利用贝叶斯方法对 Rathje 等人研究的复刻检验

周方茹 郑玉洁 李赵辰泽 舒绮 (按随机顺序排列)

(南京师范大学心理学院, 南京 210097)

摘 要 为了深入理解个体在判断政治新闻真伪时的心理机制，并为减少虚假信息的传播提供指导，本研究通过实验探讨了准确性激励和信息来源对虚假信息信念的影响。通过提供金钱激励奖励正确识别标题的被试，并移除党派来源线索，研究者检验了这些因素对政治新闻判断准确性的影响。在数据分析中，我们采用了传统贝叶斯模型和基于信号检测论的贝叶斯模型。结果显示，准确性激励显著提高了被试对政治新闻的准确判断，支持了假设 1。然而，我们并未观察到信息来源对被试判断准确性的显著影响，同时准确性激励和信息来源之间的交互作用也不显著。综上所述，本研究为深入理解虚假信息信念的形成提供了实证依据，并为减少虚假信息传播提供了理论支持和实践指导。

关键词 虚假信息；新闻来源；信号检测论；贝叶斯模型

1 引言

1.1 研究背景

虚假信息指的是那些捏造的新闻报道、虚假的谣言、阴谋论或错误信息等(Lewandowsky et al., 2021)。这些信息的广泛传播对社会产生严重的负面影响，例如导致“疫苗犹豫”等现象。因此，近年来研究者们对理解虚假信息信念的心理机制以及如何减少其传播产生了浓厚的兴趣。过去的研究表明，个体在判断信息真伪时，存在显著的党派差异。人们更容易相信与自己政治立场一致的新闻或来源于相同党派的消息源。然而，这种党派分歧的存在有多种可能的原因。一种解释认为，人们的判断经常由追求准确性驱动（即准确性动机），但同时他们也受到社会动机的引导，如群体归属和地位等。这些社会动机可能会干扰准确性动机(Taber & Lodge, 2006)。另一种解释认为，由于接触不同党派新闻渠道和社交媒体，党派成员具有不同的先前知识或信念(Pennycook & Rand, 2021)。

然而，除非通过实验操纵准确性动机或社会动机，否则很难区分这两种解释。如果虚假信息信念在某种程度上反映了动机因素，那么实验操纵人们的准确性动机或社会动机应该会改变他们对虚假信息的判断。相反，如果虚假信息信念仅仅反映了个体不同的先前信念，那么这些实验操纵应该不会改变他们对虚假信息的判断。

为了进一步探讨这一问题，Rathje 等人(2023)进行了一系列实验。通过提供正确识别标题的金钱激励，研究者检验了准确性动机在政治新闻真假判断中的作用。同时，研究者还检验了当从帖子中移除党派来源线索时，金钱激励的作用是否会减弱。通过这些实验，研究者希望能够更深入地理解个体在判断政治新闻真伪时的心理机制，并为减少虚假信息的传播提供理论支持和实践指导。

1.2 研究假设

为了重复检验 Rathje 等人的研究结论，同时解决他们在讨论部分提出的由“样本量不足”导致无法检验准确激励和党派信息来源交互效应的问题，我们采用贝叶斯的方法对 Rathje 等人的研究数据进行了分析，并提出以下假设：

假设 1：准确性激励会影响被试对政治新闻判断的准确性，存在准确性激励时判断更准确。

假设 2：党派信息来源会影响被试对政治新闻判断的准确性，有党派信息来源时的判断更不准确。

假设 3：准确性激励和党派信息来源对被试判断准确性的影响存在交互作用。

2 研究方法

2.1 数据来源与被试

本研究数据来源于 Rathje 等人研究三的研究结果。该研究每个子研究的研究方法与样本量均进行了预注册(注册时间：https://osf.io/75sqf.; 注册时间：2021 年 10 月 13 日)。实验过程、实验材料、数据、分析脚本均可在 https://osf.io/75sqf.中得到。

本研究共选用了 921 名被试(其中 435 名为男性，12 名为跨性别/非二元性别/其他)的数据，被试平均年龄为 40 岁(SD = 14.67)，542 名支持自由党、379 名支持保守党。

准确性激励的数据来自于被试每次选择时是否存在金钱激励(“No Control” = 存在金钱激励，“Control” = 不存在金钱激励)，党派信息来源的数据来自被试每次选择时是否显示有党派信息的信息来源(“Source” = 显示党派信息来源，“No Source” = 不显示党派信息来源)，被试判断准确性的数据涉及两个变量：被试每次对新闻真假的判断(“T” = 真，“F” = 假)，以及新闻本身的真假(“T” = 真，“F” = 假)。

2.2 数据分析

本研究采用两种方法进行数据分析：线性模型的方法和信号检测论的方法。

2.2.1 线性模型的方法

为了分析每个自变量如何影响被试判断的准确性，我们共定义了 4 个模型，模型 1 只包括准确性激励的作用；模型 2 只包括党派信息来源的作用；模型 3 包括党派信息来源和准确性激励的作用，但不包括两者的交互作用；模型 4 包括准确性激励和党派信息来源的作用，同时包括两者的交互作用。

在建立模型后，我们用 logit 连接函数把线性模型映射到 0-1 之间以符合二分变量的分布，对各自变量效应量(即各自变量对应的 β 值)的先验分布进行了设置，并利用 MCMC 采样得到了模型效应量的后验分布。参考后验分布 94%的 HDI 区间，对各自变量的主效应、交互效应是否存在进行了检验。并结合后验预测误差(MAE)、后验预测区间和留一法交叉验证对 4 个模型的预测效果进行了比较，以探究不同的自变量以怎样的方式组合能实现对被试判断准确性的最佳预测效果。

不同模型的详细数学描述、定义和先验设置可以在 Notebook 中找到。所有模型均在编程语言 Python 中实现，代码可以也在 Notebook 中找到。

2.2.2 信号检测论的方法

在线性模型的方法中，各系数的意义就仅代表效应的大小，无法转换成更加实际的意义。由于被试进行判断的过程可以类比为一个信号检测的过程(“新闻为真、判断为真”为击中，以此类推)，而已有研究提供了一种可以运用贝叶斯原理搭建模型、直接检验信号检测论指标的方法(详见：https://mvuorre.github.io/posts/2017-10-09-bayesian-estimation-of-signal-detection-theory-models/)，我们也利用这种方法对研究结果进行了补充。

我们将连接函数改为 probit 函数，通过这个函数的转换，模型的截距和斜率可以分别对应为信号检测论中的 C 值和 d’ 值。在此基础上，我们分别提取自变量在不同水平的数据(如存在准确性激励、不存在准确性激励)，分别搭建模型，这样便可以比较自变量在不同水平情况下被试 C 值和 d’值上的变化。

2.3 数据处理

本研究涉及两个主要的数据处理：

首先，原研究建立了一个算术式作为衡量被试判断准确性的指标，即因变量。但是原本的因变量有正有负，在采取了连接函数后还是不能做到很好的拟合。因此在对原假设进行检验的过程中，我们将因变量改为被试判断结果为真或假，设置题目本身真或假为自变量。这样设置后，在正确判断的情况下，题目本身的真假会对因变量存在主效应(即题目为真，被试更倾向于判断为真)；而题目本身真假与其他自变量的交互作用可以展现其他自变量对被试判断准确性影响的指标（如果存在正向的交互作用，说明该自变量能提升被试判断的准确性。

其次，我们对原研究的所有数据以“-0.5，+0.5”的方式进行了编码。在存在交互项的情况下，编码方式不同，当自变量取特定值时，其与 β 值组合成的表达式也会发生变化，而各自变量的主效应、交互效应由这些表达式计算而得。因此，自变量的编码方式会影响 β 值所代表的含义，进而影响其是否能代表不同自变量的主效应。而相对“0，1”编码，采用“-0.5，+0.5”编码能使各 β 值代表相应自变量的主效应。

3 研究结果

3.1 线性模型方法研究结果

3.1.1 后验解释

通过 MCMC 的采样，得到模型 1 中 β_0 = -0.464，β_1 = 0.142，β_2 = 1.274，β_4 = 0.308，所有 β 值在 94%的 HDI 中都不包括 0，准确性激励的作用显著；模型 2 中 β_0 = -0.470，β_3 = 1.271，β_2 = -0.119，β_5 = 0.243，β_2 和 β_5 在 94%的 HDI 中包括 0，党派信息来源的作用不显著；模型 3 中 β_0 = -0.466，β_1 = 0.146，β_2 = 1.280，β_3 = -0.123，β_4 = 0.300，β_5 = 0.234，β_3 和 β_5 在 94%的 HDI 中包括 0，准确性激励的作用显著，党派信息来源的作用不显著；模型 4 中 β_0 = -0.467，β_1 = 0.144，β_2 = 1.280，β_3 = -0.123，β_4 = 0.304，β_5 = 0.235，β_6 = 0.082，β_3, β_5 和 β_6 在 94%的 HDI 中包括 0，准确性激励的作用显著，党派信息来源的作用不显著，准确性激励和党派信息来源的交互作用不显著。

对后验参数进行解释，β_0 = -0.467，$e^{\beta 0}$ = 0.63，表明所有 X 值不存在时，个体将新闻判断为真与判断为假的概率比的可能性为 0.63。β_1 = 0.14，$e^{\beta 1}$ = 1.15，相比控制情况下，有准确性准确性激励的情况下，该概率上升为原来的 1.15 倍。β_2 = 1.30，$e^{\beta 2}$= 3.67，相比原题目时假的情况下，题目是真的情况下，该概率上升为原来的 3.67 倍。β_4 = 0.30，表明被试对新闻的真假进行判断是，不仅受到新闻本身真假的影响，还受到是否存在准确性激励条件的影响：$e^{\beta 4x1=-0.5}$ = 0.86，在控制条件下，当新闻为真时，该概率比缩小为原来的 0.86 倍，$e^{\beta 4x1=0.5}$ = 0.86，在准确性激励条件下，当新闻为真时，该概率比扩大为原来的 1.16 倍，验证了本研究的假设 1。然而，其他的 β 值在 94%的 HDI 中皆包含 0，说明不存在其他显著的主效应和交互效应。

3.1.2 模型比较

从后验预测的结果看，四个模型都能较为准确地描绘出观测数据的形态。模型 1，2，3，4 的 MAE 的结果分别是 0.428，0.458，0.427，0.428，模型 3 的预测误差最小，即模型 3 的预测能力最优，其次分别是模型 1 和模型 4，模型 2 的预测能力最差，这与后验参数的解释保持了一致：信息党派信息来源的主效应不显著，准确性准确性激励和信息源的交互作用不显著。通过比较 4 个模型的 elpd，得出模型 3 的 elpd_loo 最大，表明它对样本数据的预测性能最好，而模型 2 的 elpd_loo 最小，表明它的预测性能最差。这些结果与我们通过 MAE 和后验预测区间得到的判断一致。需要注意的是，arviz 提供的结果包括了 elpd se，这使得我们可以判断两个模型的预测差异 elpd_diff 是否超过两至三个标准误 se。从现在的结果看，四个模型间的差异均小于两个标准误 se 很小，表明四个模型的预测性能差异并不显著。

综合对后验参数的理解比较，根据奥斯卡姆剃刀原则，我们选择模型 1 作为最后适用模型。

3.2 信号检测论方法研究结果

同时，我们又通过信号检测论的两个指标来观察是否有准确性激励对被试的影响，得到在准确性准确性激励条件下：辨别力指数 d’ = 0.69，反应偏向 C = -0.33；在控制条件下：辨别力指数 d’= 0.89，反应偏向 C = -0.24。可以看到，和没有准确性激励相比，被试辨别力指数提高，做出“否”反应的倾向减弱了，这说明被试会更加仔细考虑新闻是否是真新闻。

之后，我们加入了用于预测的题目，得到准确性: 0.51，敏感性: 0.64，特异性: 0.67，有较好的预测效果。

4 结论与讨论

4.1 研究结论

总的来说，我们采用两种思路来对文章中研究三的内容进行了复现。

第一种方法采用常规的广义线性模型的思路，用 logit 连接函数把线性模型映射到 0-1 之间以符合二分变量的分布，再运用贝叶斯原理，设定先验后采用 MCMC 采样，最后得到各个系数的后验分布。在绘制四个模型以后，考虑预测能力的基础上，根据奥斯卡姆剃刀原则选择最为精简的模型 1。

第二种方法采用信号检测论的思路，在方法一的基础上更改连接函数为 probit 函数，由此一来模型的截距和斜率便分别对应为信号检测论中的 C 值和 d’ 值。再把数据根据有无激励分成两个部分分别来做模型，这样便可以比较有无激励时被试 C 值和 d’值上的变化。同时采用传统的信号检测论的方法对数据进行了分析，得到了和贝叶斯方法几乎一样的结果。

我们的复现结果证明了假设 1。如分析结果部分所言，两种方法的结果都趋于一致，即当被试在判断真假新闻时，有激励的被试辨别力会增加，能够更好的做出判断。并且，在移除信息来源的线索后，模型的预测能力并没有减弱(对比模型 1，3)，这进一步证明了就金钱激励对被试动机的影响。对于这一点，文章的结果与我们的相同，得到了复现。

4.2 讨论

通过对结果与模型进行分析比较，我们认为两种方法相比，各有优劣之处。方法一可以建立多个模型相互之间进行比较，也更贴近传统的广义线性模型，结果上更容易理解，但是各系数的意义就仅代表效应的大小，无法转换成更加实际的意义。但方法二则弥补了这以缺点，在经过中间函数的转换之后，截距代表了被试的反应倾向，而斜率则代表着被试的辨别力指数，通过对系数的探索，我们便能够更加直观的了解在实验处理的不同水平下，被试的认知特征是如何变化的。如在金钱激励的条件下，被试的 d’值增加，这说明了有金钱激励的结果时，被试的辨别力提高。而贝叶斯主义通过定义先验和似然以后进行抽样，直接得到系数的后验分布，这使我们能够更加直观的了解到系数的真实分布，与传统的频率主义通过反证法来证明存在差异相比，更加直观也更加可信，一定程度上也更加能够帮助我们理解文章中的统计检验力不足的问题。

从结果来看，金钱激励对被试的影响说明了准确性动机在政治新闻真假的判断中起到关键的作用。实验操纵人们的准确性动机能够改变他们对虚假信息的判断，虚假信息信念在某种程度上反映了动机因素。但是我们并未观察到信息来源的主效应及相关交互作用显著的结果，而文章中发现信息来源的主效应显著，在这一方面我们没有和作者得出相同的结论。没有充分的证据支持假设 2 和假设 3 的成立。

总体而言，贝叶斯主义的复现结果显示在金钱激励条件下，被试的辨别力提高，这强烈支持了金钱激励对准确性动机的影响的结论。这是一点非常有价值的发现，因为它深化了我们对政治新闻真假判断的心理机制的理解。然而，与原文不一致的地方在于未观察到信息来源的主效应及相关交互作用显著。这可能源于多种原因，包括实验条件的微小差异、参与者的心理状态、文化背景等。这体现了我们的工作的深远的科学和实践价值。首先，通过复现关键实证研究，通过贝叶斯主义的方法，我们巩固了先前研究的可靠性，为科学领域提供了重要的验证，但也发现了先前文章中可能的不恰当的结果。其次，展示了开放透明的科研理念，为学术合作和知识分享创造了有利条件。此外，通过创新地结合广义线性模型和信号检测论，为实验心理学提供了新的理论视角，丰富了学科讨论。最重要的是，我们的工作直接影响了决策者和社会，提供了对政治信息判断的实证支持，对媒体从业者、政治决策者和公众都具有指导意义。综合来看，我们的浮现结果在推动科学进步、促进学科发展以及为决策

者提供实用指导方面发挥了重要作用。

参考文献

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353-369.

Pennycook, G. & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388-402.

Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature human behavior, 7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Taber, C. S. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769.