

浙江大学

本科生毕业论文（设计）



题目 遮挡条件下的大小统计知觉

姓名与学号 史博皓 3120100969

指导教师 沈模卫 教授

年级与专业 2012 级 心理学

所在学院 心理与行为科学系

第一部分 文献综述及开题报告

第二部分 毕业论文正文

浙 江 大 学

本 科 生 毕 业 论 文（设计）
文 献 综 述 和 开 题 报 告



题目 遮挡条件下的大小统计知觉

姓名与学号 史博皓 3120100969

指导教师 沈模卫 教授

年级与专业 2012 级 心理学

所在学院 心理与行为科学系

一、题目：遮挡条件下的大小统计知觉

二、指导教师对文献综述和开题报告的具体内容要求：

指导教师（签名）_____

2016 年 1 月 16 日

目录

文献综述 7

1. 工作记忆 7

 1.1 工作记忆与知觉交互模型 7

 1.2 工作记忆的整合 8

2. 统计知觉 8

 2.1 统计知觉的普遍性 8

3. 问题的提出和意义 9

参考文献: 10

开题报告 13

1. 问题提出背景 13

2. 研究意义 13

3. 可行性分析 14

4. 调研报告 14

5. 研究方案 14

6. 预期结论 15

参考文献 15

文献翻译 19

外文原文 23

一、文献综述

(一) 工作记忆

工作记忆是一种容量有限的记忆系统,它为复杂的认知任务提供了临时储存空间以存放加工所需的信息(Baddeley, 2003; Baddeley & Hitch, 1974)。Baddeley 和 Hitch 在 1974 年最早提出了工作记忆的多成分结构模型,包括语音环路、视觉空间模板和中央执行系统(Baddeley & Hitch, 1974)。此后,大量研究以言语信息和视觉信息为记忆材料对工作记忆的机制进行了探索。

1 工作记忆与知觉交互模型

在早期,利用视觉信息进行的研究一直将工作记忆看作是储存知觉加工产物的心理结构。Luck 和 Vogel 提出,视觉工作记忆以整合的客体形式储存视觉信息,且容量大约为 3-4 个客体(Luck & Vogel, 1997)。随后大量研究探讨了视觉工作记忆的容量、表征精度和储存方式(Vogel & Luck, 2001; Olson & Jiang, 2002; Wheeler & Treisman, 2002)。近年来,不少研究者开始将工作记忆看作是知觉加工进程中非常关键的一个部分。有研究者认为,视觉工作记忆的一个重要作用是当视觉信息变化或是被打断时,保持知觉的连贯和稳定。例如,一项眼动研究发现,工作记忆可以在视觉搜索被目标无关信息打断之后,快速恢复对先前目标相关信息的记忆,从而保证知觉的连贯性(Hollingworth, Richard, & Luck, 2008)。还有观点则更进一步,提出无论知觉加工中断与否,视觉工作记忆都会与视知觉产生交互。例如有研究提出工作记忆中存储的信息会影响个体知觉(Kang, Hong, Blake, & Woodman, 2011),眼动研究发现工作记忆与眼跳之间存在交互作用(Hollingworth, Matsukura & Luck, 2013)等。在此基础上,高涛等人提出了视觉工作记忆与知觉的交互模型。该模型认为,工作记忆与知觉系统之间存在密切交互,工作记忆参与视知觉加工过程并为加工过程所涉及的视觉表征提供暂时存储。该研究随后进行了一系列实验证明了工作记忆中确实存在不同的机制对不同加工程度的知觉信息进行存储,从而有效地支持了其假设模型(Gao, Gao, Li, Sun, & Shen, 2011)。且整合方式类似知觉加工,有力地支持了视觉工作记忆与知觉的交互模型理论。但该研究也指出,其结论不适用于艾宾浩斯错觉。说明视觉工作记忆的整合受到它的容量限制。最新研究还通过 CDA 幅值证明了客体在视觉工作记忆中可以根据格式塔原则进行组织,并在行为指标上表现出记忆效率的提高。该结果也说明信息在工作记忆中可以发生类似于知觉的整合,支持了交互模型理论(Gao, Gao, Tang, Shui, & Shen, under review)。然而现实世界充斥着大量复杂多变的信息,远远超过了视觉工作记忆仅有的 3 到 4 个客体容量。在交互模型下,个体如何通过有限的视觉工作记忆资源,将大量复杂视觉信息整合成完整

视觉表征，上述研究尚未涉及。

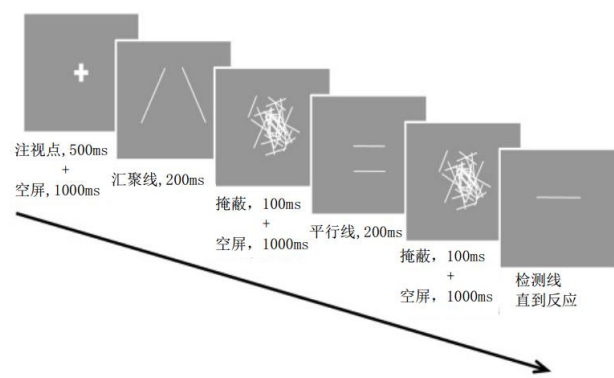


图 1 工作记忆中蓬佐错觉整合实验流程示意图(Shen et al., 2015)

2 工作记忆的整合

根据工作记忆与知觉交互模型，存储在视觉工作记忆中的信息依然可能得到进一步的知觉加工。沈模卫等人对此进行的研究发现(Shen et al., 2015)，即使将形成蓬佐错觉的知觉信息分离成两部分，使之先后进入并储存于视觉工作记忆中，错觉仍然会发生，具体实验见图 1。该结果表明，视觉工作记忆中存在对已存储信息的整合，将大量复杂视觉信息整合成完整视觉表征，上述研究尚未涉及。

（二）统计知觉

统计知觉是视觉系统突破认知资源有限性的有效方式之一。当面对一系列相似客体时，个体会自动将信息进行压缩，忽略客体细节特征，从而快速精确地获得总体统计属性(Alvarez, 2011)。

1 统计知觉的普遍性

近年大量研究探讨了统计知觉的普遍性和加工机制。最具里程碑意义的是 Ariely 的一项研究，发现被试虽然不能很好地记忆圆组中单个圆的大小，但可以相当精确地估计一组圆的平均大小(Ariely, 2001)。由此，统计知觉开始受到研究者的关注。随后越来越多的实验证明了统计知觉在视觉中普遍存在：例如平均大小 (Ariely, 2001)，方向 (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001)，空间格局 (Alvareza & Olivab, 2009)，甚至是一组人脸的平均表情(Haberman & Whitney, 2007) 等都能得到准确估计。有研究发现，给被试呈现一组动态变化的客体或一个随时间变化的客体，被试都能准确地估计其平均大小，且其精度与对静态客体平均大小的判断没有差别(Albrecht & Scholl, 2010)。这为统计知觉如何适用于动态变化的现实世界提出了可能。此外，还有研究通过给被试同时呈现听觉和视觉信息，发现被试对两者均值都能进行准确估计，证明统计知觉同样存在于听觉通道，存在跨通道的一致性 (Albrecht, Scholl, & Chun,

2012) 。 2. 统计知觉的加工机制 在统计知觉的加工机制方面,有研究发现,被试在 50ms、100ms、200ms 和 1000ms 四种呈现时间下,对 12 个圆的平均大小估计的准确性没有显著差异。且在这四种呈现条件下对两组圆大小比较的绩效与对两个圆大小比较的绩效相似,表明统计知觉的加工可能不需要集中注意的参与,且是一个自动的平行加工过程(Chong & Treisman, 2003)。Brady 等人研究工作记忆中存储的统计知觉信息发现,当用颜色对同时呈现的一堆圆进行分组时,被试记忆的个体圆大小会偏向所有圆的平均大小或同色圆平均大小,进一步说明了统计知觉是不受控制的自动加工过程(Brady & Alvarez, 2011)。在上述研究基础上,张军伟等人对统计知觉的加工机制做了进一步研究。该研究根据格式塔邻近性原则将圆分成个数和平均大小都不相同的两组,发现被试对总体平均值的估计会介于真实的总体平均值与二次平均(即对两组圆的均值进行平均)的总体平均值之间。因此,作者认为统计知觉的计算过程采用逐层加权平均法,且各组统计知觉会受总体影响(张军伟, 2014)。由此可见,对统计知觉加工机制的研究现已有不小进展,但工作记忆中统计表征的整合机制是否与知觉类似目前尚不明确。

(三) 问题提出和意义

现实生活中,大量片段信息往往不会同时被视觉系统所接收。为了保证知觉的完整性,信息会被临时储存在工作记忆中,并在其中进行整合,从而形成完整表征。但视觉工作记忆的容量存在限制,不可能储存所有细节信息。而前人对工作记忆整合的研究都局限于对简单刺激或者单一客体信息的整合,不能完全解释个体如何整合现实生活中大量复杂的知觉信息。而统计知觉相关研究表明,对单个总体的统计知觉准确性不受总体中客体数目多少的影响(Attarha, Moore, & Vecera, 2014)。因此抽取各个片段信息的统计表征存储于工作记忆,并基于储存的表征进行整合,从而获得完整知觉,应是突破视觉工作记忆容量限制的一种有效方式。本研究拟利用统计知觉自动平行加工和逐层加权平均的特征,探究统计表征是否能在工作记忆中整合,并考察其加工机制与统计知觉加工机制是否相同。本研究将有助于深入理解认知系统如何高效完整地表征现实世界的复杂场景,为人们如何使用有限的工作记忆容量整合大量的信息提出一种可能的解释;也可从统计表征领域,为工作记忆与知觉交互模型提供进一步证据。

尽管对单一整体的统计知觉不受整体内部客体数量的限制,但是对多个整体的统计知觉研究发现,记忆 4 组圆各自均值的绩效显著低于记忆两组圆的均值(Attarha et al., 2014)。因此本研究使用两组圆作为刺激,将要求被试记忆的统计表征数量控制在 3 种,防止因工作记忆容量限制引起统计表征无法在其中整合。两组圆材料中一组为 5 个(以下称为 A 组),另一组为 13 个(以下称为 B 组)。根据统计知觉逐层加权平均的加工方式,

大脑在计算总体均值时，对于圆数量不同的组，给予的权重不同(张军伟，2014)。因此，若两组圆的统计表征在工作记忆中发生整合，且整合方式与知觉相同，则无论 A 组还是 B 组圆均值大，对总体均值的估算将更多地受到 B 组圆均值的影响。但由于实验材料中每组客体数量都超过视觉工作记忆容量，被试对前后两帧圆组可能只能记住平均大小，而不能记住其他信息。因此被试对总体均值的估计，也可能运用二次平均的方式，即只存储两组圆的统计特征，并在报告总体均值时依据记忆的两组均值进行平均。为了更好地探究工作记忆中统计表征的整合方式，实验应该分两种情况进行，一种情况为 A 组均值大于 B 组（实验一），另一种情况为 A 组均值小于 B 组（实验二）。

参考文献：

张军伟.(2014). 统计知觉的层次计算模型. 硕士学位论文. 浙江大学.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*, 21(4), 560-567.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, 74(810-815).

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3).

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of the Sciences*, 106(18), 7345-7350.

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, 12(2), 157-162.

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary Statistics of Size: Fixed Processing Capacity for Multiple Ensembles but Unlimited Processing Capacity for Single Ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1440-1449.

- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751–R753.
- Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual Working Memory Modulates Rapid Eye Movements to Simple Onset Targets. *Psychological Science*, 24(5), 790-769.
- Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the Function of Visual Short-Term Memory: Transsaccadic Memory, Object Correspondence, and Gaze Correction. *Journal of Experimental Psychology*, 137(1), 163-181.
- Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, 18(860-869).
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279-281.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, 64(7), 1055-1067.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739-744.
- 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, 26-35.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48.
- Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of vision*, 15(4), 16-16.

Shi Z, Church RM, & Meck WH. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 17(11), 556-64.

二、开题报告

1 问题提出背景

在人的日常生活中，经常需要对大量的事物进行视觉加工和记忆，而人们用于这些加工过程的认知资源是相当有限的，比如说，人的工作记忆容量仅仅是三到四个简单客体。但仅凭这些有限的认知资源，人就能迅速加工并识别出事物的总体属性，比如人群的密度，大量物体的大致大小等等。人的这一认知能力被概括为统计知觉（Ariely, 2001）。

统计知觉是视觉系统突破认知资源有限性的有效方式之一。当面对一系列相似客体时，个体会自动将信息进行压缩，忽略客体细节特征，从而快速精确地获得总体统计属性（Alvarez, 2011）。如在Ariely的一项研究中，发现被试虽然不能很好地记忆圆组中单个圆的大小，但可以相当精确地估计一组圆的平均大小（Ariely, 2001）随后越来越多的实验证明了统计知觉在视觉中普遍存在：例如平均大小（Ariely, 2001），方向（Parkes, Lund, Angelucci, Solomon, & Morgan, 2001），空间格局（Alvarez & Oliva, 2009），甚至是一组人脸的平均表情（Haberman & Whitney, 2007）等都能得到准确估计。有研究发现，给被试呈现一组动态变化的客体或一个随时间变化的客体，被试都能准确地估计其平均大小，且其精度与对静态客体平均大小的判断没有差别（Albrecht & Scholl, 2010）。这为统计知觉如何适用于动态变化的现实世界提出了可能。

对于统计知觉的机制，已经有了不少进展（如Chong & Treisman, 2003; Brady & Alvarez, 2011），但实验材料多为实验室设计的简单材料。但在现实情境中，人的视觉加工往往受到多种环境因素的制约，比如物体之间的遮挡。而当遮挡条件存在时，物体的属性，如大小会存在一定的不确定性。此时，人的视觉系统是仅仅加工物体的可见部分的属性，还是推测出物体被遮挡部分的属性，并与可见部分作为一个整体进行加工，这是一个需要重视的问题。

因此，本研究拟提出以下问题：在存在遮挡的条件下，人是否能够正确推测被遮挡物体的大小，并且较好地估计整组物体的平均大小。

2 课题意义

2.1 理论意义

尽管近10年来，对于统计知觉有不少心理学研究，也取得了不少进展，但还有很多问题尚不明确。尤其是进行实验室实验时，很少考虑人在现实中面临的环境因素。而本研究准备研究的遮挡条件正是对统计知觉现象研究的一种补充。因为以往的研究中，被试需要进行判断的事物的统计属性，如圆的平均大小，箭头的平均朝向，甚至是人脸的平均表情，

都是完全确定的，而遮挡条件存在时，在人的眼中，有些客体的属性就具有了不确定性，如一个被遮住的三角形，其大小有多种可能，那么人们在判断这些图形的平均大小时，有没有考虑到这些不确定性呢？有没有对其属性进行推测？推测依据什么呢？这些都是有理论意义的问题，对人们理解统计知觉这一过程有重要的意义。

2.2 实践意义

实践意义上，正因为不确定性是一个在实际场景中经常遇到的问题，所以本研究的成果也可以应用于实际场景中。由于本研究的问题是由现实中的现象引出的，因而也可能在现实中得到应用。本研究的结论可以应用到界面设计，视频游戏等需要对一些客体的平均大小进行知觉的任务中。而如何设计客体之间的遮挡关系也可能是工效学应用中的一个潜在的问题。

3 可行性分析

可行性分析主要从以下三个方面展开，分别是理论背景，操作方法和测量工具。

就理论背景而言，本研究拥有比较坚实的文献基础，对统计知觉现象的研究从2001年开始，一直是心理学中的一个较为热门的领域，十多年的研究积累了大量的研究成果，人们对于统计知觉现象的机制也存在若干假说和证据。本人在设计本研究的方案之前已经阅读了大量关于统计知觉的文献资料，对统计知觉的理论有了较清晰的认识和理解，对其研究和实验方法也较为熟悉，为进行这项研究打下了理论基础。

就操作方法而言，通过进行心理学实验室进行心理物理学实验的方法进行研究，该研究能在浙江大学心理学实验室进行实验并获取数据。而实验法具有精确、定量、重复性好、统计信效度高的优点，能够较好地支持该研究得出合理的结果。

就实验工具而言，本人打算使用matlab软件编制实验程序，并以实验程序中记录的反应时、正确率等指标进行统计分析并得出实验结论。这些工具与指标简单、客观、有效，而且非常容易实施。本人具有一定的编程能力和经验，可以根据研究的目的和设计方案编制出符合要求的实验程序为研究所用。本人也熟悉SPSS等统计分析软件的使用，以便于分析实验数据。

综合以上所说的三点，使得研究能够快速有效地进行。

4 调研报告

近年有大量研究探讨了统计知觉的普遍性和加工机制。最具里程碑意义的是Ariely的一项研究，发现被试虽然不能很好地记忆圆组中单个圆的大小，但可以相当精确地估计

一组圆的平均大小(Ariely,2001)。由此,统计知觉开始受到研究者们的关注。随后越来越多的实验证明了统计知觉在视觉中普遍存在:例如平均大小(Ariely,2001),方向(Parkes, Lund, Angelucci, Solomon,& Morgan,2001),空间格局(Alvareza & Olivab,2009),甚至是一组人脸的平均表情(Haberman & Whitney, 2007)等都能得到准确估计。有研究发现,给被试呈现一组动态变化的客体或一个随时间变化的客体,被试都能准确地估计其平均大小,且其精度与对静态客体平均大小的判断没有差别(Albrecht & Scholl, 2010)。这为统计知觉如何适用于动态变化的现实世界提出了可能。此外,还有研究通过给被试同时呈现听觉和视觉信息,发现被试对两者均值都能进行准确估计,证明统计知觉同样存在于听觉通道,存在跨通道的一致性(Albrecht, Scholl, & Chun, 2012)。

在统计知觉的加工机制方面,有研究发现,被试在50ms、100ms、200ms和1000ms四种呈现时间下,对12个圆的平均大小估计的准确性没有显著差异。且在这四种呈现条件下对两组圆大小比较的绩效与对两个圆大小比较的绩效相似,表明统计知觉的加工可能不需要集中注意的参与,且是一个自动的平行加工过程(Chong & Treisman, 2003)。Brady等人研究工作记忆中存储的统计知觉信息发现,当用颜色对同时呈现的一堆圆进行分组时,被试记忆的个体圆大小会偏向所有圆的平均大小或同色圆平均大小,进一步说明了统计知觉是不受控制的自动加工过程(Brady & Alvarez, 2011)。在上述研究基础上,张军伟等人对统计知觉的加工机制做了进一步研究。该研究根据格式塔邻近性原则将圆分成个数和平均大小都不相同的两组,发现被试对总体平均值的估计会介于真实的总体平均值与二次平均(即对两组圆的均值进行平均)的总体平均值之间。

5 研究方案

5.1研究内容和假设

研究一:不同确定性的遮挡条件下的大小统计知觉

本研究的目的是探讨遮挡条件下,人能否对于大小不确定的物体进行正确的加工,进而正确知觉出整组物体的平均大小。

主要假设:

(1)在遮挡条件下,人能够对被遮挡物体的大小进行推测,借此知觉出一组物体的平均大小;

(2)对被遮挡的物体大小的推测主要借助于可见物体的信息。

研究二:遮挡中可见物体对大小统计知觉的影响

本研究的目的是沿着研究一的逻辑。若人们能对被遮挡的大小不确定的物体进行正

确的估计，那么这个过程是否借助了其他可见物体的信息。本研究拟使用贝叶斯方法对此进行分析。

主要假设：

遮挡条件下，可见物体的大小被作为被遮挡物体大小估计的先验（prior）信息。

5.2 实施计划

研究计划主要分成四个阶段进行实施。第一，阅读文献，确定实验方案，编制实验程序。第二，选择被试者，并按方案进行实验。第三，数据处理，采用SPSS20.0进行描述统计，多元方差检验及独立样本t检验等统计检验方法，以确认实验结论。第四，结合前人文献和获得的数据结果进行讨论、推测，总结并得出结论。

6 预期结论

结论一：在遮挡条件下，人能够对被遮挡物体的大小进行推测，借此知觉出一组物体的平均大小；

结论二：对被遮挡的物体大小的推测主要借助于可见物体的信息，可见物体的信息作为被遮挡物体大小估计的先验信息。

参考文献：

张军伟.(2014). 统计知觉的层次计算模型. 硕士学位论文. 浙江大学.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*, 21(4), 560-567.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, 74(810-815).

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3).

Alvareza, G. A., & Olivab, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of the Sciences*, 106(18), 7345-7530.

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, 12(2), 157-162.

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary Statistics of Size: Fixed Processing Capacity for Multiple Ensembles but Unlimited Processing Capacity for Single Ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1440-1449.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751–R753.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual Working Memory Modulates Rapid Eye Movements to Simple Onset Targets. *Psychological Science*, 24(5), 790-769.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the Function of Visual Short-Term Memory: Transsaccadic Memory, Object Correspondence, and Gaze Correction. *Journal of Experimental Psychology*, 137(1), 163-181.

Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, 18(860-869).

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

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, 64(7), 1055-1067.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739-744.

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, 26-35.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and*

Performance, 27(1), 92.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48.

三、文献翻译

用一个总体表征多个客体能增强视觉认知

George A. Alvarez

摘要：人类的视觉系统一次只能精确表征少数几个客体。那我们如何应对这种严苛的容量限制？一种可能是利用选择性注意来加工最重要的信息。另一种补充的策略是将一组客体作为整体加工（例如加工客体的平均大小）。最近的研究已经表明，视觉系统能在多种特征层面上精确计算客体的总体表征，而当下的研究旨在确定这些表征是如何计算、为什么要计算以及在大脑的哪个位置编码的。总体表征能在很多方面增强视觉认知，而使总体编码成为了一种应对多种视觉加工局限的重要机制。

总体表征的优势

不像实验室实验中没有可靠模式的人工显示材料，真实世界是高度结构化的以及可预测的。例如，在客体层面，人的视野中往往有类似客体的集合——人群里的人脸、灌木上的浆果。在最初级的特征层面上，自然图像在对比度、光强分布、颜色分布、反射光谱和空间结构上是高度规律的。有结构的地方就会有冗余信息，而有冗余信息就提供了形成压缩的、有效率地表征信息的机会。一种利用这种结构和冗余性质的方法是用更高层次的描述来表征这些客体或特征，即将这些客体以总体表征，而不是以个体表征。

总体表征就是由个体的测量指标进行计算，要么进行压缩，要么在时间或空间维度上进行合并。举例来说，任何概括性统计描述（如平均值）是一种总体表征，因为它把各个个体的测量值压缩成了一个值。人类在计算平均值上表现相当突出，包括平均大小、平均亮度、平均朝向和一堆客体的平均位置；还有一群人脸的平均表情、性别和一致性；还有一堆短暂呈现的数的平均值。这些都是描述一组客体的集中趋势的。其他描述一个集合的统计指标，比如方差，偏度和峰度，也同样是总体表征，尽管对这些统计指标的最近研究较少。最后，总体表征的概念可以由一级的概括性统计描述，引申到更高等级的概括性统计描述。

总体表征曾在文献中以多种命名得到研究，包括“全局特征”、“整体”或“构型”特征、“集”或者“统计性质”、“统计概括”。这些所有的词条都包含着把多个测量值整合到一个更高水平的描述的概念。而“总体表征”这个词在这里用作一个涵盖性的术语，以包含这些不同的说法。尽管暂时还没有一个关于总体表征的统一模型，最近的关于总体表征的研究由以下的一个简单的原则统一：用一个总体来表征多个客体能增强视觉认知

（的能力）。

平均的作用

怎样计算出总体表征以打破视觉系统的容量限制呢？答案就在于平均的力量：带噪音的测量值的平均值往往比单个的测量值本身要准确得多。例如，尽管进行了四舍五入，一个人能够以毫秒的精度来精确测量反应时。同样的原则也适用于“集体智慧”效应，在人们猜一头牛的重量时，往往所有人的反应的平均值更接近正确答案。这些优势的产生是由于当测量值被平均时，单个测量值中的随机误差将会倾向于互相抵消。那么，平均的优势要取决于单个测量的噪音之间的相关程度（相关程度越低，优势越大）以及待平均的测量值的数量（数量越多，优势越大）。平均操作的优势还有，可以在数学上标准化，可以做出可检验的假设等。

如果人类的视觉系统有能力进行平均操作，那么观察者应该能够比判断单个个体大小更为准确地判断整体的平均大小。这就是 Dan Ariely 在他的关于人类是否能知觉到一系列客体的平均大小的研究中所证明的一点。Ariely 的研究表明被试可以以很高的准确度估计一组客体的平均大小，尽管他们并不能报告单个客体的大小。

这种平均操作提供了一种对抗注意过程的局限的可能的机制。注意似乎是一种流动的灵活的资源：我们可以对单个物体给予足够的注意并以高精度来表征它，我们也可以将我们的注意以较低的精度分散在多个物体上。总的来说，注意焦点外的客体被以更低的清晰度和对比度和更弱的高频反应知觉到。基本上视野中的所有客体都被根据其获得注意的多少，用不同的精确度进行了表征。某些情况下，注意焦点之外的客体被表征得很弱以至于我们不能得到它们的任何有用信息。然而，现象表明这些不精确的信息可以被用来精确描述一组客体的性质。

图 1 阐明了注意如何影响总体表征的可能性。在注意的焦点内（红色光束表示），单个客体会被更加精确地表征。而如果按照关于平均的优势的期望，这些客体的平均会则甚至被表征得更精确。对于注意焦点之外的客体，我们假设它必须被注意到一定的程度才能被知觉到。例如，非注意视盲研究的结果展现了：没有注意的时候，视觉信息几乎没有被意识参与表征。这些研究以让被试彻底从测试客体上移走注意为目的，在有些情况下被试也完全把注意焦点外的信息忽略了。然而，当被试知道他们会被问到关于注意焦点外的客体的问题时，很有可能他们还是会把注意扩散到这些客体上的。图 1 暗示了一种含有多注意焦点，以及分布在注意焦点之外的扩散式注意的平行系统的存在。然而，一个类似的结果可以被简单的注意聚光灯模型解释，在这个模型里，聚光灯（注意）照亮某些位置的时间会长一些。不管哪种解释中，扩散式注意会导致对单个物体的极度不精确的表征，而平均能形成相对精确的总体表征，即使只有三个不精确的测量

值。如果一个足够大的样本被平均到一起，那么注意焦点外的客体的总体表征就与注意焦点内的客体的总体表征差不多精确了。

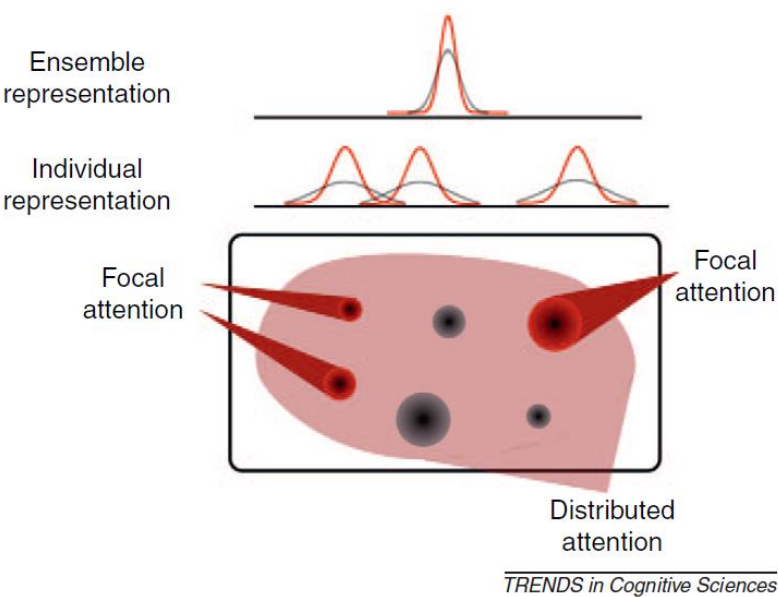


图 1 注意与总体表征

总体表征的机制

尽管人类观察者能够准确地表征总体特征，但更多的问题聚焦在这些总体表征是“怎样”形成的上，包括（1）是先进行对个体的表征再将其合并成总体表征，还是直接进行总体表征？（2）如果形成了对于单个客体的表征，那么当总体表征形成时，单个客体的表征就被丢弃了吗？（3）当计算平均值使时有多少单个客体被取样了？是一少部分还是全部？（4）所有客体被计算成平均值时权重是一样的吗？

计算总体表征的其他优势

本文已经聚焦于总体表征的一个首要优势：将不精确的单个个体表征构建成为准确的一组客体的总体表征。然而，计算总体表征将会带来更多有关的优势。

（一）作为统计推理和异常值检测的基础

统计推理，包括估计概率分布的参数（均值，标准差，范围，形状等），在参数估计中确定置信区间，对客体进行分类等。一个特殊的分类案例就是异常值检测，总体表征是一个达成目标的理想方法。例如，如果一个集合服从一个特定维度的特定分布，均值为 20，标准差为 3，那么一个值为 30 的个体就不太可能属于这个总体了。总体表征能够给一些值打上异常值的标签，甚至将它分到其他组里。

（二）建立一个“大致”表征并引导注意

尽管在局部是不精确的，但总体表征能够提供对注意焦点外的更高水平的特征和规则的精确表征。这些特征和规则对人所看到的场景是具有高度的诊断性的，因此总体表

征对于确定人所处的环境是很有用的。经验上看，人们似乎能够学习到这些总体表征与视野中客体的位置之间的联系。举例来说，人们似乎会利用全局的环境信息来把自己的注意部署到可能存在视觉搜索的目标的区域里去。因此，视觉系统进行总体表征的另一作用是形成“大致”表征，引导注意的部署，使其分配在相对重要的区域里去。

（三）对一个场景建立层级表征

对一个场景建立层级表征有很多计算上的优势。特别是，通过多级表征对信息进行整合，有可能会提高低水平表征的精确度。在被要求在工作记忆中保持一个场景时，人似乎会自动地构建这种表征。例如，当回忆一幅图像中的单个物体大小时，回忆的大小会偏向整体的平均大小。这个结果可以被一个多层的贝叶斯模型较好的拟合。

结论

在视觉认知的传统研究中，使用随机的、非结构的显示材料，能够发现视知觉和记忆在表征单个客体时所受的限制。然而，在现实世界图像中，客体之间存在大量的结构和联系，这给用一个总体来表征多个客体提供了可能。总体表征能够概括多个客体的特征，削弱单个客体测量的噪音，并打破注意的限制。目前关于总体表征的研究主要集中在总体表征的机制及其与个体表征的关系。总体表征具有比单个表征更精确、利于进行统计推断、引导注意的分配、为层级表征建立基础等优势。

四、外文原文

Review



Representing multiple objects as an ensemble enhances visual cognition

George A. Alvarez

Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, William James Hall, Room 760, Cambridge, MA 02138, USA

The visual system can only accurately represent a handful of objects at once. How do we cope with this severe capacity limitation? One possibility is to use selective attention to process only the most relevant incoming information. A complementary strategy is to represent sets of objects as a group or ensemble (e.g. represent the average size of items). Recent studies have established that the visual system computes accurate ensemble representations across a variety of feature domains and current research aims to determine how these representations are computed, why they are computed and where they are coded in the brain. Ensemble representations enhance visual cognition in many ways, making ensemble coding a crucial mechanism for coping with the limitations on visual processing.

Benefits of ensemble representation

Unlike artificial displays used in laboratory experiments, where there is no reliable pattern across individual items, the real world is highly structured and predictable [1,2]. For instance, at the object level, the visual field often consists of collections of similar objects – faces in a crowd, berries on a bush. At a more primitive feature level, natural images are highly regular in terms of their contrast and intensity distributions [3,4], color distributions [5–8], reflectance spectra [9,10] and spatial structure [2,11–14]. Where there is structure, there is redundancy, and where there is redundancy, there is an opportunity to form a compressed and efficient representation of information [15–17]. One way to capitalize on this structure and redundancy is to represent collections of objects or features at a higher level of description, describing distributions or sets of objects as an ensemble rather than as individuals.

An ensemble representation is any representation that is computed from multiple individual measurements, either by collapsing across them or by combining them across space and/or time. For instance, any summary statistic (e.g. the mean) is an ensemble representation because it collapses across individual measurements to provide a single description of the set. People are remarkably accurate at computing averages, including the mean size [18,19], brightness [20], orientation [18,21,22] and location of a collection of objects [23]; the average emotion [24], gender [24] and identity [25] of faces in a crowd; and the average number for a set of symbolically presented numbers [26,27]. These are all measures of central tendency for

a collection of objects. Other statistics that describe a set, such as variance [28], skew and kurtosis, are also ensemble representations, although the ability to compute and represent these statistics has been the focus of less attention in recent research (but see [29,30] for reviews on earlier research). Finally, the concept of ensemble representations can be extended beyond first-order summary statistics, to include higher-order summary statistics [31–33].

Ensemble representations have been explored under various names in the literature, including ‘global features’ [32,34,35], ‘(w)holistic’ or ‘configural’ features [36–38], ‘sets’ [18,39] and ‘statistical properties’ or ‘statistical summaries’ [19,40]. Each of these terms shares the notion that multiple measurements are combined to give rise to a higher level description. The term ‘ensemble representation’ is used here as an umbrella term encompassing these different ideas. Although there is, as yet, no unifying model of ensemble representation across these domains, recent research on ensemble representation is unified by a common principle: representing multiple objects as an ensemble enhances visual cognition.

The power of averaging

How can computing ensemble representations help overcome the severe capacity limitations of our visual system? The answer lies in the power of averaging: simply put, the average of multiple noisy measurements can be much more precise than the individual measurements themselves. For instance, one can measure reaction time with millisecond precision even when rounding reaction times to the nearest 100 ms (Box 1). The same principle is at play in the ‘wisdom of crowds’ effect, in which people guess the weight of an ox and the average response is closer to the correct answer than are the individual guesses on average [41]. These benefits arise because, when measurements are averaged, random error in one individual measurement will tend to cancel out uncorrelated random error in another measurement. Thus, the benefits of averaging depend on the extent to which the noise in individual measurements is correlated (less correlated, more benefit) and the number of individual measurements averaged (more measurements, more benefit). The benefit of averaging can be formalized mathematically, given certain assumptions regarding the noise in the individual measurements (Figure 1).

If the human visual system is capable of averaging, then observers should be able to judge the average size of a set more accurately than they can judge the individuals in the set. This is exactly what was demonstrated by Dan Ariely’s

Corresponding author: Alvarez, G.A. (alvarez@wjh.harvard.edu).

Review

Trends in Cognitive Sciences March 2011, Vol. 15, No. 3

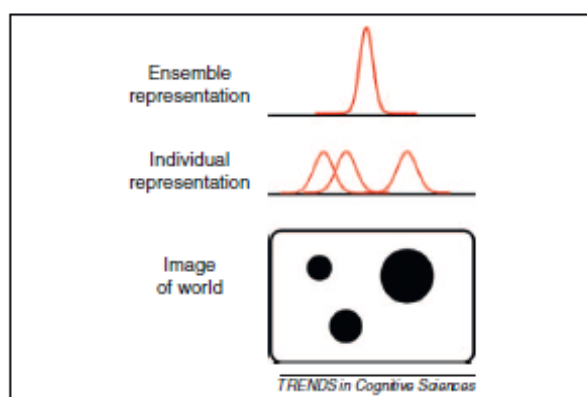


Figure 1. Gaining precision at a higher level of abstraction. By taking individual measurements and averaging them, it is possible to extract a higher-level ensemble representation. If error is independent between the individual representations, then the ensemble average will be more precisely represented than the individuals in the set. This benefit can be quantified after making certain assumptions. For instance, if each individual were represented with the same degree of independent, Gaussian noise (standard deviation = σ), then the average of these individual estimates would have less noise, with a standard deviation equal to σ/\sqrt{n} , where n is the number of individual measurements. The process is depicted for the representation of object size, but the logic holds for any feature dimension.

attention, plus diffuse attention spread over items outside the foci of attention. However, a similar result could be modeled with a single spotlight of attention that spends more time in some locations than others. Either way this diffuse attention results in extremely imprecise representations of the individual items, and yet averaging even just three imprecise measurements results in a fairly precise representation of the ensemble. If a large enough sample of items is averaged together, then the ensemble representation for items outside the focus of attention can be nearly as accurate as the ensemble representation for items inside the focus of attention.

The mechanisms of averaging

Although there is general agreement that human observers can accurately represent ensemble features, many questions remain regarding 'how' these ensemble representations are computed, including (i) Are individual representations computed and then combined to form an ensemble representation, or are ensemble representations somehow computed without computing individuals? (ii) If individual representations are computed, are they discarded once the ensemble has been computed? (iii) How many individual items are sampled and included in the calculation of the mean? Is it just a few or could it be all of them? (iv) Do all items contribute to the mean equally?

Are ensembles built up from representations of individuals?

Ariely [18] proposed that the visual system performs a type of compression, by creating an ensemble representation and then discarding individual representations. Some have interpreted this proposal to mean that the ensemble representation is computed without first directly computing individual measurements. For instance, it is possible that there is a 'total activation map' and a 'number map'

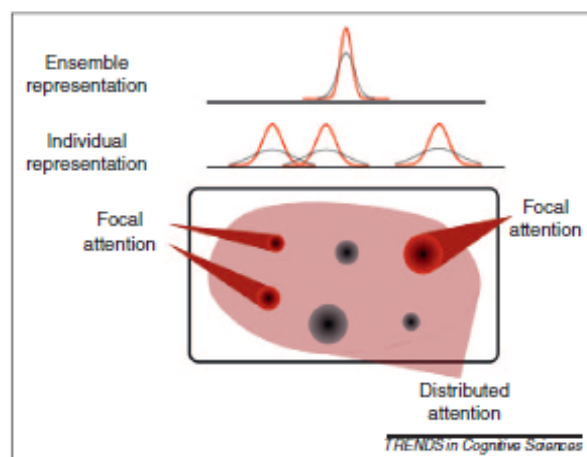


Figure 2. Effect of attention on the fidelity of ensemble representations. Two sets of items are depicted: one set inside the focus of attention (red beams) and one set diffusely attended outside the focus of attention (pink region). For illustrative purposes, both sets are composed of identical individuals, and thus both sets have the same individual and mean representations. For items inside the focus of attention, individual representations will be relatively precise (red curves). The ensemble representation of the items inside the focus of attention will be even more precise, owing to the benefits of averaging. For items outside the focus of attention which are diffusely attended, the individual representations will be very imprecise (gray curves). However, the benefits of averaging are so great that the ensemble representation will be fairly precise, even when a relatively small number of individual representations are averaged (just three in this example).

and that mean size is computed by taking the total activation and dividing it by the number of items [52]. However, Ariely's use of the term 'discard' suggests that his intended meaning was that the individual properties are computed, combined and then discarded. This type of averaging model has been supported by research on the computation of mean orientation [21]. Addressing this question empirically is a challenge because it is possible to compute accurate ensemble representations even from very imprecise individual measurements. Consequently, a poor representation of individual items cannot be used as evidence for mean computation without computing individuals – unless the mean can be shown to be represented more accurately than expected based on the number and fidelity of individual items represented.

Are individual representations discarded?

How do we explain such poor performance when observers are required to report the properties of individual members of a set? One possibility is that these properties are computed and then discarded. An important alternative possibility is that the individual representations are not discarded, but are simply so noisy and inaccurate that observers cannot consistently identify individuals from the set owing to this high level of noise. Alvarez and Oliva found support for this possibility by modeling their results [23], consistently finding that the accuracy of ensemble judgments is perfectly predicted from the accuracy of individual judgments – even when individuals appear to be judged with near chance accuracy. This alternative possibility fits with a framework in which the representation of an image is hierarchical, retaining information at multiple levels of abstraction [35,53].

Review

Trends in Cognitive Sciences March 2011, Vol. 15, No. 3

Although it would be interesting if ensemble representations could enable rapid outlier detection, this finding is not necessary to support the idea that ensemble representations play an important role in classifying and grouping items. For instance, a face with a unique facial expression does not pop-out in a visual search task [88]. However, recent research shows that an outlier face is given reduced weight in the ensemble representation of a group of faces [58], even though observers often fail to perceive the outlier. This finding is consistent with the possibility that the ensemble representation enables labeling of items, but could also indicate that the ensemble computation gives outliers lower weight without attaching a classification label. The role of ensemble representations in determining set membership has not yet been extensively studied, and research in this area can potentially bridge the gap between study on ensemble representation, statistical inference and perceptual grouping.

Building a 'gist' representation that can guide the focus of attention

As detailed in previous sections, the power of averaging makes it possible to combine imprecise local measurements to yield a relatively precise representation of the ensemble (Figure 1). Moreover, it is possible to combine individual measurements to describe spatial patterns of information (Figure 5). A primary benefit of computing either type of ensemble representation is to provide a precise and accurate representation of the 'gist' of information outside the focus of attention. Without focused attention, our representations of visual information are highly imprecise [23]. If we were to simply discard or ignore these noisy representations, our conscious visual experience would be limited to only those items currently within the focus of attention. Indeed, some have argued that this is the nature of conscious visual experience [89,90]. In such a system, attention would be 'flying blind', without access to any information about what location or region to focus on next.

Although locally imprecise, ensemble representations provide an accurate representation of higher-level patterns and regularities outside the focus of attention [23,31]. These patterns and regularities are highly diagnostic of the type of scene one is viewing [14], and therefore they are useful for determining which environment one is currently located within. Over experience, observers appear to learn associations between these ensemble representations and the location of objects in the visual field. For instance, observers appear to use global contextual information to guide the deployment of attention to locations likely to contain the target of a visual search task [33,91–93]. Thus, rather than flying blind, the visual system can compute ensemble representations, providing a sense of the gist of information outside the focus of attention, and guiding the deployment of attention to important regions of a scene.

In terms of forming a complete representation of a scene, gist representation and outlier detection probably work in tandem. For instance, when holding a scene in working memory, observers appear to encode the gist of the scene plus individual items that cannot be incorporated into the summary for the rest of the scene (i.e. outliers) [78].

Benefits of building a hierarchical representation of a scene

There are distinct computational advantages to building a hierarchical representation of a scene. In particular, by integrating information across levels of representation, it is possible to increase the accuracy of lower-level representations. It appears that observers automatically construct this type of representation when asked to hold a scene in working memory [77,78]. For instance, when recalling the size of an individual item from a display, the remembered size was biased towards the mean size of the set of items in the same color, and towards the overall mean size of all items in the display [77]. These results were well captured by a Bayesian model in which observers integrate information at multiple levels of abstraction to inform their judgment about the size of the tested item.

Concluding remarks

Traditional research on visual cognition has typically assessed the limits of visual perception and memory for individual objects, often using random and unstructured displays. However, there is a great deal of structure and redundancy in real-world images, presenting an opportunity to represent groups of objects as an ensemble. Because ensemble representations summarize the properties of a group, they are necessarily spatially and temporally imprecise. Nevertheless, such ensemble representations confer several important benefits. Much of the previous research on ensemble representation has focused on the fact that the human visual system is capable of computing accurate ensemble representations. However, the field is moving towards a focus on investigating the mechanisms that enable ensemble coding, the nature of the ensemble representation, the utility of ensemble representations and the neural mechanisms underlying ensemble coding. This future research promises to uncover important new properties of the representations underlying visual cognition and to further demonstrate how representing ensembles enhances visual cognition.

Acknowledgments

For helpful conversation and/or comments on earlier drafts, I thank Talia Konkle, Jason Haberman and Jordan Suchow. G.A.A. was supported by the National Science Foundation (Career Award BCS-0953730).

References

- 1 Kersten, D. (1987) Predictability and redundancy of natural images. *J. Opt. Soc. Am. A* 4, 2395–2400
- 2 Field, D.J. (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394
- 3 Brady, N. and Field, D.J. (2000) Local contrast in natural images: normalisation and coding efficiency. *Perception* 29, 1041–1055
- 4 Frazor, R.A. and Geisler, W.S. (2006) Local luminance and contrast in natural images. *Vis. Res.* 46, 1585–1598
- 5 Webster, M.A. and Mollon, J.D. (1997) Adaptation and the color statistics of natural images. *Vis. Res.* 37, 3283–3298
- 6 Hyvärinen, A. and Hoyer, P.O. (2000) Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* 12, 1705–1720
- 7 Judd, D.B. et al. (1964) Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am. A* 54, 1031–1040

Review

Trends in Cognitive Sciences March 2011, Vol. 15, No. 3

- 8 Long, F. et al. (2006) Spectral statistics in natural scenes predict hue, saturation, and brightness. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6013-6018
- 9 Maloney, L.T. (1986) Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A* 3, 1673-1683
- 10 Maloney, L.T. and Wandell, B.A. (1986) Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A* 3, 29-33
- 11 Field, D.J. (1989) What the statistics of natural images tell us about visual coding. *SPIE: Hum. Vis. Vis. Process. Dig.* Display 1077, 269-276
- 12 Burton, G.J. and Moorehead, I.R. (1987) Color and spatial structure in natural scenes. *Appl. Opt.* 26, 157-170
- 13 Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167-192
- 14 Torralba, A. and Oliva, A. (2003) Statistics of natural image categories. *Network* 14, 391-412
- 15 Huffman, D.A. (1952) A method for construction of minimum redundancy codes. *Proc. IRE* 40, 1098-1101
- 16 Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*, The University of Illinois Press
- 17 Atick, J.J. (1992) Could information theory provide an ecological theory of sensory processing? *Network: Comput. Neural Syst.* 3, 213-251
- 18 Ariely, D. (2001) Seeing sets: representation by statistical properties. *Psychol. Sci.* 12, 157-162
- 19 Chong, S.C. and Treisman, A. (2003) Representation of statistical properties. *Vis. Res.* 43, 393-404
- 20 Bauer, B. (2009) Does Steven's power law for brightness extend to perceptual brightness averaging? *Psychol. Res.* 59, 171-186
- 21 Parkes, L. et al. (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4, 739-744
- 22 Dakin, S.C. and Watt, R.J. (1997) The computation of orientation statistics from visual texture. *Vis. Res.* 37, 3181-3192
- 23 Alvarez, G.A. and Oliva, A. (2008) The representation of simple ensemble visual features outside the focus of attention. *Psychol. Sci.* 19, 392-398
- 24 Haberman, J. and Whitney, D. (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751-R753
- 25 de Fockert, J. and Wolfenstein, C. (2009) Rapid extraction of mean identity from sets of faces. *Q. J. Exp. Psychol. (Colchester)* 62, 1716-1722
- 26 Spencer, J. (1961) Estimating averages. *Ergonomics* 4, 317-328
- 27 Smith, A.R. and Price, P.C. (2010) Sample size bias in the estimation of means. *Psychon. Bull. Rev.* 17, 499-503
- 28 Morgan, M. et al. (2008) A 'dipper' function for texture discrimination based on orientation variance. *J. Vis.* 8, 1-8
- 29 Peterson, C.R. and Beach, L.R. (1967) Man as an intuitive statistician. *Psychol. Bull.* 68, 29-46
- 30 Pollard, P. (1984) Intuitive judgments of proportions, means, and variances: a review. *Curr. Psychol.* 3, 5-18
- 31 Alvarez, G.A. and Oliva, A. (2009) Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7345-7350
- 32 Oliva, A. and Torralba, A. (2006) Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* 155, 23-36
- 33 Oliva, A. and Torralba, A. (2007) The role of context in object recognition. *Trends Cogn. Sci.* 11, 520-527
- 34 Oliva, A. and Torralba, A. (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 145-175
- 35 Navon, D. (1977) Forest before trees: the precedence of global features in visual perception. *Cognit. Psychol.* 9, 353-383
- 36 Kimchi, R. (1992) Primacy of wholistic processing and global/local paradigm: a critical review. *Psychol. Bull.* 112, 24-38
- 37 Thompson, P. (1980) Margaret Thatcher: a new illusion. *Perception* 9, 483-484
- 38 Young, A.W. et al. (1987) Configurational information in face perception. *Perception* 16, 747-759
- 39 Halberda, J. et al. (2006) Multiple spatially overlapping sets can be enumerated in parallel. *Psychol. Sci.* 17, 572-576
- 40 Albrecht, A.R. and Scholl, B.J. (2010) Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychol. Sci.* 21, 560-567
- 41 Galton, F. (1907) Vox populi. *Nature* 75, 450-451
- 42 Palmer, J. (1990) Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 332-350
- 43 Alvarez, G.A. and Franconeri, S.L. (2007) How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *J. Vis.* 7, 1-10
- 44 Franconeri, S.L. et al. (2007) How many locations can be selected at once? *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1003-1012
- 45 Titchener, E.B. (1908) *Lectures on the Elementary Psychology of Feeling and Attention*, Macmillan
- 46 Carrasco, M. et al. (2004) Attention alters appearance. *Nat. Neurosci.* 7, 308-313
- 47 Carrasco, M. et al. (2002) Covert attention increases spatial resolution with or without masks: support for signal enhancement. *J. Vis.* 2, 467-479
- 48 Yeshurun, Y. and Carrasco, M. (1998) Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396, 72-75
- 49 Mack, A. and Rock, I. (1998) *Inattention Blindness*, The MIT Press
- 50 Neisser, U. and Becklen, R. (1975) Selective looking: attending to visually specified events. *Cognit. Psychol.* 7, 480-494
- 51 Most, S.B. et al. (2005) What you see is what you set: sustained inattention blindness and the capture of awareness. *Psychol. Rev.* 112, 217-242
- 52 Setic, M. et al. (2007) Modelling the statistical processing of visual information. *Neurocomputing* 70, 1808-1812
- 53 Kinchla, R.A. and Wolfe, J.M. (1979) The order of visual processing: "Top-down", "bottom-up", or "middle-out". *Percept. Psychophys.* 25, 225-231
- 54 Myczek, K. and Simons, D.J. (2006) Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Percept. Psychophys.* 70, 772-788
- 55 Chong, S.C. and Treisman, A. (2005) Attentional spread in the statistical processing of visual displays. *Percept. Psychophys.* 67, 1-13
- 56 Haberman, J. et al. (2009) Averaging facial expression overtime. *J. Vis.* 9, 1-13
- 57 Chong, S.C. et al. (2008) Statistical processing: not so implausible after all. *Percept. Psychophys.* 70, 1327-1334
- 58 Haberman, J. and Whitney, D. (2010) The visual system discounts emotional deviants when extracting average expression. *Atten. Percept. Psychophys.* 72, 1825-1838
- 59 Kersten, D. and Yuille, A. (2003) Bayesian models of object perception. *Curr. Opin. Neurobiol.* 13, 150-158
- 60 Vul, E. and Pashler, H. (2008) Measuring the crowd within: probabilistic representations within individuals. *Psychol. Sci.* 19, 645-647
- 61 Vul, E. and Rich, A.N. (2010) Independent sampling of features enables conscious perception of bound objects. *Psychol. Sci.* 21, 1168-1175
- 62 Mareschal, I. et al. (2010) Attentional modulation of crowding. *Vis. Res.* 50, 805-809
- 63 de Fockert, J.W. and Marchant, A.P. (2008) Attention modulates set representation by statistical properties. *Percept. Psychophys.* 70, 789-794
- 64 Dehaene, S. et al. (1998) Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355-361
- 65 Feigenson, L. et al. (2004) Core systems of number. *Trends Cogn. Sci.* 8, 307-314
- 66 Whalen, J. et al. (1999) Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130-137
- 67 Burr, D. and Ross, J. (2008) A visual sense of number. *Curr. Biol.* 18, 425-428
- 68 Geisler, W.S. et al. (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vis. Res.* 41, 711-724
- 69 Chandler, D.M. and Field, D.J. (2007) Estimates of the information content and dimensionality of natural scenes from proximity distributions. *J. Opt. Soc. Am. A* 24, 922-941
- 70 Barlow, H.B. and Foldiak, P. (1989) Adaptation and decorrelation in the cortex. In *The Computing Neuron* (Durbin, R. et al., eds), pp. 54-72, Addison-Wesley
- 71 Lewicki, M.S. (2002) Efficient coding of natural sounds. *Nat. Neurosci.* 5, 356-363
- 72 Olshausen, B.A. and Field, D.J. (1996) Natural image statistics and efficient coding. *Network* 7, 333-339

Review

Trends in Cognitive Sciences March 2011, Vol. 15, No. 3

- 73 Balas, B. *et al.* (2009) A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9, 13-18
- 74 Bulakowski, P.F. *et al.* Reexamining the possible benefits of visual crowding: dissociating crowding from ensemble percepts. *Atten. Percept. Psychophys.* (in press)
- 75 Piazza, M. and Izard, V. (2009) How humans count: numerosity and the parietal cortex. *Neuroscientist* 15, 261-273
- 76 Cavanagh, P. (2001) Seeing the forest but not the trees. *Nat. Neurosci.* 4, 673-674
- 77 Brady, T.F. and Alvarez, G.A. Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol. Sci.* (in press)
- 78 Brady, T.F. and Tenenbaum, J.B. (2010) Encoding higher-order structure in visual working-memory: a probabilistic model. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Ohlsson, S. and Catrambone, R., eds), pp. 411-416, Cognitive Science
- 79 Olshausen, B.A. and Field, D.J. (2004) Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481-487
- 80 Olshausen, B.A. and Field, D.J. (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37, 3311-3325
- 81 Wilshaw, D.J. *et al.* (1969) Non-holographic associative memory. *Nature (Lond.)* 222, 960-962
- 82 Zetsche, C. (1990) Sparse coding: the link between low level vision and associative memory. In *Parallel Processing in Neural Systems and Computers* (Eckmiller, R. *et al.*, eds), pp. 273-276, Elsevier Science
- 83 Rosenholtz, R. (2000) Significantly different textures: a computational model of pre-attentive texture segmentation. In *Proceedings of the 6th European Conference on Computer Vision* (Vernon, D., ed.), pp. 197-211, Springer-Verlag
- 84 Rosenholtz, R. (1999) A simple saliency model predicts a number of motion popout phenomena. *Vis. Res.* 39, 3157-3163
- 85 Itti, L. and Koch, C. (2001) Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194-203
- 86 Wolfe, J.M. (1994) Guided search 2.0: a revised model of visual search. *Psychon. Bull. Rev.* 1, 202-238
- 87 Itti, L. and Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* 40, 1489-1506
- 88 Nothdurft, H.C. (1993) Faces and facial expressions do not pop out. *Perception* 22, 1287-1298
- 89 Nöe, A. and O'Regan, J.K. (2000) Perception, attention and the grand illusion. *Psyche* 6 (<http://psyche.cs.monash.edu.au/v6/psyche-6-15-noe.html>)
- 90 O'Regan, J.K. (1992) Solving the "real" mysteries of visual perception: the world as an outside memory. *Can. J. Psychol.* 46, 461-488
- 91 Torralba, A. *et al.* (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766-786
- 92 Ehinger, K.A. *et al.* (2009) Modeling search for people in 900 scenes: a combined source model of eye guidance. *Vis. Cogn.* 17, 945-978
- 93 Chun, M.M. (2000) Contextual cueing of visual attention. *Trends Cogn. Sci.* 4, 170-178
- 94 Haberman, J. and Whitney, D. (2009) Seeing the mean: ensemble coding for sets of faces. *Hum. Percept. Perform.* 35, 718-734

毕业论文（设计）文献综述和开题报告考核

一、对文献综述、外文翻译和开题报告评语及成绩评定：

成绩比例	文献综述 占（10%）	开题报告 占（20%）	外文翻译 占（10%）
分 值			

开题报告答辩小组负责人（签名）_____

年 月 日

浙 江 大 学

本 科 生 毕 业 论 文（设计）



题目 遮挡条件下的大小统计知觉

姓名与学号 史博皓 3120100969

指导教师 沈模卫 教授

年级与专业 2012 级 心理学

所在学院 心理与行为科学系

本科生毕业论文（设计）任务书

一、题目：大学生社会性问题解决能力特点及其与人格、情绪的关系

二、指导教师对毕业论文（设计）的进度安排及任务要求：

- 1、2016 年 1 月:查找资料确定所要研究的题目。
- 2、2016 年 1 月 22 日：提交文献综述
- 3、2016 年 2 月-2016 年 4 月:对在校大学生发放社会性问题解决问卷、情绪问卷和大五人格问卷，并收集数据。
- 4、2016 年 3 月 11 日：提交开题报告
- 5、2016 年 3 月 20 日： 开题报告答辩
- 6、2016 年 4 月 12 日：提交文献翻译、文献原文
- 7、2016 年 5 月:进行相关数据的整理和统计分析。根据统计分析结果，撰写论文，修改并定稿。
- 8、2016 年 5 月 12 日：毕业论文全文定稿
- 9、2016 年 5 月 20 日：毕业论文全文交心理系 209 办公室
- 10、2016 年 5 月 27 日：参加毕业论文答辩。

起讫日期 2015 年 12 月 26 日 至 2016 年 5 月 27 日

指导教师（签名）_____职称_____

三、系或研究所审核意见：

负责人（签名）

年 月 日

五、毕业论文正文

目录

摘要： 29

Abstract:..... 32

1. 引言 31

 1.1 统计知觉 31

 1.2 问题的提出 33

 1.3 研究假设及构思 34

 1.4 研究意义 35

2. 实验设计与结果 36

 2.1 实验方法 36

 2.2.实验结果及讨论 38

3. 综合讨论 40

4. 结论和展望 42

参考文献： 43

致谢： 45

遮挡条件下的大小统计知觉

中文摘要

人们可在短时间内估计出大量客体的平均大小，这一现象被称为大小统计知觉。以往关于大小平均大小的研究对象多是离散的客体，相互间不存在重叠；而现实情境中，大量客体一起出现时往往存在重叠、相互遮挡等情况，导致无法确定被遮挡客体的准确大小。本研究考察遮挡条件下，人们是否能准确知觉出一组客体的平均大小。

结论：在遮挡条件下：确定条件下，人能准确估计出客体实际平均大小；不确定条件下，被试对客体平均大小的主观判断集中在可能范围的均值附近。不确定性增大导致被试判断的变异程度增加。

关键词：大小统计知觉 遮挡

Statistical summary representations of size in occlusion condition

Abstract:

People are able to estimate the mean size of a set of objects and this phenomenon is called Statistical summary representations. The target of previous studies about Statistical summary representations is always separate objects without overlapping. But in actual situations, a set of objects may overlap with each other, so it may be not possible to judge the actual sizes of objects being overlapped. The current study wonders if people can percept the mean size of a set of objects in occlusion condition.

Conclusion:

When there is occlusion, people can accurately estimate the actual mean size of a group of objects in determinate condition; people's subject judgement of the mean size of a group of objects is centralize on the average value of the possible range of the size. A reduced nondeterminacy can lead to a reduced variance in people's judgement.

Key words: statistical summary perception, occlusion condition

1 引言

人类的视觉系统一次只能精确表征少数几个客体。那我们如何应对这种严苛的容量限制？一种方式是利用选择性注意来加工最重要的信息。另一种方法则是将一组客体作为整体加工（例如加工客体的平均大小）。最近的研究已经表明，视觉系统能在多种特征层面上精确计算客体的总体表征，总体表征能在很多方面增强视觉认知，而使总体编码成为了一种应对多种视觉加工局限的重要机制。

统计知觉现象的发现为人类如何处理一系列类似客体的统计属性提供了理论基础。而在现实场景中，多个客体之间的位置关系往往是互相重叠、遮挡的，也存在各种各样的导致客体形状大小不确定的情况，这些情况下视觉系统如何抽取客体统计属性尚未有研究涉及。本研究拟在统计知觉任务中加入遮挡的条件，以探究不确定情景对人类统计知觉任务的影响。

1.1 统计知觉

在人的日常生活中，经常需要对大量的事物进行视觉加工和记忆，而人们用于这些加工过程的认知资源是相当有限的，比如说，人的工作记忆容量仅仅是三到四个简单客体。但仅凭这些有限的认知资源，人就能迅速加工并识别出事物的总体属性，比如人群的密度，大量物体的大致大小等等。该能力被称为统计知觉(Ariely, 2001)。统计知觉是视觉系统突破认知资源有限性的有效方式之一。当面对一系列相似客体时，视觉系统通过为成组客体建立整体层面统计表征的形式将信息进行压缩，忽略客体细节特征，从而以较少资源表征大量信息。

统计知觉之所以能够突破人的精确加工的容量限制，是因为：在实验室实验中的人工显示材料中，并没有特定的模式或者结构；而在真实世界是高度结构化的以及可预测的。例如，在客体层面，人的视野中往往有类似客体的集合——比如人群里的多张人脸、灌木上的浆果。在最初级的特征层面上，自然图像在对比度、光强分布、颜色分布、反射光谱和空间结构上是高度规律的。而有结构的地方就会有冗余信息，而有冗余信息就提供了形成压缩的、有效率地表征信息的机会。一种利用这种结构和冗余性质的方法就是用更高层次的描述来表征这些客体或特征，即将这些客体以总体表征，而不是以单个个体表征。统计知觉就是这样的总体表征的一种，将单个客体的特征用更高层次的统计属性来表示，比如平均大小，平均颜色等。

统计知觉的优势就在于增大了表征的精确度。比如平均知觉中，带噪音的测量值的平均值往往比单个的测量值本身要准确得多。例如，尽管进行了四舍五入，在平均之后一个人的反应时的精确度仍是毫秒级的。同样的原则也适用于“集体智慧”效应，在人们猜一头牛的重量时，往往所有人的反应的平均值更接近正确答案。这些优势的产生是由于当测量值被进行平均计算时，单个测量值中的随机误差将会倾向于互相抵消。那么，

平均的优势要取决于单个测量的噪音之间的相关程度（相关程度越低，优势越大）以及待平均的测量值的数量（数量越多，优势越大）。平均操作的优势还有，可以在数学上标准化，可以做出可检验的假设等。

除了将不精确的单个个体表征构建成为准确的一组客体的总体表征之外，统计知觉还有其他的优势：（1）可以作为统计推理和异常值检测的基础。统计推理，包括估计概率分布的参数（均值，标准差，范围，形状等），在参数估计中确定置信区间，对客体进行分类等。一个特殊的分类案例就是异常值检测，而统计知觉就是一个达成目标的理想方法。例如，如果一个总体服从一个的特定分布，均值为 20，标准差为 3，那么一个值为 30 的个体就不太可能属于这个总体了。统计知觉能够给一些值打上异常值的标签，甚至将它归入其他的类别里。例如，在一堆沙子里，我们能够很快地找出不属于其中的石块，就是因为石块不服从我们形成的统计表征中的沙子的分布；（2）建立一个“大致”表征并引导注意。尽管在局部是不精确的，但总体表征能够提供对注意焦点外的客体可证有更精确的表征。这些特征和规则对所看到的场景是具有高度的诊断性的，因此统计知觉对于确定人所处的环境是很有用的。从生活经验中看，人们似乎能够学习到这些总体表征与视野中客体的位置之间的联系。举例来说，人们似乎会利用全局的环境信息来把自己的注意部署到可能存在视觉搜索的目标的区域里去，比如把注意力放在在平均身高更矮的一群人，以找到一个小孩子。（三）对一个场景建立层级表征。建立层级表征有很多计算上的优势。其中最重要的是，通过多级表征对信息进行整合，有可能会提高低水平表征的精确度。在被要求在工作记忆中保持一个场景时，人似乎会自动地构建这种层次表征。例如，当回忆一幅图像中的单个物体大小时，回忆的大小会偏向整体的平均大小。这个结果可以被一个多层的贝叶斯模型较好的拟合。而且结论可以运用到空间位置、空间大小、运动等多个问题中去。

统计知觉同时也提供了一种打破注意过程的局限的可能机制。注意似乎是一种流动而灵活的资源：我们可以对单个物体给予足够的注意并以高精度来表征它，我们也可以将我们的注意以较低的精度分散在多个物体上。总的来说，注意焦点外的客体被以更低的清晰度及对比度知觉到。基本上视野中的所有客体都被根据其获得注意的多少，用不同的精确度进行了表征。某些情况下，注意焦点之外的客体被表征得很弱以至于我们不能得到它们的任何有用信息。然而，现象表明这些不精确的信息也可以被用来精确描述一组客体的性质。

图 1 阐明了总体表征与注意的关系。在注意的焦点内（红色光点表示），单个客体会被更加精确地表征。而如果按照关于平均的效应的期望，这些客体的平均值则甚至会被表征得更精确。对于注意焦点之外的客体，我们假设它必须被注意到一定的程度才能

被知觉到。例如，非注意视盲研究的结果表明：没有得到集中注意的时候，视觉信息几乎没有被有意识地表征。这些研究中，研究者让被试彻底从测试客体上移走注意，在有些情况下被试完全把注意焦点外的信息忽略了。然而，当被试知道他们会被问到关于注意焦点外的客体的问题时，很有可能他们还是会把注意扩散到这些客体上的。图 1 暗示了一种含有多个注意焦点，以及分布在注意焦点之外的扩散式注意的平行系统的存在。然而，一个类似的结果可以被简单的注意聚光灯模型解释，在这个模型里，聚光灯（注意）照亮某些位置的时间会长一些。而不管哪种解释中，扩散式注意会导致对单个物体的极度不精确的表征，而平均知觉能形成相对精确的总体表征，即使只有少数个不精确的测量值。而如果一个足够大的样本被平均到一起，那么注意焦点外的客体的总体表征就与注意焦点内的客体的总体表征差不多精确了。

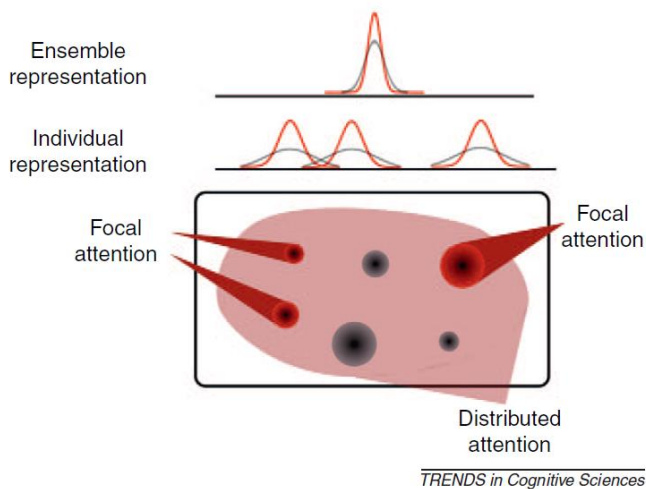


图 1 注意与总体表征

近年大量研究探讨了统计知觉的普遍性和加工机制。最具里程碑意义的是Ariely的一项研究，发现被试虽然不能很好地记忆圆组中单个圆的大小，但可以相当精确地估计一组圆的平均大小(Ariely, 2001)。由此，统计知觉开始受到研究者们的关注。随后越来越多的实验证明了统计知觉在视觉中普遍存在:例如平均大小(Ariely, 2001)，方向(Parkes, Lund, Angelucci, Solomon, & Morgan, 2001)，空间格局(Alvarez & Oliva, 2009)，甚至是一组人脸的平均表情(Haberman & Whitney, 2007)等都能得到准确估计。统计知觉还可以在时间维度上进行。有研究发现，给被试呈现一组动态变化的客体或一个随时间变化的客体，被试都能准确地估计其平均大小，且其精度与对静态客体平均大小的判断没有差别(Albrecht & Scholl, 2010)。这为统计知觉如何适用于动态变化的现实世界提出了可能。此外，还有研究通过给被试同时呈现听觉和视觉信息，发现被试对两者均值都能进行准确估计，证明统计知觉同样存在于听觉通道，存在跨通道的一致性(Albrecht, Scholl, & Chun, 2012)。

在统计知觉的加工机制方面，有研究发现，被试在50ms、100ms、200ms和1000ms

四种呈现时间下，对12个圆的平均大小估计的准确性没有显著差异。且在这四种呈现条件下对两组圆大小比较的绩效与对两个圆大小比较的绩效相似，表明统计知觉的加工可能不需要集中注意的参与，且是一个自动的平行加工过程(Chong & Treisman, 2003)。

Brady等人研究工作记忆中存储的统计知觉信息发现，当用颜色对同时呈现的一堆圆进行分组时，被试记忆的个体圆大小会偏向所有圆的平均大小或同色圆平均大小，进一步说明了统计知觉是不受控制的自动加工过程(Brady & Alvarez, 2011)。

1.2 问题提出

尽管对于统计知觉的机制的研究，已经有了不少进展(如Chong & Treisman, 2003; Brady & Alvarez, 2011)，但上述研究中所采用的实验材料多为实验室设计的简单刺激，即没有重叠的离散客体。但在现实情境中，人的视觉加工往往受到多种环境因素的制约，比如物体之间的遮挡。而当遮挡条件存在时，物体的属性，如大小会存在一定的不确定性，比如遮住一个三角形使其只露出一个角，其大小就不是唯一确定的了，而是可能在一定的范围内变动。此时，人的视觉系统是仅仅加工物体的可见部分的属性，还是推测出物体被遮挡部分的属性，并与可见部分作为一个整体进行加工，这是一个需要重视的问题。

遮挡是产生不确定性的因素之一。现实场景中，一系列可见的客体中可能存在大小不确定的情况，比如模糊、残缺、遮挡、重叠等导致客体形状的不完整。在这样的不确定情况下，个体是否还能准确地推测出完整的客体的形状并估计出所有客体的平均大小呢？例如，对于一棵树上的树叶，普通人能在很短的时间内估计出树叶的平均大小。但实际上，由于树叶之间的遮挡关系，大部分树叶的完整形状及大小是无法被知觉到的。这一现象暗示，存在遮挡关系时，人类可以推测并估计被遮挡客体的完整形状及大小。然而，尚未有实证研究对遮挡条件下的统计知觉进行系统考察。对于不确定条件下的大小统计知觉的探索，需要大量精细的实验提供证据，并非一步就能解决。

遮挡可以创造出不确定性程度不同的条件。存在遮挡时，意味着单个客体的形状和大小是不能被直接知觉的。而当客体的形状不能被直接知觉时，在某些情况下，客体的完整形状仍然是可以被准确地推测的；而某些情况下，客体的完整形状不能被准确地推测，或者说存在一个可能的范围，需要一个概率分布描述。以下的一个例子就描述了这两种情况。

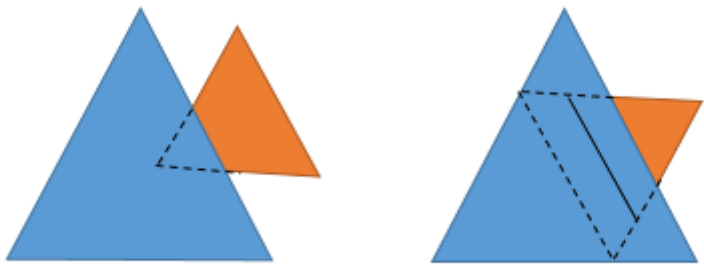


图2 遮挡与不确定性的例子

在图2中，有两对互相遮挡的正三角形，在左边的一对三角形中，被遮挡的橙色三角形虽然形状是不完整的，但是由于其可能的形状只有一种可能（即虚线所画的形状），其大小也是确定的，可以迅速地推测出被遮挡三角形的实际大小，这是一种不确定性较弱的情况，本文中称为确定条件；而在右边的一对三角形中，不仅被遮挡的橙色三角形是不完整的，而且其形状是不能确定的，它的形状可以在虚线所示的范围内变化，而人无法准确推知它的实际形状和大小。但这种情况下，尽管没有一个确定的实际形状，但可以根据最有可能或者估计的理论偏差最小来确定一个理想的大小，这个例子中橙色三角形的理想的大小应当是大小的最大值与最小值的平均，我们在本文中将它称为大小的可能范围的“均值”，由图中的实线表示（数学上很容易证明，选择这一大小时，判断的偏差的数学期望最小）。这是一种不确定性较强的情况，在本文中称为不确定条件。

而当人在统计知觉任务中遇到类似的情况时表现会如何呢？人类可能是仅仅基于客体的可见部分进行大小的平均知觉，也可能是对客体的完整形状及大小进行推测然后再进行平均。不同的不确定性程度的条件，也可能对人类的大小平均知觉有影响。这也是以下的实验要探讨的问题。

1.3 研究构思

基于现实生活中的现象和之前的研究，本研究旨在检验如下问题：

- (1)在遮挡条件下，人是能够对被遮挡物体的大小进行推测，再估计客体的平均大小，还是仅仅借助客体的可见部分估计平均大小；
- (2)在无法准确推测出客体实际大小的不确定情况下，人类的平均大小估计遵循怎样的原则。

而遮挡是各种不确定情况中的一种较为常见和简单的情况，通过对遮挡操作可以实现不同程度的不确定性。本研究将采用遮挡条件，通过不同的遮挡关系设计材料，并以此创造出不同强弱程度不确定性的条件。然后让被试完成大小统计知觉任务，通过条件之间的绩效对比验证实验假设。

在大小统计知觉研究中，有较多的研究采用心理物理学中的恒定刺激法来进行实验。恒定刺激法是心理物理学中最准确、应用最广的方法，通过以相同的次数呈现少数几个恒定的刺激，通过被试觉察到每个刺激的次数来确定阈限。本研究也将选择恒定刺激法，并拟合出被试的反应曲线，获得反应曲线的参数，以及用直线内插法得到被试知觉到的平均大小，再进行条件间的对比，得出遮挡和不确定性条件的影响。

1.4 研究意义

理论意义上，不确定性是统计知觉研究中的一个重要问题，因为人在实际场景中进行大小统计知觉时往往容易遇到带有不确定性的场景。尽管近 10 年来，对于统计知觉有不少心理学研究，也取得了不少进展，但还有很多问题尚不明确。尤其是进行实验室实验时，很少考虑人在现实中面临的环境因素。而本研究准备研究的遮挡条件正是对统计知觉现象研究的一种补充。因为以往的研究中，被试需要进行判断的事物的统计属性，如圆的平均大小，箭头的平均朝向，甚至是人脸的平均表情，都是完全确定的，而遮挡条件存在时，在人的眼中，有些客体的属性就具有了不确定性，如一个被遮住的三角形，其大小有多种可能，那么人们在判断这些图形的平均大小时，有没有考虑到这些不确定性呢？有没有对其属性进行推测？推测依据什么呢？这些都是有理论意义的问题，对人们理解统计知觉这一过程有重要的意义。

实践意义上，正因为不确定性是一个在实际场景中经常遇到的问题，所以本研究的成果也可以应用于实际场景中。由于本研究的问题是由现实中的现象引出的，因而也可能在现实中得到应用。本研究的结论可以应用到界面设计，视频游戏等需要对一些客体的平均大小进行知觉的任务中。而如何设计客体之间的遮挡关系也可能是工效学应用中的一个潜在的问题。

2 实验设计与结果

2.1 实验方法

实验采用大小统计知觉的经典范式，实验流程图见图 3。

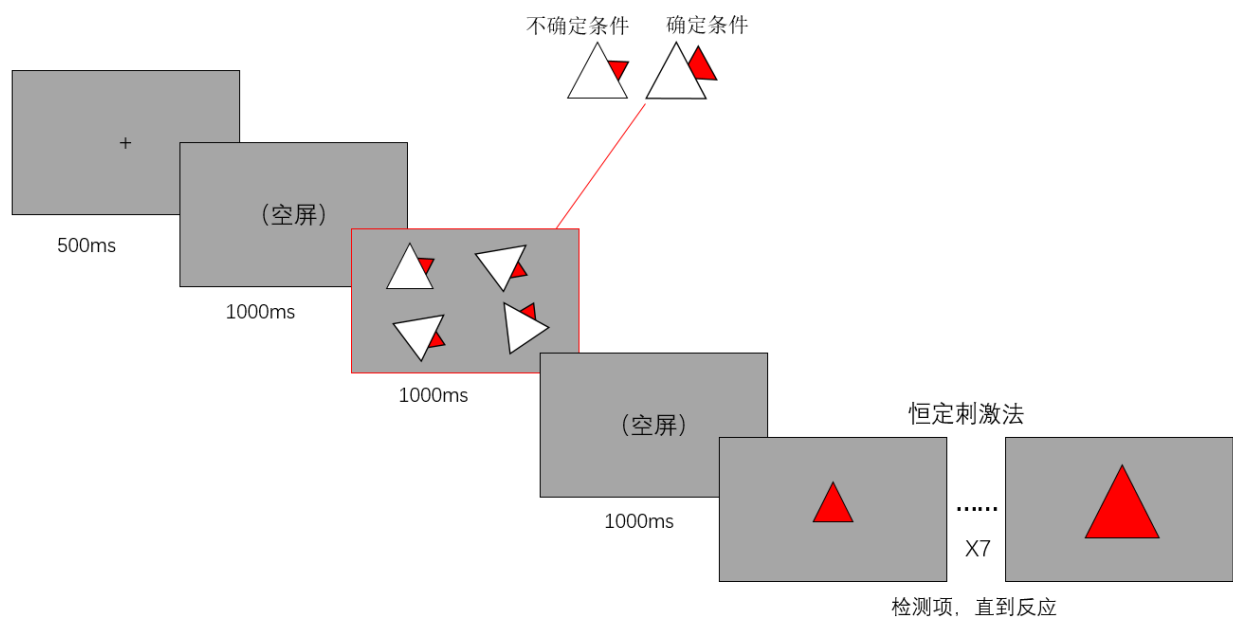


图 3 实验流程图

每个试次开始时，先呈现一个注视点，让被试将视点集中在上面。在一段 1000ms 的空屏之后，会呈现四对正三角形，每对三角形都存在遮挡关系，其中白色的大三角形遮住了红色的小三角形的一部分，而不同对之间的三角形不会相互遮挡。每对三角形之间的遮挡关系分为两种情况，一种是小三角形只露出一个角，另一种是小三角形露出两个角。两个三角形之间所成的角度分为 30°、60°、90° 三种水平。前面的讨论已经提到，这样的操作可以创造两种不确定性不同的条件，其中只露出小三角形一个角的条件是不确定条件，露出被遮三角形两个角的是确定条件。在又一段 1000ms 空屏之后，呈现一个正三角形，被试需要判断这个正三角形与前面呈现的红色被遮挡三角形的平均大小的关系，这是个双项迫选任务。检测项正三角形的大小由前面呈现的被遮挡三角形的平均大小决定，分为 85%、90%、95%、100%、105%、110%、115% 七个水平，百分数表示与前面被遮挡的三角形的平均大小的比例关系。在确定条件下，被遮挡三角形的平均大小由其实际大小直接计算得到；在不确定条件下，被遮挡三角形的平均大小由可能的最大值和最小值的平均值决定。实验中三角形的边长、大小三角形的大小比例等参数均在一定范围内进行随机的并在处理数据时使用标准化处理，这样处理是考虑到，如果对刺激进行太多的限制和固定处理，会导致刺激的“人工性”太强，削弱了刺激的生态性。实验采用恒定刺激法进行。

实验共完成有效被试 13 名，其中男性 6 名，女性 7 名，均为浙江大学本科生，年龄

在 20 岁-22 岁之间。实验程序采用 Matlab 软件进行编写，实验时显示器分辨率为 1024×768，刷新率为 85Hz。整个实验时间持续 90 分钟，被试将得到 45 元报酬。

2.2 实验结果与讨论

采用 Weibull 函数拟合分别拟合 13 名被试的反应曲线。将 13 名被试的反应曲线参数进行平均后得到确定和不确定两种条件的平均参数。通过这两组参数作出反应曲线图如图 3 所示。

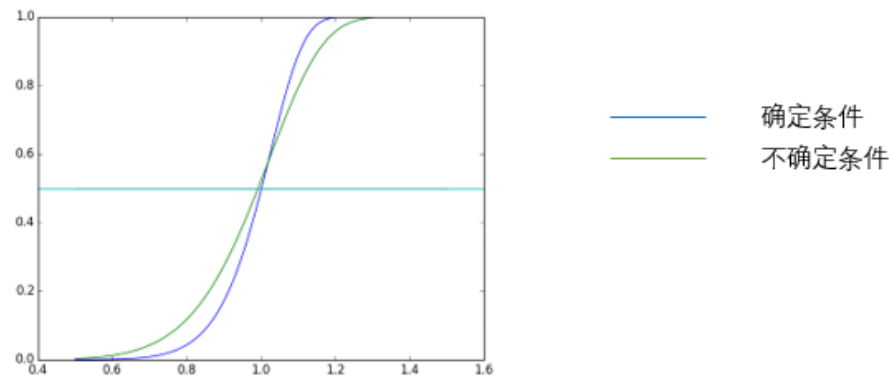


图 4 两种条件下的被试反应曲线

图 4 表明，不确定条件下，反应曲线较为平缓，即 Weibull 分布更加分散，表明被试的判断更不稳定；确定条件下，反应曲线较为陡峭，即 Weibull 分布更加集中，表明被试的判断更加稳定。

Weibull 函数有两个参数， α 和 β 。 α 为尺度参数，主要决定 Weibull 分布的范围； β 为形状参数，决定了 Weibull 分布的形状。一般来说，当 α 固定时， β 的增大使 Weibull 分布更为集中。对两种条件下的曲线参数进行配对样本 T 检验，结果发现 α 的差异不显著 ($t(12) = 0.522, p > 0.1$)， β 的差异显著 ($t(12) = -7.603, p < 0.01$)，而不确定条件下的 β 大于确定条件下的 β 。如图 5 所示。在 α 没有显著差异的前提下，不确定条件下的 β 较大，进一步说明了当不确定性较强时，被试的反应比较分散。

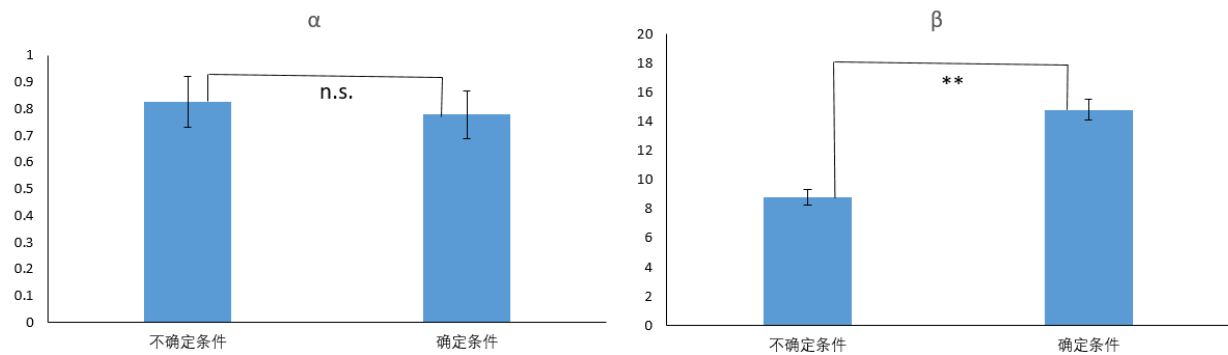


图 5 不同条件下的反应曲线参数

将两种条件下的反应曲线用直线内插法处理，取 50% 的反应概率与反应曲线的交点，作为被试对被遮挡三角形的平均大小的主观判断。对这一主观判断值进行配对样本

T 检验，无显著差异 ($t(12) = -1.332, p > 0.1$)，将两种条件下的主观判断值分别与标准值 1 进行独立样本 T 检验，均无显著差异。该结果说明：（1）在确定条件下，人能准确地推测并知觉出客体的实际平均大小；（2）在不确定条件下，人能准确地推测并知觉出客体平均大小的可能范围内最有可能的大小。

3 综合讨论

本研究通过遮挡条件的操作创造出两种不同程度的不确定性条件，并在这两种条件下对被试的反应特点进行了比较。结果发现，确定条件下，被试能够准确地知觉到被遮挡客体的实际平均大小，而且反应相对比较集中、稳定；不确定条件下，被试的主观判断值集中在判断的可能范围的均值附近，反应相对较分散。结果表明，在有遮挡存在的条件下，虽然客体的形状是不完整的，但人仍然会推测出客体的完整形状，进行“完型”并准确估计客体的实际平均大小，甚至在不能直接推测出客体实际形状和大小的情况下，也能推测出客体最有可能的形状及平均大小，反映为可能范围的均值。

在以往的大小统计知觉研究中，刺激材料往往是离散、平铺的。也就是说，是不存在不确定性的。然而，在生活场景中，人时时刻刻都将面临着带有不确定性的场景，比如估计一棵树上的树叶的平均大小，模糊图片里的字体大小等。带有不确定性的情景是统计知觉研究中必将考虑到的，因为不确定性的引入不仅将提高刺激材料选取的生态性，也因更贴近人在实际场景中的表现而更加接近人的认知过程的实质。不过，本研究的刺激与实际场景中人类面对的情景仍有较大差距，在之后的研究中应更进一步。

在大小统计知觉研究领域，比较流行的观点是对各个客体的大小进行类似于加权平均的计算(如 Chong & Treisman, 2003)，但在存在不确定性的情况下，单纯的加权平均计算并不能够实现（如果客体的大小是不确定的，那么便无法直接进行平均计算），其中必然需要加入概率推理的过程，这是构建人的统计知觉模型时必须要考虑到的问题。同时，传统的加权平均计算模型也不能解释不确定性增大时，被试判断的变异程度也增大的现象。有研究（Haberman & Whitney, 2015）探索了材料的变异程度对统计知觉的影响，该研究用表情人脸来研究表情的统计知觉。结果发现，人脸表情的变异程度

（Variance）越大，被试的判断的变异程度也随之增大。变异程度增大也可以看做不确定性的一个方面，表明了统计知觉中不确定性确有一定的影响。

当被遮挡的三角形的大小不确定时，对其平均大小的估计可能需要其上层的遮挡三角形的大小信息（就像对整棵树的树叶的平均大小的估计可以借助表面的未被遮住的树叶大小的估计来完成）。而由于在本实验中使用的材料决定了，遮挡三角形的大小一定会大于被遮挡三角形，因此可以推测，不确定条件下对被遮挡三角形的平均大小估计会稍微偏大。这种预测来源于贝叶斯推理（Bayesian Inference），贝叶斯推理针对于存在不确定性的场景中，认为对事物的判断并不只基于事物本身的性质，而是与先验的知识经验（Prior）有关。关于贝叶斯推理的研究在知觉领域中取得了一些成果，比如应用贝叶斯推理可以很好地解释时间知觉的回归效应(Regression effect, Shi et.al., 2013)。在本研

究的情境下，即对被遮挡的三角形的大小判断，不仅取决于被遮挡三角形的大小信息，还受到其他信息的影响，而最有可能作为先验信息的就是遮挡的大三角形，导致被试实际的主观判断值偏大于可能的大小范围的均值。但实际的实验结果表明，对被遮挡三角形的平均大小判断并未产生偏离。我们认为可能是以下原因导致：由于本实验中的材料，一对中的两个三角形设置了不同的颜色，且任务是只要关注被遮挡三角形的，那么被遮挡的三角形与遮挡三角形就容易被分成截然的两组，互相的联系相当微弱，导致被试不容易将遮挡三角形作为先验信息。若要进一步验证在不确定性条件下是否存在类似贝叶斯推理的过程，需要在以后的研究中进行改进。

4 结论与展望

本研究得出以下结论：在遮挡条件下，人依然具有准确地估计被遮挡客体的平均大小的能力。其中：(1)在存在遮挡但大小确定的条件下，能准确推知客体实际大小的情形下，人能够准确估计被遮挡客体的实际平均大小；(2)存在遮挡且被遮挡客体大小不确定的条件下，人对被遮挡客体的平均大小的主观判断值与大小的可能范围的均值一致；(3)不确定性的增强会导致人对被遮挡客体的平均大小估计更为分散，即变异程度更大。

本研究得到了有意义的结果，但对于不确定条件下的统计知觉研究，显然需要更多精细的研究。本研究选取了使客体大小产生不确定的其中一种方式——遮挡。而实际上，有更多的方式能产生类似的效果，如模糊、残缺等，用这些方式进行平均大小知觉的研究也是有意义的。

本研究虽然较之前统计知觉领域的研究，在刺激材料的选取上考虑了生态性，但实际上，所用的刺激仍然“很不生态”，互相遮挡的三角形虽然考虑了遮挡这一现实情境因素，但还是处于人工刺激材料的范畴中。实验中的客体与实际场景中的客体有较大不同，创设出的场景也离现实场景有很大的差距。之后的研究应更进一步，使实验的材料和场景更接近于人在现实生活中所见的刺激和环境，以更接近地获得人在实际环境中的认知特点。

虽然本研究获得了人在不确定条件下的平均大小知觉的一些特点，但单凭本研究无法建立一个完整统一的模型。前文提到，传统的加权平均模型在不确定条件下会遇到一些困难，而对不确定性条件，需要引入一些关于概率推理的过程。笔者认为，建立产生式模型是解决这一问题的一个好的选择，即先根据现象对人的解决统计知觉的认知系统做一定的假定，建立一个模型，再通过实验结果验证。概率过程可以较好地蕴含在建立的模型中，而且根据建立模型的特点，较容易设计实验进行验证并改进模型。未来会进行较多该方面的尝试。

参考文献:

张军伟.(2014). 统计知觉的层次计算模型. 硕士学位论文. 浙江大学.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*, 21(4), 560-567.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, 74(810-815).

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3).

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of the Sciences*, 106(18), 7345-7530.

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, 12(2), 157-162.

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary Statistics of Size: Fixed Processing Capacity for Multiple Ensembles but Unlimited Processing Capacity for Single Ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1440-1449.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751–R753.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual Working Memory Modulates

Rapid Eye Movements to Simple Onset Targets. *Psychological Science*, 24(5), 790-769.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the Function of Visual Short-Term Memory: Transsaccadic Memory, Object Correspondence, and Gaze Correction. *Journal of Experimental Psychology*, 137(1), 163-181.

Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, 18(860-869).

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

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, 64(7), 1055-1067.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001).

Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739-744.

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, 26-35.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1), 48.

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of vision*, 15(4), 16-16.

Shi Z, Church RM, & Meck WH. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 17(11), 556-64.

致谢:

终于迎来了本科生涯的尾声，回想这四年，多少难忘的经历犹在脑海。在毕业之际，首先要感谢我的导师，沈模卫教授。沈老师治学严谨，心系学术的作风给了我很大的影响，并坚定了我跟随他继续深造的决心。沈老师在这几年给了我很多指导与帮助，使我获益匪浅，尤其是在做心理学研究方面，让我取得了很大进步。沈老师，言语无法表达对您的感谢。

感谢周吉帆老师对我的指导，周老师为我的选题和研究实施提供了很多重要的建议和帮助，跟他的多次讨论也让我获益匪浅。向您致以谢意。

感谢徐昊骅、唐宁等各位师兄师姐，感谢他们给我提供的帮助和支持。

感谢我的同学和朋友们，很感激和你们共度了难忘的青春。

最后，感谢一路走来所经历的幸福和挫败，让我成长。

史博皓

2016年5月9日

毕 业 论 文（设计） 考 核

一、指导教师对毕业论文（设计）的评语：

指导教师(签名) _____

2015 年 5 月 22 日

二、答辩小组对毕业论文（设计）的答辩评语及总评成绩：

成 绩比例	文 献 综述 占 (10%)	开 题 报告 占 (20%)	外 文 翻译 占 (10%)	毕业论文（设计） 质量及答辩 占（60%）	总 评成绩
分 值					

答辩小组负责人（签名） _____

2016 年 5 月 27 日

浙江大学本科生毕业论文（设计）诚信承诺书

- 1.本人郑重地承诺所呈交的毕业论文（设计），是在指导教师的指导下严格按照学校和学院有关规定完成的。
- 2.本人在毕业论文（设计）中引用他人的观点和参考资料均加以注释和说明。
3. 本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。
4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

毕业论文（设计）作者签名：

2016 年 5 月 20 日