arXiv:2305.01124v1 [cs.AI] 1 May 2023

人类对自适应机器的适应会收敛至博弈论均衡

Benjamin J. Chasnov、Lillian J. Ratliff、Samuel A. Burden

华盛顿大学电气与计算机工程系，美国华盛顿州
西雅图，邮编 98195

摘要

自适应机器有可能在多种情境下协助或干扰人类行为，从认知决策（Mehrabi等，2021；Sutton等，2020）到实体设备辅助（Felt等，2015；Slade等，2022；Zhang等，2017）。因此，理解机器学习算法如何影响人类行动至关重要，尤其是在机器目标与人类目标不一致的情况下（Thomas等，2019）。人类会结合显性和隐性策略，不断适应周围环境（Heald等，2021；Taylor等，2014）；因此，当环境中存在自适应机器时，人类与机器便是在进行一场博弈（Başar 和 Olsder，1998；冯·诺伊曼和 Morgenstern，1947）。博弈论是一个成熟的框架，用于建模两个或更多决策者之间的互动，已广泛应用于经济市场（Varian，1992）和机器算法（Goodfellow等，2014）。然而，现有方法对个体人类如何因与自适应机器交互而改变适应行为，往往只是作出假设，而非通过实证检验（Madduri等，2021；Nikolaidis等，2017）。在此，我们测试了机器与人类受试者进行一般和博弈时所用的学习算法。我们的算法使机器能够从行动与策略空间中一系列博弈论均衡里，选择共同适应交互的结果。重要的是，这些机器学习算法直接依据对人类行动的观察开展工作，无需像以往研究那样求解逆问题来估计人的效用函数（Li等，2019；Ng和Russell，2000）。令人惊讶的是，其中一种算法能够将人机交互引导至机器的最优点，实际上控制人类的行动，即便人类是根据自己感知到的成本景观作出最佳响应。我们的研究结果表明，博弈论可用于预测并设计智能人类与机器之间共同适应交互的结果。

我们研究了人类 H 与机器 M之间进行的博弈。这些博弈由二次函数定义，函数将每个人类 h 和机器 m 的标量行动映射为成本 $c_H(h,m)$ 和 $c_M(h,m)$。博弈在一系列试次中连续进行，机器在试次内或试次之间进行适应。如图1a所示，人类行动 h 由手动输入设备（鼠标或触摸屏）决定；如图1b所示，机器动作 m 则根据机器的成本函数 c_M 和人类的行动 h 通过算法决定。人类的成本$c_H(h,m)$通过计算机显示屏上矩形的高度持续显示给人类受试者，如图1a所示；受试者被要求“尽可能缩小矩形”，而机器的行动则不予显示。

博弈论均衡

此处报告的实验基于一种一般和博弈，这意味着分配给人类和机器的成本函数既不一致，也不相反。一般和博弈并不存在单一的“解”概念——与纯优化问题不同，博弈参与者无法选择所有决定其成本的决策变量。尽管每个参与者都追求自己偏好的结果，博弈结果通常会体现参与者相互冲突的目标之间的折中。我们将纳什 (纳什，1950年）、斯塔克尔伯格 (冯·斯塔克尔伯格，1934年）、一致猜测变分 (鲍利，1924）和逆斯塔克尔伯格 (Ho 等，1982均衡（定义见 Başar 和 Olsder 的第 4.1 、 4.6 、 4.9 和 7.1 条 Başar 和 Olsder(1998 ））以及每个参与者的全局最优解都作为实验中可能出现的结果。补充材料的第S1节给出了这些博弈论概念的正式定义；下一段则以通俗语言进行说明。表 1 列出了定义本文所研究博弈的成本函数表达式，以及由此得到的博弈论均衡的数值。

纳什均衡（纳什，1950年）出现在同时行动的博弈中，是联合行动空间中的一些点；从这些点出发，任何一方都没有动机偏离（参见 Başar 和 Olsder（1998）第 4.2 节）。在序贯行动的博弈中，一方（领导者）选择行动时假定另一方（跟随者）将采取其最佳响应，此时可能出现斯塔克尔伯格均衡（冯·斯塔克尔伯格，1934年）。在这种情况下，领导者会对跟随者的策略作出猜测，即一个从领导者行动映射到跟随者行动的函数；这一猜测与跟随者进行博弈的方式一致（Başar 和 Olsder（1998）第 4.5 节）。领导者的猜测可视为跟随者的内部模型（Huang等，2018；Nikolaidis等，2017；Wolpert等，1995）。在我们的二次情形中，从纳什均衡转向斯塔克尔伯格均衡通常有利于领导者，因为其成本会降低。当然，跟随者随后也可能对领导者的行动方式形成自己的猜测，而博弈参与者可能会根据对手的行动，迭代更新自己的策略和猜测。在我们考察的博弈中，这一迭代会收敛到一致猜测变分均衡（鲍利，1924）；该均衡根据行动和猜测来定义：每个参与者的猜测都等于对手的策略，并且每个参与者的策略相对于其对对手的猜测而言都是最优的（Başar 和 Olsder（1998）第 7.1 节）。最后，如果一方意识到自己的策略选择会如何影响另一方，就可以设计一种激励，引导博弈走向自己偏好的结果，这种结果称为逆斯塔克尔伯格均衡（Ho 等，1982）（Başar 和 Olsder（1998）第 7.4.4 节）。

实验结果

我们使用一对二次成本函数，对不同人类受试者群体进行了三项实验 c_H， c_M 如图 1a、b 所示。这些函数经过设计，使得行动空间和策略空间中均存在不同的博弈论均衡。这些通过解析方法确定的均衡，与每项实验中人类和机器在一系列试次中所达到的行动和策略的经验分布进行了比较。在全部三项实验中，我们发现，实测行动或策略均收敛到了预测的博弈论值。

在我们的第一个实验（图 1）中，机器在试次内使用可以说是最简单的优化方案——梯度下降来调整其行动（Chasnov 等，2020；Ma 等人，2019）。我们针对每位人类受试者测试了梯度下降算法的七种适应率，如图 $\alpha \geq 0$ 1c、d、e 所示；每种适应率重复两次，且适应率的出现顺序随机。我们发现，在本实验中， $n = 20$ 人类受试者群体的中位数行动向量分布，随着适应率从最慢变为最快，从纳什均衡（东北）转向人类主导的斯塔克尔伯格均衡（东南）。

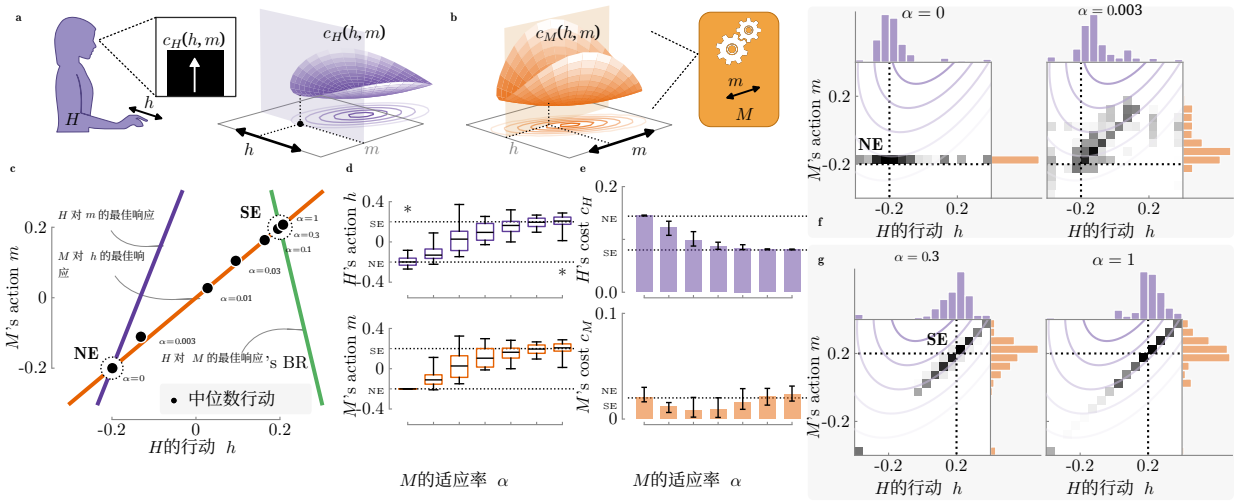

图1：**行动空间中的梯度下降（实验 1，** $n = 20$）。（**a**）每位人类受试者 H 都被要求通过手动输入 h，尽可能缩小计算机显示屏上的黑色条形。条形的高度表示规定的成本值c_H。（**b**）机器 M 有其自身的成本 c_M，其设定使得不同参与者之间以及各参与者与其全局最优解之间的博弈论均衡互不相同。机器知道自己的成本，并观察人类行动 h。在本实验中，机器以其成本$\frac{1}{2}m^2 - hm + h^2$的梯度下降来更新其行动，适应率为 α。（**c**）各机器适应率 α 下的中位数联合行动，并叠加显示博弈论均衡以及定义纳什均衡与斯塔克尔伯格均衡（分别为 NE 和 SE）的最佳响应（BR）曲线。（**d**）各机器适应率下的行动分布，以箱线图展示，并标出第 5、25、50、75 和 95 百分位数。统计显著性（$*$）通过与 NE（显示在分布下方）和 SE（显示在分布上方）进行比较来确定，采用双侧 t检验（$^*P \leq 0.05$）。（**e**）各机器适应率下的成本分布，以箱线图展示，误差线标出第 25、50 和 75 百分位数。（**f,g**）不同适应率下的行动一维和二维直方图（（f）中为$\alpha \in \{0, 0.003\}$，（g）中为 $\alpha \in \{0.3, 1\}$），并叠加显示博弈论均衡（（f）中为 NE，（g）中为 SE）。

适应率（图 1c）。重要的是，如果人类也使用梯度下降来调整其行动，就不会得到这一结果，因为在同步梯度博弈中，仅改变适应率并不会改变驻点（Chasnov 等，2020）。我们观察到从纳什均衡向斯塔克尔伯格均衡的转变有利于人类（图 1e），且具有统计显著性：在最慢适应率下，行动分布与东南不同但与东北无异；而在最快适应率下则恰好相反（图 1d；$^*P \leq 0.05$；双侧 t检验，自由度（df）19；确切统计量见表 S1）。发现人类的实证博弈与根据其既定成本理论预测的最佳响应函数相一致，这一点十分重要，因为这一发现促使我们在后续实验中将机器的博弈从行动空间提升到对其策略空间进行推理，即从人类行动到机器行动的函数。

在我们的第二项实验中（图2），机器采用仿射策略（即 m 由 h的仿射函数决定），并通过观察人类的响应来调整策略。试次成对进行，每对试次中机器的策略仅常数项不同。每对试次结束后，机器利用这对试次中的中位数行动向量来估计关于人类策略的猜测（Bowley，1924；Figuières等，2004）（或内部模型（Huang等，2018；Nikolaidis等，2017；Wolpert等，1995）），并更新机器策略，使其针对该猜测达到最优。毫不意外，人类也相应调整了自己的策略。随着这一过程不断迭代，一组 $n = 20$ 人类受试者（不同于第一项实验中的受试者群体）的中位数行动向量分布，从人类主导的斯塔克尔伯格均衡（东南）向行动与策略空间中的一致猜测变分均衡（CCVE）移动（图2a）。从首次迭代到最后一次迭代，我们观察到分布从东南向 CCVE 的移动在策略空间中具有统计显著性（图2c；$^*P \leq 0.05$；双侧 t检验，自由度（df）为 19；确切统计量见表 S1），但在行动空间中不显著（图2b；$^*P \leq 0.05$；

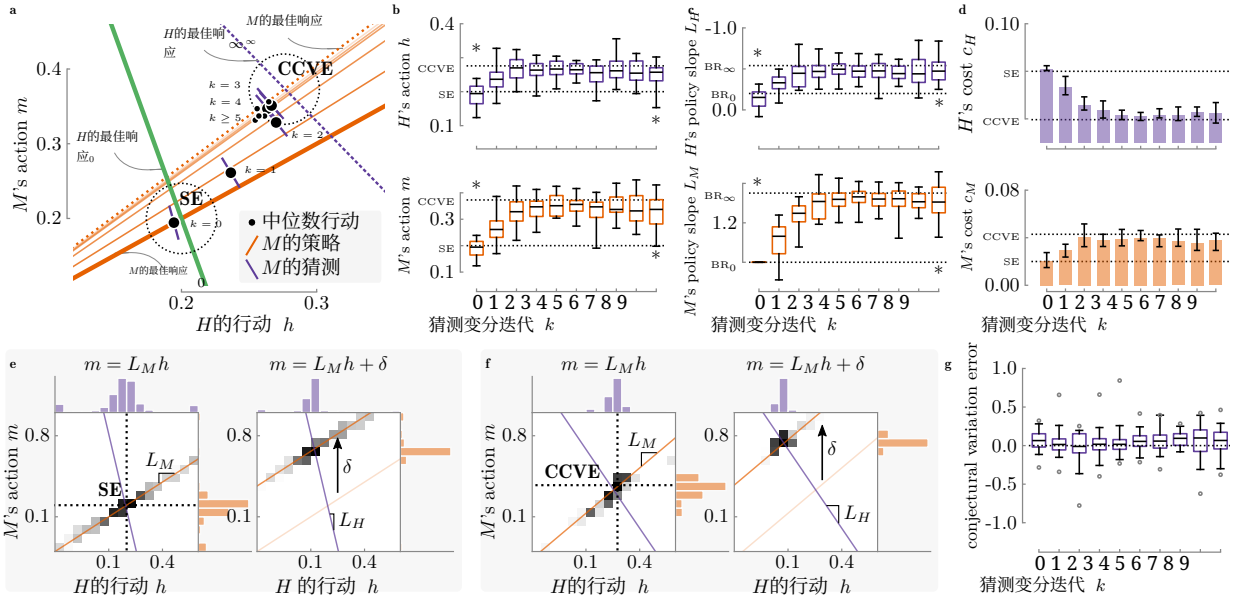

图2：策略空间中的猜测变分（实验 2，$n = 20$）。实验设置和成本与图1a、b相同，但机器采用了不同的适应算法：在本实验中，M 迭代实现并更新仿射策略 $m = L_M h$, $m = L_M + \delta$，以测量人类策略的猜测并对此作出最佳回应。（a）每次猜测变分迭代 k 的中位数行动、猜测和策略，并叠加显示与初始迭代和极限迭代时最佳响应（BR）相对应的博弈论均衡（分别为 BR_0 和 BR_∞）；这些均衡分别根据博弈的斯塔克尔伯格均衡和一致猜想变动均衡（SE 和 CCVE）预测得出。（b）每次迭代的行动分布，以箱线图显示，格式同图1d；统计显著性（$*$）按照相同的检验方法确定，分别与 SE（显示在分布下方）和 CCVE（显示在上方）进行比较。（c）每次迭代的策略斜率分布，显示方式与（b）相同；请注意，为与其他图保持一致，顶部 y轴的符号反转。统计显著性（$*$）按照（b）中的方法确定，通过双侧 t检验（$^*P \leq 0.05$）分别与初始最佳响应（显示在分布下方）和极限最佳响应（显示在上方）进行比较。（d）每次迭代的成本分布，以箱线图显示，格式同图1e。（e、f）不同迭代的行动一维和二维直方图（$k = 0$ 对应（e），$k = 9$ 对应（f））并叠加显示策略和博弈论均衡（（e）中为 SE 和 BR_0，（f）中为 CCVE 和 BR_∞）。（g）每次迭代中测得的机器对人类策略的猜测与理论预测值之间的误差，以箱线图显示，格式同（b、c）。

双侧 t检验，df 19；确切统计量见表 S1）。这种变化有利于人类，但代价由机器承担（图2d）。在所有猜测变分迭代中，机器的经验猜测与人类策略的理论预测均无显著差异（图2g；$P > 0.05$；双侧 t检验，df 19；确切统计量见表 S1），这表明人类和机器都对对手作出了一致的猜测。

在我们的第三项实验（图3）中，机器采用策略梯度方法调整其仿射策略（Chasnov等，2020）。试次仍以配对形式进行，不过这次每对试次中机器策略的差异仅在线性项上。完成一对试次后，我们用这两个试次的成本中位数估算机器成本相对于其策略中线性项的梯度，并沿着与梯度相反的方向调整线性项，以降低成本。通过反复迭代这一过程，人类受试者群体（不同于前两项实验中的群体）的中位数行动向量分布，从人类主导的斯塔克尔伯格均衡（东南）向机器的全局最优解移动（图3a）；后者也可视为逆斯塔克尔伯格均衡（Ho 等，1982）（RSE），此时优化机器的成本，而代价由人类承担（图3d）。从首次迭代到最后一次迭代，我们观察到行动空间中的分布显著地从东南移向逆斯塔克尔伯格均衡（图3b；$^*P \leq 0.05$；双侧 t 检验，自由度为19；精确统计量见表 S1），而最终的策略分布与东南和逆斯塔克尔伯格均衡策略均显著不同（图3c；$^*P \leq 0.05$；双侧 t 检验，自由度为19；精确统计量

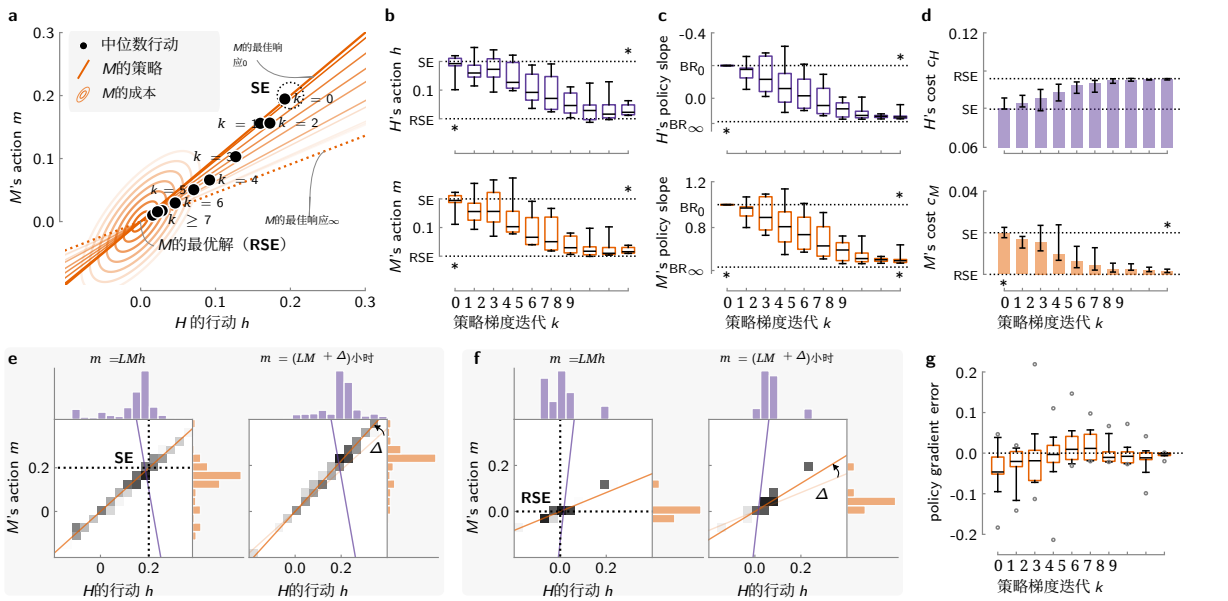

图3：策略空间中的梯度下降（实验3，$n = 20$）。实验设置和成本与图1a、b相同，但机器采用了不同的适应算法：在本实验中，M迭代实施线性策略 $m = L_M h$、$m = (L_M + \Delta)h$，以测量其成本相对于策略斜率参数 L_M 的梯度，并更新该参数以沿成本景观下降。(a) 每次策略梯度迭代 k 的中位数行动和策略，叠加了对应于机器在初始和极限迭代时最佳响应（最佳响应）的博弈论均衡（分别为 BR_0 和 BR_∞），这些均衡分别由斯塔克尔伯格均衡（东南）和机器的全局最优解（RSE）预测。(b) 每次迭代的动作分布，以与图1d相同的箱线图显示；统计显著性（$*$）采用相同检验，以双侧 t检验（$^*P \leq 0.05$）分别与东南（分布上方所示）和 M的最优解（分布下方所示）进行比较来确定；(c) 每次迭代的策略斜率分布，采用与(b)相同的标记方式显示；注意，为与其他图保持一致，顶部子图的y轴符号反转。统计显著性（$*$）按(b)中的方法，通过双侧 t检验（$^*P \leq 0.05$）分别与东南（分布上方所示）和 RSE（下方）进行比较来确定。(d) 每次迭代的成本分布，采用与图1e和图2d相同的箱线图显示。(e、f) 不同迭代（(e)中为$k = 0$，(f)中为 $k = 9$）的行动一维和二维直方图，叠加了策略和博弈论均衡（(e)中为东南，(f)中为 RSE）。(g) 每次迭代测得的策略斜率与理论预测的策略斜率之间的误差，采用与(b、c)相同的箱线图显示。

见表 S1）。然而，机器的实测策略梯度与理论预测值并无显著差异（图 3g；$P > 0.05$；双侧 t 检验，自由度为 19；精确统计量见表 S1），机器成本的最终分布与最优值也无显著差异（图 3d；$P > 0.05$；单侧 t 检验，自由度为19；精确统计量见表 S1），这表明机器能够准确估算其策略梯度并将成本降至最低。从本质上说，在这项实验中，机器通过在策略空间中进行推理来提升自身的博弈水平，从而引导博弈结果达到其在联合行动空间中期望的位置。我们在扩展数据（章节 B.1、B.2）中报告了采用不同初始化和机器最优解的该实验变体结果。

讨论

当机器在我们的实验中采取任何策略时（即机器的行动 m 被确定为人类行动的函数时 h），它实际上对人的优化问题施加了约束。策略可能是间接产生的，例如在第一个实验中，机器以很快的速度沿着其成本的梯度下降；也可能是直接采用的，例如在第二和第三个实验中。在所有三个实验中，人类行动或策略的经验分布都与针对每种机器策略的人类约束优化问题的解析解一致（图 1d；图2b、c；图3b、c）。这一发现意义重大，因为它表明，人类行为的最优性对于我们设定的成本以及

机器施加的约束具有稳健性，这表明我们的结果或可推广到人们（近似地）优化自身效用函数的其他情境。补充材料（第B.3节）报告了三项实验在采用非二次成本函数时的变体结果。

自适应机器有望作为远程机器人或协作机器人（Nikolaidis等，2017）、计算机与大脑或身体之间的接口（De Santis，2021；Perdikis和d. R. Millán，2020），以及外骨骼或假肢等设备（Felt等，2015；Slade等，2022；Zhang等，2017），协助人类完成工作和日常生活活动。然而，如何设计能够与人类良好协作的自适应算法，仍是机器人学、神经工程学和机器学习领域尚待解决的问题；人类会不断从周围世界中学习并适应环境（Nikolaidis等，2017；Perdikis和d. R. Millán，2020；Recht，2019）。我们验证了博弈论方法：机器可在与人类伙伴共同适应交互的过程中，通过塑造结果来提供辅助。重要的是，我们的方法不需要求解逆优化问题（Li等，2019；Ng和Russell，2000）——机器无需估计人类的成本函数，而是直接从人类行动中学习。这一特点或许有助于新兴的辅助设备身体／人在回路优化范式（Felt等，2015；Slade等，2022；Zhang等，2017），在这种范式中，机器的成本函数会特意设定为尊重人类的代谢能量消耗（Abram等，2022）或其他偏好（Ingraham等，2022）。

我们的结果展示了机器在人类对手参与的共同适应交互中所展现的能力。尽管人类在每项实验中都通过选择最优行动，在一个层面上作出了理性回应，但在三项实验过程中，机器通过在策略空间中进行高层次博弈，得以“智胜”对手。如果人类提升到相同的推理层次，这种机器优势或可被削弱，但机器还可以进一步提升，从理论上可能导致一种众所周知的无限回归（Harsanyi，1967）。我们在实践中并未观察到这种回归，这可能是因为人类受试者和机器可用的计算资源都存在上限（Gershmanet al.，2015）。

结论

随着机器算法渗透到日常生活的更多方面，了解它们能够对人类施加何种影响，对于防止不良行为、确保问责并最大限度地增进个人和社会福祉至关重要（Cooper et al.，2022；Thomas et al.，2019）。尽管人类和机器的能力都受到可用资源的限制，但人类理性存在众所周知的局限（Tverskyand Kahneman，1974），而机器则受益于计算资源、训练数据和算法创新的持续增长（Hilbertand López，2011；Jordan andMitchell，2015）。本文展示了，在与人类受试者进行的共同适应博弈中，机器能够单方面改变其学习策略，从理论预测的多种结果中进行选择。因此，机器学习算法可能有助于人类伙伴，例如支持决策，或在人们行动受限时提供帮助。然而，当机器目标与人的目标不一致时，可能有必要对算法施加限制，以确保人的安全、自主性和福祉。

参考文献

Sabrina J Abram、Katherine L Poggensee、Natalia Sánchez、Surabhi N Simha、James M Finley、Steven H Collins 和 J Maxwell Donelan。总体变异性促使个体朝着最优运动策略进行特异性适应。《当代生物学》*: CB*，32(10):2222–2232.e5，2022年5月。doi: 10.1016/ j.cub.2022.04.015。

Tamer Başar 和 Geert Jan Olsder。《动态非合作博弈论》。SIAM，1998。Tamer Başar 和 Hasan Selbuz。用于多层系统最优控制的闭环斯塔克尔伯格策略。《IEEE 自动控制汇刊》，24(2):166–179，1979。Dimitri P Bertsekas。《非线性规划》。Athena Scientific，第2版，1999。Arthur Lyon Bowley。《经济学的数学基础：导论性论著》。Claren- don Press，1924。Benjamin Chasnov、Lillian Ratliff、Eric Mazumdar 和 Samuel A. Burden。连续博弈中基于梯度学习的收敛性分析。载于《人工智能不确定性大会（UAI）》，《机器学习研究论文集》第115卷，第935–944页，2020。A Feder Cooper、Emanuel Moss、Benjamin Laufer 和 Helen Nissenbaum。算法社会中的问责：机器学习中的关系性、责任与鲁棒性。载于《ACM 公平、问责与透明度大会（FAccT）》，第864–876页，2022年6月。doi: 10.1145/3531146.3533150。Dalia De Santis。身体－机器接口中共同适应优化的框架。《神经机器人学前沿》，15:40，2021。doi: 10.3389/fnbot.2021.662181。Gerard Debreu。估值均衡与帕累托最优。《美国国家科学院院刊》，40(7):588–592，1954。doi: 10.1073/pnas.40.7.588。Wyatt Felt、Jessica C Selinger、J Maxwell Donelan 和 C David Remy。“身体在环”：利用瞬时能量成本指标优化设备参数。《PLoS ONE》，10 (8):e0135342，2015。doi: 10.1371/journal.pone.0135342。Tanner Fiez、Benjamin Chasnov 和 Lillian Ratliff。斯塔克尔伯格博弈中的隐式学习动态：均衡刻画、收敛性分析与实证研究。载于《ACM 国际机器学习大会（ICML）》，第3133–3144页。PMLR，2020。Charles Figuières、Alain Jean-Marie、Nicolas Quérou 和 Mabel Tidball。《猜测变动理论》。World Scientific，2004。ISBN 9789812387363。Samuel J Gershman、Eric J Horvitz 和 Joshua B Tenenbaum。计算理性：大脑、心智与机器智能的一种趋同范式。《科学》，349(6245):273–278，2015。Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。载于 Z Ghahramani、M Welling、C Cortes、N D Lawrence 和 K Q Weinberger 编，《神经信息处理系统进展（NeurIPS）》，第27卷，第2672–2680页。Curran Associates, Inc.，2014。Nore Berta Groot、Bart De Schutter 和 Johannes Hellendoorn。《逆向斯塔克尔伯格博弈：交通控制中的理论与应用》。博士论文，代尔夫特理工大学，2013。John C Harsanyi。“贝叶斯”参与者进行的不完全信息博弈，I–III。《管理科学》，8:159–182，320–334，486–502，1967。Sandra G Hart 和 Lowell E Staveland。NASA-TLX（任务负荷指数）的开发：实证与理论研究结果。《心理学进展》，52:139–183，1988。

James B Heald、Máté Lengyel 和 Daniel M Wolpert。情境推断是学习感觉运动行为库的基础。《自然》，600(7889):489–493，2021年。Martin Hilbert 和 Priscila López。全球存储、通信和计算信息的技术能力。《科学》，332(6025):60–65，2011年4月。doi: 10.1126/science.1200970。Yu-Chi Ho、P Luh 和 R Muralidharan。信息结构、斯塔克尔伯格博弈与激励可控性。*IEEE* 自动控制汇刊，26(2):454–460，1981年4月。doi: 10.1109/TAC.1981.1102652。Yu-Chi Ho、Peter B Luh 和 Geert Jan Olsder。激励的控制理论视角。《自动化》，18(2):167–179，1982年3月。doi: 10.1016/0005-1098(82)90106-6。J Huang、A Isidori、L Marconi、M Mischiati、E Sontag 和 W M Wonham。控制、生物学和神经科学中的内模。载于*IEEE*决策与控制会议 *(CDC)*，第5370–5390页，2018年12月。doi: 10.1109/CDC.2018.8619624。Kimberly A Ingraham、C David Remy 和 Elliott J Rouse。用户偏好在机器人外骨骼定制控制中的作用。《科学·机器人学》，7(64):eabj3487，2022年3月。doi: 10.1126/scirobotics.abj3487。Michael I Jordan 和 Tom M Mitchell。机器学习：趋势、观点与前景。《科学》，349(6245):255–260，2015年7月。doi: 10.1126/science.aaa8415。Yanan Li、Gerolamo Carboni、Franck Gonzalez、Domenico Campolo 和 Etienne Burdet。用于多功能物理人机交互的微分博弈论。《自然·机器智能》，1(1):36–43，2019年。Yi-An Ma、Yuansi Chen、Chi Jin、Nicolas Flammarion 和 Michael I Jordan。采样可能比优化更快。《美国国家科学院院刊》，116(42):20881–20885，2019年10月。doi: 10.1073/pnas.1820003116。Maneeshika M Madduri、Samuel A Burden 和 Amy L Orsborn。用于共适应脑机接口的博弈论模型。载于*IEEE/EMBS*神经工程会议 *(NER)*，第327–330页。IEEE，2021年。Ninareh Mehrabi、Fred Morstatter、Nripsuta Saxena、Kristina Lerman 和 Aram Galstyan。机器学习中的偏见与公平性综述。《*ACM*计算调查》，54(6):1–35，2021年。John F Nash。 N人博弈中的均衡点。《美国国家科学院院刊》，36(1):48–49，1950年1月。doi: 10.1073/pnas.36.1.48。Andrew Y Ng 和 Stuart J Russell。逆向强化学习算法。载于*ACM* 国际机器学习大会 *(ICML)*，第663–670页，2000年。Stefanos Nikolaidis、Swaprava Nath、Ariel D Procaccia 和 Siddhartha Srinivasa。人机协作中人类适应的博弈论建模。载于*ACM/IEEE*人机交互会议 *(HRI)*，第323–331页，2017年。Stefan Palan 和 Christian Schitter。Prolific.ac——在线实验的受试者样本库。《行为与实验金融学杂志》，17:22–27，2018年3月。ISSN 2214-6350。doi: 10.1016/j.jbef. 2017.12.004。

Serafeim Perdikis 和 José d. R. Millán。脑机接口：两个学习者的故事。*IEEE*《系统、人类与控制论杂志》，6(3):12–19，2020年7月。doi: 10.1109/MSMC.2019. 2958200。Lillian J Ratliff、Samuel A Burden 和 S Shankar Sastry。连续博弈中局部纳什均衡的刻画。*IEEE* 自动控制汇刊，61(8):2301–2307，2016年。Benjamin Recht。强化学习概览：连续控制视角。《控制、机器人与自主系统年度评论》*(ARCRAS)*，2(1):253–279，2019年5月。doi: 10.1146/annurev-control-053018-023825。Patrick Slade、Mykel J. Kochenderfer、Scott L. Delp 和 Steven H. Collins。在真实世界行走时个性化外骨骼辅助。《自然》，610(79317931):277–282，2022年10月。ISSN 1476-4687。doi: 10.1038/s41586-022-05191-1。Reed T Sutton、David Pincock、Daniel C Baumgart、Daniel C Sadowski、Richard N Fedorak 和 Karen I Kroeker。临床决策支持系统概述：益处、风险与成功策略。*NPJ*《数字医学》，3(1):1–10，2020年。Jordan A Taylor、John W Krakauer 和 Richard B Ivry。感觉运动适应任务中对学习的显性与隐性贡献。《神经科学杂志》，34(8):3023–3032，2014年。Philip S Thomas、Bruno Castro da Silva、Andrew G Barto、Stephen Giguere、Yuriy Brun 和 Emma Brunskill。防止智能机器产生不良行为。《科学》，366(6468): 999–1004，2019年11月。doi: 10.1126/science.aag3311。Amos Tversky 和 Daniel Kahneman。不确定情境下的判断：启发式与偏差。《科学》，185(4157):1124–1131，1974年9月。doi: 10.1126/science.185.4157.1124。Guido Van Rossum 和 Fred L. Drake。《*Python 3* 参考手册》。CreateSpace，美国加利福尼亚州斯科茨谷，2009年。ISBN 1441412697。Hal R Varian。微观经济分析。诺顿公司，1992年。John Von Neumann 和奥斯卡·摩根斯特恩。博弈论与经济行为。普林斯顿大学出版社，1947年。海因里希·冯·施塔克尔贝格。《市场结构与均衡》。施普林格，1934年。Daniel M Wolpert、Zoubin Ghahramani 和 Michael I Jordan。感觉运动整合的内部模型。《科学》，269(5232):1880–1882，1995年。doi: 10.1126/science.7569931。Juanjuan Zhang、Pieter Fiers、Kirby A Witte、Rachel W Jackson、Katherine L Poggensee、Christo-pher G Atkeson 和 Steven H Collins。行走过程中外骨骼辅助的人在回路优化。《科学》，356(6344):1280–1284，2017年6月。doi: 10.1126/science.aal5054。Ying-Ping Zheng 和 Tamer Başar。部分信息下斯塔克尔伯格博弈最优仿射激励方案的存在性与推导：一种几何方法。《国际控制杂志》，35(6):997–1011，1982年。

方法

实验方案。 人类受试者通过在线众包研究平台 *Prolific*（Palan 和 Schitter，2018年）招募。实验采用经华盛顿大学机构审查委员会（UW IRB STUDY00013524）批准的程序进行。参与者数据收集于安全的网络服务器上。每项实验均由一系列试次组成：第一次实验为14个试次，第二次和第三次实验各为20个试次。在每个试次中，参与者使用网络浏览器查看图形界面，并通过鼠标或触摸屏进行手动输入，以持续确定标量动作的值 $h \in \mathbb{R}$。该光标输入按参与者网络浏览器窗口的宽度进行缩放，使得 $h = -1$对应左边缘，而 $h = +1$ 对应右边缘。第一次实验中，每个试次持续40秒；第二次和第三次实验中，每个试次持续20秒，数据采集速率均为每秒60个样本。人类受试者从Prolific上所有可用国家的"标准样本"研究分布中选取。每名受试者仅参与三项实验中的一项。未采用其他筛选标准。

在每项实验开始时，都会向参与者展示介绍界面，其中包含任务描述和用户说明。在每个试次开始时，参与者需要将光标移动到随机确定的位置。此程序用于在实验初始化中引入随机性，并评估参与者注意力。在整个试次期间，矩形的高度显示人类成本的当前值 $c_H(h,\ m)$，并要求参与者通过选择行动 $h \in \mathbb{R}$，同时机器更新其行动 $m \in \mathbb{R}$，来"将此 [矩形] 保持得 尽可能 小"。对成本值应用平方根函数，使参与者更容易感知低成本值中的细微差异。经过固定时长后，一个试次结束，下一个试次开始。每三个试次之间，参与者都有机会休息半分钟。实验在完成固定数量的试次后结束。随后，参与者填写任务负荷调查（Hart 和 Staveland，1988）和可选的反馈表。每项实验持续约 10–14 分钟，参与者获得 2 美元的固定补偿（所有数据均于 2020 年收集）。Movie S1 展示了实验 1 最初三个试次的视频。所有实验中，向人类受试者呈现的用户界面均完全相同。然而，机器在每项实验期间都会调整其行动和策略，并且每项实验所使用的适应算法各不相同。

成本函数。 在实验 1、2 和 3 中，参与者采用了二次成本函数

$$c_H(h,m) = \tfrac{1}{2}h^2 + \tfrac{7}{30}m^2 - \tfrac{1}{3}hm + \tfrac{2}{15}h - \tfrac{22}{75}m + \tfrac{12}{125}; \tag{1}$$

机器优化了二次成本函数

$$c_M(h,m) = \tfrac{1}{2}m^2 + h^2 - hm. \tag{2}$$

这些成本经过设计，使博弈参与者的最优点以及相关博弈论均衡的分布位置各不相同，如表 1 所列。在每次实验的每个试次中，记录各试次行动的时间序列，分别作为人类行动 h_0、. . .、h_t、. . .、h_T 和机器行动 m_0、. . .、m_t、. . .、m_T，采样数量固定为 T。在时间 t，博弈参与者经历的成本为 $c_H(h_t, m_t)$ 和 $c_M(h_t, m_t)$。相关博弈论均衡的形式化定义见补充材料 S1 节，成本参数的选择方法见补充材料 S2 节。

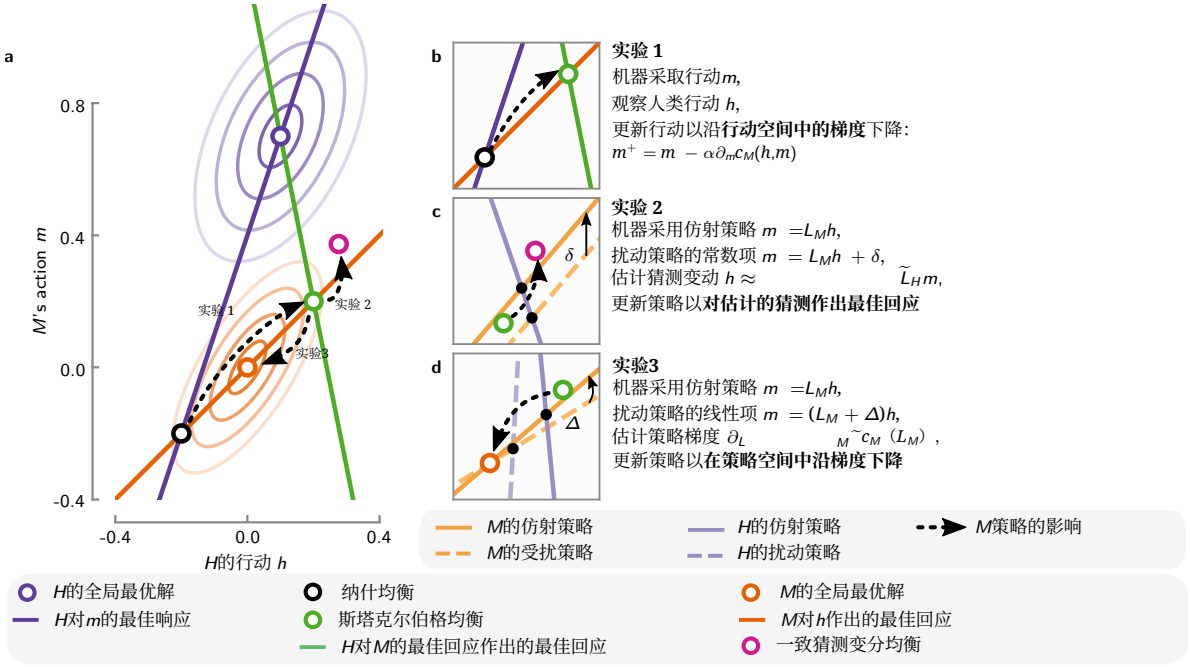

图 4：**人类与机器共同适应实验概览**。人类受试者 H 被要求提供手动输入 h，使计算机显示屏上的黑色条形尽可能小。机器 M 有其自身预先设定的成本 c_M，该成本经过选择，使博弈论均衡彼此不同，并且与每个参与者的全局最优解不同。（**a**）联合行动空间，展示博弈论均衡以及根据为人类和机器设定的成本确定的响应函数：全局最优解通过对两个变量同时求最小值来定义；最佳响应函数则通过固定一个变量并对另一个变量求最小值来定义。机器在三个实验中采用不同的策略：（**b**）在实验 *1*中使用梯度下降；（**c**）在实验 *2*中使用猜测变动；（**d**）在实验*3*中使用策略梯度下降。

实验 1：行动空间中的梯度下降。在第一个实验中，机器使用梯度下降来调整其行动，

$$m^+ = m - \alpha\, \partial_m c_M(h, m), \tag{3}$$

采用七种不同的适应率选择之一 $\alpha \in \{0, 0.003, 0.01, 0.03, 0.1, 0.3, 1\}$。在最慢的适应率下 $\alpha = 0$，机器执行常数策略 $m = -0.2$，这是机器在博弈纳什均衡中的组成部分。在最快的适应率下 $\alpha = 1$，（3）中的梯度下降迭代使机器执行线性策略 $m = h$。每位人类受试者在每种条件下体验两次，分别对应每种对称性（见下一段），顺序随机。

为避免人类受试者记住博弈均衡的位置，每个试次开始时，均从$\{-1, +1\}$ 中均匀随机选取一个变量 s，并在整个试次期间将映射 $h \mapsto s\,h$应用于人类受试者的手动输入。当该变量的值为 $s = -1$ 时，输入会呈现“镜像”对称性。联合行动在正方形 $[-0.4, +0.4] \times [-0.4, +0.4] \subset \mathbb{R}^2$ 内均匀随机初始化。每个试次持续 40 秒。

实验 2：策略空间中的猜测变分。 在第二个实验中，机器通过估计关于人类策略的猜测来调整自己的策略。为了收集用于形成估计的数据，机器在连续两个试次中采用了一个仿射策略，

这两个试次的策略仅在常数项上有所不同，

$$\text{nominal policy } m = L_M h, \tag{4a}$$
$$\text{perturbed policy } m' = L_M h' + \delta. \tag{4b}$$

机器从这对试次中取中位数行动向量（$h, \widetilde{m}$）、（$h', \widetilde{m}'$），并利用差值之比来估计对人类策略的猜测，

$$\widetilde{L}_H = \frac{\widetilde{h}' - \widetilde{h}}{\widetilde{m}' - \widetilde{m}}, \tag{5}$$

如补充材料第 4 命题（见补充材料第 S3.2 节）所示，这一估计值反映了人类行动对机器行动的响应变化。机器利用对人类策略的这一估计更新自身策略，更新方式如下

$$L_M^+ = \frac{1 - 2\widetilde{L}_H}{1 - \widetilde{L}_H}, \tag{6a}$$

如补充材料第 S3 节所示，这就是基于机器对人类策略的猜测所得到的机器最佳响应。在下一对试次中，机器采用 $m = L_M^+ h + \ell_M^+$ 作为其策略。该猜测变动过程从初始猜测$\sim L_H = 0$开始迭代 10 次，得到初始最佳响应策略 $m = h$。

在本实验中，每次猜测变动迭代 $k \in \{0, \ldots, K-1\}$均记录机器策略斜率 $L_{M,0}, L_{M,1}, \ldots, L_{M,k}, \ldots, L_{M,K-1}$ 以及机器对人类策略斜率的猜测 $L_{H,0}, L_{H,1}, \ldots, L_{H,k}, \ldots, L_{H,K-1}$，其中 $K = 10$ 次迭代。此外，还像第一次实验那样记录每个试次内行动的时间序列，不过现在每个试次仅持续 20 秒，产生 $T = 1200$个样本，用于计算式（5）中使用的中位数行动向量。

实验3：策略空间中的梯度下降。在第三个实验中，机器通过采用策略梯度策略来调整其策略：在两个连续的试次中执行仿射策略，这两个试次的区别仅在于线性项，

$$\text{nominal policy } m = L_M h, \tag{7a}$$
$$\text{perturbed policy } m' = (L_M + \Delta) h'. \tag{7b}$$

机器使用这两个试次中的中位数行动向量（h，$\widetilde{m}$）和（h'，$\widetilde{m}'$）来估计机器策略中线性项对应的机器成本梯度，并调整该线性项以降低成本。具体而言，定义辅助成本为

$$\widetilde{c_M}(L_M) := c_M\left(h, L_M(h - h_M^*) + m_M^*\right), \tag{8}$$

并利用这两个试次获得机器策略斜率对应的机器成本梯度的有限差分估计，

$$\partial_{L_M} \widetilde{c_M}(L_M) \approx \frac{1}{\Delta}\left(\widetilde{c_M}(L_M + \Delta) - \widetilde{c_M}(L_M)\right). \tag{9}$$

机器利用这一导数估计，通过沿成本梯度下降来更新其策略中的线性项，

$$L_M^+ = L_M - \gamma\, \partial_{L_M} \widetilde{c_M}(L_M) \tag{10}$$

其中 γ 是策略梯度适应率参数（在本实验中为 $\gamma = 2$）。

成本函数与博弈论均衡

H**的成本函数**	M**的成本函数**
$c_H(h,m) = \frac{1}{2}h^2 + \frac{7}{30}m^2 - \frac{1}{3}hm + \frac{2}{15}h - \frac{22}{75}m + \frac{12}{125}$	$c_M(h,m) = \frac{1}{2}m^2 + h^2 - hm$

博弈论均衡	H**和** M**的行动**	H **和** M **的策略斜率**
H 的最优值	$(h_H^*, m_H^*) = (+0.1, +0.7)$	
M 的最优值	$(h_M^*, m_M^*) = (0, 0)$	
纳什均衡	$(h^{\mathrm{NE}}, m^{\mathrm{NE}}) = (-0.2, -0.2)$	
人类主导的斯塔克尔伯格均衡	$(h^{\mathrm{SE}}, m^{\mathrm{SE}}) = (+0.2, +0.2)$	$L_H^{\mathrm{SE}} = -0.2, \quad L_M^{\mathrm{SE}} = 1$
一致猜测变分均衡	$(h^{\mathrm{CCVE}}, m^{\mathrm{CCVE}}) \approx (0.276, 0.373)$	$L_H^{\mathrm{CCVE}} \approx -0.54, \; L_M^{\mathrm{CCVE}} \approx +1.35$
机器主导的反向斯塔克尔伯格均衡 (等于 M 的最优值)	$(h^{\mathrm{RSE}}, m^{\mathrm{RSE}}) = (0, 0)$	$L_H^{\mathrm{RSE}} = 1/7, \quad L_M^{\mathrm{RSE}} = 5/11$

表 1：**实验 1、2 和 3 中所研究博弈的成本函数和博弈论均衡**。 补充材料详细说明
成本的选择方式：章节S2 介绍总体方法，第S2.7则针对本文研究的博弈进行了具体说明。

统计分析。 为确定结果的统计显著性，我们对包含$n = 20$ 名受试者的群体中位数数据分布进行单侧或双侧 t 检验，阈值为 $P \leq 0.05$。为估算效应量，我们计算了 Cohen's d：从分布均值中减去均衡值，再除以该分布的标准差。

数据可用性

所有数据均可通过 Code Ocean capsule 公开获取， `codeocean.com/capsule/6975866`。

代码可用性

复现主论文和补充材料中报告的所有图表及统计结果所需的数据和分析脚本，已公开发布于 Code Ocean capsule 中， `codeocean.com/capsule/6975866`。用于在 Prolific 平台上开展实验的源代码已公开发布于 GitHub， `github.com/dynams/web`。

致谢

本研究由美国国家科学基金会（资助编号：#1836819、#2045014）资助。Benjamin Chasnov 部分经费由计算神经科学研究生培训计划（美国国立卫生研究院 5T90DA032436-09 MPI）资助。

作者贡献

B.J.C.、L.J.R. 和 S.A.B. 负责方法设计和论文撰写；B.J.C. 收集并分析了实验数据，并制作了图表。

利益冲突

作者声明不存在利益冲突。

补充信息

本文提供补充信息。如需通信和索取材料，请联系 S.A.B.（电子邮件： sburden@uw.edu）。

以下内容的补充信息人类对自适应机器的适应收敛于博弈论均衡

Benjamin J. Chasnov、Lillian J. Ratliff、Samuel A. Burden

电气与计算机工程系 华盛顿大学，西雅图，华盛顿州 98195，美国

补充材料列表:

S1 节–S7 图 S1–S6 表 S1–S4 协议 S1–S3 源代码 S0–S3 Movie S1

补充材料摘要

本补充信息为主论文中的论述提供支持。

博弈论均衡解的正式数学定义见第S1节。一对二次成本的参数由第S2节中的均衡解确定。主论文中博弈的分析见第S3节。行动空间中梯度下降的实验 1 分析见第S3.1节。策略空间中猜测变分的实验 2 分析见第S3.2节。策略空间中梯度下降的实验3分析见第S3.3节。

猜想迭代的解释见第S4节。比较静态分析的相关经济学思想将在第S4.1 节中介绍，并在第S4.2节中使用泰勒近似刻画一致猜测。

章节

S1 博弈论定义 **18**
S1.1 纳什均衡与斯塔克尔伯格均衡.18S1.2 一致猜想变动均衡.18S1.3 反向斯塔克尔伯格均衡.19

S2 博弈设计 **20**
S2.1 全局最优解.20S2.2 纳什均衡.21S2.3 人类主导的斯塔克尔伯格均衡.21S2.4 k层猜测变动均衡.21S2.5 一致猜测变分均衡.22S2.6 机器主导的逆斯塔克尔伯格均衡.22S2.7 为具有单维行动的双人博弈选择参数 23

主论文中二次博弈的 S3 分析 **25**
S3.1 实验 1：行动空间中的梯度下降.25S3.2 实验 2：策略空间中的猜测变分.26S3.3 实验3：策略空间中的梯度下降.30

S4 一致猜测变分的解释 **32**
S4.1 比较静态分析.33S4.2 通过泰勒级数近似确定一致性阶数.33

A 补充方法 **35**
A.1 实验 1：行动空间中的梯度下降.36A.2 实验 2：策略空间中的猜测变分.36A.3 实验3：策略空间中的梯度下降.36

B 附加实验结果 **36**
B.1 机器初始化（实验3）.36B.2 机器最优点（实验3）.36B.3 非二次成本（修改后的实验 1、2 和 3）.37B.4 数值模拟.37B.5 一致性与帕累托最优性.37B.6 数值模拟.41

C 任务负荷 cyp 调查与反馈表 **44**
C.1 任务负荷调查.44C.2 可选反馈.45

图

1 实验 1.32 实验 2.43 实验3.54 人类与机器之间的协同适应博弈.11S1 采用不同初始策略的实验 3.46S2 采用不同机器最优解的实验3.46S3 采用非二次成本的修改后的实验 1、2 和 3.47S4 实验 1、2 和 3 的仿真.47S5 比较帕累托最优性与猜想一致性.48

表格

1 成本函数和博弈均衡.13S1 统计检验的精确值.38S2 共同适应博弈的符号和术语.40S3 三项实验的符号和术语.40S4 任务负荷调查结果.44S5 反馈结果.45

协议

S1 实验 1 的算法描述.39S2 实验 2 的算法描述.39S3 实验3 的算法描述.39

源代码

S0 两位参与者的参数、成本函数和梯度定义.42S1 实验 1 的数值模拟.42S2 实验 2 的数值模拟.43S3 实验3 的数值模拟.43

S1 博弈论定义

我们使用博弈论对人类与机器之间的共同适应进行建模（Başar 和 Olsder，1998；冯·诺伊曼和摩根斯特恩，1947年）。在该模型中，人类 H 选择行动 $h \in \mathcal{H}$，而机器 M 选择行动 $m \in \mathcal{M}$，以最小化各自的成本函数 c_H、c_M：$\mathcal{H} \times \mathcal{M} \to \mathbb{R}$，

$$\min_h \; c_H(h, m), \tag{11a}$$

$$\min_m \; c_M(h, m). \tag{11b}$$

需要注意的是，(11) 中的优化问题是相互耦合的。由于必须同时考虑这两个问题，因此没有明显适用的"解"概念（这与纯优化问题不同，在纯优化问题中，单个成本函数的（局部）极小化解显然就是目标）。因此，我们设计了实验，以研究在不同情境下自然出现的各种候选解概念。我们展示了纳什均衡、斯塔克尔伯格均衡、一致猜想变动均衡以及博弈参与者的全局最优解都可能是实验结果。

S1.1 纳什和斯塔克尔伯格均衡

在参与者同时行动且不对其他参与者的策略形成猜测的博弈中，一个自然的候选解概念是纳什均衡 （定义4.1，见（Başar 和 Olsder，1998）。

定义： 联合行动 $(h^{\mathrm{NE}}, m^{\mathrm{NE}}) \in \mathcal{H} \times \mathcal{M}$ 构成一个纳什均衡（东北），当且仅当

$$h^{\mathrm{NE}} = \arg\min_h c_H(h, m^{\mathrm{NE}}), \tag{12a}$$

$$m^{\mathrm{NE}} = \arg\min_m c_M(h^{\mathrm{NE}}, m). \tag{12b}$$

在序贯行动的博弈中，如果领导者（例如人类）知道跟随者（例如机器）在其选择自身行动时会如何响应，那么一个自然的候选解概念是（由人类领导的）斯塔克尔伯格均衡 （定义4.6，见（Başar 和 Olsder，1998）。

定义： 联合行动 $(h^{\mathrm{SE}}, m^{\mathrm{SE}}) \in$ 人类$\times \mathcal{M}$ 构成（人类主导的）斯塔克尔伯格均衡（SE），当且仅当

$$h^{\mathrm{SE}} = \arg\min_h \left\{ c_H\left(h, m\right) \mid m = \arg\min_{m'} c_M(h, m') \right\}, \tag{13a}$$

$$m^{\mathrm{SE}} = \arg\min_m \; c_M(h^{\mathrm{SE}}, m). \tag{13b}$$

斯塔克尔伯格均衡是一种解概念，适用于一位参与者（领导者）预判或建模另一位参与者（跟随者）的最佳响应的情形。

S1.2 一致猜测变异均衡

在每个参与者都能观察另一方行动和策略的重复博弈中，参与者可能会形成对另一方预期如何行动的内部模型或猜测。一个自然的候选方案

此情况下的解概念是一致猜测变动均衡 (定义 4.9，见 (Başar 和 Olsder，1998)。

对于给定的一对[1] $(v_H^{\mathrm{CCVE}}, v_M^{\mathrm{CCVE}}) \in \{\mathcal{M} \to \mathcal{H}\} \times \{\mathcal{H} \to \mathcal{M}\}$，记满足以下条件的唯一不动点$(h^{\mathrm{CCVE}}, m^{\mathrm{CCVE}}) \in \mathcal{H} \times \mathcal{M}$ 为

$$h^{\mathrm{CCVE}} = v_H^{\mathrm{CCVE}} \circ v_M^{\mathrm{CCVE}}(h^{\mathrm{CCVE}}), \tag{14a}$$

$$m^{\mathrm{CCVE}} = v_M^{\mathrm{CCVE}} \circ v_H^{\mathrm{CCVE}}(m^{\mathrm{CCVE}}). \tag{14b}$$

Let

$$\Delta v_H^{\mathrm{CCVE}}(m) = v_H^{\mathrm{CCVE}}(m) - v_H^{\mathrm{CCVE}}(m^{\mathrm{CCVE}}), \tag{15a}$$

$$\Delta v_M^{\mathrm{CCVE}}(h) = v_M^{\mathrm{CCVE}}(h) - v_M^{\mathrm{CCVE}}(h^{\mathrm{CCVE}}), \tag{15b}$$

它们表示每个参与者在其策略（$v_H^{\mathrm{CCVE}}, v_M^{\mathrm{CCVE}}$）下，对联合行动（$h^{\mathrm{CCVE}}, m^{\mathrm{CCVE}}$）偏离至（$m, h$）时的微分反应。

定义： 联合行动（$h^{\mathrm{CCVE}}, m^{\mathrm{CCVE}}$）$\in \mathcal{H} \times \mathcal{M}$ 连同猜测 v_M^{CCVE}:$\mathcal{H} \to \mathcal{M}$、$v_H^{\mathrm{CCVE}}$:$\mathcal{M} \to \mathcal{H}$ 构成一个一致猜测变分均衡（CCVE），当行动满足一致性时，

$$h^{\mathrm{CCVE}} = \arg\min_h \left\{ c_H(h, m) \mid m = v_M^{\mathrm{CCVE}}(h) \right\},$$

$$m^{\mathrm{CCVE}} = \arg\min_m \left\{ c_M(h, m) \mid h = v_H^{\mathrm{CCVE}}(m) \right\},$$

且策略满足一致性

$$v_H^{\mathrm{CCVE}}(m) = \arg\min_h \; c_H(h, m + \Delta v_M^{\mathrm{CCVE}}(h)),$$

$$v_M^{\mathrm{CCVE}}(h) = \arg\min_m \; c_M(h + \Delta v_H^{\mathrm{CCVE}}(m), m).$$

一致猜测变分均衡是一种解概念，它源于博弈参与者对彼此行动及其反应的预期。

S1.3 逆斯塔克尔伯格均衡

在一个参与者（领导者）能够先于另一参与者（跟随者）施加策略，而跟随者对该策略作出响应的博弈中，适用于此情形的候选解概念是逆斯塔克尔伯格均衡（Ho 等，1981，1982）。在该博弈中，机器充当领导者，并宣布策略为 $\pi : \mathcal{H} \to \mathcal{M}$。假设人类对机器策略的最优响应 π 为$r : (\mathcal{H} \to \mathcal{M}) \to \mathcal{H}$ 由一个约束优化问题给出：

$$r(\pi) := \arg\min_h \left\{ c_H(h, m) \mid m = \pi(h) \right\}.$$

定义： 联合行动（$h^{\mathrm{RSE}}, m^{\mathrm{RSE}}$）$\in \mathcal{H} \times \mathcal{M}$ 与机器策略 π^{RSE}: $\mathcal{H} \to \mathcal{M}$共同构成一个逆斯塔克尔伯格均衡（RSE），当且仅当

$$\pi^{\mathrm{RSE}} = \arg\min_\pi \left\{ c_H(h, m) \mid m = \pi(h),\ h = r(\pi) \right\}, \tag{16a}$$

$$h^{\mathrm{RSE}} = r(\pi^{\mathrm{RSE}}), \tag{16b}$$

$$m^{\mathrm{RSE}} = \pi^{\mathrm{RSE}}(h^{\mathrm{RSE}}). \tag{16c}$$

如果逆斯塔克尔伯格问题是激励可控的（Ho 等，1981），那么逆斯塔克尔伯格均衡就是机器的全局最优解。

[1]我们用简写 $\{A \to B\}$ 表示从 A 到 B的函数集合。

S2 博弈设计

在本节中，通过求解线性方程并满足特定的二阶条件和稳定性条件，推导出均衡点。一般二次成本为

$$c_H(h,m) = \tfrac{1}{2}h^\top A_H h + h^\top B_H m + \tfrac{1}{2}m^\top D_H m + b_H^\top h + d_H^\top m + a_H, \tag{17a}$$
$$c_M(h,m) = \tfrac{1}{2}m^\top A_M m + m^\top B_M h + \tfrac{1}{2}h^\top D_M h + b_M^\top m + d_M^\top h + a_M. \tag{17b}$$

其中，行动 $h \in \mathbb{R}^p$、 $m \in \mathbb{R}^q$ 是向量，具有 $p \geq 1$ 和 $q \geq 1$；成本参数 $A_H \in \mathbb{R}^{p\times p}$、 $D_H \in \mathbb{R}^{q\times q}$、 $A_M \in \mathbb{R}^{q\times q}$、 $D_M \in \mathbb{R}^{p\times p}$ 是对称矩阵， $B_H \in \mathbb{R}^{p\times q}$、 $B_M \in \mathbb{R}^{q\times p}$ 是矩阵，$b_H \in \mathbb{R}^p$、 $d_H \in \mathbb{R}^q$、 $b_M \in \mathbb{R}^p$、 $d_M \in \mathbb{R}^q$ 是向量， $a_H \in \mathbb{R}$、 $a_M \in \mathbb{R}$ 是标量。

选择成本参数，使均衡点位于行动空间中指定的位置。不失一般性，令 A_H 和 A_M 为单位矩阵，以设定每位参与者成本的（任意）尺度。随后，确定 a_H、 a_M，使两位参与者的最低成本值均为 0。最后，同样不失一般性，确定 $b_M = d_M = 0$ ，使机器的成本以联合行动空间的原点为中心。剩下的六个待确定系数为 B_H、 B_M、 D_H、 D_M、 b_H、 d_H。这些参数将决定该博弈均衡解的位置。

在主论文中，行动空间是标量，即 $p = q = 1$。人类的参数选定为 $A_H = 1$, $B_H = -1/3$, $D_H = 7/15$, $b_H = 2/15$, $d_H = -22/75$ ，机器的参数选定为 A_M =1, $B_M = -1$, $D_M = 2$, $b_M = 0$, $d_M = 0$ 。该博弈中参与者的最优点为

$$(h_H^*, m_H^*) = (0.1, 0.7),$$
$$(h_M^*, m_M^*) = (0, 0),$$

并且博弈论均衡是

$$(h^{\mathrm{NE}}, m^{\mathrm{NE}}) = (-0.2, -0.2),$$
$$(h^{\mathrm{SE}}, m^{\mathrm{SE}}) = (0.2, 0.2),$$
$$(h^{\mathrm{CCVE}}, m^{\mathrm{CCVE}}) \approx (0.276, 0.373),$$
$$(h^{\mathrm{RSE}}, m^{\mathrm{RSE}}) = (0, 0).$$

在以下小节中，针对成本，写出优化问题解的一阶和二阶条件 c_H， c_M在 (17a) 和 (17b)。

S2.1 全局最优解

两个参与者的全局优化问题是

$$(h_H^*, m_H^*) = \operatorname*{argmin}_{h,m} \; c_H(h,m),$$
$$(h_M^*, m_M^*) = \operatorname*{argmin}_{h,m} \; c_M(h,m)$$

其满足一阶条件

$$\begin{bmatrix} A_H & B_H \\ B_H^\top & D_H \end{bmatrix} \begin{bmatrix} h_H^* \\ m_H^* \end{bmatrix} + \begin{bmatrix} b_H \\ d_H \end{bmatrix} = 0 \text{ and } \begin{bmatrix} D_M & B_M^\top \\ B_M & A_M \end{bmatrix} \begin{bmatrix} h_M^* \\ m_M^* \end{bmatrix} + \begin{bmatrix} d_M \\ b_M \end{bmatrix} = 0,$$

以及满足以下条件的二阶条件 $\begin{bmatrix} A_H & B_H \\ B_H^\top & D_H \end{bmatrix}$ 和 $\begin{bmatrix} D_M & B_M^\top \\ B_M & A_M \end{bmatrix}$ 均为半正定。参见（Bertsekas，1999）中对这些条件的正式表述。

S2.2 纳什均衡

纳什均衡（$h^{\mathrm{NE}}, m^{\mathrm{NE}}$）对应的耦合优化问题为

$$h^{\mathrm{NE}} = \underset{h}{\operatorname{argmin}}\ c_H(h, m^{\mathrm{NE}}),$$
$$m^{\mathrm{NE}} = \underset{m}{\operatorname{argmin}}\ c_M(h^{\mathrm{NE}}, m),$$

其满足一阶条件

$$\begin{bmatrix} A_H & B_H \\ B_M & A_M \end{bmatrix} \begin{bmatrix} h^{\mathrm{NE}} \\ m^{\mathrm{NE}} \end{bmatrix} + \begin{bmatrix} b_H \\ b_M \end{bmatrix} = 0$$

并满足二阶条件 $A_H \geq 0$ 和 $A_M \geq 0$。如果雅可比矩阵 $\begin{bmatrix} A_H & B_H \\ B_M & A_M \end{bmatrix}$ 的特征值实部均为正，那么在梯度博弈下，纳什均衡是稳定的。

关于局部纳什均衡的必要条件，请参见（Ratliff 等，2016）中的命题 1；关于连续时间梯度博弈动态 $\dot{h} = -\partial_h c_H(h, m)$、$\dot{m} = -\partial_m c_M(h, m)$ 的稳定性结果，也请参见该命题。关于相应的离散时间梯度博弈动态 $h^+ = h - \beta \partial_h c_H(h, m)$、$m^+ = m - \alpha \partial_M c_M(h, m)$，以及学习率 α、$\beta > 0$ 和学习率比值 $\tau = \alpha/\beta$，请参见（Chasnov 等，2020）中的命题 2。随着学习率比值 τ 趋近于 ∞，机器的行动 m 适应速度快于 h，这使两个参与者之间形成时间尺度分离。

S2.3 人类主导的斯塔克尔伯格均衡

耦合的　　人类主导的斯塔克尔伯格均衡 $(h^{\mathrm{SE}}, m^{\mathrm{SE}})$ 的优化问题是

$$h^{\mathrm{SE}} = \underset{h}{\operatorname{argmin}} \left\{ c_H(h, m') |\ m' = \underset{m}{\operatorname{argmin}}\ c_H(h, m) \right\},$$
$$m^{\mathrm{SE}} = \underset{m}{\operatorname{argmin}}\ c_M(h^{\mathrm{SE}}, m),$$

其一阶条件为

$$\begin{bmatrix} A_H + L_{M,0}^\top B_H^\top & B_H + L_{M,0}^\top D_H \\ B_M & A_M \end{bmatrix} \begin{bmatrix} h^{\mathrm{SE}} \\ m^{\mathrm{SE}} \end{bmatrix} + \begin{bmatrix} b_H + L_{M,0}^\top d_H \\ b_M \end{bmatrix} = 0$$

满足 $L_{M,\,0} = -A_M^{-1} B_M$，以及二阶条件 $A_M > 0$，$A_H - B_H A_M^{-1} B_M > 0$。关于斯塔克尔伯格均衡的二次博弈表述，参见（Başar 和 Olsder，1998）中的命题 4.3，该表述仅允许纯策略斯塔克尔伯格均衡。关于局部斯塔克尔伯格均衡的条件，参见（Fiez 等人，2020）中的命题 1。

S2.4 k层级猜测变化均衡

对于中间猜测变动均衡，人类对机器保持一致猜测时的耦合优化问题为

$$h_{k+1}^{\mathrm{CVE}} = \underset{h}{\operatorname{argmin}}\ \{c_H(h, m') |\ m' = L_{M,k}(h - h_M^*) + m_M^*\},$$
$$m_k^{\mathrm{CVE}} = \underset{m}{\operatorname{argmin}}\ \{c_M(h', m) |\ h' = L_{H,k-1}(m - m_H^*) + h_H^*\},$$

其一阶最优性条件为

$$\begin{bmatrix} A_H + L_{M,k}^\top B_H^\top & B_H + L_{M,k}^\top D_H \\ B_M + L_{H,k-1}^\top D_M & A_M + L_{H,k-1}^\top B_M^\top \end{bmatrix} \begin{bmatrix} h_{k+1}^{\mathrm{CVE}} \\ m_k^{\mathrm{CVE}} \end{bmatrix} + \begin{bmatrix} b_H + L_{M,k}^\top d_H \\ b_M + L_{H,k-1}^\top d_M \end{bmatrix} = 0$$

以 $L_{M.0} = -A_M^{-1} B_M$ 为初始条件进行迭代

$$\begin{aligned} L_{H,k+1} &= -(A_H + L_{M,k}^\top B_H^\top)^{-1}(B_H + L_{M,k}^\top D_H) \\ L_{M,k} &= -(A_M + L_{H,k-1}^\top B_M^\top)^{-1}(B_M + L_{H,k-1}^\top D_M) \end{aligned}$$

对于 $k = 0$，1，2，.....在假设 $A_H + B_H L_{M、k}$和 $A_M + B_M L_{H、k-1}$均可逆的条件下。有关该迭代在主要实验所用成本的特定参数下收敛的条件，详见第 S3 节。

S2.5 一致的猜测变动均衡

根据（Başar 和 Olsder，1998中的定义 4.9），一致的猜测变动均衡对应的耦合优化问题为

$$\begin{aligned} h^{\mathrm{CCVE}} &= \underset{h}{\operatorname{argmin}} \ \{c_H(h, m') \mid m' = L_M^{\mathrm{CCVE}}(h - h_M^*) + m_M^*\} \\ m^{\mathrm{CCVE}} &= \underset{m}{\operatorname{argmin}} \ \{c_M(h', m) \mid h' = L_H^{\mathrm{CCVE}}(m - m_H^*) + h_M^*\} \end{aligned}$$

其中 L_M^{CCVE}，L_H^{CCVE}满足策略空间方程中的最优性条件（见（Başar 和 Olsder，1998中的定义 4.10）：

$$\begin{aligned} A_M L_M^{\mathrm{CCVE}} + {L_H^{\mathrm{CCVE}}}^\top B_M^\top L_M^{\mathrm{CCVE}} + {L_H^{\mathrm{CCVE}}}^\top D_M + B_M = 0, \\ A_H L_H^{\mathrm{CCVE}} + {L_M^{\mathrm{CCVE}}}^\top B_H^\top L_H^{\mathrm{CCVE}} + {L_M^{\mathrm{CCVE}}}^\top D_H + B_H = 0. \end{aligned}$$

一阶最优　　耦合优化问题的行动空间中的最优性条件 ar　　e

$$\begin{bmatrix} A_H + {L_M^{\mathrm{CCVE}}}^\top B_H^\top & B_H + {L_M^{\mathrm{CCVE}}}^\top D_H \\ B_M + {L_H^{\mathrm{CCVE}}}^\top D_M & A_M + {L_H^{\mathrm{CCVE}}}^\top B_M \end{bmatrix} \begin{bmatrix} h^{\mathrm{CCVE}} \\ m^{\mathrm{CCVE}} \end{bmatrix} + \begin{bmatrix} b_H + {L_M^{\mathrm{CCVE}}}^\top d_H \\ b_M + {L_H^{\mathrm{CCVE}}}^\top d_M \end{bmatrix} = 0.$$

命题4.5（Başar 和 Olsder，1998）指出，如果一个博弈存在唯一的纳什均衡，那么该纳什均衡同时也是一个 CCVE，其中纳什行动构成常值策略。

S2.6 机器主导的反向斯塔克尔伯格均衡

对应于机器主导的反向斯塔克尔伯格均衡的耦合优化问题如下：

$$\begin{aligned} r_H^{\mathrm{RSE}}(L_M) &= \underset{h}{\operatorname{argmin}} \ \{c_H(h, m') \mid m' = L_M(h - h_M^*) + m_M^*\} \\ L_M^{\mathrm{RSE}} &= \underset{L_M}{\operatorname{argmin}} \ \{c_M(r_H^{\mathrm{RSE}}(L_M), m') \mid m' = L_M(r_H^{\mathrm{RSE}}(L_M) - h_M^*) + m_M^*)\} \end{aligned}$$

其中，人类对机器形成一致猜测，而机器则假设人类会对机器策略斜率作出最优响应。逆斯塔克尔伯格均衡

是$(h^{\mathrm{RSE}},\ m^{\mathrm{RSE}})$，根据（Başar 和 塞尔布兹，1979；Groot 等，2013），满足与机器最优点相同的条件，即

$$\begin{bmatrix} A_M & B_M \\ B_M^\top & D_M \end{bmatrix} \begin{bmatrix} h^{\mathrm{RSE}} \\ m^{\mathrm{RSE}} \end{bmatrix} + \begin{bmatrix} b_M \\ d_M \end{bmatrix} = 0$$

以及一阶最优性条件

$$\begin{bmatrix} A_H + L_M^{\mathrm{RSE}\top} B_H^\top & B_M + L_M^{\mathrm{RSE}\top} D_H \\ -L_M^{\mathrm{RSE}} & I \end{bmatrix} \begin{bmatrix} h^{\mathrm{RSE}} \\ m^{\mathrm{RSE}} \end{bmatrix} + \begin{bmatrix} b_H + L_M^{\mathrm{RSE}\top} d_H \\ m_M^* - L_M^{\mathrm{RSE}\top} h_M^* \end{bmatrix} = 0$$

其中，我们还需要确保雅可比矩阵是稳定的。二阶条件为 $A_H + B_H L_M^{\mathrm{RSE}} > 0$。关于求解逆斯塔克尔伯格问题的方法，参见（Ho 等，1981），该方法依赖于线性激励可控性这一性质。关于相关结果的综述以及最优策略的计算，参见（Groot 等，2013）。关于最优仿射领导者策略的存在性，参见（郑和 Başar，1982）中的命题 1。

S2.7 双人博弈中单维动作的参数选择

给定具有标量动作 $h \in \mathbb{R}$、 $m \in \mathbb{R}$ 的二次成本，

$$\begin{aligned} c_H(h,m) &= \tfrac{1}{2}A_H h^2 + B_H hm + \tfrac{1}{2}D_H m^2 + b_H h + d_H m + a_H, \\ c_M(h,m) &= \tfrac{1}{2}A_M m^2 + B_M hm + \tfrac{1}{2}D_M h^2 + b_M m + d_M h + a_M. \end{aligned}$$

不失一般性，令 $A_H = 1$ 和 $A_M = 1$ 分别为各参与者成本的尺度。用最优点（h_H^*、 m_H^*）和（h_M^*、 m_M^*）表示的参数为

$$\begin{aligned} a_H &= \tfrac{1}{2}A_H {h_H^*}^2 + B_H h_H^* m_H^* + \tfrac{1}{2}D_H {m_H^*}^2, \quad b_H = -A_H h_H^* - B_H m_H^*, \quad d_H = -B_H h_H^* - D_H m_H^*, \\ a_M &= \tfrac{1}{2}A_M {m_M^*}^2 + B_M h_M^* m_M^* + \tfrac{1}{2}D_M {h_M^*}^2, \quad b_M = -A_M m_M^* - B_M h_M^*, \quad d_M = -B_M m_M^* - D_M h_M^*. \end{aligned}$$

参数以最优点和纳什均衡（$h^{\mathrm{NE}}, m^{\mathrm{NE}}$）表示为

$$B_H = -\frac{h_H^* - h^{\mathrm{NE}}}{m_H^* - m^{\mathrm{NE}}},\ B_M = -\frac{m_M^* - m^{\mathrm{NE}}}{h_M^* - h^{\mathrm{NE}}}.$$

以最优点和人类主导的斯塔克尔伯格均衡（$h^{\mathrm{SE}}, m^{\mathrm{SE}}$）表示的参数为

$$\begin{aligned} D_H = &\frac{B_H(h_M^* m_H^* + h_H^* m_M^* - (m_H^* + m_M^* - m^{\mathrm{SE}})h^{\mathrm{SE}} - (h_H^* + h_M^* - h^{\mathrm{SE}})m^{\mathrm{SE}})}{(m_H^* - m^{\mathrm{SE}})(m_M^* - m^{\mathrm{SE}})} \\ &+ \frac{(h_H^* - h^{\mathrm{SE}})(h_M^* - h^{\mathrm{SE}})}{(m_H^* - m^{\mathrm{SE}})(m_M^* - m^{\mathrm{SE}})} \end{aligned}$$

且 $A_H - B_H A_M^{-1} B_M$ 必须为正定矩阵。

剩下需要选择的参数是 D_M。它必须满足以下条件：

$$\begin{aligned} (A_H A_M - D_H D_M)^2 - 4(A_M B_H - B_M D_H)(A_H B_M - B_H D_M) \geq 0, \\ (A_M B_H - B_M D_H)(A_H B_M - B_H D_M) \neq 0 \end{aligned}$$

CCVE 由两个二次方程的解确定。每个代理人的策略斜率为

$$L_H^{\mathrm{CCVE}} = \frac{D_H D_M - A_H A_M \pm \sqrt{4(A_M B_H - B_M D_H)(B_H D_M - A_H B_M) + (A_H A_M - D_H D_M)^2}}{2A_H B_M - 2B_H D_M},$$
$$L_M^{\mathrm{CCVE}} = \frac{D_H D_M - A_H A_M \pm \sqrt{4(A_M B_H - B_M D_H)(B_H D_M - A_H B_M) + (A_H A_M - D_H D_M)^2}}{2A_M B_H - 2B_M D_H},$$

且行动为

$$\begin{bmatrix} h^{\mathrm{CCVE}} \\ m^{\mathrm{CCVE}} \end{bmatrix} = \begin{bmatrix} A_H + L_M^{\mathrm{CCVE}} B_H & B_M + L_M^{\mathrm{CCVE}} D_H \\ B_M + L_H^{\mathrm{CCVE}} D_H & A_M + L_H^{\mathrm{CCVE}} B_M \end{bmatrix}^{-1} \begin{bmatrix} b_H + L_M^{\mathrm{CCVE}} d_H \\ b_M + L_H^{\mathrm{CCVE}} d_M \end{bmatrix}$$

逆斯塔克尔伯格均衡由策略斜率决定

$$L_H^{\mathrm{RSE}} = \frac{h_H^* - h_M^*}{m_H^* - m_M^*},\ L_M^{\mathrm{RSE}} = -\frac{A_H L_H^{\mathrm{RSE}} + B_H}{B_H L_H^{\mathrm{RSE}} + D_H},$$

以及行动 $h^{\mathrm{RSE}} = h_M^*$、$m^{\mathrm{RSE}} = m_M^*$。

主论文中的二次博弈 S3 分析

本节给出了关于双人博弈（c_H， c_M）的数学表述，其中每位参与者的目标都是优化以下函数：

$$c_H(h,m) = \tfrac{1}{2}h^2 + \tfrac{7}{30}m^2 - \tfrac{1}{3}hm + \tfrac{2}{15}h - \tfrac{22}{75}m + \tfrac{12}{125}. \tag{1}$$

针对人类和

$$c_M(h,m) = \tfrac{1}{2}m^2 + h^2 - hm. \tag{2}$$

针对机器。在实验 1 中，机器通过梯度下降优化其行动。在实验 2 中，机器通过猜测变分优化其策略。在实验3 中，机器通过梯度下降优化其策略。在所有实验中，人类通过尽可能降低成本 c_H (h,m) 来更新其行动h。

本节分析了论文中的三项主要实验。通过耦合优化问题的均衡解预测实验结果。三个小节分别包含数学命题，用以证明与相应实验有关的论断。命题 1 和 2 适用于实验 1。它们证明了系统会收敛到唯一的纳什均衡解和斯塔克尔伯格均衡解。命题3、4、5、6 和 7 适用于实验 2。它们证明，机器可以扰动自身策略，以估计人类的猜测变动，并进而利用该估计形成最佳响应迭代，使其收敛到一致猜测变分均衡。命题 8、10、9、11 适用于实验3。它们证明，机器可以扰动自身策略，以估计其策略梯度，并进而利用该估计更新策略，使其收敛到全局最优解。均衡解的正式定义见第 S1节。

人机协同适应博弈是由两个成本函数决定的双人重复博弈，每位参与者各对应一个成本函数。博弈按以下方式进行：在每个时间步 t，人类选择行动 $h_t \in \mathcal{H}$。机器通过选择行动 $m_t \in \mathcal{M}$作出最佳响应。人类通过界面观察成本 $c_H(h_t, m_t)$。在下一个时间步 $t+1$，下一组行动 (h_{t+1}, m_{t+1}) 将被选定；这一过程持续固定的步数 T。在我们的每项实验中，机器用来更新其行动的方法各不相同。

S3.1 实验 1：行动空间中的梯度下降

以下命题 1 描述了 $\alpha = 0$ 实验 1 的情形，其中博弈的结果是唯一的稳定纳什均衡 $(m,\ h) = (-1/5,\ -1/5)$。这一结果在实证中得到观察（主论文的图2）。

命题 1。 给定一个由成本函数 (1)和 (2)决定的人机协同适应博弈，若机器的行动为 $m = -1/5$，则人类的最佳响应为 $h = -1/5$。

*Pr*唉。 从人类的角度来看，目标是解决这个优化问题

$$\min_h \ c_H(h,m) \tag{18}$$

（18）的二阶条件是

$$\partial_h^2 c_H(h,m) = 1 > 0.$$

优化问题（18）的一阶条件为

$$\partial_h c_H(h,m) = h - \tfrac{1}{3}m + \tfrac{2}{15} = 0. \tag{19}$$

通过求解（ h 中的19），人类对 m 的最优响应为

$$h = \frac{1}{3}m - \frac{2}{15}.$$

求解 h 可得人类的最优响应 $h = \frac{1}{3}m - \frac{2}{15}$。因此，如果 $m = -\frac{1}{5}$，那么 $h = -\frac{1}{5}$。 □

以下命题 2 描述了 $\alpha = 1$ （或“无穷”）情况下的实验 1，此时结果是在 $(m,\ h) = (1/5,\ 1/5)$ 处该博弈唯一稳定的人类主导的斯塔克尔伯格均衡。该结果已通过实证观察到（主论文中的图2）。

命题 2。 给定一个由成本函数 （1） 和 （2）决定的人机协同适应博弈，若机器策略为 $m = h$，则人类的最优响应为 $h = 1/5$。

证明。 从人类的角度来看，优化问题为

$$\min_h \{c_H(h, m) \mid m = h\} \tag{20}$$

人类承担的成本为

$$c_H(h, h) = \frac{2}{5}h^2 - \frac{4}{25}h + \frac{12}{125}$$

（20）的一阶条件为

$$\partial_h c_H(h, h) = \frac{4}{5}h - \frac{4}{25} = 0$$

解得 h，得到 $h = \frac{1}{5}$。 □

备注 1。 给定一个由成本函数决定的人机协同适应博弈 (1) 和 (2)，如果 $0 < \alpha \leq 1$ 并且机器更新其行动 $m_{t+1} = m_t - \alpha\partial_m c_M(h_t,\ m_t)$，那么 m_{t+1} 会趋近于 h_t，当 t 增大时。通过将更新写成 $m_{t+1} = (1-\alpha)m_t + \alpha h_t$，可以证明这一结果；由此可知，序列 m_t、m_{t+1}、……由时变信号 h_t 的指数平滑滤波器生成。

备注 1 在主论文的图2二维直方图中有所体现：当 $m = h$ 取较大 α 值时，点分布在等值线上。

S3.2 实验 2：策略空间中的猜测变分

在实验 2 中，机器在策略空间中迭代猜测变分。从人类的角度来看，目标是选择 h，以优化 $c_H(h, m)$。但是，m 的确定方式会影响耦合优化问题的解。从机器的角度来看，目标是选择 m，以优化 $c_M(h, m)$。同样，对 h 的假设会影响机器的响应。机器估计猜测变动，以描述 h 如何受 m 变化的影响。

以下命题 3描述了实验 1 中机器的策略扰动。人类的响应与机器的常数扰动呈线性关系 δ，但与机器策略斜率呈非线性关系 L。

命题 3。 给定一个由成本函数 (1)和 (2)确定的人机协同适应博弈，若机器策略为 $m = Lh + \delta$ 且 L 满足 $\frac{7}{15}L^2 - \frac{2}{3}L + 1 > 0$，则人类的最优响应为

$$h = \frac{22L - 10 - (35L - 25)\delta}{35L^2 - 50L + 75}$$

证明。 人的优化问题为

$$\min_h \{c_H(h,m) \mid m = Lh + \delta\} \tag{21}$$

式（21）的二阶条件为

$$\tfrac{7}{15}L^2 - \tfrac{2}{3}L + 1 > 0.$$

式（21）的一阶条件为

$$(\tfrac{7}{15}L^2 - \tfrac{2}{3}L + 1)h - \tfrac{22}{75}L + \tfrac{2}{15} - (\tfrac{7}{15}L_M + \tfrac{1}{3})\delta = 0$$

求解 h 可得结果。

□

下面的命题 4 说明机器如何通过扰动机器策略的常数项所生成的两个点来估计人类策略的斜率。

命题 4。 给定一个由成本函数 （1）和 （2）确定的人机协同适应博弈，若机器策略为 $m = Lh$ 和 $m' = Lh' + \delta$，且人类最优响应为 h 和 h'，则

$$\frac{h' - h}{m' - m} = \frac{7L - 5}{5L - 15}$$

证明。 根据命题 3，对于 h' 和 h,

$$h' - h = -\frac{35L - 25}{35L^2 - 50L + 75}\delta.$$

根据 m' 和 m的定义,

$$m' - m = L(h' - h) + \delta.$$

因此，差值之比为

$$\frac{h' - h}{m' - m} = \frac{-\left(\frac{35L-25}{35L^2-50L+75}\delta\right)}{-L\left(\frac{35L-25}{35L^2-50L+75}\delta\right) + \delta} = \frac{35L - 25}{L(35L - 25) - (35L^2 - 50L + 75)} = \frac{7L - 5}{5L - 15}.$$

□

注释 2。 在主论文中，人类策略斜率为 L_H，机器策略斜率为 L_M。对于实验*2*和实验*3*中的机器策略 $m = Lh$，这些项之间的关系为

$$L_M = L,$$
$$L_H = \frac{7L - 5}{5L - 15}.$$

在这种情况下，人类对机器的猜测与机器策略一致。均衡解由线性方程描述

$$m = L_M h + \ell_M$$
$$h = L_H m + \ell_H$$

其中 $\ell_M = 0$和 $\ell_H = -\frac{22L-10}{25L-75}$。

备注 2 可以通过沿实数轴扫描 S5 ，得到图 H 对 M 有一致猜测时的实线椭圆曲线。 L 。

以下命题5 描述了在人类采用基于命题 4中猜测变化的策略时，机器的最佳响应。

命题 5. 给定一个由成本函数 (1)和 (2)确定的人机协同适应博弈，若人类策略为 $h=\left(\frac{7L-5}{5L-15}\right)m+\ell$，其中 ℓ,—则机器的最佳响应为

$$m=\frac{9L+5}{2L+10}h$$

证明。 机器的优化问题为

$$\min_m\left\{c_M(h,m)\mid h=\left(\tfrac{7L-5}{5L-15}\right)m+\ell\right\}. \tag{22}$$

(22) 的一阶条件为

$$\partial_m c_M(h,m)+\partial_h c_M(h,m)\left(\tfrac{7L-5}{5L-15}\right)=0. \tag{23}$$

二阶条件为

$$2\left(\tfrac{7L-5}{5L-15}\right)^2-2\left(\tfrac{7L-5}{5L-15}\right)+1>0.$$

取（23）中的一阶条件，得方程

$$m-h+(2h-m)\left(\tfrac{7L-5}{5L-15}\right)=0$$

对 m 求解可得机器的最佳响应

$$m=\frac{9L+5}{2L+10}h$$

□

注释 3. 常数项 ℓ 可从命题 5中的联合行动测量估算得出。不过，要得出方程 (23)中的最优性条件，并不需要这样做。

下面的命题 6 证明了一致猜测变分均衡的存在性。均衡解概念在第S1节中定义。它描述了双方的行动和策略均保持一致的情形。

命题 6. 给定一个由成本函数 (1)和 (2)确定的人机协同适应博弈，存在两个由机器响应斜率唯一确定的一致猜测变分均衡解

$$L=\frac{-1\pm\sqrt{41}}{4}.$$

证明。 根据 Başar 和 Olsder（1998）中定义 4.10 的方程（1）和（1'），策略空间中一致猜测变动的驻定条件为

$$L-L\left(\tfrac{7L-5}{5L-15}\right)+2\left(\tfrac{7L-5}{5L-15}\right)-1=0, \tag{24}$$

化简（24）的分子后，以下二次方程定义了机器的一致策略斜率：

$$2L^2+L-5=0.$$

二次型方程的解给出了结果。 □

说明 4. 将 $L = \frac{-1\pm\sqrt{41}}{4}$ 代入即可确定人类策略斜率，结果为

$$\frac{7L-5}{5L-15} = \frac{1\mp\sqrt{41}}{10}.$$

因此，两个一致的猜测变动策略为

$$m = \frac{-1\pm\sqrt{41}}{4}h$$
$$h = \frac{1\mp\sqrt{41}}{10}m - \frac{3+7\sqrt{41}}{100}$$

以及解该线性方程的行动（m， h）。

以下 命题 7 表明，实验 2 收敛于一个稳定均衡 .

命题 7。 给定一个由成本函数 （1）和 （2）所确定的人机协同适应博弈，如果机器使用差分方程更新其策略 $L^+ = \frac{9L+5}{2L+10}$ 那么

$$L^* = \frac{-1+\sqrt{41}}{4}$$

是该迭代的局部指数稳定不动点。

证明。 定义映射 $F\colon \mathbb{R}\to\mathbb{R}$ 为

$$F(L) := \frac{9L+5}{2L+10} \tag{25}$$

为评估实验 2 的收敛性，将确定（25）的不动点及其稳定性。不动点 L^* 满足

$$L^* = F(L^*)$$

由二次方程的解决定

$$2L^2 + L - 5 = 0. \tag{26}$$

方程（26）有两个解，且它们都是真实且不相同的。不动点为

$$\frac{-1\pm\sqrt{41}}{4}.$$

恰有一个不动点是稳定的，并且它是反复应用 F时的稳定吸引子。可以通过在特定不动点处对（25）进行线性化，并确保其导数的模小于 1，来确定其稳定性。在不动点 F 处对 L^* 进行线性化，结果为

$$F(L) \approx \partial F(L^*)(L - L^*) \tag{27}$$

其中

$$\partial F(L) = \frac{20}{(5+L)^2}$$

如果 $L^* = \frac{-1+\sqrt{41}}{4}$ ，那么 $|\partial F(L^*)| \approx 0.5 < 1$，因此不动点 L^* 是稳定的。另一方面，如果 $L^* = \frac{-1-\sqrt{41}}{4}$ ，那么 $|\partial F(L^*)| > 1$，因此不动点 L^* 是不稳定的。 □

对于具有单维动作的二次博弈，存在两个一致猜想变动均衡。一个是稳定的，另一个是不稳定的。

备注5。 评估不动点映射 (25) 收敛性的另一种方法，是检查线性分式变换的标准形。式(25) 的标准形为

$$\frac{F(L)-L^*}{F(L)-L^{**}}=\lambda\frac{L-L^*}{L-L^{**}} \tag{28}$$

其中， L^* 和 L^{**} 是不动点， F 而 λ 是一个实数，其定义为

$$\lambda=\frac{-19+\sqrt{41}}{-19-\sqrt{41}} \tag{29}$$

由于 $|\lambda|\approx 0.5<1$，不动点 L^* 是半全局稳定的。

备注5 是一个 基于复分析和共形映射理论中的一个已知结果 .

S3.3 实验3：策略空间中的梯度下降

在实验3中，机器在策略空间中实施梯度下降。机器利用一对试次的成本测量来估计策略梯度。机器的成本取决于其自身策略以及人类对该策略的最佳响应。

以下命题 8描述了实验3中的机器策略扰动。人类的行动响应呈非线性变化。

命题 8。 给定一个由成本函数 (1)和 (2)确定的人机协同适应博弈，若机器策略为 $m=(L+\Delta)h$，且 L,Δ满足$\frac{7}{15}(L+\Delta)^2-\frac{2}{3}(L+\Delta)+1>0$，则人类的最佳响应为

$$h=\frac{22(L+\Delta)-10}{35(L+\Delta)^2-50(L+\Delta)+75}$$

证明。 人的优化问题为

$$\min_h\{c_H(h,m)\mid m=(L+\Delta)h\}. \tag{30}$$

式（30）的二阶条件为

$$\tfrac{7}{15}(L+\Delta)^2-\tfrac{2}{3}(L+\Delta)+1>0.$$

式（30）的一阶条件为

$$(\tfrac{7}{15}(L+\Delta)^2-\tfrac{2}{3}(L+\Delta)+1)h-\tfrac{22}{75}(L+\Delta)+\tfrac{2}{15}=0$$

解出 h 即可得到人类的响应

$$h=\frac{22(L+\Delta)-10}{35(L+\Delta)^2-50(L+\Delta)+75}. \tag{31}$$

□

以下命题 9 描述了如何像实验3中那样，使用两个试次来估计策略梯度。假设机器采用策略 $m = Lh$，那么人类的响应为

$$r(L) := \frac{22L - 10}{35L^2 - 50L + 75}$$

as 由命题决定3 或命题8 ，并将扰动设为零 o.

命题 9。 给定一个由成本函数 (1)和 (2)确定的人机协同适应博弈，若机器策略为 $m = Lh$和 $m' = (L+\Delta)h'$，且人类的最优响应为 $h = r(L)$和 $h' = r(L+\Delta)$，则

$$\lim_{\Delta \to 0} \frac{c_M(h', m') - c_M(h, m)}{\Delta} = D_L c_M(r(L), Lr(L))$$

证明。根据命题3，如果机器策略是 $m = Lh$ 且人类的最优响应是

$$h = \frac{22L - 10}{35L^2 - 50L + 75}.$$

机器的成本表示为 L 的函数是

$$\begin{aligned} c_M(h, m) = c_M(r(L), Lr(L)) &= \tfrac{1}{2}L^2 r(L)^2 + r(L)^2 - Lr(L)^2 \\ &= \tfrac{1}{2}(L^2 - 2L + 2)r(L)^2 \\ &= \frac{(L^2 - 2L + 2)(22L - 10)^2}{2(35L^2 - 50L + 75)^2} \end{aligned}$$

差异项为

$$c_M(h', m') - c_M(h, m) = c_M(r(L+\Delta), Lr(L+\Delta)) - c_M(r(L), Lr(L))$$

展开各项，并忽略阶数为 Δ^2 或更高的项，我们得到

$$\begin{aligned} c_M(h', m') - c_M(h, m) &= \frac{((L+\Delta)^2 - 2(L+\Delta) + 2)(22(L+\Delta) - 10)^2}{2(35(L+\Delta)^2 - 50(L+\Delta) + 75)^2} - \frac{(L^2 - 2L + 2)(22L - 10)^2}{2(35L^2 - 50L + 75)^2} \\ &= \frac{4(11L - 5)(2L^3 + 181L^2 - 380L + 305)}{25(7L^2 - 10L + 15)^3}\Delta + (\cdots)\,\Delta^2 + \cdots \end{aligned}$$

两边除以 Δ 并令 Δ 趋于零，得到的表达式与直接计算成本的导数相同：

$$\partial_L c_M(r(L), Lr(L)) = \frac{4(11L - 5)(2L^3 + 181L^2 - 380L + 305)}{25(7L^2 - 10L + 15)^3}.$$

因此，我们得到了所需的结果。 □

以下命题10 表明，该博弈存在唯一的机器主导的反向斯塔克尔伯格均衡。均衡解概念在第S1节中定义。它描述了领导者宣布一项策略、跟随者对该策略作出响应的情形。相比之下，命题 2 中的斯塔克尔伯格均衡描述的是领导者宣布一项行动、跟随者对该行动作出响应的情形。

命题 10. 对于由成本函数 (1) 和 (2)确定的人机协同适应博弈，存在一个逆斯塔克尔伯格均衡。

证明。 机器的全局最优解满足

$$\min_{h,m} c_M(h,m).$$

机器的全局最优解为（h,m） =(0,0)。

Su假设机器策略为 $m = Lh$，那么人的优化问题为

$$\min_{h}\{c_H(h,m) \mid m = Lh\}$$

最佳响应为

$$h = r(L) = \frac{22L - 10}{35L^2 - 50L + 75}$$

机器希望促使人类采取 $0 = r(L)$。因此，机器选择 $L = 5/11$。二阶条件为

$$\tfrac{7}{15}L^2 - \tfrac{2}{3}L + 1 > 0.$$

该条件由 $L = 5/11$ 满足。因此，(0, 0) 是机器主导的逆斯塔克尔伯格均衡。 □

以下命题11 表明实验3收敛到稳定均衡。

命题 11。 给定一个由成本函数 (1)和(2)确定的人机协同适应博弈，如果机器采用策略 $m = Lh$ 且人类以 $h = r(L)$ 作出响应，并且机器通过梯度下降更新其策略，

$$L_{k+1} = L_k - \alpha \partial_L c_M(r(L_k), L_k r(L_k))$$

那么，对于所有充分小的 $\alpha > 0$ ， $L^* = 5/11$ 是该迭代的局部指数稳定不动点。

证明。 根 o f $\partial_L c_M(r(L_k), L_k r(L_k)) = 0$ 由四次方程的解确定

$$(11L - 5)(2L^3 + 181L^2 - 380L + 305) = 0. \tag{32}$$

方程 (32) 有两个实数解，第一个解 $L^* = \frac{5}{11}$ 可通过观察得到，第二个解约为 $L^{**} \approx -92.6$。

稳定性由在特定不动点处进行线性化并确保二阶导数为正来确定。在根 L_M^* 处，线性化后的导数为

$$\partial_L c_M(r(L), Lr(L)) \approx \partial_L{}^2 c_M(r(L^*), L^* r(L^*))(L - L^*) \tag{33}$$

在 L^* 处计算的二阶导数 $\partial^2_{L_M} c_M \approx 0.18$ 为正，因此不动点 L_M^* 稳定。在 L^{**} 处计算的二阶导数为负，因此不动点不稳定。 □

一致猜测变分的 S4 解释

本节将阐释关于猜测变动的一致性条件。这些条件与经济学和非合作动态博弈中出现的偏微分方程有关。

S4.1 比较静态分析

比较静态分析（或更广义的敏感性分析）是微观经济学中的一项典型工具，用于在外生参数或干预发生变化时比较经济结果（Varian，1992）。如果表达式 $f(x,y) = 0$ 定义了某个经济体的均衡条件，其中 x 是内生参数（例如某种产品的价格），而 y 是外生参数（例如某种产品的需求），那么由 y 的（微小）变化所引起的 x 的变化，在一阶近似下必须满足 $\partial_x f \cdot \mathrm{d}x + \partial_y f \cdot \mathrm{d}y = 0$；在满足充分正则性的条件下，我们可以写成 $\mathrm{d}x/\mathrm{d}y = -(\partial_x f)^{-1} \cdot \partial_y f$。比较静态分析也可以应用于优化问题的均衡条件。

这里正是这样使用的：将比较静态分析应用于以下内容的一阶最优性条件：

$$\arg\min_m \{c_H(h,m) \mid m = \pi_M(h)\} \tag{34}$$

其中，将机器的行动视为干预。具体而言，给定一个仿射策略$\pi_M(h) = L_M h + \ell_M$和 (34)，我们使用这一微观经济学分析工具来理解 m 的变化如何引起 h 的变化，并使其符合 (34)的最优性条件。由此，我们可以推导出一个表达式，以策略参数 (L_M, ℓ_M) 和机器相应的行动 m表示人类的（最优）响应。 (34) 的一阶最优性条件为

$$0 = \partial_h c_H(h, \pi_M(h))|_{\pi_M(h)=m} + \partial_m c_H(h, \pi_M(h))|_{\pi_M(h)=m} \cdot \partial_h \pi_M(h), \tag{35a}$$
$$= \partial_h c_H(h, \pi_M(h))|_{\pi_M(h)=m} + \partial_m c_H(h, \pi_M(h))|_{\pi_M(h)=m} \cdot L_M. \tag{35b}$$

根据上述比较静态分析，我们有

$$0 = \partial_h^2 c_H(h,m)\mathrm{d}h + \partial_{hm}^2 c_H(h,m)\mathrm{d}m + (\partial_{hm}^2 c_H(h,m)\mathrm{d}h + \partial_m^2 c_H(h,m)\mathrm{d}m)L_M. \tag{36}$$

因此，我们推断

$$L_H := \frac{\mathrm{d}h}{\mathrm{d}m} = -(\partial_h^2 c_H + \partial_{hm} c_H \cdot L_M)^{-1}(\partial_{hm} c_H + L_M^\top \cdot \partial_m^2 c_H), \tag{37a}$$
$$= -(A_H + L_M^\top B_H)^{-1}(B_H + L_M^\top D_H). \tag{37b}$$

^在实验 2 中，我们将介绍一种估计人类响应的程序 h ，该响应是 m 的函数，方法是对 $\pi_M(h) = L_M h + \ell_M$ 进行仿射扰动。随后，机器将人类响应的估计值用作其对以下内容的猜测

$$\arg\min_m \{c_M(h,m) |\ h = L_H m + \ell_H\} \tag{38}$$

并获得其应在下一级实施的策略。

S4.2 通过泰勒级数近似确定相容阶

Basar 和 Olsder（Başar 和 Olsder，1998）通过将猜测进行三阶泰勒展开，推导出不同阶数的一致猜测变分均衡。设 (h^c, m^c) 为一致猜测变分均衡，(L_H^c, L_M^c) 为一致猜测策略斜率。设 $\ell_H^c = h^c - L_H^c m^c$和 $\ell_M^c = m^c - L_M^c h^c$。猜测的一阶表示，即仿射猜测

$$h^c \approx L_H^c m + \ell_H^c + \mathcal{O}(m^2),$$
$$m^c \approx L_M^c h + \ell_M^c + \mathcal{O}(h^2)$$

描述驻定性的偏微分方程为

$$\frac{\partial c_H(h,m)}{\partial h}+\frac{\partial c_H(h,m)}{\partial m}\cdot\frac{\partial(L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}})}{\partial h}=0,\ \text{for } h=L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}},$$
$$\frac{\partial c_M(h,m)}{\partial m}+\frac{\partial c_M(h,m)}{\partial h}\cdot\frac{\partial(L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}})}{\partial m}=0,\ \text{for } m=L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}},$$

将 basar 所称的“一级”CCVE 写出后，可得到驻定条件

$$\frac{\partial^2 c_H}{\partial h^2}\cdot\frac{\partial(L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}})}{\partial m}+\frac{\partial^2 c_H}{\partial h\partial m}\left(1+\frac{\partial(L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}})}{\partial m}\cdot\frac{\partial(L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}})}{\partial h}\right)+\frac{\partial^2 c_H}{\partial m^2}\cdot\frac{\partial(L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}})}{\partial h}=0,$$
$$\frac{\partial^2 c_M}{\partial m^2}\cdot\frac{\partial(L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}})}{\partial h}+\frac{\partial^2 c_M}{\partial m\partial h}\left(1+\frac{\partial(L_M^{\mathsf{c}}h+\ell_M^{\mathsf{c}})}{\partial h}\cdot\frac{\partial(L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}})}{\partial m}\right)+\frac{\partial^2 c_M}{\partial h^2}\cdot\frac{\partial(L_H^{\mathsf{c}}m+\ell_H^{\mathsf{c}})}{\partial m}=0,$$

其参数为 $(h,m)=(h^{\mathsf{c}},m^{\mathsf{c}})$。因此

$$A_HL_H^{\mathsf{c}}+B_H(1+L_H^{\mathsf{c}}L_M^{\mathsf{c}})+D_HL_M^{\mathsf{c}}=0,$$
$$A_ML_M^{\mathsf{c}}+B_M(1+L_M^{\mathsf{c}}L_H^{\mathsf{c}})+D_ML_H^{\mathsf{c}}=0,$$

由上述方程求解 $L_H^{\mathsf{c}},L_M^{\mathsf{c}}$，可得

$$L_H^{\mathsf{c}}=-\frac{B_H+L_M^{\mathsf{c}}D_H}{A_H+L_M^{\mathsf{c}}B_M},$$
$$L_M^{\mathsf{c}}=-\frac{B_M+L_H^{\mathsf{c}}D_M}{A_M+L_H^{\mathsf{c}}B_H}$$

这表明 $L_H^{\mathsf{c}},L_M^{\mathsf{c}}$ 是猜想迭代的不动点。

扩展数据部分

补充方法见第 A 节。实验 1、2 和 3 的详细信息见第 A.1 节、第 A.2 节和第 A.3 节。实验 1、2 和 3 中使用的自适应算法的数值模拟见第 B.6 节。第 B 节中的补充实验表明，这些实验具有可推广性；其中对实验参数和成本结构进行了变动。用户研究的任务负荷调查和反馈表见第 C 节。

A 补充方法

补充材料中报告了结果但未在主论文中报告的其他实验，采用了不同的二次型和非二次成本，以展示实验和理论的普适性。首先（第B.1节），使用不同的机器策略初始化重复了实验3：机器策略不是初始化为 $m = h$，而是初始化为 $m = 0$。接着（第B.2节），使用机器的不同全局最优解重复了9次实验3：机器的二次成本重新参数化为

$$c_M(h, m) = \frac{1}{2}(m - m_M^*)^2 - (m - m_M^*)(h - h_M^*) + (h - h_H^*)^2$$

使用 $h_M^* \in \{-0.1, 0, +0.1\}$ 和 $m_M^* \in \{-0.1, 0, +0.1\}$ 来测试机器是否能将行为引导至联合行动空间中有限个点中的任意一个，并测试逆向斯塔克尔伯格均衡（$h^{\mathrm{RSE}}, m^{\mathrm{RSE}}$）$= (h_M^*, m_M^*)$是否为策略梯度的稳定均衡。

随后（第B.3节），使用柯布-道格拉斯 形式的非二次成本函数重复了实验 1、2 和 3（根据（Figuières等人，2004）第 C.2 节中的示例修改而来）：

$$c_H(h, m) = 1 - 2(1 - h)^{0.175}(h + 1.1m)^{0.5} \tag{39}$$

用于实验 1、2 和 3 的重复试验中；

$$c_M(h, m) = 1 - 2(1 - m)^{0.2}(m + 1.1h)^{0.5} \tag{40}$$

用于实验 1 和 2 的重复试验中，并且

$$c_M(h, m) = (m - m_M^*)^2 + (h - h_M^*)^2 \text{ with } (m_M^*, h_M^*) = (0.5, 0.5) \tag{41}$$

用于实验3的重复组。将 c_H 来自(39) 与 c_M 来自(40) 配对后，可得实验1和实验2重复组中的以下博弈论均衡：

$$\begin{aligned}(h^{\mathrm{NE}}, m^{\mathrm{NE}}) &\approx (0.590, 0.529),\\ (h^{\mathrm{SE}}, m^{\mathrm{SE}}) &\approx (0.429, 0.579),\\ (h^{\mathsf{c}}, m^{\mathsf{c}}) &\approx (0.392, 0.336).\end{aligned}$$

将 c_H 来自(39) 与 c_M 来自(41) 结合，得出实验 3 重复试验中的以下均衡：

$$(h^{\mathrm{RSE}}, m^{\mathrm{RSE}}) = (0.5, 0.5).$$

在这些实验复现中，人的行动被限制在 $[0.2, 0.8]$ 的范围内，因此手动输入也相应地归一化到这一范围。机器的行动被限制在 $[0, 1]$。各实验方案设计的具体变更将在后续小节中说明。

A.1 实验 1：行动空间中的梯度下降

方案 S1概述了实验 1 的流程。

第 B.3 节中针对非二次成本的实验对前述方法进行了如下修改：对于情况 $\alpha = \infty$，所实施的策略为 $m = -\frac{77}{270}h + \frac{20}{27}$；联合行动在正方形 $[0.3, 0.7] \times [0.3,0.7] \subset \mathbb{R}^2$ 中均匀随机初始化。

A.2 实验 2：策略空间中的猜测变分

协议 S2概述了实验 2 的流程。

第 B.3 节中针对非二次成本的实验对前述方法进行了如下修改：给定柯布-道格拉斯形式的非二次成本

$$c_M(h, m) = 1 - 2(1 - m)^{a_M}(m + d_M h)^{b_M} \tag{42}$$

w其中 a_M、 $b_M > 0$ 和 $d_M \geq 1$，机器的猜测变分迭代为

$$L_{M,k+1} = -\frac{a_M d_M}{a_M + b_M + b_M d_M L_H}, \tag{43a}$$

$$\ell_{M,k+1} = \frac{b_M + b_M d_M L_H}{a_M + b_M + b_M d_M L_H}. \tag{43b}$$

A.3 实验3：策略空间中的梯度下降

协议S3总结了实验3的程序。

关于策略梯度估计和收敛的理论结果，请参见第 9 和 11 条命题，见第 S3.3 节。

B 额外实验结果

使用不同的二次型和非二次型成本进行了额外实验，以展示实验结果和理论结果的普适性。

B.1 机器初始化（实验3）

为证明机器策略梯度自适应算法的结果不依赖于机器策略的初始化，我们以初始策略斜率为 $L_M = 0$重复了实验3。反复迭代策略梯度后，人类受试者群体的中位数行动向量分布移至机器的全局最优解（图S1）。

B.2 机器最优点（实验3）

为证明只要联合行动配置稳定，机器就能将人类行动驱动到行动空间中的任意点，研究开展了三项具有不同机器极小值的实验。测试了一组机器极小值 $h_M^* \in \{-0.1, 0, +0.1\}$ 和 $m_M^* \in \{-0.1, 0, +0.1\}$。反复进行策略梯度下降，将人类受试者群体的中位数行动向量分布移向机器的全局最优解（图 S2）。

B.3 非二次成本（修改后的实验 1、2 和 3）

为证明实验和理论的普适性，我们采用非二次成本进行了修改后的实验 1、2 和 3。在实验 1 中，人类受试者群体的中位数行动向量分布，以最慢的速率从纳什均衡移向人类主导的斯塔克尔伯格均衡，以最快的适应率达到该均衡（图S3A）。在实验 2 中，通过反复迭代猜测变动的估计过程，人类受试者群体的中位数行动向量分布从人类主导的斯塔克尔伯格均衡移向一致猜测变分均衡（图 S3B）。在实验 3 中，通过反复迭代策略梯度下降，人类受试者群体的中位数行动向量分布移向机器的全局最小值（图 S3C）。

B.4 数值模拟

对这三个实验进行了数值模拟。模拟结果叠加在主论文中小提琴图的数据图上（图 S4）。在实验 1 中，模拟捕捉到了从速率最慢时的纳什均衡到速率最快时的人类主导的斯塔克尔伯格均衡的转变（图 S4A）。在实验 2 中，模拟捕捉到了从人类主导的斯塔克尔伯格均衡到一致猜测变分均衡的转变（图S4B）。在实验 3 中，模拟捕捉到了从人类主导的斯塔克尔伯格均衡到机器的全局最优解的转变（图 S4C）。

B.5 一致性与帕累托最优性

为证明实验中达到的均衡点并非帕累托最优解（机器的全局最小值除外），我们将这些集合与一致猜测条件进行比较（图 S5）。帕累托最优的行动集合满足

$$\min_{h,m} \ \gamma c_H(h,m) \ + (1-\gamma) c_M(h,m) \tag{44}$$

对于 γ 介于 0 和 1 之间。帕累托最优性的定义见（Debreu，1954）。当一位参与者的猜测等于另一位参与者的策略时，满足一致性条件（参见（Başar 和 Olsder，1998）中的定义 4.9）。主论文中实验2和实验3的数据，以及 B.1 节中采用不同初始化的实验3数据，绘于图B.1 见S5。这些数据与人类的猜测和机器策略相一致的曲线重叠。

实验 1、2 和 3 的统计检验结果，包括 P-值、 t-统计量和 Cohen's d.

实验 1				
H_0**: 初始人类行动分布的均值等于**	$h^{\mathbf{NE}}$	$P = 0.20$	$t = +1.3$	$d = +0.2$	
H_0**: 初始机器行动分布的均值等于**	$m^{\mathbf{NE}}$	$P = 1.00$	$t = +0.0$	$d = -1.0$	
H_0: 初始人类行动分布的均值等于	h^{SE}	$P = 0.00$	$t = -26.9$	$d = -4.2$	$\star$
H_0: 初始机器行动分布的均值等于	m^{SE}	$P = 0.00$	$t = -\infty$	$d = -\infty$	$\star$
H_0: 最终人类行动分布的均值等于	h^{NE}	$P = 0.00$	$t = +21.2$	$d = +3.4$	$\star$
H_0: 最终机器行动分布的均值等于	m^{NE}	$P = 0.00$	$t = +21.2$	$d = +3.4$	$\star$
H_0**: 最终人类行动分布的均值等于**	$h^{\mathbf{SE}}$	$P = 0.49$	$t = -0.7$	$d = -0.1$	
H_0**: 最终机器行动分布的均值等于**	$m^{\mathbf{SE}}$	$P = 0.49$	$t = -0.7$	$d = -0.1$	
实验 2					
H_0**: 初始人类行动分布的均值等于**	$h^{\mathbf{SE}}$	$P = 0.24$	$t = -1.2$	$d = -0.3$	
H_0**: 初始机器行动分布的均值等于**	$m^{\mathbf{SE}}$	$P = 0.24$	$t = -1.2$	$d = -0.3$	
H_0**: 初始人类策略分布的均值等于**	$L_H^{\mathbf{SE}}$	$P = 0.10$	$t = +1.7$	$d = +0.4$	
H_0**: 初始机器策略分布的均值等于**	$L_M^{\mathbf{SE}}$	$P = 1.00$	$t = +0.0$	$d =$ NaN	
H_0: 初始人类行动分布的均值等于	h^{CCVE}	$P = 0.00$	$t = -10.0$	$d = -2.3$	$\star$
H_0: 初始机器行动分布的均值等于	m^{CCVE}	$P = 0.00$	$t = -21.3$	$d = -4.9$	$\star$
H_0: 初始人类策略分布的均值等于	L_H^{CCVE}	$P = 0.00$	$t = +12.1$	$d = +2.8$	$\star$
H_0: 初始机器策略分布的均值等于	L_M^{CCVE}	$P = 0.00$	$t = -\infty$	$d =$ 非数字	$\star$
H_0: 最终人类行动分布的均值等于	h^{SE}	$P = 0.00$	$t = +4.9$	$d = +1.1$	$\star$
H_0: 最终机器行动分布的均值等于	m^{SE}	$P = 0.00$	$t = +7.6$	$d = +1.7$	$\star$
H_0: 最终人类策略分布的均值等于	L_H^{SE}	$P = 0.00$	$t = -6.4$	$d = -1.5$	$\star$
H_0: 最终机器策略分布的均值等于	L_M^{SE}	$P = 0.00$	$t = +13.0$	$d = +3.0$	$\star$
H_0**: 最终人类行动分布的均值等于**	$h^{\mathbf{CCVE}}$	$P = 0.02$	$t = -2.6$	$d = -0.6$	$\star$
H_0**: 最终机器行动分布的均值等于**	$m^{\mathbf{CCVE}}$	$P = 0.02$	$t = -2.5$	$d = -0.6$	$\star$
H_0**: 最终人类策略分布的均值等于**	$L_H^{\mathbf{CCVE}}$	$P = 0.31$	$t = +1.0$	$d = +0.2$	
H_0**: 最终机器策略分布的均值等于**	$L_M^{\mathbf{CCVE}}$	$P = 0.13$	$t = -1.6$	$d = -0.4$	
实验3					
H_0**: 初始人类行动分布的均值等于**	$h^{\mathbf{SE}}$	$P = 0.27$	$t = -1.2$	$d = -0.4$	
H_0**: 初始机器行动分布的均值等于**	$m^{\mathbf{SE}}$	$P = 0.33$	$t = -1.0$	$d = -0.3$	
H_0**: 初始人类策略分布的均值等于**	$L_H^{\mathbf{SE}}$	$P = 1.00$	$t = +0.0$	$d = +1.0$	
H_0**: 初始机器策略分布的均值等于**	$L_M^{\mathbf{SE}}$	$P = 1.00$	$t = +0.0$	$d =$ NaN	
H_0**: 初始机器成本分布的均值等于**	$c_M^{\mathbf{SE}}$	$P = 0.74$	$t = -0.3$	$d = -0.1$	
H_0: 初始人类行动分布的均值等于	h^{RSE}	$P = 0.00$	$t = +7.9$	$d = +2.6$	$\star$
H_0: 初始机器行动分布的均值等于	m^{RSE}	$P = 0.00$	$t = +8.4$	$d = +2.8$	$\star$
H_0: 初始人类策略分布的均值等于	L_H^{RSE}	$P = 0.00$	$t = -\infty$	$d = -\infty$	$\star$
H_0: 初始机器策略分布的均值等于	L_M^{RSE}	$P = 0.00$	$t = +\infty$	$d =$ NaN	$\star$
H_0: 初始机器成本分布的均值等于	c_M^{RSE}	$P = 0.00$	$t = +7.7$	$d = +2.6$	$\star$
H_0: 最终人类行动分布的均值等于	h^{SE}	$P = 0.00$	$t = -7.5$	$d = -2.5$	$\star$
H_0: 最终机器行动分布的均值等于	m^{SE}	$P = 0.00$	$t = -11.9$	$d = -4.0$	$\star$
H_0: 最终人类策略分布的均值等于	L_H^{SE}	$P = 0.00$	$t = +22.9$	$d = +7.6$	$\star$
H_0: 最终机器策略分布的均值等于	L_M^{SE}	$P = 0.00$	$t = -19.4$	$d = -6.5$	$\star$
H_0: 最终机器成本分布的均值等于	c_M^{SE}	$P = 0.00$	$t = -6.3$	$d = -2.1$	$\star$
H_0**: 最终人类行动分布的均值等于**	$h^{\mathbf{RSE}}$	$P = 0.07$	$t = +2.1$	$d = +0.7$	
H_0**: 最终机器行动分布的均值等于**	$m^{\mathbf{RSE}}$	$P = 0.06$	$t = +2.1$	$d = +0.7$	
H_0**: 最终人类策略分布的均值等于**	$L_H^{\mathbf{RSE}}$	$P = 0.01$	$t = -3.1$	$d = -1.0$	$\star$
H_0**: 最终机器策略分布的均值等于**	$L_M^{\mathbf{RSE}}$	$P = 0.01$	$t = +3.3$	$d = +1.1$	$\star$
H_0**: 最终机器成本分布的均值等于**	$c_M^{\mathbf{RSE}}$	$P = 0.07$	$t = +1.7$	$d = +0.6$	

表 S1：实验 1、2 和 3 中用于 t 检验的零假设和统计量精确值（P 值、 t 统计量和 Cohen 的 d 效应量）。所有检验的自由度均为 19。统计显著性（$*$）通过将 P 值与 0.05 的置信阈值进行比较来确定。针对行动和策略的检验为双侧检验，针对成本的检验为单侧检验。粗体行是博弈论分析预测的结果。

```
repeat:
  pick adaptation rate α and sign s randomly
  initialize actions h_0, m_0 randomly
  for t in {1,...,T}:
    h_t = s*get_manual_input(t)
    display_cost(c_H(h_t, m_t))
    if α = 0:
      m_{t+1} = m^NE
    else if 0 < α < ∞:
      m_{t+1} = m_t − α∂_m c_M(h_t, m_t)
    else if α = ∞:
      m_{t+1} = L_{M,0} h_t + ℓ_{M,0}
```

协议 S1： 实验 1 的算法描述。

```
function运行_试次(L_M, ℓ_M):
  initialize h_0 randomly
  for t in{1,...,T}:
    h_t =get_manual_input(t)
    m_t = L_M h_t + ℓ_M
    display_cost(c_H(h_t, m_t))
  return median of h_t and m_t
```

```
initialize L_{M,0} and ℓ_{M,0}
for k in{0,...,K − 1}:
    ~(h, m) ←运行_试次(L_{M,k}, ℓ_{M,k}):
    ~(h′, m′) ←运行_试次(L_{M,k}, ℓ_{M,k} + δ):
  L̃_{H, k+1} =( h̃′ − h̃)/(m̃′ − m̃)
  L_{M, k+1} = −(B_M + L̃_{H,k+1} D_M)/(A_M + L̃_{H,k+1} B_M)
  ℓ_{M,k+1} = −(b_M + <escape>L_{H,k+1} d_M)/(A_M + L̃_{H,k+1} B_M)
</escape>
end experiment
```

方案 S2：实验 2 的算法描述。

```
function运行_试次(L_M, h*_M, m*_M):
  initialize h_0 randomly
  for t in{1,...,T}:
    h_t =get_manual_input(t)
    m_t = L_M(h_t − h*_M) + m*_M
    display_cost(c_H(h_t, m_t))
  return median of c_M(h_t, m_t)
```

```
initialize L_{M,0} and(m*_M, h*_M)
for k in{0,...,K − 1}:
  c̃_M ←运行_试次(L_{M,k}, h*_M, m*_M)
  c̃_M′ ←运行_试次(L_{M,k} + Δ, h*_M, m*_M)
  grad_M =( c̃_M′ − c̃_M)/Δ
  L_{M,k+1} = L_{M,k} − γ*grad_M
end experiment
```

协议 S3：实验3的算法描述。

人类 H	机器 M	
$\boldsymbol{H} = [-1,1] \subset \mathbb{R}$ $h \in \boldsymbol{H}$ c_H： 人类 ×机器 $\to \mathbb{R}$	$\boldsymbol{M} = \mathbb{R}$ $m \in$ 机器 c_M： 人类 ×机器 $\to \mathbb{R}$	玩家行动空间 玩家行动 玩家成本

表 S2：人类与机器协同适应博弈的符号和术语。

符号	描述
$T > 0$	时间范围
$t \in \{0, 1, \ldots, T\}$	时间（离散步）
$h_t \in \boldsymbol{H} = [-1,1]$	H在时间 t 的行动
$m_t \in \boldsymbol{M} = \mathbb{R}$	M在时间 t 的行动
$c_H(h_t, m_t) \in \mathbb{R}$	H在时间 t 的成本
$c_M(h_t, m_t) \in \mathbb{R}$	M 在时间 t 的成本
实验 1：	
$\alpha \in [0, \infty]$	M 的适应率
$\partial_m c_M(h, m) \in \mathbb{R}$	M 的成本关于 m 的导数
$L_{M,0}(\cdot) + \ell_{M,0} \in \mathbb{R} \to \mathbb{R}$	M的纳什策略
$(h^{\mathsf{NE}}, m^{\mathsf{NE}}) \in \mathcal{H} \times \mathcal{M}$	纳什均衡
$(h^{\mathsf{SE}}, m^{\mathsf{SE}}) \in \mathcal{H} \times \mathcal{M}$	人类主导的斯塔克尔伯格均衡
实验 2：	
$k \in \{0, \ldots, K\}$	猜测变分迭代
$\delta \in \mathbb{R}$	M策略的常数项扰动
$\widetilde{L}_{M,k} \in \mathbb{R}$	在迭代 k时，M对 H策略斜率的估计
$L_{M,k}(\cdot) + \ell_{M,k} \in \mathbb{R} \to \mathbb{R}$	M在迭代 k时的策略
$(h^{\mathsf{CCVE}}, m^{\mathsf{CCVE}}) \in \mathcal{H} \times \mathcal{M}$	一致猜测变分均衡
实验3：	
$k \in \{0, \ldots, K\}$	策略梯度迭代
$\Delta \in \mathbb{R}$	对 M策略斜率项的扰动
$\partial_{L_M} \widetilde{c}_M(L_M) \in \mathbb{R}$	M的策略梯度估计
$L_{M,k}(\cdot) + \ell_{M,k} \in \mathbb{R} \to \mathbb{R}$	M在第 k次迭代时的策略
$(h^{\mathsf{RSE}}, m^{\mathsf{RSE}}) \in \mathcal{H} \times \mathcal{M}$	机器主导的反向斯塔克尔伯格均衡
$(h_M^*, m_M^*) \in \mathcal{H} \times \mathcal{M}$	M 的全局最小值

表 S3：三项实验中所用博弈的符号和术语。

B.6 数值模拟

为了针对三个实验中各自观察到的结果提供简单的描述性模型，我们使用 Python 3.8 实施了数值模拟（Van Rossum 和 Drake，2009）。共享的参数、成本和梯度定义包含在源代码 S0 中。

实验 1 为了预测在两种极限情况之间的适应速率范围内（即对于 $0 < \alpha < \infty$）会发生什么，我们基于近似梯度下降实现了人类行为模拟。该人类模型仅使用有限差分来估计其成本相对于其行动的导数 (c_H)（h），然后调整其行动以沿成本梯度下降。重要的是，假设人类执行导数估计和梯度下降过程的速度比机器慢，即机器每执行 K 步，人类才执行一次梯度更新。由于机器的步进速率为每秒 60 个样本，这种时间尺度差异意味着人类的步进速率为每秒 $60/K$ 个样本。用于模拟实验 1 的 Python 代码收录于源代码 S1 中。

实验 2 为了预测当机器扰动其策略的常数项，并利用结果估计人类策略斜率时会发生什么，我们基于猜测变分迭代对双方的行为进行了模拟。机器对人类的策略作出最佳响应。人类模型使用其成本的导数 (c_H)，并假设机器的行动 (m) 与其自身行动 (h) 之间存在猜测变动关系 ($L_{M,k}$)，然后调整自身行动以沿成本梯度下降。假设机器会观测人类~和机器的行动，以计算对人类策略斜率的估计值 ($L_{H,k}$)。用于模拟实验 2 的 Python 代码包含在源代码 S2 中。

实验3 为了预测机器扰动其策略线性项时会发生什么，我们基于策略梯度实现了一项模拟。人类模型与实验 2 的前一项模拟相同。机器使用观测到的成本的梯度估计，无需像前一项实验那样观测人类的行动或策略。用于模拟实验3的 Python 代码见源代码 S3。

```
T = 10000   # time samples

# human's cost parameters
AH, BH, DH, hH, mH = 1, -1/3, 7/15, 1/10, 7/10

# machine's cost parameters
AM, BM, DM, hM, mM = 1, -1, 2, 0, 0

def cost_H(h, m):      # H's cost
    return AH*(h-hH)**2/2 + (h-hH)*BH*(m-mH) + DH*(m-mH)**2/2

def cost_M(h, m):      # M's cost
    return AM*(m-mM)**2/2 + (h-hM)*BM*(m-mM) + DM*(h-hM)**2/2

def grad_H(h, m, LM): # H's gradient
    return AH*(h-hH) + BH*(m-mH) + LM*(BH*(h-hH) + DH*(m-mH))

def grad_M(h, m, LH): # M's gradient
    return AM*(m-mM) + BM*(h-hM) + LH*(BH*(h-hH) + DH*(m-mH))

def ceil(x):
    return int(x) if int(x)==x else int(x+1)
```

源代码 S0：两位参与者的参数、成本函数和梯度的定义。

```
# machine's adaptation rates
alphas = [3*10**(i/10) for i in range(-29,-9)]
beta = 0.003 # human's adaptation rate (assumed)
delta = 1e-5 # perturbation size of constant term of H's policy

results = []
for alpha in alphas:
    K = ceil(alpha/beta)          # ratio of M iterations to H iterations
    N = ceil(T/K)*K+1             # number of total iterations
    h,m = [0]*N, [0]*N            # initialize actions

    for t in range(0, T, K):      # gradient descent loop
        c_H = []                  # H's observed cost

        for d in [delta, 0]:

            for k in range(t, t+K):
                # perturb H's action
                h[k]   = h[t] + d
                # update M's action
                m[k+1] = m[k] - alpha*grad_M(h[k],m[k],0)
            c_H.append(cost_H(h[k],m[k]))

        gradH = (c_H[0]-c_H[1])/2/delta # estimate H's gradient

        h[t+K] = h[t] - K*beta*gradH    # update H's action
        m[t+K] = m[k+1]
    results.append([h[-1],m[-1]])
```

源代码 S1：实验 1 的数值模拟。

```
K = 10        # total conjectural variations iterations
delta = 1e-1 # perturbation size (of constant term of M's policy)

h,m = [0]*(K*T+1), [0]*(K*T+1) # initialize actions
LH, LM = [0]*(K+1),  [0]*(K+1)    # initialize policy slopes
LM[0] = -BM/AM                 # initialize M's policy

# conjectural variations iteration loop
for k in range(K):
    h_, m_ = [], []        # steady state actions

    for d in [delta,0]:   # run a pair of trials

        for t in range(k*T, k*T + T):
            # update H's action
            h[t+1] = h[t] - beta*grad_H(h[t], m[t],  LM[k])
            # update M's action
            m[t+1] = LM[k]*(h[t]-hM) + mM + d

        h_.append(h[t+1])
        m_.append(m[t+1])

    # estimate H's policy slope
    LH[k+1] = (h_[1] - h_[0])/(m_[1] - m_[0])

    # update M's policy slope
    LM[k+1] = -(BM + LH[k+1]*DM)/(AM + LH[k+1]*BM)
```

源代码 S2：实验 2 的数值模拟。

```
K = 10        # total policy gradient iterations
Delta = 1e-1 # perturbation size (of slope term of M's policy)
beta = 3e-3  # human's learning rate
gamma = 2     # policy gradient step size

# initialize actions and policies
h,m = [0]*(K*T+1), [0]*(K*T+1) # initialize actions
LH, LM = [0]*(K+1),  [0]*(K+1)    # initialize policy slopes
LM[0] = -BM/AM

# policy gradient loop
for k in range(K):
    c_M = []                      # M's steady state cost

    for D in [Delta, 0]:     # run pair of trials

        for t in range(k*T, k*T+T):
            # update H's action
            h[t+1] = h[t] - beta*grad_H(h[t], m[t],  LM[k] + D)
            # update M's action
            m[t+1] = (LM[k] + D)*(h[t] - hM) + mM

        c_M.append(cost_H(h[t],m[t]))

    # estimate M's policy gradient
    gradM = (c_M[0] - c_M[1])/Delta/2

    # update M's policy slope
    LM[k+1] = LM[k] - gamma*gradM
```

源代码 S3：实验3的数值模拟。

C 任务负荷调查和反馈表

每位参与者在完成实验后都填写了任务负荷调查和可选的反馈表。

C.1 任务负荷调查

NASA任务负荷指数（Hart 和 Staveland，1988）用于评估参与者在执行任务时的心理、身体和时间需求。问题如下：
1. 心理需求：这项任务对脑力的要求有多高？非常低（-10）–非常高（10）**2. 身体需求：**这项任务对体力的要求有多高？非常低（-10）–非常高（10）**3. 时间需求：**这项任务的节奏有多匆忙或紧迫？非常低（-10）–非常高（10）**4. 绩效：**您完成所要求任务的效果如何？完美（-10）–失败（10）**5. 努力程度：**为了达到您的绩效水平，您必须付出多大努力？非常低（-10）–非常高（10）**6. 挫折感：**您感到不安、气馁、恼火、压力大和烦躁的程度如何？非常低（-10）–非常高（10）

表 S4 提供了所有参与者的问卷数据。

	25%四分位数	中位数	75%四分位数
心理需求	-8	-5	0
身体需求	-9	-6	-2
时间需求	-8	-5	-1
绩效	-9	-6	-2
努力程度	-6	-2	3
挫折感	-9	-4	2

表S4：在两种博弈成本条件下，对三项实验进行的任务负荷调查结果。每项实验有20名参与者，共计120名参与者。

C.2 可选反馈

参与者还可以选择提供补充反馈。**有任何反馈吗？请在此告知我们：** [文本框]表 S5 提供了参与者提交的反馈。

实验	反馈
实验 1（quadratic）	没有，继续加油，也谢谢你给我这个机会 :)
实验 1（quadratic）	很酷的测试
实验 1（quadratic）	我觉得这项研究与我参加过的其他研究非常不同 在 Prolific 上。也更具挑战性。
实验 1（quadratic）	一切都好！！
实验 1（二次型）	只要光标保持不动，“保持这个小任务”的设计就很容易被滥用。
实验 1（二次型）	一切都进行得非常顺利，感谢邀请我！
实验 2（二次型）	No
实验 2（二次型）	这个实验很有意思，但当填充区块的选项 在我来得及相应操作之前就切换得太快，这让我有点沮丧。
实验 2（二次型）	N/A
实验 2（二次型）	我觉得这项任务很简单。
实验 2（二次型）	这是个有趣的任务！谢谢
实验3（二次型）	这是项有趣的研究，我很愿意再次参与
实验3（二次型）	NA
实验3（二次型）	我喜欢这项任务
实验3（二次型）	问卷很简单，只需要专注力。
实验3（二次型）	这项任务耗时太长
实验 1（非二次）	我觉得人眼是
实验 1（非二次）	这项研究还不错，但有点慢。
实验 1（非二次）	如果讲解能更详细一些，并且能够.....就更好了 当方框缩到尽可能小时，能够舔一下，谢谢。 再次进行该研究
实验 1（非二次）	这让我感到焦虑，但这是件好事。
实验 2（非二次）	或许有一些说明会更好。
实验 2（非二次）	我不理解这项研究的目的，但玩玩总是挺好的。
实验 2（非二次）	无反馈
实验 2（非二次）	一切都很完美。
实验 2（非二次）	NA
实验 2（非二次）	不太明白为什么在这 20 道练习中，下一项任务要等待这么久 不过体验不错
实验 2（非二次）	一开始我没注意到休息时间是计时的，导致我有几次没能完成 任务。
实验3（非二次）	要么说明不够清楚，要么任务之间的间隔时间实在太 长了。除非这是研究的一部分..... :O

表 S5：参与者提供的书面反馈。可选填写。

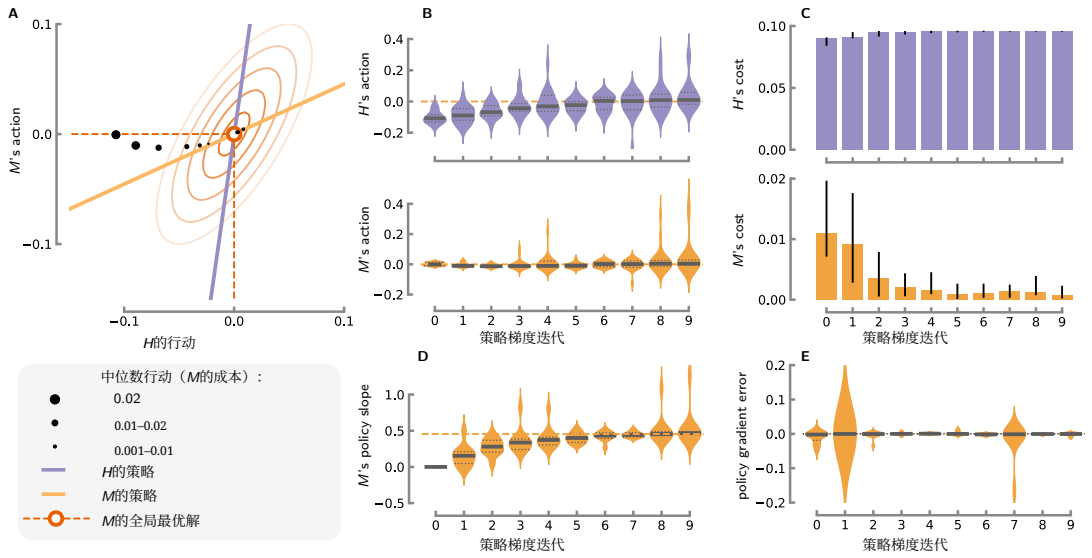

图 S1：**初始策略不同的实验3** ($n = 20$)：针对不同的初始机器策略，在策略空间中进行梯度下降。（**A**）博弈论均衡和最佳响应函数。（**B**）决策向量分布。（**C**）成本分布。（**D**）机器策略斜率。（**E**）机器策略梯度的估计误差。（B）中的行动四分位距在第4至第9次迭代时包含机器的最小值。（D）中的机器策略斜率分布四分位距达到理论预测的斜率，该斜率会使机器的最小值成为博弈结果。（E）中的机器策略梯度四分位距在每次策略梯度迭代中都包含理论梯度。

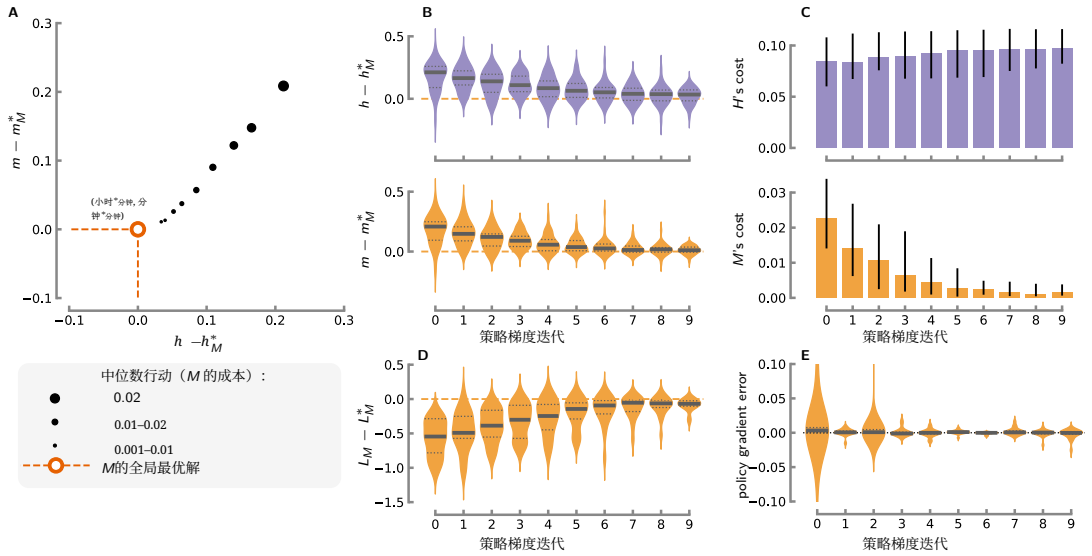

图 S2：**机器最优解不同的实验3** ($n = 18$)：针对不同的机器最优解，在策略空间中进行梯度下降。（**A**）博弈论均衡和最佳响应函数。（**B**）决策向量分布。（**C**）成本分布。（**D**）机器策略斜率。（**E**）机器策略梯度的估计误差。（B）中的行动四分位距在第7至第9次迭代时包含机器的最小值。（D）中的机器策略斜率分布四分位距趋近于理论预测的斜率，该斜率会使机器的最小值成为博弈结果。

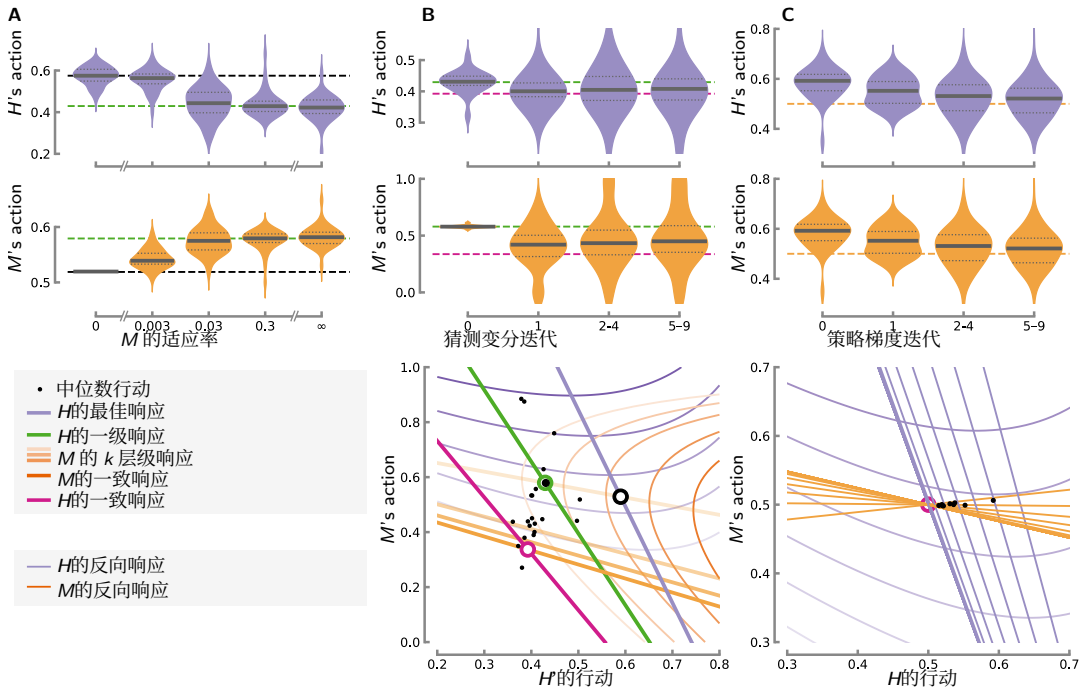

图 S3: **采用非二次成本的修改后的实验 1、2 和 3** ($n = 20 \times 3$)**:** 非二次成本。**(A)** 行动空间中的梯度下降；决策向量分布。**(B)** 策略空间中的猜测变分；决策向量分布、博弈论均衡和最佳响应函数。**(C)** 策略空间中的梯度下降；决策向量分布和策略梯度迭代。

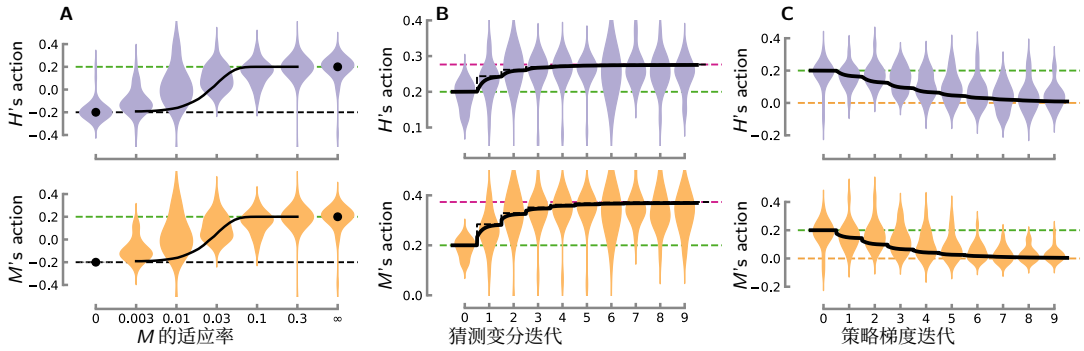

图 S4: **实验 1、2 和 3 的仿真**: 实线和圆点表示模拟数据，叠加在主论文的小提琴图上。虚线表示解析预测。（**A**）行动空间中的梯度下降。（**B**）策略空间中的猜测变分。（**C**）策略空间中的梯度下降。

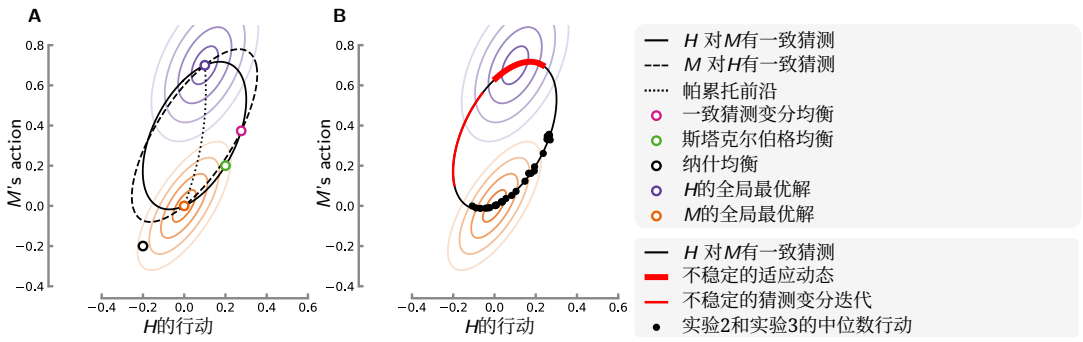

图 S5：**比较帕累托最优性与猜想一致性**（A）人类对机器、机器对人类均持有一致猜测时的均衡连续统解析解，与帕累托最优点的比较。**(**B）实验2和实验3中的中位数行动与对应于人类对机器持有一致猜测的椭圆重合。