arXiv:2601.06709v1 [physics.soc-ph] 10 Jan 2026

B旁观者效应源于个体的心理前景

Tiffanie Ng*

数学与统计学系，

经济学系，凯尼恩学院

Sara M. Clifton†

数学系，丹尼森大学（日期：2026年1月13日）

摘要

旁观者效应是一种社会心理现象，指当有其他人在场时，个体帮助可能需要帮助者的可能性会降低。社会学家和心理学家提出了多种合理解释，试图说明旁观者效应为何发生，从情境模糊性和社会传染，到责任分散和相互否认。基于这些社会心理学假说，并借鉴前景理论的观点，我们构建了一个新模型，用以描述个体在旁观者情境中是否进行干预的决策。该模型首次表明，在情境模糊的旁观者情境中，旁观者效应源于彼此不协调的个体对社会风险的感知。在此静态模型的基础上，我们进一步探讨社会学习的影响：个体在经历或目击以往干预所带来的社会影响后，会更新自己对干预风险的感知。该模型的一项新颖发现是，社会学习会加剧旁观者效应。我们使用一个全新的数据库对这些模型进行了验证，该数据库收录了涵盖广泛旁观者情境的42项实验研究和观察性研究。研究结果为这一观察到的现象提供了简明且具有普适性的解释，并可能有助于针对具体的旁观者情境制定有效的干预措施。

关键词：行为经济学、数学模型、复杂系统、动力系统、分段光滑、社会学习

* ng2@kenyon.edu

† cliftons@denison.edu

一、引言

旁观者效应（也称旁观者冷漠 或旁观者抑制）是一种社会现象：随着目击者数量增加，个体向可能需要帮助的人提供援助的可能性会降低 [1]。社会心理学家 John M. Darley 和 Bibb Latan´e 开创了旁观者效应的研究。他们在20世纪60年代发表了一系列论文，探讨了这一效应产生的各种情境和心理因素[2–5]。

关于旁观者效应背后的社会心理机制，早期理论包括责任分散（所有目击者之间对义务进行理性分担）、观众抑制（在人群较大时害怕当众尴尬）以及社会影响（除非其他人也在采取行动，否则不愿采取行动） [6]。然而，具体机制可能取决于所考察的旁观者情境类型。旁观者情境大致可根据危险的程度和类型（如有危险）、情境模糊性的程度，以及目击者提供帮助的能力程度进行分类 [6]。

在这些更广泛的类别中，研究考察了受害者/目击者人口统计特征（例如年龄 [7–9], 种族 [10–12], 性别 [13–16]）、环境因素（例如农村与城市 [17], 实验室与实地 [18]），以及目击者或受害者之间的社会或物理联系（例如朋友与陌生人 [19–21], 能够交流与彼此隔离 [22, 23], 面对面与在线 [24–27]）的影响。关注的指标不仅包括干预率，还包括反应延迟 [4, 28] 以及所采取行动的程度 [29, 30]。

尽管旁观者效应是在各种情境中观察到的稳健现象 [6],，但该效应并非普遍存在。事实上，当情境明显危险且目击者具备提供帮助的能力时，旁观者效应会减弱或逆转 [31, 32]。旁观者效应最常见于情境模糊 [33, 34] 且不存在身体危险的情况下，这表明该现象可能主要由社会风险而非身体风险驱动[35, 36]。因此，我们将研究重点放在模糊、非危险的旁观者情境上；在这些情境中，目击者不确定何种行为才是恰当的，并且目击者之间可能无法就采取行动是否恰当达成一致。

A. 背景

已有许多关于旁观者效应的定性模型得到发展和检验，其中包括五步旁观者干预模型 [37], 。在该模型中，目击者必须：（1）注意到事件；（2）判断某人可能需要帮助；（3）决定是否承担个人责任；（4）判断自己是否能够提供帮助；（5）有意识地决定采取行动或不采取行动。任何阶段的决策都可能导致旁观者效应；具体而言，情境模糊性会在第2步发挥作用，责任分散或社会影响会在第3步发挥作用，目击者能力会在第4步发挥作用，而观众抑制会在第5步发挥作用。在旁观者干预模型的基础上，Piliavin、Dovidio、Gaertner和Clark提出了唤醒—成本—回报模型 1981 [38, 39], ，认为人们看到有人需要帮助时会感到不适，并会通过选择最高奖励和/或最低成本的方式来有效减轻这种不适。人们考虑的行动或不行动的成本与回报主要是情绪性的（例如自豪感、尴尬、内疚、恐惧），而非理性的（例如金钱、时间、安全），并且这些决策可能并非有意识作出。

关于旁观者效应的定量模型相当有限，但现有的少数模型采用了各种数学框架，复杂程度范围广泛；这些模型通常包含与旁观者干预模型或唤醒—成本—回报模型相一致的决策过程。在较复杂的模型中，个体决策状态的时序因果网络模型在定性上与预期的唤醒水平动态相符 [40]。在较简单的建模方法中，一个博弈论模型表明，采取行动或不采取行动的理性成本与收益会产生旁观者效应，这为缩小班级规模以减少欺凌提供了依据 [41]。最简单的旁观者效应数学模型显示，美国俄亥俄州哥伦布市一家牛排馆的小费率遵循幂律，幂指数为 ≈ 0.22 [42]，这被归因于责任分散。集体决策的定量模型则更为常见（例如， [43]），尽管这些模型并未明确讨论旁观者效应。

由于旁观者情境往往要求人们在不确定情况下迅速做出决策（且可能面临社交风险），我们构建了一个纳入行为经济学概念的模型 [44]。尤其是，前景理论不同于传统经济框架，它承认人们面对不确定性以及潜在收益或损失时，并不会做出纯粹理性的决策 [45]。前景理论基于几项具有普遍性的观察：（1）人们会以现状（即不采取行动）为参照做出决策；（2）损失带来的痛苦大于同等收益带来的愉悦（这一概念称为“损失厌恶”）；（3）人们往往会高估较小概率、低估较大概率。前景理论承认这些观察到的非理性行为

这些行为打破了更传统的期望效用理论所假设的对称性（例如，行动与不行动之间，或收益与损失之间的对称性）。

II. 方法

为了理解旁观者效应可能的起源，我们基于干预的社会成本与收益，构建了一个模型，用于描述个体在社会情境模糊且不存在危险的旁观者情境中，是否选择干预（第 II A 节）。在模型的简单版本中，我们假设所有个体都认为干预同样恰当，且这些看法是固定的。我们在第 II B 节探讨这一模型。接着，我们假设个体会从旁观者情境的结果中学习（第 II C 节），在此基础上扩展静态旁观者模型；讨论用于模拟动态变化的各种框架（第 II D 节）；构建并探讨一个简单的全连接网络框架（第 II E 节）；并检验模型结果对参数的敏感性（第 II F 节）。最后，我们介绍一个新数据库，其中汇编了观察性和实验性旁观者效应数据（第 II G 节），并将模型行为与这些数据进行比较。

A. 旁观者模型

设想这样一种社会情境：有人似乎需要帮助，例如一位可能年迈或怀孕的人，在拥挤的公交车上或许需要一个座位。在这种情况下，目击者可以选择是否采取行动。[1] 在这种社会情境模糊的情况下采取行动，伴随着心理风险。如果这一行动被认为恰当，他们会获得自豪感和正义感带来的情绪收益。然而，如果这一行动被认为不恰当，他们可能会感到尴尬或羞耻。另一方面，选择不采取行动可能会让人感觉更接近中性。

采取行动的心理风险可视为一种前景（心理赌博），默认选项是不采取行动 [45]。尽管选择不采取行动不会带来净心理收益，但决定采取行动就意味着承担风险：行动可能得到正面回应（例如感谢或点头），也可能得到负面回应（例如皱眉或谴责）。每个决定采取行动的人都会评估他人（目击者或需要帮助的人）

[1] 在我们的方法中，“采取行动”意味着某人决定采取行动，无论目击者是否需要实际完成该行动。举例来说，假设有三个人决定让出公交车座位，但只有第一个起身的人让出了座位。在这种情况下，我们认为三个人都采取了行动。

是否会对该行动感到不满，我们用数值 $r \in [0,1]$ 表示。至少一名目击者表现出不满的感知概率计算为 $s = 1-(1-r)^N$，其中 N 是目击者数量[2]；这里我们假设表现出不满的概率彼此独立。随后，我们针对某位目击者，将采取行动的期望值计算为 $V = B(1-s) - Hs$，其中 B 和 H 分别代表行动带来的心理收益或伤害。为减少参数数量（并且由于 B 和 H 没有公认的计量单位），我们将期望值方程除以 B，得到 $v = 1-s-sx$，其中 $x = H/B$ 是损失厌恶比率。根据前景理论的原则 [45], ，如果采取行动的期望值大于不采取行动的价值（$v > 0$），目击者就会选择干预；否则，他们不会采取行动。

原则上，每位目击者（以 i 为索引）都可能感知到不同的风险值 r_i。不过，我们先考察目击者的风险评估具有同质性的情况；这是一个特殊情形，在此情形下，社会所有成员都认同干预的适当性。此外，目击者的损失厌恶比率可能各不相同 x_i，而平均损失厌恶比率通常已知介于 1.5 和 2.5 [48–50].[3] 之间。

鉴于我们预期群体层面的损失厌恶大致呈钟形分布，同时不会出现负的损失厌恶，因此假设损失厌恶比率的种群分布服从伽马分布[4]:

$$x_i \sim \Gamma\left(\alpha = \left(\frac{\mu}{\sigma}\right)^2, \beta = \frac{\sigma^2}{\mu}\right),$$

其中，μ 和 σ 分别是伽马分布的均值和标准差。损失厌恶比率的概率密度函数（PDF）为

$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad x > 0$$

采取行动的预期心理价值的概率密度函数为 $f_V(v) = \frac{1}{s} f_X\left(\frac{1-s-v}{s}\right)$。

在给定情境下，对于包含 N 名目击者的情况，可以通过求值分布中为正的部分所占的比例，来确定将采取行动的目击者比例:

$$I = \int_0^\infty f_V(v)\,\mathrm{d}v = 1 - \frac{y^\alpha E_{1-\alpha}(y)}{\Gamma(\alpha)} = 1 - \frac{\Gamma(\alpha, y)}{\Gamma(\alpha)}, \tag{1}$$

[2]此模型假设，情境中有一个可能需要帮助的个体；该个体可能会谴责他人的行动，但并非目击者，因为他们无法干预以帮助自己。若要适用于不存在此类个体的情境（例如烟雾渗入房间 [3]），$s = 1-(1-r)^{N-1}$。若要适用于存在多个此类个体的情境（例如袭击或欺凌 [18, 30, 31, 46, 47]），则按理论上可能谴责某个个体干预行为的其他相关人员数量，对 N 进行调整。由于这类旁观者情境的变化只会调整 N，我们仅探讨最常见的情形，即有一名非目击者。[3] 这些研究涉及金钱的获得或损失风险。尽管我们从定性层面知道，人群中的干预倾向各不相同 [51], 但尚不清楚心理风险在定性或定量上是否与货币风险相似。[4] 采用对数正态分布或左侧负尾可忽略的正态分布时，结果相似。5

其中，$y = \frac{(\quad 1-r^{\ N})\,(\ )^N}{\beta(1-\ 1-r}$，$E_a(x,x)$ 是指数积分，而 $\Gamma(a,x)$ 是不完全伽马函数。我们看到，随着 N 增加，进行干预的目击者比例会下降；这是一种涌现的旁观者效应，即使所有目击者都认同采取某项行动是适当的，这种效应仍然会发生（图 1）。

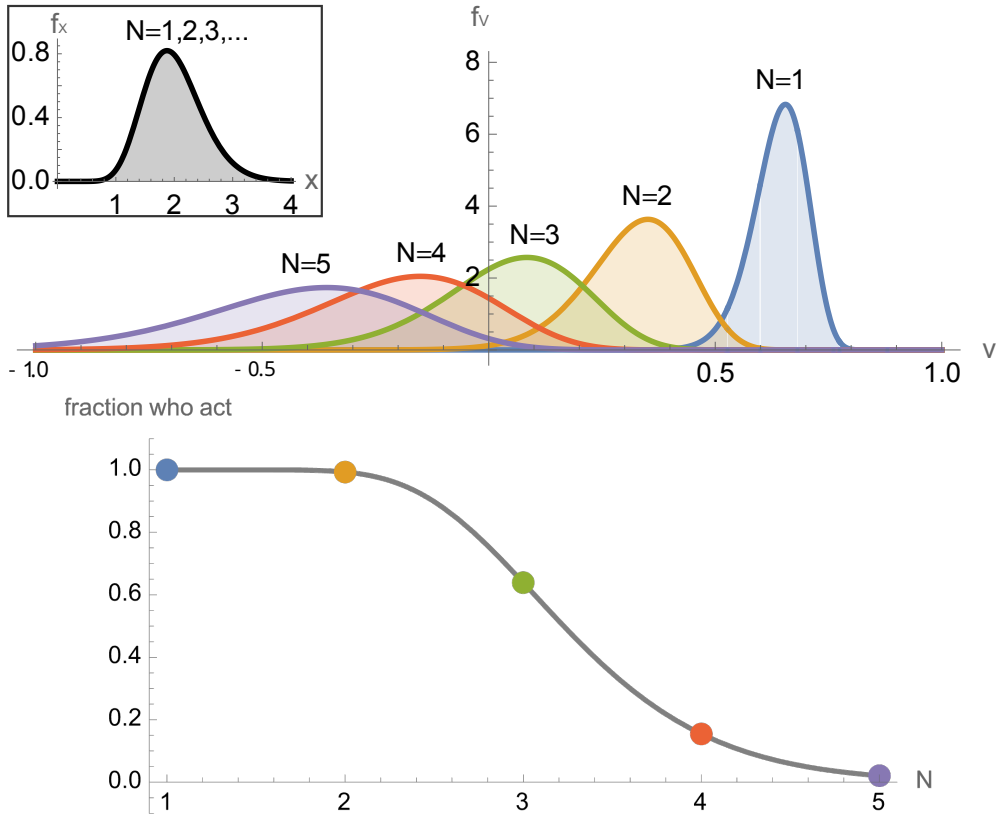

图 1. 假设损失厌恶比率服从伽马分布，且 $\mu = 2$ 和 $\sigma = 0.5$ （顶部插图）符合实验测定值 [48–50], ，则相应的行动期望值分布会随着人群规模 N 增加而向左移动（顶部）。在此例中，$r = 0.12$ 是某个人认为旁观者情境下会谴责干预的概率。落在范围$v \in (0,\ \infty)$内的 f_V 比例代表会进行干预的个体比例（I）；该比例会随人群规模增大而下降（底部）。

尽管许多社会学实验记录了在旁观者情境中采取干预行动的个体比例（这是我们关注的指标），但其他实验只记录受害者至少得到一位目击者帮助的旁观者情境比例。为换算成这一指标，我们假设个体采取干预行动的概率为 p；那么至少有一位目击者采取干预行动（受害者得到帮助）的概率为 $1-(1-p)^N$，这与 Latan´e 和 Nida [22] 使用的公式相同。图 2 中绘出了两条曲线，分别对应 $r = 0.15$，$\mu = 2$，$\sigma = 0.5$ 。我们看到，

关系　　人群规模与干预比例／受助受害者比例之间的关系相似　　r.

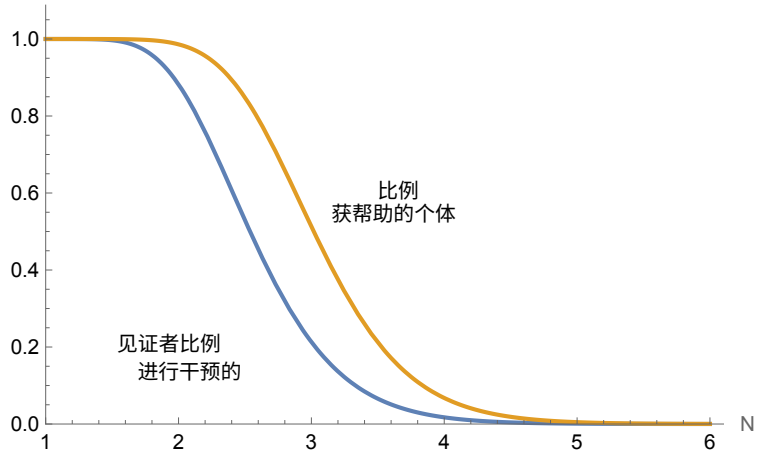

图 2. 旁观者情境中将进行干预的目击者比例与旁观者情境中将获得帮助的受害者比例之间的关系；图中显示的是损失厌恶比率服从伽马分布、其参数为 $\mu = 2$ 和 $\sigma = 0.5$，且行动适当性为 $r = 0.15$的情况。旁观者曲线保持相同的定性形状，但向右移动（对于任意给定的目击者数量，目击者获得帮助的概率大于任意给定目击者提供帮助的概率）。

B. StaticModel行为

在假设干预的适当性恒定且同质的前提下，我们探究模型行为如何随行动风险 r、损失厌恶比率分布的均值 μ 和标准差 σ 变化（见图 3）。在其他条件不变的情况下，增加行动风险 r 会使旁观者曲线左移。这意味着，干预感觉越适当，旁观者在人群规模相同的情况下就越愿意采取行动（例如，扶着门让提着杂货的人通过等社会适当行为更有可能发生，而未经他人允许触碰对方等社会风险行为则不太可能发生）。

同样，在其他条件不变的情况下，提高平均损失厌恶比率会使旁观者曲线向左收缩。这意味着，公开羞辱带来的痛苦越强烈，旁观者在任何特定情境下就越不可能进行干预。

更令人惊讶的是，损失厌恶标准差增大，会导致大规模-N 干预比例上升，而小规模-N 干预比例下降。这表明，随着人群在损失厌恶方面变得不那么同质，旁观者效应会减弱（也就是说，随着人群规模扩大，愿意干预的个体比例下降得更慢）。

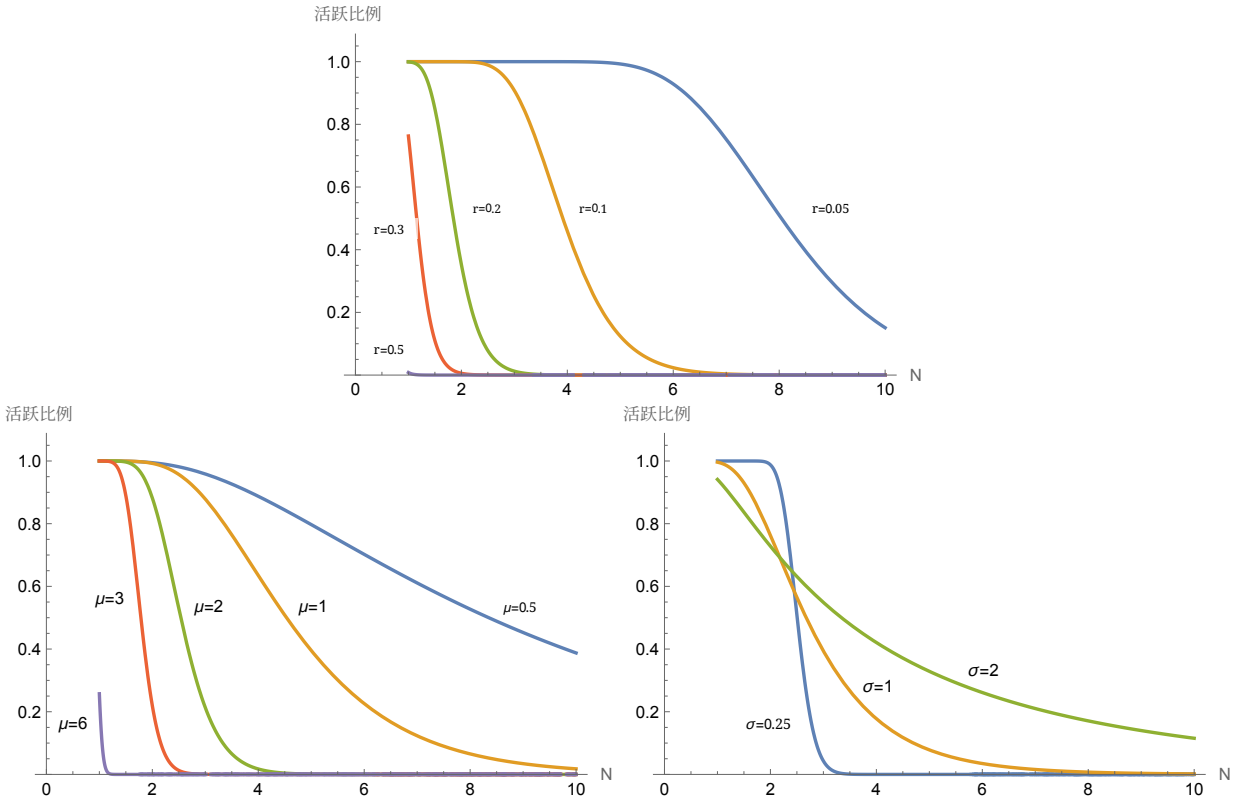

图 3. 采取干预的旁观者比例随旁观者人数的变化 N。除非另有说明，我们假设损失厌恶比率服从伽马分布，且 $\mu = 2$ 和 $\sigma = 0.5$ [48–50]，行动适当性为 $r = 0.15$。其他条件不变时，提高行动适当性 r 会使旁观者曲线左移（上方中间）。提高平均损失厌恶比率会使旁观者曲线向左压缩（左下）。提高损失厌恶标准差会增加大N 时的干预比例，同时降低小N 时的干预比例（右下）。

C. 社会学习模型

由于对是否适合干预的认识是通过反复接触旁观者情境而习得的，接下来我们将探讨个体如何从自身行为和他人的行为中学习，以及他们的感知风险 r_i 如何随之变化。假设目击者在以下四种情境下更新其风险 r_i：目击者进行了干预，但结果反响不佳或良好；或者目击者没有干预，但其他人进行了干预，且结果反响不佳或良好。如果个体或他人的行为反响不佳，目击者就会提高其行动风险；反之，反响良好的行为则会降低行动风险。我们通过假设以下内容来纳入这四种学习机制：

eac个体的风险根据微分方程持续更新 n

$$\frac{\mathrm{d}r_i}{\mathrm{d}t} = \underbrace{r_i(1-r_i)}_{\text{restrict } r\in[0,1]} \Bigg(\underbrace{a\,P_i^{(I)}(1-P_i^{(w)})}_{\text{act received poorly}} - \underbrace{b\,P_i^{(I)}P_i^{(w)}}_{\text{act received well}} + \underbrace{c\,(1-P_i^{(I)})\tilde{P}_i^{(p)}}_{\text{others' act received poorly}} - \underbrace{d\,(1-P_i^{(I)})\tilde{P}_i^{(w)}}_{\text{others' act received well}} \Bigg), \tag{2}$$

其中， a， b， c， d 分别是上述四种情境下的学习率，而 $P_i^{(I)}$、 $P_i^{(w)}$、 $\tilde{P}_i^{(p)}$ 和 $\tilde{P}_i^{(w)}$ 分别是个体 i 进行干预、个体 i的行动受到欢迎，以及另一名个体的行动受到负面或正面评价的概率，具体如下所述。

个体 i 进行干预的概率为

$$P_i^{(I)} = \begin{cases} 1, & 1-(1+x_i)s_i > 0 \\ 0, & 1-(1+x_i)s_i \le 0 \end{cases} = H(1-(1+x_i)s_i)$$

其中， H 是海维赛德函数， $s_i = 1-(1-r_i)^N$，这表示每个个体都认为其他所有人的风险水平与自己相同（没人会读心）。

概率 在个体采取行动的前提下，其行动受到欢迎的概率为 s

$$P_i^{(w)} = \prod_{j\neq i}\Big(P_j^{(I)} + \big(1-P_j^{(I)}\big)(1-r_j)\Big)$$

我们在补充材料中推导了这一概率。

在个体 i 未进行干预的情况下，另一位目击者的行动受到欢迎的概率为

$$\tilde{P}_i^{(w)} = \prod_{j\neq i}\Big(P_j^{(I)} + \big(1-P_j^{(I)}\big)(1-r_j)\Big) - \prod_{j\neq i}\big(1-P_j^{(I)}\big)(1-r_j),$$

其中，第一个乘积涵盖了无人因其行为而遭到“公开指责”的所有可能结果，第二个乘积则对无人采取行动、因而个体 i 无法从这种情境中学习这一结果进行了修正。

最后，在个体 i 未进行干预的情况下，另一位目击者的行动不受欢迎的概率为

$$\tilde{P}_i^{(p)} = 1-\tilde{P}_i^{(w)} - \prod_{j\neq i}\big(1-P_j^{(I)}\big),$$

其中，乘积项校正了无人采取行动的结果，因此个体 i 无法从这种情境中学习。关于这些概率的图示推导，请参阅补充材料。

D. 动态模型结构

至少有三种合理的方法可以模拟模型的动态。最简单的策略是假设人群由 N 个持续互动的个体组成（即，个体之间通过全连接社交网络相连）；这会得到一个由方程（2）规定的 N 常微分方程组。本文重点讨论这一简单方案，并在补充材料中探讨其他方案。

更贴近现实的假设是：一个规模庞大的有限种群在随机选取的（但静态的）网络上相互作用；由此得到一个动态网络模型，其中每个种群成员对应一个节点。我们在 S1B 节探讨这一方案。还可以假设网络结构本身是动态的；对此进行分析超出了本研究的范围。

最后，还可以假设 N 是连续统总体中的一个参数，由此得到一个关于总体风险分布 r 的偏微分方程。我们在 S1 C 节推导该偏微分方程，但将其分析留待未来研究。

E. 动态模型行为

如果 N 个体通过经历有 N 名目击者的旁观者情境而不断学习，那么该系统可视为一个动态全连接网络（即完全图）。我们首先在一个简化假设下，开始探讨模型行为：相较于从自身行动中学习 ($c = d = 0$)[5]，从他人行为中学习的影响可以忽略不计。

在一次典型的模拟中（见图 4），通过自身行动进行社会学习会使个人风险 r_i 增加，直到个体停止行动，因而也停止学习。个体并非同时达到各自的稳态风险水平，也不是因为其行动产生正面或负面反应的概率开始趋于平衡而停止行动；相反，随着风险逐渐增加，当行动的价值越过零阈值时，个体采取行动的概率会骤然降至零。这种行为在分段光滑动力系统中很常见 [69]。

对于不同的随机初始条件，所有个体的风险也可能呈指数级衰减至零。当所有个体一开始的风险都足够低，以至于他们最初都会采取行动，并且被其他所有人谴责的概率始终为零时，就会出现这种情况。事实上，系统最终达到的稳态对初始条件相当敏感（参见敏感性分析，II F 节）。

[5] 认为观察学习可以忽略不计并不合理 [64–66]。我们探讨这种情况，只是为了在加入观察学习的作用之前，先检验实践学习的作用。

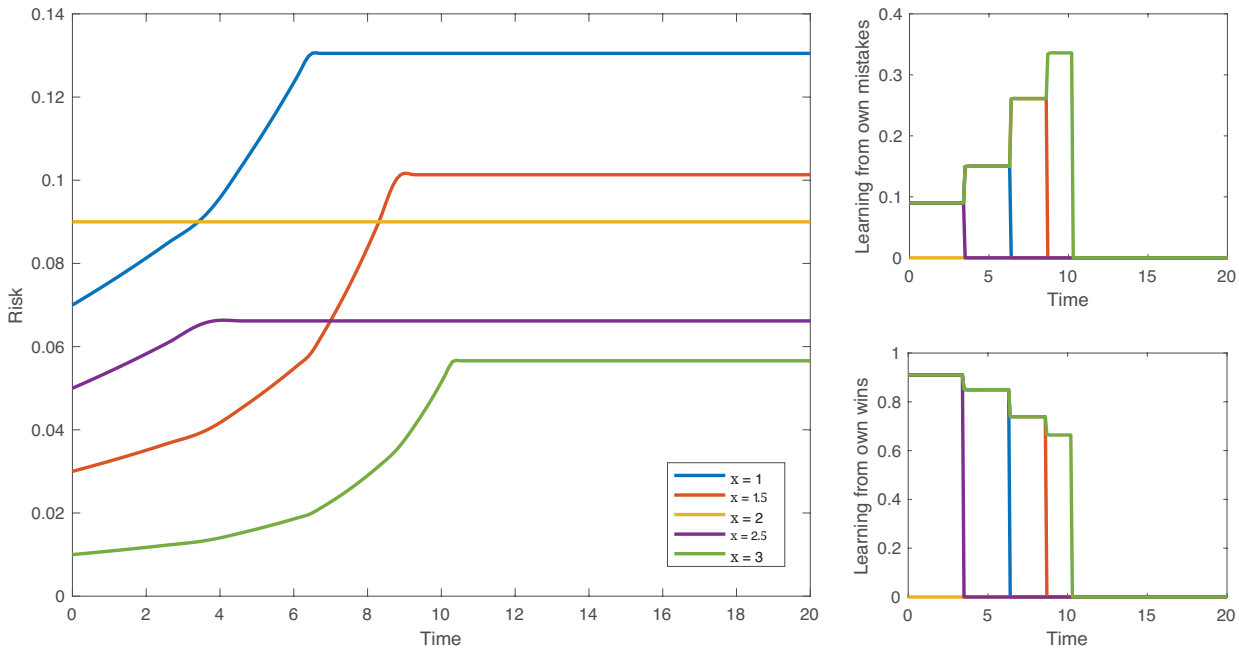

图 4. 代表性模拟，从 $r_i(0) \sim \mathcal{U}(0,0.1)$ 中抽样初始风险，且仅从自身行为中学习： $a = 1, b = 0.01$, $c = d = 0$ （左）。从自身反响不佳的行为（右上）和自身反响良好的行为（右下）进行社会学习的贡献，以这些结果发生的概率衡量。我们看到，学习在时间上呈分段式，仅当个体采取行动时才会发生学习。例如，“黄色”个体以高风险起步，从不采取行动，因此从不学习；而“绿色”个体以低风险起步，并从逐渐上升的受谴责概率中学习，直到最终停止行动。

简化假设认为个体只能从自己的行为中学习，这会导致一个不切实际的结果：不采取行动的个体无法更新自己的风险水平。在更现实的假设下，个体既从自己的行为中学习，也从他人的行为中学习；目击者会从只依据自身行为学习，转而从他人的行为中学习。由于采取行动的概率不是 0 就是 1，目击者无法同时从采取行动和不采取行动中学习。在一次典型的模拟中（见图 5），目击者起初会从自己的行为中学习，直到学到的风险达到某个阈值，从而停止采取行动；每位目击者停止行动后，仍会通过观察他人因采取行动而遭到谴责，继续学习并提高对风险的认知。

在其他条件不变的情况下，稳态习得风险水平 r_i^* 对目击者数量 N 呈非单调依赖关系（见图 6）。在表 I 所列参数下，人群规模足够小时，所有目击者的风险都会趋于零。当人群规模达到某个阈值时（在我们的参数取值下，约为 3 至 6 名目击者），稳态风险中位数会跃升至初始风险中位数以上。人群规模超过该阈值后，稳态风险中位数会下降

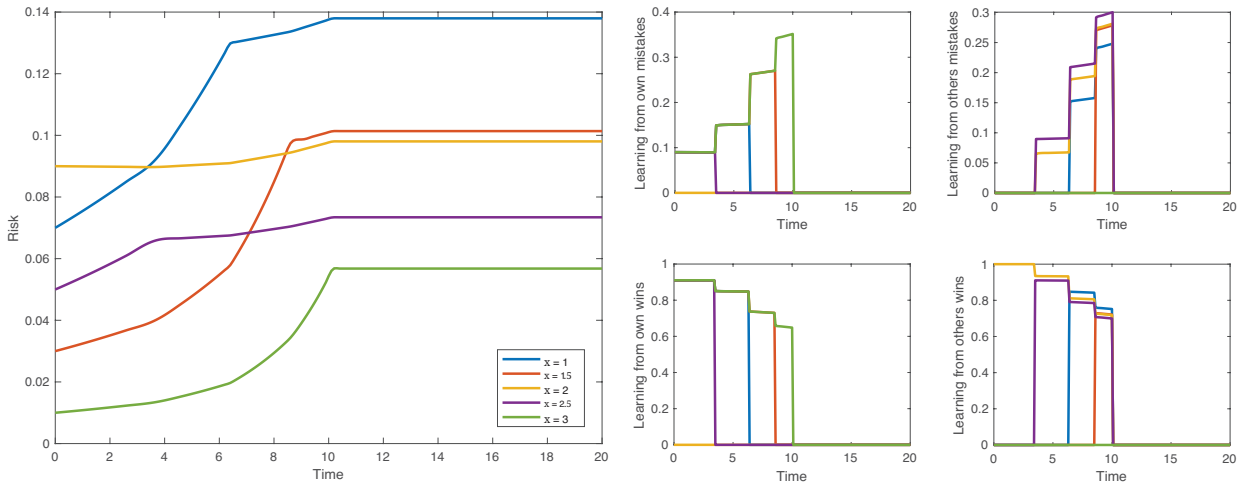

图5. 代表性模拟：从 $r_i(0) \sim \mathcal{U}(0,\ 0。1)$ 中抽取初始风险 $N = 5$，并学习自身及他人行动所引发的反应： $a = 1,\ b = 0.01,\ c = 0.1,\ d = 0.001$ （左），依据的证据表明，从经验中学习比从观察中学习更快 [57–60]，从负面结果中学习比从正面结果中学习更快 [52–56]。自身反响不佳的行为（中上）、自身反响良好的行为（中下）、他人反响不佳的行为（右上）以及他人反响良好的行为（右下）对社会学习的贡献，以这些结果发生的概率衡量。当个体也从他人的行为中学习时，我们会看到目击者从只学习自身行为，转而学习他人的行为。例如，“蓝色”最初通过自身行为学习，直到其习得的风险达到阈值并停止行动；此后，“蓝色”继续通过观察他人因采取行动而受到谴责，学习到风险在增加。

少许，并在人群规模达到 10 至 12 人后基本保持不变（这与心理学研究的结论一致：人类无法快速感知超过少数几个的数量 [70]）。

F. 敏感性分析

由于旁观者情境的社会和文化背景差异很大，我们进行敏感性与不确定性分析，以了解各模型参数对所关注观测结果——行动的稳态感知风险（我们使用人口风险中位数）——的影响。对于任意给定的人群规模 N，参数与稳态风险水平中位数之间的关系通常是非线性的，但呈单调变化，因此我们采用拉丁超立方抽样（LHS），从参数空间中抽取 $n = 1000$ 个样本，并计算偏秩相关系数（PRCCs）[71]。

敏感性分析（图 8）表明，无论人群规模如何，稳态风险水平

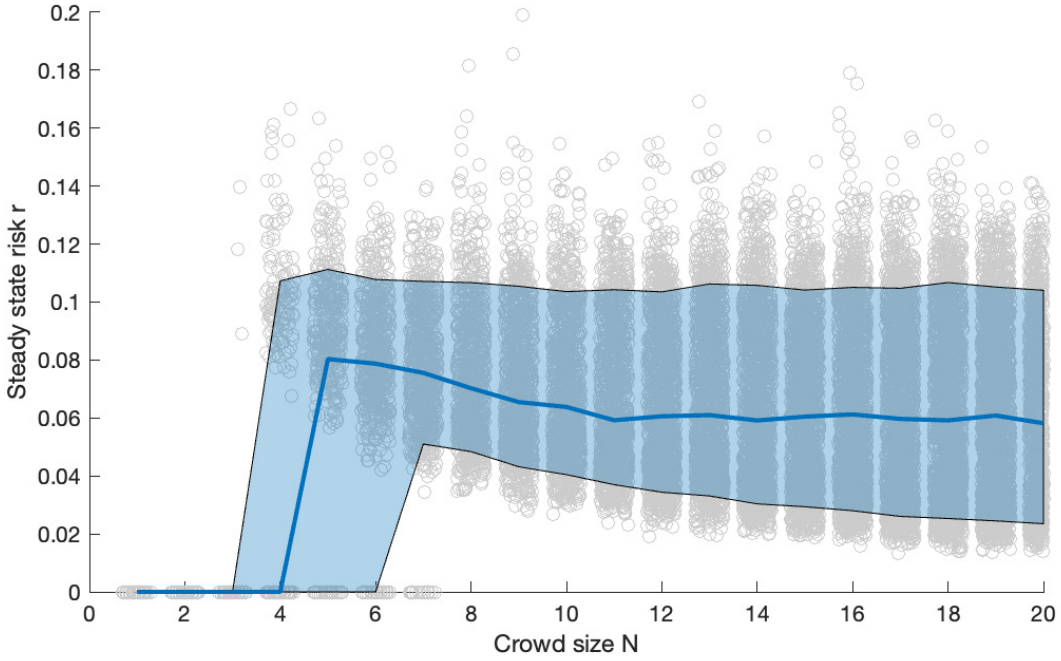

图 6. 在全连接网络中，稳态习得风险随目击者数量 N 的变化，其他条件均保持为表 I 中的基准值。灰点表示 100 次模型模拟中的个体目击者稳态风险水平 r_i，并在水平方向上添加抖动以便于观察。每个 N 的稳态风险中位数（蓝线）会随着 N 的增加呈现非单调变化：当人群规模足够小时，所有目击者的风险都趋近于零；超过临界人群规模后，习得风险会跃升至初始风险中位数之上；当人群规模进一步超过该阈值水平时，风险略有下降，随后保持近似不变。习得风险的方差往往会随着 N 增加而增大（每个 N 的模拟结果的第 10% 至第 90% 分位数以蓝色阴影表示）。

对最大初始风险（$r_{\max}$）最为敏感，其中初始条件从$r_i(0) \sim \mathcal{U}(0,\ r_{\max})$中采样。较大的初始风险中位数会导致较大的稳态风险中位数。这是因为，当社会认为某项行动不恰当时，个体进行干预的可能性较低；当人们不采取行动时，就没有人能够从行动结果中学习，因此风险会维持在较高水平。

此外，对于任何人群规模，提高从他人错误中学习的学习率（c）都会对稳态风险产生正面影响；这是合理的，因为如果人们通过观察干预者出丑而更快地吸取教训，他们就不太可能进行干预。当无人被“公开指责”时，从旁观者正面结果中学习的学习率对稳态风险的影响较小；随着人群规模扩大，从他人正面结果中学习（d）会变得更加重要（因为可供学习的干预行为更多），但从自身正面结果中学习（b）则只在

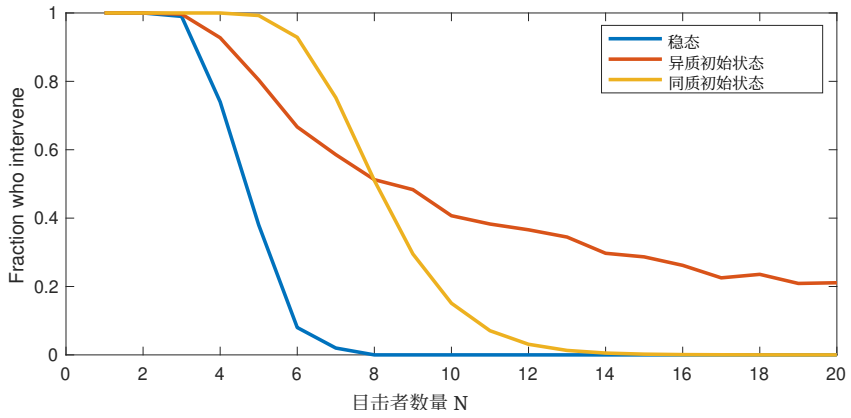

图 7. 在其他条件均固定为表 I 中基准值的情况下，初始与稳态干预比例随目击者数量 N 的变化。经过社会学习（公式 (2)）后，旁观者曲线从初始状态（红色）向左移至稳态（蓝色），这表明社会学习会加剧旁观者效应。我们还展示了初始旁观者曲线，假设所有旁观者最初具有相同的风险感知，且该感知等于初始风险均值（黄色），这说明同质风险假设无法很好地近似异质风险。

人群规模较小时才显著（因为在人数较多的人群中，个体采取行动的可能性较低）。

有趣的是，损失厌恶比率的分布对小规模人群的稳态风险影响最大，但随着人群规模增大，其影响的幅度和显著性都会减弱。对于从 $x_i \sim \Gamma\left(\alpha = \left(\frac{\mu}{\sigma}\right)^2\right.$ 中抽取的损失厌恶比率，分布的均值会正向影响稳态风险，而标准差会负向影响风险；但随着人群规模增大，这些影响会减弱。一个合理的解释是，规模很大的人群（此处指十几人或更多）会使大多数目击者不愿采取行动，无论其损失厌恶程度如何；向少数愿意冒险干预的目击者学习则重要得多。

G. 数据

为验证我们的模型，我们汇编了一个数据库，收录了有关旁观者效应及其在各种条件下表现形式的观察性和实验数据。我们在 PubMed、SageJournals、APA PsycNet、JSTOR、Wiley 和 Google Scholar 等数据库中进行了文献检索，使用的关键词包括“旁观者干预”、“模糊性”、“群体规模”、“责任分散”、“旁观者冷漠”、“社会影响”、“模型”，以及这些词语的各种组合。若一项研究包含两个关键要素，我们便将其纳入： (1) 群体规模；(2) 一个量化的

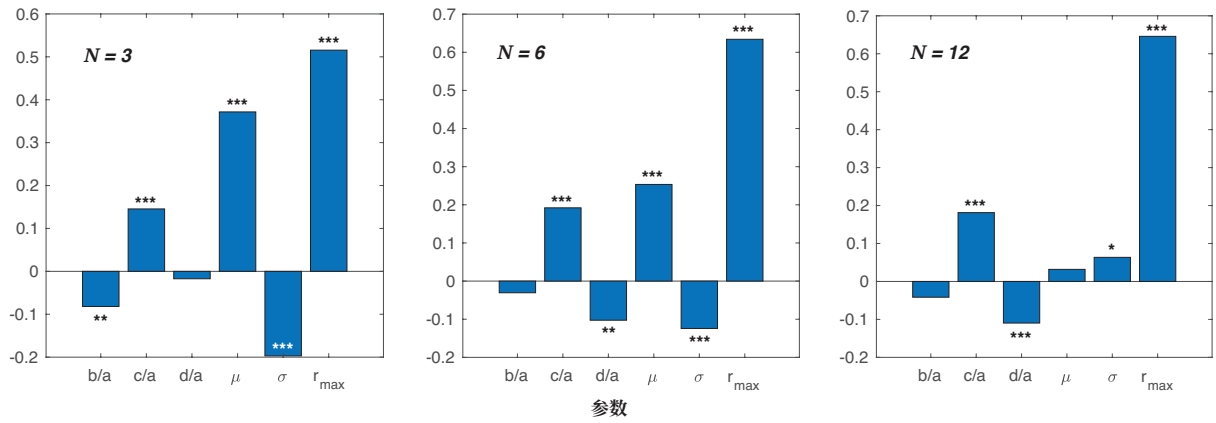

图 8. 参数对稳态感知风险水平中位数的敏感性 r_i。敏感性分析采用参数空间的拉丁超立方抽样（LHS）以及偏秩相关系数（PRCC）[71]。所有参数值均取自表 I 中基准值附近。初始条件从$r_i(0) \sim \mathcal{U}(0, r_{\max})$中抽样，其中 $r_{\max}$ 也是一个参数。星号表示在 1000 次模拟中具有显著性（$^{*}p < 0.05$，$^{**}p < 0.01$，$^{***}p < 0.001$）。我们发现，无论人群规模如何 N，稳态风险水平对最大初始风险（$r_{\max}$）最为敏感。对于较小的人群规模，损失厌恶比率的分布会显著影响习得的风险；对于较大的人群规模，学习率对习得的风险影响更大。

干预频率的指示指标[6]。

我们根据情境中受害者、施害者和目击者的数量计算群体规模；目击者可以是焦点个体（即受试者），也可以是实验同谋（即假扮受试者的实验人员）。在虚拟情境中，人群规模无法直接观察，而是通过受试者可见的数字表示。若提供的人群规模为一个范围，对于有限范围，我们采用其平均值；对于无上限范围，则采用其最小值。

干预频率通常以受害者获得帮助的情境比例，或采取干预行动的目击者比例来表示；当只有一名受试者作为目击者时，这两个比例相同。少数研究采用其他指标来衡量干预（例如，平均餐馆小费率，或李克特量表上的平均干预意愿）。在合理的情况下，我们会将这些替代指标换算为干预比例。

为便于可视化，我们将研究划分为几个大类：非危险性金钱情境（捐赠慈善、留下小费等）、掉落物品、烟雾渗入房间、虚拟请求、伴有或不伴有求救呼喊的跌倒、从轻微到严重的伤害、医疗紧急情况，以及犯罪行为（盗窃、袭击等）。显示似乎存在旁观者效应的情境见图 9

[6] 我们排除了两项研究，因为其中作为研究焦点的个体是尚未发展出心智理论能力的年幼儿童。

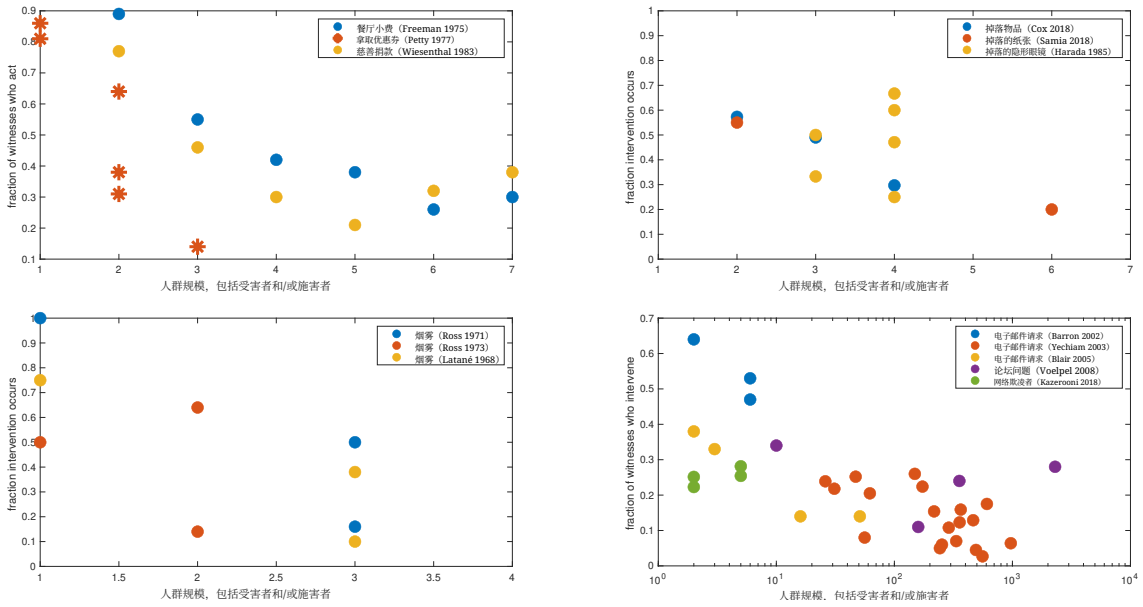

图9。按情境类型划分的旁观者情境数据，其中可能观察到旁观者效应：财务情境（左上） [42, 72, 73], 掉落物品（右上） [14, 74, 75], 房间内烟雾（左下）[3, 76, 77], 虚拟情境（右下） [25, 26, 29, 47, 78]。助人情境（左上）。在非危险情境中，可能观察到定性旁观者效应，但在紧急情况下，旁观者效应并不明显。由于每项研究的数据点数量较少，我们无法针对每种情境拟合模型，尽管这些趋势与模型一致。请注意，有些研究提供的是个体目击者的干预概率，而另一些研究提供的是发生干预的概率（无论由哪位目击者实施）。优惠券研究中的星号数据点表示，与图中其他研究不同，该研究测量的是干预发生的概率。

，显示似乎不存在旁观者效应的情境见图10。

III. 结果

由于大多数旁观者研究都是为假设检验而设计的，因此多数研究只纳入两种或三种不同的人群规模。因此，我们无法合理地为每种情境拟合数学模型。我们可以合理地选择将模型拟合到相似情境的集合，但哪些情境的行动适当性和损失厌恶分布足够相似，并不明确。鉴于数据限制，我们不拟合模型，而是仅展示：对于纳入至少六种不同人群规模的研究，我们的模型与某些情境相符，与另一些情境则较不相符（见图11）。

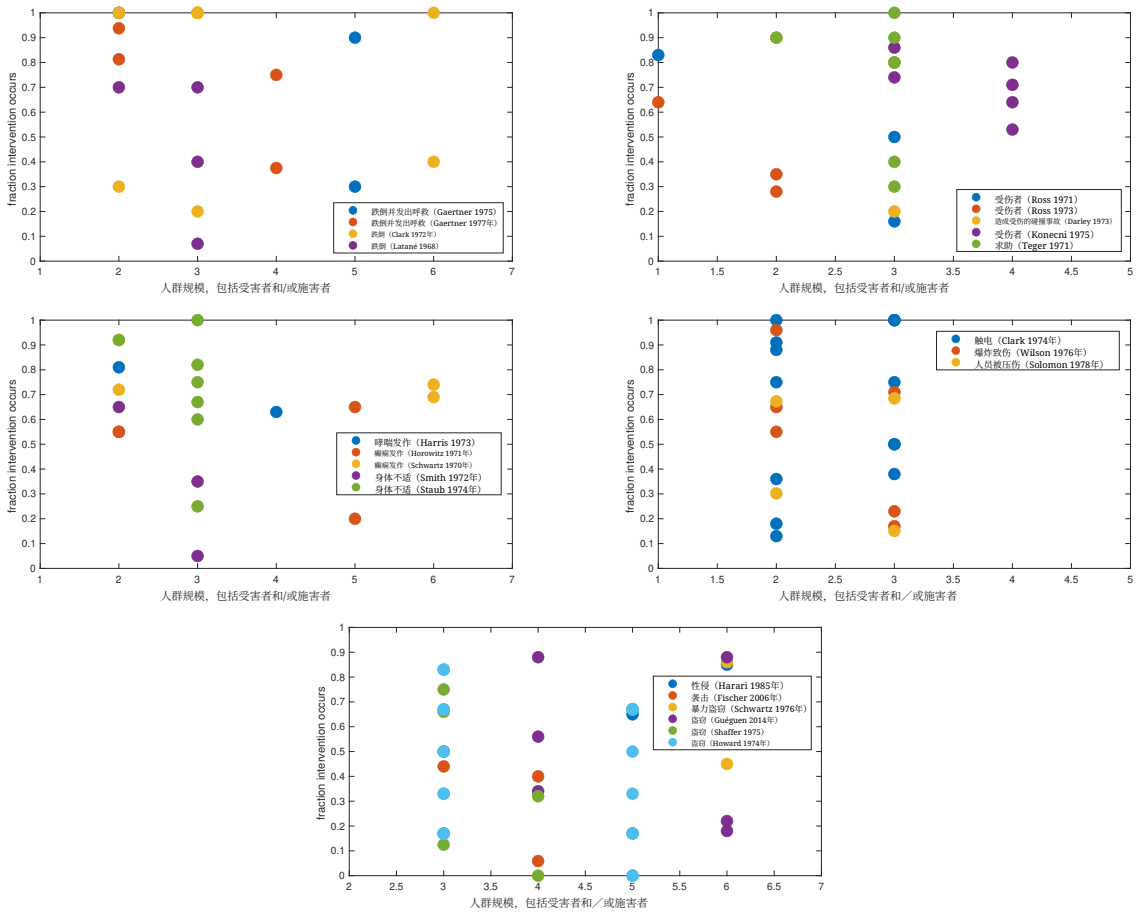

图10. 按情境类型划分的旁观者情境数据，其中可能看不到旁观者效应：跌倒（左上） [19, 79–81]，轻伤（右上） [76, 77, 82–84]，医疗紧急情况（左中） [85–89]，重伤（右中） [33, 90, 91]，犯罪情境（下方） [18, 28, 30, 31, 46, 92]。在非危险情境中，可能会观察到定性旁观者效应，但在紧急情况下，旁观者效应并不明显。由于每项研究的数据点数量较少，我们无法针对每种情境拟合模型，尽管这些趋势与我们的模型一致。请注意，一些研究提供的是个体目击者的干预概率，而另一些研究提供的是发生干预的概率（无论由哪位目击者实施）。优惠券研究中的星号数据点表明，与图中其他研究不同，该研究测量的是发生干预的概率。

尽管由于小型数据集造成的实际可辨识性问题，我们无法拟合模型，但仍能观察到若干普遍模式。在呈现旁观者效应的数据集中（例如非危险情境），随着人群规模增加，采取干预行动的目击者比例往往呈现递减且向上凹的趋势。在假设行动适宜性均匀分布的静态模型（1）中，出现这种情况的原因可能是行动风险较小但不可忽略，也可能是损失厌恶的变异较大（标准差接近所假设伽马分布的均值

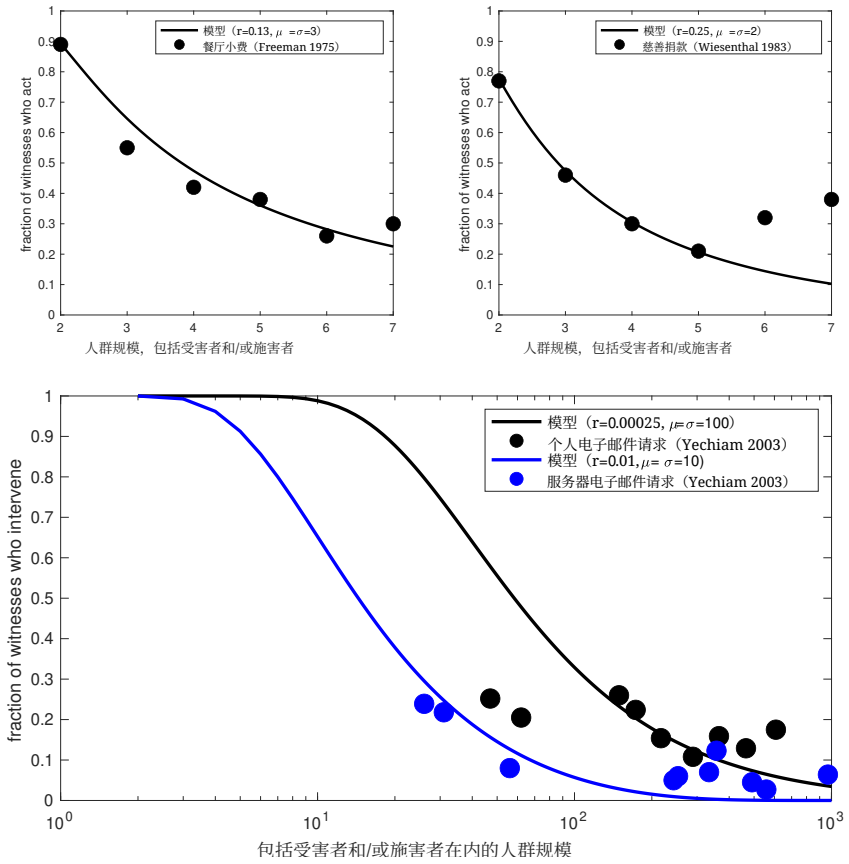

图 11。利用来自三种不同观察性或实验性旁观者情境的数据检验模型的合理性：餐厅小费金额随聚会人数的变化 [42], ，在所有目击者采取行动时为 20%、没有目击者采取行动时为 10%之间进行线性插值（左上）；在公共活动中，捐赠救济基金的人所占比例随群体规模的变化 [73] （右上）；以及回应电子邮件请求的人所占比例 [78] （下）。由于存在实际可辨识性问题，静态模型（1）未进行拟合，因此模型曲线仅显示合理的结果。

分布，从而形成指数分布）。在至少包含六种不同人群规模的研究中，所有非正式拟合都需要较低的风险和较大的损失厌恶变异（等于损失厌恶的均值），才能与数据相符。这表明，人群中的情绪前景方差可能大于金钱损失厌恶的方差（见图12）。虽然模型（1）对金融情境的拟合相当好，但对虚拟情境的拟合并不好；这可能是因为虚拟情境使用了数量庞大的虚拟见证者，而人们或许无法理解这么大的数量 [70, 93]。

如果不假设行动适当性服从均匀分布，那么数据也与损失厌恶程度的变异较小这一情况相符。假设损失厌恶服从伽马分布，且行动风险服从 $\mu = 2, \sigma = 0.5$ [48–50], 均匀分布 $U(0, 0.1)$，则会产生类似的、递减且向上凹的形状

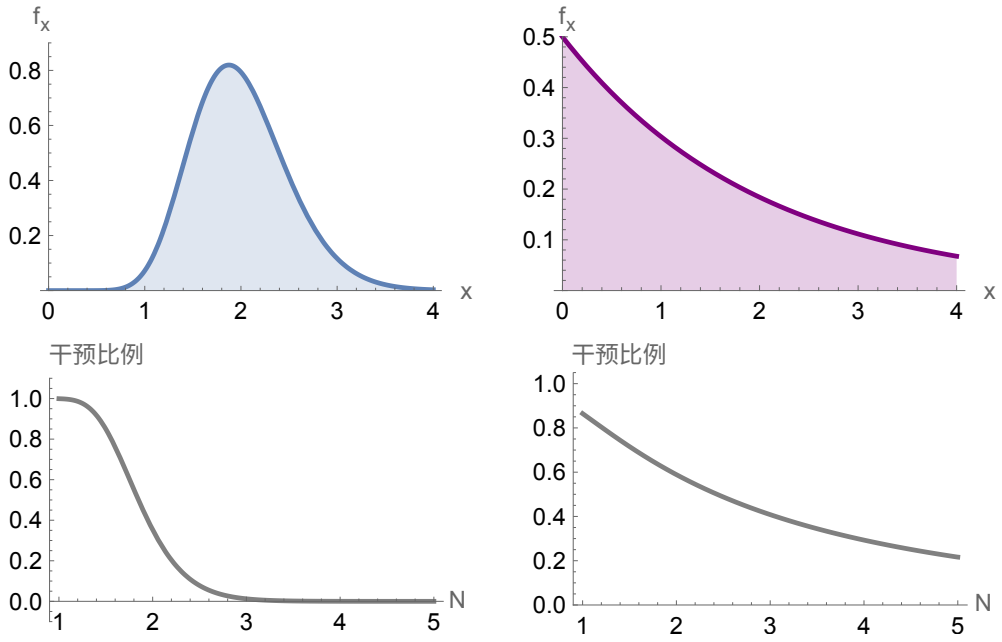

图12. 涌现的旁观者曲线（下）随损失厌恶分布（上）的变化。如果损失厌恶比率服从参数为 $\mu = 2$, $\sigma = 0.5$ 的伽马分布（左上），那么旁观者曲线呈S形（左下）；这在旁观者研究中并不常见。如果损失厌恶比率服从参数为 $\mu = 2$, $\sigma = 2$的伽马分布，即指数分布（右上），那么旁观者曲线呈衰减趋势（右下）；这在旁观者研究中更为常见。

这种形状见于许多数据集（参见图 7）；它与假设所有个体都持有该分布的平均风险水平有显著差异。数据与动态模型 (2) 的稳态不太吻合；该稳态夸大了随着人群规模扩大，干预概率在高低之间的切换（参见图 7）。这表明社会学习速度较慢，尚未达到稳态。

IV. 讨论

针对所观察到的旁观者效应，人们提出了几种定性机制，其中最突出的是责任分散、观众抑制和社会影响。本文提出的静态模型（1）定量检验了观众抑制假说的影响，而动态模型（2）纳入了社会影响（尽管其时间尺度长于心理学文献中通常考察的时间尺度）。无论采用静态模型还是动态模型，个体对公开羞辱的恐惧都会导致集体干预概率随着人群规模的扩大而降低。

有充分证据表明，即使情境很明确，但采取行动是否恰当可能存在歧义，社交尴尬也会抑制旁观者采取行动（例如，同事的牙齿上粘着食物，或脸上沾有墨水） [94]。事实上，许多研究表明，在很多情况下，由后悔引发的现状偏差使人们更倾向于不作为而非行动 [95, 96]；行动与不作为（即现状）之间根本不存在对称性，简单的责任分散模型无法捕捉到这一点，但我们的观众抑制模型通过损失厌恶对此进行了刻画（前景理论假设不行动效用为零，而由行动导致的损失或收益效用则取决于个人的损失厌恶比率）。

尽管如此，许多旁观者数据集似乎也与一种简单的责任分散模型相符，例如，干预者所占比例与 $1/N$ （或 $1/N$ 的某个幂次）成正比。鉴于数据如此有限（通常每项旁观者研究只有 2–3 个数据点），使用模型选择技术（例如赤池信息准则或交叉验证）来区分观众抑制模型和责任分散模型是不合理的。不过，有证据表明，简单的责任分散并不能普遍解释旁观者效应，因为旁观者在某些情境后报告说，他们原本没料到其他人会干预 [28]。此外，在非助人情境中也能观察到旁观者效应（例如，人们可以自由领取数量有限的优惠券），而在这类情境中，责任分散并不适用。

最后，许多旁观者数据集也可能与社会影响假说相符：如果他人采取行动，个体更可能采取行动；如果他人未能采取行动，个体更可能不作为。我们的动态模型（2）捕捉了他人的行为对类似情境中未来行动或不作为的影响，不过我们没有直接对单个旁观者情境中的行动／不行动级联进行建模。更复杂的模型版本将不再采用我们当前的假设，即在单个旁观者情境中，旁观者的决策彼此独立。

假设我们的模型已经捕捉到旁观者情境决策的重要特征，那么旁观者干预曲线（人群规模与干预概率的关系）表明，只有在某些情况下才会出现与非危险旁观者情境相符的旁观者效应：（1）干预风险的平均水平相对较低；（2）人群中的干预风险存在一定差异；和/或（3）人群中的损失厌恶程度差异较大。第一个条件合情合理，因为大多数旁观者实验的设计都认为，采取干预比不干预在文化上更容易被接受。第二和第三个条件意味着，人群中的异质性是观察到旁观者效应的关键驱动因素。如果所有人都认同采取行动是否合宜，并且对损失的厌恶程度相同，那么每种旁观者情境都会存在一个明确的人群规模干预阈值：低于该阈值时，所有人都会采取行动；高于该阈值时，没有人会

人们会采取行动。没有一项旁观者实验显示出这种行为。

A. 局限性与未来步骤

与所有简化模型一样，关于旁观者效应的静态模型（1）和动态模型（2）都简化了现实。在静态模型中，我们假设旁观者同时目击某个情境，并独立决定是否采取行动；综述论文 [97]也采用了这一假设。在某些需要迅速作出决定的情境中（例如回应挥手的人，或阻止笔记本电脑遭窃），这或许是很好的近似；但如果旁观者有时间考虑自己的决定并观察他人的行为（例如回复电子邮件请求），这一假设就不太适用。虽然我们通过动态社会学习模型纳入了社会影响，但静态模型的扩展版本可以在单个旁观者情境中纳入干预依赖性和响应时间。

静态模型还假设，人们会根据自身面临的个人风险，准确评估干预所带来的社会风险。然而，人们往往会高估小概率事件，无论是罕见的好事（例如中彩票），还是罕见的坏事（例如遭遇空难）[45]。如果将前景理论的这一特征纳入考虑，在假设估计风险与实际风险之间的关系是单调的（这一点已通过实验得到证实 [98]）的情况下，我们的结果在定量上会有所变化，但在定性上不会改变。

我们还假设，人们能够准确统计旁观者情境中的目击者数量；虽然人类能够快速、准确地数出约 4 或 5 件物品，但数量更多时，计数会变得费力且缓慢 [70]。在大多数人群规模较小的旁观者情境中，这不会影响我们的结果，但在某些人们采用数量的定性测量的情境中（例如，收件人众多的电子邮件请求 [78]），可能会降低模型的适用性。不过，当人群规模超过 10 至 12 人时，我们的模型结果几乎完全相同（见图 6），因此，对较大人群规模的测量误差很可能可以忽略不计。

我们的模型做出的另一项简化是，在旁观者情境中，目击者将所有其他个体一视同仁，忽略了受害者／施害者可能扮演不同角色这一点，也忽略了人们之间可能存在人际关系这一点。据我们所知，尚无研究探讨当参与者的角色发生变化时（例如，目击者转而成为施害者），目击者的行为会如何改变，这可能是因为情境本身也随之改变。另一方面，个体之间的关系已有相关研究；一项关于助人旁观情境的研究发现，在情况模糊时，若旁观者彼此认识，干预率较低；但在

情况明确时，熟人之间的干预率较高 [75]。换言之，情况不明时，两位朋友会抑制彼此的助人行为；情况明确时，则会相互支持对方助人。由于大多数旁观者研究涉及陌生人，我们的模型未能体现这一细节。

接下来，我们仅考虑行动的情绪成本/收益，而忽略行动的有形成本（例如金钱、时间和身体伤害）。在我们汇总的大多数旁观者研究中，有形成本都很低（例如回应友好手势 [56]或回复电子邮件 [26, 29]）。不过，有些研究涉及实际金钱支出（例如向救济基金捐款 [73] 或给服务员小费 [42]），也有些可能需要耗费大量时间（例如帮忙捡起一叠掉落的纸张 [14] 或寻找丢失的隐形眼镜[75]）。如果这些金钱或时间成本不受人群规模影响，那么我们可以将旁观者曲线（人群规模与干预概率之间的关系）向下平移一定比例，以纳入这些成本。如果成本取决于人群规模 N （例如由目击者分担一项耗时较长的任务），那么只要已知 N 与个体成本之间的关系，助人概率的降幅就可以随 N 而变化。金钱或时间成本固然是重要的考量因素，但只有在人群规模增加时，成本下降幅度远大于干预风险上升幅度的情况下，我们的结果才会发生定性变化；在这种情况下，旁观者效应可能会在达到一定人群规模后消失，甚至逆转。在某些紧急或危险的旁观者情境中，可以观察到这种效应（例如性侵 [30] 或言语/身体虐待 [31]）：旁观者干预比例稍有提高，就可能大幅降低每位旁观者面临的身体风险。我们可以在更一般的模型中纳入有形成本函数，但这一步留待未来研究。

另一项简化假设是，损失厌恶是静态的，且与社会学习率或初始行动适当性无关。然而，已知损失厌恶 [48, 49, 99]和易尴尬性 [100–102] 与文化、性别、种族/族裔、收入和阶层（以及其他人口统计特征）相关。更复杂的模型版本可以探讨模型参数之间相关性的影响，甚至允许参数随时间演变。

我们还忽略了行动适当性可能直接取决于人群规模的情形；例如，如果一名实习生向整个公司发送邮件寻求帮助，那么回复该邮件可能不太适当（例如，实习生打扰的人太多，不应得到帮助），也可能更适当（例如，回复实习生以纠正其行为）。由于这种取决于 N 的效应会因文化和情境而异，因此可能无法对其进行客观建模。

最后，我们的模型假设个体的人群规模是恒定的，但现实中，人们

会经历并从人群规模各异的旁观者情境中学习。由于在全连接网络和（静态）小世界网络中，稳态下学到的风险在性质上有所不同，我们预计，在动态社交网络结构上进行社会学习（即每个节点的度会随时间变化）也会产生不同的结果。我们将这一扩展留待未来研究。

五、结论

旁观者效应指的是，在社会情境中，随着人群规模扩大，助人行为会减少。这种效应并非普遍存在，往往出现在情境不危及生命时。在一些情况明确的助人情境中，责任分散可能是旁观者效应的一种作用机制；而在情境模糊、“正确”的做法并不明确时，社会影响和观众抑制则可能是更为可信的作用机制。

我们构建了两种旁观者情境模型：一种是基于前景理论的行动风险静态模型（1），用于刻画观众抑制；另一种是刻画社会学习的动态模型，用于体现社会影响（尽管其时间尺度比最初提出的更长）。静态模型呈现出涌现的旁观者效应，而动态模型则会加剧这种效应；该效应的特征（例如平均助人程度、助人行为下降的速度）取决于习得的行动适当性以及人群中损失厌恶的分布。尽管数据限制使我们无法在关于旁观者效应的三个主要假设中进行模型选择，但在大多数非危险情境下，数据与我们的模型相符。

通过这些模型，我们了解到，与数据一致的旁观者效应只有在某些情况下才会涌现：（1）人群中的损失厌恶具有较高的变异性；（2）干预风险的平均水平相对较低；（3）社会学习足够缓慢，以至于无法达到稳态。如果涌现的旁观者效应并不可取，该模型还提出了几种通过降低行动风险来减轻这一效应的潜在干预措施：提高行动的收益（例如，表扬采取行动的帮助者）、降低不当行动的成本（例如，以富有同理心的方式纠正不当行为，而不使其蒙羞），或直接降低行动风险（例如，通过标识或公益公告向公众讲授适当的行动方式）。

VI. 数据可用性

所有数据（.xlsx 文件）和软件（Matlab .m 文件及 Mathematica .nb 文件）都将通过 Harvard Dataverse 向公众公开。

VII. 致谢

作者感谢 Neer Bhardwaj、Nupoor Gandhi、AJ Ingram、Yiheng Li、Jonathan Morales 和 Dayo Ogunmodede 对模型早期探索所作的贡献。作者还感谢伊利诺伊几何实验室（现为伊利诺伊数学实验室）以及 Mathways 资助项目 No. DMS-1449269（SMC）提供研究支持。

---

[1] S. A. Nida, 《旁观者冷漠》, 08 2020年。[2] J. Darley 和 B. Latan´e, 《人们何时会在危机中提供帮助? 》, 《今日心理学》, 第 2 卷, 01 1968年。[3] B. Latan´e 和 J. M. Darley, 《紧急情况下旁观者干预的群体抑制》, 《人格与社会心理学杂志》, 第 10 卷, 第 3 期, 第 215–221 页, 1968年。[4] J. M. Darley 和 B. Latan´e, 《紧急情况下的旁观者干预：责任分散》, 《人格与社会心理学杂志》, 第 8 卷, 第 4 期, 第 377–383 页, 1968年。[5] B. Latan´e 和 J. M. Darley, 《旁观者“冷漠”》, 《美国科学家》, 第 57 卷, 第 2 期, 第 244–268 页, 1969年。[6] P. Fischer、J. I. Krueger、T. Greitemeyer、C. Vogrincic、A. Kastenm¨uller、D. Frey、M. Heene、M. Wicher 和 M. Kainbacher, 《旁观者效应：关于危险和非危险紧急情况下旁观者干预的元分析综述》, 《心理学公报》, 第 137 卷, 第 4 期, 第 517–537 页, 2011年。[7] E. Staub, 《陷入困境的儿童：年龄和目击者数量对儿童助人尝试的影响》, 《人格与社会心理学杂志》, 第 14 卷, 第 2 期, 第 130 页, 1970年。[8] M. Pl¨otner、H. Over、M. Carpenter 和 M. Tomasello, 《幼儿在助人情境中表现出旁观者效应》, 《心理科学》, 第 26 卷, 第 4 期, 第 499–506 页, 2015年。[9] S. Bauman、J. Yoon、C. Iurino 和 L. Hackett, 《青少年目击者经历同伴受害：旁观者效应》, 《学校心理学杂志》, 第 80 卷, 第 1–14 页, 2020年。[10] S. L. Gaertner、J. F. Dovidio 和 G. Johnson, 《受害者种族、无反应的旁观者与助人行为》, 《社会心理学杂志》, 第 117 卷, 第 1 期, 第 69–77 页, 1982年。[11] E. York Cornwell 和 A. Currit, 《美国街道医疗紧急情况下旁观者援助中的种族和社会差异》, 《美国公共卫生杂志》, 第 106 卷, 第 6 期, 第 1049–1051 页, 2016年。

[12] R. A. Garcia、J. A. Spertus、S. Girotra、B. K. Nallamothu、K. F. Kennedy、B. F. McNally、K. Breathett、M. Del Rios、C. Sasson 和 P. S. Chan，《目击心脏骤停时旁观者心肺复苏中的种族和族裔差异》，《新英格兰医学杂志》，第 387 卷，第 17 期，第 1569–1578 页，2022年。[13] B. Latan´e 和 J. M. Dabbs Jr，《三座城市中的性别、群体规模与助人》，《社会测量学》，第 180–194 页，1975年。[14] A. Cox 和 A. Adam，《非紧急情况下的旁观者效应：性别和群体规模的影响》，《现代心理学研究》，第 23 卷，第 2 期，第 3 页，2018年。[15] D. M. Tice 和 R. F. Baumeister，《男性气质抑制紧急情况下的助人：人格确实能够预测旁观者效应》，《人格与社会心理学杂志》，第 49 卷，第 2 期，第 420 页，1985年。[16] L. N. Jenkins 和 A. B. Nickerson，《欺凌参与者角色和性别对旁观者干预的预测作用》，《攻击性行为》，第 43 卷，第 3 期，第 281–290 页，2017年。[17] M. R. Merrens，《不同规模社区中的非紧急助人行为》，《社会心理学杂志》，第 90 卷，第 2 期，第 327–328 页，1973年。[18] D. R. Shaffer、M. Rogle 和 C. Hendrlck，《图书馆中的干预：责任增加对旁观者阻止盗窃意愿的影响》，《应用社会心理学杂志》，第 5 卷，第 4 期，第 303–319 页，1975年。[19] B. Latan´e 和 J. Rodin，《陷入困境的女士：朋友和陌生人对旁观者干预的抑制作用》，《实验社会心理学杂志》，第 5 卷，第 2 期，第 189–202 页，1969年。[20] S. Seo、T. H. Witte、D. M. Casper 和 S. Owen，《在风险情境中帮助朋友和陌生人：性侵和约会暴力旁观者干预的结果》，《攻击、虐待与创伤杂志》，第 31 卷，第 4 期，第 540–561 页，2022年。[21] S. Bennett 和 V. L. Banyard，《朋友真的会帮助朋友吗？关系因素和感知严重程度对性暴力旁观者认知的影响》，《暴力心理学》，第 6 卷，第 1 期，第 64 页，2016年。[22] B. Latan´e 和 S. Nida，《群体规模与助人研究十年》，《心理学公报》，第 89 卷，第 2 期，第 308 页，1981年。[23] J. I. Krueger 和 A. L. Massey，《对不当行为的理性重构》，《社会认知》，第 27 卷，第 5 期，第 786–812 页，2009年。[24] L. You 和 Y.-H. Lee，《社交网络网站网络欺凌中的旁观者效应：匿名性、群体规模和干预意图》，《远程信息处理与信息学》，第 45 卷，第 101284 页，2019年。[25] S. C. Voelpel、R. A. Eckhoff 和 J. F¨orster，《大卫对歌利亚？虚拟知识共享中的群体规模与旁观者效应》，《人际关系》，第 61 卷，第 2 期，第 271–295 页，2008年。[26] C. A. Blair、L. Foster Thompson 和 K. L. Wuensch，《电子助人行为：他人的虚拟存在会产生影响》，基础与应用社会心理学，第 27 卷，第 2 期，第 171–178 页，2005年。

[27] P. M. Markey，“计算机媒介通信中的旁观者干预”，《人类行为中的计算机》，第16卷第2期，第183–188页，2000年。[28] S. H. Schwartz 和 A. Gottlieb，“对暴力盗窃事件的旁观者反应：耶路撒冷的犯罪”，《人格与社会心理学杂志》，第34卷第6期，第1188–1199页，1976年。[29] G. Barron 和 E. Yechiam，“私人电子邮件请求与责任分散”，《人类行为中的计算机》，第18卷第5期，第507–520页，2002年。[30] H. Harari、O. Harari 和 R. V. White，“美国男性旁观者对强奸事件的反应”，《社会心理学杂志》，第125卷第5期，第653–658页，1985年。[31] P. Fischer、T. Greitemeyer、F. Pollozek 和 D. Frey，“无反应的旁观者：在危险紧急事件中，旁观者会更积极响应吗？”，《欧洲社会心理学杂志》，第36卷，第267–278页，2006年。[32] A. Franzen，“志愿者困境：理论模型与实证证据”，载于《解决社会困境》，第135–148页，Psychology Press，2013年。[33] R. D. Clark 和 L. E. Word，“冷漠的旁观者在哪里？紧急事件的情境特征。”，《人格与社会心理学杂志》，第29卷第3期，第279页，1974年。[34] L. Z. Solomon、H. Solomon 和 R. Stone，“助人行为与旁观者人数及紧急事件模糊性的关系”，《人格与社会心理学通报》，第4卷第2期，第318–321页，1978年。[35] L. M. Karakashian、M. I. Walter、A. N. Christopher 和 T. Lucas，“负面评价恐惧会影响助人行为：重新审视旁观者效应。”，《北美心理学杂志》，第8卷第1期，2006年。[36] E. Siligato、G. Iuele、M. Barbera、F. Bruno、G. Tordonato、A. Mautone 和 A. Rizzo，“冻结效应与旁观者效应：重叠与差异”，*Psych*，第6卷第1期，第273–287页，2024年。[37] B. Latan´e 和 J. Darley，“无反应的旁观者：为什么他不帮忙？；1970年”，纽约，*Appleton-Century-Croft*，1970年。[38] J. Piliavin、J. F. Dovidio、S. L. Gaertner 和 R. D. Clark III，“合作与助人行为”，*Academic Press*，纽约州纽约市，第279–304页，1982年。[39] J. Dovidio、J. Piliavin、S. Gaertner、D. Schroeder 和 R. III，“唤醒—成本—收益模型与干预过程：证据综述。”，《亲社会行为》，第12卷，1991年1月。[40] J. van den Heuvel 和 J. Treur，“帮还是不帮：旁观者效应的网络建模方法”，载于生物启发认知架构会议，第527–540页，Springer，2020年。[41] Y. Isada、N. Igaki 和 A. Shibata，“旁观者行为的经济模型”，《应用数学与物理学杂志》，第4卷第1期，第33–38页，2016年。[42] S. Freeman、M. R. Walker、R. Borden 和 B. Latan´e，“责任分散与餐厅小费：人多更省钱”，《人格与社会心理学通报》，第1卷第4期，第584–587页，1975年。

[43] A. V. Banerjee，“群体行为的简单模型”，《经济学季刊》，第107卷第3期，第797–817页，1992年。[44] R. H. Frank 和 E. Cartwright，《微观经济学与行为》，第8卷。麦格劳-希尔，纽约，2010年。[45] D. Kahneman 和 A. Tversky，“前景理论：风险下决策的分析”，《计量经济学》，第47卷第2期，第263–291页，1979年。[46] N. Gu´eguen、M. Dupr´e、P. Georget 和 C. S´en´emeaud，“承诺、犯罪与响应型旁观者：承诺形式和从众的影响”，《心理学、犯罪与法律》，第21卷第1期，第1–8页，2015年。[47] F. Kazerooni、S. H. Taylor、N. N. Bazarova 和 J. Whitlock，“网络欺凌中的旁观者干预：施害者人数和转发可预测帮助网络欺凌受害者的可能性”，《计算机中介传播学刊》，第23卷第3期，第146–162页，2018年。[48] M. Arora 和 S. Kumari，“金融决策中的风险承担与年龄、性别的关系：损失厌恶和后悔的中介作用”，《国际应用心理学杂志》，第5卷第4期，第83–89页，2015年。[49] D. Blake、E. Cannon 和 D. Wright，“量化损失厌恶：来自英国人口调查的证据”，《风险与不确定性杂志》，第63卷第1期，第27–57页，2021年。[50] D. Kahneman，《思考，快与慢》。麦克米伦，2011年。[51] R. Hortensius 和 B. De Gelder，“从同理心到冷漠：重新审视旁观者效应”，《心理科学最新进展》，第27卷第4期，第249–256页，2018年。[52] C. Yin、B. Li 和 T. Gao，“奖励与惩罚对基于强化的运动学习及其泛化的差异性影响”，《神经生理学杂志》，第130卷第5期，第1150–1161页，2023年。[53] T. W¨achter、O. V. Lungu、T. Liu、D. T. Willingham 和 J. Ashe，“奖励与惩罚对程序性学习的差异性影响”，《神经科学杂志》，第29卷第2期，第436–443页，2009年。[54] J. B. Sidowski、L. B. Wyckoff 和 L. Tabory，“强化与惩罚在最小社会情境中的影响。”，《异常与社会心理学杂志》，第52卷第1期，第115页，1956年。[55] F. Gregory Ashby 和 J. R. B. O’Brien，“正向反馈与负向反馈对信息整合类别学习的影响”，《知觉与心理物理学》，第69卷第6期，第865–878页，2007年。[56] D. L. Jones、J. D. Nelson 和 B. Opitz，“焦虑程度升高与从负向反馈中更好地学习相关”，《心理学学习与教学》，第20卷第1期，第76–90页，2021年。[57] E. M. B. d. Moraes、D. R. d. S. Carvalho、J. Sandars、T. M. Ozahata、R. Patel、D. Cecilio-Fernandes 和 T. M. Santos，“解析基于模拟的教育中观察与主动参与的差异”，《医学教师》，第47卷第6期，第991–996页，2025年。[58] J. Riley 和 K. Ward，“会计信息系统课程中的主动学习、合作主动学习和被动学习方法”，《会计教育问题》，第32卷第2期，第1–16页，2017年。

[59] A. Blani´e、S. Gorse、P. Roulleau、S. Figueiredo 和 D. Benhamou，“学习者角色（主动参与观察者或仅观察者）对麻醉学高仿真模拟课程学习成果的影响：一项单中心、前瞻性随机研究”，麻醉、重症监护与疼痛医学，第37卷，第5期，第417–422页，2018年。[60] P. P´eruch 和 P. N. Wilson，“在复杂的户外建成环境中主动学习与被动学习的比较”，《认知加工》，第5卷，第4期，第218–227页，2004年。[61] B. Eppinger 和 J. Kray，“选择还是回避：从正向和负向反馈中学习的年龄差异”，《认知神经科学杂志》，第23卷，第1期，第41–52页，2011年。[62] A. C. Van Duijvenvoorde、K. Zanolie、S. A. Rombouts、M. E. Raijmakers 和 E. A. Crone，“评估负面还是重视正面？支持跨发育阶段基于反馈学习的神经机制”，《神经科学杂志》，第28卷，第38期，第9495–9503页，2008年。[63] A. Petzold、F. Plessow、T. Goschke 和 C. Kirschbaum，“压力会降低在基于反馈的学习任务中对负向反馈的利用。”，《行为神经科学》，第124卷，第2期，第248页，2010年。[64] M. H. Reime、T. Johnsgaard、F. I. Kvam、M. Aarflot、J. M. Engeberg、M. Breivik 和 G. Brattebø，“观察学习与实践学习：跨专业模拟培训中观察者与参与者体验的比较研究”，《跨专业护理杂志》，第31卷，第1期，第51–58页，2017年。[65] P. Haidet、R. O. Morgan、K. O’malley、B. J. Moran 和 B. F. Richards，“大型团体环境中主动与被动学习策略的对照试验”，《健康科学教育进展》，第9卷，第1期，第15–27页，2004年。[66] D. K. Chalmers 和 M. E. Rosenbaum，“观察学习与实践学习。”，《教育心理学杂志》，第66卷，第2期，第216页，1974年。[67] R. S. Miller，“共情性尴尬：对他人尴尬产生反应的情境与个人决定因素。”，《人格与社会心理学杂志》，第53卷，第6期，第1061页，1987年。[68] S. T. Hawk、A. H. Fischer 和 G. A. Van Kleef，“设身处地还是感同身受：产生共情性尴尬的两种途径。”，《情绪》，第11卷，第3期，第502页，2011年。[69] M. Bernardo、C. Budd、A. R. Champneys 和 P. Kowalczyk，分段光滑动力系统：理论与应用，第163卷。施普林格科学与商业媒体，2008年。[70] L. M. Trick 和 Z. W. Pylyshyn，“为什么小数目和大数目采用不同的计数方式？视觉中的有限容量前注意阶段。”，《心理学评论》，第101卷，第1期，第80页，1994年。[71] S. Marino、I. B. Hogue、C. J. Ray 和 D. E. Kirschner，“在系统生物学中进行全局不确定性分析和敏感性分析的方法”，《理论生物学杂志》，第254卷，第1期，第178–196页，2008年。[72] R. E. Petty、K. D. Williams、S. G. Harkins 和 B. Latan´e，“抑制自我助人行为的社会因素：旁观者对芝士汉堡的反应”，《人格与社会心理学通报》，第3卷，第4期，第575–578页，1977年。

[73] D. L. Wiesenthal、D. Austrom 和 I. Silverman，“慈善捐赠中的责任分散”，基础与应用社会心理学，第4卷，第1期，第17–27页，1983年。[74] S. Samia 和 M. Khanam，“旁观者人数、时间压力和性别对利他行为的影响”，达卡大学生物科学学报，第27卷，第1期，第1–7页，2018年。[75] J. Harada，“旁观者干预：助人情境的模糊性及旁观者之间人际关系的影响”，日本心理学研究，第27卷，第4期，第177–184页，1985年。[76] A. S. Ross，“责任增加对旁观者干预的影响：儿童在场。”，《人格与社会心理学杂志》，第19卷，第3期，第306页，1971年。[77] A. S. Ross 和 J. Braband，“责任增加对旁观者干预的影响：II. 盲人的线索价值。”，《人格与社会心理学杂志》，第25卷，第2期，第254页，1973年。[78] E. Yechiam 和 G. Barron，“学习忽略在线求助请求”，计算与数学组织理论，第9卷，第4期，第327–339页，2003年。[79] S. L. Gaertner，“种族态度在助人行为中的作用”，《社会心理学杂志》，第97卷，第1期，第95–101页，1975年。[80] S. L. Gaertner 和 J. F. Dovidio，“白人种族主义的隐蔽性、唤醒与助人行为。”，《人格与社会心理学杂志》，第35卷，第10期，第691页，1977年。[81] R. D. Clark 和 L. E. Word，“旁观者为什么不帮忙？是因为模糊性吗？”，《人格与社会心理学杂志》，第24卷，第3期，第392页，1972年。[82] J. M. Darley、A. I. Teger 和 L. D. Lewis，“群体是否总会抑制个体对潜在紧急情况的反应？”，《人格与社会心理学杂志》，第26卷，第3期，第395页，1973年。[83] V. J. Koneˇcni 和 E. B. Ebbesen，“儿童在场对成人助人行为和顺从的影响：两项现场研究”，《社会心理学杂志》，第97卷，第2期，第181–193页，1975年。[84] A. I. Teger 等，“考察紧急情况下旁观者干预的社会影响假说。”，1971年。[85] V. A. Harris 和 C. E. Robinson，“旁观者干预：群体规模与受害者身份”，心理学学会通报，第2卷，第1期，第8–10页，1973年。[86] I. A. Horowitz，“群体规范对旁观者干预的影响”，《社会心理学杂志》，第83卷，第2期，第265–273页，1971年。[87] S. H. Schwartz 和 G. T. Clausen，“紧急事件中的责任、规范与助人行为。”，《人格与社会心理学杂志》，第16卷，第2期，第299页，1970年。[88] R. E. Smith、L. Smythe 和 D. Lien，“相似或不同的无反应同伴旁观者对助人行为的抑制。”，1972年。[89] E. Staub，“帮助陷入困境的人：社会、人格与刺激决定因素”，载于《实验社会心理学进展》，第7卷，第293–341页，爱思唯尔，1974年。[90] L. Z. Solomon、H. Solomon 和 R. Stone，“助人行为与旁观者人数及紧急事件模糊性的关系”，《人格与社会心理学通报》，第4卷，第2期，第318–321页，1978年。

[91] J. P. Wilson，“动机、建模与利他主义：一种个体—情境分析”，《人格与社会心理学杂志》，第34卷，第6期，第1078页，1976年。[92] W. Howard 和 W. D. Crano，“性别、交谈、地点和观察者群体规模对高风险情境中旁观者干预的影响”，《社会测量学》，第491–507页，1974年。[93] R. D. Lytle、T. M. Bratton 和 H. K. Hudson，“社交媒体时代的旁观者冷漠与干预”，载于《翡翠国际技术促成暴力与虐待手册》，第711–728页，翡翠出版有限公司，2021年。[94] P. M. Zoccola、M. C. Green、E. Karoutsos、S. M. Katona 和 J. Sabini，“尴尬的旁观者：易尴尬性与助人抑制”，《人格与个体差异》，第51卷，第8期，第925–929页，2011年。[95] A. Nicolle、S. M. Fleming、D. R. Bach、J. Driver 和 R. J. Dolan，“后悔诱发的现状偏差”，《神经科学杂志》，第31卷，第9期，第3320–3327页，2011年。[96] Q. Han、S. Quadflieg 和 C. J. Ludwig，“决策回避与决策后悔：一项系统综述与荟萃分析”，*PLOS ONE*，第18卷，第10期，第e0292857页，2023年。[97] B. Latan´e 和 S. Nida，“群体规模与助人行为的十年研究”，《心理学公报》，第89卷，第2期，第308–324页，1981年。[98] M. W. Khaw、L. Stevens 和 M. Woodford，“概率感知中的个体差异”，*PLOS*计算生物学，第17卷，第4期，第e1008871页，2021年。[99] C. Dawson，“乐观主义、损失厌恶和风险态度方面的性别差异”，《英国心理学杂志》，第114卷，第4期，第928–944页，2023年。[100] K. M. Kelly 和 W. H. Jones，“特质性易尴尬性的评估”，《焦虑、压力与应对》，第10卷，第4期，第307–333页，1997年。[101] T. M. Singelis 和 W. F. Sharkey，“文化、自我建构与易尴尬性”，《跨文化心理学杂志》，第26卷，第6期，第622–644页，1995年。[102] S. G. Hofmann、M. Anu Asnaani 和 D. E. Hinton，“社交焦虑与社交焦虑障碍的文化层面”，《抑郁与焦虑》，第27卷，第12期，第1117–1127页，2010年。[103] D. J. Watts 和 S. H. Strogatz，“‘小世界’网络的集体动力学”，《自然》，第393卷，第6684期，第440–442页，1998年。[104] D. Kirschner，“不确定性与敏感性函数及其实现”，2008年。[105] M. McKay、R. Beckman 和 W. Conover，“比较三种方法，以选择计算机代码输出分析中的输入变量取值”，《技术计量学》，第21卷，第2期，第239–245页，1979年。[106] A. Saltelli，“充分利用模型评估来计算敏感性指数”，《计算机物理通讯》，第145卷，第2期，第280–297页，2002年。

S1. 补充材料

A. 概率推导

我们指出，在个体采取了行动的前提下，该个体的行动受到欢迎的概率为

$$P_i^{(w)} = \prod_{j \neq i} \left(P_j^{(I)} + \left(1 - P_j^{(I)}\right)(1 - r_j) \right)$$

我们以四名旁观者为例，说明这一概率的推导过程，并聚焦于一名已经决定进行干预的旁观者（称其为“旁观者4”）。已知旁观者4已经采取行动后，我们绘出其他三名旁观者所有可能的结果（图 S1）。请注意，为避免符号繁杂，我们将省略所有干预概率的上标 (*I*)。

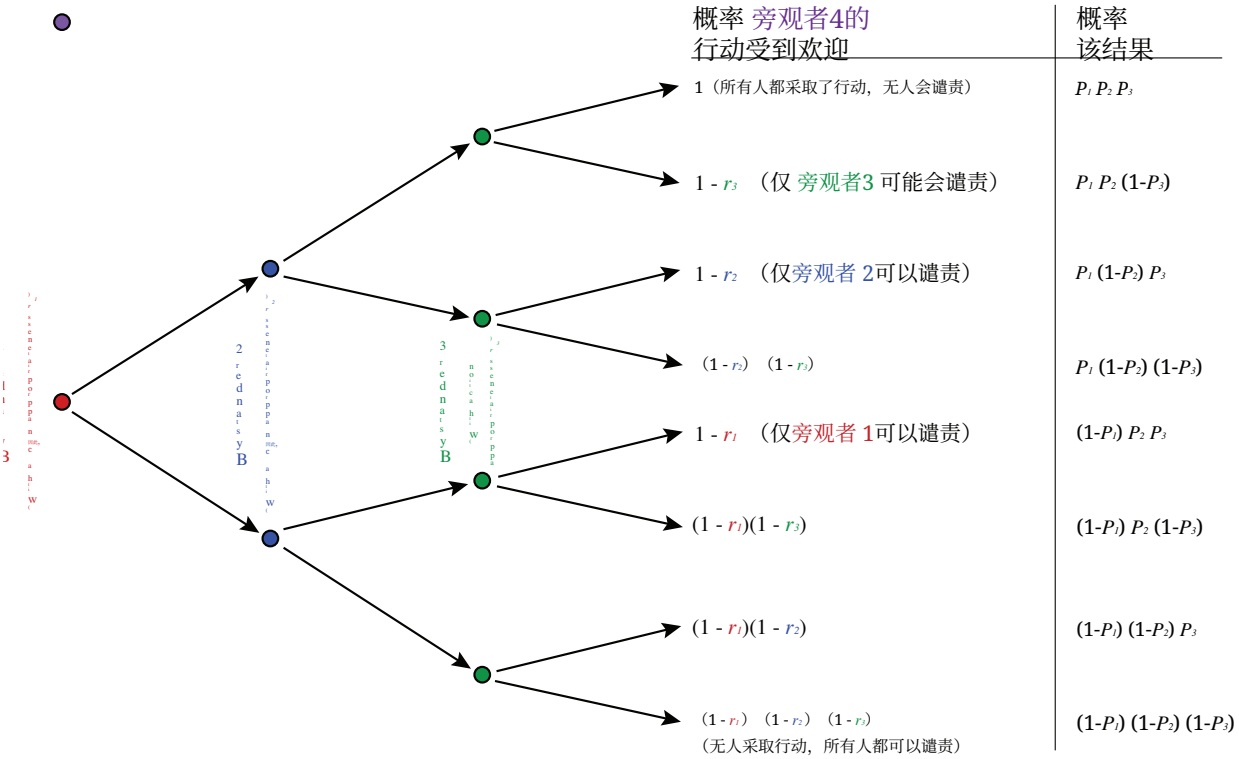

图 S1. 四名旁观者的概率树图示例，其中一人已决定采取行动。假设采取行动的所有概率 P_i 彼此独立，且取决于各自的行动风险（r_i）。

在每一种可能的分支中，旁观者4的干预受到欢迎的概率，等于所有未采取行动的旁观者所承担的风险的乘积；如果所有旁观者都采取了行动（顶部的分支），那么就没有人会谴责这一行动，因此该行动受到欢迎的概率为1。

任何特定分支最终成为结果的概率，等于所有行动或不作为概率的乘积。因此，一项行动受到欢迎的总概率是所有条件概率之和：

$$
\begin{aligned}
P_4^{(w)} =& 1 \cdot P_1P_2P_3 + (1-r_3) \cdot P_1P_2(1-P_3) \\
&+ (1-r_2) \cdot P_1(1-P_2)P_3 + (1-r_2)(1-r_3) \cdot P_1(1-P_2)(1-P_3) \\
&+ (1-r_1) \cdot (1-P_1)P_2P_3 + (1-r_1)(1-r_3) \cdot (1-P_1)P_2(1-P_3) \\
&+ (1-r_1)(1-r_2) \cdot (1-P_1)(1-P_2)P_3 + (1-r_1)(1-r_2)(1-r_3) \cdot (1-P_1)(1-P_2)(1-P_3)
\end{aligned}
$$

我们尽可能提取出 P_1 和 $(1-P_1)$，并合并同类项：

$$
\begin{aligned}
P_4^{(w)} =& P_1\Big[P_2P_3 + (1-r_3) \cdot P_2(1-P_3) + (1-r_2) \cdot (1-P_2)P_3 + (1-r_2)(1-r_3) \cdot (1-P_2)(1-P_3)\Big] \\
&+ (1-P_1)\Big[(1-r_1) \cdot P_2P_3 + (1-r_1)(1-r_3) \cdot P_2(1-P_3) \\
&\qquad + (1-r_1)(1-r_2) \cdot (1-P_2)P_3 + (1-r_1)(1-r_2)(1-r_3) \cdot (1-P_2)(1-P_3)\Big]
\end{aligned}
$$

递归地，我们继续这一过程（先提取涉及 P_2 的项，再提取涉及P_3的项，直到得到最后一个目击者概率），从而得到以下结构：

$$
\begin{aligned}
P_4^{(w)} =& P_1\Big[P_2\Big[P_3 + (1-r_3) \cdot (1-P_3)\Big] + (1-P_2)\Big[(1-r_2) \cdot P_3 + (1-r_2)(1-r_3) \cdot (1-P_3)\Big]\Big] \\
&+ (1-P_1)\Big[P_2\Big[(1-r_1) \cdot P_3 + (1-r_1)(1-r_3) \cdot (1-P_3)\Big] \\
&\qquad + (1-P_2)\Big[(1-r_1)(1-r_2) \cdot P_3 + (1-r_1)(1-r_2)(1-r_3) \cdot (1-P_3)\Big]\Big]
\end{aligned}
$$

然后，我们尽可能提取出所有风险项（$1-r_i$）：

$$
\begin{aligned}
P_4^{(w)} =& P_1\Big[P_2\Big[P_3 + (1-r_3)(1-P_3)\Big] + (1-r_2)(1-P_2)\Big[P_3 + (1-r_3)(1-P_3)\Big]\Big] \\
&+ (1-r_1)(1-P_1)\Big[P_2\Big[P_3 + (1-r_3)(1-P_3)\Big] + (1-r_2)(1-P_2)\Big[P_3 + (1-r_3)(1-P_3)\Big]\Big]
\end{aligned}
$$

最后，递归地提取出公因子 $P_3 + (1-r_3)(1-P_3)$，然后提取出公因子$P_2 +(1-r_2)(1-P_2)$：

$$
P_4^{(w)} = \Big[P_1 + (1-r_1)(1-P_1)\Big]\Big[P_2 + (1-r_2)(1-P_2)\Big]\Big[P_3 + (1-r_3)(1-P_3)\Big],
$$

这就推广得到了我们推导出的公式。

我们还声称，在个体 i 未进行干预的情况下，另一名目击者的行动受到良好反响的概率为

$$
\tilde{P}_i^{(w)} = \prod_{j \neq i}\Big(P_j^{(I)} + \big(1-P_j^{(I)}\big)(1-r_j)\Big) - \prod_{j \neq i}\big(1-P_j^{(I)}\big)(1-r_j)
$$

第一个乘积的推导与上述相同。第二个乘积对无人采取行动的结果进行修正，因此个体 i 无法从这一情境中学习。第一次推导中未考虑这一结果，因为当时假设个体 i 已经采取了行动。

B. 小世界网络中的社会学习

为检验第 IIE 节中全连接网络结果的稳健性，我们研究了小世界网络中的社会学习：该网络包含 $n = 1000$ 个体，这些个体最初分别与 $k = 4$ 个邻居相连，重连概率为 $\beta = 1$，以实现通常涵盖 $N = 2$ 或 3 至 $N = 11$ 或 12 的度分布。网络中的每个个体（节点）都遵循公式（2）的动态过程，仅从其邻居处学习；这会导致人群中群体规模的分布 N（度为 $+1$）（见图 S2）。

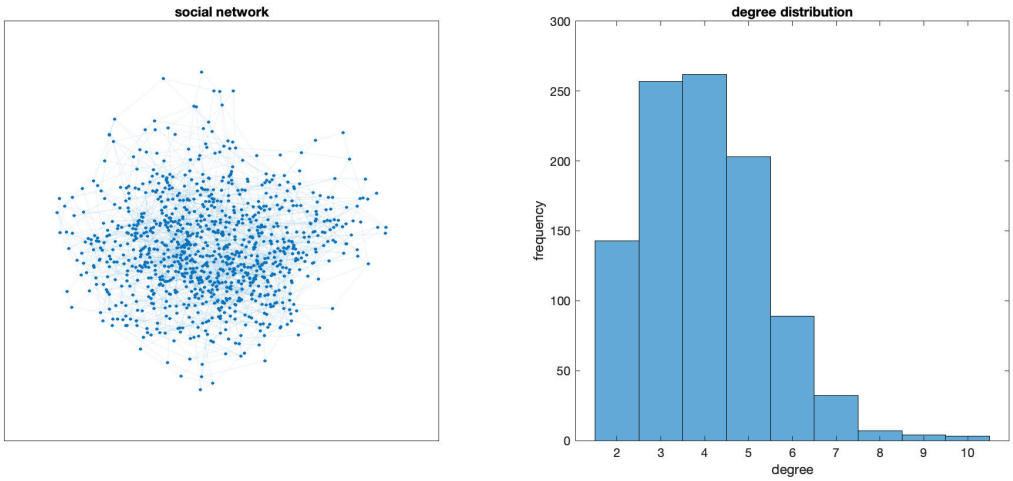

图 S2. 包含 1000 个节点的小世界网络，这些节点最初分别与 $k = 4$ 个邻居相连，重连概率为 $\beta = 1$（左）。重连后，度分布（右）大致涵盖 2 至 10（N 涵盖 3 至 11）。

在典型模拟中，每个节点的风险水平都会趋于稳态，就像在更简单的全连接网络中一样。不同于全连接网络，从度数不同的邻居处学习会导致显著不同的集体行为，尤其是在群体规模较小时（见图 S3）。

在人群规模较小时，大多数个体会趋于较低的风险水平，但也有一些个体会趋于最高风险水平 1；相比之下，在人群规模较小的全连接网络中，大多数个体会趋于最低风险水平 0，而只有少数个体（如果有的话）会趋于较低的正风险水平。

[7] 使用无标度网络时，结果在定性上相似。

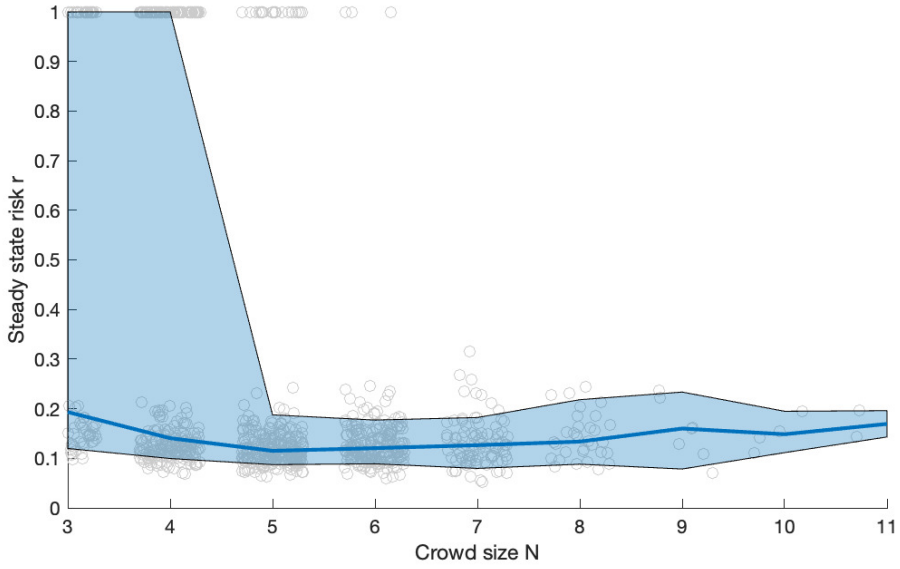

图 S3. 在表 I 中所有其他参数均保持基准值不变时，小世界网络中的稳态习得风险随目击者数量 N 的变化。灰点表示整个由 1000 个节点组成的网络中各个目击者的稳态风险水平 r_i，并在水平方向上添加了抖动以便于观察。每个 N 的稳态风险中位数（蓝线）基本保持不变，但方差会随着 N 减小（蓝色阴影表示每个 N 的模拟结果中第 10% 至第 90% 分位数）。

对于更大的人群规模，全连接网络和小世界网络中的集体行为在定性上相似。在这两种情况下，个体的风险水平都会收敛到较小的正值；然而，小世界网络的稳态风险中位数高于全连接网络。

这些结果表明，全连接网络收敛至零风险的结果对于更贴近现实的网络结构并不稳健。在这种网络中，许多身处社交网络、且所处人群规模各异的个体，会习得较低的行动风险。然而，个体所在的人群规模越小，他们最终越可能习得干预风险过高、根本不值得尝试（这些个体是小型-N 节点中的少数）。

尽管如此，固定的小世界网络结构仍不如现实世界中动态重连的社交网络真实；在现实世界中，个体在日常生活中会与数量随机的陌生人互动。

C. 连续谱上的社会学习

如果我们假设目击者数量 N 是连续统总体中的一个参数，就会得到一个关于总体风险分布 r 的偏微分方程。由于概率

是守恒的 d，人群中的风险分布 r 遵循连续性方程 :

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial r}\left(p\frac{\mathrm{d}r}{\mathrm{d}t}\right),$$

其中， $p(r,\ t)$ 是 r 的概率密度函数，$\mathrm{d}r\ \mathrm{d}t$ 由不含索引 i 的公式（2）给出，并且在连续极限下，概率变为

$$\begin{aligned}
P^{(I)} &= H\Big(1-(1+x)\big(1-(1-r)^N\big)\Big)\\
P^{(w)} &= (1-r)^{N-1}\\
\tilde{P}^{(w)} &= \Big(P^{(I)}+\big(1-P^{(I)}\big)(1-r)\Big)^{N-1} - \Big(\big(1-P^{(I)}\big)(1-r)\Big)^{N-1}\\
\tilde{P}^{(p)} &= 1-\tilde{P}^{(w)} - \big(1-P^{(I)}\big)^{N-1}
\end{aligned}$$

我们将这一探索留待未来工作。

D. 敏感性分析：技术细节

全局不确定性和敏感性分析采用 Marino 等人 [71] 概述的方法进行。分析所用的基础代码可从作者 Denise Kirschner 的网站 [104]免费获取。

简而言之，该分析采用参数空间的拉丁超立方抽样（拉丁超立方抽样），以模拟模型参数的不确定性。与简单随机抽样相比，拉丁超立方抽样所需的模型模拟次数更少，且不会引入偏差 [105]。我们对表 I 中给出的基准值附近的每个参数进行均匀抽样。均匀抽样的范围可在我们的代码中找到。

由于模型输出对模型输入的依赖关系是单调的，但并非线性趋势，因此我们使用偏秩相关系数（PRCCs）来检验模型输出对参数不确定性的敏感性。正如 Marino 等人所指出的， [71], 对于线性趋势，我们可以使用皮尔逊相关系数（CC）、偏相关系数（PCCs）或标准化回归系数（SRC）。如果我们的趋势是非单调的，我们则会使用 Sobol 方法或其众多扩展方法之一 [106]。

图 8 显示的敏感性是模型输出（稳态风险中位数 r）与模型输入（所有模型参数，包括初始条件，但不包括人群规模 N）之间的偏秩相关系数；我们必须针对每个 N 分别进行敏感性分析，因为常微分方程组的阶数会随 N变化。

正如 Marino 等人 [71], 所述，偏秩相关刻画的是移除其他输入对 y 的影响后，输入 x_j 与输出 y 之间的单调关系。偏秩相关系数的取值介于 -1 和 1 之间，其中 1 表示最强的正秩相关，而 -1 表示最强的负秩相关。显著性表示秩相关为零的概率（即，显著性越高，表明 x_j 与 y之间不存在关系）。

表 I. 社会学习模型 (2) 中模型参数的描述。

参数	含义	范围	基线来源	
N	可能需要帮助的个体的目击者数量	[1, 20]	变量	—
a	帮助[†] 从自己反响不佳的 感知到的行为	0 或 1	1[‡]	—
b	从自己反响良好的行为中学习时的风险更新率 行为	[0, 100]	0.01[§]	[52–56]
c	从他人反响不佳的行为中学习时的风险更新率 行为	[0, 100]	0.1[⋆]	[57–60]
d	从他人反响良好的行为中学习时的风险更新率 行为	[0, 100]	0.001[§]	[52–56]
x_i	从伽马分布中抽取的个体 i的损失厌恶比率 分布 $x_i \sim \Gamma\left(\alpha = \left(\frac{\mu}{\sigma}\right)^2, \beta = \frac{\sigma^2}{\mu}\right)$	[0,100]	变量	[45]
μ	损失厌恶比率分布的均值	[0, 100]	2	[48–50]
σ	损失厌恶比率分布的标准差	[0, 100]	0.5	[48–50]
$r_{\max}$	最大初始风险水平，其中初始风险取样自 均匀分布 $r_i(0) \sim \mathcal{U}(0, r_{\max})$	[0,1]	0.1	—

[†] 在我们的模型中，我们假设只有一个能够谴责干预的个体（例如，不是动物或婴儿）。如果旁观者情境中没有受害者（例如，烟雾正渗入房间），则从 N 中减去 1。如果情境中有不止一名非目击者（例如，发生欺凌或家庭争吵），则将额外的非目击者计入 N。

[‡] 不失一般性，可将速率设为 1，从而确定模型的时间尺度[§] 关于正向反馈与负向反馈下学习率的研究结论并不一致。在包含身体奖励和惩罚的身体任务中，许多研究表明，负强化下的学习速度更快（但所学内容可能较难推广） [52–54]。在涉及言语反馈的研究中，人们也普遍发现，相比于对事实性问题给出正确回答后获得的反馈，人们从针对错误回答的反馈中学习得更快 [55], ，但这一效应的强度可能取决于年龄和性别 [61, 62]。在考察反馈所激发情绪的研究中，焦虑是最常测量的情绪；一些研究表明，焦虑会增强从负向反馈中学习的效果 [56], ，另一些研究则得出相反结论 [63]。据我们所知，目前没有研究量化我们试图理解的这类社会情境中，正向反馈与负向反馈所对应的相对学习速率。因此，我们推测正向反馈后的学习速度慢 100 倍（尤其是因为正向反馈可能仅仅意味着没有负向反馈），并在 II F 节中检验模型结果对这一设定的敏感性。⋆ 关于“亲身实践”与“观察”所对应学习速率的研究尚无定论。许多研究表明，主动学习后表现（尤其是在身体任务中的表现）更好 [57–60], ，但另一些研究发现，在主动与被动学习环境中，表现并无显著差异（偶尔也会出现相反的效果）[64–66]。据我们所知，目前没有研究测量学习率（只测量最终表现），也没有研究测量我们试图理解的这类社会情境中，个人尴尬与共情性尴尬所对应的相对学习速率 [67, 68]。因此，我们推测，从观察到的行为中学习的速度比从实施的行为中学习慢 10 倍，并在 II F 节中检验模型结果对这一设定的敏感性。