



第四军医大学

The Fourth Military Medical University

分类号 B849

U D C 159.98

密 级 公开

博士学位论文

基于全息项目 Bifactor 模型探讨统计计算能力在决策中的认知优势及其机制

张家喜

培 养 类 别 全日制

学 位 类 型 学术学位

一级学科(专业类) 心理学

二级学科(专业) 应用心理学

研 究 方 向 心理测量与决策

指 导 教 师 苗丹民 教授

培 养 单 位 医学心理学系

二〇一五年五月

独 创 性 声 明

秉承学校严谨的学风与优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，不包含本人或他人已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了致谢。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名：_____日期：_____

保 护 知 识 产 权 声 明

本人完全了解第四军医大学有关保护知识产权的规定，即：研究生在校攻读学位期间论文工作的知识产权单位为第四军医大学。本人保证毕业离校后，发表论文等使用本论文工作成果时第一署名单位仍然为第四军医大学。学校可以公布论文的全部或部分内容（含电子版，保密内容除外），可以采用影印，缩印或其他复制手段保存论文；学校有权允许论文被查阅和借阅，并在校园网上提供论文内容的浏览和下载服务。同意学校将论文加入《中国优秀博硕士学位论文全文数据库》和编入《中国知识资源总库》等，并可浏览和下载，同时享受《中国优秀博硕士学位论文全文数据库出版章程》规定的相关权益。

论文作者签名：_____导师签名：_____日期：_____

目 录

缩略语表	1
中文摘要	2
英文摘要	5
前 言	9
文献回顾	12
正 文	29
第一部分 全息 Bifactor 模型中统计计算能力的测量.....	29
1 问题提出	29
2 研究方法	31
2.1 研究被试及其程序	31
2.2 测量工具	31
2.3 数据分析	32
3 研究结果	34
3.1 验证性因素分析（CFA）结果	34
3.2 CRT 和 BNT 联合测验的 IRT 分析	35
3.3 Bifactor 模型中 CRT 和 BNT 联合测验公共因子的信度分析	35
3.3 CRT 和 BNT 联合测验公共因子的效度分析	36
4 讨论	37
4.1 公共因子测量统计计算能力符合统计计算能力理论构想	37
4.2 公共因子测量统计计算能力信效度较好	38
第二部分 基于全息项目因素分析探讨统计计算能力在决策能力中的作用	40
1 问题提出	40
2 研究方法	42
2.1 研究被试及其研究程序	42
2.2 测量工具	42

2.3 数据分析	43
3 研究结果	45
3.1 验证性因素分析 (CFA) 结果	45
3.2 A-DMC 的全息项目 Bifactor 因素分析	46
3.3 决策能力公共因子与决策能力元认知的关系	48
3.4 不同决策技能之间的关系	49
3.4 统计计算能力与决策能力的关系	49
4 讨论	50
4.1 决策能力测量的双因子模型结构	50
4.2 公共因子掩盖不同决策技能的潜在关系	51
4.3 统计计算能力是规范决策中的核心因素之一	52
第三部分 统计计算能力对决策任务中数字信息把控的认知优势研究	54
1 问题的提出	54
2 方法	56
2.1 被试	56
2.2 实验材料	56
2.3 实验设计和程序	56
2.4 统计计算能力测量	57
2.5 数据处理	57
3 结果	57
3.1 不同统计计算能力对百分比变化的权重	57
3.2 不同统计计算能力对核心数字变化把握能力比较	58
4 讨论	60
4.1 高统计计算能力被试的数字信息权重线性加工优势	61
4.2 不同统计计算能力被试对核心数字把握能力	61
第四部分 统计计算能力对决策任务中启发性信息加工的认知优势	63
1 问题提出	63
2 方法	64
2.1 被试	64

2.2 实验材料	64
2.3 实验设计	66
2.4 统计计算能力测量	66
2.5 数据处理	66
3 结果	66
3.1 不同统计计算能力对基础概率的把握	66
3.2 不同统计计算能力独立事件统计规则把握的差异	68
3.3 不同统计计算能力条件概率启发式表征的差异	68
3.4 不同统计计算能力在信念偏差任务中的差异	69
4 讨论	71
4.1 高统计计算能力对基础概率信息的较好把握	71
4.2 高统计计算能力对独立事件统计规则的把握能力	72
4.3 高统计计算能力对条件概率统计规则的把握能力	72
4.4 高统计计算能力对系统 1 的控制和驾驭能力	73
第五部分 统计计算能力决策认知优势的心理机制	74
1 问题提出	74
2 方法	76
2.1 被试	76
2.2 实验材料	76
2.3 数据处理	77
3 实验结果	78
3.1 预测因子的调节效应分析	78
3.2 预测因子与启发与偏差任务的方差分析	79
4 讨论	82
4.1 高统计计算能力的认知优势依赖于高思维特质和高一般能力	82
4.2 思维特质担当统计计算能力的教练员角色	83
4.3 一般能力是统计计算能力和思维特质认知优势的保障条件	84
第六部分 模拟现实决策任务中对统计计算能力认知优势的检验	85
1 问题提出	85

2 方法	86
2.1 被试	86
2.2 实验材料及程序	86
2.3 数据分析	87
3 结果	87
4 讨论	88
4.1 高统计计算能力被试概率评估能力相对较好	88
4.2 高统计计算能力被试风险调节能力强	88
4.3 高统计计算能力冲动指数低	89
小 结	90
参考文献	94
附 录	104
个人简历和研究成果	110
致 谢	112

缩略语表

缩略词	英文全称	中文全称
BNT	Berlin Numeracy Test	柏林数学能力测验
CRT	Cognitive Reflection Test	认知反思测验
HB	Heuristics and Biases	启发与偏差
CTT	Classical Test Theory	经典测验理论
IRT	Item Response Theory	项目反应理论
TIF	Test Information Function	测验信息函数
CGT	Cambridge Gambling Task	剑桥博弈任务
TD	Thinking Dispositions	思维特质
A-DMC	Adult Decision-Making Competence	成人决策能力问卷
CFA	Confirmatory Factor Analysis	验证性因素分析
SEM	Structure Equation Modeling	结构方程模型
SPM	Raven's Standard Progressive Matrices	标准瑞文智力测验
FIFA	Full-Information item Factor Analysis	全息项目因素分析
DPT	Dual Process Theories	双加工理论

基于全息 Bifactor 模型探讨统计计算能力在决策中的 认知优势及其机制

博士研究生：张家喜

导 师：苗丹民 教授

辅 导 教 师：肖 玮 副教授

第四军医大学医学心理学系，西安 710032

资助基金项目：全军医学科技“十二五”重大项目（军人群体信息损伤预警与防护，
批准书号 AWS13J003）

中文摘要

研究背景与目的：决策信息通常包含概率信息，概率信息的处理和加工是决策过程中基本的认知过程。对概率的精确估计是良好决策的重要基础。统计计算能力作为运用概率和统计运算解决问题的能力，是高级判断和决策的最佳预测指标。现有研究对统计计算能力在决策中的认知优势和机制缺乏系统研究。其中一个原因是还未能很好的解决统计计算能力个体差异如何测量及如何准确测量的问题，对非理性行为在决策认知机制中的作用也缺乏系统性研究。为此，本研究首先探讨统计计算能力多维项目反应理论（MIRT）测量方法，以此为基础探讨统计计算能力在决策能力中的作用，进一步采用启发与偏差任务研究统计计算能力在决策中的认知优势和机制。最后采用现实情景任务对统计计算能力的认知优势进行验证。

研究方法：（1）应用全息 Bifactor 模型探讨统计计算能力的测量，采用 Berlin 数学能力测验（BNT）和认知反思测验（CRT）为主要测量工具，提取两者的公共因子，并以被试在公共因子上的得分测量被试的统计计算能力，应用测验信息函数、标准瑞文推理测验、数字表征任务和 Lipkus 数学能力测验等工具或任务验证信度和

效度；（2）采用全息 Bifactor 模型剖析决策能力的潜在结构，以此为基础探讨决策能力公共因子及其特殊因子与统计计算能力之间的关系，明确统计计算能力在决策中的作用；（3）设计数字信息权重把握任务、核心数字把握任务结合启发与偏差任务探讨不同统计计算能力个体在数字信息权重规律、核心数字把握能力和避免发生各类启发式偏差中的差异，系统研究统计计算能力的认知优势；（4）应用 Stanovich 分析思维加工理论，结合启发与偏差任务探讨统计计算能力认知优势的发生机制；（5）自编剑桥博弈任务程序软件，对统计计算能力在现实决策情景中的认知优势进行验证。

研究结果：（1）Bifactor 模型对 BNT 和 CRT 联合测验反应数据拟合指标最佳，提示 Bifactor 模型是最优拟合模型。BNT 和 CRT 联合测验公共因子在 IRT 中的信度系数理想，测验信息函数在均数上下测验信息量最大，提示公共因子测量统计计算能力具有较好的信度。同时该公共因子可以较好的区分被试数字精确表征能力，与瑞文推理测验、数学能力测验相关较高，提示测量内容是统计计算能力的核心特征，说明效度也较好；（2）成人决策能力测验（A-DMC）符合 Bifactor 模型结构，其公共因子（决策能力元认知）与统计计算能力的相关大于特殊因子与统计计算能力的相关，提示统计计算能力是决策能力的核心因素之一；（3）高统计计算能力被试对百分比的权重更符合线性加工规律，对核心数字在不同表面数字信息中有比较一致的把握，揭示高统计计算能力被试对数字信息权重线性加工和对核心数字把握的认知优势；（4）高统计计算能力被试在忽视基础概率、博弈谬误、协方差判断和信念偏差任务中，能更好的避免上述任务中的启发式偏差，提示高统计计算能力被试对基础概率信息进行准确表征、统计规则把握较好和对自动加工系统的监控和驾驭能力较好等认知优势；（5）思维特质、一般能力在统计计算能力和规范决策关系中的调节效应显著。方差分析进一步表明低一般能力时，思维特质对统计计算能力在规范决策中作用的调节效应不显著，高一般能力时，思维特质强化高统计计算能力的认知优势，低思维特质限制统计计算能力的认知优势，初步揭示统计计算能力在决策中的认知优势发生机制；（6）不同统计计算能力被试在剑桥博弈任务各类指标中，高统计计算能力被试决策质量和风险调节能力显著高于低统计计算能力被试，低统计计算能力被试冲动指数更高，统计计算能力的认知优势在现实情景任务中得到初步验证。

研究结论: (1) 全息 Bifactor 模型拟合的 BNT 和 CRT 测验公共因子测量统计计算能力具有较好的信度和效度; (2) 决策能力测量应是双维的, 一是决策能力元认知, 另一个是特殊决策技能; (3) 统计计算能力是决策能力的核心因素之一; (4) 高统计计算能力被试在决策认知过程中有数字信息权重线性加工优势, 核心数字信息把握能力、基础概率信息准确表征能力、统计规则把握能力、自动加工过程的监控能力较好等认知优势; (5) 统计计算能力在决策中的认知优势存在以下机制: 高一般能力是高统计计算能力和思维特质认知优势的保障条件、高思维特质担当监控分析思维过程、决定是否调用统计计算能力分析加工过程的教练员角色。

关键词: 统计计算能力, Bifactor 模型, 双加工理论, 全息因素分析, 决策能力, 启发与偏差任务, 项目反应理论

The Cognitive Superiority and Mechanism of Statistical Numeracy in Decision Making: Evidence from Full Information Item Bifactor Analysis

Candidate for doctor: Zhang Jiayi

Supervisor: Miao Danmin

Tutor: Xiao Wei

Faculty of Medical Psychology, Fourth Military Medical University,

Xi'an 710032, China

Sponsored Programs: the Major Project of Medicine Science and Technology of PLA

(Grant No. AWS13J003)

Abstract

Background and Aims: Numbers are important to people's decision making. The managing and processing of number information is the primary cognitive process of decision making. The precise estimation to probability is an important basis of good decision-making. Statistical numeracy is regard as the capability of probability and statistical operation which is the best predictive index of superior judgment and decision making. Recent studies were deficient in discussing the cognitive superiority and mechanism of statistical numeracy in decision making. One of the reasons was no proper resolution to the measurement and precision measurement of individual difference of the statistical numeracy had been found yet. On the other hand, researchers pay little attention on irrational behavior of decision making to explore the cognitive mechanisms of statistical numeracy. Therefore, this study firstly described the Multidimensional Item Response Theory (MIRT) method to measure the statistical numeracy, and then explored the effect of statistical numeracy in decision making, next studied the cognitive superiority and mechanism of statistical numeracy in decision making, and finally confirmed the

cognitive superiority of statistical numeracy with practical situational task.

Methods: 1. Apply the full information item analysis bi-factor model to explore the measurement of statistical numeracy. The general factor of Berlin Numeracy Testing and Cognition Reflective Testing was extracted by Bifactor Model. The statistical numeracy of participants were measured with their scores on the general factor. The reliability and validity of above measurement method was confirmed by its Test Information Function, the relationship between Raven's Standard Progressive Matrices, numerical representation tasks and Lipkus numeracy testing and the statistical numeracy scores. 2. The latent structure of decision making was analyzed by Bifactor model. Then the latent structure was used as the basis to discuss the relation between competence of decision making and statistical numeracy and the effect of statistical numeracy in decision making. 3. For the purpose of study the cognitive superiority of statistical numeracy, the number weighting controlling task, core number controlling task and Heuristics and Biases Tasks was design to analyze the difference of number information weighting regulation, core number controlling capability, and the ability to avoid various heuristic biases in different statistical numeracy. 4. The heuristics and biases tasks were adopted to analyze the mechanism of cognitive superiority of statistical numeracy on the base of Stanovich analysis process theory. 5. The Program of Cambridge gambling task was designed to confirm the cognitive superiority of statistical numeracy in practical decision making.

Results: 1. Bifactor model fit most to the response data of joint BNT and CRT test which indicate that Bifactor model is the best fitting model. The index of reliability for general factor scores is good with testing information is maximum around mean theta, which showed good reliability of the statistical numeracy in Bifactor model. The general factor of joint test could successfully distinguish the ability of precise numeracy representation and the correlation between Raven's Standard Progressive Matrices tasks, numeracy testing and statistical numeracy, which further stated that this method can measure the core trait of statistical numeracy and showed good validity. 2. A-DMC consists with Bifactor model. And the correlation between general factor—metacognition of decision making—and statistical numeracy is higher than the correlation for special factors, which indicated

that statistical numeracy is one of core factor of decision making. 3. Participants with high statistical numeracy were more consistent with the linearity processing on percentage weighting and they were also more capable of catching the core numbers in different surface number information, which revealed the number information weighting linearity of statistical numeracy and the cognitive superiority of core numbers catching. 4. Participants with high statistical numeracy have in better performance on Heuristics and Biases Tasks such as Base Rate Neglect Problems, Gambler's Fallacy, Covariation Judgment Problems and Belief Bias tasks, which showed their ability to avoid heuristic bias, further indicated cognitive superiority of statistical numeracy in decision making such as the precise characterization of basic probability information, the ability to seize statistical regulation, the ability of monitoring and controlling ability for auto operation. 5. The moderation effect of thinking disposition and general ability on the relationship between statistical numeracy and normal decision making were remarkable. ANOVA further proclaimed that when the general ability was low, the moderation effect of thinking disposition on the relationship between statistical numeracy and normal decision making was nonsignificant, whereas when the general ability was high, thinking disposition reinforced the cognitive superiority of high statistical numeracy in decision making. 6. Among various kinds of indexes in Cambridge gambling task, participants with high statistical numeracy showed higher decision quality and risk adjusting ability than participants with low statistical numeracy in decision making. However, participants with low statistical numeracy were higher in impulsion index. All of this confirmed the cognitive superiority of statistical numeracy in practical situational tasks.

Conclusion: 1. Adopting the full information bifactor model and extracting the general factor of BNT and CRT questionnaire to measure statistical numeracy had good reliability and validity. 2. The measurement of decision making is bi-dimension. One is the metacognition of decision making, the other is special decision making ability. 3. Statistical numeracy is one of the core factors of decision making. 4. The statistical numeracy has following cognitive superiorities in decision making, linearity process of number information weighting, the ability to seize the core numbers, precise

characterization of basic probability information, the ability to master statistical regulations and better monitoring ability in auto operation. 5. The following mechanisms of cognitive superiorities in decision making exists, high general ability emerge as the precondition to cognitive superiorities of high statistical numeracy and thinking disposition, in another word, high general ability is the guarantee. On the other hand, high thinking disposition plays the role of coach in decision making which monitor and control the analytical processing, determine whether to switch on process of statistical analysis or not .

Key words: Statistic Numeracy; Bifactor Model; Dual Process Theories; Full-Information item Factor Analysis; Decision-Making Competence; Heuristics and Biases; Item Response Theory

前言

概率信息和我们各方面的决策密切相关，小到我们的衣食住行，大到一场战争的军事决策，都包含各类概率信息。著名战争理论学家克劳塞维茨指出，战争是充满不确定性的领域，战争所依据的情况有 3 / 4 隐藏在云雾里一样，存在或多或少的不确定性。概率信息的处理和加工在军事决策中的作用可见一斑。从军事决策的要素来看，不论是量化作战条件、分析作战能力还是作战决心方案的制定、目标损伤效果的评估都存在一定程度的不确定性，都和概率信息的处理和加工密切相关。而现代信息化战争更是充斥着海量冗杂信息，给原本不确定的战争再蒙上一层迷雾，需要军事指挥员具有更好的信息处理加工能力。这也是本研究最大的军事意义所在。

从决策的认知过程来看，决策主要包含成本-效益的风险评估和可能结果的概率评估，对应风险调节能力和概率评估能力，都跟概率信息加工密切相关。西蒙进一步指出，决策信息存在很多的不确定性，包括信息的不完备性，事务关系的易变性，人们价值偏好的多样性等特点，因而，决策信息加工离不开概率信息加工，对概率信息的处理和加工能力成为个体是否能做出合理决策的重要基础^[1]。人们理解和运用数字信息能力的个体差异是客观存在的，即便是受过良好教育的人在决策时对数字信息的理解和把握能力也不见得特别强^[2]。而决策问题中的统计和概率信息更是让许多人望而生畏，事实上，对事件概率的精确估计是良好决策的重要基础，可见对决策问题中统计和概率信息的处理和加工能力是决策认知加工过程中基本要求。因而依据有限决策信息做出判断和决策更需要一定统计和概率推理能力作为支撑^[3]。

Fischhoff 曾批判性地指出决策研究存在孤立地研究决策的认知加工过程的缺陷^[4]。虽然这样能对决策的认知过程有深度认识，但忽略了个体决策技能与其他认知能力或个体变量的关系。影响决策个体差异的变量还包括个体认知风格、认知能力、人格特征、年龄等诸多变量^[5, 6]。近年来，研究者提出与数学有关的认知能力是影响决策的核心因素^[7]，这些认知能力也越来越受到研究者的重视。推动研究从个体差异的视角研究认知能力在决策中的具体作用和机制。统计计算能力作为运用概率和统计运算解决问题的能力，是高级判断和决策的最佳预测指标（不论是在包含数字的

任务还是不包含数字的任务中)。为此,基于个体差异的视角以统计计算能力为主要预测变量研究概率信息的处理和加工能力在决策中的具体认知优势和机制是可行之道。但是有关统计计算能力在决策中作用的研究还有诸多问题未能得到很好的解决。

一是统计计算能力个体差异如何测量和如何准确测量的问题。传统上,决策研究中对数学能力的测量大多基于经典测验理论(CTT),但是CTT存在真分数与观测分数之间不符合线性关系、误差与真分数独立的假设难以满足等现实。决策研究者试图寄希望于现代测量理论,如项目反应理论(IRT),并基于IRT开发了一些专门测量统计计算能力的测验。即便如此,研究者对测量内容和维度、模型选择都存在不同的认识。如有研究者将数学能力等同于统计计算能力,采用数学能力测验测量统计计算能力。有研究试图应用单维IRT Rasch模型测量统计计算能力,但是在测量题目的选择上却是两类不同质的题目,如采用相对难的认知反思测验、相对简单的数学测验。而事实上,统计计算能力是一个多维度的概念,简单应用单维IRT模型恐怕难以取得更好的效果。为此,需要探寻适应测量统计计算能力多维特点的IRT方法。

二是对非理性行为在决策认知机制中的作用缺乏系统性研究。建立在“经济人”假说之上的完全理性决策理论模型不断受到非理性行为表现的挑战,决策研究也从绝对理性走向有限理性,逐渐关注人们的非理性行为对认识决策机制的作用。人类在决策行为中的个体差异也是客观存在的,也就是说所并不是所有人、或某人在所有情况下都是非理性或是理性的。西蒙为证实问题解决过程中的有限理性,首先发现了人类的认知启发式加工特征。Amos Tversky 和 Daniel Kahneman 进一步证实人们在决策时会走思维捷径^[8],虽然启发式思维是一种捷径,会帮助人们快速地做出判断,但也容易导致各种启发式偏差。Kahneman 进一步指出,启发式表征偏差也不是随机的,而是系统性的,与人们的认知能力和认知风格有很大关系。因而,也为研究不同认知能力个体的认知优势打开了窗口^[8]。为此,基于启发式表征研究统计计算能力的认知优势也是本研究的基本出发点。现有关于统计计算能力在决策中认知优势的研究所采用的决策任务各异,以发现不同组间的差异为目的采用不同的决策情景。如有采用框架描述研究不同统计计算能力组间差异的,也有采用贝叶斯推理任务比较统计计算能力认知优势的。缺乏设计系列非理性决策任务系统研究某一认知特质在决策中的作用和机制。为此,本研究试图以双加工理论为基础,设计相

关决策任务，从分析思维加工过程和自动加工过程两个方面系统考察统计计算能力的认知优势。

三是对统计计算能力认知优势的发挥机制缺乏足够的研究。现有研究中大多是应用决策相关认知理论解释不同统计计算能力的组间差异，如双加工理论、模糊痕迹理论等。但是应用现有理论开展验证性研究，探讨不同认知能力在决策中机制研究相对较少。Stanovich 等人依据双加工理论开展了一系列的研究^[9]，并结合认知反思测验提出思维特质对分析性思维的监控和管理功能，进一步深化双加工理论，提高了认知能力在决策加工过程中的解释力，同时对研究特殊认知能力在决策加工的机制方面具有启示作用。出于现有研究对统计计算能力认知优势的机制还没有足够的认识，本研究拟采用 Stanovich 等人的分析思维加工理论，探讨统计计算能力认知优势在决策加工过程的认知机制。

针对上述研究现实，本研究从探讨统计计算能力个体差异的测量方法出发，进一步探讨统计计算能力在决策能力中的地位和作用，设计系列决策任务探讨统计计算能力在决策中的认知优势，同时依据 Stanovich 的分析思维加工理论探讨统计计算能力在决策中认知优势发挥机制。最后，采用现实情景任务验证统计计算能力的认知优势。以期弥补现有研究的不足。

文献回顾

本文重点回顾了统计计算能力的概念及其在决策中的作用，对与统计计算能力密切相关的数学能力等概念及其对决策的作用也做了介绍。同时重点回顾了统计计算能力的主要测量方法。考虑到决策认知理论对本研究的指导意义，这方面也有相关介绍。最后重点介绍了贯穿本文的主要测量方法——全息 Bifactor 模型及其在心理和教育测量中的应用情况。

1 统计计算能力及其在决策中的作用

研究者在探讨数学能力与决策关系的时候发现统计计算能力实际上是数学能力的一个子成分^[10]。相比于数学能力，统计计算能力对规范决策和风险感知能力具有同样的预测性。决策研究中，也有研究者将统计计算能力等同于数学能力，研究报告中也有将两者混用的现象，事实上数学能力相对于统计计算能力是一个外延更大的概念。为了更明确的阐述统计计算能力及其在决策中的作用，本部分首先介绍数学能力及其在决策中的作用。

1.1 数学能力

数学能力是指人们对数字及其相关概念进行表征、推理和转换等加工过程中所需要的一种能力^[11]。简单说来是指理解和运用数字的能力。数学能力是一个复杂的概念，包含多个功能性成分^[12]。研究者认为数学能力是一个多维度概念体系，并不局限于数学推理和数学技能的简单读写^[13]。最基本的数学能力包括对实数、时间、度量及数字估计的理解。较高水平的数学基本技能包括简单四则运算、数字大小的比较。更高一级水平的数学能力则包含基本的数学逻辑推理、多级运算，比率、分数、百分比等概念的理解和运用^[11]。从内容上上看，可以划分为数字表征能力、概率推理能力（统计计算能力）和算术能力等^[14]。从形式上看，Liberali 等人将数学能力划分为主观数学能力（subjective numeracy）和客观数学能力（objective numeracy）^[15]。主观数学能力是指个体对数学能力的自我认知及其对数字呈现形式的主观偏好。客观数学能力是指对数学概念、概率的理解和运用能力，包括在风险评估中应用百分比、比率等形式开展数学运算，及其不同比率形式的相互转化的能力^[14, 16]。

人们理解和运用数字信息能力的个体差异是客观存在的，即便是教育程度较好的被试在决策时对数字信息的理解和把握也不见得特别强^[15, 17]。最近研究发现，人们的数学能力是存在较大差异的，甚至有很多人在数学能力上是文盲。美国国民文化调查（National Adult Literacy Survey）发现^[17]，约一半的美国人缺乏数字信息运算的基本技能，即便是日常生活中的常用数学运算（比如正确计算购买日用品的找零）。而西班牙人、非裔美国人、65 岁以上的老人和低教育程度者在数学能力上的得分更低，女性比男性得分要低^[17, 18]。

1.2 统计计算能力

近年来，决策领域中的研究者倾向于将数学能力特指为数学能力中的某些成分，如统计计算能力（statistical numeracy）^[19]。作为数学能力的一个成分，统计计算能力与智力存在一定程度的相关，但又区别于智力因素。这在大脑皮层功能上也发现了相关证据，脑损伤病人研究发现，下顶叶区病变损害了患者的数字认知能力，但是对非数字的认知能力却没有影响^[20]。

研究者指出，数学能力中的统计成分是决策中风险评估的关键特征^[21]，在健康质量评价和医患交流中的风险判断都有主导作用^[22]。需要指出的是，风险（risk）也是一个广义概念，如概率分布中的方差、不确定性对事件结果的影响、损失概率、死亡率等^[23]。因而，统计计算能力是风险感知能力的一个较好的预测指标^[24]。经济 and 心理学理论一致认为决策包含了“统计概率（statistical probabilities）”计算和定量概率推理（quantitative probabilistic reasoning）等心理成分。可见，统计计算能力是风险判断的重要因素，如精确解读风险信息并基于风险信息选择恰当的行为方式。事实上，不论是在经济领域、健康领域还是消费行为领域^[18, 19]，研究证实统计计算能力是决策中统计策略选择、情感反应、标准化决策选择的一个重要预测指标^[24]。可见，统计计算能力是影响决策能力的一个非常重要的认知成分^[25]。统计计算能力的具体定义是指运用概率和统计运算解决问题的能力，包括对随机投币概念的理解和对百分比数字进行基本的数学运算、概率和比率的比较和转换^[10]。

1.3 统计计算能力在决策中作用

于决策而言，人们很难忽视数字信息而做出判断和决策。数字信息是抽象符号，理解起来有一定的难度，而背景信息对人们理解数字信息的性质和量值也会产生影

响。人们往往以主观的方式对数字信息的含义进行理解和判断。因而，对数字信息的接近性和理解力往往是做出规范决策的基本要求^[16]。研究同样发现，统计计算能力是高级判断和决策的最佳预测指标^[22]，不论是在数字任务还是非数字任务中。相比于流体智力、认知反思和注意控制，统计计算能力也是重要的预测指标^[1, 11, 12]。研究同样发现，统计计算能力在高级决策中的作用，不局限于抽象推理或是经典的规范决策。事实上，在启发式思考、情感数字直觉和直觉推理或是决策都有较好的预测性^[26]。概括起来，现有研究对统计计算能力在决策中作用有如下几个方面：

低统计计算能力的被试似乎更少的运用数字信息做出决策，对数字信息的理解程度也要低，而非数字信息、冗余信息在他们决策中的作用似乎更大。在一项包含 152 名被试的实验中^[27]，要求被试对医院的好坏做出评价，低统计计算能力者评价医院的服务质量时大多基于文字信息引发的个人情绪而非医院质量评价的一些数量化客观指标，而高统计计算能力者似乎较少的受个人情绪的影响，更多采用客观数量指标做出评价。在另一项研究中^[28]，如果被试在两个瓶子摸中一个彩球就能得到奖品，瓶子 A 中有 9 个彩球，91 个白球，瓶子 B 中 1 个彩球 9 个白球。显然，从瓶 B 中摸中彩球的概率要更大。尽管如此，33%的低统计计算能力者更愿意从瓶 A 中抽奖，同时只有 5%的高统计计算能力者愿意从瓶 A 中摸奖。在这个实验中，数字信息被赋予情感含义，直觉思维和理性分析思维相互竞争，而概率较低的选项具有更大的吸引力。但是低统计计算能力者对情感性信息赋予更高的权重，而放弃选择更理性的选项^[29, 30]。而高统计计算能力者对赢取奖品的概率信息有更清楚的认识，从而较少的发生概率偏差。由此可见，统计概率信息的采纳对规范决策的作用。Peter 等人同时发现，不同统计计算能力的被试采纳的信息可能是不同的^[5, 31]。尽管上述研究是基于双加工模型或理性-情感理论，但是该研究还不能有效的解释情感因素本身（或者说直觉加工）就是导致不同能力者决策差异的根源。值得肯定的是，这项研究中情绪和概率偏差表明低统计计算能力者，相比较于高能力者，更多的受外部表征线索的影响，而客观概率信息对其的影响似乎要小。

也有研究指出，不同统计计算能力的被试对数字表征的方式是不同的，这对个体决策也会产生一定的影响。从心理感受性的角度来看，人们对数字信息的感知也应该符合韦伯定律。研究发现，个体数量表征感知曲线对线性函数的偏离水平同样随着数学能力或统计计算能力的不同而有所差异。统计能力高的被试对数量感知表

征的曲线更接近于线性函数^[32]。这一数量感知的差异会影响到被试的决策。如 peter 等人研究发现^[11]，当要求被试在“立即获得 10 美元”还是“一个月后获得”15 美元”做出选择时，数字感知更符合线性的被试更偏好延迟获得 15 美元^[31]。Furlong 和 Opfer（2009）设计的囚徒困境实验中也发现线性模型对实验结果方差的解释高达 95%^[33]。由此可见，不同数字感知函数对人们决策行为的影响。高统计计算能力被试在这方面似乎有更大的优势。

不同统计计算能力对被试决策影响还体现在对风险的感知上。在现实情景决策中，低统计计算能力被试更容易高估风险，而高统计计算能力的被试则能更精确的对风险情况做出判断和估计。在一项要求被试根据一份虚拟情报估计某地发生恐怖袭击概率估计的实验中，低计算能力的被试对风险的感知要显著大于高计算能力的被试，事实上高计算能力被试对风险的估计要更准确^[34]。健康领域的研究也得到了类似的结果，低计算能力的被试倾向于高估乳腺癌的患病风险^[16]。此外，高统计计算能力个体更倾向于信赖数字性风险信息，而低统计计算能力的被试更倾向于信任文字性的风险信息，显然，文字性风险信息是粗略估计，而数字性风险信息更为精确^[24]。因而，他们对风险信息的感知程度是不一致的。

不同统计计算能力对决策的影响也体现在对决策偏差抵御上。Tversky 和 Kahneman 最早报道了不同认知能力的被试在启发式表征上的差异，最终导致系统错误和偏差^[8, 35]。对常见决策偏差的研究也发现，低统计计算能力的更容易出现框架效应、成本沉没。如 Peter 等人发现^[22]，高计算能力的被试更不容易受到信息框架的影响。Johnson 等人研究发现，低统计计算能力被试在概率推理任务中的决策结果容易出现前后矛盾，对同一问题采用不同信息呈现方式得到不同的决策结果^[23]。Cokely 和 Kelley 等人发现高统计计算能力的被试决策结果更符合高期望价值选择^[36]。可见，不同统计计算能力被试在避免决策偏差上也有个体差异。

统计计算能力还可能对个体决策自动加工过程中的启发式表征产生影响。Thomas 和 Morwitz（2008）研究发现^[37]，个体的算术和统计知识在决策时可以自动激活，从而影响消费者的决策判断。比如，在计算 7.99 元和 9.01 元的差别时，被试会自发的计算整数位的差值（9-7=2），从而错误的认为原价 9.01，折后价为 7.99 元的商品比原价是 9.02 折后价是 8.00 元的商品更优惠。Thomas 等人将这一谬误定义为左侧数字锚定效应，此外数字偏差中还有代表性启发和便利性启发^[38]。比如价格

标签的字体大小会对人们对价格高低感知和购买欲产生无意识的影响（字体较小时，相对于较大字体，人们对价格的估计要低），就是代表性启发。人们对整百的数字估计往往要比非整百的数字要大（如认为 38300 大于 38311）^[39]。高统计计算能力的被试能更好的避免上述启发式偏差，而低统计计算能力的被试发生此类偏差的比例要更大。

当然，统计计算能力在决策中的作用也存在一些争议，这包括统计计算能力在决策中是促进还是阻碍作用。认知能力高的个体是不是意味着决策能力更强，一直是研究者思考的问题。绝大多数研究支持数学能力在决策中的积极作用。同样引起研究者注意的是，高水平的被试中，相当比例的被试不能做出标准决策。更有研究发现，算术能力高的被试在某些决策任务上的成绩甚至不如低数学能力的被试。Peter 等人研究发现，高算术能力的被试更偏好“7/36 的概率赢得 9 美元，29/36 的概率 5 美分”，而不是期望收益更高的“7/36 的概率赢得 9 美元，否则不损失一分钱”^[5]。依据双加工理论，算术能力高的被试更偏好精细加工，但是同样也有可能面临像上述决策情景中带来一定的负面作用。在模拟现实决策情景中也发现同样的尴尬（如爱荷华博弈任务），统计计算能力对决策成绩并没有显著预测作用^[40]。可见，高统计计算能力的个体其决策并不必然优于低决策能力者，统计计算能力在决策中的认知优势可能是存在一定条件的。

1.4 统计计算能力与文化程度的关系

无从否认，人们的学历越高，其统计计算能力往往也会越高。但是，学历高并不意味着其统计计算能力就好。也有研究者认为，统计计算能力并不是完全后天习得的一种能力。如 Rothman 等人研究了 200 名受过良好教育的初级护理病人他们对食物标签的理解力，尽管 77% 的被试至少具备高中以上的读写能力，63% 的被试却达不到 9 年级学生应有的数学能力水平^[41]。在另一项包含 357 名大学生被试的研究中，超过半数的被试在 Schwartz 开发的数学能力测验中，3 题只做对了 1 题甚至 1 题都没有对^[15]。在不同学历的人群考察统计计算能力高低的时候也存在同样的尴尬。如有研究发现^[42]，临床医生、医学生等在计算相对风险估计中也存在不同程度的困难。Yamgishi 发现大学生在解释简单的死亡率数据时都存在困难，他们在比较“10000 人中死亡 1286 人死亡率高还是 100 人死亡 24.14 人的死亡率高”都有一定困难^[43]。

Lipkus 等人研究发现受过良好教育的人在回答比率、比例或百分比等问题上，即使是相对简单的问题也有麻烦，如仅有 20% 的被试能将“1000 中的 1 个”转换为百分数^[2]。Reyna 等人发现，医生和病人对理解基因风险都倾向于犯类似的错误，尽管二者的医学背景知识相距甚远^[1,15]。由此可见，统计计算能力高低不能与学历高低完全划上等号。

2 统计计算能力的测量

本节依据统计计算能力测量工具的理论基础将统计计算能力的测量划分为经典测量理论（CTT）方法和项目反应理论（IRT）方法。

2.1 经典测量理论（CTT）的方法

在经典测量理论框架下，统计计算能力的测量大致可以分为三类，间接测量法、主观评价法和客观测量法。间接测量法主要通过考察被试教育程度、学业成绩、认知能力或认知风格等与数学能力相关的心理变量对统计计算能力进行间接估计；主观评价法主要是应用临床和健康相关领域相关话题的问题情景，采用自我报告的方法对自身统计计算能力做自我评价^[16]，从而实现统计计算能力的测量。而最常用的方法是采用客观的心理测验对统计计算能力进行直接测量，通过设计具体的统计概率能力测验题，直接测量被试的统计计算能力的客观水平。虽然间接测量的方法在一定程度上能反应统计计算能力高低，研究者普遍认为间接测量所考察的因素较多，混杂因素也很难排除，因而一般不为研究者采用。本文重点介绍统计计算能力测量的主观评价法和客观评价法。需要指出的是，很多研究者对统计计算能力和数学能力没有严格区分，存在将数学能力测验和统计计算能力测验混用的局面，专门的统计计算能力测验也兴起于近几年。为此，本研究对统计计算能力的测量和数学能力测验一并介绍。

2.1.1 主观评价法

数学能力的主观测量最常见的是 Fagerlin 等人开发的主观数学能力量表（Subjective Numeracy Scale）^[16]。该量表采用自陈式报告法测量被试在完成数学任务时自我感知到的能力，同时考察被试对数字信息或文字信息的偏好程度^[16, 44]。该量表共 8 题，答案没有正确与错误之分，采用 Likert 6 点量表的方式进行评分。其中 4 题要求被试评估自己在不同情景中的数学能力，4 题要求被试在呈现数字或概率信

息时表达他们的喜好。Fagerlin 等人报告,该量表与 Lipkus 客观数学能力测验有较高的相关。同时,研究发现,该量表在数学能力测量方面有较好的信度和效度。在健康和决策领域都有较多的应用。

2.1.2 客观评价法

数学能力的客观评价方面,从现有的文献来看,主要是 Schwartz 等人开发的系类 Lipkus 数学能力测验和 Cokely 等人开发的 Berlin 数学能力测验,而 Berlin 数学能力测验也包含了 IRT 的版本。许多研究报告都应用上述量表测验被试的统计计算能力。

Lipkus 数学能力测验。Schwartz 与其同事于 1997 年首先开发一个包含 3 个题目的测验测量客观数学能力^[45]。此测验在最初设计时主要是用于低数学能力被试的考察。该测验主要考查数字信息的基本理解和运用能力,从测量内容上来看,主要包括将分数转换为百分比、估计抛硬币时出现正面或反面的概率,因而有研究者认为该量表考察的事实上就是统计计算能力。研究者进一步对其进行扩充和发展。Lipkus 等人于 2001 年在此基础上扩充了 8 题^[2],发展为 11 题的量表。该问卷自一开发就受到研究者的高度关注,其应用也比较广泛,在临床和决策领域都有较多的应用。可以说是数学能力测量方面的经典量表。其他许多研究者对原始量表信效度展开了广泛研究,都得到比较一致的结论。从现有文献来看,大多数研究者采用该量表测量数学能力或统计计算能力,研究数学能力在决策中的作用(如风险决策、框架效应等)。Peter 等人于 2007 年对该量表进一步进行扩充,添加了基础概率问题和复杂比率计算的题目,报告了 15 题的量表,有选择题和填空题两种形式^[22]。该版本与前两个版本相比,难度有所增加,同样具有理想的信度和效度。Weller 等人应用 Rash 模型的方法,对该量表进一步进行简化,报道了 8 题的版本同样具有 15 题版的信度和效度^[46]。

2.2 项目反应理论 (Item Response Theory, IRT) 的方法

经历了数十年采用 CTT 方法测量统计计算能力,决策科学的研究者开始将注意力转向现代测量理论,如项目反应理论 (IRT)。与 CTT 不同,IRT 需要应用概率分布模型拟合测验中的每一个条目的被试作答。因而测验开发的重点是每一个条目的项目参数。当然,基于 IRT 的测验能更好的估计被试的统计计算能力,减少了条目

的冗余。基于 IRT 的数学能力测验为数不多，而基于 IRT 的统计计算能力测验更是少之又少。如 Schapira 等人(2012)开发了基于 IRT 医学数学能力理解测验(Numeracy Understanding in Medicine Instrument, NUMi)^[18]，该测验共 20 题，参数估计采用两参数模型，包含四个方面数学能力，如图表理解、统计计算能力等。研究发现该测验在评估低教育程度病人的基本健康数学能力方面具有较好的灵敏度和特异度，对病人医学决策也有较好的预测效度^[47]。Weller 等人(2013)年基于 Rash 模型开发了包含 8 题统计计算能力测验^[46]，该测验题目的主要来源是现有的数学能力测验和认知反思测验，该测验在美国大众群体中参数估计稳健，同时还发现该量表在风险决策方面也有较好的预测效度。目前使用最广泛的是 Berlin 统计计算能力测验，下面对这一测验重点介绍。

Berlin 统计计算能力测验。也有研究者将其应用与数学能力的测验。该测验是 Cokely 等人于 2012 年开发^[21]，能快速实现统计计算能力和风险感知能力的测量，能广泛适应于不同能力等级的被试测量。研究发现，相比于其他认知测验或数学能力测验，该量表是日常生活的风险感知有效的预测工具。也有研究报告该量表是其他数学能力或认知测验预测效度的两倍，可见其独特的解释力。该测验也有多种测量方式，既有纸笔测验、也有计算机自适应测验，有选择题版本也有选择题版本，更有单题版，每一个版本的题数都不一样。其中纸笔测验版本是 4 题，被试 5 分钟左右能完成，计算机自适应版本依据被试的作答情况 2-3 题不等，而单题版主要用于均数划界。该测验自问世以来，就翻译为多种语言，在许多不同国家都采集了相关效度数据，为多国决策研究者广泛使用。

3 决策的认知理论

近年来，有关统计计算能力和决策的关系研究也取得一些显著的成果，但是依然缺乏具体理论的指导。现存的有关决策认知理论框架可以帮助理解统计计算能力在决策中的认知机制，也有研究者在做这方面的尝试。可以解释数学能力或者说统计计算能力在决策中作用的相关理论大致有 4 种。

3.1 信息加工理论

该理论强调准确、精密的分析思维在决策中的作用，强调为实现准确判断所做出的加工努力^[48]。该理论认为，工作记忆的局限性影响数字加工。因而，减少工作

记忆负荷可以有效提高数字加工的绩效。由此可以推断，增加决策问题的易懂性和可接近性，减少决策过程中的认知努力，比如减少无关信息的干扰，可以避免决策偏差。同样，信息加工的精确性必然对决策绩效有显著影响。与该理论比较一致的研究结果是，要求被试列举偏好某个决定的原因或绘制各个备选条件的风险概率条形图，增强其数字信息加工的主动性，可以提高数字信息的理解和运用的效率，减少无关信息的干扰，也可以提高决策质量。也有研究者发现，要求决策者陈述选择的理由，可以增加数字信息的权重，同时降低无关、非数字信息（如情感性信息）的权重。

3.2 进化理论

该理论强调自然加工，如频数加工，认为频率信息更容易得到加工^[49]。进化理论认为有两种量化信息加工，一种是伴随进化而产生的频率信息加工，即初级的数量上的区分；另一种是正式的数学知识（如概率知识）加工。研究者比较了儿童和成人在数字加工方面的差异，但是这方面的研究并没有得出数学能力在这两方面加工差异的明确结论。支持该理论证据是人们在加工频率信息的时候比概率信息更为容易，研究初步发现频率信息和概率信息对于决策问题加工效率有一定的差异^[22]。

3.3 双加工理论

该理论区分了两个相互独立又相互影响的加工系统，自动加工（又称系统 1）和分析性加工（又称系统 2）^[50, 51]。自动加工过程与任务或是问题情景有关，不需要意识过程参与自动激活记忆或相关资源。而分析思维过程需要意识参与，需要一定的认知负荷，同时负责监控自动加工过程的认知表征并判断自动加工过程能否胜任当前的任务，判断是否需要驾驭自动加工过程做进一步的任务分析，是否需要启动更多认知资源和高级认知过程（如决策能力）以更好的完成当前的任务。当分析性思维占主导时，问题解决主要由分析性思维完成认知策略选择、认知操作或更复杂的推理和运算^[52]。系统 1 加工具有快速、自动化的特点。而系统 2 的加工速度慢、需要意志努力，依赖于规则加工。双加工理论认为系统 1 和系统 2 在决策过程中都起作用，但是两者的地位和作用不同。人们在面对决策情景时，直觉加工即时做出默认反应，而分析性加工则会基于一定的规则做出推理和判断，同时对直觉加工进行干预。最终的决策结果更多的由分析加工决定，但最终的决策并非总是服从于分析

加工。如果直觉加工占优势，则以直觉加工为主体；若分析加工占优势，则以分析加工的结果作为决策的主要依据。认知反思（cognitive reflection）的相关文献证实了两个加工系统的存在。Epstein 等人设计一系列有关直觉思维和分析思维的研究为该理论积累了很好的实验证据^[53]。双加工理论在冲突模式的推理和决策中也有较好的解释力。

该理论认为直觉思维是数字加工过程中偏差和错误的本源，错误的发生往往是直觉加工的结果。而分析思维是对数字进行精确、客观加工的基础。当然，也有研究者指出^[54]，直觉思维也并不是一定就会发生错误和偏差。情绪或情感效价也认为是直觉思维的一个方面。

3.4 模糊痕迹理论

该理论也是为填补其他理论在解释诸如工作记忆与推理成绩没有显著关系时的缺陷而应运而生^[55, 56]。模糊痕迹理论在双加工理论的基础上进一步区分了整体直觉加工和逐字分析性加工，同时强调整体直觉加工是一种更高级的加工过程。该理论进一步认为个体在表征信息时存在两种不同的方法^[57]，一是字面信息的表征，另一种是实质含义的表征，该表征与个体经验、知识等相关，但是没有字面表征具体，是头脑中的“模糊痕迹（fuzzy trace）”又称之为整体表征（gist representation）。对数字信息的表征，同样存在以上两种不同的表征方式。比如三种不同价格的葡萄酒 A、B、C，分别是 98 元、198 元和 298 元。个体对上述信息的整体表征为“C 葡萄酒的价格最高”。研究同时发现，人们在决策时，更偏爱模糊痕迹表征，而不是具体细节表征。该理论认为，决策偏差的产生也正是因为个体采用整体表征，忽视了具体细节信息。人们出现框架效应也是因为人们在决策时是基于总体上的整体加工^[58]。当被试面临“小明在考试时的正确率是 74%”这样的陈述，要求判断小明的成绩如何时。被试的整体表征可能是“小明在考试中回答正确的可能性较大”。而当面临“小明在考试时的错误率是 26%”的陈述时，整体表征则为“小明在考试中会发生错误”，因而对前一陈述的估计更为乐观，而对后者的估计则较为悲观，从而出现框架效应。该理论指出，整体加工是伴随人们知识和经验的增多而更多的采用的一种加工方式。研究表明，相比较于知识和经验相对匮乏的儿童，成人更多依赖整体信息做出决策。这种加工方式，对于成人来说，可以避免信息过载减少认知负荷，从而应对字面信

息表征能力下降的现实。因而，框架效应从儿童到成人有增加的趋势。对于启发与偏差任务而言，也有同样的发展趋势。研究发现^[59]，成人专家更多的是基于整体概况做出决策，而新手更多采用局部加工的方式。模糊痕迹理论对统计计算能力的在决策认知加工过程中的机制具有较好的解释力。比如，概率偏差问题（人们倾向于选择概率较少但是分子较大的选项）就是一个典型的类包含错觉，验证了该理论回溯加工假设和表征假设。任何比率（包括概率），都涉及分组重合，因而容易引发误解。如果人们只关注分子中的数字（如 A 盘中 9 个彩球）而忽视分母（A 盘共有 100 个球）的权重，就容易产生概率偏差现象^[5]。正如提醒被试不同框架条件下的非等价性能减少框架效应一样，提醒被试关注分母的意义也能减少概率偏差。类包含错觉在文化素质高的人群中同样普遍存在^[60]。比如，外科医生和高中生在忽视基础概率（另一种类包含错觉）情景中的成绩没有太大差异。在这项研究中，被试在告知基础概率的条件下，要求被试判断检查为阳性时病人发生某疾病的概率^[61]。概而言之，虽然模糊痕迹理论同样是一种双加工模型，整体加工类似于双加工系统中的直觉加工，字面表征则类似于需要更多认知负荷的分析加工^[9]。双加工模型将分析思维等同于精确加工，认为直觉加工是一种原始的、跟进化有关的加工系统，受到分析思维的调控。而模糊痕迹理论认为知识和经验更丰富的决策者在决策时更多的依赖简单、定性的整体信息，是一种相对高级、进化后的加工，虽然同样不可避免的发生类包含错觉^[62]。事实上，枚举选择的理由并提醒被试关注数字信息的细节以对抗整体加工的影响并不能显著提高推理和决策任务中的成绩。对知识和经验丰富的决策者而言，依靠整体加工反而能在决策任务中获益。因为，研究发现，与年轻人相比，在运用整体加工时，年轻人的认知加工优势减弱，从而使老年人和年轻人在决策任务中的成绩差异减少^[63, 64]。模糊痕迹理论为解释决策过程中的个体差异提供了一种新视角，但是和双加工理论的争议也不少，有待于进一步深入探讨。近年来，研究者试图将情感研究的结论和模糊痕迹理论的原理整合，将情感因素整合为一种整体加工，对双加工理论的相关指标和模糊痕迹理论的相关结论做比较研究^[65]。

4 全息项目 Bifactor 分析模型及其应用

4.1 Bifactor 模型简介

双因子结构早在 70 年前就有人提出，但是直到今天，测量学家才重新发现 Bifactor 模型在多维结构相关的模型拟合中的重要价值^[47, 66, 67]。

心理结构通常有多层级或多层面的结构。也就是说同一心理特质在不同的行为中有不同的表现，一种行为可能由多种心理特质共同支配^[68]。就目前而言，心理特质只能通过对行为的间接测量来反应该心理品质。因而在测量该心理特质的时候需要测量不同的行为反应，以便有效的反应该心理特质的全部内涵^[69]。所以，在测量心理特质的时候通常需要采用多个子维度的行为反应来测量。以生活满意度的测量为例，研究者需要从多个维度测量生活满意度，包括工作满意度、业余时间的满意度、家庭满意度、健康、收入等社会生活各个方面^[70]。尽管该量表的设计初衷是测量单一心理变量——生活满意度，这种两阶段的因子抽样程序（如，先构建生活质量包含的维度，再确定维度需要哪些条目）经常使量表具有多维度的特征^[71, 72]。在教育测量中，这一情况也非常普遍。比如，英语阅读理解测验，通常包含好几个段落，每一个段落又包含数个问题。每一个段落使用的是不同的阅读内容。这样，在实际的测试过程中，受试者比较熟悉的内容段落他答对的可能性将会更大。因而，段落内条目之间的相关性更高。其结果是不同段落之间的题目不再具有局部独立性，违反局部独立性就排除了采用单维模型对项目反应进行分析，这与单维模型的假设是不一致的。可见，在心理量表的构建时，至少包含两个方面的方差，其一是反应该心理品质的公共方差（每一个题目都会有的方差），另一个是反应子维度的方差。从这个角度来看，心理量表很难有严格维度上的单维或多维结构^[67, 73]，更多的是一种双因子结构。目前心理测量中因素分析的模型包括相关因素模型、二阶模型和 Bifactor 模型^[71]。从现有模型来看，Bifactor 模型能较好的反应不同维度之间的相关性。因为公共因素和特殊因素之间的正交性，只有 Bifactor 模型能较好的分离这两类方差^[74]。从另一个角度来看，如果能将这两类方差进行有效分离，对理解心理特质的结构、评价和改善现有心理测量都有非常重要的价值^[75]。

Bifactor 模型早在 1938 年由 Hozinger 等人提出^[76, 77]。又称一般-特殊因子模型、嵌套模型，是拟合多层面结构的第三种重要的方法。该模型最先应用于智力的相关

研究, 最近逐渐应用于人格和健康测量的研究^[78, 79]。Bifactor 模型假设 1) 所有层面(维度)有一个公共因子解释他们的共同变异; 2) 除公共因子外, 还有多个特殊因子, 这些特殊因子解释特殊成分所造成的方差^[70, 73]。假设测验包含 s 个因素(1 个公共因子, $s-1$ 特殊因子), 共 n 个测验题目。Bifactor 模型中, 对题目 j 来说, 在公共因子上有一个非零负荷 α_{j1} , 不超过 $s-1$ 个特殊因子负荷 ($\alpha_{jk}, k=2, \dots, s$)。对一个包含 4 个题目 3 个因子的测验来说, 其 Bifactor 模式矩阵如下:

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & 0 & \alpha_{33} \\ \alpha_{41} & 0 & \alpha_{43} \end{bmatrix}$$

Bifactor 结构模型认为条目反应的协方差可以分解为公共因子和特殊因子。其中, 公共因子反应所有条目的共同变异, 特殊因子反应具有典型、高度相似特征的一组条目的额外共同变异。Bifactor 模型能将似然估计降为二维(不管量表的初始维度是多少), Bifactor 模型假设公共因子和所有的特殊因子及特殊因子之间是非正交关系, 即彼此不存在相关性^[80]。公共因素代表量表需要测量的概念上更为宽泛的目标内容结构, 特殊因子代表概念上更狭小的子结构。因而, Bifactor 模型在分析使用结构内容相关的多个不同子维度测量同一心理品质的量表中非常适合。研究者首先将其应用于因素分析或验证性因素分析^[74]。Gibbons 等开发了二级计分和多级计分的全息 Bifactor 模型^[81], 将其应用范围进一步扩大。在 Gibbons 等人开发的 Bifactor 极大边际似然参数估计及其实证研究中发现, 该方法能对 Bifactor 模型似然等式做较大程度的简化: 1) 能对较多的特殊因子进行模型拟合和分析; 2) 能满足特殊因子不同条目之间局部独立性假设; 3) 相比于传统的非限制性全息项目因素分析, 能提供更为简洁的因素分析方法^[82]。

4.2 Bifactor 模型的优点

Bifactor 模型在数据分析和模型拟合方面具有明显的优势。首先是可以同时分析公共因子(公共潜在变量)和从混合方差中分离出来的特殊因子对预测变量的关系。尤其需要指出的是, Bifactor 模型在处理特殊因子和其他变量的关系优势明显, 因为特殊因子是直接以独立因素呈现的^[83-85]。第二是 Bifactor 模型可以探明在考虑了公共因子对心理结构的影响后, 特殊因素是否还可以作为一个独立的预测变量存在。Spearman 智力理论结构就是典型的例子。该智力理论认为, 除了一般智力外, 还存

在智力相关的特殊因素，如言语能力、空间能力、数学能力和分析能力。现在假设言语能力只是一般智力，而空间能力、数学能力和分析能力除一般智力以外的，还是特殊能力。因而，在 Bifactor 模型中，言语智力在分离公共因子后，其在特殊因子上的载荷将不显著。空间能力、数学能力和分析能力抽取一般智力这一公共因子后还有显著的方差冗余，依然可以作为特殊因子存在，而言语智力在 Bifactor 模型中却不再作为特殊因子存在，从而验证一般智力和特殊智力的理论结构。在 Bifactor 模型中，如果强行把言语能力作为特殊因子，将会因为因素过度萃取而出现模型识别问题。可以表现为以下两种形式，一是在言语能力这一特殊因子上的载荷会非常小并且不显著；二是言语能力特殊因素的方差会不显著。可见，Bifactor 模型可以明确区分公共因子和特殊因子之间的方差变异，从而更清楚的明确不同心理变量之间的关系。这是二阶模型和相关模型所不能实现的^[68, 86]。第三，在 Bifactor 模型中，可以通过观测因素载荷直接考察特殊因子和对应条目的关系强度^[87, 88]。这一点在二阶模型中是不能直接检验的^[47]，因为二阶模型对冗余方差是以残差的方式报告的。第四，除了可以检验一般因素（公共因子）在不同组别的测量不变性外，还可以检验特殊因子跨组测量不变性。这也是其他模型无法实现的^[89]，如二阶模型，只能检验二阶因素的组间不变性，而特殊因子是以残差的方式报告的，因而在处理特殊因子的组间不变性时难以实现^[90]。

在拟合多层面心理结构数据方面，与 Bifactor 模型比较相似的是二阶模型。二阶模型假设存在一个高阶因子解释不同层面的公共变异。与 Bifactor 模型类似，二阶模型也可以处理几个相关的多层面数据。与二阶模型相比，Bifactor 模型同样存在显著的优势。首先，只有 Bifactor 模型能有效分离公共因子和特殊因子。在 Bifactor 模型中，可以计算条目对特殊因子的载荷，从而可以体现特殊因子和条目之间关系的强度。然而，在二阶模型中，条目和特殊因子之间的关系是无法检验的，因为条目方差拟合后的冗余方差是以一阶因子的残差（或误差）的方式报告的^[91, 92]。其次，如前所述，只有 Bifactor 模型能够探明在分离公共因子后是否还存在特殊因子。第三，只有 Bifactor 模型能检验在抽取公共因子后，不同维度层面的差异。在二阶模型中，只有二阶潜在变量能直接比较，此时特殊因子同样作为残差处理^[93]。第四，Bifactor 模型在处理特殊因子和外部变量的关系时应用性更强^[78]。因为在 Bifactor 模型中，特殊因子作为独立的因素进行分离和报告的，而不是误差^[84, 87]。当然，在二阶模型

中，也有研究者研究除二阶因素外，残差和预测变量之间的关系，这需要采用非标准化的结构方程模型，也就是说必须将一阶因子的残差作为一个预测变量，这本身是存在争议的同时在方法上也是很难实现^[94, 95]。

4.3 Bifactor 模型的应用

在海量的 Thurstone 相关因素模型应用面前，Bifactor 模型的价值显得有些暗淡。近年来，测量学家重新发现其是一种非常重要的多维结构模型，成为项目反应理论（IRT）和结构方程模型（SEM）应用和研究中的重要课题^[96, 97]，逐渐成为一个研究热点。以下几点可以充分说明 **Bifactor** 模型在现代心理测量中的重要价值^[98]：

- 1) 主流人格和测量杂志经常发表 Bifactor 模型应用性的文章。
- 2) 运用 Bifactor 解决心理结构如何科学测量和心理结构的理论构建。
- 3) 将 Bifactor 模型和其他竞争模型进行比较，解决具有挑战性的测量估计问题；应用 Bifactor 模型开发自适应测验（CAT），评估项目功能差异（DIF）。

在心理研究中，多层次结构非常普遍。Bifactor 模型非常适合有效分离不同维度及其共有心理结构的方差。如，Watson 和 Clark（1992）^[99]检验了负性情感和其四个维度（悲伤、恐惧、愤怒和愧疚）之间的关系。这四个维度是否能和一般负性情感进行有效分离？结果显示，在应用 Bifactor 模型将悲伤等维度在提取负性情感这一公共因子之后，依然能显著预测外部变量。证实悲伤、恐惧等特殊因子的存在，同时说明他们可以从公共结构中进行分离^[99, 100]。

第二，Bifactor 模型可以应用于检验有关心理结构的多个维度和预测校标之间的关系，或者说对现有理论进行实证性检验^[79, 101, 102]。例如，研究者一直被外倾性和正性情感之间稳健的关系困扰。气质模型认为人格和情感的个体差异最终反映的是共同的内在心理结构^[103]。气质模型的观点认为正性情感的个体差异包含了外倾性的核心本质特征。因此，正性情感和外倾性在高阶水平或是公共因子水平上有较高的相关。因而，在 Bifactor 模型就可以直接检验他们在公共因子上的关系。如果假设该假设成立，他们在公共因子上的相关就应该高于在特殊因子上的相关^[100]。同理，工具模型（instrumental model）认为，大五人格中的宜人性和责任心与情感的关系应该是在低阶水平或是特殊因子之间的关系。因为宜人性是责任心的部分特征，并非整体心理结构，有可能是在特定的情景和生活经历中带来特别的情感反应。那么在

Bifactor 模型中, 宜人性和责任心与正性负性情绪在公共因子上的相关要比在人格和情感特殊因子上的相关要小^[104]。

第三, Bifactor 模型特别适应于心理结构的组间差异(如性别)之间的比较^[90]。应用 Bifactor 模型的最新智力相关研究对现有关于智力性别差异的研究成果提出了挑战。一项关于高中生数学能力的元分析表明,男生在数学能力上高于女生,但是效应量比较小^[105, 106]。基于 Bifactor 模型的研究将一般认知能力和特殊数学能力进行分离得到了不同的研究结论,数学能力的性别差异非常大。然而,同一批数据,在使用传统 CTT 分析方法时,数学能力的性别差异比较小。究其原因,是数学能力混杂在认知能力中,而又鲜有证据表明认知能力的性别差异,因而不能发现数学能力的真实差异。这一研究结果表明, 在报告多维结构量表分数时,汇报量表总分和各维度分都有可能造成结果的偏倚^[107, 108]。总量表分没有考虑每一个维度对总分的贡献性(即特殊因素的方差对总分的作用),而维度分数不能将各维度的共同方差从中分离(即该维度分数可能包含共同心理结构的影响)^[109]。

最后, Bifactor 模型在量表开发和评估方面也能有较好的应用^[67, 110]。在开发或评价一个新的多维测验时,特殊因子和公共因子上的载荷可以用于指导条目选择。理想的情况应该是,公共因子上的载荷应该高于(至少不低于)特殊因子上的载荷。如果某些维度上的条目的载荷主要在公共因子之上,那么可以认为没有这一特殊因子。而如果一些条目的载荷主要是在特殊因子上,那么这些条目应该从该量表中删除,因为他们对该量表需要测量的心理结构贡献不大。

需要指出的是, Bifactor 模型在不只在能力测验,在人格和健康评定测验中也在最近有较多的应用。Reise 等人^[103]证实了增加公共因素后(在量表理论构建的因素之外)的 Bifactor 模型在人格和健康评定测验中有如下几个方面的优势: 1) 如果单维量表采用多维模型估计, Bifactor 模型可以评估参数估计的扭曲程度; 2) 提供了另一种测量个体差异非层级的多维模型分析方法; 3) 研究者可以检验构建分量表的有效性^[69, 70, 72]。Chen 等人比较了非能力测验中 Bifactor 模型相比于二阶模型的显著优势^[89]。Gibbons 等人报告了生活质量测量中, 所有条目在公共因子上有显著负荷,而在特殊因子上同样存在不同程度负荷^[71]。Gibbons 等人同样发现 全息 Bifactor 模型在慢性精神疾病患者在焦虑抑郁量表上有更好的模型拟合^[80, 82, 111], 他们采用 626 个有关情绪和焦虑障碍的条目,使用 Bifactor 模型开发了自适应测验对心理健康进行测

量,模拟研究发现,全息 Bifactor 模型在同等条件下,能更准确的进行条目参数估计和被试心理健康水平估计^[112]。可见,如果可以确认特殊因子之外存在公共方差,相比如传统因素分析模型和单维模型,Bifactor 可以提供更为简洁明了的模型拟合和数据分析。

正文

第一部分 全息 Bifactor 模型中统计计算能力的测量

1 问题提出

过去近 20 年的研究发现, 统计计算能力是规范决策和风险感知能力的最佳预测指标^[10, 21]。因而, 实现统计计算能力的准确测量对于研究统计计算能力在规范决策中的作用也显得非常重要。目前, 对统计计算能力的测量, 最常见的方式是通过设计数学测验题考察被试在数学测验中的成绩。数学测验大多基于经典测量理论 (Classical Testing Theory, CTT)。如决策领域最常用的、Schwartz 等人开发的 Lipkus 数学能力测验 (Lipkus Numeracy Test)。然而, CTT 存在真分数与观测分数线性关系不符合事实、误差与真分数独立的假设难以满足等诸多理论上的缺陷。在测验等值、适应性测验开发和标准参照测验的编制方面都存在较多问题。决策科学中的研究者逐渐将研究兴趣转向现代测量理论, 如 IRT, 以期更准确的测量统计计算能力。Schapira 等人(2012)基于 IRT 开发了医学数学理解能力测验(Numeracy Understanding in Medicine Instrument, NUMi)^[18]。研究发现, 该测验在评估病人被试基本健康数学能力方面具有较好的信度和效度。Cokely 等人(2012)开发了可以实现计算自适应测验的 Berlin 数学能力测验 (Berlin Numeracy testing)^[21], 测验开发者指出该测验可以实现文化程度较高被试群体的统计计算能力测量。Weller 等人 (2013) 基于 Rasch 模型开发了一个包含 8 个条目的数学能力测验, 而其测验题目的来源主要是数学能力量表和认知反思测验^[46]。相比于以往经典理论开发的数学能力测验, 在美国一般人群中具有更好的区分效度, 对风险决策判断也有更好的预测效度。也有研究者认为 Weller 等人的测验中条目的来源是两个不同质的测验, 其内容效度需要重新考虑。但是也有研究者认为, 统计算术能力是一个包罗万象的概念, 除一般意义上的计算能力以外, 统计计算能力的测量还包括 1) 自我调节能力和自知力 (如元认知); 2) 分析思维过程中的情感驾驭能力 (如情感的精确性); 3) 有意义直觉的建立 (如模

糊-要点表征, fuzzy-gist representation) 等等。综合上述内容能更好的测量统计计算能力的全部特征。值得研究者注意的是, Rasch 模型的基本假设之一是单维性, 即测验中的每一个项目都测量同一潜在特质。虽然测验开发者认为两者是测量同一心理品质——统计计算能力。即便如此, 统计计算能力是一个多维的心理结构, 简单应用单维模型对数据进行拟合违背了模型的基本假设。

采用数学能力测验考察统计计算能力的另一个问题是, 数学能力测验本身就包含两个方面的技能, 一是运用数学运算原理和概念的技能, 二是统计计算能力(如统计和条件概率的计算)。运用数学测验的方法测量统计计算能力可能包含测量数学运算在内的混杂因素。如何有效提纯是研究需要思考的问题。而另一个同样采用数学测验测量被试认知反思能力的测验——认知反思测验(Cognitive Reflection Test, CRT)引起了研究者的注意。虽然该测验在测量认知反思方面具有较好的信度和效度, 但是, 测验开发者 Frederick (2005) 本人也承认 CRT 和决策任务的相关关系也可能存在其他方面的解释^[13]。CRT 也许测量的是计算能力甚至是统计计算能力或学业成绩等其他技能而不是认知反思。数学能力调节认知反思和决策之间的关系也有三个重要事实^[14]: 1) 解决 CRT 问题需要一定的数学知识和统计计算能力; 2) CRT 和数学能力及学业成绩存在显著相关; 3) 研究 CRT 与决策之间关系的研究中所设计的决策任务大都包含一定的统计计算。而不需要数学计算的决策任务发现, 数学能力不是认知反思和规范决策的调节变量。也就是说 CRT 的测量与统计计算能力的测量事实上存在很高程度的相关。可见, CRT 和 BNT 的虽然测量目的不同, 但是在关注统计计算能力在决策中的作用有异曲同工之妙。考虑到 CRT 和 BNT 都是数学测验试题、两类数学题都需要一定的统计计算能力以及在实证研究中二者的高相关。结合统计计算能力的定义, 本研究认为 BNT 作为一个数学能力测验测量统计计算能力包含较多的混杂因素, 还不能反应统计计算能力的核心特征, 可以应用 Bifactor 模型结合 CRT 提取两者的公共因子, 二者公共因素可以测量统计计算能力的核心特征。笔者认为两者的公共因子能更好的反应统计计算能力的本质, 可以剔除不必要无关因素的测量误差。因而, 本研究基于全息 Bifactor 模型, 探讨统计计算能力的测量方法。并提出以下假设:

- ① Bifactor 模型是 BNT 和 CRT 联合测验的最优拟合模型;
- ② BNT 和 CRT 联合测验对能力中等水平被试的测量具有较大的信息量;

③ BNT 和 CRT 联合测验测量统计计算能力有较好的信效度。

2 研究方法

2.1 研究被试及其程序

不同学历层次的被试群体共计 564 人, 年龄 18~32 (22.14 ± 4.37) 岁, 其中初中及以下文化程度 126 人, 占 22.34%; 高中 196 人, 占 34.75%; 大学及以上文化程度 242 人, 占 42.91%。

所有被试了解本研究的研究背景、目的和意义, 并签署知情同意书。收回有效数据 548 份, 回收有效率为 97.16%。

2.2 测量工具

2.2.1 Berlin 数学能力测验 (Berlin Numeracy Testing, BNT)

BNT 由 Cokely 等人于 2012 年编制, 见附录。Cokely 等人认为, 该量表设计的主要目的在于采用数学能力测验考察被试的统计计算能力 (Statistical Numeracy)。该测验采用 3 水平分类树法设计了 4 道数学能力测验题, 回答正确计 1 分, 错误不计分。多国大样本中研究发现, 该测验在预测规范风险决策方面具有较好的信度和效度。

2.2.2 认知反思测验 (Cognitive Reflection Test, CRT)

该测验由 Frederick 于 2005 年开发^[113], 包含 3 道数学计算题, 见附录。用于测量被试与分析加工性加工 (系统 2) 过程相关的有意识付出意志努力去解决推理和决策问题的倾向性, 测验总分可以反映被试在解决决策问题时对认知反思和开放性思维的态度和倾向性, 分数越高说明被试更倾向于对不同解决方案采取开发态度、避免使用直觉思维而采用分析性思维解决问题。研究发现, 该测验对于预测较低水平信息处理过程中的认知偏见具有较好的信度和效度。Guillermo Campitelli 等人的研究指出^[114], CRT 测量的是一种思维特质或认知能力 (Thinking Disposition), 与主动开放思维 (Actively Open-Minded Thinking) 有关。该思维特质是知识经验、环境特征和启发思维三者综合作用有关的思维倾向, 为决策者适应不同的决策环境和情景中有非常重要的作用。本研究中主要用 CRT 结合 BNT 探讨统计计算能力的测量方法。

2.2.3 标准瑞文推理测验 (Raven's Standard Progressive Matrices, SPM)

标准瑞文推理测验由 Raven 于 1938 年开发^[115], 是国际通用的流体智力测验工具, 本研究中作为被试一般能力中的推理能力相关指标。整套测验共 60 题, 分为 5 组, 每组 12 题。

2.2.4 Lipkus 数学能力测验 (Lipkus Numeracy Test)

该测验由 Lipkus 等人于 2001 年编制^[2], 示例见附录。该测验共 11 题, 主要用于测量被试理解概率和数学概念的能力, 回答正确计 1 分, 错误不计分。该测验具有较好的信度和效度, 在决策研究中是数学能力的测量应用广泛的工具。

2.2.5 数字表征估计测验

为了验证 Bifactor 模型中统计计算能力的测验效度, 本实验还设计了数字表征估计任务 (见附录), 考察被试对数字表征估计能力。实验材料是饼图, 分为阴影部分和白色部分, 要求被试准确估计饼图中阴影部分或白色部分的百分比。正确计 1 分, 错误不计分。

2.3 数据分析

本研究沿用 Weller 等人对统计计算能力定义和测量方法^[46], 采用 BNT 和 CRT 联合测验考察被试的统计计算能力。数据分析采用以下步骤:

(1) 为了探索联合测验的最优拟合模型, 采用稳健加权最小二乘法 (Weighted Least Squares with Adjusted Mean and Variance, WLSMV) 进行了一系列验证性因素分析。比较了联合测验以下可能测量模型: 单维模型、两维模型和 Bifactor 模型 (Fig. 1) 的拟合程度。其中两维模型是将 BNT 和 CRT 各视为一个维度, 二阶模型是在两维模型的基础上另外还有一个高阶因子。Bifactor 模型是层级模型, 每一个条目有两个因素负荷, 一是公共因子负荷, 二是特殊因子负荷。本研究中, Bifactor 模型可以视为单维模型和双维模型的统合体, 同时包含 2 个特殊因子和一个公共因子。

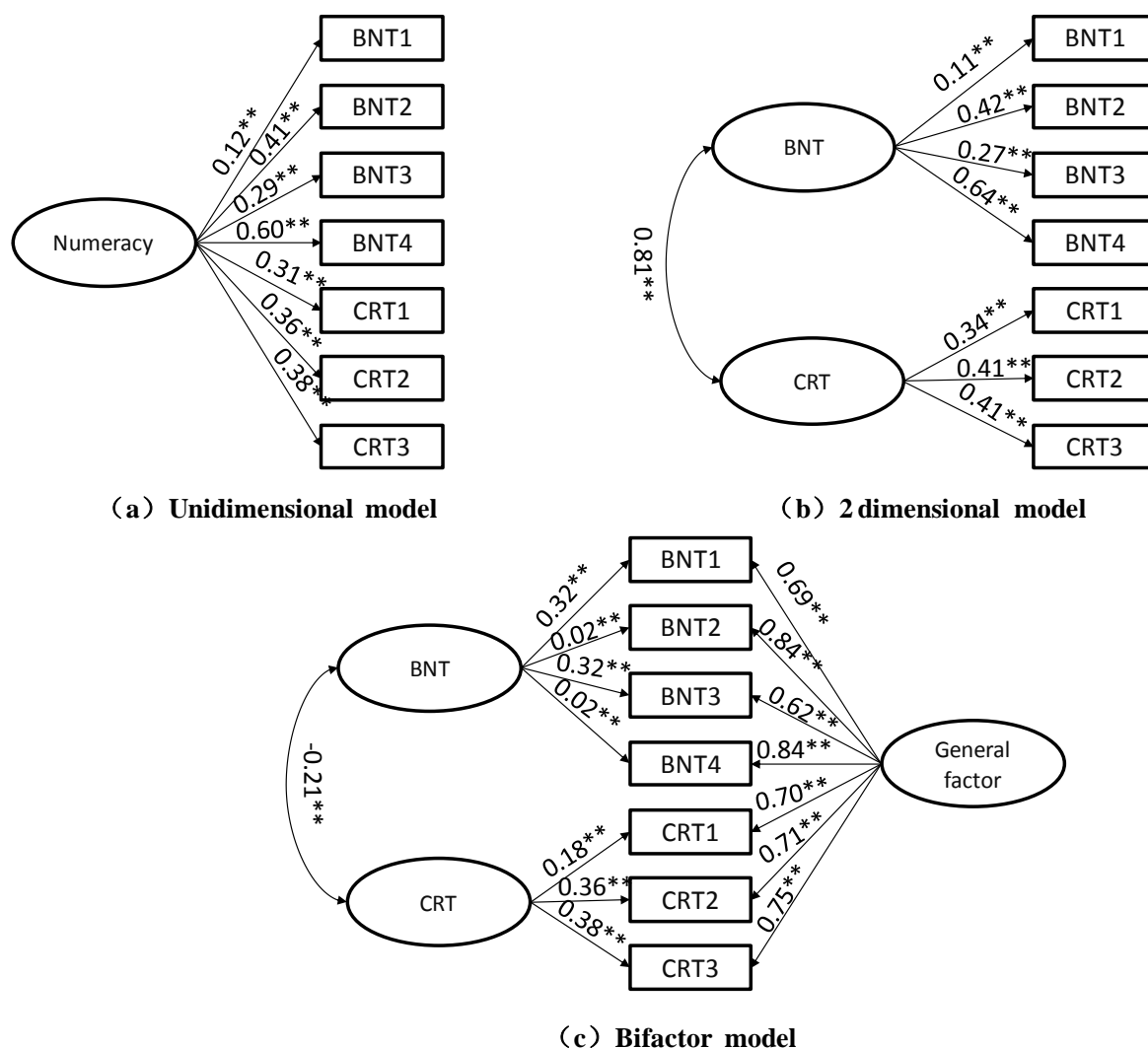


Fig.1 the CFA Models and Factor Loading of BNT and CRT Joint Measurement

Note: the factor loadings of (c) were from FIFA; **: $p < 0.01$

(2) 全息项目因素分析 (Full Information Item Factor Analysis, FIFA)。如果 Bifactor 模型是联合测验最优拟合模型的假设得到证实, 进一步对联合测验进行全息因素分析。本研究采用 FIFA 估计各条目的因素载荷、难度、区分度等参数, 计算条目测验信息量 (Test Information), 绘制测验信息函数图。同时估计被试在各个因子上的得分。因为全息 Bifactor 模型假设公共因子和特殊因子之间的相关为 0, 实现了公共因子和特殊因子的有效分离。从而, 各条目在公共因子上的载荷可以视为各题对公共因子的贡献率。同时, 特殊因子上的载荷也可以直接视为题目对特殊因子的贡献。被试在公共因子和特殊因子上的得分可以视为被试在统计计算能力和特殊能力上的得分。这也为进一步信效度验证提供了基础。

(3) 信效度验证。本研究分别计算了联合测验公共因子、特殊因子的信度系数,

同时以测验信息函数反映测验的总体信度情况。为了进一步考察联合测验公共因子测量统计计算能力的效度，还分别考察了公共因子与瑞文推理测验、Lipkus 数学能力测验之间的相关系数情况。同时以平均数为界，高于均数定义为高统计计算能力组，低于均数为低统计计算能力组。考察两组被试在数字表征估计任务中的成绩差异，作为区分效度指标。

本研究中，以下指标用于评价模型的拟合程度：1) 相对拟合度 (Comparative Fit Index, CFI) > 0.95, 2) 近似误差均方根 (the Root Mean Square Error of Approximation, RMSEA) < 0.06; 3) 标准化残差均方根 (Standardized Root Mean Square Residual, SRMR) < 0.06。如果模型满足以上指标要求，则认为该模型拟合较好。

描述统计、相关分析和独立样本 *t* 检验采用 SPSS16.0 软件进行，CFA 采用 Mplus 7.0 软件，全息项目因素分析采用 POLYBIF 软件包进行。

3 研究结果

3.1 验证性因素分析 (CFA) 结果

本研究应用 Mplus 7.0 分别对单维模型、两维模型和 Bifactor 模型进行验证性因素分析 (CFA)，探讨最优拟合模型。模型拟合指数见 Table 1。两维模型中，BNT 和 CRT 的相关相对较高 ($r=0.81$, $p<0.01$, Fig. 1)，说明两者的测量包含公共成分，需要应用 Bifactor 模型进一步拟合联合测验的数据。从模型拟合指数来看，单维模型、两维模型和 Bifactor 模型都能基本满足模型拟合要求。综合比较三个模型的拟合指数，Bifactor 模型的拟合指数最好，提示该模型是最优拟合模型。与其他模型相比，Bifactor 模型相对于其他模型的方差最小，同样也提供了 Bifactor 模型是最优拟合模型的直接证据。为此，本研究应用全息 Bifactor 模型进一步分析被试在公共因子和特殊因子上的得分情况。

Table 1 Model Fit Comparison of CFA Models

Models	df	χ^2	CFI	RSMEA	SRMR	$\Delta\chi^2$
Unidimensional	21	205.23	0.91	0.064	0.044	101.78
2-dimensional	20	201.31	0.92	0.043	0.021	97.86
BifactorModel	11	103.45	0.98	0.021	0.019	-

3.2 CRT 和 BNT 联合测验的 IRT 分析

本研究采用 PLOYBIF 程序对 CRT 和 BNT 联合测验进行 IRT 分析, 具体结果见 Table 2。BNT 和 CRT 联合测验所有条目在公共因子上的载荷都要大于其在特殊因子上的载荷, 可见, BNT 和 CRT 联合测验测量的内容更多的是两者公共因子上(统计计算能力), 而不是其特殊因子。证实了以往研究关于 BNT 和 CRT 两个量表在测量内容上有很程度上的一致性。对 BNT 和 CRT 抽取公共因子, 排除了数学测验中运算法则、特定数学情景等无关因素的干扰, 因而也考察了统计计算能力的核心特征, 为进一步计算被试统计计算能力提供了理论基础。

Table 2 Main Parameters of Bifactor Analysis for BNT and CRT Joint Measurement

Items	Factor Loading _{general factor}	Factor Loading _{BNT}	Factor Loading _{CRT}	Slope _{general factor}	Slope _{BNT}	Slope _{CRT}
1	0.689	0.321	-	1.213	-0.110	-
2	0.884	-0.022	-	1.543	1.243	-
3	0.662	0.318	-	1.307	-0.031	-
4	0.839	0.017	-	1.020	0.739	-
5	0.702	-	0.179	0.948	-	0.670
6	0.710	-	0.363	1.074	-	0.299
7	0.748	-	0.381	1.647	-	0.180

3.3 Bifactor 模型中 CRT 和 BNT 联合测验公共因子的信度分析

同时, 本研究应用 PLOYBIF 程序计算了 IRT 中的信度系数, 联合测验公共因子和特殊因子上的信度分别为 $r_{\text{general factor}}=0.76$, $r_{\text{BNT}}=0.14$, $r_{\text{CRT}}=0.19$ 。可见, 联合测验在公共因子上的信度系数更高, 说明公共因子测量的稳定性更好。同时, 应用各条目在公共因子上的 IRT 参数绘制了联合测验公共因子的测验信息函数(见 Fig. 2)。从信息函数上看, 公共因子对能力值处于平均值上下被试的统计计算能力测量的信息量更大, 换句话说, 对这类被试的测量更为准确。而对处于极端能力水平(极大或极小)被试的测量相对不准确。

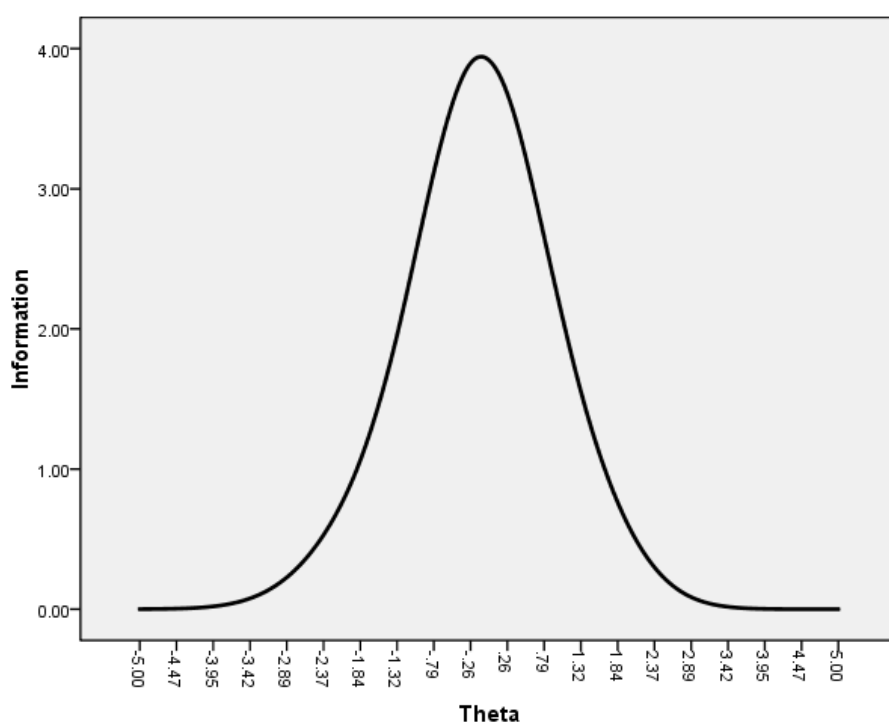


Fig.2 the Test Information of General Factor for BNT and CRT Joint Measurement

3.3 CRT 和 BNT 联合测验公共因子的效度分析

为了进一步明确 BNT 和 CRT 联合测验公共因子的测验内容效度, 本研究分别计算了公共因子和 BNT、CRT、瑞文推理测验总分、Lipkus 数学能力测验总分的相关。结果见 Table 3。公共因子和数学能力、瑞文推理测验的总分相关显著, 说明公共因子的测量包含了数学能力和推理能力测量。更进一步分析发现, 公共因子与瑞文推理总分的相关比瑞文推理总分与 BNT 和 CRT 的相关都要高, 说明公共因子能更好的测量与瑞文推理测验相关的推理能力。

Table 3 Correlations between General Factor and External Predictor

	1	2	3	4	5
1 General Factor	1				
2 BNT	0.93**	1			
3 CRT	0.83**	0.62**	1		
4 Numeracy	0.43**	0.44**	0.35**	1	
5 SPM	0.62**	0.58**	0.57**	0.46**	1

Note: ** $p < 0.01$.

同时, 本研究还比较了不同统计计算能力被试在数字精确表征上的差异, 结果见 Table 4。独立样本 t 检验表明, 高统计计算能力对数字精确表征的成绩显著高于低统计计算能力被试, 说明公共因子在数字表征任务上有较好的区分效度。至此, 假设 3 得到验证。

Table 4 the Difference between Different Statistical Numeracy in Numerical Representation Tasks

Statistical Numeracy	N	$\bar{X} \pm SD$	t	p
High	98	1.48 \pm 0.50	6.85	<0.001
Low	68	0.82 \pm 0.73		

4 讨论

本研究采用全息 Bifactor 因素分析的方法, 探讨了统计计算能力的多维 IRT 测量方法。研究结果支持 Bifactor 模型中 BNT 和 CRT 联合测验的公共因子测量统计计算能力具有较好的信度和效度。CFA 模型比较结果发现, 相比于其他可能模型, Bifactor 模型是 BNT 和 CRT 联合测验的最优拟合模型并且 χ^2 值最小, 支持联合测验符合 Bifactor 模型理论结构。通过对联合测验的 IRT 分析发现, 联合测验在公共因子的负荷大于特殊因子上的负荷, 说明联合测验测量的心理品质更多的是公共因子而不是特殊因子, 进一步支持两个量表测量的是同一心理品质。而 IRT 的信度系数表明公共因子的测量信度较好。公共因子与瑞文推理测验、数学能力的相关分析表明, 公共因子同时测量了数学能力和推理能力, 这与统计计算能力的概念符合程度比较高, 根据两个测验的性质和目的, 同时公共因子在数字精确表征方面具有较好的区分效度, 说明联合测验公共因子测验效度较好。综上, 可以推断, 公共因子对统计计算能力的测量具有较好的信度和效度。

4.1 公共因子测量统计计算能力符合统计计算能力理论构想

从统计计算能力的概念和定义来看, 其是一个多维的心理结构。包含了数学能力和推理能力等其他认知成分。因而单纯应用数学能力测验恐怕难以测量统计计算能力的核心本质特征, 如 BNT。以往研究中, 如 weller 等人, 也曾经基于 IRT Rasch 模型同时从数学能力测验和认知反思测验选择相关题目开发测量统计计算能力的量表^[46,116, 117]。这与本研究的理论构想思路一致。但是, 数学能力测验和认知反思测验

项目是两类不同质的题目（虽然两者都是采用数学测验，但是两类测验的理论构想是不同的），违背了单维 IRT 模型单维性的基本假设，违背了单维 IRT 模型局部独立性的假设，其研究结果和结论当然受到研究者的质疑。为此，本研究采用多维 IRT 模型拟合统计计算能力测量数据，更符合统计计算能力多维性的特点。本研究中验证性因素分析也表明，两维模型的拟合指标更好，证明了统计计算能力多维理论结构。需要应用多维 IRT 模型对数据进行拟合，因为多维 IRT 能更有效地估计被试能力和项目参数，对统计计算能力的测量也更为准确。

Bifactor 模型能实现项目反应协方差公共因子和特殊因子的分离^[105]，因而也可以更清楚的理解条目对公共因子和特殊因子贡献，考察测验条目所测查的真实内容和目的。在本研究中，BNT 和 CRT 的条目在公共因子上的载荷都要大于特殊因子上的载荷，说明 BNT 测验和 CRT 测验在测量内容上更多的是两者的共有成分，或者说统计计算能力。这与 Matthew B. Welsh 等人研究证实的 CRT 测量的更多的是数学能力、反映了理性认知风格的部分特征的研究结果是一致的^[118]。Frederick 本人也承认 CRT 测量中包含数学知识理解和统计计算能力的测量。当然，不同研究者采用不同的实验任务得出的结论也不一致，也有研究者认为 CRT 测量的是与主动开放思维有关的认知反思。从本研究的结果来看，CRT 和 BNT 一样，测量内容反映的同一心理结构——统计计算能力。也有研究者指出，Bifactor 模型中公共因子与 CTT 中总分的计算一样，反映的是测验中所有条目的全部心理结构^[89]。也就是说公共因子与测验总体上的能力水平是一致的。但与 CTT 不同的是，Bifactor 模型中的公共因子所测查的能力值具备等级量表的性质，突破了 CTT 理论本身的缺陷，因而该分数应用范围更广，所测量的内容也更为精确。可见，公共因子是 BNT 中的统计计算能力测量和 CRT 中数学能力、理性思维风格的综合分数，这与统计计算能力的概念更贴近。因而，应用 Bifactor 模型公共因子对统计计算能力的测量是目前比较合理的解决方案。

4.2 公共因子测量统计计算能力信效度较好

同时，本研究还证实，采用 Bifactor 模型公共因子测量统计计算能力具有较好的信度和效度。公共因子在 IRT 中的信度系数达到了理想的水平，测验信息函数进一步表明，公共因子对中等能力水平的被试测验精度更好，对区分中等水平统计计算

能力的被试具有较好的测量精度。公共因子同 BNT 和 CRT 一样，与数学能力存在较程度的相关。而公共因子相比于单独的 BNT 或 CRT 与瑞文推理测验的相关要更高，可见，统计计算能力在推理能力的预测上可解释的方差要更多。在数字表征任务上，高统计计算能力被试对数字精确表征成绩显著高于低统计计算能力被试。这些都是公共因子测量统计计算能力的直接效度证据。可见，本研究应用全息 Bifactor 模型技术，通过抽取 BNT 和 CRT 的公共因素，由此计算被试的统计计算能力分数。可以排除基本数学技能等干扰因素，能反应统计计算能力的核心特征，实现对统计计算能力更为精确的测量。

第二部分 基于全息项目因素分析探讨统计计算能力在决策能力中的作用

1 问题提出

如前所述,统计计算能力可以较好的预测被试的规范决策和风险感知能力。Hamm 等人(2003)发现高计算能力被试在前列腺癌概率估计中的精确性更高,对医疗方式的决策选择也更为理性和合理^[119]。Gurmankin, Baron 和 Armstrong (2004)发现低算术能力被试更偏好语义描述的风险信息^[120],而高算术能力被试更偏好数字性的风险信息,而两者在决策质量和速度方面都有明显差异。Paulos 认为被试不能理性处理“概率低但是后果严重”的决策问题容易导致对政府政策的误解、个体决策的混乱、更容易相信伪科学等^[121]。Peters (2006)研究发现,不同数学能力对决策的影响不是因为智力等一般因素的作用造成的,数学能力在对抵御决策中的框架效应、对决策中风险信息的感知、情感信息理解以及不同风险等级的理解和把握等方面都有重要作用^[5]。

然而,规范决策不仅仅是“做做数学题”。一个好的决策往往需要理解和处理各类复杂关系,如取舍权衡、风险收益平衡、考虑方案的可操作性等等^[122, 123]。西蒙认为,信息不完备性、价值偏好的多样性、决策能力的有限性、事务关系的易变性等复杂因素,都会影响人们在资源有限的条件下做出最优选择。推动研究者开始从决策能力的视角考虑人们在规范决策过程中的个体差异^[3, 124]。Fischhoff 等认为学习能力独立于决策能力之外,当人们面临全新的问题或者毫无学习经验时,仍然需要普遍适用的决策技能,包括摘取有用信息、应用价值与具体情境设置、整合信息片段为决策规则等^[4, 125, 126]。决策能力即核心决策技能(Core Decision-Making Skill)。Fischhoff 等整合前人的研究成果,总结出四项核心决策能力:信念评估(Assessing Beliefs),即洞察事件发生概率的能力;价值评估(Assessing Values),即评估选项后果,对决策相关信息改变敏感,而对决策不相关信息不敏感;整合能力(Integration),即对决策规则的认知;能力的元认知(Metacognition of One's Abilities),即对自己的能力

有清楚的认知与判断^[4, 40, 127, 128]。决策能力相关研究发现，统计计算能力和规范决策的紧密关系是因为二者类似的认知成分。从决策能力的定义上来看，决策能力包含了个体做出规范决策需要的所有核心特征。因而，研究统计计算能力和决策能力的关系，可以探讨统计计算能力在规范决策中的地位和作用。

全息 Bifactor 模型在临床、教育和心理测量中有非常广泛的应用，如有效分离不同维度及其共有心理结构的方差、心理结构的组间差异（如性别）之间的比较等。本部分重点回顾 Bifactor 模型应用于检验心理结构的多个维度和预测校标之间的关系，或者说 Bifactor 模型可以对现有理论进行直接验证。例如，研究者一直被外倾性和正性情感之间稳健的相关关系所困扰。气质模型的观点提出正性情感的个体差异包含了外倾性的核心本质特征，从而使正性情感和外倾性表现为较高度度的相关。研究者应用 Bifactor 模型对这一理论模型进行了直接检验^[105]：正性情感和外倾性在公共因子上的相关高于特殊因子上的相关，从而证实气质模型的观点。另一个明显的例子是，工具模型（Instrumental Model）认为，大五人格中的宜人性和责任心与情感的关系应该是在低阶水平或是特殊因子之间的关系。因为宜人性是责任心的部分特征，并非整体心理结构，情绪特征只是在特定的情景和生活经历中带来特殊的情感反应^[100]。如果应用 Bifactor 模型对这一理论进行直接验证，宜人性和责任心与正性负性情绪的公共因子上与正负情绪公共因子上的相关要比在人格和情感特殊因子上的相关要小^[103]。同理，目前研究者也为数学能力或统计计算能力与决策能力之间的高相关性所困扰，也有研究者从理论的角度提出数学能力或统计计算能力是决策能力的核心成分，但研究者始终未能从实证的角度对这一理论假设进行科学的验证。统计计算能力和决策能力之间的高相关性到底如何解释？如上所述，统计计算能力到底是决策能力的核心成分还只是决策能力的部分特征呢？本研究发现这个问题同样可以应用 Bifactor 模型直接进行验证，从现有文献来看，还没有研究者应用 Bifactor 模型验证这一科学问题。这是本部分试图回答的问题之一。

因此，本部分的目的在于应用全息项目因素分析（FIFA）评价决策能力的维度，探讨不同决策能力之间的潜在关系，以此为基础探明统计计算能力和决策能力之间的特殊关系，进一步推断统计计算能力在规范决策中的作用。本研究假设被试完成一道决策能力测验题由两个因素决定，一是公共因素（决策元认知成分）和特殊因素（如信念评估能力、价值评估能力、整合能力）。全息双因子（Bifactor）模型能实

现决策能力量表被试反应数据方差的有效拆分，一是公共方差，二是特殊因子方差，从而将决策能力一般因素和特殊因素分离开来。为实现上述目标，本研究首先采用验证性因素分析的方法，探明决策能力不同维度之间可能的潜在结构模型。本研究比较单维模型、4 维模型、二阶模型和 Bifactor 模型的最优拟合度，对不同理论模型进行模型比较，提出最优拟合模型。如果 Bifactor 模型是最优拟合模型的假设得到验证，本研究首先基于 IRT 分别计算被试在每一题上公共因子和特殊因子上的决策能力得分情况，以此为基础探讨不同决策能力的潜在关系。之后，进一步分析统计计算能力和决策能力及决策能力各维度之间的关系。分别计算决策能力特殊因子和公共因子与统计计算能力的相关，探讨统计计算能力在决策中的作用。基于以上计算和分析，本研究可以为深入理解不同决策能力之间的潜在关系提供一定的数据支持，并进一步深入探讨决策能力和统计计算能力之间的关系。

2 研究方法

2.1 研究被试及其研究程序

采用整群抽样的方法选取不同学历层次的被试群体共计 462 人，年龄 18~25 (20.36 \pm 2.56)岁，其中初中及以下文化程度 62 人，占 13.42%；高中 226 人，占 48.92%；大学及以上文化程度 174 人，占 37.66%。

被试首先在计算机上完成本研究自编的《决策能力测量》程序测验，随后完成纸笔测验的统计计算能力测验，收回有效数据 460 份。

2.2 测量工具

2.2.1 统计计算能力测量

同第一部分。

2.2.2 决策能力测量

决策能力测量采用卡梅隆大学 Fischhoff 等人编制的《成人决策能力问卷》(Adult Decision Making Competence, A-DMC, 示例见附录)^[4, 125, 129]。该问卷采用 6 种决策任务测量信念评估、价值评估、整合能力和能力元认知 4 种核心决策技能。其中信念评估采用风险感知一致性(20 题)和社会规范识别(16 题)两类任务测量。风险感知一致性要求被试判断指定事件(如车祸)在指定时间(1 年或 5 年)内发生的概

率。计分采用配对的方式进行计分,即被试认为同一事件在 5 年内发生的概率大于 1 年内发生的概率记 1 分,反之不计分。被试认为低概率事件发生的可能性小于大概率事件发生的可能性记 1 分,反之不计分。社会规范识别包含两方面的任务,首先要求个体回答是否接受某一负性事件(如偷窃),然后请被试判断有多大百分比的人会接受这一负性事件。被试接受负性事件的概率和人们普遍接受的概率二者的相关值即为被试的得分。价值评估采用抵御框架效应(14 题)和抵御成本沉没效应(10 题)两类任务进行测量。抵御框架效应包含 7 道风险框架效应和 7 道属性框架效应问题。两类问题分别采用正性和负性方式进行描述,请被试分别在 1-6 点量表上判断方案或人、物的喜好程度,两类描述方式中被试喜好程度的决策差值即被试受框架效应影响的程度,反向计分即为被试决策能力得分。抵御成本沉没主要考查被试是否容易受到已经发生的事件对当前决策的影响,被试越不容易受过去事件的影响得分越高。整合能力采用决策规则任务(15 题)进行测量。决策规则任务要求被试依据决策规则对不同商品做出选择,选择符合决策规则的选项记 1 分,否则不计分。能力元认知是指一个好的决策者应该知道自己的知识和能力能否应对当前的决策任务,既不盲目自信、缺乏自省,也不因为信心不足而举棋不定,主要采用自信任务(34 题)进行测量。该类任务要求判断某一命题是否正确(香烟中的尼古丁可以用于缓解抑郁症状),然后要求被试判断自己回答这道题时的把握程度(从 50%到 100%),1 减去全部题目的正确率与把握程度的绝对差值即为被试的得分。

因为社会规则识别只计算总体分数,不便于计算各题的得分情况,本研究暂不进行全息因素分析。而自信任务本身就是能力的元认知,与其他任务能力测量交叉较大,本研究将其视为公共因子的一个效标。为此,本研究选取抵御成本沉没、抵御框架效应、决策规则和风险感知一致性 4 类决策任务进行模型拟合研究和进一步的相关数据分析。

2.3 数据分析

数据分析采用以下步骤:

(1) 为了探索 A-DMC 被试反应数据的最优拟合模型,采用稳健加权最小二乘法(Weighted Least Squares with Adjusted Mean and Variance, WLSMV)进行验证性因素分析。本研究比较了 A-DMC 以下测量模型,单维模型、四维模型、二阶模型和

Bifactor 模型 (Fig. 3) 的拟合程度。如前所述, Bifactor 模型是层级模型, 本研究中, Bifactor 模型可以视为单维模型和四维模型的统合体, 同时包含 4 个特殊因子和一个公共因子。

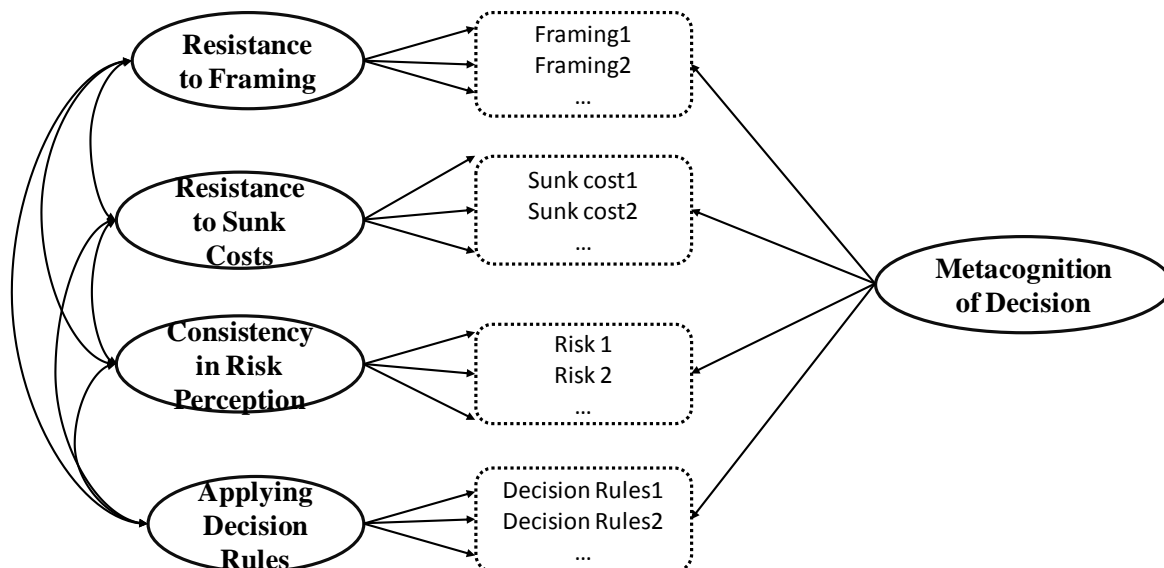


Fig. 3 A-DMC Bifactor Model

(2) 采用全息 Bifactor 因素分析进一步明确 A-DMC 测验的结构, 同时估计被试在各个因子上的得分。应用全息 Bifactor 因素分析, 可以将决策能力方差变异区分为特殊因子和公共因子。因而各题项在公共因子上的载荷可以视为各题对决策能力公共成分的贡献, 在特殊因子上的载荷可以视为各题对特殊决策技能的贡献。同理, 被试在公共因子和特殊因子上的得分可以视为被试在能力元认知和特殊决策能力上的得分。由此, 依据模型拟合程度和载荷系数大小可以明确决策能力的认知结构。为了验证抵御成本沉没、抵御框架效应、风险感知一致性和决策规则公共因子的内容效度, 探讨其具体认知成分, 本研究还计算了公共因子与决策能力元认知的相关。

同时采用 FIFA 估计因素载荷、难度、区分度等条目参数, 计算 A-DMC 的测验信息量 (Test Information), 绘制测验信息函数图, 作为公共因子的测验信度。

(3) 为了探明各决策能力之间的关系, 本研究分计算了决策技能在 CTT 和 Bifactor 理论结构中的相关。Bifactor 理论结构中, 因为特殊因子和公共因子已经分离, 被试在特殊因子上的得分是剔除决策能力元认知后的更为纯净的特殊决策能力, 因而可以更明确的揭示不同决策技能之间的关系。

(4) 为了进一步明确统计计算能力和决策能力之间的关系。本部分采用前述统

计算能力的测量方法，计算被试统计计算能力得分，考察决策能力和统计计算能力之间的相关。如若统计计算能力在决策能力公共因子上的相关要低于特殊能力上的相关，则提示统计计算能力是决策能力的部分特征。统计计算能力在决策能力公共因子上的相关大于特殊因子上的相关，则提示统计计算能力是决策能力的核心成分。同时，通过相关系数的大小也可以侧面反应统计计算能力和各决策能力关系的紧密程度。

本部分中，以下指标用于评价模型的拟合程度^[130-132]：1)相对拟合度(Comparative Fit Index, CFI) >0.95；2)近似误差均方根(the Root Mean Square Error of Approximation, RMSEA) <0.06；3)标准化残差均方根(Standardized Root Mean Square Residual, SRMR) <0.06。模型满足以上指标要求，则认为该模型拟合较好。

描述统计和相关分析采用 SPSS16.0 软件进行，模型拟合检验采用 Mplus 7.0 软件，FIFA 采用 POLYBIF 软件包。

3 研究结果

3.1 验证性因素分析(CFA)结果

本研究应用 Mplus 7.0 分别对单维模型、4 维模型、二阶模型和 Bifactor 模型进行验证性因素分析(CFA)，探讨最优拟合模型。模型拟合指数见 Table 5。从拟合指数来看，单维模型达不到模型拟合要求，4 维模型、二阶模型和 Bifactor 模型都能满足模型拟合要求。综合比较模型的拟合指数不难发现，Bifactor 模型的拟合指数最好，说明 Bifactor 模型对被试反应数据的拟合最好，上述四个模型中，Bifactor 模型是最优拟合模型。同时发现，与其他模型相比，Bifactor 模型相对于其他模型的方差最小，更进一步说明 Bifactor 模型是最优拟合模型。

Table 5 Model Fit Comparison of CFA A-DMC Models

Model	df	χ^2	CFI	RSMEA	SRMR	$\Delta\chi^2$
Unidimensional	768	5963.26	0.74	0.068	0.082	3644.81
4-dimensional	762	3846.43	0.92	0.053	0.071	1527.98
Second order Model	743	3574.71	0.94	0.050	0.054	1256.26
Bifactor model	721	2318.45	0.96	0.041	0.039	-

3.2 A-DMC 的全息项目 Bifactor 因素分析

接下来,采用 PLOYBIF 程序计算 A-DMC 各条目的 IRT 参数和被试在各因子上的得分。A-DMC 各条目的因素载荷见 Table 6。全息 Bifactor 因素分析发现,所有条目在公共因子上都有显著负荷,说明公共因子可以解释不同决策能力的共同方差。除公共因子外,绝大多数条目在其相应的特殊因子上也有中等到较大程度不等的因素载荷。可见,决策能力还有公共因子不能解释的特殊方差变异。说明决策能力的测量包含公共方差变异和特殊因子方差,单一特殊因子或公共因子尚不能解释所有方差变异。需要指出的是,除抵御框架效应外,其他因子上都有条目在特殊因子上的负荷是负或是零,这说明这些条目与原始特殊因子可能是负相关关系。

Table 6 Factor Loading of Each Item for A-DMC

Item	General Factor	Resistance to Sunk Costs	Resistance to Framing	Consistency in Risk Perception	Applying Decision Rules
1	0.186	-0.720	-	-	-
2	0.385	0.117	-	-	-
3	0.097	-0.094	-	-	-
4	0.416	0.906	-	-	-
5	0.224	0.162	-	-	-
6	0.022	0.038	-	-	-
7	0.009	0.252	-	-	-
8	0.321	0.387	-	-	-
9	0.467	0.623	-	-	-
10	0.967	-0.240	-	-	-
11	0.489	-	0.494	-	-
12	0.470	-	0.514	-	-
13	0.479	-	0.632	-	-
14	0.346	-	0.521	-	-
15	0.194	-	0.333	-	-
16	0.477	-	0.439	-	-
17	0.353	-	0.304	-	-
18	0.467	-	0.382	-	-
19	0.522	-	0.548	-	-
20	0.435	-	0.377	-	-
21	0.433	-	0.531	-	-
22	0.674	-	0.738	-	-
23	0.467	-	0.406	-	-
24	0.608	-	0.514	-	-
25	0.432	-	-	-0.085	-

Table 6 Factor Loading of Each Item for A-DMC (continuing)

Item	General Factor	Resistance to Sunk Costs	Resistance to Framing	Consistency in Risk Perception	Applying Decision Rules
26	0.326	-	-	0.016	-
27	0.488	-	-	0.128	-
28	0.314	-	-	0.015	-
29	0.413	-	-	0.22	-
30	0.189	-	-	0.293	-
31	0.124	-	-	-0.136	-
32	0.269	-	-	-0.075	-
33	0.028	-	-	-0.049	-
34	0.255	-	-	0.150	-
35	0.020	-	-	0.672	-
36	0.151	-	-	0.988	-
37	0.359	-	-	0	-
38	0.032	-	-	0.172	-
39	0.356	-	-	0.029	-
40	0.121	-	-	-	-0.153
41	0.451	-	-	-	0.761
42	0.282	-	-	-	-0.120
43	0.282	-	-	-	-0.120
44	0.282	-	-	-	-0.120
45	0.318	-	-	-	-0.312
46	0.553	-	-	-	0.344
47	0.521	-	-	-	0.388
48	0.405	-	-	-	-0.070
49	0.500	-	-	-	0.199
50	0.557	-	-	-	0.226
51	0.078	-	-	-	0.148
52	0.606	-	-	-	0.483
53	0.645	-	-	-	0.509
54	0.649	-	-	-	0.298
55	0.914	-	-	-	0.282
56	0.912	-	-	-	0.353
57	0.511	-	-	-	0.631
58	0.339	-	-	-	0.931
59	0.369	-	-	-	0.902

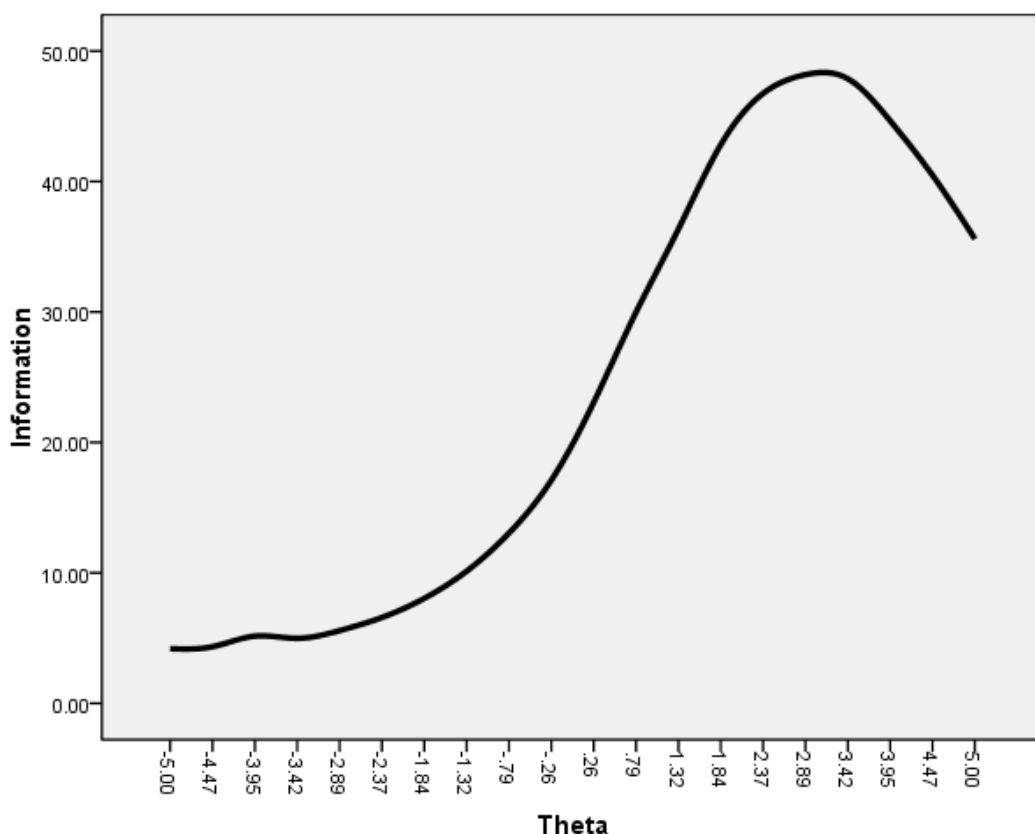


Fig.4 the Test Information of General Factor for A-DMC

应用各条目在公共因子上的 IRT 参数绘制了 A-DMC 公共因子的测验信息函数。测验信息函数在 IRT 中是反应测量精度的信度指标。从 Fig. 4 发现, A-DMC 公共因子在对 Theta 值在 0 以上的被试测验信息量在 20 以上, 说明 A-DMC 公共因子对高于平均水平的被试决策能力公共因素测量精度更大, 而对低决策能力被试在公共因子上的测量精度相对要小。换句话说, A-DMC 公共因子更适宜于区分高决策能力被试。

3.3 决策能力公共因子与决策能力元认知的关系

为了进一步探明决策能力公共因子内在认知成分, 本研究计算了决策能力公共因子和特殊因子与决策能力元认知——过分自信/不自信的相关, 结果见 Table 7。决策能力公共因子与过分自信/不自信显著相关, 说明决策能力公共因子是与决策能力元认知高相关的一种思维特质。同时, 通过 Table 7 还发现, 在经典测验理论下, 抵御成本沉没、抵御框架效应、风险感知一致性和决策规则与过分自信/不自信显著相关, 即与决策能力元认知相关显著。在提取公共因子后, 过分自信/不自信与四者的

相关关系不再显著，更进一步说明公共因子提取的是四者的决策能力元认知。

Table 7 Correlations between General Factor and Under/overconfidence

	General factor	Resistance to Sunk Costs	Resistance to Framing	Consistency in Risk Perception	Applying Decision Rules
Under/overconfidence	0.425**	0.228** (0.038)	0.290** (0.030)	0.197** (-0.125)	0.329** (0.059)

Note: the correlation coefficient in bracket was the correlation between different decision making skills and Under/overconfidence; * $p < 0.05$; ** $p < 0.01$.

3.4 不同决策技能之间的关系

为了探明不同决策技能之间潜在的关系，本研究比较了 CTT 和 Bifactor 结构中，不同决策技能之间的相关系数变化情况。从 Table 8 可以看到提取公共因子前后两类相关系数的变化：一是相关关系显著增强，抵御成本沉没和抵御框架效应的由相关不显著变为显著负相关；二是相关减弱，决策规则与抵御成本沉没、抵御框架效应和风险感知一致性由显著相关变为相关不显著。可见，因为不同决策技能中因为包含了决策能力元认知，掩盖了不同决策技能之间的潜在关系。

Table 8 The Correlation Among Different Decision Making Skills in CTT and Bifactor Analysis

General factor		1 Resistance to Sunk Costs (SF RSC)	2 Resistance to Framing (SF RF)	3 Consistency in Risk Perception (SF RSP)
1	1.00			
2	.095 (-.188*)	1.00		
3	.125 (-.024)	-.153* (-.139*)	1.00	
4	-.329** (-.024)	.475** (-.066)	.131* (-.063)	1.00

Note: SF RSC =specific factor for Resistance to Sunk Costs; SF RF =specific factor for Resistance to Framing; SF RSP =specific factor for Consistency in Risk Perception; the correlation coefficient in bracket was the correlation in CTT, and the correlation coefficient above bracket was the correlation in IRT; * $p < 0.05$, ** $p < 0.01$.

3.4 统计计算能力与决策能力的关系

为探讨统计计算能力在决策能力中的作用，本研究分别计算了统计计算能力与决策能力公共因子、特殊因子之间的相关（结果见 Table 9）。结果发现，统计计算能

力与公共因子的相关显著大于统计计算能力与决策能力特殊因子之间的相关，依据前述推论，本研究可以推断统计计算能力是决策能力的核心因素之一。

Table 9 the Correlation between Different Decision Making Skills and Statistical Numeracy

	General factor	Resistance to Sunk Costs	Resistance to Framing	Consistency in Risk Perception	Applying Decision Rules
Statistical Numeracy	0.462**	-0.022	0.027	0.354**	-0.039
Z	-	7.936	5.453	1.967	8.163
P	-	0.000	0.000	0.049	0.000

Note: * $p < 0.05$, ** $p < 0.01$; The value of Z and p are the difference between the correlation of statistical numeracy and the general factor of A-DMC, and the correlation of statistical numeracy and the specific factor of A-DMC respectively.

4 讨论

本研究首先基于 FIFA 研究不同决策能力的潜在结构，明确不同决策技能之间的关系。以此为基础，探讨了统计计算能力在决策能力中的作用。CFA 和 FIFA 数据结果支持决策能力测量的 Bifactor 模型结构、决策能力公共因子与自信/不自信的相关分析支持决策能力公共因子是决策能力元认知，而抽取公共因子前后决策技能特殊因子相关系数的变化支持公共因子的存在掩盖了不同决策技能之间的潜在关系；统计计算能力与决策能力公共因子的相关显著大于其与特殊决策技能的相关支持统计计算能力是规范决策中的核心认知因素之一。

4.1 决策能力测量的双因子模型结构

本研究初步表明，对决策能力的测量应该是双维（Bi-dimensional）的，其中一个维度测量决策能力元认知，另一个维度测量特殊决策技能（如抵御框架效应、抵御成本沉没等等）。决策能力元认知的存在可以更好的解释被试在不同决策任务中成绩的一致性。如研究发现被试在抵御框架效应、抵御成本沉没、运用决策规则都有较大程度的一致性，即在一类任务中成绩优秀时，在另一类任务中往往也表现不错。Stanovich and West (1998, 2000)报告了不同决策任务中被试成绩的显著正相关^[133, 134]。Guillermo Campitelli 等人研究发现决策能力元认知是与主动开放思维相关的一种思维特质^[136]，是现有知识经验、启发思维和环境特征综合作用的结果，该特质监

控和驾驭启发思维过程，同时也是在合适的情景中调动恰当的启发思维特质，对于决策者的环境适应能力有重要作用。从本研究的结果来看，不同决策任务中的一致性表现是因为决策能力元认知对不同决策能力的监控和驾驭能力，元认知高的个体在不同的决策任务中对不同的决策能力具有同样高的调用和驾驭能力，能更好的运用启发思维特质和过程，因而在各类决策任务中的表现都比较好，而低元认知个体则正好相反。从而解释了不同决策任务中个体成绩的跨情境一致性，这当然是特殊决策技能所不能回答的问题^[135]。可见，对决策能力的测量应该是双维的，研究者不但要关注特殊决策技能在决策中的作用，同样需要关注决策元认知的跨情景一致性。

尤其需要指出的是，特殊决策技能的有效发挥依赖于决策能力元认知的主动开放思维，元认知可以促进不同决策情景中的特殊决策技能的启动。虽然 Fishhoff 本人在开发 A-DMC 时考虑到决策能力本身应该包含决策能力元认知成分的存在并设计了决策能力元认知测量的维度，但是忽视了特殊决策技能的测量本身就存在决策能力的元认知，因而对元认知的测量存在较大误差。当然，这与 Fishhoff 开发决策能力测验是基于 CTT 分不开的，虽然 CTT 中被试总分的估计能对被试决策能力的总体状况有一个基本的认识，但是对被试的决策能力结构缺乏足够的认识。基于 IRT 的 Bifactor 模型能实现决策能力元认知与特殊决策能力的有效分离，研究者可以对决策能力的结构有更明确的认识，对于人们决策能力评价有一个更为清晰的评估策略。当然也是有必要将具体决策技能和决策能力元认知进行分离，从而对被试决策能力结构有一个更为科学的测量和明确的认识。

4.2 公共因子掩盖不同决策技能的潜在关系

提取公共因子前后，不同决策技能的相关关系发生较大的变化。这一结果很大程度上是因为决策元认知公共因子的存在掩盖了不同决策技能之间潜在的关系。在 CTT 中，能反应观测变量之间的外在关系，但是对心理变量间潜在关系揭示却不够。本研究同 Fishhoff 的研究结果类似^[4]，不同决策技能在 CTT 中存在不同程度的相关。Bifactor 模型能更好的揭示不同决策技能的潜在结构，其中最明显的例子是运用决策规则与抵御成本沉没、抵御框架效应和风险感知一致性在 CTT 中有显著相关，但是在 Bifactor 模型中抽取公共因子之后，运用决策规则和上述三者之间的相关不再显著，也就是说他们之间的相关是因为公共因子——决策能力元认知的存在。从测验

目的和内容来看,运用决策规则目的在于考察被试能否正确运用决策规则做出合理的判断和选择,而抵御成本沉没目的在于考察被试在当前决策时不考虑过去投资,即排除过去投资对当前决策的影响。抵御框架效应考察的是决策者不受决策问题描述的形式对决策的影响的能力;风险感知一致性考察的是被试对风险概率规律的把握情况。从理论上来看,他们之间理应没有直接关系。可见,Bifactor 模型下揭示的关系更为合理,公共因子的存在混淆了他们之间真实的潜在关系,我们的其他同类研究中(如不同应对方式)也发现了同样类似的现象^[137]。同样的道理是抵御成本沉没和抵御框架效应两者相关系数的变化。同 Fishhoff 报告的一样,CTT 中,抵御成本沉没和抵御框架效应没有显著关系。需要指出的是,因为两类题型计分方式的不同,抵御框架效应需要反向计分,虽然在 Bifactor 模型中数据上是负相关关系,事实上两者应该是显著正相关关系。依据卡尼曼的前景理论对框架效应的解释^[138],框架效应的出现是因为损失和获益的价值曲线不同,即同等程度的损益因为不同的描述方式对被试产生的心理价值是不同的,从而出现框架效应。现有理论对成本沉没的解释,不论是前景理论还是“消极偏向”的观点^[139],都认为成本沉没现象的出现是因为个体在决策过程中,消极信息的作用要大于同等强度的积极信息。过去投资的损失对被试来说是一个消极消息。可见,被试在有效抵御这两类心理偏差过程中,存在相同的认知过程、需要调用类似的相关认知资源,因而这两类决策技能是存在一定程度的相关的。Bifactor 模型中通过提取决策能力公共因子,更明确的揭示了二者的关系。综上所述,将决策能力公共因子和特殊因子分离,可以更明确的揭示不同决策技能之间的关系。在 Bifactor 模型中,剥离公共因子在决策中的作用,单纯研究不同决策技能与其他心理变量的关系或探讨决策技能的作用也有一定的积极意义。

4.3 统计计算能力是规范决策中的核心因素之一

通过全息项目 Bifactor 因素分析的方法,对决策能力的心理结构有更为清楚的认识,同时也为研究决策能力和统计计算能力的关系积累了基础。统计计算能力和决策能力的相关分析提示统计计算能力是决策能力的核心因素之一。这一论断的证据是统计计算能力在决策能力公共因子上的相关要显著高于统计计算能力与决策特殊能力之间的相关,说明统计计算能力和决策能力之间的相关更多的是在其核心特征

之上。为此，可以推测统计计算能力是决策能力的核心因素之一。这与前述研究者应用 Bifactor 模型验证正性情感是外倾性的本质特征的推理过程是一致的。

决策心理学家通过不同决策任务与数学能力的关系研究提出数学能力是影响决策的核心因素^[6]，比如，高数学能力意味着较小受框架效应的影响^[140]、非数字信息的影响更少（比如情绪信息）^[141]、对数字所蕴含的风险更为敏感等等^[142, 143]。本研究也从一定程度上证明了这一观点。因为统计计算能力与数学能力存在共有的认知成分。研究还发现^[60]，高认知能力的被试在任务的准备上倾向于花更多的时间，对信息的编码和加工付出更多的努力，从而建立更有利于问题解决的认知表征。而高统计计算能力被试在考虑决策概率情况和对不同的选项进行比较的时候，能启动更多的认知努力，在决策过程中启动更细致、完整的认知过程，进而发生更少的决策偏差^[144]。应用 CRT 研究认知反思在决策中的作用的相关研究同样证明，反思能力高的个体因为他们对启发系统的监控和驾驭能力更强^[114]，因而他们的决策更接近规范决策^[15, 145, 146]。CRT 测验中所测量的认知结构同数学能力一样，也是规范决策的预测因子。而统计计算能力作为两者公共成分，依据双加工理论的观点，统计计算能力不但参与系统 1 的决策认知过程中，还参与系统 2 的认知过程。Edward T. Cokely 等人研究发现^[36]，决策元认知能力能对许多决策问题启动更多的知识经验，因而能更好的避免知识局限性和冲动决策。本研究发现高统计计算能力和决策能力公共因子的相关要大于其与特殊决策能力之间的相关，说明统计计算能力在决策能力中的作用主要是其核心因素上，而非特殊决策技能。从而证实统计计算能力是决策能力的核心因素之一。

也有研究指出，高数学能力的被试能更好的通过分析系统对启发系统的反应进行监控从而在决策任务中的成绩更好^[11, 34, 147]。而统计计算能力与决策能力元认知的高相关，同样证明高统计计算能力个体对系统 1 进行监控和驾驭的能力。同时，本研究还发现，与数学能力不同，统计计算能力和决策能力特殊因子还存在一定程度的相关，说明统计计算能力不但在核心因素上，在特殊决策技能上也是有一定的促进作用。

第三部分 统计计算能力对决策任务中数字信息把控的 认知优势研究

本研究第二部分采用全息因素分析的方法发现统计计算能力是决策能力的核心因素之一。其认知优势如何还不得而知。为此，本部分拟探讨不同统计计算能力被试对数字权重规律和核心数字把握情况，初步探讨不同统计计算能力被试对数字信息把控的认知优势。

1 问题的提出

研究发现，人们在做概率判断时，往往过分关注分子，甚至有时候忽略分母。比如，被试认为 10/13 的概率比 9/11 的概率中奖的可能性要大^[28]。再比如，被试认为患某癌症的病人中，10000 人中有 1286 人（12.86%）死亡，比 100 个人中有 24.14（24.14%）的人死亡的危险性要大^[148]。类似的问题是虽然同一个比率，被试似乎认为大数字（100 个中有 10 个）的比率比小数字（如 10 个中有 1 个）的比率要大^[149]。模糊痕迹理论认为，人们主要依赖于整体信息做出判断和决策。在面对概率和风险判断时，需要同时从分子和分母中提取整体信息，这就显得有些困难了。因为分子中的分类（如某手术的存活率）是包含在分母（做某手术的总人数）中的，这容易产生一定的混淆，因而，人们往往只关注比率中明显的整体信息，或者说分子，从而忽略分母^[150]。比如，多少人在某手术中存活了，这是与被试关心的结果息息相关的信息，但是忽略了到底有多少人做了这个手术。换句话说，人们容易受表面数字信息的影响，对核心数字信息的把握出现了偏差。研究者将这一决策中的偏差称之为“忽略分母”偏差。与此偏差相类似的是“100%”效应^[151]。即人们更偏好 100% 概率信息，对其赋予的权重往往也越大。比如研究发现，相比于“对 100% 的病毒 70% 有效”的疫苗，被试更偏好“对 70% 的病毒 100% 有效”疫苗。而事实是两类疫苗的净有效率是一致的。人们似乎只关注 100% 有效，而忽略了只是对 70% 的病毒有效。对此类偏好的另一种解释是人们对百分数信息的权重是非线性的，在 100% 附近的是

一个凸面曲线，即对 100% 赋予更高的权重。Li and Chapman (2009, 2013)采用实验法证实了这一假设，这与期望理论提出的人们对概率信息的权重是倒 S 型曲线有异曲同工之妙^[12, 28]。

同时决策研究发现，风险决策中概率数字信息的运用能力存在个体差异。Gurmankin, Baron 和 Armstrong (2004)发现高数学能力的被试更信任数字性的风险信息^[120]，而低数学能力的被试更偏好文字描述的风险信息。Peters 等人 (2006)发现^[5, 34]，概率信息呈现方式的不同对高数学能力被试的影响相对较小，高数学能力的被试更善于提取和利用合适的数学原理，从而在包含数字信息的决策情景中更少的出现框架效应。Chapman 和 Liu (2009)采用贝叶斯推理决策任务同样发现，高数学能力被试对包含计算任务的决策中的成绩更好^[12]。Stanovich 和 West发现^[152]，高认知能力的被试出现分母忽略的可能性要小。统计计算能力作为决策能力的核心因素之一，本研究认为统计计算能力不同的被试对数字信息加工的权重规律及核心数字信息的把握上都会存在差异，因而在“100%效应”和“分母忽略”中高统计计算能力的被试在涉及此类偏差的决策任务中成绩也会更好，由此挖掘统计计算能力高的被试在决策中对数字信息把控的认知优势。为此，本部分设计如下实验条件：A 市基础雾霾天数 20 天，B 市基础雾霾天数是 10 天，那么 A 市雾霾天数增长（或下降）10%和 B 市增长（或下降）20%，其实质雾霾天数变化是一致的。由此设计不同增长百分比条件下，核心数字信息变化相同的实验条件。以此考察不同统计计算能力被试如何受表面百分比的变化的影响，以及人们是如何对百分比赋予权重的和对核心数字信息的把握能力。同时提出以下几个假设：

（1）不同统计计算能力被试百分比权重规律不同，高统计计算能力被试的权重规律更趋于线性；

（2）人们普遍受表面比率变化的影响，对实际雾霾天数的变化关注较少，但是高统计计算能力被试受到表面信息的影响相对较少；

（3）不同统计计算能力被试对核心数字信息的把握程度不同，高统计计算能力被试能更好的应用核心数字信息进行决策。

2 方法

2.1 被试

被试为参加航天员决策能力选拔预研项目的被试 248 人。为尽可能反应被试决策能力和统计计算能力的全距,本研究整群抽取三类人群,工程师、重点大学生和基层官兵。涵盖从小学文化程度到重点大学生不同学历层次的被试群体,年龄 18~25 (21.46 ± 2.67) 岁,其中初中及以下文化程度 42 人,占 16.93%;高中 86 人,占 34.67%;大学及以上文化程度 174 人,占 48.38%。所有被试了解本研究的研究背景、目的和意义,并给予一定的实验报酬,收回有效数据 243 份。

2.2 实验材料

实验任务为请被试不考虑其他因素的影响,随着 A、B 两市雾霾天数的变化,评价自己到某市工作的意愿,采用 likert 1-8 点评分,其中 1 为非常不愿意,8 为非常愿意。其中 A 市的基础雾霾天数是每 100 天中有 20 天,B 市的基础雾霾天数是 100 天中有 10 天。并以柱状图展示两市雾霾天数的比率及其差异(实验材料见附录)。然后假设 A、B 两市的雾霾天数分别提高(或下降)25%、40%、50%、80%、95%和 100%。A 市雾霾天数变化的 25%、40%和 50%与对应 B 市的 50%、80%和 100%相等,分别是 5 天、8 天和 10 天。这样,雾霾天数的相对比率变化嵌套于两种基础比率(大比率和小比率)之中。本实验采用被试内设计重点考察的是 3 种雾霾天数的绝对变化 \times 两类比率(大比率和小比率)6 种条件下,被试到该市工作的意愿(其他条件相同),考察被试对空气质量百分比变化的感知权重,考察不同条件下数字信息权重对被试决策的影响。需要说明的是,实验材料中,除上述提及的 A、B 两市的百分比数据用于分析以外,其他数据如 A 市 25%、80%仅用于计算被试对百分比数字权重规律的计算,其真正目的在于掩盖本实验的真实目的,以便更准确的探讨被试的评价是基于字面数字信息还是核心数字信息。

2.3 实验设计和程序

本实验采用 2(大比例和小比例) \times 2(高、低统计计算能力) \times 3(雾霾天数绝对变化 5 天、8 天和 10 天)3 因素混合实验设计,其中基础比例和雾霾天数变化为被试内因素,统计计算能力为被试间因素。被试分别评价两类比例条件下不同雾霾天

数变化去不同城市工作的意愿。而后完成统计计算能力测量试题,采用第一部分 FIFA 方法估计被试统计计算能力,同时将被试以平均数为界划分为高统计计算能力组和低统计计算能力组。

2.4 统计计算能力测量

同第一部分。

2.5 数据处理

采用 POLYBIF 软件包计算被试统计计算能力分数,采用 SPSS 16.0 for windows 运用重复测量分析方法进行数据分析。

3 结果

3.1 不同统计计算能力对百分比变化的权重

本研究首先采用重复测量分析方法比较了不同统计计算能力被试对百分比变化的权重规律。其中, A 市和 B 市对同一百分比的工作意愿的评价取其平均值,百分数作为被试内自变量,不同统计计算能力作为被试间变量,被试工作意愿的评价作为因变量。不同百分比条件下高、低统计计算能力被试权重的描述统计结果见 Table 10。从总体上看,百分数越大被试的给予的权重也越高。重复测量方差分析发现,不同百分比增长主效应显著 ($F_{1, 243}=19.81, p < 0.0001$, 偏 $\eta^2=0.203$),说明百分比变化不同,被试到该市工作的意愿也不同,即对不同百分比数字的权重不同。统计计算能力主效应不显著 ($F_{1, 243}=1.35, p = 0.249$, 偏 $\eta^2=0.017$),统计计算能力与百分比变化交互作用显著 ($F_{1, 243}=6.32, p = 0.014$, 偏 $\eta^2=0.075$),说明不同统计计算能力被试在不同百分比变化条件下,其权重存在差异。从 Fig. 5 可以进一步发现,高统计计算能力被试对 25%-100%变化的百分数变化权重更趋于线性,而低统计计算能力被试的百分数变化权重无明显规律性。初步验证了本实验提出的研究假设 1。

Table 10 the Weighting of Percentage for Different Statistical Numeracy

Statistical Numeracy	25%	40%	50%	80%	95%	100%
High	3.93±2.01	4.32±2.18	5.12±2.04	5.83±2.33	6.15±2.13	6.41±2.73
Low	4.33±1.58	4.79±1.34	4.79±1.64	4.85±2.07	5.18±2.33	5.13±2.66

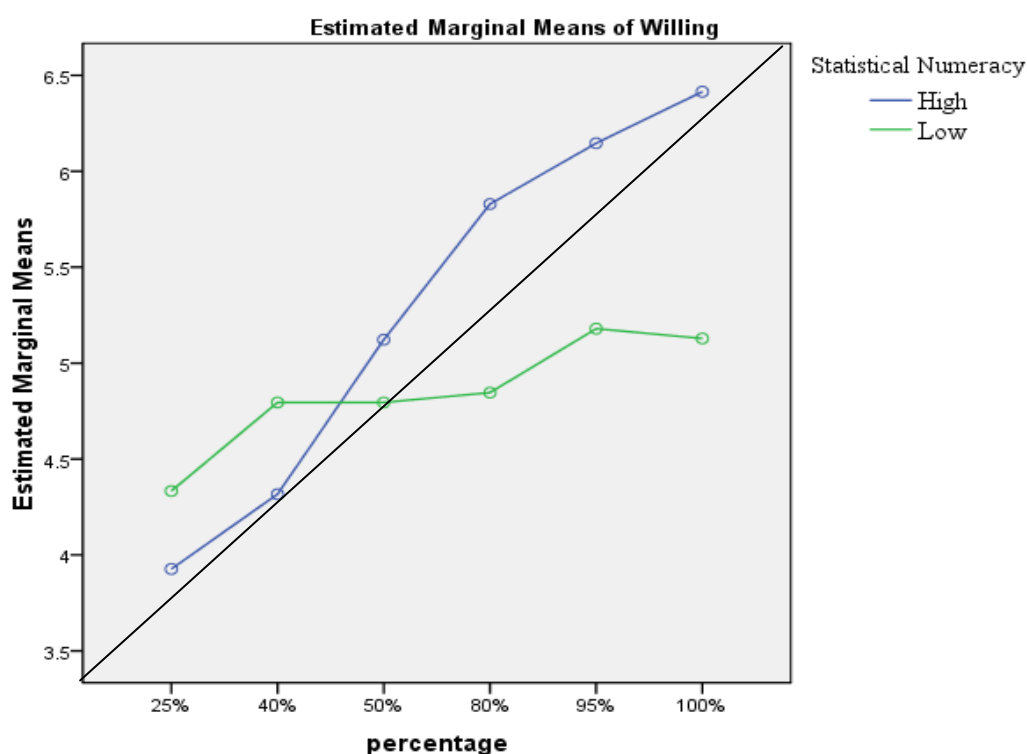


Fig.5 the Weighting Curve of Percentage for Different Statistical Numeracy

3.2 不同统计计算能力对核心数字变化把握能力比较

本研究采用重复测量分析方法比较了不同统计计算能力被试高、低基础比例条件下核心雾霾天数实际变化 5 天、8 天和 10 天 6 种条件下被试到某城市工作的意愿，考察被试对核心数字的把握能力。其中，雾霾天数增加条件时的被试工作意愿反向计分后与降低时的工作意愿取平均数，三种雾霾天数实际变化和高低基础比例作为被试内变量，被试工作意愿的评价作为因变量。不同条件下被试的工作意愿描述统计结果见 Table 11。

重复测量方差分析发现（结果见 Table 12），雾霾天数主效应显著（ $F_{2, 243}=128.25$, $p < 0.001$, 偏 $\eta^2=0.347$ ），雾霾天数和基础比例（ $F_{2, 243}=0.16$, $p < 0.85$, 偏 $\eta^2=0.001$ ）、统计计算能力（ $F_{2, 243}=1.42$, $p = 0.32$, 偏 $\eta^2=0.005$ ）交互作用不显著，说明空气质量不同，被试到该市工作的意愿也不同。统计计算能力主效应显著（ $F_{1, 243}=323.17$, $p < 0.001$, 偏 $\eta^2=0.931$ ），基础比例主效应显著（ $F_{1, 243}=122.88$, $p < 0.001$, 偏 $\eta^2=0.338$ ）同时统计计算能力与不同基础比例交互作用显著（ $F_{1, 243}=17.63$, $p < 0.001$, 偏 $\eta^2=0.068$ ），说明不同统计计算能力被试在不同基础比例条件下，其工作意愿存在显

著差异。

Table 11 the Weighting of Core Number for Difference Statistical Numeracy in Difference Base Rate

Statistical Numeracy	Base rate	5 days			8 days			10 days		
		N	\bar{X}	$\pm SD$	N	\bar{X}	$\pm SD$	N	\bar{X}	$\pm SD$
High	Big	132	4.28	± 2.35	132	4.83	± 2.36	132	5.34	± 2.32
	Small	132	5.92	± 1.72	132	6.77	± 1.50	132	7.36	± 1.59
Low	Big	111	4.13	± 1.62	111	4.76	± 1.73	111	5.28	± 1.72
	Small	111	5.10	± 1.83	111	5.57	± 2.19	111	6.02	± 2.38

Table 12 ANOVA of Repeated Measurement in Different Condition of Core Number, Base Rate and Statistical Numeracy

Source of Variation	F	df	<i>p</i>	Effect Size (Partial η^2)
Statistical Numeracy (S)	323.17	1	<0.001	0.931
Core Numerber (D)	128.25	2	<0.001	0.347
Base Rate (R)	122.88	1	<0.001	0.338
D×R	0.163	2	0.85	0.001
D×S	1.42	2	0.32	0.005
R×S	17.63	1	<0.001	0.068
D×R×S	2.67	2	0.07	0.11

从 Fig. 6 可以进一步发现交互作用差异来源,高统计计算能力被试在不同基础比例条件下(即不同表面数字)对同一核心数字的差异不显著,而低统计计算能力被试的在不同基础比例条件下同一核心数字差异显著。初步说明高统计计算能力被试不容易受表面数字的干扰,对核心数字在不同表面数字信息中能有比较一致的把握。而低统计计算能力被试受到表面数字大小(即百分比数字大小)的影响,对同一核心数字的把握评价出现显著差异,验证了本实验提出的研究假设 2 和 3。

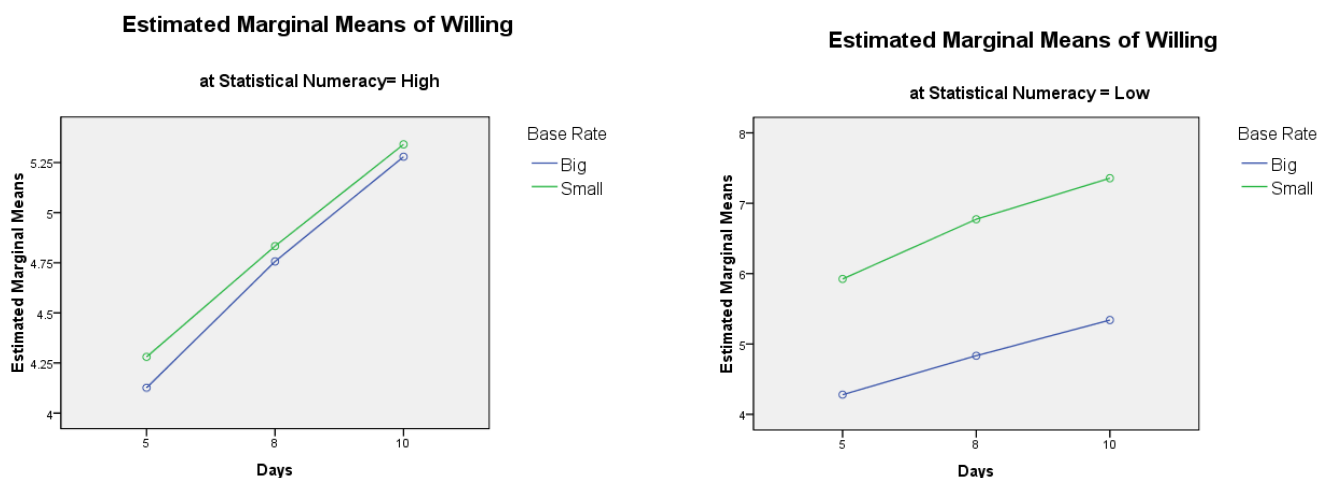


Fig. 6 Difference Working Willing for Different Statistical Numeracy under Different Base Rate

4 讨论

随着城市的高速发展，空气污染、雾霾等环境问题受到人们越来越多的关注。本实验以人们关注的热点环境问题为素材，编制虚拟城市雾霾天数的相对比率变化场景，研究统计计算能力的认知加工优势。本研究初步发现，不同统计计算能力的被试在百分数权重规律方面存在显著差异，高统计计算能力被试的百分比数字权重更趋于线性。低统计计算能力被试对百分比数字权重无明显线性规律。重复测量方差分析进一步发现，高统计计算能力被试不容易受到表面数字的影响，对同一核心数字赋予的权重基本一致，而低统计计算能力被试则更容易受到表面数字的影响，对不同基础比例条件下相同数字赋予不同的权重。以往研究发现，人们容易受到表面数字的影响，从而忽视核心数字信息在决策中的作用^[153]。本研究进一步发现，表面数字对决策的影响也是存在个体差异的，不同统计计算能力的被试在核心数字能力的把握上存在差异。

提取相关数字信息进行决策是数学能力的基本技能。以往研究大多将注意集中于人们提取不同类别的数字信息对规范决策的影响^[22, 153]。事实上研究数字信息在决策中的作用，除了考虑人们提取不同数字信息对决策的影响外，还需要考虑数字信息的权重在决策过程中的作用。因为规范决策的做出，不但要提取决策中的核心数字信息，还要赋予其合适的权重。为此，本研究从核心数字的提取和数字信息权重

两个方面考量不同统计计算能力被试在决策中数字信息利用中的差异，挖掘高统计计算能力被试在决策中的认知优势所在。

4.1 高统计计算能力被试的数字信息权重线性加工优势

以往研究更多的是探讨数字信息本身对人们数字信息权重感知的影响，对数字信息的加工对决策的影响研究较少。Michael Ranney 等认为人们对数字信息的权重主要依赖于认知和表征线索^[154]，并提出影响数字信息权重的 5 类线索：信息的可接近性、信息与期望的差异程度、信息获取的方式、数字信息表征的精确性、信息不确定性的感知。研究者认为，数学能力是感知以上 5 类认知和表征线索的基本技能^[155, 156]。研究者同时发现不同数学能力的被试在表征线索的感知上存在差异，从而在数字信息的权重方面存在差异。事实上，本研究是对以往研究的拓展。本研究重点考察统计计算能力不同的被试认知加工过程本身。统计计算能力作为数学能力的重要部分，本研究也发现不同统计计算能力的被试在数字信息的权重方面存在差异。本研究进一步发现，高统计计算能力被试对数字信息的权重更符合线性规律，而低统计计算能力被试对数字信息的权重感知无明显线性关系，从而说明不同统计计算能力的被试在数字信息的权重规律方面表现出差异。由此，可以推断，高统计计算能力的被试在数字信息的感知方面更为理性和客观，因而也能更多的做出规范和理性决策。由此，本研究提出高统计计算能力被试的认知优势之一，对数字信息权重感知的线性加工。

4.2 不同统计计算能力被试对核心数字把握能力

本研究发现，核心数字相同时，相比较于大概率中的小数字，低统计计算能力被试偏好小基础比例中的大数字。这与 Li 和 Chapman (2009) 在同类任务中的发现一致^[28]。但是高统计计算能力的被试受到表面数字信息干扰的影响更少，表现为对核心数字的准确把握能力。有研究者指出，人们发生“相比较于大鱼缸中鱼，更偏好小鱼缸中的同样大小的鱼”这一认知偏见的原因之一是人们在数字处理能力方面的欠缺^[28]。如果这一假设成立，数学能力差的被试在此类任务中的认知偏差将要大于数学能力高的被试。研究却发现，数学能力与此类认知偏差没有明显关系^[145]。与此研究结果矛盾的是研究发现学业成绩（包含数学成绩）是抵御“分母忽略”偏差的重要预测变量^[157]。事实上，本研究中设计的实验任务与“分母忽略”、比率偏差

是类似的认知偏差任务。稍微有不同的是“分母忽略”任务中考察的是被试在计算百分比时是不是忽略分母的作用，而本研究中的任务是考察被试在计算核心数字时是否考虑基础比率的作用。人们可能会忽视基础比例的作用从而发生与“分母忽略”相类似的认知偏差。

需要承认的是数学能力测量和学业成绩测量所考察的认知能力有所不同。对此矛盾的结果可以从两个方面理解，一是“分母忽略”认知偏差和本研究中的认知偏差可能是两类不同的认知偏差，前者可能是被试根本就没有注意分母的信息，从而在核心数字上的把握上出现偏差^[51, 158]；而后者需要被试同时考虑基础比例和百分比变化情况，低统计计算能力的被试在两类基础比例数字信息或百分比变化赋予的权重不同（见 Fig. 5），从而使不同统计计算能力的被试表现为核心数字的把握上出现差异，可见高统计计算能力被试数字信息权重线性加工的认知优势也表现在核心数字信息的把握上的作用。另一种可能的解释是，数学能力所测查的内容和范围相比于统计计算能力包含了更多的混杂因素（如数学概念的理解、数学运算规则等）^[11]，使其规范决策中的预测能力下降，数学能力上的差异不足以区分被试在此类任务中的差异。而统计计算能力作为决策能力的核心因素之一，在决策任务中的预测效度更高，事实上，相比较于“分母忽略”任务，本研究设计的实验任务更不容易察觉核心数字任务。高统计计算能力的被试更善于计算核心数字信息，从而表现出其认知优势。

第四部分 统计计算能力对决策任务中启发性信息加工 的认知优势

1 问题提出

Kahneman 指出，人们往往容易基于过去的经验和直觉形成启发式表征做出判断和决策，而不是总是依据逻辑或推理过程^[159]。在过往的数十年中，大量研究证实这一倾向使被试在许多经典推理和决策任务中的成绩受到影响。即便如此，基于过去经验建立的启发式表征并非全是带来如标准理论所预测的非标准决策行为。如 Frederick 等人设计的实验任务中请被试在肯定获得 100\$ 和 75% 的概率获得 \$200 中做出选择，期望理论预测被试会选择风险规避从而选择肯定获得 100\$^[113]。然而，认知反思测验（CRT）得分高的被试中，即能建立合适的启发表征的被试选择了后者，而低分被试却选择了前者。可见，基于已有的经验建立适应当前任务的启发式表征不仅可以节约认知计算资源，对规范决策也是有益的^[160]。启发式表征同样能为判断和决策提供高质量、有效的解决方案，虽然难免发生一定的认知偏差。因而采用启发与偏差任务（Heuristics and Biases Tasks）考察被试是否能够避免决策偏差是决策能力强弱的重要表现^[152]。

也有研究者认为决策的个体差异不一定发生于规范决策模型中分析系统的期望-价值高级运算过程，而是因为启发式表征或是对自动加工过程的监控和驾驭能力的差异造成的。因为分析系统（系统 2）和自动加工系统（系统 1）的算法不同，后者通过将复杂任务的概率评估和价值判断简化为简单运算，从而将复杂问题简化，且的在于节约认知资源、更快的解决问题^[161]。可见，两者在问题解决中的问题表征存在必然的联系。个体差异使不同的人在系统 1 中形成的启发式表征不同。正如 Kahneman 所言，来自启发式表征的错误判断也不是随机的，而是系统性的^[162]。不同认知能力形成的启发式表征对决策问题的解决可能是完全不同的结果^[163]。可见，采用启发和偏差任务评价个体认知偏见的易感性，可以研究不同类型个体的认知优

势。综上所述，本研究有充分的理由认为，统计计算能力的个体差异在决策中的认知优势不完全发生于计算能力方面的差异，在自动加工的系统 1 中已经形成有利于决策问题的启发式表征^[164]。因而，本研究假设统计计算能力的认知优势还存在于建立适应当前决策任务的优势启发式表征，对系统 1 更好的监控和控制能力。为此，本研究拟采用启发和偏差任务（如忽视基础概率、观念偏差任务等）进一步探索高统计计算能力被试在规范决策中的认知优势。并提出以下实验假设：

- （1）高统计计算能力被试有对基础概率信息进行准确表征的认知优势；
- （2）高统计计算能力被试有在启发式表征中融入独立事件统计规则的认知优势；
- （3）高统计计算能力被试有在启发式表征中融入条件概率规则的认知优势；
- （4）高统计计算能力被试对系统 1 有较好的监控和控制能力。

2 方法

2.1 被试

本研究采用方便取样法抽取两类人群，某重点大学学生和基层官兵共计 670 人。年龄 18~34 (22.46 ± 6.84) 岁，初中及以下文化程度 136 人，占 20.30%；高中 186 人，占 27.76%；大学及以上文化程度 348 人，占 51.94%。被试随机参与本研究中 5 类启发和偏差任务，全部被试需要完成统计计算能力的测验题目。所有被试了解本研究的研究背景、目的和意义，并给予一定的实验报酬，收回有效数据 664 份。

2.2 实验材料

本研究中，采用 4 类不同的启发与偏差任务（忽视基础概率、博弈谬误、协方差判断和信念偏差，见附录）考察不同统计计算能力在启发式加工的认知偏差情况。每类任务 4-8 题不等，分别计算每一类任务的正确题数，然后求和得出被试在启发与偏差任务中的综合成绩。下面对本研究中这四类任务进行简要介绍：

（1）忽视基础概率（Base Rate Neglect Problems）任务^[165]。该任务的标准范式是：每一个问题包含两类基本信息，一是基础概率数据，描述两类群体的基础比例情况；二是假设从这两类群体中随机选择一人的刻板印象文字描述信息（两类群体都有明显的特征，如男性和女性）。被试的任务是判断随机选择的这个人应该属于哪

个群体(如男性还是女性)。被试的正确作答应该是基于基础概率选择概率较大那类,而不是根据文字描述选择刻板印象的群体。如果文字描述信息是低比例人群刻板印象,称之为冲突情景。如果文字描述是高比例群体特征称之为非冲突情景。本实验中包含 8 个忽视基础概率问题,其中冲突情景 4 题,非冲突情景 4 题(其中两题文字信息为大概率群体特征,另两道题是与两类群体特征无关的中性描述。)

(2) 博弈谬误 (Gambler's Fallacy) 任务^[166, 167]。经典博弈谬误任务通常采用以下形式:假设随机抛硬币 N 次,前三次都是正面,请被试判断第四次是正面还是反面。如果被试认为第四次是反面的可能性会更大,则认为博弈谬误现象发生,因为第四次中正面和反面出现的概率依然是相等的。Tversky 和 Kahneman 认为博弈谬误是因为启发式认知表征偏差导致的^[162, 168]。本研究设计了 4 个与上述赌徒谬误任务类似的问题。依据标准决策理论范式,如果被试认为目标任务(如第四次抛硬币)的概率大于或小于标准概率,则认为被试发生了博弈谬误,同时在该题上不计分。反之则计 1 分。

(3) 协方差判断 (Covariation Judgment Problems) 任务^[169]。该类问题为被试提供一个四格表,请被试判断变量 A 对变量 B 是不是有效。四个表中依次呈现:实验组/有效(a)、实验组/无效(b)、控制组/有效(c)、控制组/无效(d)。变量 A 对变量 B 的效果可以计算 Phi (ϕ)或是比较条件概率 $[a/(a+b)-c/(c+d)]$ 。当 a 的数值比 b 和 c 的数值显著要大的时候,被试往往容易运用简单的判断策略,即简单比较 a 和 b 或 c 数值大小就做出变量 A 是不是在变量 B 上有效的判断。在本研究中同样设计了冲突任务和非冲突任务,如果 a 的数值比 b 和 c 的数值差别非常明显,但是 Phi (ϕ)系数是负的(还不能充分说明 A 对 B 有效),即基于 a、b、c 三者之间数字的比较做出的判断与计算条件概率所得出的结论是不一致的,称之为冲突任务;而如果 a 的数值比 b 和 c 的数值差别不明显,并且 Phi (ϕ)系数是负的,定义为非冲突任务。考察不同统计计算能力被试在条件概率运用能力的差异。本研究中,两类任务都是无效的。请被试 6 点量表上,请被试判断 A 对 B 是不是有效,其中 1 表示倾向于无效,计 5 分;6 表示倾向于有效,计 0 分。被试评分越高,说明抵御此类偏差的能力越高。

(4) 信念偏差 (Belief Bias) 任务^[170]。研究者大多主要采用三段论推理任务研究信念偏差。本研究中设计了 8 个两类三段论推理任务,1) 结论在经验上是正确的,但是推理过程是错误的,4 题;2) 结论在经验上是错误的,但是推理过程是符

合逻辑的，4 题。实验任务是要求被试依据逻辑推理过程判断三段论是不是有效，对符合逻辑的答“是”计 1 分，对不符合逻辑的答“否”计 1 分。这两类任务推理过程中的有效性和实际经验都是冲突的，目的在于考察被试是否是依据知识经验进行判断还是运用分析推理过程进行判断。

2.3 实验设计

除博弈任务采用单因素实验设计外，其他三类任务采用 2 任务类别（冲突、非冲突） \times 2（高、低统计计算能力）两因素混合实验设计，其中任务类别为被试内变量，统计计算能力为被试间变量。

被试完成启发与偏差任务后，完成统计计算能力测量试题。同时将被试以高于平均数一个标准差定义为高统计计算能力组，低于平均数一个标准差定义为低统计计算能力组。

2.4 统计计算能力测量

统计计算能力的测量和计分方法同第一部分。

2.5 数据处理

采用 POLYBIF 软件包计算被试统计计算能力分数，SPSS 16.0 for windows 进行描述统计、t 检验和重复测量方差分析。

3 结果

3.1 不同统计计算能力对基础概率的把握

采用重复测量分析方法比较不同统计计算能力被试在忽略基础概率任务情景中的正确作答，考察被试对基础概率的把握能力。不同条件下被试的正确作答情况描述统计结果见 Table 13。重复测量方差分析发现（结果见 Table 14），任务类别主效应显著（ $F_{1, 155}=69.08$, $p < 0.001$, 偏 $\eta^2=0.311$ ），被试在非冲突类忽略基础概率任务中成绩好于冲突类任务。统计计算能力（ $F_{1, 155}=21.84$, $p < 0.001$, 偏 $\eta^2=0.144$ ）主效应显著，高统计计算能力被试成绩相对较好，任务类别和统计计算能力交互作用效应显著（ $F_{1, 155}=10.78$, $p = 0.01$, 偏 $\eta^2=0.066$ ），Fig.7 进一步揭示交互作用的差异来源，不同统计计算能力的被试在冲突类任务中成绩差异显著，而在非冲突类任务中两者差异不显著。冲突和非冲突任务根本差异在于，冲突类任务中被试只有依据

基础概率信息才能做出判断和决策。可见高统计计算能力被试能排除语言描述信息的干扰，抓住基础概率信息做出判断，而低统计计算能力被试更多的是依据语言描述信息做出判断而忽视了基础概率信息。由此推断，高统计计算能力的在启发与偏差任务中的认知优势之一是对基础概率信息的把握。

Table 13 Descriptive Statistics of Correct Items for Different Statistical Numeracy in Different Base Rate Neglect Problems

Statistical Numeracy	Conflict Task			Non-Conflict Task		
	N	\bar{X}	$\pm SD$	N	\bar{X}	$\pm SD$
High	64	2.94	± 1.47	64	3.61	± 0.63
Low	91	1.73	± 1.62	91	3.29	± 0.87

Table 14 ANOVA of Repeated Measurement in Different Condition of Task and Statistical Numeracy in Base Rate Neglect Problems

	Source of Variation	F	df	p	Effect Size (Partial η^2)
Between-Subject Factor	Statistical Numeracy (S)	21.84	1	<0.001	0.144
	Task Categories (TC)	69.08	1	<0.001	0.311
Whthin-Subject Factors	S \times TC	10.78	1	0.01	0.066

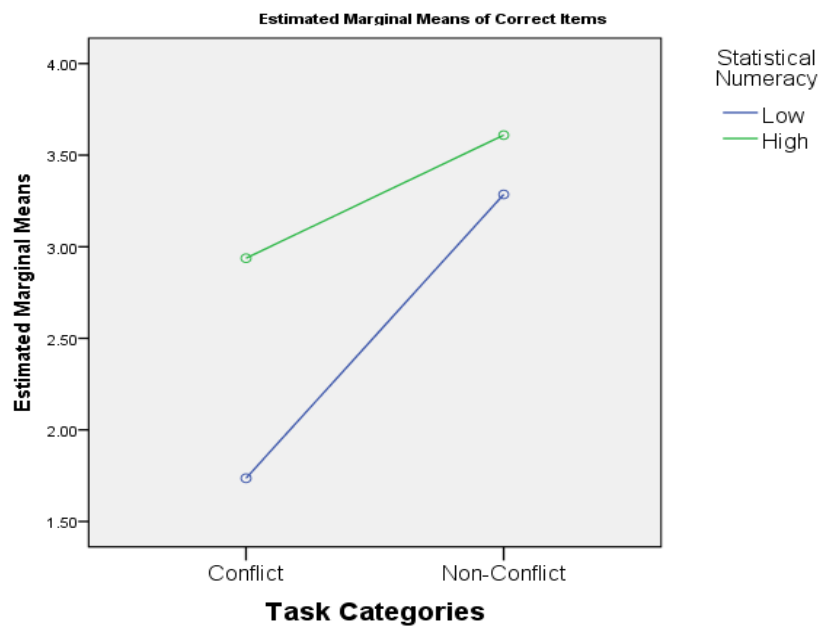


Fig.7 the Interaction between Statistical Numeracy and Task Categories of Base Rate Neglect Problems

3.2 不同统计计算能力独立事件统计规则把握的差异

独立样本 t 检验表明 (Table 15), 高统计计算能力的被试在博弈谬误中的成绩显著高于低统计计算能力被试。说明高统计计算能力被试更多应用独立事件统计规则做判断, 而不是依据小样本统计数据形成的即时表征做出判断和决策。

Table 15 Difference of Correct Items for Different Statistical Numeracy in Gambler's Fallacy Problems

Statistical Numeracy	N	\bar{X} \pm SD	t	p
High	60	3.82 \pm 0.50	4.03	<0.001
Low	94	3.17 \pm 1.21		

3.3 不同统计计算能力条件概率启发式表征的差异

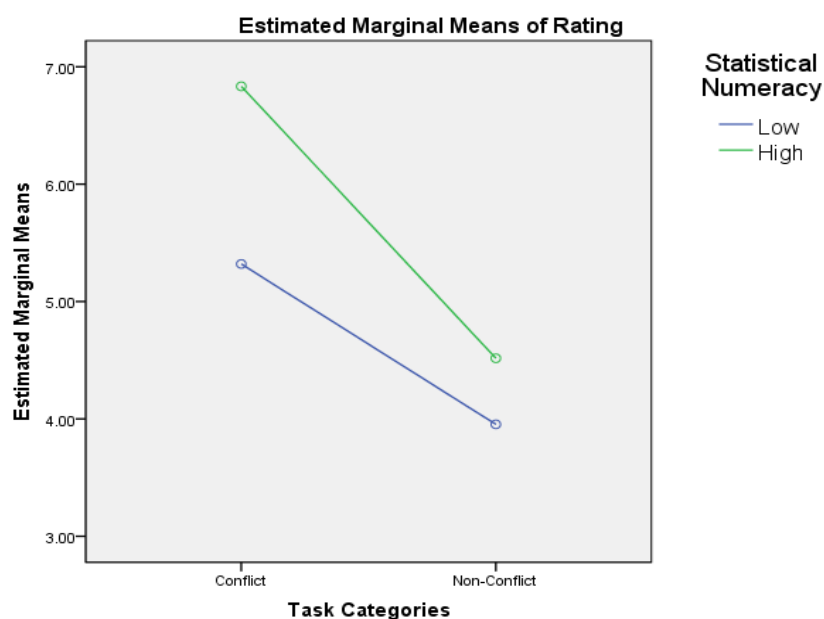
采用重复测量分析方法比较不同统计计算能力被试在协方差判断任务中的作答情况, 考察被试对启发式表征中条件概率统计规则的掌控能力。被试在冲突和非冲突任务中的得分情况描述统计结果如 Table 16 所示。重复测量方差分析发现 (Table 17), 任务类别主效应显著 ($F_{1, 154}=77.69$, $p < 0.001$, 偏 $\eta^2=0.236$), 被试在冲突类协方差判断任务中得分高于非冲突类任务。统计计算能力 ($F_{1, 154}=12.02$, $p < 0.001$, 偏 $\eta^2=0.046$) 主效应显著, 可见高统计计算能力协方差判断任务中的得分更高, 任务类别和统计计算能力交互作用效应显著 ($F_{1, 154}=5.18$, $p = 0.02$, 偏 $\eta^2=0.020$), 图 8 进一步揭示统计计算能力不同的被试在此类任务中成绩差异主要来源于冲突类任务, 而在非冲突类任务中两者没有显著差异。两类任务根本差异在于冲突类任务需要被试依据条件概率信息做出判断和决策, 可见高统计计算能力被试能运用条件概率信息做出推理和判断, 而不是表面数字信息; 低统计计算能力被试在冲突任务中对条件概率信息把握不够, 因而在此类任务中的得分较低。可见, 对条件概率统计规则的把握也是高统计计算能力的认知优势之一。

Table 16 Descriptive Statistics of Correct Items for Different Statistical Numeracy in Covariation Judgment Problems

Statistical Numeracy	Conflict Task		Non-Conflict Task	
	N	\bar{X} \pm SD	N	\bar{X} \pm SD
High	60	6.83 \pm 1.71	60	4.52 \pm 2.41
Low	94	5.32 \pm 2.63	94	3.95 \pm 2.52

Table 17 ANOVA of Repeated Measurement in Different Condition of Task and Statistical Numeracy in Covariation Judgment Problems

	Source of Variation	F	df	<i>p</i>	Effect Size (Partial η^2)
Between-Subject					
Factor	Statistical Numeracy (S)	12.02	1	0.001	0.046
Whthin-Subject					
Task Categories (TC)		77.69	1	<0.001	0.236
Factors	S×TC	5.18	1	0.02	0.020

**Fig.8 the Interaction between Statistical Numeracy and Task Categories of Covariation Judgment Problems**

3.4 不同统计计算能力在信念偏差任务中的差异

采用重复测量分析方法考察不同统计计算能力被试对信念偏差的驾驭能力，结果发现(见 Table18、19)，任务类别和统计计算能力交互作用效应不显著($F_{1, 154}=0.30$, $p=0.58$, 偏 $\eta^2=0.001$)，任务类别主效应显著($F_{1, 154}=5.15$, $p < 0.05$, 偏 $\eta^2=0.236$)，统计计算能力($F_{1, 154}=89.72$, $p < 0.001$, 偏 $\eta^2=0.263$)主效应显著，可见不论是结论不符合常理而逻辑推理正确的三段推理任务，还是结论符合常理但逻辑推理过程不正确的冲突的任务，高统计计能力被试的得分都要更高，Fig. 9 进一步发现这一现象。说明高统计计算能力被试更多的是应用逻辑推理过程进行判断，而不是依据已有的知识经验或固有观念进行推理判断。

Table 18 Descriptive Statistics of Correct Items for Different Statistical Numeracy in Belief Bias Task

Statistical Numeracy	Conclusions conform to common sense but logical is wrong			Conclusion don't conform to common sense but logical is right		
	N	\bar{X}	$\pm SD$	N	\bar{X}	$\pm SD$
High	60	3.38	± 1.11	60	3.60	± 0.69
Low	94	1.77	± 1.60	94	2.13	± 1.41

Table 19 ANOVA of Repeated Measurement in Different Condition of Task and Statistical Numeracy in Belief Bias Task

	Source of Variation	F	df	p	Effect Size (Partial η^2)
Between-Subject Factor	Statistical Numeracy (S)	89.72	1	0.001	0.263
Whthin-Subject Factors	Task Categories (TC)	5.15	1	0.024	0.236
	S \times TC	0.30	1	0.582	0.001

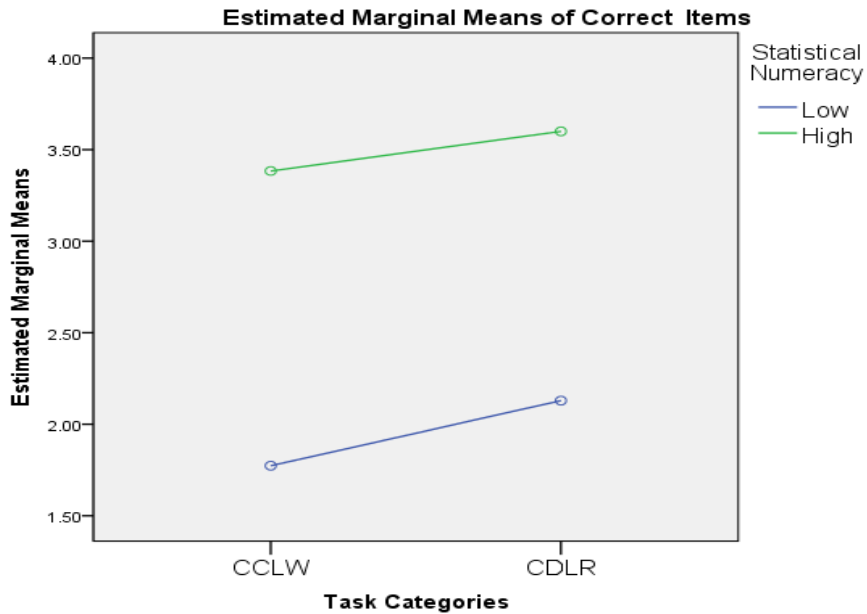


Fig.9 the Interaction between Statistical Numeracy and Task Categories of Belief Bias Task

Note: CCLW=Conclusions conform to common sense but logical is wrong; CDLR=Conclusion don't conform to common sense but logical is right

4 讨论

本实验通过 4 个小实验研究不同统计计算能力被试在启发和偏差任务中的成绩差异,考察高统计计算能力的在启发式加工过程中的认知优势。研究结果支持高统计计算能力被试能更好的避免启发式偏差。有研究证实优势启发过程表征 (priority heuristic process model)是认知能力和高级决策的中调节变量^[171]。这一启发式表征可以预测被试的决策行为。因而,可以推断高统计计算能力被试能更好的建立适应当前任务的启发式表征,具备一定的认知优势,从而避免决策偏差。

4.1 高统计计算能力对基础概率信息的较好把握

本实验中,采用经典忽略基础概率任务范式,编制冲突和非冲突情景,研究不同统计计算能力被试的启发式特征。这一实验范式可以直接检验被试对文字描述信息和基础概率信息所形成冲突的敏感性,而冲突情景和非冲突情景可以直接检验被试是否考虑基础概率信息。Kahneman 和 Evans 推测,被试甚至很少注意到基础比率的存在^[172],他们的判断和决策完全是依据文字描述信息,从而发生忽视基础概率偏差。本研究中,低统计计算能力的被试在非冲突情景中的成绩更好,似乎印证了 Kahneman 的推测,低统计计算能力的被试很少感知两类信息所带来的冲突性答案,只是依据文字描述信息做出推理和判断。同时本研究发现,高统计计算能力的被试在冲突情景中的成绩更好,说明他们的判断和决策不但感知到了基础概率信息的存在,同时更多的采纳了基础概率信息,从而能更好的避免忽略基础概率偏差^[173]。事实上,本研究第三部分还发现,高统计计算能力被试在数字信息的精确表征方面也有明显的优势 (Table 6),能形成与数字比例大体相当的认知表征图像。而低统计计算能力的被试在数字信息的图像表征方面似乎是模糊的,他们对饼图中阴影部分的数字表征的感知是模糊的。这样看来,即使他们感知到了基础概率信息的存在,对其认知加工也是模糊的,从而文字描述信息所推测的答案容易占主导^[174, 175]。可见,高统计计算能力被试能更好的感知两类信息的冲突,同时能对数字信息精确表征,从而更好的避免决策偏差^[176]。本研究的贡献在于发现忽略基础概率效应是存在个体差异的,高统计计算能力被试能更好的感知基础概率信息,感知不同信息源的冲突,同时对数字信息有精确表征,也体现其在启发式加工过程中的认知优势。

4.2 高统计计算能力对独立事件统计规则的把握能力

本研究采用博弈谬误任务发现,高统计计算能力被试在博弈谬误中的成绩更好,也就是说更少地发生博弈谬误效应,可见博弈谬误效应同样存在个体差异。Tversky 和 Kahneman 认为博弈谬误发生的原因是人们倾向于认为小样本的结果也要符合理论概率分布^[177]。可见,高统计计算能力的被试对统计规则的把握更为准确,而较少的依据小样本数据结果推论事件下次发生的概率^[178]。而低统计计算能力的被试更多的依据小样本概率信息形成的启发式表征做出判断,系统 1 的快速启动加工认知表征占优,而对统计规则的把握出现了偏差,容易在此类任务中犯错^[179]。事实上,依据贝叶斯定理(Bayes' theorem)此类任务被试也不需要通过复杂的计算^[12, 180],被试在启发式表征中只要能够认识到此类任务中的事件(如抛硬币)是独立事件而不是条件概率,就很容易做出理性判断和选择。虽然在本研究中所采用的任务是系列发生的概率事件,研究证实博弈谬误在横断维度也是普遍存在的,研究证实横断维度的博弈谬误的发生也是被试认为数据在某个范围中的分布应该是平均的^[181],被试不能认为他们是独立事件。事实上,被试在本研究中也无法进行复杂的概率推理计算。由此,我们可以进一步推断高统计计算能力的认知优势还在于对其启发式表征中更多的融入更多统计规则,对统计规则的把握更为准确。

4.3 高统计计算能力对条件概率统计规则的把握能力

本研究采用协方差判断任务发现,高统计计算能力被试能更好的应用条件概率推理做出判断和决策。从现有文献来看,研究者对被试在协方差判断任务中的成绩差异有不一致的结论,如有研究者认为数字的精确表征是影响该类任务正确与否的重要因素,也有研究者认为协方差任务的正确作答是多种认知规则综合作用的结果^[181]。同样有研究者认为,对除非被试掌握条件概率规则,否则不能在此类任务中正确作答。而比较一致的结论是被试必须要付出一定的认知努力对协方差判断任务进行认知加工才能在此类任务中取得成功,而其中最重要的是掌握协方差判断任务中的统计规则^[182]。本研究中所设计的实验任务,需要被试掌握一定的条件概率规则才能正确作答。高统计计算能力的被试在协方差判断任务中成绩更好,相比于低统计计算能力的被试,可见高统计计算能力的被试对条件概率规则有一定的掌握并融入其认知表征中。

也有研究者认为协方差判断冲突任务中,人们的第一反应(系统1)往往认为A对B是有效的,但是通过条件概率的计算得到的是另一个结论,系统1和系统2得出的结论是矛盾的,前者容易产生启发式偏差^[60, 175]。也就是说这类任务往往需要通过一定的条件概率计算才能得出正确的结论,因此也有研究认为该任务考察的是被试对系统1的控制和驾驭能力,即对第一反应(系统1)的反省能力^[54, 56]。当然,这是无法否认的,事实上这也是高统计计算能力的元认知优势之一,这一点在本研究中信念偏差任务中得到了证实。但是单凭对系统1的监控和驾驭能力还不足以胜任协方差任务,被试还需要进一步的条件概率计算。这也再次证明统计计算能力对条件概率规则掌握的认知优势。

4.4 高统计计算能力对系统1的控制和驾驭能力

本研究采用信念偏差任务发现,被试在已有知识经验和逻辑推理过程得出的结论是矛盾的三段论推理任务中,高统计计算能力被试成绩更好。本实验的真正目的不在于考察被试对三段推理规则的掌握情况。本实验中所有三段论推理中被试现有知识经验与逻辑推理过程是冲突的,目的在于考察被试是否能够监控和驾驭系统1的第一反应,第一反应会出现认知偏差,只有被试在控制第一反应后进行简单的逻辑推理过程才能得出正确的结论。虽然三段论推理规则的掌握会对被试的成绩产生一定的影响,但是如果单纯运用知识经验进行判断,其结论肯定是错误的。以往研究同样发现,人们倾向于依据已有的知识经验和固有观念判断演绎推理逻辑有效性^[163, 170]。尤其是三段论的结论(如猫是动物)在经验上是正确时,人们倾向于认为三段是逻辑有效的,而不管推理过程如何;而对结论是错误的倾向于判断逻辑推理是错误的。本研究中低统计计算能力的被试倾向于根据已有知识经验对逻辑推理过程进行推断,说明其倾向于依据自动加工系统的启发式表征做出判断和决策,从而出现了这一认知偏差。而高统计计算能力则能力对第一反应的启发式表征进行较好的控制和驾驭,削弱系统1在决策过程中的权重,增加分析系统的权重^[183],从而在这类任务中能有较好的表现,避免认知偏差的出现。由此,我们可以推断,高统计计算能力的认知优势也体现在元认知上,即对系统1表现较好的监控和驾驭能力,因而其在此类三段推理任务中能取得相对较好的成绩。这一实验也帮助我们认识到高统计计算能力被试在决策中的元认知优势。

第五部分 统计计算能力决策认知优势的心理机制

1 问题提出

如本研究第四部分所述，研究启发和偏差任务中个体认知偏见的易感性是研究个体认知优势的方法。而研究其易感性的发生条件，同样为探讨统计计算能力认知优势发生机制的创造了机会。为此，本研究拟沿袭第四部分的思路，采用启发和偏差任务为抓手研究统计计算能力的认知优势发生机制。

为了解释启发与偏差任务中被试为何容易出现各类认知偏差。Baron (2008) 提出了“有限理性”假设^[120]。他认为启发与偏差任务中的认知偏差是由于人们的理性思维受信息处理过程（如工作记忆，编码速度）限制导致^[183]。与该假设一致的研究结果是，智力一般因素（通常是信息处理过程有限性的指标）与许多启发与偏差任务成绩成正相关。智力测量在一定程度上也能反映认知能力的个体差异，这些研究结果一定程度上支持上述“有限理性”假设^[133]。由此看来，一般能力因素也是克服认知偏差的一个重要因素。然而，一般能力和启发与偏差任务只有中等程度的相关，启发与偏差任务成绩的绝大部分方差得不到合理的解释。研究显示，一般能力上的差异大约能解释启发与偏差任务中 20%左右的方差^[40, 133]。由此看来，除了一般能力，还有其他一些个体差异因素需要研究者进一步考虑和研究。**Spearman** 双因素智力理论自提出以来一直是经典理论，特殊能力同样是智力的重要组成部分。可见，除了一般能力还是需要考虑特殊能力在决策中的作用，如统计计算能力。本文采用全息 **Bifactor** 模型技术证实统计计算能力在决策中的核心作用，并采用实证研究提出统计计算能力在决策中的认知优势。可见，统计计算能力也是决策中的重要影响因素。广义上讲，统计计算能力是包含一系列特殊算法的微观策略，如个体对统计概念理解和掌握水平。如本文前述研究所指出的，启发与偏差任务需要有对统计概念和规则有最基本的理解。事实上，既往很多研究（包括本研究）发现，它可以解释启发与偏差任务中除一般能力以外的部分方差^[9, 152]。

为了更清楚的研究决策中的分析思维过程，研究者基于一般能力和思维特质提

出了一些理论模型。最常见的理论是双加工模型。该模型提出的两种不同的认知加工过程，即自动加工过程（系统 1）和分析性思维过程（系统 2）。Stanovich 进一步提出，分析思维过程由两类相关的思维过程构成^[146, 184]：反思思维（Reflective Mind）和算法思维（Algorithmic Mind），分别代表两种不同的分析思维水平。反思思维水平的操作运算，通常以认识论的理解和思维特质为指标，监控或管理算法思维水平的认知活动，因而也可以视为一种元认知^[128]。算法思维水平，通常以智力、一般认知能力、信息处理效率（如工作记忆）、推理能力（演绎或归纳推理）、运算和逻辑规则的策略和能力为指标^[134]。算法思维水平又可以进一步分为两个层次，一般能力水平和特殊能力水平。相对于特殊能力，一般认知能力处于更高层级水平。因为如果缺乏足够的一般认知资源，个体即算是有再高的特殊能力也无法充分发挥。如果针对具体的启发与偏差任务，个体即便是再高的特殊能力也会因为一般能力不足而出现判断或决策错误。因而，反思思维水平、一般运算思维水平和特殊运算思维水平三者之间的关系可以进一步做如下理解，一是沉思水平是评判自动加工中表征和线索的关键，决定是否启用分析思维过程，同时具备理解任务的认知需求、选择恰当的算法技能、监控运算操作和评估结果的功能，因而，在启发与偏见任务中做出正确的判断和决策首先需要有较高的沉思水平运算能力。二是尽管一般运算思维水平相对于反思思维水平处于低阶层级，可获得的认知资源也能影响反思思维水平绩效的发挥，因为反思思维也是需要基于一定的认知资源和能力水平才能有效发挥；三是，一般运算思维水平也会影响特殊运算思维水平绩效的发挥。如此一来，分析思维阶段的高统计计算能力认知优势的发挥依赖于高思维特质和高一般认知能力^[185]。因而，统计计算能力的认知优势是存在一定的发生条件的，统计计算能力、一般能力和思维特质三者在决策思维中的功能和作用如何值得研究者进一步考虑。

本研究试图对上述理论模型对统计计算能力在决策思维过程中的应用做初步验证，即同时考察思维特质、一般认知能力和统计计算能力三者在决策思维过程中的具体作用和相互关系，进而对统计计算能力认知优势的发挥机制进行探讨。综合考虑反思思维水平、一般能力水平和统计计算能力三者的关系：沉思思维水平和一般能力水平是决定在启发与偏差任务中是否采用统计计算能力生成规范决策的前提，可见二者是统计计算能力在启发与偏差任务中能否发挥认知优势的重要前提，二者同时主导统计计算能力水平的运算。因而本研究需要验证以下三个基本假设：

(1) 思维特质和一般能力对统计计算能力在启发与偏差任务中的认知优势具有调节作用;

(2) 思维特质、一般能力和统计计算能力同时影响被试在启发与偏差任务中的成绩。被试在三个预测因子都高时,在启发与偏差任务中的成绩最好,而在三者都低时成绩最差;

(3) 高思维特质和高一般能力会强化统计计算能力在启发和偏差任务中的认知优势,低思维特质和低一般能力会限制统计计算能力认知优势的发挥。

2 方法

2.1 被试

同前面研究一致,为尽可能反应被试统计计算能力的全距。本研究选取了两类人群,某重点大学学生和基层官兵共计 360 人。年龄 18~34 (22.46 ± 6.84) 岁,初中及以下文化程度 92 人,占 25.26%;高中 124 人,占 34.44%;大学及以上文化程度 144 人,占 44.00%。所有被试了解本研究的研究背景、目的和意义,并给予一定的实验报酬,收回有效数据 357 份,剔除无效数据 7 份,最终收回有效数据 350 份。

2.2 实验材料

2.2.1 思维特质 (Thinking Disposition, TD)

思维特质采用 Stanovich 等人开发的思维特质问卷测量被试反思思维水平^[185],本研究翻译为中文并依据中国文化特点进行适当的修订。该问卷共包含 5 个维度:1) 思维灵活性,共 10 题,用于评价被试是否愿意采用更多的方法解决问题,对问题或价值观是否持有包容态度。2) 反思 VS 直觉性,共 10 题,用于测量是运用逻辑思维还是依靠直觉主导决策的态度。3) 认知需求,18 题,用于测量被试对智力挑战、复杂思维和逻辑推理的喜好程度。4) 冲动决策,共 14 题,用于测量被试对冲动性决策(如不考虑决策的后果和更多可能性)的倾向性。5) 认知调节,共 8 题,用于考察被试对“冲突观念能通过可获得的最佳证据解决”这一概念的认识。所有题目都采用 Likert 1 (非常不同意) -6 (非常同意) 点评分量表计分。量表开发者在后期的数据处理中不计算各因子的因子分,而采用量表总分考察被试思维特质的情况,本研究亦借用此法。

2.2.2 一般能力

本研究采用标准瑞文智力测验作为被试一般能力的指标。测验介绍同第一部分。

2.2.3 统计计算能力

同第一部分。

2.2.4 启发与偏差任务

本研究中，采用 4 类不同的启发与偏差任务（忽视基础概率、大数法则、比率偏差和协方差判断），每类任务 2-4 题不等，以正确题数作为被试在启发与偏差任务中的综合成绩。其中忽视基础概率和协方差判断任务原理同第四部分，但具体测验题目有所差异。下面对本研究中大数法则和概率偏差任务做简要介绍：

（1）大数法则（Law of Large Numbers）^[186]（示例见附录）。这类问题需要被试阅读两方面的信息，一是大样本实证性数字中得出的结论性信息，另一类是小样本但是与被试实际生活经验更为接近、相对更为鲜活的直接经验中得出的结论信息。这两类信息分别支持不同的决策。要求被试依据这两类信息做出决策，用于考察被试是否会有选择偏向。在本测验中，依据大样本实证数据得出的结论是正确答案。

（2）概率偏差^[187]（示例见附录）。此类问题中，要求被试判断从以下两类样本中抽取目标物体的可能性更大，一是从分子相对比较大、同时分母也相对比较大（如 100 个球中，9 个是红球，91 个蓝球）但是概率相对较低的样本，二是从分子相对较小、分母相对也比较小的（如 10 个球中，1 个是红球，9 个蓝球）但是概率相对较大的样本。如果被试认为从分子和分母相对比较大的样本中抽取目标物体的概率更大时，则发生了概率偏差。本研究中，大分母大分子的样本中目标物体的概率小于小分母小分子中目标物体的概率。

2.3 数据处理

本研究首先采用相关分析探讨思维特质、一般能力和统计计算能力在启发与偏差任务中的作用。然后采用层级回归分析探讨思维特质、一般能力在统计计算能力和启发与偏差任务关系中的调节效应。最后应用多因素方差分析进一步揭示统计计算能力的认知优势的条件。

依据温忠麟等人提出的调节效应的标准化估计方法，本研究标准化调节效应估计采用以下步骤：①对自变量和因变量做标准化处理，计算各自的 Z 分数；②计

算交互项乘积；③检验主效应是否显著；④检验交互作用是否显著。

采用 POLYBIF 软件包计算被试统计计算能力分数，描述统计、相关分析和层级回归采用 SPSS16.0 软件进行。

3 实验结果

3.1 预测因子的调节效应分析

本研究首先计算了思维特质、一般能力、统计计算能力与启发与偏差任务的相关（结果见 Table 20）。相关分析发现，各预测因子与启发和偏差任务相关显著，并且各预测因子之间相关也显著。

Table 20 Descriptive Statistics of Composite Scores and Correlation among Predictor Variables (N=350)

	\bar{X}	$\pm SD$	1	2	3	4
1 Correct Items in Heuristics and Biases Tasks (HB)	7.45	± 1.99	1			
2 Thinking Disposition (TD)	238.30	± 20.79	0.21 [*]	1		
3 General Ability (GA)	40.43	± 5.75	0.49 ^{**}	0.24 ^{**}	1	
4 Statistical Numeracy (SN)	0.01	± 0.86	0.37 ^{**}	0.13 ^{**}	0.60 ^{**}	1

Note: ^{*} $p < 0.05$, ^{**} $p < 0.01$; the mean of Statistical Numeracy is 0 because it were standardized.

Table 21 Hierarchical Regression Analysis on Composite Scores in Heuristics and Biases Tasks

	Predictors	β	t	p	ΔF	R^2	ΔR^2
The First Step	Thinking Disposition (TD)	0.181	3.252	0.001			
	General Ability (GA)	0.209	2.379	0.001	35.796	0.237	0.237
	Statistical Numeracy (SN)	0.218	3.352	0.001			
The Second Step	TD \times GA	0.095	1.419	0.157			
	TD \times SN	0.004	0.065	0.948	0.379	0.239	0.003
	GA \times SN	0.126	1.838	0.067			
The Third Step	TD \times GA \times SN	0.165	2.592	0.010	6.718	0.254	0.015

为了检验思维特质和一般能力是否调节统计计算能力在启发和偏差任务中认知优势的发挥，同时检验思维特质是否调节一般能力的发挥，本研究应用层级多元回归检验思维特质和一般能力的调节作用。思维特质、一般能力和统计计算能力作为第一层首先放入回归方程，两因素交互作用（即思维特质 \times 一般能力、思维特质 \times 统计计算能力、一般能力 \times 统计计算能力）作为第二层，三因素交互作用作为第三层，结果见 Table 21。结果发现，思维特质、一般能力和统计计算能力对规范决策的主效应显著，三者对规范决策都有显著正向预测作用。同时还发现三因素交互作用对规范决策回归系数也显著，说明统计计算能力和规范决策的关系同时受到思维特质和一般能力的调节，换句话说统计计算能力在规范决策中认知优势的发挥依赖于思维特质和一般能力。假设 1 得到验证。

3.2 预测因子与启发与偏差任务的方差分析

虽然上述调节效应结果初步表明统计计算能力和规范决策的关系受到不同思维特质和一般能力的影响。为了更进一步验证不同思维特质和一般能力在统计计算能力对规范决策的调节作用。本研究对各启发与偏差任务中的预测因子进行标准化处理，计算思维特质、一般能力和统计计算能力的 Z 分数，同时以平均数为界划分为高、低能力组，依据三类预测因子分数高低，将被试划分为 8 组，如高思维特质高一般能力高统计计算能力组。计算各组在启发与偏差任务中的平均成绩，同时进行方差分析，探讨不同条件下的被试在启发与偏差任务中的成绩差异，结果见 Table 22。

Table 22 Descriptive Statistics of Composite Scores for Different Group of Predictor Variables in Heuristics and Biases Tasks

Thinking Disposition	General Ability	Low Statistical Numeracy		High Statistical Numeracy	
		N	$\bar{X} \pm SD$	N	$\bar{X} \pm SD$
Low	Low	46	5.89 \pm 2.56	42	6.83 \pm 2.89
	High	45	7.90 \pm 1.79	46	7.85 \pm 1.52
High	Low	44	6.81 \pm 2.02	40	7.90 \pm 1.79
	High	43	8.03 \pm 1.27	44	8.47 \pm 1.04

方差分析显示,不同组别之间方差异显著($F_{7,350}=12.87, p < 0.001$, 偏 $\eta^2=0.214$),说明从总体上看,不同预测因子组合条件下,被试在启发与偏差任务中的成绩不同。

通过 Fig. 10 可以进一步发现,被试的成绩在三个预测因子都高时成绩最好,即高统计计算能力被试只有在高思维特质和高一般能力的时候,其在启发和偏差任务中的成绩最好,而在三个因子都低时被试的成绩最差,说明思维特质和一般能力在总体上调节统计计算能力和规范决策的关系,而假设 2 也得到验证。同时还发现,8 中不同条件下被试在启发与偏差任务中的成绩大致可以分为四类,即 1 为一类,2、3 为一类(两个因子中只有一个因子高),4、5、6、7(三个因子中有两个高,5 除外)为一类,8 为一类,可见 8 个因子中只有一个高时,被试的成绩也不同,还需要进一步探讨不同思维特质和一般能力各水平在统计计算能力与规范决策中的调节作用。

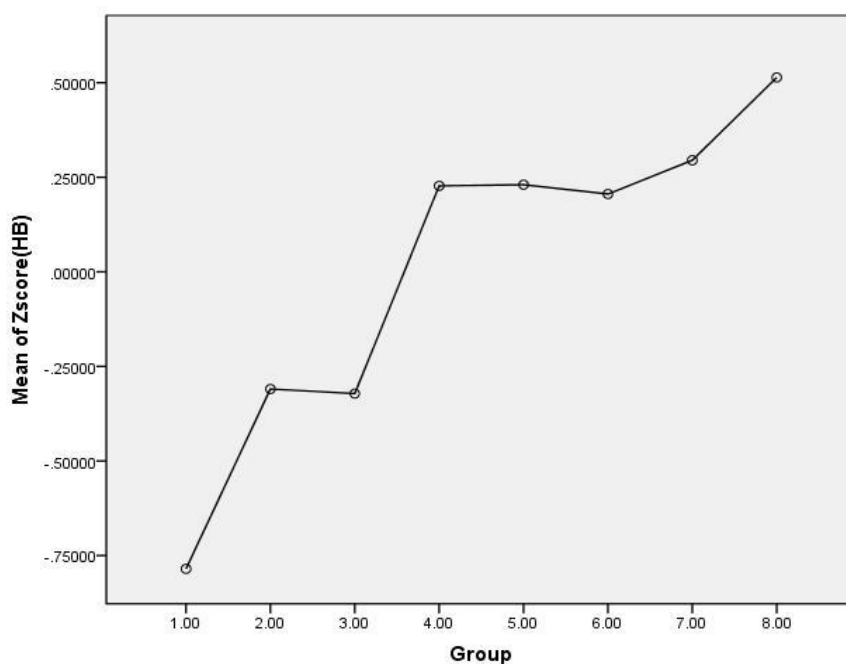


Fig. 10 Different Composite Scores among Different Group of Predictor

Variables in Heuristics and Biases Tasks

Note: 1= Low general ability, low thinking disposition and low statistical numeracy; 2= Low general ability, low thinking disposition and high statistical numeracy; 3= Low general ability, high thinking disposition and low statistical numeracy; 4= Low general ability, high thinking disposition and high statistical numeracy; 5=High general ability, low thinking disposition and low statistical numeracy; 6= High general ability, low thinking disposition and high statistical numeracy; 7=High general ability, high thinking disposition and low statistical numeracy; 8= High general ability, high thinking disposition and high statistical numeracy.

为此,本研究应用上述分组方法采用 2 (高、低思维特质) \times 2 (高、低一般能力) \times 2 (高、低统计计算能力) 方差分析检验预测因子不同水平时对启发与偏差任务成绩的影响,结果见 Table 23。方差分析显示思维特质、一般能力和统计计算能力主效应显著,两两交互作用不显著,但是三因素交互作用显著 ($F_{1, 350}=5.47$, $p < 0.05$, 偏 $\eta^2=0.013$)。

Table 23 ANOVA of Correct Items for Predictor Variables in Heuristics and Biases Tasks

	Source of Variation	F	df	<i>p</i>	Effect Size (Partial η^2)
Main Effect	Thinking Disposition (TD)	7.354	1	0.007	0.022
	General Ability (GA)	23.112	1	<0.001	0.065
	Statistical Numeracy (SN)	5.792	1	0.017	0.017
Interaction Effect	TD \times GA	1.537	1	0.216	0.005
	TD \times SN	0.392	1	0.532	0.001
	GA \times SN	2.689	1	0.102	0.008
	TD \times GA \times SN	5.468	1	0.014	0.013

从 Fig.11 可以进一步发现,低一般能力的时候,思维特质和统计能力的交互作用不显著 ($F_{1, 350}=0.77$, $p > 0.05$, 偏 $\eta^2 < 0.001$)。结合调节效应分析结果,可以推断低一般能力时,思维特质对统计计算能力在规范决策中作用的调节作用不显著;而当一般能力高时,高思维特质能力进一步强化了高统计计算能力被试在启发与偏差任务中的认知优势 ($F_{1, 350}=26.15$, $p < 0.05$, 偏 $\eta^2=0.025$)。而低思维特质限制了统计计算能力的认知优势的发挥 ($F_{1, 350}=1.92$, $p > 0.05$, 偏 $\eta^2=0.006$)。进一步验证了假设 3。

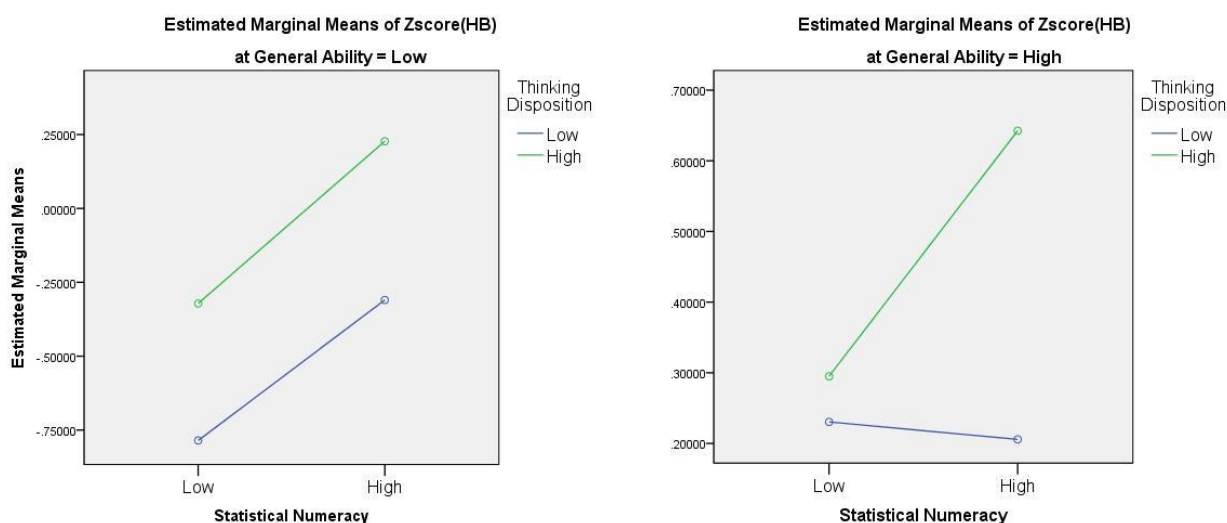


Fig. 11 Different Composite Scores of Different Statistical Numeracy and Thinking Disposition when General Ability was different

4 讨论

本研究依据 Stanovich 等人提出的决策思维加工过程,采用 4 类启发和偏差任务,研究统计计算能力的认知优势发挥机制。发现,高思维特质、高一般能力、高统计计算能力的被试在启发和偏差任务中的成绩更好,而在三者都低的时候成绩最差。进一步调节效应检验发现,一般能力的高低决定思维特质对统计计算能力的认知优势是否有调节作用,而当一般能力高的时候,高思维特质增强统计计算能力的认知优势,低思维特质限制统计计算能力认知优势的发挥。

4.1 高统计计算能力的认知优势依赖于高思维特质和高一般能力

本研究发现,被试在启发和偏差任务中的成绩只有在高思维特质、高一般能力和高统计计算能力的时候成绩最好,而在思维特质或一般能力有一个低的时候,被试的成绩都会受到影响。可见,高统计计算能力的认知优势依赖于高思维特质和高一般能力。这与 Klaczynski 等人对数学能力的思维加工过程的研究结果类似^[175, 187]。他们研究发现,当思维特质低的时候,数学能力的高低与被试在启发和偏差任务中的规范决策没有关系;而当一般能力低的时候,数学能力也与规范决策没有关系。依据 Stanovich 分析思维过程理论,低反思水平能力,会限制算法水平的功能发挥,也就说会影响一般能力和特殊能力的发挥^[146]。本研究中的统计计算能力与 Stanovich

的理论假设基本一致。统计计算能力作为一种特殊能力，高反思思维和一般能力是统计计算能力认知优势发挥的必要条件^[9]。本研究的意义在于以下两个方面，一是本研究突破以往研究采用单一启发和偏差任务研究特殊能力认知优势的发挥机制。本研究结果发现，特殊能力的认知优势机制在 4 类概率推理任务综合分数有明显效应，可见统计计算能力的认知优势在决策中的总体预测性。二是，本研究不但应证第三、第四部分提出的关于高统计计算能力在规范决策中有认知优势，并且发现认知优势发生依赖于高思维特质和高一般能力，二者是统计计算能力认知优势发生的必要条件。

4.2 思维特质担当统计计算能力的教练员角色

本研究中另一个发现是思维特质的管理和监控功能。在高一般能力的前提下，高思维特质进一步强化高统计计算能力的认知优势，而低思维特质则会限制统计计算能力的认知优势。思维特质就像一个教练员，而高统计计算能力是一个优秀的运动员。依据 Stanovich 的理论，系统 1 的快速自动加工过程会激活人们的决策行为反应^[152]。如果思维特质低，人们将采用系统 1 的启发式表征解决决策问题，不会调用一般能力和特殊能力对决策问题进行精细加工，就像教练员不让运动员上场，被试再高的统计计算能力都无从发挥，从而限制了高统计计算能力的认知优势。而如果人们的思维特质高，善于启动分析思维过程并调用恰当的认知资源对决策问题进行分析性加工，则可以进一步强化统计计算能力的认知优势。就像教练员在合适的时机给予运动员合适的上场机会，运动员的绩效也会发挥得更佳。而对低统计计算能力的被试来说，因为其对决策中的统计概率问题没有能力进行科学把握^[60]，因而其在本研究中的启发和偏差任务中成绩都比较差。以往研究也从不同角度证实高思维特质是特殊能力优势发挥的前提条件。如前述的 Klaczynski 等人发现被试只有在中等以上思维特质时，数学能力才会在规范决策中发挥作用^[187]。Handley 等人发现决策行为有时是自动激活的，其对分析性思维过程的决策结果添加额外的权重，强调思维特质在规范决策中的重要性^[188]。可见，在高一般认知能力的前提下，人们的规范决策同时依赖于高统计计算能力和高思维特质。这一发现对于是对特殊能力的认知优势发生机制有了更进一步深入的认识，思维特质在规范决策中，担当统计计算能力认知优势发挥的教练员角色。

4.3 一般能力是统计计算能力和思维特质认知优势的保障条件

本研究中一般能力和思维特质在统计计算能力和规范决策关系中调节作用的合理解释是，高思维特质对系统 1 的启发式加工过程进行监控和驾驭的时候，还需要高一般认知资源（如短时记忆能力）做保障。而 Stanovich 同样指出^[152]，思维特质和一般能力是相互作用的，反思思维监控算法思维的认知过程，但是算法思维同样需要足够的认知资源。这也是本研究的理论基础。一般能力是高思维特质使高统计计算能力被试在规范决策中发生认知优势的保障条件。在要求被试依据逻辑推理过程进行决策的实验的相关报道中^[156]，决策结果依据启发式表征做出判断的成分下降，依据分析思维进行决策的成分增加，但是这一现象只有在高一般能力的被试中才会出现。可见，低水平的一般认知能力同样会限制思维特质强化统计计算能力功能的有效发挥，影响被试在启发与偏差任务中的成绩。而高统计计算能力的发挥同样需要一般能力认知资源做保障，如长时记忆资源的提取。本研究中同样发现，低一般能力的时候，高统计计算能力的被试在启发与偏差任务的成绩同样不高（Fig. 10 中的组 4）。可见，一般能力是统计计算能力和思维特质认知优势的保障条件。

第六部分 模拟现实决策任务中对统计计算能力认知

优势的检验

1 问题提出

通过前 5 部分的研究我们初步发现了统计计算能力在规范决策中存在认知优势，同时对认知优势的机制进行了探讨。但是高统计计算能力被试的认知优势在现实情景中是否经得起检验呢？这是本部分试图解决的问题。本研究拟采用剑桥博弈任务（Cambridge Gambling Task）研究现实情景中统计计算能力的决策功能如何。

决策是人们对不同可能结果的行为选择过程。然而，决策存在一定的不确定性，因而决策过程往往也包含一定的风险。现实生活中成功的决策需要在效益和成本之间做出平衡，同时也需要考虑实现决策结果的可能性^[189]。可见，风险调节（遵循风险规律做出合适的决策行为）和概率评估是日常生活成功决策中的重要能力。剑桥博弈任务通过被试的博弈行为考察被试的决策质量和风险调节能力，对决策的不同方面进行分离，如风险选择、押注行为、反应时和风险调节等^[190]。该任务对决策过程需要考虑的信息都采用外显的方式呈现，如赢和输的概率都以红色和蓝色的比例明显的呈现，该任务能评价被试的外显风险决策能力，而不是模糊决策能力。该任务的另一个优势是刺激强化学习、反转学习和工作记忆的要求或负荷是最小的，因而可以排除学习因素的影响直接研究被试的决策质量^[191, 192]。在该任务中，赌注的比率是以升序或降序的方式依次呈现的，这样被试的真实风险偏好和冲动行为就分开。（如在升序序列中，风险偏好的被试需要等待高比例的赌注出现），因而可以直接考察被试的风险调节能力。研究发现，剑桥博弈任务与现实情景中的决策非常相似，对人们现实生活中的决策有较好的预测性。该任务也是研究异常人群决策障碍的常用工具，如成瘾人群决策障碍、赌博障碍人群决策偏差、过度网络使用者的决策行为等^[193]。本研究拟采用剑桥赌博任务研究不同统计计算能力被试在决策质量和风险调节方面的差异，验证统计计算能力在现实决策情景中的认知优势。

2 方法

2.1 被试

被试为参加航天员决策能力选拔预研项目的被试 187 人。为尽可能反应统计计算能力的全距,本研究涵盖了从小学文化程度到重点大学生不同学历层次的被试群体,年龄 19~28 (22.37 ± 1.97) 岁,其中初中及以下文化程度 55 人,占 29.41%;高中 67 人,占 35.83%;大学及以上文化程度 65 人,占 34.76%。所有被试了解本研究的研究背景、目的和意义。依据被试在剑桥博弈任务中获得代币的数量发放实验报酬。收回有效数据 186 份。

2.2 实验材料及程序

2.2.1 剑桥博弈任务

电脑屏幕上呈现 10 个小方块,这 10 个方块有红色和蓝色两种颜色,其中有一个特殊的方格,它的另一面有黄色的代币。实验情景如 Fig. 12 所示。被试的任务是猜哪一种颜色的方格下藏有黄色的代币,并要求下一定比率的赌注。如果被试所猜的那种颜色下面藏有黄色的代币,则被试赢取相应下注点数,反之则输掉他所下注点数。红蓝方块的比例从 1:9 到 9:1 随机分布,黄色代币也是随机隐藏于 10 个方块中。试验程序为,请被试先猜哪种颜色下可能会有代币并选择相应的颜色,然后被试需要下一定比例的赌注,可能赌注的比例(5%, 25%, 50%, 75%和 95%)呈现在屏幕的右方,如果被试点击相应比例则表示被试下相应比例的赌注。赌注比例以升序或是降序的方式依次呈现,升序和降序的次数相等。被试选择相应的赌注后,屏幕随机显示黄色代币的位置,反馈被试在此次中是赢还是输,并奖励或扣除相应的点数。

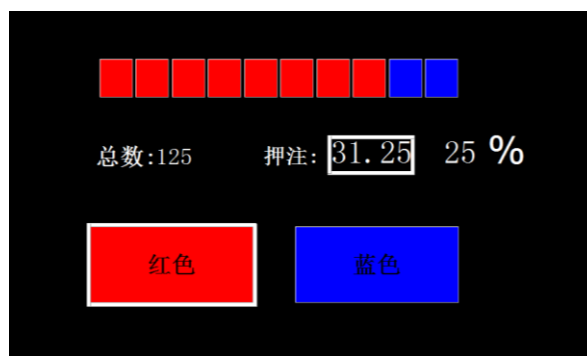


Fig. 12 Cambridge Gambling Task

为了更真实的体现现实决策,告知被试尽可能在剑桥任务中获得更多的金币,并告知被试我们按获取金币数量的 1/10 发放实验报酬。此后,完成统计计算能力测量相关测量题目。

本研究重点记录如下 3 个指标,1) 被试选择比例大颜色的次数,作为决策质量的指标; 2) 降序中下注的均值减去升序中下注的均值,作为被试冲动指数的指标; 3) 并采用下式计算被试的风险调节指标。

$$(2 \times 9:1 \text{ 时的下注} + 8:2 \text{ 时的下注} - 7:3 \text{ 时的下注} - 2 \times 6:4 \text{ 时的下注}) / \text{平均下注}$$

2.2.2 统计计算能力测量

测量方式和能力估计方法同本研究第一部分。

2.3 数据分析

采用 POLYBIF 软件包计算被试统计计算能力分数,采用 SPSS16.0 软件进行描述统计、独立样本 t 检验。

3 结果

本研究以平均数为界,将被试分为高、低统计计算能力两组,比较两组在剑桥博弈任务中决策质量、冲动指数和风险调节方面的差异,结果见 Table 24。独立样本 t 检验发现,高统计计算能力组决策质量显著高于低统计计算能力组($t=3.88, p<0.01$),说明高统计计算能力组的决策质量更高。在冲动指数上,低统计计算能力组显著高于高统计计算能力组($t=6.19, p<0.01$),说明低统计计算能力的被试在决策时更为冲动。风险调节上,高统计计算能力组显示更高的风险调节能力($t=3.36, p<0.01$)。

Table 24 the Difference of Quality of Decision Making, Risk Taking, Risk Adjustment between Different Statistical Numeracy (N=186)

	Statistical Numeracy	$\bar{X} \pm SD$	t	p
Quality of Decision Making	Low	25.96 \pm 6.18	-3.88	<0.001
	High	29.17 \pm 5.03		
Risk Taking	Low	0.48 \pm 0.43	6.19	<0.001
	High	0.19 \pm 0.18		
Risk Adjustment	Low	2.23 \pm 4.17	-3.36	0.001
	High	4.54 \pm 5.11		

4 讨论

本实验的目的在于验证高统计计算能力被试在现实情景中的认知优势。研究结果支持高统计计算能力被试风险调节能力强、决策质量要好。如前所述，人们现实生活中决策的好坏主要取决于概率评估和风险调节两个方面的能力，高统计计算能力被试似乎在这两个方面都有明显的优势。同时高统计计算能力被试还表现为决策过程中冲动指数相对较低，这也验证了第四部分采用启发与偏差任务中关于高统计计算能力对系统 1 的控制和驾驭能力。

4.1 高统计计算能力被试概率评估能力相对较好

通过本部分的实验研究，高统计计算能力被试在决策质量上的表现更好。本实验通过模拟现实实验情景，验证了高统计计算能力的被试在现实生活中的决策确实存在其明显的认知优势。如前所述，虽然本研究还不是完全真实的生活场景决策，但是剑桥博弈任务与现实情景中的决策非常相似，或者说是现实生活决策的实验室研究。剑桥博弈任务一个显著的特征考察被试的外显风险决策能力，在面对清楚的输赢概率，高统计计算能力的被试能清楚的选择胜算较大的颜色，表现为较好的概率评估能力，从而在剑桥博弈任务中能取得相对较好的成绩^[192]。

4.2 高统计计算能力被试风险调节能力强

风险调节能力是现实生活决策好坏的另一个方面，高统计计算能力的被试在风险调节能力方面也有较好的表现。剑桥博弈任务可以将被试的风险偏好和冲动行为有效分离，直接计算被试的风险调节能力^[189]。高统计计算能力被试在风险小（如红蓝比为 9:1）时倾向于下高比例的赌注，但是对风险较大（如红蓝比为 6:4）时，倾向于下低比例的赌注，从而表现为较好的风险调节能力。结合本研究的前期发现，是因为高统计计算能力的被试对概率信息准确表征，依据准确概率信息进行不同比例的下注，表现为更为理性的下注。而低统计计算能力的被试对概率信息的表征是模糊的^[10]，他们可能不能有效表征 9:1 和 8:2 或 7:3 概率信息的差别，对风险感知不能有效表征，因而在下注时也存在混乱，不能遵循风险规律做合理的下注。这充分印证了在现实情景决策中，高统计计算能力被试在数字信息加工和概率信息把握上的认知优势。

4.3 高统计计算能力冲动指数低

本研究前期采用冲突性三段论推理任务发现，高统计计算能力被试对系统 1 的自动加工过程有较好的控制和驾驭能力，这在模拟现实的剑桥博弈任务中也得到了验证。高统计计算能力的被试不但更多的依据分析性系统判断的结果选择合适的下注比例，还能较好的控制系统 1 的快速加工的行为判断^[9]，其有足够的理由耐心等待所要下注的比例，表现为对冲动行为的较好控制。而低统计计算能力的被试系统 1 和系统 2 的竞争性结果中，系统 1 的权重更高，系统 1 加工所得到的决策结果往往是模糊的^[145]，因而不论是在升序还是降序比例中，都快速的选择下注比例，因而其冲动指数较高。从另一个角度来看，也在现实情景决策中印证高统计计算能力被试对自动加工过程较好的控制和驾驭能力，而低统计计算能力被试依赖系统 1 的加工结果做决策^[194]。

小 结

统计计算能力作为规范决策和风险感知能力的最佳预测指标，与现实生活中人们的各类决策及军事决策都有密不可分的关系。然而，从现有文献来看，对统计计算能力在决策中的作用、统计计算能力的认知优势及其机制都缺乏足够的认识和研究，本研究立足于填补这一方面的空白。为此，本研究从探讨统计计算能力的测量工具和方法出发，探讨统计计算能力在决策能力中的地位，进一步探明统计计算能力的认知优势，依据双加工理论对统计计算能力认知优势的发挥机制进行验证性研究。并采用模拟现实决策任务验证统计计算能力在现实情景中的认知优势。研究结果初步发现，Bifactor 模型是 BNT 和 CRT 联合测验的最优拟合模型，BNT 和 CRT 测验二者的公共因子与数学能力、推理能力有较高相关，对数字精确表征也较好的区分性，同时测验信息量也能满足测量学信度要求，说明 BNT 和 CRT 联合测验公共因子测量统计计算能力有较好的信度和效度。接下来，本研究采用全息因素分析方法（FIFA）进一步发现统计计算能力和决策能力的相关主要体现在决策能力公共因子上，证明统计计算能力是决策能力的核心因素之一，进一步证实以往研究关于统计计算能力在规范决策中的核心预测作用。本研究还设计了两个具体实验探讨统计计算能力的认知优势所在。发现高统计计算能力被试对百分比变化的权重更符合线性加工特点，在不同基础比例条件下对核心数字把握能力相对较强，表明高统计计算能力被试具有明显的数字信息权重线性加工优势、对核心数字把握能力较强；同时应用忽视基础概率、博弈谬误、协方差判断和信念偏差 4 类启发与偏差任务还发现，高统计计算能力被试能更好的排除语义信息干扰、抓取基础概率信息，更好的避免选择偏向和信念偏差，更好的统计规则表征，从而证明高统计计算能力被试基础概率信息准确表征、对独立事件和条件概率统计规则都能较好的把握的认知优势；除此之外，高统计计算能力在对系统 1 的控制和驾驭能力方面也表现出优势。同时本研究依据 Stanovich 的分析思维加工理论，采用调节效应分析方法结合多因素方差分析探讨统计计算能力、思维特质和一般能力三者的关系和作用，发现思维特质和一般能力显著调节统计计算能力和规范决策之间的关系，方差分析进一步揭示不同思维特质和一般能力水平对不同统计计算能力在规范决策中的不同作用。明确

统计计算能力的认知优势的发挥机制，发现高统计计算能力的认知优势依赖于高思维特质和高一般能力，高思维特质担当高统计计算能力认知优势的教练员角色，决定是否运用统计计算能力精细认知加工过程，而高一般能力是高思维特质认知优势发挥的保障基础，离开一般能力的支撑，高思维特质和高统计计算能力的认知优势难以实现。最后，本研究采用与现实情景非常相似的剑桥博弈任务验证现实情景决策中统计计算能力的认知优势，发现高统计计算能力被试在现实决策中决策质量和风险调节能力显著高于低统计能力被试，而冲动指数却显著低于低统计计算能力被试，从而在现实任务中证实统计计算能力的认知优势。本研究有如下基本结论：

（1）采用全息 Bifactor 模型提取 BNT 和 CRT 的公共因子测量统计计算能力具有较好的信度和效度；

（2）统计计算能力是决策能力的核心因素之一；

（3）高统计计算能力在规范决策中有对自动加工过程较好的控制和驾驭能力、数字信息权重线性加工、统计规则把握较好等认知加工优势；

（4）高思维特质是高统计计算能力的教练员，担当是否启用统计计算能力的精细认知加工，高一般能力是高统计计算能力认知优势的保障。

同时本研究在以下几个方面做了一些创新性工作：

（1）实现了统计计算能力的多维 IRT 测量。本研究采用多维项目反应（MIRT）技术测量统计计算能力，运用 MIRT Bifactor 模型弥补以往研究对统计计算能力测量无法科学测量和准确测量的不足，在方法学上有一定的创新性；

（2）发现统计计算能力是决策能力的核心因素之一。本研究采用全息因素分析技术，发现统计计算能力是决策能力的核心因素之一；提出统计计算能力在决策中的核心作用，是对以往研究关于统计计算能力在决策中作用的进一步深化，是一种新见解。

（3）比较全面的探讨了统计计算能力在决策中的认知优势及其心理机制。本研究采用 Bifactor 模型结合启发与偏差任务，系统研究统计计算能力在决策中的认知优势和机制，发现一般能力是统计计算能力认知优势的保障、思维特质担当监控分析和决定是否调用统计计算能力分析思维过程。具有一定的理论创新；

虽然本研究取得通过一系列实验对统计计算能力的测量、统计计算能力的认知优势及其机制有一定发现，但是还存在一些研究不足，统计计算能力在规范决策中

作用和机制也有进一步研究的空间需要研究者进一步探索。

(1) 影响决策的认知因素很多,如认知风格、一般能力特征等都有可能是影响决策的重要因素,但是限于博士期间的时间和精力,本研究不能对影响决策全部认知因素进行系统研究,今后的研究中可以进一步探讨其他认知因素在决策中的作用及其机制。

(2) 虽然本研究对统计计算能力的认知优势在模拟现实任务中进行了验证,但是剑桥博弈任务与现实生活决策及军事决策情景还有一定的差异。今后的研究中可以进一步探讨统计计算能力真实生活情境及军事决策中的认知优势。

(3) 本研究所采集的数据来自于行为学、问卷调查,虽然行为学和问卷调查的也有其本身的优势。但是本研究缺乏神经电生理学、神经影像学等方面相关证据的支持统计计算能力的认知优势。比如高统计计算能力被试在对系统 1 的控制和加工方面的优势是否可以考虑应用事件相关电位(ERPs)的方法中 Nogo 刺激的 N2 成分寻找高统计计算能力对冲突监控的神经生理学证据。今后的研究中可以从神经生理学、神经影像学思考实验设计,探讨统计计算能力在规范决策中的认知优势。

(4) 虽然本研究对统计计算能力的认知机制做了初步研究和探讨,依据 Stanovich 决策思维分析过程明确了统计计算能力和一般能力、思维特质的关系。但是人们的决策思维过程是非常精细和复杂的认知过程,探讨如何将自动加工过程和分析思维过程采用如加法反应时或减法反应时的方法,对其进行更为明确的区分,探讨统计计算能力在各个思维阶段的具体作用和意义还需要在今后的研究中进一步开展。

(5) 虽然本研究采用相关实验设计,对统计计算能力的认知优势有所研究。但是对其学习过程和成长发展机制缺乏足够的认识,即对统计计算能力的提高途径和方法缺乏足够的研究。这对帮助指导提升人们的统计计算能力,从而更好的做出规范决策具有非常重要的现实指导意义,因而也是今后研究中一个非常重要的方向。

(6) 虽然本研究对统计计算能力在决策中认知优势有比较系统的研究,但是对低统计计算能力的个体如何改善其决策缺乏实证性的研究。事实上,研究发现统计计算能力并不是短时间内可以迅速提高的一种特殊能力,因而如何帮助低统计计算能力的个体做出良好的决策也是一个非常重要的课题。如研究发现,改变数字信息的呈现方式(将概率形式改变为频率形式)可以有效帮助低数学能力者改善决策。

在今后的研究中也可以加强探讨适应于低统计计算能力者认知规律的信息呈现方式和决策过程，帮助他们在处理决策问题时，做出更好的判断。

参考文献

- [1] Reyna VF, Brainerd CJ. The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making[J]. *Learning and Individual Differences*, 2007;17:147-159.
- [2] Lipkus IM, Samsa G, Rimer BK. General Performance on a Numeracy Scale among Highly Educated Participants[J]. *Medical Decision Making*, 2001;21:37-44.
- [3] Simon HA. A behavioral model of rational choice[J]. *The quarterly journal of economics*, 1955:99-118.
- [4] Fischhoff B. Assessing adolescent decision-making competence[J]. *Developmental Review*, 2008;28:12-28.
- [5] Peters E, Vastfjall D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and Decision Making[J]. *Psychological Science*, 2006;17:407-413.
- [6] Peters E, Levin IP. Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options[J]. *Judgment and Decision Making*, 2008;3:435-448.
- [7] Peters E, Hart PS, Fraenkel L. Informing Patients The Influence of Numeracy, Framing, and Format of Side Effect Information on Risk Perceptions[J]. *Medical Decision Making*, 2011;31:432-436.
- [8] Tversky A, Kahneman D. The framing of decisions and the psychology of choice[J]. *Science*, 1981;211:453-458.
- [9] Stanovich KE, West RF. Natural myside bias is independent of cognitive ability[J]. *Thinking & Reasoning*, 2007;13:225-247.
- [10] Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated participants: Numeracy and metacognition[J]. *Judgment and Decision Making*, 2014;9:15-34.
- [11] Peters E. Beyond comprehension the role of numeracy in judgments and decisions[J]. *Current Directions in Psychological Science*, 2012;21:31-35.
- [12] Chapman GB, Liu J. Numeracy, frequency, and Bayesian reasoning[J]. *Judgment and Decision Making*, 2009;4:34-40.
- [13] Peters E, Hart PS, Fraenkel L. Informing Patients: The Influence of Numeracy, Framing, and Format of Side Effect Information on Risk Perceptions[J]. *Medical Decision Making*, 2010;31:432-436.
- [14] Ancker JS, Kaufman D. Rethinking Health Numeracy: A Multidisciplinary Literature Review[J]. *Journal of the American Medical Informatics Association*, 2007;14:713-721.
- [15] Liberali JM, Reyna VF, Furlan S, Stein LM, Pardo ST. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment[J]. *Journal of Behavioral Decision Making*, 2012;25:361-381.
- [16] Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitations[J]. *Medical Decision Making*, 2007;27:663-671.
- [17] Hamilton M, Barton D. The International Adult Literacy Survey: what does it really measure?[J]. *International review of education*, 2000;46:377-389.
- [18] Schapira MM, Walker CM, Cappaert KJ, 等¹. The Numeracy Understanding in Medicine Instrument A Measure of Health Numeracy Developed Using Item Response Theory[J]. *Medical Decision Making*, 2012;32:851-865.

- [19] Galesic M, Garcia-Retamero R. Statistical numeracy for health: A cross-cultural comparison with probabilistic national participants[J]. Archives of Internal Medicine, 2010;170:462-468.
- [20] Dehaene S, Cohen L. Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic[J]. Cortex, 1997;33:219-250.
- [21] Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: The Berlin numeracy test[J]. Judgment and Decision Making, 2012;7:25-47.
- [22] Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice[J]. Annals of Behavioral Medicine, 2008;35:261-274.
- [23] Schultz JV. A framework for military decision making under risks. DTIC Document; 1997.
- [24] Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making[J]. Psychological bulletin, 2009;135:943-948.
- [25] Dickert S, Kleber J, Peters E, Slovic P. Numeracy as a precursor to pro-social behavior: The impact of numeracy and presentation format on the cognitive mechanisms underlying donation decisions[J]. Judgment and Decision Making, 2011;6:638-650.
- [26] Slovic P, Peters E. Risk perception and affect[J]. Current Directions in Psychological Science, 2006;15:322-325.
- [27] Peters E. Numeracy and the perception and communication of risk[J]. Annals of the New York Academy of Sciences, 2008;1128:1-7.
- [28] Li M, Chapman GB. A big fish or a small pond? Framing effects in percentages[J]. Organizational Behavior and Human Decision Processes, 2013;122:190-199.
- [29] Kray L, Gonzalez R. Differential weighting in choice versus advice: I'll do this, you do that[J]. Journal of Behavioral Decision Making, 1999;12:207-218.
- [30] Kray LJ. Contingent weighting in self-other decision making[J]. Organizational Behavior and Human Decision Processes, 2000;83:82-106.
- [31] Peters E, Slovic P, Västfjäll D, Mertz C. Intuitive numbers guide decisions[J]. Judgment and Decision Making, 2008;3:619-635.
- [32] Dehaene S, Izard V, Spelke E, Pica P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures[J]. Science, 2008;320:1217-1220.
- [33] Furlong EE, Opfer JE. Cognitive constraints on how economic rewards affect cooperation[J]. Psychological Science, 2009;20:11-16.
- [34] Dieckmann NF, Slovic P, Peters EM. The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy[J]. Risk Analysis, 2009;29:1473-1488.
- [35] Tversky A, Kahneman D. Rational choice and the framing of decisions[J]. Journal of business, 1986;S251-S78.
- [36] Kelley CM. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation Edward T. Cokely Max Planck Institute for Human Development Center for Adaptive Behavior and Cognition[J]. Judgment and Decision Making, 2009;4:20-33.
- [37] Ülkümen G, Thomas M, Morwitz VG. Will I spend more in 12 months or a year? The effect of ease of estimation and confidence on budget estimates[J]. Journal of Consumer Research, 2008;35:245-256.
- [38] Strack F, Mussweiler T. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility[J]. Journal of personality and social psychology, 1997;73:437-446.
- [39] Thomas M, Simon DH, Kadiyali V. Do consumers perceive precise prices to be lower than round prices? Evidence from laboratory and market data[J]. Evidence from Laboratory and Market Data (September 2007)

Johnson School at Cornell University Research Paper, 2007.

- [40] Toplak ME, Sorge GB, Benoit A, West RF, Stanovich KE. Decision-making and cognitive abilities: A review of associations between Iowa Gambling Task performance, executive functions, and intelligence[J]. *Clinical psychology review*, 2010;30:562-581.
- [41] Detweiler JB, Bedell BT, Salovey P, Pronin E, Rothman AJ. Message framing and sunscreen use: gain-framed messages motivate beach-goers[J]. *Health psychology*, 1999;18:189-198.
- [42] Nelson W, Reyna V, Fagerlin A, Lipkus I, Peters E. Clinical Implications of Numeracy: Theory and Practice[J]. *Annals of Behavioral Medicine*, 2008;35:261-274.
- [43] Price PC, Matthews TV. From group diffusion to ratio bias: Effects of denominator and numerator salience on intuitive risk and likelihood judgments[J]. *Judgment and Decision Making*, 2009;4:436-446.
- [44] Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale[J]. *Medical Decision Making*, 2007;27:672-680.
- [45] Schwartz B, Ward A, Monterosso J, Lyubomirsky S, White K, Lehman DR. Maximizing versus satisficing: Happiness is a matter of choice[J]. *Journal of personality and social psychology*, 2002;83:1178-1197.
- [46] Weller JA, Dieckmann NF, Tusler M, Mertz CK, Burns WJ, Peters E. Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach[J]. *Journal of Behavioral Decision Making*, 2013;26:198-212.
- [47] Cai L. A Two-Tier Full-Information Item Factor Analysis Model with Applications[J]. *Psychometrika*, 2010;75:581-612.
- [48] Natter HM, Berry DC. Effects of active information processing on the understanding of risk information[J]. *Applied Cognitive Psychology*, 2005;19:123-135.
- [49] Gigerenzer G. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa)[J]. *Subjective probability*, 1994:129-161.
- [50] Bechara A. Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective[J]. *Nature neuroscience*, 2005;8:1458-1463.
- [51] Reyna VF. How people make decisions that involve risk a dual-processes approach[J]. *Current Directions in Psychological Science*, 2004;13:60-68.
- [52] Gerrard M, Gibbons FX, Houlihan AE, Stock ML, Pomery EA. A dual-process approach to health risk decision making: The prototype willingness model[J]. *Developmental Review*, 2008;28:29-61.
- [53] Epstein S. Cognitive-experiential self-theory: A dual-process personality theory with implications for diagnosis and psychotherapy. In: Masling RFBJM, editor. *Empirical perspectives on the psychoanalytic unconscious*. Washington, DC, US: American Psychological Association; 1998. p. 99-140.
- [54] Evans JSBT. How many dual-process theories do we need? One, two, or many? In *two minds: Dual processes and beyond*. New York, NY, US: Oxford University Press; 2009. p. 33-54.
- [55] Reyna VF. A theory of medical decision making and health: fuzzy trace theory[J]. *Medical Decision Making*, 2008;28:56-78.
- [56] Reyna VF, Brainerd CJ. Dual processes in decision making and developmental neuroscience: A fuzzy-trace model[J]. *Developmental Review*, 2011;31:180-206.
- [57] Reyna VF. Fuzzy-Trace Theory, Judgment, and Decision-Making: A Dual-Processes Approach. In: Ohta CIN, editor. *Human learning and memory: Advances in theory and application: The 4th Tsukuba International Conference on Memory*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2005. p.

239-256.

[58] Reyna VF, Ellis SC. Fuzzy-trace theory and framing effects in children's risky decision making[J]. *Psychological Science*, 1994;5:275-279.

[59] Reyna VF, Farley F. Risk and rationality in adolescent decision making implications for theory, practice, and public policy[J]. *Psychological science in the public interest*, 2006;7:1-44.

[60] Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making[J]. *Psychological bulletin*, 2009;135:943-973.

[61] Reyna VF, Lloyd FJ. Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing[J]. *Journal of Experimental Psychology: Applied*, 2006;12:179-195.

[62] Reyna VF, Brainerd CJ. Fuzzy-trace theory and false memory: New frontiers[J]. *Journal of Experimental Child Psychology*, 1998;71:194-209.

[63] Reyna VF, Brainerd C. Fuzzy-trace theory and children's acquisition of mathematical and scientific concepts[J]. *Learning and Individual Differences*, 1991;3:27-59.

[64] Brainerd C, Reyna V. Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development[J]. *Developmental Review*, 1990;10:365-403.

[65] Reyna V, Brainerd C. Fuzzy-trace theory: Some foundational issues[J]. *Learning and Individual Differences*, 1995;7:145-162.

[66] Chen FF, Hayes A, Carver CS, Laurenceau JP, Zhang Z. Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches[J]. *Journal of Personality*, 2012;80:219-251.

[67] Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures[J]. *Quality of Life Research*, 2007;16:19-31.

[68] Immekus JC, Imbrie PK. Dimensionality Assessment Using the Full-Information Item Bifactor Analysis for Graded Response Data: An Illustration With the State Metacognitive Inventory[J]. *Educational and Psychological Measurement*, 2008;68:695-709.

[69] Reise SP, Waller NG. Item response theory and clinical measurement[J]. *Annual review of clinical psychology*, 2009;5:27-48.

[70] Reise SP. The Rediscovery of Bifactor Measurement Models[J]. *Multivariate Behavioral Research*, 2012;47:667-696.

[71] Chen FF, West SG, Sousa KH. A comparison of bifactor and second-order models of quality of life[J]. *Multivariate Behavioral Research*, 2006;41:189-225.

[72] Simms LJ, Grös DF, Watson D, O'Hara MW. Parsing the general and specific components of depression and anxiety with bifactor modeling[J]. *Depression and Anxiety*, 2008;25:E34-E46.

[73] Bell V, Studerus E, Gamma A, Vollenweider FX. Psychometric Evaluation of the Altered States of Consciousness Rating Scale (OAV)[J]. *PLoS ONE*, 2010;5:e12412.

[74] Al-Turkait FA, Ohaeri JU, El-Abbasi A-HM, Naguy A. Relationship between Symptoms of Anxiety and Depression in a Sample of Arab College Students Using the Hopkins Symptom Checklist 25[J]. *Psychopathology*, 2011;44:230-241.

[75] Jennrich RI, Bentler PM. Exploratory Bi-Factor Analysis[J]. *Psychometrika*, 2011;76:537-549.

[76] Gibbons RD, Hedeker DR. Full-information item bi-factor analysis[J]. *Psychometrika*, 1992;57:423-36.

[77] Holzinger KJ, Swineford F. The bi-factor method[J]. *Psychometrika*, 1937;2:41-54.

[78] Ebesutani C, Smith A, Bernstein A, Chorpita BF, Higa-McMillan C, Nakamura B. A bifactor model of negative affectivity: Fear and distress components among younger and older youth[J]. *Psychological*

Assessment, 2011;23:679-691.

[79] Tackett JL, Daoud SL, De Bolle M, Burt SA. Is relational aggression part of the externalizing spectrum? A bifactor model of youth antisocial behavior[J]. Aggressive behavior, 2013;62:45-61.

[80] Brodbeck J, Abbott RA, Goodyer IM, Croudace TJ. General and specific components of depression and anxiety in an adolescent population[J]. BMC Psychiatry, 2011;11:191-209.

[81] Gibbons RD, Bock RD, Hedeker D, 等. Full-Information Item Bifactor Analysis of Graded Response Data[J]. Applied Psychological Measurement, 2007;31:4-19.

[82] Simms LJ, Prisciandaro JJ, Krueger RF, Goldberg DP. The structure of depression, anxiety and somatic symptoms in primary care[J]. Psychological Medicine, 2011;42:15-28.

[83] Brouwer D, Meijer RR, Zevalkink J. On the factor structure of the Beck Depression Inventory-II: G is the key[J]. Psychological Assessment, 2013;25:136-145.

[84] Yang FM, Tommet D, Jones RN. Disparities in self-reported geriatric depressive symptoms due to sociodemographic differences: An extension of the bi-factor item response theory model for use in differential item functioning[J]. Journal of Psychiatric Research, 2009;43:1025-1035.

[85] Yang Y, Sun Y, Zhang Y, 等. Bifactor Item Response Theory Model of Acute Stress Response[J]. PLoS ONE, 2013;8:e65291.

[86] Brown AR, Finney SJ, France MK. Using the Bifactor Model to Assess the Dimensionality of the Hong Psychological Reactance Scale[J]. Educational and Psychological Measurement, 2010;71:170-185.

[87] Van Dam NT, Earleywine M, Borders A. Measuring mindfulness? An Item Response Theory analysis of the Mindful Attention Awareness Scale[J]. Personality and Individual Differences, 2010;49:805-810.

[88] Xie J, Bi Q, Shang W, 等. Positive and Negative Relationship between Anxiety and Depression of Patients in Pain: A Bifactor Model Analysis[J]. PLoS ONE, 2012;7:e47577.

[89] Chen FF, Hayes A, Carver CS, Laurenceau J-P, Zhang Z. Modeling General and Specific Variance in Multifaceted Constructs: A Comparison of the Bifactor Model to Other Approaches[J]. Journal of Personality, 2012;80:219-251.

[90] Jeon M, Rijmen F, Rabe-Hesketh S. Modeling Differential Item Functioning Using a Generalization of the Multiple-Group Bifactor Model[J]. Journal of Educational and Behavioral Statistics, 2012;38:32-60.

[91] Ebesutani C, Reise SP, Chorpita BF, 等. The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor[J]. Psychological Assessment, 2012;24:833-845.

[92] Kõrber A, Takács Á, Urbán R. The Bifactor Model of the Strengths and Difficulties Questionnaire[J]. European Journal of Psychological Assessment, 2013;1:1-9.

[93] Kramer MD, Patrick CJ, Krueger RF, Gasperi M. Delineating physiologic defensive reactivity in the domain of self-report: phenotypic and etiologic structure of dispositional fear[J]. Psychological Medicine, 2011;42:1305-1320.

[94] Fayers PM, Hjermstad MJ, Klepstad P, 等. The dimensionality of pain: Palliative care and chronic pain patients differ in their reports of pain intensity and pain interference[J]. Pain, 2011;152:1608-1620.

[95] Li Y, Rupp AA. Performance of the S - 2 Statistic for Full-Information Bifactor Models[J]. Educational and Psychological Measurement, 2011;71:986-1005.

[96] Fukuhara H, Kamata A. A Bifactor Multidimensional Item Response Theory Model for Differential Item Functioning Analysis on Testlet-Based Items[J]. Applied Psychological Measurement, 2011;35:604-22.

[97] Thomas ML. Rewards of bridging the divide between measurement and clinical theory: Demonstration of a bifactor model for the Brief Symptom Inventory[J]. Psychological Assessment, 2012;24:101-113.

- [98] Luciano JV, Barrada JR, Aguado J, Osma J, García-Campayo J. Bifactor Analysis and Construct Validity of the HADS: A Cross-Sectional and Longitudinal Study in Fibromyalgia Patients[J]. Psychological Assessment, 2013;No Pagination Specified.
- [99] Clark LA, Watson D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications[J]. Journal of abnormal psychology, 1991;100:316-336.
- [100] Revelle W, Wilt J. The General Factor of Personality: A General Critique[J]. Journal of Research in Personality, 2013;45:36-97.
- [101] Martel MM, Roberts B, Gremillion M, Eye A, Nigg JT. External Validation of Bifactor Model of ADHD: Explaining Heterogeneity in Psychiatric Comorbidity, Cognitive Control, and Personality Trait Profiles Within DSM-IV ADHD[J]. Journal of Abnormal Child Psychology, 2011;39:1111-1123.
- [102] Palm KM, Strong DR, MacPherson L. Evaluating symptom expression as a function of a posttraumatic stress disorder severity[J]. Journal of Anxiety Disorders, 2009;23:27-37.
- [103] Reise SP, Scheines R, Widaman KF, Haviland MG. Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling A Bifactor Perspective[J]. Educational and Psychological Measurement, 2013;73:5-26.
- [104] van der Linden D, te Nijenhuis J, Bakker AB. The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study[J]. Journal of Research in Personality, 2010;44:315-327.
- [105] Cai L, Yang JS, Hansen M. Generalized full-information item bifactor analysis[J]. Psychological Methods, 2011;16:221-248.
- [106] Mészáros V, Ádám S, Szabó M, Szigeti R, Urbán R. The Bifactor Model of the Maslach Burnout Inventory–Human Services Survey (MBI - HSS)—An Alternative Measurement Model of Burnout[J]. Stress and Health, 2013;24:14-33.
- [107] Reininghaus U, McCabe R, Burns T, Croudace T, Priebe S. Measuring patients' views: a bifactor model of distinct patient-reported outcomes in psychosis[J]. Psychological Medicine, 2010;41:277-89.
- [108] Gomez R, Vance A, Gomez RM. Validity of the ADHD Bifactor Model in General Community Participants of Adolescents and Adults, and a Clinic-Referred Sample of Children and Adolescents[J]. Journal of attention disorders, 2013;35:26-34.
- [109] Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption[J]. Quality of Life Research, 2009;18:447-460.
- [110] Norton S, Cosco T, Doyle F, Done J, Sacker A. The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis[J]. Journal of Psychosomatic Research, 2012;46:43-51.
- [111] Gibbins C, Toplak ME, Flora DB, Weiss MD, Tannock R. Evidence for a general factor model of ADHD in adults[J]. Journal of attention disorders, 2012;16:635-644.
- [112] Garin O, Ferrer M, Pont À, 等. Evidence on the global measurement model of the minnesota living with heart failure questionnaire[J]. Quality of Life Research, 2013:1-10.
- [113] Frederick S. Cognitive reflection and decision making[J]. Journal of Economic perspectives, 2005:25-42.
- [114] Campitelli G, Labollita M. Correlations of cognitive reflection with judgments and choices[J]. Judgment and Decision Making, 2010;5:182-191.
- [115] Raven JC, De Lemos MM. Standard progressive matrices: Lewis London; 1958.
- [116] Irwin KC, Britt MS. The algebraic nature of students' numerical manipulation in the New Zealand

- Numeracy Project[J]. Educational Studies in Mathematics, 2005;58:169-188.
- [117] Irwin KC, Irwin RJ. Assessing development in numeracy of students from different socio-economic areas: a Rasch analysis of three fundamental tasks[J]. Educational Studies in Mathematics, 2005;58:283-298.
- [118] Welsh M, Burns N, Delfabbro P. The Cognitive Reflection Test: how much more than numerical ability. Proceedings of the 35th Annual Conference of the Cognitive Science Society 2013. p. 1587-1592.
- [119] Hamm R, Bard D, Scheid D. Influence of numeracy upon patient's prostate cancer screening outcome probability judgments. annual meeting of the Society for Judgment and Decision Making, Vancouver, British Columbia, Canada 2003.
- [120] Gurmankin AD, Baron J, Armstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication[J]. Medical Decision Making, 2004;24:265-271.
- [121] Caesar J, These A. Innumeracy: Mathematical Illiteracy and its Consequences[J]. vinculum, 2009;1:37-45.
- [122] Fasolo B, Bana e Costa CA. Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words[J]. Omega, 2014;44:83-90.
- [123] Xiao W, Wu Q, Yang Q, 等. Moral Hypocrisy on the Basis of Construal Level: To Be a Utilitarian Personal Decision Maker or to Be a Moral Advisor?[J]. PLoS ONE, 2015;10:e0117540.
- [124] Brooker S, Rowlands M, Haller L, Savioli L, Bundy DA. Towards an atlas of human helminth infection in sub-Saharan Africa: the use of geographical information systems (GIS)[J]. Parasitol Today, 2000;16:303-317.
- [125] Bruine de Bruin W, Parker AM, Fischhoff B. Individual differences in adult decision-making competence[J]. Journal of personality and social psychology, 2007;92:938-956.
- [126] Del Missier F, Mäntylä T, Bruin WB. Decision-making Competence, Executive Functioning, and General Cognitive Abilities[J]. Journal of Behavioral Decision Making, 2012;25:331-351.
- [127] Appelt KC, Milch KF, Handgraaf MJ, Weber EU. The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research[J]. Judgment and Decision Making, 2011;6:252-262.
- [128] Thompson VA, Prowse Turner JA, Pennycook G. Intuition, reason, and metacognition[J]. Cognitive psychology, 2011;63:107-140.
- [129] Moyer J, Marson DC. Assessment of decision-making capacity in older adults: an emerging area of practice and research[J]. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 2007;62:3-11.
- [130] Zhang J, Miao D, Sun Y, 等. The Impacts of Attributional Styles and Dispositional Optimism on Subject Well-Being: A Structural Equation Modelling Analysis[J]. Social Indicators Research, 2013:1-13.
- [131] Zhang J, Miao D, Sun Y, 等. The Impacts of Attributional Styles and Dispositional Optimism on Subject Well-Being: A Structural Equation Modelling Analysis[J]. Social Indicators Research, 2014;119:757-769.
- [132] Zhang J, Wu Q, Miao D, Yan X, Peng J. The Impact of Core Self-evaluations on Job Satisfaction: The Mediator Role of Career Commitment[J]. Social Indicators Research, 2014;116:809-822.
- [133] Stanovich KE, West RF. Individual differences in rational thought[J]. Journal of experimental psychology: general, 1998;127:161-174.
- [134] Stanovich KE, West RF. Advancing the rationality debate[J]. Behavioral and brain sciences, 2000;23:701-717.
- [135] West RF, Toplak ME, Stanovich KE. Heuristics and biases as measures of critical thinking:

- Associations with cognitive ability and thinking dispositions[J]. *Journal of Educational Psychology*, 2008;100:930-941.
- [136] Campitelli G, Gerrans P. Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach[J]. *Memory & Cognition*, 2014;42:434-447.
- [137] Zhang J, Cui M, Wang W, 等. The Coexistence of Coping Resources and Specific Coping Styles in Stress: Evidence from Full Information Item Bifactor Analysis[J]. *PLoS ONE*, 2014;9:e96451.
- [138] Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk[J]. *Econometrica: Journal of the Econometric Society*, 1979:263-291.
- [139] Ito TA, Larsen JT, Smith NK, Cacioppo JT. Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations[J]. *Journal of personality and social psychology*, 1998;75:887-900.
- [140] Garcia-Retamero R, Galesic M. How to reduce the effect of framing on messages about health[J]. *Journal of general internal medicine*, 2010;25:1323-1329.
- [141] Gamliel E, Kreiner H. Is a picture worth a thousand words? The interaction of visual display and attribute representation in attenuating framing bias[J]. *Judgment and Decision Making*, 2013;8:482-491.
- [142] Gambará H, Piñón A. A meta-analytic review of framing effect: Risky, Attribute and Goal framing[J]. *Psicothema*, 2005;17:325-331.
- [143] Hardisty DJ, Johnson EJ, Weber EU. A dirty word or a dirty world? Attribute framing, political affiliation, and query theory[J]. *Psychological Science*, 2010;21:86-92.
- [144] Kleber J, Dickert S, Peters E, Florack A. Same numbers, different meanings: How numeracy influences the importance of numbers for pro-social behavior[J]. *Journal of Experimental Social Psychology*, 2013;49:699-705.
- [145] Teovanović P, Knežević G, Stankov L. Individual differences in cognitive biases: Evidence against one-factor theory of rationality[J]. *Intelligence*, 2015;50:75-86.
- [146] Toplak ME, West RF, Stanovich KE. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks[J]. *Memory & Cognition*, 2011;39:1275-1289.
- [147] Riege AH, Teigen KH. Additivity neglect in probability estimates: Effects of numeracy and response format[J]. *Organizational Behavior and Human Decision Processes*, 2013;121:41-52.
- [148] Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication[J]. *Applied Cognitive Psychology*, 1997;11:495-506.
- [149] Pacini R, Epstein S. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon[J]. *Journal of personality and social psychology*, 1999;76:972-987.
- [150] Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability[J]. *Learning and Individual Differences*, 2008;18:89-107.
- [151] Li M, Chapman GB. “100% of anything looks good”: The appeal of one hundred percent[J]. *Psychonomic bulletin & review*, 2009;16:156-162.
- [152] Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability[J]. *Journal of personality and social psychology*, 2008;94:672-695.
- [153] Dieckmann N. Numeracy: A Review of the Literature[J]. 2008;64:34-52.
- [154] Ranney SH, Fincel MJ, Wuellner MR, VanDeHey JA, Brown ML. Assessing length-related bias and the need for data standardization in the development of standard weight equations[J]. *North American Journal of Fisheries Management*, 2010;30:655-664.

- [155] Brase GL. Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats[J]. *Journal of Behavioral Decision Making*, 2002;15:381-401.
- [156] Shah AK, Oppenheimer DM. Easy does it: The role of fluency in cue weighting[J]. *Judgment and Decision Making*, 2007;2:371-379.
- [157] Okan Y, Garcia - Retamero R, Cokely ET, Maldonado A. Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension[J]. *Journal of Behavioral Decision Making*, 2012;25:390-401.
- [158] Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future[J]. *Science*, 2011;333:1393-400.
- [159] Slovic P, Finucane ML, Peters E, MacGregor DG. The affect heuristic[J]. *European journal of operational research*, 2007;177:1333-1352.
- [160] Hoppe EI, Kusterer DJ. Behavioral biases and cognitive reflection[J]. *Economics Letters*, 2011;110:97-100.
- [161] Evans JSB. In two minds: dual-process accounts of reasoning[J]. *Trends in Cognitive Sciences*, 2003;7:454-459.
- [162] Kahneman D, Frederick S. Frames and brains: elicitation and control of response tendencies[J]. *Trends in Cognitive Sciences*, 2007;11:45-56.
- [163] Osman M. An evaluation of dual-process theories of reasoning[J]. *Psychonomic bulletin & review*, 2004;11:988-1010.
- [164] Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning[J]. *Advances in health sciences education*, 2009;14:27-35.
- [165] Bar-Hillel M. The base-rate fallacy in probability judgments[J]. *Acta Psychologica*, 1980;44:211-233.
- [166] Clotfelter CT, Cook PJ. Notes: The “Gambler's Fallacy” in Lottery Play[J]. *Management Science*, 1993;39:1521-1535.
- [167] Ayton P, Fischer I. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?[J]. *Memory & Cognition*, 2004;32:1369-1378.
- [168] Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability[J]. *Cognitive psychology*, 1973;5:207-232.
- [169] Shaklee H, Tucker D. A rule analysis of judgments of covariation between events[J]. *Memory & Cognition*, 1980;8:459-467.
- [170] Evans JSB, Curtis-Holmes J. Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning[J]. *Thinking & Reasoning*, 2005;11:382-389.
- [171] Brandstätter E, Gigerenzer G, Hertwig R. The priority heuristic: Making choices without trade-offs[J]. *Psychological review*, 2006;113:409-432.
- [172] Evans JSB. The heuristic-analytic theory of reasoning: Extension and evaluation[J]. *Psychonomic bulletin & review*, 2006;13:378-395.
- [173] Fiedler K, Brinkmann B, Betsch T, Wild B. A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format[J]. *Journal of experimental psychology: general*, 2000;129:399-418.
- [174] Pennycook G, Thompson VA. Reasoning with base rates is routine, relatively effortless, and context dependent[J]. *Psychonomic bulletin & review*, 2012;19:528-534.
- [175] Pennycook G, Trippas D, Handley SJ, Thompson VA. Base rates: Both neglected and intuitive[J]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2014;40:544-554.

- [176] Welsh MB, Navarro DJ. Seeing is believing: Priors, trust, and base rate neglect[J]. *Organizational Behavior and Human Decision Processes*, 2012;119:1-14.
- [177] Kahneman D, Tversky A. Subjective probability: A judgment of representativeness. *The Concept of Probability in Psychological Experiments*: Springer; 1974. p. 25-48.
- [178] Xue G, He Q, Lei X, 等. The gambler's fallacy is associated with weak affective decision making but strong cognitive ability[J]. *PLoS ONE*, 2012;7:e47019.
- [179] Shenhav A, Rand DG, Greene JD. Divine intuition: Cognitive style influences belief in God[J]. *Journal of experimental psychology: general*, 2012;141:423-428.
- [180] Johnson ED, Tubau E. Words, numbers, & numeracy: Diminishing individual differences in Bayesian reasoning[J]. *Learning and Individual Differences*, 2013;28:34-40.
- [181] Lien JW, Yuan J. The cross-sectional "Gambler's Fallacy": Set representativeness in lottery number choices[J]. *Journal of Economic Behavior & Organization*, 2015;109:163-172.
- [182] Shaklee H, Holt P, Elek S, Hall L. Covariation judgment: Improving rule use among children, adolescents, and adults[J]. *Child Development*, 1988;755-768.
- [183] Evans JSB, Stanovich KE. Dual-process theories of higher cognition advancing the debate[J]. *Perspectives on Psychological Science*, 2013;8:223-241.
- [184] S á WC, Kelley CN, Ho C, Stanovich KE. Thinking about personal theories: Individual differences in the coordination of theory and evidence[J]. *Personality and Individual Differences*, 2005;38:1149-1161.
- [185] West RF, Toplak ME, Stanovich KE. Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions[J]. *Journal of Educational Psychology*, 2008;100:930-938.
- [186] Etemadi N. An elementary proof of the strong law of large numbers[J]. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1981;55:119-122.
- [187] Klaczynski PA. Heuristics and biases: interactions among numeracy, ability, and reflectiveness predict normative responding[J]. *Frontiers in psychology*, 2014;5:457-459.
- [188] Pennycook G, Trippas D, Handley SJ, Thompson VA. Base rates: Both neglected and intuitive[J]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2014;40:544-551.
- [189] Van Leijenhorst L, Westenberg PM, Crone EA. A developmental study of risky decisions on the cake gambling task: age and gender analyses of probability estimation and reward evaluation[J]. *Developmental neuropsychology*, 2008;33:179-196.
- [190] Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans[J]. *Cerebral Cortex*, 2005;15:58-63.
- [191] Clark L, Bechara A, Damasio H, Aitken M, Sahakian B, Robbins T. Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making[J]. *Brain*, 2008;131:1311-1322.
- [192] Rahman S, Sahakian BJ, Hodges JR, Rogers RD, Robbins TW. Specific cognitive deficits in mild frontal variant frontotemporal dementia[J]. *Brain*, 1999;122:1469-1493.
- [193] DeVito EE, Blackwell AD, Kent L, 等. The effects of methylphenidate on decision making in attention-deficit/hyperactivity disorder[J]. *Biological Psychiatry*, 2008;64:636-639.
- [194] Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability[J]. *Journal of personality and social psychology*, 2008;94:672-688.

附 录

1 Ploybif 程序中主要语句示例

BNT and CRT joined measurement

Statistical Numeracy

7 2 3 4 3 6 5 7 30 2 1 0 1 0 0 0

1 2 3 4

5 6 7

(7i1)

2222222

1010101

0101001

1000100

.....

2 A-DMC 示例

本测验主要考察你的**决策能力和特点**。一些题目看起来相似，但并不相同，请你按照测验**顺序**和**要求**对每道题**独立**做出判断，请**尽快**完成本测验，测验用时将用于计算测验成绩。

指导语：

下列每一个问题都有 A、B 两套选择方案。每一个问题的回答采用 6 级评定：

1：最倾向 A 2：较倾向 A 3：一般倾向 A

4：一般倾向 B 5：较倾向 B 6：最倾向 B

请通过在相应数字上画“√”来表示你的选择，每个问题只能选择一个数字。

1. 试想一下：近期证据表明，有一种杀虫剂正威胁着 1200 只濒危动物的生命。现有两套救护方案可供选择：

如果采用 A 方案，600 只动物一定能获救

如果采用 B 方案，有 75% 的可能性 800 只动物能获救，25% 的可能性没有一只动物能获救。

1 2 3 4 5 6

最倾向 A

最倾向 B

2. 由于税法改革，你有可能得到 12000 元的退税。有两套方案来使你充分利用好这

个机会：

如果采用 A 方案，在可能获得的 12000 元中，你肯定得到 4000 元；

如果采用 B 方案，你有 $\frac{1}{3}$ 的可能性得到 12000 元，也有 $\frac{2}{3}$ 的可能性得到 0 元。

1 2 3 4 5 6

最倾向 A

最倾向 B

3. 试想一下：在下一个学年中，本市将有 1000 名学生辍学。现有两套方案可以解决这个问题，但只有一种方案能被采用。基于其它市的处理经验，两套方案的结果可以被精确估计：

如果采用 A 方案，1000 名学生中将有 400 名可以继续上学；

如果采用 B 方案，有 40% 的可能性 1000 名学生都可以继续上学，而有 60% 的可能性 1000 名学生中没有人可以继续上学。

1 2 3 4 5 6

最倾向 A

最倾向 B

3 BNT 测验

扔一个只有 5 个面的骰子 50 次，平均起来在这 50 次中有多少次是奇数 (1, 3, 5) 朝上的？ _____

小镇有 1000 居民，其中 500 人是合唱团成员，合唱团中有 100 人是男性。非合唱团员的 500 人中有 300 人是男性。随机抽取一名小镇男性是合唱团员的概率是 _____ %

有这样一个骰子 (6 面)，正面出现 6 点概率是其他数字的两倍。平均起来，如果扔 70 次这样的骰子，出现 6 点的次数是多少？ _____

树林中 20% 的蘑菇是红色，50% 是棕色，30% 是白色。红色蘑菇有毒的概率是 20%，其他蘑菇有毒的概率是 5%。请问该树林里毒蘑菇是红色的概率是多少？ _____ %

4 CRT 测验

1. 扔一个只有 5 个面的骰子 50 次，平均起来在这 50 次中有多少次是奇数 (1, 3, 5) 朝上的？

2. 小镇有 1000 居民，其中 500 人是合唱团成员，合唱团中有 100 人是男性。非合唱团员的 500 人中有 300 人是男性。随机抽取一名小镇男性是合唱团员的概率是 _____ %

3. 有这样一个骰子（6 面），正面出现 6 点概率是其他数字的两倍。平均起来，如果扔 70 次这样的骰子，出现 6 点的次数是多少？

4. 树林中 20%的蘑菇是红色，50%是棕色，30%是白色。红色蘑菇有毒的概率是 20%，其他蘑菇有毒的概率是 5%。请问该树林里毒蘑菇是红色的概率是多少？ %

5 Lipkus 数学能力测验示例

在一次抽奖活动中，1000 张彩票中会有一张抽中汽车，请问一张彩票赢得汽车的概率是多少？ _____

请问如下数字中哪个患病概率意味着最大的风险：a：100 个中 1 个，b：1000 个中 1 个，c：105 个中 1 个？ _____

请问如下数字中哪个患病概率意味着最大的患病风险：a：1%，b：10%，c：5% ？

6 启发与偏差任务示例

（1）忽视基础概率任务：

在某个被试群体中，共有 1000 人，其中男性 4 人，女性 996 人。小李是随机从中选择的一个。小李今年 23 岁，获得工程学学位，周末晚上经常会去健身，喜欢喝啤酒和听摇滚音乐。请问小李最有可能是

A. 男性 B. 女性

（2）博弈谬误

在一次抛硬币游戏中（正面和反面出现的概率都是 50%），小李已经抛了 4 次，前四次出现的都是正面。那么请问第五次是正面的几率有多大？

A. 0.18 B. 0.27 C. 0.50 D. 0.73 E. 0.91

（3）协方差判断任务

张老师研究出一种提高学生物理成绩的教学方法。为了证明他的方法是不是有效。他选择了两个班的同学做了如下实验，一班 200 人采用他的新方法学习，150 人的成绩明显提高，二班 60 人仍用老方法学习，45 人的成绩明显提高，得到的结果如

下：

	成绩显著提高	没有明显提高
新方法	150	50
老方法	45	15

请问你在多大程度上认为他的方法有效？

1	2	3	4	5	6
肯定 无效	比较倾向于 无效	可能 无效	可能 有效	比较倾向于 有效	肯定 有效

(4) 信念偏差任务

下面的每个题目都由三个句子组成，先出现的两个句子称为前提，最后一个句子称为结论。所有的题目都要求你判断能不能从前面的两个前提中合乎逻辑地得出后面的结论。在回答问题时，你要先假定先给出的两个前提都是真实的，并在此基础上回答问题。如果你认为后面的结论能够从前面的前提中**必然地**得出，请选择“A. 能得出”，否则，请选择“B. 不能得出”。

1.

前提 1：所有冒烟的东西都是对健康有益的；

前提 2：香烟是冒烟的；

结论： 香烟是对身体有益的。

请问，在前两个前提的基础上能否必然得出后面的结论？

A. 能得出 B. 不能得出

(5) 大数法则任务

张老师是全国知名的英语教授。他所在学校每年大约有 2000 人会参加英语四级考试，往年该校英语四级一次性通过率是 82%。在今年的四级考试后，他随机询问了 12 名学生的四级英语成绩，其中只有 2 名学生成绩合格。为此，张老师推断，今年的英语四级考试通过率要低于往年。请问，你是否同意张老师的看法？

A. 同意 B. 不同意

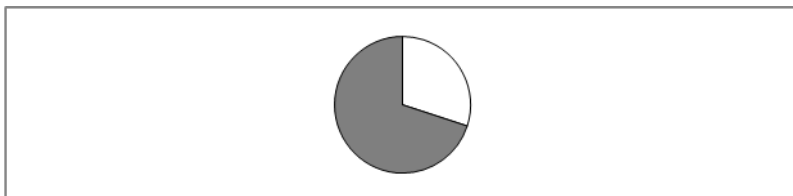
(6) 概率偏差任务

假设你和你朋友打赌。现在他给你两摞牌。第一摞牌中有 50 张牌，其中有 9 张红桃，41 张是其他花色。第二摞中有 5 张牌，其中 1 张是红桃，4 张是其他花色。如果你从任意一摞牌中抽中红桃花色，他就请你吃饭。他说你可以从这两摞牌中任意一摞抽取一张牌，你更愿意从哪一摞牌中抽？

A. 第一摞牌 B. 第二摞牌

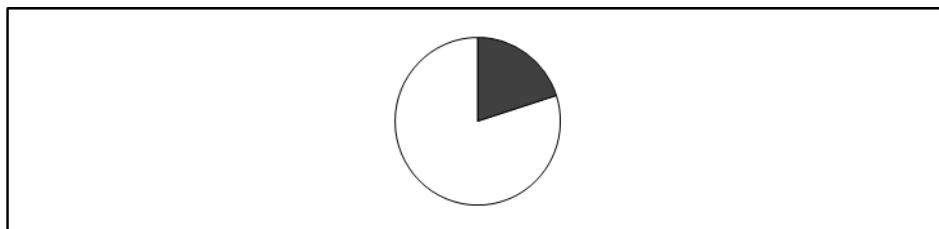
6 数字表征任务

1. 假设你是一名医生，经销商向你推荐某药品，声称他们药物的有效率如下图黑色部分所示，但是他并没有给你提供相关数据，请你估计该药物的有效率



A. 73% B. 70% C. 67% D. 64%

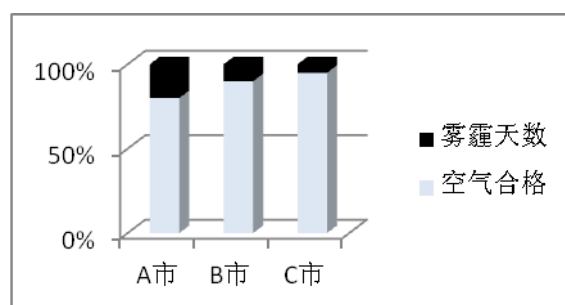
2. 假设某疾病的死亡率如下图黑色部分所示，因为原始数据丢失，需要你估计该疾病的死亡率，请选择你认为最可能的答案



A. 16% B. 18% C. 20% D. 22%

7 第三部分实验材料示例

假设即将面临毕业或正在考虑换工作，A 市、B 市和 C 市都能给你提供较理想的工作岗位，就工作而言三市在薪水和个人发展上都没有差异，可是三座城市在气候环境方面却有一定程度的不同，现在你不得不考虑气候因素对你的健康的影响。A、B 和 C 市过去雾霾天气的资料如下：气象专家对过去一年三市雾霾天进行抽样检测，在 100 天有效观测中，A 市出现雾霾天气 100 天中有 20 天，B 市出现雾霾天气 100 天中有 10 天，C 市出现雾霾天气 100 天中有 5 天。



A、B、C 三市历史气象资料对比图

请根据以上信息回答下列问题：请在相应数字上画“√”表示你的意愿。

1. 据气象专家预测，未来一年中 **A 市/B 市** 空气质量指数合格天数预计会提升 **25%/40%/50%/80%/95%/100%**。如果不考虑其他因素，请问你在该市就职的意愿有多大？

1	2	3	4	5	6	7	8
非常 不愿意				不确定			非常 愿意

个人简历和研究成果

张家喜，男，1982年4月出生，湖南省隆回人。2000年-2004年就读于湖南科技大学应用心理学专业，2004年获理学学士学位。本科毕业后特招入伍，现供职于第二炮兵96164部队。2009年-2012年就读于第四军医大学应用心理学专业，2012年获教育学硕士学位。2012年-2015年于第四军医大学攻读应用心理学博士学位。

攻读博士学位期间主要研究成果：

1. **Zhang, J.**, Cui, M., Wang, W., Lu, H., Wu, Q., Zhu, X., . . . Xiao, W. (2014). The Coexistence of Coping Resources and Specific Coping Styles in Stress: Evidence from Full Information Item Bifactor Analysis. PLoS ONE, 9(5), e96451. doi: 10.1371/journal.pone.0096451. (SCI, IF=3.53, 第一作者)
2. Xiao, W., Wu, Q., Yang, Q., Zhou, L., Jiang, Y., **Zhang, J.**, . . . Peng, J. (2015). Moral Hypocrisy on the Basis of Construal Level: To Be a Utilitarian Personal Decision Maker or to Be a Moral Advisor? PLoS ONE, 10(2), e0117540. doi: 10.1371/journal.pone.0117540. (SCI, IF=3.53, 通讯作者)
3. Peng, J., Li, D., Zhang, Z., Tian, Y., Miao, D., Xiao, W., & **Zhang, J.** (2014). How can core self-evaluations influence job burnout? The key roles of organizational commitment and job satisfaction. Journal of Health Psychology, 1359105314521478. (SSCI, IF=1.84)
4. **Zhang, J.**, Wang, W., Tan, Z., Wu, Q., Xiao, W., Shang, L., . . . Miao, D. (2014). Spatial analysis of schizotypal personality traits in Chinese male youths: evidence from a GIS-based analysis of Sichuan. International journal of mental health systems, 8(1), 1-9. (SSCI, IF=1.08, 第一作者)
5. Xiao, W., Zhou, L., Wu, Q., Zhang, Y., Miao, D., **Zhang, J.**, & Peng, J. (2014). Effects of person-vocation fit and core self-evaluation on career commitment of medical university students: the mediator roles of anxiety and carrier satisfaction. International journal of mental health systems, 8(1), 8. (SSCI, IF=1.08, 通讯作者)

6. **Zhang, J.**, Wu, Q., Miao, D., Yan, X., & Peng, J. (2014). The Impact of Core Self-evaluations on Job Satisfaction: The Mediator Role of Career Commitment. *Social Indicators Research*, 116(3), 809-822. doi: 10.1007/s11205-013-0328-5.(SSCI, IF=1.264, 第一作者)
7. **Zhang, J.**, Miao, D., Sun, Y., Xiao, R., Ren, L., Xiao, W., & Peng, J. (2014). The Impacts of Attributional Styles and Dispositional Optimism on Subject Well-Being: A Structural Equation Modelling Analysis. *Social Indicators Research*, 119(2), 757-769. doi: 10.1007/s11205-013-0520-7 (SSCI, IF=1.264, 第一作者)
8. Peng, J., Jiang, X., **Zhang, J.**, Xiao, R., Song, Y., Feng, X., . . . Miao, D. (2013). The Impact of Psychological Capital on Job Burnout of Chinese Nurses: The Mediator Role of Organizational Commitment. *PLoS ONE*, 8(12), e84193. DOI: 2013, 10.1371/journal.pone.0084193. (SCI, IF=3.74,共同第一)

致 谢

博士三年时光尤如白驹过隙，却满载而归。一路走来如沐浴春风夏雨，得到太多人太多的关心和帮助，感激之情无以言表，谨借文末聊表寸心。

首先感谢我的导师苗丹民教授，您是严师更像慈父。从硕士到博士 6 年来，您的言传身教如春风化雨润物无声，温暖如此、历久弥新。您诲人不倦的高尚师德、奋发图强的人生追求、严谨求实的科学态度、顾全大局的团队协作精神细水长流滋润心田。您更像慈父关心学生的点滴成长和进步，竭尽所能为学生创造发展环境和条件，包括我的调动您都呕心沥血。桃李不言，下自成蹊，能有幸成为您的学生是幸福的，谨诚挚的道一声：老师，谢谢您。

衷心感谢我的辅导老师肖玮副教授，您像情同手足的兄长 6 年来一直关心照顾我的学习和生活，大到人生方向的选择小到文章标点您都事无巨细提供指导、鼓励和帮助，我同样是如此幸运能碰到您这样的良师益友。

衷心感谢刘旭峰教授在我求学过程中给予的不尽关怀和无私帮助。不论是课题中的问题还是生活中的困惑，您都细大不捐不吝赐教。不限于日常点拨更有多个彻夜长谈让我明晓人生的哲理、化解生活的困惑、明晰科研思路。

感谢朱霞副教授一直以来对我个人和课题研究的关心、支持、鼓励和帮助。您总是特别热情的为我想点子出主意明思路，我的几个重要节点上您都给了我莫大的支持和帮助。

感谢杨群副教授、施旺红教授、武圣君副教授、王伟高级实验师、王秀超博士、冯曦老师、张琰老师在给予课题和生活上无私的关心和帮助。

感谢决策研究课题组彭嘉熙博士、周梁硕士、冯田硕士在课题设计和数据处理通力配合和帮助，我也非常感激这支团队的凝聚力让我们的学习高效、充实。

感谢杨业兵博士、孙云峰博士、戴久淳博士、吴青博士、廖扬博士、路惠捷博士、江曦华博士、黄鹏博士、石蕊博士、杨志兵博士、关玮楼硕士、吴迪硕士、曹菲硕士、任慧娟硕士以及医学心理学系所有老师和同学给予我学习研究中的各项关心帮助和支持。

感谢我的父母、妻子、小孩在我上博士期间为我做的牺牲和奉献。感谢 96164 部队曹耀生政委为我上博士创造的便利条件和博士期间给予的无限关怀和帮助。