在对话式检索情境下，检索任务与认知风格对大学生检索策略和行为的影响

林兴波、张轩、郭铭康

**一、概述**

**1、研究背景**

近几年，随着大语言模型和人工智能的应用，诸如ChatGPT、文心一言、New Bing等生成式搜索引擎得到研发使用。有别于传统的检索情境，在海量数据加持了，嵌入人工智能的生成式检索模式推动实现了从“检索”到“检索+生成”的升级，这种技术升级带来了检索结果的更多可能。由于生成式搜索引擎强大的功能，在大学校园也受到学生们的欢迎和使用。关于对话式搜索引擎的研究也越来越多。

**2、研究动机与研究目的**

目前关于对话式检索情境的研究多集中在技术的实现，关于交互式信息交互、人机交互的研究维度尚不完善，因此我们小组希望能够从两个维度探究用户与搜索引擎的交互关系，更加注重用户的角度通过实证研究探究其行为的差异。在这一研究动机之下，我们确定了两个主要的研究目的：（1）在对话式检索情境下，检索任务对大学生检索策略和行为的影响；（2）在对话式检索情境下，认知风格对大学生检索策略和行为的影响。

**二、文献回顾**

在当今研究现状下，在交互式信息检索领域，关于对话式检索的研究相对较少，但在近些年随着ChatGPT等大语言模型的爆火，又有许多研究者从这个方面进行研究。从对话式检索系统的角度，刘亚丽、范逢春等人基于扎根理论探讨了ChatGPT-AIGC用户风险感知维度识别与治理，Sakirin和 Said调查了用户对ChatGPT支持的会话界面与传统方法的偏好；从用户行为的角度，卢新元、雷晓鹏等人以对话式搜索引擎为例探究了人工智能生成内容环境下用户的信息行为，Frazier等人的研究表明用户发现会话代理有助于获取与软件相关的探索性搜索的有用信息，周涛,吴晓颖,邓胜利建模解释了用户从问答社区到生成式的知识问答转移行为，Ghosh等人研究的用户实验使用了WOZ的研究方法，确定了12个语言行为和4个搜索动作；从模型检索评估的角度，刘萍、杨志伟等人研究了认知建构视角下的交互式信息检索模型，Balog提出了一种基于用户模拟的对话式检索评估方法，刘畅与刘涵蕊探讨了认知风格对信息交互行为的影响，并提出一些刻量指标，帅训波,石文昌,冯梅提出了一种面向用户体验增强的信息检索动态评估模型，韩正彪,翟冉冉等人提取出了多种信息行为模型中情感的概念化，并对各种概念进行了细分和解释，揭示情感在用户信息行为模型中的作用机制。Belkin的知识异构状态理论和Kulthau提出的信息搜索过程理论也在对话式检索系统和检索行为评估上发挥着重要作用。

希望我们的研究能填补在对话式检索下用户检索行为研究的空缺，探讨用户认知风格和检索任务对检索行为的影响，为之后的研究提供理论和数据，在交互式检索领域中发挥出自己的力量。

**三、研究方法**

**1、测试系统**

本次实验我们选择的测试系统是New Bing。

2023年2月7日，微软正式宣布将ChatGPT集成进新版必应（New Bing）和Microsoft Edge浏览器中。微软将集成后的Bing称之为“新必应”（New Bing），集成后的聊天机器人称为“必应聊天”（Bing Chat）。New Bing采用OpenAI模型的升级版，即GPT-4，这比ChatGPT3.5版本更强大，模型更优，能够使用最新信息和注释答案更好地回答搜索查询，也可以切进或者切出聊天模式。它的搜索查询框最多可以接受1000个字符。New Bing会有三种模式，“精确”“平衡”“创意”，用户可以在这三种模式之间进行切换，以体验不同的聊天语气。综合来看，New Bing在对话式智能检索领域整体效果比较多，且系统稳定，比较适合进行测试和实验。

**2、研究设计**

**（1）操作流程**

本次实验主要包含三个具体的步骤：

A、首先被试需要完成一个基本背景的调查问卷和认知风格测试题。

B、被试需要阅读理解搜索任务，对搜索任务进行解读分析，并且可以在提供的纸张上记录分析自己的检索思路。接着开始正式检索，不限制检索的时间和检索次数，检索过程中被试可以选择手动打字或者复制粘贴的方式将检索的结果组织并记录到指定的记事本中，当被试认为自己已经获取所需要的所有信息时（或者认为已经完成检索任务时）即可停止检索。每个任务结束后，被试关闭浏览器和记事本，回答搜索后的问卷对任务完成的满意度进行评估。

C、当完成上述内容即可进入下一个检索任务，重复上述实验过程直至完成所有任务。

**（2）检索任务**

文本检索任务根据不同的目的和难度进行分类，可以分为信息查找型和探索分析型。对此我们分别就两种类型设计两道一共四道的检索问题。

A、信息查找型：这类任务已知检索任务的主题或者具体检索目标，通常是事实性的，比如日期、地点、人物、定义等。

任务一：中国长城是世界上规模最 大的文化遗产。由国家文 物局和国家测绘局联合历时两年2007 年 2 月 2009 年 4 月）完成的长城资源调查项目，首次对外发布了有关明长城的一系列权威调查和测量数据。你对这一项目和长城很感兴趣，想要该项目组发布的明长城总长度、起止点、省域分布等具体数据，以及明长城北京段的详细数据，了解长城的历史和作用。

任务二：随着当前信息资源的爆炸性增长和海量化，大数据时代已经到来，新兴的数据科学正在快速崛起。与此相适应，科学研究也进入到第四范式阶段。你想要了解什么是“第四范式”？最早由谁提出？它和以前的研究范式各有何特征与区别？

B、探索分析型：这类任务需要人类的智力劳动，通常是主观性的，比如批评、评价、评论、态度、情感等。

任务三：由于学习和工作的原因，你计划本月购买一台新的电脑，你听说Mac book Pro很好，但你的一位朋友建议你购买MacBook Air，而不是MacBook Pro。你想要搜索并找出什么是MacBook Air和MacBook Pro的主要区别和优劣，以及其他用户对它们的评价和观点，从而考虑买哪台电脑。

任务四：你打算参加学校组织的新生舞会，但此前没有参加过类似活动的你觉得迷茫，不知道该如何准备，你希望检索该如何在新生舞会上表现得体，事先应该做好怎样的准备工作，同时你也想了解一下新生舞会这种活动会有什么内容和流程。

**（3）被试样本的确定**

根据所学知识，结合本次实验的自身特点，我们采用拉丁方的方式，减少因为问题顺序而带来的对实验的误差，同时可以确定我们所需要的被试数量。经过计算，我们需要四名被试即可，考虑到数据量，我们最终确定八名被试样本进行实验，他们对应的测试问题顺序如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 被试样本 | 测试问题 | | | |
| S1 | 问题一 | 问题四 | 问题二 | 问题三 |
| S2 | 问题二 | 问题一 | 问题三 | 问题四 |
| S3 | 问题三 | 问题二 | 问题四 | 问题一 |
| S4 | 问题四 | 问题三 | 问题一 | 问题二 |
| S5 | 问题一 | 问题四 | 问题二 | 问题三 |
| S6 | 问题二 | 问题一 | 问题三 | 问题四 |
| S7 | 问题三 | 问题二 | 问题四 | 问题一 |
| S8 | 问题四 | 问题三 | 问题一 | 问题二 |

**3、操作化设计**

**（1）认知风格**

镶嵌图形检测用于学生场独与场依存的认知风格分类，又叫隐蔽图形测验，美国心理学家威特金等人在研究知觉时发现，场独立性的人比场依存性的人更容易离析出知觉单元，基于此原理编制了此量表。量表分为三个部分：第一部分，9道题提供给被试熟悉题型练习使用，不计入成绩；第二、三部分各10道题，要求正确画出指定的简单图形，其中各部分的1、2题0.5分，3、4题1分，5-10题1.5分，满分24分（测试量表附在后面）。被试得分越高，说明其更善于排除周围环境的干扰，倾向于场独立认知风格，反之则属于场依存认知风格。

具体计算公式为：；。（其中，成年总体的常规的常模分数为9.76，常模标准差为4.57）

根据公式计算，T大于50则表明倾向于场独立，小于50则倾向于场依存。

**（2）评估指标**

将搜索交互行为分为搜索行为、阅读行为和记录行为。

A、搜索行为的相关指标[[1]](#footnote-1)：搜索对话记录；搜索生成结果阅读（停留）总时间、平均时间；搜索生成结果利用率（每个生成结果中打开链接或直接产生记录行为的次数）

B、阅读行为的相关指标：内容页面数量、内容页面总阅读时间、内容页面平均阅读时间；重复阅读率；有效内容页面数量；有效内容页面占比

C、记录行为的相关指标：记录次数；记录总时间；平均记录时间

**（3）搜索对话记录**

首先根据已有文献中的十二个言语行为对对话记录中用户的行为进行标签，发现主要能够产生检索操作的言语行为可以分为四种：Question or Seek，Inform or Declare，Instruct，Repeat。

在本实验中，规定Question型为直接向系统进行提问，提问式可以不是问句形式；Declare型为用户进行追加信息或相关提问，与Question型的主要区分点在于其能否单独提供检索必要的信息，以及是否和上一提问式具有承接关系；Instruct型为用户指导系统进行检索，在本次实验中并未出现故忽略；Repeat型为用户因某些缘故如对生成结果的精度不满意，从而重复对一个提问式或片段再次提问，与Declare型的区别在于Repeat型应不具有不存在于原始提问式的额外信息。

其次，还研究了用户提问式的语法结构。在这一点上将提问式分为两种：问句和非问句，后者包括陈述句和短语。

再者注意到用户在提问式构建时有不同的习惯，因此从提问式的构建行为上又将其分为两种：自行构建和外源构建。自行构建指用户充分阅读了任务或分析结果后再完全自行构造提问式，而相对的，外源构建指用户通过对于任务文本的片段进行复制粘贴以及进行少量修改，或采用系统提供的提问式进行检索。这一点可能可以体现用户在分析和操作上的重心。

**4、数据分析方法**

本实验采用双因素方差分析方法研究认知风格与任务类型对用户对话式检索流程中策略和行为的影响。数据分析流程以认知风格与任务类型两个分类变量作为因子，为使研究结果具有普适性，尽可能采用平均、占比等方式标准化数据，主要将这些处理后字段作为因变量，每一个session单独作为一条数据进行处理，总计32条数据。判定因子是否对因变量具有影响主要从两个方面考虑：差异是否显著以及是否具有统计学意义。对于每一组分析（即两个因子对应一个选定的因变量）首先进行因子分类下的描述性统计，获取不同分类下因变量的均值和方差，并且绘制估计边际均值图像以显性观测差异的显著性；其次进行方差分析，分别对于两个因子以及因子交互进行计算从而得到F值与值，根据后者判断差异是否具有统计学意义。通常而言当值大于0.05时即认为不具有统计学意义，然而由于本实验中数据量较小且具有较大的异质性，因此我们放宽条件，并根据值大小对差异的统计学意义进行排序：时差异具有强统计学意义，时差异具有弱统计学意义，时差异具有极弱统计学意义。

在上文中方差分析中提到了我们不仅要分析两个因子各自对因变量的影响，还要考虑两个因子间是否具有交互作用，即因子是否共同对因变量有影响。一种简单的判别方式是观察绘制的估计边际均值图像：若代表同一因子不同类别的线段或线段的延长线具有交点，则可以认为因子间具有交互作用；若线段间近似平行，则认为因子间不具有交互作用。而具体的交互作用是否有统计学意义还要检查因子交互的值。

**四、测试结果**

根据我们的统计结果，收集了八位被试样本的基本数据。基本数据如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 被试 | 年龄 | 性别 | 年级 | 专业 | 是否用过New Bing |
| S1 | 男 | 21 | 大三 | 经院 | 是 |
| S2 | 男 | 22 | 大三 | 法学院 | 是 |
| S3 | 男 | 20 | 大三 | 政管 | 是 |
| S4 | 女 | 21 | 大三 | 信管 | 是 |
| S5 | 男 | 20 | 大三 | 信管 | 是 |
| S6 | 男 | 21 | 大三 | 信管 | 否 |
| S7 | 男 | 21 | 大三 | 信管 | 否 |
| S8 | 女 | 20 | 大三 | 信管 | 是 |

按照实验设计，我们要求被试在检索回答每一道问题之后对回答进行一个基本评判，采取五刻度量表，分越高，认为完成地越满意，根据统计，四道题的满意度均值分别为3.88、4.5、3.75、3.38，均高于平均值，可以认为数据质量有一定保障。

在实验数据的基础上，我们对数据分析指标进行了一定的调整。由于在实验中，仅由一人实际点击了链接进入结果内容页面，因此在数据分析中不再考虑内容页面相关属性，将内容页面阅读时间合计入搜索结果阅读时间。调整后的分析指标可以根据其意义大致分为三类：检索效率、时间分配和提问策略相关。结果分析重点关注相对具有统计学意义的一些差异，对于可能出现问题的分析结果有备注。

**（1）检索效率**

检索效率部分主要分析检索行为概况的因变量，包括检索总时间、提问次数、记录次数、有效结果比率。具体的描述统计与方差分析结果见下表。

表格1检索效率因变量的描述统计与方差分析结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 认知风格 | | 任务类型 | | 交互作用 |
| 场依存 | 场独立 | 事实 | 探索 |
| 总时间 | Mean | 145.56 | 211.5 | 177.87 | 179.19 |  |
| std | 85.926 | 91.694 | 108.712 | 79.314 |  |
| F | 4.203 | | 0.002 | | 0.619 |
| sigma | 0.05 | | 0.968 | | 0.438 |
| 提问次数 | Mean | 1.88 | 2.06 | 2 | 1.94 |  |
| std | 0.885 | 1.124 | 1.211 | 0.772 |  |
| F | 0.257 | | 0.029 | | 0.029 |
| sigma | 0.616 | | 0.867 | | 0.867 |
| 有效结果比率 | Mean | 0.520164 | 0.735417 | 0.641369 | 0.614211 |  |
| std | 0.299769 | 0.305619 | 0.314443 | 0.329708 |  |
| F | 3.792 | | 0.06 | | 0.06 |
| sigma | 0.062 | | 0.808 | | 0.809 |
| 记录次数 | Mean | 1.69 | 1.63 | 1.69 | 1.63 |  |
| std | 0.793 | 0.619 | 0.793 | 0.619 |  |
| F | 0.058 | | 0.058 | | 0.058 |
| sigma | 0.812 | | 0.812 | | 0.812 |

可以看出总时间和有效结果比率在因子认知风格上，前者具有强统计学意义，后者具有弱统计学意义。具体而言，场独立型用户的检索总时间整体较长，说明场独立型用户更倾向于投入更多时间进行检索任务；同样的，场独立型用户的检索有效结果比率更高，结合不同认知风格用户的记录次数基本相近的结果，说明场独立型用户的提问式构建可能更有效率。综合两者，可以推测场独立型用户倾向于或能够在检索任务中获取更多信息。

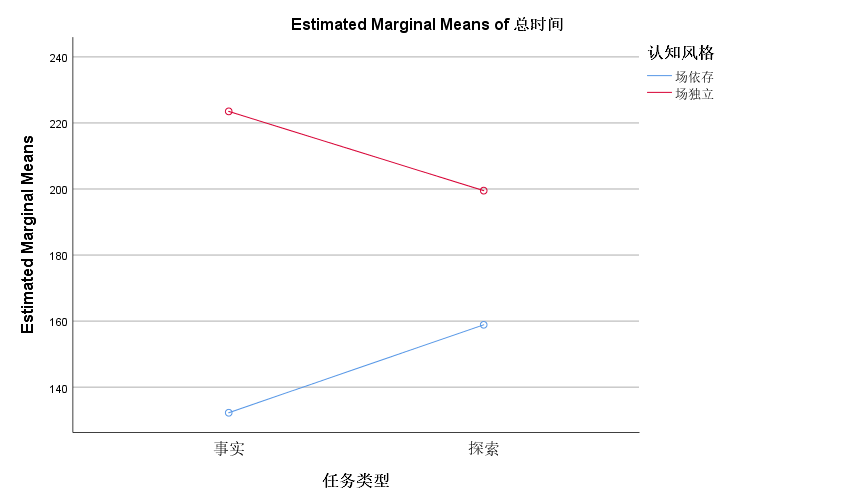


图1总时间的估计边际均值

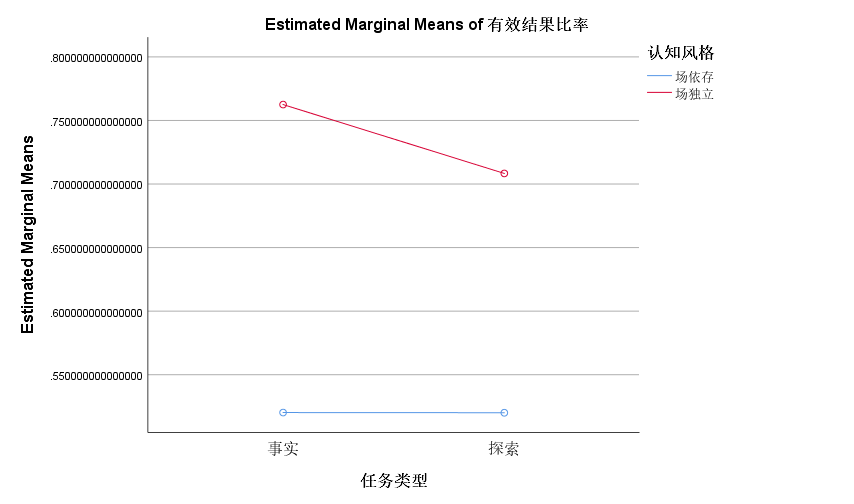


图2有效结果比率的估计边际均值

**（2）时间分配**

时间分配部分主要关注用户的平均时间开销以及时间分配方式，研究指标包括：平均搜索结果阅读时间、平均记录时间、阅读记录时间比值、阅读记录平均时间比值、阅读时间占比、记录时间占比。具体的描述统计与方差分析结果见下表。

表格2:时间分配因变量的统计描述与方差分析

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 认知风格 | | 任务类型 | | 交互作用 |
| 场依存 | 场独立 | 事实 | 探索 |
| 平均搜索结果阅读时间 | Mean | 16.25357 | 51.02708 | 35.18616 | 32.09449 |  |
| std | 13.32546 | 33.27485 | 33.24997 | 28.6187 |  |
| F | 14.364 | | 0.114 | | 0.503 |
| sigma | 0.001 | | 0.739 | | 0.484 |
| 平均记录时间 | Mean | 12.20833 | 28.44792 | 18.42708 | 22.22917 |  |
| std | 9.489858 | 23.10313 | 19.91128 | 18.98146 |  |
| F | 6.392 | | 0.35 | | 0 |
| sigma | 0.017 | | 0.559 | | 0.991 |
| 阅读记录时间比值 | Mean | 4.090676 | 3.955659 | 4.235633 | 3.810702 |  |
| std | 3.968433 | 3.765427 | 3.524279 | 4.173701 |  |
| F | 0.009 | | 0.092 | | 0.473 |
| sigma | 0.924 | | 0.764 | | 0.497 |
| 平均阅读记录时间比值 | Mean | 1.964788 | 2.38111 | 2.321771 | 2.024126 |  |
| std | 2.441168 | 2.044848 | 2.04759 | 2.448116 |  |
| F | 0.272 | | 0.139 | | 1.708 |
| sigma | 0.606 | | 0.712 | | 0.202 |
| 阅读时间占比 | Mean | 0.722772 | 0.702593 | 0.752191 | 0.673174 |  |
| std | 0.137722 | 0.174599 | 0.116618 | 0.180942 |  |
| F | 0.132 | | 2.022 | | 0.013 |
| sigma | 0.719 | | 0.166 | | 0.91 |
| 记录时间占比 | Mean | 0.277228 | 0.297407 | 0.247809 | 0.326826 |  |
| std | 0.137722 | 0.174599 | 0.116618 | 0.180942 |  |
| F | 0.132 | | 2.022 | | 0.013 |
| sigma | 0.719 | | 0.166 | | 0.91 |

可以观察到因子认知风格对平均搜索结果阅读时间和平均记录时间的差异有强统计学意义，因子任务类型对阅读时间占比和记录时间占比的差异有极弱统计学意义，两个因子的交互作用对阅读记录平均时间比值有极弱统计学意义。

在实际检查数据时发现每个用户个体的记录时间具有很大差异，通过对实验录屏的分析发现这是由于用户的记录方式不同导致的，实验中多数用户选择使用复制粘贴以进行记录，而有少数用户打字整理记录，后者的记录时长远大于前者，此时进行双因素方差分析会明显违反其前提假设：任一分类中不存在显著异常值、任一分类中残差近似正态分布。而由于原本数据过少，去除离群点的处理方法不可信，因此关于记录时长的相关指标分析不具有统计学意义，在此部分不再讨论。规范实验记录方式或研究记录策略是本实验改进的要点之一。

对于平均搜索结果阅读时间，场独立型用户大于场依存型用户，说明前者可能更注重于结果分析。

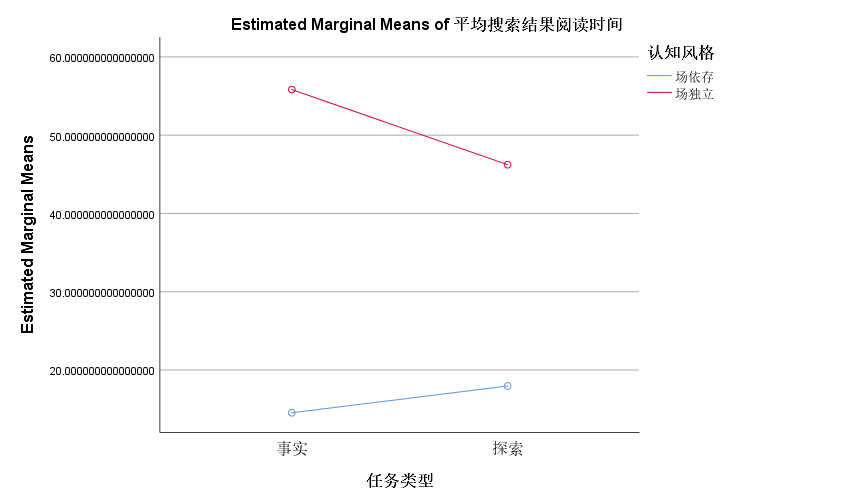


图3平均搜索结果阅读时间的估计边际均值

**（3）提问策略**

此部分主要分析用户的提问策略，关注用户提问式的构建。主要包括四种类型的提问方式占比、两种提问语法结构分类的占比和两种提问式构建来源的占比。由于实验中未出现Instruct型提问式，因此我们在分析中忽略其影响。各因变量的描述性统计与方差分析结果见下表。

表格3提问策略因变量的描述统计与方差结果分析

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 认知风格 | | 任务类型 | | 交互作用 |
| 场依存 | 场独立 | 事实 | 探索 |
| Question占比 | Mean | 0.817708 | 0.835417 | 0.845833 | 0.807292 |  |
| std | 0.285804 | 0.266797 | 0.283333 | 0.268257 |  |
| F | 0.031 | | 0.146 | | 0.117 |
| sigma | 0.862 | | 0.705 | | 0.735 |
| Declare占比 | Mean | 0.182292 | 0.077083 | 0.097917 | 0.161458 |  |
| std | 0.285804 | 0.168531 | 0.216185 | 0.259038 |  |
| F | 1.532 | | 0.559 | | 0.012 |
| sigma | 0.226 | | 0.461 | | 0.913 |
| Repeat占比 | Mean | 0 | 0.088 | 0.056 | 0.031 |  |
| std | 0 | 0.1893 | 0.1548 | 0.125 |  |
| F | 3.251 | | 0.265 | | 0.265 |
| sigma | 0.082 | | 0.61 | | 0.61 |
| 问句占比 | Mean | 0.572917 | 0.425 | 0.497917 | 0.5 |  |
| std | 0.403543 | 0.465475 | 0.473829 | 0.408248 |  |
| F | 0.888 | | 0 | | 0.888 |
| sigma | 0.354 | | 0.99 | | 0.354 |
| 非问句占比 | Mean | 0.427083 | 0.575 | 0.502083 | 0.5 |  |
| std | 0.403543 | 0.465475 | 0.473829 | 0.408248 |  |
| F | 0.888 | | 0 | | 0.888 |
| sigma | 0.354 | | 0.99 | | 0.354 |
| 自行构建占比 | Mean | 0.46875 | 0.182292 | 0.234375 | 0.416667 |  |
| std | 0.464354 | 0.313609 | 0.40279 | 0.421637 |  |
| F | 4.246 | | 1.719 | | 0.743 |
| sigma | 0.049 | | 0.2 | | 0.396 |
| 外源构建占比 | Mean | 0.53125 | 0.817708 | 0.765625 | 0.583333 |  |
| std | 0.464354 | 0.313609 | 0.40279 | 0.421637 |  |
| F | 4.246 | | 1.719 | | 0.743 |
| sigma | 0.049 | | 0.2 | | 0.396 |

可以看到因子认知风格对于Declare型（极弱）、Repeat型（弱）、自行构建提问式和外源构建提问式占比（强）的差异有统计学意义，而因子任务类型对两种提问式构建来源占比的差异有极弱统计学意义，而交互作用没有明显统计学意义。

具体而言，场依存型用户的Declare型提问占比整体略高于场独立型用户，即说明前者略倾向于向提供额外信息进行继续检索而非重新构造提问式。关于Repeat型提问的占比由于出现这种提问类型的用户数过少，因此认为其差异分析是无效的，不再陈述。关于自行构建和外源构建提问式两种提问式来源，由于互补性，因此认为对两者的方差分析是等效的。场依存性用户自行构建提问式的倾向更大，且探索型任务用户自行构建检索式的概率更高，但两者的交互作用并不具有统计学意义。

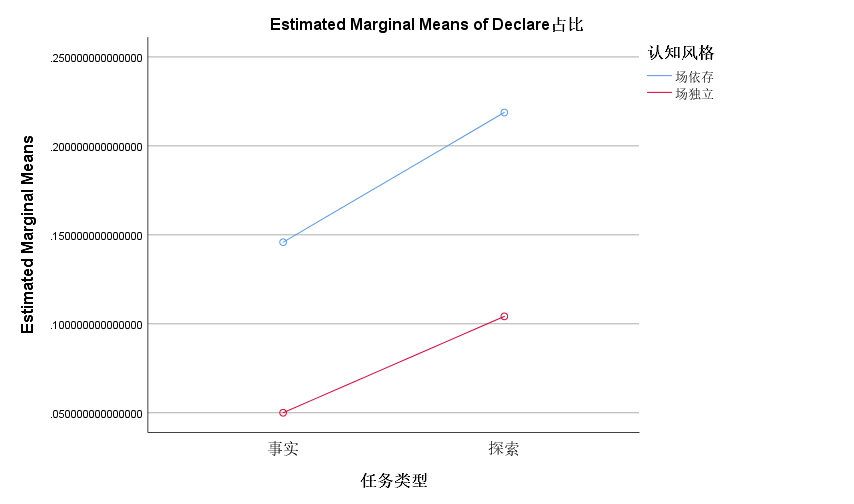


图4:Declare型提问占比的估计边际均值

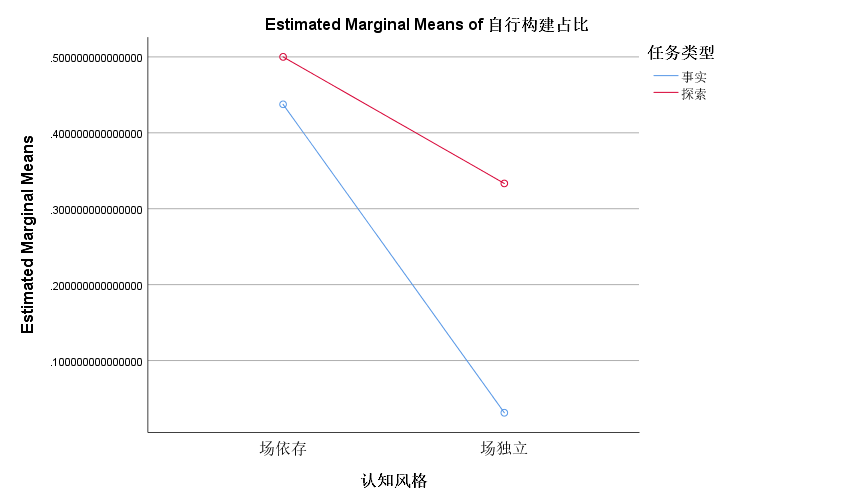
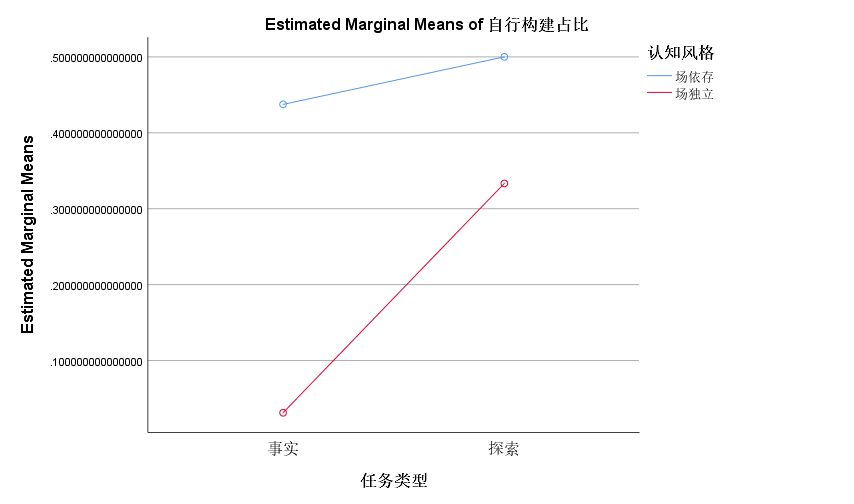


图5自行构建提问式占比的估计边际均值

**（4）数据分析总结**

基于上述的实验结果，可以得出一些较为有效的实验结论：

* 场独立型用户的检索总时间整体较长；
* 场独立型用户的检索有效结果比率更高；
* 场独立型用户平均搜索结果阅读时间大于场依存型用户；
* 场依存型用户的Declare型提问占比整体略高于场独立型用户；
* 场依存性用户自行构建提问式的倾向更大；

极弱统计学意义的结论：

* 探索型任务用户自行构建检索式的概率更高；

可以发现在已有结论中因子对于用户在对话式检索中的影响主要来自认知风格，而两个因子之间基本不具有统计学意义上的交互作用。

**五、研究不足与展望**

在本次研究中我们整体设计思路比较明确，对指标定义和刻量说明做了说明，并且通过数据分析得到一些结论。但是限于我们成员知识水平和操作能力，本次实验也有诸多不足之处：（1）研究被试的选定，在统计数据中也可以看出，被试的误差（或变量控制）没有得到完善，这有可能会对我们的数据产生一定的影响；（2）研究问题的设定，研究问题是我们小组成员参考一定的文献进行自主设计，问题代表性、问题难度、问题检验度需要更多数据的支撑；（3）实验的流程，我们参考文献提供的思路，针对我们的实验目的，基本采用时间逻辑顺序跟进，但是不能克服实验进程中各种变量对我们实验带来的影响；（4）此数据分析过程存在一些问题，人工实验、数据处理分析过程具有主观误差；数据量小、异质性大，统计性意义可能不明显；可能并没有满足双因素方差分析的假设4-6：任一分类中不存在显著异常值、残差近似正态分布、具有等方差性；（5）在本次研究中，我们小组对“检索策略”这一概念没有明确定义，在指标度量上没有将“策略”和“行为”分开说明；（6）对于对话记录这一重要研究角度，我们的研究方式和研究内容尚缺理论性说明。

悟已往之不谏，知来者之可追。正视整个实验中的问题与不足，我们希望能够有机会研究和学习更多相关内容的文章、期刊，丰满我们的羽翼，在老师的指导下能够重新审视研究问题和研究方法，完善研究流程，减少实验中的误差对数据的影响，通过更加科学有效的指标度量我们的研究问题，如果有机会，也希望试图提出一定的模型以供探讨和交流。

**六、小组分工**

|  |  |
| --- | --- |
| 成员 | 分工 |
| 林兴波（组长） | 实验计划和报告内容规划、实验数据收集、撰写研究问题陈述、实验设计、总结等内容，组织最终报告 |
| 张轩 | 实验计划和报告内容规划、实验数据收集、撰写文献综述、审查最终报告 |
| 郭铭康 | 实验计划和报告内容规划、实验数据分析、实验结果可视化、审查最终报告 |

**七、参考文献**

[1]S. Ghosh, S. Ghosh and C. Shah, "Toward Connecting Speech Acts and Search Actions in Conversational Search Tasks," 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Santa Fe, NM, USA, 2023, pp. 119-131, doi: 10.1109/JCDL57899.2023.00027.

[2]韩正彪,翟冉冉.用户信息行为模型中情感的作用机制研究[J/OL].情报理论与实践:1-13[2023-12-07].http://kns.cnki.net/kcms/detail/11.1762.G3.20220809.0936.002.html.

[3]M. Frazier, S. Kumar, K. Damevski and L. Pollock, "Investigating User Perceptions of Conversational Agents for Software-related ExploratoryWeb Search," 2022 IEEE/ACM 44th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER), Pittsburgh, PA, USA, 2022, pp. 51-55, doi: 10.1145/3510455.3512778.

[4]帅训波,石文昌,冯梅,等. 面向用户体验增强的信息检索评估模型研究[J]. 电子技术应用,2023,49(8):88-92. DOI:10.16157/j.issn.0258-7998.233743.

[5]刘亚丽,范逢春.ChatGPT-AIGC用户风险感知维度识别与治理研究——基于扎根理论的探索性分析[J/OL].情报理论与实践:1-13[2023-12-07].http://kns.cnki.net/kcms/detail/11.1762.G3.20231031.1540.004.html.

[6]周涛,吴晓颖,邓胜利.用户知识问答转移行为研究:从问答社区到生成式AI[J/OL].情报杂志:1-8[2023-12-07].http://kns.cnki.net/kcms/detail/61.1167.G3.20231017.1328.002.html.

[7]Sakirin, T., & Said, R. B. (2023). User preferences for ChatGPT-powered conversational interfaces versus traditional methods. Mesopotamian Journal of Computer Science, 2023, 24-31.

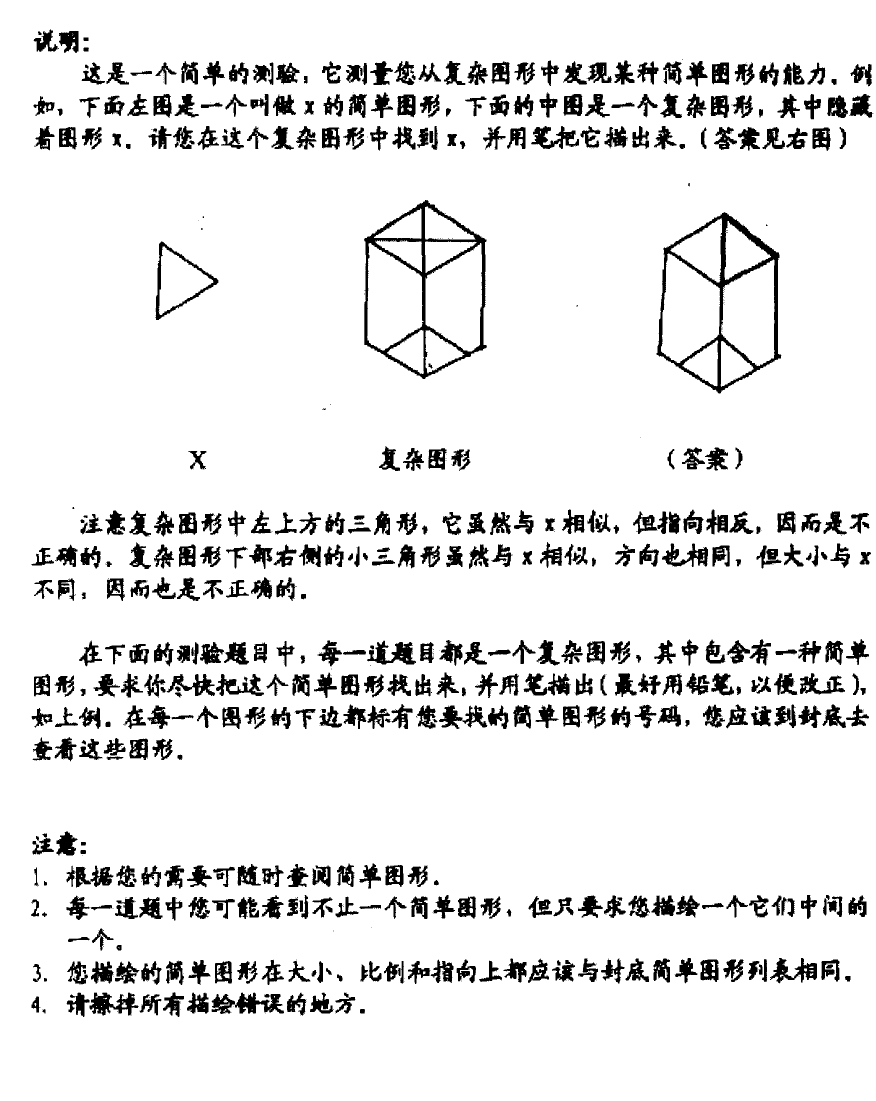
[8]Balog, K. (2021). Conversational AI from an information retrieval perspective: Remaining challenges and a case for user simulation.

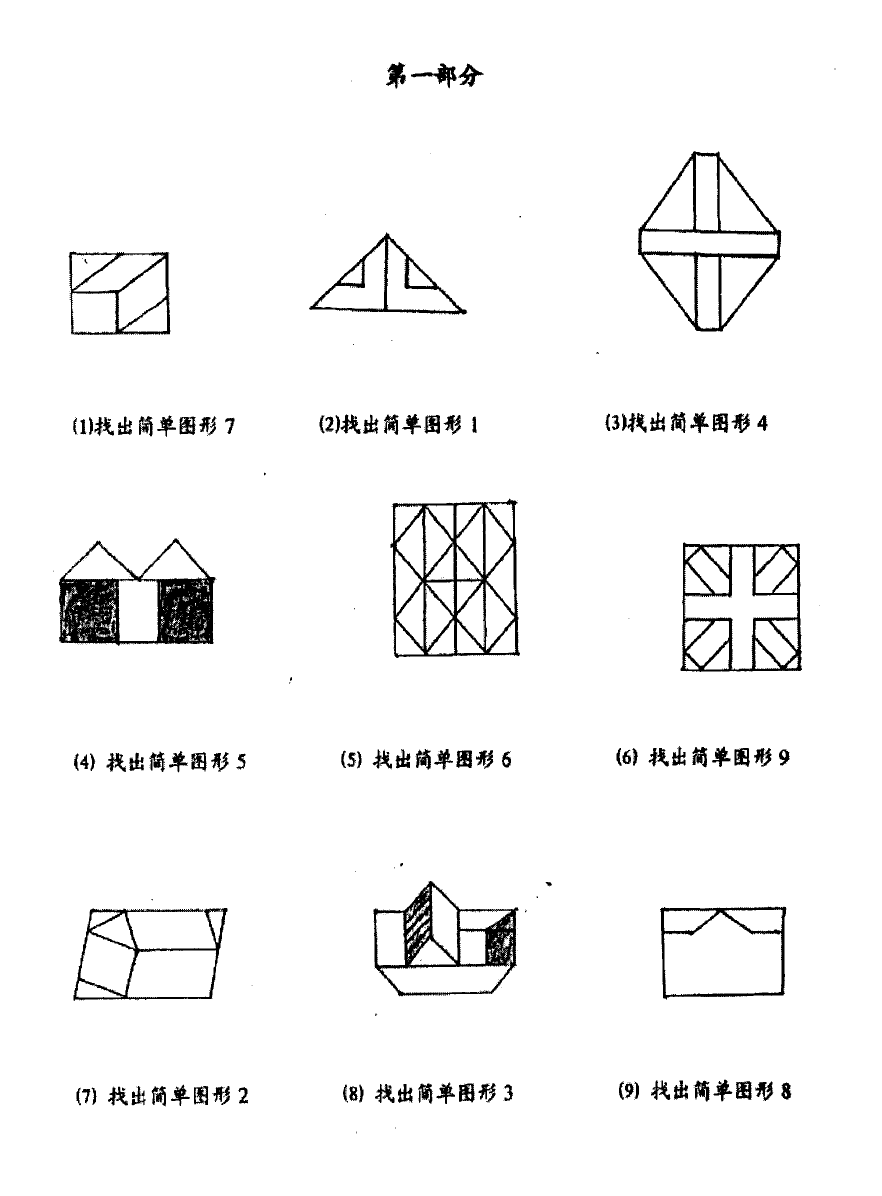
**八、附录**

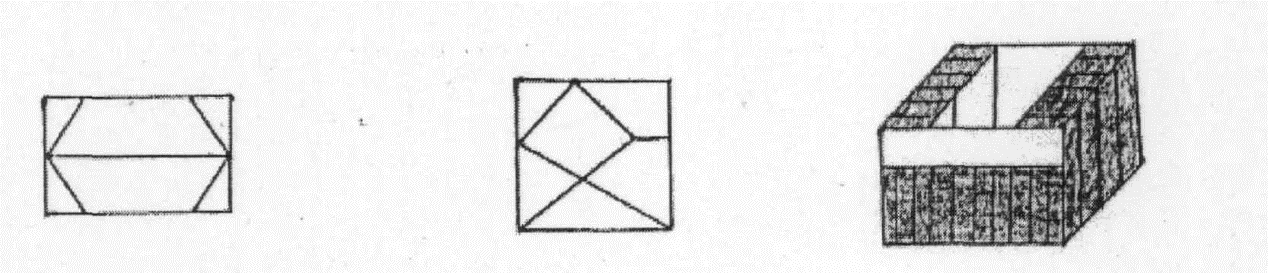
1、镶嵌图形测试

**Cognitive Style Figure Test**

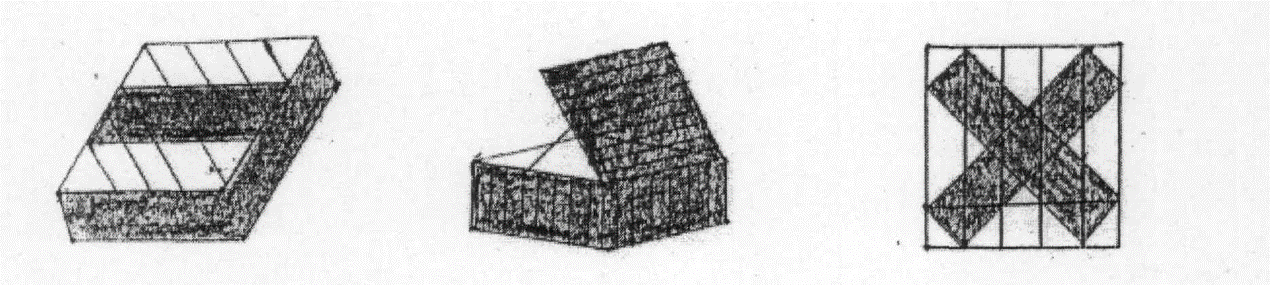
**镶嵌图形测试**



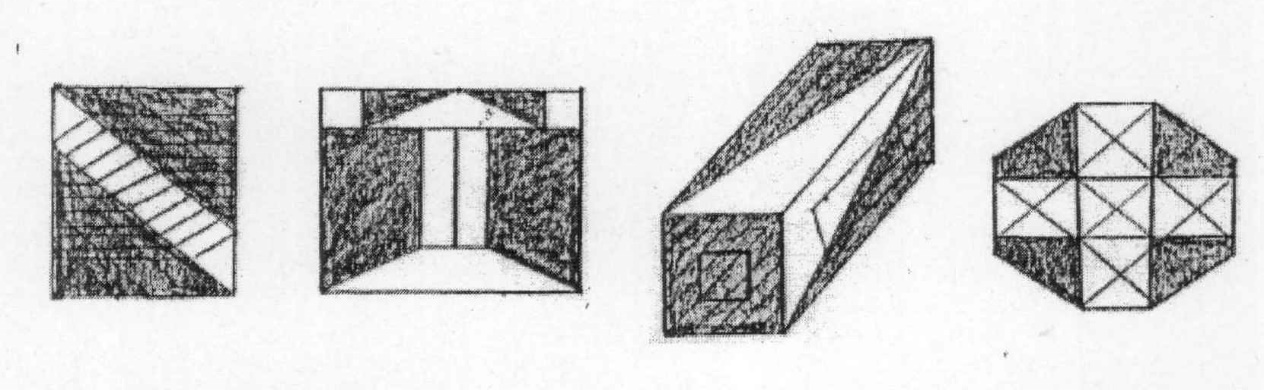




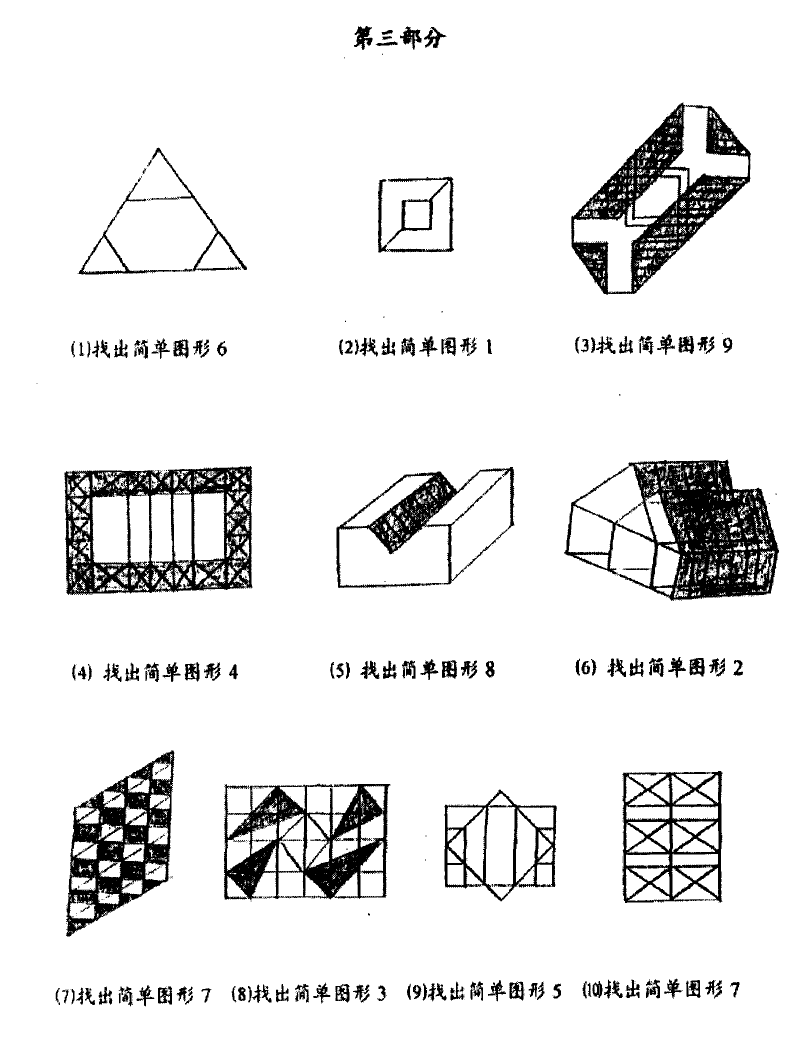
（1）找出简单图形6 （2）找出简单图形 5 （3）找出简单图形9

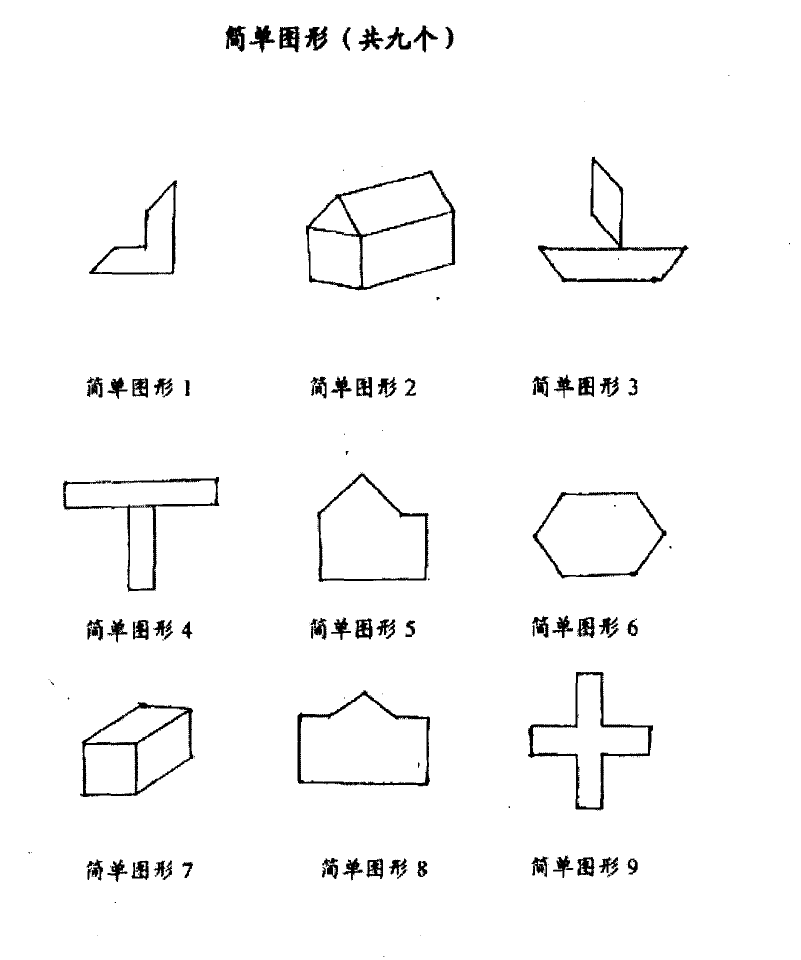


（4）找出简单图形6 （5）找出简单图形2 （6）找出简单图形3



（7）找出简单图形8 （8）找出简单图形4 （9）找出简单图形1 （10）找出简单图形7





1. 刘涵蕊,刘畅.认知风格与话题熟悉度对学习型任务下搜索交互行为的影响研究[J].情报理论与实践,2018,41(04):56-62.DOI:10.16353/j.cnki.1000-7490.2018.04.011. [↑](#footnote-ref-1)