

# 《心理学报》审稿意见与作者回应

题目：多目标追踪中基于运动信息的分组效应

作者：胡路明, 苏晶, 魏柳青, 张学民

---

## 第一轮

### 审稿专家 1 意见回应

《多目标追踪中基于运动信息的分组效应》一文的研究选题对视觉多目标追踪中运动信息的作用进行了研究,通过三个实验观察到运动信息能够导致多目标追踪出现分组效应,并且表面特征能在一定程度上影响运动信息的分组效应。该研究的成果对于探究运动客体的视觉加工机制,拓展现有的理论具有推动作用。论文的文献引用能体现相关研究主题的发展脉络及现状,实验设计的思路清晰,实验数据能够为研究结论提供支持。论文全文已趋近完善,但有如下建议供作者参考:

**意见 1:** 该研究中,“多目标追踪中存在基于运动信息的自动化分组效应”是核心的结论,但论文对于实验结果如何为“自动化”加工的理论解释提供支持,其中的推论逻辑如何,虽然在论文“综合讨论”部分有所涉及,并给出了相应的参考文献来源,但还是建议作者在问题提出部分就先给出一个更为清晰的解释,以使作者能够更顺利地理解各个实验结果及结论。

**回应:** 十分感谢审稿专家的意见。作者已经根据审稿专家的建议在正文前言问题提出部分以及讨论部分新增了更为清晰的推论解释。具体修改请参见正文第 5 页红色字体部分和正文第 18 页“5.2 多目标追踪中基于运动信息的自动化分组机制”红色字体部分。

**意见 2:** 实验 1 的客体运动速度为 8.5 %s,而实验 2 和 3 的运动速度为 10.2 %s,在实验中运动速度这一实验参数是如何确定的?

**回应:** 十分感谢审稿专家的意见。实验 1 和 2 中的速度是参照课题组之前的实验以及预实验结果来确定的。在正式实验中,各个物体的运行速度在某一初始速度的上下一段速度范围内随机变化。8.5 %s 和 10.2 %s 则是运动后计算的平均运动速度。实验 3 的设计类似于实验 2,因此为了和实验 2 具有一致性,继续保持了实验 2 的速度。此外,此前的研究也大多有一定的速度差异范围,如 Suganuma 和 Yokosawa (2006)的实验一中物体运动范围在 0.5 %s-4.5 %s

内。Scholl, Pylyshyn 和 Feldman (2001) 在研究中采用的速度范围为 0 %s-7.02 %s 等。因此，对于本研究而言，2 %s 左右的速度差异并不会影响当前结论。

参考文献

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

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

意见 3：有几处文字编辑的细节请作者检查核实。一处为“2.5”部分第一段落中，“ $\eta_p^2 = 0.6.$ ”应该以中文句号结尾；“4.5”部分第一段落中，“ $\eta_p^2 = 0.95.$ 。”应该是多一句号；“5.1”部分第三段落中，“此类特征能够在目标内部分组时促进多目标追踪表现，但不会在目标与非目标间分组时并不会损害多目标追踪表现”，应该是多用了“不会”二字，此外用“成一组”来替代“分组”，可能更恰当；同样在 5.1”部分第三段落，“第四类是自动分组的加工特征 (automatic nongrouping features)”中“nongrouping”应该有错。

回应：非常感谢审稿专家所提的以上细节。作者已对正文相应部分进行了修正，并进一步仔细校对了全文。具体修改请见正文第 8 页“2.5”部分第一段落红色字体部分，第 15 页“4.5”部分第一段落红色字体部分。第 18 页“5.2”部分(调整了顺序)第一段落红色字体部分。

.....

审稿专家 2 意见回应

该研究发现，目标的运动方式相同可促进追踪，目标与干扰子的运动方式相同可损害追踪，而且，此干扰和损害量因表面特征的出现而缩减。研究具有一定价值，对多目标追踪领域的研究有良好的推进作用。但该研究还存在以下一些问题。

意见 1：对前人研究(Suganuma & Yokosawa, 2006)和此研究的关系界定不是很准确。文中表述为“然而，该研究着重于考察运动信息在多目标追踪中建立客体连续性的作用，尚未详细探究运动信息是否可以作为一类分组特征来影响多目标追踪”，但事实上，Suganuma & Yokosawa (2006)的文章标题就说的是“grouping”分组。区别是在于 Suganuma & Yokosawa (2006)研究中只做了目标和干扰子的共同运动，此研究增加了目标和目标共同运动的条件，以及客体的表面特征。

回应：十分感谢审稿专家的意见。此处表述有些许偏差。首先，Suganuma & Yokosawa (2006) 的研究确实提及到了“grouping”分组效应，也在讨论部分有所论述，但是这一研究并非重点关注“grouping”效应，在其问题提出部分也并未关注“grouping”效应的问题。从全篇来看，该研究论述“grouping”效应主要是用来解释其结果，论述高水平的知觉组织在维持标签-目标时的作用。也即，Suganuma & Yokosawa (2006)的对“grouping”效应的关注是先现象后理论的。而本研究中则直接从“grouping”效应开始进行推论，重点考察运动信息作为分组特征时对多目标追踪产生的影响，从理论到现象演绎。这一点上文章的思路类似于 Erlikhman 等(2013)的研究。是以，作者认为 Suganuma & Yokosawa (2006)一文并未如 Erlikhman 等(2013)的研究一般，即直接关注特征本身是否可以产生“grouping”效应。其次，本文中分组效应的概念也与 Suganuma & Yokosawa (2006)的有所不同。Suganuma & Yokosawa (2006)主要认为相同运动信息的客体可作为 grouping 的线索，进而出现 grouping 效应。这一点只是延续了 Yantis (1992)的早期观点。而本文实际上想探讨的 grouping 效应应该类似于 Erlikhman 等(2013)，一方面目标群内部的分组能够促进追踪成绩，另一方面目标与非目标间的分组会损害追踪成绩。也即试图区分出 grouping 效应在不同情形下的影响。虽然两篇文章中都涉及对 grouping 效应的论述，但本文中对分组效应的分析更为直接和更为具体。最后，作者已经根据审稿专家的意见，进行了更为准确清晰的界定。具体修改请参见正文第 3 页第二自然段红色字体部分。

#### 参考文献

- Suganuma, M., Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: impairments due to common item motions. *Perception*, 35, 483–495.
- 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 & Performance*, 39, 1625–1637.

意见 2：图 2 的图例应介绍得更详细以便读者理解，因为此研究也是借用的这一范式。

回应：十分感谢审稿专家的意见。作者已对图 1 和图 2 均进行了更为详细的注解，以便于读者理解。具体修改请参见正文第 4 页图 1 和第 7 页图 2 图例部分。

意见 3：客体是在运动的一开始就进入配对运动，还是先开始随机运动一段时间然后进入配对运动？

**回应：**十分感谢审稿专家的意见。本研究中所有的配对运动都是从运动阶段开始即进入配对运动状态，持续 10s，并在运动阶段结束前 1s 解除配对关系，进入随机运动状态，以防止被试采用策略进行选择。作者已在正文中增加了更为详细的描述。具体修改请参见正文 7 页“2.3 实验设计”第二段的红色字体部分和正文 8 页“2.4 实验程序”的红色字体部分。

**意见 4：**为什么配对运动条件下客体间接触无反弹，而随机运动条件下客体不接触即提前反弹？这个实验条件设置的不对等可能如何影响实验结果？

**回应：**十分感谢审稿专家的意见。这一问题在设计实验时也经过讨论和考虑。首先，本实验中运动客体有 12 个，在预实验中发现若配对追赶运动时允许客体间相互碰撞反弹，会导致碰撞反弹的现象大面积出现，影响配对追踪效果。因此，配对追踪运动时设置为接触无反弹的情形。而随机运动条件下设置不接触提前反弹的情形是考虑到配对追赶运动时前后运动客体间的间距为三倍客体直径，即这部分客体之间也相当于无接触的状态。为了抵消这一现象可能带来的增益效应，进而产生出差异显著的假阳性结果，实验中将随机运动设置为不接触提前反弹的状态。当然，如此一来也会导致前后设置不对等。然而，即使将随机运动条件也设置为接触无反弹，这一设置也是不对等。因此，在上述两种不对等情形下，将随机运动设置为不接触提前反弹可能会比设置为接触无反弹所得的结果更为保守，更不易造成假阳性结果。最后，从实际结果来看，配对运动状态与随机运动状态的对比差异十分显著， $p < 0.01$ ，Cohen's  $d > 0.75$ 。这表明，配对运动状态对多目标追踪结果的影响是极其显著和稳定的。作者再次感谢审稿专家的如此详细建议。

**意见 5：**实验之后是否有询问被试有否发现三种不同的运动条件？

**回应：**十分感谢审稿专家的意见。实验前并未告知被试当前实验的具体设置。因此在实验结束时，实验主试与被试进行过简短交流，绝大多数被试均反馈只发现实验条件有所差异但并未特别在意，也有个别被试有发现具体差异，但其反应出的追踪方式并不具有策略性。此外，本实验对三种实验条件进行了随机化处理，三种实验条件混合随机呈现，被试只能在追踪开始后才可能知晓当前运动条件。在一定程度上，也避免被试产生策略和练习效应。

**意见 6：**图的纵轴应标明其所表示的变量。

**回应：**十分感谢审稿专家的意见。作者已将所有数据结果图添加上纵坐标变量名。具体修改请分别参见正文第 9 页图 3，正文第 12 页图 6 和正文第 16 页图 9。

意见 7：文中提到的图的编号与图题中标示的编号不对应。

回应：十分感谢审稿专家的意见。作者已对全文图表编号和正文引用进行检查和修改。所有图表及其引用均已正确。

意见 8：三个实验的被试量均偏少，而且数量不等。被试量是如何确定的？至少应各个实验的被试量基本相同，以便比较。

回应：十分感谢审稿专家的意见。首先，实验 1 和实验 2 的被试量都是参照实验组和此前国内外相关实验被试的数量大致确定的。如 Scholl, Pylyshyn 和 Feldman (2001)在这一研究中的每个条件为 9 个被试。Suganuma 和 Yokosawa (2006)在其三个实验中的被试量分别为 10 个,12 个和 12 个。此外，在实验正式开始前，也采用 G\*Power 软件预估了被试量，如实验 1，在 Effect size  $f = 0.4$  (大效应量)的情况下，当  $\alpha = 0.05$  时需要达到 power  $(1-\alpha) = 0.8$  所需的被试量约为 12 人。

Test family

F tests

Statistical test

ANOVA: Repeated measures, within factors

Type of power analysis

A priori: Compute required sample size - given  $\alpha$ , power, and effect size

Input Parameters

Determine =>

Effect size f

0.4

$\alpha$  err prob

0.05

Power ( $1-\beta$  err prob)

0.8

Number of groups

1

Number of measurements

3

Corr among rep measures

0.5

Nonsphericity correction  $\epsilon$

1

Output Parameters

Noncentrality parameter  $\lambda$

11.5200000

Critical F

3.4433568

Numerator df

2.0000000

Denominator df

22.0000000

Total sample size

12

Actual power

0.8162038

而实验 2 和实验 3 相同，在 Effect size  $f = 0.4$  (大效应量)的情况下，当  $\alpha = 0.05$  时需要达到 power  $(1-\alpha) = 0.8$  所需的被试量约为 8 人。

Test family		Statistical test	
F tests		ANOVA: Repeated measures, within factors	
Type of power analysis			
A priori: Compute required sample size – given $\alpha$ , power, and effect size			
Input Parameters		Output Parameters	
Determine =>	Effect size f	Noncentrality parameter $\lambda$	15.3600000
	$\alpha$ err prob	Critical F	2.4851432
	Power (1- $\beta$ err prob)	Numerator df	5.0000000
	Number of groups	Denominator df	35.0000000
	Number of measurements	Total sample size	8
	Corr among rep measures	Actual power	0.8111719
	Nonsphericity correction $\epsilon$		

其次，当前各个实验的被试量不等是因为在实验设计时并未设想着相互做对比。此前，当前三个实验是主题上一致，但每一部分都可视为独立，且正文中也并未相互作出对比。当然，也确实可以如审稿专家所言，设置为相同被试量，但其结果也将与当前所得结果一致。因此，当前实验的被试量也是较为合理的。但审稿专家此处所提建议，作者将在此后的研究中对被试样本问题加以更充分考虑和控制。非常感谢专家细致的建议！

**意见 9：**表 2 与表 1 的内容基本相同，可略去，仅留图 7 即可。

**回应：**十分感谢审稿专家的意见。已经正文中表 2 内容删除。具体修改请见正文第 14 页“4.3 实验设计”红色字体修改部分。

**意见 10：**实验 2,3 的检测阶段是怎样的？所有客体都呈现同样颜色和形状，还是保持各自在追踪阶段的颜色和形状？应明确说明实验设置，及其对结果的可能影响。

**回应：**十分感谢审稿专家的意见。实验 2 和实验 3 在检测阶段均呈现相同的颜色和形状，采用相同大小的黑色实心圆(30 pixel)掩蔽。作者已经正文中加以详细阐述。具体修改请参见正文第 11 页“3.4 实验程序”红色字体修改部分。

**意见 11：**在综合讨论 5.1 部分，作者认为“视觉索引标签也被绑定在一起，或者说可以使用同一视觉索引标签指向两个视觉客体”，以此来解释基于运动的分组效应，这个解释的说服力不强。如果 TT 追踪条件下 6 个目标确实被两两绑定，只用 3 个视觉索引标签即可追踪，其追踪绩效将远远好于随机条件下用 6 个索引标签的情况，但实际差异结果并没那么大。

**回应：**十分感谢审稿专家的意见。审稿专家所提意见确实十分精准。作者已对此处表述进行

修改，并结合审稿专家的意见 12 作出了统一的整合解释。具体修改请参见正文第 18 页“5.2 多目标追踪中基于运动信息的自动化分组机制”的红色字体修改部分。

**意见 12：**作者对实验结果的解释还没有形成一个统一的系统。除视觉索引解释外，作者还用了格式塔法则、工作记忆占用进行解释。各种解释有其各自的道理，把它们放在一起，多目标可能是怎样的一个过程？如理论解释能有所整合，则研究的理论价值还可有所提升。

**回应：**十分感谢审稿专家的意见。在涉及到 grouping 效应和工作记忆的多目标追踪任务仅使用单一理论可能并不能做出很系统和全面的解释，因此作者做出了多个理论上的阐述。为了更进一步提升研究的理论价值，作者根据审稿专家所提建议对多个理论的解释进行了相应整合，更详细的阐述了多目标追踪过程。多目标追踪在不同阶段应用了不同的加工过程，在初级阶段预分配了视觉索引标签，在运动过程中继续维持这些视觉索引标签，并且还受到格式塔知觉原则和工作记忆的影响，导致可维持的视觉索引标签数量受到影响，追踪表现也随之受到影响。具体修改请参见正文第 18 页“5.2 多目标追踪中基于运动信息的自动化分组机制”的红色字体修改部分。

**回复总结：**本轮修改已按照两位审稿专家的建议仔细修正和补充。十分感谢两位审稿专家在这一轮修改中提出的修改意见和建议，使得本研究的成果更进一步地得到修改和完善。对此，再次对审稿专家们的付出、工作和细致的建议致以诚挚的谢意。

---

## 第二轮

### 审稿人 1 意见

**意见 1：**作者的修改基本可接受，仍存在以下问题，尤其是样本量需增加。作者对实验所需被试样本量的预估过于宽松。在实验前假定  $f = 0.4$  (大效应量)，设定  $\alpha = 0.05$ ， $\text{power}(1-\beta) = 0.8$  是一些传统研究的做法，但不是最佳的做法。这是把这三项标准都设到了可接受下限，计算出来的样本量也是最低值，很可能造成结果的不稳定。建议作者用更严格的标准预估所需样本量。另外，作者每名被试每种条件下的实验次数也不多，同样可能带来结果的不稳定。作者现在的样本量对实验 1 和 2 的结果可能影响不大，但很可能影响实验 3 的结果。因为实验 1 和 2 的效果量显示均较大，而实验 3 的效果量较小。尤其现在实验 3 的交互作用是边缘

显著，而作者还是按显著进行了简单效应检验及后续的解释和讨论。作者至少应参考现在的交互作用的效果量重新估计所需样本量，将样本量补到该数目，再看结果到底如何。如结果还不明确，则需考虑调整解释和讨论。

**回应：**十分感谢审稿专家的意见。专家对样本量的关心确实非常重要。作者之所以在实验前假定  $f = 0.4$  (大效应量)， $\alpha = 0.05$ ，Power  $(1-\beta) = 0.8$ ，主要是因为参考了之前此类同等实验中所得出的效应量和 Power 值。不过，作者也很认可和接受审稿专家的建议。因此，作者根据专家的建议在样本量估计和效力计算上做了如下四方面的工作。第一，在完全先验的条件下，使用三种不同的先验参数来预估所需的样本量。第二，基于部分后验信息，即参照当前结果中的效应量(主效应或交互效应的效应量)，以及更严格的 Power 值(Power  $(1-\beta) = 0.95$ )来预估所需的样本量。第三，基于完全后验信息，在当前样本量和效应量下计算 Power 值，核查当前实验的统计效力。第四，根据审稿专家的建议，增加实验 3 的样本量到严格标准所需的样本量 24 人（实际增加至 40 人）。整体而言，第一项工作反映出，当前实验实际样本量处在较严标准和严格标准之间，远高于传统标准所估计的样本量。也即，当前样本量可被认为处于可接受标准范围之内。第二项工作反映出，若以当前实验所得效应量来估计所需样本量，那么当前实验实际样本量要远超于在以当前实验得出的效应量为前提下预估的样本量。第三项工作反映出，基于当前三个实验的样本量和效应量等参数所计算出的对应的 Power 值都在维持在 0.99 的水平。这说明接受当前研究结果的把握是非常高的。前三项工作共同表明了当前样本量已经能够比较明确地体现出实验效应，且具有较高的统计功效。最后，第四项工作还反映出，增加样本量后所得的实验结果和此前所得实验结果相一致，且更显著和稳定。也即，再一次证明了当前实验所得实验效应和结果是相对稳定和可靠的。

**第一项工作：在先验信息下预估样本量(G\*Power)**

表 1 三种不同先验参数标准下的样本量估计

参数	实验 1			实验 2			实验 3		
	传统标准	较严标准	严格标准	传统标准	较严标准	严格标准	传统标准	较严标准	严格标准
Effect Size $f$	0.40	0.40	0.25	0.40	0.40	0.25	0.40	0.40	0.25
$\alpha$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Power $(1-\beta)$	0.8	0.9	0.9	0.8	0.9	0.9	0.8	0.9	0.9
Estimated Sample Size	12	15	36	8	10	24	8	10	24
Actual Sample Size		16			18			13	



表 1 呈现出了三种不同先验参数标准下的样本量估计情况。由表可知，当前三个实验的样本量都略高于较严标准的预估样本量，远高于传统标准的预估样本量，与严格标准下的预估样本量之间存在一定差距。也即，当前三个实验的样本量都并非采用传统标准的下限数额来进行实验，而是介于较严标准和严格标准之间，是可接受的估算标准。因此，当前三个实验中的样本量也是可以接受的。

## 第二项工作：基于当前实验的实验效应的效应量进行样本量估计

G\*Power 中计算  $F$  检验 Power 值时使用的效应量指标是 Cohen's  $f$  值。因此，需根据先将 SPSS 结果输出中的偏  $\eta^2$  转为此处的  $f$  值后再计算 Power 值。转换操作可直接在 G\*Power 软件中的 Determine 选项卡中予以计算，如图 1-图 3。

表 2 基于当期实验效应量值和  $\alpha = 0.05$ , Power  $(1-\beta) = 0.95$  的前提下的样本量估计

参数	实验 1	实验 2	实验 3
Partial $\eta^2$	0.595	0.357	0.219
Effect Size $f$	1.212	0.745	0.530
$\alpha$	0.05	0.05	0.05
Power $(1-\beta)$	0.95	0.95	0.95
Estimated Sample Size	3	5	8
Actual Sample Size	16	18	13

表 2 在以当前实验效应量为前提的基础上，以更严格的标准(Power 提高到 0.95)来估算样本量，其估算结果表明当前实验的样本量也是充分的。

附实验 1：单因素被试内设计，采用主效应的效应量值( $\eta_p^2 = 0.595$ )进行样本量估计

图 1 实验 1 在  $\eta_p^2 = 0.595$ (对应  $f = 1.21$ ),  $\alpha = 0.05$ , Power  $(1-\beta) = 0.95$  的前提下估算样本量

实验 2：两因素被试内设计，采用交互效应的效应量值( $\eta_p^2 = 0.357$ )进行样本量估计

图 2 实验 2 在  $\eta_p^2 = 0.357$ (对应  $f = 0.75$ )， $\alpha = 0.05$ ，Power  $(1-\beta) = 0.95$  的前提下估算样本量

实验 3：两因素被试内设计，采用交互效应的效应量值( $\eta_p^2 = 0.357$ )进行样本量估计

图 3 实验 3 在  $\eta_p^2 = 0.357$ (对应  $f = 0.75$ )， $\alpha = 0.05$ ，Power  $(1-\beta) = 0.95$  的前提下估算样本量

第三项工作：根据当前实验的样本量和实际效应量参数计算 Power 值

表 3 基于当期实验的样本量和效应量值在 $\alpha = 0.05$ 的前提下所得 Power 值			
参数	实验 1	实验 2	实验 3
Partial $\eta^2$	0.595	0.357	0.219
Effect Size $f$	1.212	0.745	0.530
$\alpha$	0.05	0.05	0.05
Actual Sample Size	16	18	13
Estimated Power $(1-\beta)$	1	1	0.99

表 3 基于当前实验的样本量和各实验的实际效应量在  $\alpha = 0.05$  的前提下估算出了当前三个实验的效应量，每一实验的效应量都是相对较高的。因此，当前三个实验的统计效力是足够和稳定的。以下同时给出了 Power 值的计算结果图，以及在此基础上的样本量变化对 Power 值的影响趋势图。

附实验 1 Power 值计算

Test family: F tests, Statistical test: ANOVA: Repeated measures, within factors

Type of power analysis: Post hoc: Compute achieved power – given  $\alpha$ , sample size, and effect size

Input Parameters: Determine => Effect size f: 1.2120791,  $\alpha$  err prob: 0.05, Total sample size: 16, Number of groups: 1, Number of measurements: 3, Corr among rep measures: 0.5, Nonsphericity correction  $\epsilon$ : 1

Output Parameters: Noncentrality parameter  $\lambda$ : 141.037, Critical F: 3.3158295, Numerator df: 2.0000000, Denominator df: 30.0000000, Power (1- $\beta$  err prob): 1.0000000

From variances: Variance explained by effect: 1.0, Variance within group: 2.0

Direct: Partial  $\eta^2$ : 0.595

Calculate, Effect size f: 1.212079

图 4a 实验 1 在样本量为 16 人,  $\eta_p^2 = 0.595$ (对应  $f = 1.21$ ),  $\alpha = 0.05$  的前提下计算 Power 值

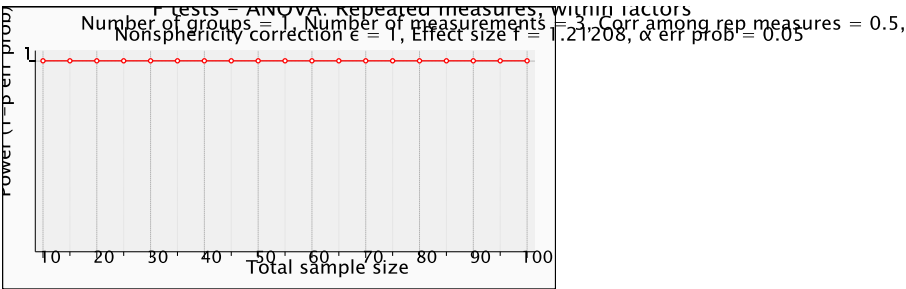


图 4b 实验 1 在  $\eta_p^2 = 0.595$ (对应  $f = 1.21$ ),  $\alpha = 0.05$  的前提下, 样本量变化对 Power 值的影响图

根据图 4a 所计算出的 Power 值可知, 实验 1 的统计效力很高, 因此当前结果的可接受程度也是很高的。此外, 图 4b 还呈现了样本量变化对 Power 值影响的趋势走向图。由图可知, 在当前实验中, 进一步增加样本量对 Power 值的影响并不明显。整体而言, 实验 1 结果的统计效力是很稳定和可靠的。

实验 2 Power 值计算

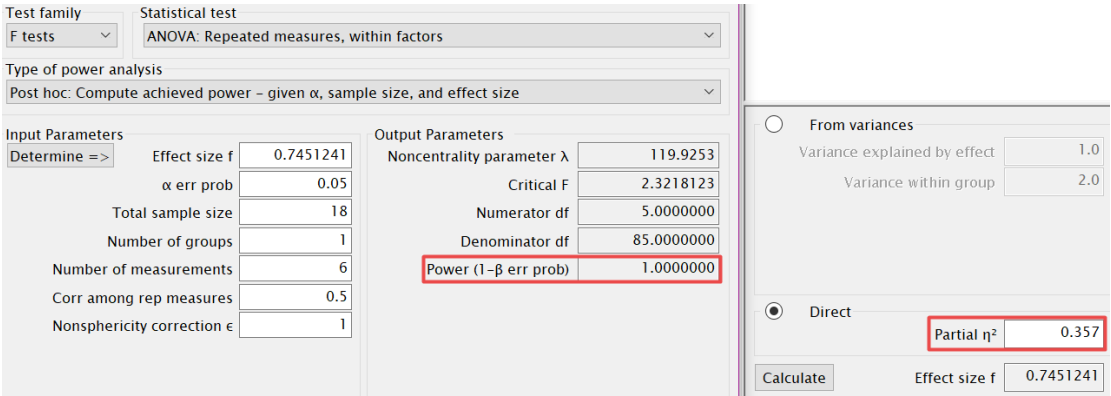


图 5a 实验 2 在样本量为 18 人， $\eta_p^2 = 0.357$ (对应  $f = 0.75$ )， $\alpha = 0.05$  的前提下计算 Power 值

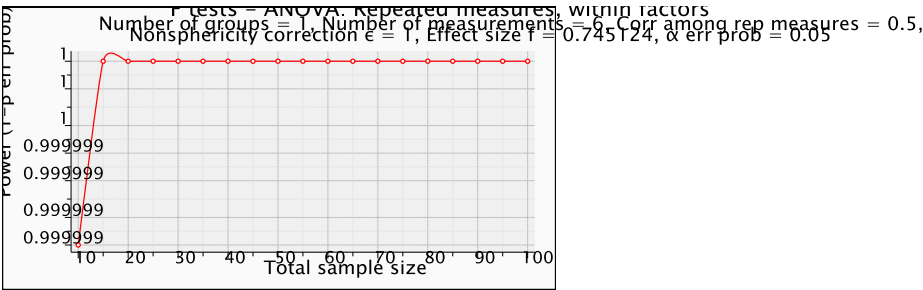


图 5b 实验 2 在  $\eta_p^2 = 0.357$ (对应  $f = 0.75$ )， $\alpha = 0.05$  的前提下，样本量变化对 Power 值的影响图

根据图 5a 所计算出的 Power 值可知，实验 2 的统计效力很高，因此当前结果的可接受程度也是很高的。此外，图 5b 也呈现了样本量变化对 Power 值影响的趋势走向图。由图可知，在当前实验中，进一步增加样本量对 Power 值的影响也并不明显。整体而言，实验 2 结果的统计效力也是很稳定和可靠的。

实验 3 Power 值计算

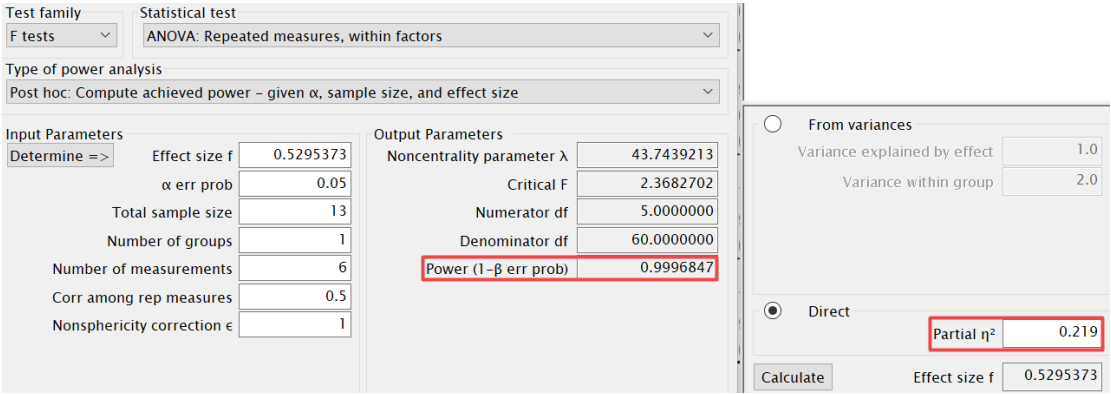


图 6a 实验 3 在样本量为 13 人， $\eta_p^2 = 0.219$ (对应  $f = 0.53$ )， $\alpha = 0.05$  的前提下计算 Power 值

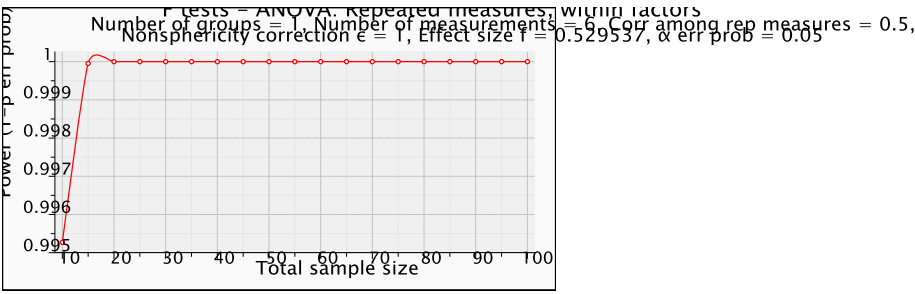


图 6b 实验 3 在  $\eta_p^2 = 0.219$ (对应  $f = 0.53$ )， $\alpha = 0.05$  的前提下，样本量变化对 Power 值的影响图

最后，根据图 6a 所计算出的 Power 值可知，实验 3 的统计效力也同样很高，因此当前结果的可接受程度也是很高的。此外，图 6b 也同样呈现了样本量变化对 Power 值影响的趋势走向图。由图可知，在当前实验中，进一步增加样本量对 Power 值的影响也同样并不明显。整体而言，实验 3 结果的统计效力也同样是很稳定和可靠的。

第四项工作：对新增样本量后的实验 3 再次进行 Power 值估算，检验其统计效力

根据严格标准  $f = 0.25$  (中效应量),  $\alpha = 0.05$ , Power  $(1-\beta) = 0.9$  所预估出的样本量为 24 人。为了达到这一要求, 避免因样本量造成的不稳定结果, 作者再次招募了 27 名被试进行实验 3。新增被试后的实验结果与未增被试时的实验结果一致, 只是更加显著和稳定了。为了解新增被试后的实验 3 结果的统计效力, 作者用新增被试后的实验 3 的效应量值再次对实验 3 进行 Power 值估算, 见图 7a 和图 7b。另, 新增被试后的具体数据和插图的变化, 请详见正文第 15-16 页“4.5 结果与分析”部分的红色字体内容。

附实验 3 补充被试后的 Power 值: 增加被试后

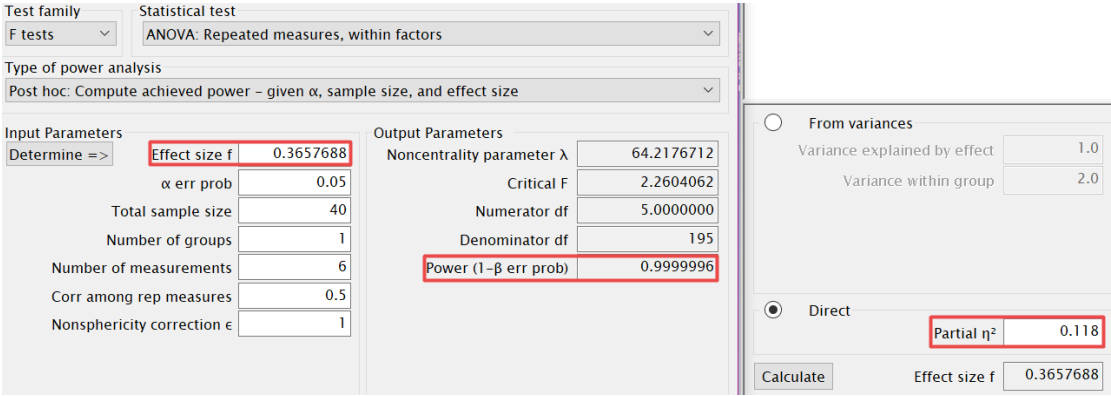


图 6a 实验 3 在样本量为 40 人,  $\eta_p^2 = 0.118$ (对应  $f = 0.37$ ),  $\alpha = 0.05$  的前提下计算 Power 值

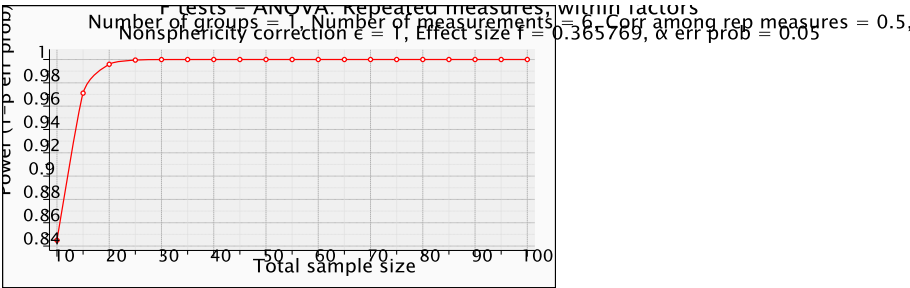


图 7b 实验 3 在  $\eta_p^2 = 0.118$ (对应  $f = 0.37$ ),  $\alpha = 0.05$  的前提下, 样本量变化对 Power 值的影响图

根据图 7a 所计算出的 Power 值(Power  $(1-\beta) \approx 1$ )可知, 新增被试后实验 3 的统计效力也同样很高, 因此当前结果的可接受程度也是很高的。此外, 图 7b 呈现的样本量变化对 Power

值影响的趋势走向图与此前的实验所得的趋势走向图基本一致。因此，此处实验 3 结果的统计效力是很稳定和可靠的。

此外，实验 2 和实验 3 的实验设计是类似的，在未增加被试前，实验 3 的效应还要远高于实验 2。因此，从增加被试后的实验 3 的结果同理可知，实验 2 的结果也应该如实验 3 一样，是较为稳定的。并且，实验 3 自身还包含着实验 1 的两种分组条件，TT chase 和 TD chase。在新增被试后的实验结果中显示，TT chase 在任一形状分组条件下的结果都仍显著高于 TD chase 条件。因此，实验 1 的结果也是相对稳定和可信的。也即，新增被试的超过严格标准样本量的实验 3 也在一定程度上从侧面映证了三个实验的结论都是稳定和可靠的。

综合全部四项工作而言，当前研究中的样本量是较为合理的，统计效力都是较好的，也都在可接受范围内。因此，有理由说明当前研究的研究结论是能够被当前实验加以证实的。

最后，关于实验次数的一点小说明。多年来国内外和本课题组的研究表明，20 试次对于多目标追踪研究是一个可接受的经验次数标准。国内外有很多实验室和研究者均以这一次数标准下做出了多项研究成果(Suganuma & Yokosawa, 2006; Makovski & Jiang, 2009; Ren, Chen, Liu, & Fu, 2009; Feria, 2010,2012; Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013; Oksama & Hyönä 2016; Wang, Zhang, Li, & Lyu, 2016)。这其中也包括作者在本研究中重点参考的两篇研究文献(Suganuma & Yokosawa, 2006; Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013)。因此，本课题组也一直采用这一经验次数标准进行实验。作者此后也会对此问题多多加以权衡和进行周到考虑。在此，十分感谢审稿专家对此的建议。

### 【参考文献】

- 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 & Performance*, 39(6), 1625-1637
- Feria, C. S. (2010). Attentional prioritizations based on spatial probabilities can be maintained on multiple moving objects. *Attention Perception & Psychophysics*, 72(4), 926-938.
- Feria, C. S. (2012). The effects of distractors in multiple object tracking are modulated by the similarity of distractor and target features. *Perception*, 41(3), 287-304.
- Makovski, T., & Jiang, Y. V. (2009). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology Human Perception & Performance*, 35(6), 1687-1697.

- Oksama, L., & Hyönä J. (2016). Position tracking and identity tracking are separate systems: Evidence from eye movements. *Cognition*, 146, 393-409. doi: 10.1016/j.cognition.2015.10.016.
- Ren, D., Chen, W., Liu, C. H., & Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision*, 9(5), 18. doi: 10.1167/9.5.18.
- Suganuma, M., & Yokosawa, K. (2006). Grouping and trajectory storage in multiple object tracking: impairments due to common item motions. *Perception*, 35(4), 483-495.
- Wang, C., Zhang, X., Li, Y., & Lyu, C. (2016). Additivity of Feature-based and Symmetry-based Grouping Effects in Multiple Object Tracking. *Frontiers in Psychology*, 7.

**意见 2:** 配对运动条件下客体间无反弹有遮挡, 随机运动条件下有反弹无遮挡。按照作者的解释来看, 该差异对结果的影响可能不大, 但以往多目标追踪的研究显示是否有遮挡会影响追踪绩效。作者最好在文中对此做简单说明。

**回应:** 非常感谢审稿专家的细心意见。作者已在正文对这一设置加以说明, 并对其可能产生的影响也一并做了详细补充。具体修改请详见正文第8页“2.4 实验程序”部分第一段落红色字体部分。

**回复总结:** 本轮修改作者已按照审稿专家的建议仔细修正和补充。十分感谢审稿专家在这一轮修改中提出的修改意见和建议, 使得本研究的成果更进一步地得到完善。对此, 再次对审稿专家们的付出、工作和细致的建议致以诚挚的谢意。

---

### 第三轮

#### 审稿人 1 意见

**意见 1:** 作者对实验 3 补充了被试, 获得了较为稳定且与原来一致的结果。文章已基本可以。还有一个关于 GPower 使用的提示。在将 SPSS 中的偏  $\eta^2$  转为 GPower 中的 f 值计算 power 时, 需在 Options 中更改设置 Effect size specification “as in SPSS”。因为 SPSS 中的偏  $\eta^2$  和 GPower 中的偏  $\eta^2$  是不同的, 前者小于后者, 不更改此设置将高估 power, 低估所需的样本量。这一问题不明显影响此研究的实验结果, 供作者在后续研究中参考。

**回应:** 十分感谢审稿专家的提示。事实上, 作者在上一轮回应中的正式操作已经对此进行过设置了, 但在回应时未能阐述详尽。因此, 此前回应中的数据结果均是在 Effect size specification “as in SPSS”之中进行的, 所以并未高估 power 值和低估样本量。



---

### 主编终审

**意见 1:** 实验程序描述还需要条理化。目前表示先后的用语太随意。

**回应:** 十分感谢主编专家的意见。作者重新梳理了实验程序描述，使其更加条理化。

**意见 2:** 在本研究中，设置了三种条件。在正文和图表中，写成“TT chase 条件、Random 条件和 TD chase 条件”。这样的写法表达意思不准确，建议作者要么选择用中文，要么选择用英文。最好是中文表达。

**回应:** 十分感谢主编专家的意见。作者根据主编专家的意见，已经将所有条件的写法统一转为中文表达。具体修改请参见正文。“TT chase 条件”改为“目标与目标配对条件”，“Random 条件”改为“随机无配对条件”，以及“TD chase 条件”改为“目标与非目标配对条件”。

**意见 3:** “摘要”“关键词”之后都无需冒号。中英文的关键词之间都用分号隔开。

**回应:** 十分感谢主编专家的意见。作者已经按照《心理学报》的发表标准进行了修正。具体修改请参见正文第 1 页“关键词”的青色字体修改部分。

**意见 4:** 请将小数点前省略的 0 补上，如“ $p < .001$ ”应为“ $p < 0.001$ ”。余同。

**回应:** 十分感谢主编专家的意见。作者已经按照《心理学报》的发表标准进行了修正。具体修改请参见正文。

**意见 5:** 文中彩图需考虑修改，黑白印刷版中各颜色之间应具有区分度。

**回应:** 十分感谢主编专家的意见。作者已经按照《心理学报》的发表标准进行了修正。具体修改请参见正文图片。