# 致 谢

感谢求学期间老师们的教导、朋友们的陪伴。

感谢张萌老师、吴明证老师和戴俊毅老师在升学过程中对我的帮助,特别感谢戴俊毅老师对我毕业设计的辅导。

最后感谢父母对我的支持。

# 摘要

近年来,大量研究显示了描述决策和经验决策之间存在系统性差异,尤其是在小概率事件的权重方面。尽管这一描述-经验差距表明两种决策可能是基于不同的认知过程,但几乎没有研究对部分反馈范式二元选择任务中的经验决策进行系统的模型比较。为此,基于 Technion 预测竞赛(Erev et al., 2010)的数据集,本研究比较了 2(效用函数)×3(选择函数)×2(选择敏感性)共12个强化学习(reinforcement learning, RL)模型,以及3个基于胜留败走(win-stay-lose-shift, WSLS)策略的启发式模型。

综合拟合性能和泛化性能的比较结果,本研究得到如下结论: (1)对于 RL 模型的函数选择,在效用函数方面,对数效用函数优于幂效用函数。在选择函数方面,softmax 选择函数优于 probit 选择函数,ratio 选择函数的适用性仍有待探究。在选择敏感性方面,尽管非试次依赖选择敏感性在多个评价指标上优于试次依赖选择敏感性,但选择敏感性的试次依赖性仍有待探究。(2)对于启发式模型,WSLS-2模型始终优于 WSLS-1和 SS 模型,这支持了决策者在胜利(win)试次和失败(loss)试次后的初始停留(stay)倾向存在差异,并且单个胜利试次和失败试次会对两组条件概率均进行调整的假设。(3)在策略使用上,启发式模型在整体层面和最优模型层面均表现出明显优势,但关于启发式模型和 RL 模型的相对优劣仍有待进一步探究。本研究为未来经验决策的行为建模工作提供了参考,并且显示了定量模型比较在提出新理论或完善已有理论方面的作用。

关键词: 部分反馈范式,强化学习,胜留败走,定量模型比较

#### **Abstract**

In recent years, numerous studies have demonstrated a systematic discrepancy between decisions from description and decisions from experience, especially in the weighting of small probabilities. Although this description-experience gap implied different cognitive processes of these two decisions, there was nearly no study conducting an integrated model comparison for decisions from experience in binary choice problems using partial-feedback paradigm. Based on the dataset from Technion prediction tournament (Erev et al., 2010), this study compared 12 reinforcement learning (RL) models consisting of 2 utility functions, 3 choice functions and 2 choice sensitivities, as well as 3 heuristic models derived from the win-stay-lose-shift strategy.

By comparing the performances of post hoc fit and generalization, the study revealed: (1) For the RL models, the logarithmic utility function outperformed the power utility function in terms of the utility function. The softmax choice function outperformed the probit choice function, while the applicability of the ratio choice function remained to be explored in terms of the choice function. The trial-independent choice sensitivity outperformed the trial-dependent choice sensitivity in many indicators, but the trial-dependence remained to be explored in terms of the choice sensitivity. (2) For the heuristic models, WSLS-2 model was always the optimal model, which supported its two key assumptions. The first was that the initial stay probabilities after win trials and loss trials were different. And the second was that a win trial or a loss trial would adjust both conditional stay probabilities. (3) In terms of the strategy usage, the heuristic models dominated in both overall performances and optimal performances, but the applicabilities of the heuristic models and the RL models remained to be explored. The current study provides a reference for future behavioral modeling works, and shows the contribution of quantitative model comparison in building new theories or improving existing theories.

**Key words:** partial-feedback paradigm, reinforcement learning, win-stay-lose-shift, quantitative model comparison

# 目 次

# 第一部分 毕业论文(设计)

承请	与书	I
致	谢	II
摘	要	.III
1号	言	1
1.1	描述-经验差距	2
1.2	部分反馈范式	2
1.3	强化学习模型	3
1.3.	1 效用函数	4
1.3.	2 更新函数	5
1.3.	3 选择函数	5
1.3.	4 选择敏感性	8
1.4	启发式模型	9
1.5	问题提出	. 11
2 模	[型拟合	.13
2.1	数据集	.13
2.2	参数估计	.14
3 抄	【合性能比较	16
3.1	评价指标	16
3.2	<i>BIC</i> 比较结果	.18
3.3	拟合 <i>MSD</i> 比较结果	.22
3.4	讨论	.25
4 泛	化性能比较	.26
4.1	评价指标	.26
4.2	$G^2$ 比较结果	.27
4.3	泛化 <i>MSD</i> 比较结果	.29
4.4	讨论	.32
5 总	计论	33

5.1 强化学习模型函数选择
5.2 启发式模型选择
5.3 策略使用差异
5.4 不足和展望
6 结论
参考文献
附录 145
附录 2
附录 3
附录 4
作者简历
《浙江大学本科生毕业论文(设计)任务书》
《浙江大学本科生毕业论文(设计)考核表》
第二部分 文献综述和开题报告
文献综述和开题报告封面
指导教师对文献综述和开题报告具体内容要求
目录
文献综述1
开题报告
外文翻译
外文原文
《浙汀大学本科生文献综术和开题报告者核表》

# 1引言

在现实生活中,人们每天要进行各式各样的风险决策,小至某天的晚饭是自己烹饪还是外出就餐,大至是否要与一个人迈入婚姻的殿堂。在某些情境中,决策者知道可选选项的所有结果及相应概率。例如,有一个博彩者正准备买彩票,现在有两种彩票可以选择:购买彩票 A 需要花费 5 元,它的中奖概率为 0.05,奖金为 100 元;购买彩票 B 需要花费 10 元,它的中奖概率为 0.01,奖金为 1000 元。这个博彩者事先知道相关信息,因此,该决策问题可以简化为在选项(-5, 0.95; 95, 0.05)和选项(-10, 0.99; 990, 0.01)之间进行选择(此处引入(x,p;y,q)来表示某一选项,该符号指选项能以概率 p 产生结果 x ,概率 q 产生结果 y ,概率 1-p-q 不产生任何收益或损失,即结果 0 )。在这类问题中,由于决策者需要通过查阅描述信息来了解选项结构,因此它被称为描述决策(decisions from description)。

然而在更多情境中,决策者事先无法得到完整的信息。例如,有一个工作者在忙碌了一天后感到十分疲惫,决定到餐馆就餐。他家附近有两家餐馆:餐馆 A 常常能提供中等质量的菜品,餐馆 B 则有时能提供高质量的菜品,有时却提供低质量的菜品(这种变化可能和食材当天的新鲜程度或厨师的心情有关)。显然,这个工作者有许多可用的策略。比如,他可以给不同质量的菜品打分(例如,分别给高、中、低三种质量的菜品 10、5、0 分),然后估计不同质量的菜品在不同餐馆中的出现概率,那么这一问题就可以和前面的"彩票问题"类似。当然,他也可以直接比较在两家餐馆最近一次的就餐经历,并选择提供更好菜品的一家。但除了少数极为简单的策略(例如,投硬币决定去哪家餐馆)外,一般的策略均需要这个工作者至少去过一次这两家餐馆,也就是依据在每家餐馆的就餐经验进行决策。在这类问题中,由于决策者需要根据自身的经验信息来推测选项结构,因此它被称为经验决策(decisions from experience)。

在这两种情境中,如果决策者采用相同的策略,那么对于同一个问题,当描述信息和经验信息等价时(例如,被告知某选项出现某结果的概率是 0.8, 或 10 次选择该选项时有 8 次得到该结果),他们应当做出类似的决策。然而,近十几年来大量研究揭示了这两类问题中决策行为的系统性差异(Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Wulff, Mergenthaler-Canseco, & Hertwig, 2018),这种差异也被称为描述-经验差距 (description-experience gap, DEG)。

#### 1.1 描述-经验差距

在实验室情境中,风险决策问题往往被简化为二元选择形式的经济问题。也就是说,决策者会面对两个选项,每个选项对应一个结果分布,这两个结果分布独立且稳定。在描述决策研究中,关于结果分布的信息通常使用可视化方式(例如,饼状图和频率分布图)或数字呈现,决策者需要做出单次选择,并且往往没有反馈(Hertwig et al., 2004; Kahneman & Tversky, 1979)。而在经验决策研究中,决策者需要通过重复选择并得到反馈,来获取关于选项结果和相应概率的信息,并基于这些信息进行对选项的评估和选择(Hertwig et al., 2004)。

面对这两类决策问题,人们表现出不同的行为特征。具体来说,在描述决策中,人们的决策行为往往显示了小概率权重过高(overweighted)。即决策者在评估某一选项时,小概率结果所占权重似乎高于客观概率本身。因此可以观察到确定性效应(certainty effect)、反射效应(reflection effect)等现象(Tversky & Kahneman, 1981)。然而在经验决策中,人们的决策行为一般却显示了小概率权重过低,可以观察到以上效应的反转(Barron & Erev, 2003; Wulff et al., 2018)。

由于这两类问题中决策行为存在系统性差异,研究者试图建立不同模型来捕捉这些相异的特征。对于描述决策,当前心理学界接受较广的理论是 Kahneman 和 Tversky (1979) 提出的前景理论(prospect theory, PT)。而对于经验决策,因为其研究历史相对较短,并且实验范式存在异质性,所以目前仍缺少一个统一的理论或模型来建构这类决策的认知过程。

# 1.2 部分反馈范式

在 DEG 研究中,主要通过三种实验范式来实现经验决策:抽样范式(sampling paradigm)、部分反馈范式(partial-feedback paradigm)和全部反馈范式(full-feedback paradigm)。Camilleri和Newell(2011)在固定经验次数的条件下,观察到这三种范式中的经验决策行为存在差异,表明这三类范式触发的心理过程可能并不相同。

本研究关注部分反馈范式下的经验决策问题。在这类范式中,决策者起初对选项没有任何先验知识,他们需要在两个选项间进行规定次数(通常较大)的选择来获取关于选项结果和相应概率的信息。在每次选择后,只有被选择选项会给出反馈,这一反馈会计入决策者的总损益(payoff)。决策者的任务是在规定的次数内尽可能最大化自己的总损益,图 1.1

简要显示了实验流程。部分反馈范式触发的心理过程较为复杂。首先,决策者需要根据直接经验(direct experience),即每次选择后的反馈来对选项做出评估。其次,决策者进行的是重复决策(repeated decisions),也就是需要在选项间做出一系列的选择(Hertwig et al., 2004)。最后,由于仅有被选择的选项给出反馈,决策者面临探索-利用权衡(exploration-exploitation tradeoff),即最大化个体损益和尽可能了解选项信息之间的权衡(Audibert, Munos, & Szepesvári, 2009)。

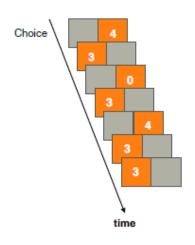


图 1.1 部分反馈范式流程图

注: 引自 Wulff 等(2018)。

对于部分反馈范式,目前较为成熟的两种模型是强化学习(reinforcement learning, RL)模型和启发式(heuristic)模型。RL模型是一种规范的模型,可以捕捉经验决策中的动态学习过程,但其每一部分的函数形式和在不同环境中的拟合效果是多样化的(Yechiam & Busemeyer, 2005)。启发式模型在某些环境中能较好地预测经验决策行为,但这类静态策略假设决策者不会发生对选项结构的学习,因此难以捕捉经验决策中的动态学习过程(Erev & Barron, 2005; Yoon, Vo, & Venkatraman, 2017)。

### 1.3 强化学习模型

RL 模型一般包含效用函数(utility function)、更新函数(updating function)和选择函数(choice function)三个部分。效用函数是将单次选择结果的客观价值转化为效用,更新函数是通过经验对每个选项的期望(expectancy)进行更新,选择函数是根据每个选项当前的期望,计算对其的选择概率(Ahn, Busemeyer, Wagenmakers, & Stout, 2008)。其中,效用函数

和选择函数是大多数描述决策理论也包含的成分(实际上, PT 中使用"价值函数"代替"效用函数",但本研究不对"效用"和"主观价值"作区分)。

# 1.3.1 效用函数

不论是描述决策还是经验决策,人们对于选择结果的评价并不是基于客观价值,而是基于效用。在PT中,Kahneman和Tversky(1982)最早使用了幂函数作为效用函数的数学形式,将其应用到部分反馈范式中,即为

$$u(t) = \begin{cases} x(t)^{\alpha} & x(t) \ge 0 \\ -\lambda |x(t)|^{\alpha} & x(t) < 0 \end{cases}$$
(1-1)

其中 u(t)是第 t 个试次选择结果的效用,x(t)是第 t 个试次的选择结果。参数  $\alpha$  控制效用函数的形状( $0 < \alpha < 1$ ,对于收益  $\alpha \to 0$ , $u(t) \to 1$ ;  $\alpha \to 1$ , $u(t) \to x(t)$ ),参数  $\lambda$  控制损失厌恶的程度( $\lambda > 1$ , $\lambda$  越大代表决策者对损失越敏感)。

另一种使用较为广泛的效用函数形式是对数函数。对数效用函数具有很多变体,其中一种较有竞争力的变体是 Scholten 和 Read(2010)提出的标准化对数效用函数(normalized logarithmic utility function),将其迁移到部分反馈范式中可表示为

$$u(t) = \begin{cases} \frac{1}{\alpha} \ln(1 + \alpha x(t)) & x(t) \ge 0\\ -\frac{\lambda}{\alpha} \ln(1 + \alpha |x(t)|) & x(t) < 0 \end{cases}$$
(1-2)

其中参数  $\alpha$  控制效用函数的形状( $\alpha > 0$ , 对于收益  $\alpha \to 0$ ,  $u(t) \to x(t)$ ;  $\alpha \to \infty$ ,  $u(t) \to 0$ ), 其他符号含义与公式 1-1 相同。

在经验决策中,目前并无研究对这两个效用函数进行比较。但在描述决策中,存在一种称为量级效应(magnitude effect)的现象。具体来说,想象两次选择,第一次是在选项(m, p)和选项(n, q)之间,第二次是在选项(am, p)和(an, q)之间,其中 0 < m < n,  $0 < q < p \le 1$ , a > 1 且 pm = qn。多数决策者在第二个问题中对安全选项(即方差更小选项,在这两个问题中分别是(m, p)和(am, p))的选择概率高于在第一个问题中,并且选择概率增加的幅度随 a 增加而增大(Markowitz, 1952; Prelec & Loewenstein, 1991; Weber & Chapman, 2005)。在使用不同的选择函数(见 1.3.3 节)时,对数效用函数相较幂效用函数都能更好地解释这一现象(具体证明见附录 1)。而 Scholten 和 Read(2014),Bouchouicha 和 Vieider(2017)使用不同指标,分别比较了采用不同效用函数的 PT 模型对量级效应相关数据的拟合性能,均发现

对数效用函数表现最优。这表明对数效用函数具有一定的测试意义。

此外,尽管无直接比较这两个效用函数的经验决策研究,但关于效用是对客观价值的非线性转换这一假设在经验决策中仍是得到支持的。Ahn 等(2008)对 Iowa 赌博任务(Iowa gambling task, IGT)和 Soochow 赌博任务(Soochow gambling task, SGT),即两种含四个选项的经验决策任务的模型比较研究显示,幂效用函数比直接将选择结果客观价值进行线性转换的效用函数表现更好。这表明,即使在经验决策中,人们对损益的主观感受仍表现出灵敏度递减(diminishing sensitivity)的特点。而在 Lejarraga 和 Hertwig(2017)对部分反馈范式的模型比较研究中,也发现相比直接用选择结果客观价值作为效用的 RL 模型,使用幂效用函数的 RL 模型对数据的拟合性能更好。他们的另一个重要发现是,对于各类效用函数,其损失厌恶参数的估计均未大于 1,这表明经验决策中可能不存在损失厌恶现象,即参数 λ 可取 0 到 1 间的数值。

#### 1.3.2 更新函数

在每次得到反馈后,决策者需要将其与过去的经验结合起来,更新对选项的期望,并基于选项的期望进行下一步的选择。更新函数实际上就是这一"学习"过程的体现,也是经验决策相对描述决策最特殊的部分。在更新函数的选择上,较为常用的一种被称为 Delta 学习规则(Rescorla & Wagner, 1972; Sutton & Barto, 2018),具体可表示为

$$E_{j}(t) = E_{j}(t-1) + A\delta_{j}(t)(u(t) - E_{j}(t-1))$$
(1-3)

其中  $E_j(t)$ 代表选项 j(j=1,2)在第 t 个试次的期望。参数 A 控制被选择选项期望更新的程度 (0 < A < 1, A 越大代表近因效应越强,即最近结果对选项期望影响越大)。 $\delta_j(t)$ 是一个哑变量,当选项 i 在第 t 个试次被选择时取 1,否则取 0。

### 1.3.3 选择函数

Stott(2006)回顾了 8 项要求决策者对同一问题进行多次选择的描述决策研究,发现其中均有部分参与者的选择发生过反转,这表明人们的选择行为本身具有随机性。但随着选项整体价值(PT 中决策依据的指标,与经验决策中的期望类似)差异的增大,人们选择较优选项的概率会越来越大,这表明选择的随机性与选项整体价值差异有关。

当前有两种重要的理论解释选择行为的随机性。第一种是 Luce 的选择公理(Luce's

choice axiom, LCA; Luce, 1959), 它定义在一组选项集合 S 中选择选项 x 的概率为

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)}$$

(1-4)

其中v(x)是选项x的响应强度(response strength)。在描述决策中,一种直接的想法是以选项的整体价值作为响应强度,而将其迁移到部分反馈范式中,可以表示为

$$P(D(t+1) = j) = \begin{cases} \frac{E_{j}(t)^{\theta(t)}}{E_{j}(t)^{\theta(t)} + E_{\sim j}(t)^{\theta(t)}} = \frac{1}{1 + (\frac{E_{\sim j}(t)}{E_{j}(t)})^{\theta(t)}} & \text{wábiling} \\ \frac{|E_{\sim j}(t)|^{\theta(t)}}{|E_{j}(t)|^{\theta(t)} + |E_{\sim j}(t)|^{\theta(t)}} = \frac{1}{1 + (\frac{E_{j}(t)}{E_{\sim j}(t)})^{\theta(t)}} & \text{fighting} \end{cases}$$

(1-5)

其中 P(D(t+1)=j)代表在第 t+1 个试次选择选项 j(j=1,2)的概率, $E_j(t)$ 代表选项 j 在第 t 个试次的期望, $E_{-j}(t)$ 代表互斥选项在第 t 个试次的期望(若 j=1,则~j=2,反之亦然)。 $\theta(t)$  代表在第 t 个试次的选择敏感性, $\theta(t)$ 越大代表在第 t 个试次的选择敏感性越强,即决策者 越倾向选择当前较优选项。收益域问题指所有可能结果非负的决策问题,损失域问题指所有可能结果非正的决策问题。

不难看出,这是一个基于选项期望比值的选择函数,因而在本研究中将其称为 ratio 选择函数。此外,直接由 LCA 导出的 ratio 选择函数没有考虑到选项期望为负数的情况。对于两个为负数的期望,如果仍沿用收益域的规则,会导致对较劣选项选择概率较高的预测,显然不符合人类决策行为。因此,对损失域问题中的选择规则做出如公式 1-5 所示调整,但该选择函数仍无法处理选项期望一正一负时的选择情况。

另一种更为常用的 LCA 型选择函数是 softmax 选择函数,它可以写为

$$P(D(t+1) = j) = \frac{e^{\theta(t)E_{j}(t)}}{e^{\theta(t)E_{j}(t)} + e^{\theta(t)E_{\sim j}(t)}} = \frac{1}{1 + e^{\theta(t)\left(E_{\sim j}(t) - E_{j}(t)\right)}}$$

$$= \operatorname{logistic}\left(\theta(t)\left(E_{j}(t) - E_{\sim j}(t)\right)\right)$$
(1-6)

其中所有符号的含义与公式 1-5 相同。并且由于在仅有两个选项时,softmax选择函数等价于将两选项期望差值和选择敏感性的乘积进行logistic转换,因此能化简成公式 1-6 最后的形式。尽管softmax也是LCA型选择函数,但它是一种基于期望差值的选择函数,同时,

它对期望的正负性没有要求。

第二种解释选择行为随机性的理论是随机效用理论(random utility theory, RUT)。RUT 可以追溯至 Thurstone(1927)的成对比较理论,并从上世纪六十年代起被广泛用于决策行为 建模(Luce & Suppes, 1965; Yellott, 1977)。该理论假设,在一组选项集合中,决策者对每个选项关键属性的评估值含有一个固定成分和一个随机成分,即

$$U(x) = u(x) + \varepsilon_x \tag{1-7}$$

其中u(x)是选项x关键属性评估值的固定成分,在描述决策中可以认为是整体价值。而 $\varepsilon x$ 是选项x关键属性评估值的随机成分,符合某个概率分布。这两个成分的加和形成某次选择前对选项x关键属性的评估值U(x),决策者在该次选择中会选择评估值最大的选项。为方便起见,选项的随机成分通常被认为服从均值为0独立同分布的正态分布,此时形成的选择函数可称为probit选择函数。在部分反馈范式中,它可表示为

$$P(D(t+1)=j) = P\left(E_j(t) + \varepsilon_j(t) > E_{\sim j}(t) + \varepsilon_{\sim j}(t)\right) = \Phi\left(\frac{E_j(t) - E_{\sim j}(t)}{\theta(t)}\right)$$
(1-8)

其中 $\Phi(\cdot)$ 是累积正态分布函数, $\varepsilon_j(t)$ 和 $\varepsilon_{-j}(t)$ 分别为选项j与其互斥选项(j=1,2)的期望在第t个试次的随机成分,其他符号的含义与公式 1-5 相同。但需要指出的是,此处 $\theta(t)$ 反映了各选项随机成分在第t个试次服从的正态分布的方差大小,其值越小代表各选项期望的评估越稳定,决策者在第t个试次的选择敏感性也越强,即越倾向选择当前较优选项。与softmax选择函数类似,probit选择函数也是一种基于期望差值的选择函数。

在经验决策研究中,softmax选择函数可以说是RL模型内最广泛使用的选择函数,因为它计算简洁,并容易扩展到多个选项的决策问题中(Schulz & Gershman, 2019)。此外,它对于探索-利用权衡的描述有一定神经生理学方面的合理性(Daw, O'doherty, Dayan, Seymour, & Dolan, 2006; Collins & Frank, 2014),并且确实在一些研究中表现出模型拟合上的优势(Daw et al., 2006; Yechiam & Busemeyer, 2005)。但其他两种选择函数则几乎没有应用,因此也缺乏直接对这三个选择函数的比较研究。

然而, ratio选择函数和probit选择函数仍有一定的测试意义。对于前者, Worthy, Maddox和Markman(2008)设计了一项选项结果分布动态变化的经验决策任务,并设计了三类问题。第一类问题为基线问题,第二类问题在基线上每次选择结果加上一个固定数量(控制选择结果的差值与第一类问题近似),第三类问题在基线上每次选择结果乘以一个固定

数量(控制选择结果的比值与第一类问题近似)。结果发现第一类和第三类问题中决策者选择行为类似而与第二类存在显著差异,这表明人们对选项差异的感知可能是基于比值而非差值的(然而他们的实验缺陷是控制选择结果差值或比值一定,不代表控制由此转化的效用,乃至选项期望的差值或比值一定),也就是说ratio选择函数可能更适合解释经验决策行为。对于后者,它是RUT中的一个代表性函数,而RUT关于人们对选项的评估服从概率分布的假设,已经在一些RL模型中得到应用(Gershman, 2018; Schulz & Gershman, 2019),因此该选择函数也有尝试的价值。

#### 1.3.4 选择敏感性

在三种选择函数中,均有一个选择敏感性参数,该参数捕捉了除选项期望差异之外影响决策者选择随机性的因素。在描述决策中,由于决策者往往进行少量选择,因此选择敏感性基本是稳定的(Stott, 2006),将这一观点迁移到部分反馈范式中,即为选择敏感性是非试次依赖的

$$\theta(t) = 3^c - 1 \tag{1-9}$$

其中参数 c 控制整个决策过程中的选择敏感性强度(c>0,c 越大代表整个决策过程中选择敏感性越强)。

然而,由于部分反馈范式中,决策者需要在两个选项间进行多次选择,可能会出现明显的练习或疲劳效应。这种动态变化过程可以由另一种形式的选择敏感性捕捉

$$\theta(t) = (\frac{t}{10})^c \tag{1-10}$$

其中参数 c 控制选择敏感性的变化趋势。当 c>0 时,随试次数增加,决策者选择敏感性增强,这反映了决策者关于自己对期望的估计越来越有信心,可以认为是一种练习效应。而当 c<0 时,随试次数增加,决策者选择敏感性减弱,这可能反映了疲劳或无聊导致选择趋向随机水平(以上推断均限于 ratio 选择函数和 softmax 选择函数,对 probit 函数则推断结果相反)。

在经验决策相关研究中, Ahn 等(2008)对 IGT 和 SGT 的数据分析显示, 在不同模型的拟合和泛化能力上, 尽管最优模型均使用了非试次依赖选择敏感性, 但试次依赖选择敏感性整体表现更优。而在 Beitz, Salthouse 和 Davis(2014)对不同年龄组被试 IGT 数据的模

型拟合中,发现对于成人以及老年人的选择行为,非试次依赖选择敏感性相关模型相较试次依赖选择敏感性相关模型表现出明显的优势。

综上,对于 RL 模型,本研究将比较 2(效用函数)×3(选择函数)×2(选择敏感性)共 12个模型。

### 1.4 启发式模型

尽管 RL 模型是一种系统并且规范的认知模型,但直观来说,人们在选项间进行重复选择时可能不会使用所假设的这种较为复杂的策略。他们更可能仅依据选项的少数属性,例如可能结果的相对大小或产生收益的概率来进行选择,也就是运用启发式策略。对于部分反馈范式,一种重要的启发式模型是胜留败走(win-stay-lose-shift, WSLS)模型。该模型假设人们在得到奖赏之后会倾向于保持自己的选择,而在未得到奖赏后会倾向改变自己的选择。该策略具体可表示为

$$P(D(t+1) = j) = \begin{cases} p(stay|win) & D(t) = j & & x(t) \ge x(t-1) \\ 1 - p(stay|win) & D(t) = \sim j & & x(t) \ge x(t-1) \\ p(shift|loss) & D(t) = \sim j & & x(t) < x(t-1) \\ 1 - p(shift|loss) & D(t) = j & & x(t) < x(t-1) \end{cases}$$

$$(1-11)$$

其中 D(t) = j 代表第 t 个试次选择了选项 j(j = 1, 2), ~j 代表与 j 互斥选项(若 j = 1, 则~j = 2, 反之亦然),x(t)是第 t 个试次的选择结果。

不难看出,该模型假设当第t个试次的选择结果大于等于第t-1个试次的选择结果时,该试次就是得到奖赏的胜利(win)试次。而在第t+1个试次中,决策者会以p(stay|win)的条件概率保持之前的选择(即停留,stay),以1-p(stay|win)的条件概率(即p(shift|win))切换到另一选项(即转移,shift)。当第t个试次的选择结果小于第t-1个试次的选择结果时,该试次就是未得到奖赏的失败(loss)试次。而在第t+1个试次中,决策者会以p(shift|loss)的条件概率shift,以1-p(shift|loss)的条件概率(loss)loss)loss)loss)loss0。因此,该模型仅含两个参数,即条件概率loss0。loss1。loss2。loss3。loss4。loss3。loss4。因此,该模型仅含两处数,即条件概率loss3。loss4。loss6。loss6。loss6。loss6。loss6。loss7。loss8。loss8。loss8。loss8。loss8。loss9 loss9。loss9 loss9。loss9 loss9 l

尽管在一些研究中, WSLS 模型表现出近似甚至优于 RL 模型的拟合性能 (Worthy, Hawthorne, & Otto, 2013; Worthy, Otto, & Maddox, 2012)。但如其他启发式模型一

样,WSLS 模型没有办法捕捉经验决策中的动态学习过程。为此,受到 Estes(1950)联想学习(associative learning)模型的启发,Worthy 和 Maddox(2014)提出了一个修正的模型,在此称为 WSLS-1 模型。WSLS-1 模型假设如果第 t 个试次是 win 试次,会使得

$$p(stay|win)_{t+1} = p(stay|win)_t + \theta_{p(stay|win)}(1 - p(stay|win)_t)$$
$$p(shift|loss)_{t+1} = p(shift|loss)_t$$

(1-12)

其中  $p(stay|win)_t$  指若第 t-1 个试次是 win 试次,第 t 个试次选择 stay 的概率。 $p(shift|loss)_t$  指若第 t-1 个试次是 loss 试次,第 t 个试次选择 shift 的概率。参数  $\theta_{p(stay|win)}$ 代表单个 win 试次对被试后续 stay 概率的调整比例( $0 < \theta_{p(stay|win)} < 1$ ,  $\theta_{p(stay|win)}$ 越大代表调整程度越大)。而如果第 t 个试次是 loss 试次,会使得

$$p(shift|loss)_{t+1} = p(shift|loss)_t + \theta_{p(shift|loss)}(1 - p(shift|loss)_t)$$

$$p(stay|win)_{t+1} = p(stay|win)_t$$
(1-13)

其中参数  $\theta_{p(shift|loss)}$ 代表单个 loss 试次对被试后续 shift 概率的调整比例( $0 < \theta_{p(shift|loss)} < 1$ , $\theta_{p(shift|loss)}$ 越大代表调整程度越大),其他符号与公式 1-12 相同。此外,该模型还允许决策者两组初始的条件概率不同,即另有两个参数  $p(stay|win)_1$  和  $p(shift|loss)_1$ 。

Worthy 和 Maddox(2014)将该模型与 RL 模型组合形成了一个加权模型,并对不同的 经验决策任务数据进行模型拟合,均显示了两种模型相近的权重,他们认为这反映了人们 会同时使用两种策略进行经验决策。

WSLS-1 模型隐含的一个关键假设是单个的 win 试次或者 loss 试次仅会调整与之相关的一组条件概率。然而,Lejarraga 和 Hertwig(2017)对其进行了修改,他们认为单个的 win 试次或者 loss 试次会对两组条件概率均进行调整,在此称为 WSLS-2 模型。WSLS-2 模型 假设如果第 t 个试次是 win 试次,会使得

$$p(stay|win)_{t+1} = p(stay|win)_t + \theta_{p(stay|win)}(1 - p(stay|win)_t)$$

$$p(shift|loss)_{t+1} = (1 - \theta_{p(shift|loss)})p(shift|loss)_t$$
(1-14)

而如果第t个试次是 loss 试次, 会使得

$$p(shift|loss)_{t+1} = p(shift|loss)_t + \theta_{p(shift|loss)}(1 - p(shift|loss)_t)$$

$$p(stay|win)_{t+1} = (1 - \theta_{p(stay|win)})p(stay|win)_t$$

(1-15)

其中所有符号含义与公式 1-12 和 1-13 相同。

Lejarraga 和 Hertwig(2017)使用 WSLS-2 模型对部分反馈的数据进行拟合,发现它相对几种 RL 模型具有明显的优势。

不论 WSLS-1 模型还是 WSLS-2 模型都设置了两组条件概率分别描述决策者在 win 试次和 loss 试次后的选择行为,然而似乎并无研究对这一假设进行验证。此外,在 Estes(1950)的联想学习模型中,个体实际只有一组反应的概率。因此,本研究提出一个新的启发式模型来探究设置两组条件概率的必要性,在此称为 simple stay(SS)模型。SS 模型 假设决策者仅存在简单的保持选择的倾向,但同样受试次是 win 或 loss 的调节。如果第 t 个试次是 win 试次,会使得

$$p(stay)_{t+1} = p(stay)_t + \theta(1 - p(stay)_t)$$
(1-16)

其中 $p(stay)_t$ 是指被试在第t个试次stay的概率,参数 $\theta$ 代表单个试次结果对被试后续stay概率的调整比例 $(0<\theta<1,\;\theta$ 越大代表调整程度越大)。而如果第t个试次是loss 试次,会使得

$$p(stay)_{t+1} = (1 - \theta)p(stay)_t$$
(1-17)

其中所有符号含义与公式 1-16 相同。该模型另含一个参数  $p(stay)_1$  代表初始的 stay 概率。实际上,SS 模型等价于一种受限的 WSLS-2 模型,即要求  $p(stay|win)_1$  和 1 -  $p(shift|loss)_1$  相等,且  $\theta_{p(stay|win)}$  和  $\theta_{p(shift|loss)}$  相等。

综上,对于启发式模型,本研究将比较 WSLS-1、WSLS-2 和 SS 三个模型。

#### 1.5 问题提出

对于目前存在的相关研究,最重要的问题是多数模型比较工作并不是针对本研究所关注的,部分反馈范式下二元选择任务中的经验决策问题,而是针对 IGT、概率学习任务 (probability learning task)等更广泛的经验决策问题。IGT 的决策问题包含 4 个选项,并且每个选项会产生一个确定的收益结果和一个可能的损失结果。概率学习任务每个试次必然

有一个选项产生收益,而另一选项不产生收益,并且收益无绝对大小(可认为选项结果只有 1 和 0)。而其他一些未形成系统范式的任务可能使用了选项结果分布动态变化的决策问题。这种不同的问题结构可能触发决策者不同的策略,因此它们的研究结论未必适用于本研究关注的决策问题。

此外,几乎没有研究系统地比较,由不同效用函数、选择函数和选择敏感性组成的RL模型对经验决策数据的拟合或泛化(generalization)性能(即使用根据已有样本数据估计出的参数去预测新样本数据的能力)。描述决策的模型比较研究(Stott, 2006)已表明,模型的不同组分之间可能存在交互作用,这说明每个组分中最优函数所组成的模型未必是整体上的最优模型。而 Ahn 等(2008)在对 IGT 数据的模型比较研究中也发现,选择敏感性表现的相对优劣受到模型中更新函数类型的调节。因此,寻找最优 RL模型仍需要尝试不同函数的所有组合。而对于启发式模型,尽管 Lejarraga 和 Hertwig(2017)已发现 WSLS-2 模型相对于部分 RL模型的优势,但是关于它设置两组条件概率,以及每个 win 或 loss 试次会对两组条件概率均进行调整的假设并未被验证过,所以有必要比较它和 WSLS-1、SS模型的优劣,以找出其中最合适的启发式模型。

最后,以往的研究大都使用单一指标比较模型的拟合性能或泛化性能。但如果一个模型确实捕捉到了经验决策的本质特征,那么不论在拟合还是泛化上,它都应该表现出优越的性能,并且这种性能不受评价指标的干扰。所以,为增强比较结果的可信性,应当设置多个指标来比较模型在拟合和泛化两方面的性能。

就以上问题,本研究将基于经验决策中的部分反馈范式,使用当前公开的数据集开展系统的模型比较工作。具体来说,本研究将使用多个指标,对由不同效用函数、选择函数和选择敏感性组成的 RL 模型,以及 WSLS-1、WSLS-2 和 SS 三个启发式模型的拟合性能和泛化性能进行比较,整合各个指标下的最优模型,分别从 RL 策略和启发式策略角度分析经验决策行为特征。并且比较 RL 最优模型和启发式最优模型,分析经验决策中决策者的策略使用差异。

在理论意义上,定量模型可以帮助研究者根据外在的行为数据推断内在的认知机制,并将不同的认知过程加以分离。此外,定量模型比较有助于提出新理论或完善已有理论。例如,本研究关注的 RL 模型各部分函数形式中,幂效用函数和 ratio 选择函数是唯一在描述决策中无法解释量级效应的组合(具体证明见附录 1),那么至少在描述决策中,研究者可以避免用这种组合的相关模型去解释数据。甚至如果本研究发现该组合的相关模型在对数据的拟合和泛化上表现较差的话,也可以推测经验决策中同样存在量级效应。而 Erev

等(2010)在讨论定量模型的意义时也指出,许多重要理论的建立来自对定量规律的观察和总结,例如 Weber-Fechner 定律就是起源于 Weber 在研究重量最小可觉差时发现的数量规律。

在实际意义上,尽管随着信息化时代的到来,人们有许多方式得到不同事件的统计信息,但对于一些个人化的选择,例如是否要创业、是否要去某家店购物,我们只能通过经验这种原始的方式来获得信息。而 DEG 的存在显示了这两种方式导致的决策结果不能等同,因此对于经验决策的模型比较研究可以帮助我们,更好地认识自己在动态环境中如何通过学习进行决策,也可以帮助有关机构更好地制定规则以助推决策行为。

## 2模型拟合

#### 2.1 数据集

本研究使用的数据来自 Erev 等(2010)组织的 Technion 预测竞赛(Technion prediction tournament, TPT),这也是当前最大的经验决策公开数据集。TPT 共提供了 120 道经验决策问题的部分反馈范式实验数据,其中每道问题包含一个确定选项和一个风险选项,风险选项有两个可能结果。在结果符号方面,这 120 道问题中有 40 道收益域问题,40 道损失域问题,40 道混合域问题。混合域问题指风险选项的可能结果一正一负,并且这其中约 20 道问题的确定选项提供收益(即结果为正),20 道问题的确定选项提供损失(即结果为负)。在概率大小方面,约有 40 道问题的风险选项包含小概率(即 p < 0.1)的高结果,40 道问题的风险选项包含小概率的低结果。具体问题以及其生成算法见附录 2、3。

这 120 道问题分为 10 组,每组包含 4 道收益域问题、4 道损失域问题和 4 道混合域问题,每道问题包含 100 个试次的选择。每组问题由 20 名被试完成,每个被试需要完成其中全部 12 道问题,对于每次的选择行为,如果选择风险选项记为 1,否则记为 0。在实验结束后,会从被试完成的试次中随机抽取一个,将其结果支付给被试,支付的货币是谢克尔(Sheqel),一种以色列货币(在这种支付形式下仍存在探索-利用权衡,因为在每个问题中选择偏好选项的次数越多,抽到选择该选项试次的概率越大)。

在不同组别的 200 名被试中,有 7 名被试在超过一半(7 道及以上)的问题中对某一选项选择次数小于等于 2 次,这种极端的选择行为可能反映了他们并未认真完成实验,其选择数据从后续分析中被排除。因此,最终的分析样本包含 193 名被试的选择数据。

#### 2.2 参数估计

本研究使用极大似然(maximum likelihood)方法估计各个模型参数。由于每个被试都完成了多个决策问题,因此本研究将收益域和损失域问题的行为数据作为拟合集,混合域问题的行为数据作为测试集,使得能够在个体层面上测试模型的泛化性能。此外,基于以下两点原因,对于拟合集每个问题,只取后 90 次的选择数据计算似然函数:第一,在选择初期,被试的选择行为有较大随机性,比如他们可能在开始的几个试次中始终选择某一选项。这种极端的选择行为或导致两选项的期望差异较大,如果使用的初始参数也较为极端(比如选择敏感性中的参数 c 较大),可能导致 RL 模型对期望较低选项,预测的选择概率的对数值低于软件处理下限(甚至在 ratio 选择函数的相关模型中,如果在一开始仅选择了某一选项,那么对另一选项预测的选择概率为 0,即对数值为负无穷),这会使得优化算法出错。为避免这一问题,本研究在限制参数取值范围和设定选项初始期望外,还对拟合数据进行了筛选,即使用从第 11 个试次开始的选择数据计算似然函数。此时被试一般对两个选项都进行了几次选择,其期望差异一般不会较为极端。第二,对于三种启发式模型,被试需要得到第二个试次的反馈,才能开始判断 win 或者 loss,因此前两个试次的选择数据本身是无法用来计算似然函数的。由于似然函数的大小和选择数据的数量也有关系,为控制各个模型使用数据个数相同,亦应从第 11 个试次开始计算似然函数。

为搜索最优参数值组合,本研究根据每个被试的选择数据对每个模型计算对数似然函数,即

$$LL_{model} = \ln L(data|model)$$

$$= \sum_{q=1}^{8} \sum_{t=11}^{100} \sum_{j=1}^{2} \ln \left( P(D(t,q) = j | X(t-1,q), Y(t-1,q)) \right) \delta_{j}(t,q)$$

(2-1)

其中q指的是拟合集的问题编号,因为每个被试共完成 4 道收益域问题和 4 道损失域问题,所以依次设为 1 至 8。Y(t-1,q)指在第q 个问题中被试从第 1 个到第t-1 个试次的选择序列,X(t-1,q)指第q 个问题中从第 1 个到第t-1 个试次的反馈序列。D(t,q)=j 指在第q 个问题的第t 个试次选择了选项j。 $\delta_j(t,q)$ 是一个哑变量,在第q 个问题中,当选项j 在第t 个试次被选择时取 1,否则取 0。

为寻找最大化对数似然函数的参数值,本研究使用网格搜索(grid search)和单纯形搜索(simplex search)方法的组合(Nelder & Mead, 1965)。该算法将每个参数取值范围的四等

分点组成网格, SS 模型网格具有 9 个格点, 其他模型网格具有 81 个格点。每个格点均作 为初始参数进行一次搜索, 并返回估计的对数似然函数最大值以及取该值时的参数值。为 避免估计的最大值是局部最大, 该算法最后比较了以不同格点为初始参数的估计最大值, 再取其中最大值所对应的参数值作为各个模型最终的参数估计。

如前所述,为避免特定模型预测的选择概率对数值低于软件处理下限,有必要对参数取值范围进行限制:对于幂效用函数, $0<\alpha<1$ ,对于对数效用函数, $0<\alpha<5$ ,同时对于两效用函数, $0<\lambda<5$ 。对于更新函数,0<A<1。对于非试次依赖选择敏感性,当使用 ratio 选择函数和 softmax 选择函数时,0<c<3,当使用 probit 选择函数时,0.25<c<3。对于试次依赖选择敏感性,当使用 ratio 选择函数和 softmax 选择函数时,-1.5<c<1.5,当使用 probit 选择函数时,-0.5<c<1.5。对于三种启发式模型,所有参数均取 0 到 1 之间的数值。

此外,如果被试选择较为极端,即前 10 个试次都只选择某个选项时,ratio 选择函数相关模型对期望较低选项,预测选择概率的对数值仍会低于软件处理下限。因此对于这类模型,在收益域问题中两选项初始期望取 0.0001,在损失域问题中两选项初始期望取 -0.0001。在 TPT 的决策问题中,这两个数值相对较小,不会对参数估计造成较大影响。而在 softmax 和 probit 选择函数的相关模型中,不会出现这一问题,因此两选项初始期望均取 0。同时,尽管对数似然函数是以多个问题的选择数据联合计算,但每个问题是独立的,即 RL 模型中的初始期望和启发式模型中的初始 stay 概率在每个问题中保持一致,不受之前经验的影响。

参数估计及后续数据分析均使用 R 4.0.0 软件完成。

根据以上方法,各模型参数估计结果如表 2.1 和表 2.2 所示。

<del></del>	
<del>***</del> 7 1	不同强化学习模型估计参数平均值
12 4.1	71191年111777778年111779817918

模型				估计	一参数	
效用函数	选择函数	选择敏感性	α	λ	A	С
	ratio	非试次依赖	0.61(0.43)	0.29(0.95)	0.57(0.30)	1.73(0.79)
	rano	试次依赖	0.65(0.43)	0.18(0.70)	0.52(0.32)	0.69(0.50)
幂效用	softmax	非试次依赖	0.47(0.41)	1.30(1.88)	0.47(0.35)	1.47(0.92)
帝双用	SOITHIAX	试次依赖	0.51(0.31)	1.02(1.64)	0.42(0.32)	0.69(0.60)
	probit	非试次依赖	0.51(0.38)	1.13(1.72)	0.43(0.34)	0.67(0.60)
		试次依赖	0.47(0.27)	1.16(1.67)	0.44(0.33)	-0.25(0.37)
	ratio	非试次依赖	3.39(2.28)	0.39(1.21)	0.44(0.32)	1.74(0.74)
	rano	试次依赖	3.13(2.39)	0.29(1.01)	0.41(0.33)	0.68(0.47)
对数效用	softmax	非试次依赖	2.29(2.32)	1.15(1.76)	0.44(0.34)	1.49(0.94)
小致双角	Soluliax	试次依赖	1.05(1.35)	0.98(1.60)	0.45(0.32)	0.64(0.64)
	probit	非试次依赖	1.41(1.68)	1.11(1.67)	0.41(0.34)	0.71(0.65)
	prooit	试次依赖	0.98(0.97)	1.10(1.66)	0.37(0.35)	-0.22(0.40)

注: 括号内数值为标准差。

表 2.2 不同启发式模型估计参数平均值

L#+ 10.0			估计参数			
模型	$p(stay win)_1$	$p(shift/loss)_1$	$ heta_{p(stay/win)}$	$ heta_{p(shift/loss)}$	$p(stay)_1$	$\theta$
WSLS-1	0.85(0.20)	0.40(0.41)	0.05(0.08)	0.29(0.29)	_	_
WSLS-2	0.82(0.13)	0.35(0.24)	0.02(0.02)	0.02(0.08)	_	
SS		_	_	_	0.75(0.28)	0.07(0.06)

注: 括号内数值为标准差

由于本研究主要关注各模型的拟合和泛化性能,而非估计参数意义,所以在此不对参数估计结果作专门分析。

# 3 拟合性能比较

## 3.1 评价指标

为评价模型的拟合性能,本研究将使用贝叶斯信息准则(Bayesian information criterion, *BIC*; Schwartz, 1978)和平均平方偏差(mean square deviation, *MSD*; Yechiam, & Busemeyer, 2005)两种统计量作为评价指标。

为计算模型的 BIC, 首先要计算其  $G^2$ , 即

$$G^2 = 2(LL_{model} - LL_{baseline})$$

(3-1)

其中  $LL_{model}$  是目标模型对数似然函数最大值,即根据所估参数计算出的对数似然函数值,而  $LL_{baseline}$  是基线模型对数似然函数最大值。对于拟合集数据,基线模型是 Bernoulli 模型,该模型假设用于计算似然函数的 720 次选择均为 Bernoulli 试验,即始终以概率 p 选择风险选项,概率 1-p 选择确定选项。不难看出,该模型仅有一个参数 p,并且当该参数等于这 720 次选择中对风险选项的选择比例时,对数似然函数可取得最大值。在不考虑参数量的情况下, $G^2$  越大,表明目标模型对数据的解释能力越好。

然而,对于拟合集数据,目标模型更好的解释能力可能仅是因为它具有更多的参数,即捕捉了选择过程中的噪声而非特征,因此 BIC 给予过多的参数以惩罚,即

$$BIC = G^2 - \Delta k \ln(N) \tag{3-2}$$

其中 $\Delta k$  是目标模型和基线模型的参数量差异,对于 SS 模型, $\Delta k = 1$ ,对于其他模型, $\Delta k$  = 3。N 是拟合数据个数,即 720。当目标模型相比基线模型对数据的解释能力更强时, BIC 为正值,并且其大小随目标模型优势的增加而增大。

虽然 BIC 考虑到了参数量的影响,但它衡量的仍是模型一步向前(one-step-ahead)预测的能力。也就是说,目标模型对每个试次选择情况的预测均依赖于该试次之前的实际选择和反馈,是一种短程预测。而 MSD 衡量的则是目标模型在无实际数据基础上对整个选择序列的长程预测能力,更具体来说,它是指基于所估参数和选项结构,模拟生成被试在某一问题中的选择序列(模拟算法见附录 4),然后比较这一序列和实际选择序列的相似程度。此外,因为在任务中每个被试最初的选择具有较大随机性,所以 MSD 最终比较的是群体实际平均选择序列和模拟平均选择序列的相似程度。由于在 TPT 中每组问题由多个被试完成,因此可将完成每个问题的被试看作一个群体,按如下方法计算 MSD

$$MSD = \frac{1}{200} \sum_{t=1}^{100} \sum_{j=1}^{2} \left( \overline{D}_{exp,j}(t) - \overline{D}_{sim,j}(t) \right)^{2}$$
$$= \frac{1}{100} \sum_{t=1}^{100} \left( \overline{D}_{exp,risk}(t) - \overline{D}_{sim,risk}(t) \right)^{2}$$

(3-3)

其中  $\overline{D}_{exp,j}(t)$ 是该问题中所有被试实际在第 t 个试次选择选项 j 的比例,  $\overline{D}_{sim,j}(t)$ 是指该问题中所有被试模拟在第 t 个试次选择选项 j 的比例。由于部分反馈范式中只有两个选项,其选择是互斥的,因此计算可以化简为公式 3-3 最后的形式,其中  $\overline{D}_{exp,risk}(t)$ 是该问题中所有被试实际在第 t 个试次选择风险选项的比例,  $\overline{D}_{sim,risk}(t)$ 是该问题中所有被试模拟在

第t个试次选择风险选项的比例。以MSD为拟合指标,则其值越小,目标模型拟合性能越好。

此外,由于 MSD 一般是针对连续变量计算的统计量,所以在计算之前还需要对实际和模拟选择数据进行平滑。对于实际数据,每个被试第 4 到第 97 次的选择用取包含该点在内附近 7 个点的移动平均代替。对于模拟数据,每个被试的选择序列进行 100 次模拟,取每个试次选择的平均值作为模拟选择。同时,为放大不同模型 MSD 的差异,本研究在计算选择比例时使用百分数而非小数。

由于两个统计量的分布可能并非正态分布,因此除比较其绝对大小外,还要考虑其相对大小。传统的非参数方法往往对所有数据进行等级转换,但 Stott(2006)使用的平均等级 (average rank)统计量可以更精确地在被试内比较模型的相对优劣。它的计算方法较为简单:选定评价指标后,对于每个被试(或问题),将不同模型根据该指标由优到劣进行等级转换(在本研究中,最优记为1,最劣记为15),然后取在所有被试(或问题)中的平均等级以衡量该模型。为从多个方面比较模型的拟合性能,在比较各模型 BIC 和 MSD 绝对大小的同时,本研究也比较了其平均等级。

### 3.2 BIC 比较结果

不同模型 BIC 统计量分布如表 3.1 所示,该表中还列举了每个模型 BIC 统计量的 Shapiro-Wilk 正态性检验(Shapiro & Wilk, 1965)结果。由于 RL 模型和启发式模型中均存在 个别模型的 BIC 统计量未通过正态性检验(p < .05),因此本研究将使用对齐等级转换 (aligned rank transform; ART)重复测量方差分析来确定两种模型中的最优模型。ART 是一种非参数方法,能够处理多因素设计中的交互作用,并且克服了传统方法因直接对数据进行等级转换而引发第一类错误膨胀的缺点(Wobbrock, Findlater, Gergle, & Higgins, 2011)。

表 3.1 不同模型 BIC 分布

模型						BIC	分布			
效用函数	选择函数	选择敏感性	10%分位数	中位数	90%分位数	平均数	标准差	平均等级	S-W 检验 显著性	优于基线 模型比例
	mati a	非试次依赖	-68.69	72.88	292.17	93.62	147.93	12.27	0.066	76.17
	ratio	试次依赖	-67.38	82.76	298.41	100.67	152.66	11.99	0.179	76.68
幂效用	softmax	非试次依赖	-31.55	154.03	336.52	144.95	147.01	7.92	0.505	84.97
帝双用	sorumax	试次依赖	-13.71	165.19	353.98	157.79	144.31	8.06	0.887	87.05
	nrohit	非试次依赖	-24.68	142.47	342.17	140.99	147.76	9.62	0.195	85.49
	probit	试次依赖	-27.08	138.78	338.31	137.58	146.38	9.80	0.424	84.97
	ratio	非试次依赖	-46.43	147.78	502.98	182.23	210.09	7.74	< .001	82.38
	Tauo	试次依赖	-32.84	143.58	431.48	167.25	186.72	8.63	0.006	84.97
对数效用	softmax	非试次依赖	-31.84	147.80	353.09	147.80	144.39	8.28	0.822	84.97
MXX用	SOITHIAX	试次依赖	-14.21	158.48	353.44	156.41	143.73	8.37	0.892	86.53
	probit	非试次依赖	-32.66	137.77	345.71	136.69	146.45	10.38	0.300	85.49
	proon	试次依赖	-29.59	130.87	339.49	136.48	145.90	10.16	0.558	84.97
WSLS-1			181.45	509.54	763.27	474.76	230.67	2.41	0.002	98.45
WSLS-2			219.72	525.82	756.87	493.63	214.02	1.89	0.003	98.45
SS			178.94	504.50	760.12	476.97	228.16	2.47	0.001	97.93

注: 所有模型优于基线 Bernoulli 模型的比例, 即 BIC > 0 的比例均显著大于  $50\%(\alpha = .05, 单侧检验)$ 。

所有统计采用 .05 的  $\alpha$  水平,以效用函数(幂效用、对数效用),选择函数(ratio、softmax、probit)和选择敏感性(非试次依赖、试次依赖)为自变量,BIC 为因变量,进行 ART 重复测量方差分析。结果显示效用函数主效应显著,F(1,2112)=92.35,p<.001。选择函数主效应显著,F(2,2112)=48.76,p<.001。效用函数和选择函数的交互作用显著,F(2,2112)=103.6,p<.001。选择函数和选择敏感性的交互作用显著,F(2,2112)=3.25,p=.039。其他效应(包括交互作用)均不显著。因此任一自变量的效应都无法单独分析。

固定选择敏感性为非试次依赖,以效用函数和选择函数为自变量,BIC 为因变量,进行 ART 重复测量方差分析。结果显示效用函数主效应显著,F(1,960)=60.88,p<.001。选择函数主效应显著,F(2,960)=17.17,p<.001。效用函数和选择函数的交互作用显著,F(2,960)=65.23,p<.001。

由于 ART 方法在检验每个效应之前都对数据进行转换,因此无法直接检验简单效应,但可以检验简单效应差异。因此,以幂效用函数相关模型平均 BIC-对数效用函数相关模型平均 BIC 作为简单效应进行差异检验(需要注意的是,差异检验时使用的仍是经对齐等级转换的 BIC, 所以此处的简单效应并不等价于原始 BIC 的平均数差异,但图 3.1 中所示

的简单效应则仍为原始 *BIC* 的平均数差异)。结果显示在选择函数为 softmax 和 probit 时,简单效应无显著差异,t(960) = -0.924,p = .356。在选择函数为 ratio 时,简单效应显著低于在选择函数为 softmax 时,t(960) = -9.40,p < .001,也显著低于在选择函数为 probit 时,t(960) = -10.32,p < .001。

图 3.1 为固定选择敏感性为非试次依赖时,不同效用函数和选择函数组合 RL 模型的 BIC 分布小提琴图(violin plot),其中显示了每个模型的 BIC 数据和相应核密度估计。结合简单效应差异检验和平均等级结果,可以推断当选择敏感性为非试次依赖时,最优模型为对数效用 + ratio + 非试次依赖 RL 模型,最劣模型为幂效用 + ratio + 非试次依赖 RL 模型。

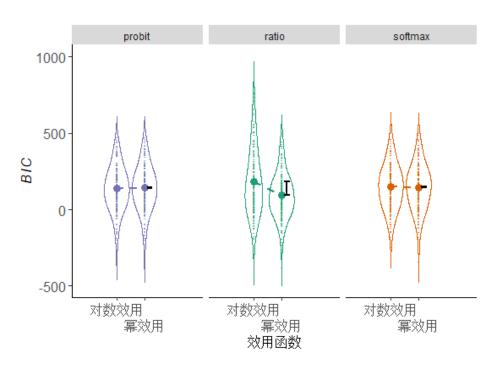


图 3.1 不同效用函数和选择函数组合 RL 模型的 BIC 分布(非试次依赖)

注:加粗圆点代表 BIC 平均数,黑色线段长度代表效用函数简单效应的绝对大小(以原始 BIC 计算)。

固定选择敏感性为试次依赖,以效用函数和选择函数为自变量,BIC为因变量,进行ART 重复测量方差分析。结果显示效用函数主效应显著,F(1,960)=31.88,p<.001。选择函数主效应显著,F(2,960)=32.65,p<.001。效用函数和选择函数的交互作用显著,F(2,960)=37.61,p<.001。

同样,以幂效用函数相关模型平均 BIC- 对数效用函数相关模型平均 BIC 作为简单效应进行差异检验。结果显示在选择函数为 softmax 和 probit 时,简单效应无显著差异,t(960) = -0.090,p = .929。在选择函数为 ratio 时,简单效应显著低于在选择函数为 softmax 时,t(960) = -7.47,p < .001,也显著低于在选择函数为 probit 时,t(960) = -7.56,p < .001。

图 3.2 为固定选择敏感性为试次依赖时,不同效用函数和选择函数组合 RL模型的 BIC 分布小提琴图。结合简单效应差异检验和平均等级结果,可以推断当选择敏感性为试次依赖时,最优模型为对数效用 + ratio + 试次依赖 RL模型,最劣模型为幂效用 + ratio + 试次依赖 RL模型。

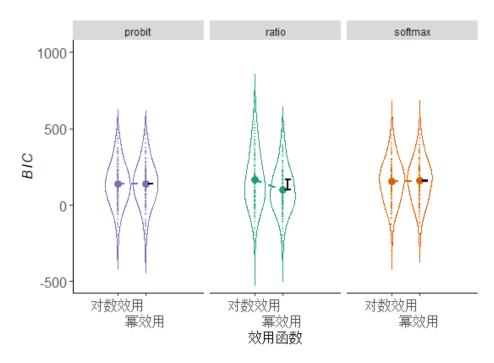


图 3.2 不同效用函数和选择函数组合 RL 模型的 BIC 分布(试次依赖)

注:加粗圆点代表 BIC 平均数,黑色线段长度代表效用函数简单效应的绝对大小(以原始 BIC 计算)。

最后,不论从原始 BIC 的平均数还是平均等级上,对数效用 + ratio + 非试次依赖 RL 模型均明显优于对数效用 + ratio + 试次依赖 RL 模型,因此 RL 模型中的最优模型应 为对数效用 + ratio + 非试次依赖 RL 模型。

所有统计采用 .05 的  $\alpha$  水平,以模型类型(WSLS-1、WSLS-2、SS)为自变量,*BIC* 为因变量,进行 ART 重复测量方差分析。结果显示模型类型效应显著,F(2,384)=29.42,p<.001。事后比较显示 WSLS-1 和 SS 模型的 *BIC* 无显著差异,t(384)=-0.71,p=.758。 WSLS-2 模型的 *BIC* 显著高于 WSLS-1 模型,t(384)=6.97,p<.001,也显著高于 SS 模型,t(384)=6.26,p<.001 (p 值经 tukey 方法调整)。图 3.3 为不同启发式模型的 *BIC* 分布小提琴图,结合事后比较和平均等级的结果,表明 WSLS-2 是启发式模型中的最优模型。

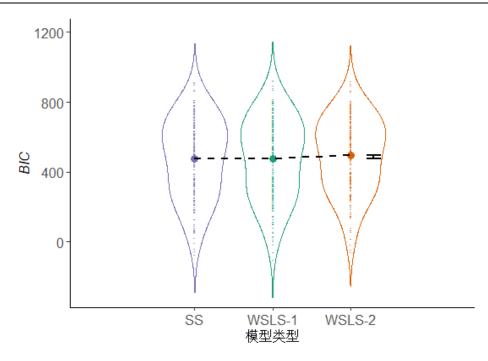


图 3.3 不同启发式模型的 BIC 分布

注:加粗圆点代表 BIC 平均数,黑色线段长度代表 WSLS-2 模型 BIC 平均数,与 WSLS-1 和 SS 模型 BIC 合并平均数的差异的绝对大小(以原始 BIC 计算)。

对 WSLS-2 模型和对数效用 + ratio + 非试次依赖 RL 模型的 BIC 进行配对样本 Wilcoxon 秩和检验,结果显示 WSLS-2 模型的 BIC 显著高于对数效用 + ratio + 非试次依赖 RL 模型的 BIC, V=14, p<.001。因此,启发式最优模型在以 BIC 为评价标准的拟合性能上表现更优。此外,在平均等级方面,启发式模型整体也优于 RL 模型。

# 3.3 拟合 MSD 比较结果

不同模型拟合 MSD 统计量分布如表 3.2 所示。

表 3.2 不同模型拟合 MSD 分布

模型			拟合 MSD 分布							
效用函数	选择函数	选择敏感性	10%分位数	中位数	90%分位数	平均数	标准差	平均等级	S-W 检验 显著性	
	4: -	非试次依赖	58.35	238.34	916.04	382.79	405.94	10.54	< .001	
	ratio	试次依赖	52.23	207.37	1002.05	351.93	340.10	9.76	< .001	
官並田	a often av	非试次依赖	43.39	155.25	784.70	307.62	323.65	8.23	< .001	
幂效用	softmax	试次依赖	37.80	164.98	577.07	250.88	245.39	6.54	< .001	
	probit	非试次依赖	48.59	177.43	853.44	312.97	327.41	8.69	< .001	
		试次依赖	46.95	165.67	679.26	279.96	288.82	7.38	< .001	
		非试次依赖	68.80	260.56	1047.25	402.91	404.41	10.98	< .001	
	ratio	试次依赖	57.76	219.66	1014.53	364.88	351.51	9.95	< .001	
北米光田	Ct	非试次依赖	44.62	167.67	771.19	289.65	306.25	7.69	< .001	
对数效用	softmax	试次依赖	38.99	162.01	566.97	245.89	241.65	6.80	< .001	
		非试次依赖	50.46	179.45	692.30	291.09	300.37	8.41	< .001	
	probit	试次依赖	47.51	174.38	640.76	278.36	274.96	8.20	< .001	
WSLS-1			31.76	108.79	456.97	202.47	266.25	5.43	< .001	
WSLS-2			37.96	105.34	471.01	189.67	257.87	5.48	< .001	
SS			38.27	112.91	513.82	224.74	272.35	5.95	< .001	

所有统计采用 .05 的  $\alpha$  水平,以效用函数(幂效用、对数效用),选择函数(ratio、softmax、probit)和选择敏感性(非试次依赖、试次依赖)为自变量,拟合 MSD 为因变量,进行 ART 重复测量方差分析,结果显示选择函数主效应显著,F(2,869)=51.36,p<.001。选择敏感性主效应显著,F(1,869)=7.23,p=.007。其他效应(包括交互作用)均不显著。

选择函数的事后比较显示,softmax 选择函数相关模型和 probit 选择函数相关模型的拟合 MSD 无显著差异,t(869) = -1.76,p = .182。softmax 选择函数相关模型的拟合 MSD 显著低于 ratio 选择函数相关模型,t(869) = -9.53,p < .001。probit 选择函数相关模型的拟合 MSD 显著低于 ratio 选择函数相关模型,t(869) = -7.76,p < .001 (p 值经 tukey 方法调整)。

选择敏感性的事后比较显示,非试次依赖选择敏感性相关模型的拟合 MSD 显著高于试次依赖选择敏感性相关模型,t(869)=2.69,p=.007。

图 3.4 为不同选择函数和选择敏感性组合 RL 模型的拟合 MSD 分布小提琴图。结合事后比较和平均等级结果, RL 模型中的最优模型应为幂效用 + softmax + 试次依赖 RL 模型和对数效用 + softmax + 试次依赖 RL 模型。

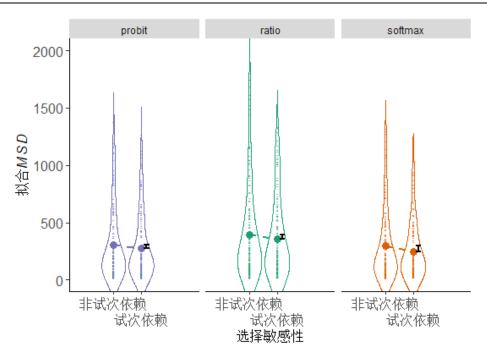


图 3.4 不同选择函数和选择敏感性组合 RL 模型的拟合 MSD 分布

注:每列散点包含幂效用函数和对数效用函数两个相关模型的数据,加粗圆点代表拟合 *MSD* 平均数,黑色线段长度代表选择敏感性简单效应的绝对大小(以原始拟合 *MSD* 计算)。

所有统计采用 .05 的  $\alpha$  水平,以模型类型(WSLS-1、WSLS-2、SS)为自变量,拟合 MSD 为因变量,进行 ART 重复测量方差分析。结果显示模型类型效应不显著,F(2,158) = 1.47, p = .232。图 3.5 为不同启发式模型的拟合 MSD 分布小提琴图,结合平均等级的结果,可以推断启发式模型中的最优模型应为 WSLS-1 和 WSLS-2 模型。

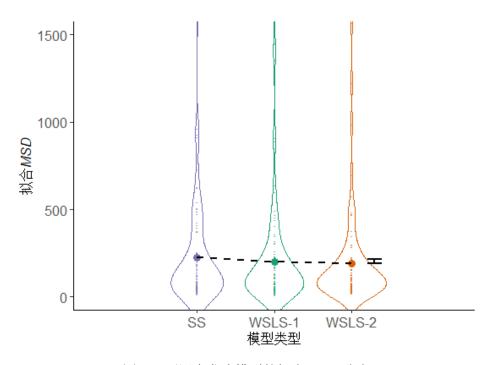


图 3.5 不同启发式模型的拟合 MSD 分布

注:加粗圆点代表拟合 *MSD* 平均数,黑色线段长度代表 WSLS-2 模型拟合 *MSD* 平均数,与 WSLS-1和 SS 模型拟合 *MSD* 合并平均数的差异的绝对大小(以原始拟合 *MSD* 计算)。

对启发式最优模型和 RL 最优模型的拟合 MSD 进行配对样本 Wilcoxon 秩和检验,结果显示 WSLS-1 模型的拟合 MSD 显著低于两个 RL 最优模型, $V_1$ =2250, $p_1$ =.003, $V_2$ =2231, $p_2$ =.003。WSLS-2 模型的拟合 MSD 也显著低于两个 RL 最优模型, $V_1$ =2144, $p_1$ =.012, $V_2$ =2162, $p_2$ =.009。

因此,启发式最优模型在以拟合 MSD 为评价标准的拟合性能上表现更优。此外,在平均等级方面,启发式模型整体也优于 RL 模型,但优势不如在以 BIC 为评价指标时那么明显。

### 3.4 讨论

对于 RL 模型的函数选择,综合两种拟合性能评价指标的比较结果,在效用函数方面,对数效用函数优于幂效用函数。在选择函数方面,softmax 选择函数略优于 probit 选择函数(这点主要由平均等级反映),而 ratio 选择函数的可用性仍有待考证。在选择敏感性方面,两种选择敏感性的优劣仍有待进一步比较。

此外,在以 BIC 为评价指标时,一个有趣的发现是观察到了效用函数和选择函数显著的交互作用,即在每种选择敏感性相关模型中,对数效用 + ratio 模型和幂效用 + ratio 模型均分别是最优和最劣模型。事实上,在本研究比较的 6 类效用函数和选择函数组合的 RL 模型中,对数效用 + ratio 相关模型是唯一能严格解释量级效应的模型,而幂效用 + ratio 相关模型则是唯一完全不能解释量级效应的模型(具体证明见附录 1),这从定量模型比较角度支持了经验决策中存在量级效应。

对于启发式模型,综合两种拟合性能评价指标的比较结果,WSLS-2 始终是最优模型,而 WSLS-1 略优于 SS 模型,这支持了 WSLS 策略设置两组条件概率的假设。而从参数估计结果(见表 2.2)也可发现,两个 WSLS 模型中被试初始的  $p(stay|win)_1$  和  $p(shift/loss)_1$  之和并不接近 1,这与以往研究一致(Worthy & Maddox, 2014; Worthy et al., 2012),即决策者在win 试次和 loss 试次后的 stay 概率存在初始差异。

至于决策者的策略使用差异,在两种评价指标下启发式模型都表现出相对 RL 模型明显的优势,尽管在以拟合 MSD 为评价指标时这种优势有所减小。这显示了决策者可能更倾向于仅依赖试次间结果的相对大小进行选择,或者这种简单策略在整个决策过程中占较

大权重。

### 4 泛化性能比较

一个复杂的模型在对拟合集数据的解释上优于一个简单的模型,既可能是因为它确实更好地捕捉了行为的特征,也可能是它把拟合集数据中的噪声当作特征捕捉。除了使用BIC 来惩罚过多的参数外,另一种有效的方法是基于已估参数直接比较模型对一批新数据的预测能力,也就是泛化性能。本质来看,拟合性能关注的是模型的后验(posterior)解释能力,因为它的预测数据是已观测数据。而泛化性能关注的是模型的先验(prior)解释能力,因为它的预测数据是未观测数据(Busemeyer & Wang, 2000)。在某种意义上,泛化性能或许比拟合性能更重要,因为多数定量模型的设计目标是在普遍情境下对行为作出精确预测,这需要模型具有跨情境的稳定性,也就是基于由已观测数据估计的参数,能对未观测数据进行良好的预测。

### 4.1 评价指标

为评价模型的泛化性能,本研究将使用  $G^2$  和 MSD 两种统计量作为评价指标。此外,为在个体层面上测试模型的泛化性能,本研究将使用每个被试在混合域问题中后 90 个试次的选择作为测试集(选用后 90 次的原因见 2.2 节)。

 $G^2$ 的计算方式见公式 2-1 和 3-1,但要做两点调整。第一,计算对数似然函数值时使用的是拟合集估计的参数和测试集的行为数据。第二,由于目标模型参数不是由测试集数据进行估计,因此基线模型要改成随机选择模型,即假设被试在每次选择时都等可能选择任一选项的模型,该模型无自由参数,对于每个被试对数似然函数值相同。之所以选用 $G^2$ ,是因为这是一批新数据,不必考虑过拟合问题,也就不必对过多的参数做出惩罚。 $G^2$ 越大,目标模型泛化性能越好。

*MSD* 的计算方法见公式 3-3,同样地,模型使用参数为拟合集所估参数,使用数据为测试集行为数据。*MSD* 越小,目标模型泛化性能越好。

此外,由于 ratio 选择函数不适合混合域问题,因此相关模型将不进行泛化测试。但在评价模型泛化性能时,仍会使用平均等级统计量,只是此时被试内最劣模型的等级为11。

#### 4.2 G2 比较结果

不同模型  $G^2$  统计量分布如表 4.1 所示。

表 4.1 不同模型 G2分布

模型				$G^2$ 分布							
效用函数	数 选择函数 选择敏感性		10%分位数	中位数	90%分位数	平均数	标准差	平均等级	S-W 检验 显著性	优于基线 模型比例	
	C	非试次依赖	-1077.36	42.82	241.54	-225.15	968.89	6.80	< .001	59.59	
幂效用	softmax	试次依赖	-614.54	28.68	241.42	-207.93	1141.27	7.29	< .001	58.03	
希双用		非试次依赖	-810.54	32.82	223.36	-273.87	1273.00	7.33	< .001	57.51	
	probit	试次依赖	-1070.44	5.50	232.04	-342.06	1689.51	7.99	< .001	51.81	
	<b>C</b>	非试次依赖	-240.82	45.02	241.94	-39.72	469.39	6.55	< .001	61.66	
对数效用	softmax	试次依赖	-501.50	25.80	233.00	-169.45	968.06	7.45	< .001	57.51	
MXX用	nrohit	非试次依赖	-384.24	47.90	217.04	-81.20	563.39	7.34	< .001	61.66	
	probit	试次依赖	-806.72	12.04	222.84	-349.43	1692.49	8.17	< .001	53.37	
WSLS-1			122.50	322.72	444.19	280.19	172.34	2.55	< .001	97.41	
WSLS-2			142.20	324.08	441.78	267.06	445.21	1.99	< .001	96.89	
SS			111.64	322.98	434.20	288.16	148.01	2.54	< .001	97.92	

注: 如果模型优于基线随机选择模型的比例,即  $G^2 > 0$  的比例显著大于  $50\%(\alpha = .05$ ,单侧检验),则 在最后一列对应单元格用灰色填充。

所有统计采用 .05 的  $\alpha$  水平,以效用函数(幂效用、对数效用),选择函数(softmax、probit)和选择敏感性(非试次依赖、试次依赖)为自变量, $G^2$  为因变量,进行 ART 重复测量方差分析,结果显示选择函数主效应显著,F(1,1344)=12.22,p<.001,选择敏感性主效应显著,F(1,1344)=10.89,p=.001。其他效应(包括交互作用)均不显著。

选择函数的事后比较显示, softmax 选择函数相关模型的  $G^2$  显著高于 probit 选择函数相关模型, t(1344)=3.50, p=.001。

选择敏感性的事后比较显示,非试次依赖选择敏感性相关模型的  $G^2$ 显著高于试次依赖选择敏感性相关模型,t(1344)=3.30,p=.001。

图 4.1 为不同选择函数和选择敏感性组合 RL 模型的  $G^2$  分布小提琴图。结合事后比较和平均等级结果,RL 模型中的最优模型应为幂效用 + softmax + 非试次依赖 RL 模型和对数效用 + softmax + 非试次依赖 RL 模型。

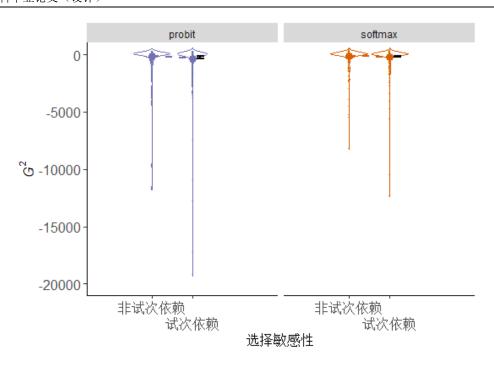


图 4.1 不同选择函数和选择敏感性组合 RL 模型的 G2分布

注:每列散点包含幂效用函数和对数效用函数两个相关模型的数据,加粗圆点代表  $G^2$  平均数,黑色线段长度代表选择敏感性简单效应的绝对大小(以原始  $G^2$  计算)。

所有统计采用 .05 的  $\alpha$  水平,以模型类型(WSLS-1、WSLS-2、SS)为自变量, $G^2$  为因变量,进行 ART 重复测量方差分析。结果显示模型类型效应显著,F(2,384)=8.28,p<.001。事后比较显示 WSLS-1 和 SS 模型的  $G^2$  无显著差异,t(384)=-1.08,p=.527。 WSLS-2 模型的  $G^2$  显著高于 WSLS-1 模型,t(384)=3.94,p<.001,也显著高于 SS 模型,t(384)=2.86,t(384)=2.86,t(384)=3.94 t(384)=3.94 t(384)=3.94

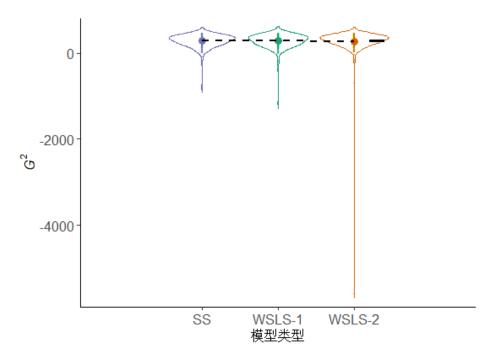


图 4.2 不同启发式模型的  $G^2$  分布

注:加粗圆点代表  $G^2$  平均数,黑色线段长度代表 WSLS-2 模型  $G^2$  平均数,与 WSLS-1 和 SS 模型  $G^2$  合并平均数的差异的绝对大小(以原始  $G^2$  计算)。

对 WSLS-2 模型和两个 RL 最优模型的  $G^2$  进行配对样本 Wilcoxon 秩和检验,结果显示 WSLS-2 模型的  $G^2$  显著高于两个 RL 最优模型, $V_1$ =705, $p_1$ <.001, $V_2$ =743, $p_2$ <.001。因此,启发式最优模型在以  $G^2$  为评价标准的泛化性能上表现更优。此外,在平均等级方面,启发式模型整体也优于 RL 模型。

对于不同 RL 模型和 WSLS-2 模型,需要指出的是它们均包含个别极小的  $G^2$ ,因此不能仅依据表 4.1 中的平均数对模型泛化性能作判断。首先,尽管不同 RL 模型  $G^2$  平均数均小于 0,但是从中位数和优于基线模型的比例上看,大部分 RL 模型相对基线模型仍能对测试集数据做出更好的解释。其次,尽管 WSLS-2 的  $G^2$  平均数在启发式模型中最低,但中位数是高于其他两个模型的,并且在平均等级方面也表现出明显的优势。由于 ART 方法和平均等级都能在一定程度上减小极端数据的影响,因此 RL 最优模型和启发式最优模型的判定仍主要依据这两个指标。

## 4.3 泛化 MSD 比较结果

不同模型的泛化 MSD 统计量分布如表 4.2 所示。

SS

表 4.2 不同模型泛化 MSD 分布										
模型				泛化 MSD 分布						
效用函数	选择函数	选择敏感性	10%分位数	中位数	90%分位数	平均数	标准差	平均等级	S-W 检验 显著性	
	softmax	非试次依赖	98.02	417.14	885.21	480.02	372.66	6.90	.001	
幂效用	softmax	试次依赖	90.38	433.40	968.06	536.32	379.79	7.85	.023	
帝双用	probit	非试次依赖	82.23	391.83	943.21	498.53	375.03	7.70	.007	
		试次依赖	88.91	430.68	941.91	526.33	389.87	7.68	.004	
	softmax	非试次依赖	77.40	319.98	886.05	433.95	346.33	5.58	.002	
对数效用	SOITHIAX	试次依赖	83.76	399.49	964.10	489.42	363.20	6.10	.008	
小蚁双用	nuchit	非试次依赖	101.63	366.53	956.58	463.78	354.25	6.85	.002	
	probit	试次依赖	85.69	373.01	922.05	477.79	363.47	6.75	.003	
WSLS-1			39.89	172.70	594.44	286.40	416.14	3.53	<.001	
WSLS-2			31.58	103.50	640.14	255.82	441.32	3.00	<.001	

所有统计采用 .05 的  $\alpha$  水平,以效用函数(幂效用、对数效用),选择函数(softmax、probit)和选择敏感性(非试次依赖、试次依赖)为自变量,泛化 MSD 为因变量,进行 ART 重复测量方差分析,结果显示效用函数主效应显著,F(1,273)=13.63,p<.001。选择敏感性主效应显著,F(1,273)=11.40,p=.001。

120.21

912.69

313.36 441.16

4.08

<.001

38.04

效用函数的事后比较显示,幂效用函数相关模型的泛化 MSD 显著高于对数效用函数相关模型,t(273)=3.69,p<.001。

选择敏感性的事后比较显示,试次依赖选择敏感性相关模型的泛化 MSD 显著高于非试次依赖选择敏感性相关模型,t(273)=3.38,p=.001。

图 4.3 为不同效用函数和选择敏感性组合 RL 模型的泛化 MSD 分布小提琴图,结合事后比较和平均等级结果,可以推断 RL 模型中的最优模型为对数效用 + softmax + 非试次依赖 RL 模型。

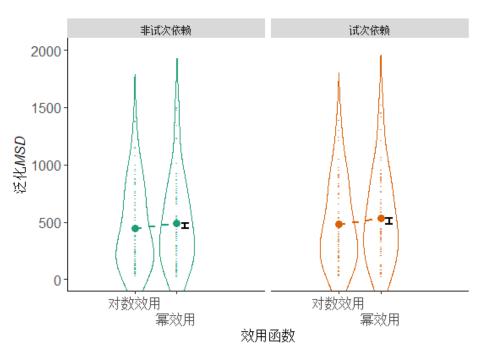


图 4.3 不同效用函数和选择敏感性组合 RL 模型的泛化 MSD 分布

注:每列散点包含 softmax 选择函数和 probit 选择函数两个相关模型的数据,加粗圆点代表泛化 *MSD* 平均数,黑色线段长度代表效用函数简单效应的绝对大小(以原始泛化 *MSD* 计算)。

所有统计采用 .05 的  $\alpha$  水平,以模型类型(WSLS-1、WSLS-2、SS)为自变量,泛化 MSD 为因变量,进行 ART 重复测量方差分析。结果显示模型类型效应显著,F(2,78)=7.61,p=.001。事后比较显示 WSLS-1 和 SS 模型的泛化 MSD 无显著差异,t(78)=-0.30,p=.951。 WSLS-2 模型的泛化 MSD 显著低于 WSLS-1 模型, t(78)=-3.22, p=.005,也显著低于 SS 模型, t(78)=-3.52, p=.002 (p 值经 tukey 方法调整)。图 4.4 为不同启发式模型的泛化 MSD 分布小提琴图,结合事后比较和平均等级结果,可以推断 WSLS-2 是启发式模型中的最优模型。

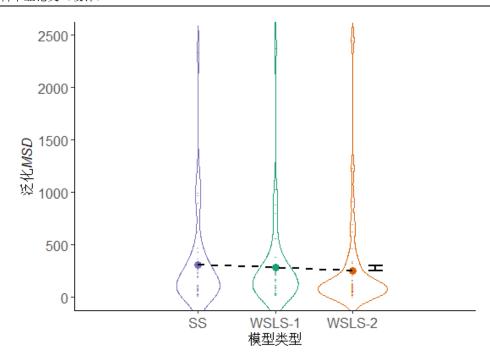


图 4.4 不同启发式模型的泛化 MSD 分布

注:加粗圆点代表泛化 *MSD* 平均数,黑色线段长度代表 WSLS-2 模型泛化 *MSD* 平均数,与 WSLS-1和 SS 模型泛化 *MSD* 合并平均数的差异的绝对大小。

对 WSLS-2 模型和对数效用 + softmax + 非试次依赖 RL 模型的泛化 MSD 进行配对样本 Wilcoxon 秩和检验,结果显示 WSLS-2 模型的泛化 MSD 显著低于对数效用 + softmax + 非试次依赖 RL 模型,V=679,p < .001。因此,启发式最优模型在以泛化 MSD 为评价标准的泛化性能上表现更优。此外,在平均等级方面,启发式模型整体也优于 RL 模型,但优势不如在以  $G^2$  为评价指标时那么明显。

#### 4.4 讨论

对于 RL 模型的函数选择,综合两种泛化性能评价指标的比较结果,在效用函数方面,对数效用函数优于幂效用函数。在选择函数方面,softmax 选择函数优于 probit 选择函数。在选择敏感性方面,非试次依赖选择敏感性优于试次依赖选择敏感性。

对于启发式模型,综合两种泛化性能评价指标的比较结果,WSLS-2 始终是最优模型,这不仅再次支持了 WSLS 策略设置两组条件概率的假设,还支持了 WSLS-2 在单个 win或 loss 试次后同时调整两组条件概率的假设。事实上,WSLS-1 模型和 SS 模型均劣于 WSLS-2 模型,但彼此之间无明显优劣的结果也在一定程度上支持了上述两个假设。因为如 1.4 节所述,SS 模型实际是一种受限的 WSLS-2 模型,它保留了 WSLS-2 模型中单个

win 或 loss 试次后对所有 stay 概率进行调整的假设,但违背了设置两组条件概率的假设。而 WSLS-1 模型所满足和违背的假设则正好相反。这两个模型在泛化性能上无明显优劣,表明 WSLS-2 模型的两个假设可能是同等重要的。

至于决策者的策略使用差异,与拟合性能比较结果类似,在两种泛化性能评价指标下启发式模型都表现出相对 RL 模型明显的优势,尽管在以泛化 MSD 为评价指标时这种优势有所减小。这显示了决策者可能更倾向于仅依赖试次间结果的相对大小进行选择,或者这种简单策略在整个决策过程中占较大权重。

# 5总讨论

## 5.1 强化学习模型函数选择

综合拟合性能和泛化性能比较结果,在 RL 模型的效用函数方面,对数效用函数相对 幂效用函数表现出明显的优势。由于本研究采用的对数效用函数是为解释描述决策中的量 级效应而被提出(Scholten & Read, 2010),并且已在描述决策研究中对该效应相关数据表现 出良好拟合性能(Bouchouicha & Vieider, 2017; Scholten & Read, 2014),因此它在本研究中 展现的优势可能反映了经验决策中也存在量级效应。事实上,尽管未对不同量级问题中安 全选项的选择比例作直接比较,但 Erev, Ert 和 Yechiam(2008),以及 Konstantinidis, Taylor 和 Newell(2018)确实在使用部分反馈范式的实验中观察到了量级效应的趋势。这一结果表 明定量模型比较确实可以为理论的提出或完善提供帮助。

在RL模型的选择函数方面,softmax 选择函数优于 probit 选择函数,而 ratio 选择函数的适用性还有待进一步考证。对于 ratio 选择函数,尽管它的使用情境有一定限制性,但是它的价值在描述决策中已得到一定体现。例如在 Stott(2006)对由不同效用函数、权重函数和选择函数组合成的 PT 模型的比较中,发现 ratio 选择函数是整体表现最优的选择函数。而在 Scholten 和 Read(2014)对量级效应相关数据的模型拟合工作中,也发现 ratio 选择函数优于常用的基于差值的选择函数。此外,ratio 选择函数符合 Weber-Fechner 定律,表明它的导出具有一定心理物理学依据。而本研究的结果显示,它在经验决策中可能也是一个可用的选择函数。

对于 softmax 选择函数,它本身是在经验决策中最广泛使用的选择函数(Ahn et al., 2008; Lejarraga & Hertwig, 2017; Yechiam & Busemeyer, 2005),而本研究再次证明了它良好

的性能。并且,虽然该函数直观上是一种 LCA 型的选择函数,但它也可由 RUT 导出。 Yellott (1977)已证明当 RUT 中的随机成分遵从独立同分布的 gumbel 分布时,决策者的选 择遵从 softmax 选择函数,而且这一结论可以推广到多个选项的决策问题。这表明 softmax 选择函数具有良好的数学性质。事实上,softmax 选择函数相对 probit 选择函数的优势可 能也是来自其函数形状上的优势。尽管标准 logistic 分布和标准正态分布形状类似,但前 者峰度系数更高,尾部更厚,这使得当选项期望相差较大时,如果两选择函数中选择敏感 性一致(即 softmax 选择函数中的  $\theta(t)$ 和 probit 选择函数中的  $1/\theta(t)$ 相等),则 softmax 选择 函数对期望较小选项预测的选择概率会大于 probit 选择函数。softmax 选择函数这种对极 端差异的不敏感性可以较好地适应经验决策中的序次依赖(sequential dependence)现象 (Erev & Barron, 2005)。该现象是指即使在选择后期,决策者仍会在不同选项间发生交替, 也就是从较优选项得到一个较好结果后仍会偶尔选择较劣选项(从 WSLS 角度,可以理解 为在 win 试次后选择 shift)。由于 softmax 选择函数对极端差异不敏感,因此相关模型允 许这种探索。但 probit 选择函数只能通过调低选择敏感性来适应这种现象(可见参数估计 结果), 这会使得相关模型对选择行为的预测更接近随机水平, 从而显示出较差的拟合和 泛化性能。此外, Daw 等(2006)使用功能性磁共振成像记录了一项经验决策任务中决策者 的皮层活动,发现它们与 softmax 相关模型对不同选项的预测概率有显著相关,表明该选 择函数有神经生理学方面的合理性,从另一角度支持了它在经验决策中的使用。

在 RL 函数的选择敏感性方面,非试次依赖选择敏感性在以 BIC、G<sup>2</sup>和泛化 MSD 为评价指标时表现均优于试次依赖选择敏感性,这与以往研究也是一致的(Ahn et al., 2008; Beitz, Salthouse, & Davis, 2014)。然而,当前对选择敏感性的试次依赖性做出判断仍为时过早。首先,在以拟合 MSD 为评价指标时,试次依赖选择敏感性相关模型相对非试次依赖选择敏感性相关模型表现出明显的优势,而且这种优势在整体层面和最优模型上均有体现。其次,在 TPT 中每个决策者完成了多个问题,虽然这使得能够在个体层面上比较模型的泛化性能,但为捕捉选择敏感性的动态变化带来了困难。例如,决策者的选择敏感性可能在较先呈现的问题中主要受练习效应影响,而在较后呈现的问题中主要受疲劳效应影响,那么将这些问题联合进行参数估计时,试次依赖选择敏感性的表现可能反而不如非试次依赖选择敏感性,因为在不同的问题中其参数值本身就是不同的。最后,即使在单个问题中,选择敏感性可能也并非是单调变化的。例如,它可能同样受练习和疲劳的双重影响而呈现倒 U 形变化,那么本研究采用的函数形式也无法捕捉这种动态过程。

## 5.2 启发式模型选择

综合拟合性能和泛化性能比较结果,在启发式模型上,WSLS-2模型优于WSLS-1和SS模型。这表示该模型设置两组条件概率,以及单个win或loss试次后对两组条件概率均进行调整的假设是必要的,而泛化性能比较结果甚至显示了这两个假设同等重要。

关于前一假设,决策者 win 试次后倾向于 stay,而在 loss 试次后倾向于 shift 是一个直观的推理结果,也是 RL 模型所预测的选择模式(Erev & Barron, 2005)。而本研究的结果表明,被试初始的 stay 倾向就存在差异,而非仅通过试次的 win 或 loss 进行调整,支持了 WSLS 策略设置两组条件概率。

关于后一假设,尽管在单个 win 或 loss 试次后仅调整与下一个试次有关的条件概率似乎是合理的,但是它不允许在 win 试次后的 stay 概率,以及在 loss 试次后的 shift 概率减小。事实上,由于经验决策中小概率权重过低的现象,人们一般在选择后期会偏好确定选项或以大概率产生高结果的风险选项。如果是后者,则偶尔出现的 loss 试次并不会对下一个试次的 stay 概率造成较大影响,这种行为特征无法被 WSLS-1 模型所捕捉。而在WSLS-2 模型中,在后期两种 stay 概率都会趋近于 1,能够较好地描述被试稳定的偏好。

# 5.3 策略使用差异

不论是拟合性能比较还是泛化性能比较,启发式模型相较于 RL 模型在整体层面和最优模型层面均表现出明显优势,这种优势可以从单一策略和混合策略两方面分析。

如果人们在经验决策中使用单一策略,那么这一结果表明优势策略是 WSLS 策略,而这一结论是得到一些研究支持的。Otto,Taylor 和 Markman(2011)使用概率学习任务,发现在无认知负荷的条件下,WSLS 模型对决策者的行为数据拟合得更好,而在认知负荷条件下,RL 模型对决策者的行为数据拟合得更好。他们的解释是 WSLS 是一种外显的、基于规则(rule-based)的策略,RL 是一种内隐的、信息集成(information-integration)的策略,认知负荷使得决策者更依赖内隐的学习过程,从而使 RL 模型能更好地捕捉其选择行为。作为相关证据,他们还发现被试对 WSLS 策略的主观偏好和 WSLS 模型的性能成正相关,但对 RL 策略的主观偏好和 RL 模型的性能无显著相关。这符合外显学习能进入意识,而内隐学习无法进入意识的规则。类似地,Worthy 等(2012)设计了两类选项结果分布动态变化,并且变化与决策者选择相关的经验决策问题。结果仍发现在无认知负荷条件下,WSLS

模型表现更好,而在认知负荷条件下,RL模型表现更好。这表明 WSLS 模型和 RL模型 在不同经验决策问题中的差异可能是普遍存在的。根据这一观点,TPT 中的决策者显然 是在无认知负荷情境下完成决策问题,因此外显学习系统占主导地位,进而使得启发式模型在模型比较中表现出优势。

然而,作为决策和学习领域的代表人物,Estes(1997, 2002)认为,人类认知的重要特点是存在多重并发过程(multiple concurrent processes)。也就是说,人们在决策时可能使用混合策略。实际上,在经验决策中已有一些研究使用 RL 和启发式的加权模型来捕捉行为特征,并取得了比单一模型更好的效果(Ahn et al., 2014; Worthy & Maddox, 2014; Worthy Pang & Byrne, 2013)。根据这一观点,启发式模型的优势显示了它在决策者的混合策略中权重较大,而这种较大的权重可能由多种原因导致。例如,在 TPT 中决策者要进行大量试次的选择,他们可能偏向于使用简单规则来较快完成任务,并保持选择的一致性。

此外,尽管启发式模型在模型比较上表现出明显优势,但这并不能说明对RL模型函 数选择的讨论重要性较低。首先,如果决策者在任务中使用了混合策略,那么单一模型实 际捕捉了对应策略的特征。当任务情境发生变化(例如,决策任务更改为 IGT),使得 RL 策略权重更高时,由合适函数组成的 RL 模型或能更好地描述选择行为。其次,相比于启 发式模型, RL 模型是一种更完备的模型。它易于迁移到经验决策任务的不同范式, 或更 广义的经验决策任务中(Ahn et al., 2008; Sutton & Barto, 2018; Yechiam & Busemeyer, 2005)。而且 RL 模型中的每个参数具有实际的理论意义,如效用函数中的参数  $\alpha$  可以表 征风险厌恶程度,参数λ可以表征损失厌恶程度。事实上,在计算精神病学领域,已有一 些研究比较了由不同人群经验决策数据拟合 RL 模型的估计参数差异, 并试图将这种差异 作为辅助诊断指标(Lane, Yechiam, & Busemeyer, 2006; Fridberg et al., 2010)。最后, 启发式 模型的某些假设可能太过绝对。第一, 启发式模型假定决策者仅依据试次间结果的相对大 小进行选择。那么想象这样两个经验决策问题,第一个问题是在选项(3,1)和选项(4,0.8) 间进行选择, 第二个问题是在选项(3, 1)和选项(400, 0.8)间进行选择。如果人们完全按照 WSLS 策略进行选择,那么在这两个问题中,至少在群体层面上选择序列应该是类似的。 但这一结论显然是反直觉的,因为在第二个问题中选项在结果绝对大小上的差异远大于第 一个问题, 决策者应当能更快地学习到选项间的差异并稳定地选择偏好选项。 第二, 启发 式模型假设当t个试次的结果等于第t-1个试次时,该试次仍是win 试次。那么想象某个 决策者连续 3 次选择了选项(-4, 0.8),得到的反馈分别是 0,-4,-4,则第 3 个试次被定为 win 试次, 而下一个试次决策者会倾向于 stay。这显然也是反直觉的, 因为决策者明确知

道该选项有一个更高的结果 0。所以, 启发式模型或许不能直接迁移到更复杂的决策问题中。

## 5.4 不足和展望

在技术方面,本研究于个体层面上使用了传统的极大似然方法进行参数估计。但 Ahn, Krawitz, Kim, Busemeyer 和 Brown(2013)在比较了不同估计方法对模拟数据的参数恢复能力后,发现考虑群体特征的分层贝叶斯方法表现更佳。他们还指出,传统方法得到的估计参数容易落在边界附近,而和真实值有较大偏离,这一现象在本研究中也有发现。由表2.1 和表2.2 可见,对于一些取值范围大于0的参数,其估计值的标准差甚至大于平均数,这反映了多数估计值落于取值范围下界,而少数估计值落于取值范围上界的分布。这种极端的参数估计情况表明在个体层面上无法得到足够信息,将群体特征纳入估计可以减少对个体信息量的要求,或能帮助得到更稳定和准确的估计参数。此外,分层贝叶斯方法本身能用于模型比较,即可以构建一层各模型以相等概率被选中的先验分布,在进行模型拟合后直接比较各模型被选中的后验概率以确定最优模型。该方法避免了传统参数方法对数据分布的要求,可以更方便地进行模型比较。同时,由于复杂模型涉及多个参数的概率分布,分层贝叶斯方法会给予它们惩罚,也就减少了参数量对模型拟合性能的影响(Kruschke, 2014)。所以,未来研究可以尝试使用分层贝叶斯方法进行参数估计和模型比较,以验证本研究结果是否有跨方法的稳定性。

在函数选择方面,ratio 选择函数是一个值得进一步探究的选择函数。虽然本研究已在损失域问题中对该函数进行了扩展,但因其不适合混合域问题所以未对其泛化性能进行测试,也就无法对它的价值作更全面的评判。未来研究可以尝试在混合域问题中完善这一函数,拓宽其使用情境,并进一步比较它和其他选择函数的相对优劣。本研究中另一个有待改进的函数是试次依赖选择敏感性。直观来说,在大量选择过程中,选择敏感性应当发生动态变化,但这种单调变化的函数形式可能过于绝对,并且在试次位置极小或极大时可能产生极端的选择敏感性数值。未来研究可以尝试使用更灵活的试次依赖选择敏感性,合理描述选择敏感性的变化过程。

在数据集方面,尽管 TPT 是当前最大的经验决策公开数据集,并且涉及了不同情境、不同概率的多种决策问题。但是决策问题中的结果变化范围较小(即从-30 到+30),同时问题结构比较单一(即风险选项和安全选项数学期望近似,同时安全选项只有一个确定结

果)。事实上,Glöckner,Hilbig,Henninger 和 Fiedler(2016)在两选项均含多个结果的决策问题中发现了反转的 DEG,这表明问题结构可能确实会影响经验决策的心理过程。此外,如 5.3 节所述,当选项之间的绝对大小差异较大时,启发式在直觉上可能并不具有优势。因此,未来研究可以尝试更多样的问题结构,以比较 RL 模型和启发式模型对复杂环境的适应能力。同时,尽管 TPT 中单个决策者完成多个问题的设计,为在个体层面上进行泛化测试提供了基础,但是完成过多的试次可能导致决策者的策略使用发生改变,也为捕捉选择敏感性的动态变化带来了困难。所以未来研究仍应控制合理的试次数量,避免过多试次带来的影响。

最后,在备选模型方面,本研究主要关注单一模型对于经验决策的解释能力。但如5.3 节所述,决策行为可能是多重并发过程的共同作用结果。因此,未来研究可以尝试构建 RL 策略和启发式策略的加权模型,并与单一模型比较,检验多过程模型是否能更好地捕捉经验决策行为特征。

## 6结论

本研究比较了 12 个 RL 模型和 3 个基于 WSLS 策略的启发式模型对部分反馈范式下经验决策数据的拟合和泛化性能,得到如下结论:

- (1)对于 RL 模型的函数选择,在效用函数方面,对数效用函数优于幂效用函数。在选择函数方面,softmax 选择函数优于 probit 选择函数,ratio 选择函数的适用性仍有待探究。在选择敏感性方面,尽管非试次依赖选择敏感性在多个评价指标上优于试次依赖选择敏感性,但选择敏感性的试次依赖性仍有待探究。
- (2)对于启发式模型, WSLS-2 模型始终优于 WSLS-1 和 SS 模型, 这支持了决策者在 win 试次和 loss 试次后的初始 stay 倾向存在差异,并且单个 win 试次和 loss 试次会对两组 条件概率均进行调整的假设。
- (3)在策略使用上,启发式模型在整体层面和最优模型层面均表现出明显优势,但关于启发式模型和 RL模型的相对优劣仍有待进一步探究。

# 参考文献

- Ahn, W., Busemeyer, J., Wagenmakers, E., & Stout, J. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science: A Multidisciplinary Journal*, 32(8), 1376-1402. doi:10.1080/03640210802352992
- Ahn, W., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2013). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Decision*, 1(S), 8-23. doi:10.1037/2325-9965.1.s.8
- Ahn, W., Vasilev, G., Lee, S., Busemeyer, J. R., Kruschke, J. K., Bechara, A., & Vassileva, J. (2014). Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology*, *5*, 849-863. doi:10.3389/fpsyg.2014.00849
- Audibert, J., Munos, R., & Szepesvári, C. (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19), 1876-1902. doi:10.1016/j.tcs.2009.01.016
- Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3), 215-233. doi:10.1002/bdm.443
- Beitz, K. M., Salthouse, T. A., & Davis, H. P. (2014). Performance on the Iowa gambling task: From 5 to 89 years of age. *Journal of Experimental Psychology: General*, *143*(4), 1677-1689. doi:10.1037/a0035823
- Bouchouicha, R., & Vieider, F. M. (2017). Accommodating stake effects under prospect theory. *Journal of Risk and Uncertainty*, 55(1), 1-28. doi:10.1007/s11166-017-9266-y
- Busemeyer, J. R., & Wang, Y. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, 44(1), 171-189. doi:10.1006/jmps.1999.1282
- Camilleri, A. R., & Newell, B. R. (2011). When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychonomic Bulletin & Review*, *18*(2), 377-384. doi:10.3758/s13423-010-0040-2

- Collins, A. G., & Frank, M. J. (2014). Opponent actor learning (Opal): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive.

  \*Psychological Review, 121(3), 337-366. doi:10.1037/a0037015
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876-879. doi:10.1038/nature04766
- Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*(4), 912-931. doi:10.1037/0033-295x.112.4.912
- Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, 21(5), 575-597. doi:10.1002/bdm.602
- Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1), 15-47. doi:10.1002/bdm.683
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*(2), 94-107. doi:10.1037/h0058559
- Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, *104*(1), 148-169. doi:10.1037/0033-295x.104.1.148
- Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, 9(1), 3-25. doi:10.3758/bf03196254
- Fridberg, D. J., Queller, S., Ahn, W., Kim, W., Bishara, A. J., Busemeyer, J. R., Stout, J. C. (2010). Cognitive mechanisms underlying risky decision-making in chronic cannabis users. *Journal of Mathematical Psychology*, *54*(1), 28-38. doi:10.1016/j.jmp.2009.10.002
- Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34-42. doi:10.1016/j.cognition.2017.12.014
- Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in

- risky choice. *Journal of Experimental Psychology: General*, *145*(4), 486-508. doi:10.1037/a0040103
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534-539. doi:10.1111/j.0956-7976.2004.00715.x
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision making under risk. *Econometrica*, 47(2), 263-291. doi:10.21236/ada045771
- Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, 246(1), 160-173. doi:10.1038/scientificamerican0182-160
- Konstantinidis, E., Taylor, R. T., & Newell, B. R. (2018). Magnitude and incentives: Revisiting the overweighting of extreme events in risky decisions from experience. *Psychonomic Bulletin & Review*, 25(5), 1925-1933. doi:10.3758/s13423-017-1383-8
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Boston, MA: Academic Press.
- Lane, S. D., Yechiam, E., & Busemeyer, J. R. (2006). Application of a computational decision model to examine acute drug effects on human risk taking. *Experimental and Clinical Psychopharmacology*, *14*(2), 254-264. doi:10.1037/1064-1297.14.2.254
- Lejarraga, T., & Hertwig, R. (2017). How the threat of losses makes people explore more than the promise of gains. *Psychonomic Bulletin & Review*, 24(3), 708-720. doi:10.3758/s13423-016-1158-7
- Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.
- Luce, R. D., & Suppes, P. P. (1965). Preference, utility, and subjective probability. In R. D.Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*(pp. 252-410). New York, NY: Wiley.
- Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, 60(2), 151-158. doi:10.1086/257177
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308-313. https://doi.org/10.1093/comjnl/7.4.308

- Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*(2), 274-279. doi:10.1016/j.cognition.2010.11.009
- Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science*, *37*(7), 770-786. doi:10.1287/mnsc.37.7.770
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York, NY: Appleton-Century-Crofts.
- Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review*, 117(3), 925-944. doi:10.1037/a0019619
- Scholten, M., & Read, D. (2014). Prospect theory and the "forgotten" fourfold pattern of risk preferences. *Journal of Risk and Uncertainty*, 48(1), 67-83. doi:10.1007/s11166-014-9183-2
- Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7-14. doi:10.1016/j.conb.2018.11.003
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464. doi:10.1214/aos/1176344136
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. doi:10.1093/biomet/52.3-4.591
- Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2), 101-130. doi:10.1007/s11166-006-8289-6
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: MIT Press.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273-286. doi:10.1037/h0070288
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458. doi:10.1126/science.7455683

- Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organizational Behavior and Human Decision Processes*, 97(1), 31-46. doi:10.1016/j.obhdp.2005.03.001
- Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the Annual Conference on Human Factors in Computing Systems 11*, 143-146. doi:10.1145/1978942.1978963
- Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review*, 20(2), 364-371. doi:10.3758/s13423-012-0324-9
- Worthy, D. A., Maddox, W. T., & Markman, A. B. (2008). Ratio and difference comparisons of expected reward in decision-making tasks. *Memory & Cognition*, *36*(8), 1460-1469. doi:10.3758/mc.36.8.1460
- Worthy, D. A., & Maddox, W. T. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. *Journal of Mathematical Psychology*, *59*, 41-49. doi:10.1016/j.jmp.2013.10.001
- Worthy, D. A., Otto, A. R., & Maddox, W. T. (2012). Working-memory load and temporal myopia in dynamic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1640-1658. doi:10.1037/a0028146
- Worthy, D. A., Pang, B., & Byrne, K. A. (2013). Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Frontiers in Psychology*, *4*, 640-648. doi:10.3389/fpsyg.2013.00640
- Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140-176. doi:10.1037/bul0000115
- Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, 12(3), 387-402. doi:10.3758/bf03193783

- Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*(2), 109-144. doi:10.1016/0022-2496(77)90026-8
- Yoon, S., Vo, K., & Venkatraman, V. (2017). Variability in decision strategies across description-based and experience-based decision making. *Journal of Behavioral Decision Making*, 30(4), 951-963. doi:10.1002/bdm.2009

## 附录1

现于 PT 框架下证明, 在使用不同的选择函数时, 对数效用函数相较幂效用函数都能 更好地解释量级效应。

PT 假设在描述决策问题中,决策者会计算每一选项的整体价值(overall value, V),并基于这一指标进行选择(整体价值类似于经验决策中的期望)。整体价值的计算方式为

$$V = \sum \pi(p_i)v(x_i)$$
 (Pf 1-1)

其中 $p_i$ 和 $x_i$ 分别指特定选项每一结果的发生概率和客观价值( $i=1,\ldots,n$ ), $\pi(p)$ 和v(x)分别是权重函数和价值函数(由于本研究不区分"效用"和"主观价值",因此以下称"价值函数"为"效用函数")。

由于量级效应仅在收益域比较稳定,因此以下证明仅限于收益域问题。低量级问题为在选项 A(m,p)和选项 B(n,q)之间选择,高量级问题为在选项 C(am,p)和选项 D(an,q)之间选择,其中 0 < m < n, $0 < q < p \le 1$ ,a > 1 且 pm = qn。量级效应指决策者在高量级问题中选择 C 的概率高于在低量级问题中选择 C 的概率,且选择概率增加的程度随变化比例 a 的增加而增大。

首先利用效用函数的弹性证明,在使用基于比值的选择函数时,幂效用函数不支持量级效应,对数效用函数支持量级效应。

幂效用函数的弹性及其导数为

$$\varepsilon_u(x) = \frac{xu'(x)}{u(x)} = \frac{\alpha x^{\alpha}}{x^{\alpha}} = \alpha$$
$$\varepsilon'_u(x) = 0$$

(附 1-2)

对数效用函数的弹性及其导数为

$$\varepsilon_{u}(x) = \frac{xu'(x)}{u(x)} = \frac{\frac{\alpha x}{1+\alpha x}}{\ln(1+\alpha x)} = \frac{\alpha x}{(1+\alpha x)\ln(1+\alpha x)}$$

$$\varepsilon'_{u}(x) = \frac{\alpha(1+\alpha x)\ln(1+\alpha x) - \alpha x\left(\alpha\ln(1+\alpha x) + \frac{\alpha(1+\alpha x)}{1+\alpha x}\right)}{[(1+\alpha x)\ln(1+\alpha x)]^{2}}$$

$$= \frac{\alpha[\ln(1+\alpha x) - \alpha x]}{[(1+\alpha x)\ln(1+\alpha x)]^{2}} < 0$$

(附 1-3)

因此幂函数弹性恒定,对数效用函数弹性随收益的增加而降低。

将高量级问题中安全选项和风险选项整体价值的比值视为变化比例 a 的函数,该函数及其导数形式为

$$f(a) = \frac{\pi(p)u(am)}{\pi(q)u(an)}$$

$$f'(a) = \frac{\pi(p)}{\pi(q)} \left( \frac{mu'(am)u(an) - nu'(an)u(am)}{u^2(an)} \right)$$

$$= \frac{\pi(p)u(am)u(an)}{a\pi(q)u^2(an)} (\varepsilon_u(am) - \varepsilon_u(an))$$

(附 1-4)

对于幂效用函数,因为弹性恒定,所以两选项整体价值的比值不随 a 改变而发生变化。 因此在使用基于比值的选择函数时,决策者在两个问题中对安全选项的选择概率始终恒 定。对于对数效用函数,因为弹性递减,所以两选项整体价值的比值随 a 增加而增大。因 此在使用基于比值的选择函数时,决策者在第二个问题中对安全选项的选择概率高于在第 一个问题中,并且增加的程度随 a 增加而增大。

其次证明在使用基于差值的选择函数时,幂效用函数在一定程度上支持量级效应,但 不允许偏好反转的发生,即不支持在低量级问题中偏好风险选项,而在高量级问题中偏好 安全选项。

将高量级问题中安全选项和风险选项整体价值的差值视为变化比例 a 的函数,当效用函数为幂函数时,该函数及其导数形式为

$$f(a) = \pi(p)u(am) - \pi(q)u(an) = \pi(p)(am)^{\alpha} - \pi(q)(an)^{\alpha}$$
$$= a^{\alpha} (\pi(p)u(m) - \pi(q)u(n))$$
$$f'(a) = a^{\alpha-1} (\pi(p)u(m) - \pi(q)u(n))$$

(附 1-5)

不难看出, a 的作用是放大了低量级两选项整体价值的差异, 并且放大程度随变化比例 a 的增大而增加。如果要用该函数组合来解释量级效应, 则需要低量级问题中安全选项的整体价值仍高于风险选项, 即决策者始终偏好安全选项。

最后证明不论是使用基于比值还是基于差值的选择函数,对数效用函数均在一定程度 上支持量级效应,并允许偏好反转的发生。

当使用基于比值的选择函数时,对数效用函数对量级效应的支持已证明,现证明这一

结论在使用基于差值的选择函数时,一定程度上也成立。

将高量级问题中安全选项和风险选项整体价值的差值视为变化比例 a 的函数,当效用函数为对数效用函数时,其形式为

$$f(a) = \pi(p)u(am) - \pi(q)u(an) = \frac{\pi(p)\ln(1 + \alpha am) - \pi(q)\ln(1 + \alpha an)}{\alpha}$$

$$f'(a) = \frac{m\pi(p)}{1 + \alpha am} - \frac{n\pi(q)}{1 + \alpha an} = \frac{m\pi(p) - n\pi(q) + \alpha amn\pi(p) - \alpha amn\pi(q)}{(1 + \alpha am)(1 + \alpha an)}$$
(\text{\text{\text{\$|\$}}} \ 1-6)

由于在PT中客观概率被映射为决策权重,因此f'(a)并不始终为正数。但不难看出,当a相对较大时,f'(a)的符号主要由 $\alpha amn\pi(p)$  -  $\alpha amn\pi(q)$ 决定,该项的值严格大于 0。而在量级效应相关研究中,变化比例a往往可取至数十到数万(Scholten, & Read, 2014; Weber & Chapman, 2005),在该条件下可近似认为f(a)为增函数,即随a增大,两选项的差异减小甚至发生反转。因此,在这种情境下对数效用函数可认为能够支持量级效应。

不论是基于差值还是基于比值的选择函数,均预测决策者对于两选项中整体价值较高的一项选择概率大于 0.5,即偏好该选项。因此,若对数效用函数支持偏好反转现象,应要求在较小量级时,风险选项的整体价值高于安全选项,而在较高量级时,安全选项的整体价值高于风险选项。已知在使用对数效用函数时,安全选项与风险选项整体价值的比值随变化比例 a(不要求 a>1)增加而增大。而当 a 相对较大时,其差值随 a 增加而减小甚至发生反转。因此只需证明在  $a\to 0$  时,风险选项整体价值高于安全选项,在  $a\to \infty$ 时,安全选项整体价值高于风险选项。

对于前者,可做如下推导:

$$\lim_{a \to 0} \frac{\pi(p)u(am)}{\pi(q)u(an)} = \lim_{a \to 0} \frac{\pi(p)\ln(1 + \alpha am)}{\pi(q)\ln(1 + \alpha an)} = \lim_{a \to 0} \frac{\pi(p)\frac{\alpha m}{1 + \alpha am}}{\pi(q)\frac{\alpha n}{1 + \alpha an}} = \frac{m\pi(p)}{n\pi(q)}$$
(\text{\text{\text{\$M\$}}}\tag{1-7})

由于 pm = qn, 且权重函数具有次可加性(subadditivity, 即  $\pi(rp) > r\pi(p)$ , 其中 0 < p, r < 1), 所以  $\pi(p)/\pi(q) < n/m$ , 即  $a \to 0$  时, 风险选项整体价值高于安全选项。

对于后者,可做如下推导

$$\lim_{a \to \infty} \frac{\pi(p)u(am)}{\pi(q)u(an)} = \lim_{a \to \infty} \frac{\pi(p)\ln(1+\alpha am)}{\pi(q)\ln(1+\alpha an)} = \lim_{a \to \infty} \frac{\pi(p)\frac{\alpha m}{1+\alpha am}}{\pi(q)\frac{\alpha n}{1+\alpha an}} = \frac{\pi(p)}{\pi(q)}$$
(Pf 1-8)

由于权重函数是概率的增函数,因此 $a \to \infty$ 时,安全选项整体价值高于风险选项。

综上,在使用基于差值的选择函数时,两种效用函数均不能严格地解释量级效应。但 考虑到量级较小时,决策者往往出现对风险选项的偏好,因此对数效用函数相对性能更好。 此外,本研究关注的效用函数和选择函数组合中,唯一能严格解释量级效应的是对数效用 函数 + ratio 选择函数, 唯一完全不能解释量级效应的是幂效用函数 + ratio 选择函数。

附录 2

附表 2.1 TPT 经验决策问题集

问题序号 -		风险选项		确定选项	风险选项		
	高结果	高结果概率	低结果	中结果	选择比例		
1*	-0.3	0.96	-2.1	-0.3	0.33		
2	-0.9	0.95	-4.2	-1.0	0.50		
3	-6.3	0.30	-15.2	-12.2	0.24		
4	-10.0	0.20	-29.2	-25.6	0.32		
5	-1.7	0.90	-3.9	-1.9	0.45		
6	-6.3	0.99	-15.7	-6.4	0.68		
7	-5.6	0.70	-20.2	-11.7	0.37		
8	-0.7	0.10	-6.5	-6.0	0.27		
9	-5.7	0.95	-16.3	-6.1	0.43		
10	-1.5	0.92	-6.4	-1.8	0.44		
11	-1.2	0.02	-12.3	-12.1	0.26		
12	-5.4	0.94	-16.8	-6.4	0.55		
13	-2.0	0.05	-10.4	-9.4	0.11		
14	-8.8	0.60	-19.5	-15.5	0.66		
15	-8.9	0.08	-26.3	-25.4	0.19		
16	-7.1	0.07	-19.6	-18.7	0.34		
17	-9.7	0.10	-24.7	-23.8	0.37		
18	-4.0	0.20	-9.3	-8.1	0.34		
19	-6.5	0.90	-17.5	-8.4	0.49		
20	-4.3	0.60	-16.1	-4.5	0.08		
21	2.0	0.10	-5.7	-4.6	0.11		
22	9.6	0.91	-6.4	8.7	0.41		
23	7.3	0.80	-3.6	5.6	0.39		
24	9.2	0.05	-9.5	-7.5	0.08		
25	7.4	0.02	-6.6	-6.4	0.19		
26	6.4	0.05	-5.3	-4.9	0.20		
27	1.6	0.93	-8.3	1.2	0.50		
28	5.9	0.80	-0.8	4.6	0.58		
29	7.9	0.92	-2.3	7.0	0.51		
30	3.0	0.91	-7.7	1.4	0.41		
31	6.7	0.95	-1.8	6.4	0.52		
32	6.7	0.93	-5.0	5.6	0.49		
33	7.3	0.96	-8.5	6.8	0.65		
34	1.3	0.05	-4.3	-4.1	0.30		
35	3.0	0.93	-7.2	2.2	0.44		
36	5.0	0.08	-9.1	-7.9	0.09		
37	2.1	0.80	-8.4	1.3	0.28		
38	6.7	0.07	-6.2	-5.1	0.29		

问题序号 -	风险选项			确定选项	风险选项
	高结果	高结果概率	低结果	中结果	选择比例
39	7.4	0.30	-8.2	-6.9	0.58
40	6.0	0.98	-1.3	5.9	0.61
41	18.8	0.80	7.6	15.5	0.52
42	17.9	0.92	7.2	17.1	0.48
43*	22.9	0.06	9.6	9.2	0.88
44	10.0	0.96	1.7	9.9	0.56
45	2.8	0.80	1.0	2.2	0.48
46	17.1	0.10	6.9	8.0	0.32
47	24.3	0.04	9.7	10.6	0.25
48	18.2	0.98	6.9	18.1	0.59
49	13.4	0.50	3.8	9.9	0.13
50	5.8	0.04	2.7	2.8	0.35
51	13.1	0.94	3.8	12.8	0.52
52	3.5	0.09	0.1	0.5	0.26
53	25.7	0.10	8.1	11.5	0.11
54	16.5	0.01	6.9	7.0	0.18
55	11.4	0.97	1.9	11.0	0.66
56	26.5	0.94	8.3	25.2	0.53
57	11.5	0.60	3.7	7.9	0.45
58	20.8	0.99	8.9	20.7	0.63
59	10.1	0.30	4.2	6.0	0.32
60	8.0	0.92	0.8	7.7	0.44
61	-8.7	0.06	-22.8	-21.4	0.25
62	-2.2	0.09	-9.6	-8.7	0.27
63	-2.0	0.10	-11.2	-9.5	0.25
64	-1.4	0.02	-9.1	-9.0	0.33
65	-0.9	0.07	-4.8	-4.7	0.37
66	-4.7	0.91	-18.1	-6.8	0.63
67	-9.7	0.06	-24.8	-24.2	0.30
68	-5.7	0.96	-20.6	-6.4	0.66
69	-5.6	0.10	-19.4	-18.1	0.31
70	-2.5	0.60	-5.5	-3.6	0.34
71	-5.8	0.97	-16.4	-6.6	0.61
72	-7.2	0.05	-16.1	-15.6	0.25
73	-1.8	0.93	-6.7	-2.0	0.44
74	-6.4	0.20	-22.4	-18.0	0.21
75*	-3.3	0.97	-10.5	-3.2	0.16
76	-9.5	0.10	-24.5	-23.5	0.39
77 77	-2.2	0.92	-11.5	-3.4	0.47
78	-1.4	0.93	-4.7	-1.7	0.41
79	-8.6	0.10	-26.5	-26.3	0.49
80	-6.9	0.06	-20.5	-20.3	0.45
81	1.8	0.60	-20.5 -4.1	1.7	0.23

		风险选项		确定选项	风险选项
问题序号 -	高结果	高结果概率	低结果	中结果	选择比例
82*	9.0	0.97	-6.7	9.1	0.14
83	5.5	0.06	-3.4	-2.6	0.28
84	1.0	0.93	-7.1	0.6	0.46
85	3.0	0.20	-1.3	-0.1	0.21
86	8.9	0.10	-1.4	-0.9	0.23
87	9.4	0.95	-6.3	8.5	0.67
88	3.3	0.91	-3.5	2.7	0.58
89	5.0	0.40	-6.9	-3.8	0.39
90	2.1	0.06	-9.4	-8.4	0.33
91*	0.9	0.20	-5.0	-5.3	0.88
92	9.9	0.05	-8.7	-7.6	0.21
93	7.7	0.02	-3.1	-3.0	0.28
94	2.5	0.96	-2.0	2.3	0.52
95	9.2	0.91	-0.7	8.2	0.56
96*	2.9	0.98	-9.4	2.9	0.34
97	2.9	0.05	-6.5	-5.7	0.30
98	7.8	0.99	-9.3	7.6	0.62
99	6.5	0.80	-4.8	6.2	0.32
100	5.0	0.90	-3.8	4.1	0.46
101	20.1	0.95	6.5	19.6	0.50
102	5.2	0.50	1.4	5.1	0.08
103	12.0	0.50	2.4	9.0	0.17
104	20.7	0.90	9.1	19.8	0.44
105	8.4	0.07	1.2	1.6	0.20
106	22.6	0.40	7.2	12.4	0.41
107	23.4	0.93	7.6	22.1	0.72
108	17.2	0.09	5.0	5.9	0.24
109	18.9	0.90	6.7	17.7	0.57
110	12.8	0.04	4.7	4.9	0.26
111	19.1	0.03	4.8	5.2	0.22
112	12.3	0.91	1.3	12.1	0.41
113	6.8	0.90	3.0	6.7	0.41
114	22.6	0.30	9.2	11.0	0.60
115	6.4	0.09	0.5	1.5	0.28
116	15.3	0.06	5.9	7.1	0.17
117	5.3	0.90	1.5	4.7	0.66
118	21.9	0.50	8.1	12.6	0.47
119	27.5	0.70	9.2	21.9	0.42
120	4.4	0.20	0.7	1.1	0.38

注:理论上确定选项结果应等于风险选项数学期望,因此记为"中结果"。但由于问题生成算法引入了随机项(见附录 3),使得在某些问题中确定选项结果大于等于风险选项高结果,或小于等于风险选项低结果,也就是存在占优选项,这些问题的序号已用星号标出。风险选项选择比例是取该问题中所有被

试所有试次选择的平均值。

## 附录3

- 120 道经验决策问题根据如下算法生成:
- 1.对于风险选项的高结果的概率 p, 首先等可能地从(0.01, 0.09), (0.10, 0.90), (0.91, 0.99) 中抽取一个区间,然后在该区间设置以 0.01 为间隔的网格,随机抽取一个格点作为 p。
- 2.从(-10, 0)区间的均匀分布中抽取数字  $X_{min}$ ,从(0, 10)区间的均匀分布中抽取数字  $X_{max}$ 。将这两个数字保留一位小数后分别记为 L'和 H',将 pH'+(1 p)L'保留一位小数后记为 m。然后将 m + e 记为 M'。其中 e 服从正态分布,该分布的均值为 0,标准差为|(m L)/2|、|(m H)/2|和 2 中的最小值。
- 3.将问题序号记为 o。如果  $1 \le o \le 20$  或  $61 \le o \le 80$ ,则 con = -H' + L';如果  $21 \le o \le 40$  或  $81 \le o \le 100$ ,则 con = 0;如果  $41 \le o \le 60$  或  $101 \le o \le 120$ ,则 con = H' L'。
- 4.将 M' + con 作为确定选项结果,将 L' + con 作为风险选项低结果,将 H' + con 作为风险选项高结果。

# 附录4

以幂效用 + softmax + 非试次依赖的 RL 模型为例,对于给定被试和决策问题,选择序列的单次模拟根据如下算法生成:

- 1.依据 2.2.2 节所述方法估计该模型的参数值。对每一选项 j(j=1,2),设置初始期望  $E_i(0)=0$ 。根据公式 1-5,该模拟被试在第一个试次以 0.5 的概率选择任一选项。
- 2.在第 t 个试次,从(0,1)区间的均匀分布中抽取一个数字,如果该数字小于等于在第 t 个试次模型预测的选择风险选项的概率,则该试次模拟被试选择风险选项,否则选择确 定选项。
- 3.如果模拟被试在第 t 个试次选择确定选项,则该试次结果即为确定选项结果。如果模拟被试在第 t 个试次选择风险选项,则再从(0,1)区间的均匀分布中抽取一个数字,如果该数字小于等于风险选项高结果概率,则该试次结果为风险选项高结果,否则为风险选项低结果。
- 4.根据公式 1-1 将第 t 个试次的结果转换为效用,然后根据公式 1-3 更新第 t 个试次两选项期望。
  - 5.根据公式 1-5 计算第 t+1 个试次两选项的选择概率。
  - 6.对 t = 1-100,依次重复步骤 2-5。

对于其他 RL 模型,模拟算法类似。但注意对于 ratio 选择函数相关模型, $E_j(0)$  在收益域问题中设为 0.0001,在损失域问题中设为-0.0001。对于启发式模型,被试前两个试次以 0.5 的概率选择任一选项,之后试次的选择依据公式 1-9 至 1-15 的规则进行模拟。