

对作者(年份)研究结果的计算可复现性检验

小组成员分工

组长	郭冰冰		
组员	孙睿忻、姜文彬、曹钟文		
分工			
数据分析	郭冰冰（33%）、孙睿忻（33%）、姜文彬（33%）	PPT 制作	郭冰冰（25%）、孙睿忻（25%）、姜文彬（25%）、曹钟文（25%）
文字报告制作	郭冰冰（25%）、孙睿忻（25%）、姜文彬（25%）、曹钟文（25%）	PPT 展示	曹钟文

\* 同一名同学可负责多个部分；如同一内容由多位同学负责，可按百分比注明贡献占比

**摘要：**通过分析鼠标移动轨迹来研究认知过程，已成为心理学研究中的一种常用方法。然而，鼠标追踪实验在设计上存在多种变体，例如参与者是否需点击确认答案、鼠标的灵敏度设定以及起始操作方式等，目前尚不清楚这些方法学差异是否会系统性影响研究结果。本研究采用典型/非典型分类任务，系统操控了三种常见设计因素：反应指示方式、鼠标灵敏度以及起始程序，以探讨其对鼠标轨迹的影响。结果发现，非典型刺激导致鼠标轨迹更偏向未选项的现象在所有条件下均能复制，但其效应强度随设计设置变化而不同。点击作答和动态起始条件下，该效应更强。轨迹形态在不同实验设置间也存在差异，如动态起始产生平滑曲线，点击作答更易出现“改变主意”式轨迹。曲率分布在部分条件下呈双峰结构，提示存在多种认知路径。本研究支持：在利用鼠标追踪法进行心理推理时，实验设计本身可能对关键指标产生实质影响，应在解释认知过程时加以审慎考量。

**关键词：**鼠标追踪、认知过程、实验设计、决策过程、反应动态、计算可复现性

1 引言

1.1 所选文献信息

表 1 文献信息表

1 文献基本信息	
所选文献	Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? <i>Behavior Research</i>

	<i>Methods</i> , 52(1), 317 – 341.		
数据来源	OSF 链接: <a href="https://osf.io/xdp7a/">https://osf.io/xdp7a/</a>		
<b>2 文献选取</b>			
文献主题是否包含不止一篇研究?	<input type="checkbox"/> 是, 且包含元分析研究 <input checked="" type="checkbox"/> 是, 但不包含元分析研究 <input type="checkbox"/> 否	文献此前被其他研究者重复过?	<input type="checkbox"/> 是(附上原文链接) <input checked="" type="checkbox"/> 否
<b>3 研究假设选取</b>			
重复的研究假设	检验了实验 3 中关于不同启动程序对鼠标轨迹影响的假设验证。 具体假设: 与静态启动相比, 动态启动和 intimax 启动程序会产生更大的轨迹曲率 (MAD 值更高)。		
重复的研究假设是否在其他研究中经过重复?	<input type="checkbox"/> 是(附上原文链接) <input checked="" type="checkbox"/> 否	文献共几个实验, 重复的研究假设是第几个实验中的?	共 3 个实验, 重复的是第 3 个实验中的假设
选择该假设的原因	(若未按照指南推荐的优先顺序, 需额外说明)		
<b>4 数据集选取</b>			
是否采用原始数据?	<input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否	是否对样本量进行修改?	<input type="checkbox"/> 是(说明原因) <input checked="" type="checkbox"/> 否
若修改, 报告原文样本量大小和修改后的样本量大小		若修改, 报告使用 G-power 计算的修改后的样本量对应的效应量	

## 1.2 文献介绍

### 1.2.1 研究背景

在心理学研究中, 鼠标追踪技术已成为探究认知过程的重要工具, 通过记录鼠标轨迹来揭示决策过程中的实时认知动态。然而, 该领域缺乏统一的实验设计标准, 不同研究中响应指示、鼠标灵敏度和启动程序等设计因素的差异可能导致数据偏差, 进而影响对认知理论的推断。尽管鼠标追踪技术已广泛应用于社会认知、动作控制、数值认知等多个领域, 但设计因素对数据的具体影响尚未得到系统实证检验。

### 1.2.2 主要研究问题

本研究系统考察鼠标追踪 (mouse-tracking) 实验中的三大核心设计因素对鼠标追踪数据的影响: 响应指示 (点击 vs. 触摸) 对轨迹曲率和典型性效应的影响; 鼠标灵敏度 (默认 vs. 低速) 对轨迹形状和冲突指标的影响; 启动程序 (static vs. rtmax vs. initmax vs. dynamic)

对轨迹双模态分布及典型性效应的影响。

### 1.2.3 研究假设

反应指示方式：假设点击条件会比触摸条件产生更大的轨迹曲率效应（如典型性效应），因为点击条件允许更极端的轨迹（如“改变主意”轨迹）。

鼠标灵敏度：假设低速设置（禁用加速）能更好地捕捉认知效应，因为手部运动与光标移动的线性关系更明确。

启动程序：假设动态启动（dynamic）和限制启动时间的程序（如 initmax）会增强典型性效应，因为这些程序强制早期运动初始化，从而更完整地反映决策过程。

### 1.2.4 实验三各变量含义

自变量 1：启动程序，包含四种具体操作方式：static 是最基础的条件，被试点击开始按钮后刺激立即呈现，可以自由选择响应时间；rtmax 在静态启动基础上增加了 2.5 秒的响应时限，促使被试加快决策速度；initmax 要求被试在刺激呈现后 0.6 秒内必须开始移动鼠标，确保早期运动参与；dynamic 则采用行为触发机制，只有被试将鼠标向上移动超过 50 像素后才会呈现刺激，强制运动与认知加工的同步性。

自变量 2：刺激类型，根据刺激的典型性分为两类：典型样例是与所属类别高度一致的原型代表（如狮子作为哺乳动物），其分类过程快速且自动化；非典型样例则具有跨类别特征（如鲸鱼同时具有哺乳动物和鱼类的特征），会引发显著的认知冲突。这两类刺激的对比能够有效分离自动化加工和控制性加工的不同认知过程。

因变量主要包含三类测量指标。因变量 1：轨迹曲率，通过最大绝对偏差（MAD）量化，计算实际轨迹与起点-终点连线的最大垂直距离，反映决策过程中的冲突强度。因变量 2：轨迹形态特征，通过将每条轨迹匹配到五种原型（直线、弯曲、连续改变主意、离散改变主意、双重改变主意）进行分类，揭示不同的决策动态模式。因变量 3：行为指标，包括反应时、正确率和运动初始化时间。

### 1.2.5 研究结果

实验一：点击响应比触摸响应产生更大的典型性效应（ $d_z = 0.61$  vs.  $0.36$ ），轨迹更易出现“离散改变主意”模式。点击条件下曲率分布呈双峰（ $BC > 0.555$ ），触摸条件下呈单峰；

实验二：鼠标灵敏度对典型性效应大小无显著影响，但默认设置下极端轨迹（dCoM）更多；慢速条件下直线轨迹比例增加，默认条件下离散改变主意轨迹更多；

实验三：initmax 条件和 rtmax 条件的正确率显著低于 static 组。initmax 组和 rtmax 组的 RTinit 显著短于 static 组；rtmax 组、initmax 组、dynamic 组的启动时间显著短于 static 组，这说明启动程序能有效操控动作触发速度。在所有实验条件下，非典型样例的聚合轨迹相较于典型样例更大程度地向未被选择选项偏离。与静态和反应时间最大化条件相比，dynamic starting 和启动限制条件通常导致垂直向上运动的时间延长。此外，典型性效应在全部四种实验条件下均显著，非典型样本的 MAD 值显著高于典型样本，在 initmax 条件下表现得尤

为明显 ( $d_z=0.91$ )。静态启动 (static) 和总反应时限制 (rtmax) 条件下的鼠标轨迹以直线轨迹为主, 动态启动 (dynamic) 和 启动限制 (initmax) 条件下曲线与改变主意轨迹更常见。

### 1.2.6 研究结论

响应指示、鼠标灵敏度和启动程序对轨迹曲率和单个轨迹的形状有显著影响。这些差异可能会导致理论构建的偏差, 进而对某些理论的支持或反对做出过早的结论。在解释鼠标追踪数据以测试理论时, 以及在规划鼠标追踪研究时, 必须仔细考虑具体的设计。

## 2 方法

### 2.1 样本

实验三纳入 245 名参与者 (162 女, 83 男), 平均年龄 21.9 岁 ( $SD=3.3$ ), 随机分配至四个起始程序组: static ( $n=59$ )、rtmax ( $n=60$ )、initmax ( $n=66$ ) 和 dynamic ( $n=60$ )。原文献通过 G\*Power 计算样本量, 确保检测中等效应量 ( $d_z=0.5$ ) 的统计功效  $\geq 0.95$ 。参与者均为德国大学生, 右手优势者占 69.8% (EHI 得分  $>60$ )。

### 2.2 原研究方法简介

#### 2.2.1 原研究设计

##### ①研究设计

实验采用 4 (启动程序: static、rtmax、initmax、dynamic)  $\times$  2 (刺激类型: 典型、非典型) 的混合设计。自变量为启动程序和刺激类型, 其中启动程序为被试间变量, 刺激类型为被试内变量, 因变量为聚合轨迹曲率和轨迹形状分布。

##### ②数据分析方法

卡方检验、描述性统计、回归分析 (广义线性混合模型、有序混合回归模型)、t 检验、方差分析 (单因素方差分析、重复测量方差分析)、事后检验 (对比分析)

##### ③使用的软件及软件包

数据采集: OpenSesame、数据分析: R 语言

软件包如下:

mousetrap: 用于处理和分析鼠标轨迹数据。

ggplot2: 提供灵活的语法进行数据可视化。

dplyr: 简化数据操作 (如筛选、排序、汇总)。

tidyr: 帮助整理和重塑数据结构 (如长宽格式转换)。

afex: 支持重复测量 ANOVA 等进阶统计分析。

MBESS: 计算效应量 (如 Cohen's  $d$ ) 和置信区间。

ordinal: 拟合有序逻辑回归模型 (如比例优势模型)。

readxl: 读取 Excel 文件 (.xls/.xlsx) 到 R 环境。

emmeans: 进行事后检验和边际均值估计 (如 ANOVA 后续分析)

#### 2.2.2 小组复现内容

本小组的复现工作主要针对原研究的第三个研究, 本小组的复现工作主要有以下四部分:

①比较四个实验条件（静态启动 static，总时间限制 rtmax，启动限制 initmax，动态启动 dynamic）下正确率及其差异；

②对以下三个变量在不同的实验条件下进行操作检验：RT<sub>init</sub>：被试将鼠标上移 50 像素所需时间——反映参与者多快完成有效初始移动；initiation time：即试验中首次动作响应时间，是参与者动作触发速度；RT：总反应时；

③完成轨迹聚合的可视化和曲率数据聚合，针对 4（组别）×2（刺激类型）的不同条件下的数据进行描述性统计检验和配对样本 t 检验，完成 MAD 均值差异的可视化并进行重复测量方差分析，探究其主效应与交互效应是否显著并进行进一步的对比分析；

④轨迹形状分布分析：计算参与者在每个实验刺激类型（典型/非典型）和实验条件（4 种启动程序）下标准化 MAD 值的双峰系数；将所有参与者的鼠标轨迹映射到前人提出的 5 种原型轨迹，量化和统计检验不同启动程序条件和不同实验刺激下 5 种原型轨迹相对频率的差异；使用有序混合回归模型（ordinal mixed regression）评估不同起始程序以及实验刺激的典型性如何影响参与者鼠标轨迹类型的分布。

## 2.3 重复思路与 R 包

### 2.3.1 准备工作

获取必要的软件和工具：安装并设置好 R、RStudio。下载数据和数据分析代码：访问 <https://osf.io/xdp7a/files/osfstorage>。

### 2.3.2 数据理解和预处理

了解数据结构：仔细阅读研究论文的“方法”部分，了解数据的来源、格式和变量定义。打开下载的数据文件，查看数据的整体结构和内容，确保对每个变量的含义有清晰的理解。

导入数据到 R，并进行数据预处理，主要处理内容包括自定义 ggplot2 主题以及自定义函数，部分代码及注解如下：

```
# 设置自定义ggplot2主题，使图表更美观
theme_set(theme_classic()+
  theme(
    axis.line = element_line(colour = "black"), # 设置坐标轴为黑色
    axis.ticks = element_line(colour = "black"), # 设置刻度线为黑色
    axis.text = element_text(colour = "black"), # 设置坐标轴文本为黑色
    panel.border = element_rect(colour = "black", fill=NA) # 设置面板边框为黑色
  ))

# 自定义函数：计算部分eta平方及其置信区间
get_partial_etas <- function(anova_table, conf.level=.90){
  partial_etas <- sapply(row.names(anova_table),function(i){
    F <- anova_table[i,"F"] # 提取F值
    df1 <- anova_table[i,"num Df"] # 提取分子自由度
    df2 <- anova_table[i,"den Df"] # 提取分母自由度
    ci <- conf.limits.ncf(F.value=F,conf.level=conf.level,df.1=df1,df.2=df2) # 计算非中心
    return(
      c(pes=((F*df1)/(F*df1+df2)), # 计算部分eta平方
        lower=ci$Lower.Limit/(ci$Lower.Limit+df1+df2+1), # 计算置信区间下限
        upper=ci$Upper.Limit/(ci$Upper.Limit+df1+df2+1))) # 计算置信区间上限
    )
  })
  return(t(partial_etas)) # 转置结果并返回
}

# 导入实验数据
raw_data <- read.csv("../data/exp3.csv") # 从CSV文件读取数据
raw_data$typicality <- factor(raw_data$Condition,levels=c("Typical","Atypical")) # 创建T
raw_data$group <- factor(raw_data$group,levels=c("static","rtmax","initmax","dynamic"))
```

### 2.3.3 设置分析环境

在 RStudio 中设置项目：打开 RStudio，创建一个新的项目，并将下载的数据和代码文件导入项目目录中。安装必要的 R 包：在 RStudio 的控制台输入并运行以下代码，安装所有必要的 R 包：dplyr、tidyr、ggplot2、mousetrap、afex、MBESS、ordinal、emmeans，部分代码及注解如下：

```
library(mousetrap) # 用于分析鼠标轨迹数据
library(ggplot2)   # 用于数据可视化
library(dplyr)     # 用于数据处理和转换
library(tidyr)     # 用于数据整理
library(afex)      # 用于高级统计分析
library(MBESS)     # 用于效应量计算
library(ordinal)   # 用于有序回归模型
library(emmeans)
```

注意，此处原文献给出的代码中并未加载 emmeans 包，导致代码在进行后续方差分析和对比分析时会报错，因此此处我们复现时加上了一些 R 包的加载。

### 2.3.4 运行数据分析代码

打开并运行 RMarkdown 文件。在 RStudio 中打开下载的 RMarkdown 文件 (.Rmd)，逐块运行代码，或点击“Run”按钮运行整个 RMarkdown 文件。确保每个代码块都成功运行，并记录下每步的输出结果。

1、正确率检验：对四种实验条件下的正确率进行描述性统计及卡方检验，并拟合广义线性混合模型，分析组别对正确率的影响，部分代码及注释如下：

```
# 分析正确率（包括所有试验）
with(raw_data, table(group, correct)/c(table(group))) # 计算每个组的正确率

# 进行卡方检验，分析组别与正确率的关系
chisq.test(with(raw_data, table(group, correct)), correct = FALSE)

# 拟合广义线性混合模型，分析组别对正确率的影响
# 使用默认的对比如编码（哑变量编码，以static组为基线）
contrasts(raw_data$group) # 显示group变量的对比编码
summary(glmer(correct~(1|subject_nr)+group, family="binomial", data=raw_data)) # 拟合模型并

# 分析正确率（排除rtmax组中超过时间限制的试验）
# 统计rtmax组中符合条件的和不符合条件的试验数量
n_eligible <- sum(with(raw_data, group=="rtmax" & response!="None")) # 符合条件的试验数
n_noneligible <- sum(with(raw_data, group=="rtmax" & response=="None")) # 不符合条件的试验
```

2、操作检验：对三个变量在不同不同实验条件下的数据进行描述性统计和方差分析及事后检验，部分代码及注释如下：

```
# 计算描述性统计量
mean_times <- agg_times %>%
  group_by(group) %>%
  summarize(
    N = n(), # 每组样本量
    M_RT_initial = mean(RT_initial), # 初始阶段平均反应时间
    SD_RT_initial = sd(RT_initial), # 初始阶段反应时间标准差
    M_IT = mean(IT), # 平均启动时间
    SD_IT = sd(IT), # 启动时间标准差
    M_RT = mean(RT), # 平均总反应时间
    SD_RT = sd(RT) # 总反应时间标准差
  ) %>%
  as.data.frame()
print(mean_times, digits=5) # 打印描述性统计结果
```



```
# 比较初始阶段反应时间
# 进行方差分析
anova_RT_initial <- aov_ez(data=agg_times,dv = "RT_initial", between = "group", id = "subject_nr") # 拟合方差分析模型
nice(anova_RT_initial,es = c("pes","ges")) # 显示方差分析结果和效应量

# 计算部分eta平方及其90%置信区间
round(get_partial_etas(anova_RT_initial$anova_table, conf.level=.90),2)

# 进行对比分析
anova_RT_initial_grid <- lsmeans(anova_RT_initial,~group) # 计算最小二乘均值
contrast(anova_RT_initial_grid,contrast_matrix_separate) # 执行指定的对比
```

3、聚合轨迹曲率：首先完成轨迹聚合可视化 and 曲率数据聚合，即绘制不同组别（group）和刺激类型（typicality）的平均运动轨迹，按被试 ID 聚合数据；再按被试（subject\_nr）和条件（typicality、group）聚合轨迹曲率（MAD）。其次，按组别（group）和刺激类型（typicality）分组，计算每组数据的  $M$  均值、 $SD$  标准差，以及配对样本  $t$  检验、 $p$  值并报告 Cohen's  $d$  值，此外还要完成 MAD 均值差异的可视化。再者，进行重复测量方差分析，其中 MAD 为因变量、启动程序为组间变量，典型性为组内变量，探究其主效应和交互效应是否显著，并进一步进行对比分析，部分代码及注释如下：

```
# 使用mt_plot_aggregate函数绘制聚合轨迹图
mt_plot_aggregate(
  mt_data, # 输入数据集（轨迹数据）
  use = "tn_trajectories", # 指定使用的轨迹数据列
  facet_col = "group", # 按"group"列分面（分组显示）
  x = "xpos", # x轴对应位置坐标
  y = "ypos", # y轴对应位置坐标
  color = "Typicality", # 按"Typicality"设置颜色（区分典型/非典型）
  subject_id = "subject_nr" # 指定被试ID列（用于聚合计算）
) +

# 添加坐标轴标签
xlab("x coordinate (px)") + # x轴标签（像素坐标）
ylab("y coordinate (px)") + # y轴标签（像素坐标）

# 自定义颜色映射
scale_color_manual(
  values = c("black", "grey60") # 设置颜色值：第一类黑色，第二类灰60%
)

2. 按被试聚合轨迹曲率数据
# 计算每个被试在不同条件下的平均轨迹曲率(MAD)
agg_mad <- mt_aggregate_per_subject(mt_data, subject_id = "subject_nr",
  use_variables = "MAD", use2_variables = c("Typicality","group"))

3. 描述性统计与配对t检验
# 按组别计算描述性统计并进行配对t检验
mad_table <- agg_mad %>%
  group_by(group) %>%
  select(MAD,group,Typicality) %>%
  summarize(
    N = length(MAD[Typicality=="Typical"]), # 样本量
    M_t = mean(MAD[Typicality=="Typical"]), # 典型刺激平均MAD
    SD_t = sd(MAD[Typicality=="Typical"]), # 典型刺激标准差
    M_a = mean(MAD[Typicality=="Atypical"]), # 非典型刺激平均MAD
    SD_a = sd(MAD[Typicality=="Atypical"]), # 非典型刺激标准差
    t = t.test(MAD[Typicality=="Atypical"],MAD[Typicality=="Typical"],paired=TRUE)$statistic, # t统计量
    p = t.test(MAD[Typicality=="Atypical"],MAD[Typicality=="Typical"],paired=TRUE)$p.value, # p值
    d = (M_a-M_t)/sd(MAD[Typicality=="Atypical"]-MAD[Typicality=="Typical"]) # Cohen's d效应量
```

```

# 打印结果表格（保留3位小数）
mad_table %>%
  as.data.frame() %>%
  print(digits=3)
4. MAD均值差异图
# 绘制不同组别和刺激类型的MAD均值比较图
ggplot(agg_mad,aes(x=Typicality,y=MAD,linetype=group,group=group))+
  geom_line(stat="summary",fun.y="mean")+ # 连接均值点
  geom_point(stat="summary",fun.y="mean")+ # 显示均值点
  geom_errorbar(stat="summary",fun.data="mean_se",width=.2,linetype=1)+ # 添加标准误差线
  scale_linetype_manual(values=c(1,2,3,4))+ # 设置线型
  coord_cartesian(ylim=c(0,700)) # 设置y轴范围
5. 混合设计方差分析(ANOVA)
# 执行混合设计ANOVA（组间变量：group，组内变量：Typicality）
anova_mad <- aov_ez(data=agg_mad, dv = "MAD", between = "group", within = "Typicality",
  id = "subject_nr")
# 输出ANOVA结果（含效应量）
nice(anova_mad,es = c("pes","ges"))

# 计算偏 $\eta^2$ 的90%置信区间
round(get_partial_etas(anova_mad$anova_table, conf.level=.90),2)
6. 对比分析
# 获取最小二乘均值网格
anova_mad_grid <- lsmeans(anova_mad,~Typicality:group)

# 定义对比矩阵
contrast_matrix_complete <- list(
  typicality_static = c(-1,1,0,0,0,0,0,0), # 静态组内典型性对比
  rtmax_static_main= c(-1,-1,1,1,0,0,0,0)/2, # RTmax组与静态组主效应
  initmax_static_main = c(-1,-1,0,0,1,1,0,0)/2, # Initmax组与静态组主效应
  dynamic_static_main = c(-1,-1,0,0,0,0,1,1)/2, # 动态组与静态组主效应
  rtmax_static_int = c(1,-1,-1,1,0,0,0,0), # RTmax组与静态组的交互
  initmax_static_int = c(1,-1,0,0,-1,1,0,0), # Initmax组与静态组的交互
  dynamic_static_int = c(1,-1,0,0,0,0,-1,1)) # 动态组与静态组的交互

# 执行对比检验
contrast(anova_mad_grid,contrast_matrix_complete)

```

#### 4、轨迹形状分布

（1）鼠标追踪范式中的双系统决策模型预测会出现低冲突试次与‘最初某选项会被强烈激活，随后发生态度转变’试次的混合情况；这将导致直线轨迹与运动路径中突然转向的轨迹并存现象。‘改变主意’轨迹会产生由系统 I 驱动倾向于某一选项（导致鼠标向该选项移动），随后被系统 II 推翻并选择另一选项。这类改变主意的反应会生成较大曲率值，而直线轨迹则产生较小曲率值。因此，综合所有试次数据，预计会出现小曲率值与大曲率值的双峰分布。根据仿真研究，0.555 被用作判定双峰分布的临界值（参见 Freeman & Ambady, 2010; Freeman & Dale, 2013），结果大于 0.555 提示数据分布可能为双峰分布。

因此，研究人员计算了参与者在每个实验刺激类型（典型/非典型）和实验条件（4 种启动程序）的标准化 MAD 值的双峰系数，旨在揭示不同实验设置下是否会因设计因素影响而表现出不同的认知过程模式，为快速验证轨迹分布形态提供了一种标准化、可量化的指标，便于跨实验比较。代码及注释如下：

```

## Bimodality coefficient
{r}
# 将每一个参与者的MAD值标准化（在被试内进行）
mt_data <- mt_standardize(mt_data, use_variables = "MAD", within = "subject_nr")

# 计算双峰系数（双峰性分析的结果，Bimodality coefficient简称BC，BC大于0.555提示标准化后的MAD值数据分布可能为双峰分布，即小曲率值和大曲率值并存）
mt_check_bimodality(mt_data, use_variables = "z_MAD",
  grouping_variables = c("group","Typicality"), methods = "BC")

```

（2）为量化和统计检验不同启动程序条件下轨迹类型频率的差异，将实验 3 收集到的轨迹映射至实验 1 和 2 中使用的同一组原型集上。再通过卡方检验，分析实验组和轨迹原型



分类之间的关联性。部分代码及注释如下，此部分改动了 3 处：

绘制轨迹原型的代码运行时出现“warning: The argument only\_ggplot is deprecated. Please use return\_type instead and consult the function documentation for its enhanced functionality.”，表示 `mt_plot()` 函数中的参数 `only_ggplot` 已过时，需改用 `return_type`。这是开发者为统一参数命名做的更新，功能完全相同。因此将 `mt_plot(..., only_ggplot = TRUE)` 改为 `mt_plot(..., return_type = "ggplot")`

```
### Plot prototypes
```{r,fig.height=1.5}
mt_plot(mt_prototypes,facet_col="mt_id",only_ggplot= TRUE)+
  geom_path()+
  facet_grid(cols = vars(factor(mt_id,levels=rownames(mt_prototypes))))+
  theme(axis.text=ggplot2::element_blank(),axis.ticks=ggplot2::element_blank())
```

```{r,fig.height=1.5}
mt_plot(mt_prototypes,facet_col="mt_id",return_type = "ggplot")+
  geom_path()+
  facet_grid(cols = vars(factor(mt_id,levels=rownames(mt_prototypes))))+
  theme(axis.text=ggplot2::element_blank(),axis.ticks=ggplot2::element_blank())
```
```

由于运动通常以恒定的采样率进行追踪，不同持续时间的 `trial` 会导致轨迹长度不同。在将轨迹映射到原型上时，需要将轨迹进行长度归一化（将轨迹插值为数量相同的等间距空间数据点。整个试次过程中的空间分辨率保持恒定，强调轨迹的形状，而不受 `trial` 中运动速度变化的影响）时，出现了 Warning: `mt_spatialize` 不再有用，改用 '`mt_length_normalize`'。

```
### Map trajectories onto prototypes 将轨迹映射到原型上
```{r}
mt_data <- mt_length_normalize(mt_data) #对原始轨迹数据进行长度标准化
mt_data <- mt_map(mt_data,prototypes = mt_prototypes,
  save_as = "measures", grouping_variables = "group")
#轨迹与原型映射，将标准化后的轨迹与原型轨迹进行相似性比对，生成量化指标。
mt_data$measures$prototype_label <- mt_data$measures$prototype_label
#将映射结果中的原型分类标签从 measures 层提取到主数据层 (data)，便于后续分析或可视化。
```
```

先计算整体分类频率，回答不同启动程序是否改变轨迹分布模式：

```
421- ### 计算每组的分类轨迹
422-
423- ### 计算相对频率
424- ```{r,fig.height=4.5}
425- prototype_percentages <- mt_data$measures %>% #计算各组原型轨迹的百分比，从主数据对象中提取轨迹数据
426-   group_by(group,prototype_label) %>% #按实验分组和原型标签分组
427-   summarise(n=n()) %>% #计算每个组合的轨迹数量
428-   mutate(Percent=paste(round(100*n/sum(n)),"%",sep="")) # 计算百分比并格式化为字符串
429-
430-
431- mt_plot(mt_data, use = "sp_trajectories", # 使用空间化轨迹数据
432-   x = "xpos", y = "ypos", # 指定坐标列
433-   facet_col = "prototype_label", #列分面：按原型分类（将图表按照prototype_label分成多列显示，每列展示该类别下的所有数据）
434-   facet_row="group", #行分面：按实验组
435-   alpha=.2)+ # 设置透明度，使轨迹半透明，避免过度重叠
436-   xlab("x coordinate (px)") + # 设置X轴标签
437-   ylab("y coordinate (px)") + # 设置Y轴标签
438-   geom_text(data=prototype_percentages, # 添加百分比标签
439-     aes(label=Percent),x=650,y=50)+
440-   scale_y_continuous(breaks=c(0,500,1000))+ # 标签定位（需根据坐标范围调整）
441-   coord_cartesian(xlim=c(-900,900)) # 控制显示范围
442-
443-
444- ```{r}
445- chisq.test(with(mt_data$measures,table(group, prototype_label))) #生成一个列联表（contingency table），统计每个 group 和 prototype_label 组合的频数
446-```
```

注意，此处原文献给出的代码中用的是“`sp_trajectories`”，但运行代码时出现报错“Warning in `extract_data(data = data, use = use)` :Instead of `sp_trajectories` only `ln_trajectories` were found. These will be used. Please note that since the `mt_spatialize` function has been replaced with the `mt_length_normalize` function, the resulting trajectories are now called `ln_trajectories`”。

instead of `sp_trajectories`.

Error in `dataset[, mt_id] <- rownames(dataset)`: 矩阵里的下标数目不对

收卷时出错: 堆叠上没有这么多的环境框

Error: no more error handlers available (recursive errors?); invoking 'abort' restart”。因此将“`sp_trajectories`”改为“`ln_trajectories`（长度标准化轨迹）”，这段代码便能正常运行。

(3) 计算各实验组在不同典型性条件下轨迹分类的相对频率，揭示典型性效应如何受实验条件调节，进行描述性统计，部分代码及注释如下：

```
### 各实验组在不同典型性条件下的轨迹分类

### 计算相对频率
library(ggplot2)
set.seed(123)
r <- matrix(rnorm(1000), nrow=10, ncol=100)
mt_data <- data.frame(r, group=rep(1:10, each=100))
rel_freq_agg <- mt_data %>%
  group_by(group, Typicality, prototype_label) %>%
  summarise(n=n()) %>%
  mutate(Percent=n/sum(n)) %>%
  spread(rel_freq_agg[, -4], "prototype_label", "Percent", fill = 0) %>%
  as.data.frame() %>%
  print(digits=2)

ggplot(rel_freq_agg, aes(x=Typicality, y=Percent,
  fill=forcats::fct_rev(prototype_label))) +
  geom_bar(stat="identity", color="black") +
  scale_fill_brewer(type="seq", name="Classification") +
  facet_grid(.~group)

# 设置 x 轴为 Typicality (典型/非典型), y 轴为 Percent;
# 按 prototype_label 填色 (fill); 使用 forcats::fct_rev() 把因子的显示顺序反转, 使 legend
# 顺序美观 (如从高频到低频)。
# 创建柱状图: 直接使用 Percent 的实际值作为高度; 直接使用 Percent 的实际值作为高度;
# 表示图例标题为 "Classification"
# 用 facet_grid() 创建子图, 每个实验组 (group) 作为一个子图水平排列
```

#### (4) 通过有序混合回归模型预测轨迹类型

因为轨迹原型 (prototype) 本质上是有序的分类变量, 研究者用有序混合回归模型中预测不同实验条件和不同刺激下的轨迹类型 (如 straight、curved、cCoM、dCoM、dCoM2 等原型分类), 控制被试间的个体差异, 并通过累积比数模型 (cumulative logit model) 估计每个原型类别的概率。旨在检验以下假设: ①典型性效应 (typicality): 非典型刺激 (atypical exemplars) 是否会导致更极端的轨迹类型 (如更多 dCoM 轨迹)? ②实验条件效应 (starting procedure): 不同启动程序 (如 dynamic vs. static) 是否会影响轨迹类型的分布。

由于模型本身没有  $z$  值, 而文中报告的全是  $z$  值和  $p$  值应该是显著性检验的结果, 经过查阅了解到: 有序混合回归分析属于广义线性混合模型 (GLMM), 其参数估计通常通过最大似然估计 (MLE) 或限制性最大似然估计 (REML) 完成。在混合模型中, 自由度 (df) 的计算复杂且存在争议 (尤其是对随机效应)。 $z$  检验是更保守且广泛接受的选择, 尤其在大样本条件下 (如实验 3 的 245 名参与者), MLE 的估计量服从渐进正态分布,  $t$  分布近似正态分布,  $z$  值与  $t$  值的差异可忽略, 因此回归系数  $\beta$  的显著性可通过 Wald 检验计算:

$$z = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

其中,  $\hat{\beta}$  是系数估计值, SE 是标准误。 $z$  值表示系数偏离 0 的标准误倍数,  $p$  值则基于标准正态分布计算。

部分代码及注释如下:

```
####有序混合回归模型
```{r}
contrasts(mt_data$data$Typicality) <- c(-0.5,0.5)
#为分类变量Typicality设置效应编码，表示将两个水平编码为-0.5和0.5
# use default contrasts for group (dummy coding with static as baseline)
contrasts(mt_data$data$group) #检查 group 变量的对比编码
summary(cglmm(prototype_label~Typicality*group+(1|subject_nr),data=mt_data$data))
#拟合一个有序混合效应模型（CLMM, Cumulative Link Mixed Model），并输出模型摘要
```
```

2.3.5 结果对比和验证

对照三人的输出结果。将分析结果与研究论文中报告的结果进行详细对比，确保各项结果一致，并根据原代码复现图表，验证其准确性。

调试和修正。如果发现任何不一致，仔细检查数据预处理、代码实现和分析步骤，找出可能的原因并进行修正。

3 结果

3.1 不同实验条件下正确率检验

3.1.1 不同实验条件下正确率检验的描述性统计

比较四个实验条件（静态启动 static，总时间限制 rtmax，启动限制 initmax，动态启动 dynamic）下正确率的描述性统计结果，见表 1。

| 表 1 四种条件下正确率的描述性统计 |        |       |         |         |
|--------------------|--------|-------|---------|---------|
|                    | static | rtmax | initmax | dynamic |
|                    | 正确率    | 正确率   | 正确率     | 正确率     |
| 原研究                |        |       |         |         |
| 报告结果               | 94.1%  | 89.1% | 89.7%   | 93.6%   |
| 本研究                | 94.1%  | 89.1% | 89.7%   | 93.6%   |
| $\delta$           | 0      | 0     | 0       | 0       |
| 评级                 | 完全一致   | 完全一致  | 完全一致    | 完全一致    |

3.1.2 不同实验条件下正确率检验的推论性统计

对不同实验条件下的正确回答试次数进行卡方检验发现，不同实验条件下的正确率存在显著差异，具体结果见表 2。

为对比不同条件的影响，采用二项联结函数（用于处理二元分类变量或比例数据的数学工具）在试次层面建立了广义线性混合模型，并包含每位被试的随机截距。起始条件作为预

测变量，其中 static 作为基准进行虚拟编码。Wald Z 检验的结果见表 3。

表 2 四种条件下正确率的卡方检验结果

|          | $\chi^2$ | $p$    |
|----------|----------|--------|
| 原文献      | 29.93    | < .001 |
| 报告结果     |          |        |
| 本研究      | 29.93    | < .001 |
| $\delta$ | 0        | 0      |
| 评级       | 完全一致     | 完全一致   |

表 3 四种条件下的 Wald Z 检验结果

|          |                    | $Z$   | $p$   |
|----------|--------------------|-------|-------|
| 原文献      | dynamic vs. static | -0.42 | 0.673 |
|          | initmax vs. static | -3.12 | 0.002 |
| 报告结果     | rtmax vs. static   | -1.18 | 0.237 |
|          | dynamic vs. static | -0.42 | 0.673 |
| 本研究      | initmax vs. static | -3.12 | 0.002 |
|          | rtmax vs. static   | -1.18 | 0.237 |
| $\delta$ |                    | 0     | 0     |
| 评级       |                    | 完全一致  | 完全一致  |

3.2 不同实验条件下的操纵检验

3.2.1 不同实验条件下的操纵检验的描述性统计

在本研究的操纵检验中，选取了三个变量进行分析，分别为（1）RTinit：被试将鼠标上移 50 像素所需时间——反映参与者多快完成有效初始移动；（2）initiation time：即试验中

首次动作响应时间，是参与者动作触发速度；（3）RT：总反应时

对不同实验条件下以上三个变量的描述性检验结果，具体结果见表 4。

表 4 操纵检验的描述性统计的结果对比

|          | static             |                  |                    | rtmax              |                  |                    | initmax            |                  |                    | dynamic            |                  |                      |
|----------|--------------------|------------------|--------------------|--------------------|------------------|--------------------|--------------------|------------------|--------------------|--------------------|------------------|----------------------|
|          | RT <sub>init</sub> | Initiation Time  | RT                 | RT <sub>init</sub> | Initiation Time  | RT                 | RT <sub>init</sub> | Initiation Time  | RT                 | RT <sub>init</sub> | Initiation Time  | RT                   |
| 原研究报告结果  | 808.5<br>(324.1)   | 508.7<br>(215.6) | 2,110.4<br>(654.1) | 650.1<br>(176.6)   | 437.3<br>(160.0) | 1,521.6<br>(183.4) | 377.4<br>(159.5)   | 243.1<br>(142.8) | 1,471.7<br>(248.6) | 773.4<br>(752.2)   | 348.7<br>(233.2) | 2,805.4<br>(1,199.8) |
| 本研究      | 808.5<br>(324.1)   | 508.7<br>(215.6) | 2,110.4<br>(654.1) | 650.1<br>(176.6)   | 437.3<br>(160.0) | 1,521.6<br>(183.4) | 377.4<br>(159.5)   | 243.1<br>(142.8) | 1,471.7<br>(248.6) | 773.4<br>(752.2)   | 348.7<br>(233.2) | 2,805.4<br>(1,199.8) |
| $\delta$ | 0                  | 0                | 0                  | 0                  | 0                | 0                  | 0                  | 0                | 0                  | 0                  | 0                | 0                    |
| 评级       | 完全一致               | 完全一致             | 完全一致               | 完全一致               | 完全一致             | 完全一致               | 完全一致               | 完全一致             | 完全一致               | 完全一致               | 完全一致             | 完全一致                 |

```
# 打印描述性统计结果
print(mean_times, digits=5)
group N M_RT_initial SD_RT_initial M_IT SD_IT M_RT SD_RT
static 59 808.47 324.06 508.69 215.60 2110.4 654.06
rtmax 60 650.12 176.61 437.25 159.98 1521.6 183.42
initmax 66 377.43 159.49 243.14 142.76 1471.7 248.62
dynamic 60 773.38 752.17 348.67 233.25 2805.4 1199.82
```

图 1 R 代码输出的描述性统计结果表

Table 4. Means (and SDs) of the per-participant aggregated timing variables in Experiment 3, presented separately for each condition (in milliseconds)

| Condition | RT <sub>init</sub> | Initiation Time | RT                |
|-----------|--------------------|-----------------|-------------------|
| static    | 808.5 (324.1)      | 508.7 (215.6)   | 2,110.4 (654.1)   |
| rtmax     | 650.1 (176.6)      | 437.3 (160.0)   | 1,521.6 (183.4)   |
| initmax   | 377.4 (159.5)      | 243.1 (142.8)   | 1,471.7 (248.6)   |
| dynamic   | 773.4 (752.2)      | 348.7 (233.2)   | 2,805.4 (1,199.8) |

RT<sub>init</sub> = Time until cursor was moved 50 pixels upward

图 2 原文献中描述性统计结果表

3.2.2 不同实验条件下的操纵检验的推断性统计



对以上三个变量在不同实验条件下分别进行方差分析，并进行事后检验，结果显示不同实验条件间的  $RT_{init}$ 、initiation time 和 RT 均存在显著差异，方差分析与对比检验结果见表 5-8。

表 5 操纵检验的推断性统计结果的比较

|          | 启动程序 vs $RT_{init}$ |       |            |             | 启动程序 vs initiation time |       |            |             | 启动程序 vs RT |       |            |             |
|----------|---------------------|-------|------------|-------------|-------------------------|-------|------------|-------------|------------|-------|------------|-------------|
|          | $F$                 | $p$   | $\eta p^2$ | 90%CI       | $F$                     | $p$   | $\eta p^2$ | 90%CI       | $F$        | $p$   | $\eta p^2$ | 90%CI       |
| 原研究      |                     |       |            |             |                         |       |            |             |            |       |            |             |
| 报告结果     | 13.64               | <.001 | 0.15       | [0.08,0.21] | 22.69                   | <.001 | 0.22       | [0.14,0.29] | 49.61      | <.001 | 0.38       | [0.30,0.44] |
| 本研究      | 13.64               | <.001 | 0.15       | [0.08,0.21] | 22.69                   | <.001 | 0.22       | [0.14,0.29] | 49.61      | <.001 | 0.38       | [0.30,0.44] |
| $\delta$ | 0                   | 0     | 0          | 0           | 0                       | 0     | 0          | 0           | 0          | 0     | 0          | 0           |
| 评级       | 完全一致                | 完全一致  | 完全一致       | 完全一致        | 完全一致                    | 完全一致  | 完全一致       | 完全一致        | 完全一致       | 完全一致  | 完全一致       | 完全一致        |

表 6 不同启动程序间  $RT_{init}$  的对比分析的推断性统计结果比较

| 启动程序 vs $RT_{init}$ |                    | $t$   | $p$   |
|---------------------|--------------------|-------|-------|
| 原文献<br>报告结果         | initmax vs. static | -5.70 | <.001 |
|                     | rtmax vs. static   | -2.05 | .042  |
|                     | dynamic vs. static | -0.45 | .651  |
| 本研究                 | initmax vs. static | -5.70 | <.001 |
|                     | rtmax vs. static   | -2.05 | .042  |
|                     | dynamic vs. static | -0.45 | .651  |
| $\delta$            |                    | 0     | 0     |
| 评级                  |                    | 完全一致  | 完全一致  |

表 7 不同启动程序间启动时间的对比分析的推断性统计结果比较

| 启动程序 vs initiation time |                    | $t$   | $p$   |
|-------------------------|--------------------|-------|-------|
| 原文献                     | initmax vs. static | -7.78 | <.001 |
| 报告结果                    | rtmax vs. static   | -2.05 | .042  |

|          |                    |       |       |
|----------|--------------------|-------|-------|
| 本研究      | dynamic vs. static | -4.58 | <.001 |
|          | initmax vs. static | -7.78 | <.001 |
|          | rtmax vs. static   | -2.05 | .042  |
|          | dynamic vs. static | -4.58 | <.001 |
| $\delta$ |                    | 0     | 0     |
| 评级       |                    | 完全一致  | 完全一致  |

表 8 不同启动程序间 RT 的对比分析的推断性统计结果比较

| 启动程序 vs RT |                    | <i>t</i> | <i>p</i> |
|------------|--------------------|----------|----------|
| 原文献        | initmax vs. static | -5.14    | <.001    |
|            | rtmax vs. static   | -4.63    | <.001    |
| 报告结果       | dynamic vs. static | 5.47     | <.001    |
|            | initmax vs. static | -5.14    | <.001    |
| 本研究        | rtmax vs. static   | -4.63    | <.001    |
|            | dynamic vs. static | 5.47     | <.001    |
| $\delta$   |                    | 0        | 0        |
| 评级         |                    | 完全一致     | 完全一致     |

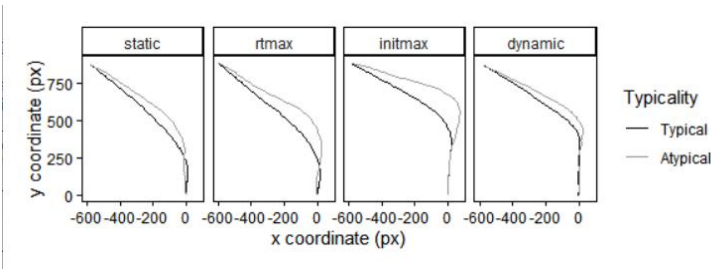
### 3.3 聚合轨迹曲率

#### 3.3.1 平均时间归一化轨迹

绘制不同组别和刺激类型的平均运动轨迹：

在所有实验条件下，非典型样例的聚合轨迹相较于典型样例更大程度地向未被选择选项偏离。与 rtmax 和 rtmax 最大化条件相比，dynamic 和 initmax 条件通常导致垂直向上运动的时间延长。

此外，典型性效应在 initmax 条件下表现得尤为明显。



原研究图表：

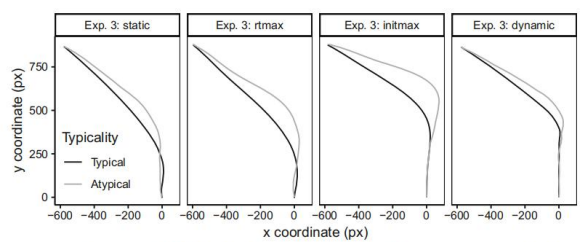


Fig. 9 Aggregate mouse trajectories for Experiment 3. All individual trajectories were flipped to the left, time-normalized, and aggregated separately per typicality and experimental condition

### 3.3.2 按参与者聚合的平均绝对偏差（MAD）比较

按被试和条件聚合数据，并进行描述性统计

表 9 描述性统计结果的比较

| Condition | N  |          | Typical |       | Atypical |       |
|-----------|----|----------|---------|-------|----------|-------|
|           |    |          | Mean    | SD    | Mean     | SD    |
| static    | 59 | 原研究      | 185.2   | 134.4 | 269.7    | 172.7 |
|           |    | 报告结果     |         |       |          |       |
|           |    | 本研究      | 185.2   | 134.4 | 269.7    | 172.7 |
|           |    | $\delta$ | 0       | 0     | 0        | 0     |
|           |    | 评级       | 完全一致    | 完全一致  | 完全一致     | 完全一致  |
| rtmax     | 60 | 原研究      | 189.8   | 150.8 | 301.5    | 197.8 |
|           |    | 报告结果     |         |       |          |       |
|           |    | 本研究      | 189.8   | 150.8 | 301.5    | 197.8 |

|         |    |          |       |       |       |       |
|---------|----|----------|-------|-------|-------|-------|
|         |    | $\delta$ | 0     | 0     | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  | 完全一致  | 完全一致  |
|         |    | 原研究      |       |       |       |       |
|         |    | 报告结果     | 304.8 | 140.7 | 470.9 | 203.2 |
| initmax | 66 | 本研究      | 304.8 | 140.7 | 470.9 | 203.2 |
|         |    | $\delta$ | 0     | 0     | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  | 完全一致  | 完全一致  |
|         |    | 原研究      |       |       |       |       |
|         |    | 报告结果     | 297.0 | 111.6 | 364.1 | 154.0 |
| dynamic | 60 | 本研究      | 297.0 | 111.6 | 364.1 | 154.0 |
|         |    | $\delta$ | 0     | 0     | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  | 完全一致  | 完全一致  |

### 3.3.3 配对样本 t 检验结果

表 10 Typical 和 Atypical 在四种条件下的  $t$  检验表

| 条件      | $t$ Test | 原研究报告结果 | 本研究    | $\delta$ | 评级   |
|---------|----------|---------|--------|----------|------|
|         | $t$      | 4.18    | 4.18   | 0        | 完全一致 |
| static  | $p$      | <0.001  | <0.001 | 0        | 完全一致 |
|         | $d$      | 0.54    | 0.54   | 0        | 完全一致 |
|         | $t$      | 4.32    | 4.32   | 0        | 完全一致 |
| rtmax   | $p$      | <0.001  | <0.001 | 0        | 完全一致 |
|         | $d$      | 0.56    | 0.56   | 0        | 完全一致 |
|         | $t$      | 7.39    | 7.39   | 0        | 完全一致 |
| initmax | $p$      | <0.001  | <0.001 | 0        | 完全一致 |
|         | $d$      | 0.91    | 0.91   | 0        | 完全一致 |

|         |     |        |        |   |      |
|---------|-----|--------|--------|---|------|
| dynamic | $t$ | 3.95   | 3.95   | 0 | 完全一致 |
|         | $p$ | <0.001 | <0.001 | 0 | 完全一致 |
|         | $d$ | 0.51   | 0.51   | 0 | 完全一致 |

**Table 2.** Descriptive statistics for maximum absolute deviation (MAD) values (in pixels) per typicality condition, and paired  $t$  test results (for a comparison of the atypical and typical conditions)

| Experiment | Condition | $N$ | Typical |       |      | Atypical |       |      | $t$ Test |        |       |
|------------|-----------|-----|---------|-------|------|----------|-------|------|----------|--------|-------|
|            |           |     | $M$     | $SD$  | $BC$ | $M$      | $SD$  | $BC$ | $t$      | $p$    | $d_z$ |
| 1          | Click     | 53  | 142.7   | 111.5 | .632 | 287.5    | 237.0 | .641 | 4.43     | < .001 | 0.61  |
| 1          | Touch     | 55  | 52.2    | 78.3  | .442 | 79.0     | 91.6  | .500 | 2.69     | .009   | 0.36  |
| 2          | Default   | 59  | 157.8   | 158.8 | .558 | 283.4    | 225.7 | .576 | 5.49     | < .001 | 0.71  |
| 2          | Slow      | 59  | 73.4    | 113.1 | .573 | 150.9    | 141.2 | .593 | 4.52     | < .001 | 0.59  |
| 3          | Static    | 59  | 185.2   | 134.4 | .520 | 269.7    | 172.7 | .548 | 4.18     | < .001 | 0.54  |
| 3          | Rtmax     | 60  | 189.8   | 150.8 | .536 | 301.5    | 197.8 | .501 | 4.32     | < .001 | 0.56  |
| 3          | Initmax   | 66  | 304.8   | 140.7 | .510 | 470.9    | 203.2 | .473 | 7.39     | < .001 | 0.91  |
| 3          | Dynamic   | 60  | 297.0   | 111.6 | .560 | 364.1    | 154.0 | .508 | 3.95     | < .001 | 0.51  |

MAD values were first aggregated per participant and typicality condition.  $BC$  = bimodality coefficient, based on the per-participant standardized MAD values

图 3 原文献中描述性统计结果及配对样本  $t$  检验图

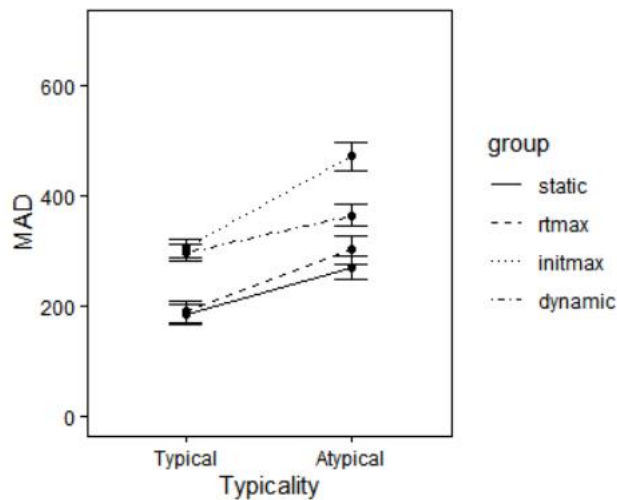
| Description: df [4 × 9] |    |          |          |          |          |          |              |           |
|-------------------------|----|----------|----------|----------|----------|----------|--------------|-----------|
| group                   | N  | M.t      | SD.t     | M.a      | SD.a     | t        | p            | d         |
| static                  | 59 | 185.2376 | 134.4211 | 269.7187 | 172.7319 | 4.176474 | 1.007751e-04 | 0.5437306 |
| rtmax                   | 60 | 189.8261 | 150.8272 | 301.4930 | 197.7812 | 4.323063 | 6.003791e-05 | 0.5581050 |
| initmax                 | 66 | 304.7886 | 140.7270 | 470.9017 | 203.2007 | 7.392753 | 3.497486e-10 | 0.9099850 |
| dynamic                 | 60 | 297.0376 | 111.6093 | 364.0986 | 153.9755 | 3.952973 | 2.091033e-04 | 0.5103266 |

4 rows

图 4 本报告中 R 代码输出的描述性统计及  $t$  检验结果表

### 3.3.4 MAD 均值差异图

绘制不同组别和刺激类型的 MAD 均值比较图：



原研究图表：



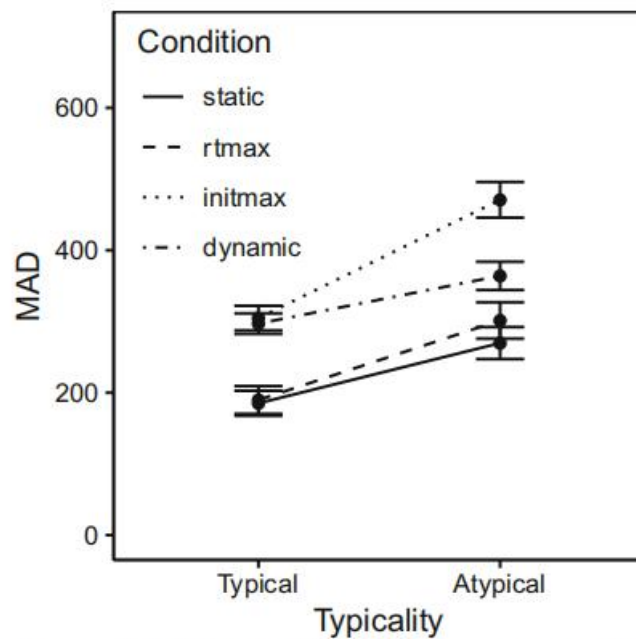


Fig. 10 Mean of maximum absolute deviation values (MAD, in pixels) for Experiment 3, shown separately per typicality and experimental condition. Error bars indicate one SEM

### 3.3.5 混合设计方差检验

采用以每位参与者聚合的 MAD 值为因变量、组内因素典型性（非典型 vs 典型）和组间因素启动程序（start vs rtmax vs initmax vs dynamic）的重复测量方差分析显示：

①典型性的主效应显著

$$F(1,241)=97.72, p<0.001, \eta_p^2 = 0.29, 90\% \text{ CI } [0.21, 0.36]$$

实验条件：非典型样例的 MAD 值显著高于典型样例。

②起始程序的主效应显著

$$F(3,241)=18.67, p<0.001, \eta_p^2 = 0.19, 90\% \text{ CI } [0.11, 0.25]$$

实验条件：不同起始程序下的 MAD 值存在整体差异。

③典型性与起始程序的交互效应显著

$$F(3,241)=4.12, p=0.007, \eta_p^2 = 0.05, 90\% \text{ CI } [0.01, 0.09]$$

表 11 交互效应分解表

|            |          | 自由度  | $p$    | $F$   | $\eta_p^2$ | 95%CI        |
|------------|----------|------|--------|-------|------------|--------------|
| starting   | 原研究报告结果  | 3    | <0.001 | 18.67 | 0.19       | [0.11, 0.25] |
|            | 本研究      | 3    | <0.001 | 18.67 | 0.19       | [0.11, 0.25] |
|            | $\delta$ | 0    | 0      | 0     | 0          | 0            |
|            | 评级       | 完全一致 | 完全一致   | 完全一致  | 完全一致       | 完全一致         |
| typicality | 原研究报告    | 1    | <0.001 | 97.72 | 0.29       | [0.21, 0.36] |

|                     |          |     |        |       |      |              |
|---------------------|----------|-----|--------|-------|------|--------------|
| 告结果                 |          |     |        |       |      |              |
| starting×typicality | 本研究      | 1   | <0.001 | 97.72 | 0.29 | [0.21, 0.36] |
|                     | $\delta$ | 0   | 0      | 0     | 0    | 0            |
|                     | 评级       | 完全一 | 完全一致   | 完全一致  | 完全一致 | 完全一致         |
|                     |          | 致   |        |       |      |              |
|                     | 原研究报     | 3   | 0.007  | 4.12  | 0.05 | [0.01, 0.09] |
|                     | 告结果      |     |        |       |      |              |
|                     | 本研究      | 3   | 0.007  | 4.12  | 0.05 | [0.01, 0.09] |
|                     | $\delta$ | 0   | 0      | 0     | 0    | 0            |
|                     | 评级       | 完全一 | 完全一致   | 完全一致  | 完全一致 | 完全一致         |
|                     |          | 致   |        |       |      |              |
| Error               | 原研究报     | 241 |        |       |      |              |
|                     | 告结果      |     |        |       |      |              |
|                     | 本研究      | 241 |        |       |      |              |
|                     | $\delta$ | 0   |        |       |      |              |
|                     | 评级       | 完全一 |        |       |      |              |
| 致                   |          |     |        |       |      |              |

```
### ANOVA
library(ez)
{r}
anova_mad <- aov_ez(data=agg_mad, dv = "MAD", between = "group", within = "Typicality",
  id = "subject_nr")
nice(anova_mad, es = c("pes", "ges"))

# 90 % confidence interval for partial eta-squared
round(get_partial_etas(anova_mad$anova_table, conf.level=.90), 2)
```

Contrasts set to contr.sum for the following variables: group

Anova Table (Type 3 tests)

Response: MAD

|   | Effect           | df     | MSE      | F         | ges  | pes  | p.value |
|---|------------------|--------|----------|-----------|------|------|---------|
| 1 | group            | 3, 241 | 37593.62 | 18.67 *** | .144 | .189 | <.001   |
| 2 | Typicality       | 1, 241 | 14412.64 | 97.72 *** | .101 | .289 | <.001   |
| 3 | group:Typicality | 3, 241 | 14412.64 | 4.12 **   | .014 | .049 | .007    |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '+' 0.1 '.' 1

|                  | pes  | lower | upper |
|------------------|------|-------|-------|
| group            | 0.19 | 0.11  | 0.25  |
| Typicality       | 0.29 | 0.21  | 0.36  |
| group:Typicality | 0.05 | 0.01  | 0.09  |

图 5 本报告中 R 代码输出的重复测量方差分析结果表（原研究中以文字形式呈现）

### 3.3.6 对比分析

针对起始程序各条件与静态条件的差异比较：

①initmax vs. static

$t(241)=6.53, p<0.001$ （显著差异）

实验条件：initmax 条件下的 MAD 值显著高于 static 条件。

②dynamic vs. static

$t(241)=4.10, p<0.001$ （显著差异）

实验条件：dynamic 条件下的 MAD 值显著高于 static 条件。

③rtmax vs. static

$t(241)=0.72, p=0.470$ （无显著差异）

实验条件：rtmax 条件与 static 条件的 MAD 值无显著差异。

表 12 对比分析表

| 条件                  | Contrast analyses | 原研究报告结果 | 本研究    | $\delta$ | 评级   |
|---------------------|-------------------|---------|--------|----------|------|
| typicality_static   | $t$               |         | 3.82   |          |      |
|                     | $p$               |         | <0.001 |          |      |
| rtmax_static_main   | $t$               | 0.72    | 0.72   | 0        | 完全一致 |
|                     | $p$               | 0.470   | 0.470  | 0        | 完全一致 |
| initmax_static_main | $t$               | 6.53    | 6.53   | 0        | 完全一致 |
|                     | $p$               | <0.001  | <0.001 | 0        | 完全一致 |
| dynamic_static_main | $t$               | 4.10    | 4.10   | 0        | 完全一致 |
|                     | $p$               | <0.001  | <0.001 | 0        | 完全一致 |
| rtmax_static_int    | $t$               | 0.87    | 0.87   | 0        | 完全一致 |
|                     | $p$               | 0.383   | 0.383  | 0        | 完全一致 |
| initmax_static_int  | $t$               | 2.68    | 2.68   | 0        | 完全一致 |
|                     | $p$               | 0.008   | 0.008  | 0        | 完全一致 |
| dynamic_static_int  | $t$               | - 0.56  | - 0.56 | 0        | 完全一致 |
|                     | $p$               | 0.576   | 0.576  | 0        | 完全一致 |

针对典型性效应在不同起始程序中的差异比较：

①initmax 条件下的典型性效应 vs. static 条件下的典型性效应

$t(241)=2.68, p=0.008$ （显著差异）

实验条件：initmax 条件下，典型性对 MAD 值的影响幅度显著大于 static 条件。

②dynamic 条件下的典型性效应 vs. static 条件下的典型性效应

$t(241)=-0.56, p=0.576$ （无显著差异）

实验条件：dynamic 条件与 static 条件中，典型性对 MAD 值的影响幅度无显著差异。

③rtmax 条件下的典型性效应 vs. static 条件下的典型性效应

$t(241)=0.87, p=0.383$ （无显著差异）

实验条件：rtmax 条件与 static 条件中，典型性对 MAD 值的影响幅度无显著差异。

```

346
347 ~ ## Contrast analyses
348 ~ ```{r}
349 # Retrieve grid
350 anova_mad_grid <- lsmeans(anova_mad,~Typicality:group)
351
352 # Specify contrasts
353 contrast_matrix_complete <- list(
354   typicality_static = c(-1,1,0,0,0,0,0,0),
355   rtmax_static_main= c(-1,-1,1,1,0,0,0,0)/2,
356   initmax_static_main = c(-1,-1,0,0,1,1,0,0)/2,
357   dynamic_static_main = c(-1,-1,0,0,0,0,1,1)/2,
358   rtmax_static_int = c(1,-1,-1,1,0,0,0,0),
359   initmax_static_int = c(1,-1,0,0,-1,1,0,0),
360   dynamic_static_int = c(1,-1,0,0,0,0,-1,1))
361
362 # Test contrasts
363 contrast(anova_mad_grid,contrast_matrix_complete)
364 ~

```

| contrast            | estimate | SE   | df  | t.ratio | p.value |
|---------------------|----------|------|-----|---------|---------|
| typicality_static   | 84.5     | 22.1 | 241 | 3.822   | 0.0002  |
| rtmax_static_main   | 18.2     | 25.1 | 241 | 0.723   | 0.4702  |
| initmax_static_main | 160.4    | 24.6 | 241 | 6.529   | <.0001  |
| dynamic_static_main | 103.1    | 25.1 | 241 | 4.101   | 0.0001  |
| rtmax_static_int    | 27.2     | 31.1 | 241 | 0.873   | 0.3833  |
| initmax_static_int  | 81.6     | 30.4 | 241 | 2.684   | 0.0078  |
| dynamic_static_int  | -17.4    | 31.1 | 241 | -0.560  | 0.5763  |

图 6 本报告中 R 代码输出的对比分析结果表（原研究中以文字形式呈现）

3.4 轨迹形状分布分析

3.4.1 参与者在每个典型性和实验条件下的标准化 MAD 值的双峰系数

比较不同实验刺激类型（典型/非典型）和启动程序（static/ rtmax / initmax /dynamic）下标准化 MAD 值双峰系数的描述性统计结果，见表 13。

表 13 四种实验条件下标准化 MAD 值的双峰系数描述性统计结果的比较

| Condition | N  | Bimodality coefficient |          |
|-----------|----|------------------------|----------|
|           |    | Typical                | Atypical |
| static    | 59 | 原研究                    | 0.520    |
|           |    | 报告结果                   | 0.548    |
|           |    | 本研究                    | 0.520    |
|           |    | $\delta$               | 0        |
|           |    | 评级                     | 完全一致     |
| rtmax     | 60 | 原研究                    | 0.536    |
|           |    | 报告结果                   | 0.501    |

|         |    |          |       |       |
|---------|----|----------|-------|-------|
| initmax | 66 | 本研究      | 0.536 | 0.501 |
|         |    | $\delta$ | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  |
|         |    | 原研究      | 0.510 | 0.473 |
|         |    | 报告结果     |       |       |
|         |    | 本研究      | 0.510 | 0.473 |
|         |    | $\delta$ | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  |
|         |    | 原研究      | 0.560 | 0.508 |
|         |    | 报告结果     |       |       |
| dynamic | 60 | 本研究      | 0.560 | 0.508 |
|         |    | $\delta$ | 0     | 0     |
|         |    | 评级       | 完全一致  | 完全一致  |

---

3.4.2 量化和统计检验不同条件下轨迹类型频率的差异

先将实验 3 收集到的轨迹映射至实验 1 和 2 中使用的同一组原型集上。  
原研究结果如下：



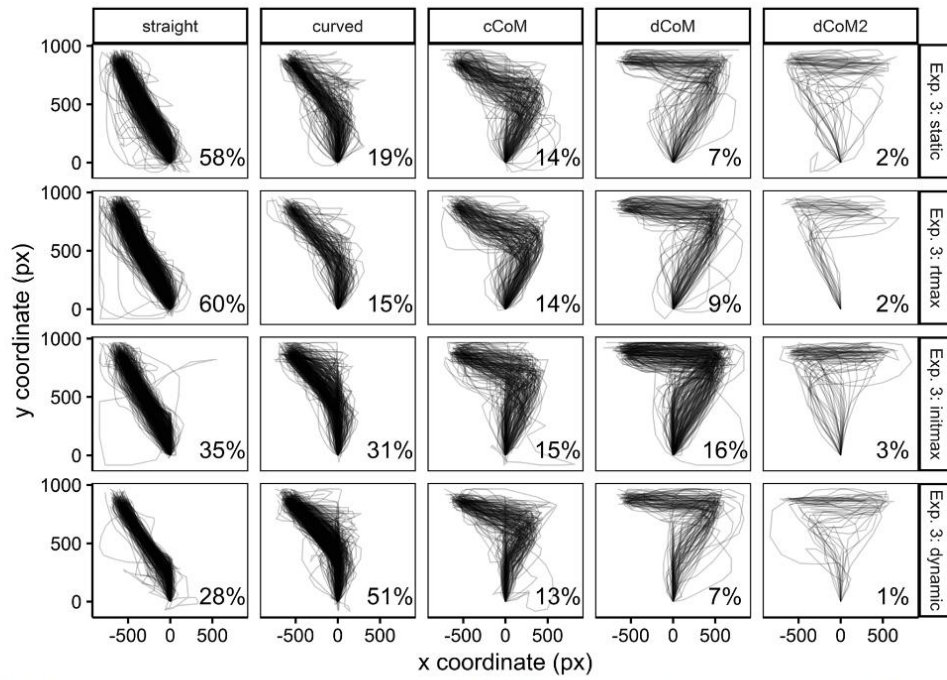


Fig. 12 Individual trajectories per assigned prototype, shown separately for the different experimental conditions in Experiment 3. For each prototype, the relative frequency of classifications per experimental condition is displayed

本研究结果如下：

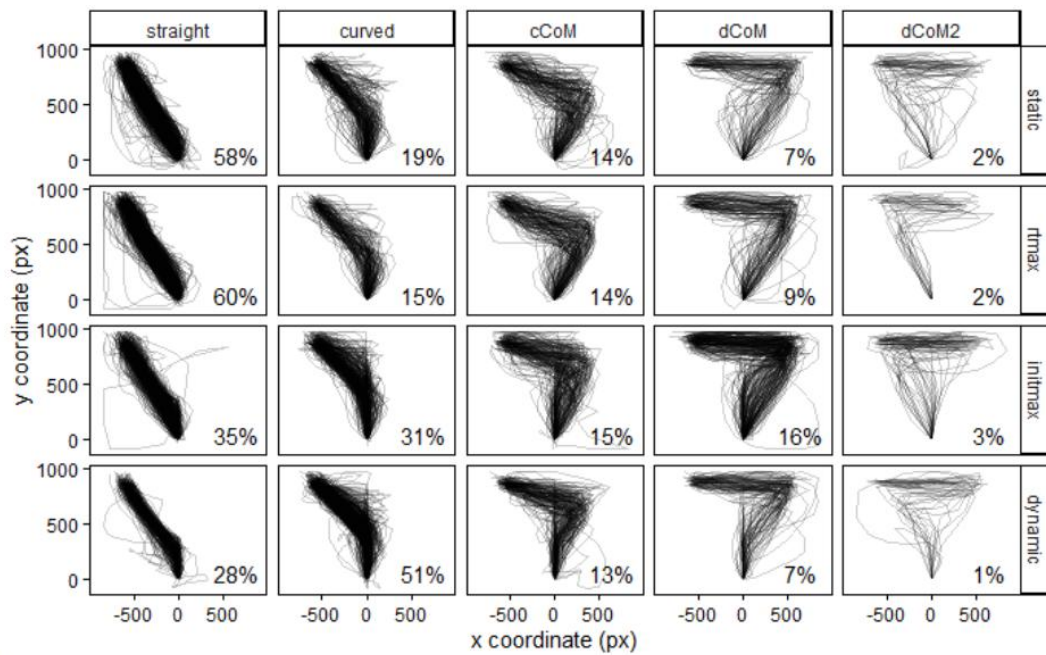


表 14 不同启动条件下各原型的分类相对频率分布描述性统计结果的比较

| Condition | Assigned Prototype |        |      |      |       |
|-----------|--------------------|--------|------|------|-------|
|           | Straight           | Curved | cCoM | dCoM | dCoM2 |

|         |          |      |      |      |      |      |
|---------|----------|------|------|------|------|------|
| static  | 原研究      | 58%  | 19%  | 14%  | 7%   | 2%   |
|         | 报告结果     |      |      |      |      |      |
|         | 本研究      | 58%  | 19%  | 14%  | 7%   | 2%   |
|         | $\delta$ | 0    | 0    | 0    | 0    | 0    |
|         | 评级       | 完全一致 | 完全一致 | 完全一致 | 完全一致 | 完全一致 |
| rtmax   | 原研究      | 60%  | 15%  | 14%  | 9%   | 2%   |
|         | 报告结果     |      |      |      |      |      |
|         | 本研究      | 60%  | 15%  | 14%  | 9%   | 2%   |
|         | $\delta$ | 0    | 0    | 0    | 0    | 0    |
|         | 评级       | 完全一致 | 完全一致 | 完全一致 | 完全一致 | 完全一致 |
| initmax | 原研究      | 35%  | 31%  | 15%  | 16%  | 3%   |
|         | 报告结果     |      |      |      |      |      |
|         | 本研究      | 35%  | 31%  | 15%  | 16%  | 3%   |
|         | $\delta$ | 0    | 0    | 0    | 0    | 0    |
|         | 评级       | 完全一致 | 完全一致 | 完全一致 | 完全一致 | 完全一致 |
| dynamic | 原研究      | 28%  | 51%  | 13%  | 7%   | 1%   |
|         | 报告结果     |      |      |      |      |      |
|         | 本研究      | 28%  | 51%  | 13%  | 7%   | 1%   |
|         | $\delta$ | 0    | 0    | 0    | 0    | 0    |
|         | 评级       | 完全一致 | 完全一致 | 完全一致 | 完全一致 | 完全一致 |

再通过卡方检验,分析实验组别（不同的启动程序）和轨迹原型分类之间的关联性。

表 15 推断性统计卡方检验结果的比较(原文献方法)

|  |
|--|
| 假设一：实验组别（group）与轨迹原型分类（prototype_label）之间存在关联，实验 |
|--|

|                         |            |              |
|-------------------------|------------|--------------|
| 组别不同会导致轨迹原型分类的分布存在显著差异。 |            |              |
|                         | 统计量        | 显著性指标( $p$ ) |
|                         | $(\chi^2)$ |              |
| 原文献                     | 535.73     | < .001       |
| 报告结果                    |            |              |
| 本研究                     | 535.73     | < .001       |
| $\delta$                | 0          | 0            |
| 评级                      | 完全一致       | 完全一致         |

结果显示，原型在不同实验条件下的出现频率存在显著差异。

3.4.3 计算各实验组在不同典型性条件（典型/非典型）下轨迹分类的相对频率

表 16 各实验组在不同典型性条件下轨迹分类的相对频率描述性统计结果的比较

| Condition |          | Assigned Prototype |        |       |      |       |
|-----------|----------|--------------------|--------|-------|------|-------|
|           |          | Straight           | Curved | cCoM  | dCoM | dCoM2 |
| static    | 原研究      | 62/49              | 19/20  | 12/18 | 5/10 | 2/4   |
|           | 报告结果     |                    |        |       |      |       |
|           | 本研究      | 62/49              | 19/20  | 12/18 | 5/10 | 2/4   |
|           | $\delta$ | 0                  | 0      | 0     | 0    | 0     |
|           | 评级       | 完全一致               | 完全一致   | 完全一致  | 完全一致 | 完全一致  |
| rtmax     | 原研究      | 64/50              | 15/15  | 13/18 | 7/14 | 1/3   |
|           | 报告结果     |                    |        |       |      |       |
|           | 本研究      | 64/50              | 15/15  | 13/18 | 7/14 | 1/3   |
|           | $\delta$ | 0                  | 0      | 0     | 0    | 0     |
|           | 评级       | 完全一致               | 完全一致   | 完全一致  | 完全一致 | 完全一致  |

|         |          |       |       |       |       |      |
|---------|----------|-------|-------|-------|-------|------|
| initmax | 原研究      |       |       |       |       |      |
|         | 报告结果     | 39/24 | 33/27 | 14/17 | 12/26 | 2/6  |
|         | 本研究      | 39/24 | 33/27 | 14/17 | 12/26 | 2/6  |
|         | $\delta$ | 0     | 0     | 0     | 0     | 0    |
|         | 评级       | 完全一致  | 完全一致  | 完全一致  | 完全一致  | 完全一致 |
| dynamic | 原研究      |       |       |       |       |      |
|         | 报告结果     | 28/26 | 53/45 | 13/15 | 6/10  | 1/4  |
|         | 本研究      | 28/26 | 53/45 | 13/15 | 6/10  | 1/4  |
|         | $\delta$ | 0     | 0     | 0     | 0     | 0    |
|         | 评级       | 完全一致  | 完全一致  | 完全一致  | 完全一致  | 完全一致 |

\*行代表 实验条件（不同的启动程序），列代表鼠标轨迹被分类到的原型类别（assigned prototype），单元格数值为典型试次百分比/非典型试次百分比，每个条件下的百分比分母是该条件下所有有效试次的总数。例如，straight:62/49 表示：典型条件下，62%的轨迹被分类为 straight；非典型条件下，49%的轨迹被分类为 straight。

#### 3.4.4 通过有序混合回归模型预测轨迹类型的推断性统计

模型包含每位参与者的随机截距和两个预测变量：典型性（非典型 = .5，典型 = - .5）以及启动程序（采用虚拟编码，其中 static start 作为基线参照）

原研究结果如下：

We again predicted the trajectory type in an **ordinal mixed regression** including a random intercept per participant and the predictors typicality (atypical = .5, typical = -.5) and experimental condition (dummy-coded, with static serving as the baseline condition). Atypical trials led to a significantly higher probability of more extreme trajectories in the static condition ( $z = 5.06, p < .001$ ). The rtmax condition did not differ significantly from the static condition,  $z = 0.31, p = .760$ . Both the initmax and the dynamic conditions led to significantly more extreme trajectories than the static condition,  $z = 5.99, p < .001$ , and  $z = 4.37, p < .001$ , respectively. For the initmax condition, more instances of both the curved and change-of-mind trajectory types occurred, whereas for the dynamic condition, there seemed to be especially more curved trajectories (see Fig. 12). With regard to the interaction of typicality and condition, we observed no significant interactions for the rtmax and initmax conditions,  $z = 0.79, p = .427$ , and  $z = 1.68, p = .094$ . The dynamic condition led to a relatively smaller increase in extreme trajectories for atypical trials than the static condition,  $z = -2.19, p = .029$ .

本研究结果如下：

```
> contrasts(mt_data$Typicality) <- c(-0.5,0.5)
> # use default contrasts for group (dummy coding with static as baseline)
> contrasts(mt_data$group)
      rtmax initmax dynamic
static      0      0      0
rtmax       1      0      0
initmax      0      1      0
dynamic      0      0      1
> summary(c1mm(prototype_label~Typicality*group+(1|subject_nr),data=mt_data$data))
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: prototype_label ~ Typicality * group + (1 | subject_nr)
data:      mt_data$data

Random effects:
Groups      Name      Variance Std.Dev.
subject_nr (Intercept) 0.6917   0.8317
Number of groups:  subject_nr 245

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
Typicality1      0.69431   0.13716   5.062 4.15e-07 ***
grouprtmax        0.05605   0.18342   0.306  0.7599
groupinitmax      1.05662   0.17657   5.984 2.18e-09 ***
groupdynamic      0.77965   0.17838   4.371 1.24e-05 ***
Typicality1:grouprtmax 0.15934   0.20078   0.794  0.4274
Typicality1:groupinitmax 0.30873   0.18414   1.677  0.0936 .
Typicality1:groupdynamic -0.39681   0.18115  -2.191  0.0285 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
              Estimate Std. Error z value
straight|curved 0.1227   0.1296   0.946
curved|cCoM     1.6248   0.1323  12.283
cCoM|dCoM       2.6964   0.1371  19.667
dCoM|dCoM2      4.6552   0.1693  27.504
> |
```

结果显示：

①在引发极端运动轨迹的概率上，非典型试次在 static 下导致更极端轨迹的概率显著更高 ( $z = 5.06, p < .001$ )。rtmax 条件与 static 无显著差异 ( $z = 0.31, p = .760$ )。



相较于 static, initmax 条件和 dynamic 均引发显著更极端的运动轨迹 ( $z = 5.99, p < .001$ ;  $z = 4.37, p < .001$ ) ;

②initmax 条件下弯曲轨迹 (curved) 与改变主意 (cCoM / dCoM/ dCoM2) 类轨迹同时增加, 而 dynamic 尤以弯曲轨迹增多为特征 (见原文中图 12 比例增加) 。

③针对典型性与条件的交互作用, rtmax 与 initmax 条件均未呈现显著交互效应 ( $z = 0.79, p = .427$ ;  $z = 1.68, p = .094$ ) ;

④dynamic 下非典型试次的极端轨迹增幅相对 static 较小 ( $z = - 2.19, p = .029$ ) 。

表 17 有序混合回归模型的推断性统计结果的比较(原文献方法)

|  |                          |                 |              |
|--|--------------------------|-----------------|--------------|
| 假设一：非典型刺激 (atypical exemplars) 会导致更极端的轨迹类型 (如更多 dCoM 轨迹) |                          |                 |              |
| 假设二：不同启动程序 (static/rtmax/initmax/dynamic) 会影响轨迹类型的分布。    |                          |                 |              |
| 固定效应系数   |                          | 效应量<br>(Wald Z) | 显著性指标( $p$ ) |
| 原文献<br>报告结果  | Typicality1              | 5.06            | < .001       |
|  | grouprtmax               | 0.31            | 0 .760       |
|  | groupinitmax             | 5.99            | < .001       |
|  | groupdynamic             | 4.37            | < .001       |
|  | Typicality1:grouprtmax   | 0.79            | 0.427        |
|  | Typicality1:groupinitmax | 1.68            | 0.094        |
|  | Typicality1:groupdynamic | - 2.19          | 0.029        |
| 本研究  | Typicality1              | 5.062           | < .001       |
|  | grouprtmax               | 0.306           | 0 .760       |
|  | groupinitmax             | 5.984           | < .001       |
|  | groupdynamic             | 4.371           | < .001       |
|  | Typicality1:grouprtmax   | 0.794           | 0.427        |
|  | Typicality1:groupinitmax | 1.677           | 0.094        |
|  | Typicality1:groupdynamic | -2.191          | 0.029        |
| $\delta$   |                          | 0               | 0            |
| 评级   |                          | 因舍入导致的<br>偏差    | 完全一致         |

3.5 对原文计算可复现性进行评估

3.5.1 使用与原文献相同方法

报告原文献的值的评级分布、推论的一致情况，整理成表格，如下表所示：

表 18 结果可复现性的评估表

| 可复现性情况                        | 数量及占比    |     |
|-------------------------------|----------|-----|
|                               | <i>N</i> | %   |
| 完全一致( $\delta = 0\%$ )        | 127      | 99% |
| 偏差较小( $0\% < \delta < 10\%$ ) | 0        | 0   |
| 偏差较大( $\delta > 10\%$ )       | 0        | 0   |
| 因舍入导致的偏差                      | 1        | 1%  |

\* 结果数量 *N* 指在重复分析中，对重复分析结果与原结果进行配对比较的次数。对于每个目标效应，结果包括一组数值，如汇总效应估计(summary estimate，如 *t* 值/ *F* 值)、置信区间界限(confidence interval bound)、效应量(effect size)样本大小(size effect)等，应将原文中报告的每个数值与重复结果进行比较。

表 19 推论的一致性的评估表(原分析方法)

| 推论的一致性 | 数量及占比      |      |
|--------|------------|------|
|        | <i>N</i> * | %    |
| 一致     | 31         | 100% |
| 不一致    | 0          | 0%   |

\* 推论数量 *N* 指在重复分析中，对效应做出统计推断的次数。例如，仅进行了一个 *t* 检验，则 *N*=1；如果进行了一个 2\*2 的方差分析，并进行简单效应分析，则有可能有 7 个统计推断：两个主效应的推论，一个交互作用的推论，四个可能的简单效应分析的推论，因此 *N*=7。如果报告的 *p* 值相对于重复的 *p* 值落在显著性水平边界的另一侧，则被归类为推论不一致；反之为推论一致。

4 讨论

4.1 计算可复现性检验结果分析

结合下表，对原文献进行分析，推测可能导致可复现性检验结果差异的原因。对于重要的原因，逐段进行展开说明。

表 20 计算上（不）可重复的原因分析表

| 可能原因 | 研究三 | …… | 研究 n |
|------|-----|----|------|
|------|-----|----|------|

|             |              |  |   |
|-------------|--------------|--|---|
| 原文献开放性问题的原因 | 一般性开放获取问题    | 几个结果的微小差异，可能是由于分析中使用了没有设置固定种子的随机数；                   | × |
|             |              | 个别结果的微小差异，可能是由于印刷或复制粘贴错误；                            | × |
|             |              | 文章文本中程序报告不明确，包括纳入亚组的标准、缺乏或不正确报告用于回归模型的变量、以及未报告的单侧分析； | × |
|             |              | 在文章的开放实践声明中对研究的模糊标记。                                 | × |
|             | OSF 开放获取特定问题 | OSF 中缺乏对数据和/或代码内容进行说明的文档(readme 文档)；                 | × |
|             |              | OSF 上的数据与代码文件不一致，如代码中对部分数据进行了操作，但这部分数据在数据文件中无对应；     | × |
|             |              | OSF 上的数据存储问题，包括文件损坏或无法下载。                            | × |
|             | 数据开放获取特定问题   | 没有提供原始数据；  | × |
|             |              | 没有提供处理后的数据；  | × |
|             |              | 没有提供数据处理过程的描述或代码。                                    | × |
| 重复过程的原因     | 代码开放获取特定问题   | 缺乏共享的分析代码或建模代码；                                      | × |
|             |              | 软件包或软件版本的问题。   | × |
|             | 重复研究与原研究的区别  | 是否使用同样的数据集；  | √ |
|             |              | 是否使用同样的数据分析软件及软件包；                                   | √ |
|             |              | 是否使用同样的数据分析方法。                                       | √ |
|             | 重复者相关因素      | 重复者此前是否有过 R 使用经验；                                    | √ |
|             |              | 重复者对关于 R 的知识或操作上存在漏洞，较难理解原文章中的部分操作(可做简单说明)。          | √ |
| 其他影响因素      | 文献年份         | 文献发表年份是否较为久远，是否在开放科学运动之前；                            | × |

|      |                     |   |
|------|---------------------|---|
| 素    | 文献引用量大小;            | × |
| 文献质量 | 是否有其他研究支持本文献结果;     | √ |
|      | 是否有其他研究对本文献结果进行了    | × |
|      | 重复, 重复结果如何(可做简单说明)。 |   |

---

本次计算可复现性检验中, 我们聚焦于文献的实验 3, 得益于以下关键因素, 成功复现了所有结果:

数据与代码的完备共享: 原文献在 OSF 平台提供了实验 3 的完整原始数据及详尽代码, 涵盖数据记录、处理到分析的全流程, 确保我们精准把握每一个细节, 这是复现成功的基石。

清晰的实验设计与操作流程描述: 文献对实验 3 的描述极为详尽, 从参与者如何分配到不同启动程序条件 (静态、rtmax、initmax 和动态), 到每种条件下具体的实验流程, 甚至包括了对鼠标移动初始化时间等细节的记录要求, 使我们能够完全按照原文献的方法进行复现。

高度一致的分析方法与软件环境: 在分析过程中, 我们严格遵循文献的方法论。我们运用与原文献相同的 R 语言及其软件包对数据进行处理与分析, 确保了分析方法的一致性。同时, 我们对实验条件、数据处理步骤和分析流程进行了精细控制, 使其与原文献保持高度一致。

综上所述, 正是由于原文献的开放科学实践、清晰的方法描述以及我们团队对细节的严格把控, 才使得实验 3 的结果得以成功复现。这一成功不仅验证了原文献的发现, 也为未来相关研究的可复现性提供了有益参考。

#### 4.2 其他思考

首先, 通过这次作业, 我们将课堂知识转化为实践能力。小组成员独立调试代码能力显著提升, 90% 的报错可自主解决。其次, 我们增加了统计知识的积累, 了解了有序混合回归模型等分析方法。最后, 本次可重复性作业让我们意识到团队协作的重要性, 可复现性是科学研究的底线, 而协作创新才是科学进步的引擎。

### 参考文献

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, 52(1), 317–341. <https://doi.org/10.3758/s13428-019-01228-y>