

# 探究基于创造过程的事件结构特征进行创造力评估的可能性

**摘 要** 创造性认知是人类重要的认知能力之一。本研究关注创造力的社会评估,试图从创造过程的事件结构中发掘可预测他人对该过程创造力评估的潜在变量与维度。本研究采用事件分割的实验范式,要求被试在观看不同的游戏通关视频时实时地进行事件分割,并在最后对该通关方法进行创造力评分。实验结果指出预先提出的过程结构变量并无法预测被试对其的创造力评估,发现被试在任务中无法达到实验设计预想的表现。通过对实验失败的分析,我们指出了未来在从事件结构视角探究创造力评估可能需要注意的事项,并提出了未来可能的研究方向与重点。

**关键词** 创造力 创造力评估 创造过程 事件分割

## 1 前言

创造力(creativity)是人类所具备的最复杂的认知能力之一。在日常生活中,我们通常将人类从无到有地创造以前不存在的事物的能力称作创造力。进一步深入这一概念,我们发现,大多数情况下,这些“创造”的起点是已经存在或至少可能已经存在的概念,但尚未有人将这些概念在一个特定的产品中明确地建立联系。正是这种“新颖”的联系,为我们带来了诸如梵高的《星空》等艺术作品:每个人都看过星空与漩涡,但从未有人用漩涡的形态来绘制星空。

在心理学研究中,创造力作为一个抽象的构念,研究者们对其概念定义尚未达成一致,并由于其概念的多维性与复杂性,研究者们创造了诸多模型以描述或概括这一概念。例如 Rhodes 在 1961 年提出了创造力的 4P 模型(4 “P’s” of creativity),认为应当从个人(person)、过程(process)、产物(product)与环境媒介(press)四个视角定义创造力;而 Kaufman 与 Beghetto 在 2009 年提出 4C 模型(Four C model),将创造力概念依照创造产物的价值不同区分出四个不同的层次加以概括描述。

尽管在创造力研究中对于哪些维度与特征定义了创造力仍存在诸多差异(e.g., Amabile, 1990; Boden, 2004; Dietrich, 2004; Finke et al., 1992; Hennessey & Amabile, 2010),在创造力概念的诸多维度中,有两个维度得到了极多的关注与认可(Jordanous & Keller, 2012, 2016)。其中,第一个维度为新颖(novel)或不寻常(unusual),第二个维度为有用(usefulness)或有意义(meaningful),只有当某一概念既新颖、又符合其创造环境的要求——即有其价值时才能被称为是有创造力的(Ward, 2007)。

在过去的研究中,已有诸多研究者对创造力的新颖、有用这两个维度在不同领域使用不同方法进行各种探索。在语义认知的框架下,有研究者通过隐喻加工任务(metaphor processing task)与修改后的替代用途测验(Alternate Uses Task, AUT)结合 ERP 研究、fMRI 研究证明了新颖且有价值的材料可以诱发在语义认知中的

概念拓展(conceptual expansion), 是创造性认知中的核心认知过程之一(Rutter et al., 2011; Rutter et al., 2012; Kröger et al., 2012; Kröger et al., 2013)。Diedrich 等人(2015)也通过发散性思维任务(divergent thinking task)收集了 1500 名被试的反应, 并要求另外 18 名独立裁判对反应的新颖性(novelty)、有用性(usefulness)和创造性(creativity)进行评估, 结果发现新颖性可以作为创造性的一阶标准, 有用性可以作为创造性的二阶标准: 如果一个想法不是新颖的, 那么它的有用性无关紧要, 但是如果一个想法是新颖的, 那么它的有用性将额外地决定它的实际创造性。但在诸多有关创造力的新颖与有用的研究中, 其评估的视角均从 Rhodes(1961)的 4P 模型中的“product”出发。但对于创造这一复杂的认知过程, 针对其产物的研究无法深入到认知过程本身, 因此本研究关注于在 4P 模型的“process”中是否存在同样的新颖、有用维度可以预测创造力大小。

同时, 除上文提及的隐喻加工任务、替代用途测验与发散性思维任务外, 在近几年兴起了利用电子游戏进行创造力测验的浪潮。相较于传统的创造力测验, 使用电子游戏为载体不仅能考虑到最终的产品, 还能同时考虑其过程(例如在通关电子游戏时的操作序列)。在某些电子游戏中, 玩家创造的过程本身即是其创造的产品, 例如最经典的电子游戏之一——推箱子(Sokoban Puzzle)。玩家的游戏目标是将箱子推至目标地点, 而玩家的创造过程抽象化得到的通关方法即是玩家创造的最终产品。因此对这样通关过程的评估可以将过程与产品的不同维度结合起来, 共同研究其对创造力评估的贡献。本研究选择的实验材料来自一款类推箱子游戏“BaBa Is You”的玩家通关视频, 同样共同探究过程与产品的不同维度对于创造力评估的预测能力。

为探究人们是如何评估他人的创造力以及创造性认知过程中是否存在某些维度与特征可以预测他人对于这一过程的评估, 我们要求被试对不同的游戏通关视频进行事件分割(event segmentation), 并在最后进行创造力评分。其中, 事件分割范式在事件知觉(event perception)研究中常被用来收集被试对于某事件的认知与预测数据。Eisenberg 等人(2018)便利用此范式结合眼动数据说明了被试做出事件分割的行为源于其对事件预期误差的短暂增加, 而这与上文提及的创造力的新颖性不谋而合——通关过程中某一新颖操作带来的预期误差的短暂增加有可能使被试在此处进行事件分割, 即事件分割的节点可能可以预测富有创造力的创造过程。此外, 也有研究者通过事件分割范式说明了人类对事件的编码具有分层的特点, 并结合被试进行不同细致程度的事件分割可量化被试对于某事件编码的层次性与一致性(Kurby & Zacks, 2011)。

综上所述, 本研究采用事件分割的实验范式, 要求被试在观看不同的游戏通关视频时实时地进行事件分割, 并在视频结束后对该通关方法进行创造力评分, 试图探究新颖性、分割的层次性、一致性等创造性产品与过程的不同维度是否能预测创造力评估。

## 2 方法

### 2.1 被试

本实验共 19 名被试，均为北京大学在读本科生。所有被试均在实验前无相关游戏经验。所有被试视力或矫正视力正常。本实验中被试获得 75 元人民币被试费作为报酬。

### 2.2 实验材料

本实验中使用的视频材料来自北京大学心理与认知科学学院计算与决策实验室的博士生陆杨帆博士在其博士项目中收集到的数据生成，指导老师为张航老师。生成视频的数据为陆杨帆博士在实验中收集的被试通关游戏的操作序列，生成得到的视频即为依次匀速执行该被试操作的视频。视频的帧率为 4 fps，每帧中玩家操纵的物体可以向上下左右某方向移动一个单位。某关卡的视频截图如图 1 所示。

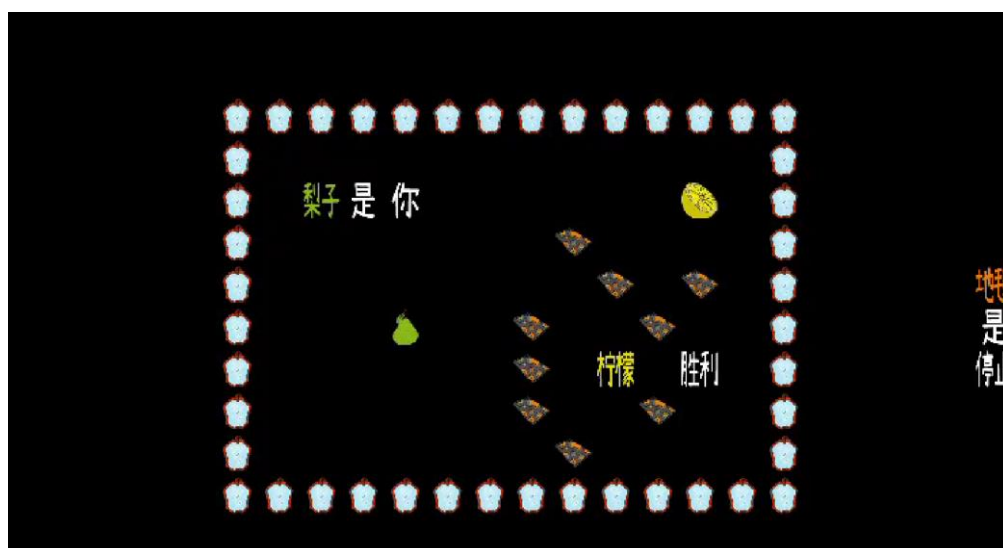


图 1 某关卡的视频截图

本实验中所有使用的视频来自五个关卡，其中每个关卡有三种不同存留 (preservation) 水平的通关方式。存留水平的定义如下：将陆杨帆博士项目中所有被试在某个关卡的通关方法绘制在如图 2 的有向图中，每个节点代表一个状态 (state)，边的粗细代表所有被试经过这一步的次数，将一条通关路径中所有具有非零访问次数的边的访问次数的均值称为存留水平 (preservation)，可以直观地认为存留水平越低，该路径或该通关方法的 novelty level 越高，一定程度上可能可以反映更高的创造力水平

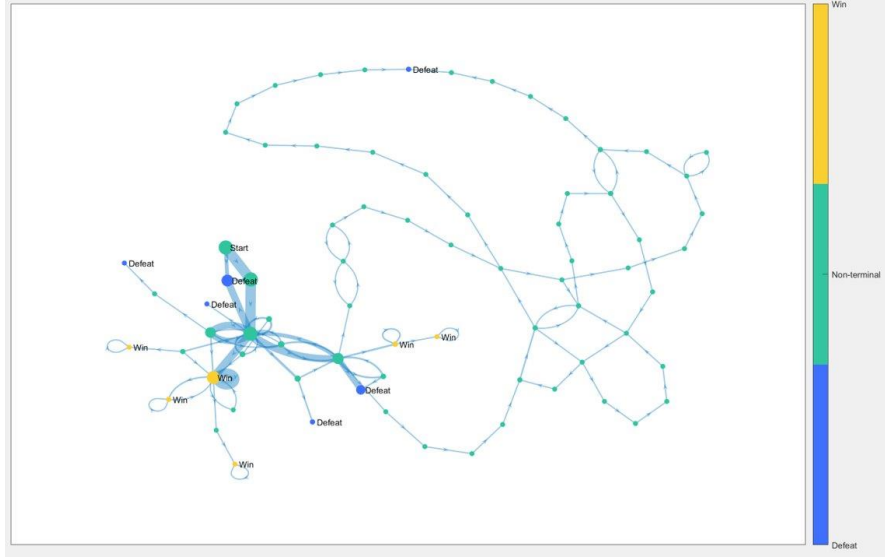


图 2 某关卡的所有被试通关方法有向图(引自陆杨帆博士的项目)

## 2.2 实验设计

本实验采用组内设计。本实验参考了 Kurby 与 Zacks(2011)有关事件结构层次感知研究, 计算分割一致性(segmentation agreement)与层级齐性(hierarchical alignment)并纳入分析。

其中, 分割一致性是指每个被试对于某视频的事件分割与群体对于该事件的分割的一致性程度。分割一致性的计算方法如下: 对于每个视频, 我们将其分为数个 1 秒的区间, 并记录每个被试是否在每个区间内标记了事件边界。对于每个视频, 我们计算每个被试在所有区间内标记事件边界的情况与被试群体在所有区间内标记事件边界的比例的点二列相关(point-biserial correlation)作为对分割一致性的测量。同时, 为了控制标记的事件边界个数的个体差异对于分割一致性的影响, 我们对计算得到的相关系数进行矫正。这是因为直接计算得到的点二列相关系数会受到被试标记的事件边界数量的限制(例如在被试标记的事件边界极少, 那么其点二列相关系数会随标记位置的不同表现为弱正相关或负相关, 但绝对不会表现为较大的相关)。我们可以使用这样的方法对点二列相关系数进行矫正:

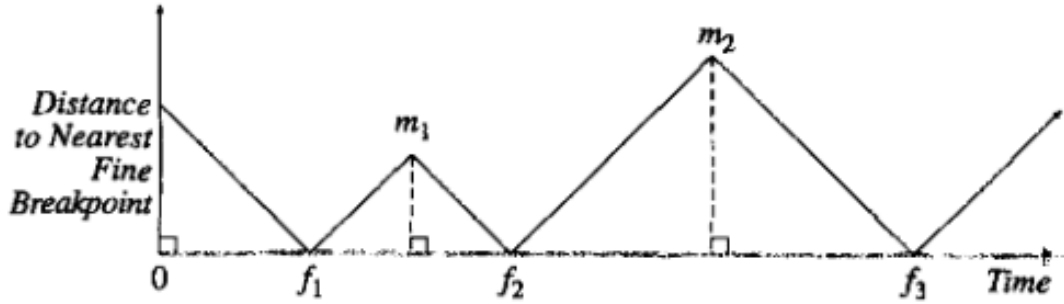
$$\text{Segmentation Agreement} = \frac{r - r_{\min}}{r_{\max} - r_{\min}}$$

其中,  $r$  表示直接计算得到的点二列相关,  $r_{\max}$  与  $r_{\min}$  分别代表给定某被试标定的事件边界数量时, 可能的最大或最小的点二列相关。

层级齐性度量了对于某视频中某被试标记的每个粗事件边界(coarse event boundary)与其最近的细事件边界(fine event boundary)的平均时间距离, 该值相对于机会水平的时间距离的大小反映了被试对于该视频事件划分的层级结构性。即该平均时间距离越小于机会水平的时间距离, 被试的分割编码越有层级结构的特征, 表明被试在粗、细编码条件下的事件边界吻合度优于机会水平。某视频的层级齐性(下面简称 *AvgDist*)的数学表达为:

$$AvgDist = \frac{\sum \text{粗事件边界与其最近细事件边界的时间距离}}{\text{粗事件边界数}}$$

为计算机会水平下的平均时间距离期望, 我们需要进行一些简单的假设。首先, 我们假设粗细时间边界位置之间相互独立, 且粗事件边界随机均匀分布。同时我们将细事件边界的位置作为给定的值, 计算随机均匀分布的粗事件边界到最近的细事件边界的时间距离, 求和并除以视频长度即为机会水平下的平均时间距离期望。记  $F = \{f_1, f_2, \dots, f_{Fine}\}$  为所有细事件边界时间点的集合, 取最后一个细事件边界即  $f_{Fine}$  为被试对于视频中事件时间长度的估计, 则任意时间点上的粗事件边界到其最近细事件边界的事件距离可以写成与时间有关的函数, 如图 3 所示。对此函数进行积分并处以  $f_{Fine}$  即为机会水平下的平均时间距离期望。



**Figure 3.** Graphical derivation of the null model prediction for the continuous analysis. The distance from each location on the time line and the nearest fine breakpoint is plotted by the fine line. Fine breakpoint locations are fixed and indicated by  $f_1, f_2, \dots$ . Midpoints between adjacent breakpoints are indicated by  $m_1, m_2, \dots$ . The total area under the fine line, divided through by the length of the stimulus, is the null model estimate.

图 3 粗事件边界到其最近细事件边界的时间距离函数示意图(Zacks et al., 2001)

因此, 机会水平下该视频中每个粗事件边界与其最近的细事件边界的平均时间距离期望(下面简称  $AvgDist_0$ )的计算公式为:

$$AvgDist_0 = \frac{\frac{f_1^2}{2} + \sum_{i=1}^{i=Fine-1} [\frac{f_{i+1} - f_i}{2}]^2}{f_{Fine}}$$

同时本实验对额外变量进行了控制, 为减少观看视频的顺序对于被试对通关方法的理解不同的影响, 被试按随机顺序观看所有通关视频。

## 2.4 实验程序

本实验中被试单独进行实验。被试由主试告知任务背景——游戏评审后, 首先需完成 BabaIsExp 的新手教程以理解游戏的基本规则。在完成新手教程之后,

被试在需阅读并确认指导语，以设计好的三个新手教程关卡视频为材料，完成事件分割的练习。指导语见附录。

正式实验中，被试需依次观看不同的游戏通关视频，并在观看的同时要求被试在其认为是某一子事件结束的时间节点按空格键反应。被试需连续观看同一视频两遍并进行细致程度不同的事件分割。第一遍事件分割称为粗分割，要求被试分割出隶属于整个通关视频的最大事件单元，即在被试发现某一最大的子事件结束时按空格键反应，此时视频暂停，被试按数字键反应自己的暂停距真实子事件结束延迟了多少秒（被试已被告知游戏角色移动一次约 0.25 秒，因此可以进行简单的估计）；第二遍事件分割称为细分割，要求被试标记隶属于整个通关视频的最小事件单元，即在每个被试认为是最小的子事件结束时按空格键反应，此时视频不会暂停，被试也无需按数字键反应。每次被试按下空格进行事件分割时，均会有“嘟”声反馈。需要指出的是，被试被告知不存在最大子事件(即粗事件)与最小子事件(即细事件)的绝对判断标准，只需自由地进行反应即可。观看一个通关视频并分别进行两次事件分割后，被试需通过滑块进行创造力评分（创造力评分的界面见图 4），之后进入下一试次，试次间被试有 15 秒的休息时间，试次内两次播放同一视频之间被试也有 10 秒的休息时间。单试次的实验流程见图 5。

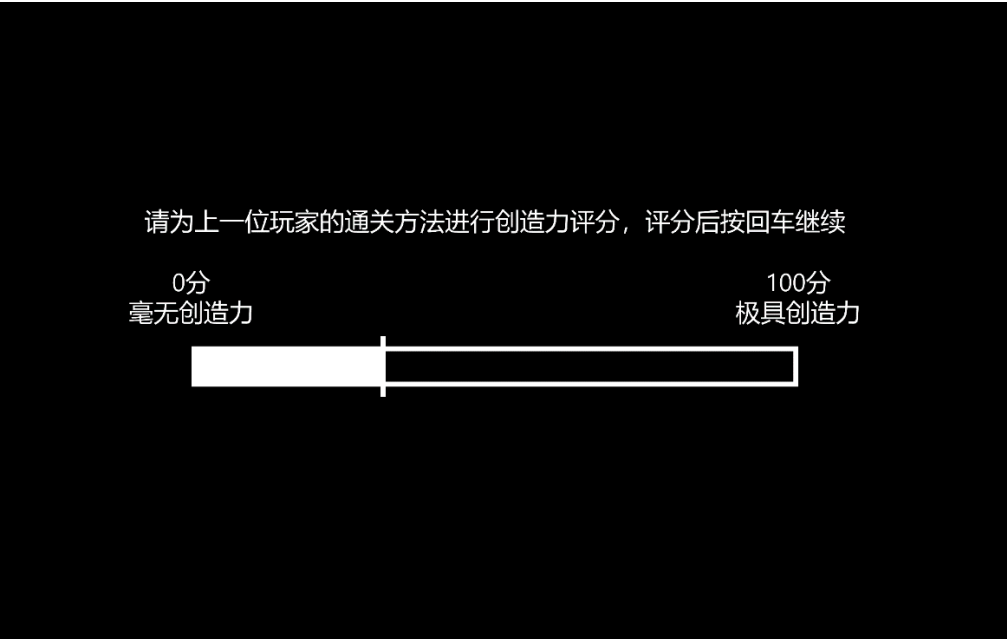


图 4 创造力评分界面

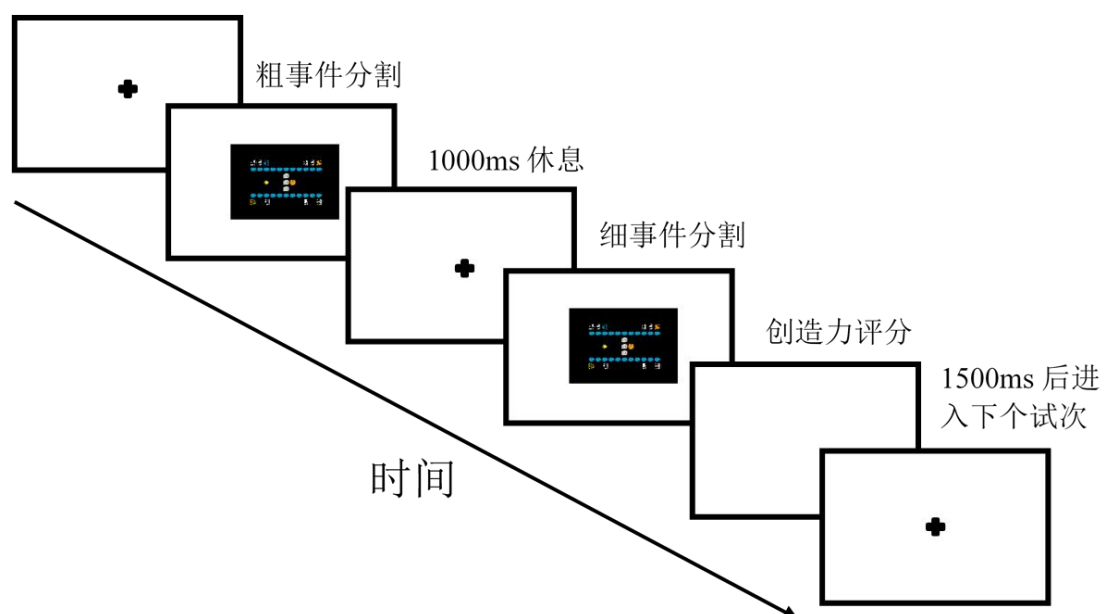


图 5 单试次实验流程图

### 3 结果

#### 3.1 事件分割结果

首先，依照 2.3 中叙述的方法计算得到每个被试在不同分割中的分割一致性，每个被试分割的层次齐性及层次齐性的期望水平，并用层次齐性/其期望水平得到被试分割的相对层次齐性，其范围为[0,1]，值越小代表被试分割的事件结构越具有分层的特点。不同关卡不同存留水平通关视频经被试分割得到的分割一致性与相对层次齐性见图 6、图 7。

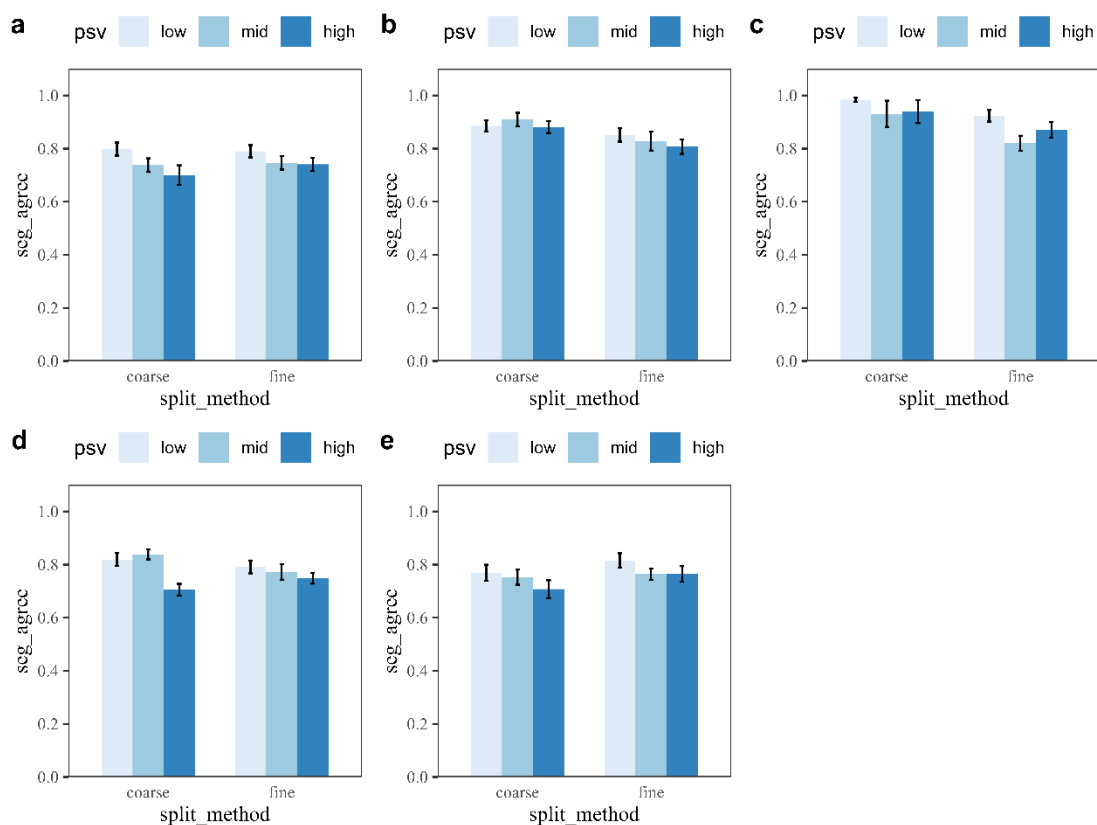


图 6 被试分割不同游戏通关视频的分割一致性

注：i) a/b/c/d/e 分别代表第一至第五关；ii) psv 代表存留水平(preservation)；  
iii) seg\_agree 代表分割一致性；iiii) 图中误差线表示标准误

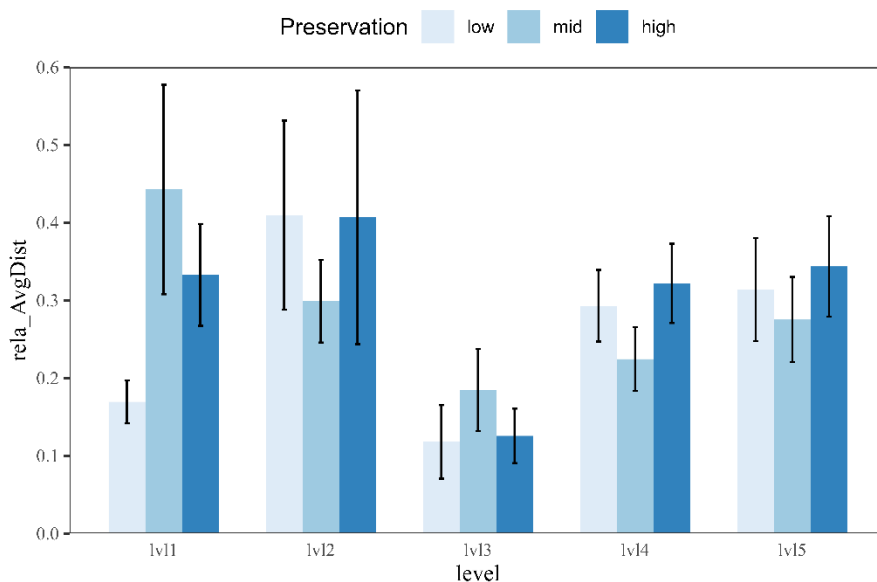


图 7 被试分割不同游戏通关视频的相对层次齐性

注：图中误差线表示标准误

从图 6 可以看出，所有被试进行事件分割的一致性分数均较高，说明被试应该正确地理解了事件分割的实验要求，可以较好地完成事件分割的实验任务。



务。而在图 7 中可以看出, 被试事件分割的相对层次齐性均远小于 1, 即被试分割的层次齐性远小于其期望水平, 说明被试对于其他玩家通关的事件过程的认知与编码具有分层的特点。

同时, 在图 6 与图 7 中都可以看出, 在某些关卡中, 分割一致性、相对层次齐性与存留水平的变化具有相似的趋势, 说明有关事件过程的某些结构变量可能与事件本身的新颖程度甚至创造力水平具有一定的关系, 事件过程的结构变量可能可以预测创造力苹果的分值。

进一步地, 以被试对不同视频进行粗分割的分割一致性为因变量, 以游戏关卡、存留水平为组内自变量进行 5(游戏关卡) $\times$ 3(存留水平: 低/中/高)的双因素重复测量方差分析。首先进行球形度检验, 游戏关卡、存留水平以及交互项均不符合球形度假设( $p < .05$ ), 因此采用 Greenhouse-Geisser 进行校正。方差分析的结果显示, 游戏关卡的主效应显著,  $F(2.414, 43.444) = 30.557, p < .001, \eta_p^2 = .629$ ; 存留水平的主效应显著,  $F(1.468, 26.425) = 9.123, p = .002, \eta_p^2 = .336$ ; 交互作用不显著,  $F(3.548, 63.864) = 1.641, p = .181$ 。对存留水平进行事后检验, 采用 Bonferroni 矫正, 存留水平低条件下分割一致性显著高于存留水平高条件,  $MD = .065, MSE = .016, t(18) = 3.984, p = .002$ ; 存留水平中条件下分割一致性与存留水平高条件差距边缘显著,  $MD = .047, MSE = .019, t(18) = 2.477, p = .070$ ; 存留水平低条件下分割一致性与存留水平中条件无显著区别,  $MD = .017, MSE = .010, t(18) = 1.668, p = .338$ 。

以被试对不同视频进行细分割的分割一致性为因变量, 以游戏关卡、存留水平为组内自变量进行 5(游戏关卡) $\times$ 3(存留水平: 低/中/高)的双因素重复测量方差分析。首先进行球形度检验, 游戏关卡、存留水平以及交互项均符合球形度假设( $p > .09$ ), 因此无需进行校正。方差分析的结果显示, 游戏关卡的主效应显著,  $F(4, 72) = 10.395, p < .001, \eta_p^2 = .366$ ; 存留水平的主效应显著,  $F(2, 36) = 5.534, p = .008, \eta_p^2 = .235$ ; 而交互作用并不显著,  $F(8, 144) = 0.614, p = .765$ 。对存留水平进行事后检验, 采用 Bonferroni 矫正, 存留水平低条件下分割一致性显著高于存留水平中条件,  $MD = .048, MSE = .016, t(18) = 3.073, p = .020$ ; 存留水平低条件下分割一致性与存留水平高条件差距边缘显著,  $MD = .048, MSE = .019, t(18) = 2.484, p = .069$ ; 存留水平中条件下分割一致性与存留水平高条件无显著区别,  $MD < .001, MSE = .015, t(18) = 0.001, p = 1.000$ 。

以被试对不同视频进行分割的相对层次齐性为因变量, 以游戏关卡、存留水平为组内自变量进行 5(游戏关卡) $\times$ 3(存留水平: 低/中/高)的双因素重复测量方差分析。首先进行球形度检验, 游戏关卡及其与存留水平的交互项均不符合球形度假设( $p < .001$ ), 因此采用 Greenhouse-Geisser 进行校正。方差分析的结果显示, 游戏关卡的主效应显著,  $F(1.787, 32.167) = 3.620, p = .043, \eta_p^2 = .167$ ; 存留水平的主效应不显著,  $F(1.949, 35.088) = 0.823, p = .445$ ; 交互作

用不显著,  $F(2.552, 45.943) = 1.284, p = .290$ 。

综上, 事件分割的相关结果可以说明, 被试可以基本理解事件分割范式的任务要求, 并且在群体水平上具有较高的分割一致性水平。同时被试对于视频中玩家通关操作的分割具有一定的层次性, 对于该创造性认知过程的编码具有分层的特点。结合方差分析的部分结果, 部分有关事件过程的结构性变量可能可以反映被试对于事件过程的认知乃至编码特点, 进而可能可以预测被试对其的创造力评估结果。

### 3.2 创造力评估结果

在确定有关事件结构的变量具有预测创造力评估的可能性后, 将被试对每个通关视频的创造力评分纳入分析。在具体分析前, 将每个被试的评分标准化, 减少不同被试打分偏好不同产生的影响。在五个关卡中, 所有被试的创造力评分与相对层次齐性/粗划分分割一致性/细划分分割一致性的分布见图 8 至图 12。

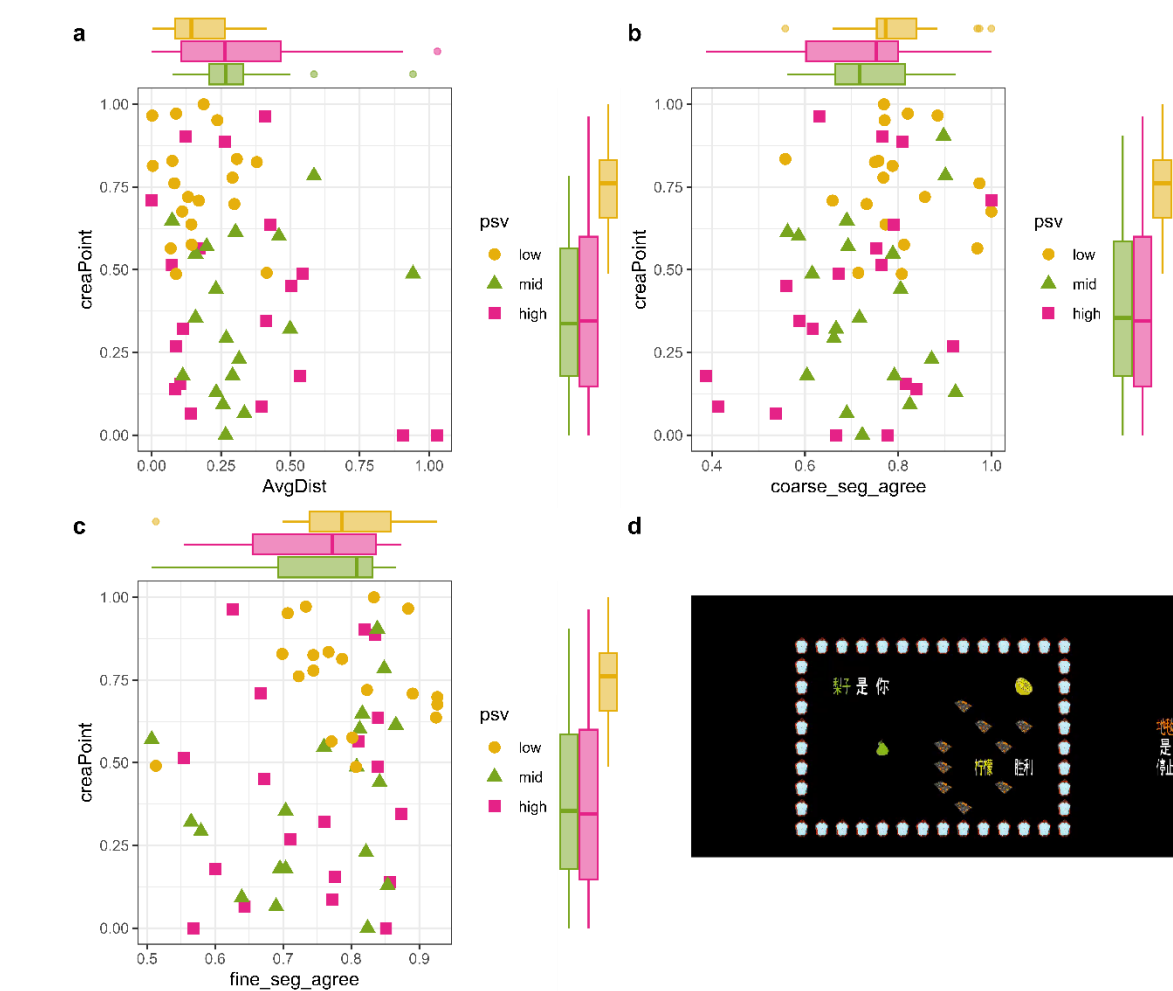


图 8 第一关视频创造力评分与事件结构变量分布

注: i) a/b/c 分别为创造力评分与相对层次齐性、粗划分一致性、细划分一致性的分布;  
ii) d 为本关初始状态示意图; iii) psv 代表存留水平(preservation);

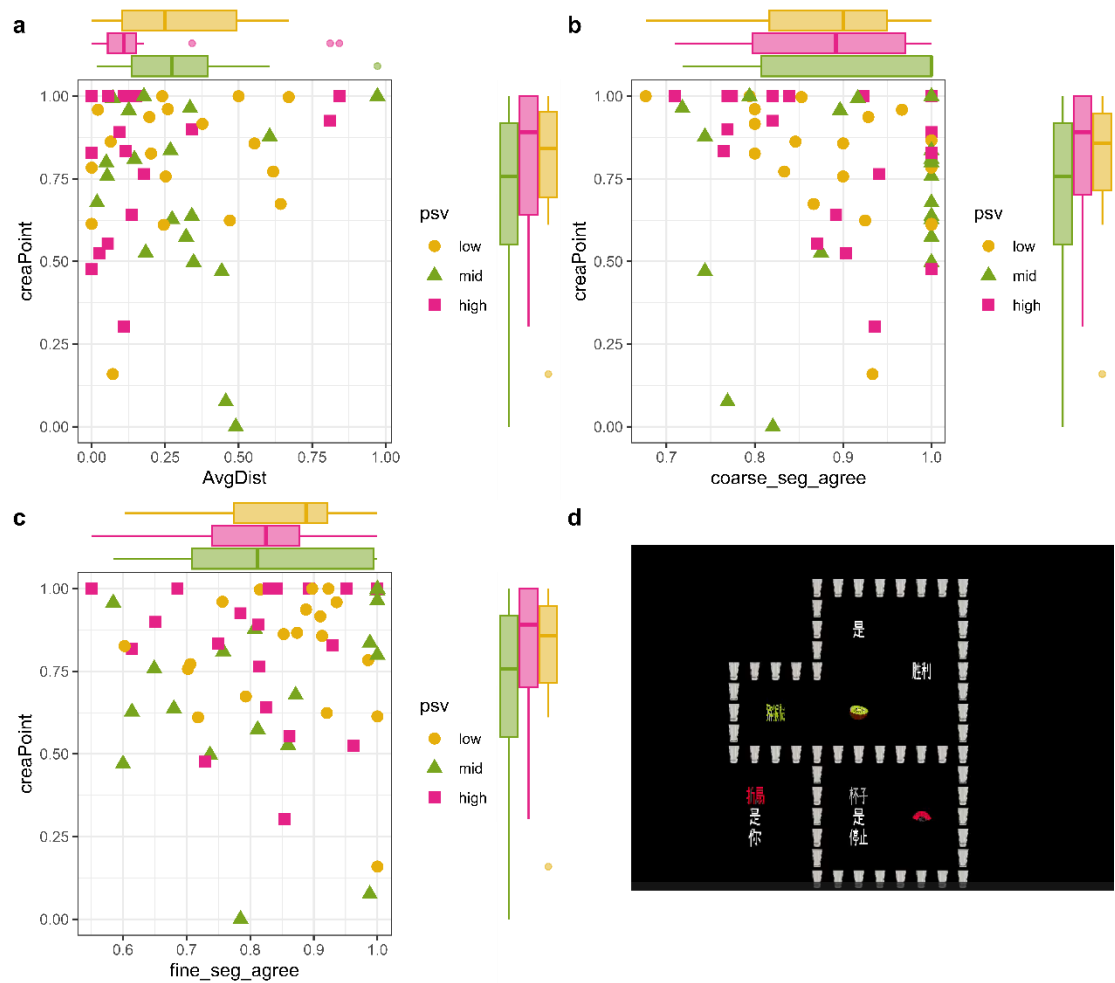


图9 第二关视频创造力评分与事件结构变量分布

注： i) a/b/c 分别为创造力评分与相对层次齐性、粗划分一致性、细划分一致性的分布；  
 ii) d 为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

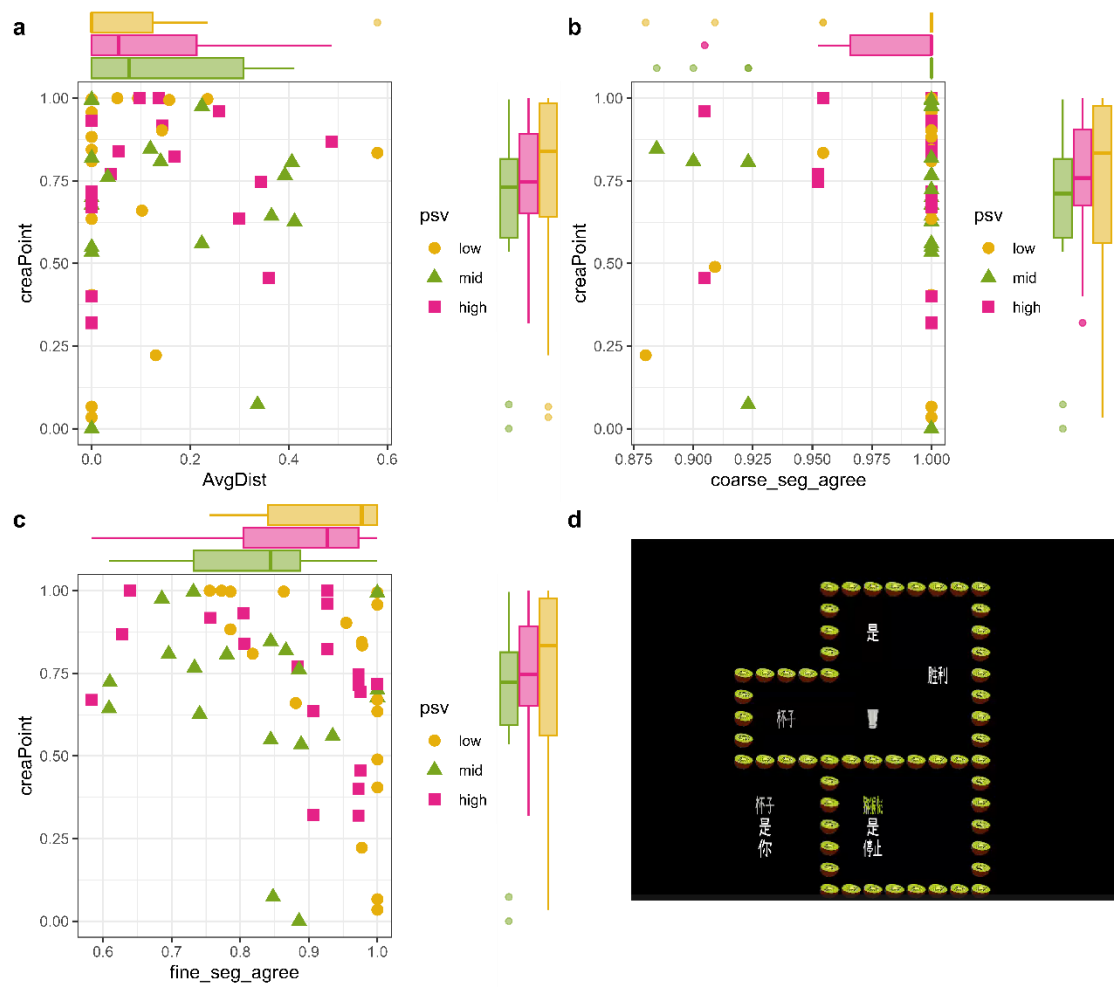


图 10 第三关视频创造力评分与事件结构变量分布

注： i) a/b/c 分别为创造力评分与相对层次齐性、粗划分一致性、细划分一致性的分布；  
 ii) d 为本关初始状态示意图； iii) psv 代表存留水平(preservation)；



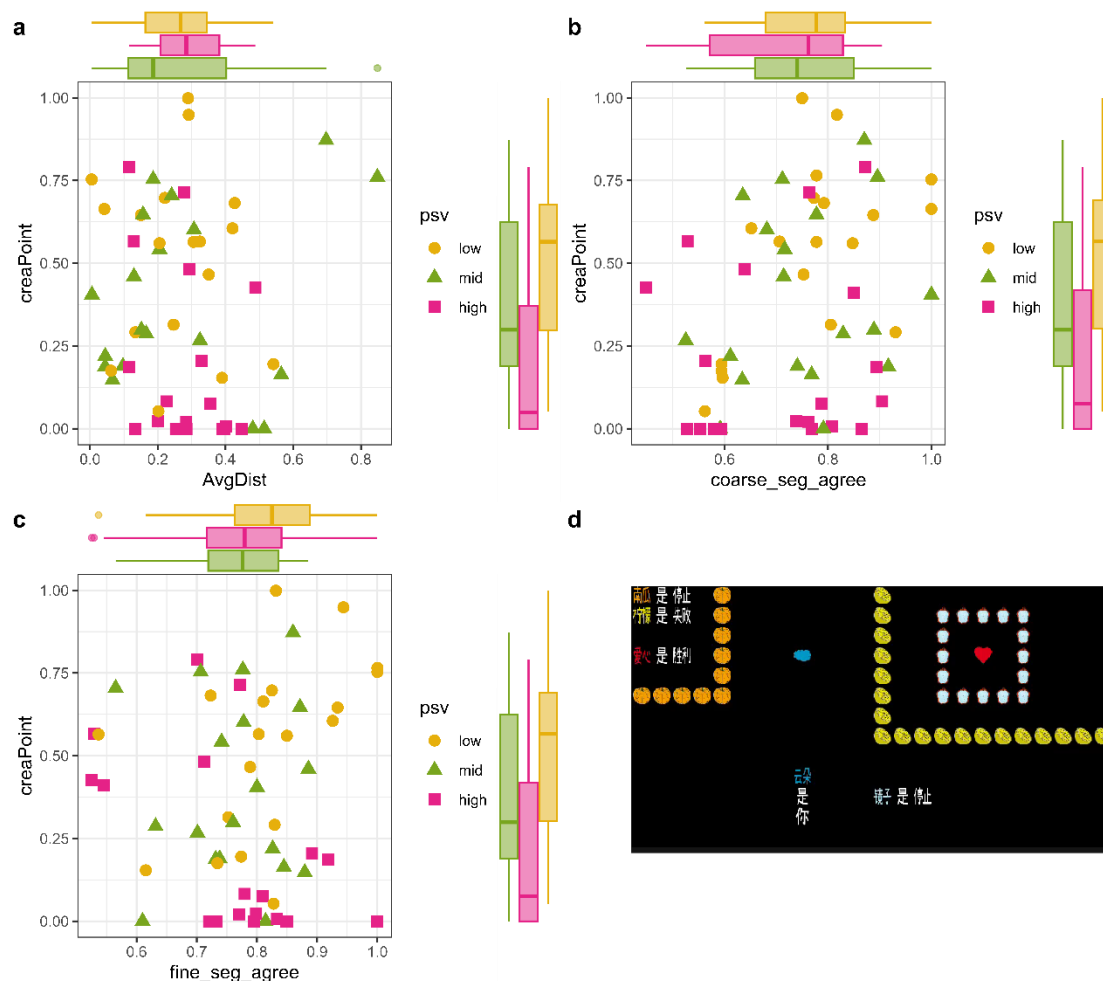


图 12 第五关视频创造力评分与事件结构变量分布

注： i) a/b/c 分别为创造力评分与相对层次齐性、粗划分一致性、细划分一致性的分布；

ii) d 为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

从上述五图中可以看出，除第二关、第三关外，各个关卡中不同存留水平通关视频下的创造力评分与事件结构变量分布相对聚合成团，并彼此间有一定的区分度，进一步验证了有关事件结构的变量具有预测创造力评估的可能性，但分布图中无法观察到明显的相关性。将创造力评分分别与相对层次齐性、粗划分一致性、细划分一致性做相关分析得到结果，发现创造力评分仅与粗划分一致性具有正相关( $Pearson's r = .343, p < .001$ )，与相对层次齐性、细划分一致性均无显著相关( $Pearson's r_1 = .001, p_1 = .821$ ;  $Pearson's r_2 = .110, p_2 = .075$ )。

而除上述事件结构变量外，基于被试做出事件分割的行为源于其对事件预期误差的短暂增加的假设，而在粗划分中对于分割延迟时长的汇报也是对事件预期误差大小的一种表征。因此被试在进行创造力评分时，也有可能参考了其在粗划分中进行汇报分割延迟的次数。因此将粗划分中被试汇报了分割具有延迟的次数也纳入分析，其在不同关卡中与创造力评分的分布见图 13 至图 17。

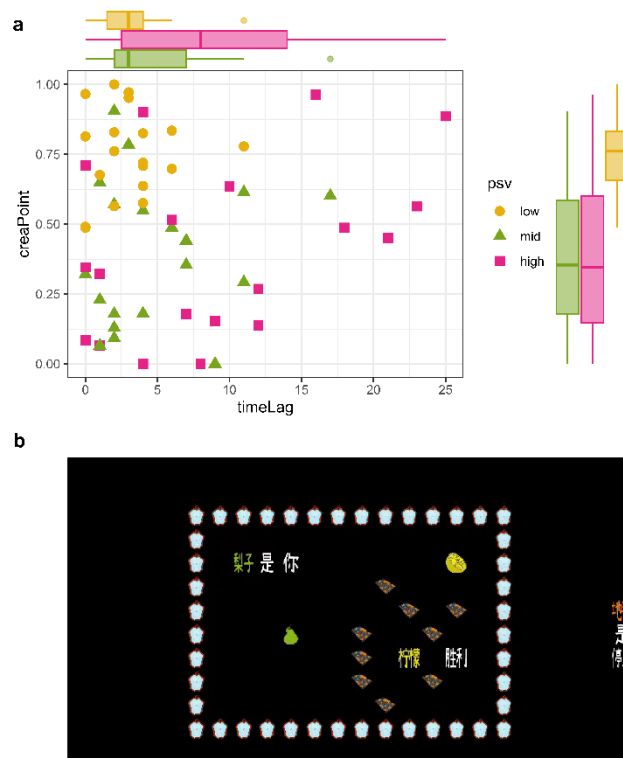


图 13 第一关视频创造力评分与粗事件分割延迟次数分布

注： i) a 图为创造力评分与粗事件分割延迟次数的分布；

ii) b 图为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

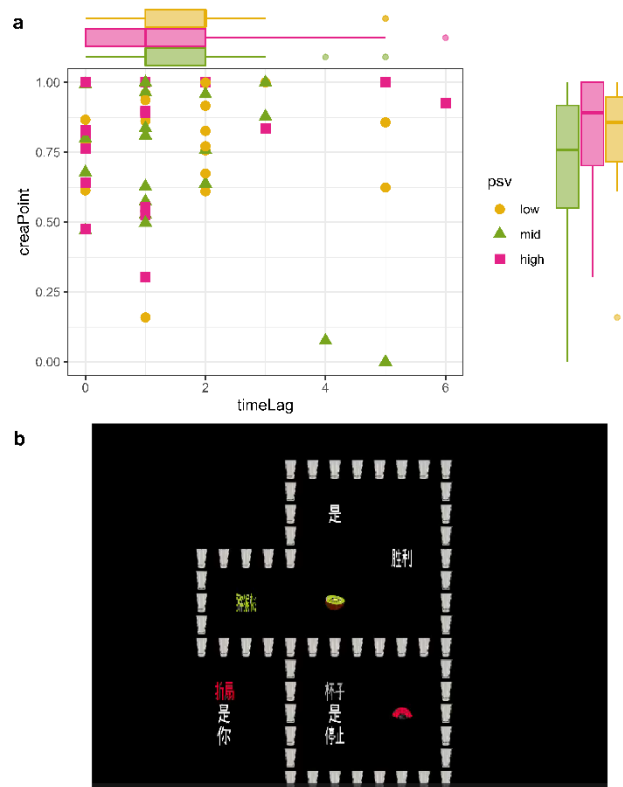


图 14 第二关视频创造力评分与粗事件分割延迟次数分布

注： i) a 图为创造力评分与粗事件分割延迟次数的分布；  
 ii) b 图为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

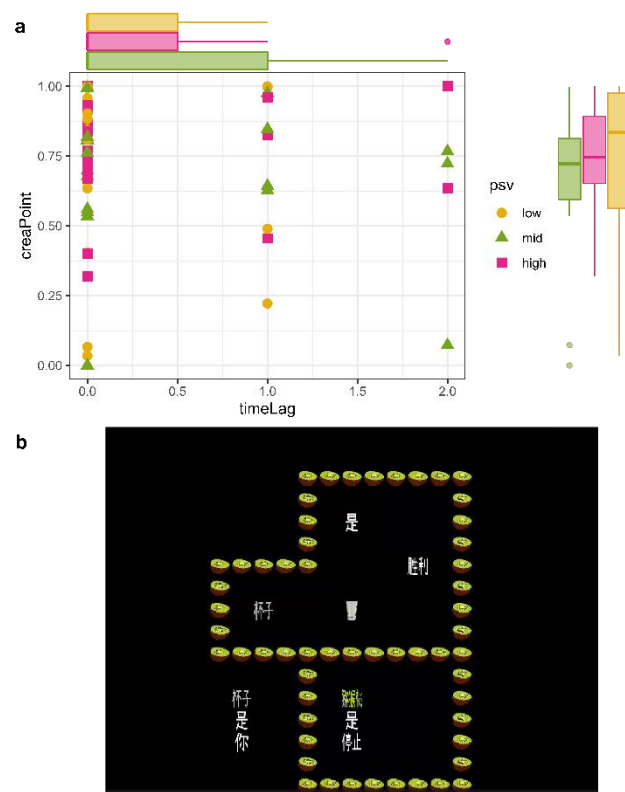


图 15 第三关视频创造力评分与粗事件分割延迟次数分布

注： i) a 图为创造力评分与粗事件分割延迟次数的分布；  
 ii) b 图为本关初始状态示意图； iii) psv 代表存留水平(preservation)；



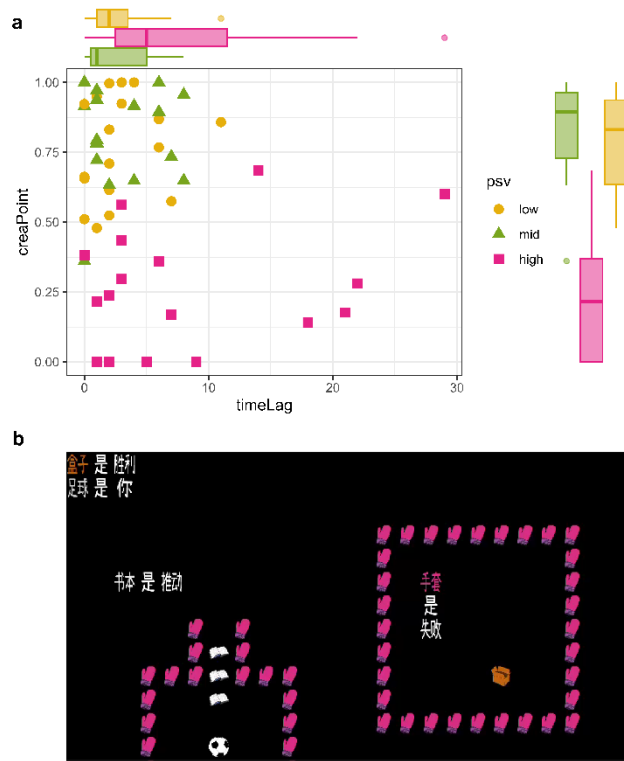


图 16 第四关视频创造力评分与粗事件分割延迟次数分布

注： i) a 图为创造力评分与粗事件分割延迟次数的分布；

ii) b 图为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

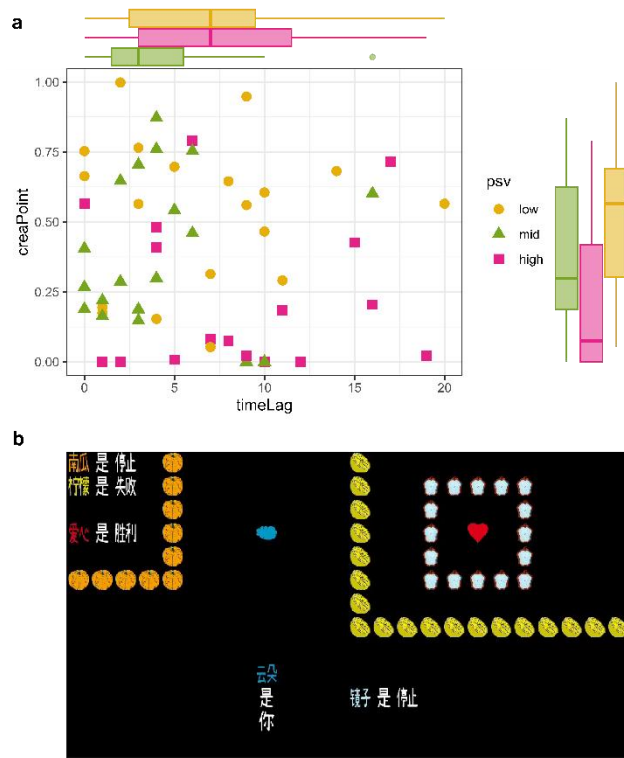


图 17 第五关视频创造力评分与粗事件分割延迟次数分布

注： i) a 图为创造力评分与粗事件分割延迟次数的分布；

ii) b 图为本关初始状态示意图； iii) psv 代表存留水平(preservation)；

可以看到除第二关、第三关外，各个关卡中不同存留水平通关视频下的创造力评分与粗事件分割延迟次数分布相对聚合成团，并彼此间也有一定的区分度，进一步验证了上文的猜想。对创造力评分与粗事件分割延迟次数做相关性分析，结果发现创造力评分与粗事件分割延迟次数具有显著的负相关( $Pearson's r = -0.230, p < .001$ )，与之前的预期不符。为避免第二三关时间过短对粗事件分割的干扰使被试无法及时做出分割与分割时是否延迟的判断，在剔除第二三关的数据后重新对创造力评分与粗事件分割延迟次数做相关性分析，创造力评分与粗事件分割延迟次数仍具有显著的负相关( $Pearson's r = -0.200, p = .002$ )，与被试做出事件分割的行为源于其对事件预期误差的短暂增加，更多的粗事件分割延迟次数将导致更高的创造力评估的假设相违背。

考虑到在选取的五个关卡中，不同关卡的通关视频的时长差距极大( $\min = 9s, \max = 270s$ )，而粗划分与细划分的可能划分次数受视频时长影响巨大，进而会影响划分一致性。尤其是在粗划分需要在按键反应后报告实际划分事件点距反应时间点的时长，也会受到不同视频的时长不同的干扰。同时不同关卡内容本身的特点也可能会干扰不同因素对于创造力评估的影响。因此以创造力评估分数为因变量进行逐步多元线性回归分析，在开始分析前假设回归方程为：创造力得分 = 存留水平+视频时长+(相对层次齐性+粗分割一致性+细分割一致性+粗分割延迟次数+每此粗分割的平均延迟时长)×游戏关卡，并基于 AIC 进行逐步回归得到最优模型为：创造力得分 = 存留水平+视频时长+相对层次齐性 +粗分割延迟次数+细分割一致性×游戏关卡。最优模型的多元线性回归模型结果如表 1 所示。

	Estimate	S.E.	<i>t</i>	<i>p</i>	
(截距)	0.847	(0.247)	3.430	< .001	***
存留水平中	-0.063	(0.037)	-1.718	.087	.
存留水平高	0.026	(0.049)	0.528	.598	
视频时长	-0.003	(0.000)	-8.314	< .001	***
相对层次齐性	0.064	(0.043)	1.482	.139	
细分割一致性	0.192	(0.308)	0.625	.533	
游戏关卡第二关	-0.115	(0.328)	-0.350	.727	
游戏关卡第三关	0.535	(0.341)	1.570	.118	
游戏关卡第四关	0.402	(0.330)	1.216	.225	
游戏关卡第五关	-0.005	(0.324)	-0.015	.988	
粗分割延迟次数	0.008	(0.004)	2.311	.022	*
细分割一致性:游戏关卡第二关	-0.077	(0.400)	-0.193	.847	
细分割一致性:游戏关卡第三关	-0.932	(0.406)	-2.294	.023	*
细分割一致性:游戏关卡第四关	-0.508	(0.427)	-1.188	.236	
细分割一致性:游戏关卡第五关	-0.214	(0.417)	-0.513	.608	

表 1 最优多元线性回归模型拟合结果

注: \*\*\*  $p < .001$ , \*  $p < .05$

从多元线性模型的逐步回归筛选结果与最终最优模型的拟合结果可以看出,对创造力评估得分最具预测效力的变量为视频时长,粗分割一致性、相对层次齐性与每此粗分割的平均延迟时长无法预测创造力评估的大小。存留水平对预测部分创造力评估大小的预测能力边缘显著。粗分割延迟的次数能显著预测创造力评估的大小,且回归系数为正,与预期相符。而细分割一致性只有在特殊关卡中才能有效预测创造力评估。

### 3.3 事件分割与创造力评估

在上部分的结果中,我们发现所选取的表示与创造性过程有关的结构变量基本不能很好地预测创造力评估的得分,不符合之前的预期。而在 3.1 中对与事件分割产生的结构变量本身的分析告诉我们被试对于事件分割任务本身的理解是较为一致的,且被试能较好地完成任务。为了解释为何事件分割的结构变量与创造力评估得分的关系,我们分析了被试对不同分割条件下、不同关卡、不同存留水平的事件分割点的分布,并对其密度分布进行了高斯拟合,粗事件分割条件下拟合的结果如图 18 至图 22 所示(细事件分割条件下的结果见附录 3)。

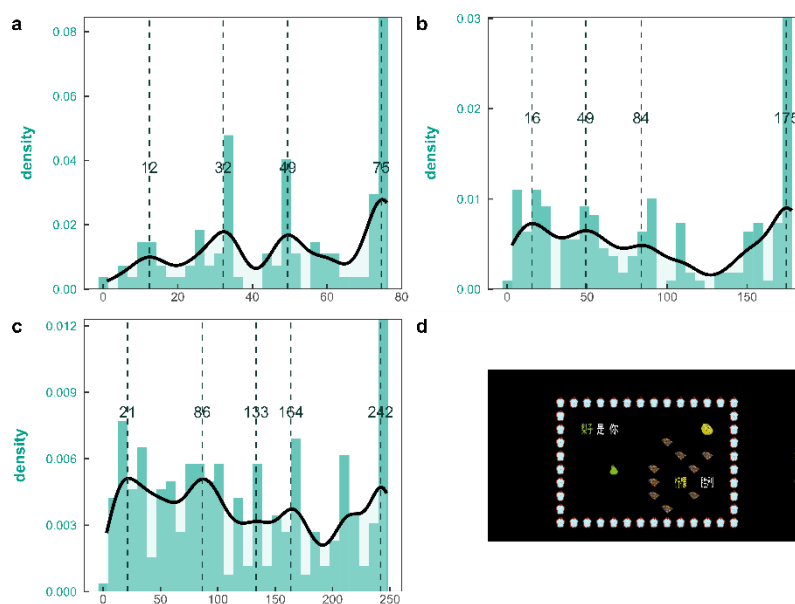


图 18 第一关中被试事件分割时间点密度图

注: i) a/b/c 分别代表存留水平低/中/高;

ii) a/b/c 三图中的竖直虚线为高斯拟合曲线的峰值所在位置, 数字为对应的时间

iii) d 图为该关卡的开始状态

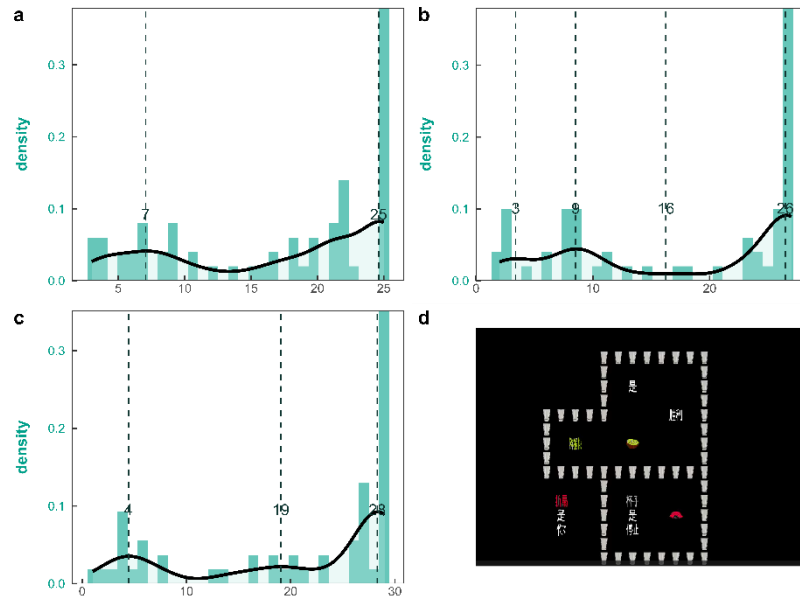


图 19 第二关中被试事件分割时间点密度图

注：i) a/b/c 分别代表存留水平低/中/高；

ii) a/b/c 三图中的竖直虚线为高斯拟合曲线的峰值所在位置，数字为对应的时间

iii) d 图为该关卡的开始状态

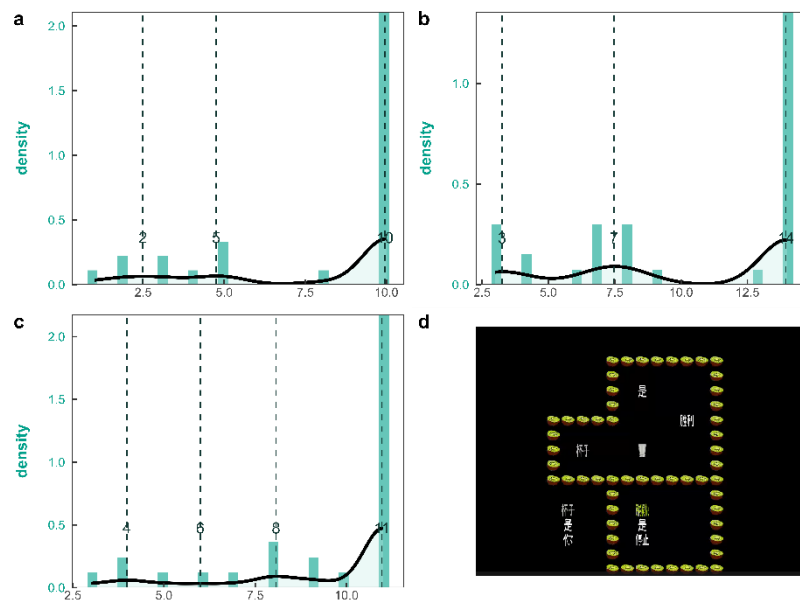


图 20 第三关中被试事件分割时间点密度图

注：i) a/b/c 分别代表存留水平低/中/高；

ii) a/b/c 三图中的竖直虚线为高斯拟合曲线的峰值所在位置，数字为对应的时间

iii) d 图为该关卡的开始状态

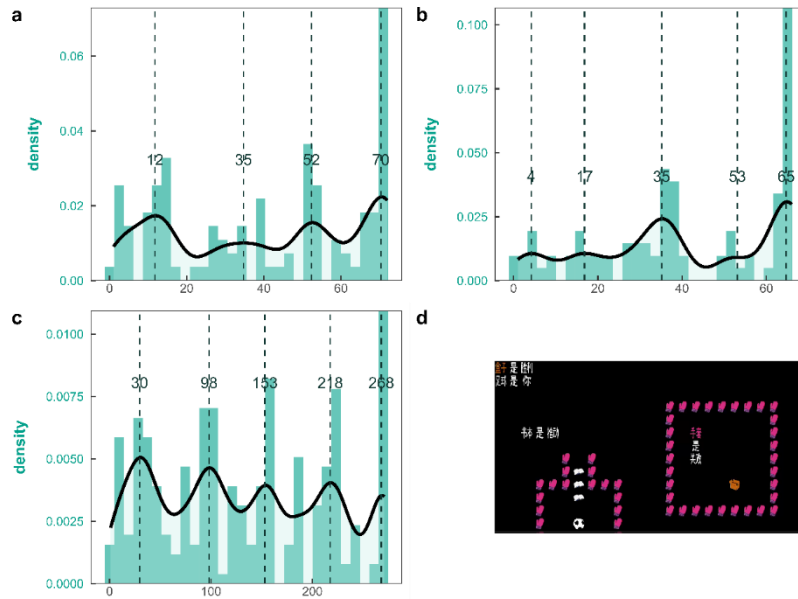


图 21 第四关中被试事件分割时间点密度图

注：i) a/b/c 分别代表存留水平低/中/高；

ii) a/b/c 三图中的竖直虚线为高斯拟合曲线的峰值所在位置，数字为对应的时间

iii) d 图为该关卡的开始状态

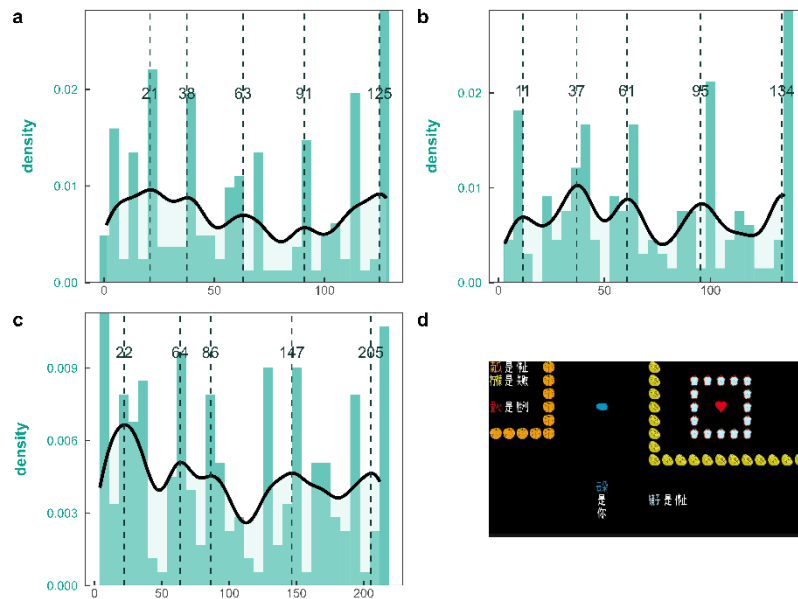


图 22 第五关中被试事件分割时间点密度图

注：i) a/b/c 分别代表存留水平低/中/高；

ii) a/b/c 三图中的竖直虚线为高斯拟合曲线的峰值所在位置，数字为对应的时间

iii) d 图为该关卡的开始状态

在与视频内容对比后，我们发现在分割中被试事件分割时间点的拟合分布曲线的分值往往是出现在被试观察到视频中的玩家撤销之前的操作、重启游戏状态、游戏失败等本身带有强烈事件结构特征的事件结束时。而对于一些通关所必需的或大多数人会想到的一些“妙手”的出现并未引发被试们的事件分

割。

因此，我们可以认为，被试虽然了解了任务的要求，但不能在观看通关视频时准确实时地识别出那些富有创造力的、或对通关至关重要的特殊操作，并在这些操作结束后做出事件分割的反应，从而导致了诸多所选取的表示与创造性过程有关的结构变量基本不能很好地预测创造力评估的得分。

## 4 讨论

在本研究中，我们采用事件分割的实验范式，要求被试在观看不同的游戏通关视频时实时地进行事件分割，并在视频结束后对该通关方法进行创造力评分。实验结果并未能验证预期的猜想，本文采用的几个分割创造性认知过程的结构变量并无法预测被试对其的创造力评估，唯有粗分割的延迟次数以及在特定关卡下的细分一致程度对于创造力评估具有一定的预测能力。对被试分割事件的时间点分布进行高斯拟合发现，被试在本实验中的分割基本根据视频中的玩家撤销之前的操作、重启游戏状态、游戏失败等本身带有强烈事件结构特征的事件的发生进行，与创造性过程中的重要子事件均无较大关系，这可能是导致无法得到预期研究成果的重要原因之一。

在 Eisenberg 等人(2018) 以及 Kurby 与 Zacks, 2011)使用事件分割范式的研究中，被试需要分割的事件均为日常生活的普通活动，例如换灯泡等。这些普通活动显然对于被试来说更为熟悉简单。而在我们的实验中虽然已让被试在正式实验前完成训练关卡以及训练试次，但我们并未在实验中检验被试是否真的理解视频中的玩家正在做什么，因此要求被试对此进行事件划分可能是一个过于困难的任务。同时，我们在收集数据时避开了具有相关游戏经验的被试，相较于日常生活中的普通活动，不具备相关游戏经验的被试很难在观看游戏视频时充分接收视频中玩家操作的信息量，其思路也难以跟上玩家的思路，从而干扰他们做出我们设计实验时所期待的反应。

除上述理由外，也有可能是在创造力评分前被试需完成繁琐的事件分割任务，干扰了被试对于玩家通关游戏方法的整体编码，导致被试对于不同时长视频的记忆强度不同，从而使视频时长更长的视频得到更低的创造力评分。或者在时长更长的通关视频中，玩家做出了更多的无关操作或重复操作，含有有效信息的操作占比较少，被试更能理解这些视频中玩家的思路，从而更感觉简单熟悉而给更低的创造力评分；而时长较短的视频播放时被试来不及步步思考，这样的视频可能带来更强的新异感，从而得到更高的创造力评分。由于每个被试都只能看到每个关卡中的三个通关视频与小于等于三种通关方法，因此也无法对玩家群体的整体通关方法有充分的了解，无法将观看的视频与玩家群体进行比较，只能根据对于视频的主观感受进行创造力评分，因此上述猜测有合理的可能性。

可以总结的是，对于较为复杂的创造性认知过程，要求被试在观看事件的

同时理解玩家的思路并做出事件的分割具有一定的难度，被试对于过程整体的主观感受与创造力评分受极多因素干扰。因此为研究在创造性认知过程中是否存在结构性的变量或某些特征可以预测他人对此的创造性评估，可以考虑人为将创造性认知过程拆分为若干子状态组成的状态集与将部分子状态映射到其他子状态的操作集，要求被试对集合内元素做出相应的反应，最终共同组成被试对于整体的认知与编码，同时也避免了时间信息对于创造力评估的影响。

此外，Miller 与 Jhala(2021)提出了一种观察者导向(observer-oriented)的创造力与美学价值评估理论框架，其核心思想是将观察者所掌握的概念空间与创造者在创造中体现的概念空间相比较，进而来得出观察者视角中创造力与美学价值的大小。而要将 Miller 与 Jhala(2021)的理论框架应用于具体计算模型搭建，核心在于如何获取观察者内隐的概念空间，并且将高维度的复杂信息量化以实现概念空间之间的比较。Petr 与 Radek(2010)有关推箱子游戏难度的计算提供了另一种思路。他们通过拟合启发式模型模拟人类的通关行为，根据模型的行为特征来表征推箱子游戏的难度；并将通关过程转化为有向图上状态的转移，将对通关过程的研究转化为对有向图图结构的研究，发现了有向图中的“瓶颈”结构，证明其对于通关游戏至关重要。作为推箱子游戏的变体，“BaBa Is You”同样可能可以通过设计某种更复杂的启发式模型以通关，并将人类被试的行为与模型相比较以构建创造力的某种指标。

综上所述，本研究虽未验证预期的猜想，但指出了在从事件结构探索创造性认知过程时需要注意的问题。并结合相关研究提出了未来实验可能可行的改进方向。

## 参 考 文 献

- Amabile, T. M. (1990). Motivation and personal histories. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 61 – 91). Newbury Park, CA: Sage Publications Inc.
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. London: Routledge.
- Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review*, 11(6), 1011–1026. <https://doi.org/10.3758/BF03196731>
- Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*, 9(1), 35–40. <https://doi.org/10.1037/a0038688>
- Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, 3. <https://doi.org/10.1186/s41235-018-0146-z>
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. The MIT Press.
- Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, 61, 569–598

- Jordanous, A., and Keller, B. (2012). What makes musical improvisation creative? *J. Interdiscip. Music Stud.* 6, 151–175. <https://doi.org/10.4407/jims.2014.02.003>
- Jordanous, A., and Keller, B. (2016). Modelling creativity: identifying key components through a corpus-based approach. *PLoS ONE* 11:e0162959. <https://doi.org/10.1371/journal.pone.0162959>
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, 13(1), 1–12. <https://doi.org/10.1037/a0013688>
- Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: Neural correlates of passive conceptual expansion. *Brain Research*, 1430, 52–61. <https://doi.org/10.1016/j.brainres.2011.10.031>
- Kröger, S., Rutter, B., Hill, H., Windmann, S., Hermann, C., & Abraham, A. (2013). An ERP study of passive creative conceptual expansion using a modified alternate uses task. *Brain Research*, 1527, 189–198. <https://doi.org/10.1016/j.brainres.2013.07.007>
- Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, 39(1), 75–91. <https://doi.org/10.3758/s13421-010-0027-2>
- Miller, C. & Jhala, A. (2021). An Observer-oriented Theory of Creativity and Aesthetic Measure. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43(43)
- Petr, J. & Radek, P.(2010). *Difficulty Rating of Sokoban Puzzle*. Proceedings of the Fifth Starting AI Researchers' Symposium. IOS Press
- Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan* 42, 305–310
- Rutter, B., Kröger, S., Stark, R., Schweckendiek, J., Windmann, S., Hermann, C., & Abraham, A. (2011). Can clouds dance? Neural correlates of passive conceptual expansion using a metaphor processing task: Implications for creative cognition. *Brain and Cognition*, 78(2), 114–122. <https://doi.org/10.1016/j.bandc.2011.11.002>
- Rutter, B., Kröger, S., Hill, H., Windmann, S., Hermann, C., & Abraham, A. (2012). Can clouds dance? Part 2: An ERP investigation of passive conceptual expansion. *Brain and Cognition*, 80(3), 301–310. <https://doi.org/10.1016/j.bandc.2012.08.003>
- Ward, T. B. (2007). Creative cognition as a window on creativity. *Methods*, 42, 28–37. <https://doi.org/10.1016/j.cogbrainres.2005.05.013>

## 附录 1 实验任务背景

### 任务背景告知：

为选出最具有创造力和想象力的玩家与通关方法，我们正在进行一项游戏通关方法的评比比赛。作为评审团的一员，你将协助我们完成一系列任务以完成这项评选。但为了让你可以更轻松地完成任务，在开始实验前我们将



一起简单熟悉本游戏的规则。

（进入游戏）

在游戏中，玩家使用“↑↓←→”键控制主人公的行动，并可以推动规则允许推动的任何物体。游戏中的句子完成了对游戏世界中世界规则的制订。游戏的目标便是使“主人公”获得“胜利”。玩家可以使用“q”键撤销上一步操作，也可以使用“z”键重置游戏局面重新开始。现在你可以自由探索游戏内容，并完成四个新手关卡。

## 附录 2 实验指导语

### 指导语：

接下来，你将会进入评估环节。在实验中，你将会看到其他玩家的通关视频，每个视频从玩家进入关卡开始，到玩家获得胜利结束。这批玩家使用的游戏界面和你使用的完全相同，因此通关的全过程中玩家可能会撤销操作、重启关卡、经历失败。

日常生活中的事件通常可以被分解为一系列事件单元。比如如果主试现在去关门，这个事件可以被分解为“转身”、“走到门前”、“拉动门”、“转身”、“走回来”5个事件单元。玩家的通关过程也是如此。我们希望通过你的评估对玩家的通关过程进行分解。因此，在观看视频的过程中，你需要实时地对通关过程进行事件单元的划分。划分没有标准的答案，你只需按照自己的理解进行反应。

在你划分出的每个事件单元中，可以包含一系列游戏的基础操作，比如上下左右按键、撤销、重启。当你认为，当前的事件单元结束，而下一个事件单元开始时，请迅速按下空格键进行反应。当你按下空格键后，会听到“嘟”声反馈。

对于每一个通关视频，你需要连续观看两遍，分别做出不同细致程度的划分，我们称之为“粗划分”和“细划分”。

第一遍观看时，你需要作出“粗划分”，将玩家的通关过程划分成最大的事件单元；第二遍观看时，你需要作出“细划分”，将视频划分为最小的事件单元。“最大”和“最小”没有固定的标准，不同的评分者会有自己的主观标准，你只需要按照自己的理解进行划分。

以日常事件为例，假设你要评估的是某人做早餐的过程。在第一遍“粗划分”时，你可能只会在“烤面包”这个事件单元的前后按空格做出划分标记，而在第二遍“细划分”时，你可能会在“烤面包”中细分出“打开一袋面包”“将面包放入烤面包机”“启动烤面包机”等多个事件单元。

针对当前你进行的划分程度的不同，在两次观看中按空格进行反应得到的“嘟”声反馈不同。

在本实验中，你的任务是：观看一系列通关视频，每个视频观看两遍，第一

遍进行“粗划分”，第二遍进行“细划分”，在观看完两遍视频后会对该视频进行创造力评分。“细划分”的方式是在当前的事件单元结束，而下一个事件单元开始时，按下空格键，按下空格键后会听到“嘟”声反馈。“粗划分”时需在发现事件单元结束时按下空格暂停视频，再按下键盘上方的数字键(0~9)报告本次的事件单元结束于几秒前(视频中 4 步操作约耗时 1 秒)(特别的，0 表示暂停时正好事件单元结束)，同样会在按下后听到“嘟”声反馈。每看完一个视频，你都可以自行休息，准备好后按鼠标左键进入下一个视频。

请注意：

- 1.【你观看的每个通关视频都来自不同的玩家】
- 2.视频中的通关过程并非完美，是玩家解谜过程的回放，可能会出现你认为无效或者错误的操作。
- 3.你需要完成的评估任务没有正确或错误的标准答案，你可以自由地按照自己的理解进行反应。

请按鼠标左键开始评估任务的练习，三组练习后进入正式实验

附录 3 细分割中被试事件分割事件点的分布

图 I 至图V 分别为关卡一至关卡五：

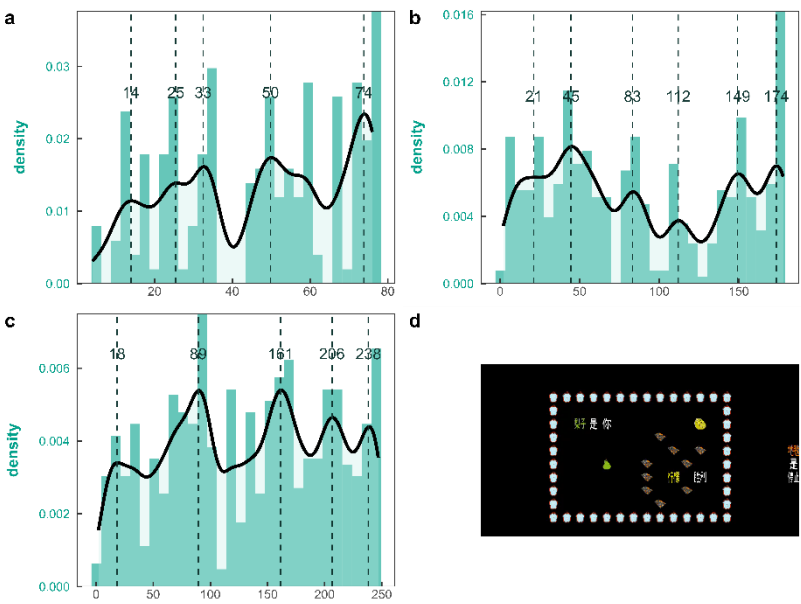
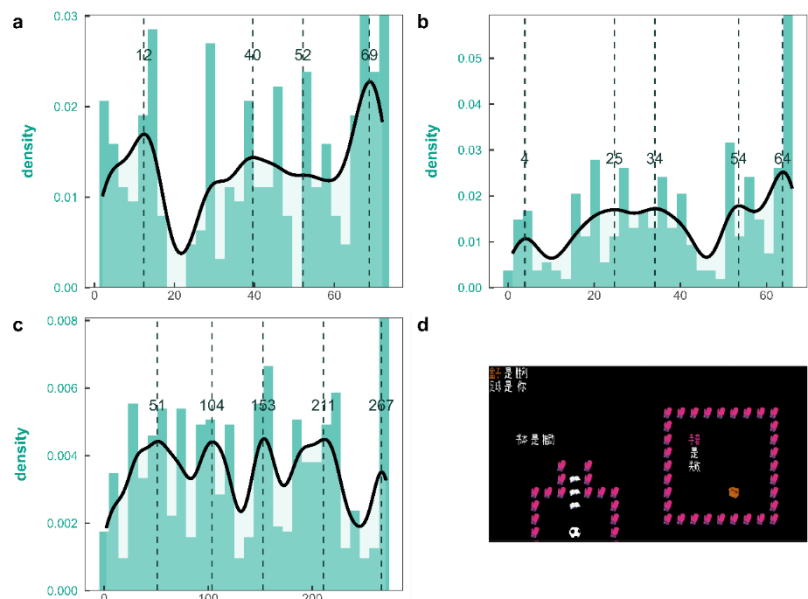
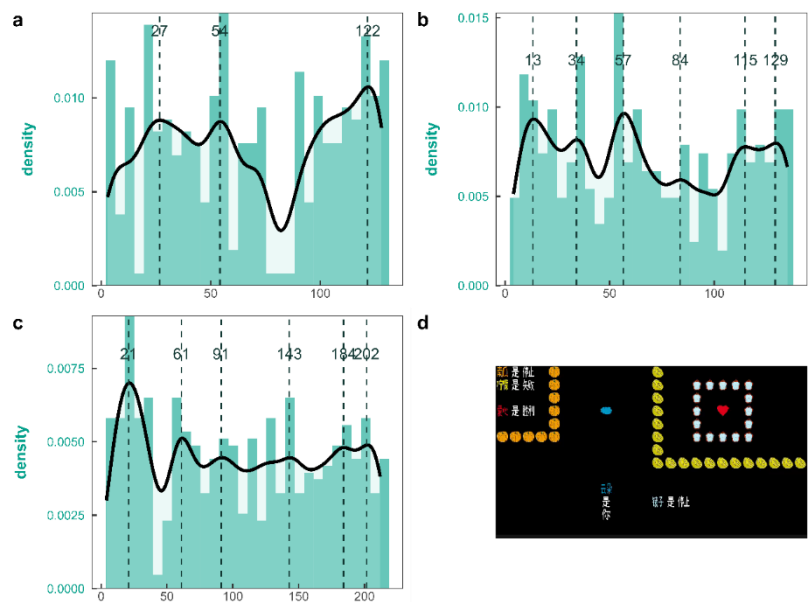


图 I





图IV



图V