**对不同文化下情境因素对道德困境中的道德判断的影响研究复刻**

周天 王妍 杨捷茜

（心理学院）

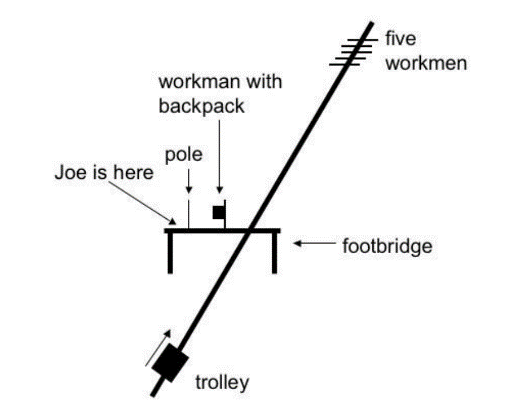
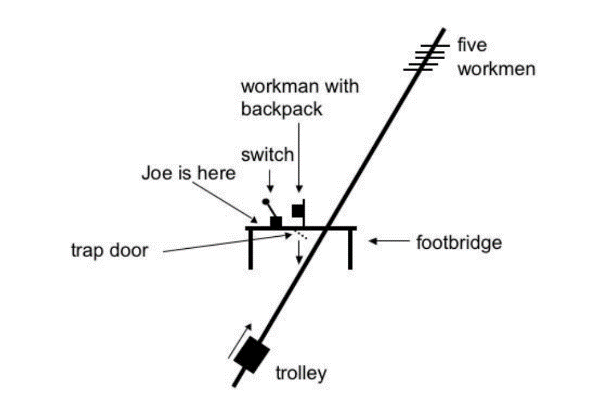
**摘要** 选取来自Nature human behaviour的Bago等人的研究，使用贝叶斯方法对原文献中研究一、研究二进行复刻，并探讨了除了在不同文化下，个人武力的使用对道德可接受程度的影响是否在不同的国家间存在差异。道德困境可以被描述为两个主要相互冲突的道德原则之间的抉择，已有研究表明不同文化背景对道德困境中的道德判断影响不同。本研究借助贝叶斯分析方法，使用python语言定义贝叶斯模型，最终验证原文献结论：（一）个人力量的使用对道德判断中道德可接受程度有显著影响，且该影响具有文化普遍性；（二）在西方文化集群中，个人力量和意图对道德判断存在交互作用。另外，在研究三中使用多层线性模型，发现道德可接受程度在不同国家间存在差异，但是个人力量的使用对道德可接受程度的影响在不同国家间不存在差异。

**关键词** 贝叶斯分析 模型定义 道德困境 多文化样本

# **1研究背景**

## 问题提出

道德困境可以被描述为两个主要相互冲突的道德原则之间的抉择，这个问题长期以来一直引起哲学家和心理学家的兴趣。在原文献中有电车和游艇两种道德困境情境，本研究选取了电车难题进行讨论，在这种情境下，举两个例子来说明困境主要内容：第一种情况是有一辆失控的手推车在铁轨上滚滚而下，前方在铁轨上有五名工人。手推车直奔他们而来，如果不采取任何措施，他们就会被杀死。Joe和一个工人站在铁轨上的天桥上，Joe的手边有一个杠杆，如果他拉动这个杠杆，那一个工人就会掉在铁轨上，这样那个工人就会被杀，但原本的五个工人就会得救（如图1）。第二种情况是将Joe手边的杠杆换成了一根杆子，Joe需要拿起这根杆子将那一个工人推下天桥才能阻止五个工人被车撞（如图2）。在这两种情境下Joe拉动杠杆或用杆子推工人的道德可接受程度即为人们的道德判断。

图1 图2

Greene等人提出了和个人力量（即行为者是否必须使用个人努力杀死受害者并拯救更多人）和伤害意图（即伤害作为手段或副作用）在道德可接受性评级上的相互作用的证据，例如在图1情况下为不需要使用个人力量，而在图2情况下是需要使用个人力量。更具体地说，当人们不得不使用个人武力杀死这个人并且需要这个人的死亡来拯救更多的人时，他们不太可能判断牺牲一个人来拯救更多的人在道德上是可以接受的（这就是意图伤害的意思）。但这一结论的大部分证据来自西方、受过教育、工业化、富裕、民主（WEIRD）的社会样本，因此这些影响具有普遍性还是独特性还需要讨论。

## 研究假设

根据现在存在的问题，原文献假设为：

（1）个人力量对道德判断的影响具有文化上的普遍性

1a：在西方集群中，个人力量对道德判断有影响；

1b：如果个人力量的影响在文化上是普遍的，那么个人力量也会对南部和东部文化集群的道德可接受评级产生影响。

（2）个人力量和意图对道德判断的交互作用具有文化上的普遍性

2a：在西方集群中，个人力量和意图之间存在相互作用。更具体地说，与没有个人力量时相比，存在个人力量时的意图因素更大；

2b：如果这种影响在文化上是普遍的，那么在南部和东部文化集群中也存在影响。

另外，在原文献文化普适性的研究基础上，本小组关注除了文化普适性，个人力量的使用对道德可接受程度的影响是否在不同的国家间存在差异，因此存在以下几种可能：

（3）国家站点的差异

3a：道德可接受程度在不同国家间存在差异，但是个人力量对道德可接受程度的影响在不同国家间不存在差异；  
 3b：国家只调节个人力量对道德可接受程度的影响，而各国家间道德可接受程度相当；  
 3c：国家可能既影响道德可接受程度，又影响个人力量对道德可接受程度的影响。

# **2研究方法**

**2.1实验数据与实验材料**

本研究数据来源于原文献公开数据，共收集了来自27个国家的50245名参与者的数据，并根据一定标准排除了一定样本。

本研究中选取电车难题进行复刻。电车难题共分为六种情境，图1、图2两种情境分别代表有、无个人力量的使用，不讨论伤害意图，剩下四种情景分别为有个人力量和伤害意图、无个人力量和伤害意图、有个人力量无伤害意图以及无个人力量有伤害意图四种情况。研究一主要针对图1、图2所代表情境，研究二主要使用后四种情境。参与者需要对每种情况下Joe牺牲了一个工人拯救了五个工人的情况的道德可接受程度进行1到9分评分，1代表接受程度最低，9代表接受程度最高。

**2.2分析方法**

研究一中自变量为个人力量，因变量为道德可接受度评级，数据符合正态分布，因此选择Normal-Normal贝叶斯模型进行模型一的定义，并进行MCMC评估以及后验回归线预测。对东部、南部文化背景下做同样的分析。

研究二中自变量为个人力量和伤害意图，因变量为道德可接受度评级。仍然选择Normal-Normal贝叶斯模型进行模型的定义。首先定义模型二不存在交互作用，并定义模型三存在交互作用，通过对两种模型的比较，可以得出是否存在个人力量与伤害意图的交互作用。对东部、南部文化背景下做同样的分析。

研究三中自变量与因变量与研究一相同，为对比不同国家站点之间的影响，使用多层线性模型，分别定义模型0为普通线性模型，即仅考虑个人力量对道德可接受程度的影响；模型1为变化截距模型，在模型0的基础上考虑道德可接受程度在不同国家的变化；模型2为变化斜率模型，在模型0的基础上不同国家间的个人力量影响的变化；模型3为变化截距和斜率模型，结合模型1和模型2，同时考虑国家对道德可接受程度以及个人力量的变化的影响。最后通过四种模型的比较，验证假设。

# **3结果分析**

**3.1描述性统计**

在研究一中对不同文化背景下个人力量的使用对道德评级的影响数据进行可视化，可以看出在使用个人力量时，1到4分评级人数多于不使用个人力量，而在5到9分评级情况正好相反。

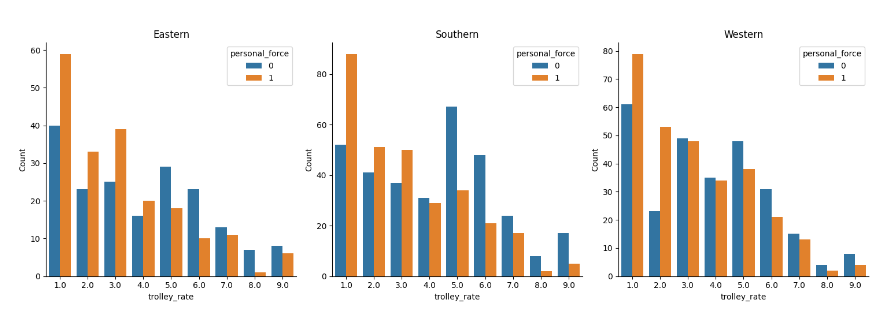


图3 研究一数据可视化

**3.2差异性分析**

研究一中通过对后验参数的分析，β0 = 3.758， 表明x1为0即不需要使用个人力量时，在西方文化背景下的被试认为这件事情从道德上可以接受的分数为3.758，β1 = -0.594，表明当在西方文化背景下的被试需要使用个人力量时，被试认为这件事情从道德上可以接受的分数将降低0.594，为3.164，并且β1的94%HDI不包括0，说明个人武力能有效预测道德接受度变化的概率。同样的在东部文化中β0 = 3.929，β1 = -0.806，β1的94%HDI不包括0，在南部文化中β0 = 4.218，β1 = -1.063，β1的94%HDI不包括0，说明这一结论在文化上具有普适性。

研究二中，对模型二后验参数的分析可得个人力量和伤害意图都能有效预测道德接受程度的变化，β1 、β2的94%HDI都不包括0，模型三中对后验参数的分析同模型二，同时代表个人力量与伤害意图交互作用的β394%HDI都不包括0，再对两个模型进行比较发现模型三好于模型二，因此在西方文化下个人力量与伤害意图存在交互作用。同样在南部文化和东部文化背景下做同样的分析，发现在南部文化下可以得出相同的结果，但在东部文化下，模型三中β394%HDI都包括0，且模型比较发现模型二优于模型三，因此在东部文化中不具有普适性。

研究三中，通过对不同国家站点的数据筛查，建立了各个模型，通过模型之间的比较，发现变化截距即道德可接受程度在不同国家之间的变化模型排名最高，也就是说道德可接受程度在不同国家间存在差异。但是个人武力对道德可接受程度的影响在不同国家间不存在差异。但是由于变化截距和变化斜率、截距模型的elpd非常接近，模型比较的结果只能作为参考。

# **4结论**

（1）个人力量的使用对道德判断中道德可接受度的有显著影响，并且具有文化普适性。（与原文献结论一致）

1. 对于研究二交互作用的分析，我们定义了两个模型，模型2不考虑交互作用，模型3考虑交互作用。模型比较的结果告诉我们：在西方、南方文化集群中，个人力量和意图对道德判断存在交互作用，但无法在东方文化集群中得到同样的结果。（与原文献一致）
2. 道德可接受程度在不同国家间存在差异，但是个人力量的使用对道德可接受程度的影响在不同国家间不存在差异。

**参考文献**

Bago, B., Kovacs, M., Protzko, J. et al. (2022). Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. *Nat Hum Behav 6*, 880–895.