arXiv:2611.28484v1 [stat.OT] 21 Nov 2026

助推中的助推：当场域反弹时

Giuseppe Alessandro Veltri

行为与实现科学干预（BISI），杨潞龄医学院，新加坡国立大学，新加坡 社会学与社会研究系，特伦托大学，特伦托，意大利 gaveltri@nus.edu.sg · ORCID 0000-0002-9472-2236

预印本，2026年11月

摘要

行为干预的设计与测试仿佛其所进入的设置会保持不变。然而，其他参与者会对同一批人施加影响，观察行为，并在行为发生变化时作出响应。本文探讨在这样的场域中，助推会变成什么样。在基于主体的模型中，当个人的倾向与其感受到的推动共同跨过阈值时，个人便会采取行动；助推会争夺有限注意力；其他参与者则在看不到助推的情况下，将行为的比率引向自身目标。模型根据已发表的效应量进行校准，复现了四项有记录的规律，其中一项通过校准得到。相同的助推会因其他参与者的行动方式而走向五种结局：衰减、瞬态、空值、反向，或盲站点——在这种情况下，一项可靠的现场实验报告出几乎完整的效应，而其站点却几乎没有获得增益。当场域作出响应时，试点自身的比较并不能预测推广结果：它看不到试点所引发的响应，而且小规模试点不会像推广那样耗尽竞争者。结局由少数几个量决定，校准后没有再进行拟合；推导结果在十张设置图谱中的八张上与模拟结果相符。在49,152 个随机设定中，每个设定都包含一个竞争者，预期效应反而是例外；结局受何种因素影响最大，取决于注意力如何分配：当最强助推先被听见时，关键在于助推的强度；当注意力被分割时，关键在于竞争助推的数量。一次真实的场域是否会作出响应，可以从试验能够记录的痕迹中读出，例如未处理组之间的漂移。

意义声明

诸如助推之类的行为干预往往在试点中奏效，却在大规模推广时令人失望。通常的解释是实现不佳和选择性发表。还有另一种可能：环境会反作用。其他参与者也有自己的目标，他们会观察行为，并在行为发生变化时作出响应。在模型中，试点自身的比较无法捕捉这种响应，因此无法预测推广后的情况；一项运行良好的试验可能报告几乎全部效应，但其所在站点获得的增益却微乎其微。同一种助推可能被削弱、消除或反转，具体取决于还有哪些人在采取行动。每种结果都会留下某种可供试验记录的痕迹，例如未处理组中的偏移。这表明，评估应记录其他参与者，而不应只记录干预。

引言

行为干预（包括助推） [1], 被用于改变人们的锻炼方式 [2]、投票行为 [3], 、能源使用方式 [4] 以及应对气候变化的行动 [5], ，政府也设立了相关部门来设计和测试这些干预 [6]。有四项发现与这些干预的设计方式格格不入。在

试点中的效应在干预扩大规模后会减弱 [7]。美国两个政府助推部门开展的试验报告称，其平均效应约为已发表学术试验中效应大小的六分之一 [6]。荟萃分析发现，不同研究的效应差异很大；考虑发表偏倚后，效应会减弱，其中一项分析甚至发现效应全无 [8–10]。此外，同一项干预在不同站点产生的效应差异，超出了随机因素所能解释的范围 [11]。

每项发现都有一种标准解读。规模损失被归因于实现 [7]；政府试验与学术试验之间的差距，以及较小的平均值，被归因于统计功效不足和选择性发表 [6, 9]；不同地点之间的差异，则被归因于未测量的调节变量 [11], ，或被归因于早期地点之所以被选中，是因为它们反应良好 [4]。这些解读共享一个假设。干预所进入的设置只是一个背景：它可能在试点与推广之间有所不同，也可能因地点而异，但不会对干预所产生的作用施加影响。人们知道存在例外。企业会撤销它们反对的默认选项 [12], ；反助推会回应助推 [13], ；市场反应可能逆转助推所预测的效应 [14], ；随着项目扩大，价格也会发生变化 [15, 16]。这类响应大多被视为规模化的威胁 [7] ，很少进一步追踪其对试验中处理组和未处理组的影响，但也存在例外 [17, 18]。

行为受到周围力量维持这一观点，比助推的出现更早。Lewin 将行为描述为驱动力与约束力之间的均衡，其中增加的压力往往会引发更大的反压力 [19]。沿用 Lewin 的说法，作用于一群人的力量，以及这些力量背后的参与者，在此被称为场域。这个观点以多种形式重新出现。一个广告活动在在线广告中的可测量提升，取决于竞争者在同一时间测试了什么 [20]。同时推出的政策会相互影响彼此的评估 [21]。在这些版本中，力量是给定的：它们不会对干预如何影响行为作出响应。

响应的场域很难用现有数据加以研究。很少有试验会记录作用于其招募对象的其他力量。对开放式资料库进行的三轮筛查，共考察了约 120 个候选数据集，却没有发现任何能够检验响应场域特有性质的数据集（补充信息附录，S1节）。

本文构建了一个模型，将一种行为干预置于一群会观察其所针对行为并作出响应的参与者之中，并提出一个问题：当场域作出反应时，助推会变成什么样？其他参与者包括竞争者、盟友和旁观者；旁观者推动另一种行为，只争夺注意力。每个参与者都追求某种行为达到特定比率的目标；他们观察该比率，却看不到干预。主要结论是，助推的结局由它进入的场域决定，而检验它的试验可能无法看出这一结局。场域会作出反应时，试点会引发与其规模成比例的响应，而试点自身的对照无法观察到这种响应。与推广不同，小规模试点也不会把其他参与者推至极限，因此无法揭示推广的结局。现场实验可能报告几乎完整的预期效应，而其站点却几乎没有增益；本文将这种结局称为盲站点。视场域而定，同一种助推也可能被衰减、变成瞬态、归零或反向。这些结局可由少数几个量通过简短推导得出，无需拟合；在十张模拟图中的八张上，推导结果与模拟一致。这些模拟图是设置的网格，其中场域的两个属性有所变化。在随机设置中，最能改变结局的因素取决于注意力如何分配：若最强的助推最先被听到，关键因素就是助推本身的强度；若注意力按比例分配，关键因素就是旁观者的数量。每种结局都会留下试验能够记录的痕迹。模型根据已发表的效应量进行校准，未拟合任何数据。它展示了会作出反应的场域可能产生什么影响，而非真实场域作出反应的频率。

结果

一个会做出反应的场域模型

模型将一种行为干预置于其他会对所见信息作出反应的参与者之中，此处称其为新助推（图 1）。每个个体都对某种

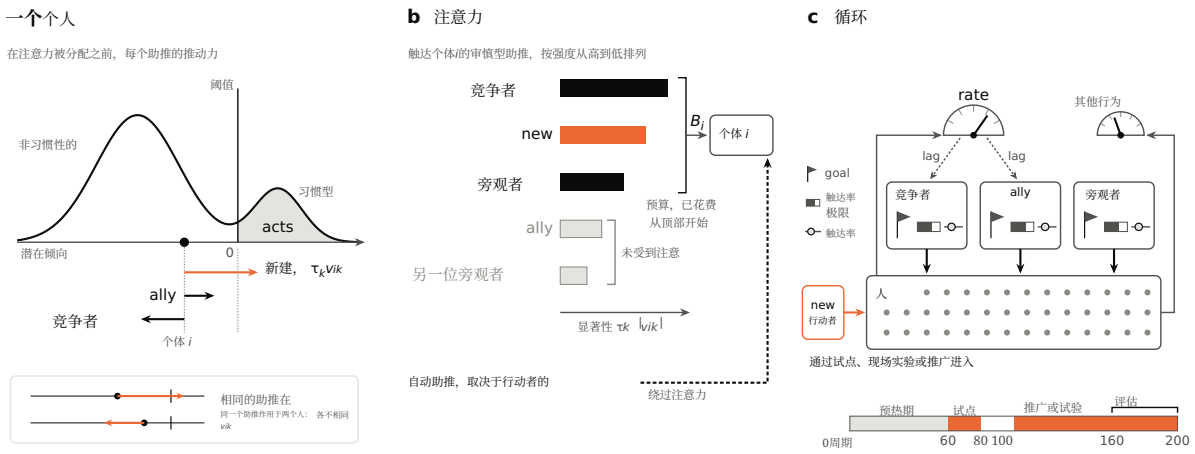


图1：助推是在阈值上施加的一种力量；它与其他助推共享个人的注意力；拥有这些助推的参与者则根据自身目标和触达上限调节行为比率，并对该比率而非新助推作出响应。 **a**，个体。曲线表示总体的基线倾向概率（非习惯性人群和习惯性人群）；在0处阈值之外的人（阴影部分）会采取行动。个体 i 是这个点。箭头表示作用于个体 i的三种助推：新助推、盟友的助推和竞争者的助推；它们各自施力，$\tau_k v_{ik}$，在注意力分配之前不会相加。插图：同一种助推作用于两个人，对其中一人持续时间较长，对另一人则发生翻转（v_{ik} 变化）。**b**，注意力。个人 i的审慎型助推（即需要注意力的助推）按显著性 $\tau_k|v_{ik}|$排序，从上至下消耗注意力预算 B_i（优先级）。超出预算的助推（浅色）不会被注意到：此处旁观者的助推挤掉了盟友的助推。自动助推无论属于哪个参与者（虚线），都能绕过预算。**c**，循环。站点中的人群决定比率（刻度盘）。竞争者、盟友和旁观者各自有一个目标（旗帜）、一个触达上限（横条）和一个触达率（滑块），它们观察存在滞后的比率（虚线），并向人群发出助推（实线）。旁观者推广不同的行为（第二个刻度盘），保持固定触达率并争夺注意力；没有任何信息反馈给它。没有箭头从新行动者（即助推的所有者）指向其他行动者。时间线以周期为单位，标出预热期、试点（60–79期）、从第100期开始的推广或试验，以及评估窗口（160–199期）；新助推在第80至99期关闭。未绘出的响应规则：恒温器将比率调节至目标；最大化者则在比率高于其基准值时持续施力。长度、密度和刻度盘仅作示意。

行为具有一种稳定倾向；当这种倾向加上触达他们的助推所施加的推动力以及一定的随机因素超过阈值时，个体便会采取行为（图 1a）。每个个体还拥有有限的注意力预算（图 1b）。需要注意力的助推（包括新助推）会争夺这一预算。个体首先关注自己感受最强烈的助推；这是默认规则，称为优先级共享。在另一种规则，即比例共享下，触达个体的每个助推都会获得相等的预算份额（方法）。自动助推（例如除非个体主动选择退出，否则便会生效的预设选项）无需注意力即可发挥作用。

一个站点有三类其他参与者（图 1c）。竞争者会推动相反方向的同一种行为，盟友会朝着与新助推相同的方向推动该行为，而旁观者会促进另一种行为，并且只通过注意力影响这种行为。每个竞争者或盟友都有一个目标，并会选择其助推的触达人数，上限为某个极限；旁观者的触达率则固定不变。竞争者或盟友会延迟观察其所推动行为的比率，即每个周期内采取该行为的人口比例；一个周期按一周计算。它们看不到新助推。竞争者要么进行调节，不断调整其助推的触达人数，直到比率达到目标；要么采取棘轮式推进，在比率仍高于其见过的最低水平时，持续扩大触达人数。有些竞争者会升级：当比率连续十个周期偏离目标且它们已达到极限时，就会启动并持续运行一项自动助推（方法部分给出了各类竞争者所遵循的规则）。场域由站点中的其他参与者及其向站点内的人发送的助推构成。

一个试点t 对一小部分随机抽取的人进行处理。现场实验则对场域中 p 的一部分人进行处理，e

并将他们与其余人进行比较；这个术语仅用于这种设计。推广面向所有人，并通过前后对比来评估。

没有其他参与者时，新助推的效应正如设计者所预期，记作 E。任何相对于 E 的短缺都可拆分为两部分；只需对同一批人使用相同的随机数运行模拟三次：没有其他参与者、其他参与者保持推广前的行为不变，以及其他参与者作出反应。其他参与者保持不动时的短缺，就是损失的静态部分：场域仅仅存在所带来的成本。其余部分在其他参与者作出反应时出现，即反应部分（方法）。

模型重现了四种已知规律

模型首先依据关于助推的四种已记录规律进行了检查（补充信息附录，图 S3）。助推效应会随规模扩大而减小 [7]。存在竞争者或盟友时，无论采用哪种注意力规则，推广都能保留试点效应的 0.62–0.74。政府试验报告的效应小于学术试验 [6]。在模型中，同一项试验在拥挤场域报告的效应也小于在平静场域报告的效应。一个拥挤场域——竞争者和盟友都达到极限所形成的僵局——是根据这一差距校准的。在该场域中，平静场域报告的效应是拥挤场域的 2.25 倍；所用尺度考虑了这种行为原本的普遍程度。以同样方式衡量，学术界与政府间的差距约为 2。不同地点的效应差异超过了偶然因素所能解释的程度 [11]。在其他参与者各不相同的六个地点间，效应的差异是单个地点内抽样误差的 11 倍。此外，未实施助推时比率略低于二分之一的地方，效应最大 [22]。模型的峰值是一个平台区，对应的比率为 0.356–0.414，接近 0.423；在该比率下，固定推力能使最多的人跨过阈值。这四项是对一致性的检查，而非模型的证据：政府试验与学术试验之间的差距是通过校准构建的；地点间的差异几乎是必然的，因为所选地点的参与者有所不同；而接近二分之一处的峰值在很大程度上源于阈值。

同一个助推，五种结局

图2的每个面板中，助推都相同。不同之处在于其他参与者、注意力有多拥挤，以及如何测试助推。每个面板都展示了设计者的预期、试验报告的结果，以及推广带来的成效；这些结果是对100次运行取平均值。

衰减。 面对一个调节比率的竞争者，推广仍保留了0.63 E （图 2a）。试点报告的结果几乎达到了预期效应的全部：助推引发的响应很小，而且处理组和未处理组受到的影响相同，因此两组比较无法发现这一效应。推广改变了该比率；几个周期内，竞争者触达了更多人，直至达到极限（补充信息附录，图 S2a）。

瞬态。 当同一个竞争者能够升级为自动助推时，效应便无法持久（图2b）。推广约十个周期后，所有运行中都会启动自动助推（补充信息附录，图S2b）。随后，效应降至零以下，达到 $-0.37\,E$；随着竞争者降低自身的触达率，效应又逐渐回升。到最后，效应被消除：数值为0.00 E。

空值。 </style}低于 E 四分之一的效应视为空值。图2c展示了一个高度拥挤的场域：一名竞争者已达到极限，另有16名旁观者争夺紧张的注意力预算。即使在试点中，效果也微乎其微（0.15 E），因为在推广开始之前，人群就已占用了新助推所需的注意力。推广带来了0.20 E 的效应。这个拥挤场域并非上文校准的僵局：它有一名竞争者和旁观者，却没有盟友（方法）；而且负载足够大，足以使竞争者达到极限。

反向。 </style}面对一个不断棘轮式推进的竞争者，新助推实施后，行为最终低于没有新助推时的水平，达到 $-0.34\,E$；然而，试点再次报告了几乎完整的效应（图 2d）。这名竞争者在推广约14个周期后开启其自动助推。在下面的图谱中（图3），一种反转在多次运行和相邻

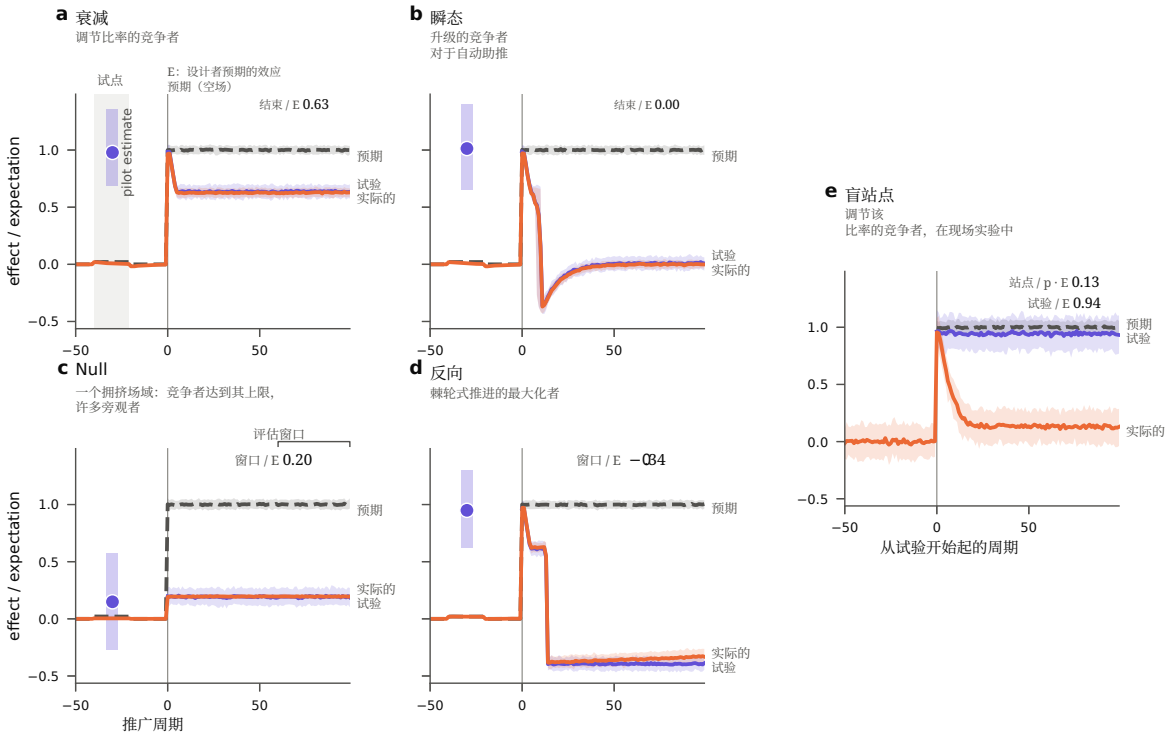


图2：同一助推被部署到其他参与者、注意力拥挤程度以及助推测试方式各不相同的场域中：设计者的预期、试验报告的结果与推广交付的结果彼此偏离，而差距的方向和大小取决于其他参与者与注意力负荷。a–d，先进行试点，再进行推广；周期从推广时的0开始计数（运行的第100个周期）。**a**，衰减：一名恒温器竞争者（即将比率调节至其目标的竞争者），响应增益为2。**b**，瞬态：同一竞争者逐步升级为强度为 τ（自动）0.30的持久性自动助推。**c**，空值：一个拥挤场域，竞争者（τ 0.35）达到其极限，16名旁观者，注意力预算为0.7。**d**，反向：一名最大化竞争者（即在比率高于其基准值时仍施加推动的竞争者），配有持久性自动助推，τ（自动）0.30；其基准值为它曾见过的最低比率。**e**，盲站点：针对**a**中的竞争者开展现场实验，站点中有40%接受处理。曲线显示了在 E期间的预期（灰色虚线）、试验报告的（紫色）和实际的（橙色）效应；其中，预期效应是设计者预期新助推在没有其他参与者的场域中对站点产生的效应。在**a–c** 中，试验报告曲线（即前后估计）位于实际曲线下方（在评估窗口内为 $0.02\,E$）；在**d** 和**e**中，两者出现分离。带有误差线的紫点表示先导试验估计：试点对比相对于 E的份额，即在试点窗口（周期 -40 至 -21，在**a**中以阴影标示）内取平均值；在**c** 中，试点已直接在拥挤场域中运行。评估窗口为第60–99周期（在**c**上方用括号标示）。每个面板都打印其结果类别所读取的数量：终点 / E，即最后10个周期的均值（**a**、**b**）；窗口 / E，即评估窗口均值（**c**、**d**）；在**e**中，站点 / $p \cdot E$，即站点效应相对于站点预期效应的比值（p，接受处理的份额），以及试验 / E，即处理组与 $-$ 对照组在 E上的对比。曲线表示重复样本均值；区间涵盖每个面板 $n = 100$ 次重复运行的第10至第90百分位数。图示的每种结局在各次运行及相邻设置中均保持一致（图3）。

设置需要这样一个竞争者。在下文所述的随机设置中，能够调节比率的竞争者也会在其中几个百分点的设置中反转效应。

盲站点。 只有现场实验能显示第五种结局（图 2e）。面对图 2a 中的竞争者，一项将站点中 40% 纳入处理组的试验报告称，每位接受治疗者获得了预期效应的 94%。整个站点获得的增益，仅相当于对其中 40% 进行处理时预期增益的 13%。该试验的比较仍然有效，但站点获得的增益很小。在随机设置中，这种结局出现在获得某种结局的现场实验的 4–16% 中（见下文）。

在图 2a–c 的推广中，前后比较将实际效应追踪到相差不超过 0.02 E；而在图 2d 和 2e 中，报告的效应与实际效应出现分歧。研究人员还寻找了另外两种结局：延迟效应和振荡效应。在图谱中，这两种效应都没有在相邻设置之间持续存在。在随机设置中，两者都很少见：无论竞争者的规则、注意力规则或试验设计如何，其出现比例在所有设置中都低于 1%；而在模型同时复现已知规律的设置中，其比例低于 2%（补充信息附录，S4 节）。

损失按场域所决定的比例分为静态部分和反应部分（补充信息附录，图 S1）。在图2c的零场中，全部损失 0.80 E 都是静态的：那里没有行动者能够作出反应。只要竞争者能够作出反应，场域仅因其身处其中而造成约 0.03 E 的成本（静态部分，在图 S1 中以 $-0.03\,E$ 绘出），而反应几乎造成了全部损失。盟友则将两者结合起来。随着比率上升，盟友会撤回其助推，效应降至 0.46 E。撤回造成了该损失的 41%；其余部分是静态损失。

决定结果的因素

在新助推保持不变的情况下，改变场域的两个属性，就能让助推经历全部五种结局（图3）。

面对调节比率的竞争者，结局取决于其强度以及它能否升级（图3a，左）。若有自动助推可供其升级至该水平，那么在图的大部分区域，效应都是瞬态的。若没有自动助推，那么强度与新助推相当的竞争者会削弱该效应。在优先级共享的情况下，更强的竞争者会使效应逐渐消退；当其强度约为新助推的 1.5 倍时，效应便不复存在。尚未达到极限的竞争者仍有余力反击，此处称为余量。其目标决定了它在推广前是否有余量。在校准情境中，目标比无行动者时的比率低 0.03；这一距离经过设定，使竞争者在推广前有余量、推广后则耗尽余量。下文的随机情境将这一距离设为 0 至 0.10。面对采用棘轮式推进的竞争者，当其自动助推足够强时，效应便会反转（图3a，右）。

在改变旁观者数量和注意力预算的图上，结局按照固定顺序变化（图3b）。随着旁观者增多、注意力预算收紧，效应依次从瞬态变为衰减，再变为空值。在竞争者达到其极限之前，效应几乎不随旁观者数量变化（在默认注意力水平下为0.39–0.42 E ）。一旦旁观者数量足以将竞争者推至其极限，效应便下降并趋近于空值。一种可能的解读是，拥挤分两个阶段发挥作用：负载首先削弱其他参与者的反应能力，随后才影响新助推本身的效应。在同一图上的现场实验中，盲站点出现在注意力充足且旁观者较少的区域，因此竞争者仍保有反应空间（图3c）。

每张图上的轮廓标出了模型同时再现规模损失和已发表效应量的设置。在轮廓内部，这些图上没有任何一组相邻设置能够产生预期效应。其定义在很大程度上预先决定了这一点：轮廓要求推广产生的结果低于试点报告值的0.8，而只有从0.75 E 起才将某一效应计为预期效应，因此，某个设置只有在试点报告值至少为0.94 E时，才可能同时满足这两个条件。图2c的零场位于轮廓之外。

对于会对其目标的任何差距作出响应的竞争者，其结局可以表示为四个量的公式，校准后无需再拟合任何参数（方法）。这四个量分别是不存在其他行动者时新助推的效应， E；竞争者的助推使比率降低了多少

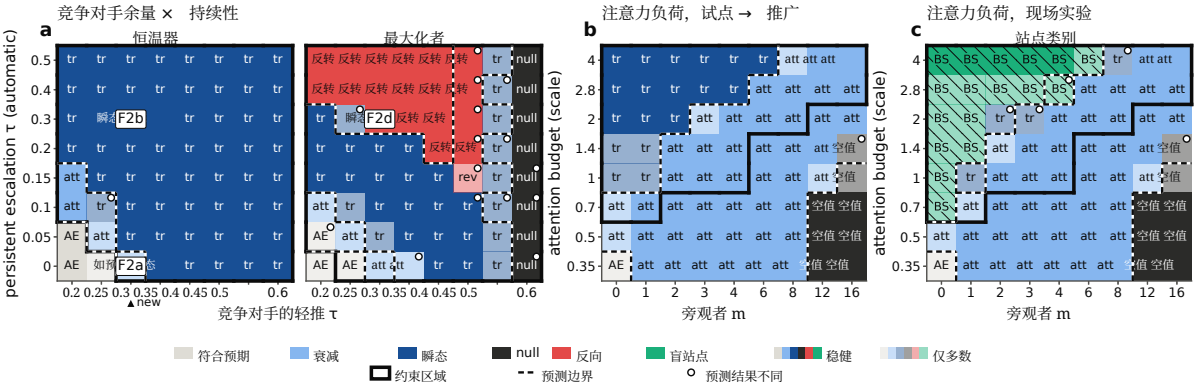


图3：助推会遭遇何种结局，由其他参与者的响应规则和竞争对手余量以及场域的注意力负荷决定；而当最强的助推首先被听见时，还取决于助推自身的强度（补充信息附录，图 S5）：改变两个场参数，就能让同一个助推在模型重现已知规律的区域内，依次经历衰减、瞬态效应、零效应、反转和盲站点。 每个单元格表示该设置下 $n = 100$ 次重复运行中的多数结果。**a**，竞争对手余量和持续性，先试点再推广：竞争者的助推强度 τ（新助推的强度为 0.30，三角形标记为“新”）相对于其持久性自动助推的强度，τ（自动），分别对应恒温器竞争者（左）和最大化者（右）。**b**，注意力负荷，先试点再推广：旁观者数量 m 与注意力预算（尺度）的关系。该尺度是人们注意力预算的平均值：每个个体的预算等于该尺度乘以一个平均值为 1 的随机乘数。模型的默认尺度为 1。**c**，现场实验中的同一张图，按站点发生的情况分类。实心单元格表示稳健：在**a**中，每个 τ 列内，沿 τ（自动）变化时其结果保持不变；在**b** 和**c**中，每个预算行内，沿 m 变化时其结果保持不变；褪色单元格表示仅在该点取得的多数结果。轮廓标出推广总量低于试点对比的 0.8 倍的单元格（四项检查中的第一项，即规模损失；补充信息附录，图 S3），并且不存在其他参与者时新助推的效应位于模型的幅度界限内（方法）。当轮廓线沿着图的边缘时，表示该区域延伸至该边缘。虚线是根据预期速率推导出的结果之间的边界，无需模拟运行（方法）；空心点标出预测结果与模拟结果在结局上不同，或大小相差超过 0.1 的单元格（以 E的份额计；在**c**中，则以站点和试验的预期效应的份额计），单元格颜色采用模拟结果。结果一致的单元格：**a**，恒温器为 64 个中的 63 个，最大化者为 64 个中的 49 个（部分为预测结果）；**b**，72 个中的 71 个；**c**，72 个中的 67 个。当一张图中至少 90% 的单元格一致时，即视为预测成功。F2a、F2b 和 F2d 标出图2a、b 和 d 的场域。坐标轴为类别型且间距不均匀，绘制为等大的单元格，不进行任何插值。单元格标签以灰度标出结局：AE，符合预期；att，衰减；tr，瞬态；null，空值；rev，反向；BS，盲站点（阴影线）。

在新助推触达的人群中，称为它对这些人的拉力；它对其他所有人的拉力；以及在没有其他参与者时，竞争者的目标比率低了多少。通过在析因设计中开启和关闭竞争者的推动，站点的试验可以估计前三项，但这需要竞争者配合。第四项无法观测，必须根据竞争者停止响应的位置推断。当这样的竞争者在推广前还有响应空间、但推广后没有时，E 保留的份额是

$$\frac{\text{effect kept}}{E} = 1 - \frac{\begin{array}{c}\text{pull on the people the new nudge reaches} \\ - \text{distance of the goal}\end{array}}{E}. \tag{1}$$

因此，除非竞争者在推广前仍有推动空间，否则推广在规模扩大后不会损失任何效果。一旦这种拉力达到新助推的效应加上竞争者目标的差距，该效应就会被消除。若竞争者能够升级，其自动助推的强度就是第五项数量。对于采用棘轮式推进的竞争者，或只有当比率偏离其目标超过一个小幅度（即容差带）时才采取行动的转向型竞争者，并不存在这样的公式：要确定助推与之对抗的结局，只需将竞争者自身的规则应用于预期速率，无需再次进行拟合。

我们按照预先确定的规则，将推导结果与模拟结果进行了比较。图上的每个设置都是一个单元格。当推导结果与模拟结果给出相同的结局，且大小相差不超过 0.1 E 时，该单元格便达到一致；当每十个单元格中至少有九个达到一致时，该图便达到一致。按照这一规则，在十张图中的八张上，推导结果与模拟结果一致（图3和补充信息附录，图 S4，虚线）。另外两张图则部分一致：优先级共享下的棘轮式竞争者（64个单元格中有49个）和容差带竞争者（64个单元格中有50个）。在这两张图中，疑似原因是随机性：比率的一次微小随机变化，就会决定是否越过竞争者的触发器，即它所见过的最低比率或其区间边界。在这两张图以及一张达到一致的图中，模拟结果给出的效应略大于推导结果；相对于 E，其中位数约为 0.05 至 0.07，但最强的竞争者除外（补充信息附录，S3节）。

这些结局经受住了规则变化，但其边界发生了移动（补充信息附录，图 S4）。当注意力按比例分配，而不是优先分配给最强者时，或者当竞争者仅在比率离开容差带后才采取行为时，相同的结局会出现在不同位置。在规则之间的 114 次比较中，每次比较考察某一设置下的结局、图中一条线上的结局序列，或某一结局首次出现的点；其中 32 次在结局和大小上均一致。在其余 82 次中，有 76 次规则在某一结局成立的位置上存在差异，另有 6 次仅在其大小上存在差异。在一个角落区域，某一结局发生了变化（补充信息附录，图 S4c）：对于一个较弱的容差带竞争者，即使给予所尝试过的最强自动助推，它也只在略超过一半的运行中反转效应，未达到结局被视为成立所需的 90%。

最后，研究利用该推导确定了49,152 个随机设定下的结局：对于竞争者的三种响应规则（朝目标引导、棘轮效应，以及仅在容差带之外进行引导）、两种注意力规则和两种设计的十二种组合，每种组合包含4,096个设定（方法）。每个设定都会随机抽取九个或十个参数。这些范围仅部分与校准结果相关，因此这些份额描述的是模型，而不是真实场域。三个结果尤为突出。第一，预期效应反而是例外。在优先级共享下，针对引导型竞争者的推广仅在9.0%的设定中实现了该效应，且没有任何组合超过11.7%；在32.1%的设定中，效应受到衰减；在31.9%的设定中为空值；在14.6%的设定中为瞬态；在2.9%的设定中发生反转。第二，什么因素最能改变结局，取决于注意力如何共享。在最强的助推首先被注意到的情况下，在全部六种组合中，最关键的都是新助推自身的强度；而预先确定的预期是，强度在其中任何一种组合中都不会排在第一位。在注意力按比例共享的情况下，最关键的是旁观者数量，新助推的强度则重要得多。第三，在获得结局的设定中，引导型竞争者会在3–6%的设定中使效应发生反转；在推广中，几乎所有这些设定（323项中的322项）都启用了其自动助推。针对这一竞争者，研究还模拟了96个随机设定。其中90个设定的模拟结果在结局和大小上与推导结果一致。推导预测有六个设定会发生反转，而模拟在这六个设定中均显示效应发生了反转。

在另外两种响应规则下，一致性较低（96项中的55项和96项中的79项）。样本抽取前已确定五项预期。其中一项得到验证；补充信息附录的S4 节报告了全部五项。

试验能够记录的特征

试验无法通过比较处理组和未处理组人群而观察到的结局，会在它能够记录的数量中留下痕迹（图4）。

盲站点会在未处理组中显现（图 4a）。当站点中有40%接受处理时，未处理组人群的比率比试验前水平低0.018，即低1.8个百分点，相当于预期每位接受处理者效应的三分之一。在同一窗口期内，试验的比较结果仍约为该预期效应的95%（图2e的运行中为94%），因此仅凭比较结果看不出任何异常。如果试验还将对照组与其自身的试验前水平进行比较，并且该站点原本会维持在这一水平，那么试验就能发现这种偏移。

只要其他参与者让站点的比率维持在原有水平，试验对其处理人群所受效应的夸大程度便可由一个会计恒等式推导得出（图 4b）。此时，未处理组的下降会抵消处理组的增益，而处理组实际获得的增益只是试验比较结果的一部分 $1-p$： [17, 23, 24]

$$\frac{\text{effect reported by the trial}}{\text{actual effect on the treated}} = \frac{1}{1-p}. \tag{2}$$

在校准情境中，模拟试验遵循这条曲线，偏差不超过 1.3%；此时站点中最多有 30% 处于处理状态。当处理组占比达到 40% 时，竞争者达到其极限，无法再维持站点的比率，数据点便落在曲线下方。推导得出，竞争者达到极限时的处理组占比约为 0.35（图 4b，垂直线）。在随机情境中，这一关系并不具有普遍性。无论处理组占比是多少，在获得结局的现场实验设置中，有 4–16% 的设置出现了盲站点；但即使在只有少数人接受处理的设置中，这一关系成立的比例也不到一半（补充信息附录，S4 节）。这似乎是校准情境所特有的性质，在该情境中，竞争者恰好维持着站点的比率。

推广后，站点的比率出现了反转（图4c）。在反转场域中，前后比较结果起初下降，显然是因为竞争者触达了更多人。推广约 14 个周期后，竞争者的自动助推开始发挥作用。当该助推的强度至少达到新助推的三分之二时，比较结果会降至推广前的水平以下；而当其强度为新助推的三分之一时，比较结果仍高于推广前水平。若过早进行比较，将会报告正向效应。

瞬态效应会在其时间进程中显现出来（图4d）。竞争者响应得越快，该效应就越早降至 E 的四分之三以下：在尝试的五种速度下，推广后分别经过 11、6、4、2 和 2 个周期。它会在 8 至 19 个周期内降至 E 的一半以下，随后自动助推会消除剩余效应。

规模更大的试点并不会观察到更多结果。在覆盖站点 2% 至 60% 的各项试点中，面对图 2a、b 和 d 中的竞争者，试点自身的比较值始终处于 E 的 0.91–1.01 倍范围内（补充信息附录，S8节）。发生变化的是试点对站点所产生的影响。在达到自身极限之前，无论试点规模如何，转向竞争者都会吸收试点对站点效应中相同的比例（在默认响应速度下为 0.45）。2% 的试点与 40% 的试点一样，都必然会引发这种响应，只是程度更小。小规模试点无法做到的是像推广那样，将竞争者驱动至其极限。规模达到 40% 或更大的试点确实会在试点期间将升级型竞争者驱动至其极限。随后，竞争者的自动助推会持续开启，紧接着进行的推广从一个升级场域中开始：其结局为空值，而非瞬态。规模更大的试点本身就是一个盲站点，规模足够大的试点还会改变原本要预测的推广。

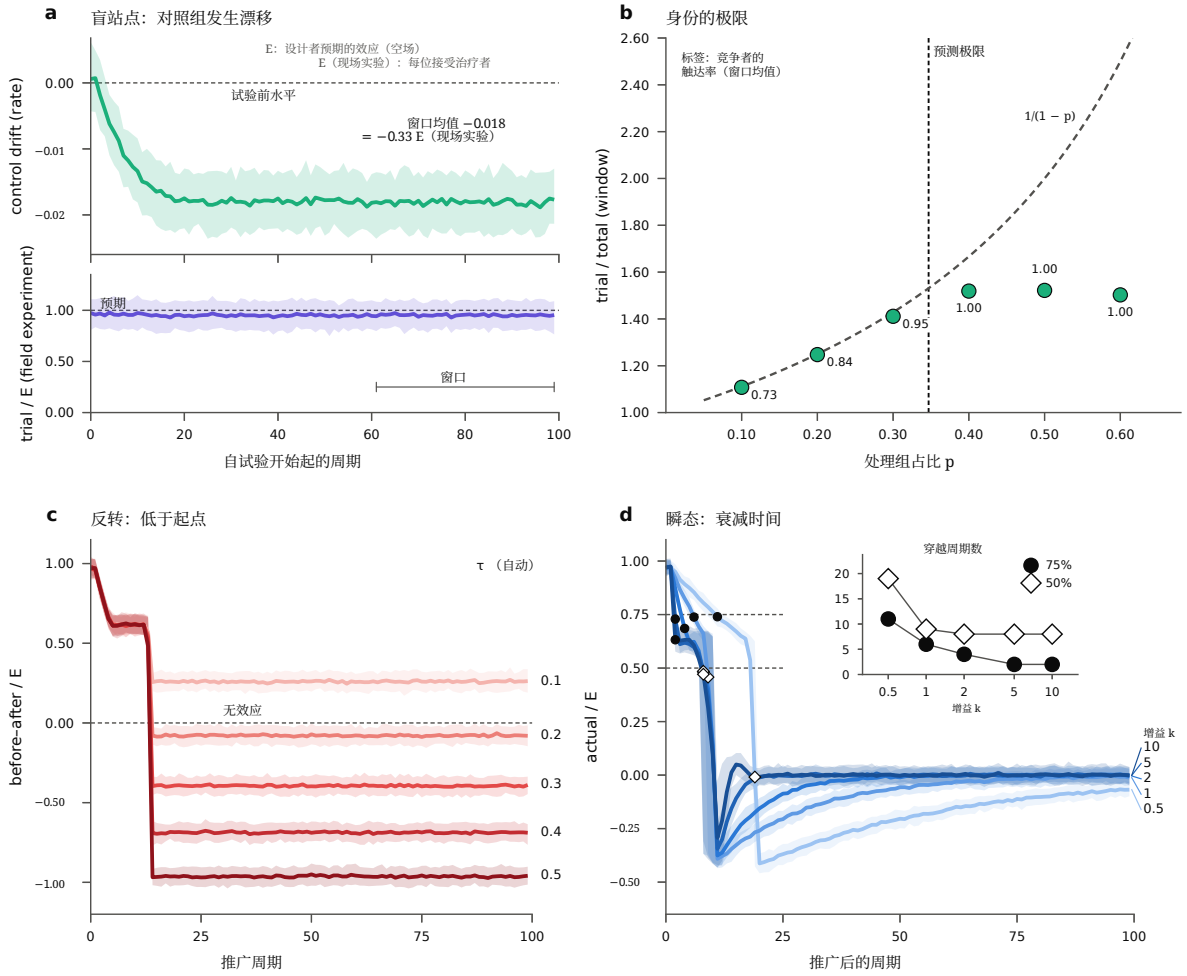


图4：试验无法通过其处理组 − 对照组对比观察到的结局，会在试验能够记录的指标中留下痕迹：对照组的变化、推广后的站点比率，以及效应的时间进程。 **a**，盲站点：在处理组占站点人数40%的现场实验中，顶部为对照组的变化，即采取行动的人群比例的变化（按构造，试验前为0）；底部为试验的处理组 − 对照组对比，占 E 的比例（现场实验）。**b**，恒等式的极限：试验对比与实际效应之比（试验数 / 总数，窗口均值）随处理组占比 p 的变化曲线 $1/(1-p)$。标签显示竞争者的窗口平均触达率；从 $p = 0.4$ 起，它达到极限。垂直线标出处理组占比约为0.35的位置；推导预测竞争者在此处达到极限（方法）。**c**，反转：推广后前后估计与最大化竞争者的对比；每条线代表其自动助推的一种强度，τ（自动）为0.1至0.5。**d**，瞬态：推广后实际效应随 E 的变化；每条线代表图2b中升级型竞争者的一种响应速度（其增益 k为0.5至10）。标记显示每条线首次低于 E的75%（实心圆）和50%（空心菱形）的位置；插图绘出到达各交叉点的周期数随 k 的变化（对数尺度）。 E 是设计者预期的效应，即新助推在没有其他参与者的场域中对站点产生的效应（**c**、**d**）； E （现场实验）是相同预期下每位处理组人员的效应（**a**）。实际（总计）：有其他参与者在场时，采用新助推的站点比率减去未采用新助推时的站点比率；在**b**中，则是每位处理组人员的比率，即采用新助推时处理组人员的比率减去未采用时的比率。交叉点指推广后达到最高点之后，某条线首次低于该水平的周期；周期从推广时开始计数。窗口为第60至第99周期，从推广或试验开始时的0起计（运行的第100周期）。线条表示重复均值；区间涵盖 $n = 100$ 重复运行的第10至第90百分位数。

讨论

在这个模型中，助推的结果取决于还有谁在采取行动。同一个助推可能会衰减、变成瞬态、归于空值或发生反转，也可能留下一个盲站点；而这几种结局之间唯一的差别就是场域：其他会观察某种行为比率并对此作出响应的参与者。助推的结局取决于它进入的场域，而非助推本身。五种结局中，只有空值结局在试点中出现；其余四种都是试点之后才出现的。试验报告的结果也可能与干预实际带来的结果不同：试点的比较无法看出它所引发的响应，推广则遇到了完整的响应，而现场实验的比较依然成立，尽管站点获得的成效甚微。

这种机制并不陌生。某个行动者会在一定极限内，针对比率与其目标之间的差距进行反作用；这就是带上限的负反馈，是控制工程中的标准对象 [25]。系统动力学将通过响应来挫败干预的系统称为政策抵抗 [26]。卢卡斯批判在经济学中也提出了类似观点 [27]：在一项政策下估算出的关系，会在政策发生变化时随之改变。这里讨论的不是反馈本身，而是助推在其中的结局，以及试验会如何报告这一结局。

模型为四种已知异常提供了一种解读。效应会随尺度增大而缩小 [7]：试点只引发较小的响应，而对照将其掩盖了；推广则会引发完整效应。政府试验报告的效应小于学术试验 [6]：同一项试验在拥挤场域中的报告效应低于在安静组中的报告效应，而在校准拥挤场中，这一差距完全源于拥挤造成的注意力变化。DellaVigna 和 Linos 认为，差距主要由选择性发表和统计功效不足造成，因此效应大小相符并不能证明拥挤效应存在；这两种解读针对的是同一个差距。不同站点之间的效应差异大于随机因素所能解释的程度 [11]：仅在其他参与者方面存在差异的站点，会产生不同的效应，而且结局往往也不同。荟萃分析发现，不同研究的效应差异很大 [8–10]：各项研究所处场域的差异也可能造成额外的离散程度。这些现象都不能排除实现、选择或发表方面的失误。即使没有这些因素，其他参与者构成的场域也能产生相同的模式。

这个模型新增了一个能够调节某种行为发生比率的行动者：它会抵消站点的增益，而试验报告的效应仍接近预期效应。Lewin 理念较早的复兴版本将其他力量视为既定条件：广告活动显示出的提升，取决于竞争者同期测试的内容 [20], ；政策的测得效应，则取决于与之并行实施的政策 [21]。组合助推可能会相互抵消 [28, 29], ，一种干预也可能分走另一种干预的注意力 [30, 31], ，就像争夺稀缺注意力的来源彼此竞争一样 [32]。其他参与者即使有所响应，大多也是对干预本身作出响应 [13, 14], ，或随着项目推进，通过价格作出响应 [15, 16, 18]。在一个配套的基于主体的模型中，同样的播种式干预会停滞、漂移或层层扩散，具体取决于人们彼此学习所遵循的规则 [33]：在那里，场域指的是人们相互影响；而在这里，场域指的是其他参与者。与这些参与者最接近的类比，来自在线市场。在那里，A/B测试的两个臂会在同一场拍卖中竞争 [34] ，并使用同一笔广告预算，按固定节奏支出，因此每个臂都会使另一个臂的结果产生偏差 [35, 36]。模型中的参与者则有所不同：他们从未看到助推，只对比率作出响应，而且触达率对处理组和未处理组人群一视同仁。

盲站点最接近这样一类试验研究：对部分人实施干预会改变其他人的结果。其核算方式已为人所知 [23, 24], ；研究发现，在处理组站点中，未处理组人员的表现会变差，而净增益很小 [17]。Munro、Kuang 和 Wager 表明，在同一市场内进行随机化的试验可以估计干预的直接效应，而价格带来的溢出效应则需要借助扩展设计中的价格弹性来估计 [18]。盲站点体现的是同一种区分：试验报告了直接效应，却没有报告总效应。其作用渠道有所不同：这里是行动者将比率调节至某个目标，并可能触及自身极限；而不是通过价格使市场出清。因此，响应有上限，也会随时间变化；要还原总效应，需要记录参与者的行为。损失的静态部分正是先前那些复兴所描述的内容。反应部分则是 Lewin 所描述的响应 [19]。

有一项性质虽未明文写入任何规则，却可由规则推知：陷入僵局的场域不会产生反应。当竞争者和盟友将同一比率调节至无法同时实现的目标时，双方都会使尽全力，因而都无法再作出响应。此时，新助推造成的损失几乎完全是静态的；推广结果大致会维持试点所显示的水平，除非注意力充足，否则不会出现盲站点（补充信息附录，图 S4h,i，以及 S7 节，运行 4）。只有在注意力充足时，新助推才会使比率超过盟友的目标，随后盟友便会退出。因此，场域是否会产生反应，主要取决于其他参与者尚有多少余量，而拥挤会最先耗尽这份余量。基于学术界与政府间差距校准的拥挤场域正是这样一种僵局，因此模型与该差距的吻合完全依赖于拥挤。

真实场域会作出反应吗？已发表的试验给出的答案有所保留。在记录到具有自身目标的行动者时，他们确实作出了反应，不过反应针对的是干预，而非某种行为的比率；而且这并非发生在助推情境中。家庭在预期会获得学校补助金时削减了自身的教育支出 [37], ；学校增加了一名合同制教师后，公务员教师则减少了投入 [38]。两者的表现都像模型中的盟友：随着比率上升，盟友会退出。政党代理人将登记违规行为从被观察到的中心转移到了附近未被观察到的中心 [39], ；未接受辅导的求职者则在接受辅导的人越多的地方表现越差 [17]。两者都呈现出盲站点的模式，但这种模式源于违规行为或可供分配的工作岗位数量固定，而不是源于会调节比率的行动者。关于助推本身，证据则指向相反方向。在 400 多个产品类别中，企业大多没有对竞争对手的促销或广告作出反应 [40]。而且，在所有发现的行为试验中，只要某个地点只有部分人接受处理，未处理组的人就会与处理组的人朝同一方向变化，而非与他们相反。未受到选民意识宣传活动拜访的女性，在邻居受到拜访的情况下投票率更高 [41]。当一个街区中有较高比例的人收到信件时，未处理组的纳税人缴纳了更多税款 [42]。在信息广告触达大多数选民的地方，未看到广告的选区也与看到广告的选区朝同一方向变化 [43]。没有收到执法通知邮件的家庭，在邻居收到邮件时也更愿意遵守规定 [44]。正向溢出效应说明人们之间存在相互影响，而这正是本模型未纳入、由配套模型加以研究的内容 [33]。目前尚未发现一个开放数据集能同时涵盖行为干预、变化的处理组占比和明确的行动者，因此助推场域是否会作出反应、反应出现的频率如何，仍有待测量。

模型有其局限。它依据已发表的量级进行校准，且未拟合任何数据，因此展示的是一个会作出反应的场域可能产生什么结果，而非真实场域中这种情况发生的频率。除了棘轮式竞争者（其触达率增益是永久的）之外，每个场域都按设计稳定在一种状态：不存在随比率上升而增强的社会规范，也不存在因重复而形成的习惯，更没有人际影响；而配套模型显示，这些因素本身就可能决定结局 [33]。参与者遵循固定响应规则，彼此之间不作建模。每次运行中都固定了三项假设：升级一旦发生便不会撤回；参与者能够无误地观察比率；在试点或现场实验中，同一批人始终留在处理组。注意力成本和预算没有已发表的锚点，因此对它们进行了变动。在优先级共享机制下，助推的强度也决定个体会多早注意到它，因此，强度最能左右结局这一发现，部分反映了该规则的影响。与校准情境和图谱中的结果相比，有两项结果在随机情境中的普适性也较弱。试验报告与处理组占比之间的关系，在不到一半的情境中成立。反转也并非只出现在棘轮式竞争者身上。在推广中，反转几乎总是伴随着竞争者的自动助推开启（在三种响应规则下，反向推广中有98.7%如此）。事后看来，这种效应之所以反转，是因为自动助推使比率下降的幅度超过了竞争者撤回另一项助推所能弥补的幅度。棘轮式竞争者的反转也取决于其回溯期有多长。如果竞争者将当前比率与其见过的最低水平，或与大约25个周期及更长时间内的最低水平进行比较，那么这种反转在各次运行中都会出现；若回溯期约为其一半，效应则会衰减。

试验可以保留两类记录。一类记录场域：清点某个站点上其他助推和参与者的情况，以及它们是否已达到极限。另一类记录结局留下的痕迹：

对照组的比率相对于试验前自身水平的变化、推广后站点的比率，以及效应随时间的变化轨迹；序列方法可以从中读出 [45]。这两类记录本身都无法识别机制：比率漂移、效应衰减和推广后的下降，也可能源于趋势、替代或习惯化。能够识别机制的检验，是在彼此相距足够远、不会共享参与者的站点间随机设定处理组占比，同时保留未处理组，并按照随机饱和设计的做法 [46]，随时间测量行为以及具名参与者的行动；数据集筛查未发现符合这种设计的公开试验（补充信息附录，S1节）。模型还提出了一项尚未开展的检验。棘轮式竞争者会保留其获得的触达率，因此，从这样的场域中撤回新助推，不应使比率恢复到先前水平；而竞争者若引导行为朝向固定目标，场域就应回到该目标。试点中对处理组和未处理组人员的比较无法显示这些情况：试点引发的响应会同时影响两组，而且小规模试点不会像推广那样耗尽竞争者。扩大试点规模也无济于事。对于覆盖站点至多 60% 的试点，比较结果仍接近完整效应，因为引导型竞争者会吸收试点对站点效应中相同比例的部分。而足以耗尽升级型竞争者的试点，会让推广面对一个已经升级的场域。如果试点还记录站点的比率和其他参与者的触达率，就能揭示更多情况。

模型对自己提出的问题给出的答案是这样的：助推的命运取决于周围参与者的余量：如果他们有回应空间，推广就会被吸收、变成瞬态或被反转，而运行良好的试验可以报告完整效应，站点却几乎没有增益；如果他们毫无余地，场域就只是背景，试点也就能预测推广情况。因此，对场域的记录和对效应的记录一样，都是评估的一部分。

材料与方法

本节提供评判结果所需的信息。完整规则以及每项检查背后的数字见补充信息附录 S1节；基于主体的模型标准协议所采用的 ODD 格式（概述、设计概念、细节）中的模型描述 [47]、每个参数及其来源，以及注明日期的运行记录，分别见 S5、S6 和 S7 节。

模型

该模型模拟一个有 $N = 20,000$ 人的站点、新助推以及其他参与者，历时 200 个周期，每个周期视为一周。

人群与助推。 每个个体 i 都有一个基线倾向 a_i，即结果对助推所针对的行为的持续倾向。它位于一个不可观测的尺度上，阈值为 0，并且取自由习惯群体（占人群的 30%）和非习惯性群体构成的混合分布。一个助推 k 有一个方向 $d_k \in \{+1, -1\}$、一个强度 $\tau_k \geq 0$ 以及一个触达率，即它在一个周期内触达站点的份额。它会使个体 i 改变 $\tau_k v_{ik}$；效价 v_{ik}，也就是个人如何接受这一助推，会因人而异，其中约 5% 的人对此持负面反应（助推会排斥他们）。个体 i 在周期 t中行为，当

$$
\begin{aligned}
Y_{it} = \mathbf{1}\Big\{ a_i + s_i^{\text{auto}} \sum_{k \in \mathcal{A}_{it}} d_k \tau_k v_{ik} \\
+ s_i^{\text{delib}} \sum_{k \in \mathcal{D}_{it}} w_{ikt}\, d_k \tau_k v_{ik} + \varepsilon_{it} > 0 \Big\}, \qquad \varepsilon_{it} \sim \mathcal{N}(0,1),
\end{aligned}
\tag{3}
$$

其中，$\mathcal{A}_{it}$ 和 $\mathcal{D}_{it}$ 分别表示在周期 t内触达 i 的自动助推和审慎助推（需要注意力），s_i^{auto}和 s_i^{delib}分别表示个人对每种类型的易感性，以及个人给予助推的注意力。噪声在不同个体和周期之间相互独立。新助推属于审慎型，能够提升行为，其 $\tau = 0.30$。在没有其他行动者的情况下，对照组参与率（即没有助推时的

比率）为 0.309，而新助推在概率单位尺度上的效应为 0.152，这一尺度考虑了该行为原本的普及程度。两者都处于根据已发表的试验为模型设定的范围内（概率单位效应为 0.05–0.4，对照组参与率为 0.05–0.9） [2, 6]。

注意力。每个个体每个周期都有一定的注意力预算，每种审慎型助推也有一定的注意力成本（这里每种助推的成本均为 1.0）。在默认的优先级共享模式下，个人会先关注触达自己的审慎型助推中感知推动力最强的那些 $\tau_k|v_{ik}|$（即图 1b 中的显著性），直到注意力预算用完；其余助推则得不到关注。在比例共享模式下，触达个人的每种审慎型助推权重相同，即将注意力预算除以这些助推的总成本，且权重最高为 1。自动助推不受注意力限制。

行动者。每个其他行动者各自拥有一种审慎型助推，并设定其触达率。竞争者会降低目标行为的发生率，盟友则会提高发生率。旁观者会促进另一种不同的行为：它会争夺注意力，但不会推动目标行为，也不会对其作出响应；从起点开始，其助推触达站点内 60% 的人，且始终不变。竞争者和盟友的初始触达率均为 0，收到首次延迟观测到的比率后才开始响应；最大化者（见下文）在预热期内保持 0.60 的触达率，预热期指助推介入前的开头几个周期。行动者只能观测到其目标行为在该站点的比率，且会延迟两个周期；它看不到其他行动者的助推，也看不到新助推。触达整个站点的行动者已达到极限；触达人数较少的行动者则还有余量。

竞争者遵循三种响应规则之一；此处运行中的盟友都是恒温器。恒温器，即结果部分所述的引导型竞争者，会将观测速率调节至目标值：每个周期，它都会根据观测速率与目标值之间的差距，乘以一个增益（默认为 2），来调整自己的触达率，并将其限制在 0 到 1 之间。竞争者的目标值是该站点在没有参与者时的速率减去 0.03，而盟友的目标值则是该速率加上 0.03。最大化者，即结果部分所述的棘轮式竞争者，会保留一个基准值，也就是它见过的最低速率；每当观测速率比基准值高出 0.01 以上时，它就提高自己的触达率，而且绝不降低触达率：这就是棘轮。阈值触发器，即结果部分所述的容差带竞争者，会以每周期 0.05 的固定步长，朝缩小差距的方向调整触达率；只要速率与目标值相差超过 0.01 就会这样做，否则便保持不变。三者中的任何一种还可以通过开启针对该行为的自动助推来升级；该助推会触达所有人，并在本次运行剩余时间内持续开启。恒温器或阈值触发器会在观测速率连续十个周期高于目标值且自身触达率已达到极限时开启自动助推；最大化者则会在触达率达到极限并且速率连续十个周期比基准值高出 0.01 以上后开启。

设计与效果

每次运行都从 60 个周期（0–59）的预热期开始，让场域在助推介入前趋于稳定。试点会在周期 60–79 期间覆盖站点中固定随机抽取的 2%。此后，新助推会暂停，直到周期 100 才开始推广，并持续覆盖所有人，直至本次运行结束。现场实验不设试点：从周期 100 起，覆盖站点中固定随机抽取的份额 p（默认为 0.4），其余部分构成对照组。结果在评估窗口（周期 160–199）内读取。

每次运行都会在三个共享参与者及其随机抽样结果的世界中重复进行：没有其他参与者的世界、将参与者保持在推广前触达率的世界，以及参与者会作出反应的世界。每个世界都会分别在有和没有新助推的情况下运行，两种情况下比率的差异就是该世界中的效应（在现场实验中，则是处理组中的效应）。 E，即预期效应，是没有其他参与者时的效应；总效应，即结果与图表中的实际效应，是参与者作出反应时的效应。损失包含一个静态部分，即在参与者保持不变时已经存在的部分；以及一个反应部分，即参与者作出反应时新增的部分。试验报告的是反应世界中的处理组 − 对照组对比（即试验中的比较，见结果）。推广没有对照组，

因此通过前后差异进行评估：评估窗口内的平均比率减去第80至99期的平均比率。

结果类别（结局）与稳健性

这些类别、各自的阈值以及稳健性规则，均在首次模拟之前记录于一份注明日期的记录中。对于每次运行，比率为总效应除以 E。我们使用该比率的两个均值：评估窗口内的均值（窗口均值）和窗口最后十个周期内的均值（末段均值）。包含 $|E| < 0.02$ 的运行将被剔除。每次运行都按以下顺序，归入其满足规则的第一个类别：

- 反向：窗口均值 ≤ -0.25；
- 盲法试验，即盲站点的个体层面对应类别（仅适用于现场实验）：试验报告显示为 E 的0.75–1.25倍，同时窗口均值低于0.5；
- 瞬态：自推广开始后的某个周期内，比率为 ≥ 0.75 ，且末段均值为 < 0.5；
- 振荡：窗口内存在一个周期（滞后为2–20个周期时，自相关峰高于0.5，且振幅高于 $0.25|E|$）；
- 延迟：比率首次达到 ≥ 0.75 20 或更高的周期数（从起点算起），且末段均值为 ≥ 0.75；
- 空值： |窗口均值| < 0.25；
- 衰减： $[0.25, 0.75]$中的末段均值；
- 符合预期： $[0.75, 1.25]$中的窗口均值；
- 其他情况均未分类。

在首次运行后进行的一项修订，确定了类别符合预期 和延迟 的现行形式；重新对保存的运行结果进行分类后，6,782 项中有 372 项改变了类别。在一项现场实验中，首先检验两个站点级类别，并与设计者预期的站点层面效应进行比较；该效应来自对份额 p进行处理，其值为每位接受治疗者效应的 p 倍。如果某站点的变化不高于该效应的 -0.25 ，则该站点属于反向站点；如果试验报告的效应为每位接受治疗者效应的 0.75–1.25 倍，而站点变化低于预期站点层面效应的一半，则该站点属于盲站点。

结果来自扫描，即在参数值网格上进行的一系列运行。当某一类别在扫描中至少两个相邻点上均有至少90%的运行符合，且这些点覆盖的扫描范围至少达到10%时，该类别便是稳健的。图3中的轮廓标出了满足以下条件的单元格：推广的总效应低于试点对比的0.8，且没有其他行动者时新助推的效应处于上述界限之内。每项预先指定的主张都在另一种注意力规则下重新运行；恒温器的图还在阈值触发器下重新绘制。消除 不是一个类别：结果部分用它表示一种瞬态效应，其运行平均终端效应为0.00 E。

对照四种已知规律的检验

在预测任何新结果之前，模型必须在方向和大致幅度上再现四项已有记录的规律。每项都是一项测试，其阈值和运行次数（每种设置10次）均预先固定：规模损失 [7], 安静场与拥挤场之间的差距 [6], 跨地点离散度 [11] 以及接近二分之一处的峰值（倒U形） [22]。四项测试均通过，但有三点局限。第一项通过运行均值的比率得到满足：2%的试点结果噪声较大，每种设置下的10次单次运行中，有1至6次未显示出这种损失。第二项是校准，而非预测：拥挤场中的旁观者数量是为满足该项测试而设定的。第四项中，每种设置运行10次只能将峰值定位为一个平台区。图2c和图3b、c背后的拥挤场并非僵局：其中有一个强度为0.35的恒温器竞争者和若干旁观者，但没有盟友。在其校准点（两名旁观者、默认预算），竞争者仍有余量，而第二项测试也只是勉强通过：报告的

效应的安静场与拥挤场之比为1.508，处于1.5–2.5的可接受范围内。如图2c所示，负载加大后，竞争者会达到其极限。补充信息附录的S1节列出了每项测试及其数据。

盲点恒等式

用 D 表示试验中处理组与 – 对照组之间的对比，用 C 表示对照组相对于未进行试验时本会发生的变化。根据核算结果，当有 p 的份额接受处理时，处理组的增益 $G_{\text{treated}} = D + C$，站点增益 $G_{\text{site}} = pD + C$。如果一个仍有余量的竞争者将站点比率维持在原有水平，即 $G_{\text{site}} = 0$，那么 $C = -pD$ 且 $G_{\text{treated}} = (1 - p)\, D$：试验报告的数值是实际处理效应的 $1/(1 - p)$ 倍（式 2）。这一关系是恒等式，而非定律：只有在竞争者仍有余量且将站点比率维持在原有水平时，它才成立。C 是相对于反事实来衡量的；只有当站点在其他情况下本会维持原状时，对照组的试验前比率才能代表它（推导见补充信息附录的 S1 节）。

解析约化

这是结果的推导。行动者的助推在每个周期独立触达每个个体，因此预期站点比率关于每个行动者的触达率呈线性。只有一个竞争者时，推导需要计算行动概率的四种平均值：两种助推都未触达个体时、新助推单独触达时、竞争者的助推单独触达时，以及两者都触达时。这些平均值给出 E 以及竞争者的拉动效应，分别针对新助推未触达和已触达的人；当竞争者目标与无行动者时的比率（0.03）之间的距离已知时，这四个量就是结果中的四个数量。对于恒温器，其静息时的触达率、推广后保留的 E 份额，以及它在现场实验中达到极限时的处理组占比，均可通过闭式解得到（见补充信息附录表 S1）；启用自动助推后，该助推的强度则成为第五个量。对于每条规则，与模拟进行对比的预测都将模拟器自身的行动者规则应用于预期比率，因此也给出静态部分和反应部分。平均值是对从模型总体中抽取的 200,000 个个体计算得出的。没有进行任何拟合。

当某个图谱单元格的预测类别与模拟多数类别一致，且预测大小与模拟值的差异不超过其 E 的0.1（在现场实验中，则不超过预期站点层面效应和每位接受治疗者的预期效应的0.1）时，该单元格即为一致；当图谱中至少90%的单元格一致时，该图谱即为预测图谱。在比较图3和补充信息附录图 S4中的十张不同的图之前，规则便已确定。约简建立之前，模拟图已知；该约简不含自由参数。文中还说明了对少数已模拟设置进行的可行性检查，以及一次未计划的模拟器运行，详见补充信息附录S3节。八张图谱通过检验。在优先级共享下的最大化者（64个单元格中有49个）和阈值触发器（64个单元格中有50个）部分通过预测检验。另有七组运行（对应图2a、b、d、e和图4，以及补充信息附录图 S1中的盟友）在所有点上均一致（33个点中有33个）。

结局对场域参数的敏感性

针对响应规则、注意力规则和设计的十二种组合中的每一种，我们从各范围内均匀分布的设置中抽取了 4,096 个样本（采用拉丁超立方体抽样）：共计 49,152 个设置，每个设置包含一名竞争者且没有盟友。助推、竞争者、注意力和参与率的九个参数，以及现场实验中的处理组占比均有所变化（范围见补充信息附录 S4 节）。根据约简对该次运行的预测（该预测由不含抽样噪声的预期速率计算得出），按照上述规则判定每个设置所属的类别。我们报告每种组合中各结局的占比，并针对每个参数，报告该参数抽取值最低三分之一与最高三分之一之间各结局占比的变化幅度。每种组合的设置中，随机抽取 24 个（共 288 个）进行模拟，并依据上述单元规则将模拟结果与该预测进行比较。范围、

统计数据和五项预期在抽取样本之前已固定；补充信息附录 S4 节报告了每项预期的登记值和实际结果，包括未能达到的预期。

预先指定

每张图背后的每次扫描，都在运行前以日期条目的形式记录在版本控制记录中，并注明情景、设计、扫描值、运行、随机种子、注意力规则、预期的各点类别，以及所依据的先验知识。该记录并非公开预注册。偏差和未达到的预期都会在条目中披露。有一项决定缩小了图的范围：图3a从竞争者强度为 0.20 开始，因为当强度为 0.15 时，恒温器在预热期达到极限，并在试点前升级；敏感性扫描则纳入了此类设置。在其他所有结果都已得出后才登记的第11次运行，将试点份额从 0.02 变到 0.60，并与图2中的三位竞争者进行比较（补充信息附录，S8节）。补充信息附录的 S1 和 S7 节列出了每张图背后的运行。

软件与可复现性

模拟器使用 Python 和 PyTorch 编写。不同设备上的随机抽样结果有所不同，因此只有在同类设备上才能精确复现一次运行；第11次运行在第二种设备上重复了图 2a、b 和 d 的设置，所得类别相同，且每个比率与第一次运行的差值均在 0.1 以内，这一容差在运行前已确定。每次运行都会保留一份记录（清单），其中包含其配置、随机种子、设备、软件版本以及代码的确切版本；如果清单显示运行未完成，或代码存在未记录的更改，绘图脚本就会拒绝使用该运行。每张图背后的每次运行都经过预先指定、完整完成，并使用有记录的代码版本；图表也由此类版本绘制，再次绘制得到的文件完全相同。

数据、材料与软件可用性

模拟器（使用 PyTorch 的 Python）、每次运行的配置和注明日期的记录、图的脚本、每个图和表背后的汇总表格、每次运行的清单（配置、随机种子、设备、软件版本和代码版本），以及每次运行的效应表（1,635 个 Parquet 文件，约 1.2 GB）均已存档至 Zenodo，数字对象标识符：10.5281/zenodo.24150612。模型根据已发表的效应量进行校准，未使用任何个体层面数据；正文中的每个数字均已在文章或补充信息附录中报告。

致谢

作者感谢杨潞龄医学院、新加坡国立大学提供的启动经费支持。

参考文献

[1] Richard H. Thaler 和 Cass R. Sunstein。《style id='2'>助推：改善有关健康、财富与幸福的决策》。耶鲁大学出版社，纽黑文，2008 年。ISBN 9780300122237。[2] Katherine L. Milkman、Dena Gromet、Hung Ho、Joseph S. Kay、Timothy W. Lee、Pepi Pandiloski 等。大型研究提升应用行为科学的影响力。《style id='5'>自然》，600(7889)：478–483，2021 年。数字对象标识符：10.1038/s41586-021-04128-4。[3] Christopher B. Mann 和 Katherine Haenschen。基于选举显著性的选民动员策略荟萃分析。《style id='8'>选举研究》，87：102729，2024 年。数字对象标识符：10.1016/j.electstud.2023.102729。

[4] Hunt Allcott。项目评估中的地点选择偏差。《style id='2'>经济学季刊》，130(3)：1117–1165，2015 年。数字对象标识符：10.1093/qje/qjv015。[5] Madalina Vlasceanu、Kimberly C. Doell、Joseph B. Bak-Coleman、Boryana Todorova、Michael Berkebile-Weinberg、Samantha J. Grayson、Yash Patel、Danielle Goldwert、Yifei Pei、Alek Chakroff 等。运用行为科学应对气候变化：63 个国家的全球干预竞赛。《style id='5'>科学进展》，10(6)：eadj5778，2024 年。数字对象标识符：10.1126/sciadv.adj5778。[6] Stefano DellaVigna 和 Elizabeth Linos。从随机对照试验到规模化：来自两个助推部门的综合证据。《style id='8'>计量经济学》，90(1)：81–116，2022 年。数字对象标识符：10.3982/ecta18709。[7] Omar Al-Ubaydli、John A. List 和 Dana Suskind。运用科学的科学：理解可扩展性面临的威胁。《style id='11'>国际经济评论》，61(4)：1387–1409，2020 年。数字对象标识符：10.1111/iere.12476。[8] Stephanie Mertens、Mario Herberz、Ulf J. J. Hahnel 和 Tobias Brosch。助推的有效性：跨行为领域选择架构干预的荟萃分析。《style id='14'>美国国家科学院院刊》，119(1)：e2107346118，2022 年。数字对象标识符：10.1073/ pnas.2107346118。[9] Maximilian Maier、František Bartoš、T. D. Stanley、David R. Shanks、Adam J. L. Harris 和 Eric-Jan Wagenmakers。调整发表偏倚后，没有证据表明助推有效。《style id='17'>美国国家科学院院刊》，119(31)：e2200300119，2022 年。数字对象标识符：10.1073/ pnas.2200300119。[10] Barnabas Szaszi、Daniel G. Goldstein、Dilip Soman 和 Susan Michie。选择架构干预的可推广性。《style id='20'>自然评论：心理学》，4(8)：518–529，2025 年。数字对象标识符：10.1038/s44159-025-00471-9。[11] Christopher J. Bryan、Elizabeth Tipton 和 David S. Yeager。没有异质性革命，行为科学不太可能改变世界。《style id='23'>自然·人类行为》，5(8)：980–989，2021 年。数字对象标识符：10.1038/s41562-021-01143-3。[12] Lauren E. Willis。助推何时失效：滑动默认选项。《style id='26'>芝加哥大学法律评论》，80(3)：1155，2013 年。URL https://chicagounbound.uchicago.edu/uclrev/vol80/iss3/4。[13] Cass R. Sunstein。失败的助推。《style id='30'>行为公共政策》，1(1)：4–25，2017 年。数字对象标识符：10.1017/bpp.2016.3。[14] Ran Spiegler。论助推的均衡效应。《style id='33'>法律研究杂志》，44(2)：389–416，2015 年。数字对象标识符：10.1086/684291。[15] James J. Heckman、Lance Lochner 和 Christopher Taber。一般均衡处理效应：学费政策研究。《style id='36'>美国经济评论》，88(2)：381–386，1998 年。[16] Jeremy Lise、Shannon Seitz 和 Jeffrey Smith。均衡政策实验与社会项目评估。工作论文 10283，美国国家经济研究局，2004 年。[17] Bruno Crépon、Esther Duflo、Marc Gurgand、Roland Rathelot 和 Philippe Zamora。劳动力市场政策是否会产生挤出效应？来自一项聚类随机实验的证据。《style id='40'>经济学季刊》，128(2)：531–580，2013 年。数字对象标识符：10.1093/qje/ qjt001。[18] Evan Munro、Xu Kuang 和 Stefan Wager。市场均衡中的处理效应。《style id='43'>美国经济评论》，115(10)：3273–3321，2025 年。数字对象标识符：10.1257/aer.20230039。

[19] Kurt Lewin。《社会科学中的场论：理论论文选集》。Harper，纽约，1951年。Dorwin Cartwright 编。[20] Caio Waisman、Navdeep S. Sahni、Harikesh S. Nair 和 Xiliang Lin。在线广告平台上的并行实验与竞争性干扰。《营销科学》，44(2)：437–456，2025年。数字对象标识符：10.1287/mksc.2022.0194。[21] Ellicott C. Matthay、Erin Hagan、Spruha Joshi、May Lynn Tan、David Vlahov、Nancy Adler 和 M. Maria Glymour。革命将难以评估：共同发生的政策变化如何影响社会政策健康影响的研究。《流行病学评论》，43(1)：19–32，2021年。数字对象标识符：10.1093/epirev/mxab009。[22] Silvia Saccardo、Hengchen Dai、Maria A. Han、Naveen Raja、Sitaram Vangala 和 Daniel M. Croymans。扩大助推规模：基线动机与替代效应的作用。工作论文，2025年4月9日版本。早期版本：“扩大助推规模：谁在行动以及如何行动”，SSRN，数字对象标识符：10.2139/ssrn.3971192，2025年。[23] Michael E. Sobel。住房流动性的随机化研究说明了什么？面对干扰的因果推断。《美国统计学会会刊》，101 (476)：1398–1407，2006年。数字对象标识符：10.1198/016214506000000636。[24] Michael G. Hudgens 和 M. Elizabeth Halloran。迈向存在干扰时的因果推断。《美国统计学会会刊》，103(482)：832–842，2008年。数字对象标识符：10.1198/ 016214508000000292。[25] Karl Johan Åström 和 Richard M. Murray。《反馈系统：科学家与工程师导论》。普林斯顿大学出版社，普林斯顿，2008年。ISBN 9781400828739。数字对象标识符：10.1515/9781400828739。[26] John D. Sterman。从复杂世界的证据中学习。《美国公共卫生杂志》，96(3)：505–514，2006年。数字对象标识符：10.2105/AJPH.2005.066043。[27] Robert E. Lucas, Jr。计量经济学政策评估：批评。《卡内基—罗切斯特公共政策会议论文集》，1：19–46，1976年。数字对象标识符：10.1016/S0167-2231(76)80003-6。[28] Alec Brandon、John A. List、Robert D. Metcalfe、Michael K. Price 和 Florian Rundhammer。检验社会助推中的挤出效应：来自电力市场自然田野实验的证据。《美国国家科学院院刊》，116(12)：5293–5298，2019年。数字对象标识符：10.1073/pnas.1802874115。[29] Marius Alt、Hendrik Bruns、Nives DellaValle 和 Ingrida Murauskaite-Bull。促进亲环境行为的干预措施协同效应——实验研究的荟萃分析。《全球环境变化》，84：102776，2024年。数字对象标识符：10.1016/j.gloenvcha.2023. 102776。[30] Matthias Rodemeier 和 Andreas Löschel。信息助推、补贴与注意力挤出：来自能源效率投资的现场证据。《欧洲经济学会会刊》，23(4)：1397–1430，2025年。数字对象标识符：10.1093/jeea/jvae058。[31] Steffen Altmann、Andreas Grunewald 和 Jonas Radbruch。干预与认知溢出。《经济学研究评论》，89(5)：2293–2328，2022年。数字对象标识符：10.1093/restud/rdab087。[32] Petra Persson。注意力操纵与信息过载。《行为公共政策》，2(1)：78–106，2018年。数字对象标识符：10.1017/bpp.2017.10。

[33] Giuseppe Alessandro Veltri 和 Alberto Acerbi。社会学习规则与行为政策的有效性：基于主体的模型。《行为公共政策》，第1–16页，2026年。数字对象标识符：10.1017/bpp.2026.10046。在线先行出版，数字对象标识符：10.1017/bpp.2026.10046。[34] Thomas Blake 和 Dominic Coey。为什么市场实验比看起来更困难：测试组—对照组干扰的作用。载于第十五届*ACM*经济学与计算会议论文集（*EC '14*），第567–582页，2014年。数字对象标识符：10.1145/2600057.2602837。[35] Min Liu、Jialiang Mao 和 Kang Kang。采用预算拆分设计，开展可信且高效的在线市场实验。载于第*27*届*ACM SIGKDD*知识发现与数据挖掘会议论文集（*KDD '21*），第3319–3329页，2021年。数字对象标识符：10.1145/3447548.3467193。[36] Luofeng Liao、Christian Kroer、Sergei Leonenkov、Okke Schrijvers、Liang Shi、Nicolas Stier-Moses 和 Congshan Zhang。一价节奏均衡之间的干扰：偏差与方差分析。载于国际学习表征会议（*ICLR*），2025年。URL https://arxiv.org/abs/2402.07322。[37] Jishnu Das、Stefan Dercon、James Habyarimana、Pramila Krishnan、Karthik Muralidharan 和 Venkatesh Sundararaman。学校投入、家庭替代与考试成绩。美国经济学杂志：应用经济学，5(2)：29–57，2013年。数字对象标识符：10.1257/app.5.2.29。[38] Esther Duflo、Pascaline Dupas 和 Michael Kremer。学校治理、教师激励与师生比：来自肯尼亚小学的实验性证据。《公共经济学杂志》，123：92–110，2015年。数字对象标识符：10.1016/j.jpubeco.2014.11.008。[39] Nahomi Ichino 和 Matthias Schündeln。遏制还是转移选举违规行为？加纳随机现场实验中观察员的溢出效应。《政治学杂志》，74(1)：292–307，2012年。数字对象标识符：10.1017/S0022381611001368。[40] Jan-Benedict E. M. Steenkamp、Vincent R. Nijs、Dominique M. Hanssens 和 Marnik G. Dekimpe。对广告与促销攻势的竞争反应。《营销科学》，24(1)：35–54，2005年。数字对象标识符：10.1287/mksc.1040.0069。[41] Xavier Giné 和 Ghazala Mansuri。我们将共同前进：关于巴基斯坦女性投票行为的实验性证据。《美国经济学杂志：应用经济学》，10(1)：207–235，2018年。数字对象标识符：10.1257/app.20130480。[42] Guillermo Cruces、Dario Tortarolo 和 Gonzalo Vazquez-Bare。部分人群实验的设计及其在税收遵从溢出效应中的应用。《经济学与统计评论》，第1–45页，2025年。数字对象标识符：10.1162/rest_a_01552。在线先行出版，数字对象标识符：10.1162/rest_a_01552。[43] José Ramón Enríquez、Horacio Larreguy、John Marshall 和 Alberto Simpser。社交媒体上的大规模政治信息：墨西哥的Facebook广告、选民饱和与选举问责。《欧洲经济学会会刊》，22(4)：1678–1722，2024年。数字对象标识符：10.1093/jeea/jvae011。[44] Francesco Drago、Friederike Mengel 和 Christian Traxler。网络中的遵从行为：来自现场实验的证据。《美国经济学杂志：应用经济学》，12(2)：96–133，2020年。数字对象标识符：10.1257/app.20170690。[45] Giuseppe Alessandro Veltri。行为公共政策中的时间敏感型随机对照试验：运用序列方法、个性化与强化学习的实用框架。《行为经济学前沿》，5：1684887，2026年。数字对象标识符：10.3389/frbhe.2026.1684887。

[46] Sarah Baird、J. Aislinn Bohren、Craig McIntosh 和 Berk Özler。存在干扰时实验的最优设计。《经济学与统计评论》，100(5)：844–860，2018。数字对象标识符：10.1162/rest_a_00716。

[47] Volker Grimm、Steven F. Railsback、Christian E. Vincenot、Uta Berger、Cara Gallagher、Donald L. DeAngelis、Bruce Edmonds、Jiaqi Ge、Jarl Giske、Jürgen Groeneveld、Alice S. A. Johnston、Alexander Milles、Jacob Nabe-Nielsen、J. Gareth Polhill、Viktoriia Radchuk、Marie- Sophie Rohwäder、Richard A. Stillman、Jan C. Thiele 和 Daniel Ayllón。《用于描述基于参与者的模型及其他仿真模型的 ODD 协议：提升清晰度、复现性和结构现实主义的第二次更新》。《人工社会与社会模拟期刊》，23（2）：7，2020 年。数字对象标识符：10.18564/jasss.4259。

补充信息附录，S1节：完整方法

本节完整介绍模型、设计、结果类别、检查和分析；正文中的“材料与方法”对其进行了总结。文中保留了模型自身的名称：恒温器、最大化者和阈值触发器分别对应“结果”部分中的转向、棘轮效应和容忍区间竞争者；空置世界、冻结世界和反应世界分别指没有其他参与者的世界、参与者在推广前保持其触达率不变的世界，以及参与者会作出反应的世界；入场券是针对已知规律的四项检查，即 T1–T4；预热期是场域在助推进入前趋于稳定的 60 个周期；参与者的极限是整个站点。

模型

该模型模拟一个包含 $N = 20,000$ 人的站点、新助推和其他参与者，历时 200 个周期，每个周期视为一周。S5节和S6节提供 ODD 描述 [1] 以及每个参数及其来源。

人。 每个个体 i 在潜在尺度上，对焦点行为具有基线倾向概率 a_i 。倾向概率服从双成分正态混合分布：一个人以 0.3 的概率属于习惯型，其取值来自样本量(0.4，0.8^2)；否则取自样本量 (-1.2, 0.8^2)。易感性 s_i^{auto}和 s_i^{delib}决定自动型助推和审慎型助推的作用大小；两者均服从对数正态分布（均值为 1，对数尺度标准差为 0.3）。每个周期的注意力预算为 $B_i = b\,u_i$，其中 u_i 服从对数正态分布（均值为 1，对数尺度标准差为 0.4），预算尺度 b 默认为 1.0。助推 k 会使个体 i 的倾向改变 $\tau_k v_{ik}$。效价 v_{ik} 对每个个体和助推各抽取一次：其幅度服从均值为 1、标准差为 0.5 的对数正态分布，符号以 0.05 的概率翻转，因此平均效价为 0.90。个体 i 在周期 t采取行为时

$$
\begin{aligned}
Y_{it} = \mathbf{1}\Big\{ & a_i + s_i^{\text{auto}} \sum_{k \in \mathcal{A}_{it}} d_k \tau_k v_{ik} \\
& + s_i^{\text{delib}} \sum_{k \in \mathcal{D}_{it}} w_{ikt}\, d_k \tau_k v_{ik} + \varepsilon_{it} > 0 \Big\}, \qquad \varepsilon_{it} \sim \mathcal{N}(0,1),
\end{aligned}
\tag{S1}
$$

其中，$\mathcal{A}_{it}$ 和$\mathcal{D}_{it}$ 分别是在周期 t触达 i 的自动助推和审慎型助推， $d_k \in \{+1, -1\}$表示助推的方向，$w_{ikt} \in [0,1]$ 表示其注意力权重。噪声在人与周期之间相互独立（概率单位），且习惯更新处于关闭状态。

助推。 助推有一个方向 d_k、一种类型（自动或审慎型）、一个强度 $\tau_k \geq 0$、一项注意力成本 c_k（仅审慎型助推有）、一个触达率 $r_{kt} \in [0,1]$ （每个周期内所触达的站点份额）以及一个所有者。行动者的助推每个周期都会重新抽取触达每个个体的概率 r_{kt}；新助推则触达固定的处理组。新助推属于审慎型助推，会提升行为，具有 $\tau = 0.30$ ，成本为1.0；除非运行设定另有说明，其他参与者的助推也具有该成本。在平静场域中，对照组参与率为0.309，新助推的Probit效应为0.152，处于设计范围之内（Probit效应为0.05–0.4，对照组参与率为0.05–0.9） [2, 3]。

注意力。 在优先级共享模式下（默认设置），个人会根据感知到的推动力对触达自己的审慎型助推进行排序，$\tau_k |v_{ik}|$，推动力最强的优先。每个助推依次获得 $w_{ikt} = \min(1,\ B^{\text{rem}}/c_k)$，其中剩余预算 B^{rem} 从 B_i 开始，并按 $w_{ikt} c_k$递减；其余助推则无人关注。在比例共享模式下，触达个人的每个审慎型助推都会获得 $w_{ikt} = \min(1,\ B_i / D_{it})$，其中 $D_{it} = \sum_k c_k$ 在这些助推之间共享。自动助推无需经过注意力分配（$w = 1$）。

参与者。每位其他参与者各自拥有一种审慎型助推，并控制其触达率。竞争者会降低焦点行为的发生率，盟友会提高该比率，而旁观者则会促进不同的行为。参与者的触达率最多可覆盖整个站点；触达率为 $r \geq 0.99$ 的参与者已达到极限，而

低于该值则还有余量。参与者只能观察站点上其目标行为的比率，且存在 $L = 2$ 个周期的滞后；它永远看不到其他参与者的助推，也看不到新助推。

恒温器通过有界积分控制，将观测速率调节至目标值 G 。给定差距 $g_t = r^{\text{obs}}_{t-L} - G$，竞争者将其触达率设为 $r_t = \min\{1,\max\{0,\, r_{t-1} + \kappa g_t\}\}$，盟友则使用 $-\kappa g_t$；默认增益为 $\kappa = 2$。目标值为该次运行的无行动者基线（没有其他参与者或新助推时，周期 0–59 的平均比率），竞争者的目标值比基线低 0.03，盟友的目标值则比基线高 0.03；在预热期，这两个偏移值都能使单独行动的参与者稳定在其极限以下。竞争者和盟友的初始触达率均为 0。若盟友的目标值被超越，它会降低触达率。

只要比率仍高于它所见过的最低水平，最大化者就会持续施压。其触达率在周期 59 之前保持为 0.60。从周期 60 起，它会维持一个基准值 q_t，初始值为周期 50–59 的观测速率平均值。当 $r^{\text{obs}}_{t-L} > q_t + 0.01$ 时，它会将触达率提高 $\kappa(r^{\text{obs}}_{t-L} - q_t)$，上限为 1，且绝不降低。随后，基准值变为 $\min(q_t, r^{\text{obs}}_{t-L})$，形成棘轮。将棘轮替换为指数平均后， $q \leftarrow (1-\lambda)q + \lambda r^{\text{obs}}_{t-L}$，在记忆约为 $1/\lambda = 25$ 个周期时，反转在 $\lambda = 0.04$ 仍然稳健；到 $\lambda = 0.08$，即 12.5 个周期（运行 2c），则转为衰减。

三个竞争者中的任何一个都可以升级。恒温器或阈值触发器在达到其极限且观测到的差距连续十个周期为正时升级；最大化者则在连续十个周期均处于其极限且高于其基准值加 0.01 后升级。随后，它会针对该行为开启一个自动助推，强度为 τ_{auto} 且触达率为 1，并在运行的其余时间保持开启。阈值触发器这一替代响应规则会在 $|g_t| > 0.01$ 时，沿着缩小差距的方向以每个周期固定 0.05 的幅度调整触达率，否则保持不变。

旁观者不会响应：其增益始终为 0，触达率始终为 0.6。旁观者的助推会按照相同的阈值规则推动其自身行为；对于焦点行为，它会以自身的强度争夺注意力，但不会施加任何推动。

设计与反事实世界

每次运行都从预热期开始，即第 0–59 个周期。除非在每次运行中，站点比率在第 40–59 个周期内的拟合斜率（反应世界，不含新助推）低于每个周期 0.002，否则扫描会停止；此处报告的每次运行均通过了该条件。试点会在第 60–79 个周期内处理站点中固定随机抽取的 2% 部分。随后，新助推会保持关闭，直到第 100 个周期；从该周期起，推广会覆盖所有人，直至运行结束。现场实验不设试点：从第 100 个周期起，它会处理站点中固定随机抽取的份额p （默认值为 0.4），其余部分构成对照组。结果在评估窗口内读取，即第 160–199 个周期。

三个世界（没有其他参与者；参与者保持不变；参与者作出反应）各自分别在有无新助推的情况下运行，每次运行使用相同的人群和各周期的随机抽样结果。空世界中没有其他行动者；在反应世界中，参与者会作出反应；冻结世界则从第 100 个周期起，将每位参与者保持在推广前的触达率（以及自动助推可达范围），即反应世界第 90–99 个周期的平均值。某个世界中两次运行的站点比率之差（在现场实验中，即处理组人员的比率）就是空场效应、冻结场效应或总效应。 E 是每次运行在评估窗口内的平均空场效应；在现场实验中，该值按每位处理组人员计算（E_{fe}）。损失的静态部分为（冻结 $-$ 空置）$/E$ ，反应部分为（总效应 $-$ 冻结）$/E$；二者均为每次运行的窗口均值，因此总效应$/E = 1 +$静态部分 $+$反应部分。

试验报告的是反应世界中处理组与 $-$ 对照组的对比（试点：第60至79期；现场实验：从第100期开始），其中包含了两臂构成上的差异，就像真实试验报告的那样。推广没有对照组，因此记录的数量是评估者会测量的前后差异：评估窗口内的平均比率减去第80至99期内的平均比率。在现场实验中，试验前基准是反应世界在第90至99期内的平均站点比率。

结果类别与稳健性

类别、阈值和稳健性规则在首次模拟前已记录并固定。对每次运行，令 $\text{ratio}(t) = \text{total}(t)/E$，其窗口均值取第160至199期，窗口末均值取第190至199期。含有 $|E| < 0.02$ 的运行将被剔除。每次运行按以下顺序归入其符合规则的第一个类别：

- 反向：窗口均值 ≤ -0.25；
- 盲法试验（仅限现场实验）：试验/$E_{\text{现场实验}}$在 $[0.75, 1.25]$ 中的窗口均值，以及窗口均值 < 0.5；
- 瞬态：推广开始后某个周期的比率 ≥ 0.75，以及窗口末尾均值 < 0.5；
- 振荡：在去趋势窗口序列中，负值之后滞后 2–20 处的自相关峰高于 0.5，且振幅高于 $0.25|E|$；
- 延迟：起点后至少 ≥ 0.75 20 个周期才首次出现的比率，以及窗口末尾均值≥ 0.75；
- 空值： |窗口均值| < 0.25；
- 衰减： $[0.25, 0.75]$中的窗口末均值；
- 符合预期： $[0.75, 1.25]$中的窗口均值；
- 否则未分类。

在前几次运行给出符合预期的结果后进行的一项修订，导致其呈现当前形式，此前区间边缘附近的噪声使一些运行被归为未分类或延迟。重新分类已保存的运行（保留原始运行）后，6,782 次运行中有 372 次发生了变化，全部是从未分类变为符合预期，或从延迟变为衰减。

在现场实验中，首先检验两个站点级类别。设计者预期对份额 p 进行处理所产生的站点层面效应为 pE_{fe}。如果窗口均值站点变化为 $\leq -0.25\,pE_{\text{fe}}$，则该运行属于反向站点；如果试验/$E_{\text{fe}}$ 的窗口均值位于 $[0.75, 1.25]$，同时窗口均值站点变化低于 $0.5\,pE_{\text{fe}}$，则该站点属于盲站点。

当某个类别在至少两个相邻扫描点构成的区段内、且该区段跨度至少为扫描范围的10%时，在每个扫描点上至少90%的运行中都成立，该类别便具有稳健性。稳健性按设计评估：图3a中沿升级强度评估，图3b、c中沿旁观者数量评估。轮廓标示出推广的总效应低于试点对比的0.8倍（如以下 T1），且新助推的静默场效应处于该设计界限内的单元格。每项预先指定的主张都在另一种注意力规则下重新运行（恒温器映射还在阈值触发器下重新运行）；若多数类别或其稳健标记不同，则该扫描点的类别为依赖规则；若只有平均比率相差超过0.1，则其大小依赖规则；否则为规则无关。

消除 不是一个类别：结果部分用它指代一种瞬态效应，其运行平均的窗口末端效应为0.00 E （图2b；图3a中恒温器半部在升级强度为零、竞争者强度为0.45时的结果，与竞争者强度为0.35时的0.39 E 相对照）。衰减时间（图4d）指推广后，在运行平均总效应/E 达到峰值的时刻或之后，效应首次降至0.75（随后降至0.5）所经历的周期。

入场券：四项已知规律检验

在预测任何新内容之前，模型必须在方向和大致大小上重现四项有据可查的规律。每项检验的阈值和运行次数（每种设置 10 次）都预先确定。

T1，规模损失 [4]。在每条注意力规则下，先试点再推广，并分别加入一个增益为 2 的恒温器竞争者或一个恒温器盟友。该统计量，即推广阶段窗口平均总量在各次运行中的均值除以试点对比在各次运行中的均值，必须低于 0.8。一个

均值比之所以被采用，是因为 2% 的试点（约 400 人）会产生噪声较大的对比。加入竞争者时，在（优先级、比例）规则下，比率分别为 0.741 和 0.671；加入盟友时，比率分别为 0.621 和 0.738；10 次单次运行中，分别有 4、3、1 和 6 次达到 0.8 或更高。

T2，安静场域与拥挤场域 [2]。统计量是现场实验的 Probit 试验效应，即安静场域中窗口平均率的（处理组）Φ^{-1}与（对照组）$-\Phi^{-1}$之比，在各次运行上取平均后，再与拥挤场域中的相应比率相比。锚点是学术界与政府间差距在概率单位尺度上的值，约为 2；优先级下的区间为 $[1.5, 2.5]$。拥挤场域中有一个恒温器竞争者和一名盟友（增益为 2，偏移量为 ∓ 0.03），另有 m 名旁观者；在预先指定集合 {2，3，4，6，8} 中，区间内最小的值为 $m = 2$，比率为 2.25。因此，T2 是校准值，而非预测值；该区间无法确定 m 的下限（$m = 0$ 和 1 对应的值分别为 1.81 和 2.16）。这里的两个恒温器都在预热期达到极限，因此没有行动者能够作出反应（即僵局场域）。另一组拥挤场域，即余量场，保留了一个会作出反应的竞争者：一个强度为 0.35 的恒温器，另有 m 名旁观者，没有盟友。其校准点（第 8 次运行）在预算尺度为 1.0 时为 $m = 2$，保有余量；T2 为 1.508，刚好落在区间下边界以内，T1 为 0.640。图2c和图3b、c来自这第二组场域，而非僵局场域。

T3，跨地点离散度 [5]。在六个地点开展的现场实验（安静组；一名竞争者；一名盟友；2名和4名旁观者；僵局场域）。窗口均值试验对比的地点间均值的标准差为0.0132，必须超过站点内标准误均值 $3 \times 0.0012 = 0.0036$的三倍。

T4，倒*U*形 [6]。没有其他行动者；新助推自动运行，且 $\tau = 0.5$；一个倾向分量，$a_i \sim \mathcal{N}(\mu, 0.8^2)$，并将 μ 设定为使对照组参与率涵盖从0.1到0.8的15个级别，以及从0.30到0.50的9个级别。推广效应达到峰值时的参与率必须位于内部、低于0.5，并且与阈值机制的预测 $\Phi(-\beta/2) = 0.423$ 相差不超过0.05，且 $\beta = \tau\,\mathbb{E}[s^{\text{auto}}]/\sqrt{1 + 0.8^2}$。峰值位于0.376，但在参与率0.356–0.414范围内，顶部在一个标准误以内保持平坦，因此在运行10次时只能确定到该平台区。

盲点恒等式

在一项对站点中占比为 p 的人实施干预的现场实验中，用 T_t 和 C_t 分别表示反应世界中处理组和对照组的比率，用 $S_t = pT_t + (1-p)C_t$ 表示站点比率，用 S 表示试验前基准值。试验报告的结果为 $T_t - C_t$，对照组漂移为 $C_t - S$，站点变化为 $S_t - S$。根据定义，

$$\text{site change} = \text{control drift} + p \times \text{trial}. \tag{S2}$$

处理组总效应为 $T_t - T_t^0$，其中 T_t^0表示在没有新助推的配对世界中，同一批人的比率。满足两个条件，才能观测到这一效应。第一，如果没有干预，站点的比率本应维持在试验前比率 $S_t^0 \approx \bar{S}$。第二，处理组是站点的随机样本，因此 $T_t^0 \approx S_t^0$。于是，总效应约为 $T_t - \bar{S}$，即试验结果加上对照组漂移。代入定义可得

$$\text{total} \approx (1-p) \times \text{trial} + \text{site change}. \tag{S3}$$

当竞争者将站点的比率维持在其目标值时，站点变化为零，且试验/总效应$= 1/(1-p)$。在运行 6 中，$p = 0.4$ 处的比率为 1.519，比 1.667 低 8.9%：竞争者已达到其极限（窗口平均触达率为 0.9995），因此站点变化不再为零。

解析约化

行动者的助推每个周期都会独立触达每个人，因此预期站点比率与每个行动者的触达率呈线性关系。只有一个竞争者时，需要计算行动概率的四种平均值：$R_\emptyset$ 当两种助推都没有触达某人时，R_o 当只有新助推触达时，R_c当只有竞争者的助推触达时，以及 R_{oc} 当两者都触达时。于是 $E = R_o - R_\emptyset$；竞争者的拉动效应 $D_0 = R_\emptyset - R_c$ 作用于新助推未触达的人，$D_1 = R_o - R_{oc}$ 作用于新助推触达的人；当新助推覆盖的份额为 p 、竞争者的触达率为 r 时，站点比率为

$R_{\emptyset}+pE-rD$（p），D（p）$=(1-p)\,D_0+pD_1$。恒温器的目标位于 $\delta=0.03$ 以下 $R_{\emptyset}$，其静息状态为 r^*（p）$=\min\{1,\max\{0,(pE+\delta)/D(p)\}\}$。根据推广前后的静息状态，在未升级的情况下，保留$E$的份额为 $1-(D_1-\delta)/E$，若 $D_0>\delta$ 且 $D_1<E+\delta$；若 $D_0>\delta$ 且 $D_1\geq E+\delta$，则为 0；若 $D_0\leq\delta$ 且 $D_1<E+\delta$，则为$1-(D_1-D_0)/E$，均为静态；其他情况下则为 $(D_0-\delta)/E\leq 0$。在现场实验中，当处理组占比达到 $p^*=(D_0-\delta)/(E+D_0-D_1)$ 时，恒温器便达到极限；低于该值时，站点变化为零，试验总效应为$/=1/(1-p)$。运行的任何阶段都可能发生升级，因此，若恒温器太弱而无法达到其目标（$D_0<\delta$），它会在预热期升级；一旦其自动助推启动，就会据此重新计算四种速率，因此其强度成为第五个数量。在静息状态附近，当增益乘以拉力小于 1 时（运行中最多为 0.51，增益为 10），恒温器规则在两期滞后下保持稳定。

最大化者的基准值是观测速率的运行最小值，该速率每个周期的波动约为 0.003。它会漂移到预期值以下，因此，若预测采用预期速率，可能会误判最大化者停止或升级的位置。另一项针对它的预测会将基准值下调至起始值减去 $\sigma\,m(n)$：σ 是它读取的速率的抽样标准差，n 是截至目前速率与无噪声运行最小值相差不超过 σ 的周期数，而 $m(n)$是 n 次标准正态抽样的预期最大值。该预测与第一项并列报告，且不纳入下方的规则；采用它时，优先级共享下最大化者图中的 64 个单元格有 53 个一致，而不采用时为 49 个（S3节）。

这些速率（僵局场域中有八个）是在模型的总体抽样方式下，从 200,000 人中抽取样本，并按照此次运行的注意力规则计算出的平均值。预测路径采用预期速率运行模拟器自身的行动者规则。没有进行任何拟合。

当预测类别为模拟多数类别，且类别规则读取的大小与该大小相差不超过0.1（作为 E的份额；在现场实验中，站点变化作为 pE_{fe} 的份额，试验作为 E_{fe}的份额）时，单元格即为一致；当一张图中至少90%的单元格一致时，即预测该图。在比较图3和图 S4中的十张不同的图之前，规则及其应用方式已确定。事先已知并在S3节中声明的内容包括：对草稿中已有的少数设置进行可行性检查，以及在构建过程中针对一个设置进行的一次未计划的模拟器运行。八张图通过了检查；优先级共享下的最大化者（64个单元格中的49个）和阈值触发器（64个单元格中的50个）均得到部分预测。另有七次运行分别对应图2a、b、d、e和图4以及图 S1中的盟友，并且在每个点上都一致（33个点中的33个）。

结局对场域参数的敏感性

对于竞争者（恒温器、最大化者或阈值触发器）、注意力规则和设计的十二种组合，每种组合都通过拉丁超立方抽样抽取了 4,096 个设置，均匀覆盖每个参数的取值范围：共 49,152 个设置，每个设置包含一个竞争者且没有盟友。新助推的强度和竞争者的强度范围为 0.10–0.60；竞争者的增益范围为 0.5–10（对数尺度；对于阈值触发器，其步长为每周期 0.01–0.10）；其目标比无行动者基线低 0 至 0.10；其滞后范围为 1–5 个周期；自动助推的强度范围为 0–0.50；预算尺度范围为 0.35–4.0（对数尺度）；旁观者数量为 0–16；对照组参与率为 0.10–0.80，通过同时调整两个倾向分量一个单位来设定；在现场实验中，处理组占比范围为 0.10–0.60。其余均与上述相同。根据约简的无噪声路径（200,000 人，每个设置使用相同的抽样结果），按照上述规则判定每个设置的类别，以及其是否位于轮廓内；若某个设置包含 $|E|<0.02$，则将其视为排除。报告每种组合中各种结局在所有设置及轮廓内设置中的占比；此外，对每个参数，报告其抽取值最低三分之一与最高三分之一之间结局占比的变化幅度（在已分类设置中计算绝对差之和的一半）。从每种组合的设置中随机抽取 24 个（总计 288 个）进行模拟：在云端 T4 GPU 上，每次运行模拟 20,000 人，共运行 10 次，并按照上述单元规则与路径进行比较。取值范围、统计量和五项预期均在抽样前确定；S4 节报告每项预期的登记内容及实际结果。

预先指定的运行与参数

每幅图背后的扫描在运行前都已作出具体规定，并作为带日期的条目记录在版本控制记录中，注明情景、设计、扫描值、运行、随机种子、注意力规则、每个点的预期类别以及所依据的先验知识。该记录并非公开预注册。偏差和未实现的预期均在条目中披露。运行 1、2 和 2b 涵盖竞争者增益 0.5–5、盟友、旁观者、升级强度 0.1–0.5，以及升级型恒温器的增益 0.5–10（图 2a、b、d、e、4c、d；图 S1、S2）；运行 1p 和 2c 则增加了比例共享和最大化者的记忆。运行 3 绘制恒温器、最大化者和阈值触发器的竞争者强度 0.20–0.60 与升级强度 0–0.50 之间的关系（图3a；图 S4a–e）。运行 4 在预算尺度 0.35–4.0 和旁观者数量 0–16 的范围内绘制僵局场域（图 S4h、i）；运行 6 则考察处理组占比 0.1–0.6（图4a、b）。运行 8 对余量场进行校准；运行 4h 则在两种设计和规则下，按照预算尺度和旁观者数量绘制这些场域（图2c、3b、c；图 S1 、 S2 、 S4f 、 g ）。运行 9 和 10 是由四项检查组成的整洁测试套件及其更精细的 T4 网格（图 S3）。记录及每个条目的日期见 S7 节。第11次运行在 2026年11月12日 最终审查后登记，考察试点在站点中的占比相对于图2所示三种竞争者的变化（S8节）。

软件与可复现性

该模拟器使用 Python 和 PyTorch 编写，并针对运行和个体进行了向量化；它在 Apple M4、NVIDIA DGX Spark 上运行，第 8 次运行则使用云端 T4 GPU。随机种子按情景、设计、参与者结构和种子基数设置；不同设备的抽样结果有所不同，因此运行结果只能在同一设备上复现。每个运行目录都包含一份清单，记录其配置、随机种子、设备、软件版本、代码提交和工作树状态。图表脚本会拒绝任何清单显示运行未完成、工作树不干净、类别规则过时或注意力规则错误的运行；此外，侧车文件还会列出每张图所用的运行和提交。图 1–4 和 S1–S4 均在干净提交上生成，第二次生成的结果与第一次逐字节一致；图 S5 也在干净提交上生成。

数据集筛查

没有找到可用于检验响应场特有现象的开放数据集，因此论文中所有针对特定场域的结果都基于模拟。我们进行了三轮筛查，每轮的资格门槛都在搜索开始前固定。第一轮搜寻已发表的 2×2 现场实验：实验具有二元结果，两项干预都会改变结果，且各臂样本量足以检测概率单位尺度上 0.2 的交互作用。共找到 9 个候选项，但没有一个同时通过两项门槛；较精确的研究有一个因素没有效应，而包含两个有效因素的研究样本量不足。第二轮在 Zenodo、Figshare、OSF、哈佛数据宇宙及另外三个档案库中，搜索三类设计：包含两种强作用力的析因设计；在不同背景下实施相同干预、且已知每个个体的倾向概率的设计；以及记录了其他来源字段的试验。我们评估了约 50 个候选项，其中两个符合用于确定参与率峰值位置的条件，但没有一个符合区分响应场与固定场的检验条件。这是因为开放试验一次只针对一种弱作用力，而具有强作用力的交叉设计则单元格样本量较小。第三轮寻找“结果”部分“试验可记录的特征”中提及的特征：各站点随机分配的处理组占比、记录下来的另一行动者响应，或先进行试点再推广，以及带有站点标识符的多站点试验。66 种设计中，没有一种符合条件。筛查揭示了三个结构性原因。大型研究逐个分配个体，因此各站点总体中的处理组占比在各处大致相同，而且占比很低。在改变处理组占比的行为试验中，对未处理组的溢出效应为零或正值；只有在明确指出固定池或响应行动者的情形下，盲站点才会出现相反符号，而且这些试验研究的并非助推。此外，开放数据记录的是另一行动者的结果，而非其行动。有些候选项出现在多轮筛查中；所有轮次合计评估了约 120 个候选项。筛查记录及所有候选项和门槛，见

该存档（数字对象标识符：10.5281/Zenodo.24150612）。

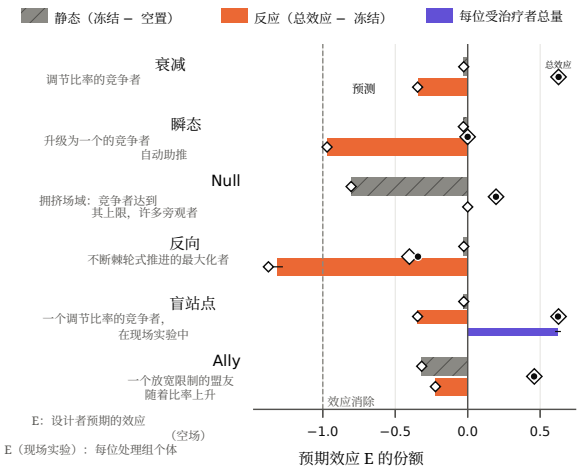


图 S1：场的效应包含一个静态部分，力场早期复兴所描述的就是这一部分；还包含一个反应部分，即 Lewin 所描述的力的响应 [7] 而复兴未涉及这一部分；哪一部分占主导由场决定：单纯的拥挤属于静态部分，具有余量的反应性行动者则属于反应部分。每个场对应一行：按图2中的顺序及其设置列出五个场，然后是一个响应增益为2的恒温器盟友。条形图将新助推在评估窗口内相对于E 的短缺分解为静态部分（灰色、阴影线：冻结的 − 空置）和反应部分（橙色：总 − 冻结），并以 E的份额表示。冻结场将其他参与者的助推保持在推广前的触达率，因此，静态部分是场仅凭存在所造成的成本，反应部分则是参与者对比率的响应所增加的部分。这一分解固定了一种顺序：冻结场就是试点所看到的场，因此静态部分是试点能够测量的，反应部分则是试点无法测量的；两部分相加恰好等于完整的短缺。 E 是新助推在空场中对站点产生的效应。由于 $1 +$静态 $+$反应 $=$总效应 $/\,E$，每行中的墨点（“总效应”）就是窗口内相对于 E的总效应。 -1处的虚线（“效应消除”）标示两部分合计抵消预期效应的位置；超过该线的部分，如反向行中的部分，会使效应的符号反转。空值行没有反应条：其在每次重复运行中的反应均为0，因为竞争者在试点前已达到极限，旁观者也不会响应这一比率。盲 site 行按每位接受治疗者计，在 E （现场实验）期间统计，并加入每位受治疗者总量（紫色）。空心菱形标出解析约化对各部分的预测值，并在墨点周围标出对总效应的预测值（S1节）；在反向行中（其反应部分），它们与模拟值的最大差异约为0.06，在其他行中则低于0.004。条形图表示重复均值。误差线覆盖每行 $n = 100$ 次重复运行的第10至第90百分位数；仅在条形足够长、能够容纳误差线时绘制，因此约为 $-0.03\,E$ 的静态条不带误差线。所有运行均已预先指定、完整，并来自一次干净提交。

补充信息附录，S2 节：补充图

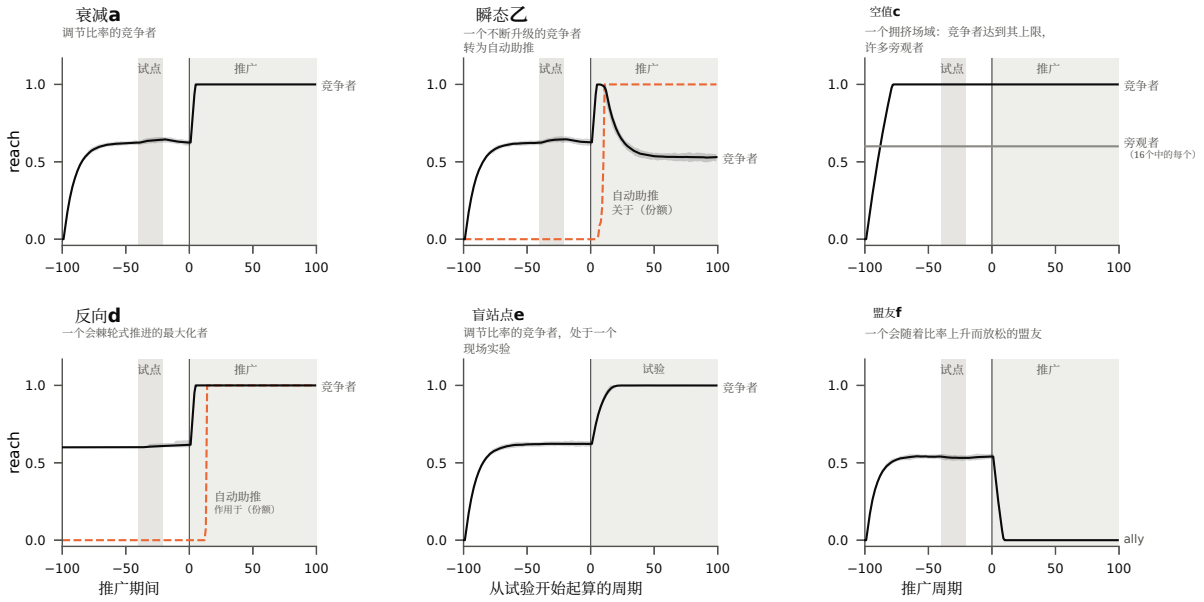


图 S2：在参与者能够响应的情况下，损失由其对比率的响应所产生；在他们无法响应时（c），损失则是静态的。 图 S1 的六个场域中，其他参与者在相应设置下的表现：**a**，衰减；**b**，瞬态；**c**，空值；**d**，反向；**e**，盲站点；**f**，盟友。每个面板绘出了其他每位参与者在整个运行期间的触达率，即从推广或试验开始起的周期 -100 至 99；试点和推广阶段以阴影标示（在现场实验**e**中没有试点，因此仅试验阶段以阴影标示）。同一行动者的副本合并为一条线取平均值（**c**中的 16 名旁观者）。虚线表示竞争者的自动助推处于开启状态的重复运行份额。**a**，竞争者在推广后的几个周期内达到极限。**b**，推广后约十个周期，自动助推在每次运行中都会开启，竞争者随之放松。**c**，竞争者在预热期早期达到极限；旁观者不响应比率（增益为 0），因此触达率保持在初始水平：没有行动者能够响应，全部损失都是静态的（图 S1）。**d**，竞争者在推广后的几个周期内达到极限，并在推广后约十四个周期自动助推开启时仍保持在该水平。**e**，竞争者在试验开始后约 20 个周期内达到极限。**f**，盟友在推广后的十个周期内放松至 0。每条触达率曲线表示重复样本均值（先在每次运行中对副本取平均值）；区间涵盖每个面板中 $n = 100$ 次重复运行的第10至第90百分位数。份额曲线没有区间，因为它是开/关指示变量的平均值。所有运行均为预先指定、完整，并基于一次干净提交。

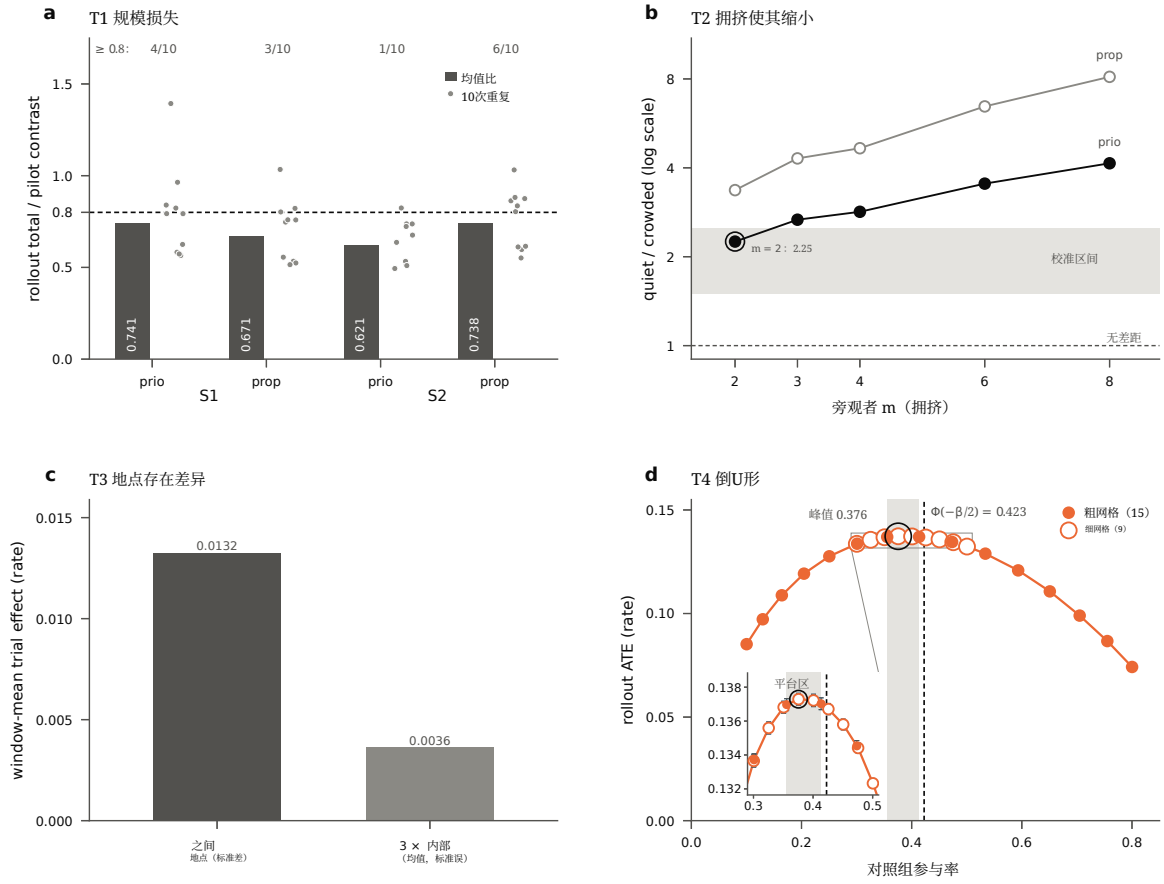


图 S3：模型在预测新的规律之前，先复现已知规律。入场券测试套件的四项测试。**a**，T1，规模损失。对于每个站点和注意力规则（S1，恒温器竞争者；S2，恒温器盟友〔情景名称，非补充信息章节〕），两者的响应增益均为 2；prio，优先级；prop，比例），每次重复运行的推广总量与试点对比之比（浅灰点）、重复均值之比（深灰条；测试统计量）以及 0.8 的界限（虚线）。顶行统计达到或超过 0.8 的运行次数；由于试点对比存在噪声，单次运行会越过该界限。**b**，T2，拥挤会使效应减弱。平静场域中的试验效应（概率单位尺度）与拥挤僵局场域中的试验效应之比（该场域中恒温器竞争者与盟友的目标无法同时实现，且二者都处于极限），随拥挤场域中的旁观者数量 m 变化（对数尺度，以 2 为底；优先级情形为实心，比例情形为空心），并标出校准区间 $[1/5, 2.5]$（对应优先级）和经校准的 $m = 2$（以圆圈标出）。**c**，T3，站点存在差异。六个站点的窗口均值试验效应的站点间标准差（深灰；统计量），与站点内平均标准误的三倍（浅灰；阈值）对比。**d**，T4，倒U形。推广平均处理效应随对照组参与率的变化：采用包含 15 个水平的粗网格（实心），以及从 0.30 到 0.50、包含 9 个水平的细网格（空心）；附有 ± 1 标准误条，峰值以圆圈标出并标注，平台区（与最大值相差不超过一个标准误的水平）以阴影表示，预言机 $\Phi(-\beta/2)$ 以虚线表示，即仅由阈值决定的峰值参与率。插图放大了参与率 0.29–0.51 的范围，在此处误差条和平台区清晰可见。该测试套件每种设置运行 $n = 10$ 次重复，次数预先固定；其他所有图均采用 100 次运行。所有运行均经过预先指定、完整完成，且来自同一个干净提交。

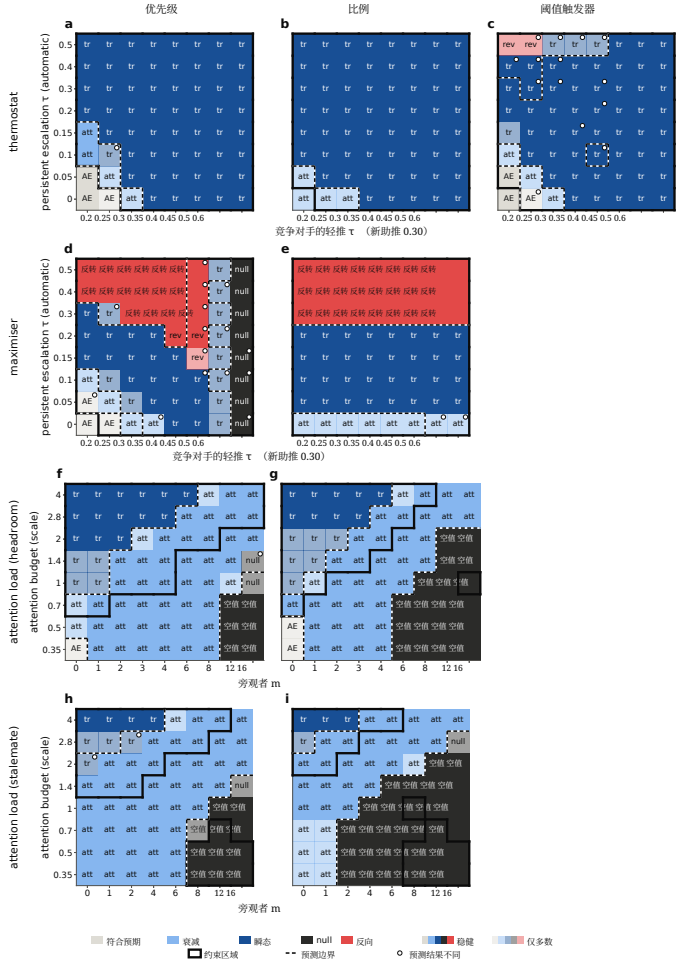


图 S4：改变注意力分配或响应规则后，各种结局依然存在；它们之间的边界会移动。九张相位图，与图3相同（每个单元格表示 $n = 100$ 次重复运行中的多数结果），按图的类型排列为行，按规则排列为列，因此同一行中的图可以逐单元格比较。列：优先级（注意力首先分配给最强的助推；默认规则）、比例分配（按比例共享注意力）和阈值触发器（只有当比率离开容差带时，竞争者才会采取行动）。第1行，恒温器竞争者：**a**，优先级；**b**，比例分配；**c**，阈值触发器。第2行，最大化者：**d**，优先级；**e**，比例分配。第3行，有余量的竞争者下的注意力负荷：**f**，优先级；**g**，比例分配。第4行，僵局场域中的注意力负荷（竞争者与盟友的目标无法同时实现，且双方都已达到极限）：**h**，优先级；**i**，比例分配。优先级图**a**、**d** 和**f** 对应图3a和3b，但不含图2中的标记。第1、2行绘制竞争者助推的强度 τ（新助推的强度为 0.30）与 τ（自动）的关系；第3、4行绘制旁观者数量 m 与注意力预算（尺度）的关系，与图3相同。第3行有九个 m 列，第4行有八个。实心和淡化单元格（稳健与否）的标记方式与图3相同。轮廓的含义与图3相同；在**e** 中，轮廓覆盖整张图。最大化者没有阈值触发变体，而阈值触发器对旁观者不起作用（增益为 0），因此第2至4行没有第三张图。在**h** 和**i**中，当预算为 1 或更低时，四项检查中的第一项由噪声决定（在 64 个单元格中，分别有 33 个和 38 个单元格的值处于 0.8 的两个标准误范围内），因此其中的轮廓区域会分裂成几块；图中按数据所示绘制。在**c**中，两个淡化单元格的结果发生反转（在最强自动助推下，对应最弱的两个竞争者；100 次运行中分别有 58 次和 52 次），虚线和空心点表示解析约化预测的边界，以及预测结果与模拟不同的单元格，与图3相同。一致的单元格：**a**，64 个中有 63 个；**b**，64 个中有 64 个；**c**，64 个中有 50 个（部分单元格可预测；在 14 个不一致的单元格中，预测效应均低于模拟值）；**d**，64 个中有 49 个（部分单元格可预测）；**e**，64 个中有 62 个；**f**，72 个中有 71 个；**g**，72 个中有 72 个；**h**，64 个中有 62 个；**i**，64 个中有 64 个。在 **c** 、 **d** 和 **e** 中，没有点的单元格仍可能存在高达 0.1 的差异：在**c** 和**e** 的每个单元格中，模拟效应都更大（ E 的差异中位数分别为 0.069 和 0.065）；在**d**中，竞争者强度不超过 0.40 时模拟效应更大（各列中位数为 0.050 至 0.067），而竞争者强度为 0.45 至 0.60 时，模拟效应大多更小。在探索的运行中，振荡、延迟、盲试验和反向位点结果均未表现出稳健性，因此未绘出。所有运行均预先指定、完整，且来自一次干净提交。

补充信息附录，S3节：解析约化

本节根据模型规则推导正文中讨论的各种结局，不进行任何拟合，并将结果与模拟图进行比较。推导遵循模拟器代码中实现的规则。其中的每个数量都来自某次运行的配置；运行之间的差异、试验比较中的噪声以及稳健性标记仍由模拟提供。

站点的比率对每个触达率都是线性的

个体 i 在周期 t 中采取行为的条件是：方法部分阈值方程中的潜在和加上标准正态噪声后为正，因此其概率为 $\Phi(\ell_{it})$，其中 ℓ_{it} 是未加噪声时的该总和。每个周期内，其他每个行动者的助推都会独立地以等于该行动者触达率的概率触达每个个体；新助推则会触达一个固定的处理组。总体固定，习惯更新关闭，旁观者保持恒定的触达率，而自动助推一旦开启，就会触达所有人。因此，任意群体的预期比率，是对其成员及其暴露状态下的 Φ取平均，并按各状态的概率加权。存在新助推和一个竞争者时，有四种状态和四种比率：

$R_{\emptyset}$ (未受助推) 也未受竞争者助推)　R_o (仅受新助推)　R_c (仅受竞争者助推)　R_{oc} (同时受两者助推)。

每个比率都是在注意力规则下，对总体成员的倾向、易感性、效价和注意力预算取平均所得；该规则会在触达个体的各项助推之间分配预算。记

$$E = R_o - R_{\emptyset}, \qquad D_0 = R_{\emptyset} - R_c, \qquad D_1 = R_o - R_{oc}$$

新助推在空场中的效应，以及竞争者对新助推未触达的人和已触达的人的拉动效应。当新助推覆盖站点的 p 份额、竞争者的触达率为 r时，预期站点比率为

$$S(r,p) = R_{\emptyset} + pE - r\,D(p), \qquad D(p) = (1-p)\,D_0 + p\,D_1.$$

该表达式在预期意义下是精确的；对于 20,000 个人，比率在周期之间的抽样标准差约为 0.003。对于 k 个响应型助推，比率关于它们的触达率是多重线性的：对这些助推的所有子集求和，取恰好该子集触达所有人时的比率，乘以该子集上的 $\prod_i r_i$，再乘以其余部分上的 $\prod_i(1-r_i)$，因此需要 2^k 个比率。旁观者的助推以其恒定的触达率进入每个比率：每个个体和助推只抽取一次，并由所有子集共享。新助推作为响应型助推进入，其触达率就是处理组占比。每个比率都根据模拟器自身的总体函数，对抽取的 200,000 个人取平均值计算得出；每种设置抽取一次。

引导型竞争者的静止触达率

恒温器竞争者设置 $r_t = \min\{1,\ \max\{0,\ r_{t-1} + \kappa(S^{\text{obs}}_{t-L} - G)\}\}$，增益为 κ，滞后为 $L = 2$，目标为 $G = R_{\emptyset} - \delta$，$\delta = 0.03$ （S1节）。在静息状态下，它观测到的比率等于其目标，除非其触达率受到截断，$S(r,\ p) = G$，因此其静止触达率为

$$r^*(p) = \min\left\{1, \max\left\{0,\ \frac{pE + \delta}{D(p)}\right\}\right\}.$$

推广前（$p = 0$），它在 $r_0 = \min\{1,\ \delta/D_0\}$；向所有人推广后（$p = 1$），在$r_1 = \min\{1,\ (E + \delta)/D_1\}$。推广前的余量为 $1 - \delta/D_0$。与模拟器中一样，触达率至少为 0.99 即达到极限。

四种情况

没有新助推时，竞争者保持在 r_0，比率为 $R_\emptyset - r_0 D_0$；有了新助推后，竞争者变为 r_1，比率变为 $R_\emptyset + E - r_1 D_1$。推广在静息状态下的效应是

$$T = E - (D_1 r_1 - D_0 r_0),$$

在没有升级的情况下，呈现四种形式之一（表 S1）。规模损失需要在推广前留出空间，$D_0 > \delta$（见下文的试点）。当竞争者对新助推所触达的人的拉动效应至少达到新助推的效应加上其目标的距离时，该效应就会被消除，$D_1 \geq E + \delta$。第四种情况要求 $D_1 - D_0 \geq E$，即竞争者对新助推所触达的人的拉力至少比对其余人的拉力强出新助推的全部效应。

表 S1：没有升级时，存在引导型竞争者的四种情况。 T/E 是静息状态下保留的预期效应份额。

Case	条件	T/E
前后都没有空间	$D_0 \leq \delta$ 和 $D_1 < E + \delta$	$1 - (D_1 - D_0)/E$，全部为静态
空间不足	$D_0 > \delta$ 和 $D_1 < E + \delta$	$1 - (D_1 - \delta)/E$
空间充足	$D_0 > \delta$ 和 $D_1 \geq E + \delta$	0
前面没有空格，后面有空格	$D_0 \leq \delta$ 和 $D_1 \geq E + \delta$	$(D_0 - \delta)/E \leq 0$

图 2a 的场域中，约简得到 $E = 0.0555$, $R_\emptyset = 0.3090$, $D_0 = 0.0484$ 和$D_1 = 0.0509$：推广前，竞争者的触达率为 0.620；这是一个空间会耗尽的情形；推广后保留了 E 的 0.623（模拟值为 0.63）。

损失的组成部分

冻结世界将竞争者维持在 r_0，此时新助推的效应为 $E - r_0(D_1 - D_0)$。因此，静态部分和反应部分为

$$\text{static} = -\frac{(D_1 - D_0)\, r_0}{E}, \qquad \text{reaction} = -\frac{D_1(r_1 - r_0)}{E},$$

以及 $1 + \text{静态} + \text{反应} = T/E$。当竞争者对处理组和未处理组人群的拉力大致相同时，静态部分很小，$D_1 \approx D_0$：如模拟结果所示，在图 2a 的场域中，它是 -0.028。在图 S1 的六行中，预测部分与模拟部分的差异最大约为 0.06，出现在反向行的反应部分；其他行的差异均小于 0.004。

试点

在站点 2% 的范围内开展试点，不会改变竞争者的影响范围。其对比为 $E - r_0(D_1 - D_0)$，接近 E，因为 $D_1 \approx D_0$。如果推广前没有余地（$r_0 = 1$），那么在表 S1 的第一种情况下，试点的对比等于推广的效应；在第四种情况下则不大于推广的效应。因此，若要出现规模损失，就需要 $D_0 > \delta$。

盲站点及其极限

在一项关于份额 p的现场实验中，竞争者在 $r^*(p)$处于静息状态。处理组比率为 $R_\emptyset + E - r^* D_1$，对照组比率为 $R_\emptyset - r^* D_0$，因此

$$\text{trial} = E - r^*(p)\,(D_1 - D_0), \qquad \text{control drift} = -\big(r^*(p) - r_0\big)\, D_0,$$
$$\text{site change} = pE - r^*(p)\, D(p) + r_0 D_0.$$

当 $r^{*}(p) < 1$ (以及 $r_0 < 1$)时，站点变化为零，而 S1节的身份给出了试验/总效应 $= 1/(1-p)$。竞争者在 $pE + \delta = D(p)$处达到其极限，在

$$p^{*} = \frac{D_0 - \delta}{E + D_0 - D_1},$$

当 $D_0 > \delta$ 和 $E + D_0 - D_1 > 0$时具有意义。对于图 4b 的设置，$p^{*} = 0.347$ 来自对 200,000 人进行的一次抽样。在 $p = 0.4$ 竞争者达到其极限时，试验对比为每位接受治疗者预期效应的 0.954 倍，对照组漂移 -0.018，站点变化为 $+0.003$。模拟试验遵循 $1/(1-p)$，在达到 $p = 0.3$ 之前的偏差不超过 1.3%，并在 $p = 0.4$处低于该值 8.9%。

升级

带有自动助推的恒温器会在达到其极限，且在运行的任何阶段中所观测到的比率连续十个周期高于其目标时升级。推广后，情况为 $D_1 < E + \delta$。推广前，情况为 $D_0 < \delta$：一个弱到无法达到其目标的竞争者会在预热期内升级，无论是否有新助推。约简在两个阶段都应用该规则。自动助推启用后，四种速率会用它重新计算（写作带上标 a），而竞争者会放松到

$$r^{**} = \frac{R_o^{a} - G}{D_1^{a}}.$$

如果 $0 \le r^{**} \le 1$ 效应以零结束，则对应图2b中被消除的效应。如果 $r^{**} > 1$ 竞争者保持在其极限，仍会有一些效应残留。如果 $r^{**} < 0$ 仅自动助推就能使比率保持在目标以下，且效应以负值结束，则为$(R_o^{a} - G)/E$。

棘轮式竞争者

最大化者的触达率在第59个周期前一直为0.60。以预期比率作为输入后，其基准值从该周期的比率开始，$q = R_{\emptyset} - 0.60\,D_0$。推广后，只要它观察到的比率高于基准值0.01以上，就会提高触达率，而且从不降低触达率。如果其触达率达到极限时的比率 R_{oc}，比 q高出0.01以上，它就会升级，效应最终为$(R_{oc}^{a} - q)/E$；自动助推足够强时，效应会变为负值。如果比率在达到极限前进入距 q 0.01以内的范围，竞争者就会停在那里，并留下一个较小的正效应，约为 $0.01/E$。这些预测采用该路径，而该路径运行模拟器自身的最大化者规则。

基准值是从第 60 个周期起观测速率的运行最小值。每个周期约 0.003 的抽样噪声会使运行最小值低于其预期值，因此模拟的基准值会低于 q。输入预期速率的路径（即用于每张图的无噪声预测）会使基准值保持在 q，因此可能会错误判断竞争者停止的位置，以及停止与升级之间的分界线。对于这个竞争者，在比较之前还确定了第二个预测。每个周期，竞争者的基准值都会降低至其起点减去$\sigma\, m(n)$，其中 σ 是其读取的速率的抽样标准差（20,000 人），n 是截至目前读取速率在无噪声运行最小值的 σ 范围内的周期数，而 $m(n)$是 n 个标准正态抽样的预期最大值。该预测仅适用于基准值确为运行最小值的情况。它与无噪声预测并列报告，不纳入比较所用的规则。

加入这一项后，优先级共享下最大化者的图中有64个单元格，其中53个一致；不加入这一项时则有49个。在竞争者强度为0.45和0.60时，除强度为0.45时有一个单元格以外，其余单元格均一致；该单元格的类别一致，但大小不同（-1.27预测，-1.17模拟）。强度为0.50时，带有自动助推的七个单元格全部不一致（三个预测为反向，而模拟为瞬态；三个预测为空值，而模拟为瞬态；还有一个仅大小不同）；不加入这一项时则有三个单元格不一致。在比例共享下，以及图2d后方的线上，两种预测对应的单元格相同。

盟友

一位盟友将调节至 $R_{\emptyset}+\delta$。令 U_0 和 U_1 分别表示新助推未触达和触达的人所获得的提升（$U=-D$）。推广前，其静息状态为 $\min\{1,\delta/U_0\}$；推广后为 $\min\{1,\max\{0,(\delta-E)/U_1\}\}$。因此，当 $E\geq\delta$ 时，它会完全退出，此时保留的 E 份额为 $1-\delta/E$。各部分遵循上文的公式，其中 $D=-U$。预测的总效应为 E 的 0.46，与模拟结果一致（图 S1）。

两个反应参与者：僵局场域

有竞争者和盟友时，站点比率分别与各自的触达率呈线性关系，因此需要八个暴露率，而静息状态则由两个有界积分器联立求解。当双方触及各自的极限时，双方触达率均达到最大值时的比率介于各自的目标之间。推广后，新助推使比率超过盟友的目标（如注意力充足时），盟友便会退出。加入旁观者并改变注意力预算后，如注意力负荷图所示，比率会成为旁观者数量和预算尺度的函数，代数关系仍然成立：当 $D_0=\delta$ 时，竞争者便失去回应空间；当 T/E 降至 0.25 时，空值结局便开始出现。

其他规则

比例共享只会改变比率中的注意力权重，其他方面均不变。阈值触发器只有在比率与目标相差超过 0.01 时才会调整触达率，因此它会停在该区间边缘；实际上，$G+0.01$ 是从上方趋近，而只有当其极限处的比率仍高于目标 0.01 以上时，它才会升级，即 $D_1<E+\delta-0.01$。

时间进程与稳定性界限

模拟器根据 $t-L$ 处观察到的比率，在形成周期t 的比率之前，设定周期 t 的触达率。将推广周期计为 0，周期 0、.....、$L-1$ 保持 r_0，而从 $t=L$ 起

$$r_t=\min\{1,\max\{0,\ r_{t-1}+\kappa\,(E+\delta-D_1 r_{t-L})\}\},\qquad \mathrm{effect}_t=E-D_1 r_t+D_0 r_0.$$

按照预期速率迭代，效应在推广后分别经过 6、4、2、2 个周期，降至 E 11的四分之三以下；对应的增益依次为 0.5、1、2、5 和 10，与模拟结果一致（图4d）。

远离极限时，触达率偏离静息状态的幅度 x_t 遵循 $x_t=x_{t-1}-k\,x_{t-L}$，其中$k=\kappa D_1$。其特征方程为 $z^L-z^{L-1}+k=0$，因此 $k=z^{L-1}\,(1-z)$。在单位圆上，$z=e^{i\theta}$ 且 $1-z=2\sin(\theta/2)\,e^{i(\theta-\pi)/2}$，因此当 $(L-1)\theta+(\theta-\pi)/2=0$ 时，k 为实数且为正，即 $\theta=\pi/(2L-1)$，并且

$$k=2\sin\!\Big(\frac{\pi}{2(2L-1)}\Big),$$

第一个使根离开单位圆盘的 $k>0$；对于较小的 $k>0$，所有根都位于单位圆盘内。一个周期的滞后值为 2，而模型滞后两个周期时为 1：当增益乘以拉力小于 1 时，响应会逐渐消失。在探索的运行中，最大值为 0.51，此时增益为 10。

比较之前已知的信息

可行性检查。 在构建约简之前，项目外部的一个脚本（未保留）按照 S1节的描述，从抽取的 200 万人中进行蒙特卡罗计算，得出了四种速率，并将上述闭式解应用于草稿中已有模拟值的设置（表 S2）。这些比较是在已知模拟数值的情况下进行的。该约简没有自由参数，因此没有需要调节的内容，但这些行属于先验知识。盟友以及现场实验在 $p=0.4$ 处的对比和对照组漂移并未计算，尽管它们可以根据这些行通过算术运算得出；没有注意负荷单元，没有比例分摊单元，也没有

整体相位图被计算出来。检查的滞后时序比模拟早一个周期出现交叉点；比较中使用的约简，其时序取自模拟器的代码。

表 S2：可行性检查是在构建约简之前，并在已知模拟值的情况下进行的。竞争者强度指竞争者助推的强度（新助推的强度为 0.30）；“自动”指其自动助推的强度。

数量	可行性检查	模拟（草稿）
$R_\emptyset$； E	0.309；0.0553	0.309；0.055
推广前竞争者的影响范围（强度 0.30）	0.623	0.626
保留效应，竞争者强度 0.20 / 0.30 / 0.35 / ≥ 0.45	1.10 / 0.625 / 0.39 / 0.00	1.10 / 0.63 / 0.39 / 0.00
静态部分，强度 0.30	-0.029	关于 -0.03
升级后的触达率（自动 0.30）	0.526	0.53
反转，自动 0.30；0.10	-0.41；$+0.25$	-0.39（结束）； $+0.26$
引导型竞争者，自动 0.50	-0.21 （不足 -0.25 定义为反向）	默认规则下的瞬态规则；在 58 和 52 中反向阈值以下运行 100 次
盲站点， p^*	0.35（在 0.3 时保持，在 0.4 时达到极限）	在两个较弱的强度下触发保持在 0.3，到了 0.4 就失效
低于 $0.75\,E$ 所需的周期，增益 0.5 / 1 / 2 / 5 / 10	10 / 5 / 3 / 1 / 1	11 / 6 / 4 / 2 / 2

一次计划外的模拟器运行。 在构建过程中，违反了构建自身的规则，模拟器在图2b的设置下运行了一次（引导型竞争者强度为 0.30，增益为 2，自动助推为 0.30），使用 20,000 人、8 次运行和随机种子 3。该竞争者的触达率与路径在三个周期进行了比较：99（路径为 0.623；运行结果为 0.608–0.648）、120（0.731；0.66–0.76）和 199（0.526；0.50–0.56）；到周期 120 时，两者的自动助推都已开启。根据实现者的陈述，这次运行是在代码及其测试编写完成且通过之后进行的，之后没有更改任何代码或测试值。没有查看过论文图表背后的运行数据表。该设置是图3a中引导型竞争者图的一个单元格。这个单元格和其他所有单元格一样计入统计；将其排除后，该图仍有 63 个单元格中的 62 个一致（98.4%，不变），而该单元格本身也一致。

谁知道了什么，以及何时知道。模拟图是在约简构建之前绘制的，作者当时也已知晓这些图。比较规则及其应用方式（比较什么、单元格的模拟侧、无需舍入容差的一致性、最大化者的第二个预测值、人数）在预测任何图之前就已确定并提交，且约简的代码已冻结。约简代码的实现者无法接触运行输出，唯一例外是上文所述的情况。在首次计算预测值与进行比较之间，对较早运行的类别规则版本的读取方式作出了一项修正：运行 1、2 和 2b 的清单是在类别规则修订之前生成的，其中没有版本信息；现在改为从每次运行旁的重新分类记录中读取，正如图表代码已经执行的那样。十张图以及两张有限尺寸地图的预测文件在修正前后完全相同（其哈希值随数据提供）。比较计算完成后未作任何更改。

比较

当预测类别等于模拟多数类别，且类别规则读取的每个大小相差不超过 0.1 时，该单元格即为一致：在先试点后推广地图中，窗口平均效应超过 E ；在现场实验地图中，站点变化超过其预期变化，且试验对比超过预期效应。一张图在无噪声预测下至少有 90% 的单元格一致时，即为通过，

表 S3 给出了十张不同的图以及另外七次运行逐点比较的结果，供参考，也包括规则之外的比较。在十张图中，664 个单元格有 624 个一致（94.0%）；这是描述性总数，并非规则所用的统计量。

表 S3：约简与模拟的比较。差异在于每个类别规则读取的大小（ E 的份额；在现场实验地图中，取站点差异和试验差异中较大的一个）。对规则之外的各点，逐点比较这些线条。

图或线条	绘制于	单元格	一致	中位数	最大值	结果
引导型竞争者，优先级	图 3a 左图；图 S4a	64	63	0.0002	0.017	通过
棘轮式竞争者，优先级	图3a 右侧；图S4d	64	49	0.060	1.336	部分
注意力负荷，先试点再推广，优先级	图3b；图S4f	72	71	0.002	0.008	通过
注意力负荷，现场实验，优先级	图3c	72	67	0.008	0.038	通过
引导型竞争者，比例	图S4b	64	64	0.0004	0.008	通过
引导型竞争者，阈值触发器	图 S4c	64	50	0.069	0.203	部分
棘轮式竞争者，比例	图 S4e	64	62	0.065	0.090	通过
注意力负荷，先试点再推广，比例	图 S4g	72	72	0.0003	0.004	通过
僵局，优先级	图 S4h	64	62	0.001	0.011	通过
僵局，比例型	图 S4i	64	64	0.0004	0.009	通过
引导型竞争者，增益	图 2a	3	3	0.001	0.005	
升级型引导竞争者	图2b	3	3	0.000	0.003	
棘轮型竞争者，记忆	图 2d、4c	10	10	0.030	0.059	
引导型竞争者，场域实验	图2e	3	3	0.012	0.013	
处理组占比	图4a、b	6	6	0.009	0.020	
升级型竞争者，增益	图4d	5	5	0.000	0.000	
Ally	图 S1（盟友行）	3	3	0.002	0.002	

在通过的八张图谱中，11 个结果不一致的单元格仅类别不同。它们的大小差异最多为 0.011，只有比例共享条件下棘轮式竞争者图谱中的两个单元格例外：模拟结果为衰减，预测结果为瞬态（0.572 对 0.493 和 0.541 对 0.462，差异为 0.079）。有几个单元格的大小本身就处于类别边界：模拟值为 0.499、预测值为 0.502（瞬态或衰减）；0.248 对 0.250（空值或衰减）；以及试验对比为 0.749 对 0.760（盲站点或瞬态）。在优先级共享条件下的棘轮式竞争者图谱中，15 个结果不一致的单元格有 9 个位于竞争者强度 0.45 和 0.50 处。在 0.45 处，模拟效应为负（六种自动助推中最弱的一种为 -0.09 ，类别为瞬态；其余五种为 -0.25 至 -1.17 ，类别为反向），而无噪声路径在升级之前便停止（$+0.16$至 $+0.17$）。在 0.50 处，三个单元格的类别不同（模拟为瞬态，预测为空值），大小约为 0.10 对 0.13。0.60 处的三个单元格类别一致（空值），大小约为 -0.006 对 $+0.096$；另外三个则分别是 0.20、0.25 和 0.35 处的单个单元格。在阈值触发器条件下，14 个结果不一致的单元格中有 13 个带有自动助推；且在全部 14 个单元格中，预测效应均低于模拟效应，低 0.05 至 0.20；在这些单元格中，预测结果通过 -0.25 的情况下，类别为反向；而模拟结果除一个外均为瞬态。

在三张图谱上，差异的符号一致，包括一致的单元格。在阈值触发器图谱上，所有 64 个单元格中模拟效应都高于预测效应（中位数 $+0.069$；范围从 $+0.008$到 $+0.203$）；在比例共享条件下的棘轮式竞争者图谱上（通过），所有 64 个单元格也都是如此（中位数 $+0.065$；范围从 $+0.047$ 到 $+0.090$）。在优先级共享条件下的棘轮式竞争者图谱上，当竞争者强度为 0.20 至 0.40 时，模拟效应高于预测效应（列中位数 $+0.050$

至 $+0.067$），而在 0.45 至 0.60 时则大多低于预测效应（24 个单元格中有 21 个）。另外七张图谱上的差异没有共同的符号，其中位数大小至多为 0.008。因此，在这三张图谱上，一致的单元格误差在 0.1 以内，但未必完全吻合。表 S4 列出了无噪声预测下所有不一致的单元格，包含两种类别和两种大小。

表 S4：在无噪声预测下存在不一致的所有单元格（40 个单元格）。 图：引导型竞争者或棘轮式竞争者，或具有余量的竞争者对应的注意力负荷（“负载”），或僵局场域；注意力规则或竞争者响应规则。轴 1 和轴 2：竞争者的强度 τ 以及其自动助推的强度 τ（自动），用于竞争者图谱；旁观者数量 m 以及注意力预算（尺度），用于负荷图谱。大小是类别规则所读取的值： E上的窗口平均效应，或在现场实验中，地点变化与其预期变化的比值（站点），以及试验对比与预期效应的比值（试验）；差异取两者中较大者。每个图的行数：1、15、1、5、14、2 和 2；比例共享下的引导型竞争者、比例共享下的注意力负荷，以及比例共享下的僵局均无对应行。棘轮式竞争者的第二种有限规模预测的各行未排版；它们位于比较的生成报告（`docs/analysis/2026-11-06-analytic-reduction-comparison/comparison.md`）中，将与数据一同存档。

Map	轴 1	轴 2	模拟	预测	模拟 size	预测 size	差异
转向，优先级	0.25	0.1	瞬态	衰减	0.499	0.502	0.003
棘轮效应，优先级	0.35	0	衰减	瞬态	0.507	0.447	0.060
棘轮效应，优先级	0.6	0	null	null	−0.005	0.095	0.101
棘轮效应，优先级	0.2	0.05	符合预期	衰减	0.770	0.720	0.050
棘轮效应，优先级	0.45	0.1	瞬态	瞬态	−0.087	0.166	0.252
棘轮效应，优先级	0.5	0.1	瞬态	null	0.105	0.130	0.025
棘轮效应，优先级	0.6	0.1	null	null	−0.007	0.096	0.103
棘轮效应，优先级	0.45	0.15	反向	瞬态	−0.250	0.166	0.415
棘轮效应，优先级	0.6	0.15	null	null	−0.006	0.096	0.102
棘轮效应，优先级	0.45	0.2	反向	瞬态	−0.400	0.167	0.567
棘轮效应，优先级	0.5	0.2	瞬态	null	0.107	0.131	0.024
棘轮效应，优先级	0.25	0.3	瞬态	反向	−0.219	−0.276	0.057
棘轮效应，优先级	0.45	0.3	反向	瞬态	−0.692	0.161	0.852
棘轮效应，优先级	0.45	0.4	反向	瞬态	−0.940	0.165	1.104
棘轮效应，优先级	0.5	0.4	瞬态	null	0.103	0.130	0.027
棘轮效应，优先级	0.45	0.5	反向	瞬态	−1.169	0.167	1.336
Load，推广	16	1.4	null	衰减	0.248	0.250	0.002
Load，场域实验	16	1.4	null	衰减	站点 0.252，试验 0.243	站点 0.250，试验 0.250	0.007
Load，场域实验	2	2	瞬态	盲站点	站点 −0.001，试验 0.749	站点 0.000，试验 0.760	0.011

表 *S4*（续）

Map	轴 1	轴 2	模拟	预测	模拟 size	预测 size	差异
Load, 场域实验	3	2	瞬态	衰减	站点 −0.002，试验 0.694	站点 0.000，试验 0.693	0.002
Load, 场域实验	4	2.8	盲站点	瞬态	站点 0.001，试验 0.752	站点 0.000，试验 0.742	0.011
Load, 场域实验	8	4	瞬态	衰减	站点 0.003，试验 0.690	站点 0.000，试验 0.698	0.008
转向阈值触发器	0.25	0	符合预期	衰减	0.824	0.725	0.099
转向阈值触发器	0.4	0.1	瞬态	反向	−0.067	−0.263	0.197
转向阈值触发器	0.35	0.15	瞬态	瞬态	−0.086	−0.215	0.129
转向阈值触发器	0.4	0.2	瞬态	瞬态	−0.066	−0.198	0.132
转向阈值触发器	0.25	0.3	瞬态	反向	−0.111	−0.313	0.203
转向阈值触发器	0.3	0.3	瞬态	瞬态	−0.098	−0.207	0.109
转向，阈值触发器	0.4	0.3	瞬态	瞬态	−0.077	−0.206	0.129
转向，阈值触发器	0.2	0.4	瞬态	反向	−0.130	−0.279	0.149
转向阈值触发器	0.25	0.4	瞬态	反向	−0.116	−0.266	0.150
转向阈值触发器	0.3	0.4	瞬态	瞬态	−0.101	−0.244	0.143
转向阈值触发器	0.25	0.5	反向	反向	−0.250	−0.350	0.100
转向阈值触发器	0.3	0.5	瞬态	反向	−0.245	−0.318	0.073
转向，阈值触发器	0.35	0.5	瞬态	反向	−0.243	−0.292	0.049
转向阈值触发器	0.4	0.5	瞬态	反向	−0.240	−0.297	0.057
棘轮效应，比例	0.5	0	衰减	瞬态	0.572	0.493	0.079
棘轮效应，比例	0.6	0	衰减	瞬态	0.541	0.462	0.079
僵局，优先级	0	2	瞬态	衰减	0.394	0.395	0.001
僵局，优先级	2	2.8	瞬态	衰减	0.390	0.390	0.000

补充信息附录，S4节：敏感性扫描

图3和图 S4 的图谱每次围绕一个校准设置，改变场域的两个参数。本节探讨参数空间中每种结局占据多大范围、由哪些参数决定，以及在随机设置下，解析约化所用来判读结局的无噪声路径是否与模拟器一致。样本、统计数据和五项预期均在抽取样本前，由一项获批的设计确定；下文将这些预期分别按已登记和实际结果报告，其中五项有四项未能实现。各份额是基于下文所列范围计算的，这些范围只有一部分受到校准的约束；它们并非真实场域中的频率。

样本

十二种组合：竞争者的响应规则（恒温器，会调节至目标；最大化者，会逐步棘轮式推进；阈值触发器，仅在比率偏离目标超过0.01时才进行调节）、注意力规则（优先级；比例）以及设计（先试点再推广；现场实验）。每种组合中都有一名竞争者、没有盟友，以及余量场中的旁观者（增益为0，触达率为0.6）。每种组合均采用拉丁超立方抽样，抽取4,096个设置（随机种子为20261108加上该组合的索引），涵盖九个参数；在现场实验中则涵盖十个参数：新助推的强度为0.10–0.60；竞争者的强度为0.10–0.60；其增益为0.5–10，呈对数均匀分布（对于阈值触发器，则为每个周期固定步长0.01–0.10）；其目标比无行动者基线低0至0.10；其滞后为1–5个周期；其自动助推的强度为0–0.50；注意力预算尺度为0.35–4.0，呈对数均匀分布；旁观者数量为0–16；对照组参与率为0.10–0.80，通过对倾向性混合的两个组件各进行一次偏移来设定；在现场实验中，处理组占比为0.10–0.60。最大化者从不读取其目标，因此该参数在其四种组合中不起作用，并作为噪声底限。其余所有参数均采用S1节中的值。样本中包含无法达到目标的弱势竞争者。图谱排除了这种情况：在运行 3 之前作出的一项裁决，将竞争者强度为0.15的情况从图谱中剔除，因为在这种情况下，恒温器会在预热期达到极限，并在试点前升级（S7 节，条目 §10）。总计49,152个设置。

计算了什么

每个设置都是在 200,000 人条件下对无噪声路径的一次运行；所有设置使用相同的人和抽样结果，因此设置之间的差异仅在于参数。该路径依据 S1 节的规则给出类别，并判定设置是否位于轮廓内（依据路径自身序列上的规模损失检验和幅度界限；现场实验则在相同设置下，依据先试点后推广路径进行首次检验），以及竞争者的自动助推是否已开启。四项统计量均事先确定，并在预注册设计中命名为 S1–S4（而非补充信息章节）：$S1$，每种结局在所有设置中的占比（将排除的（$|E| < 0.02$）和未分类设置计入），以及在轮廓内已分类设置中的占比（表 S5）；$S2$，对于每个参数，按其抽取值的最低三分位与最高三分位（按排名划分三分位，同值时按设置打破平局）对应的结局分布之间的全变差距离；统计范围为已分类设置，并据此对参数排名（图 S5）；$S3$，在现场实验中，按处理组占比的三分位统计盲点占比，以及试验/总效应落在 $1/(1-p)$ 的 10% 以内的设置占比；$S4$，检查。从每种组合中随机抽取 24 个设置作为样本（总计 288 个；随机种子为 20261108），并在云端 T4 GPU 上各模拟 10 次，每次 20,000 人。若模拟中的多数类别与路径给出的类别相同，且大小之差不超过 0.1，则该设置判为一致；在现场实验中，大小取两者中的较大值，而仅在一侧被排除的设置判为不一致（约简的单元规则；表 S6）。在 288 个模拟设置中，有 20 个设置的每次运行都被排除，其存储的类别表缺少大小列；比较代码在补入空值列之前无法处理这些表。补入操作未改变任何值或行（模拟器未重新运行；两侧均排除的设置判为一致）。

已登记的预期与实际发现

五项预期均在抽取样本前撰写并签字批准。签字批准后、抽取样本前，又对其表述作了五处修订。E1 被限定为自动助推从未开启的设置范围，即约简表格所涵盖的范围；最初的表述未明确这一点。只有当自动助推的强度大于零时，才算开启，因为强度为零时不会产生推动作用。现场实验中的反转包括反向站点。E3 中的“每个现场参数”指竞争者的五个参数、注意力预算和旁观者，因此参与率和处理组占比不纳入其中。E1 的占比按四种引导组合汇总读取，同时列出各组合的占比。最后四处修订在构建过程中确定，并由作者在运行后批准。这些修订均未改变任何预期的判定，但其中两项结果依赖于单个组合。若不采用第一处修订，E1 在汇总结果和全部四种引导组合中都会不成立，因为升级设置会被计入（如下）：在优先级共享的推广中，按修订后的 E1，3,747 个设置中有 930 个违反规则。所有抽取设置的自动助推强度均为正，因此第二处修订不影响任何设置。若不采用第三处修订，E2 在引导推广和阈值触发推广实施中都不成立，因为这两类推广不读取站点；因此，在测试的八种组合中，有四种不成立，而非八种都不成立。参与率和处理组占比对结局的影响最多分别为 0.20 和 0.28，低于每种组合中表现最好的现场参数（至少为 0.40）。单独考察时，E1 在四种引导组合中的三种不成立。看到结果后，样本、范围、统计量和规则均未作更改。五项预期中有一项得到满足。

E1，约简表的全局化版本 （引导型竞争者）。已登记：在自动助推从未启动的分类设置中，符合预期仅出现在竞争者在推广前没有空间的情况下，而空值仅出现在竞争者在推广后有空间的情况下，违反规则的此类设置不超过 2%。实际发现：**未满足**。汇总四种引导组合后，15.3% 的设置违反规则（2,277 个中有 349 个）；按组合分别为：推广中的优先级共享为 1.0%（489 个中有 5 个；满足），现场实验中的优先级为 11.9%（945 个中有 112 个），推广中的比例共享为 16.1%（205 个中有 33 个），现场实验中的比例共享为 31.2%（638 个中有 199 个）。违规情况分为两类：推广前有空间时却符合预期，仅见于现场实验（66 和 63 个设置）；推广后没有空间时却为零值，则见于每种组合（分别为 5、46、33 和 136 个设置）。表格完成后才作出的解读，并非预先登记：该表是针对推广的总效应推导的；在现场实验中，当竞争者有空间时，若试验报告了预期效应，这就属于盲站机制，因此 E1 本应仅适用于推广；此外，38 个推广违规设置在推广后均没有空间，且都有 3–16 个旁观者。旁观者负载下的效应与没有旁观者时的效应之比，中位数为 0.21；其中 33 个属于比例共享，因此这类设置反映的是拥挤，而非竞争者是否有空间。未登记、也不作此主张的，是将 E1 限定于没有旁观者的推广设置。升级设置不属于该表的适用范围，因为升级情形有其自身的闭式解；这些设置按已登记的方式报告，但未进行检验：在 3,258、2,815、3,562 和 3,115 个分类设置中，分别有 925、632、1,936 和 1,562 个违反该表规则，几乎全都是推广后没有空间时出现空值。

E2，反转发生在此处。 已登记：面对棘轮式竞争者时会发生反转；面对引导型竞争者时，在已分类的设置中不会发生反转；面对阈值触发器时，反转仅限于自动助推强度最高的三分之一范围内（图 S4c 的拐角）。结果：**八种测试组合均未满足**。面对引导型竞争者时，在已分类设置中，助推效应分别有 3.2%、6.2%、5.4% 和 5.5% 发生反转（优先级推广、优先级现场实验、比例推广、比例现场实验中分别为 3,747 项中的 119 项、3,764 项中的 235 项、3,767 项中的 204 项和 3,756 项中的 207 项）。面对阈值触发器时，发生反转的设置中，仅有 56%、41%、54% 和 42% 位于自动助推强度最高的三分之一范围内（分别为 225 项中的 126 项、357 项中的 148 项、326 项中的 176 项和 313 项中的 133 项）。棘轮式竞争者的结果未报告检验；其助推效应在已分类设置中分别有 24.3%、17.1%、26.9% 和 5.5% 发生反转。在引导型推广中，323 个反转案例里有 322 个开启了自动助推，强度介于 0.14 至 0.50 之间（中位数为 0.39）；在三种竞争者的所有推广中，98.7% 的反转案例开启了自动助推（2,805 项中的 2,768 项）。在现场实验中，自动助推的开启频率较低：面对引导型竞争者时，442 个发生反转的站点中有 399 个开启；面对三种竞争者时，1,963 个站点中有 1,697 个（86.4%）开启。

一种解释（未经登记）：当自动助推的力度超过撤去竞争者主动助推所能弥补的程度时，由于触达率不可能低于零，其比率最终会低于推广前的水平。图谱中未显示这一点，因为在相应的竞争者强度和增益下，撤去主动助推恰好抵消了自动推动；阈值触发器的拐角以及图 2d 中的棘轮式竞争者，都是这种情况。检查结果证实了路径预测的反转：在路径预测会发生反转的六种引导设置中（五种反转，一处反向站点），模拟均显示效应反转；六种设置的类别均吻合，其中四种的大小差异在 0.1 以内（另外两种相差 0.59 和 0.14）。在所有组合中，25 个预测反转中有 24 个、11 个预测反向站点中有 10 个通过模拟得到。

E3，决定结果的因素。已登记：新助推的强度在任何组合中都未排名第一；这里的“第一”是指其数值比每个场参数的数值都高出 0.03 以上；面对引导型竞争者时，排名第一的场参数是其强度、自动助推或目标。结果：在优先级共享下，**未满足**；在比例共享下，**已满足**。在优先级共享下，新助推的强度在全部六种组合中均为 0.54–0.63，而最佳场参数为 0.40–0.52（差值为 0.07–0.15）。在比例共享下，新助推的强度为 0.07–0.21，而旁观者数量在全部六种组合中均排名第一，为 0.58–0.74；面对引导型竞争者时，新助推的强度排名第七，在推广中为九项中的第七，在现场实验中为十项中的第七。第二部分在四种引导组合中的一种中得到满足（优先级、推广：竞争者的强度）；在另外三种组合中，旁观者排名第一。棘轮式竞争者的惰性目标为 0.01–0.04，即噪声底限。一项未登记的解读：在优先级共享下，较强的助推会首先吸引个体的注意力，因此新助推相对于竞争者的强度决定了个体听到哪一个；在比例共享下则不会。

E4，跨田间的盲站点。已登记：随着处理组占比上升，盲点占比下降；对于处理组占比最低三分位中的大多数设置以及最高三分位中的少数设置，盲点恒等式——试验/总效应在 $1/(1-p)$ 的10%以内——成立。结果：**未达到**。在六种现场实验组合中的一种（引导型竞争者、比例共享）中，盲点占比从最低三分位到最高三分位由0.11降至0.08；其余组合中分别为0.15和0.15、0.10和0.15、0.04和0.07、0.10和0.14、0.06和0.07。对于最低三分位，恒等式成立的比例分别为0.45、0.34、0.11、0.05、0.23和0.19，从未达到多数；在全部六种组合中，最高三分位中成立的情况更少。在随机场中，盲站点在处理组占比的每个三分位中，均出现在分类后的现场实验设置的4%–16%里。一项未登记的解读：图 4b中的恒等式是校准情境的属性，在该情境中，竞争者恰好保持站点的比率不变。

E5，检查。已登记：至少 90% 的引导设置一致；棘轮效应与阈值触发组合按实际结果报告，预期低于标准，并预期呈现约化映射的单侧差异（模拟值高于预测值）。结果：**达标**。引导设置在 24 项中分别有 24、22、22 和 22 项一致，汇总为 96 项中的 90 项（93.75%）。棘轮效应组合的一致率为 37.5–79.2%（96 项中的 55 项），符合预期，低于标准。阈值触发组合的一致率为 75.0–91.7%（96 项中的 79 项，82.3%）：汇总结果及三个组合低于标准，但在一次推广中，比例共享设置下高于标准（24 项中的 22 项），因此已登记的子预期“低于标准”未能成立。单侧差异也未再次出现，因此这一子预期同样未能成立（表 S6）；按已登记的标准，E5 达标。六项转向分歧中，两项类别不同（一处盲站点被模拟为衰减；一次运行五比五平局时，空值被模拟为反向站点），另外四项仅大小不同，差异为 0.14至0.59。

决定结果的因素：抽样结果

图 S5 展示了推广过程中，在优先级共享下竞争者的三种响应规则对应的 S2 抽样结果：各参数最低、中间和最高三分位中的结局占比。其余九种组合见表 S5（份额），以及包含数据的生成报告（距离）。

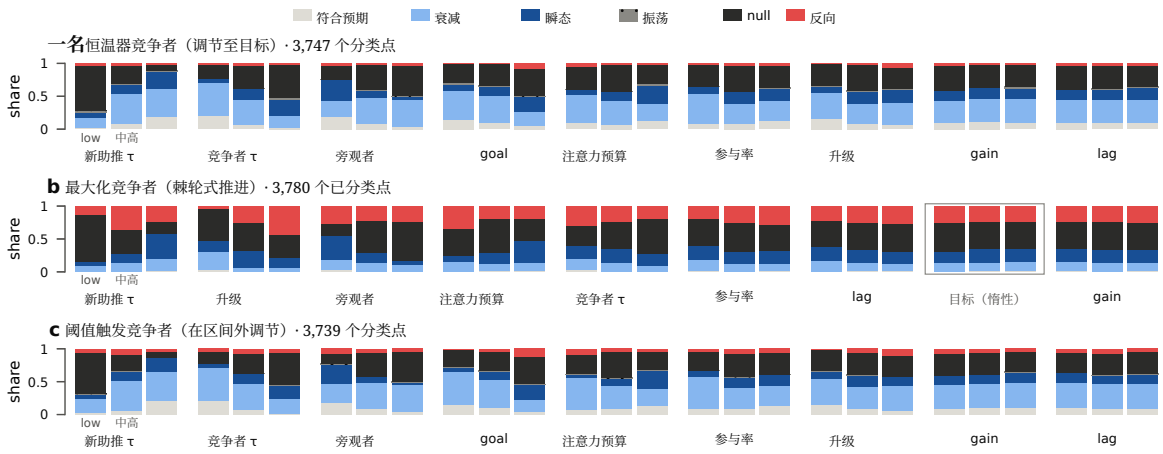


图 S5：在默认注意力规则下推广时，结局受助推自身强度的影响最大，其次是竞争者的强度（对于棘轮式竞争者，即其自动助推）、旁观者、竞争者的目标（若其有目标），以及注意力预算；参与率、竞争者的增益或滞后对结局的影响则很小。每个面板对应竞争者的一种响应规则，先试点再推广，优先级共享：**a**，恒温器（调节至目标），3,747 个已分类设置；**b**，最大化者（棘轮式推进），3,780 个；**c**，阈值触发器（在容差带之外进行调节），3,739 个；每种设置均抽取 4,096 次。对于每个参数，三个堆叠条形图分别显示其抽取值最低（低）、中间（中）和最高（高）三分位中，已分类设置的结局占比；排除的设置和未分类设置不计入。各组按最低与最高三分位之间的结局占比变化幅度排序，幅度最大者排在最前：新助推的强度分别为 0.63、0.56 和 0.56（**a**、**b**、**c**）；注意力预算分别为 0.28、0.26 和 0.27；参与率分别为 0.16、0.09 和 0.17；竞争者的增益分别为 0.07、0.02 和 0.08；其滞后分别为 0.04、0.08 和 0.03。第一句所述各组之间还存在三项变化：最大化者的强度（0.22，**b**），以及自动助推对恒温器（0.15，**a**）和阈值触发器（0.11，**c**）的强度。最大化者从不读取其目标，因此在 **b** 中，目标不起作用（以方框标出），其离散度为 0.04，是三分位组的噪声底限；**b** 中没有振荡设置。颜色同图 2 和图 3。结局根据 200,000 人时解析约化的无噪声路径读取；所有设置使用相同的抽样结果。

表格

表 S5：每种组合的结局占比（统计量 S1），以百分比表示。“全部”：对于该组合的 4,096 种设置，将排除的设置（$|E| < 0.02$）和未分类设置作为类别，并在其旁列出轮廓（区域）内的份额以及预热期尚未稳定的份额（未稳定）。“区域”：对于轮廓内的分类设置。AE，符合预期；att，衰减；tr，瞬态；osc，振荡；rev，反向；RS，反向站点；BS，盲站点（仅限现场实验；现场实验的反转在站点层面读取，因此其 rev 列为 0）。盲法试验（1 种设置）和延迟（2 种设置）不予显示。推广：先试点再推广；现场实验：现场实验。根据定义，在轮廓内，推广中的符合预期项均属于排除项（结果）。根据 `shares.csv` 正式运行排版。

组合	范围	AE	att	瞬态	osc	空值	rev	RS	盲点	excl.	uncl.	区域 unsettled
引导，优先级，推广	全部	9.0	32.1	14.6	0.9	31.9	2.9	–	–	8.4	0.1	49.9	1.7
	区域	0.9	14.1	30.2	1.9	46.9	6.0	–	–	–	–	–	–
引导，优先级，场域 exp.	all	9.3	40.4		1.2	0.3 20.8	0.0	5.7	14.2	7.9	0.2	49.2	1.9
	区域	2.8	31.7		2.4	0.6 23.9	0.0	10.6	28.0	–	–	–	–
转向、比例、推广	all	1.8	25.1		9.3	0.2 50.6	5.0	–	–	7.9	0.1	40.6	2.0
	区域	0.5	11.5	23.4	0.6	51.5 12.6		–	–	–	–	–	–
转向、比例、现场实验	all	2.9	34.3		0.4	0.2 40.4	0.0	5.1	8.5	8.1	0.2	40.5	1.9
	区域	3.0	34.9		1.0	0.5 28.7	0.0	10.5	21.4	–	–	–	–
棘轮效应，优先级，推广	all	0.9	12.3	18.1	0.0	38.6 22.4		–	–	7.7	0.0	71.4	0.0
	区域	0.5	11.4	25.4	0.0	31.2 31.5		–	–	–	–	–	–
棘轮效应，优先级，现场实验	all	6.5	38.8		0.7	0.0 18.3	0.0	15.8	12.1	7.9	0.0	70.1	0.0
	区域	8.3	49.1		1.0	0.0 2.1	0.0	22.6	16.8	–	–	–	–
棘轮效应，比例，推广	all	0.1	8.5		7.3	0.0 51.2 24.8		–	–	8.0	0.0	61.9	0.0
	区域	0.0	7.4	11.9	0.0	40.3 40.3		–	–	–	–	–	–
棘轮效应，比例，现场实验	all	4.6	37.8		0.3	0.0 37.8	0.0	5.0	6.0	8.3	0.3	62.3	0.0
	区域	7.4	54.1		0.5	0.0 20.1	0.0	8.1	9.7	–	–	–	–
阈值触发器，优先级、推广	all	9.0	33.8	12.9	0.4	29.6	5.5	–	–	8.3	0.4	44.0	4.8
	区域	0.9	15.3	29.9	1.0	40.2 12.7		–	–	–	–	–	–
阈值触发器，优先级，现场实验	all	11.7	40.5		0.6	0.0 19.1	0.0	8.7	11.1	8.1	0.2	43.4	5.6
	区域	5.1	36.2		1.3	0.1 14.4	0.0	18.0	24.9	–	–	–	–
阈值触发器，比例，推广	all	1.7	25.5		7.6	0.2 49.0	8.0	–	–	7.8	0.1	38.1	6.5
	区域	0.2	12.1	20.4	0.5	45.6 21.3		–	–	–	–	–	–
阈值触发器，比例，现场实验	all	3.6	36.0		0.3	0.0 38.3	0.0	7.6	6.0	8.0	0.1	38.5	6.7
	区域	5.6	37.1		0.9	0.1 24.0	0.0	16.4	15.9	–	–	–	–

表 S6：每种组合的模拟检验（统计量 S4）。每种组合从样本中随机抽取 24 种设置，每种设置均模拟运行 10 次，每次包含 20,000 人。当模拟多数类别与路径的多数类别相同，且两者大小相差不超过 0.1 时，视为一致（在现场实验中，取站点与试验的差异中较大者）。差异（diff.）为模拟值减去预测值，单位为类别规则所读取的大小（ E的份额）；“最大值”指最大的 |diff.|。90% 门槛仅为引导型竞争者预先登记；其余行按实际结果报告。根据正式运行的 `check_summary.csv`排版。

组合	一致	%	中位数 |diff.|	最大值	中位数差异
引导，优先级，推广	24 项，共 24 项	100.0	0.001	0.034	−0.0001
转向，优先级，现场实验	24 项，共 24 项	91.7	0.022	0.238	−0.0063
转向，比例，推广	24 项，共 24 项	91.7	0.001	0.590	−0.0006
转向、比例、现场实验	24 项中的 22 项	91.7	0.017	0.143	−0.0100
棘轮效应、优先级、推广	24 项中的 19 项	79.2	0.033	0.993	+0.0041
棘轮效应、优先级、现场实验	24 项中的 12 项	50.0	0.180	1.320	−0.1796
棘轮效应、比例、推广	24 个中的 15 个	62.5	0.051	2.268	+0.0234
棘轮效应、比例、现场实验	24 个中的 9 个	37.5	0.187	1.377	−0.1784
阈值触发器、优先级、推广	24 个中的 20 个	83.3	0.009	0.167	−0.0007
阈值触发器、优先级、现场实验	24 个中的 18 个	75.0	0.038	0.777	−0.0308
阈值触发器、比例、推广	24 个中的 22 个	91.7	0.013	0.144	−0.0005
阈值触发器、比例、现场实验	24 个中的 19 个	79.2	0.024	4.984	+0.0029
转向，全部四项（汇总）	96项中的 90项	93.8	0.007	0.590	−0.0006

补充信息附录，S5 节：ODD格式中的模型

本节复现了项目对 ODD格式（概述、设计概念、细节）中模型的描述 [1]。补充信息附录中复现的文档写于项目开展期间，并采用了项目当时使用的名称。F1–F4 对应图 1–4，ED1–ED4 对应图 S1–S4，SF1 对应图 S5；“我们的助推”指新助推；T1–T4 是四项已知规律检验；“规范”“简报”和“尖峰注释”是代码库中的设计文档，引用时标注了路径和章节。记录中的图号和分图标签，是条目登记时计划使用的名称（“扩展数据”是补充信息附图的旧称）；每幅已发表图背后的运行均列于 S1节。R1–R9 是确定模型规则的裁定，其中 R1–R8 见尖峰注释，R9 是修订后的结果类别规则；G1–G5 是设计文档中的目标。构建阶段对模型进行了测试，探索阶段则运行了已登记扫描。机器：M4，一台苹果笔记本电脑（设备“mps”）；Spark，一台英伟达 DGX Spark（“CUDA”）；Modal，一项使用 T4 GPU 的云服务。“cls_combined”是某次运行的合并结果类别列；“级别”是已保存运行的比率水平，为进行 Probit 对比而根据其清单重新运行；“孪生运行”指在相同设置下采用另一种设计的对应运行；T4 的“预言机”是解析峰值 $\Phi(-\beta/2)$；拥挤、僵局和余量系列是 S1节中的拥挤场域；“S1_竞争者”“S2_盟友”等是场景文件，而非补充信息章节。来源：`docs/nudgefield/odd-protocol.md`，最后于 2026-10-29 在提交 5e32b61 时修改；原样复现。

本说明以 `nudgefield` ODD格式（概述、设计概念、细节；Grimm等（2020），*JASSS* 23(2) 7）介绍 `module.function` 相对于 `code/src/nudgefield/`的模型。参数值及其来源见 `docs/nudgefield/parameters.md`，此处不再重复。“规范”是设计文档；“尖峰注释”是 `docs/analysis/2026-10-26-spike-note.md`，其第1节至第3节在任何运行之前确定了规则。

1. 概述

1.1 目的与模式

该模型探讨这样一种情况：一个为中性场域设计的行为干预（“我们的助推”）进入了一个由其他参与者的助推构成的场域，而这些助推的所有者会根据其观察到的行为率做出反应（规范第1节）。模型比较每次运行中的四种效应：设计者预期的效应（空场）、随机试验报告的效应、场域的静态效应（冻结场）以及真实总效应（场域可自由反应）。结果根据尖峰注释第1.3节中的预注册类别（符合预期、衰减、空值、反向、延迟、瞬态、振荡、盲法试验）以及规则R8的站点级类别（`reversed_site`、 `blind_site`）进行分类。

模型必须复现的模式（规格说明 §6； `code/tests/nudgefield/test_targets.py`）：从试点到推广的 T1 效果损失；安静组与拥挤组的概率单位比率约为 2（已校准，因此是约束而非预测；规范第12节）；跨地点离散度高于抽样误差；对照组参与率呈倒U形，在低于一半处达到峰值。

1.2 个体、状态变量与尺度

个体 (`people.Population`)，每次运行 `N` 。状态变量均为形状为 `(R, N)` 的张量：基线倾向 `a` （双组件正态混合）；注意力预算 `B` （对数正态）；易感性 `s_auto`、 `s_delib` （对数正态，均值为 1）； `a`、 `B` 和 `s_delib` 的三分位标签，在抽样时固定。对于每个个体和助推，均有一个效价 `v_ik` （幅度服从均值为 1 的对数正态分布，符号以 `flip_prob` 的概率翻转），首次使用时抽取，并按助推名称缓存。对于每个个体和旁观者，均有一个针对该旁观者自身行为的倾向概率。只有 `a` 会随时间变化，并且仅在习惯开启时变化。

助推 (`nudges.Nudge`)。名称、方向 $d \in \{+1, -1\}$、类型（`auto` 或 `delib`），以及每次运行的张量 `(R,)`：潜在强度 τ、注意力成本 `c` 和显著性（优先级规则据此进行排序的强度；除旁观者的助推外，该值等于 τ，见第3.3.3节）。触达率由所有者持有，而不是由助推持有。

参与者（`actors.base.ActorSpec`，标量描述； `actors.base.Actor`，每次运行的状态）。角色（`competitor`、`ally`、 `bystander`）、响应规则（来自注册表 `actors.RULES`）、一个审慎型助推、滞后、增益、目标偏移量以及其他规则参数。状态，在每个 `(R,)` 中：触达率、目标、恒温器先前的差距、最大化者的基准值、升级计数器、自动助推开关及其触发周期。我们自己的行动者没有规则：其助推的触达率由研究设计设定。

世界 (`world.World`)。一个总体（共享的，或在启用习惯时的副本）、参与者列表、我们的助推或无助推、设计及其处理组集合，以及记录的历史记录。一个世界是 `free` （参与者会作出反应）、 `frozen` （参与者

保持不变）或 empty （无行动者）。

尺度。一个周期为一周（规范第4节）。一次运行持续 T = 200 个周期。一次运行对应一个包含 N 人的站点（默认为20,000人；Spark目标为50,000人）； R 次运行构成张量的首个轴。没有空间，也没有网络：人们只通过参与者关注的聚合速率进行互动（规范第2节）。站点不是模拟器的一个维度；多站点比较会运行多个情景配置（计划决策；T3）。

1.3 流程概述与调度

每个世界在 world.World.step中一次推进一个周期，顺序如下（尖峰的顺序）：

1. **规则**。依次处理每个行动者（步骤1–3），其响应规则读取 t − lag 时的观测速率并设定其触达率。2. **升级**。规则R6更新行动者的计数器，并可能开启其自动助推。3. **暴露掩码**。当行动者 k 触达个体 i 时，会进行一次新的均匀抽样 u_ik < reach_k （每个周期重新抽取成员）；启用的自动助推会触达所有人；我们的助推会触达设计指定的处理组（designs.Design.our_mask）。4. **注意力**。分享规则为每个人和每项审慎型助推计算一个权重。5. **潜在价值**。 a + s_auto · Σ_auto d τ v + s_delib · Σ_delib w d τ v，针对所有触达该个体的助推。6. **行为**。 y = 1{latent + ε > 0}, ε ~ N(0, 1)。7. **记录**。站点比率；随机化周期中的处理组和对照组比率；按三分位组划分的比率；每个行动者的触达率；评估窗口内 y 的累计和。8. **自身行为**。每位旁观者自身行为的比率（仅限自由世界）。9. **观测**。每个行动者在本周期的观测速率：站点比率（旁观者则为其自身比率）加上 obs_noise_sd · ε_obs。10. **习惯**。 a ← a + habit_rate · (2y − 1)；默认 habit_rate = 0时不执行任何操作。

配对世界和通过(simulate.run_paired)。六个世界共享同一个总体和同一个抽样流：

世界	参与者	我们的助推
empty_none	none	no
empty	none	yes
none	free	no
free	free	yes
frozen_none	held	no
frozen	held	yes

第1轮在所有周期中执行步长 empty_none 和 empty 。每次运行中，周期 0–59 的平均 empty_none 站点比率即为无行动者基线，并将每位焦点观测行动者的目标设为基线 + actors.base. 目标偏移量（；裁决 R）。第1轮会重置抽样流，重放相同的抽样结果并执行步长和。在rollout_start （周期 100），根据自由世界参与者的克隆副本构建两个冻结世界，每个参与者都固定在其周期 90–99 的平均触达率（和自动触达率）(simulate._held_clones, actors.base.Actor.hold)，并从此时起继续执行步长。

Actor.set_goal 1 2 none free

设计(designs.Design；尖峰注释第3.3节，规范第4.4节)。所有设计的预热期均为 0–59。 pilot_rollout：在 60–79 期间固定处理组为 2%，80–99 期间无人接受处理，100 起所有人接受处理。 field_experiment：从 100 起固定处理组为 40%，无试点。 pilot_field_experiment：先进行试点，然后从 100 起采用独立抽取的 40% 处理组。当前处理组以外的所有人均为对照组。评估窗口为 160–199。

2. 设计概念

基本原则。 以勒温的力场为灵感：行为是相互作用的力所形成的均衡（规范第1节）。力在潜在尺度上相加，而行为是一次阈值跨越（概率单位），因此效应取决于人们相对于阈值所处的位置。审慎型助推会争夺有限的注意力预算；自动助推则绕过注意力预算（规范第4.1节）。参与者是观测速率的控制器，而非战略家（规范第2节）。

涌现。 预期效应、试验报告效应、冻结效应与总效应之间的差距，进而结果类别，源于行为与参与者规则之间的反馈；没有任何规则编码衰减、反转或盲法试验。

适应。 参与者会根据观测速率调整触达率（第3.3.4节），并可能不可逆地切换为持久性自动助推（第3.3.5节）。人不会适应，除非在习惯机制开启时通过习惯适应。

目标。 恒温器和触发器角色以比率为目标：竞争者的基线为 − 0.03，盟友的基线为 + 0.03（裁定 R1）；旁观者观察自身行为，并且在已注册场景中，当增益为 0 时不会做出反应。会做出反应的旁观者（非零 gain 或 gain_p）在 v1（actors.base.ActorSpec）中会被拒绝：其目标将

是绝对的 goal_offset，而不是相对于其自身行为基线的比率。最大化者具有方向性（规范 §4.3，规则 R7）：它希望焦点速率更低，并且绝不放弃触达率。

学习ng. **预测**无。最大化者的基准值是其达到的最佳比率的记忆（一个棘轮），而不是学习

tion. 无。参与者仅对滞后观测作出反应。

Sensing. 感知。行动者感知一个数量：其站点上目标行为的聚合速率，带有滞后（默认为 2）以及可选的高斯观测噪声（规范 §4.3、§12）。它不会感知其他参与者的助推（作者、规范 §10）。人只能感知触达他们的助推。

Interaction. 交互作用。仅限间接交互。参与者通过他们共同关注的比率进行交互。助推在每个个体内部通过潜在和以及注意力竞争发生交互。不存在个体之间的交互。

随机性。 每次运行时抽取一次：总体（people.Population.draw）、每个助推名称对应的效价（people.Population.valence）、旁观者自身的倾向（{v5）以及我们的处理组（designs.draw_treated）。每个周期（draws.DrawStream.period）按以下顺序抽取：行为噪声 ε、用于我们助推的均匀随机数 u_ours （每个周期都会抽取，以固定顺序；已预注册的设计会触达固定处理组，因此不会读取该值）、行为体暴露的均匀随机数 u_actor、自动助推的均匀随机数 u_auto、旁观者自身行为的噪声 eps_own 以及观测噪声 obs。每次配对运行只有一个抽样流，六个世界共享该流，因此配对差异不包含来自独立抽样的噪声。无论哪个世界使用，每个名称对应的抽样结果都会在每个周期生成。

随机种子。 扫描会按参与者结构对网格点分组；每组包含 points × n_rep 次运行，使用一个随机种子；stable_hash("scenario|design|structure|seed_base") 63 sweep.batch_ 会被屏蔽为比特（；是一个 -seed stable_hash 8 2 run.max_batch_runs > 0 字节的 blake b）。批次较大时会分割成块，块 i 使用同一字符串并附加后缀 |chunk{i} 作为随机种子；能够装入一个批次的则保留未加后缀的随机种子。总体种子和处理集生成器由批次种子派生（draws.population_seed draws. ,）；抽样流则直接使用该种 treated_generator 子。批次内的运行是同一张量中的各行，因此彼此独立。抽样结果可在同一设备上复现，但无法跨设备复现。清单记录每个批次的随机种子及其运行。

集合。 无（臂和三分位组是记录组）。

观测。 第3.1节（输出）和第3.3.10节（指标）。

3. 细节

3.1 实现与输出

PyTorch，逐人张量使用 float32，所有逐期运算均围绕 (R, N)进行向量化；绝不会分配带有跨个体时间轴的张量。设备由标志指定（cpu、 mps、 cuda、 auto； device.resolve_device）。历史记录以 (T, R) 的形式保存在设备上，并在最后转换为 CPU float64（world.to_cpu64）。一次扫描（sweep.run_sweep）会将 effects.parquet （长表 t、 replicate、 quantity、 value）、 classes.parquet （每次运行一行）、 robust.parquet 和 manifest.json （配置、随机种子、批次、Git 提交、设备、环境）写入 code/outputs/<scenario>_<design>_<YYYYMMDD_HHMMSS>/；每行都包含 scenario、 variant、 design、 point、 param_name run 、 param_value、 replicate、 seed, 。场景文件位于 code/conf/nudgefield/scenario/，并通过 scenarios.load_scenario 2 code/conf/nudgefield/figures/ 加载；F 类别到情景的图为。 F2.yaml

3.2 初始化与输入数据

在周期 0：人是根据总体参数抽取的；竞争者和盟友的触达率从 0 开始，并从周期 lag 起作出响应；旁观者的触达率从 0.6 开始并保持不变（增益为 0）；最大化者保持在 0.60，直到周期 60（R7）。目标在第2轮开始之前，根据第1轮的基线设定。预热期为周期 0–59。**预热稳定性检验 (**world.burn_in_settled；尖峰注释 §1.2)：在每次运行中， none 世界站点比率在周期 40–59 内的最小二乘斜率绝对值必须低于每个周期 0.002；否则扫描会在该批次之后停止，写入包含失败运行的清单，并触发 sweep.BurnInNotSettledError。

输入数据： 无。模型依据已发表量级进行校准（规格说明 §6），而非拟合；每个值和 i ts 来源均在 docs/nudgefield/parameters.md。

3.3 子模型

3.3.1 总体抽样 people.Population.draw。使用同一个带种子生成器，按以下顺序抽取：组成指标、 a、 B、 s_auto、 s_delib；根据运行内排名确定三分位标签。效价（people.Population.valence）由一个以总体种子与助推名称的哈希值异或结果为种子的生成器生成，因此不受调用顺序影响。由尖峰注释§3.1和§1.6的效价解读固定；规范第4.1节。

3.3.2 潜在价值与行为 attention.latent, attention.behave。规范 §4.1 的 Probit 阈值采用推动力 d_k τ_k v_ik，自动助推按 s_auto 缩放，审慎型助推按 s_delib · w 缩放。依据规范 §4.1 和尖峰注释§3.1 固定。

3.3.3 注意力共享 注册表 `attention.ATTENTION_RULES`；规则由 `run.attention` 选择。- `attention.priority`（默认；作者于 2026-10-27 作出的作者决定，规范第12节）：每个个体按感知到的推动力 `salience_k · |v_ik|`（`salience = τ`），将触达自己的审慎型助推从强到弱排序；若强度相同，则按接触顺序决定；每个个体依次接受 `w = min(1, remaining / c_k)`，直至用完预算 `B`。按感知到的推动力排序意味着强度相同也不会出现并列。- **旁观者显著性**（控制器对任务 4a 的裁决，提交 6309d9a，依据规范 §4.3“只通过注意力影响我们的行为”）：旁观者的助推以 τ =0 进入焦点行为（即不产生推动力），其显著性等于自身的 τ，并使用自身名称，因此也有自身的效价；所以在优先级规则下，它会以个体感知到的强度（`actors.base.Actor.focal_nudge`）参与注意力竞争。- `attention.proportional`（尖峰规则，尖峰说明 §3.2；规范 §4.1）： `w = min(1, B / D)`, `D` 触达该个体的审慎型助推的成本总和。奇偶性运行采用此规则（尖峰结果已按此规则登记）；两项规则均保留在敏感性扫描中。

3.3.4 行动者响应规则规则注册表 `actors.RULES`；每条规则读取 `observed[t − lag]`，就地更新触达率，并返回 R6 计数的每次运行“失利”标记。- `actors.rules.thermostat`： `gap = r[t − lag]− goal`；竞争者将触达率改变 `+(gain · gap + gain_p · Δgap)`，盟友和旁观者则改变其相反数，并截断至 $[0, 1]$。在 `gain_p = 0` 下，它就是尖峰积分规则（spike note §2）；比例项是规范 §4.3 的比例-积分形式，且在首次观测时为零（周期 `lag`： Δgap 从该周期的差距开始，因此不会产生冲击）。- `actors.rules.maximiser`：规则 R7（尖峰说明 §1.6）。从周期 60 开始， `ref` 取周期 50–59 的平均观测速率；当 `r[t − 2] > ref + tolerance`时，触达率只升不降，每次上升 `gain · (r − ref)`； `ref ← min(ref, r)`（已登记的棘轮），或采用带有 `ref_smoothing > 0` 的指数平均（尖峰说明 §8 (ii)，探索）。仅适用于竞争者。- `actors.rules.threshold_trigger`：规范 §4.3 的备选规则；当 `|gap| > tolerance` 时，触达率每个周期改变 `gain`。该规则列于规则注册表中，用于规则形式的敏感性扫描（规范 §5 SI）；没有已登记的情景使用它。

3.3.5 升级为持久性自动助推 `actors.escalation.escalate`；规则R6（尖峰说明 §1.6）以及规则R7的升级条件。应用规则后，连续“失利”周期的计数器会递增（对最大化者而言，仅在触达率也达到上限时才递增）或重置；在触达率为 ≥ 0.99 且计数为 ≥ 10 -1 `escalation_` 的首个周期，`tau` 1 朝竞争者方向的自动助推（适用于每个已登记情景）会启动，强度为 ，触达率会切换并保持开启。仅适用于竞争者。

3.3.6 观测与噪声 `world.World.step` `obs_noise_sd`。观测 · ε `obs` ε `obs ~ N(0, 1)` 2026 10 27 12 速率等于真实速率加上来自共享流的 ，按每次运行、每个周期计算。作者决定采用 - -（规范 § 实现规范 §4.3 中的“滞后与噪声”）；奇偶性的默认值为 0，在“探索”中进行遍历。没有它，行动者的敏感度就会成为 `N` 的人为产物。

3.3.7 旁观者自身的行为 `world.World._own_rate`、 `people.Population.own_propensity`。规范第4.1节的简化形式：每位旁观者的行为都有其自身的倾向（按 a的方式抽取），并使用相同的阈值机制；该机制仅由旁观者自身的助推推动，其注意力权重取自该助推在焦点尺度上获得的注意力权重。仅保留其比率；这就是旁观者所观察到的内容。仅在自由世界中进行模拟，与尖峰中的做法相同（spike note §2； `code/spike/worlds.py`）。

3.3.8 研究设计与处理组集合 `designs.Design`、 `designs.draw_treated`。每次运行恰好有`round(share · N)` 人，每次运行从标记为 `pilot` 或 fe的生成器中均匀抽取一次。由尖峰注释 §2（“组别分配”）和第3.3节确定；参见规范第4.4节。时间线和触达率位于`conf/nudgefield/config.yaml` 的 `design_params` 部分（默认采用尖峰注释 §3.3中的设置；CLI 的 `design=` 用于命名类型）。

3.3.9 反事实世界与习惯 `simulate.run_paired`, `simulate._held_clones`。冻结场将每位行动者在 90–99 期间的触达率固定为其自由世界平均触达（尖峰注释第1.1节；规范第4.4节）。习惯关闭时，在周期 100 分支的结果等同于从周期 0 开始运行（人没有记忆）。**习惯作用下的冻结世界反事实**（作者于 2026-10-27 作出的决定，规范第12节）：对于 `habit_rate > 0`，每个世界都拥有一份人口副本；冻结对在周期 100 从其来源世界的人口状态副本分支（`free` 对应 `frozen`， `none` 对应 `frozen_none` `people.`），随后固 `Population.update_habit` 1 11 定场：直到推广前的历史相同，此后场不再作出反应。习惯在 v 中关闭（作者，规范第 ），这是一个敏感性开关。

3.3.10 指标、类别与稳健性

- `metrics.effects.effects`：尖峰注释 §1.1 的逐期量。试点 → 推广： `total`（免费 − 无），`frozen`、 `empty`、 `trial`（试点）、 `before_after`。现场实验中每位接受治疗者： `total`、 `frozen`、 `empty`（`E_fe`，裁决 R2）、 `trial`、 `control_drift`（裁定 R3）、 `site_effect`（规则 R8）。所有设计：按 `a`、 `B`、 `s_delib` 的三分位组划分，计算每位行动者的触达率和亚组对比。• `metrics.classes.classify`：尖峰注释 §1.3 和 §
1 5 4 `metrics.classes.oscillation` 8 `metrics.classes.` 中的类别、阈 `site_class` `metrics.classes.subgroup_flip` 值、评估顺序及读数。（按裁决 R 规定的振荡滞后，），R 站点类别（）以及已记录但未分类的。

- `metrics.robust.robust`：尖峰注释 §1.4（每个 ≥ 2 相邻扫描点上，$\geq 90\%$ 的重复试验）。**将稳健性阈值设为参数**（作者决定 2026 10 27, 12 `min_points` `min_` - - 规范 §）：由 和 设 `range_share` `run.robust_min_points` `run.robust_min_range_share` 2 0 10 定，并且（默认选项和 。由于作者重述 2026 10 29, 13 ≥ 2 $\geq$ - - 规范 §：一段相邻 10 2 0 0 合格点，其跨度占扫描范围的 %；以及 。这是尖峰的规则； `conf/nudgefield/config.yaml`），传递给两个标签的调用并记录在清单中（`run` 和 `robust_rule`）。在多参数扫描中，该规则沿 `param_value` 运行，在其他扫描路径的每个切片内执行（每个切片和类别各占一行 `robust.parquet` ）。

3.3.11 扫描 `sweep.expand`、 `sweep.run_sweep`。情景的网格是其扫描列表的笛卡尔积；共享同一参与者结构的数据点构成一个批次，其扫描浮点数会在各次运行中作为 `(R,)` 覆盖值变化；扫描一个 `count` 会改变结构，因此会形成单独的批次。按批次设定随机种子（计划决策）可避免每个张量每个周期进行 `R` 内核启动。

补充信息附录，S6 节：每个参数及其来源 e

本节复现项目的参数表格：每个值、其单位、来源，以及扫描是否会改变该参数。工作名称在 S5 节开头进行了说明。来源：docs/nudgefield/parameters.md，最后一次更改于提交 63f16a0，2026-11-01；原样复现。

PopulationConfig、 NudgeSpec （我们的和其他参与者的）以及 ActorSpec、DesignConfig和 RunConfig（code/src/nudgefield/config.py、 code/src/nudgefield/actors/base.py）的每个参数各占一行，另加校准拥挤场。**来源** 包括以下之一：尖峰注释 §3（或确定该参数的尖峰说明 §1.6 裁决）、规范第6节中的幅度、作者决定（规范第12节）、“已扫描，无锚点”（规范第6节：无已发布锚点）或“已校准至 T2 区间”。**探索** 给出存在时为探索阶段预设的扫描范围；“待注册范围”表示该参数已预设用于扫描，但尚未确定范围。潜变量单位采用 Y = 1{a + F + ε > 0}、 ε ~ N(0, 1)的概率单位尺度。周期以周为单位（规范第4节）。

1. 总体（PopulationConfig）

参数	值	Unit	来源	探索
n	20 000_	每个站点的人数	尖峰注释§3.1	50_000在 Spark（规范第8节）；一种尺度，而非模型参数
t_total	200	周期	尖峰注释 §3.3	—
mix_share	0.3	.....的份额	尖峰注释§3.1；规格说明 §6	—（T4 使用 0：一个组件)
		高倾向性成分	幅度（对照组参与率 0.05–0.9：实际静默控制组参与率 0.309，见§6)	
mu_low	−1.2	潜在变量	尖峰注释§3.1	—（T4 对其进行扫描，见§5）
mu_high	0.4	潜在变量	尖峰注释§3.1	—
sd_a	0.8	潜在变量	尖峰注释§3.1	—
budget_mean	1.0	注意力成本每个周期的单位	尖峰注释§3.1；已扫描，无锚点	注意力预算：待注册（规范§5 F3 名称旁观者 ×预算）；每次运行可扫描，因为 1898717（扫描路径 population.budget_mean，一个情景的 population: {budget_mean: [..]}）
budget_sd_log	0.4	对数尺度	尖峰注释§3.1；已扫描，无锚点	待注册范围
susc_sd_log	0.3	对数尺度（均值 1)	尖峰注释§3.1	—
valence_sd	0.5	标准差	尖峰注释§3.1和效价 §1.6的效价解读	—
		对数正态幅度（均值 1)		
flip_prob	0.05	每个	尖峰注释§3.1（平均效价 0.90）	—
habit_rate	0.0	个人提示每个周期	规范第4.1节（作者：习惯关闭在 v1 中）；全局约束	将习惯开启作为开关；范围至已登记

2. 助推（NudgeSpec）

参数	值	Unit	来源	探索
我们的: direction	+1	—	尖峰说明 §3.2	—
我们的: kind	delib	—	尖峰说明 §3.2	—（T4 使用 auto, tau 0.5）

参数	值	Unit	来源	探索
我们的: tau	0.30	潜在变量	尖峰说明 §3.2；规格说明 §6 幅度（概率单位效应 0.05–0.4：实际安静现场实验的概率单位效应 0.152，见 §6）	—
我们的: cost	1.0	注意力成本单位	尖峰说明 第3.2节；已扫过，无锚点	注意力成本：范围待注册
其他: tau	0.30	潜在变量	尖峰说明 §3.2	escalation tau下方_
其他: cost	1.0	注意力成本单位	尖峰说明 §3.2；已扫过，无锚点	注意力成本：范围待注册

3. 参与者（ActorSpec默认选项；场景文件可按参与者覆盖）

参数	默认	Unit	来源	探索
rule	thermostat	—	spike note §2（积分恒温器）；maximiser R7; threshold_trigger 规范 §4.3	其中的规则形式敏感性扫描（规范 §5 SI)
gain	2.0	单位覆盖范围每周期速率差	尖峰注释 §2（S1/S2 扫描 k ∈ {0.5, 2, 5}）；已扫描，无锚点	待登记范围
gain_p	0.0	周期单位覆盖范围.....的变化	规范 §4.3（比例积分形式）；0 = 尖峰积分规则	gain_p: 待登记范围已登记
goal_offset	0.0	gap rate	场景文件：竞争者 −0.03、盟友 +0.03（裁定 R1，尖峰说明 §1.6）	—
lag	2	周期	尖峰说明 §3.2	—
reach0	0.0	人份额尖峰说明§3.2（旁观者 0.6, 预热期内最大化者 0.60，R7）		—
tolerance	0.01	rate	R7（三个每周期采样标准差在 n = 20_000 时的站点比率）	与 obs noise sd_ _关联
ref_window	（50，60）	周期, 半开	R7	—
start	60	周期	R7	—
ref_smoothing	0.0	每周期权重周期（0 = R7棘轮）	R7；尖峰注释第8节 (ii)	ref_smoothing λ > 0（平滑参考）：待注册范围
escalation_tau	无（无升级）	潜在变量	R6、R7（登记扫描 0.10、0.20, 0.30)	escalation tau到_0.40及以上（尖峰注释 第8节（i））
patience	10	周期	R6	—
reach_threshold	0.99	人口比例（覆盖上限）	R6	—
count	1	副本	场景文件（S3 m ∈{0, 2, 4, 8, 16}）	旁观者数量，使用预算（规范第5节 F3)
obs_noise_sd	0.0	rate	作者决定 2026-10-27（规范第12节；0 表示尖峰奇偶性）	obs_noise_sd: 范围待注册

4. 时间线和运行设置（DesignConfig, RunConfig）

时间线各行对应 YAML 部分 design_params （DesignParams, DesignConfig 的字段，但不包括其 kind）；其他部分则为 run。

参数	值	Unit	来源	探索
pilot_start, pilot_end	60, 80	周期（半开: 60–79）	尖峰注释 第3.3节	—

参数	值	Unit	来源	探索
pilot_reach	0.02	人口比例（从固定集合中抽取在 60 时）	尖峰注释 §3.3；规范 第4.4节（1–5%）	—
rollout_start	100	周期	尖峰注释 §3.3	—
fe_reach	0.4	人口比例（从固定集合中抽取 100 个）	尖峰注释 §3.3；规范第4.4节（20–50%）	处理组占比（尖峰注释 披露：探针概率为 0.2–0.5）；每次运行可扫描，因为 b7dfa39（扫描路径 design.fe_reach，一个情景的 design: {fe_reach: [..]}；仅适用于现场实验类型，pilot_field_的 fe集合 experiment；R8 地点类别使用每次运行的份额；清单的 design.fe_reach 保持此默认值）
eval_window	40	周期 (160–199)	尖峰注释 第3.3节	—
frozen_window	(90, 100)	周期，半开	尖峰注释 第3.3节，裁定 R3	—
settle_window	(40, 60)	周期，半开	尖峰注释 第3.3节	—
settle_tol	0.002	每周期速率（绝对斜率）	尖峰注释 §3.3	—
n_rep	10	每个点的重复测量点	spike note §2	关于 Spark 的更多内容（规范 §8）
seed_base	20261027	—	计划（全局约束，播种）	—
attention	priority	规则名称	作者决定 2026-10-27（规范第12节）； proportional对于奇偶性扫描	敏感性分析中的两条规则
max_batch_runs	0（不分割）	runs	计划（任务6）	—
robust_min_points	2	相邻扫描数据点	尖峰注释 第1节.4（已登记规则）；用范围术语重新表述 2026-10-29（规范第13节）	—
robust_min_range_share	0.10	.....的份额扫描范围	规范第7节（每个扫描范围的 ≥ 10% 参数范围）；作者裁定 2026-10-29（规范 第13节）；此前为 0.0	仅适用于更细的网格 超过 10 个数据点
类别规则	r9	—	R9（尖峰说明 §1.6）	run.class rulesv1 用于_移植的规则
resume	false	—	计划 2026-10-30（Spark 的一半，任务1）：继续一个已停止的从其 parts/ 发起调用	—

5。校准拥挤场（code/conf/nudgefield/scenario/crowded.yaml）

参数	值	Unit	来源	探索
拥挤：竞争者	恒温器，增益 2， goal_offset −0.03	—	spike note §2 S1 位于 k = 2，裁决 R1	—
拥挤：盟友	恒温器，增益 2， goal_offset +0.03	—	spike note §2 S2，k = 2 处的裁决 R1	—
拥挤：旁观者 m	2 （增益 0, reach0 0.6）	旁观者	校准至 T2 区间 [1.5, 2.5]（规范第12节），默认规则 priority	—
安静	无行动者	—	规范第6节T2	—

校准运行。 test_crowded_calibration 在 code/tests/nudgefield/test_targets.py中，输出为 code/outputs/targets_20261027_201230/calibration/ ，表格在 calibration.parquet 处。场域实验设计，n = 20_000，10次重复，设备 mps。Probit 试验效应是每次重复中自由世界窗口均值（周期 160–199）的 Φ^{-1}(treated) $-\Phi^{-1}$(control)，再对重复试验取平均；比率为安静组 / 拥挤组。安静组：处理组 0.3645，对照组 0.3090，Probit效应 0.1524（在两条规则下相同：一个审慎型助推，因此两条规则都给出 w = min(1, B)）。该规则选择满足 priority 比率为 [1.5、2.5]的最小 m 。Probit效应背后的级别，是每个运行目录的 levels.parquet，自 86b6317 起由包工具 nudgefield.levels 写入（在此之前，则由 test_targets.py的私有辅助函数写入）；T2 测试将 priority 下的m = 2 比率固定为 2.25 ± 0.01（在 86b6317 时为 2.2506）。

m	拥挤 概率单位 (priority)	比率 (priority)	在区间内	拥挤 概率单位 (proportional)	比率 (proportional)
2	0.0677	**2.25**	yes	0.0453	3.36
3	0.0571	2.67	no	0.0354	4.30
4	0.0537	2.84	no	0.0327	4.66
6	0.0431	3.54	no	0.0236	6.46
8	0.0368	4.14	no	0.0187	8.13

在两条规则下，比率都随 m 上升； m = 2是唯一处于该区间内的候选项。在 proportional 下，每个比率都高于该区间（该区间仅适用于默认规则；方向保持不变）。

T2 位于 m = 0，而 1 位于 priority （2026-10-27 的修复波次）。相同的校准代码（_quiet_crowded，使用 test_targets.py中的 _crowded_swept((0, 1))），仅 priority，设置和种子推导相同（情景 crowded、 seed_base 20261027； m = 0 和 1 是新的行动者结构，因此使用新的批次种子），设备 mps，代码版本为 90a1370（git_dirty: false）。输出位于新目录 code/outputs/targets_20261027_213643/ calibration/priority/{quiet,crowded}/ calibration_m01.parquet targets_ 中，表格也在那里；已提交的运行未发 20261027_201230 0 1524
生变化。安静组运行复现了已提交的安静状态 Probit 效应，..

m 拥挤的概率单位（priority）	比率（priority）	处于区间内
0 0.0841	1.81	yes
1 0.0704	2.16	yes

该区间不会从下方限制 m： 仅竞争者和盟友（m = 0）就已使比率落入 [1.5, 2.5], , 而{0, 1, 2}中的每个 m 都在区间内；只有 m ≥ 3 被排除。 m = 2 是已登记候选集{2, 3, 4, 6, 8}中处于区间内的最小成员，而不是区间内最小的 m 。**作者裁定（2026-10-29，规范 §13）：m = 2 维持不变**。这是已登记候选集上已登记规则的结果；旁观者数量在F3阶段图中逐一变化，并在其中展示敏感性。 crowded.yaml 不变。

T4设置（仅用于测试，不是基线）。无行动者；我们的助推自动， tau 0.5； mix_share 0, sd_a 0.8；mu_low 在参与率0.1至0.8之间七个等间距取值上， mu = sqrt(1 + 0.8^2) · Φ^{-1}(q)： −1.641, −1 188 −0 735 −0 282, 0 171, 0 625, 1 078 .,.,.....

余量族（第 8 次运行，注册表第14节；2026-10-31）。第二个拥挤族， crowded_headroom （S1 竞争者位于 nudge.tau 0. 35，加上 m 个旁观者，没有盟友），按第14节的联合单元规则，在 m = 2、预算 1.0 的条件下进行校准：安静组 / 拥挤组 Probit 比率为 1.508（处于该区间的下边界；相邻单元 m = 3 给出 1 733 1 0 640, 0 735, 100 100, docs/analysis/2026- .），T . 竞 10-31-run8-calibration-map.md 4, 1 2
争者预热可达范围 . 在优先级（；Modal T CUDA）下衰减。这是 T 和 T 与会作出反应的竞争者并存的族；上述 crowded.yaml 校准仍作为僵局情形（其两个恒温器在预热期均达到完全触达率； docs/analysis/2026-10-31-t1-probe-two-actor-field.md）。运行 4h（注册表第15节）绘制了余量族在 m × 预算下的图谱。

6. 对照规范第6节和 T1差异进行幅度检查

基于上述运行结果中的 levels.parquet 和 effects.parquet （窗口均值、重复样本均值） ：

- 平静场域的对照组参与率为 0.309（规格说明 §6：0.05–0.9）。• 我们的助推在安静现场实验中的 Probit效应为 0.152；在拥挤场域中为 0.068（priority）（规格说明 §6：0.05–0.4）。• 在 priority下，安静/拥挤比为 2.25：这是一个约束，而非预测，因为 m 是为满足该约束而选定的（规范第12节）。

T1，规模化效应损失 （试点 → 推广； total 在160–199期间的均值除以试点 trial 在60–79期间的均值；来自 T1/<rule>/<site>/effects.parquet，在 T1_<site>_<rule>.parquet中逐重复）：

站点 / 规则	比率 重复实验 均值（检验, < 0.8)	均值 每次重复实验的 比率	重复试验 ≥ 0.8	每次重复实验的 最小值–最大值	试点对比均值 （重复试验间的标准差）	推广 `total`
S1 k $= 2$ / `priority`	0.741	0.795	4/10	0.565–1.393	0.0467（0.0117）	0.0346
S1 k $= 2$ / `proportional`	0.671	0.705	3/10	0.515–1.034	0.0462 (0.0106)	0.0310
S2 k $= 2$ / `priority`	0.621	0.639	1/10	0.494–0.824	0.0405 (0.0070)	0.0252
S2 k $= 2$ / `proportional`	0.738	0.767	6/10	0.552–1.031	0.0341 (0.0068)	0.0252

检验统计量是重复均值之比，因为对20,000人进行2%的试点（约400名处理组）会产生噪声较大的对比（标准差为0.007–0.012，而均值为0.034–0.047），而推广 `total` 在各重复实验之间变化很小（标准差为 ≤ 0.0007）。逐重复观察结果较弱：在四个单元格中的三个单元格里，至少有一个重复实验的推广 $\geq$ 试点（比率 ≥ 1：S1/`priority`、S1/`proportional` 和S2/`proportional`中各有一个重复实验），而在S2/`proportional` 6 的10次重复中，比率达到或高于0.8。

7. 运行来源

25 `manifest.json` 下的文件记录了 `git_commit` 5882f76，并带有`git_dirty: true`，因为运行执行时 `quiet.yaml`、`crowded.yaml` 和 `test_targets.py` 均未被跟踪。提交 c47b114 保存了生成这些文件的代码，而任务 8 fix-round-1 提交 9bcf9d6 则保存了包含命名常量和索引对齐检查的版本；两者都没有改变运行输入（T4 网格和每个随机种子均未改变），修复后重新运行的 targets 测试套件通过率为 9/9。

干净重跑（第9次运行，注册表 §16，2026-11-01）。 `code/outputs/targets_20261101_200002/` 保存了在 c220bcf 上重新运行的 targets 测试套件，所有 33 份清单均包含 `git_dirty: false` （M4， `mps`），只有一项已登记的输入变更：`T4_POINTS` 15 （原为 7）。重复均值之比在小数点后三位复现了上方表格（0.741、0.671、0.621、0.738），每次重复实验的数量 ≥ 0保持不变。8（10 次中分别为 4、3、1、6）；优先级下 T2 为 2.2506；T3 对比 0.0036 为 0.0132；T4 峰值在对照组参与率 0.414 时达到，相比预言机的 0.423。ED3 取自此目录；上方未清理的构建阶段运行仍是该校准来源的记录。

精细 T4 网格（第10次运行，注册表 §17，2026-11-01）。 `code/outputs/targets_20261101_223633/` 在 4b1b94e（干净；42 份清单）上重新运行测试套件，并在相同随机种子下使用包含 9 个参与率水平（0.300–0.500）的第二个 T4 网格；粗网格表与第9次运行完全相同。细网格的峰值出现在参与率 0.376 处，与预言机 0.423 相差 0.047（在 0.05 容差范围内，但不在一个网格步长之内：倒U形曲线的顶部较平——在参与率 0.356–0.414 范围内，峰值与最大值相差不超过一个重复样本标准误；在 0.351–0.426 范围内，与最大值相差不超过 0.0006——因此在样本量为_重复 10 次时，峰值只能确定到该平台区）；注册表 §17 记录了这一点，以及 `ate_sd` 上的另一项未达成预期，并将其列为偏差。现在两个 T4 表都包含 `ate_sd` 和 `n_rep`。ED3 面板 (d) 根据此目录绘制。

补充信息附录，S7 节：注明日期的运行记录

本节重现了这样一份记录：每次扫描在运行前都已明确规定，并列出了每个条目的日期、预期，以及事后披露的偏差和未达成预期。这是一个版本控制文件，并非公开预注册；每个条目的日期即其提交日期。工作名称已在 S5 节开头说明。来源：`docs/nudgefield/explore-registry.md`，最近一次更改为提交 dae2903，日期为 2026-11-12；重现时未作更改。

每次 Explore 扫描都会在运行**之前**作为带日期的条目登记在此：场景文件、设计、扫描值、重复试验、总体规模、种子基数、注意力规则、输出目录、预期每个数据点呈现的类别，以及哪些内容计为先验知识。条目提交后，由作者签字确认，随后才运行命令。运行后对条目的任何修改都必须作为新的带日期条目，并说明原因；原条目予以保留。结果不写在此处：它们会写入运行目录的 `classes.parquet` / `robust.parquet` 以及分析笔记中。图中数据点的选择（`code/conf/nudgefield/figures/F2.yaml` 和 F3/F4 YAML 文件）是运行完成后的另一个经批准的步骤。

每个条目均适用的规则：尖峰注释 §1–§3及裁定 R1–R8；规范 §13重新表述的稳健性规则（一次运行中跨越扫描范围 $\geq 10\%$、由 ≥ 2 个相邻数据点组成的范围内，每个数据点的重复试验中有 $\geq 90\%$满足条件；`robust_min_points` 2 `run.robust_min_range_share` 0 10 `run.attention = priority` `run.` ，。）；（软件包默认值），除非条目另有说明；`run.max_batch_runs = 0`（不进行分块：分块会改变批次种子，进而改变抽样结果）。设备不是已登记的参数（快速测试套件将 `cpu`、`mps` 和 `cuda` 固定为相同结果，且处于 float32 容差范围内）。如果一个批次无法装入记忆，回退方案是 `run.max_batch_runs` = 每个数据点对应的一组运行（100），并记录在条目的台账行中；同时报告该运行是在该设置下完成的。

0. 运行 1、2 和 2b 共享的设置（已登记 2026-10-29）

设置	值	原因
`run.n_rep`	100	简报 §3：n_rep = 100 时的最终图
`population.n`	20,000	简报 第9节；在这个大小下，M4 已足够
`run.seed_base`	20261029	探索阶段的新基数，使其抽样结果与构建阶段运行区分开来（种子基数 20261027）；批次以随机种子初始化 `scenario|design|structure|seed_base`
`run.attention`	优先级	默认规则；F2 数据点正在被在其下重新建立（F2.yaml 标头）
`run.max_batch_runs`	0	每种参与者结构一个批次
窗口	`eval_window` 40 （周期 160–199），`frozen_window` [90, 100]， `settle_window` [40, 60]， `settle_tol` 0.002	与 `config.yaml` 相同
`design_params`	触达率为 0.02 时，试点 60–80；推广 / 场域现场实验从 100 开始； `fe_reach` 0.4	不变
输出	`code/outputs/explore_20261029/<run>/<scenario>_<design>/`通过 `out_dir=`	每次扫描调用对应一个目录；图 YAML 此处的点
code	包含此条目的提交（记录于每个 `manifest.json` as `git_commit`， `git_dirty: false`）	简报 §3 的来源规则

1. 运行 1：优先级下已登记的扫描（数据源 F2、ED1、ED2）

目的：在默认注意力规则下，以 n_rep = 100，重新确定尖峰中已登记的扫描各数据点的结果类别（在 n_rep = 10 下采用比例共享）。这些是 F2 候选数据点。场景文件中的内容均不作更改。

Call	场景文件	设计	扫描值	Runs
1a	`S1_competitor`	试点_推广	`competitor.gain` {0.5, 2, 5}	300
1b	`S1_competitor`	场域_实验	`competitor.gain`{0.5, 2, 5}	300
1c	`S2_ally`	试点_推广	`ally.gain` {0.5, 2, 5}	300

Call	场景文件	设计	扫描值	Runs
1d	S3_bystanders	试点_推广	bystander.count{0, 2, 4, 8, 16}	5 个批次 × 100（一个结构每 m）
1e	S1_persistent_auto	试点_推广	competitor.escalation_tau {0.1, 0.2、0.3}	300
1f	S1_persistent_auto	场域_实验	competitor.escalation_tau {0.1, 0.2, 0.3}	300
1g	S1_maximiser	试点_推广	competitor.escalation_tau {0.1, 0.2, 0.3}	300
1h	S1_maximiser	场域_实验	competitor.escalation_tau {0.1, 0.2, 0.3}	300

范围检查：三个等间距的数据点中，两个相邻数据点跨越范围的 50%（1a–1c、1e–1h）；对于 1d，最小相邻间隔（$0 \to 2$）为 0–16 的 12.5%。该规则在所有地方均按原文执行。

运行前说明的预期类别（先验知识：比例共享下尖峰的结果，尖峰注释 §4–§7；优先级下 1,002 次 S1现场实验运行的 Spark 运行，盲_站点 1,002/1,002，衰减 1,002/1,002， ratio_window 0.63；奇偶性运行 P1–P6；目标运行的 T1差异， parameters.md §6）：

- 1a：在每个 k 下均衰减（尖峰：每个 k 均为 10/10；在优先级下，Spark 现场实验运行给出 E 的 0.63，因此试点 → 推广应接近尖峰的 0.6–0.7）。预期稳健。• 1b：在每个 k 下均为 cls_combined 盲_站点（Spark：1,002/1,002）。预期稳健。 cls 衰减。

- 1c：衰减或符合预期；尖峰（比例）发现盟友放松至零并移除约十分之一；在优先级条件下，盟友对我方方向的复制可能以不同方式分配注意力（尖峰注释 §4.6 第 3 项需在此处重新确立）。预期：在 $k = 2$、5 处衰减，ratio_window 0.85–0.95；在 $k = 0.5$ 处可能符合预期。未表述为稳健。
- 1d：在 $m = 0$ 处符合预期（按构造；单个点）；在 $m = 2$、4 处衰减；在 $m = 8$、16 处为空值（比例条件下的尖峰：10/10；在优先级条件下，拥挤更为尖锐， parameters.md §5，因此空值可能从 $m = 4$ 开始）。简报关于优先级是否会产生一条临界线的问题，首先在此处得到回答。• 1e：在 0.1、0.2、0.3 处为瞬态（尖峰：每处均为 10/10，稳健）。稳健地符合预期。• 1f：在 0.1 和 0.2 处为 cls_combined 盲_站点；在 0.3 处为盲_站点或反向_站点（R8 的预期；尖峰的站点窗口均值为 0.0002 / 0.0001 / -0.0040）。• 1g：0.1 处为瞬态或空值；0.2 处为瞬态（尖峰： $-0.12 \cdot E$，高于 -0.25 阈值）；0.3 处反转（尖峰：10/10， $-0.44 \cdot E$）。仅在一个点发生反转，因此本次运行中并不稳健；运行 2 扩展了轴。• 1h：在 0.1、0.2 处为 cls_combined 盲_站点；在 0.3 处为盲_站点或反向_站点。

哪些情况会推翻模型的说法：如果某个类别在已登记的 F2 点处因规则不同而发生翻转（例如，1e 中没有瞬态，或 1b 中没有盲_站点），就意味着 F2 条目依赖规则，图中必须对此作出说明（ED4）。

2. 运行 2：反转稳健性与参考规则（反向数据源：F2、F4b）

场景文件 S1_maximiser_smoothing 7 S1_maximiser competitor. escalation_tau 0 1, 0 2, 0 3, 0 4, 0 5 competitor.ref_smoothing ∈ 0, 0 1 （新增，本次提交）：R 最大化者将范围扩展至 {.....} 和 {.}，以此作为切片轴；两种设计；每种设计运行一批，共 1,000 次。

Call	设计	稳健性轴	切片	Runs
2p	试点_推广	escalation tau 0.1 . . . 0.5	ref_smoothing 0(R7 棘轮), 0.1（指数平均）	1,000
2f	场域_实验	same	same	1,000

选择与理由。上限 *0.5*： 简报的提议；按照 R5 所采用的相同判断，高于规格说明 §6 中校准后的单臂 Probit 范围（0.05–0.4）（升级中的行动者可能使效应超过单臂效应的一半；R5 取值为 0.60）。间距： 0.1，因此相邻的两个数据点覆盖该范围的 25%。 $\lambda = 0.1$： 指数平均 ref← 0.9 ref + 0.1 r[t − 2]的记忆约为十个周期，与升级的耐心期限（10 个周期）相同；它消除了尖峰说明 §7.4 和 §7.7 所指出的棘轮效应，而该效应正是最大化者的触达率在推广之前逐渐上升的原因。取一个值，而不是进行扫描：问题在于反转是否依赖棘轮效应，而不是它如何随 λ变化。

运行前所述的预期类别（先验知识：尖峰的 S1max 和 S1maxfe，说明 §7；运行 1g/1h 尚未运行）：

- 2p, $\lambda = 0$：0.1 为空值或瞬态；0.2 为瞬态（临界）；0.3、0.4、0.5 为反向。如果 0.3–0.5 各自有 $\geq 90\%$ 的反向，则该类别是稳健的（三个相邻数据点，占范围的 50%），且 F2 的“反转”面板会失去其“单点”标记。反转的深度应随 τ_自动推动而增加（在尖峰中，0.3 处约为 $-0.44 \cdot$E；自动推动作用于所有人）。• 2p, $\lambda = 0.1$：推广前的触达率攀升应更小（不会棘轮式地降至噪声最低点），因此推出前状态更接近已登记的

0 60 ≥ 0 99 ref + 。；触 0 01 0 01

发器（触达率达到 . 且连续十个周期高于 .）在推广后仍应触发，因为推广使比率提升超过 .，而十期平均值会滞后于它。预期：与 $\lambda = 0$ 在 0.3–0.5 处相同的类别；在 0.1–0.2 处相同，或向“符合预期”迈进一步； classes.parquet （trigger_free_competitor@auto）中的触发周期更晚。如果反转在 $\lambda = 0.1$ 下消失，则结果表明 R7 的反转依赖于棘轮效应，论文对此会予以说明。• 2f：在 0.1 和 0.2 处为 cls_combined 盲_站点；在 0.3 或更高处，一旦自动推动超过站点上的 p · E_fe，站点即为反向_（根据两种 λ，从 0.3 起均符合预期）。整个过程中，试验对比保持接近 E_fe （身份试验/总量 $\approx 1/(1-\text{p})$，而竞争者会使站点比率趋于中性）。

3. 运行 2b：针对竞争者的增益的衰减时间（数据源 F4c）

场景文件 S1_persistent_auto_gain （新的，本次提交）： S1_persistent_auto 在 escalation_tau 0.3 处的 R6 竞争者，使用 competitor.gain ∈ {0.5, 1, 2, 5, 10}；仅进行试点_推广；一批 500 次运行。

理由。简报 §7c：衰减时间由滞后和增益决定，而不是由 τ_自动决定（尖峰说明 §6.4：当 τ_自动为 0.1–0.3 时，75% / 50% 交叉点落在相同周期）。0.5–5 覆盖了尖峰的已登记增益；10 则延伸至仅在披露的探测实验中 k $= 20$–40 时观察到的振荡区域。间距不均匀；最小相邻间隔（$0.5 \to 1$）占范围的 5.3%，因此从 1 开始的任意两个相邻数据点均满足规则规定的 10% 份额（$1 \to 2$ 为 10.5%）；仅 $0.5 \to 1$ 这一对不符合要求。 escalation_tau 0.3，因为那是已登记的瞬态点（F2.yaml），也是尖峰最明显的衰减点。

运行前已说明的预期：在 k $= 1$、2、5、10 时出现瞬态（自动推动会抹去推广效果，而恒温器会抵消自身的推动）；在 k $= 0.5$ 时出现瞬态或衰减，因为响应较慢的恒温器可能无法在评估窗口开始前达到上限并持续十个周期的差距（尖峰首次达到 75% 交叉点时，k $= 0.5$ 对应第 109 个周期，因此瞬态更有可能）。达到 75% 和 50% 交叉点所需的时间会随增益增加而缩短，并在 k $= 5$ 时趋于平缓（尖峰对应的周期数：k $= 0.5$、2、5 时分别为 109、103、102）。预计 k $= 20$ 以下不会出现振荡。

4. 已登记规则：注意力预算三分位组上的子组反转（作者，2026-10- 29）

适用 于此条目之后的每次 Explore 运行；在运行 4（F4d）中首次读取，也可能在运行 1d 中读取 .

规则。在一次重复实验中，对于分组变量 v ∈ {budget （已登记）， a （SI 变体）}，根据周期 0 的总体（按照 metrics/effects.py 的分组方式）划分出三分位组 g∈ {0, 1, 2}。当 group_total_v_g 的窗口均值（周期 160–199）最大的三分位组与 group_empty_v_g 的窗口均值最大的三分位组不同时，翻转指标为 1；否则为 0；若任一组的序列缺失或其窗口均值不是有限值，也记为 0。按照重述的稳健性规则，翻转与类别一样计数：当 $\geq 90\%$ 的重复实验发生翻转，且翻转出现在相邻的 ≥ 2 个扫描点上、覆盖范围的 $\geq 10\%$ 时，该扫描点即为稳健。预测（*F4d*）：在优先级共享条件下，注意力预算较少的人会最先失去我们的助推，因此，最大值三分位组将从空场中预算较低或相等的组，转向拥挤场域中的高预算组；对于 a，不预测会发生翻转。

实现。 classes.parquet 会 subgroup_flip a （metrics/classes.py， FLIP_VAR = "a"）。预算变体根据 effects.parquet的 group_total_budget_<g>和 group_empty_budget_<g> 序列计算得出，而每次运行都会写入这些序列，因此运行 1–2b 无需更改代码；在运行 4 之前，先以测试优先的方式向 classes.parquet 添加一列 subgroup_flip_budget ，并在相同序列上应用相同规则进行计算。凡是报告其中一列的地方，都会同时报告两列。在 63af731 之前分类的运行目录（运行 1、1p、2、2b、2c）仅包含subgroup_flip ，除非使用 --force重新分类。已在 63af731 实现：在 metrics/classes.py 中， subgroup_flip(wide, win,var) 给出 subgroup_flip （a）和 subgroup_flip_budget （budget，紧随其后），而robust.parquet 和 robust_r9.parquet则分别为每个切片保留一行，包含 label = 列和 cls = "flip" （metrics.robust.robust_flag）。

5. 运行 1、2 和 2b 的先验知识（已披露）

- 比例共享下尖峰的已登记结果（尖峰注释 §4–§7），以及 §1.6 中披露的探测实验（类似 R5 的强度递增；增益 20 和 40 分别以 2/5 和 1/5 振荡；处理组占比为 0.2–0.5）。
- 构建阶段奇偶性运行（docs/analysis/2026-10-27-nudgefield-parity.md，比例共享下的六个固定值）以及优先级下的 Spark 运行（docs/analysis/2026-10-27-spark-timing.md：S1 现场实验，效应衰减且盲态_站点 1,002/1,002, ratio_window 0.63）。
- 目标运行（parameters.md §5–§6）：T1 在优先级下仍表现为均值比，10 次 S1 重复试验中有 4 次 ≥ 0.8；当 m 为 $= 0$ / 1 / 2 时，安静/拥挤比率分别为 1.81 / 2.16 / 2.25。
- 在此条目之前，没有运行过 S1_maximiser_smoothing 或 S1_persistent_auto_gain ；任何运行文件夹或助手的会话记录中，都不存在 ref_smoothing > 0 的探针，也不存在增益为 10 的探针。

6. 命令（来自 code/、M4、设备 mps）

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps"O=outputs/explore_20261029
uv run python -m nudgefield scenario=S1_competitor      design=pilot_rollout    $R
out_dir=$O/run1/S1_competitor_pilot_rollout, →
uv run python -m nudgefield scenario=S1_competitor      design=field_experiment $R
out_dir=$O/run1/S1_competitor_field_experiment, →
uv run python -m nudgefield scenario=S2_ally            design=pilot_rollout    $R
out_dir=$O/run1/S2_ally_pilot_rollout, →
uv run python -m nudgefield scenario=S3_bystanders      design=pilot_rollout    $R
out_dir=$O/run1/S3_bystanders_pilot_rollout, →
uv run python -m nudgefield scenario=S1_persistent_auto design=pilot_rollout    $R
out_dir=$O/run1/S1_persistent_auto_pilot_rollout, →
uv run python -m nudgefield scenario=S1_persistent_auto design=field_experiment $R
out_dir=$O/run1/S1_persistent_auto_field_experiment, →
uv run python -m nudgefield scenario=S1_maximiser       design=pilot_rollout    $R
out_dir=$O/run1/S1_maximiser_pilot_rollout, →
uv run python -m nudgefield scenario=S1_maximiser       design=field_experiment $R
out_dir=$O/run1/S1_maximiser_field_experiment, →
uv run python -m nudgefield scenario=S1_maximiser_smoothing  design=pilot_rollout    $R
out_dir=$O/run2/S1_maximiser_smoothing_pilot_rollout, →
uv run python -m nudgefield scenario=S1_maximiser_smoothing  design=field_experiment $R
out_dir=$O/run2/S1_maximiser_smoothing_field_experiment, →
uv run python -m nudgefield scenario=S1_persistent_auto_gain design=pilot_rollout    $R
out_dir=$O/run2b/S1_persistent_auto_gain_pilot_rollout, →
```

成本估算（M4 mps，计时说明 §4：样本量为 $= 20{,}000$ 时，每 10 次运行耗时 2.5 秒，耗时与运行次数成正比）：运行 1 约 2,600 次运行 ≈ 11 分钟；运行 2 约 2,000 次运行 ≈ 8 分钟；运行 2b 约 500 次运行 ≈ 2 分钟；总计约 25 分钟，另加 S3 m $= 16$ 批次的额外参与者。

7. 台账

Date	条目	提交	状态
2026-10-29	第0节–§6：运行 1、2、2b 已登记；翻转规则 §4	316abf0	已登记
2026-10-29	作者已签字确认（“Spark 现在很忙，所以我们可以在这里运行”）；运行 1、2、2b 按 §6 顺序在 M4（MPS）上开始运行	d8b0118	run
2026-10-29	运行 1、2、2b 完成（调用 11 次，23 分钟，每份清单均在 d8b0118 中保持干净，无分块）；结果见 docs/analysis/2026-10-29-explore-runs-1-2.md；除以下情况外均符合预期：盟友比率 0.46（而非 0.85–0.95），m $= 16$处仅为空值，未出现反转在参考_平滑值为 0.1 时（触发器从未触发）；F2.yaml 已更新，由作者批准	此提交	done
2026-10-29	第8节：运行 1p 已登记（运行 1 和 2 的十次调用在 run.attention=proportional，输出 explore_20261029/run1p/）；等待作者签字批准	f3329d2 (已修订 1565a75)	已登记
2026-10-29	第9节：运行 2c 已登记（S1_maximiser_memory, λ 0.02–0.16 $\times$ τ_自动 {0.3, 0.5}，两种设计，输出 explore_20261029/run2c/）；等待作者签字批准	f3329d2 （已修订 1565a75）	已登记
2026-10-29	作者确认第8节和第9节按原文保留（ $\lambda = 0.1$ 切片的 1pp/1pq 仅作描述性报告，未登记任何预期）；运行 1p 和 2c 从 M4（MPS）起点开始	fc1d622	run
2026-10-29	运行 1p 和 2c 完成（十二次调用，18分钟，每份清单均位于 fc1d622 中且无误）；结果见 docs/analysis/2026-10-29-explore-runs-1p-2c.md；表格 v1 位于 docs/analysis/2026-10-29-rule-independence-v1.md；反转的边界位于 λ之间 τ_处均为 0.04（反转稳健）和 0.08（衰减稳健）自动，比预测的 0.06–0.10 低一个步长（反转不稳健：在 0.06 处为 68 / 100 和 69 / 100），而现场实验的盲法_站点在无 λ时稳健；7 项主张（21 个数据点中的 19 个）规则无关，类别相关时为空值依赖规则，τ_自动 0.3 处的持久自动站点类别相对于其登记预期；F2.yaml 未更改	f476b2e	已分析

Date	条目	提交	状态
2026-10-30	作者批准了 R9 重新分类差异表 `docs/analysis/2026-10-29-r9-reclassification.md`按原样（规范 §4.6 步长 1；6,782 次重复试验中的 372 次，即预计算集合）； `F2.yaml` 正如_预期，条目一直保持到运行 3 的恒温器半部，延迟说明重申（“截至目前尚未出现；运行 5 是已登记搜索”）；反转文本 =条件与测量边界；Spark 半部设计 §6 裁定 1–4（`docs/superpowers/specs/2026-10-30-modelling-solidity-spark-design.md`）	此提交	已批准
2026-10-30	§10：运行 3 已登记（`S1_headroom_thermostat`, `S1_headroom_maximiser`, `S1_headroom_trigger`；τ 0.20–0.60 × τ_自动 0–0.50，试点 → 推广；五次调用， `run.max_batch_runs` 2,000, Spark CUDA；输出 `explore_20261029/run3/`）；等待作者签字批准	此提交	已登记
2026-10-30	§11：运行 4 已登记（`crowded_load`；预算 0.35–4.0 × 分钟 0–16，两种设计、两条规则；四次调用， `run.max_batch_runs` 800 = 一个 m 批次，因此未分块；Spark CUDA；输出 `explore_20261029/run4/`）；等待作者签字批准	此提交	已登记
2026-10-30	第12节：运行 5 已登记（`S2_ally_cycles`；盟友增益 0.5–40 ×滞后 {1, 2, 5, 10} 通过覆盖，两条规则；八次调用，M4 mps；输出 `explore_20261029/run5/`）；等待作者签字批准	此次提交	已登记
2026-10-30	第13节：运行 6 已登记（`S1_treated_share`；p 0.1–0.6，现场实验，均适用两条规则；两次调用，M4 mps；输出 `explore_20261029/run6/`）；等待作者签字批准	此次提交	已登记
2026-10-30	作者签署确认§10、第12节和第13节按原文执行，第11节则附带三项“计入”条款已移除（每个条目单独列出，并附其优缺点）；运行3和4从 Spark（CUDA、tmux）开始，运行5和6在 M4（MPS）上，于此次提交中	此次提交	run
2026-10-30	运行 3–6 已完成（共十九次调用；未恢复，未停止预热期；每份清单 `status`已完成， `burn_in_settled` 为真；清单位于 181c143（签核提交中的仅文档提交 65914f2；两者之间无代码更改：HANDOFF.md 以及仅每日记录）， `git_dirty`为假）。Spark（CUDA、一个 tmux 脚本、顺序执行，3 小时 40 分钟； `elapsed_s` / 峰值记忆 `torch.cuda.max_memory_allocated`）：3a 1,759 秒 / 7.04 GiB，3b 258 秒 / 4.28 GiB，3c 1,764 秒 / 7.04 GiB，3d 258 秒 / 4.28 GiB，3e 1,765 秒 / 7.04 GiB（四个分块，每块分别为 2,000/2,000/2,000/400）；4a 3,110 秒 / 19.36 GiB，4b 532 秒 / 7.36 GiB，4c 3,122 秒 / 19.35 GiB，4d 594 秒 / 7.36 GiB（八个未分块的 m 批次，每批 800；该 m = 16 优先批次为 20.79 GB，预测值为 20.75 GB，无回退）。M4（MPS，35 分钟）：5a–5d 分别为 448 / 445 / 445 / 445 秒，5e–5h 各 36 秒，6a 为 133 秒，6b 为 30 秒。清单已跟踪（运行 5 和 6 在 f0a1384 中，运行 3 和 4 在此处）；Parquet 已加入 Git 忽略，Spark 文件已复制到相同路径，且 SHA-256 校验值相同，在两台机器上	此提交	done
2026-10-30	运行 7（`targets` 在运行 3–6 中；运行 6 的 T1 广播来自运行 1 / 1p S1 增益 2 由 `--twin-cell` 提供，faf42c5），规则独立性表 v2（`docs/analysis/2026-10-30-rule-independence-v2.md`：82 行，28 条规则无关，48 条类别相关，6 条在大小；v1 的行保持不变），未找到列表（`docs/analysis/2026-10-30-not-found.md`：振荡、延迟、盲_试验，反向_站点稳健无处）以及结果说明 `docs/analysis/2026-10-30-explore-runs-3-6.md`产生了（八个偏差，第1节；F2.yaml 未更改：没有稳健的如_预期单元格约束区域内的 3a）	此次提交	分析了

Date	条目	提交	状态
2026-10-31	T1 探测（docs/analysis/2026-10-31-t1-probe-two-actor-field.md）：拥挤族的恒温器在 预热期、反应 ≈ 0，T1 ≈ 1 按构造，T2 的 2.25 为静态； 十次重复可行性探测（说明 §7，已披露）：竞争者 + 旁观者仅将余量维持到 m ≈ 2，更强的 竞争者将其抹去。作者选择校准映射作为一个 可行性研究；**第 8 次运行已登记（第14节），情景 crowded_headroom；停止以等待作者签字批准**；工程事项 nudgefield.levels 运行前	此提交	已签署 按 2026-10-31 所写； levels 工具 next
2026-10-31	第 8 次运行完成（六次调用、9,800 次运行，以及 levels 重跑中的 8c、8d、8q、8qp；Modal，每次调用使用一个 T4 容器， run.device CUDA，第14节设备偏差；每份清单均为 59e8b74, status 完成， git_dirtyfalse； elapsed_s 8a 2,077 秒，8b 211 秒，8c 2,072 秒，8d 249 秒，8q 30 秒，8qp 8 秒；于 15:42 启动，退出 当地时间 16:53）。 docs/analysis/2026-10-31-run8- 的结果 calibration-map.md（code/scripts/calibration_map.py）：a **优先级下存在联合单元格**：（τ 0.35，m 2）和（0.35，3） 符合判据 1–5；校准点（0.35，2），T2 为 1.508，位于 区间下边界处（标准误 0.022），T1 为 0.640，余量为 0.735， 衰减 100；比例：（0.40，2）和（0.45，2）符合 1–4， 无联合单元（描述性）。决策规则：第一分支（T2 校准 crowded_headroom于（0.35，2）；运行 4h 其自身 条目； crowded.yaml 仍作为僵局情形），尚未采取	59e8b74	已分析
2026-10-31（晚上）	第 8 次运行的结果按第14节的决策规则应用：余量 族 crowded_headroom 在竞争者 τ 0.35 处已校准 在 m = 2 （T2 1.508，区间边界；m 3：1.733）处，僵局 族的校准保持不变；**run 4h 已登记（第15节），情景 crowded_headroom_load，由作者签署确认（裁定 docs/superpowers/specs/2026-10-31-run4h-draft-中的 1–4 entry.md）**；接下来在 Spark 上运行	此次提交	已签署； Spark 运行 next
2026-10-31（夜间）	**run 4h 完成**（四次调用，28,800 次运行，Spark GB10 CUDA，17:36–19:43 本地时间，7,641 秒；每份清单均为 ce1107d 完成、干净，预热期已稳定；图形处理器最高 77 °C，守护程序从未 等待； targets 四次调用）。结果见 docs/analysis/2026-10-31-run4h-headroom-load.md (code/scripts/headroom_load_map.py）：按已登记顺序排列的两条等值线，在 两条规则下（m = 0、1、3、6、12、16、–、– 时饱和，以及在 12、12、12、16、16、–、–、– 的预算下按优先级为空值；0、1、2、3、6、8、12、16，以及在比例规则下为 6、6、6、8、12、12、–、–）；T1 区域 37 / 72（优先级，每个单元格内部）和 26 / 72（比例）； 在两条规则下，衰减和瞬态内部的类别 → 简报 §6 规则满足：**来自 run 4h 的 F3b；运行 4 的结果见扩展数据**； 在现场实验中，盲_站点在预算为 4.0 时稳健（两条 条规则均满足）；四项预期偏差（T1 在 (1.0, 4) 时也成立； 预算为 0.35–0.5 时 T1 区域为空置；预算为 0.7–1.0 时，盲_站点在 m 为 0； 一个比例单元中有 3 个计数反对项， (1.0, 16)，在上限处为空值，T1 为 0.732，标准误为 0.136）；表格 v2 +32 行（114；旧行保持不变）	此提交	已分析； F3b 决策 已提取
2026-11-01	§16：第9次运行已登记（入场券，干净； T4_POINTS 15），并且 已由作者按原文签字确认	0639070	已登记
2026-11-01	§16：命令已更正（NUDGEFIELD_TARGETS_OUT，绝对 路径）；首次执行的偏差已披露	c220bcf	已更正
2026-11-01	§16：记录的第9次运行（M4 mps，20:00–20:03；33 份清单干净；9 通过；所有预期均已满足；表格、级别、清单、日志 已在 outputs/targets_20261101_200002/ 下提交）	8eab24e	运行；数据源 ED3
2026-11-01	§17：第10次运行已起草——峰值附近的精细 T4 网格（9 个级别， 参与率 0.30–0.50），待签字批准	a913951	已起草
2026-11-01	第17节：运行 10 已由作者按原文签字确认（精细 T4 网格，9 级别 0.30–0.50）	4b1b94e	已登记

Date	条目	提交	状态
2026-11-01	§17：记录的第10次运行（M4 mps，22:36–22:53；42份清单无误；10项通过；粗网格表 = 第9次运行的；参与率为 0.376 时的精细峰值，预言机给出的 0.047：在容差范围内，但不在一个网格内步长——已说明偏差，参见修订行）	dddc1dd（制品），6589748（运行行）	运行；数据源ED3
2026-11-01	§17：阅读段落已修订（平台区 → ED3 的计算值单标准误范围为 0.356–0.414，比 0.351–0.426 的范围高 0.0006；日期标记）	d9b93e6, 0bcec5a	已修订
2026-11-01	§17：最终全分支审查后，运行行已修订——ate_sd 预期仅在峰值附近得到满足（四个外部相邻差异均超过该值）：第二偏差，已注明；时序条款；parameters.md §7 已对齐	63f16a0	已修订
2026-11-02	§17：打印的首个相邻差异为 0.0020，表格给出 0.00195 → 0.0019（限定范围的重新审查，轻微）；分类未更改	66458d5	已更正

8. 运行 1p：比例共享下的已登记数据点（已登记 2026-10-29；是规则独立性表 v1 的数据源）

目的：建模稳健性规范 §4.2（G2，前半部分）。运行 1 的八次调用和运行 2 的两次调用，均采用 run.attention=proportional，其他条件完全相同，从而确保每项已登记主张（F2 条目、盟友点、τ_ 0 3, 2 S1_maximiser_smoothing auto . 下的持久自动 $\lambda = 0$ 站点类别，以及 τ_auto . – . 下的 F 反向数据点）在另一种注意力规则下都有对应项。不更改场景文件。规则独立性表 v1（python -m nudgefield.reports.rule_independence, code/conf/nudgefield/reports/rule_independence.yaml 中的主张，其规则 B 路径为下方的输出目录）将这些目录视为规则 B，将运行 1–2 视为规则 A；该表在运行后生成，并非此条目的命令之一。

设置：除比例外，§ 0 run.n_rep 100 population.n 20 000 run.seed_base 20261029 run.max_batch_ 始终使用（、、、、、相同的窗口和设计参数）。批次种子不包含注意力规则（scenario|design|structure|seed_base），因此每次调用的批次种子分别是其运行 1 / 运行 2 对应副本的批次种子。类别规则： run.class_rules r9（自 9466eed 起的软件包默认值；尖峰说明 §1.6 R 9 cls_v1 cls 1 2 9 classes_），并记录在 旁边；运行 –r9.parquet robust_r9.parquet 的副本通过其 R 附件进行比较（、）。代码：运行开始时所对应的干净提交，记录在每个 json manifest. 中。

Call	场景文件	设计	扫描值	Runs	输出（code/outputs/explore_20261029/run1p/...）
1pa	S1_competitor	试点_推广	competitor.gain {0.5, 2, 5}	300	S1_competitor_pilot_rollout
1pb	S1_competitor	场域_实验	competitor.gain {0.5, 2, 5}	300	S1_competitor_field_experiment
1pc	S2_ally	试点_推广	ally.gain {0.5, 2, 5}	300	S2_ally_pilot_rollout
1pd	S3_bystanders	试点_推广	bystander.count {0, 2、4、8、16}	5 个批次 × 100	S3_bystanders_pilot_rollout
1pe	S1_persistent_auto	试点_推广	competitor.escalation_tau {0.1, 0.2, 0.3}	300	S1_persistent_auto_pilot_rollout
1pf	S1_persistent_auto	场域_实验	competitor.escalation_tau {0.1, 0.2, 0.3}	300	S1_persistent_auto_field_experiment
1pg	S1_maximiser	试点_推广	competitor.escalation_tau {0.1, 0.2, 0.3}	300	S1_maximiser_pilot_rollout
1ph	S1_maximiser	场域_实验	competitor.escalation_tau {0.1, 0.2, 0.3}	300	S1_maximiser_field_experiment
1pp	S1_maximiser_smoothing	试点_推广	escalation tau 0.1_ ... 0.5 × ref_smoothing {0, 0.1}	1,000	S1_maximiser_smoothing_pilot_rollout

Call	场景文件	设计	扫描值	Runs	输出（code/outputs/explore_20261029/run1p/...）
1pq	S1_maximiser_smoothing	场域_实验	same	1,000	S1_maximiser_smoothing_field_experiment

4,600次运行；在M4上约需18分钟（规范§4.2）。范围检查：与第1节和第2节一致（网格未更改）；重述的稳健性规则按原文适用。

运行前已说明的预期类别。规范第4节 .2，逐字如下：

> 根据比例（n_rep 10）下尖峰的结果以及奇偶性固定点登记了预期：S1 在 0.56 处衰减（优先级 0.63：大小差异低于 0.1 阈值）；盟友为 0.45–0.47（与优先级相同）；空值位于 m = 8 和 16（优先级：仅 16：存在稳健性差异）；每个 τ_自动均为瞬态；反向位于 0.3，并据此在棘轮机制下延伸至 0.4–0.5；每个 S1 和 S1max 现场实验点均为盲站点；持续自动现场实验位于 τ 0.3 分割处（尖峰：盲_站点 5、反向_站点 4、衰减 10 个中的 1 个；优先级：25 / 75），因此预期该点的站点类别依赖于规则。

每次调用。每个数据点的预期有三个来源之一，并在其单元格中注明：规范 = 所引用的§4.2句子（该句子提及的每个数据点）；*§1 / §2* = 本文件中为运行1–2登记的预期（已登记为优先级预期，依据n_rep 10时尖峰的比例结果表述），用于规范句子未指明类别的情况（1pd m = 0，2，4 1 τ_ 0 1, 0 2 1 τ_ 0 1, 0 2 1 $\lambda = 0$ τ_ 0 1, 0 2 = ；pf 自动 . .；pg 自动 . .；pp 位于自动 . .）；无 *</style}*未登记比例预期：*1pp*和*1pq*的*.1*切片（*46*个数据点中的*10*个）。第*2*节的*.1*预期（与相同的类别）在优先级下被运行*2*证伪，因此不予沿用；这*10*个数据点仅作描述性报告，不纳入规则独立性表。按照此处的解读，规范中的"每个*S1max*现场实验点均设有盲站点"并不涵盖*.1*处的*1pq*。作者可以在签字批准时，将任何*§1 / §2* 或无 单元格替换为已说明的预期。一个单元格包含推导结论，并标记为此类（1pc）。规则A是优先级下运行1 / 运行2的结果，R 9 docs/analysis/2026-10-29-explore-runs-1-2.md docs/analysis/2026-10- 类别（，）。29-r9-reclassification.md

Call	比例预期	规则A（优先级）结果
1pa	每个 k 均衰减，比率约为 0.56（规范）	衰减 100 / 100 / 100，比率 0.62–0.63
1pb	cls_combined盲法_站点在每个 k（规范）	盲法_站点 100 / 98 / 100
1pc	每个 k 处的比率为 0.45–0.47（规范）；推导出的衰减类别在 R9 下，规范未说明（窗口均值 0.46 在.....之外 [0.75，1.25], 末端比率 $\leq$ 0.75)	衰减 96 / 99 / 97，比率 0.46
1pd	在 m = 8 和 16 处为空值（规范）；规范未提及 m = 0、2、4 规范：符合预期，在 m = 0通过构造，衰减在 m = 2，4（第1节，1d，比例下的尖峰）	符合_预期 100；衰减 100；衰减 100；衰减 100；空值 100
1pe	每个 τ_自动均为瞬态（规范）	瞬态 100 / 100 / 100
1pf	τ_自动 0.3：在盲法_站点与反向_站点之间分割，站点类别预期依赖规则（规范）；0.1、0.2 未由 该规范规定：盲法_站点（第1节，1f）	cls_combined 盲法_站点 99 / 93 / 25，反向_站点 0 / 6 / 75
1pg	在 0.3 处反向（规范）；0.1、0.2 未由该规范命名：0.1 瞬态或空值，0.2 瞬态（第1节，1g）	瞬态 100 / 100 / 1，反向 0 / 0 / 99
1ph	cls_combined 盲_站点在每个 τ_自动（规范）	盲_站点 98 / 97 / 99
1pp	$\lambda = 0$：在 0.3 处反向，并由此延伸至 0.4–0.5（规范）；0.1 空值或瞬态，0.2 为瞬态（§2）。 $\lambda = 0.1$：无（没有在比例规则下登记的预期；§2 的 $\lambda = 0.1$ 在优先级规则下的预期被运行 2 推翻；报告为描述性结果，不在规则独立性表中）	$\lambda = 0$：瞬态 100 / 100，反向 99 / 100 / 100； $\lambda = 0.1$：每个点均衰减至 100 τ_自动，比例 0.61，自动助推从未触发
1pq	$\lambda = 0$： cls_combined 盲法_站点，每个 τ_自动（规范："每个 S1max 现场实验点"）。 $\lambda = 0.1$：无（因为 1pp $\lambda = 0.1$；该规范的"每个 S1max 场域现场实验点"确实未涵盖它）	cls combined： $\lambda = 0$：盲_ _站点 97 / 99 / 98 / 97 / 96； $\lambda = 0.1$：正如_所预期的，在每个位置均为 100 τ_自动（R9）

规则独立性标签（规范 §4.2：类别相关、大小相关，比例为 $>$ 0.1 时依赖规则；否则规则无关；优先匹配）由表格脚本根据目录计算得出，本文未予说明。引文所隐含的标签为：S1 衰减，规则无关（大小差异低于 0.1

阈值）；盟友规则无关；空值在类别上依赖规则（稳健性差异）；在 τ_auto 0.3 下，持久自动站点类别依赖规则。

哪些情况会反驳模型的叙述：一个已注册的 F2 类别（衰减、盲站点、瞬态、反向）在不同规则下改变多数类别或稳健性；此时该主张会被标记为依赖规则，文本中也会明确说明。

先验知识（已披露）。第 5 节中的全部内容；运行 1、2 和 2b 在优先级（ `docs/analysis/2026-10-29-explore-runs-1-2.md` 9 `docs/analysis/2026-10-29-r9-` ）下的结果；其 R 重新分类（）；尖峰在 n_rep 下的比例结果，以及奇偶性固定点 P–P（比例）。尚未进行 `reclassification.md` 10 1 6 `S1_maximiser_smoothing` 在比例条件下的运行；对于十次调用中的任何一次，都不存在 n_rep 100 下的比例运行。

命令（来自 `code/`、M4、设备 `mps`）：

使用 `bash` 运行该代码块（将其粘贴到 `bash` shell 中，或将其封装为 `bash <<'EOF' ..... EOF`）： `$R` 未加引号的内容会被 Bash 按空格分割，但不会被 Z shell（macOS 终端的默认 shell）分割。

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps run.attention=proportional"
O=outputs/explore_20261029/   1
  runuv run python -m nudgefield scenario=S1_competitor      design=pilot_rollout    $R
out_dir=$O/S1_competitor_pilot_rollout, →
uv run python -m nudgefield scenario=S1_competitor      design=field_experiment $R
out_dir=$O/S1_competitor_field_experiment, →
uv run python -m nudgefield scenario=S2_ally            design=pilot_rollout    $R
out_dir=$O/S2_ally_pilot_rollout, →
uv run python -m nudgefield scenario=S3_bystanders      design=pilot_rollout    $R
out_dir=$O/S3_bystanders_pilot_rollout, →
uv run python -m nudgefield scenario=S1_persistent_auto design=pilot_rollout    $R
out_dir=$O/S1_persistent_auto_pilot_rollout, →
uv run python -m nudgefield scenario=S1_persistent_auto design=field_experiment $R
out_dir=$O/S1_persistent_auto_field_experiment, →
uv run python -m nudgefield scenario=S1_maximiser       design=pilot_rollout    $R
out_dir=$O/S1_maximiser_pilot_rollout, →
uv run python -m nudgefield scenario=S1_maximiser       design=field_experiment $R
out_dir=$O/S1_maximiser_field_experiment, →
uv run python -m nudgefield scenario=S1_maximiser_smoothing design=pilot_rollout    $R
out_dir=$O/S1_maximiser_smoothing_pilot_rollout, →
uv run python -m nudgefield scenario=S1_maximiser_smoothing design=field_experiment $R
out_dir=$O/S1_maximiser_smoothing_field_experiment, →
```

9. 运行 2c：竞争者记忆中反转的边界（已登记 2026- 10-29；作为参考规则的 G1 和扩展数据面板的数据源）

目的：建模稳健性规范 §4.3（G1）。测量最大化者的指数参考在何种记忆长度下反转会停止：据此将反转表述为带有边界的条件。本次运行替代运行 2 中的 $\lambda = 0.1$ 部分，作为参考规则的扩展数据面板；运行 2 仍作为已登记的 τ_自动稳健性运行。

场景文件 `S1_maximiser_memory`（新建，本次提交）： `S1_maximiser` 的 R7 最大化者（最大化者，增益 2，滞后 2，reach0 0.60，R6 的持续自动升级），以 `competitor.escalation_tau` $\in$ {0.3, 0.5} 作为切片，以 `competitor.ref_smoothing` （λ） $\in$ {0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16} 作为稳健性轴（最后一条扫描路径）；16 个数据点；两种设计；每种设计一个包含 1,600 次运行的批次。设置：第0节（优先注意力，n_rep 100，样本量 20,000，种子基数 20261029，不进行分块）；类别规则 R9（`run.class_rules` r9， `cls_v1` 记录在旁）。

Call	设计	稳健性轴	切片	Runs	输出（`code/outputs/explore_20261029/run2c/...`）
2cp	试点_推广	`ref_smoothing` 0.02 ... 0.16，步长 0.02（记忆为 1/λ，从 50 到约 6 个周期）	`escalation_tau` 0.3、0.5	1,600	`S1_maximiser_memory_pilot_rollout`
2cf	场域_实验	same	same	1,600	`S1_maximiser_memory_field_experiment`

在 M4 上每种设计约需 7 分钟（规范 §4.3）。范围检查：八个等间距数据点；两个相邻数据点覆盖 0.02 / 0.14 = 14% 的范围，因此，已注册稳健性规则（`run.robust_min_points` 2, `run.` `robust_min_range_share` 0 10 λ $\geq$ 90 .）按原文适用：如果某个类别在至少两个相邻数据点上均达到 % 的重复试验份额，则该类别在这两个或更多相邻数据点上是稳健的。网格采用作者裁定 2 1，0 2026 10 29 6 0，0 01，0 02，0 05，0 2，0 5 10 4 3 的 - -（规范第 § 节）；较早的网格 { . . . . . } 在 % 的范围占比要求下无法使任何边界稳健（规范 § . ）。

基准情况，规范 §4.3 原文如下：

> 棘轮（$\lambda = 0$）是一种不同的更新规则（`ref ← min(ref, r)`，即运行最小值，而不是平均值的 $\lambda \to 0$ 上限；后者会成为冻结的基准值），并且不是该轴上的一个点：其结果是运行 2 的 $\lambda = 0$ 切片，作为基准情况报告在扫描旁边。

运行 2 在本次运行的两个 τ_自动 $\lambda = 0$ 处的切片：试点 → 推广在 0.3 处反转了 100 个中的 99 个（比率 −0.34），而 100 100 0 5 −0 91 114 `cls_` 在 . 处（比率 . ）；自动助推在每次重`combined` 98 0 3, 96 0.5 复实验的周期均开启；现场实验盲站点在 . 处以及在 . 处。_

运行前陈述的预期类别。规范第4.3节，逐字引用（预期与机制）：

> 运行前陈述的预期。只有在观测速率高于 `ref` + 0.01 且触达率达到其上限后，自动助推才会在连续十个周期后触发。在运行 2 的 $\lambda = 0.1$ 条件下，触达率在周期中位数 107 时达到其上限（棘轮条件下为 105），因此十周期计数会在推广后 15–17 个周期结束；指数参考值已将推广约 0.05 的差距缩小至 0 05 1 − λ 16, 0 05 1 − λ 16 。 ( ) ˆ, = 0 01 λ ≈ 0 096, 而竞争者自身触达率的上升会进一步缩小该差距。求解 . ( ) ˆ . 得到 ., 而上升会将边界进一步压低。预测：在 $\lambda \leq 0.06$ 时反转（稳健），相较于 $\lambda \geq 0.10$ 有所衰减，其中 0.08 为不确定点；在试点 → 推广中，两个 τ均如此。现场实验则会在同一边界上，从盲站点（小 λ）转为符合预期（大 λ，在 R9 条件下）。该边界报告为：反转稳健时的最大 λ，以及衰减稳健时的最小 λ；记忆长度 1/λ 是文本所使用的数值。随后，文本将该行动者描述为其基准值是历史最高点（棘轮）或具有长记忆的行动者，并将边界作为条件；命名这类现实参与者（例如锚定于过去最佳周期的销售目标、会棘轮式上调的预算）则是作者在写作阶段的任务。

每个点的预测都相同：

Call	λ 0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16
2cp, τ_自动 0.3 和 0.5 (`cls`)	反向（稳健）	反向（稳健）	反向（稳健）	不确定	衰减（稳健）	衰减（稳健）	衰减（稳健）	衰减（稳健）
2cf, τ_auto 0.3 和 0.5（`cls_combined`）	盲目 _site	盲审 _site	盲目 _site	不确定	as_ 预期	as_ 预期	as_ 预期	as_ 预期

预测边界：介于 λ 0.06 和 0.10 之间（记忆为 1/λ，约为 17 至 10 个周期）。报告：反向稳健时的最大值，以及衰减稳健时的最小值，并注明每个点的触发周期（`trigger_free_competitor@auto`, `classes.parquet`）。

哪些情况会反驳该预测：在 $\lambda \geq 0.10$ 时反向稳健，或在 $\lambda \leq 0.06$ 时不稳健，且两者均适用于 τ_自动；不存在单一边界（反向和衰减沿着 λ交错出现）；在现场实验中（`cls_combined`），盲_站点在 $\lambda \leq 0.06$ 时不稳健，或到 $\lambda \geq 0.10$ 时没有如_预期般发生偏移。以上任何一种情况均报告为已发现，且文本中的条件取自运行结果，而非预测。

先验知识（已披露）。在优先级（`docs/analysis/2026-10-29-explore-runs-1-2.md` §2）下运行 2：使用棘轮时，反向在 τ_自动 0.3–0.5范围内稳健；使用 $\lambda = 0.1$ 时，自动助推从未触发，并且在比率 0.61 下每个点均为衰减（试点 → 推广），且在现场实验中符合 R9 的预期。运行 2 的 $\lambda = 0.1$ 在 τ_自动 0.3 和 0.5 时，与本次运行在不同情景名称下的 $\lambda = 0.10$ 属于同一参数点，因此批次种子不同。触达率达到其上限时的周期中位数（ 107 $\lambda = 0$ 1, 在棘轮下的 .）来自运行结果，并进入上述推导。105 2 $\lambda \in$ 0 02, 0 04, 0 06, 0 08, 任何运行文件夹中都不存在带有 { . . . .0 12, 0 14, 0 16 `S1_maximiser_memory` . . . } 的运行；在此条目之前也未进行过任何运行。

命令（来自 `code/`、M4、设备 `mps`）：

运行该 代码块，使用 `bash`（将其粘贴到 `bash` shell 中，或将其包装为 `bash <<'EOF'` 。 。 。 `EOF`）：`$R` 不带引号 is 按单词分割 Bash 支持，但 macOS 终端默认使用的 Z shell 不支持。

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps"
O=outputs/explore_20261029/run2c
uv run python -m nudgefield scenario=S1_maximiser_memory design=pilot_rollout    $R
out_dir=$O/S1_maximiser_memory_pilot_rollout,→
uv run python -m nudgefield scenario=S1_maximiser_memory design=field_experiment $R
out_dir=$O/S1_maximiser_memory_field_experiment,→
```

10. 运行 3：余量 × 持续性，三个半（注册于2026-10-30；数据源 F3a、F2 的符合预期点、ED4 以及表格 v2 的响应规则行）

目的：Spark-half 设计 `docs/superpowers/specs/2026-10-30-modelling-solidity-spark-design.md` 第4.4节（G4）。针对速率目标竞争者（恒温器），将竞争者的余量与其自动推动的持续性进行映射；

最大化可达性的竞争者（R7 最大化器），以及采用第二种响应规则（阈值触发器）的恒温器，从而确定每个余量切片中沿持续性的类别，以及在两种注意力规则下的类别。2026-10-30 的作者裁定（设计第6节）：1，`competitor.nudge.tau` 0.20–0.60（删除 0.15：恒温器在预热期达到饱和，其升级在试点之前触发；设计第7节发现3）；4，阈值触发增益 0.05，一个值。同次会议的裁定（c）：F2 的按预期条目保持不变，直至本次运行的恒温器半部运行完毕。

场景文件（本次提交新增），一个网格，仅进行试点 → 推广：

- `S1_headroom_thermostat`：S1 竞争者（恒温器，增益 2，目标_偏移量 −0.03，滞后 2，reach0 0），并启用升级。
- `S1_headroom_maximiser` 7 2, 2, 0 0 60, `ref_smoothing`：R 0 最大化者（增益 滞后 触达率 . 棘轮参考 ）并启用升级。• `S1_headroom_trigger`：恒温器半部，参数为 `rule: threshold_trigger`、 `gain: 0.05`、 `tolerance` 0.01（ActorSpec 默认值）。• 每个网格中的数据为：切片 `competitor.nudge.tau` ∈ {0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.60}；稳健性轴（最后一条扫描路径） `competitor.escalation_tau` ∈ {0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50}； `sweep: [competitor.nudge.tau, competitor.escalation_tau]` 64 6,400 `escalation_tau` ；数据点，每次调用运行次数。0 会在强度为零时开启升级，因此仍会记录触发器（设计 §7，已通过构建器检查确认）。

设置：§0（`run.n_rep` 100, `population.n` 20,000, `run.seed_base` 20261029，相同的窗口和设计参数），但以下项除外：`run.max_batch_runs` 2,000 和 `run.device` cuda（Spark），每次调用 `run.attention` 。分块：每次调用对应一种参与者结构，因此一个批次包含 6,400 次运行；按单个行动者运行约 4 MB 计算（M4 清单、设计 §3），一个批次约为 25 GB，超过设计 §3 的 20 GB 分块上限，因此将其分为四块，每块包含 2,000, 2,000, 2,000 400 `scenario|design|structure|seed_` 次运行，并设置随机种子（`c...`）。这与序言中的 `base|chunk<c>` = 0 3 `run.max_batch_runs` = 0 不同：各块的随机种子不同于未分块批次的随机种子，因此抽样结果并非来自未分块调用；此前没有运行使用这些场景文件，所以没有任何比较依赖于此。4 MB 是一个 MPS 数值；CUDA 斜率仅针对拥挤族测量（任务 7，`docs/analysis/2026-10-27-spark-timing.md` §7），因此没有采用设计中的替代方案（若 CUDA 测量值低于 MPS 数值，则使用一个批次）。批次种子不包含注意力规则，因此每次比例调用的块随机种子与其优先级对应项相同。类别规则 R 9 `run.class_(r`，记录在旁边)。代码：运行开始时所用的干净提交，记录在每个 `manifest.json` 中。 `rules_9_cls_v1`

Call	场景文件	设计	注意力	Runs	输出（`code/outputs/explore_20261029/run3/...`）
3a	`S1_headroom_thermostat`	试点_推广	优先级	6,400 (4 块)	`S1_headroom_thermostat_pilot_rollout`
3b	`S1_headroom_thermostat`	试点_推广	比例	6,400（4 块）	`S1_headroom_thermostat_pilot_rollout_proportional`
3c	`S1_headroom_maximiser`	试点_推广	优先级	6,400（4 块）	`S1_headroom_maximiser_pilot_rollout`
3d	`S1_headroom_maximiser`	试点_推广	比例	6,400（4 块）	`S1_headroom_maximiser_pilot_rollout_proportional`
3e	`S1_headroom_trigger`	试点_推广	优先级	6,400（4 块）	`S1_headroom_trigger_pilot_rollout`

触发半部仅按优先级运行：表格 v2 的响应规则分支，将优先级条件下的恒温器半部作为对照读取（3a 对比 3e）；其比例版本将构成第四个半部，但尚未登记（设计 §4.4）。32,000 次运行。耗时（设计 §4.4：Spark 上优先级条件下，单个行动者每次运行需 0.29 秒；比例条件下需 0.05–0.06 秒）：两个优先级半部（3a、3c）约需 1.0 小时，触发半部（3e）需 0.5 小时，两个比例半部（3b、3d）需 0.2 小时；总计约 1.7 小时。

范围检查：在稳健性轴 `escalation_tau` 0–0.50 上，最小相邻间隔为 0.05，占范围的 10%，恰好等于已登记的份额； `metrics/robust.py`将跨度与容差 10^{-9} (`_EPS`) 进行比较，因此任意两个相邻数据点都符合条件，重述后的稳健性规则（ `run.robust_min_points` 2 `run.robust_min_` ,.）按原样适用于每个余量切片。 `range_share` 0 10

运行前声明的预期类别。设计 §4.4，原文：

> 恒温器半部：在 τ_自动 0 时符合预期，或在 tau 0.20 时衰减（触达率为 0.9–1 的竞争者施加的推力，是其在 0.30 时推力的三分之二，这使比率接近 0.75 的区间边缘；映射答案属于哪一侧），并且从 tau 0.25 起开始衰减（tau 0.30 时已登记为 0.63）；一旦 τ_自动 $\geq$ 0.10，在每个 tau 下均为瞬态（tau 0.30 时已登记；尖峰 §6.6：速率定位主体抵消了自身的持续推力，因此没有单元格发生反转）。最大化器半部：在 τ_自动 0 时衰减（运行 2

在平滑处理下：0.61，且无触发器），在 0.10–0.20 时为瞬态（运行 1），在 tau 0.30 时从 0.30 起反向（运行 1、运行 2），随着 tau 增大，从较低的 τ_自动 值开始反向；随着 tau 减小，则从较高的值开始反向（自动推动为 τ_自动 · 触达率 · 易感性；只有当它超过我们的助推所增加的量时，才会发生反转），在 tau 0.20 下，该列可能在任何 τ_自动 值下都无法反转。触发半部：恒温器半部的类别逐单元格对应（表格 v2 的响应规则行）；在步长为 0.05 时，预计不会发生振荡（开关控制超调只有一个步长，低于振荡检测器的振幅）。F2 符合预期点是恒温器半部中，在 τ_自动 0 时距离 tau 0.30 最近的稳健的符合预期单元格（如果存在）（简报第 5 节；交接事项中的作者条目（c））。

每次调用。来源：规范 = 引用的第4.4节原文；无 = 原文未为该单元格指定类别（未登记预期；以描述性方式报告；作者可在签字批准时说明一个类别）。原文未指定注意力规则；本条目将其解读为适用于两条规则（3a与3b同样适用，3c与3d同样适用）。类别 `cls` (R9)，试点 →推广。

Call	τ_自动 0	τ_自动 0.05	τ_自动 0.10，0.15, 0.20	τ_自动 0.30，0.40, 0.50
3a、3b（恒温器）	τ 0.20：如_预期般或衰减；τ 0.25–0.60：衰减（规范）	none	瞬态发生于每个 τ（规范）	瞬态于每个 τ；无单元格反向（规范）
3c、3d（最大化者）	衰减（规范；来源、运行 2 的 $\lambda = 0.1$ 切片位于 τ 0.30；每个 τ 的内容均在此读取，因为在强度为 0 时没有任何反转）	none	τ 0.30：瞬态（规范，运行 1）。其他 τ：参见起始 row	τ 0.30：反向（规范，运行 1 和 2）。其他 τ：参见起始行
3c, 3d, 起始	最小的 τ _自动，此时该列反向列在 τ 高于 0.30 时较低，而在 τ 低于 0.30 时较高；在 τ 0.20 时，该列可能在任何 τ_自动（规范）中都不反转。没有逐单元格类别是根据 τ 0.30 列推断得出的：这段文字给出了排序，而非单元格			
3e（触发器）	同一单元格中 3a 的类别，逐个单元格对应（规范）	类别 3a 中的同一单元格（规范）	相同（规范）	相同（规范）
3e, 处处	没有任何单元格振荡（规范）			

F2 的符合预期点（裁决 (c)）：在 τ_自动 0 时，3a 中最接近 τ 0.30 的稳健符合预期单元格 _；如果不存在，则由作者在运行后决定回退方案（设计第6节，裁决 (c)）。

哪些情况会与预期相悖：稳健多数类别与所述类别不同的单元格、恒温器半部中的反向单元格、最大化器半部中未随 τ 单调移动的反转起点、3e 中的振荡，或类别与 3a 不同的 3e 单元格。发现的情况均如实报告。

先验知识（已披露）。第5节和第8节中的所有内容；运行 1、2、2b、1p 和 2c（`docs/analysis/2026-10-29-explore-runs-1-2.md` `docs/analysis/2026-10-29-explore-runs-1p-2c.md` `docs/analysis/2026-10-29-r9-reclassification.md`），其 τ 数据点以其他情景名称位于此网格上（e / pe：τ_自动 0.1–0.3 时的恒温器；1g / 1pg 和 2p / 1pp $\lambda = 0.1$：τ_自动 0.1–0.5 时的最大化者）；设计第7节的构建器检查 CPU 探针（τ 0.15 和 0.20 时的恒温器；τ_自动 0 时的升级）。在任何运行文件夹中，都不存在这三个场景文件的运行、τ 不等于 0.30 的竞争者运行、`escalation_tau` 0 或 0.05 的运行，或 `threshold_trigger` 运行（2026-10-30 检查的 `code/outputs/` 下每个 `manifest.json`）。

命令（来自 `~/MultiActorBPP/code` Spark 上、tmux 中的克隆副本；位于记录作者签字批准的提交，且 `git status` 干净；设备 `cuda`）。使用 `bash` 运行此代码块（将其粘贴到 `bash` shell 中，或将其包装为 `bash <<'EOF' ..... EOF`）：`$R` 不加引号时，Bash 会按单词拆分，但 Z shell 不会。停止的调用会通过添加 `run.resume=true`，由同一命令继续（设计 §4.1）；清单会记录此事。运行 3 和运行 4 在 Spark 上依次运行，先运行 3。

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=cuda run.max_batch_runs=2000"
O=outputs/explore_20261029/run3
~/.local/bin/uv run python -m nudgefield scenario=S1_headroom_thermostat design=pilot_rollout $R
out_dir=$O/S1_headroom_thermostat_pilot_rollout,→
~/.local/bin/uv run python -m nudgefield scenario=S1_headroom_thermostat design=pilot_rollout $R
run.attention=proportional out_dir=$O/S1_headroom_thermostat_pilot_rollout_proportional ,→
~/.local/bin/uv run python -m nudgefield scenario=S1_headroom_maximiser  design=pilot_rollout $R
out_dir=$O/S1_headroom_maximiser_pilot_rollout,→
```

```
~/.local/bin/uv run python -m nudgefield scenario=S1_headroom_maximiser  design=pilot_rollout $R
run.attention=proportional out_dir=$O/S1_headroom_maximiser_pilot_rollout_proportional ,→
~/.local/bin/uv run python -m nudgefield scenario=S1_headroom_trigger    design=pilot_rollout $R
out_dir=$O/S1_headroom_trigger_pilot_rollout,→
```

11. 运行 4：拥挤场域中的注意力负荷（注册于2026-10-30；数据源为 F3b、F4d、ED4 以及表格 v2 的临界转变行）

目的：设计 §4.5（G4）。在两种注意力规则和两种设计下，描绘拥挤场域的类别随旁观者数量和总体注意力预算变化的情况，并回答优先级共享是否会产生比例共享所没有的临界线。2026-10-30 作者裁定 2（设计第6节）：预算网格为 0.35–4.0，保留 0.35 边界切片，并报告其排除份额。

场景文件 `crowded_load`（新增，本次提交）：由 `crowded.yaml`（恒温器竞争者和盟友的增益为 2, ∓0 03, 2, 0 0 0, 0 0 6 `population:`，目标偏移量 . 滞后 触达率；旁观者的增益为 触达率 .）构成的校准后的拥挤族，切片为 (因子；网格上有 T 点 m，预算 .)，稳健性轴为 `bystander.count` ∈ {0, 1, 2, 4, 6, 8, 12, 16}（最后一条扫描路径；已登记的空值主张沿 m 延伸）；`sweep: [population.budget_mean, bystander.count]`；64 个数据点；两种设计。

`{budget_mean: [0.35, 0.5, 0.7, 1.0, 1.4, 2.0, 2.8, 4.0]}` √2 2 = 2 1 0

设置：第0节（`run.n_rep` 100、`population.n` 20、000、`run.seed_base` 20261029，相同的窗口和设计参数，`fe_reach` 0.4），但以下内容除外：`run.max_batch_runs` 800 和 `run.device` CUDA（Spark），每次调用 `run.attention` 。 `bystander.count` 8 × 进入批次结构，因此 100 = 800 800 7 每次调用包含八个批次，每个 m 对应一个批次，预算的重复试验运行。上限来自任务测量（`docs/analysis/2026-10-27-spark-timing.md` 7 = 16 20,000, 25 9 0 806 §：拥挤 m、现场实验，N CUDA：每次优先级运行 . MB 和 . s；比例运行则为 9.8 MB 和 0.160 s；20 GiB / 25.9 MB = 828.6 次运行，向下取整为 800）。它等于一个 m 批次，因此运行 4 实际上是未分块的：每个批次作为一个分块运行，使用未分块批次种子 `scenario|design|structure|seed_base`，与前言中的 `run.max_batch_runs = 0` 一致（在场景文件上已检查：`expand(..., max_batch_runs=800)` 给出了八个包含 800 个项目的批次，且 `chunk` 为 None）。m = 16 优先批次的预测峰值：16 MB + 800 × 25.9 MB = 20.75 GB（19.32 GiB），处于 Spark 的 124 GB 统一内存之内，并为分配器开销留有余量，比十进制 20 GB 高出约 4%（补充说明 §7，注意事项）。该测量采用现场实验设计；试点 → 推广调用未进行测量（张量形状取决于 R、N 和行动者数量，而非设计，因此预期斜率可以沿用，但未经检查；补充说明 §7）。如果某个批次无法容纳，则适用前言中的回退规则（`run.max_batch_runs` 100，已记录在台账中）。批次种子不包含注意力规则，因此每个比例调用的批次种子都与其优先级对应项相同。逐次运行预算覆盖会保留每个批次的抽样结果（a、 s_auto、 s_delib，即对数正态预算因子），只调整预算（设计 §3、§4.3）；不声称将单元格成对匹配。类别规则 R9；现场实验调用读取cls_combined。 `classes.parquet` 携带 `subgroup_flip`（a）和 `subgroup_flip_budget`（§4）。

Call	设计	注意力	Runs	输出（`code/outputs/explore_20261029/run4/...`）
4a	试点_推广	优先级	6,400（8 批次 × 800）	`crowded_load_pilot_rollout`
4b	试点_推广	比例	6,400（8 批次 × 800）	`crowded_load_pilot_rollout_proportional`
4c	场域_实验	优先级	6,400（8 批次 × 800）	`crowded_load_field_experiment`
4d	场域_实验	比例	6,400（8 批次 × 800）	`crowded_load_field_experiment_proportional`

25,600 次运行。时间按任务 7 的单次运行成本计算，将 m = 16 处的成本应用于每个 m（这是上限：较小的 m 成本更低）：6,400 × 0.806 秒 = 5,158 秒，优先级调用约需 1.4 小时；6,400 × 0.160 秒 = 1,024 秒，比例调用约需 17 分钟；四次调用总计约需 3.4 小时。根据设计 §3 的两小时规则（按实测时间计算），任何一次调用的预算都不会超过两小时。

范围检查：在稳健性轴 m = 0–16 上，相邻间隔为 1、1、2、2、2、4、4，即占范围的 6.25%、6.25%、12.5%、12.5%、12.5%、25%、25%。仅配对 (0, 1) 和 (1, 2) 不满足 10% 的份额要求；三个数据点 0、1、2 覆盖 12.5%，满足要求；从 m = 2 开始，任意两个相邻数据点都满足要求。重述后的稳健性规则按原文应用于每个预算切片。

边界切片（预算 0.35），这是已声明的偏差。网格的下限由排除规则设定，而非由 y 选择。设 计 §4.5，原文如下：

> 较低端由排除规则而非选择决定：构建器检查在预算 0.25 处的探针，在两条规则下、m = 0 和 m = 16 的每次重复实验中均给出 `E` ≈ 0.016，低于 `SMALL_E` 0.02，因此该切片的每个单元格都会被 `excluded`（即使在平静场域中，预算匮乏者也会放弃我们自己的助推），而且其概率单位 0.044 无论如何都超出幅度界限；在 0.35 `E` ≈ 0.022

以及概率单位 0.062，这两个值都刚好位于界限内。0.35 切片作为边界切片保留，并按单元格报告其排除份额；在多数为excluded的任何单元格中， `tipping` 主张（§4.8）均未定义。

因此，0.35 切片可以在任意单元格中容纳 `excluded` 重复试验，这与网格中的其余部分不同；其排除份额按单元格报告，且只有在多数派不是 `excluded` 时，这些单元格才纳入临界主张。

运行前声明的预期类别。设计 §4.5，原文照录：

试点 → 推广：在每个预算水平下，$m = 0$ 处均为衰减（仅看两名反应参与者时也为衰减：运行 1 的 S1 和 S2 均为衰减，而盟友的身份会使数值减少 0.03）；中间区域均为衰减；当我们的权重低于零效应阈值时为空值；在优先级规则下，这是一条 (m, 预算) 上的线（我们的助推成本为 1、τ 0.30，每当个体的预算被高显著性助推耗尽时，就会舍弃我们的助推：这是一条临界线）；在比例规则下，则是平缓下降，在 $m \geq 8$ 、预算为 1.0 时达到空值（运行 1p），预算较小时则在更小的 m 值处达到空值；这张图上没有任何符合预期的情况（F2 的符合预期点来自运行 3，而非这里）。现场实验：只要竞争者抵消了站点比率、而我们的助推仍能触达处理组，就属于盲站点（m 较小、预算不限）；在 T2 点（$m = 2$，预算为 1.0：相对于平静场域，校准比率为 2.25）时为符合预期或衰减。在两种标签下，大 m 值且预算较小时均为空值。优先级规则是否会产生比例规则下不存在的临界线，正是这张图要回答的问题（简报 §6b）；主张的文本将根据答案确定，表格 v2 中对应行则按规则标为“存在／不存在临界线”。

子 组翻转，已登记规则 §4（作者，2026-10-29），本次运行（F4d）首先读取，照录 §4：

预测（*F4d*）：在优先级共享下，注意力预算较少的人会最先失去我们的助推，因此最大值三分位组会从空场中预算较低或相同的人群转向拥挤场域中预算较高的人群；不会对 `a` 预测翻转。

每次调用。来源：规范 = 所引用的 §4.5 段落；*§4* = 已登记的翻转规则；无 =没有为该单元格命名的类别。该段落说明的是区域，而非单元边界：其中提到“小 m”“大 m”或“中间”时，本条目并未固定任何 m 或预算；本次运行会报告边界所在位置。

Call	预期
4a（试点 → 推广、优先级）	$m = 0$：在每个预算下均衰减（规范）；在中部均衰减（规范）；在一个由一条线（临界线）界定的（m，预算）区域内为空值（规范）；在任何单元格中均如_预期（规范）。在拥挤场域中，预算（`subgroup_flip_budget`）发生翻转； `a` 不发生翻转（§4；该规则名称中不含 m 或预算）
4b（试点 → 推广，比例）	$m = 0$：在每个预算下均衰减（规范）；在中间范围内衰减（规范）；平缓下降在预算为 1.0 时于 $m \geq 8$ 达到空值，预算越小则在更小的 m 下达到空值（规范）；正如_预期，在任何单元格中都不会出现（规范）。翻转：没有预测（§4 说明只有在优先级下才有一项预测）
4c（场域实验，优先级）	`cls_combined`盲法_站点在较小的 m 下适用于任意预算（规范）；在_预期或衰减，因为 T2 点 $m = 2$，预算为 1.0（规范）；在两种标签中，较大的 m 和较小的预算下均为空值（规范）。翻转为 4a（§4；该规则未指定设计）
4d（场域实验，比例）	作为 4c（规范；该段给出了场域实验的预期，但没有规则）。翻转：否预测（§4）
临界转变（4a 对照 4b；4c 相对于 4d）	在优先级规则下为一条临界线，在比例规范下为平缓下降；表格 v2 的行根据规则，显示为“存在/不存在临界线”，以图中所示为准（规范）

哪些情况会与预期相悖：在任何试点推广单元格中，均应如预期般稳健；在优先级规则下不应存在零值单元格。每项均按实际发现进行报告，而临界主张的文字无论如何都遵循该图（作者在签字批准时，即 2026-10-30，删除了三个条款；这些条款增加了设计中未陈述的稳健性条件和规则差异：在 $m = 0$ 时为衰减稳健、小 m 时盲点站点稳健，以及临界主张中的一项规则差异）。

先验知识（已披露）。§5 和第8节中的全部内容；1d / 1pd 运行（`S3_bystanders`，不含两名反应参与者的旁观者：仅在优先级规则下、$m = 16$ 时为空值，第8节表格；在比例规则下、$m = 8$ 和 16 时为空值，设计 §4.5 和 §4.2）；目标运行（`parameters.md` §5–§6：$m = 2$ 时校准后的安静/拥挤比例为 2.25；预算为 1.0）；构建器检查对预算 0.25–0.35 下拥挤场域进行的 CPU 探针检查（设计 §7 发现 2）；任务 7 测量（`docs/analysis/2026-10-27-spark-timing.md` §7： `crowded`，现场实验，$m = 16$，预算 1.0，样本量 20,000, 100 400 `code/`，其中每一 `outputs/targets_20261027_* parameters.md` 5 1 0 $\in$ 0, 1 个该运行在两条规则下均为空值；并非已登记运行）。目标运行（，§）在预算 . 下、m { } 时运行了拥挤族，并带有 { 3, 4, 6, 8 `crowded_load` 1 0, $= 12$ }。在任何运行文件夹中，都不存在关于、预算不是 . 的运行，也不存在拥挤族在 m 时的运行（每个 `manifest.json` 均在 `code/outputs/` 上于 2026-10-30 进行了检查：全程为 `budget_mean` 1.0）。

命令（在 `~/MultiActorBPP/code` 上的 Spark 中，通过 tmux，在运行 3 之后，克隆副本位于记录作者通过 `git status` clean 签字批准的提交处；设备 `cuda`；使用 `bash` 运行代码块；通过添加 `run.resume=true` 继续已停止的调用）：

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=cuda run.max_batch_runs=800"
O=outputs/explore_20261029/run4
~/.local/bin/uv run python -m nudgefield scenario=crowded_load design=pilot_rollout    $R
out_dir=$O/crowded_load_pilot_rollout,→
~/.local/bin/uv run python -m nudgefield scenario=crowded_load design=pilot_rollout    $R
run.attention=proportional out_dir=$O/crowded_load_pilot_rollout_proportional ,→
~/.local/bin/uv run python -m nudgefield scenario=crowded_load design=field_experiment $R
out_dir=$O/crowded_load_field_experiment,→
~/.local/bin/uv run python -m nudgefield scenario=crowded_load design=field_experiment $R
run.attention=proportional out_dir=$O/crowded_load_field_experiment_proportional ,→
```

12. 运行 5：盟友增益 × 滞后（注册于2026-10-30；数据源为 F3c、F4e（如果有周期），以及未找到列表）

目的：设计 §4.6（G4、G5）。在竞争者作为固定背景的情况下，搜索盟友的增益和滞后，以考察尚未出现的两种类别：振荡和延迟（2026-10-30 的裁决 (c)：延迟说明写道"迄今尚未出现；运行 5 是已登记的搜索"）。

场景文件 `S2_ally_cycles`（新增，本次提交）：将增益为 2 的 S1恒温器竞争者（目标_偏移 -0.03，滞后 2，reach0 0，不升级）设为行动者 0，将 S2 盟友（恒温器，目标_偏移 $+0.03$，reach0 0）设为行动者 1，并将 `ally.gain` $\in$ {0.5, 1, 2, 5, 10, 20, 30, 40} 作为稳健性轴；`sweep: [ally.gain]`；8 个数据点，每次调用运行 800 次；试点 $\rightarrow$ 推广。盟友的 `lag` 不可扫描（观测历史索引中的整数，设计 §4.3）：每个滞后 $\in$ {1, 2, 5, 10} 调用一次，并使用 Hydra 覆盖参数 `scenario.actors.1.lag=<l>`（盟友是行动者索引 1；通过组合四个覆盖值并验证结果进行检查），每个滞后使用一个输出目录。竞争者的滞后保持为 2。

设置：§ 0 `run.n_rep` 100 `population.n` 20 000 `run.seed_base` 20261029 `run.max_batch_` 全程（，，，，：每次调用使用一个未分块批次）在 M （ MPS ）上逐次调用。批次种子 `scenario|design|structure|seed_base` 不包含滞后或注意力规则，因此八次调用共享一个批次种子：四个滞后值和两条规则使用相同的抽样结果运行。没有任何声明将单元格配对。类别规则 R9。

runs 0 4 `run.device` `run.attention`

Call	盟友延迟	注意力	Runs	输出（`code/outputs/explore_20261029/run5/...`）
5a	1	优先级	800	`S2_ally_cycles_pilot_rollout_lag1`
5b	2	优先级	800	`S2_ally_cycles_pilot_rollout_lag2`
5c	5	优先级	800	`S2_ally_cycles_pilot_rollout_lag5`
5d	10	优先级	800	`S2_ally_cycles_pilot_rollout_lag10`
5e	1	比例	800	`S2_ally_cycles_pilot_rollout_lag1_proportional`
5f	2	比例	800	`S2_ally_cycles_pilot_rollout_lag2_proportional`
5g	5	比例	800	`S2_ally_cycles_pilot_rollout_lag5_proportional`
5h	10	比例	800	`S2_ally_cycles_pilot_rollout_lag10_proportional`

6,400 次运行；在 M4 上约需 20 分钟（设计 §4.6），与 Spark 上的运行 3 和 4 并行进行。

范围检查：在 `ally.gain` 0.5–40（范围为 39.5）内，相邻间隔为 0.5、1、3、5、10、10、10，分别占范围的 1.3%、.52%、7.6%、12.7%、25.3%、25.3%、25.3%。增益低于 5 的任何一对都无法单独达到 10% 的份额；以下运行 {1、2、5}（10.1%）和 {0.5、1、2、5}（11.4%）可以；增益从 5 起的任意两个相邻数据点都可以。仅限于 g含 0.5–2 的类别在此网格上无法稳健；重述的稳健性规则按原文适用于每个滞后目录。**运行前声明的预期类别**。设计 §4.6，原文如下：

> 在增益 ≤ 10 下的每个滞后期均衰减（盟友趋于零，而竞争者保持不变：运行 1 的 0.46 身份）；在增益 ≥ 20 且滞后期为 ≥ 5 时可能发生振荡（尖峰 §1.6 的披露探测发现，在 k $= 20$–40 且滞后期为 2 的情况下，5 次重复试验中有 2 次、另有 5 次重复试验中有 1 次观察到振荡：长滞后期下的积分规则会过冲）；任何情况下均不延迟（R9：一个在末期到达但确实存在的晚到事件；盟友的松弛是即时的）。如果在约束区域内，振荡或延迟均不具备稳健性，则以运行 6 的处理份额映射（简报 §6c）替代 F3c，并将这些类别连同该区域列入未找到列表。

每次调用。来源：规范 =引述的§4.6段落；无 =未为该单元格指定类别。该段落没有提及注意力规则；本条目将其解读为同时适用于两者（5a–5d和5e–5h均如此）。

调用	增益 0.5、1、2、5、10	增益 20、30、40
5a、5b、5e、5f（滞后期 1、2）	衰减（规范）	无；未延迟到任何地方（规范）

调用	增益 0.5、1、2、5、10	增益 20、30、40
5c、5d、5g、5h（滞后 5、10）	衰减（规范）	可能振荡，“如果存在于任何位置”（规范；未说明（稳健）；无处延迟（规范）

规范规定，任何调用中的任何单元格都不预期出现延迟。回退（规范）：如果在约束区域内，振荡和延迟都不稳健（运行 7 的 `targets.parquet`，设计§4.8），则以运行 6 的处理份额映射替代F3c，并将两个类别连同该区域一起列入未找到列表。

哪些结果会与预期相悖：在某些滞后下，增益为 ≤ 10 时效果衰减而非稳健；滞后小于 5 或增益小于 20 时出现振荡；任何情况下都存在稳健的延迟。每项均报告为已发现。

先验知识（已披露）。第5节和第8节中的所有内容；运行 1c / 1pc（`S2_ally` 增益为 0.5、2、5，滞后为 2，且不含竞争者：效果衰减，优先级下比率为 0.46；第8节表格）；尖峰 §1.6 的已披露探测实验（k 为 $= 20$ 和 40、滞后为 2 时，5 次重复试验中分别有 2 次和 1 次出现振荡）；构建器检查中对盟友增益 0.5–40 $\times$ 、滞后 1–10 的 CPU 探测（设计 §7：每次运行都通过了预热测试；未引用其中的类别）。m 为 $= 0$ 的拥挤族（目标运行，`code/outputs/targets_20261027_*`，预算为 1.0）是竞争者，也是增益为 2、滞后为 2 的盟友：该网格中滞后为 2 时的增益点为 2，但使用了另一情景名称（安静/拥挤比率为 1.81，§5）。任何运行文件夹中都不存在 `S2_ally_cycles` 的运行、盟友滞后不是 2 的运行，或盟友增益高于 5 的运行（每个 `manifest.json` 下的 `code/outputs/` 均已于 2026-10-30 检查）。

命令（来自 `code/`、M4、设备 `mps`；使用 `bash` 运行此区块）：

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps"O=outputs/explore_20261029/run5
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=1  $R
out_dir=$O/S2_ally_cycles_pilot_rollout_lag1,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=2  $R
out_dir=$O/S2_ally_cycles_pilot_rollout_lag2,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=5  $R
out_dir=$O/S2_ally_cycles_pilot_rollout_lag5,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=10 $R
out_dir=$O/S2_ally_cycles_pilot_rollout_lag10,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=1  $R
run.attention=proportional out_dir=$O/S2_ally_cycles_pilot_rollout_lag1_proportional ,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=2  $R
run.attention=proportional out_dir=$O/S2_ally_cycles_pilot_rollout_lag2_proportional ,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=5  $R
run.attention=proportional out_dir=$O/S2_ally_cycles_pilot_rollout_lag5_proportional ,→
uv run python -m nudgefield scenario=S2_ally_cycles design=pilot_rollout scenario.actors.1.lag=10 $R
run.attention=proportional out_dir=$O/S2_ally_cycles_pilot_rollout_lag10_proportional ,→
```

13. 运行 6：处理组占比（注册于2026-10-30；数据源为 F4a 和 F3c 回退）

目的：设计第4.7节。衡量现场实验的站点类别及其试验组/总组比例相对于处理组占比的表现 p，在两条注意力规则下。2026-10-30 第3号作者裁定（设计第6节）：本阶段运行 6。

场景文件 `S1_treated_share`（新提交）：`S1_competitor` 在增益 2 时的 S1 竞争者（恒温器，目标_偏移 -0.03，滞后 2，reach0 0，无升级）在现场实验中的配置，其中 `design: {fe_reach: [0.1, 0.2、0.3、0.4、0.5, 0.6]}` 作为唯一扫描路径（`sweep: [design.fe_reach]`，沿 p 的稳健性；简报中的四个值加上 0.3 和 0.5）；6 个数据点，每次调用运行 600 次；仅限现场实验（`design.fe_reach` 仅为现场实验类型定义）。扫描的占比取代第0节中每次运行的 `fe_reach` 0.4；每行的处理组是一次均匀抽样中排名最低的 `round(p · N)` 4 3 `S1_` 个对象 `competitor` 1 = 0 4（设计第 §。）。场景名称不同于，因此批次种子以及抽样结果，即使在 p 。。时，也不同于运行 b。

设置：全程采用第0节（`run.n_rep` 100、`population.n` 20、000、`run.seed_base` 20261029、`run.max_batch_runs` 0：每次调用使用一个未分块批次），运行于 M4（`run.device` MPS），每次调用 `run.attention`；比例调用与优先级调用共享批次种子。类别规则 R9；站点类别为 `cls_combined`，根据每行的份额计算（`site_ratio = site_w / (share_r · e)`，设计§4.3）。

Call	设计	注意力	Runs	输出（`code/outputs/explore_20261029/run6/...`）
6a	场域_实验	优先级	600	`S1_treated_share_field_experiment`
6b	场域_实验	比例	600	`S1_treated_share_field_experiment_proportional`

1,200 次运行；在 M4 上约需 5 分钟（设计 §4.7）。范围检查：在 0.1–0.6 上取六个等间距数据点；两个相邻数据点跨越范围的 0.1 / 0.5 = 20%，因此按原文所述适用重述后的稳健性规则。**运行前声明的预期类别**。设计 §4.7，原文如下：

预期：盲站点在 p $\leq$ 0.4 以上保持稳健（运行 1：在各个 k 下 p $=$ 0.4 时为 100/98/100），`trial / total` 在该处沿 1/(1 − p) 曲线运行，并在 p $=$ 0.6 处分离（简报第7a节；竞争者在 p $=$ 0.4 时的触达率为 0.998，且已无余量）。

每次调用。来源：规范 $=$ 所引段落；无 $=$ 未注明该单元格所属类别。该段落未提及注意力规则；本条目将其解读为适用于两者（6a 和 6b 均适用）。

Call	p 0.1、0.2、0.3、0.4	p 0.5	p 0.6
6a, 6b	`cls_combined` 盲态_站点，四个数据点均稳健 数据点（规范）；`trial / total` ≈ 1/(1 − p) （规范）	none	类别无；`trial / total` 偏离 1/(1 −p) （规范）

哪些结果会与预期相悖：盲法_站点在 p $\leq$ 0.4 时不稳健，或在 p $\leq$ 0.4 时 `trial / total` 偏离 1/(1 − p) 曲线，或在 p $=$ 0.6 时位于曲线上。每项均报告为已发现。

先验知识（已披露）。第5节和第8节中的所有内容；运行 1b / 1pb（`S1_competitor`，现场实验，p $=$ 0.4，增益 0.5、2、5：优先级下盲法_站点为 100 / 98 / 100；第8节表格）；尖峰 §1.6 的已披露探测实验（处理组占比 0.2–0.5）。任何运行文件夹中都不存在 p $\neq$ 0.4 的运行，也不存在 `S1_treated_share` 的运行（2026-10-30 检查了每个 `manifest.json` 的 `code/outputs/`：始终为 `fe_reach` 0.4）。

命令 （来自 `code/`、M4、设备 `mps`；使用 `bash` 运行该区块）:

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps"
O=outputs/explore_20261029/run6
uv run python -m nudgefield scenario=S1_treated_share design=field_experiment $R
out_dir=$O/S1_treated_share_field_experiment,→
uv run python -m nudgefield scenario=S1_treated_share design=field_experiment $R
run.attention=proportional out_dir=$O/S1_treated_share_field_experiment_proportional ,→
```

14. 第 8 次运行：余量校准图，竞争者强度 × 旁观者（2026-10-31 已登记；一项可行性研究；若存在联合单元格，则作为 T2 重新校准的数据源；无论如何均用于 ED3）

目的：T1 探测（`docs/analysis/2026-10-31-t1-probe-two-actor-field.md`，第3节–第9节；作者于 2026-10-31 的选择：最站得住脚的方案，即将校准映射作为一项可行性研究）。校准后的拥挤族（`crowded.yaml`）在其 T2 点上没有会作出反应的行动者：两个恒温器在预热期便达到触达率上限（点目标之间的一场不可能双双达成的军备竞赛），总效应中的反应部分为零，且 T1 因构造原因在此失效。本次运行将依据此处确定的判据，检验任何存在会作出反应的竞争者的拥挤场域是否能同时复现 T1 和 T2。无论结果如何都会给出结论：若存在稳健联合单元，则据此定义一个新的拥挤族，并在其上重新校准 T2 点；若不存在联合单元，则将此作为研究发现报告，说明这两种规律属于两种场域状态，其原因在于该机制（注意力拥挤会先耗尽参与者的余量，再消除助推）。

场景文件 `crowded_headroom` （新增，本次提交）：S1 竞争者（恒温器，增益 2，目标_偏移 −0.03，滞后 2，reach0 0），其助推强度 `competitor.nudge.tau` ∈ {0.30, 0.35, 0.40, 0.45} 作为切片，另以 `bystander.count` ∈ 1, 2, 3, 4, 6, 8 0, 0 0 6, `sweep`: m { }（增益、触达率、仅注意力）作为稳健性轴；；数据点，每次调用运行次数；两种设计；无盟友。预算尺度保持默认值 1.0（T2 校准的设置；`parameters.md` §5）。m $=$ 0 不在网格中：它是运行 1 的 S1（τ 0.30）和运行 3 的 τ_自动 $=$ 0 列（τ 0.35–0.45），两者均已登记。网格中的 τ 值是运行 3 中 τ 值的子集。已在本次提交中检查：文件可加载，并扩展为每种设计各六批、每批 400 次运行。 `[competitor.nudge.tau, bystander.count]` 24 2,400

设置：M4（`run.device` MPS）上的第0节（`run.n_rep` 100, `population.n` 20,000, `run.seed_base` 20261029, `run.max_batch_runs` 0，窗口和设计参数相同，`fe_reach` 0.4），每次调用 `run.attention` 。`bystander.count` 进入批次结构，因此每次调用包含六个批次（每个 m 一个），每个批次有 4 个 τ × 100 = 400 运行；一个批次中的四个 τ 值共用其抽样结果（沿 τ 配对，正如运行 4 中的预算一样）。类别规则 R9。

Call	场景文件	设计	注意力	Runs	输出（`code/outputs/explore_20261029/run8/...`）
8a	`crowded_headroom`	试点_推广	优先级	2,400（6 批次 × 400）	`crowded_headroom_pilot_rollout`
8b	`crowded_headroom`	试点_推广	比例	2,400	`crowded_headroom_pilot_rollout_proportional`
8c	`crowded_headroom`	场域_实验	优先级	2,400	`crowded_headroom_field_experiment`
8d	`crowded_headroom`	场域_实验	比例	2,400	`crowded_headroom_field_experiment_proportional`

Call	场景文件	设计	注意力	Runs	输出（code/outputs/explore_20261029/run8/...）
8q	quiet	场域_实验	优先级	100	quiet_field_experiment
8qp	quiet	场域_实验	比例	100	quiet_field_experiment_proportional

9,800 次运行，另加 8c、8d、8q 和 8qp 的级别重跑（下文的工程事项：5,000 次运行）。在 M4 上的耗时，根据构建阶段的目标运行数估算（code/outputs/targets_20261027_201230：在现场实验中，拥挤族在 m = 2–4 时，优先级模式下每次运行约需 0.5–0.7 秒，比例模式下约为其五分之一）：每次优先级调用约 25 分钟，每次比例调用约 6 分钟；加上级别重跑，总计约 1.5 小时。记忆：在 m = 8 且每批次 400 次运行时，远低于 M4 的极限（运行 1 的 S3 m = 16 批次包含 100 次运行，峰值为 4.1 GiB；自 623f481 起，各批次部分会写入文件）。

范围检查：在稳健性轴 m = 1–8（范围为 7）上，相邻间隔分别为 1、1、1、2、2，即分别占范围的 14.3%、14.3%、14.3%、28.6%、28.6%；任意两个相邻数据点均满足 10% 的份额要求，且重述后的稳健性规则（run.robust_min_points 2, run.robust_min_range_share 0.10）在每个 τ 切片内按原文适用。

运行前的工程工作（一个事项，测试先行）。T2 比率是处理组和对照组窗口比率的 Probit 对比，而扫描输出并未记录这一项（effects.parquet 记录的是差异）。构建阶段校准使用一个私有测试辅助程序重新计算了这些比率（tests/nudgefield/test_targets.py::_levels：使用相同的情景、设计、设置、随机种子和设备，重新运行该目录中的批次，保留 none、free、处理组和对照组比率的窗口均值，并检查重新运行能否复现 effects.parquet 至 10^{-9}）。随后，该辅助程序成为包工具 nudgefield.levels，以 uv run python -m nudgefield.levels <run_dir> 的形式运行：它读取目录中的 manifest.json，重新运行其中的批次，并在复现检查通过后，将 levels.parquet 写入原始文件旁边（列包括 point、param_value、slice 列、replicate、seed、run、none_rate、free_rate、free_treated、free_control；元数据包括：工具提交、来源提交）；如果目录中 src/nudgefield 下的任何文件的代码提交与该工具的代码提交不同（遵循 run.resume 的部分规则），它就会拒绝处理该目录。测试套件中的辅助程序从包中导入；T2 测试的数值必须保持不变（在优先级下对校准表进行固定断言）。扫描的类别、阈值和输出列均不变。

联合单元格判据，在运行前固定。每个单元格 (tau, m) 中，按默认规则优先级 y:

1. 余量：竞争者在预热期 90 99 8 0 99 reach_ – 内的重 competitor effects.parquet 0 99 复样本平均覆盖率低于。（在 中；达到或超过 的重复试验份额会在旁边报告）。2. *T1*： t1_ratio < 0.8，在 8a 中（nudgefield.targets.t1_cells，与运行 7 相同）。3. *T2*：该单元格处安静组 / 拥挤组的 Probit 试验效应 probit(8q) / probit(8c)，位于 $[1.5, 2.5]$, 其中probit 是 Φ^{-1}（处理组） $-\Phi^{-1}$（对照组）窗口均值的重复样本均值（levels.parquet；构建阶段定义见 parameters.md 第5节；该区间由作者设定，规范第12节，保持不变）。4. 类别：该单元格在 8a 中的多数类别（R9 cls）为衰减或瞬态（非空值、非排除、非预期：T1 所指的电压降是部分的）。5. 稳健性：条件 1–4 在同一 τ 切片中沿 m 方向的相邻单元格也成立（重述后的规则：两个相邻数据点）。

满足 1–5 的单元格是一个**联合单元格**。如果存在校准点，则它是具有联合单元格的最小 m，其次是最小 τ。还会计算并报告按比例（8b、8d、8qp）下的条件 1–4，仅作描述；校准采用默认规则，与构建阶段校准相同。脚本 code/scripts/calibration_map.py会从目录中生成 24×2 单元格表（每项条件的值和判定），并与结果说明一同提交。

运行后的决策规则，在此固定。如果存在联合单元格：点 (τ, m) 将登记为新系列在该 τ 处的 T2 校准 crowded_headroom， parameters.md §5 中的 T2 区间表述将予以扩展（而非替换： crowded.yaml 校准仍保留为僵局情形），并将新系列上的 m × 预算图（运行 4h）作为独立条目登记，并设有各自的预期；随后，F3b 的候选项为运行 4 和运行 4h。如果不存在联合单元格：不再进行拥挤系列运行；结果说明指出，在该模型中，T1 和 T2 由两种场域状态再现，并解释原因；F3b 保持为运行 4，图中注明其参与者的饱和情况；ED3 指出，T1 来自单竞争者场，T2 来自拥挤场。

运行前陈述的预期。来源：T1 探测说明 §7 中的十次重复探测（已披露，情景 crowded 中盟友处于惰性状态，未登记，文件不在代码库中）、运行 1 的 S1、运行 3 的 τ_自动 $= 0$ 列、运行 4。

切片	余量（条件 1）	T1（条件 2）	T2 区间（条件 3）	类别（条件 4）	联合单元
τ 0.30	m ≤ 2 处的内部（探针：0.76，0.93），在从 m 起的上限处 = 3或 4	< 0.8，样本量为 m = 1 仅限 (0.73)；0.88 样本量为 m = 2；≥ 1 来自 m = 4	由代理指标得出 E / 试点（m =时为 1.44 2，m 时为 2.14 = 4）：m = 3–4，可能 6	在 m ≤ 4–6 时衰减；到 m = 8 时或为空值 超出（m = 16 空值位于探针）	**none 预期**
τ 0.45	每个 m 内部（m 处为 0.52–0.72 2–8）	≈ 0.01–0.35 各处	m = 2–4 (1.67, 2.39)	m 2–8 时为空值（窗口总效应 ≤ 0.09 E）；条件4失败	**none 预期**
τ 0.35，0.40	无探针；介于两者之间：内部至比在……时更大的 m 0.30	低于 0.30；竞争者在 m = 0 处抹去（运行 3：瞬态，比率 0.39 和 0.17）	m = 2–4	瞬态或空值 at m 较小时，衰减 在……处可能出现拥挤将……限制在上限 竞争者部分地	**在……时可能 第 2–4 个月；未 表述为 预期**
成比例的（8b、8d）	无探测	未知	未知	未知	描述性 only

哪些结果会反驳这些探测及本条目的解读：τ 0.30 处出现稳健联合单元（探测结果显示，竞争者在 m = 3–4 时达到上限）；某些 m ≤ 8 时，竞争者在 τ 0.45 处达到上限；τ 0.45 处出现衰减稳健结果；m = 1 时，按优先级出现 T2 区间。这些结果均报告为已发现。τ 0.35–0.40 处出现联合单元并不与任何内容矛盾；问题就在于它是否存在。

先验知识（已披露）。§ 5 8 1 1 0 30, = 0 1 0 64, 0 63 → 1 00 3 0 35 0 60 = 0 0 35, 0 60 4 和 § 中的所有内容；运行（S 在 τ . m: T. 触达率. . ）；运行（τ . – . 在 m: 在优先级下从 . 空值于 . 开始的瞬态）；运行（僵局族：每个单元格中，从周期 50 起，竞争者和盟友的触达率均为 1.00）；构建阶段校准（parameters.md §5：安静状态 Probit 0.152；优先级下拥挤 m = 2 比率 2.25；m = 0：1.81，m = 1：2.16，均有盟友在场）；T1 探测说明 §2（单主体和双主体场的静态／反应划分）以及 §7（τ 0.30、0.45、0.60 × m 1–16 的十次重复探测，已列于上方）。在任何运行文件夹中，都不存在 `crowded_headroom` 的运行、不存在没有盟友且 n_rep 100 的竞争者加旁观者场运行，也不存在超出构建阶段目标运行的 `levels.parquet` 运行（每个 `manifest.json` 均在 2026-10-31 检查了 `code/outputs/` ）。

命令（来自 `code/`、M4、设备 `mps`；使用 `bash` 运行此区块；在 `nudgefield.levels` 工具提交且作者签署此条目后；每次场域实验调用后都会重新运行这些级别）：

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=mps"
O=outputs/explore_20261029/run8
uv run python -m nudgefield scenario=crowded_headroom design=pilot_rollout     $R
out_dir=$O/crowded_headroom_pilot_rollout,→
uv run python -m nudgefield scenario=crowded_headroom design=pilot_rollout     $R
run.attention=proportional out_dir=$O/crowded_headroom_pilot_rollout_proportional ,→
uv run python -m nudgefield scenario=crowded_headroom design=field_experiment  $R
out_dir=$O/crowded_headroom_field_experiment,→
uv run python -m nudgefield scenario=crowded_headroom design=field_experiment  $R
run.attention=proportional out_dir=$O/crowded_headroom_field_experiment_proportional ,→
uv run python -m nudgefield scenario=quiet design=field_experiment             $R
out_dir=$O/quiet_field_experiment,→
uv run python -m nudgefield scenario=quiet design=field_experiment             $R
run.attention=proportional out_dir=$O/quiet_field_experiment_proportional ,→
for d in crowded_headroom_field_experiment crowded_headroom_field_experiment_proportional
quiet_field_experiment quiet_field_experiment_proportional; do ,→
uv run python -m nudgefield.levels $O/$d_done
```

签字批准：本条目到此为止，等待作者对网格、判据（条件 1–5）、决策规则和预期进行签字批准，然后再构建 levels 工具并运行调用。

作者于 2026-10-31 按原文签字批准（“可以继续”）。

已声明的偏差（2026-10-31，运行前）：设备和机器。该条目登记了 M4（`run.device` MPS）。在 M4 上的第一次尝试中，一个调用的四个批次运行完后笔记本过热（作者指示：停止并转移到 Modal）；所有部分生成的 M4 目录均已丢弃（代码库中没有任何目录，也未使用）。六个调用和四个级别重跑均改在 Modal 上运行（`code/scripts/modal_run8.py`：代码库在包含本段落的提交处的浅克隆，`uv sync --frozen`，每个调用使用一个 T4 GPU 容器，`run.device=cuda`，Torch 2.14.1+cu130，驱动程序 610.57；输出位于卷 `multiactorbpp-run8` 上，使用 `modal volume get`获取）。其他所有设置均未改变。后果是：批次种子不依赖于设备，但抽样结果依赖于设备（CUDA 流并非 MPS 流），因此这些单元格无法与 T1 探测说明 §7 中的 M4 探测配对

（十次重复试验，仅使用先验知识）；六个调用共享一个设备，因此，按照已登记的设置，批次内沿 τ 的配对以及级别重跑（与清单使用相同设备）均成立。慢速测试的 T2 固定值（2.25）是一个 MPS 值，本次运行未对其进行处理。顺序为：`nudgefield.levels` 工具（测试优先，已审查），然后是在该工具提交处进行的六个调用和四个级别重跑，接着是 `scripts/calibration_map.py` 和结果说明 `docs/analysis/<date>-run8-calibration-map.md`，最后是决策规则。

在 59e8b74 上运行 2026-10-31（Modal T4，CUDA；levels 工具的合并 c631cbc 加上 `scripts/modal_run8.py`和设备段落，在 `code/src/nudgefield` `docs/analysis/2026-10-31-run8-` 下没有变化）；结果见 。 `calibration-map.md`

15. run 4h：余量族中的注意力负荷（2026-10-31 晚间已登记；数据源为余量族中的 F3b、ED4、T2 重新校准；运行 4 成为僵局对比）

目的：§ 14, 8 `docs/analysis/2026-10-31-run8-` 分支“存在联合单元格”的决策规则（运行： ）。余量族的校准点是竞争者 。在预算为 . 时有 m = 2 1 0 2 1 508, = 3 1 733 1 个旁观者（区间下边界处的 T 比率为 .；相邻单元格 m 给出 .；T . 和 .；竞争者预热期触达率为 . 和 .；衰减的 为 ）。运行 h 按旁观者数量和 × 注意力预算绘制该族图谱，就像运行 4 绘制僵局族一样，这样 F3b 就能在一张图上显示拥挤情况和竞争者的反应，并将两种拥挤状态（竞争者的饱和边界和零效应边界）呈现为两条等值线。 `nudge.tau` 0 35 0 640 0 698 0 735 0 828 100 100 4

场景文件 `crowded_headroom_load`（新增，本次提交；已检查：负载数量正确，并扩展为每种设计各包含九批、每批 800 个）：S1 竞争者设为 `nudge.tau` 0.35（恒温器，增益 2，目标_偏移 −0.03，滞后 2，reach0 为 0），另有旁观者（增益 0，reach0 为 0.6）；以 `population: {budget_mean: [0.35, 0.5, 0.7, 1.0, 1.4, 2.0, 2.8, 4.0]}` 作为切片，以 `bystander.count` ∈ {0, 1, 2, 3, 4, 6, 8, 12, 16} 作为稳健性轴（运行 4 的网格加上 m = 3 2 72 2，裁决）；`sweep: [population.budget_mean, bystander.count]`；数据点；两种设计；无盟友。T 点（预算 1.0，m = 2）和运行 8 的列（预算 1.0，m 1–8 减去 3）都在网格上；预算为 1.0 时的 m = 0，对应运行 3 中 τ 为 0.35 的列，在 τ_自动 0 处（瞬态，比率 0.39）。

设置：§ 0 `run.n_rep` 100 `population.n` 20 000 `run.seed_base` 20261029 `fe_reach` 0 4 `run.max_batch_` (, , , , .), `runs` 0 8 × 100 = 800 9, `run.max_batch_runs` 0 (每次调用包含九个未分块批次，每批 800)。成本按运行 4 的 Spark 速率计算（优先级下 m = 16时每次运行 0.806 秒，比例调用下为 0.160 秒）：每次优先级调用约 1.6 小时，每次比例调用约 20 分钟，顺序运行总计约 4 小时。记忆：运行 4 中 m = 16 优先批次的峰值为 19.4 GiB（总计 124 GB）。

(每个 m 一个批次：预算运行)，类别规则 R 和两项注意力规则。机器：DGX Spark（下文裁决 3）；

调用：4ha 试点 → 推广优先级；4hb 试点 → 推广比例；4hc 现场实验优先级；4hd 现场实验比例。输出 `code/outputs/explore_20261029/run4h/crowded_headroom_load_<design>[_proportional]/` 7 `targets` → 。运行 's producer 对约束区域的全部四项（现场实验调用及其对应的试点推广调用）执行；不重新运行级别（T2 不是此图的逐单元格判据）。

范围检查：参见裁决 2（九个计数；m = 4 以下的各对数据点需要三个相邻数据点才能覆盖范围的 10%；从 m = 4 开始，任意两个相邻数据点即可）。

机器（裁决 3）：DGX Spark，运行 4 在这台机器上执行（GB10、CUDA、124 GB 统一内存；运行 4 的峰值为 19.4 GiB，四次调用耗时 3 小时 40 分钟）。使用标准随机种子的未分块批次，`run.device` CUDA，通过一个 tmux 作业依次运行四次调用（约四小时），然后在 M4 上通过 `targets` scp。运行前：将签核提交中的一组 `main` 文件传送到 Spark（`git bundle create`，通过 `~/nudgefield.bundle`、`git fetch origin`、`git checkout -B main origin/main`、`uv sync --frozen`使用 scp），并检查在 `src/nudgefield`、`conf/nudgefield` 和 `uv.lock` 下，`git status` 保持干净（暂存的维护文件不在该范围内）。温度保护程序：日志记录器在整个作业期间每 60 秒将温度、功耗和利用率写入 `code/outputs/explore_20261029/run4h/gpu.log`的 `nvidia-smi`；当图形处理器读数高于 85 °C 时，作业会在每次调用前等待；结果说明的来源信息中会注明日志记录的最大值。

运行前给出的预期类别（来源：运行 8 在预算 1.0 时 τ 为 0.35 的行；其他预算下的运行 4；运行 3 的 τ_自动 0 列；以及对 T1 探测说明第8节的双区域解读）。- 预算为 1.0（运行 8 的列，其他抽样结果）：m = 0 瞬态（运行 3：比率 0.39，稳健）；m 1 时为瞬态或衰减（运行 8：瞬态 75）；m 2–4 时衰减，竞争者在 m 4 时处于内部，从 m 6 起达到上限；T1 < 0.8 位于 m ≤ 3；优先级规则下，m 12 之前不会出现空值（运行 8：m 8 时为衰减，比率 0.33）；按比例规则下，从 m 8 起为空值（运行 8：m 8 时为空值 100）。- 较小预算（0.35–0.7）：静态损失增大，竞争者在较小的 m 值时便达到饱和（它的助推会随着我们的助推而失去注意力）；两种规则下，空值出现的 m 值都小于预算为 1.0 时的 m 值（运行 4 的模式）；T1 区域缩小至 m ≤ 1–2。- 较大预算（1.4–4.0）：竞争者在更大的 m 值下仍有余量，我们的效应受到的拥挤较少；而 τ 为 0.35 的竞争者在保有余量时会抹去更多效应：较小 m 值时为瞬态（运行 3 在预算为 1.0、m = 0 时已呈瞬态），竞争者因拥挤而达到上限的区域内为衰减；预算为 ≥ 2.8 时，m ≤ 16 处没有空值。- 现场实验：在静态机制下，与试点 → 推广孪生实验相同；当 m 值较小且预算较大、竞争者仍有余量且试验对比接近 E_ 现场实验时，盲法_站点是可行的（预算为 ≤ 1.0 时则不预期如此：此时试验对比已经衰减）。- 两条等值线：竞争者的饱和边界（预热期触达率为 ≥ 0.99）从小预算下较小的 m 值延伸至大预算下较大的 m 值；每个预算下，零效应边界都位于该边界之外。T1 仅在饱和边界内成立。

哪些情况会构成反证：竞争者处于内部且空值稳健的单元格（存在反应但效应消失，意味着竞争者需要将其抹去，即呈瞬态而非空值）；按比例规则时的饱和边界比按优先级规则时位于更大的 m 值；竞争者达到上限的区域内仍符合 T1。

表格 v2 行： `F3b_headroom_pilot_rollout` 以及每个预算对应的 `F3b_headroom_field_experiment` 次运行；用于零效应边界的提示行；新的 `saturation` 主张类型未登记（该边界取自单元格表的余量列，而不是作为一个类别）。

运行后对 F3b 的决定：如果该图在两条规则下的约束区域内均显示至少两个类别，则保留该图（简报 § 6 4 parameters. 规则）；运行（僵局族）移至扩展数据，md 5 2 作为比较对象。§ 增加余量族的 T 校准，与僵局族的校准并列（而非替代后者）。

作者的签字批准事项：(1) τ 0.35、m = 2 处的族，按校准结果确定（或 m = 3，即稳健对中的更中心单元格：裁决已明确）；(2) 网格（与运行 4 相同）；(3) 使用分块的 Modal，或使用更大的图形处理器；(4) 按原文所述的预期。

裁定（作者，2026-10-31 晚间，逐项进行）：(1) **A**：校准单元格为（τ 0.35，m = 2），这是预先声明的规则给出的答案；凡引用校准之处，都注明 T2 边界（1.508 对比 1.5），并在旁边报告 m = 3 单元格；run 4h 的网格涵盖两者。理由：看到结果后朝着增强结果的方向更改预先声明的规则，正是注册表旨在防止的模式。(2) **B**：run 4 的网格加上 m = 3（九个取值：0、1、2、3、4、6、8、12、16；72 个单元格），这样校准后的数对会出现在图上，饱和边界所在之处也会增加一个点；run 4 的每个单元格仍然保留，以便并排比较。范围检查：间隔为 1、1、1、1、2、2、4、4，范围为 16 → 6.25%、12.5% ×2、25% ×2；低于 m = 4 的数对需要三次才能达到 10%，与 run 4 一样；批次：每次调用九个（每个 m 一个），每批 800 次运行。(3) **C**：Spark，并启用上述温度保护程序；理由：run 4h 是 run 4 的孪生版本，而 Spark 消除了两者制作方式上的所有差异，场域除外。(4) **A**：按原文所写的预期。

命令（在 Spark 上，从 `~/MultiActorBPP/code` 记录此条目的提交开始，使用 `gitstatus` clean，在 `src/nudgefield`、`conf/nudgefield`、 `uv.lock`下；在 tmux 中；使用 `bash`运行该代码块；GPU 日志记录器和thermalguard都是作业的一部分）：

```
R="run.n_rep=100 population.n=20000 run.seed_base=20261029 run.device=cuda"
O=outputs/explore_20261029/run4hmkdir -p $Onvidia-smi --query-gpu=timestamp, temperature.gpu, power.draw,
utilization.gpu, memory.used --format=csv -l 60> $O/gpu.log 2>&1 & , →LOGGER=$!
cool() { while [ "$(nvidia-smi --query-gpu=temperature.gpu --format=csv,noheader,nounits)" -gt 85 ]; do
echo "$(date +%T) GPU above 85 C,  waiting"; sleep 60; done; } , →
cool; uv run python -m nudgefield scenario=crowded_headroom_load design=pilot_rollout    $R
out_dir=$O/crowded_headroom_load_pilot_rollout, →
cool; uv run python -m nudgefield scenario=crowded_headroom_load design=pilot_rollout    $R
run.attention=proportional out_dir=$O/crowded_headroom_load_pilot_rollout_proportional , →
cool; uv run python -m nudgefield scenario=crowded_headroom_load design=field_experiment $R
out_dir=$O/crowded_headroom_load_field_experiment, →
cool; uv run python -m nudgefield scenario=crowded_headroom_load design=field_experiment $R
run.attention=proportional out_dir=$O/crowded_headroom_load_field_experiment_proportional , →kill $LOGGER
```

然后在 M4 上，在四个目录的 `scp` 之后以及 `gpu.log`：对每个试点 → 推广目录（`--x population.budget_mean --y bystander.count`）和每个带有其试点 → 推广配对目录的现场实验目录（ `--twin` 7 2 docs/analysis/<date>-run4h-headroom- ）执行 `targets`，如运行；先列出表 load.md 格 v 中的行和带有偏差的结果说明。

作者已签字确认 **2026 10 31** 1 4 docs/superpowers/specs/2026- **--晚间**，裁定记 10-31-run4h-draft-entry.md 1 0 35, 2 2 4 录于：( ) 校准单元格 (.) 按规则执行，注明边界；( ) 运行 's 网格加 m = 3；(3) Spark、未分块的温度保护程序和 GPU 日志；(4) 按原文记录的预期。

运行（2026-10-31 晚间）。四次调用按已登记的顺序，在 Spark 上的一个 tmux 作业中于 ce1107d 执行（`git status` 范围内干净； `~/run4h.sh`，日志 `~/run4h.log`），当地时间 17:36–19:43（Spark 的时钟显示 2026-11-01 00:35–02:43：它比当地时间快七小时，此为运行结束后测得），总计 7,641 秒（2 小时 7 分；原估计为 4 小时，该估计对每个批次都采用运行 4 的 m = 16 比率）；所有清单均完整，`git_dirty` 为 false，CUDA，预热期已稳定； `gpu.log` 最大值为 77 °C（平均 69.7 °C，128 个样本），温度保护程序从未等待；优先级峰值内存 18 5 `targets` docs/analysis/2026-10-31- 。GiB。四项 run4h-headroom-load.md 均如上。结果：（`code/scripts/headroom_load_map.py`）。**依据上述规则作出裁决：**在两条规则下，约束区域内均有两个类别（衰减、瞬态），因此 F3b 取自 run 4h，而运行 4 移至扩展数据，作为僵局对比。

16. 第9次运行：入场券，干净（已于 2026-11-01 晚间登记并签核；正式运行时间为 20:00–20:03，地点为 c220bcf；数据源为 ED3）

目的：简报的第8节 ED3 和第9节第9次运行。构建阶段的目标运行（`code/outputs/targets_20261027_201230/`）带有 `git_dirty: true` （`parameters.md` 第7节：生成这些文件时，有三个文件尚未纳入版本控制）；图表规则（简报 §3）只允许使用干净的运行目录。第9次运行会在干净提交上重新运行目标测试套件（`code/tests/nudgefield/test_`

targets.py，这些测试：每个站点和规则对应的 T1、两种规则下的 T2、T3、T4，以及拥挤的校准检查），并将测试套件生成的所有表格以 Parquet 格式写入一个已登记的目录，从而确保 ED3 的四个图板（T1–T4）和 T3 表格均有干净的来源记录。

唯一的输入变更： T4_POINTS 7 → 15 （第8节简报中，15级网格上倒U形曲线与对照组参与率的关系）； T4_TAKE_UP (0.1, 0.8)、 T4_TAU 0.5、 T4_SD_A 0.8 和预言机 Φ(−β/2) 均保持不变，因此 15 mu 值在相同的两端之间等距分布。阈值均未移动： T1_MAX 0.8、 T2_BAND [1.5、2.5]、 T2_PIN (m = 2、2.25、±0 01 T3_FACTOR 3 T4_PEAK_MAX 0 5 T4_ORACLE_TOL 0 05 .)、 ,.,.。

设置：按测试套件中的写法（N_POP 20,000，n_rep 10，即测试套件的随机种子； priority 默认值，T1 和 T2 也在 proportional下； crowded.yaml 取其校准后的 m）。机器：M4（mps）；此前的慢速测试套件运行 15 项测试约需 3 分钟，T4 在 15 个级别下会使其自身部分的耗时约翻倍。输出：通过环境变量 NUDGEFIELD_TARGETS_OUT 输出 code/outputs/targets_<YYYYMMDD_HHMMSS>/ （测试套件的 TARGETS_OUT 常量为其命名；测试套件会将每次运行的目录及 _save 都写入其中；此处使用绝对路径，因为测试可能会更改目录）。这些表格（T1_<site>_<rule>.parquet、 T2_<rule>.parquet、 T3.parquet、 T4.parquet、每次运行的 levels.parquet）以及运行清单均与 git add -f 一并提交（内容较少；已登记来源信息）；效应 Parquet 文件仍被忽略。

预期值，运行前已声明。所有测试均通过。在相同随机种子下，重复均值之比的 T1 比率与 parameters.md §6 相差不超过 0.05（S1/优先级 0.741，S1/比例 0.671，S2/优先级 0.621，S2/比例 0.738）；每次重复的离散程度如该处所披露（优先级条件下，10 次 S1 重复试验中有 4 次 ≥ 0.8）。优先级条件下，T2 位于区间内，m = 2 0 01 2 25, > 1 3 > 3 比率与 . 相差不超过 .；比例条件下亦然。T 站点间标准差：平均站点内标准误。T：推广平均处理效应在 -级网格上呈现内部峰值；该峰值对应的对照组参与率低于 0.5，且与 Φ(−β/2) 相差不超过 0.05（β = 0.5 / √(1 + 0.8²)）；旧有的 7 个级别仅在取值上（经四舍五入后）构成子集，因此峰值位置是在更细网格上重新读取的。

以下任何情况都将构成不利证据：任何测试失败；在相同随机种子下，T1 比率变化超过 0.05（这意味着代码路径发生了变化，而非场域发生了变化）；T4 峰值跨过 0.5 线，或在更细网格上偏离预言机。

命令（来自 code/，在记录此条目的提交中，并在 git status clean、 src/nudgefield、conf/nudgefield、 tests/nudgefield/test_targets.py、 uv.lock 下； bash）：

```
export NUDGEFIELD_TARGETS_OUT=$PWD/outputs/targets_$(date +%Y%m%d_%H%M%S)
mkdir -p $NUDGEFIELD_TARGETS_OUT
uv run pytest -m slow tests/nudgefield/test_targets.py -q -p no:cacheprovider 2>&1 | tee
$NUDGEFIELD_TARGETS_OUT/pytest.log. →
```

然后：检查 manifest.json 下的每个 $NUDGEFIELD_TARGETS_OUT （状态为 complete， git_dirty 为 false，只有一个提交）； git add -f 表格、清单和 pytest.log； parameters.md §7 纳入干净运行的段落；台账行；ED3（figures/ed3.py）仅读取此目录。

作者的签字批准事项：(1) T4_POINTS 15 ，如简报所述（或者保留 7，并在 Build 网格上绘制 ED3 的 T4）；(2) 按照原文写出的预期；(3) 使用 -f 提交表格和清单。

作者于 2026-11-01 晚间按原文签字批准 （一个问题，三项内容）： T4_POINTS 15；按草案写出的预期；强制添加表格和清单。此次运行位于计划第3部分中任务0的重构提交之后（因此在 src/nudgefield 下工作树是干净的），并且早于 F4 模块工作。

偏差（2026-11-01，当地时间 19:52–19:55，正式运行之前）。 签字批准的命令块中指定了输出变量 TARGETS_OUT；测试套件读取的是 NUDGEFIELD_TARGETS_OUT （ test_targets.py 中的常量 TARGETS_OUT 保存了该名称）。因此，首次执行在 0639070 上运行（范围干净；9 项测试通过，用时 179 秒），但它将表格和 33 份运行清单写入了 pytest 的临时目录（pytest-81），而不是已登记的目录。输入、随机种子和测试套件均未发生任何变化。本节中的命令块现已更正（使用变量的实际名称，并采用绝对 $PWD/... 路径，因为测试可能会更改目录）；更正后的命令块没有重新签字批准：它没有更改任何输入，本文对此予以披露。正式运行是在包含本说明的提交（c220bcf）上通过更正后的命令重新运行所得；从临时目录中找回的首次执行表格仅用于交叉核对，不会提交。

正式运行（2026-11-01，本地时间 20:00–20:03，M4 mps）。 uv run pytest -m slow tests/nudgefield/test_targets.py 在 c220bcf 处（git status 范围内干净）， NUDGEFIELD_TARGETS_OUT=code/outputs/targets_20261101_200002/：9 次运行通过，耗时 179 秒；33 份运行清单，全部 status 完整， git_dirty 为 false，提交版本为 c220bcf。对照预期：重复均值之比 T1 比率为 0.741 / 0.671 / 0.621 / 0.738（S1/优先级、S1/比例、S2/优先级、S2/比例），与 parameters.md §6 中保留三位小数的结果相同，其中 10 次重复中分别有 4 / 3 / 1 / 6 次重复试验 ≥ 08，正如该节所披露；T2 安静组 / 拥挤组的 Probit 比率在优先级方案下为 2.2506（区间 [1.5，2.5]；达到 2.25 ± 0.01 的要求），在比例方案下为 3.3614（> 1）；T3 站点间标准差为 0.0132，对比 3 × 站点内标准误 0 0036 4 15 = 7 0 414 < 0 5 Φ −β 2 = 0 423 | |.；各水平的 T：内部 0 009 ≤ 0 05 峰值位于 j，对照组参与率为 .（.），预言机（/）为.（差异..）。所有预期均已达到；计数反证列表中没有任何项目。首次执行恢复的表格（上述偏差）与正式运行的 T1–T4 表格及校准表完全相同（最大 |差异| 0）。已提交以下内容

-f 1 4 levels.parquet 33 pytest.log run9. : T –T 和校准表、log effects.parquet classes.parquet robust.parquet 5 8 3 每次运行的、清单、，；每次运行的、、仍被忽略（磁盘上全部为 . MB）。ED 只读取此目录。

17. 第10次运行：峰值附近的精细 T4 网格（2026-11-01 夜间起草，待签字批准；供 ED3 面板 d 使用）

目的：ED3 的 T4 判定（注册表 §16，第9次运行）只在网格分辨率下成立：峰值位于参与率水平 7（0.414），水平 6（0.356）的平均处理效应仅低 0.00005；而水平 6 与预言机 $\Phi(-\beta/2)$ = 0 423, 0 05 0 058 相差 0.067，超出了 . 容差；峰值附近粗网格的步长（.）大于容差，且表格未列出平均处理效应的离散程度。作者于 2026-11-01 夜间裁定，应在峰值附近使用已登记的更细网格，而非添加图注中的限定说明。

新增内容（不改变任何现有输入）： 在相同情景名称 `T4_auto_quiet` 下新增第二个 T4 测试 `test_T4_peak_location_fine`，在对照组参与率 `T4_FINE_TAKE_UP` $=$ (0.30, 0.50) 之间等距设置 `T4_FINE_POINTS` $= 9$ 个级别（步长为 0.025，即容差的一半；预言机 0.4226 位于 0.400 和 0.425 两个级别之间），因此使用与粗网格相同的批次种子；两张 T4 表均新增 `ate_sd`（每次重复的窗口平均推出效应的标准差）和 `n_rep`。阈值不变：`T4_PEAK_MAX` 0. 5, `T4_ORACLE_TOL` 0.05；粗网格（`T4_TAKE_UP` (0.1, 0.8)，`T4_POINTS` 15）保持不变（快速测试已将其固定）。

设置：同第1§（按原样编写的测试套件；n_rep 10，样本量 20,000，测试套件的随机种子；M4 mps）。将整个测试套件重新运行到新目录 `code/outputs/targets_<YYYYMMDD_HHMMSS>/` 中，以便 ED3 在一次干净提交中读取一个目录；粗网格表（T1–T4、校准）预计与第9次运行的结果完全一致。使用 `-f`提交：每个表格、每次运行的 `levels.parquet`、 $33 + 9$ 清单、 `pytest.log`、 `run10.log`。

预期结果，在运行前已明确。 每项测试均通过，粗网格测试的表格与第9次运行的结果达到机器精度级别的一致。细网格的峰值位于内部，其参与率与预言机的差异在 0.025 以内（一个网格步长），且低于 0.5。两个网格的并集的峰值位于细网格上，并处于容差范围内。 `ate_sd` 在细网格上大于相邻细网格级别之间的平均处理效应差异（因此判定依据是网格，而不是级别之间的区间）。

哪些情况会构成反证：细网格的峰值超出容差，或不在内部；粗网格表与第9次运行的表格不同。

命令（来自 `code/`，在记录此条目的提交中， `git status` clean 处于 `src/nudgefield,conf/nudgefield, tests/nudgefield/test_targets.py, uv.lock`之下； `bash`）：

```
export NUDGEFIELD_TARGETS_OUT=$PWD/outputs/targets_$(date +%Y%m%d_%H%M%S)
mkdir -p $NUDGEFIELD_TARGETS_OUT
uv run pytest -m slow tests/nudgefield/test_targets.py -q -p no:cacheprovider 2>&1 | tee
$NUDGEFIELD_TARGETS_OUT/pytest.log↩
```

然后：检查 `manifest.json` 下的每个 `$NUDGEFIELD_TARGETS_OUT`（完整、干净、只有一个提交）；将粗网格表与第 9 `git add -f parameters.md` 7 3 `figures/ed3`. 次运行的表格进行比较；检查制品；py §；台账中的行；ED ( ) 只读取此目录。

作者的签字批准事项：（1）所述细网格（9 个级别，0.30–0.50）；（2）按所写内容设定的预期；（3）将整个测试套件重新运行，并输出到一个新目录。

作者于 2026-11-01 夜间按所写内容签字批准（一个问题，三项内容）。运行前记录的实现说明：粗网格 `T4_MU` 在 μ 上等距分布，因此此处按所述方式构建细网格，在 参与率 上等距分布（μ_j $= s{\cdot}\Phi^{-1}(0.30 + 0.025\ j)$，四舍五入至小数点后三位；参与率为 0.300–0.500；预言机位于第 4 和第 5 个细网格级别之间）；两张 T4 表在 `ate`之后都增加了 `ate_sd` 和 `n_rep`，因此按列名将第10次运行的粗网格表与第9次运行的表进行比较。慢速测试套件现在有 10 项测试。

正式运行（2026-11-01，本地时间 22:36–22:53，M4 mps；测试套件自身计时 190 s——清单显示测试之间机器空闲，时间为 22:39–22:47 和 22:47–22:53）。登记的命令在 4 1 94 `git status` `NUDGEFIELD_TARGETS_OUT=code/outputs/targets_20261101_` b b e（范围 223633/ 10 190 42 `status` `git_dirty` 4 1 94 内干净），：在 s 内通过；运行清单全部完整，false，提交 b b e。粗网格表（T1 × 4、T2 × 2、T3、校准、T4）在所有共有列上均与第9次运行一致，达到机器精度（最大 | | 0 **9** **0 300 0 500** 0 1336, 0 1356, 0 1368, 0 1373, 0 1372, 差异）。**细网格（级别，参与率 . – .）：** 平均处理效应 . . . . . ；. – .（均值误约为 ., n 次重复实验）。峰值位于级别 3，参与率为 0.376（内部；低于 0.5）； |峰值 − 预言机| $= 0.047$，处于容差 0.05 以内，因此测试通过；并集的峰值位于相同的细网格级别。**与预期的偏差：** 精细峰值与预言机 0.423 相差超过一个网格步长（0.025）——相差 0.047，即两个步长。解读，而非规则（2026-11-01 夜间修订，时间在 ED3 绘制之后：第一版将平台区写为 0.35–0.43，差值为 ≤ 0.0005；后更正为 ED3 计算所得的单标准误范围；运行数据未变）：倒U形顶部平坦——在参与率 0.356–0.414 范围内，与最大值的差距不超过一个重复样本标准误（约 0.0003）（ED3 计算所得的平台区）；在 0.351–0.426 范围内，与之相差不超过 0.0006——因此，峰值位置只能确定到该平台区，样本量_重复实验 10；根据登记的容差，“成立”这一判定仍然有效，但正处于容差边缘（平台区下限 0.356 与预言机相差 0.067）。 `ate_sd` 超过相邻差值的预期仅在峰值附近得到满足，在网格两端则不成立——这是第二项偏差（2026-11-01 夜间修订，在最终全分支审查中发现；第一版写道“满足（网格两端为 0.0009 对 ≤ 0.0021，平台区为 ≤ 0.0006）”，前后矛盾）：八个相邻差值为 0.0019、0.0012、0.0005、0.0001、0.0005、0.0009、0.0014、0.0021，对照`ate_sd` 0.0009–0.0012，因此四个内侧差值（参与率 0.351–0.451）低于两个级别的 `ate_sd`，而四个外侧差值（两端各两项）则高于该值。这不影响判定（测试读取最大值位置；计数反证列表未触发），而上述解读（平台区）正是预期成立之处。计数反证列表中没有其他内容

。使用 -f提交：所有表格（两张 T4 表，含 ate_sd、 n_rep）、每次运行的 levels.parquet、42 份清单、 pytest.log、 run10.log；忽略效应/类别/稳健 Parquet（磁盘占用 6.9 MB）。ED3 只读取此目录；第9次运行的目录仍作为§16的记录。

18. 第11次运行：试点在站点中的份额（2026-11-12 起草，待签字批准；回应最终审查的 M4 点）

目的：2026-11-12 的审稿人式评审（M4 点）指出，“试点无法预测推广”这一说法，前提是试点只占竞争者所关注总体的一小部分（2%），而且这一份额从未改变。本次运行将改变这一份额。问题不只是更大的试点是否会激起竞争者的响应（它必然会按其改变的比率作出响应），还在于试点随后报告什么：竞争者的响应会同样触达处理组和未处理组人群，因此，试点中处理组与对照组的 — 对比不应随着份额增加而减弱；相反，试点应变成一个盲站点，其站点层面增益下降，而对比效应保持不变。若确实如此，更大的试点无法通过自身的比较来预测推广，论文的主张仍然成立，只是明确了其适用条件。对于棘轮式竞争者，大型试点还应损害随后进行的推广，因为竞争者在试点期间获得的触达率永远不会归还。

设置：情景 S1_competitor （恒温器；其扫描 competitor.gain {0.5, 2, 5}）、 S1_persistent_auto （升级型恒温器；
competitor.escalation_tau 0 1, 0 2, 0 3 S1_maximiser competitor. {...}）和（棘轮效应；{...}）；设计采用 {... escalation_tau 0 1, 0 2, 0 3 pilot_rollout design_params.pilot_reach ∈ 0 02, 0 05, 0 10, 0 20, 0 40, 0 60 0 02 run.attention run.n_rep 100 population.n 20 000 ...} (. 为论文中的试点）；优先级；；，；
run.
seed_base 20261112 1 2 3 × 3 × 6 × 100 = 5 400 ；其他所有内容均采用该情景已登记的值（§ –§）。, 运行。

机器：DGX Spark（GB10， cuda），与运行 3、4 和 4h 一样，通过 SSH 从 Linux机器启动；与之前一样，Spark 的克隆副本从 Git 捆绑包中获取注册提交。运行 4h 的计时结果（28,800 次运行耗时 7,641 秒）：约 25 分钟。**设备，事先声明：**图 2a、2b、2d 来自 M4（mps）上的运行 1 和 2；不同设备的抽样结果不同，因此本次运行的 0.02 列也作为本次运行自身的对照组：其类别和比率必须符合稳健性规则（衰减 / 瞬态 / 反向，比率与 F2 值 0.63、0.00 末端、 −0.34 的差异在 0.1 E 以内），否则设备差异具有实质影响，本次运行将附带此项说明予以报告。（本条目的初稿曾写明 Linux机器的 CPU，其基准测试结果为每次运行 0.7 秒；作者选择了 Spark。）

统计数据，按（情景、点、份额）预先固定，来自 effects.parquet 试点窗口 t ∈[60, 80) 和 classes.parquet：- P1，试点对 E 的对比效应：窗口期及重复试验中 trial 的均值，除以 E（来自 classes.E的人均效应）。- P2，试点的站点效应相对于预期地点效应：窗口期内 total 的均值，除以 pilot_reach × E。- P3，竞争者在试点期间的响应：t ∈ [60, 80) 内 reach_competitor 的均值减去 t ∈ [40, 60) 内的均值。- P4，推广： ratio_window、 cls，以及每个份额下的稳健类别。- P5，预测差距：P1 减去 P4 的 ratio_window。

预期，在运行前陈述：- E1（对比效应保持不变）：对于每个情景和点，在所有份额下，P1 与其在 0.02 处的值之差均不超过 0.10。若到份额 0.4 时下降超过 0.10，则意味着更大的试点确实通过自身对比效应预测了推广结果，这与论文的主张相悖。- E2（试点变成盲站点）：对于增益为 2 和 5 的恒温器，P2 随份额增加而下降，并在份额 0.4 时低于 0.75；在增益 0.5 时，下降幅度较小（响应缓慢的竞争者在二十个周期内的响应较弱）。- E3（响应按比例被激发）：在每个恒温器情景中，P3 随份额单调上升。- E4a（恒温器：推广结果不变）：对于S1_competitor 和 S1_persistent_auto，P4 的类别在每个份额下都相同，并且 ratio_window 的变化小于 0.10（在 80–99 期间自动助推关闭，恒温器回归静息状态）。- E4b（棘轮具有记忆）：对于 S1_maximiser ratio_window 0 10 0 02 0 6 escalation_ ，其值随份 tau 0 1,
额下降，在 . 与 . 之间的 . 处下降超过 ，并且在最大份额下，自动助推会更早启动，或已在试点期间启动。

什么情况会构成反证：E1 在任何地方都失败（该主张的措辞改为“一个小规模试点”）；E4a 失败（届时，试点会改变它本应预测的推广，这是另一种值得报告的机制）。

已声明的先验知识：在撰写本条目前，于 2026-11-12 在这台机器上进行过一次探索性运行，S1_competitor 样本量为 = 2,000，仅重复一次，份额为 0.4, seed_base 1：增益为 2 时，试点的对比为 0.073，而预期值 E 为 0.053（高于 E: 对照组下降）；站点效应为 0.011，而预期值为 0.021（P2 ≈ 0.5）；试点期间，竞争者的影响范围从 0.68 上升至 0.89。以上预期是在看到这次运行结果后写下的；因此，对于那个单元格而言，它们是完整规模下的确认，而非预测。

命令（来自 code/，在记录本条的提交时， git status 在 src/nudgefield 和conf/nudgefield下保持干净；bash）：

```
O=outputs/explore_20261112/run11; R="run.n_rep=100 population.n=20000 run.seed_base=20261112
run.device=cuda". →
for s in S1_competitor S1_persistent_auto S1_maximiser; do for p in 0.02 0.05 0.10 0.20 0.40 0.60; do
uv run python -m nudgefield scenario=$s design=pilot_rollout $R design_params.pilot_reach=$p
out_dir=$O/${s}_pilot$p, →done; done
```

然后：检查每个 manifest.json（状态为 complete， git_dirty false，一个提交，设备为 CUDA）；一份结果说明 docs/analysis/2026-11-12-run11-pilot-share.md ，其中每个单元格都列出 P1–P5，以及已登记和实际发现的 E1–E4；受版本控制的汇总 CSV 放在运行文件旁；只有作者阅读结果说明后，才能修改论文正文。

供作者签字批准的事项：(1) 六个份额和三个情景；(2) 在 Spark 上以 N = 20,000 运行 n_rep 100（cuda，约 25 分钟），并将 0.02 这一列作为设备对照；(3) 按原文列出预期 E1–E4b；(4) 声明这一次探索性运行属于先验知识。

作者于 2026-11-12 下午按原文签署确认（“签署确认”；这四项）。在签署提交的后续版本（本说明）上于 Spark 启动，已在 tmux run11 code/outputs/ session 中分离，log。explore_20261112/run11/launch.log

正式运行（2026-11-12，15:30–15:56 本地时间，DGX Spark cuda，提交 6fd2e39）。18 次调用，5,400 次运行，25 分 42 秒；每份清单均已完成，git_dirty false，一次提交，设备为 CUDA。汇总、清单和启动日志均在 code/outputs/explore_20261112/run11/ 下跟踪；脚本；结果说明。code/scripts/analyse_run11.py docs/analysis/2026-11-12-run11-pilot-share.md 1

登记时及实际结果：E 已满足（试点的对比在所有情况下都与其 2% 值相差不超过 0.08）；E2 已满足（份额为 0.4 时，P2 为 0.45 和 0.28）；E3 对自动 0.2 和 0.3 时的升级型恒温器未满足（在份额为 0.4 或更高的试点期间，竞争者会升级并降低其触达率）；E4a 对自动 0.3 时的升级型恒温器未满足（相同的窗口比率，在份额 0.4–0.6 时为类别空值而非瞬态：试点期间开启的自动助推会持续到全面推广阶段）；E4b 未满足（最大化者的窗口比率在不同份额下完全相同；其全面推广结果由自动助推决定；所用统计量不合适）。未预先命名的发现：P2 随份额保持平稳，直到竞争者的极限生效，因此引导型竞争者会吸收任何试点站点效应的相同比例。设备对照：2% 列在规则允许范围内与图 2a、b、d 相符。中心 Mac 上的 Parquet 文件副本尚未就绪（运行结束时处于离线状态）。

补充信息附录，S8节：试点在站点中的份额

文章中的试点处理了站点的 2%。第11次运行在最终审查后登记（S7 节、第18节），并于 2026年11月12日运行；针对图2中的三种竞争者，将该份额设为 0.02、0.05、0.10、0.20、0.40 和 0.60：增益为 0.5、2 和 5 的转向恒温器，自动助推强度为 0.1、0.2 和 0.3 的升级型恒温器，以及强度相同的棘轮式最大化器；先试点再推广，优先级共享，每组包含 100 次重复实验、每次涵盖 20,000 人，种子基数为 20261112，在 DGX Spark（cuda）上运行。运行前固定了五项统计量，按单元格计算，覆盖试点窗口期（第 60–79 周期）及各次重复实验：P1，试点中处理组与 $-$ 对照组在 E 上的对比；P2，试点的站点层面效应与按该份额处理后预期的站点层面效应之比（份额 $\times$ E）；P3，竞争者在试点窗口期的平均触达率减去其第 40–59 周期的平均触达率；P4，全面推广的评估窗口比率及类别；P5，P1 减去 P4 的比率（见跟踪汇总）。登记了五项预期；其中两项得到满足。

- E1，符合：在每个单元格中，P1与份额为0.02时的值相差不超过0.10（最大变动为0.08）。更大的试点无法通过自身的比较来预测推广。
- E2，符合：对于增益为2和5的转向恒温器，P2在份额达到0.4时降至0.75以下（分别为0.45和0.28）。
- E3，对于强度为0.2和0.3的升级型恒温器，不符合：P3随份额增加而上升，直至0.4，并在0.6时下降。解读：份额达到0.4或更高的试点，会在试点范围内将恒温器保持在极限状态十个周期，因此它会在试点期间升级，并降低其审慎覆盖范围；当强度为0.3、份额为0.6时，试点的站点效应为负（P 2 -0.09）。
- E4a，未满足：对于强度为 0.3 的升级型恒温器，各份额下全面推广的窗口比率保持不变（范围为 0.001），但在份额为 0.4 和 0.6 时，其类别为“空值”而非“瞬态”（100 次运行中分别有 87 次和 100 次）。解读：试点期间开启的自动助推具有持续性，因此全面推广一开始就处于升级场域，且始终无法达到 E 的 0.75。

- E4b，未满足：最大化者的窗口比率在各份额下完全相同；其全面推广结果由自动助推决定，而自动助推在所有情况下都会开启，因此已登记统计量不可能发生变化。出于与 E4a 相同的原因，其类别会随份额变化。

未预先命名：在竞争者的极限生效之前，P2 在不同份额下保持平稳（增益为 0.5 时，每种份额下均为 0.79；增益为 2 时，在 0.02 至 0.40 之间为 0.45；增益为 5 时，在最高 0.20 时为 0.20）。引导型竞争者会吸收试点站点效应中相同的比例，不受试点大小影响，因为它会按差距的比例作出响应。设备对照：0.02 列重复了图 2a、b 和 d 中的设置，cuda 与 M4 的 mps 对比：0.625 对 0.63，-0.001 对 0.00（终点），-0.346 对 -0.34，符合稳健性规则。

表 S7：第11次运行：试点的站点份额。每个单元格显示基于 100 次重复试验的均值：P1，试点相对于 E的对比；P2，试点站点效应与预期地点效应之比；P3，试点期间竞争者的触达增益；P4，推广的评估窗口比率及其众数类别，以及属于该类别的运行占比。

竞争者、设置	试点份额	P1	P2	P3	P4 比率	P4 类别（运行占比）
转向，增益 0.5	0.02	1.00	0.79	0.08	0.62	衰减（1.00）
	0.05	0.97	0.79	0.08	0.62	衰减（1.00）
	0.10	0.98	0.80	0.09	0.62	衰减（1.00）
	0.20	0.98	0.80	0.11	0.62	衰减（1.00）
	0.40	0.97	0.80	0.16	0.62	衰减（1.00）
	0.60	0.97	0.79	0.20	0.62	衰减（1.00）
转向，增益 2	0.02	0.98	0.45	0.02	0.62	衰减（1.00）
	0.05	0.98	0.45	0.03	0.62	衰减（1.00）
	0.10	0.97	0.45	0.06	0.62	衰减（1.00）
	0.20	0.96	0.45	0.12	0.62	衰减（1.00）

表 *S7*（续）

竞争者、设置	试点份额	P1	P2	P3	P4 比率	P4 类别（运行占比）
	0.40	0.96	0.45	0.24	0.62	衰减（1.00）
	0.60	0.96	0.54	0.29	0.62	衰减（1.00）
转向，增益 5	0.02	1.01	0.20	0.02	0.62	衰减（1.00）
	0.05	0.98	0.20	0.05	0.62	衰减（1.00）
	0.10	0.98	0.20	0.09	0.62	衰减（1.00）
	0.20	0.97	0.20	0.18	0.62	衰减（1.00）
	0.40	0.95	0.28	0.31	0.62	衰减 (1.00)
	0.60	0.96	0.48	0.33	0.62	衰减 (1.00)
棘轮效应，τ（自动）0.1	0.02	1.01	0.94	0.00	0.31	瞬态 (1.00)
	0.05	0.98	0.89	0.01	0.31	瞬态 (1.00)
	0.10	0.96	0.81	0.02	0.31	瞬态 (1.00)
	0.20	0.97	0.60	0.09	0.31	瞬态（1.00）
	0.40	0.96	0.47	0.23	0.31	衰减（0.87）
	0.60	0.96	0.48	0.30	0.31	衰减（1.00）
棘轮效应，τ（自动）0.2	0.02	0.95	0.91	0.00	-0.02	瞬态（1.00）
	0.05	0.98	0.88	0.01	-0.02	瞬态（1.00）
	0.10	0.97	0.78	0.03	-0.02	瞬态 (1.00)
	0.20	0.96	0.58	0.09	-0.02	瞬态 (1.00)
	0.40	0.96	0.46	0.23	-0.02	空值 (0.95)
	0.60	0.96	0.44	0.30	-0.02	空值 (0.99)
棘轮效应，τ（自动）0.3	0.02	0.97	0.92	0.00	-0.35	反向 (1.00)
	0.05	0.96	0.89	0.01	-0.35	反向（1.00）
	0.10	0.95	0.81	0.02	-0.35	反向（1.00）
	0.20	0.96	0.59	0.09	-0.35	反向（1.00）
	0.40	0.96	0.46	0.23	-0.35	反向（1.00）
	0.60	0.94	0.40	0.30	-0.35	反向（1.00）
升级，τ（自动）0.1	0.02	0.96	0.45	0.01	0.27	瞬态 (1.00)
	0.05	0.95	0.45	0.03	0.27	瞬态 (1.00)
	0.10	0.95	0.45	0.06	0.27	瞬态 (1.00)
	0.20	0.95	0.45	0.12	0.27	瞬态 (1.00)
	0.40	0.94	0.35	0.24	0.27	瞬态 (1.00)
	0.60	0.93	0.28	0.27	0.27	瞬态 (1.00)
升级，τ（自动）0.2	0.02	0.97	0.45	0.01	-0.00	瞬态 (1.00)
	0.05	0.96	0.45	0.03	-0.00	瞬态 (1.00)
	0.10	0.96	0.45	0.06	-0.00	瞬态 (1.00)
	0.20	0.95	0.45	0.12	-0.00	瞬态 (1.00)
	0.40	0.95	0.27	0.23	0.00	瞬态 (0.98)
	0.60	0.94	0.09	0.22	0.00	瞬态（0.98）
升级，τ（自动）0.3	0.02	0.98	0.45	0.02	-0.00	瞬态（1.00）
	0.05	0.98	0.45	0.03	-0.00	瞬态（1.00）
	0.10	0.97	0.45	0.06	-0.00	瞬态（1.00）
	0.20	0.98	0.45	0.12	-0.00	瞬态（1.00）
	0.40	0.94	0.14	0.23	0.00	空值（0.87）
	0.60	0.91	-0.09	0.17	0.00	空值（1.00）

SI 参考文献

[1] Volker Grimm、Steven F. Railsback、Christian E. Vincenot、Uta Berger、Cara Gallagher、Donald L. DeAngelis、Bruce Edmonds、Jiaqi Ge、Jarl Giske、Jürgen Groeneveld、Alice S. A. Johnston、Alexander Milles、Jacob Nabe-Nielsen、J. Gareth Polhill、Viktoriia Radchuk、Marie-Sophie Rohwäder、Richard A. Stillman、Jan C. Thiele 和 Daniel Ayllón。《用于描述基于智能体及其他仿真模型的 ODD 协议：为提升清晰度、复现性和结构现实主义而进行的第二次更新》。《人工社会与社会模拟期刊》，23(2):7，2020。数字对象标识符：10.18564/jasss.4259。[2] Stefano DellaVigna 和 Elizabeth Linos。《将随机对照试验推广至更大规模：来自两个助推部门的综合证据》。《计量经济学》，90(1):81–116，2022。数字对象标识符：10.3982/ecta18709。[3] Katherine L. Milkman、Dena Gromet、Hung Ho、Joseph S. Kay、Timothy W. Lee、Pepi Pandiloski 等。《大型研究提升应用行为科学的影响力》。《自然》，600(7889):478–483，2021。数字对象标识符：10.1038/s41586-021-04128-4。[4] Omar Al-Ubaydli、John A. List 和 Dana Suskind。《运用科学的科学：理解可扩展性面临的威胁》。《国际经济评论》，61(4):1387–1409，2020。数字对象标识符：10.1111/iere.12476。[5] Christopher J. Bryan、Elizabeth Tipton 和 David S. Yeager。《如果没有异质性革命，行为科学不太可能改变世界》。《自然·人类行为》，5(8):980–989，2021。数字对象标识符：10.1038/s41562-021-01143-3。[6] Silvia Saccardo、Hengchen Dai、Maria A. Han、Naveen Raja、Sitaram Vangala 和 Daniel M. Croymans。《扩大助推的规模：基线动机与替代效应的作用》。工作论文，2025年4月9日版本。早期版本：《扩大助推的规模：谁会行动以及如何行动》，SSRN，数字对象标识符：10.2139/ssrn.3971192，2025年。[7] Kurt Lewin。《社会科学中的场论：理论论文选集》。Harper，纽约，1951年。Dorwin Cartwright 编。