附录一、研究报告格式(中文)

# 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 | 范超鸿 | | |
| 组员 | 尹子涵、牛至旭 | | |
| 分工 | | | |
| 数据分析 | 范超鸿、尹子涵 | PPT 制作 | 牛至旭、尹子涵 |
| 文字报告制作 | 范超鸿、尹子涵、牛至旭 | PPT 展示 | 范超鸿 |

* 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：**尽管人们从社会关系中获得了实质性的好处，但他们经常避免与陌生人交谈，因为他们对这种谈话的结果抱有悲观的态度（例如，他们认为自己会被拒绝或在交谈中不知道该说什么）。之前的研究试图让人们意识到，他们夸大了对与陌生人交谈的担忧。为了减少人们的担忧，我们设计了一种干预方法，让参与者玩一个为期一周的寻宝游戏，其中包括反复寻找、接近和与陌生人交谈。与对照组相比，这种最小的、容易复制的治疗使人们对被拒绝的可能性不那么悲观，对自己的对话能力更乐观——这些好处在研究结束后至少持续了一周。每日报告显示，人们的期望越来越积极和准确，强调了反复的经验在改善人们与陌生人交谈的态度方面的重要性。

**关键词：**社会互动，谈话，干预，社会联系，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 | Sandstrom, G. M. , Boothby, E. J. , & Cooney, G. . Talking to strangers: a week-long intervention reduces psychological barriers to |

| | | | |
|---|---|---|---|
| | social connection. Journal of Experimental Social Psychology, 102. | | |
| 数据来源 | https://osf.io/b76gf/?view_only=a1baa2407bf249eaa3376fabb2c63246 | | |
| **2 文献选取** | | | |
| 文献主题是否包含不止一篇研究？ | □ 是，且包含元分析研究<br>□ 是，但不包含元分析研究<br>√否 | 文献此前被其他研究者重复过？ | □ 是(附上原文链接)<br>√否 |
| **3 研究假设选取** | | | |
| 重复的研究假设 | (指出检验了原文献哪部分的哪些假设结果，如研究一、研究二/ 行为数据、电生理数据/ A 任务、B 任务等) | | |
| 重复的研究假设是否在其他研究中经过重复？ | □ 是(附上原文链接)<br>√否 | 文献共几个实验，重复的研究假设是第几个实验中的？ | 文献共一个实验，该实验中一共有 6 个因变量，我们选取了其中具有代表性的 2 个，其中一个为连续型变量 ability，另一个为离散型变量 rejection，其他因变量的分析方法与这两个因变量的分析一致。 |
| 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) | | |
| **4 数据集选取** | | | |
| 是否采用原始数据？ | √ 是<br>□ 否 | 是否对样本量进行修改？ | □ 是(说明原因)<br>√否 |
| 若修改，报告原文样本量大小和修改后的样本量大小 | 无 | 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 | 无 |

## 1.2 文献介绍

### 1.2.1 研究背景

社会互动对人们健康和幸福非常重要，但人们很少主动与陌生人交谈。并且，人们带着耳机为了避免说话，在公共场所盯着手机，这些行为可能的原因是：人们对与陌生人交谈的

很多方面都感到悲观。

①认为别人对于交谈不感兴趣；②低估他人对自己的喜爱程度；③怀疑自己启动和维持交谈的这种能力。

### 1.2.2 主要研究问题及假设

1、研究问题：

（1）验证过去的研究结论：人们会低估陌生人对社交的积极反应；

（2）干预人们根深蒂固的悲观信念：让人们在连续的几天里反复交谈，让人们适应同陌生人交谈的情况，减少他们的恐惧。

2、研究假设：

（1）和以往的研究结果相同，人们会低估陌生人对他们之间社交的积极反应。

（2）在干预实验中，参与“寻宝游戏”的干预组比对照组对自己的对话能力更为乐观，对被拒绝交谈的恐惧感更低。

### 1.2.3 研究结果和结论

1、研究结果：在干预实验中，与对照组相比，参与了为期一周的“寻宝游戏”的参与者在这种最小的、容易复制的治疗中使他们对被拒绝的可能性不那么悲观，对自己的对话能力更乐观。这些改变在研究结束后至少持续了一周。

2、研究结论；反复的经验在改善人们与陌生人交谈的态度方面有重要作用。

# 2 方法

## 2.1 样本

共有 454 人开始参与研究。在数据分析前，移除了 21 名实验错误，68 名完成的任务少于四天，68 名未能通过“诚实检查”的被试。最终分析的数据来自 286 名参与者，其中 75 名男性，209 名女性，2 名其他或不愿透露，平均年龄 20.1 岁，标准差 2.1 岁。参与分析的 286 名被试中干预组 198 名，控制组 88 名。

## 2.2 原研究方法简介

### 2.2.1 研究设计

本实验采用 2（条件：干预组、对照组）×3（时间：实验开始、实验结束、结束一周后）混合实验设计，探究干预方法（重复与陌生人交谈）是否能降低人们对社交连接的悲观预期（例如，担心被拒绝）。

本实验通过寻宝游戏进行。共有 29 个寻宝游戏“任务”供被试选择。每个任务的目标是

找到具有某些特征的陌生人（例如，“找到穿着有趣鞋子的人”或“找到正在喝咖啡的人”）。参与者每完成一项任务即可获得积分和奖品。参与者可以查看他们与其他参与者相比的表现（完成的任务数量）。在一周的时间里（周一到周五）被试每天进行寻宝游戏，其中包括找到一个陌生人，然后与那个人交谈（干预组）或观察那个人（治疗组）。

本实验对条件（干预组、对照组）的操纵如下。干预组在完成寻宝任务时，需要真正地与陌生人进行对话（每天寻找、接近并与至少一名陌生人交谈）。干预组与陌生人进行共 1336 次对话。对照组不需要与陌生人交谈，只进行观察（每天寻找、接近但仅观察至少一名陌生人）。

本实验对时间的操纵体现在测验安排上。每位被试完成“一般”调查（捕捉干预引起的更广泛的态度和行为变化）和“每日”调查（检查干预进程及干预组心理过程）。被试分别在周一（干预开始）、周五（干预结束）、干预结束一周后完成“一般”调查，由此形成三个时间点。“每日”调查在被试与陌生人对话前后分别进行一次，因此干预五天中共有五次“每日”调查。对话前调查（pre-conversation survey）为被试对接下来对话体验的预测。对话后调查（post-conversation survey）为被试对实际对话体验的报告。为确保所有被试都有相似的经历，对照组也需完成每日调查。对照组在游戏前调查中报告了他们当前的情绪，在游戏后调查描述观察到的人。

本研究因变量——社交担忧从以下 6 个方面进行测量。

（1）拒绝担忧（Rejection）：在“一般”调查中及“每日”调查的对话前调查中，被试需估计他们在接触多少陌生人后才能找到愿意对话的，若认为接触第一个人就会与他们交谈，请输入“1”。在“每日”调查的对话后调查中，被试需报告实际接触多少人后，能够找到与之交谈的陌生人（人数包括对话的陌生人）。报告人数减 1 则为预期或实际被拒绝数，例如预期需要接近两个人意味着预期被一个人拒绝。

（2）对话能力（Conversational ability）：被试需对以下问题进行七点评分。“一般”调查（$\alpha start\_of\_study = 0.60$, $\alpha end\_of\_study = 0.63$, $\alpha follow\text{-}up = 0.70$）：It is hard [to start a conversation / to keep a conversation going / to end a conversation] with a stranger。（和一个陌生人很难[开始谈话/保持谈话进行下去/结束谈话]。）“每日”调查：It [will be / was] hard [to start a conversation / to keep a conversation going / to end a conversation。（开始一个谈话/保持一个谈话进行下去/结束一个谈话是很困难的。）

（3）尴尬程度（Awkwardness）：被试需对关于与陌生人交谈时 Awkwardness、enjoyment 的问题评分。“一般”调查为七点评分，“每日”调查为五点评分。

（4）积极印象 (Positive impression)：被试需对关于与陌生人交谈时自己会给对方留下什么印象的问题评分。“一般”调查为七点评分，“每日”调查为五点评分。

（5）与陌生人开启对话（Initiating conversations with strangers）：询问被试“您在过去 7 天内与多少陌生人开始交谈？”以测试被试行为变化。为测量被试与陌生人的自发对话而不是参与研究所需的对话，该项目只被包含在一般调查的干预开始时和结束一周时，没有被纳入在研究结束时的调查中。

（6）注意到与陌生人交谈的机会（Noticing opportunities to talk to strangers）：在一般调查中加入了一项 7 分制的探索性措施：“我注意到与陌生人交谈的机会。

2.2.2 数据分析工具及方法

本研究使用 R 进行数据分析。分析方法为构建混合效应模型，具体而言，对于“一般”调查结果，将条件（干预、对照）与时间（实验开始、实验结束、结束一周后）以及条件与时间的交互作用作为预测因子（自变量），并包含参与者随机截距。对于“每日”调查结果，将评价类型（预测、实际）与时间（周一至周五）以及二者的交互作用作为预测因子（自变量），并包含参与者随机截距。由于因变量数据类型不同，因此对于不同因变量具体选用的分析方法不同。拒绝担忧（rejection）为计数数据，过度分散（over-dispersed），因此本研究使用负二项回归（negative binomial regression）进行分析。对话能力（Conversational ability）、尴尬程度（Awkwardness）、享受程度（Enjoyment）、留下积极印象（Positive impression）为连续变量，本研究使用混合效应回归（mixed-effects regression）进行分析。

2.2.3 使用的 R 包

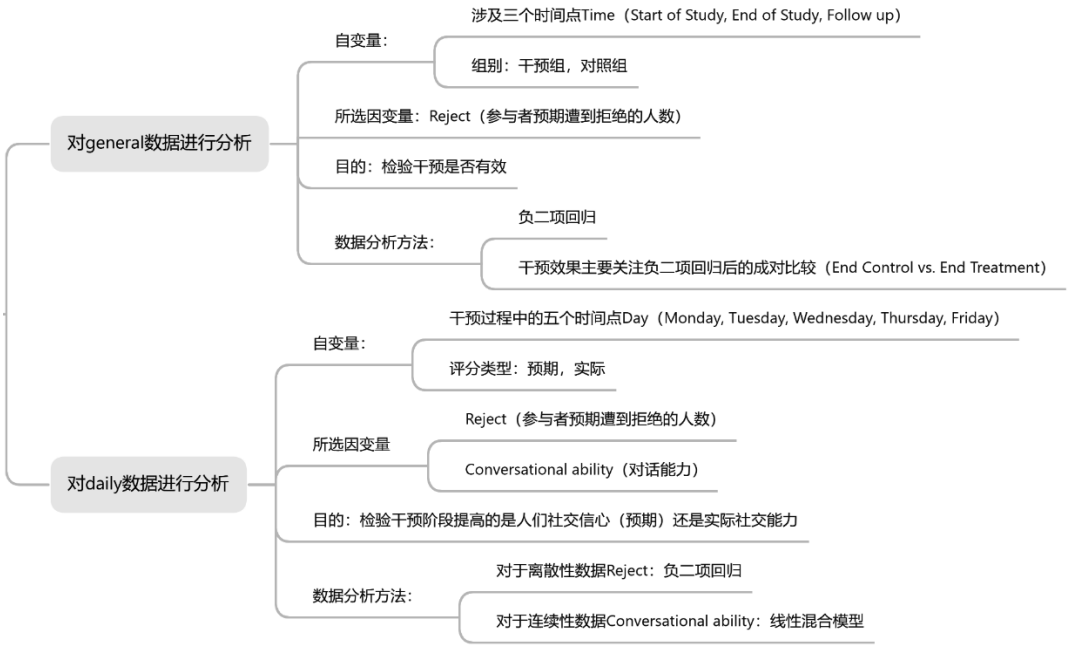
**表 2 使用的 R 包**

| R 包 | 用途 |
|---|---|
| sqldf 包 | 使用 SQL 查询语句来处理数据框，常用于筛选数据 |
| lme4 | 使用了 lme4 包中的函数 glmer.nb()，用于拟合负二项（negative binomial）广义线性混合模型（GLMM） |
| MASS | glmer.nb 函数属于 lme4 包，但它需要依赖 MASS 包来实现负二项分布的广义线性混合模型 |
| doBy | 使用了 doBy 包中的 summaryBy 函数，用于分组计算描述性统计量 |
| emmeans | 用于计算边际均值（估计边际均值）并进行成对比较 |

| | |
|---|---|
| | 使用 emtrends()语句计算：连续变量×分类变量各分组内的斜率 |
| lmerTest | 在使用 glmer.nb() 函数时，该包提供了额外的功能，会为固定效应添加显著性检验，并显示 p 值 |
| dplyr | 使用 mutate()函数：创建/修改列，使用管道操作符%>%（传递数据流） |
| ggplot2 | 可视化结果 |

### 2.3 重复思路说明

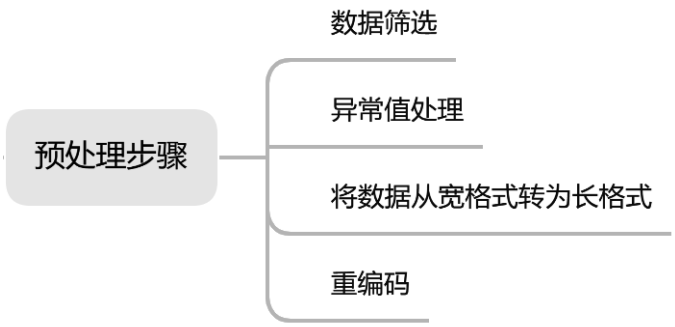
**图 1 重复思路总览**




#### 2.3.1 general 数据

##### 2.3.1.1 数据预处理

**图 2 分析步骤**




1. 数据筛选

```
#使用SQL语法筛选数据
#排除没有完成至少 4 天任务的人+被发现“撒谎”说完成了任务但没完成的人
#排除了自述重大事件影响状态的被试（这一步先前已经完成）
dfs2_general_filtered <- sqldf("select * from dfs2_general
             where T_Completed_NumDays>=4
             AND ES_Honesty_NoMission=0")
nrow(dfs2_general_filtered)
```

2. 异常值处理

分别对研究开始时和研究随访阶段的变量进行异常值处理，分别设置 2SD、2.5SD、3SD 的阈值，创建三个版本的变量（trimmed）用于后续分析控制异常值。

```
#对研究开始时的对话数量进行异常值处理
#设置 2/2.5/3 标准差的上限作为异常值门槛
limit2sd <- mean(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)+
2*sd(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)
limit25sd <- mean(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)
+2.5*sd(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)
limit3sd <- mean(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)+
3*sd(dfs2_general_filtered$SS_NumConvos_LastWk_Clean,na.rm=TRUE)
#创建新的变量来标记这些异常值（2、2.5、3个标准差）
dfs2_general_filtered$SS_NumConvos_LastWk_Clean_Trim2SD <-
ifelse(dfs2_general_filtered$SS_NumConvos_LastWk_Clean<=limit2sd,
dfs2_general_filtered$SS_NumConvos_LastWk_Clean, NA)
dfs2_general_filtered$SS_NumConvos_LastWk_Clean_Trim25SD <-
ifelse(dfs2_general_filtered$SS_NumConvos_LastWk_Clean<=limit25sd,
dfs2_general_filtered$SS_NumConvos_LastWk_Clean, NA)
dfs2_general_filtered$SS_NumConvos_LastWk_Clean_Trim3SD <-
ifelse(dfs2_general_filtered$SS_NumConvos_LastWk_Clean<=limit3sd,
dfs2_general_filtered$SS_NumConvos_LastWk_Clean, NA)
```

```
#对研究随访阶段的对话数量进行异常值处理
#设置 2/2.5/3 标准差的上限作为异常值门槛
limit2sd <- mean(dfs2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)+2*sd(df
s2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)
limit25sd <- mean(dfs2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)+2.5*sd
(dfs2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)
limit3sd <- mean(dfs2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)+3*sd(df
s2_general_filtered$FU_NumConvos_Clean,na.rm=TRUE)
#创建新的变量来标记这些异常值（2、2.5、3个标准差）
dfs2_general_filtered$FU_NumConvos_LastWk_Clean_Trim2SD <-
ifelse(dfs2_general_filtered$FU_NumConvos_Clean<=limit2sd,
dfs2_general_filtered$FU_NumConvos_Clean, NA)
dfs2_general_filtered$FU_NumConvos_LastWk_Clean_Trim25SD <-
ifelse(dfs2_general_filtered$FU_NumConvos_Clean<=limit25sd,
dfs2_general_filtered$FU_NumConvos_Clean, NA)
dfs2_general_filtered$FU_NumConvos_LastWk_Clean_Trim3SD <-
ifelse(dfs2_general_filtered$FU_NumConvos_Clean<=limit3sd,
dfs2_general_filtered$FU_NumConvos_Clean, NA)
```

3. 将宽格式转换为长格式

每个参与者一行变成每人三行，每行代表一个时间点 Time = 1/2/3，使用 SQL 的 UNION 语法拼接三个时间点的数据（start, end, follow up）

```
#将数据集从宽格式转换为长格式
#每人一行变成每人三行，拼接三个时间点time的数据（1=start, 2=end, 3=follow up）
```

```
dfs2_general_filtered_long <- sqldf("
    select 1 as Time, GooseChaseId_Fixed, Uni,T_Condition2,T_Condition3,
      SS_Demog_Age,SS_Demog_SexFemale,
      SS_Trait_SocConn_Avg7 as SS_Trait_SocConn_Avg,
      SS_Trait_SHS_Avg7 as SS_Trait_SHS_Avg,
      SS_SocialCuriosity_Avg,SS_InteractionAnxiety_Avg,SS_Shy_Avg,SS_SelfEst
eem_Avg,
      SS_Predict_NumToApp_Clean as NumApp,SS_Predict_Reject_Clean as
Reject,SS_Predict_ConvoLen_Clean as ConvoLen,
      SS_Predict_Study_ValAroConn_Avg ValAroConn,
      SS_General_Ability_Avg7 as Ability,
      SS_HardStart as HardStart,SS_HardStart_rev as AbleStart,
      SS_HardMaintain as HardMaintain,SS_HardMaintain_rev as AbleMaintain,
      SS_HardEnd as HardEnd,SS_HardEnd_rev as AbleEnd,
      SS_General_Awk_Avg7 as Awk_Avg,SS_General_Enj_Avg7 as Enj_Avg,
      SS_General_PartnerPerception_Avg7 as PartnerPerception,
      SS_DV_Avg_z as DV_Avg_z,
      SS_Strangers_Trust as Strangers_Trust,SS_Strangers_FeelWarmly as
Strangers_FeelWarmly,SS_Strangers_FeelConnected as
Strangers_FeelConnected,SS_notice_opportunities as notice_opportunities,
      '' as Num_ContactInfo, '' as Num_Communicated
      , SS_NumConvos_LastWk_Clean_Trim2SD as NumConvos_LastWk_Clean_Trim2SD
      , SS_NumConvos_LastWk_Clean_Trim25SD as
NumConvos_LastWk_Clean_Trim25SD
      , SS_NumConvos_LastWk_Clean_Trim3SD as NumConvos_LastWk_Clean_Trim3SD
    from dfs2_general_filtered
```

```
    union
    select 2 as Time, GooseChaseId_Fixed, Uni,T_Condition2,T_Condition3,
      SS_Demog_Age,SS_Demog_SexFemale,SS_Trait_SocConn_Avg7,SS_Trait_SHS_Avg7,
      SS_SocialCuriosity_Avg,SS_InteractionAnxiety_Avg,SS_Shy_Avg,SS_SelfEstee
m_Avg,
      ES_Predict_NumToApp_Clean,ES_Predict_Reject_Clean,ES_Predict_ConvoLen_Cl
ean,
      ES_Predict_ValAroConn_Avg,
      ES_General_Ability_Avg7,
      ES_HardStart,ES_HardStart_rev,ES_HardMaintain,ES_HardMaintain_rev,ES_Har
dEnd,ES_HardEnd_rev,
      ES_General_Awk_Avg7,ES_General_Enj_Avg7,
ES_General_PartnerPerception_Avg7,'',
      ES_Strangers_Trust,ES_Strangers_FeelWarmly,ES_Strangers_FeelConnected,ES
_notice_opportunities,
      ES_Num_ContactInfo_Clean,ES_Num_Communicated_Clean
      ,'','',''
    from dfs2_general_filtered
```

```
    union
      select 3 as Time, GooseChaseId_Fixed, Uni,T_Condition2,T_Condition3,
      SS_Demog_Age,SS_Demog_SexFemale,SS_Trait_SocConn_Avg7,SS_Trait_SHS_Avg7,
      SS_SocialCuriosity_Avg,SS_InteractionAnxiety_Avg,SS_Shy_Avg,SS_SelfEstee
m_Avg,
      FU_Predict_NumToApp_Clean,FU_Predict_Reject_Clean,FU_Predict_ConvoLen_Cl
ean,
      FU_Predict_ValAroConn_Avg,
      FU_General_Ability_Avg7,
      FU_HardStart,FU_HardStart_rev,FU_HardMaintain,FU_HardMaintain_rev,FU_Har
dEnd,FU_HardEnd_rev,
      FU_General_Awk_Avg7,FU_General_Enj_Avg7,
FU_General_PartnerPerception_Avg7,'',
      FU_Strangers_Trust,FU_Strangers_FeelWarmly,FU_Strangers_FeelConnected,FU
_notice_opportunities,
      '','',
      FU_NumConvos_LastWk_Clean_Trim2SD,FU_NumConvos_LastWk_Clean_Trim25SD,FU_
NumConvos_LastWk_Clean_Trim3SD
    from dfs2_general_filtered")
nrow(dfs2_general_filtered_long)
```

4. 重编码

将实验组别和时间这两个变量转换为因子

```
#把实验组别和时间变量变成有标签的因子变量

#实验组别T_Condition2：“0”=对照组，“1”=干预组
dfs2_general_filtered_long$T_Condition2 <-
as.factor(dfs2_general_filtered_long$T_Condition2)
dfs2_general_filtered_long$T_Condition2 <-
revalue(dfs2_general_filtered_long$T_Condition2,
c("0"="control","1"="treatment"))

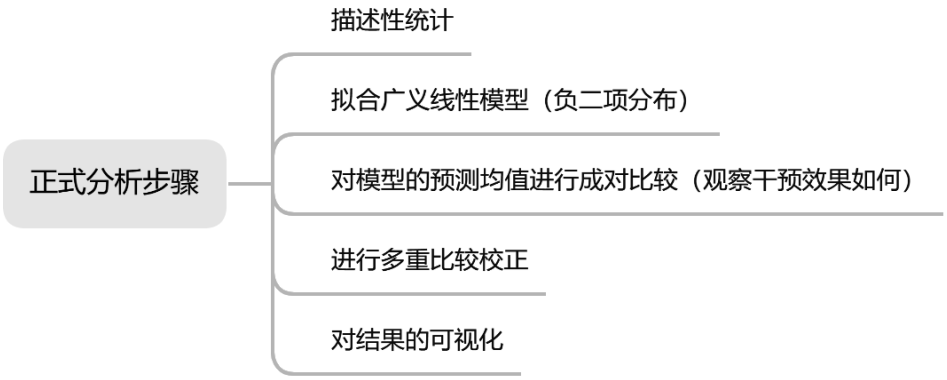
#创建一个新的变量Timei，是 Time 的数值编码版本，用于需要连续时间处理的模型
dfs2_general_filtered_long$Timei <-
as.numeric(dfs2_general_filtered_long$Time)

#时间变量Time：“1”=“start”, “2”=“end”, “3”=“followup”
dfs2_general_filtered_long$Time <- as.factor(dfs2_general_filtered_long$Time)
dfs2_general_filtered_long$Time <- revalue(dfs2_general_filtered_long$Time,
c("1"="start", "2"="end", "3"="followup"))
```

#### 2.3.1.2 数据分析

**关注：**在干预前、干预后、随访阶段三个时间点对“预期被拒绝”变量进行分析

**图 3 分析步骤**




1. 描述性统计

```
#非模型的描述性统计，按照Time和Condition进行分组
summaryBy(Reject ~ Time + T_Condition2, data = dfs2_general_filtered_long,
FUN=c(mean,sd), na.rm=TRUE)
```

**表 3 输出结果**

Description: df [6 × 4]

| | Time <fctr> | T_Condition2 <fctr> | Reject.mean <dbl> | Reject.sd <dbl> |
|---|---|---|---|---|
| 1 | start | control | NaN | *NA* |
| 2 | start | treatment | 1.2461538 | 1.578993 |
| 3 | end | control | 1.3011364 | 1.598382 |
| 4 | end | treatment | 0.3535354 | 1.010736 |
| 5 | followup | control | 1.4375000 | 1.455774 |
| 6 | followup | treatment | 0.5099338 | 1.591090 |

2. 建模及推断性统计

（1）使用使用广义线性混合模型拟合一个负二项回归模型，分析在干预前、干预后、随访阶段三个时间点干预组和对照组在“预期被拒绝”上是否存在差异

**选择负二项模型的理由：**当数据是计数且存在过度离散（方差大于均值）时，负二项比泊松回归更合适

**选择合适的模型：**

```
> mnb.general.reject <- glmer.nb(formula = Reject ~ T_Condition2*Time + (1+Time|GooseChaseId_F
ixed), data = dfs2_general_filtered_long)
```

```
错误: number of observations (=704) < number of random effects (=858) for term (1 + Time | Goo
seChaseId_Fixed); the random-effects parameters are probably unidentifiable
```

尝试将模型改为：Reject ~ T_Condition2*Time + (1+Time|GooseChaseId_Fixed)，发现模型过拟合（模型实际需要 858 个随机参数，但只有 704 个观测值），故按照文献原方法使用简化模型：Reject ~ T_Condition2*Time + (1|GooseChaseId_Fixed)

```
#########################################################
# General: Change in Fear of Rejection
#########################################################

#建立一个广义线性混合模型GLMM，使用负二项分布
#考虑组别和时间的交互作用，将每个参与者作为随机截距，考虑起始时的个体差异
mnb.general.reject <- glmer.nb(formula = Reject ~ T_Condition2*Time +
(1|GooseChaseId_Fixed), data = dfs2_general_filtered_long)
#模型摘要
summary(mnb.general.reject)
```

```
#计算边际均值，基于模型的预测均值（不是原始样本均值）
means.general.reject <- emmeans(mnb.general.reject, specs = ~ Time*T_Condition2,
data=dfs2_general_filtered_long, type = "response")
means.general.reject

#成对比较，对所有 Time × Condition 的组合进行成对比较，一共有组
#使用Wald Z检验进行成对比较
lsmlist <- contrast(means.general.reject, method = "pairwise", adjust = "none")
```

输出结果

```
> means.general.reject
 Time     T_Condition2 response     SE  df asymp.LCL asymp.UCL
 start    control        nonEst     NA  NA        NA        NA
 end      control         0.987 0.1440 Inf     0.743     1.313
 followup control         1.110 0.1710 Inf     0.821     1.500
 start    treatment       0.893 0.0930 Inf     0.728     1.095
 end      treatment       0.234 0.0354 Inf     0.174     0.315
 followup treatment       0.318 0.0489 Inf     0.235     0.430
```

```
#保留最关心的7组对比
#(6)end control vs. followup control,关注control 组随时间变化
#(7)end control vs. start treatment,关注基线对比
#(8)end control vs. end treatment,关注干预效果
#(10)followup control vs. start treatment,关注长期效果
#(12)followup control vs. followup treatment,关注干预延续性
#(13)start treatment vs. end treatment,关注干预前后变化
#(14) start treatment vs. followup treatment,关注干预延续效果
lsmlist <- lsmlist[c(6,7,8,10,12,13,14)]
#使用多变量t分布multivariate t（mvt)进行多重比较校正
mydiffs = update(lsmlist, pri.vars = "contrast", by.vars = NULL,adjust="mvt")
mydiffs

#计算置信区间
confint(mydiffs, adjust = "mvt")
```

输出结果

```
 contrast                                ratio    SE  df null z.ratio p.value
 end control / followup control           0.89 0.149 Inf    1  -0.699  0.9450
 end control / start treatment            1.11 0.192 Inf    1   0.581  0.9715
 end control / end treatment              4.22 0.858 Inf    1   7.076  <.0001
 followup control / start treatment       1.24 0.225 Inf    1   1.202  0.7046
 followup control / followup treatment    3.49 0.739 Inf    1   5.912  <.0001
 start treatment / end treatment          3.82 0.580 Inf    1   8.813  <.0001
 start treatment / followup treatment     2.81 0.434 Inf    1   6.692  <.0001

P value adjustment: mvt method for 7 tests
Tests are performed on the log scale
 contrast                                ratio    SE  df asymp.LCL asymp.UCL
 end control / followup control           0.89 0.149 Inf     0.574      1.38
 end control / start treatment            1.11 0.192 Inf     0.701      1.74
 end control / end treatment              4.22 0.858 Inf     2.471      7.20
 followup control / start treatment       1.24 0.225 Inf     0.773      2.00
 followup control / followup treatment    3.49 0.739 Inf     2.003      6.09
 start treatment / end treatment          3.82 0.580 Inf     2.559      5.69
 start treatment / followup treatment     2.81 0.434 Inf     1.873      4.22

Confidence level used: 0.95
Conf-level adjustment: mvt method for 7 estimates
Intervals are back-transformed from the log scale
```

（1）**Treatment** Start vs. End 关注干预前后变化：对于治疗组，为期一周的干预结束时，参与者预期遭到拒绝的人数少于研究开始时（ratio = 3.82, $p < 0.001$），说明干预有效

（2）**Treatment** Start vs. Follow-up 关注干预延续效果：对于治疗组，干预结束一周后，参与者预期遭到拒绝的人数仍然少于研究开始时（ratio = 2.81, $p < 0.001$），说明干预产生了持久的效果

（3）End Control vs. Start Treatment 关注基线对比：(控制组实验结束时预测 & 实验组在初始阶段预测)的被拒绝的人数一样多（ratio = 1.11, $p = 0.97$）

（4）**Control** End vs. Follow-up 关注 control 组随时间变化（对照）：对照组参与者的被拒绝信念从研究结束到随访期间没有发生变化（ratio = 0.89, $p = 0.95$）

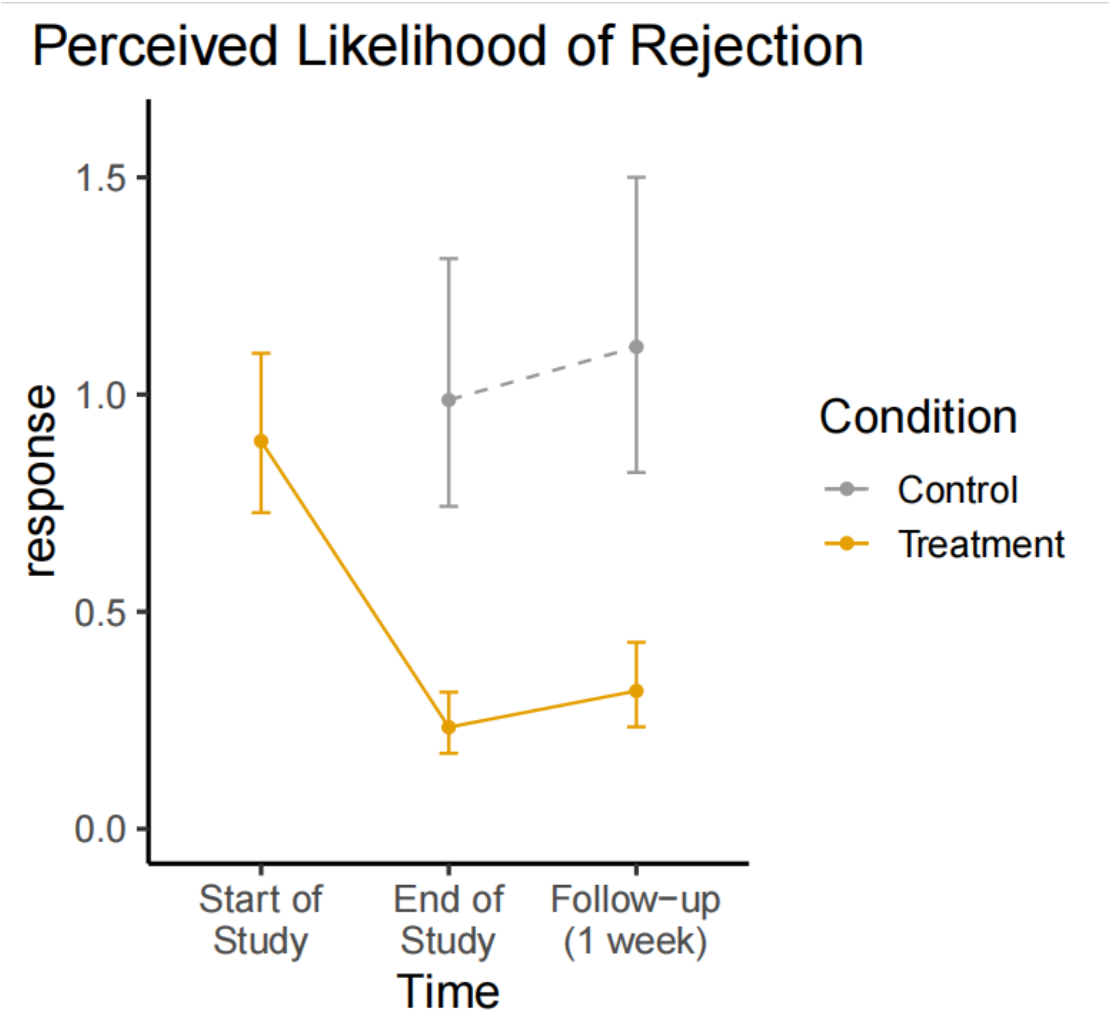
（5）**End** Control vs. Treatment **关注干预效果**：研究结束时，治疗组的参与者预期遭到拒绝的人数显著少于对照组（ratio = 4.22, $p < 0.001$）

（6）**Follow-up** Control vs. Treatment 对比干预延续性：随访时期，治疗组的参与者预期遭到拒绝的人数显著少于对照组（ratio = 3.49, $p < 0.001$），干预效果至少持续了一周

**总体而言，干预组的变化显著大于对照组的变化，干预有效**

3. 可视化

**图 4 不同条件下预期被拒绝的可能性**




### 2.3.2 daily 数据

#### 2.3.1.1 数据预处理

首先用 SQL 语法筛选数据，其次对变量进行清洗与重编码

```
## daily analyses数据预处理
```{r}
dfs2_daily_filtered <- sqldf("select * from dfs2_daily
            where T_Pilot_ONo1Yes=0 #只保留非试验性（pilot）数据
            AND T_Completed_NumDays>=4 #只保留完成至少 4 天任务的参与者
            AND ES_Honesty_NoMission=0 #排除没有完成诚实任务的参与者
            AND SS_Include_NoYes=1") #只保留通过质量筛选的参与者
nrow(dfs2_daily_filtered) #检测筛选后的样本数量

#变量清洗与重新编码
dfs2_daily_filtered$RatingType <- as.factor(dfs2_daily_filtered$RatingType) #把 RatingType 变量从数字类型转换为因子
dfs2_daily_filtered$RatingType <- revalue(dfs2_daily_filtered$RatingType, c("0"="prediction", "1"="actual"))
#重命名因子水平为更有意义的标签
dfs2_daily_filtered$Day <- as.numeric(dfs2_daily_filtered$Day) #确保 Day 是数值型
dfs2_daily_filtered$Dayf <- as.factor(dfs2_daily_filtered$Day) #创建因子变量 Dayf（星期几）
dfs2_daily_filtered$Dayf <- revalue(dfs2_daily_filtered$Dayf, c("1"="Monday", "2"="Tuesday", "3"="Wednesday", "4"="Thursday",
"5"="Friday")) ##重命名因子水平为具体日期
```

#### 2.3.1.2 正式分析

##### 2.3.1.2.1 因变量为 Reject

1. 描述性统计（当把天数 Day 作为连续变量处理时）计算 Reject_Clean 在不同天数与评分类型（预测 vs 实际）下的描述性平均值与标准差。
```

```r
summaryBy(Reject_Clean ~ Day + RatingType, data = dfs2_daily_filtered, FUN=c(mean,sd), na.rm=TRUE) #先计算 Reject_Clean
在不同天数与评分类型（预测 vs 实际）下的描述性平均值与标准差
means.daily.reject <- emmeans(mnb.daily.reject, specs = ~ RatingType*Day, type = "response") #
means.daily.reject
```

输出结果

```
> summaryBy(Reject_Clean ~ Day + RatingType, data = dfs2_daily_filtered, FUN=c(mean,sd), na.rm=TRUE)
   Day RatingType Reject_Clean.mean Reject_Clean.sd
1    1 prediction         0.9717949       1.1689604
2    1     actual         0.1024735       0.3767548
3    2 prediction         0.7458101       1.0213572
4    2     actual         0.2627737       0.8876035
5    3 prediction         0.5828571       0.8042580
6    3     actual         0.2377049       1.3150179
7    4 prediction         0.6988950       1.2579520
8    4     actual         0.1547619       0.5537424
9    5 prediction         0.4869792       0.9229972
10   5     actual         0.2037736       0.5806091
```

2. 建模及推断性统计：

（1）使用广义线性混合模型拟合一个负二项回归模型,分析每天的拒绝数是否受到 RatingType（预测 vs 实际）和 Day（星期几）及其交互作用的影响。

并使用 confint 函数计算混合模型中各个参数的置信区间，使用 summ 函数输出结果更美观整洁。

输出结果显示

```r
mnb.daily.reject <- glmer.nb(formula = Reject_Clean ~ RatingType*Day + (1|GooseChaseId_Fixed), data = dfs2_daily_filtered)
#使用广义线性混合模型拟合一个负二项回归模型,分析 每天的拒绝数是否受到 RatingType（预测 vs 实际）和
Day（星期几）及其交互作用的影响
confint(mnb.daily.reject) #计算主效应、交互效应的置信区间
summ(mnb.daily.reject,digit=4)
```

```
> summ(mnb.daily.reject,digit=4)
MODEL INFO:
Observations: 2240
Dependent Variable: Reject_Clean
Type: Mixed effects generalized linear regression
Error Distribution: Negative Binomial(2.8168)
Link function: log

MODEL FIT:
AIC = 3172.2057, BIC = 3206.4911
Pseudo-R² (fixed effects) = 0.1475
Pseudo-R² (total) = 0.4786

FIXED EFFECTS:
----------------------------------------------------------------
                               Est.     S.E.     z val.        p
-------------------------- --------- -------- ---------- --------
(Intercept)                  -0.3127   0.1299    -2.4076   0.0161
RatingTypeactual             -1.9915   0.1921   -10.3654   0.0000
Day                          -0.1818   0.0332    -5.4716   0.0000
RatingTypeactual:Day          0.2251   0.0582     3.8665   0.0001
----------------------------------------------------------------

RANDOM EFFECTS:
-------------------------------------------------
       Group           Parameter    Std. Dev.
------------------- ------------- -----------
 GooseChaseId_Fixed   (Intercept)    1.0074
-------------------------------------------------

Grouping variables:
-----------------------------------------
       Group          # groups     ICC
------------------- ---------- --------
 GooseChaseId_Fixed     198      0.4939
-----------------------------------------
```

```
> confint(mnb.daily.reject)
Computing profile confidence intervals ...
                              2.5 %      97.5 %
.sig01                    0.8637278  1.17762165
(Intercept)              -0.5712415 -0.06136974
RatingTypeactual         -2.3731073 -1.61941654
Day                      -0.2472679 -0.11697468
RatingTypeactual:Day      0.1111341  0.33941752
>
```

（2）使用（emmeans）获取基于模型估计的预测边际均值

```
> means.daily.reject <- emmeans(mnb.daily.reject, specs = ~ RatingType*Day, type = "response")
> means.daily.reject
 RatingType  Day response     SE  df asymp.LCL asymp.UCL
 prediction 2.97    0.426 0.0405 Inf    0.3536     0.513
 actual     2.97    0.114 0.0125 Inf    0.0915     0.141

Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

输出结果表明：在平均 Day 值（约星期三）下，prediction 比 actual 更容易出现 Reject 行为，且差异显著（置信区间不重叠）。

（3）检验预测拒绝数与实际拒绝数是否有显著差异

通过（emmeans）获得边际均值（means. daily. reject）后，**对不同条件之间的差异进行事后比较（pairwise contrast）**，并输出其置信区间（confidence intervals）

```r
contr.all <- contrast(means.daily.reject, method = "pairwise", adjust = "none")
contr.all
confint(contr.all, adjust = "none")
```

输出结果

```
> contr.all <- contrast(means.daily.reject, method = "pairwise", adjust = "none")
> contr.all
 contrast                                                ratio    SE  df null z.ratio p.value
 prediction Day2.97232142857143 / actual Day2.97232142857143  3.75 0.319 Inf    1  15.578  <.0001

Tests are performed on the log scale
> confint(contr.all, adjust = "none")
 contrast                                                ratio    SE  df asymp.LCL asymp.UCL
 prediction Day2.97232142857143 / actual Day2.97232142857143  3.75 0.319 Inf      3.18      4.43

Confidence level used: 0.95
Intervals are back-transformed from the log scale
>
```

结果表明：在 Day≈2.97（大约是周三）那天，平均预测拒绝数（prediction）是平均实际拒绝数 actual 的 3.75 倍，且这个差异显著（$z = 15.578$ $p < .0001$），95%CI 为 [3.18, 4.43]。

（4）简单斜率检验：RatingType × Day 的**交互效应**是否显著

检验预测拒绝数与实际拒绝数在天数（Day）上的**斜率（趋势）是否显著不同**。

```r
#简单斜率检验
emtrends(mnb.daily.reject, ~ RatingType, var="Day")
test(emtrends(mnb.daily.reject, ~ RatingType, var="Day"))
```

输出结果

```
> emtrends(mnb.daily.reject, ~ RatingType, var="Day")
 RatingType Day.trend     SE  df asymp.LCL asymp.UCL
 prediction   -0.1818 0.0332 Inf   -0.2469    -0.117
 actual        0.0433 0.0478 Inf   -0.0505     0.137

Confidence level used: 0.95
> #source for emtrends: https://stats.idre.ucla.edu/r/seminars/interactions-r/#s4b
> test(emtrends(mnb.daily.reject, ~ RatingType, var="Day"))
 RatingType Day.trend     SE  df z.ratio p.value
 prediction   -0.1818 0.0332 Inf  -5.472  <.0001
 actual        0.0433 0.0478 Inf   0.904  0.3659
```

Predicted的斜率

Actual的斜率

结果表明：预测被拒绝数随干预进程（Day）显著降低；而实际被拒绝数随干预进程（Day）变化不显著。

3. 把 Day 转化为因子重建模型。将 Day 设为因子（Dayf）后，每一天就可以被视为一个独立水平。论文最终选择 Day 为因子时，2（Rating type）×5（Day）条件下，拒绝次数的预测均值和标准差（emmeans），以此报告每个“具体时间点”（如星期几）的差异和平均数。

cc

```r
# 把day由数字转为因子，再次建模分析
mnb.daily.reject.categorical <- glmer.nb(formula = Reject_Clean ~ RatingType*Dayf + (1|GooseChaseId_Fixed), data = dfs2_daily_filtered)
means.daily.reject.categorical <- emmeans(mnb.daily.reject.categorical, specs = ~ RatingType*Dayf, type = "response")
means.daily.reject.categorical_RatingType <- emmeans(mnb.daily.reject.categorical, specs = ~ RatingType, type = "response")
```

输出结果

```
> means.daily.reject.categorical
 RatingType Dayf      response     SE  df asymp.LCL asymp.UCL
 prediction Monday      0.6522 0.0780 Inf    0.5159    0.8245
 actual     Monday      0.0644 0.0136 Inf    0.0425    0.0974
 prediction Tuesday     0.4678 0.0616 Inf    0.3614    0.6057
 actual     Tuesday     0.1528 0.0237 Inf    0.1128    0.2070
 prediction Wednesday   0.3644 0.0514 Inf    0.2764    0.4804
 actual     Wednesday   0.1433 0.0240 Inf    0.1032    0.1991
 prediction Thursday    0.4110 0.0553 Inf    0.3157    0.5351
 actual     Thursday    0.0952 0.0181 Inf    0.0656    0.1382
 prediction Friday      0.2899 0.0418 Inf    0.2185    0.3845
 actual     Friday      0.1165 0.0199 Inf    0.0834    0.1627

Confidence level used: 0.95
Intervals are back-transformed from the log scale
> means.daily.reject.categorical_RatingType
 RatingType response     SE  df asymp.LCL asymp.UCL
 prediction    0.421 0.0400 Inf    0.3496     0.507
 actual        0.109 0.0122 Inf    0.0879     0.136

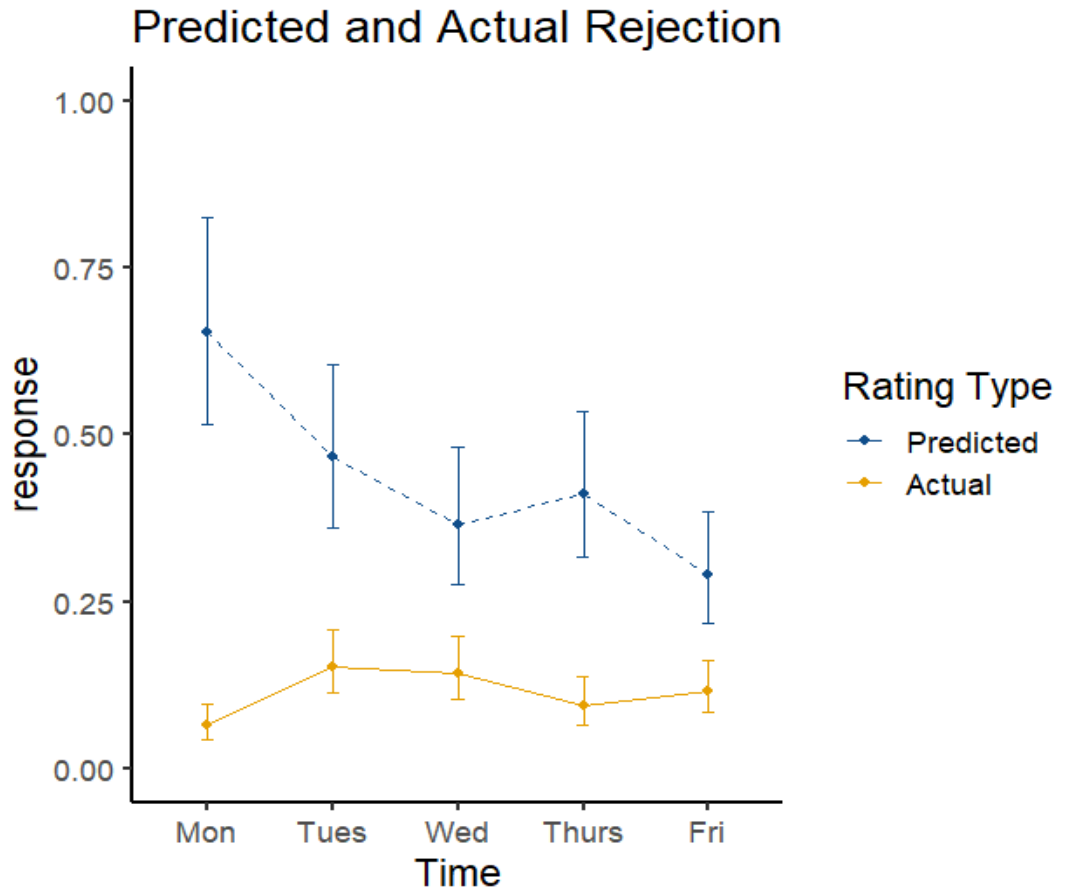
Results are averaged over the levels of: Dayf
Confidence level used: 0.95
Intervals are back-transformed from the log scale
>
```

4. 可视化

```
##可视化
```{r}
# Panel A: Rejection by day, talkers' daily predictions vs. experience
plot_daily_reject <- ggplot(as.data.frame(means.daily.reject.categorical), aes(x = Dayf, y = response, group=RatingType)) +
  geom_line(aes(color = RatingType, linetype=RatingType)) + geom_point(aes(color = RatingType)) +
  geom_errorbar(aes(ymin=asymp.LCL, ymax=asymp.UCL, colour = RatingType), width=.1) +
  scale_color_manual("Rating Type",values=c('dodgerblue4','#E69F00'),labels=c("Predicted","Actual")) +
  scale_linetype_manual("Rating Type",values=c("dashed","solid"),labels=c("Predicted","Actual")) +
  theme_classic(base_size = 15) + theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle("Predicted and Actual Rejection")+
  scale_x_discrete("Time",labels=c("Monday"="Mon", "Tuesday"="Tues", "Wednesday"="Wed", "Thursday"="Thurs", "Friday"="Fri")) +
  scale_y_continuous(limits=c(0,1))
```
```

输出结果

**图 5**




**2.3.1.2.2 因变量为 Ability**

1. 描述性统计

描述性统计（当把天数 Day 作为连续变量处理时）计算 Ability_Avg7 在不同天数与评分类型（预测 vs 实际）下的描述性平均值与标准差。

```
#计算 Ability_Avg7 在不同天数与评分类型（预测 vs 实际）下的描述性平均值与标准差
summaryBy(Ability_Avg7 ~ Day + RatingType, data = dfs2_daily_filtered, FUN=c(mean,sd), na.rm=TRUE)
```

输出结果

```
> summaryBy(Ability_Avg7 ~ Day + RatingType, data = dfs2_daily_filtered, FUN=c(mean,sd), na.rm=TRUE)
   Day RatingType Ability_Avg7.mean Ability_Avg7.sd
1    1 prediction          4.176068        1.141379
2    1     actual          5.013498        1.116107
3    2 prediction          4.387037        1.056542
4    2     actual          5.098182        1.158757
5    3 prediction          4.718456        1.073668
6    3     actual          5.306122        1.154379
7    4 prediction          4.717033        1.168991
8    4     actual          5.312336        1.196355
9    5 prediction          4.944444        1.165399
10   5     actual          5.323270        1.192315
```

2. 建模及推断性统计

（1）建立新模型

由于因变量“会话能力”Ability 为连续变量，使用混合效应模型。

原文献模型：Ability_Avg7 ~ RatingType*Day + (1|GooseChaseId_Fixed)

建立新模型的考虑：

1. 不同参与者的能力随时间的变化趋势可能不同，故将 Day 考虑为随机斜率；

2. 二分变量 RatingType 的编码设置为-0.5 和 0.5 时（新变量 R_T），线性混合模型输出的固定效应即为真实的主效应和交互效应。

新模型如下：

Ability_Avg7 ~ R_T*Day + (1+Day|GooseChaseId_Fixed)

使用 anova 语句进行模型对比，看模型是否有显著差异。

```
# ability
#使用线性混合效应模型，分析每天的会话能力Ability是否受到 RatingType（预测 vs 实际）和
Day（星期几）及其交互作用的影响
#原初模型：仅考虑被试作为随机截距
lm.daily.ability1 <- lmer(formula = Ability_Avg7 ~ RatingType*Day +
(1|GooseChaseId_Fixed), data = dfs2_daily_filtered)
#使用summ函数输出原初模型摘要
summ(lm.daily.ability1)
#使用confint函数计算混合模型中各个参数的置信区间
confint(lm.daily.ability1)

#新模型
#将 T_Condition2 列中的 0 和 1 编码转换为 -0.5 和 0.5
library(dplyr)
dfs2_daily_filtered <- dfs2_daily_filtered %>%
  mutate(R_T = recode(RatingType, "prediction" = -0.5, "actual" = 0.5))
#新模型：考虑Day 的效应（即时间趋势）在不同被试间存在显著差异
lm.daily.ability2 <- lmer(formula = Ability_Avg7 ~ R_T*Day +
(1+Day|GooseChaseId_Fixed), data = dfs2_daily_filtered)
#使用summ函数输出新模型摘要
summ(lm.daily.ability2)
#使用confint函数计算混合模型中各个参数的置信区间
confint(lm.daily.ability2)
#模型比较
anova(lm.daily.ability1,lm.daily.ability2)
```

模型比较输出结果：

$Chi^2(2)=74.476$, $p < 0.001$，似然比检验结果表明，**模型 2 显著优于模型 1**

模型 2 有更低的 AIC/BIC；更高的对数似然 loglik；极显著的似然比检验（$p < 0.001$）

```
> #模型比较
> anova(lm.daily.ability1,lm.daily.ability2)
refitting model(s) with ML (instead of REML)
Data: dfs2_daily_filtered
Models:
lm.daily.ability1: Ability_Avg7 ~ RatingType * Day + (1 | GooseChaseId_Fixed)
lm.daily.ability2: Ability_Avg7 ~ R_T * Day + (1 + Day | GooseChaseId_Fixed)
                  npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
lm.daily.ability1    6 6285.8 6320.1 -3136.9    6273.8
lm.daily.ability2    8 6215.3 6261.0 -3099.7    6199.3 74.476  2  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

新模型输出结果：

```
> #使用summ函数输出新模型摘要
> summ(lm.daily.ability2)
```

MODEL INFO:
*Observations:* 2249
*Dependent Variable:* Ability_Avg7
*Type:* Mixed effects linear regression

MODEL FIT:
*AIC* = 6235.68, *BIC* = 6281.43
*Pseudo-R² (fixed effects)* = 0.09
*Pseudo-R² (total)* = 0.49

FIXED EFFECTS:

| | Est. | S.E. | t val. | d.f. | p |
|---|---|---|---|---|---|
| (Intercept) | 4.47 | 0.07 | 63.75 | 194.12 | 0.00 |
| R_T | 0.89 | 0.09 | 10.46 | 1928.25 | 0.00 |
| Day | 0.14 | 0.02 | 7.28 | 191.68 | 0.00 |
| R_T:Day | -0.10 | 0.03 | -3.79 | 1933.64 | 0.00 |

*p values calculated using Satterthwaite d.f.*

RANDOM EFFECTS:

| Group | Parameter | Std. Dev. |
|---|---|---|
| GooseChaseId_Fixed | (Intercept) | 0.78 |
| GooseChaseId_Fixed | Day | 0.19 |
| Residual | | 0.85 |

Grouping variables:

| Group | # groups | ICC |
|---|---|---|
| GooseChaseId_Fixed | 198 | 0.46 |

```
> #使用confint函数计算混合模型中各个参数的置信区间
> confint(lm.daily.ability2)
```

```
Computing profile confidence intervals ...
```

```
                  2.5 %      97.5 %
.sig01        0.6589364  0.90739223
.sig02       -0.6390470 -0.29395050
.sig03        0.1574143  0.23104855
.sigma        0.8195359  0.87411247
(Intercept)   4.3297543  4.60506881
R_T           0.7257220  1.06032977
Day           0.1007384  0.17520750
R_T:Day      -0.1487624 -0.04739118
```

（2）使用（emmeans）获取新模型估计的预测边际均值

```
#使用（emmeans）获取基于模型估计的预测边际均值
means.daily.ability2 <- emmeans(lm.daily.ability2, specs = ~ R_T*Day, type =
"response")
means.daily.ability2
```

输出结果

```
> means.daily.ability2 <- emmeans(lm.daily.ability2, specs = ~ R_T*Day, type = "response")
> means.daily.ability2
  R_T  Day emmean     SE  df lower.CL upper.CL
 -0.5 2.97   4.58 0.0574 261     4.46     4.69
  0.5 2.97   5.18 0.0559 235     5.07     5.29
```

输出结果表明：在平均 Day 值（约星期三）下，实际会话能力比预测会话能力更强，且差异显著（置信区间不重叠），与原模型和原文献结果一致。

（3）检验预测会话能力与实际会话能力是否有显著差异

```
#通过 emmeans() 获得边际均值（means.daily.ability）后，对不同条件之间的差异进行事后比较
（pairwise contrast），并输出其置信区间（confidence intervals）
contr.all <- contrast(means.daily.ability2, method = "pairwise", adjust = "none")
contr.all
confint(contr.all, adjust = "none")
```

输出结果

```
> contr.all <- contrast(means.daily.ability2, method = "pairwise", adjust = "none")
> contr.all
 contrast                                                  estimate     SE   df
 (R_T-0.5 Day2.97198755002223) - R_T0.5 Day2.97198755002223   -0.601 0.0373 1896
 t.ratio p.value
 -16.127  <.0001

Degrees-of-freedom method: kenward-roger
> confint(contr.all, adjust = "none")
 contrast                                                  estimate     SE   df
 (R_T-0.5 Day2.97198755002223) - R_T0.5 Day2.97198755002223   -0.601 0.0373 1896
 lower.CL upper.CL
   -0.675   -0.528

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
```

结果表明：在 Day≈2.97（大约是周三）那天，平均预测会话能力（prediction）比平均实际会话能力（actual）低 0.6，且这个差异显著 $t(1896) = -16.13, p < 0.0001$，95%CI 为 [-0.68, -0.53]，与原模型（$\Delta M = 0.61$，$t(2069) = -15.55, p < 0.0001$，95%CI 为 [-0.69, -0.53]）几乎一致。

（4）简单斜率检验：检验预测会话能力与实际会话能力在天数（Day）上的**斜率（趋势）是否显著不同**。

```
#简单斜率检验:检验预测会话能力与实际会话能力在天数（Day）上的斜率（趋势）是否显著不同
#source for emtrends: https://stats.idre.ucla.edu/r/seminars/interactions-r/#s4b
test(emtrends(lm.daily.ability2, ~ R_T, var="Day"))

#置信区间
emtrends(lm.daily.ability2, ~ R_T, var="Day")
```

输出结果：

```
> #简单斜率检验:检验预测会话能力与实际会话能力在天数（Day）上的斜率（趋势）是否显著不同
> #source for emtrends: https://stats.idre.ucla.edu/r/seminars/interactions-r/#s4b
> test(emtrends(lm.daily.ability2, ~ R_T, var="Day"))
  R_T Day.trend     SE  df t.ratio p.value
 -0.5     0.187 0.0239 484   7.841  <.0001
  0.5     0.089 0.0220 336   4.043  0.0001

Degrees-of-freedom method: kenward-roger
>
> #置信区间
> emtrends(lm.daily.ability2, ~ R_T, var="Day")
  R_T Day.trend     SE  df lower.CL upper.CL
 -0.5     0.187 0.0239 484   0.1402    0.234
  0.5     0.089 0.0220 336   0.0457    0.132

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
```

Predicted的斜率

Actual的斜率

结果表明：预测会话能力随干预进程（Day）显著提高 $t(484) = 7.84, p < 0.0001$；实际会话能力也显著提高 $t(336) = 4.04, p < 0.0001$

新的模型自由度大幅减少，是混合模型中引入随机斜率后常见且合理的现象，它反映更谨慎的统计推断；从参数估计、显著性与置信区间的结果来看，和原模型结果一致，支持新模型的可靠性。

3. 把 Day 转化为因子重建模型。将 Day 设为因子（Dayf）后，每一天就可以被视为一个独立水平，报告每个“具体时间点”（如星期几）的平均预测会话能力和平均实际会话能力。

```
#把Day转化为因子重建模型。将 Day 设为因子（Dayf）后，每一天就可以被视为一个独立水平
#报告每个“具体时间点”（如星期几）的平均预测会话能力和平均实际会话能力
lm.daily.ability2.categorical <- lmer(formula = Ability_Avg7 ~ RatingType*Dayf +
(1+Day|GooseChaseId_Fixed), data = dfs2_daily_filtered)
means.daily.ability2.categorical <- emmeans(lm.daily.ability2.categorical, specs = ~
RatingType*Dayf, type = "response")
print(means.daily.ability2.categorical)
```

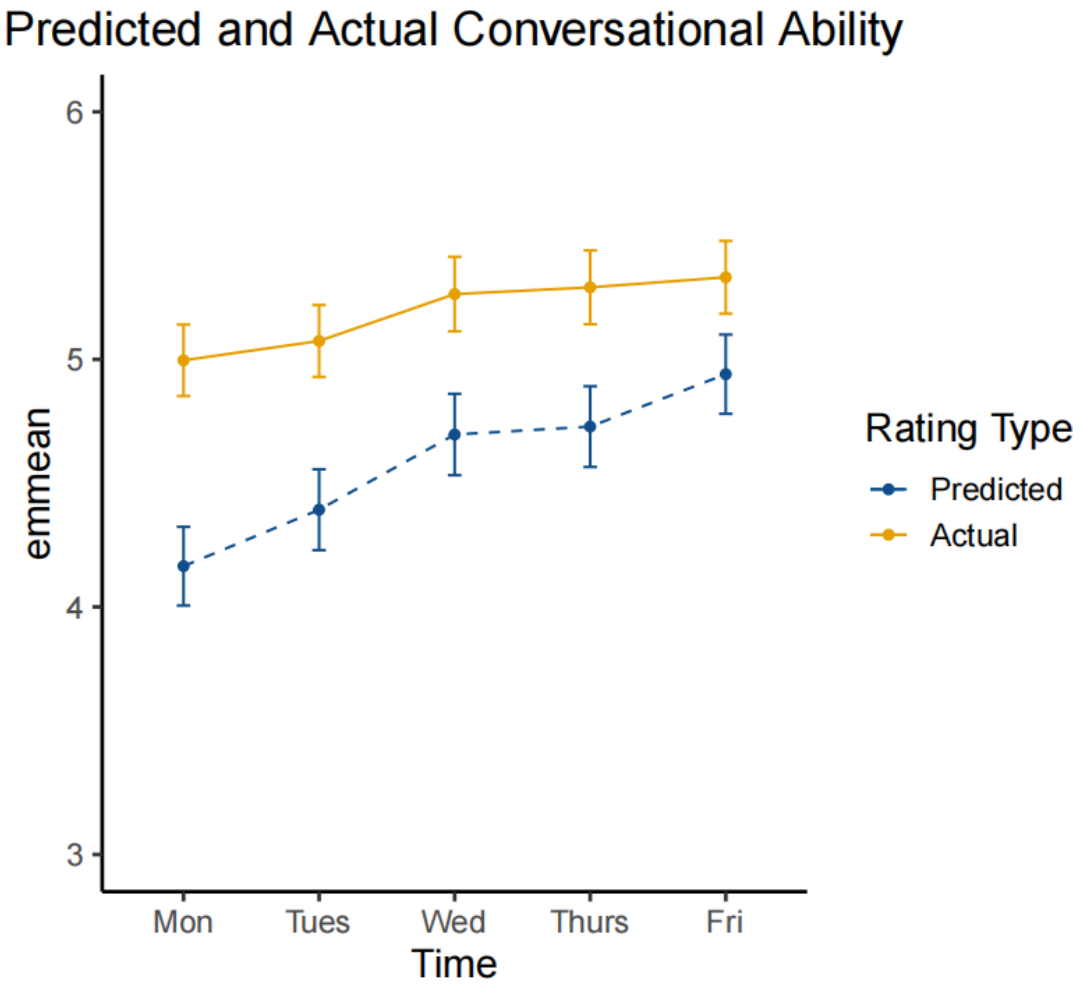
输出结果

| RatingType | Dayf | emmean | SE | df | lower.CL | upper.CL |
|---|---|---|---|---|---|---|
| prediction | Monday | 4.17 | 0.0787 | 582 | 4.01 | 4.32 |
| actual | Monday | 4.98 | 0.0726 | 418 | 4.84 | 5.12 |
| prediction | Tuesday | 4.40 | 0.0796 | 850 | 4.24 | 4.55 |
| actual | Tuesday | 5.07 | 0.0711 | 584 | 4.93 | 5.21 |
| prediction | Wednesday | 4.69 | 0.0812 | 885 | 4.54 | 4.85 |
| actual | Wednesday | 5.26 | 0.0747 | 680 | 5.12 | 5.41 |
| prediction | Thursday | 4.72 | 0.0841 | 685 | 4.56 | 4.89 |
| actual | Thursday | 5.28 | 0.0782 | 531 | 5.13 | 5.44 |
| prediction | Friday | 4.94 | 0.0885 | 451 | 4.76 | 5.11 |
| actual | Friday | 5.31 | 0.0838 | 362 | 5.15 | 5.48 |

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

4. 可视化

图 6




# 3 结果

## 3.1 描述性统计

对原文献描述性统计进行重复的结果，并汇总表格：

**表 4 因变量 Reject（general）描述性统计结果**

| Time | Condition | Reject | | |
|---|---|---|---|---|
| | | *N* | Mean | *SD* |
| Start | Control | 88 | NaN | NA |
| Start | Treatment | 198 | 1.25 | 1.58 |
| End | Control | 88 | 1.30 | 1.60 |
| End | Treatment | 198 | 0.35 | 1.01 |
| Follow up | Control | 88 | 1.44 | 1.46 |
| Follow up | Treatment | 198 | 0.51 | 1.59 |

**表 5 因变量 Reject（daily）描述性统计结果**

| Day | Rating Type | | | | | |
|---|---|---|---|---|---|---|
| | Predicted | | | Actual | | |
| | *N* | Mean | *SD* | *N* | Mean | *SD* |
| Monday | 198 | 0.97 | 1.17 | 198 | 0.10 | 0.38 |
| Tuesday | 198 | 0.10 | 0.38 | 198 | 0.75 | 1.02 |
| Wednesday | 198 | 0.75 | 1.02 | 198 | 0.26 | 0.89 |
| Thursday | 198 | 0.26 | 0.89 | 198 | 0.58 | 0.80 |
| Friday | 198 | 0.58 | 0.80 | 198 | 0.24 | 1.32 |
| All | 198 | 0.70 | 1.06 | 198 | 0.20 | 0.80 |

**表 6 general 分析：因变量 Reject 描述性统计结果（模型估计的边际均值）**

| | | Reject | | | | | |
|---|---|---|---|---|---|---|---|
| | | Treatment | | | Control | | |
| | | *N* | Mean | *SE* | *N* | Mean | *SE* |
| Start of Study | 原研究 | 198 | 0.89 | 0.09 | 88 | n/a | n/a |
| | 本研究 | 198 | 0.893 | 0.093 | 88 | n/a | n/a |
| | $\delta$ | 0% | 0% | 0% | 0% | 0% | 0% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| End of Study | 原研究 | 198 | 0.23 | 0.04 | 88 | 0.99 | 0.14 |
| | 本研究 | 198 | 0.234 | 0.035 | 88 | 0.987 | 0.144 |
| | $\delta$ | 0% | 0% | 0% | 0% | 0% | 0% |
| Follow-up | 原研究 | 198 | 0.32 | 0.05 | 88 | 1.11 | 0.17 |
| | 本研究 | 198 | 0.318 | 0.048 | 88 | 1.110 | 0.171 |
| | $\delta$ | 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 |

**表 7 daily 分析：因变量 Reject 描述性统计结果（模型估计的边际均值）**

| | Daily experiences of rejection | | | | | |
|---|---|---|---|---|---|---|
| | Predicted | | | Actual | | |
| | *N* | Mean | *SD* | *N* | Mean | *SD* |
| 原研究报告结果 | 198 | 0.43 | 0.04 | 198 | 0.11 | 0.01 |
| 本研究 | 198 | 0.43 | 0.04 | 198 | 0.11 | 0.01 |
| $\delta$ | 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 |

**表 8 daily 分析：因变量 Ability 的描述性统计结果（模型估计的边际均值）（创新方法）**

| | Ability | | | | | |
|---|---|---|---|---|---|---|
| | Predicted | | | Actual | | |
| | *N* | Mean | *SD* | *N* | Mean | *SD* |
| 原研究报告结果 | 198 | 4.58 | 0.06 | 198 | 5.19 | 0.06 |
| 本研究 | 198 | 4.58 | 0.057 | 198 | 5.19 | 0.056 |
| $\delta$ | 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 |

## 3.2 推断性统计

### 3.2.1 使用与原文献相同方法的推断性统计

报告对原文献推断性统计进行重复的结果，并汇总表格：

**表 9 general-不同条件下被拒绝人数的成对比较**

| | | 不同条件的比较观测干预效果 | | | |
|---|---|---|---|---|---|
| | | df | 统计量 (Ratio) | 置信区间 (95%CI) | 显著性指标(*p*) |
| **Treatment** | 原文献 | Inf | 3.82 | [2.56, 5.69] | < 0.001 |
| Start vs. | 本研究 | Inf | 3.82 | [2.559, 5.69] | <.0001 |
| End | $\delta$ | 0% | 0% | 0% | 0% |
| **Treatment** | 原文献 | Inf | 2.81 | [1.87, 4.22] | < 0.001 |
| Start vs. | 本研究 | Inf | 2.81 | [1.873, 4.22] | <.0001 |
| Follow-up | $\delta$ | 0% | 0% | 0% | 0% |
| End Control | 原文献 | Inf | 1.11 | [0.70, 1.74] | 0.97 |
| vs. Start | 本研究 | Inf | 1.11 | [0.701, 1.74] | 0.9715 |
| Treatment | $\delta$ | 0% | 0% | 0% | 0% |
| **Control** | 原文献 | Inf | 0.89 | [0.57, 1.38] | 0.95 |
| End vs. | 本研究 | Inf | 0.89 | [0.574, 1.38] | 0.9450 |
| Follow-up | $\delta$ | 0% | 0% | 0% | 0% |
| **End** | 原文献 | Inf | 4.22 | [2.47, 7.20] | < 0.001 |
| Control vs. | 本研究 | Inf | 4.22 | [2.472, 7.20] | <.0001 |
| Treatment | $\delta$ | 0% | 0% | 0% | 0% |
| **Follow-up** | 原文献 | Inf | 3.49 | [2.00, 6.09] | < 0.001 |
| Control vs. | 本研究 | Inf | 3.49 | [2.003, 6.09] | <.0001 |
| Treatment | $\delta$ | 0% | 0% | 0% | 0% |
| 评级 | | 完全一致 | 完全一致 | 完全一致 | 完全一致 |

**表 10　daily-Reject 一般线性模型的推断性统计结果比较**

| | | *b* | *SE* | *z* | *p* | 95%CI |
|---|---|---|---|---|---|---|
| 原文献 | Rating Type:Day | 0.23 | 0.06 | 3.87 | <.001 | [0.11, 0.34] |

| | | | | | | |
|---|---|---|---|---|---|---|
| 报告结果 | Predicted | -0.18 | 0.03 | -5.47 | <.001 | [-0.25, -0.12] |
| | Actual | 0.04 | 0.05 | 0.90 | 0.37 | [-0.05, 0.14] |
| | Rating Type:Day | 0.23 | 0.06 | 3.87 | <.001 | [0.11, 0.34] |
| 本研究 | Predicted | -0.18 | 0.03 | -5.47 | <.001 | [-0.25, -0.12] |
| | Actual | 0.04 | 0.05 | 0.90 | 0.37 | [-0.05, 0.14] |
| *δ* | | 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 |

### 3.2.2 使用与原文献不同方法的推断性统计

对采用新方法进行数据分析的描述性或推断性结果进行描述，并记录在表格中。

**表 11 因变量 Ability 的推断性统计结果比较(创新方法)**

| | | *b* | *SE* | *df* | *t* | *p* | 95%CI |
|---|---|---|---|---|---|---|---|
| Rating Type:Day | 原文献报告结果 | -0.10 | 0.03 | 2057.23 | -3.72 | <.001 | [-0.15, -0.05] |
| | 本研究 | -0.10 | 0.03 | 1933.64 | -3.79 | <.001 | [-0.15, -0.05] |
| | *δ* | 0% | 0% | 6.391% | 1.847% | 0% | 0% |
| | 评级 | 完全一致 | 完全一致 | 次要偏差 | 次要偏差 | 完全一致 | 完全一致 |
| Predicted | 原文献报告结果 | 0.19 | 0.02 | 2052 | 9.23 | <.001 | [0.15, 0.23] |
| | 本研究 | 0.19 | 0.02 | 484 | 7.84 | <.001 | [0.14, 0.23] |
| | *δ* | 0% | 0% | 324% | 17.73% | 0% | 7.143%/0% |
| | 评级 | 完全一致 | 完全一致 | 主要偏差 | 主要偏差 | 完全一致 | 次要偏差/完全一致 |
| Actual | 原文献报告结果 | 0.09 | 0.02 | 2065 | 5.12 | <.001 | [0.06, 0.12] |
| | 本研究 | 0.09 | 0.02 | 336 | 4.04 | <.001 | [0.05, 0.13] |
| | *δ* | 0% | 0% | 514.58% | 26.73% | 0% | 20%/7.692% |
| | 评级 | 完全一致 | 完全一致 | 主要偏差 | 主要偏差 | 完全一致 | 主要偏差/次要偏差 |

### 3.3 对原文计算可复现性进行评估

#### 3.3.1 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 12　结果可复现性的评估表**

| 可复现性情况 | 数量及占比 | |
|---|---|---|
| | *N* | % |
| 完全一致($\delta$ = 0%) | 95 | 100 |
| 偏差较小(0% < $\delta$ < 10%) | 0 | 0 |
| 偏差较小($\delta$ ＞ 10%) | 0 | 0 |
| 因舍入导致的偏差 | 0 | 0 |

* 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。例如，在一个 *t* 检验中，原文献报告了 *t* 值、95%置信区间、cohen's d 和样本大小，则这个效应中 *N*＝4。将各效应的 *N* 求和即为全体数量。

**表 13　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 | |
|---|---|---|
| | *N* * | % |
| 一致 | 12 | 100 |
| 不一致 | 0 | 0 |

#### 3.3.2 使用与原文献不同方法

报告采用新方法后，推论与原文献推论的一致情况，整理成表格，如下表所示

**表 14　推论的一致性的评估表(创新方法)**

| 推论的一致性 | 数量及占比 | |
|---|---|---|
| | *N* * | % |
| 一致 | 3 | 100 |
| 不一致 | 0 | 0 |

## 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 15 计算上（不）可重复的原因分析表**

| | 可能原因 | | 研究一 |
|---|---|---|---|
| | 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；<br>个别结果的微小差异，可能是由于印刷或复制粘贴错误；<br>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；<br>在文章的开放实践声明中对研究的模糊标记。 | |
| 原文献开放性问题 | 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)；<br>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；<br>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；<br>没有提供处理后的数据；<br>没有提供数据处理过程的描述或代码。 | √ |

| | | |
|---|---|---|
| | **代码开放获取特定问题** | 缺乏共享的分析代码或建模代码；<br>软件包或软件版本的问题。 |
| | **重复研究与院研究的区别** | 是否使用同样的数据集；<br>是否使用同样的数据分析软件及软件包；<br>是否使用同样的数据分析方法。 |
| **重复过程的原因** | **重复者相关因素** | 重复者此前是否有过 R 使用经验；<br>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； |
| | **文献质量** | 文献引用量大小；<br>是否有其他研究支持本文献结果；<br>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)。 |

**本研究进行完全一致的复现有以下原因：**

首先，原文献分别提供了 general 原始数据与 daily 原始数据，且数据预处理思路清晰，有明确注释。更重要的是，原文献数据分析思路清晰明了有明确注释，且构建模型并不复杂。因此，在理解原代码的基础上，我们使用与原文献相同的原始数据、R 包与代码能够完全复现原结果。

### 4.2 其他思考

（1）学习使用方面

本次课程带我们认识、实践了心理学统计分析常用的 R 代码。此学习过程让我们认识到，一方面，“在做中学”是学习 R 语言最有效的方法。在遇到编程问题时，应主动动手编写代码并多次尝试，查阅帮助文档，或查阅使用相关分析方法的文献，借鉴学习文献公开的代码并尝试运行。只有通过实际操作，才能熟练使用常见函数、参数和数据结构，培养解决实

际问题的能力。另一方面，学习 R 语言不能只靠临时抱佛脚，平时的积累也同样重要。对 R 包用法的熟练可以大大缩减编程与排查错误的时间。并且 R 包数量众多，社区活跃，时常会有新工具、新函数涌现，R 时常有着超越想象的功能。因此，我们应该养成在平时阅读关于 R 推文的习惯，从博客、社区资源等渠道不断获取 R 知识，记录有用的代码片段与技巧，整理自己的学习笔记。这些零碎知识通过长期积累，可以帮助我们完善知识体系，也方便我们在需要时迅速调用，提升效率。

（2）实践心得方面

1. 通过本次 R 实践，我们对数据分析的整个过程有了更深入的理解。从预处理到模型建立，从模型比较到效应解释，每一步紧密相扣，相辅相成。这不仅是对 R 工具的熟悉过程，更是对统计建模逻辑的系统性训练。在分析开始之前，数据预处理是不可或缺的环节。通过 dplyr、tidyr 等高效的数据处理工具进行变量筛选、缺失值处理、变量类型转换以及必要的标准化操作，快速清洗数据并构建符合建模需求的变量结构。本次在分析 daily 数据中会话能力随干预的变化趋势时，我们尝试使用不同的建模方法，将二分变量 RatingType 的编码由原来的 0，1 改为 0.5，-0.5，以使线性混合模型的输出结果显示真实的主效应和交互效应。此外，考虑到每位被试会话能力的随时间的变化趋势可能不同，我们还添加了 Day 作为被试的固定斜率。比较原文使用的模型（Ability_Avg7 ~ RatingType*Day + (1|GooseChaseId_Fixed) ）与我们构建的新模型（Ability_Avg7 ~ RatingType*Day + (1+Day|GooseChaseId_Fixed)），$\chi 2$ 显示新模型与原文模型有显著差异，AIC, BIC 等指标显示新模型显著优于原文模型。通过本次实践，我们更深入理解使用数据分析时“理论指导、尝试验证”的思路。当存在多个理论上合理的模型假设，可以选择“都跑一跑多尝试”的策略，通过观察、比较它们的拟合情况最终选择一个最合适的模型。这可以避免因先验偏好遗漏重要模型，也能够使分析灵活地适应数据的实际结构。

2. 我们对广义线性混合模型（GLMM）有了更深入的理解，尤其是在处理计数型数据以及变量建模策略方面。针对过度离散计数数据的建模策略，使用的是 glmer.nb()。在实际数据分析中，计数型响应变量（如人数）往往存在过度离散的情况，即数据的方差远大于其均值，导致传统的泊松回归（Poisson regression）模型拟合不佳。通过本次作业，我们学习使用了负二项广义线性混合模型（Negative Binomial GLMM），借助 R 中的 glmer.nb()函数进行建模。相比于 glmer()默认的泊松假设，glmer.nb()允许误差项具有更大的灵活性，从而更好地处理现实中常见的离散程度更高的数据，提高模型的拟合优度与预测能力。

对连续变量与因子变量建模存在差异。在对时间变量如 Day 进行建模时，我认识到将

其视为连续变量与因子变量会导致截然不同的解释结果：

当 Day 作为连续变量进入模型时（如 Day），模型拟合的是趋势性的变化，适合用来检验“随着时间推移，反应变量是否呈现某种线性或非线性变化趋势”，适用于探索性的假设检验或趋势分析。

而将 Day 设为因子变量（如 factor(Day)）时，模型则会分别估计每一天的边际效应，从而可以对不同天数之间的显著性差异进行比较。这种方式特别适合结果展示与可视化，能直观呈现各时间点的响应变化，有助于撰写研究报告或撰写论文时进行组间比较与可视化呈现（如边际估计图）。

## 参考文献(APA 格式)

Sandstrom, G. M., Boothby, E. J., & Cooney, G. (2022). Talking to strangers: A week-long intervention reduces psychological barriers to social connection. *Journal of Experimental Social Psychology*, *102*, 104356. https://doi.org/10.1016/j.jesp.2022.104356