以下结论：

（1） 运动信息潜在层级结构的变化影响被试的行为绩效，表明视觉加工中形成了对运动信息的层级表征。

（2） 视觉对运动信息的层级表征不受运动信息的时间长度、任务线索和任务目的的影响，具有跨情景的一致性。

（3） 带有社会信息的运动，同样能在视觉加工中以层级结构加以表征。

（4） 视觉对运动信息的层级表征具备因果性，不仅描述了运动的形式，同时描述了运动的产生过程。

（5） 视觉加工基于所构建的层级表征，通过逆向工程的计算完成对运动场景的识别、理解和预测。

上述结果不仅为视觉加工中运动信息层级表征的存在提供了坚实的证据，同时揭示了该层级表征的稳定性和普遍性。此外，本研究的计算模型进一步模拟了视觉系统利用层级表征执行后续加工的过程，为现有人工智能系统向人类智能逼近提供了有益的尝试。

6.2 进一步研究设想

在本研究的基础上，笔者提出以下进一步研究设想：

（1） 人类认知活动的生物基础是大脑的神经活动，探索认知活动的神经基础对深入理解人类认知加工过程有重要意义。后续研究可针对性地考察与运动信息视觉层级表征的构建和进一步加工相关的神经过程。

（2） 本研究中被试均作为观察者，站在第三者视角执行任务。日常生活中，个体很可能作为活动的参与者卷入到运动过程中，此时对整个运动场景的表征如何构建，是一个值得进一步探讨的问题。

（3） 相对稳定的表征有利于认知加工过程的平稳进行，本研究的部分结果

确实反应了这一特性。然而现实生活中，视觉对象的运动规律可能存在变化：一方面，对运动产生限制的外部环境容易发生改变，譬如风的方向发生变化导致风筝的运动模式改变；另一方面，运动对象自身的意图也可能发生变化，譬如捕食者在追逐猎物的中途遭遇天敌，其目标转变为逃生。视觉能否处理这样灵活的场景？表征如何在潜在运动模式变化时维持相对稳定？对上述问题的进一步研究探讨，将有助于加深对视觉表征及其计算过程的理解。

（4） 层级表征的构建需要对因果关系具有一定程度的理解，同时受到个体对运动的产生过程的经验和偏好的影响。因此，在个体发展的角度上考察层级表征是一个有意思的问题，一方面能够揭示表征的发展规律，另一方面也能够探索在复杂的社会场景下，经验对底层视觉加工的影响。

参考文献

Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *Quarterly Journal of Experimental Psychology*, *59*, 1101-1116.

Alvarez, G. A., Arsenio, H. C., Horowitz, T. S., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 643–667.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13): 14, 1_10.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*(3), 708–717.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Beutter, B. R., & Stone, L. S. (2000). Motion coherence affects human perception and pursuit similarly. *Visual Neuroscience*, *17*(1), 139–153.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(2), 353–69.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*(1), 47–73.

Blei, D., Griffiths, T. L., Jordan, M. I., & Tenenbaum, J. B. (2004). Hierarchical topic models and the nested Chinese restaurant process. *Advances in Neural Information Processing Systems*, *16*, 106-113.

Botterill, K., Allen, R., & McGeorge, P. (2011). Multiple-object tracking the binding of spatial location and featural identity. *Experimental Psychology*, *58*(3), 196–200.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109.

Braund, M. J. (2008). The structures of perception: An ecological perspective. *Kritike: An Online Journal of Psychology*, *2*(1), 123–144.

Caicedo, J. C., & Lazebnik, S. (2015). Active object localization with deep reinforcement learning. *Proceedings of the IEEE International Conference on Computer Vision*, 2488–2496.

Chi, Z., & Geman, S. (1998). Estimation of probabilistic context-free grammars. *Computational Linguistics*, *24*(2), 299–305.

Chomsky, N. (1964). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(1), 109–127.

Clair, R., Huff, M., & Seiffert, A. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, *10*(4), 1–13.

Dasser, V., Ulbaek, I., & Premack, D. (1989). The perception of intention. *Science*, *243*(4889), 365–367.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*,

19(1), 158–164.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition, 80*(1-2), 61-95.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, *113*(4), 501–17.

Duncker, K. (1929). Uber induzierte Bewegung - Ein Beitrag zur Theorie optisch wahrgenommener Bewegung. *Psychologische Forschung*, *12*(1), 180–259.

Endress, A. D., Korjoukov, I., & Bonatti, L. L. (2017). Category-based grouping in working memory and multiple object tracking. *Visual Cognition*, *25*(9–10), 868–887.

Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., & Kellman, P. J. (2013). Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1625–1637.

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, *69*(4), 567–577.

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, *107*(3), 904–931.

Froyen, V., Feldman, J., Singh, M., Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping : Perceptual grouping as mixture estimation. *Psychological Review*, *122*(4), 575–597.

Gao, T., Gao, Z., Li, J., Sun, Z., & Shen, M. (2011). The perceptual root of object-based storage: An interactive model of perception and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1803–1823.

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154–179.

Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of

animacy. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 669–684.

Gao, Z., Gao, Q., Tang, N., Shui, R., & Shen, M. (2016). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition, 146*, 277–288.

Gao, Z., Yin, J., Xu, H., Shui, R., & Shen, M. (2011). Tracking object number or information load in visual working memory: Revisiting the cognitive implication of contralateral delay activity. *Biological Psychology*, *87*(2), 296-302.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research*, *126*, 232–241.

Gorniak, P., & Roy, D. (2004). Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research, 21*, 429-470.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychiatry*, *57*, 243–259.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception and Psychophysics*, *69*(2), 172–184.

Humboldt, W. von. (1836). *On language: The diversity of human language-structure and its influence on the mental development ofmankind*. Cambridge, England, UK: Cambridge University Press.

Humphreys, G. W., & Riddoch, M. J. (2001). Detection by action: Neuropsychological evidence for action-defined templates in search. *Nature Neuroscience, 4*(1), 84–88.

Ilg, U. J., & Churan, J. (2004). Motion perception without explicit activity in areas MT and MST. *Journal of Neurophysiology*, *92*(3), 1512–1523.

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), Language and space. Language, speech, and communication (pp. 1–30). Cambridge, MA: MIT Press.

Jara-ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2015). The naïve utility calculus : Joint inferences about the costs and rewards of actions. *Proceedings of the 37th*

Annual Conference of the Cognitive Science Society, 974–979.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.

Johansson, G. (1975). Visual motion perception. *Scientific American*, *232*(6), 76–88.

Johnson, S. G. B., & Keil, F. C. (2014). Causal inference and the hierarchical structure of experience. *Journal of Experimental Psychology: General*, *143*(6), 2223–2241.

Joo, J., Wang, S., Zhu, S.-C., & Wagemans, J. (2012). *"Hierarchical organization by And-Or Tree" in Book Chapter in Handbook of Perceptual Organization*, New York, NY, USA:Springer, 1–17

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.

Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin and Review*, *22*(6), 1784–1790.

Karpathy, A., & Li, F. F. (2015). Deep visual-semantic alignments for generating image descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 664–676.

Kiley Hamlin, J., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Kim, I.-K., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*(3), 339–362.

Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, *31*(1), 18–30.

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal Of Neurophysiology*, *91*(2), 591–603.

Laverick, R., Wulff, M., Honisch, J. J., Chua, W. L., Wing, A. M., & Rotshtein, P. (2015). Selecting object pairs for action: Is the active object always first? *Experimental Brain Research*, *233*(8), 2269–2281.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, *15*(4), 10.

Lisberger, S. G., & Ferrera, V. P. (1997). Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. *The Journal of Neuroscience*, *17*(19), 7490–7502.

Liverence, B. M., & Scholl, B. J. (2011). Selective attention warps spatial representation: Parallel but opposing effects on attended versus inhibited objects. *Psychological Science*, *22*(12), 1600–1608.

Luck, S., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjuctions. *Nature*, *390*(6657), 279–281.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information. San Francisco, CA, Chicago: Henry Holt & Company.*

Masson, G., Proteau, L., & Mestre, D. R. (1995). Effects of stationary and moving textured backgrounds on the visuo-oculo-manual tracking in humans. *Vision Research*, *35*(6), 837–852.

McKeefry, D. J., Burton, M. P., & Vakrou, C. (2007). Speed selectivity in visual short term memory for motion. *Vision Research*, *47*(18), 2418–2425.

Michotte, A. E. (1950). The emotions regarded as functional connections. In M. Reymert (Ed.), Feelings and emotions: The Mooseheart symposium (pp. 114–125). New York: McGraw-Hill. [Reprinted in Thinès, G., Costall, A., & Butterworth, G. (Eds.).(1991). Michotte's experimental phenomenology of perception (pp. 103–116). Hillsdale, NJ: Erlbaum.]

Minchotte, A. E. (1963). The perception of causality. *Perception*, (11), 173–186.

Narasimhan, S., Tripathy, S. P., & Barrett, B. T. (2009). Loss of positional information when tracking multiple moving dots: The role of visual memory. *Vision Research*, *49*(1), 10–27.

Neisser, U. (1967). *Cognitive psychology*: Classic edition. Psychology Press.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186–1192.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*(4), 441–474.

Pearl, J. (2009). *Causality*. *Cambridge University Press*.

Pomerantz, J., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*. (pp.36-1–36-46). New York, NY: Wiley.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, *4*(5), 197–207.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822.

Pylyshyn, Z. W., & Annan, V. (2006). Dynamics of target selection in Multiple Object Tracking (MOT). *Spatial Vision*, *19*(6), 485–504.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.

Rensink, R. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.

Russell, J. S., & Norvig, P. (2003). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259–290.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1–2), 159–177.

Schultz, J., Friston, K. J., O'Doherty, J., Wolpert, D. M., & Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, *45*(4), 625–635.

Shen, M., Xu, H., Zhang, H., Shui, R., Zhang, M., & Zhou, J. (2015). The working memory Ponzo illusion: Involuntary integration of visuospatial information stored in visual working memory. *Cognition*, *141*(8), 26-35.

Shooner, C., Tripathy, S. P., Bedell, H. E., & Ogmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision*, *10*(6), 8.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214–241.

Smith, K., Battaglia, P., & Vul, E. (2013). Consistent physics underlying ballistic motion prediction. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 3426–3431.

Southgate, V., & Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Developmental Psychology*, *45*(6), 1794–1798.

Sperber, D., Premack, D., & Premack, A. J. (1995). *Causal cognition: A multidisciplinary debate*. Oxford: Clarendon Press.

Spering, M., & Gegenfurtner, K. R. (2007). Contextual effects on smooth-pursuit eye movements. *Journal of Neurophysiology*, *97*(2), 1353–1367.

Spering, M., & Montagnini, A. (2011). Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye

movements: A review. *Vision Research*, *51*(8), 836–852.

Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: Impairments due to common item motions. *Perception*, *35*(4), 483–495.

Sun, Z., Huang, Y., Yu, W., Zhang, M., Shui, R., Gao, T. (2015). How to break the configuration of moving objects ? Geometric invariance in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1247–1259.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*(1), 49–100.

Tenenbaum, J. B. (1999). *A Bayesian framework for concept learning.* (Doctoral dissertation, Massachusetts Institute of Technology).

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*(8), 943–951.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*(2), 58–64.

Wallach, H. (1935). Uber visuell wahrgenommene Bewegungsrichtung. *Psychological Research*, *20*(1), 325–380.

Wang, S. H., & Baillargeon, R. (2006). Infants' physical knowledge affects their change detection. *Developmental Science*, *9*(2), 173–181.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*(1), 301–350.

Xu, H. K., Tang, N., Zhou, J. F., Shen, M. W., & Gao, T. (2017). Seeing "what" through "why": Evidence from probing the causal structure of hierarchical motion. *Journal of Experimental Psychology: General*, *146*(6), 896-909.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual

organization. *Cognitive Psychology*, *24*(3), 295–340.

Yin, J., Ding, X., Zhou, J., Shui, R., Li, X., & Shen, M. (2013). Social grouping: Perceptual grouping of objects by cooperative but not competitive relationships in dynamic chase. *Cognition*, *129*(1), 194–204.

Yin, J., Xu, H., Ding, X., Liang, J., Shui, R., & Shen, M. (2016). Social constraints from an observer' perspective: Coordinated actions make an agent's position more predictable. *Cognition*, *151*, 10-17.

Yin, J., Zhou, J., Xu, H., Liang, J., Gao, Z., & Shen, M. (2012). Does high memory load kick task-irrelevant information out of visual working memory? *Psychonomic Bulletin and Review*, *19*(2), 218-224.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*, 80–84.

Zhao, L., Gao, Q., Ye, Y., Zhou, J., Shui, R., & Shen, M. (2014). The role of spatial configuration in multiple identity tracking. *PLoS ONE, 9*(4), e93835.

Zhou, J., Huang, X., Jin, X., Liang, J., Shui, R., & Shen, M. (2012). Perceived causalities of physical events are influenced by social cues. *Journal of Experimental Psychology: Human Performance and Perception, 38*(6), 1465-1475.

Zhou, J., Zhang, H., Ding, X., Shui, R., & Shen, M. (2016). Object formation in visual working memory: Evidence from object-based attention. *Cognition*, *154*, 95–101.

Zhu, S. C., & Mumford, D. (2006). A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, *2*(4), 259–362.

Zhu, S. C. (1999). Embedding gestalt laws in markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(11), 1170–1187.

Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision*, *11*(14), 2.Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.

周吉帆, 徐昊骅, 唐宁, 史博皓, 赵阳, 高涛, … 沈模卫. (2016). “强认知”的心理学研究：来自AlphaGo的启示. 应用心理学, *22*(1), 3–11.

唐宁，安玮，徐昊骅，周吉帆，高涛，沈模卫. (2018). 从数据到表征：人类认知对人工智能的启发. 应用心理学, *24*(1), 3–14.

程少哲, 史博皓, 赵阳, 徐昊骅, 唐宁, 高涛, 周吉帆沈模卫. (2017). 对注意的再思考：一个注意的强化学习模型. 应用心理学, *23*(1), 3–12.

附录

附录一：运动的产生与层级结构的确认

层级结构的产生通过嵌套中国餐馆过程（nested Chinese Restaurant Process, nCRP）实现，该过程是中国餐馆过程（Chinese Restaurant Process, CRP）的迭代。CRP 描述了每一个客体进入到层级结构的某一层时，将被分配到哪个节点的概率，具体计算公式如下：

$$\mathrm{P}(C_{nd}=j|\boldsymbol{C}_{1:n-1})=\begin{cases}\frac{M_j}{n-1+\gamma}, & j\le J\\ \frac{\gamma}{n-1+\gamma}, & j=J+1\end{cases} \quad \text{……（1）}$$

其中C_{nd}表示客体 n 在 d 层的分配，$\boldsymbol{C}_{1:n-1}$表示客体 n 之前的 n-1 个客体的分配情况。j 是对该层已有节点的编号，J 表示该层已有节点的总数。M_j表示在节点 j 上已经分配的客体总数，γ是 CRP 的参数，用于调节产生一个新节点作为对客体 n 的分配的概率，本研究中γ取值为 1，代表不对是否产生一个新节点进行任何有偏的假设。上述过程表明，客体 n 的分配依赖于已有节点上分配好的客体的数量，同时有一定概率产生一个新节点。通过迭代调用 CRP 并计算所有可能的分配情况，可以得到总数为 n 的客体对应的所有可能的层级结构的概率（终止节点为空的层级结构将不予考虑）。

理论上，层级结构可以具有无限的深度，但对于固定总数的客体来说，超过某个深度的层级结构不再具有外在表现（本研究中为运动信息）上的区别，因此将可能的最大深度固定为与客体总数相同。nCRP 将产生所有深度为最大深度的层级结构，但在实际中，必须考虑深度较浅的层级结构，因此采用马尔科夫随机场的方式对深度进行采样，不同深度的概率可由如下公示计算：

$$\mathrm{P}(\boldsymbol{d}) \propto exp\{\alpha \sum_{m=1}^{N} \sum_{n>m}^{N} \blacksquare[d_m = d_n] - \rho \sum_{n=1}^{N} d_n\} \quad \text{（2）}$$

其中■[]是一个判断函数，当括号内条件成立时返回 1，否则返回 0。$\boldsymbol{d}$是描述所有客体深度的向量，d_m和d_n分别表示客体 m 和客体 n 的深度，参数α调节了客体间深度相同的概率，参数ρ则调节了客体深度的具体深浅。本研究中将参数设定为$\alpha = 1$和$\rho = 0.1$，与以往研究一致。

每一时刻客体的位置 $\boldsymbol{s}_t$ 由当前时刻的运动 f 与上一时刻的位置 $\boldsymbol{s}_{t-1}$ 相加得到。向层级结构中节点分配运动分量的过程通过高斯过程实现，高斯过程的均值和协方差矩阵如下：

$$\mathrm{P}(\boldsymbol{f}) = \mathrm{GP}(\boldsymbol{f}; \mathrm{m}, \mathrm{k}),\ \mathrm{m}(\boldsymbol{s}) = 0,\ \mathrm{k}(\mathrm{s}, \mathrm{s}') = \tau\, exp\left\{-\frac{\|s-s'\|^2}{2\lambda}\right\} \quad \text{（3）}$$

此处 s 和 s’只考虑具有相同父节点的客体运动分量。参数τ 和 λ 调控运动的平滑程度，本研究中根据以往研究将其设为τ = 1、λ = 100。

上述过程描述了产生特定的层级结构，并基于该层级结构产生运动的概率分布。对运动背后潜在层级结构的确认通过逆向的贝叶斯推理实现：

$$\mathrm{P}(\boldsymbol{c}, \boldsymbol{d}|\boldsymbol{s}) \propto \mathrm{P}(\boldsymbol{c})\mathrm{P}(\boldsymbol{d})\mathrm{P}(\boldsymbol{s}|\boldsymbol{c}, \boldsymbol{d}) \quad \text{（4）}$$

其中$\boldsymbol{c}$客体在层级结构中的分配情况，$\boldsymbol{d}$表示层级结构的深度，它们共同描述了层级结构的具体形式。$\boldsymbol{s}$是客体在真实场景中的位置，即一段时间内的客体运动轨迹。贝叶斯推理过程中后验概率的具体计算通过随即采样过程实现，本研究中采用了 Gibbs 采样，并将概率最大值作为推理的最终结果。

附录二：位置预测任务的模型模拟

对位置预测任务的模拟包括两部分，一部分是在完全可见阶段将运动信息编码为特定的表征形式并储存，另一部分是在部分可见阶段根据仍旧可见的运动信息和储存的表征形式，计算不可见客体的运动。两类模型的区别在于对运动信息采用不同的编码表征方式，并进而导致后续计算过程的差异。

层级模型采用层级结构描述运动，采用如附录一中相同的方法计算运动对应的层级结构，在部分可见阶段则通过如下公式计算消失客体可能的位置分布：

$$\mathrm{P}(\boldsymbol{s_n}|\boldsymbol{s_{1:n-1}},\boldsymbol{c},\boldsymbol{d}) \propto \mathrm{P}(\boldsymbol{c},\boldsymbol{d})\mathrm{P}(\boldsymbol{s}|\boldsymbol{c},\boldsymbol{d}) \quad (5)$$

其中$\mathrm{P}(\boldsymbol{c},\boldsymbol{d})$即为完全可见阶段计算出的层级结构。

相关模型采用全连接网络描述运动，运动由高维高斯过程产生，以两个客体为例，其均值和协方差矩阵如下：

$$\mu = [\mu_1, \mu_2],\ \Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix},\ \Lambda = \Sigma^{-1} = \begin{bmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix} \quad (6)$$

协方差矩阵描述了客体间运动的相关程度。在部分可见阶段，对消失客体的位置预测符合如下概率：

$$p(x_1|x_2) \sim Normal(x_1|\mu_{1|2}, \Sigma_{1|2}), \quad (7)$$

其中 $\Sigma_{1|2} = \Lambda_{11}^{-1}$, $\mu_{1|2} = \Sigma_{1|2} \times (\Lambda_{11}\mu_1 - \Lambda_{12} \times (x_2 - \mu_2))$

两类模型均得到位置的概率分布，通过对所得分布进行抽样的方法模拟单个被试的单个试次，并对所得结果进行与被试行为结果完全相同的分析过程。

附录三：意图识别任务的模型模拟

意图识别任务是行为规划的反过程，后者在已知意图的情况下选择行为，在计算上通过马尔科夫决策过程（Markov Decision Process, MDP）实现。MDP 的核心假设是生命体由意图驱动的运动遵循完全理性的行为模式。以追逐场景为例，每一时刻狼的运动（A）从五种可能的行为中产生：向右、向左、向上、向下或静止。同时，狼能够观测到场景中所有有意义的状态（S），包括狼自身的真实位置和运动、羊的真实位置和运动、干扰子的真实位置和运动，以及运动范围的边界位置。狼的奖赏函数（R）由狼和羊的距离决定，距离越近，奖励越大。基于此，狼在特定状态下对行为的选择可由下面的公式计算：

$$\mathrm{V(s)} = \max_{a}[R(s,a) + \gamma \sum_{s \in S} T(s,a,s')V(s')] \quad \text{（8）}$$

其中 V(s)表示在状态 s 下的行为选择策略网络，R(s,a)表示在状态 s 下执行行为 a 的奖赏，T(s,a,s')描述了在 s'状态下执行行为 a 到达 s 状态的概率，γ 则描述了上一步（过去时间）的情况对当前决策的贡献比例。

上述过程可以得到在对应状态和意图下产生行为的概率分布，基于此可通过贝叶斯定理反推观测到的行为背后的意图：

$$\mathrm{P}(\boldsymbol{s},\boldsymbol{g}|\boldsymbol{a}) \propto \mathrm{P}(\boldsymbol{a},\boldsymbol{s}|\boldsymbol{g}) \quad \text{（9）}$$

公式右边由 MDP 过程得到，行为 a 和状态 s 是一个互相迭代更新的过程。公式左边的推理需要同时求解状态 s 和意图 g，因此这是一个联合推理，其求解通过 Gibbs 采样完成。

两类模型的区别在于对状态 s 的表达。层级模型将状态 s 中所有客体的运动表达为层级结构，并依据附录一中的概率计算过程进行 MDP 计算。一般模型则将 s 表达为全连接网络，具体计算过程与附录二相似。

作者简历

教育背景

2013.9 — 2018.6	浙江大学	心理系应用心理学	博士学位
2008.9 — 2013.6	浙江大学	心理系应用心理学 竺可桢学院理科平台	学士学位

期刊文章

Xu, H. K., Tang, N., Zhou, J. F.*, Shen, M. W.*, & Gao, T.* (2017). Seeing "what" through "why": Evidence from probing the causal structure of hierarchical motion. *Journal of Experimental Psychology: General*, 146(6): 896-909.

Yin, J., **Xu, H. K**., Duan, J. P., & Shen, M. W. (2018). Object-based attention on social units: Visual selection of hands performing a social interaction. *Psychological Science*. (In press).

Yin, J., **Xu, H. K.**, Ding, X. W., Liang, J. Y., Shui, R. D., & Shen, M. W. * (2016). Social constraints from an observer's perspective: Coordinated actions make an agent's position more predictable. *Cognition.* 151: 10-17.

周吉帆，**徐昊骅**，唐宁，史博皓，赵阳，高涛，沈模卫*.“强认知”的心理学研究：来自 AlphaGo 的启示[J]. 应用心理学, 2016, 22(1): 3-11.

Shen, M. W., **Xu, H. K.,** Zhang, H. H., Shui, R. D., Zhang, M., & Zhou, J. F.*** (**2015**).** The working memory Ponzo illusion: Involuntary integration of visuospatial information stored in visual working memory. *Cognition.* 141: 26-35.

卢剑刚，**徐昊骅**，尹军，高在峰，沈模卫*. 视觉客体在工作记忆中的累积构建[J]. 应用心理学, 2010, 16(3): 195-200.

唐宁，安玮，**徐昊骅**，周吉帆，高涛*，沈模卫*. 从数据到表征：人类认知对人工智能的启发[J]. 应用心理学, 2018, 24(1): 3-14.

Yin, J., Ding, X. W., **Xu, H. K.**, Zhang, F., & Shen, M. W. (2017). Social Coordination Information in Dynamic Chase Modulates EEG Mu Rhythm. *Scientific Reports, 7(1), 4782*.

Li, J., Shao, N., **Xu, H. K.**, Shui, R. D., & Shen, M. W.* (2013). Does visual working memory work as a few fixed slots? *The Quarterly Journal of Experimental Psychology, 66(11): 2103-2117.*

Yin, J., Zhou, J. F., **Xu, H. K.**, Liang, J. Y., Gao, Z. F., & Shen, M. W.* (2012). Does high memory load kick task-irrelevant information out of visual working memory? *Psychonomic Bulletin & Review, 19: 218-224.*

Gao, Z. F., Yin, J., **Xu, H. K.**, Shui, R. D., & Shen, M. W.* (2011). Tracking object number or information load in visual working memory: Revisiting the cognitive implication of contralateral delay activity. *Biological Psychology, 87: 296-302*

程少哲，史博皓，赵阳，**徐昊骅**，唐宁，高涛，周吉帆*，沈模卫*. 对注意的再思考：一个注意的强化学习模型[J]. 应用心理学, 2017, 23(1): 3-12.

学术会议

Xu, H. K., Tang, N., Zhou, J. F., Shui, R. D., Shen, M. W., & Gao, T. (2018). Perceiving animacy with causal constraints: A “leash resistance” effect in chasing detection. Poster present at the 18th Annual Visual Sciences Society Meeting.

Cheng, S. Z., **Xu, H. K**., Tang, N., Zhou, J. F., Chen, H., Liang, J. Y., Gao, T. & Shen, M. W. (2018). Hierarchical constrains on the distribution of attention in dynamic displays. Poster present at the 30th Annual Convention of Association for Psychological Science.

Xu, H. K., Zhou, J. F., Tang, N., & Shen, M. W. (2016). Hierarchical representation of motion scene: Closer relationship in the motion tree makes the position of moving object more predictable. Talk present at the 31st International Congress of Psychology.

徐昊骅, 史博皓, 周吉帆, 沈模卫. (2017). 动态场景中意图识别的计算模型. 口头报告, 第二十届全国心理学大会.

Yin, J., **Xu, H. K**., Ding, X. W., Shui, R. D., & Shen, M. W. (2015). Social constraints from an observer's perspective: Coordinated actions make agent's position more predictable. Poster present at the 6th Joint Action Meeting.

Xu, H. K., Yin, J., Wu, F., Wang, X., & Shen, M. W. (2013). Cooperation information affects distance perception. Poster present at the 2013 Annual Conference of Society for Social Neuroscience (S4SN).

Xu, H. K., Yin, J., Ding, X. W., & Shen, M. W. (2013). Cooperation and competition affect distance perception. Poster present at the 9th Asia-Pacific Conference on Vision.