

# 任务情境和证据信息对青少年贝叶斯推理的影响研究<sup>\*</sup>

史滋福<sup>1,2</sup> 邱江<sup>2</sup> 张庆林<sup>\*\*2</sup>( <sup>1</sup>湖南师范大学数学与计算机科学学院, 长沙, 410081)( <sup>2</sup>西南大学心理学院, 重庆, 400715)

**摘 要** 采用生活情境调查任务和经典调查任务探讨了任务情境对青少年贝叶斯推理的影响, 以及生活情境调查任务中不同证据信息对青少年贝叶斯判断的影响作用。结果表明: (1) 在生活情境调查任务中, 从小学六年级到大学二年级, 被试的贝叶斯推理能力稳步缓慢提升(相邻的两个被试组之间差异不显著, 而不相邻的两个被试组之间差异更容易达到显著水平), 发展没有出现加速期, 而经典调查任务情境下没有表现出年龄差异; (2) 贝叶斯推理作为条件概率的判断不仅受任务情境的影响, 而且同一任务情境中不同证据信息也会影响贝叶斯判断。当证据信息与先验信念一致时, 被试可以充分利用线索进行推理。

**关键词:** 贝叶斯推理 条件概率 任务情境 证据信息

## 1 前言

日常生活中人们需要不断地做出决策, 需要不断地估计各种选择行为结果的可能性, 推导各种已知结果的根本原因。随着思考者年龄的增长, 这类应用于不确定事件的概率判断变得越来越重要<sup>[1]</sup>。贝叶斯推理作为逻辑和认知研究的一个重要方面, 已备受数学家、逻辑学家和心理学家的关注。研究表明, 人们在贝叶斯推理时经常犯错误, 例如大多数推理者包括那些名牌医学院的学生, 在著名的医学疾病诊断问题上表现得都很差, 正确率低于 10%<sup>[2]</sup>。由于出现“地板效应”, 因此很少有关于儿童、青少年贝叶斯推理能力发展的研究。

目前心理学家对贝叶斯推理的研究关注的是人们对问题进行直觉推理时是否遵循贝叶斯定理, 即人们的直觉推理过程是否就是贝叶斯公式的算法过程。采用的实验范式大多是以具体的数据呈现一个二元假设(binary hypothesis)(如患病和没有患病)及二元资料(binary data)(如检测呈阳性或阴性), 要求被试估计在给定证据信息的条件下假设的可能性(如假设检测呈阳性的条件下患病的可能性)。统计分析的方法是将人们的估计结果与贝叶斯公式计算的结果做比较<sup>[2-4]</sup>。

综观目前的研究, 可能存在这样的问题: 第一, 任务的数据呈现方式(类似于教材中讲解贝叶斯公式时采用的习题形式)影响了被试所采用的策略。目前有关贝叶斯推理策略的研究表明, 被试所采用的策略很多是对任务中数据进行各种组合的策略, 如联合同现(joint occurrence)策略、保守主义(conservatism)策略、只依靠证据(evidence-only)策略和前贝叶斯(pre-Bayes)策略等<sup>[3,4]</sup>。第二, 在贝叶斯推理的研究中采用类似于数学情境的贝叶斯推理问题可能引发被试的数学焦虑, 使其内心感到迷惑和无助(乳腺癌问题在很多概率论或高等数学的教材中都作为贝叶斯推理的经典例题呈现)。Ho Hsiu-Zu 等(2000)的研究表明, 数学焦虑会妨碍人们日常生活和学习情境中对数字的处理以及对数学问题的解决和应用<sup>[5]</sup>。第三, 类似于数学情境的贝叶斯推理问题不适合用

来贝叶斯推理能力发展的研究。一方面, Zhuliqi 等指出用概率表征信息时会涉及到大量计算, 当信息以概率的形式呈现给小学四、五、六年级学生时, 他们估计贝叶斯后验概率的能力为零<sup>[4]</sup>。

因此, 探测贝叶斯直觉推理的发展状况不能局限于领会公式的应用, 而应该让被试从公式的变量替换及其计算中解脱出来, 重点考虑任务中所给证据信息条件下后验概率及公式涉及到的另两个条件概率的直觉判断, 然后考察其间的关系是否符合贝叶斯公式。已有的关于蜜蜂、大黄蜂、鸟等动物的研究表明, 动物是相当好的贝叶斯主义者<sup>[6]</sup>, 这个结果似乎表明有关人类的研究需要调整。

事实上, 日常生活中我们对于事件的发生赋以粗略概率的方式不尽相同, 对事件发生之间的联系所指派的概率甚至更为不同, 但生活中我们确实很少出错, 说明人们都在利用某种规则暗中修正概率估计, 但很少有人会利用规则的形式文本。为此, 本研究试图采用不同于经典调查任务的生活情境调查任务, 促使被试在贝叶斯推理过程中完全依靠个人生活经验做出各种证据信息条件下事件的可能性估计, 根本不涉及规则的文本形式。我们假设, 和经典调查任务情境相比较, 这一任务情境可以比较真实地反映出不同年龄被试的生活经验对贝叶斯推理的推理能力, 表现出一定的发展趋势, 同时可以探测到不同证据信息(即推理条件)对贝叶斯推理的影响。这一研究有助于进一步明确贝叶斯推理的认知机制和加工规律, 并为最终改善人们的概率推理、判断与决策等思维活动的质量提供理论依据和实践指导。

## 2 方法

### 2.1 被试

随机抽取普通中小学和某大学被试 199 人, 其中小学六年级(简称小六)43 人、初二 45 人、高二 50 人 and 大二 61 人。所有的学生都不熟悉贝叶斯推理规则。

### 2.2 实验材料

实验材料包括装订在一起的两份问卷, 问卷一(为“经典调查任务”)含指导语和经典乳腺癌问题, 问卷二(为“生活情境

<sup>\*</sup> 本研究得到国家重点学科西南大学基础心理学项目资助。

<sup>\*\*</sup> 通讯作者: 张庆林, 男, E-mail: zhangql@swu.edu.cn

测量任务:李明—小刚任务”)含指导语、两个人物的特征描述和相关的9个问题(见附录,下文统一称为“李明—小刚任务”)。

其中“李明—小刚任务”给被试呈现两个虚构人物:李明和小刚的一些特征描述,然后要求被试对有关二者的各种陈述做出可能性估计(第1—6题),即估计概率 $P(\text{陈述}|\text{李明})$ 和 $P(\text{陈述}|\text{小刚})$ ,接下来要被试在假定关于一个不确定的人 $X(X$ 是李明或小刚,即隐含了先验概率 $P(\text{李明})$ 和 $P(\text{小刚})$ 分别等于0.5)的一些证据信息的条件下,估计 $X$ 是李明的概率(第7—9题),即估计 $P(\text{李明}|\text{陈述})$ 。

由于贝叶斯推理的本质是探讨新证据对原假设可能性的影响作用,即条件概率,因此在本任务中设置了可能支持假设、可能不太支持假设的证据以及二者合取共三种证据信息(是会计师、业余喜欢踢足球、是会计师且业余喜欢踢足球),以探测这三种证据信息对贝叶斯推理精确性的影响。

2.3 实验设计

采用 $4(\text{年级})\times 4(\text{贝叶斯推理类型})$ 的混合实验设计,其中年级为被试间变量(有4个水平:小六/初二/高二/大二),贝叶斯推理类型为被试内变量(有4个水平:经典乳癌问题和李明—小刚任务中三种证据信息条件下的推理)。因变量为贝叶斯推理的准确性,以估计值(第7—9题)与公式计算出的“内隐标准值”之间的距离为指标。距离越小,表明被试估计准确性越高;反之,距离越大,表明被试估计准确性越低。

关于“经典测量任务”中“内隐标准值”的计算可以直接通过公式计算得到,而“李明—小刚任务”中“内隐标准值”的计算,我们以计算 $P(\text{李明}|\text{会计师})$ 为例:将被试的估计值 $P(\text{陈述}|\text{李明})$ (对问题1的估计值)和 $P(\text{陈述}|\text{小刚})$ (对问题

3的估计值)代入贝叶斯公式:

$$\frac{P(\text{李明})P(\text{会计师}|\text{李明})}{P(\text{李明})P(\text{会计师}|\text{李明})+P(\text{小刚})P(\text{会计师}|\text{小刚})}$$
$$=\frac{0.5\times\text{问题1的估计值}}{0.5\times\text{问题1的估计值}+0.5\times\text{问题3的估计值}}\tag{1}$$

计算即得 $P(\text{李明}|\text{会计师})$ 的“内隐标准值”,类似用问题4、5和问题2、6的估计值分别可以计算 $P(\text{李明}|\text{业余喜欢踢足球})$ 及 $P(\text{李明}|\text{会计师和业余喜欢踢足球})$ 的“内隐标准值”。

2.4 实验程序

利用上课时间进行集体测试,将装订好的实验材料(一半是“经典测量任务”在前,另一半是“经典测量任务”在后)按1:1的比例混合后随机发放给被试。需要填写的个人信息包括性别、年龄、专业。被试阅读指导语后,主试用一个与实验无关的简单概率估计任务作示范,要求被试用0—100之间的数字估计概率,被试理解实验要求然后开始完成后面的实验任务。提醒被试认真独立完成,随后将材料收回,完成任务大约需要十五分钟。

3 结果分析

在计算“李明—小刚任务”的“内隐标准值”时发现,有6名被试(小六3人、初二、高二和大二各1人)的估计值代入贝叶斯公式后使得一个或多个式子的分母为0,从而无法得出“内隐标准值”,所以实际参与数据分析的有效问卷共193份。将所有有效数据输入计算机,使用SPSS11.5统计软件包进行处理。四个年级被试的四种贝叶斯推理的估值及根据贝叶斯公式计算所得“内隐标准值”之差的绝对值的平均数见表1。

表1 不同年级四种贝叶斯推理的成绩 ( $M\pm SD$ )

年级	$P(E A)$	$P(E B)$	$P(E A\&B)$	$P(H D)$
	$P(\text{李明} \text{会计师})$	$P(\text{李明} \text{业余喜欢踢足球})$	$P(\text{李明} \text{会计师和业余喜欢踢足球})$	$P(\text{患乳癌} \text{检测呈阳性})$
小六	29.93±21.39	24.73±22.40	27.18±26.10	44.37±23.33
初二	26.40±25.83	21.86±19.94	23.58±23.50	41.16±21.92
高二	18.94±20.31	26.32±25.53	19.86±22.84	37.90±20.56
大二	17.59±16.20	18.31±17.04	15.07±16.11	39.74±26.76
合计	22.50±21.26	22.48±21.29	20.73±22.20	40.25±23.44

对四个年级关于经典乳癌问题贝叶斯推理的“被试估计值”及“内隐标准值”之间的距离进行单因素方差分析,结果表明,年级主效应不显著( $F(3,189)=0.677, p>0.05$ )。说明经典测量任务的推理成绩不能像“李明—小刚任务”那样反映出贝叶斯推理能力的发展。

对“李明—小刚任务”的测量结果进行 $4(\text{年级})\times 3(\text{条件判断类型})$ 重复测量的方差分析,结果见表2。方差分析的结果表明,条件判断类型主效应不显著( $F(2,189)=0.428, p>0.05$ ),而年级类型主效应极其显著( $F(3,189)=4.778, p<0.01$ ),条件判断类型和年级之间的交互作用不显著( $F$

表2 不同年级被试和三种贝叶斯推理方式对贝叶斯推理成绩的影响

变异来源	$df$	$MS$	$F$
年级类型(A)	3	2790.28	4.778**
条件判断类型(B)	2	166.63	0.428
$A\times B$	6	423.84	1.088

注: \*  $p<0.05$ , \*\*  $p<0.01$

(6,189)=1.088,  $p>0.05$ )。

为了进一步分析年级类型主效应,对年级变量做多重比较,结果见表3。多重比较的结果表明,小六和大二之间( $p<0.001$ ),初二和大二之间( $p<0.05$ )达到差异显著性,小六和高二之间( $p=0.062$ )、高二和大二之间( $p=0.081$ )达到边缘差异显著,而小六和初二之间,以及初二和高二之间差异不显著。这说明从小六到大二被试的内隐贝叶斯推理能力稳步缓慢提升(相邻的两个被试组之间差异不显著,而不相邻的两个被试组之间差异更容易达到显著水平),发展没有出现加速期。

4 讨论

4.1 青少年贝叶斯推理能力的发展趋势

表2和表3似乎表明,使用如“李明—小刚任务”这样的生活情境任务探测出被试贝叶斯推理能力随年级增长稳步缓慢地提高、精确性逐渐地得到改善的发展趋势,这与使用

表3 年级的LSD多重比较

小六一初二	小六一高二	小六一大二	初二一高二	初二一高二	高二一高二
3.33	5.57	10.29***	2.24	6.95*	4.71

注: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 。

“经典测查任务”没有得出随年龄增长内隐贝叶斯推理能力改善的趋势是不一致的。

双加工理论(dual-process theory of reasoning)似乎也可以解释贝叶斯推理能力随年级增长有逐渐改善的趋势。这一理论认为推理存在两个系统,系统一是前意识的、自动的、快速的、基于直观判断的,允许快速决策而无需任何认知努力。它使用启发策略,如代表性启发,对概率的估计是根据如典型性或相似性这些概念进行的。系统二则是更多分析的、需要努力的、有意识的系统,或许要通过构建心理模型或通过使用逻辑规则起作用。这两个系统的整合是一个渐进的过程,在成人时期达到极点<sup>[7]</sup>。面对“李明一小刚”这一推理任务,低年级被试可能很少有相关技巧和专业背景,而且由于日常学习过程中遇到的通常都是结构良好的、封闭的、适宜通过传统模式解决的问题,因此他们做贝叶斯推理时可能更多地关注所提供信息的细节,采用对认知要求较少的启发式策略,利用典型性或相似性来进行判断,如从李明的特征描述中感觉他更具有会计师的典型性,而业余喜欢踢足球似乎不太符合李明。在后来的生活中人们逐渐形成更成熟的思维,能够识别和考虑多种观点,因此更高学历阶段的被试可能更加有选择地关注所提供的部分信息,并以一种更多分析的方式整合信息,导致了贝叶斯推理成绩年级之间的差异,表现出随年级提高贝叶斯推理能力稳步缓慢发展的趋势。

此外,从表1可以看出,每种判断条件和每个年级被试的平均估值实际上都偏离了标准正确值,这也说明,即使在生活情境中人们进行直觉判断出现偏差也是比较普遍的。

#### 4.2 不同证据信息对贝叶斯推理能力发展趋势的影响

实验设计中为了探测可能支持假设和可能不太支持假设的证据以及二者的结合对假设的影响作用,设置了三种证据信息条件下的贝叶斯推理。由表2方差分析的结果可知,条件判断类型主效应不显著( $p > 0.05$ ),说明三种证据信息条件下的判断类型具有比较高的同质性。为了进一步探测可能支持假设和可能不太支持假设的证据以及二者的结合对贝叶斯推理能力发展趋势的影响作用,对三种证据信息条件下的概率判断分别做了以年级为被试间变量的单因素方差分析。结果表明,第一种条件( $X$ 是会计师)和第三种条件( $X$ 是会计师且业余喜欢踢足球)下概率判断表现出显著的年级差异( $F(3, 189) = 3.808, p < 0.05$ ;  $F(3, 189) = 2.768, p < 0.05$ ),而第二种条件( $X$ 业余喜欢踢足球)下概率判断却没有表现出显著的年级差异( $F(3, 189) = 1.469, p > 0.05$ )。这似乎说明,贝叶斯推理作为条件概率的判断不仅受任务情境的影响,而且同一任务情境中不同证据信息也会导致其所做贝叶斯推理有一定的差异。

在Chasseigne等(1997)的研究中,采用多重线索概率学习(MCPL)范式(其中某一标准变量的结果与许多线索变量的值有关)表明,当线索变量与标准变量之间的关系一致时,

被试可以有效地利用线索<sup>[8]</sup>。本研究中,附加的信息D1(即会计师)通常与H(即李明)的特征描述一致,而D2(即业余喜欢踢足球)则不太符合李明的特征描述,从而在第一和三种条件的判断中被试很好地利用了信息线索,表现出和总的任务中一致的发展趋势(稳步缓慢地增长),而在第二种条件的判断中可能由于被试没有充分利用信息,使得推理能力受到影响,表现出和总的任务不同的结果(没有年龄差异)。

为了更深入地分析“李明一小刚任务”中数据信息线索对贝叶斯推理发展趋势的影响,对三种条件下概率判断做以“被试估计值”和“内隐标准值”间的距离为因变量,以年级为自变量的直线回归分析,得到相应的贝叶斯推理精确性和年级之间的线性关系:

$l_1$ : 估计值和内隐标准值之间的距离 =  $34.157 - 4.369 \times \text{年级}$  (2)

$l_2$ : 估计值和内隐标准值之间的距离 =  $26.909 - 1.658 \times \text{年级}$  (3)

$l_3$ : 估计值和内隐标准值之间的距离 =  $31.518 - 4.041 \times \text{年级}$  (4)

容易看出,直线 $l_1$ 和 $l_3$ 几乎是平行的(其倾角分别为 $102^\circ$ 和 $103^\circ$ ,直线 $l_2$ 的倾角为 $121^\circ$ ),这也可能说明了被试在进行贝叶斯推理时,线索变量“ $X$ 是会计师”与被试的先验信念“ $X$ 是李明”相一致时,促使被试有效地利用了线索,表现出其对第一种条件下的判断和第三种条件下的判断的影响作用是相当的。由于保存在内隐记忆中的恰好是事物的常见的显著的属性,而不完全是其必要属性,这也在一定程度上,说明了内隐记忆对思维的影响。

#### 4.3 数学焦虑和刻板印象对贝叶斯推理的影响

对学生进行个别访谈发现,他们面对“经典测查任务”表现出明显的焦虑和无助,基本上靠猜测得出结论;面对如“李明一小刚任务”这样的生活情境测查任务表现得要轻松多了,表现得很有兴趣。因此,使用“经典测查任务”没有表现出年级之间贝叶斯推理成绩的差异,而使用后者却表现得相当明显,这在一定程度上可能是由于数学焦虑这种消极的情绪妨碍了人们的思维,阻碍和掩盖了真实的贝叶斯推理能力,也说明了生活情境测查任务在探测贝叶斯推理能力时可能更适当、更敏感。

另外,一些比较直接的证据表明,刻板印象会促使人们对处于不同群体但却具有相同信息的个体产生不同的内隐推理<sup>[9]</sup>。“李明一小刚任务”中的描述也可能使被试在无意识或无控制的状态下形成刻板印象,从而唤醒了其各种心理图景,做出个性的推理。因此,研究这些心理图景对推理的影响作用,对以后贝叶斯推理研究是否考虑或者屏蔽任务情境中的刻板印象以及考察不同刻板印象对贝叶斯推理的影响具有重要的作用,同时对更深入探索影响贝叶斯推理的内在机制具有深远的意义。

5 结论

采用生活情境测查任务和经典测查任务探讨了青少年贝叶斯推理能力的发展状况,以及生活情境测查任务中不同证据信息对青少年贝叶斯判断的影响作用。结果表明:第一,在生活情境测查任务中,从小学六年级到大学二年级,被试的贝叶斯推理能力稳步缓慢提升(相邻的两个被试组之间差异不显著,而不相邻的两个被试组之间差异更容易达到显著水平),发展没有出现加速期,而经典测查任务情境下没有表现出年龄差异;第二,贝叶斯推理作为条件概率的判断不仅受任务情境的影响,而且同一任务情境中不同证据信息也会影响贝叶斯判断。当证据信息与先验信念一致时,被试可以充分利用线索进行推理。

此外,数学焦虑可能妨碍人们的思维,阻碍和掩盖了真实的贝叶斯推理能力,使用非数学情境的任务似乎可以屏蔽数学焦虑对推理成绩的影响。刻板印象也可能影响人们的贝叶斯推理成绩,因此进一步研究和考证这些因素对贝叶斯推理的影响作用,对更深入探察影响贝叶斯推理的内在机制、提高人类条件推理判断能力具有深远的意义。

6 参考文献

1 Fisk J E. Age and probabilistic reasoning: biases in conjunctive, disjunctive and Bayesian judgments in early and late adulthood. *Journal of Behavioral Decision Making*, 2005, 18(1): 55—82

2 Eddy D M. Probabilistic reasoning in clinical medicine in judgment under uncertainty: Heuristics and biases. In: Kahneman D, Slovic P, Tversky A. (Eds.). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press, 1982, 249—268

3 Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 1995, 102: 684—704

4 Zhu L Q, Gigerenzer G. Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, 2006, 98(3): 287—308

5 Ho H Z, Chiu S Y, Wang Ch P, et al. The affective and cognitive dimensions of math anxiety: a cross-national Study. *Journal for Research in Mathematics Education*, 2000, 31(3): 362

6 Real L A. Animal choice behavior and the evolution of cognitive architecture. *Science*, 1991, 253: 980—986

7 Epstein S, Lipson A, Holstein C, et al. Irrational reactions to negative outcomes: evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 1992, 62: 328—339

8 Chasseigne G, Mullet E, Stewart T R. Aging and multiple cue probability learning: the case of inverse relationships. *Acta Psychologica*, 1997, 97: 235—252

9 Kunda Z, Sherman W B. Stereotypes and the construal of individuation information. *Personality and Social Psychology Bulletin*, 1993, 19: 90—99

Effects of Task Context and Evidence Information  
on the Bayesian Inference of Adolescents

Shi Zifu<sup>1,2</sup>, Qiu Jiang<sup>2</sup>, Zhang Qinglin<sup>2</sup>

(<sup>1</sup> College of Mathematics and Computer Science, Hu'nan Normal University, Hu'nan, Changsha, 410081)

(<sup>2</sup> School of Psychology, Southwest University, Chongqing, 400715)

**Abstract** Life context and classical tasks were used to discuss the effects of different task and evidence information on the Bayesian inference of adolescents. The results indicated that (1) Compared to classical mammography problems, the ability of Bayesian reasoning had an apparent improvement from the sixth grade in elementary school to the sophomore year at college, which was reflected by the more accuracy results. No such thing was found in the situation of classical tasks; (2) Bayesian inference was affected not only by task context but also by different evidence information.

**Key words:** Bayesian reasoning, conditional probability, task context, evidence information