arXiv:2603.20617v1 [econ.TH] 21 Mar 2026

人工智能裁员陷阱

Brett Hemenway Falk、Gerry Tsoukalas*

2026年3月2日

摘要

如果人工智能替代人类劳动者的速度快于经济重新吸纳他们的速度，就有可能侵蚀企业赖以生存的消费者需求。我们指出，仅仅意识到这一点，并不足以让企业停止替代。在竞争性任务模型中，需求外部性会使理性的企业陷入自动化军备竞赛，导致劳动者被替代的程度远超集体最优水平。由此造成的损失既伤害劳动者，也损害企业所有者。竞争加剧和"更好"的人工智能会放大这种过度替代；工资调整和自由进入也无法消除它。资本所得税、劳动者股权参与、全民基本收入、技能提升或科斯式谈判同样无济于事。只有庇古式自动化税能够解决这一问题。研究结果表明，政策不仅应应对人工智能导致的劳动力替代所带来的后果，也应着力解决推动这一趋势的竞争性激励。

关键词： 人工智能、自动化、劳动力替代、庇古税。

1 引言

技术会取代工人的担忧，至少可以追溯到工业革命时期（Ricardo，1821年；Keynes，1930年；Leontief，1982年）。从历史上看，替代在很大程度上会自行得到纠正：现有任务的自动化，会被新任务和新职业的创造所抵消。Acemoglu 和 Restrepo（2018年，2019年）所称的复位效应，一直有助于稳定劳动力市场。在人工智能时代，这种平衡能否维持仍是一个悬而未决的问题：Autor 等人（2024年）发现，过去四十年来替代现象日益加剧，而新工作的创造并非总能跟上；早期迹象表明，当前这波变化对入门级劳动者的影响尤为显著（Brynjolfsson 等人，2025a）。

即使最终实现了复职，过程中仍会出现一个问题：被替代工人也是消费者，如果他们失去的收入得不到补偿，每一轮裁员都会削弱所有企业赖以生存的购买力。极端情况下，这会走向自我毁灭：企业不断自动化

*Hemenway Falk：宾夕法尼亚大学（fbrett@cis.upenn.edu），Tsoukalas：波士顿大学（ger-ryt@bu.edu）

，直至生产率无止境地提高、需求降至零。公众讨论越来越倾向于将这一动态视为一种不可避免、没有自然刹车机制的过程（Shah，2026年）。但理性且有远见的企业应该成为刹车；如果所有人都看得见前方的悬崖，它们为什么还要争先恐后地冲过去?

然而，证据表明，企业正朝着这一方向发展。2026年2月，Block裁减了其1万名劳动力中近一半人员；首席执行官杰克·多西表示，人工智能已使其中许多岗位变得不再必要，并称“在未来一年内，大多数公司都将得出同样的结论”（CNBC，2026b）。仅在2025年，就有超过10万名科技行业工人被裁员；在超过半数案例中，人工智能被列为主要驱动因素，裁员主要集中于客户支持、运营和中层管理岗位（CNBC，2025b）。[1] 这种影响并不局限于科技行业：Eloundou等人（2024年）估计，约80%的美国工人所从事的工作包含易受大型语言模型自动化影响的任务。这一切并非不为人知。在此背景下，我们探讨：在何种条件下，理性和完全预见足以防止竞争性过度自动化；当它们不足以做到这一点时，扭曲的规模由何决定；以及所提出的哪些政策应对措施能够加以纠正。

为回答这些问题，我们构建了一个受Acemoglu 和 Restrepo （2018年）启发的、基于任务的自动化模型，但将关注点从劳动力市场转向产品市场：当自动化取代工人时，他们失去的支出会减少每家企业的收入。若干家对称企业中的每一家都会选择用人工智能替代其劳动力的比例。自动化任务的成本更低，但整合摩擦使得每一项后续任务都更难实现自动化。在需求侧，工人将其收入的一部分用于购买该行业的产出；企业所有者的支出较少，在基准情形中被标准化为零。一部分被取代的工资收入会通过再就业或转移支付得到恢复，但其余部分将从该行业中流失。该模型经过有意简化，以清晰展示这一传导渠道，并让所有企业都能看见前方的需求悬崖。基准情形保持工资固定，并关闭资本收入再循环；扩展部分则放松了这两项以及其他若干基准情形假设。尽管简约，该框架仍涵盖了一系列政策工具和稳健性检验。

我们证明，竞争会造成需求外部性，使企业陷入困境：采用自动化的企业可以获得全部成本节约，但在竞争性定价下，只需承担由此造成的总需求萎缩的一部分，其余部分则由竞争对手承担。每家企业的利润最大化自动化率都是严格占优策略，并且高于合作效率水平，因此，仅凭预见性无法阻止企业竞相走向悬崖。竞争越激烈，这种扭曲就越严重：垄断者

[1]个别案例揭示了这一规模：Salesforce 用智能体人工智能取代了 4,000 名客户支持代理人（CNBC，2025c）；Cognition 的 Devin 已部署于高盛和印孚瑟斯，使一名资深工程师能够完成一个五人团队的工作（CNBC，2025a；印孚瑟斯，2026年）。

会完全内化这种外部性，而碎片化市场中的差距最大。在无摩擦极限下，由于每项任务都同样容易实现自动化，博弈会演变为囚徒困境：每家企业都会用人工智能取代全部人类劳动力，尽管集体克制本可以提高所有企业的利润。由此造成的剩余损失并非从工人向企业所有者的转移，而是会同时损害双方的无谓损失。

由于损失由双方共同承担，一个自然的问题是，政策能否予以纠正。我们根据外部性边际评估了六种政策工具。技能提升和劳动者股权参与可以缩小楔子，却无法将其消除。科斯式谈判也无济于事：由于自动化是占优策略，企业之间任何自愿达成的协议都无法自我执行。资本所得税不会改变均衡自动化率，因为它作用于利润水平，而非外部性所在的单任务边际。全民基本收入也同样无效：它能提高生活水平的底线，却不会改变自动化激励。只有将税额设为每项任务未被内部化的需求损失的庇古式自动化税，才能实现合作最优；税收收入可用于资助再培训，提高收入替代水平，从而逐步缩小外部性，并有可能使税收最终不再必要。

核心结论对于若干推广情形同样稳健。更高的人工智能生产率不会消除楔子，反而会使其扩大：每家企业都认为，自己比竞争对手多实现自动化就能获得市场份额收益；但在对称均衡下，这些收益会相互抵消，只留下额外的扭曲。这种红皇后效应意味着，“更好的”人工智能非但不能缓解外部性，反而会加剧它。内生工资调整是 Acemoglu 和 Restrepo (2018年框架中的一个关键自我纠正渠道。它会提高外部性开始显现的阈值，但外部性一旦显现，就无法消除楔子：工资灵活性改变的是问题 何时产生影响，而非 是否存在。自由进入、资本收入再循环以及更丰富的产品市场结构，同样无法消除这种扭曲。

我们的研究对多个领域的文献有所贡献。我们以自动化的任务型方法为基础（Zeira，1998年；Autor 等人，2003年；Acemoglu 和 Restrepo，2018年，2019年），该方法强调在劳动被替代后恢复劳动需求的抵消力量，尤其是新任务的创造和能够自我纠正的工资渠道。Acemoglu （2025年）在这一框架下评估了人工智能对总体生产率的影响。这些研究关注劳动力市场是否以及如何重新平衡；我们则探讨，当重新平衡缓慢或不完全时，产品市场会发生什么。

越来越多的研究指出，自动化可能过度。与我们的研究设定最接近的是Beraja 和 Zorzi (2025年)，他们指出，在劳动力重新配置过程中，如果被替代工人面临借贷约束，自动化就会缺乏效率。他们的机制通过劳动力市场发挥作用：企业忽视了给受信贷约束工人造成的福利成本。我们的机制则通过

产品市场发挥作用：企业忽视了自身给竞争企业造成的需求损失。即使只有一家企业独立经营，他们所说的低效率也会出现；而我们的机制则需要竞争，在垄断下便会消失。此外，他们的规划者会调整自动化以保护工人福利；而我们的规划者即使对工人赋予零权重，也会减少自动化，因为过度自动化本身会损害企业利润。

其他导致自动化过度的渠道也有一个共同特点：即使单个企业孤立决策，这些因素也会扭曲其决策：技术生态系统可能偏向“不优不劣”的自动化，这类自动化会取代工人，却无法带来显著的生产率提升（Acemoglu 和 Restrepo，2020年）；自动化可能会过度针对高租金任务，耗散工人剩余，而非提高产出（Acemoglu 和 Restrepo，即将出版）；此外，过渡性摩擦（Guerreiro 等人，2022年）和分配问题（Costinot and Werning，2023年）也被用来论证征收矫正性税收的合理性。相比之下，我们所研究的外部性只会在竞争环境下出现；即使自动化能大幅提升生产率、信贷市场完备，且规划者不看重分配问题，这种外部性依然持续存在。

我们研究的需求外部性属于总需求溢出这一类现象，最早由 Rosenstein-Rodan (1943年提出，并由 Murphy 等人 (1989年)加以形式化。在他们的“大推进”模型中，部门间的需求互补性可能导致单独来看无利可图的投资无法实现，尽管同时采用这些投资对整体而言是有利可图的。我们的机制恰好相反： 单独来看有利可图的自动化会损害整体利益，因为每家企业节约成本的同时，也侵蚀了所有企业共同依赖的收入基础。 Cooper and John (1988年)提出了总需求外部性导致协调失灵的经典框架；我们的博弈与其结构相同，但会产生唯一的占优策略均衡，因此这是一个真正的外部性问题，而非可以通过沟通解决的协调失灵。关于自动化与需求的相关研究，包括 Benzell 等人 (2015年)对重叠世代框架下机器人采用的研究，以及 Korinek and Stiglitz (2019年)对人工智能驱动的收入再分配的研究，都没有对企业间的策略互动进行建模，而这正是产生我们所识别外部性的原因。

信息系统文献已证实，人工智能系统能显著提升生产率（Brynjolfsson 等人，2025b；Brynjolfsson and McAfee，2014年），并且越来越多地被部署于定价等策略性角色，在这些场景中，算法可能自发学会合谋（Banchio andMantegazza，2022年；Keppo 等人，2026年）。在采用方面，Li 等人(2025年)发现，受到劳工问题审查的企业会专门投资人工智能自动化，而非其他形式的信息技术；Bastaniand Cachon(2025年)则发现，随着人工智能可靠性提升，激励有效人工监督的成本会高到令人难以承受，从而削弱了制约自动化的一项关键机制。这些文献尚未建模的是，这些分别得到证实的现象如何在企业之间相互作用：每项采用决策单独看来都合乎理性，但从整体上看，它们会侵蚀所有企业赖以生存的消费者需求

。我们提出了这一模型，将信息系统文献所记录的微观层面证据与宏观层面的市场失灵联系起来，而这种失灵是任何单个企业都无法阻止的。

本文其余部分安排如下。第 2 节介绍模型。第 3节推导均衡与过度自动化楔子。第 4节评估政策工具。第 5节将模型扩展至人工智能生产率提升、内生进入、内生工资、资本收入再循环以及更丰富的产品市场互动。第 6 节讨论其影响与局限。

2 模型

基准情形在最简单且足以支持该机制的环境中，剥离出自动化对需求的影响：对称企业、单一行业和外生工资。我们先描述供给侧（成本结构和自动化选择），再讨论需求侧（替代如何反馈至收入），最后说明企业所参与的博弈。第 5节将放宽每项假设。

考虑一个行业，其中有 $N \geq 2$ 家对称企业，编号为 $i = 1, \ldots, N$。稍后我们会发现，将每家企业视为拥有一位所有者很有用，例如有权获得企业经营利润的股权持有人。

按照Acemoglu 和 Restrepo (2018年) 的基于任务的框架，每家企业都拥有 $L > 0$ 个任务岗位。最初，所有任务都由人类工人完成；随后出现一项新的技术冲击，例如智能体人工智能，每家企业都必须决定替换多少劳动力。具体而言，企业 i选择自动化率 $\alpha_i \in [0, 1]$：任务 $z \in [0, \alpha_i]$ 由人工智能以每项任务 c 的成本完成，而任务 $z \in (\alpha_i,\ 1]$ 则由人类劳动者以每项任务 w 的工资完成，其中$0 \leq c \leq w$。由于每项自动化任务都会替代一名工人， α_i 既是自动化率，也是被裁员劳动力所占的比例；我们交替使用这两种表述。基准情形下，工资是外生给定的；第5.3 节将工资设为内生变量。

在Acemoglu 和 Restrepo (2018年) 的 CES 任务聚合函数的完全替代极限下，无论采用哪种方式，每项任务都产出一个单位的产品，因此企业产出为 $Y_i = L$；第 5.1 节放宽这一假设，允许人工智能不仅降低成本，还能增加企业产出。这一标准化处理排除了生产率和质量边际的影响，因此基准情形仅刻画劳动力替代对支出的影响。

我们遵循相关文献的假设，按照比较优势对任务进行排序，使边际任务的整合难度逐步增加；我们通过凸型整合成本来刻画这一点 $\frac{k}{2} L \alpha_i^2$其中 $k \geq 0$，采用标准的二次调整成本设定（Lucas、1967年；Hamermesh

和 Pfann，1996年）。企业 i的总生产成本因此为

$$C_i(\alpha_i) = L\big(\alpha_i c + (1-\alpha_i)w\big) + \tfrac{k}{2}\,L\,\alpha_i^2. \tag{1}$$

将自动化带来的每项任务成本节省定义为 $s := w - c$，成本方程可改写为 $C_i = L(w - s\,\alpha_i) + \frac{k}{2}\,L\,\alpha_i^2$：每项自动化任务可节省 s 劳动力成本，但会产生整合摩擦。

在需求侧，工人的边际消费倾向（边际消费倾向）高于所有者（卡尔多，1956年；Mian等人，2021年）；工人将其收入的一部分 $\lambda \in (0,1]$ 用于购买该行业的商品，从而产生了 Murphy 等人（1989年）所分析的企业间需求联系。相比之下，在基准情形下，所有者不会将其收入用于该行业（第5.4节放宽了这一假设）。这种边际消费倾向不对称意味着，当自动化取代工人时，收入会转向行业边际消费倾向较低的代理人，从而减少该行业的总支出。

当企业 j 将其任务中的α_j 比例实现自动化时，$\alpha_j L$ 名工人被替代。被替代工人的工资收入中，有 $\eta \in [0,1]$的部分通过再就业、转移支付或其他来源得到补偿（Jacobson 等人，1993年）；剩余部分，即每名被替代工人 $(1-\eta)\,w$，则从该行业中消失。

在所有 N 企业中，被替代工人的总数为 $\sum_j \alpha_j L$，因此因替代而损失的工资收入为 $(1-\eta)\,w \sum_j \alpha_j L$。因此，该行业的劳动收入总额为 $wLN - (1-\eta)w\sum_j \alpha_j L$，其中有 λ 的比例用于购买该行业的商品。加上自主需求 $A > 0$（来自行业外部或资本收入），该行业的总支出为

$$D(\boldsymbol{\alpha}) = A + \lambda w L\Big[N - (1-\eta)\textstyle\sum_j \alpha_j\Big]. \tag{2}$$

令 $\bar{\alpha} := \frac{1}{N}\sum_j \alpha_j$表示平均自动化率，则可写为 $D = A + \lambda w L N[1-(1-\eta)\bar{\alpha}]$。将每项自动化任务造成的有效需求损失定义为

$$\ell := \lambda(1-\eta)w, \tag{3}$$

这可简化为 $D = A + \lambda w L N - \ell L N \bar{\alpha}$：需求随平均自动化率线性下降。

企业以统一价格在产品市场上出售其产出，该价格使总供给与需求相等。由于所有企业的产出都相同 $Y_i = L$，总供给为 NL，市场出清价格为 $p = D/(NL)$。每家企业的收入为 收入$_i = p \cdot Y_i = D/N$，将（2）代入后可得

$$\mathrm{Rev}_i = \frac{A}{N} + \lambda w L - \ell L \bar{\alpha}. \tag{4}$$

企业 i的利润为 $\pi_i = \text{收入}_i - C_i$。将（4）和（1）代入：

$$\pi_i = \Pi_0 + L\big(s\,\alpha_i - \ell\bar{\alpha} - \tfrac{k}{2}\alpha_i^2\big)\,, \tag{5}$$

其中，$\Pi_0 := A/N + (\lambda - 1)wL$ 表示没有企业采用自动化时的单个企业利润。写作 $\bar{\alpha} = (\alpha_i + \sum_{j\neq i}\alpha_j)/N$ 以单独表示企业 i自身的行动：

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s - \frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\sum_{j\neq i}\alpha_j\right]. \tag{6}$$

企业进行一次性同时行动博弈，各自选择 α_i 以最大化 π_i；随后，给定自动化配置，产品市场便机械地出清。[2] 解概念为纳什均衡。

定义所有者剩余总额 $\mathcal{K}$ 和工人收入总额 $\mathcal{W}$：

$$\mathcal{K} := \sum_i \pi_i$$
$$\mathcal{W} := wLN[1 - (1-\eta)\bar{\alpha}].$$

我们依据两个基准来衡量过度自动化：其一是使 $\mathcal{K}$最大化的合作最优；其二是使下式最大化的广义社会规划者

$$S(\mu) \;:=\; \mu\,\mathcal{W} + (1-\mu)\,\mathcal{K}$$

其中，工人的权重为 $\mu \in [0,1]$ 。

请注意，环境假设完全透明：每家企业都能直接观察到自动化如何导致工人收入减少和总支出下降。第 3节要回答的问题是：在竞争环境中，仅有这种可见性是否足以促使企业遏制自动化。

[2]另一种做法是采用两阶段博弈：企业首先选择自动化率，然后再围绕价格或数量展开竞争。我们不探讨这种更复杂的第二阶段产品市场竞争，因为相关战略效应已得到充分研究，而且会掩盖我们在此要剖析的新机制：完全透明条件下自动化带来的需求外部性。定性结果很可能对更丰富的产品市场互动依然稳健，但封闭形式的解会变得复杂得多。

3 均衡与过度自动化

本节推导均衡，表明相较于合作最优，企业会过度自动化，并量化由此造成的剩余损失。所有证明汇总于第 A节。

3.1 均衡与过度自动化楔子

为了刻画均衡，考虑企业 i进行自动化的边际激励。回顾（3）可知，$\ell(w) = \lambda(1-\eta)w$ 是每位被替代工人所导致的需求损失，其与工资成正比，因为被替代工人放弃的支出会随其收入同比例变化。（当工资保持不变时，我们简单写作 ℓ；第5.3节将工资内生化为 w。）由（6）可知，企业 i从自动化中获得的边际利润为

$$\frac{\partial \pi_i}{\partial \alpha_i} = L\left(s - \frac{\ell}{N} - k\alpha_i\right).$$

自动化程度的边际增加可节省 s 的劳动力成本，但会产生摩擦成本 $k\alpha_i$，并使企业收入减少 ℓ/N。收入损失为 ℓ/N 而不是 ℓ，因为竞争性定价会在对称企业之间平均分配收入（4）：企业 i的自动化会使总需求减少 ℓL，但其中只有 $\ell L/N$ 由企业 i 自身承担。因此，每家企业都会低估其自动化的社会成本，这表明均衡中存在系统性的过度自动化。以下命题证实了这一点，并量化了这一差距。

命题1（均衡与过度自动化）。在第 2 节定义的模型中，定义自动化阈值

$$N^* := \frac{\ell}{s} = \frac{\lambda(1-\eta)w}{w-c}. \tag{7}$$

如果 $N \le N^*$，没有企业进行自动化（$\alpha^{NE} = 0$）。

如果 $N > N^*$（等价地，$s > \ell/N$）：

(i) 每家企业的严格占优策略为 $\alpha^{NE} = \min\left((s-\ell/N)/k, 1\right)$；*(ii)* 合作最优值为 $\alpha^{CO} = \min(\max(0, (s-\ell)/k), 1)$；*(iii)* 如果 $\ell < s < k + \ell/N$ 那么 α^{NE} 和 α^{CO} 均为内部解，且过度自动化楔子 is

$$\alpha^{NE} - \alpha^{CO} = \frac{\ell(1-1/N)}{k} > 0.$$

这随 N 和 ℓ严格递增，随 k递减。

(iv) 如果 $s \le \ell$，那么 $\alpha^{CO} = 0$，因此楔子为 α^{NE}。所以，如果 $s < k + \ell/N$，那么 $\alpha^{NE} = (s-\ell/N)/k$，楔子为 $(s-\ell/N)/k$。另一方面，如果 $k + \ell/N \le s$，那么 $\alpha^{NE} = 1$，

因此楔子为 *1*。

该命题由上文推导出的私人一阶条件及其合作对应条件直接得到：规划者为所有企业设定统一的合作率时，每个 ℓ 自动化任务所面临的是完整的需求损失，而不是 ℓ/N 每个企业所感知到的需求损失，因此得到 $\alpha^{CO} = (s-\ell)/k$。由于竞争对手的合作率在（6）中仅通过 $-(\ell/N)\sum_{j\neq i}\alpha_j$这一项进入，而该项与 α_i无关，因此均衡率是一项严格占优策略：即使完全预见每个竞争对手的行为，每个企业仍会过度自动化。

情形结构之所以出现，是因为 α^{NE} 和 α^{CO} 都处于 $[0,1]$：根据成本节约 s 与需求损失和摩擦成本参数的比较结果，二者都可能处于不自动化、内部解或完全自动化状态。过度自动化楔子在合作率为零、但个别企业仍认为自动化在私人层面有利时最大；当两个合作率达到同一边界时，该楔子则消失。第（iii）–（iv）列举了相关组合；贯穿始终的经济力量相同。

该楔子随 N 严格递增：竞争性行业越多，自动化差距越大。这与传统直觉背道而驰。传统直觉认为，竞争会约束企业，使其采取符合消费者利益的行为；但在这里，竞争越激烈，每家企业分摊的需求损失越少，从而削弱了企业克制自动化的私人激励。垄断者（$N=1$）会将外部性（$\alpha^{NE}=\alpha^{CO}$）完全内部化；随着$N\to\infty$，该楔子趋近其最大值 ℓ/k。

根据命题 1，只有当 $N > N^* = \ell/s$ 时，企业才会采用自动化：竞争者数量必须足够多，使每家企业承担的需求损失份额 ℓ/N 低于其成本节约 s。随着人工智能成本下降（$c\to 0$），$N^*\to\lambda(1-\eta)\leq 1$：过度自动化区域扩大，几乎涵盖所有 $N\geq 2$ 的市场。对于示例参数（$c/w=0.30$，$\lambda=0.5$，$\eta=0.30$，$N\to\infty$），楔子等于 $\ell/k=\alpha^{CO}$：竞争性市场中的企业采用自动化的比率是合作有效率的两倍。

图1展示了这些比较静态结果。在每个面板中，虚线标出了以下 $N=N^*$ 边界：企业不会在其下方进行自动化；阴影颜色越深，表示楔子越大。占优模式是楔子随 N增大而增大；其他维度上的非单调性反映了 $s=\ell$处的机制转换，在此处，合作最优值从零转为内部解。

当摩擦成本为正时，调整成本会抑制均衡自动化率。下一小节将说明，当摩擦成本消失（$k\to 0$）时，这种抑制作用也随之消失，博弈便简化为囚徒困境：要么完全自动化，要么完全不自动化。

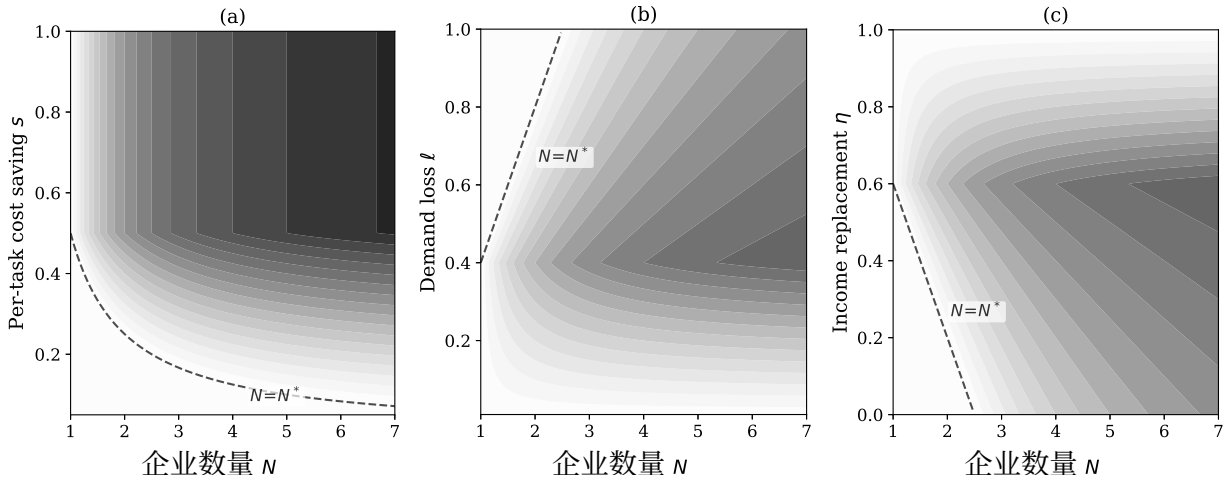


图1：参数空间中的过度自动化楔子 $\alpha^{NE} - \alpha^{CO}$ 。阴影由白色（楔子为零）渐变至黑色（楔子为 ≥ 0.50）。虚线标示无企业进行自动化的边界 $N = N^*$ 。**(a)** 固定工资（$\lambda = 0.5$、 $\eta = 0$、$k = 1$）时，企业数量与成本节约 $s = w - c$ 的关系。**(b)** 企业数量与需求损失 $\ell = \lambda\,(1-\eta)\,w$ （$c = 0.60$、 $\eta = 0$、 $k = 1$）的关系。**(c)** 企业数量与收入替代（$c = 0.60$、 $\lambda = 1$、 $k = 1$）的关系。

3.2 无摩擦自动化作为囚徒困境

当 $k = 0$时，边际利润变为常数 $L(s - \ell/N)$，与自动化水平无关，结果呈现全有或全无。如果 $N \leq N^*$，没有企业会采用自动化。如果 $N > N^*$，采用自动化是严格占优策略，但会损害集体利益：

推论1 (无摩擦极限)。 假设调整摩擦消失（$k = 0$），且企业数量超过临界阈值（$N > N^*$）。

（*i*）完全自动化（$\alpha_i = 1$）对每家企业而言都是严格占优策略。（*ii*）如果成本节约还小于每项任务的需求损失（$s < \ell$），那么合作最优是不自动化（$\alpha_i = 0$，适用于所有 i），此时单个企业利润为 Π_0；均衡则为$\Pi_0 + L\,(s - \ell) < \Pi_0$。总无谓损失为 $NL\,(\ell - s)$。

在第 (ii) （$s < \ell$）所述条件下，囚徒困境的结构清楚地揭示了自愿克制为何会失败。单方面克制的企业（选择 $\alpha_i = 0$）仍会承受竞争对手自动化造成的收入下降，却放弃了可抵消这一下降的成本节约；偏离克制的企业（选择 $\alpha_i = 1$）则能获得成本节约，同时仅让自身承担 $1/N$ 份额的需求损失。由此形成的收益矩阵具有经典形式：相互克制时，每家企业的收益为 Π_0 ；相互自动化时，每家企业的收益为 $\Pi_0 + L\,(s - \ell) < \Pi_0$；然而，无论其他企业如何选择，背离克制在个体层面都是理性的。由于自动化是严格占优策略（而不仅仅是在其他企业选择自动化时的最佳回应），非约束性协议无法恢复效率。在

Crawford 和 Sobel 所说的Crawford 和 Sobel （1982年）意义上，沟通只是廉价交谈：即使所有企业都承认，集体克制会提高利润，每家企业的个体最优行动仍不会改变。这使自动化外部性有别于纯粹的协调失灵（在这类情形中，企业只需就选择哪一种均衡达成一致），也说明了第 4.5 节分析科斯式谈判的动机。

3.3 过度自动化作为无谓损失

过度自动化楔子仅仅是从工人向企业所有者的再分配，还是会降低总剩余？回顾模型部分引入的广义规划者，其目标是最大化

$$S(\mu) \;=\; \mu\,\mathcal{W} + (1-\mu)\,\mathcal{K} \tag{8}$$

对工人赋予权重 $\mu \in [0,1]$ 。

Prop命题 2 （广义规划者与剩余损失）。 假设 $k > 0$ 且 $N > N^*$.

（*i*） μ-规划者的最优自动化率为

$$\alpha^{SP}(\mu) \coloneqq \frac{s-\ell}{k} - \frac{\mu\,\ell}{\lambda(1-\mu)\,k},$$

其中，和往常一样， α^{SP}（μ）表示自动化率，因此被限制在区间 $[0,1]$。当$\mu = 0$ 时，这简化为 α^{CO} ，见命题 *1*。

(ii) 与规划者最优值相比，纳什均衡造成的剩余损失为

$$S\big(\mu;\alpha^{SP}\big) - S\big(\mu;\alpha^{NE}\big) = \frac{(1-\mu)\,NLk}{2}\left[\alpha^{NE} - \alpha^{SP}(\mu)\right]^2.$$

(iii) **(帕累托支配。)** $\alpha^{NE} > \alpha^{SP}(\mu)$ 对每个 $\mu \in [0,1)$均成立。纳什均衡处于合作最优的帕累托劣势：工人和企业所有者双方的境况都严格变差。

过度自动化并非将财富从工人转移给所有者：它会造成无谓损失，使双方都受到损害（第（iii）部分）。工人因被替代而直接损失工资收入。企业所有者虽然降低了每项自动化任务的成本，但也会蒙受损失：集体性的替代侵蚀了需求，以至于每家企业的均衡利润都低于其合作最优利润。无论如何在这两个群体之间进行再分配，都无法使纳什结果变得有效率。

当双方 α^{N} ^{E} 和 α^{SP} 均为内部解时，均衡与规划者的最优值之间的总楔子 s

可分解为两个不同来源：

$$\underbrace{\alpha^{NE} - \alpha^{SP}(\mu)}_{\text{total wedge}} = \underbrace{\frac{\ell(1-1/N)}{k}}_{\text{demand externality}} + \underbrace{\frac{\mu\,\ell}{\lambda(1-\mu)\,k}}_{\text{distributional}}. \tag{9}$$

第一项是命题 1(iii)中的未内化需求外部性：即使规划者对工人的权重为零（$\mu = 0$），只关注总利润时，它依然存在。它随 N 增长，并在 $N \to \infty$ 时趋近于 ℓ/k，因此碎片化市场受到的影响尤为严重。第二项是分配溢价：重视工人收入（$\mu > 0$）的规划者会在利润最大化基准之上进一步降低自动化率。它与 N 无关，但会随着 $\mu \to 1$ 增大而无限增长；当 $\bar{\mu} := \lambda k\,\alpha^{CO}/(\ell + \lambda k\,\alpha^{CO})$ 规划者会完全禁止自动化。(ii) 中的剩余损失与这一总楔子呈二次关系，并随 NL 变化，因此碎片化和市场规模都会加剧福利成本。

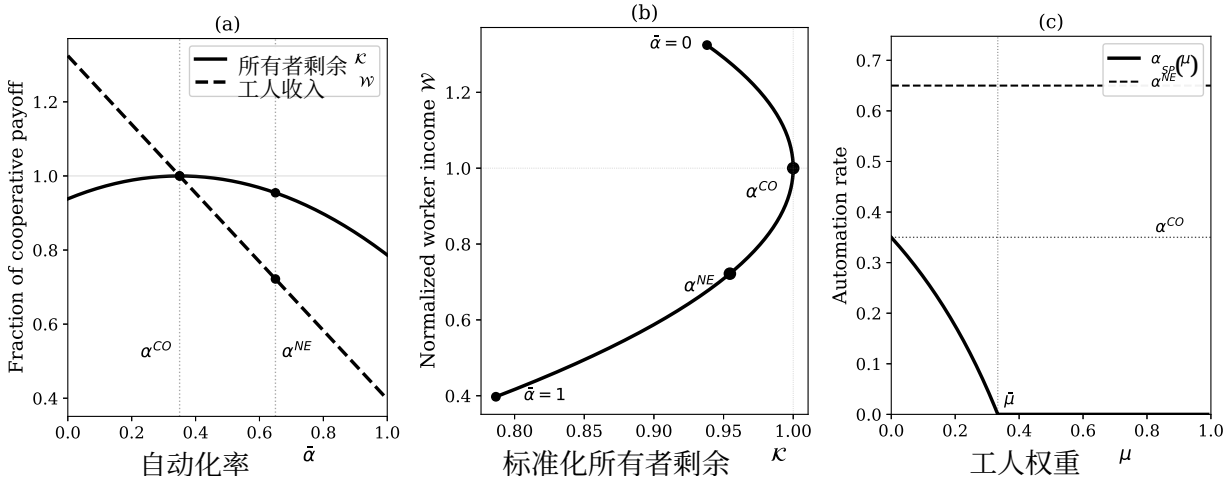


图 2：过度自动化的福利后果。所有面板均采用 $c = 0.30$, $\lambda = 0.5$, $\eta = 0.30$,$N = 7$, $k = 1$, $A = 10$。(a) 标准化所有者剩余 $\mathcal{K}$ 和工人收入 $\mathcal{W}$ 随自动化率变化的情况（$1 =$ 合作基准）：两组人在 α^{NE}时均受损。(b) 要素收益前沿：α^{CO} 位于 (1，1)，α^{NE} 严格位于其西南方（帕累托劣势）。(c) 规划者最优值 $\alpha^{SP}(\mu)$ 随工人权重 μ变化的情况（未标准化）：与 α^{NE} 的差距即为过度自动化楔子，在 $\mu = 0$时也为正。

图 2 展示了帕累托支配关系及其分解。为使两组的损失具有可比性，面板 (a) 和 (b) 分别将各收益除以其在 α^{CO}处的值进行标准化，因此 1 对应于合作基准。面板 (a) 绘制了标准化收益与共同自动化率 $\bar{\alpha}$之间的关系：两条曲线均在 α^{CO} 处或之前达到峰值，并且在 α^{NE}处均低于 1 。关键观察是，均衡率位于总利润峰值的右侧，因此所有者剩余和工人收入均低于合作情形。工人承担了

更大的损失，因为其收入随 $\bar{\alpha}$线性下降，而利润曲线呈凹形，下降得较为缓慢。面板 (b) 将相同信息重新表示为要素收益前沿：曲线上的每一点都对应一个不同的共同自动化率，随着 $\bar{\alpha}$增加，描绘出 $(\mathcal{K}, \mathcal{W})$ 收益对。合作率位于 (1,1) 处，而 α^{NE} 严格位于其西南方，这证实了从均衡率转向合作率将使两组人的境况都得到改善。面板 (c) 将 (9) 中的分解可视化：水平线表示 α^{NE}，下降曲线则是规划者最优值 $\alpha^{SP}(\mu)$。即使在 $\mu = 0$处，差距仍为正（仅需求外部性项就足以造成该差距），并且随着分配溢价随 μ扩大，所需的修正幅度也会进一步增加。

由于过度自动化楔子是一种会损害两类要素的结构性外部性，一个自然的问题是，政策能否消除它。

4 政策工具

原则上，有几种工具可以应对这种外部性；问题在于，哪些工具能作用于正确的边际。为回答这一问题，我们以合作最优 α^{CO}作为基准；该最优值在不直接赋予工人福利权重的情况下最大化总利润。这是对干预最不利的情形：命题 2 表明，仅需求外部性就会同时降低企业利润和工人收入，而对工人赋予任何正权重（$\mu > 0$）只会进一步扩大楔子。因此，采用效率基准既简化了阐述，也为干预提供了最保守的情形。

表 1概述了结果：只有庇古式自动化税能够完全纠正这种扭曲；其余工具只能缓冲受损者的损失，或部分缩小楔子，但没有一种能将其消除。

T表 1：政策工具及其对过度自动化外部性的影响。

	章节	变化 N^*?	变化楔子?	修正外部性?
技能提升/再培训（η）	4.1	Yes	Yes	部分地
全民基本收入（A）	4.2	No	No	No
资本所得税（t）	4.3	No	No	No
劳动者股权（ϵ）	4.4	Yes	Yes	部分
科斯式谈判（$M < N$）	4.5	No	部分	No
自动化税（τ）	4.6	Yes	Yes	**Yes**

注。 每一行记录了 该政策工具是否改变自动化阈值 N^*、过度自动化楔子 $\ell(1 - 1/N)/k$，以及是否完全纠正需求外部性。

需要指出一个局限：分析将每种政策工具都针对单一边际进行评估，即第 3 节所识别的需求外部性，并假定经济的其他所有特征保持不变。在实践中，每种工具在模型之外还伴有额外的成本和收益（行政

负担、劳动力市场扭曲、政治可行性），全面的福利分析需要对此加以权衡。尽管如此，无论某种工具在其他方面表现如何，若其无法作用于外部性边际，就无法纠正这种扭曲；下文将区分能够发挥作用的工具与不能发挥作用的工具。

4.1 替代与技能提升

需求损失参数 $\ell = \lambda(1-\eta)w$ 决定了外部性的大小。在基准模型中，$\eta \in [0,1]$ 表示通过再就业、转移支付或其他来源弥补的被替代工人工资收入比例：η 越高，ℓ越小，从而过度自动化楔子也越小。

但这个参数自然可以扩展到大于 1 的范围。当 $\eta > 1$ 时，技能提升和再吸纳会使被替代工人进入薪酬更高的岗位，自动化会提高总劳动收入，损失 ℓ 转为负值，我们可以将其理解为收益。这正是人工智能乐观派所设想的情形：技术性岗位替代是迈向更好工作的垫脚石。如下文推论所示，ℓ 的符号反转会使外部性本身也发生反转。

推论2（外部性的符号）。对于任意 $\eta \geq 0$ 和 $s > \ell$，过度自动化楔子为（$\alpha^{NE} - \alpha^{CO}$） $= \ell$（$1 - 1/N$）$/k$，并在 $\eta = 0$ 处达到最大值（此时 $\ell = \lambda w$）；当所有 $\eta < 1$时为正，当 $\eta = 1$ 时为零（此时 $\alpha^{NE} = \alpha^{CO} = s/k$），而当 $\eta > 1$时为负（自动化不足）。

逻辑是对称的。当 $\eta < 1$时，替代会摧毁需求，而每家企业只承担损失的 $1/N$，从而导致过度自动化。当 $\eta > 1$时，替代会创造需求，因为再就业工资更高，而每家企业只获得收益的 $1/N$，从而导致自动化不足。在这两种情况下，扭曲程度都随着 N 增大：竞争越激烈，每家企业从外部性中获得的份额就越小，无论这种外部性是负面还是正面。垄断者（$N = 1$）在所有情况下都会完全内化外部性。

竞争力量完全相同；不同的只是需求外部性的符号。正如第4.6 节将要说明的，同一种纠正性工具可以处理这两种情况：当 $\eta < 1$时征税，当 $\eta > 1$时补贴。

这种情况 $\eta > 1$ 并非纯属理论推演。历史上的技术转型往往最终会以更高的工资重新吸纳被替代工人（Acemoglu 和 Restrepo，2019年），而当前的人工智能基础设施建设提供了一条具体途径：数据中心、能源基础设施和人工智能相关服务的扩张，正在创造技能型岗位，其薪酬可能高于自动化所替代的岗位。如果再吸纳的速度足够快，使 η 高于 1，竞争性企业的自动化速度就会过慢。然而，过去的岗位替代事件一再造成 $\eta < 1$：被替代工人

会遭受巨大且持久的收入损失（Jacobson 等人，1993年），目前几乎没有证据表明人工智能驱动的岗位替代会有所不同，因此大多数经济体都稳居过度自动化区间。

政策启示是，通过再培训计划、工资保险和鼓励新企业创立来提高 η，并非只是对被替代工人的暂时纾困，而是直接作用于外部性的杠杆：η 每向 1 提高一个单位，就会缩小 ℓ，收窄过度自动化楔子，并减轻下文所分析的纠正性工具所承受的负担。将 η推高至超过 1 会使扭曲转为自动化不足，但这远没有那么迫切：在这种情况下，被替代工人已经在薪酬更高的岗位上发展得很好。

4.2 全民基本收入

针对自动化驱动的岗位替代，讨论最多的应对措施之一是全民基本收入。在模型中，由一般财政收入资助的全民基本收入对应于自主需求的增加 A：由于这项转移支付不附带条件，就业工人和被替代工人获得的支付相同，从而为总支出增加一个常数，却不改变岗位替代造成的边际收入损失。这使全民基本收入有别于针对岗位替代的转移支付（工资保险、离职补偿金），后者会提高收入替代率 η 并直接缩小 ℓ；参见第4.1节。下文的结果针对的是这一建模对象，不应被视为对所有全民基本收入方案的评判。

由于全民基本收入会在需求中增加一个常数项，因此它只通过 $\Pi_0 = A/N + (\lambda - 1)wL$进入企业利润，即没有任何企业进行自动化时的基准利润。该项会从一阶条件 $s - \ell/N - k\alpha_i = 0$中消失：更高的 A 会提高利润底线，但既不会改变决定自动化率的成本节约 s，也不会改变需求损失 ℓ。因此，全民基本收入既不会改变自动化阈值 $N^* = \ell/s$，也不会改变过度自动化楔子 $\ell(1 - 1/N)/k$。用博弈论的语言来说，全民基本收入改变了收益水平，但不会改变驱动策略行为的收益差异。更一般地说，作用于利润水平的工具可以重新分配收入，却无法纠正外部性；只有改变每项任务自动化边际的工具才能做到这一点。

尽管无法纠正外部性，全民基本收入仍发挥着互补作用。更高的 Π_0 可以缓冲过度自动化造成的利润损失，而这项转移支付本身则提高了工人生活水平的底线，为作用于正确边际的纠正性工具争取时间。

然而，当企业数量是内生的时，UBI 可能带来意想不到的副作用。基准情形固定 N，但如果企业可以自由进入（第5.2 节对此进行了形式化说明），更高的利润会吸引新进入者，从而使市场分散化。由于过度自动化楔子随 N增加，UBI 诱发的进入反而可能扩大外部性，部分抵消基准消费提高带来的福利增益；第5.2 节将进一步阐述完整论证。

在模型中，UBI 是自动化税的补充，而非替代：仅依赖 UBI 的社会仍会以相同的速度过度自动化，生活水平的底线虽更高，但外部性依然不变。

4.3 资本所得税

如果无条件转移支付不会改变自动化激励，一个自然的替代方案是直接对自动化的收益 征税。考虑对资本收入（利润）征收比例税 $t \in (0, 1)$，并将税收收入再分配给工人。企业 i 现在最大化 $(1-t)\pi_i$，但由于 $(1-t)$ 是一个正的标量，它会在一阶条件中抵消：均衡自动化率、阈值 N^*以及过度自动化楔子均保持不变。

从收入端看，情况也没有好转。总额转移支付会提高自主需求 A，但 A只通过常数项 $\Pi_0 = A/N + (\lambda-1)wL$进入单个企业利润，而该项不会出现在一阶条件中。如果税收收入转而用于提供能提高 η的失业保险，外部性会通过 ℓ缩小，但起作用的渠道是 η，而不是利润税本身。

这一区别很重要，因为在政策辩论中，资本所得税经常与机器人税混为一谈。文献中研究的机器人税（例如 Guerreiro 等人，2022年）是针对采用行为征收的单位税，作用于单任务边际；比例资本所得税则是根本不同的工具，它将整个利润函数乘以 $(1-t)$，并在最优条件中抵消。这种政策失效在结构上与全民基本收入（第4.2节）完全相同：两种工具都会改变利润水平，而不是作用于外部性所在的边际。

4.4 工人股权参与

一种源自利润分享文献的市场化替代方案（Weitzman，1985年）可以让工人直接分享自动化所带来的利润。全民基本收入只通过水平项进入需求，而利润分享则通过利润函数发挥作用，因此会与自动化决策相互影响。假设每家 $\epsilon \in [0,1]$ 企业将其部分利润分配给工人（通过员工持股计划、股权授予或共同决策要求）。工人会将这部分收入中的 λ份额花在本行业，从而让利润分享将资本收入重新转化为需求。

总需求现在满足一个不动点条件： $D = A + \lambda[\text{工资收入} + \epsilon \sum_i \pi_i]$，其中 $\sum_i \pi_i = D - \sum_i C_i$。由于利润取决于 D，需求与自动化决策同时确定；证明将明确求解这一不动点。

命题3 (劳动者股权减少但无法消除楔子**)**。 令 $\epsilon \in [0,1]$ 和 $k > 0$，并假设均衡为内部解。定义 $N_\epsilon := N - \lambda\epsilon(N-1)$。

(i) 合作最优保持不变： $\alpha^{CO}(\epsilon) = (s-\ell)/k$，其中，和往常一样， α^{CO} 被限制在 $[0,1]$内。 *(ii)* 纳什均衡自动化率为 $\alpha^{NE}(\epsilon) = (s - \ell/N_\epsilon)/k$ （限制在 $[0,1]$内）。*(iii)* 当 α^{NE} 和 α^{CO} 均为内部解时，过度自动化楔子

$$\alpha^{NE}(\epsilon) - \alpha^{CO} = \frac{\ell(N-1)(1-\lambda\epsilon)}{k\, N_\epsilon}$$

随 ϵ 严格递减，但对所有 $\epsilon < 1/\lambda$均严格为正。该楔子仅在$\epsilon = 1/\lambda$时消失；这要求每当 $\lambda < 1$时， $\epsilon > 1$ 。

第(i)点并非显而易见：利润分享会将资本收入再循环至工人支出，从而改变需求函数，因此人们可能会认为这会改变规划者最优值。这个结论之所以成立，是因为规划者已经控制了所有 N 企业，因而完全内化了需求外部性。在规划者的一阶条件中，利润分享项相互抵消：修正后的需求损失参数 $\ell_\epsilon = \ell - \lambda\epsilon s$ 与需求乘数 $1/(1-\lambda\epsilon)$ 恰好抵消，剩下$k\alpha = s - \ell$ ，且与 ϵ无关。

尽管合作最优不受影响，纳什均衡却会发生变化。直观来说，工人持有股权时，因替代而损失的部分需求会通过利润份额再循环回来；因此，与基准情形相比，每家企业都会认为自身自动化带来的有效需求损失更大，并相应地加以克制。这一变化的幅度由复合参数 $N_\epsilon = N - \lambda\epsilon($ $N-1)$ 决定。该参数衡量有效需求泄漏除数：在 $\epsilon = 0$ 时，它等于 N，由此回归基准情形，即每家企业认为每项自动化任务会造成 ℓ/N 的需求损失。随着 ϵ 上升， N_ϵ 会趋近于 1，从而推动 α^{NE} 趋向合作最优。

尽管有所改善，只要 $\lambda < 1$，再循环就无法完全消除楔子。要消除楔子，乘积 $\lambda\epsilon$ 必须达到 1，也就是 $\epsilon = 1/\lambda$。当 $\lambda < 1$ 时，这一要求超出了可行范围 $\epsilon \in [0,1]$：每向工人再循环一单位利润，只能产生 λ 单位的部门需求，因此，要弥补泄漏就必须分享超过企业全部利润的部分。即使在 $\epsilon = 1$ （全面利润分享）时，楔子也仅降至 $\ell(N-1)(1-\lambda)/[k(N-\lambda(N-1))]$, , 仍严格大于零。（临界情形 $\lambda = 1$ 是例外：全面利润分享确实能消除楔子，但这只是因为这一极端假设消除了导致外部性的支出泄漏。）

结构性限制在于，这种外部性从根本上是多边的：每家企业的自-

动化都会压低所有 N 企业的需求，而一家企业与其自身工人之间的双边安排，无法触及流向竞争对手的需求。

另一个问题是，利润分享是否会自发出现。

推论 3 (不会自发出现自愿利润分享)。 如果每家企业都独立选择自己的利润分享率 $\epsilon_i \in [0,1]$，以最大化留存利润 $(1-\epsilon_i)\pi_i$，那么 $\epsilon_i = 0$ 就是占优策略。

分享的边际成本为 π_i（留存收益等额减少），而边际需求收益仅为 $\lambda\pi_i/N$：工人会将分享利润中的 λ 用于该行业的消费，而企业 i 只能获得由此带来的需求增长中的 $1/N$。由于 $\lambda/N < 1$对任意 $N \geq 2$均成立，成本严格大于收益。这是在自动化外部性之上叠加的二阶协调失灵，体现了第3.2节囚徒困境的结构。

因此，必须强制实施利润分享，才能产生任何效果；即便如此，它也无法替代矫正税：利润分享可以缩小楔子，却无法消除楔子；而且，与矫正税（第 4.6节）不同，它不会产生可用于资助再培训计划的政府收入，而这些计划本可提高 η。

4.5 科斯式谈判

迄今讨论的各种工具都无法完全消除过度自动化楔子，而劳动者股权也不会自愿产生（推论3）。一个自然的问题是，私人治理能否在这些工具未能奏效之处取得成功。根据科斯定理（科斯，1960年），如果外部性的产权界定明确，且交易成本足够低，那么无需政府干预，谈判也能实现合作最优。当代版本的这一论点设想，每位工人都配备一个人工智能代理，代表自己进行谈判，并依据 Arrieta-Ibarra 等人（2018年）和 Posner 和 Weyl（2018年）提出的训练数据产权。要评估这种可能性，不妨将问题分为两个：企业与其员工之间的谈判能否纠正外部性？企业之间的谈判能否做到这一点？如下文所示，两者都不能。

企业与其员工之间的谈判。 如果被替代工人能够就补偿进行谈判，例如按任务支付的遣散费 σ，企业的有效成本节约将从 s 降至$s-\sigma$，均衡自动化率也会下降。但遣散费还会将收入重新转化为需求：被替代工人在该行业将其补偿中的一部分 λ 用于支出，使有效需求损失参数从 ℓ 降至 $\ell - \lambda\sigma$。因此，过度自动化楔子变为 $(1-1/N)(\ell-\lambda\sigma)/k$，虽有所缩小，但仍为正值。在操作上，这等同于将收入替代率 η 提高 σ/w；通过以下方式持有股权份额时，同样的逻辑也适用

利润分享参数 ϵ （第 4.4节分析了这一点）。正如第 4.1 节所示，提高 η 会缩小楔子，但无法将其消除：只要被替代工人的支出未得到完全替代，外部性就会持续存在；这要求 $\lambda\sigma = \ell$，即仅通过谈判实现完全的收入替代。

此外，未被内部化的那部分外部性完全不会由企业 i自身的工人承担。企业 i 进行自动化时，需求损失 $\ell(1 - 1/N)L$ 会降低竞争企业的收入，从而减少其所有者利润（方程式（4））。竞争企业中保住职位的工人仍然能按每项任务赚取 w；他们不会直接受到损害，也没有理由与企业 i进行谈判。因此，这种外部性是通过产品市场传导的企业间渠道，而不是双边谈判能够触及的企业与工人之间的渠道。

企业间谈判。 由于外部性跨企业传导，我们考虑由 $M \leq N$家企业组成的联盟，它们共同选择自动化率以最大化合计利润，而其余$N - M$ 家企业则采取纳什策略。

命题 4 （部分联盟无法消除楔子）。 设 $k > 0$，并假设均衡为内部解。由 M 家企业组成的联盟共同最大化其成员的合计利润，并选择共同自动化率

$$\alpha^M = \frac{s - M\ell/N}{k}.$$

相对于合作最优值的剩余过度自动化楔子为

$$\alpha^M - \alpha^{CO} = \frac{\ell(1 - M/N)}{k} > 0 \quad for\ all\ M < N.$$

楔子仅在 $M = N$时消失：只有大联盟才能实现合作最优值 .

直观地说， M 家企业组成的联盟会将 M/N 部分总需求损失内化，但剩余的$(1 - M/N)$ 部分仍会外部化给非成员。自动化外部性的四个特点阻碍了大联盟的形成。首先，自愿协议无法自我执行：在无摩擦极限下（推论1），自动化是严格占优策略，因此无论其他成员是否遵守协议，联盟成员都能从违约中获益。这并非沟通能够解决的协调失败；囚徒困境的结构意味着，任何不具约束力的安排都不稳定。在凸成本下（$k > 0$），违约激励虽是连续的，但仍然为正。其次，外部性具有多边性且分散广泛。科斯定理的经典应用涉及双边或少数参与者的情形；而在这里，每家 N 企业都会给其余所有 $N - 1$ 家企业造成需求损失。每家企业对需求损失的个体贡献仅为 $\ell L/N$，小到

不足以促成任何单独谈判，却又在总体上大到无法忽视。这正是科斯在(1960年) 本人承认私人谈判会失灵的大规模情形。第三，企业之间无法对自动化率 α_i 订立可执行的合同：它属于企业内部的组织选择，竞争企业无法观察或核实，因此难以达成具有约束力的私人协议。第四，自动化决策涉及大量沉没成本，且基本不可逆，因此即使在重复博弈中，触发策略惩罚也无法撤销违约行为；竞争对手推进自动化时，选择延迟的企业会失去市场份额（见下文第5.1 节）；而较大的 N 会使违约更难察觉，也让惩罚更难维持。

总而言之，本文研究的需求外部性并非私人治理能够纠正的市场失灵。工人方谈判作用于企业内部渠道（η, ϵ），无法触及外部性所在的企业间边际；企业间谈判虽然针对了正确的边际，却无法维持消除这一楔子所需的大联盟。根本障碍在于激励相容，而非交易成本：即使谈判没有成本，自动化博弈仍保有其占优策略结构。因此，纠正外部性需要一种不依赖自愿协议、而是直接改变各企业自动化边际激励的工具。

4.6 庇古式自动化税

纠正负外部性的经典方法是征收庇古税：按单位征收的税额等于边际外部成本，从而使每个主体的私人激励与社会成本相一致（庇古，1920年）。与许多教科书中的外部性不同，后者受损方位于企业市场之外（例如污染）；在这里，受损方是工人，而他们的收入构成了企业自身的需求。这意味着税率、税收收入和税负归宿都通过同一个劳动力市场渠道相互作用，由此带来比标准情形更复杂的政策设计问题。

命题 5 (庇古式自动化税)。 令 $\tau \geq 0$ 为按任务征收的自动化税，$k > 0$，并假设 $N > N^*$ 且均衡为内部解。

(i) 纳什均衡自动化率为 $\alpha^{NE}(\tau) = (s - \tau - \ell/N)/k$。该比率

$$\tau^* = \ell\left(1 - \frac{1}{N}\right)$$

实现了 $\alpha^{NE} = \alpha^{CO} = (s - \ell)/k$。

(ii) 在不返还的情况下，税率为 τ^* 时，每家企业的收益为 $\pi^{\text{tax}} = \pi^{CO} - \tau^* L \alpha^{CO}$。将税收收入一次性返还后，每家企业恰好获得 π^{CO}。

最优税率具有直观的经济解释：每家企业已经承担了 ℓ/N 自身自动化造成的需求损失；该税则让它为施加于竞争对手的剩余 $\ell(1-1/N)$ 买单。当 N较大时， $\tau^* \approx \ell = \lambda(1-\eta)w$，因此设定税率只需行业层面可观测量。然而，征税需要观测企业层面的自动化率，这在实践中颇具挑战；不过，随着人工智能采用产生可观测的采购记录，这一难题或许正在缓解（Guerreiro 等人，2022年）。不同于科斯式谈判情境中的竞争企业（第4.5节），税务机关可以通过强制报告、工资记录和采购审计，强制企业披露信息，因此即便无法进行私人核验，也能实现近似测量。由于楔子造成的福利损失呈二次函数关系（命题 2），即使税收的针对性不够精确，也能带来一阶收益。

税收收入分配。 命题 5 确定了税率；剩下的设计问题是如何处理这笔收入。由于外部性通过劳动力市场传导，这一选择会影响支配扭曲的结构参数，而不仅仅是影响收益的分配。

向企业提供一次性返还可以完全恢复合作利润（第 (ii) 部分），但这会把收入返还给实行自动化的企业，却让受到损害的被替代工人得不到补偿。更自然的做法是将收入用于帮助这些工人。有两种渠道可选，其激励效应各不相同。

直接转移支付（工资保险、离职补贴）通过弥补收入损失，机械性地提高 η。企业的自动化激励不受影响：无论收入流向何处，企业每执行一项自动化任务都要支付 τ，因此其一阶条件不会改变。但标准的道德风险问题依然存在：慷慨的收入替代可能会削弱工人再培训或重新配置的动力，使 η 依靠转移支付而非生产性再吸纳得以维持。

为再培训计划提供资金也会提高 η，但这是通过人力资本投资，而非收入替代来实现的。这种渠道见效较慢，也更难实施，但它能提升工人的能力，使其重返劳动力市场并获得相当或更高的工资，从而让 η 的收益能够自我维持。原则上，由此产生的动态过程会自我强化，与第 4.1节的分析相辅相成：税收为能够提高 η的计划提供资金，进而降低 ℓ，从而减少未来时期的 τ^*。如果再吸纳取得成功，所需的纠正措施会随时间推移而缩减，税收也就只是过渡性的，而非永久性的。这与 Guerreiro 等人 Guerreiro 等人 (2022年) 的研究结论相呼应：随着被替代群体退休，以及新工人在充分了解自动化的情况下选择职业，最优机器人税会逐渐降至零。

在实践中，两者结合起来可能是最优选择：短期内通过转移支付缓解替代带来的冲击，同时通过长期再培训，在 η中积累持久收益。因此，庇古税有望

一举两得：在边际上纠正外部性，并将其收入重新分配，以逐步减少扭曲。

5 扩展

基准情形在能够支持需求外部性的最简单环境中，将其单独分离出来。一个自然的担忧是，该结果取决于哪些因素被固定不变：内生工资调整可能会弥合楔子，自由进入可能会促使市场达到有效规模，更高的人工智能生产率可能通过扩大整体经济规模来解决需求问题，而资本收入再循环可能抵消因替代而损失的支出。本节将逐一讨论这些质疑，同时分析更丰富的产品市场互动，并表明外部性对所有这些变化都具有稳健性，在某些情况下甚至会被放大。表 2 预览了相关结果。

Tab表 2：基准模型的扩展及其对过度自动化结果的影响 t.

	章节	是否修改了 N^*?	楔子效应	消除外部性?
人工智能生产率（$\phi > 1$）	5.1	No	扩大	No
内生进入	5.2	No	持续存在（可能扩大）	No
内生工资	5.3	是（提高）	缩小	No
资本收入再循环（$\hat{}\eta > 0$）	5.4	是（提高）	缩小	部分地
不完全竞争	5.5	—	持续存在	No

注。 每一行都记录了该推广如何影响自动化阈值 N^*、过度自动化楔子，以及是否消除了需求外部性。在经验上合理的参数下，五种推广情形均未能消除这一外部性。

5.1 人工智能生产率

在基准情形中，人工智能和人类劳动者每项任务的产出相同，因此自动化激励完全由成本驱动。实际上，人工智能在替代人类的同时，也能提高每项任务的产出（例如，自主智能体编程代理、吞吐量更高的客服机器人）。为体现这一点，我们在成本节约之外加入生产率优势。一个自然的猜想是，这一产出渠道通过提高经济生产率来缓解需求问题。我们表明，事实恰恰相反：人工智能每项任务的产出越高，过度自动化楔子就越大。

设人工智能执行的任务产出 $\phi \geq 1$ 单位产出，而人类执行的任务产出1单位。在 $\phi > 1$下，企业 i的产出变为

$$Y_i(\alpha_i) = [\phi\alpha_i + (1 - \alpha_i)]L = [1 + (\phi - 1)\alpha_i]L.$$

竞争中，收入按产出份额分配：$\text{Rev}_i = D \cdot Y_i / \sum_j Y_j$。在

在完美共 a

¯对称策略组合下，所有企业的产出均为 $Y = [1 + (\phi - 1)\alpha]L$，因此 $\text{Rev}_i = D/N$ 与基准情形相同。对 α_i 求导，并在对称策略组合处求值，得到

$$\left.\frac{\partial \text{Rev}_i}{\partial \alpha_i}\right|_{\text{sym}} = \underbrace{-\frac{\ell L}{N}}_{\text{demand loss}} + \underbrace{\frac{D(\phi-1)(N-1)}{N^2[1+(\phi-1)\alpha]}}_{\text{market-share gain}}. \tag{10}$$

第一项是基准需求外部性，它仅取决于 ℓ 和 N，因此与 ϕ无关。第二项则是新增的：偏离企业提高产出，超过竞争对手的产出，从而占据更大的支出份额。只要 $\phi > 1$，这种市场份额收益就是正的，会使自动化的私人激励高于基准情形。

为了量化这一影响，我们将成本节约、需求损失和市场份额收益纳入一阶条件。对称均衡下，边际整合成本等于综合收益：

$$k\alpha = s - \frac{\ell}{N} + \frac{D(\alpha)\,(\phi-1)(N-1)}{N^2[1+(\phi-1)\alpha]\,L}.$$

由于 $D(\alpha)$关于 α是线性的，通分后可得一个二次方程，其正根即为唯一均衡；然而，所得表达式不如基准公式直观，因此下文的比较静态结论将通过单调交叉论证得出。

命题6 （人工智能生产率会扩大过度自动化楔子）。 设 $k > 0$ 并假设均衡为内部解。

（*i*）纳什均衡自动化率随人工智能生产率上升而提高： $\alpha^{NE}(\phi) > \alpha^{NE}(1)$ 对于$\phi > 1$成立。（*ii*）然而，合作最优和广义规划者最优均不变：$\alpha^{CO}(\phi) = \alpha^{CO}(1)$ 以及 $\alpha^{SP}(\mu;\ \phi) = \alpha^{SP}(\mu;$ **1)** 对所有 $\mu \in [0,1]$均成立。（*iii*）因此，对于每个 $\mu \in [0,1]$，过度自动化楔子 $\alpha^{NE}(\phi) - \alpha^{SP}(\mu;$ $\phi)$ 都随 ϕ 严格增加。

该机制属于红皇后效应：每家企业都认为，自动化程度超过竞争对手将带来市场份额收益；但在对称均衡下，所有企业都会同等扩张，因此这些收益彼此抵消。相比之下，成本节约 s会以相同方式进入每家企业的利润，而不受竞争对手选择的影响，因此会使α^{NE} 和 α^{CO} 同等变动，并使楔子保持不变。

第（ii）部分成立，是因为在市场出清 D 条件下，部门总收入等于总支出，而 D （2）取决于工人收入，而非产出：更高的 ϕ 会提高产出，但会按相同比例降低价格，从而使规划者目标对 ϕ保持不变。结合第（i）部分，可得

第（iii）部分：更先进的人工智能会提高均衡自动化率，却不会改变效率基准，因此，扭曲程度会随着人工智能能力的增强而扩大。

更大的楔子还带来一项政策含义。由于市场份额动机在需求外部性之上又增加了第二重扭曲，基准庇古税率 $\tau^* = \ell(1-1/N)$ 已不再足够：实施 α^{CO} 需要额外进行校正，校正幅度等于在合作率下评估的（10）中的市场份额项；只要 $\phi > 1$，该项就严格为正。尽管如此，较低的价格意味着每一单位支出能够购买更多的实物产出，因此，由名义流量构建的福利度量 $S(\mu)$低估了较高 ϕ 所带来的实际消费收益。该命题揭示的是一种策略性扭曲，而不是声称提高人工智能生产率会降低总体净福利。

5.2 内生进入

到目前为止，企业数量一直是外生给定的。在自由进入条件下，人们可能会预期，过度自动化问题能够自我纠正：剩余侵蚀降低利润，边际企业退出，剩余行业最终达到有效规模。然而，这一逻辑能否成立，取决于进入边际如何与自动化决策相互作用。

考虑一个两阶段博弈：企业支付固定成本 $\kappa \geq 0$ 进入市场，然后同时选择自动化率。给定 N 个新进入者，第 2 阶段的纳什均衡会产生单个企业经营利润$\Pi^*(N)$。纯策略**自由进入均衡** 是一个整数 $N \geq 1$，满足

$$\Pi^*(N) \geq \kappa \qquad \text{and} \qquad \Pi^*(N+1) \leq \kappa : \tag{11}$$

在位企业弱偏好于继续经营，而额外的新进入者无法收回固定成本。

由于无摩擦（$k = 0$）和凸成本（$k > 0$）机制对利润函数的影响在性质上截然不同，我们分别讨论。无摩擦情形令 $\lambda = 1$，以便得到简洁的闭式解；对于任意 $\lambda \in$，定性结果仍然成立（$0, 1],$，因为这一假设只影响利润水平，不影响进入机制的结构）。

在无摩擦基准（$k = 0$）中，自动化是全有或全无的。当 $\ell > s$时，利润函数在 N^*处离散下降：低于该阈值时，没有企业采用自动化；高于该阈值时，全面自动化是占优策略（推论 1），且单个企业利润下降 $\Delta := L(\ell - s) > 0$（附录中的图 4 对此作了说明）。令 $m := \lfloor N^* \rfloor$ 表示不超过 N^*的最大整数。

命题7（无摩擦基准下的内生进入）。假设 $k = 0, \quad \lambda = 1, \quad \ell > s,$

且 $0 < \kappa < A$。单企业均衡利润函数

$$\Pi^*(N) = \begin{cases} A/N & \text{if } N \le N^*, \\ A/N - \Delta & \text{if } N > N^*, \end{cases}$$

在 $\mathbb{N}$ 上严格递减。唯一的纯策略自由进入均衡中，企业数量为 $N^{FE} = \max\{N \in \mathbb{N} : \Pi^*(N) \ge \kappa\}$。根据进入成本的不同，会出现三种情形：

(i) 低进入成本 *($\kappa \le A/(m+1) - \Delta$)*：$N^{FE} = \lfloor A/(\kappa + \Delta) \rfloor \ge m + 1$。每家企业都实现完全自动化。

(ii) 中等进入成本 *($A/(m+1) - \Delta < \kappa < A/m$)*：$N^{FE} = m$。没有企业实现自动化，但每家企业都获得严格为正的利润 $A/m > \kappa$。自动化的威胁 阻止了更多企业进入：再增加一家企业就会触发完全自动化，使单个企业利润降至 $A/(m+1) - \Delta < \kappa$。

(iii) 高进入成本 *($\kappa \ge A/m$)*：$N^{FE} = \lfloor A/\kappa \rfloor \le m$。没有企业实现自动化。限制企业数量的是进入成本，而非自动化外部性。

当情形（i）出现时，推论1 中的囚徒困境便会在自由进入下显现：所有企业都进行自动化，需求收缩，而如果没有任何企业进行自动化，每家企业都会更好。情形（iii）是标准的自由进入结果：进入成本足够高，以至于市场始终不会接近自动化阈值，外部性也无关紧要。情形（ii）最为独特：自动化的威胁充当内生进入壁垒，在实际上并未发生任何自动化的情况下维持正利润，代价则是维持市场势力。

凸成本情形没有那么极端，但更为稳健。当 $k > 0$ 时，自动化率 $\alpha^{NE}(N) = (s - \ell/N)/k$ 随 N 连续变化，因此利润函数不再在 N^* 处跳跃，命题 7(ii) 中的进入威慑机制也不会出现。

命题8 (凸成本下的内生进入**)**。假设 $k > 0$，$\ell > s$，且 $\Pi_0(1) > \kappa$（市场可行）。存在满足 N^{FE} 的自由进入均衡 (11)。若 $N^{FE} > N^*$，则 $\alpha^{NE}(N^{FE}) > \alpha^{CO}$：过度自动化在自由进入下持续存在。

一般而言，当阈值处的零自动化利润 $\Pi_0(N^*)$ 超过 κ 时，N^{FE} 就大于 N^*。[3] 自由进入随后确定企业数量，但不会改变自动化子博弈中的战略激励：每家企业仍只承担需求损失的一部分 $1/N^{FE}$，

[3] 在涵盖 $c/w \in \{0.1, \dots, 0.5\}$、$\lambda \in \{0.3, \dots, 1\}$、$\eta \in \{0, \dots, 0.3\}$、$k \in \{0.5, 1, 2\}$ 和 $\kappa \in \{0.1, \dots, 5\}$ 的数值网格中，在满足该命题条件的参数设定中，超过 % 的设定满足 N^{FE} exceeds N^*；例外情况仅出现在进入成本足够高、以致市场勉强能够支撑超过 94 家企业时。

过度自动化楔子 $\ell(1-1/N^{FE})/k$持续存在。如果 $N^{FE} \leq N^*$，则没有企业进行自动化，结果是有效率的，但这仅仅是因为市场过于集中，私人自动化激励无法发挥作用。

综合来看，这两个命题传达了一个共同的启示：自由进入会重塑过度自动化问题，却无法解决它。事实上，过度进入的普遍倾向（Mankiw 和 Whinston，1986年）会进一步分割市场，从而扩大楔子。

进入边际还揭示了全民基本收入（第 4.2 节）的一种意外副作用。通过提高自主需求 A，全民基本收入会在任意给定的 $\Pi_0 = A/N + (\lambda - 1)wL$ 下提高单个企业利润 N，吸引更多新进入者，直到零利润条件（11）在更大的 N^{FE}处成立。由于过度自动化楔子 $\ell(1-1/N)/k$ 随 N增加，旨在缓冲替代影响的政策，反而可能扩大导致替代的那种外部性。

5.3 内生工资

Acemoglu 和 Restrepo (2018年)的一个核心洞见是，内生工资调整能够稳定自动化进程：随着企业推进自动化，被替代工人会增加劳动力供给，压低工资；工资下降会缩小自动化带来的成本节约，从而抑制进一步替代。这种自我纠正的反馈机制，自然可以用来解决上文所指出的需求外部性。我们将表明，它会提高外部性开始显现的阈值，但一旦外部性显现，就无法消除楔子。

在Acemoglu 和 Restrepo （2018年）中，工资由完整一般均衡中的劳动力市场出清决定；我们采用一种简化形式的表述，以捕捉其机制的关键定性特征。令工资取决于总体自动化率： w（$\bar{\alpha}$）其中 w（$\bar{\alpha}$）$> c$ 且w'（$\bar{\alpha}$）≤ 0；企业在选择 α_i时，将现行工资视为既定。这一设定只要求总体劳动力市场闲置程度上升时工资下降；效率工资模型也具有这一性质，即无偷懒工资会随失业率下降（Shapiro 和 Stiglitz，1984年），经验工资曲线亦然：在针对十多个国家的研究中，Blanchflower 和 Oswald （1995）发现工资与失业之间存在稳健的负相关关系。在我们的设定中，自动化将工人挤入劳动力池，恰好造成这种闲置。

成本节约 $s(w) = w - c$ 和需求损失参数 $\ell(w) = \lambda(1-\eta)w$ 都随工资上升而增加，因此工资下降会影响自动化边际的两端：一方面削弱自动化的私人激励（自我纠正渠道），另一方面降低每项自动化任务造成的需求损失。因此，均衡自动化率是一个不动点，其中自动化、工资和外部性共同决定。尽管存在这种更复杂的反馈，阈值

$N^* = \ell/s = \lambda(1-\eta)w/(w-c)$会随着工资下降而上升，因为成本节约 $s = w - c$收缩得比需求损失 $\ell = \lambda(1-\eta)w$更快。然而，扭曲的结构性来源并未改变。**命题 9** （对工资调整的稳健性）。 令 $k > 0$，令 $w\colon [0,1] \to (c,\ \infty)$ 可微，且满足 $w'(\bar{\alpha}) \le 0$，并假设企业是工资接受者。

(i) 在任何满足 $N > N^*(w(\bar{\alpha}))$ 的对称均衡中，纳什自动化率都高于合作最优： $\alpha^{NE} > \alpha^{CO}$。 *(ii)* 对所有 $\bar{\alpha} \in [0,1]$，内生工资调整都会提高阈值： $N^*(w(\bar{\alpha})) \ge N^*(w(0))$；当且仅当 $w(\bar{\alpha}) < w(0)$ 时，不等式严格成立。

无论工资水平如何，竞争性定价都会将收入分配为 $\text{Rev}_i = D/N$，因此无论 w 是高还是低，每家企业都只承担其自动化所造成的需求破坏的一部分。工资调整会改变幅度，但不会改变每家企业内化的比例；该比例取决于市场结构，而非要素价格。

自我纠正论点最有力的版本是，工资可能会下降到足以彻底消除外部性的程度。随着 $w \to c$，成本节约 $s \to 0$，而 $N^* \to \infty$：最终 N^* 超过 N，没有企业会认为自动化在私人层面值得采用。但这是一种得不偿失的解决方案。当工资被压低到接近人工智能成本时，保住工作的工人，其收入只比取代他们的机器高一点点；总购买力也会因工资压低而崩塌，而非因替代而崩塌。外部性之所以消失，并不是因为需求问题得到了解决，而是因为每名工人所剩的收入太少，私人激励与社会激励之间的楔子变得微不足道：一个只能靠让劳动力陷入贫困来“自我纠正”的劳动力市场，不过是把替代转化成了生活水平下降。更一般地说，工资灵活性改变的是外部性何时产生影响，而不是是否存在。

上述分析设定 $\mu = 0$，因此规划者只关心企业利润。若规划者也重视工人福利（$\mu > 0$），就不会认为工资压低比替代更可接受，因而会要求更大幅度的纠正；然而，工资调整对 ℓ的压缩幅度相同。因此，内生工资只能弥合较小一部分差距，作为纠正机制就更加不足。附录中的推论4 证实了这一点：在内生工资条件下，过度自动化结果适用于任何$\mu \in [0,1)$。

5.4 资本收入再循环

第4.1 节表明，提高 η（工人收回的被替代收入比例）会降低需求损失参数 ℓ 并缩小过度自动化楔子。资本方面一个自然的对应做法是

所有者将利润用于支出：如果他们的消费抵消了替代导致的支出损失，需求外部性或许会消失。我们表明，再循环会缩小楔子，但在经验上合理的参数范围内无法将其消除。

假设资本所有者在该行业中将其资本收入的 $\hat{\eta} \in [0, 1)$ 用于消费。部门利润总额为 $\Pi = D - NL(w - s\bar{\alpha})$。将资本消费 $\hat{\eta}\Pi$ 加入总需求，并求解 D，得到

$$D = \frac{A + (\lambda - \hat{\eta})wLN}{1 - \hat{\eta}} - \frac{\ell_{\hat{\eta}} LN}{1 - \hat{\eta}}\,\bar{\alpha}, \tag{12}$$

其中

$$\ell_{\hat{\eta}} = \ell - \hat{\eta}s$$

是有效需求损失参数：每项自动化任务会导致 ℓ 的工人支出损失，但所有者会将每项任务的成本节省 $\hat{\eta}$ 的部分 s 再投入需求。当 $\hat{\eta} = 0$时，$\ell_{\hat{\eta}} = \ell$ 且方程（12）化简为方程（2）。

竞争性定价仍给出 $\text{Rev}_i = D/N$。一阶条件变为 $L(s - \ell_{\hat{\eta}}/[N(1 - \hat{\eta})])$，从而得到修正阈值

$$N_{\hat{\eta}} := \frac{\ell_{\hat{\eta}}}{s(1 - \hat{\eta})}.$$

命题 10（资本收入再循环）。假设 $k = 0$ 和 $\ell_{\hat{\eta}} > 0$。当资本收入以 $\hat{\eta}$ 的比率进行再循环时，

（*i*）当且仅当 $N > N_{\hat{\eta}}$ 时，完全自动化是占优策略。
（*ii*）外部性仅在 $\hat{\eta} \geq \ell/s$ 时消失（$\ell_{\hat{\eta}} \leq 0$）。

第 (ii) 部分要求$\hat{\eta} \geq \ell/s = \lambda(1 - \eta)w/(w - c)$：所有者必须将每项任务的部分成本节约进行再循环，以弥补被替代工人原本会创造的需求。当 $\ell > s$时，所需比率超过 1，因此再循环恰恰在外部性危害最大之处不起作用，因为 $\ell > s$ 意味着 $\alpha^{CO} = 0$，企业会进行自动化，而规划者则更倾向于完全不进行自动化。当$\ell < s$时，从原则上说，消除是可行的，但规划者本已更倾向于进行正向自动化（命题 1），且该楔子在数量上更小。

无摩擦情形给出了最精确的结果，但当摩擦成本为正时，这一结构依然成立。命题10 的证明将结果推广到 $k > 0$。纳什均衡推广为 $\alpha^{NE} = (s - \ell/\hat{N})/k$，其中$\hat{N} = N(1 - \hat{\eta}) + \hat{\eta}$ 是一个有效市场规模，取值介于 N (不再循环)与1 (完全回收)之间，使每家企业的行为仿佛面对的竞争者更少。不过，合作最优保持不变：$1/(1 - \hat{\eta})$ 乘数会按比例放大总利润，但不会改变最优解。

结论与第4.1节类似：再循环提高了每家企业内化的需求损失比例，从ˆ $1/N$ 提高到 $1/N$，但无法使其达到1。讨论收入如何使用可以缩小楔子，却无法消除它，因为企业之间的稀释仍然存在。

5.5 不完全产品市场竞争与任务互补性

基准情形假设采用竞争性定价，且各项任务之间完全替代。虽然未作正式论述，但我们认为，更丰富的产品市场互动和任务互补性会使分析更加复杂，却不应消除需求外部性。

第二阶段的价格或数量竞争。 如果企业在选择自动化率后进行数量竞争（古诺）或价格竞争（伯特兰），就会产生两种新的作用力。第一种是需求分配效应：策略性竞争会改变既定支出水平在企业之间的分配方式。第二种是市场份额动机：自动化程度高于竞争对手的企业可以降低价格或扩大产出，从而获取更大的支出份额。这些作用力对过度自动化楔子大小的影响各不相同，但都不会消除需求外部性本身，因为它们都没有改变这样一个事实：自动化通过工人替代降低了总支出的水平 D。

需求分配效应显而易见。无论收入是通过竞争性定价、古诺市场份额，还是伯特兰式压价来分配，每家企业仍只承担其自动化所导致的 D 降幅的一部分。对于差异化产品，未内部化的需求损失与企业未 占据的市场份额成正比；因此，只要企业不具备完全市场势力，过度自动化就会持续存在。

市场份额动机更加微妙。在对称均衡下，各企业的市场份额收益相互抵消，这与第 5.1 节的红皇后结构相呼应。然而，在古诺模型下，扩大市场份额的企业也会承担其造成的更大一部分需求损失，从而部分抵消这一动机；对楔子的净影响可能取决于市场结构的具体细节。

CES任务聚合。 Acemoglu 和 Restrepo 的基于任务的框架(2018年)通过 CES生产函数聚合任务；为便于分析，我们的基准情形采用完全替代极限。在替代弹性大于 1 的一般 CES 聚合函数下（任务彼此可替代，但并非完全替代），自动化边际任务带来的产出增益递减，因此即使没有凸型整合成本，也会出现内部解 k。由于这改变了生产函数的形态，却没有改变需求侧，因此需求外部性仍会在

自动化边际发挥作用，而过度自动化楔子持续存在，不过它会较小，因为边际收益递减已从供给侧抑制了自动化。

当任务具有互补性（弹性小于一）时，约束更强：生产函数本身会限制自动化，因为消除人类任务会降低产出。即便如此，只要被替代工人失去收入，需求外部性仍然为正。

在每种情况下，更丰富的模型假设都会改变过度自动化的程度，但不会改变其根源：只要企业没有完全内化替代造成的需求损失，这一楔子就会持续存在。刻画策略性定价、市场份额动机与需求外部性之间的确切相互作用，是未来研究的一个有前景方向；基准情形的简化假设以最鲜明的形式凸显了核心机制。

6 讨论

本文构建了一个简单的模型，并得出了一个简单却鲜明的洞见。即使人工智能驱动的裁员席卷各个行业，即使每家企业都认识到，工资收入的消失意味着客户的消失，也没有一家企业会停下来。每家企业都获得了替代自身工人所带来的全部节省，却只承担了其造成的需求损失中的一小部分；其余损失则落在竞争对手身上。没有一家企业能够承受成为那个选择克制的企业。这就是陷阱：一场自动化军备竞赛，随着人工智能的进步只会愈演愈烈，使工人和企业所有者都陷入更糟的境地，而且没有任何市场力量能够打破它。最后，我们将讨论对实证研究和政策的影响，以及分析的范围与局限性。

实证启示。Anthropic 首席执行官 Dario Amodei 曾警告，人工智能驱动的岗位替代将比以往的技术冲击“痛苦得多”“波及范围广得多”且“发生得快得多”（CNBC，2026a）。如果这一判断被证明正确，且收入替代仍不充分，那么模型或许会得出一个有悖直觉的结论，指出问题最严重之处：并非占优策略的科技企业，而是部署最强大人工智能的分散型行业（命题 1 和6）。其独特的实证特征将是利润侵蚀。标准竞争模型预测，降低成本的技术会提高利润；若利润侵蚀与大规模裁员同时发生，就很难在不考虑外部性的情况下解释这一现象（命题 2）。不过，这一特征需要岗位替代达到迄今尚未出现的规模和速度。如果再吸纳跟得上自动化步伐，外部性可能仍小到难以察觉，而本文的贡献在于指出一种结构性脆弱性，而不是诊断一场正在发生的危机。人工智能驱动的岗位替代已然展开的三个领域，为此提供了具体的切入点：客户支持，

数千家企业正同时以智能体人工智能取代代理人（CNBC，2025c）；软件服务领域，能够让一名工程师取代一支多人团队的工具（CNBC，2025a）正在带来可衡量的人数与产出比变化；以及竞争中的金融机构的后台运营领域，在这里，监管报告使采用率和收入结果都格外透明。

更广泛地说，相关文献表明，人工智能能带来显著的生产率提升（Brynjolfsson 等人，2025b），竞争压力会加速采用（Li 等人，2025年），而随着人工智能能力提升，有效的人工监督也愈发难以实现（Bastani and Cachon，2025年）。这些发现都表明，企业层面存在强烈的自动化激励。我们的模型揭示了当每家企业同时响应这一激励时会发生什么：任何一家企业所观察到的回报，都没有计入所有企业共同造成的需求损失，因此，人工智能带来的私人回报会系统性地高估其对整个经济的回报。

政策影响。围绕人工智能驱动的岗位替代展开的政策辩论，大多聚焦于事后如何应对，例如通过再培训、收入支持或监管。我们的研究结果重新界定了问题：竞争性激励是否会驱使企业过度自动化，超出集体最优水平？即使规划者完全不考虑工人福利，也会将自动化率降至均衡水平以下（命题 2）。问题不在于企业以牺牲工人为代价获利，而在于过度自动化损害了两个群体，因此，纠正这一问题的关键是消除浪费，而非重新分配收益。全民基本收入或许是讨论最广泛的应对措施，它能提高生活水平，却不会改变任何一家企业的自动化激励（第 4.2节）。集体谈判也面临同样的障碍：由于自动化是一种占优策略，企业之间任何自愿达成的裁员限制协议都无法自我执行（第4.5节）。根据廷伯根原则，不同的市场失灵需要不同的政策工具；只有庇古式自动化税能够提供这样的工具（表 1）。再多的再培训、收入支持或谈判，都无法减缓这场军备竞赛；只有对自动化本身征税，才能改变驱动这场竞赛的决策逻辑。

实施方面有一项实际考量：该模型是一个封闭行业博弈，单边征收自动化税可能会推动采用转移至海外，从而增强了采取多边协调或类似于碳政策中所用边境调整机制的理由。

范围、局限与未来方向。该模型刻意保持简单：一个行业、一个时期、对称企业。这些设定都较为保守，也就是说，现实问题很可能比我们展示的更加严重。

只考虑一个行业会低估外部性。在多部门经济中，一个行业的裁员会减少对所有行业产出的支出，从而形成相互强化的需求螺旋。平台生态系统让这一点更加具体：当平台将卖家支持、零工物流或内容审核自动化时，支出减少的影响会蔓延至整个互补者生态系统。

静态设定忽略了两种方向相反的动态力量。人工智能投资大多不可逆，而命题 7 表明，即便只是自动化的威胁，也可能在发生任何替代之前重塑市场结构，从而强化尽早采取政策行动的理由。另一方面，收入替代率 η 会随着被替代工人接受再培训、新职业出现而逐渐上升（Acemoglu 和 Restrepo，2019年），因此随着经济调整，最优税应逐步降低（第4.6节）。

对称性排除了企业和工人之间的异质性，而将人工智能发展内生化可能会加剧这一问题：竞相推进自动化的企业可能会过度投资于替代劳动型人工智能，而非增强劳动型人工智能（Acemoglu 和 Restrepo，2018年），从而助长模型所揭示的军备竞赛。

这些扩展都指向同一个结论：问题会变得更大，而非更小。尽管如此，推进这些研究，并开展上述实证检验，都是未来研究的有益方向。

参考文献

Daron Acemoglu。《人工智能的简单宏观经济学》。《经济政策》，40(121):13–58，2025年。Daron Acemoglu 和 Pascual Restrepo。《人与机器之间的竞赛：技术对增长、要素份额和就业的影响》。《美国经济评论》，108(6):1488–1542，2018年。Daron Acemoglu 和 Pascual Restrepo。《自动化与新任务：技术如何取代并重新创造劳动》。《经济展望杂志》，33(2):3–30，2019年。Daron Acemoglu 和 Pascual Restrepo。《错误类型的人工智能?人工智能与劳动需求的未来》。《剑桥区域、经济与社会杂志》，13(1):25–35，2020年。Daron Acemoglu 和 Pascual Restrepo。《自动化与租金耗散：对工资、不平等和生产率的影响》。《经济学季刊》，即将出版。Imanol Arrieta-Ibarra、Leonard Goff、Diego Jiménez-Hernández、Jaron Lanier 和 E. Glen Weyl。《我们应将数据视为劳动吗？超越“免费”》。《美国经济学会论文与会议录》，108:38–42，2018年。David H. Autor、Frank Levy 和 Richard J. Murnane。《近期技术变革的技能构成：实证探索》。《经济学季刊》，118(4):1279–1333，2003年。David H. Autor、Caroline Chin、Anna Salomons 和 Bryan Seegmiller。《新前沿：新工作的起源与内容，1940–2018年》。《经济学季刊》，139(3):1399–1465，2024年。

Martino Banchio 和 Giacomo Mantegazza。《人工智能与自发合谋》。工作论文，博科尼大学和南加州大学马歇尔商学院，2022年。Hamsa Bastani 和 Gérard P. Cachon。《人类与人工智能契约悖论》。工作论文，沃顿商学院、宾夕法尼亚大学，2025年12月。SSRN 5962739。Seth G. Benzell、Laurence J. Kotlikoff、Guillermo LaGarda 和 Jeffrey D. Sachs。《机器人就是我们：人类替代的经济学》。工作论文 20941，美国国家经济研究局，2015年2月。Martin Beraja 和 Nathan Zorzi。《低效自动化》。《经济研究评论》，92(1):69–96，2025年。David G. Blanchflower 和 Andrew J. Oswald。《工资曲线导论》。《经济展望杂志》，9(3):153–167，1995年。Erik Brynjolfsson 和 Andrew McAfee。《第二次机器时代：辉煌技术时代的工作、进步与繁荣》。W. W. Norton，纽约，2014年。Erik Brynjolfsson、Bharat Chandar 和 Ruyu Chen。《煤矿里的金丝雀？人工智能对近期就业影响的六个事实》。工作论文，斯坦福经济政策研究所，2025a年8月。Erik Brynjolfsson、Danielle Li 和 Lindsey Raymond。《生成式人工智能在工作中的应用》。《经济学季刊》，140 (2):889–942，2025b年。CNBC。《高盛正在试点其首个自主编码器：华尔街人工智能里程碑》。
https://www.cnbc.com/2025/07/11/goldman-sachs-autonomous-coder-pilot-marks-major-ai-milestone.html，2025a年。访问日期：2026年2月28日。CNBC。《人工智能裁员：亚马逊、微软等公司称人工智能是2025年裁员的原因》。
https://www.cnbc.com/2025/12/21/ai-job-cuts-amazon-microsoft-and-more-cite-ai-for-2025-layoffs.html，2025b年。访问日期：2026年2月28日。CNBC。《Salesforce首席执行官确认因“我需要更少的人手”而进行4,000人裁员，并归因于人工智能》。
https://www.cnbc.com/2025/09/02/salesforce-ceo-confirms-4000-layoffs-because-i-need-less-heads-with-ai.html，2025c年。访问日期：2026年2月28日。CNBC。《Anthropic首席执行官Dario Amodei警告称，人工智能可能对就业造成“异常痛苦”的冲击》。
https://www.cnbc.com/2026/01/27/dario-amodei-warns-ai-cause-unusually-painful-disruption-jobs.html，2026a年。访问日期：2026年3月20日。CNBC。《Block裁员约4,000人，接近其劳动力的一半》。
https://www.cnbc.com/2026/02/26/block-laying-off-about-4000-employees-nearly-half-of-its-workforce.html，2026b年。访问日期：2026年2月28日。R. H. Coase。《社会成本问题》。《法律与经济学杂志》，3:1–44，1960年。Russell Cooper 和 Andrew John。《凯恩斯主义模型中的协调协调失灵》。《经济学季刊》，103(3):441–463，1988年。Arnaud Costinot 和 Iván Werning。《机器人、贸易与卢德主义：最优技术监管的充分统计量方法》。《经济研究评论》，90(5):2261–2291，2023年。Vincent P. Crawford 和 Joel Sobel。《策略性信息传递》。《计量经济学》，50(6):1431–1451，1982年。Tyna Eloundou、Sam Manning、Pamela Mishkin 和 Daniel Rock。《生成式预训练变换器就是生成式预训练变换器：大型语言模型对劳动力市场的潜在影响》。《科学》，384:1306–1308，2024年。

João Guerreiro、Sergio Rebelo 和 Pedro Teles。《机器人应该被征税吗？》《经济研究评论》，89(1):279–311，2022年。Daniel S. Hamermesh 和 Gerard A. Pfann。要素需求的调整成本。《经济文献杂志》，34 (3):1264–1292，1996年。印孚瑟斯。印孚瑟斯与 Cognition 宣布开展战略合作，加速全球企业的人工智能价值之旅。新闻稿，2026年1月7日。
h
t
t
p
s://
www.prnewswire.com/news-releases/infosys-and-cognition-announce-strategic-collaboration-to-accelerate-the-ai-value-journey-for-glob
al-enterprises-
html，访问日期：2026年3月20日。Louis S. Jacobson、Robert J. LaLonde 和 Daniel G. Sullivan。被替代工人的收入损失。《美国经济评论》，83(4):685–709，1993年。Nicholas Kaldor。分配理论的其他观点。《经济研究评论》，23(2):83–100，1956年。Jussi Keppo、Yuze Li、Gerry Tsoukalas 和 Nuo Yuan。人工智能智能体合谋的脆弱性。工作论文，波士顿大学和新加坡国立大学，2026年1月。SSRN 5386338。约翰·梅纳德·凯恩斯。我们子孙后代的经济前景。载于《劝说文集》，第358–373页。麦克米伦，伦敦，1930年。Anton Korinek 和 Joseph E. Stiglitz。人工智能及其对收入分配和失业的影响。载于 Ajay Agrawal、Joshua Gans 和 Avi Goldfarb 编，《人工智能经济学：研究议程》。芝加哥大学出版社，芝加哥，2019年。Wassily Leontief。工作与收入分配。《科学美国人》，247(3):188–204，1982年。Boshuo Li、Ni Huang 和 Wei Shi。被迫改变？劳工问题的媒体曝光与企业人工智能投资。《信息系统研究》，2025年。doi: 10.1287/isre.2022.0402。在线优先发表文章。Robert E. Lucas。调整成本与供给理论。《政治经济学杂志》，75(4):321–334，1967年。N. Gregory Mankiw 和 Michael D. Whinston。自由进入与社会无效率。《兰德经济学杂志》，17 (1):48–58，1986年。Atif Mian、Ludwig Straub 和 Amir Sufi。负债需求。《经济学季刊》，136(4):2243–2307，2021年。Kevin M. Murphy、Andrei Shleifer 和 Robert W. Vishny。工业化与大推进。《政治经济学杂志》，97(5):1003–1026，1989年。Arthur Cecil Pigou。《福利经济学》。麦克米伦，伦敦，1920年。Eric A. Posner 和 E. Glen Weyl。《激进市场：为公正社会铲除资本主义与民主制度》。普林斯顿大学出版社，普林斯顿，2018年。David Ricardo。《政治经济学及赋税原理》。John Murray，伦敦，第三版，1821年。Paul N. Rosenstein-Rodan。东欧和东南欧工业化问题。《经济学杂志》，53(210/211):202–211，1943年。Alap Shah。《2028年全球智能危机》。Citrini Research，2026年2月。网址 https://www.citriniresearch.com/p/2028gic。Carl Shapiro 和 Joseph E. Stiglitz。作为工人纪律约束机制的均衡失业。《美国经济评论》，74(3):433–444，1984年。

Martin L. Weitzman. 利润分享的简单宏观经济学. 《美国经济评论》, 75(5):937–953, 1985年. Joseph Zeira. 工人、机器与经济增长. 《经济学季刊》, 113(4):1091–1117, 1998年.

A 证明

命题*1*的证明。

序言及第 *(i)* 部分：纳什均衡。 根据方程 (6)，我们有

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s - \frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\sum_{j\neq i}\alpha_j\right]$$

对 α_i 求导（只有涉及 α_i 的项会产生贡献，因为从企业 i的角度看，竞争对手总和$\sum_{j\neq i}\alpha_j$ 是常数）：

$$\frac{\partial \pi_i}{\partial \alpha_i} = L\left[\left(s - \frac{\ell}{N}\right) - k\alpha_i\right]. \tag{13}$$

二阶导数为

$$\frac{\partial^2 \pi_i}{\partial \alpha_i^2} = -kL < 0,$$

因此， π_i 关于 α_i是严格凹的。

令一阶条件等于零:

$$L\left(s - \frac{\ell}{N} - k\alpha_i\right) = 0 \qquad \Longrightarrow \qquad \alpha_i^* = \frac{s - \ell/N}{k}.$$

Now, α_i 的取值范围为 $[0,1]$（因为它是自动化工作的比例）。

如果 $\alpha_i^* > 1$，那么由于 $\frac{\partial^2 \pi_i}{\partial \alpha_i^2} < 0$， $\frac{\partial \pi_i}{\partial \alpha_i} > 0$ 对于 $\alpha_i < \alpha_i^*$成立，因此在这种情况下， $\pi_i(\alpha_i)$ 的最大值在 $\alpha_i = 1$处取得。

同理，如果 $\alpha_i^* < 0$，那么对于 $\alpha_i > \alpha_i^*$， $\frac{\partial \pi_i}{\partial \alpha_i} < 0$ 成立，因此在这种情况下， $\pi_i(\alpha_i)$ 的最大值在 $\alpha_i = 0$处取得。注意，当且仅当 $s < \ell/N$时， $\alpha_i^* < 0$，即 $N < \ell/s = N^*$；这证明了前言中的断言：当 $N \leq N^*$时，没有企业会采用自动化。

注意， α_i^* 不取决于任何竞争对手的选择α_j $(j \neq i)$：竞争对手的自动化水平仅通过加项 $-(\ell/N)\sum_{j\neq i}\alpha_j$产生影响，而该项不会影响一阶条件。因此，企业 i的最优策略独立于 α_{-i} （占优策略），即设定

$$\alpha_i = \frac{s - \ell/N}{k},$$

当右侧位于 $[0,1]$ 内时成立；否则在相应边界处为 $\alpha_i = 0$ 或 $\alpha_i = 1$ 。由于每个企业求解的是同一个问题，唯一的对称纳什均衡满足对所有 i， $\alpha_i = \alpha^{NE}$。

（*ii*）部分：合作最优。将所有企业的单个企业利润（5）相加 N 后，行业的总利润为

$$\begin{aligned}\pi_{tot} &:= \sum_{i=1}^{N} \pi_i \\ &= \sum_{i=1}^{N} \left[\Pi_0 + L\left(s\alpha_i - \ell\bar{\alpha} - \frac{k}{2}\alpha_i^2 \right) \right] \\ &= N\Pi_0 + Ls\sum_i \alpha_i - L\ell N\bar{\alpha} - \frac{kL}{2}\sum_{i=1}^{N} \alpha_i^2.\end{aligned}$$

利用 $\sum_i \alpha_i = N\bar{\alpha}$，可将其简化为

$$\pi_{tot} = N\Pi_0 + NL\bar{\alpha}(s-\ell) - \frac{kL}{2}\sum_{i=1}^{N} \alpha_i^2.$$

前两项仅通过 $\bar{\alpha}$ 依赖于 $\boldsymbol{\alpha}$；最后一项 $-\frac{kL}{2}\sum_i \alpha_i^2$ 在 α_i 尽可能相等时达到最大值（即负值最小），这是由 x^2 的凸性决定的。具体而言，对于任意固定的 $\bar{\alpha}$，$\sum_i \alpha_i^2 \geq N\bar{\alpha}^2$，且当且仅当对所有 i 都有 $\alpha_i = \bar{\alpha}$ 时等号成立（根据均方算术平均不等式）。因此，最优值是对称的：对所有 i 都有 $\alpha_i = \alpha$，且 $\bar{\alpha} = \alpha$。代入：

$$\pi_{tot} = N\Pi_0 + NL\alpha(s-\ell) - \frac{kNL\alpha^2}{2}. \tag{14}$$

这是关于 α 的凹二次函数：

$$\frac{\partial \pi_{tot}}{\partial \alpha} = NL(s-\ell) - kNL\alpha, \qquad \frac{\partial^2 \pi_{tot}}{\partial \alpha^2} = -kNL < 0.$$

令一阶条件等于零：

$$NL(s-\ell) - kNL\alpha = 0 \quad \Rightarrow \quad \alpha^* = \frac{s-\ell}{k}.$$

由于 π_{tot} 是严格凹函数，因此它在 $\mathbb{R}$上有唯一的全局最大值。将范围限制在 $\alpha \in [0,1]$：如果 $(s-\ell)/k > 1$，严格凹性意味着 $\partial \pi_{tot}/\partial \alpha > 0$ 在 $[0,1]$, 上成立，因此约束最大值为 $\alpha = 1$；如果 $(s-\ell)/k < 0$，那么 $\partial \pi_{tot}/\partial \alpha < 0$ 在 $[0,1]$, 上成立，因此约束最大值为 $\alpha = 0$。因此

$$\alpha^{CO} = \frac{s-\ell}{k},$$

当表达式超出 $[0,1]$时，边界值为 0 或 1。

第（*iii*）部分：过度自动化楔子。设 $N > N^*$，因此 $\alpha^{NE} = (s - \ell/N)/k > 0$由（i）可得。

情形 $s > \ell$。$(s - \ell/N)/k > 0$ 和 $(s - \ell)/k > 0$均成立，因此两个比率都没有被限制在零点。相减得：

$$\alpha^{NE} - \alpha^{CO} = \frac{s - \ell/N}{k} - \frac{s - \ell}{k} = \frac{\ell - \ell/N}{k} = \frac{\ell\,(1 - 1/N)}{k}.$$

由于 $\ell > 0$ 且 $1 - 1/N > 0$，因此这对 $N \geq 2$ 严格为正。

对于比较静态，将楔子写作 $W = \ell\,(1 - 1/N)/k$:

$$\begin{aligned}
\frac{\partial W}{\partial N} &= \frac{\ell}{kN^2} > 0 && \text{(increasing in } N\text{)},\\
\frac{\partial W}{\partial \ell} &= \frac{1 - 1/N}{k} > 0 && \text{(increasing in } \ell\text{)},\\
\frac{\partial W}{\partial k} &= -\frac{\ell(1 - 1/N)}{k^2} < 0 && \text{(decreasing in } k\text{)}.
\end{aligned}$$

第（*iv*）部分：边界情形。

当 $s \leq \ell$时，则 $(s - \ell)/k \leq 0$，因此 $\alpha^{CO} = 0$。所以楔子为 $\alpha^{NE} - \alpha^{CO} = \alpha^{NE}$。由于我们处于 $N > N^*$的情形，因此有 $(s - \ell/N)/k > 0$。然而，$(s - \ell/N)/k > 1$仍有可能。

若 $s < k + \ell/N$，则 $\alpha^{NE} = (s - \ell/N)/k$，这就是楔子。若 $s > k + \ell/N$，则 α^{NE}是一个角点情形，$\alpha^{NE} = 1$，这就是楔子。 □

引理 1 (边界情形)。将 $\delta(\mu) := \lambda^{\mu\ell}(1-\mu)$ 记为超出α^{CO}的额外规划者修正。于是，对于 $k > 0$

(i) α^{NE}是内部解，当且仅当 $\ell/N < s < k + \ell/N$；注意，$N > N^*$等价于 $s > \ell/N$。

(ii) α^{CO} 是内部解，当且仅当 $\ell < s < k + \ell$;

(iii) $\alpha^{SP}(\mu)$ 是内部解，当且仅当 $\ell + \delta(\mu) < s < k + \ell + \delta(\mu)$。

每个区间的宽度均为 k。由于对于 $\mu > 0$，$\ell/N < \ell < \ell + \delta(\mu)$，这三个窗口逐渐向右移动：在纳什行为下，内部自动化在成本节约最低时出现；在合作下，在成本节约处于中间水平时出现；而在社会规划者情形下，只有在成本节约最高时才会出现。

引理的证明 *1*。每个断言都可通过检查原始公式何时严格落在 (0,1) 中得到证明。

(i) 纳什均衡。根据命题 1(i)，$\alpha^{NE} = (s - \ell/N)/k$。原始公式 $(s - \ell/N)/k$严格为正当且仅当 $s > \ell/N$（即当 $N > N^*$ 时），且严格小于 1 当且仅当 $s < k + \ell/N$。

（*ii*）合作最优。根据命题1(ii)，$\alpha^{CO} = (s - \ell)/k$。原始表达式$(s - \ell)/k$严格大于零，当且仅当 $s > \ell$；严格小于一，当且仅当 $s < k + \ell$。

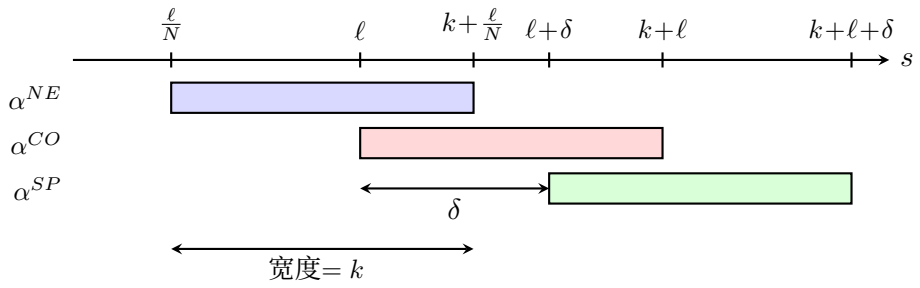


图 3： α^{NE}、 α^{CO}和 $\alpha^{SP}(\mu)$的内部区域随成本节约 s变化的情况，其中$\delta = \mu\ell/[\lambda(1-\mu)]$。每条阴影横条覆盖了相应自动化率严格处于内部解时 s 的取值范围。三个区间的宽度均为 k；随着各目标函数所考虑的需求损失项逐渐增多，这些区间依次向右平移。

（*iii*）社会规划者。根据命题2(i)， $\alpha^{SP}(\mu) = (s-\ell)/k - \mu\ell/[\lambda(1-\mu)k]$。合并各项后， $\alpha^{SP}(\mu) = \big(s-\ell-\delta(\mu)\big)/k$，严格大于零，当且仅当 $s > \ell + \delta(\mu)$；严格小于一，当且仅当 $s < k + \ell + \delta(\mu)$。

最后， $\ell/N < \ell$ 成立，因为 $N > N^* \geq 1$ （事实上，在任何策略环境中都有 $N \geq 2$ ），并且对于 $\mu > 0$ ， $\ell < \ell + \delta(\mu)$，因为 $\delta(\mu) > 0$。 □

推论的证明*1* （无摩擦极限，命题*1*的推论）。

（i）当 $k = 0$时，企业 i从（6）式得到的利润变为

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s - \frac{\ell}{N}\right) - \frac{\ell}{N}\sum_{j\neq i}\alpha_j\right].$$

这是关于 α_i的仿射函数，斜率为 $L(s-\ell/N)$。关键是，该斜率不取决于竞争对手的选择 α_{-i} —— 竞争对手的自动化会影响企业 i的利润水平（通过最后一项），但不会影响 α_i的边际收益。因此，最优的 α_i 不取决于 α_{-i}，使其成为占优策略。

当 $N > N^*$（等价地， $s > \ell/N$）时，斜率严格为正。由于 π_i 在 $[0,1]$, 上关于 α_i 是线性递增的，因此唯一最优值位于上边界 $\alpha_i^* = 1$。这对所有企业同时成立，因此对所有 i ， $\alpha_i = 1$ 是唯一的纳什均衡。

（ii）另外假设成本节约小于需求损失： $s < \ell$。当 $k = 0$时， （5）中的总利润关于 $\bar{\alpha}$是线性的： $\sum_i \pi_i = N\Pi_0 + NL(s-\ell)\bar{\alpha}$。由于 $s - \ell < 0$，该利润关于 $\bar{\alpha}$严格递减，因此利润最大化的合作结果是 $\bar{\alpha} = 0$（不自动化），此时单个企业利润为 Π_0。

在纳什均衡（$\bar{\alpha}=1$）下，单个企业利润为 $\Pi_0+L\,(s-\ell)$。由于 $s-\ell<0$，我们有 $\Pi_0+L\,(s-\ell)$：每个企业获得的利润都严格低于合作时的利润。单个企业的利润损失为 $\Pi_0-[\Pi_0+L\,(s-\ell)\,]=L\,(\ell-s)$。对于 N家企业，总无谓损失为 $NL\,(\ell-s)$。对于需求损失，根据方程（2）：

$$D=A+\lambda wLN-\ell LN\,\bar{\alpha}.$$

在 $\bar{\alpha}=0$: $D^{CO}=A+\lambda wLN$。在 $\bar{\alpha}=1$: $D^{NE}=A+\lambda wLN-\ell LN$。因此 $D^{CO}-D^{NE}=\ell LN$。

□

命题的证明2。

(i) 规划者 μ-选择 $\bar{\alpha}$ 以最大化 $S(\mu)\;=\mu\,\mathcal{W}+(1-\mu)\,\mathcal{K}$，其依据是（8）。我们依次计算各个导数。回想一下，$\mathcal{W}=wLN[1-(1-\eta)\bar{\alpha}]$，因此

$$\frac{d\mathcal{W}}{d\bar{\alpha}}=-wNL(1-\eta). \tag{15}$$

由于 $\ell=\lambda(1-\eta)w$，这等于 $-\ell\,NL/\lambda$。回想一下，$\mathcal{K}=N[\Pi_0+L((s-\ell)\bar{\alpha}-\frac{k}{2}\bar{\alpha}^2)]$（例如，方程（14））$,$ 因此

$$\frac{d\mathcal{K}}{d\bar{\alpha}}=NL\big[(s-\ell)-k\bar{\alpha}\big]. \tag{16}$$

合并后:

$$\frac{dS}{d\bar{\alpha}}=\mu\,\frac{d\mathcal{W}}{d\bar{\alpha}}+(1-\mu)\,\frac{d\mathcal{K}}{d\bar{\alpha}}=-\frac{\mu\,\ell\,NL}{\lambda}+(1-\mu)\,NL\big[(s-\ell)-k\bar{\alpha}\big].$$

令其等于零并除以 $NL>0$:

$$(1-\mu)\big[(s-\ell)-k\bar{\alpha}\big]=\frac{\mu\,\ell}{\lambda}.$$

两边同时除以 $(1-\mu)\;>0\,$，并整理出 $k\bar{\alpha}$:

$$(s-\ell)-k\bar{\alpha}=\frac{\mu\,\ell}{\lambda(1-\mu)},\qquad\Longrightarrow\qquad k\bar{\alpha}=(s-\ell)-\frac{\mu\,\ell}{\lambda(1-\mu)}.$$

两边除以 k，可得

$$\alpha^{SP}(\mu)=\frac{s-\ell}{k}-\frac{\mu\,\ell}{\lambda\cdot(1-\mu)\cdot k} \tag{17}$$

只要满足 $0 \le \alpha^{SP} \le 1$，这个最优值就有效。要确定界限，令 $\bar{\mu}$ 为满足 $\alpha^{SP}(\bar{\mu}) = 0$ 的值。然后，解出 $\bar{\mu}$，可得

$$s - \ell = \frac{\bar{\mu}\ell}{\lambda \cdot (1 - \bar{\mu})}$$

$$\Downarrow$$

$$\bar{\mu} = \frac{\lambda(s - \ell)}{\ell + \lambda(s - \ell)}$$

由方程（17），我们知道 α^{SP} 关于 μ递减，因此如果 $\mu > \bar{\mu}$，那么方程（17）将为负值。类似地，将 μ 设为 $\alpha^{SP}(\mu) = 1$所对应的值，我们有

$$s - \ell - k = \frac{\underline{\mu}\ell}{\lambda \cdot (1 - \underline{\mu})}$$

$$\Downarrow$$

$$\underline{\mu} = \frac{\lambda(s - \ell - k)}{\ell + \lambda(s - \ell - k)}$$

那么对于 $\mu \le \underline{\mu}$，方程（17）大于 1。

（ii）回顾

$$\begin{aligned}
S &= \mu \cdot \mathcal{W} + (1 - \mu) \cdot \mathcal{K} \qquad \text{(eq. (8))} \\
&= \mu \cdot wLN\left[1 - (1 - \eta)\bar{\alpha}\right] + (1 - \mu) \cdot N\left[\Pi_0 + L\left((s - \ell)\bar{\alpha} - \tfrac{k}{2}\bar{\alpha}^2\right)\right] \qquad \text{(eqs. (8) and (14))} \\
&= N\left[\mu\, wL\left(1 - (1 - \eta)\bar{\alpha}\right) + (1 - \mu)\left(\Pi_0 + L\left((s - \ell)\bar{\alpha} - \tfrac{k}{2}\bar{\alpha}^2\right)\right)\right] \\
&= \underbrace{N\left[\mu\, wL + (1 - \mu)\Pi_0\right]}_{a} + \underbrace{NL\left(-\mu\,\frac{\ell}{\lambda} + (1 - \mu)(s - \ell)\right)}_{b}\bar{\alpha} + \underbrace{\left(-\frac{(1 - \mu)NLk}{2}\right)}_{\frac{1}{2}\gamma}\bar{\alpha}^2
\end{aligned}$$

在线性项中，我们使用了 $w(1 - \eta) = \ell/\lambda$，来自（3）。这是一个二次函数 $S = a + b\,\bar{\alpha} + \frac{1}{2}\gamma\,\bar{\alpha}^2$，其 $\gamma = -(1 - \mu)\,NLk < 0$严格为负。在最大值 $x^* = -b/\gamma$附近配方：

$$f(x) = a + bx + \tfrac{1}{2}\gamma\, x^2 = f(x^*) + \tfrac{1}{2}\gamma\,(x - x^*)^2,$$

因此

$$f(x^*) - f(x) = -\tfrac{1}{2}\gamma\,(x - x^*)^2 = \tfrac{1}{2}|\gamma|\,(x - x^*)^2,$$

其中 la st 等式使用 $\gamma \le 0$。将其应用于 $x^* = \alpha^{SP}(\mu)$、$x = \alpha^{NE}$ 和 $|\gamma| =$

$(1-\mu)\ NLk$:

$$S(\alpha^{SP}) - S(\alpha^{NE}) = \frac{(1-\mu)\,NLk}{2}\,[\alpha^{NE} - \alpha^{SP}(\mu)]^2.$$

(iii) 回顾（9），楔子可分解为:

$$\alpha^{NE} - \alpha^{SP}(\mu) = \frac{\ell(1-1/N)}{k} + \frac{\mu\,\ell}{\lambda(1-\mu)\,k}.$$

第一项 $\ell(1-1/N)/k$ 严格为正，因为 $\ell > 0$ 且 $N \geq 2$。第二项 $\mu\,\ell/[\lambda(1-\mu)\,k]$ 对任意 $\mu \in (0,\ 1)$ 都严格为正，并在 $\mu = 0$ 时等于零。因此，对于所有 $\mu \in [0, 1)$，$\alpha^{NE} > \alpha^{SP}(\mu) \geq \alpha^{CO}$。

对于工人收入：由方程（15）可知，$\mathcal{W} = wLN[1-(1-\eta)\bar{\alpha}]$ 是关于 $\bar{\alpha}$ 的仿射函数且严格递减（因为 $(1-\eta) > 0$ 在假设 $\eta < 1$下成立）。因此，$\alpha^{NE} > \alpha^{CO}$意味着 $\mathcal{W}(\alpha^{NE}) < \mathcal{W}(\alpha^{CO})$。

对于所有者剩余：由方程（16）可知，$d\mathcal{K}/d\bar{\alpha} = NL[(s-\ell) - k\bar{\alpha}]$，其在 $\bar{\alpha} = (s-\ell)/k = \alpha^{CO}$ 处等于零，并且当 $\bar{\alpha} > \alpha^{CO}$ 时严格为负（因为 $k > 0$）。因此，$\mathcal{K}$ 在 $(\alpha^{CO}, \alpha^{NE}]$上严格递减，所以 $\mathcal{K}(\alpha^{NE}) < \mathcal{K}(\alpha^{CO})$。

由于在纳什均衡下，$\mathcal{W}$ 和 $\mathcal{K}$ 均严格低于合作最优水平，因此从 α^{CO} 转向 α^{NE}无法使任何一类变得更好：纳什均衡处于帕累托劣势。

□

推论2的证明（外部性的符号）。楔子公式 $\alpha^{NE} - \alpha^{CO} = \ell\ (1-1/N)\ /k$已在命题 1(iii)中针对 $\ell = \lambda\ (1-\eta)\ w$ 和 $s > \ell$得到确立。三种符号情形由 ℓ的符号决定：$\eta < 1$ 得到 $\ell > 0$（正楔子），$\eta = 1$ 得到 $\ell = 0$（零楔子，且两个比率均降至 s/k），而 $\eta > 1$ 得到 $\ell < 0$（负楔子）。当 $\eta = 0$时，$\ell = \lambda w$ 达到最大值，因此楔子$\lambda w\ (1-1/N)\ /k$ 也达到最大值。注意，$\eta > 1$ 意味着 $\ell < 0 < s$，因此内部条件 $s > \ell$自动成立。

□

命题的证明 *2*（劳动者股权）。采用利润分享 ϵ时，总需求满足 s

$$D = A + \lambda\big[\underbrace{wLN(1-(1-\eta)\bar{\alpha})}_{\text{wage income}} + \epsilon\,\Pi\big],$$

其中总利润为 $\Pi = D - \sum_i C_i = D - NL(w - s\bar{\alpha} + \frac{k}{2}\bar{\alpha}^2)$。将 Π 的表达式代入

需求方程，并将 D 项移至左侧：

$$D(1-\lambda\epsilon) = A + \lambda\, wLN(1-(1-\eta)\bar{\alpha}) - \lambda\epsilon\, NL\big(w - s\bar{\alpha} + \tfrac{k}{2}\bar{\alpha}^2\big)\,.$$

展开右侧，并按 $\bar{\alpha}$ 的幂次合并：

$$D = \frac{A + \lambda(1-\epsilon)wLN - \ell_\epsilon LN\bar{\alpha} - \lambda\epsilon\frac{k}{2}NL\bar{\alpha}^2}{1-\lambda\epsilon},$$

其中 $\ell_\epsilon := \ell - \lambda\epsilon s = \lambda(1-\eta)w - \lambda\epsilon(w-c)$。

(i) 合作最优。规划者通过选择对称策略组合下的共同自动化率 α，使总利润 $\Pi = D - \sum_i C_i$ 最大化。对 α 求 D（由上式得出）的导数：

$$\frac{\partial D}{\partial \alpha} = \frac{-\ell_\epsilon\, LN - \lambda\epsilon kNL\alpha}{1-\lambda\epsilon}.$$

对总成本 $\sum_i C_i = NL$ 求导（$w - s\alpha + \frac{k}{2}\alpha^2$）：

$$\frac{\partial(\sum_i C_i)}{\partial \alpha} = NL(-s + k\alpha).$$

令 $\partial\Pi/\partial\alpha = \partial D/\partial\alpha - \partial(\sum_i C_i)/\partial\alpha = 0$，并重新整理：

$$NL(s - k\alpha) = \frac{\ell_\epsilon NL + \lambda\epsilon kNL\alpha}{1-\lambda\epsilon}.$$

两边同时乘以（$1-\lambda\epsilon$）：

$$(s-k\alpha)(1-\lambda\epsilon) = \ell_\epsilon + \lambda\epsilon k\alpha.$$

展开左侧并合并 α 项：

$$k\alpha \underbrace{\big[(1-\lambda\epsilon) + \lambda\epsilon\big]}_{=1} = s(1-\lambda\epsilon) - \ell_\epsilon.$$

代入 $\ell_\epsilon = \ell - \lambda\epsilon s$后，右侧变为 $s - s\lambda\epsilon - \ell + \lambda\epsilon s = s - \ell$。因此 $k\alpha = s - \ell$，得到 $\alpha^{CO}(\epsilon) = (s-\ell)/k$，与 ϵ无关。

(ii) 纳什均衡。竞争性定价给出 $\mathrm{Rev}_i = D/N$。所有者最大化 $(1-\epsilon)\pi_i$；由于 $(1-\epsilon) > 0$，一阶条件简化为 $\partial\pi_i/\partial\alpha_i = 0$。由于 D 仅通过 $\bar{\alpha}$（以及 $\bar{\alpha}^2$项）依赖于 α_i，且 $\partial\bar{\alpha}/\partial\alpha_i = 1/N$，对 α_i求 $\mathrm{Rev}_i = D/N$的导数，并在

对称策略组合处求值得到

$$\frac{\partial \mathrm{Rev}_i}{\partial \alpha_i} = \frac{1}{N}\cdot\frac{\partial D}{\partial \alpha_i} = \frac{-\ell_\epsilon L - \lambda\epsilon k L\alpha}{N(1-\lambda\epsilon)}.$$

结合成本导数 $\partial C_i/\partial\alpha_i = L(-s+k\alpha)$，对称一阶条件 $\partial \mathrm{Rev}_i/\partial\alpha_i - \partial C_i/\partial\alpha_i = 0$ 变为：

$$s - k\alpha - \frac{\ell_\epsilon + \lambda\epsilon k\alpha}{N(1-\lambda\epsilon)} = 0.$$

定义 $N_\epsilon := N(1-\lambda\epsilon) + \lambda\epsilon = N - \lambda\epsilon(N-1)$。将 α 项移至左侧：

$$k\alpha \underbrace{\left[\frac{N(1-\lambda\epsilon)+\lambda\epsilon}{N(1-\lambda\epsilon)}\right]}_{=N_\epsilon/[N(1-\lambda\epsilon)]} = s - \frac{\ell_\epsilon}{N(1-\lambda\epsilon)}.$$

两边同时乘以 $N(1-\lambda\epsilon)/N_\epsilon$：

$$k\alpha = \frac{s\,N(1-\lambda\epsilon) - \ell_\epsilon}{N_\epsilon}.$$

将 $\ell_\epsilon = \ell - \lambda\epsilon s$ 代入分子：

$$s\,N(1-\lambda\epsilon) - (\ell - \lambda\epsilon s) = s\big[N(1-\lambda\epsilon)+\lambda\epsilon\big] - \ell = s\,N_\epsilon - \ell.$$

因此 $k\alpha = (sN_\epsilon - \ell)/N_\epsilon = s - \ell/N_\epsilon$，得到 $\alpha^{NE}(\epsilon) = (s - \ell/N_\epsilon)/k$。

(iii) 楔子。用纳什率减去合作最优：

$$\alpha^{NE}(\epsilon) - \alpha^{CO} = \frac{s-\ell/N_\epsilon}{k} - \frac{s-\ell}{k} = \frac{\ell - \ell/N_\epsilon}{k} = \frac{\ell(N_\epsilon - 1)}{k\,N_\epsilon}.$$

由于 $N_\epsilon = N - \lambda\epsilon(N-1)$，我们有 $N_\epsilon - 1 = (N-1) - \lambda\epsilon(N-1) = (N-1)(1-\lambda\epsilon)$，因此楔子变为

$$W(\epsilon) = \frac{\ell(N-1)(1-\lambda\epsilon)}{k\,N_\epsilon}.$$

为证明楔子关于 ϵ 严格递减，写作 $W = \ell(N-1)(1-\lambda\epsilon)/[k(N-\lambda\epsilon(N-1))]$，并应用商法则。分子为 $f(\epsilon) = 1-\lambda\epsilon$，其中 $f' = -\lambda$；分母为 $\tilde{g}(\epsilon) = N - \lambda\epsilon(N-1)$，其中 $\tilde{g}' = -\lambda(N-1)$。根据商法则：

$$\frac{d}{d\epsilon}\left(\frac{f}{\tilde{g}}\right) = \frac{f'\tilde{g} - f\tilde{g}'}{\tilde{g}^2} = \frac{-\lambda[N-\lambda\epsilon(N-1)] + \lambda(N-1)(1-\lambda\epsilon)}{N_\epsilon^2} = \frac{-\lambda}{N_\epsilon^2} < 0.$$

因此 $\partial W/\partial\epsilon = -\lambda\ell(N-1)/(kN_\epsilon^2) < 0$：楔子严格递减。令 $W=0$，则需要满足 $1-\lambda\epsilon = 0$，即 $\epsilon = 1/\lambda$；当 $\lambda \le 1$ 时，$\epsilon \ge 1$ 成立。 □

推论 *3* （自愿利润分享）。企业 i 选择 $\epsilon_i \in [0,1]$ 以最大化留存利润 $(1-\epsilon_i)\,\pi_i$。对 ϵ_i 求导：

$$\frac{d}{d\epsilon_i}\big[(1-\epsilon_i)\pi_i\big] = -\pi_i + (1-\epsilon_i)\frac{\partial \pi_i}{\partial \epsilon_i}.$$

当企业 i 将其利润的 ϵ_i 分给工人时，工人会收到 $\epsilon_i\pi_i$，并在该行业支出 $\lambda\epsilon_i\pi_i$。企业 i获得由此带来的需求增长的 $1/N$，因此 $\partial\pi_i/\partial\epsilon_i = \lambda\pi_i/N$。代入 $\epsilon_i = 0$进行计算：

$$-\pi_i + \frac{\lambda\pi_i}{N} = \pi_i\left(\frac{\lambda}{N} - 1\right) < 0,$$

由于 $\lambda \le 1$且 $N \ge 2$。导数在左端点处为负，因此无论其他企业如何选择，$\epsilon_i = 0$ 都是最优的。

□

命题 *4* 的证明。将 N 家企业划分为一个联盟 $\mathcal{M}$，其中包含 M 家企业，以及一个由 $N-M$家非成员企业组成的外围。令

$$\alpha^F := \frac{1}{N-M}\cdot\sum_{i\notin\mathcal{M}}\alpha_i \tag{18}$$

为外围企业的平均自动化率。在所有联盟成员均选择 α^M的对称策略组合下，平均自动化率为

$$\bar{\alpha} = \frac{M\alpha^M + (N-M)\alpha^F}{N}.$$

根据（6），联盟成员 i的利润为

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s - \frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\sum_{j\neq i}\alpha_j\right].$$

Then

$$\begin{aligned}\pi_i &= \Pi_0 + L\left[\alpha_i\left(s-\frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\left(\sum_{j\in\mathcal{M}\setminus\{i\}}\alpha_j + \sum_{j\notin\mathcal{M}}\alpha_j\right)\right]\\ &= \Pi_0 + L\left[\alpha_i\left(s-\frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\left(\sum_{j\in\mathcal{M}\setminus\{i\}}\alpha_j + (N-M)\alpha^F\right)\right].\end{aligned}$$

在对称策略下，联盟中的所有企业都采用相同的策略 $\alpha_i = \alpha^M$，我们

have

$$\begin{aligned}\pi_i &= \Pi_0 + L\left[\alpha^M\left(s-\frac{\ell}{N}\right)-\frac{k}{2}\left(\alpha^M\right)^2-\frac{\ell}{N}\left((M-1)\alpha^M+(N-M)\alpha^F\right)\right]\\ &= \Pi_0 + L\left[\alpha^M\left(s-\frac{\ell M}{N}\right)-\frac{k}{2}\left(\alpha^M\right)^2+\frac{\ell}{N}(N-M)\alpha^F\right]\\ &= \Pi_0 + L\frac{\ell(N-M)\alpha^F}{N}+L\left[\alpha^M\left(s-\frac{\ell M}{N}\right)-\frac{k}{2}\left(\alpha^M\right)^2\right]\end{aligned}$$

联盟选择 α^M 以最大化 $\sum_{i\in\mathcal{M}}\pi_i$。在对称均衡下， $\sum_{i\in\mathcal{M}}\pi_i = M\pi_i$，因此，最大化联盟中单个企业的利润也会最大化联盟利润。求导可得

$$\frac{\partial \pi_i}{\partial \alpha^M} = L\left(s-\frac{M\ell}{N}-k\alpha^M\right) = L\left(s-\frac{\ell}{N}-k\alpha^M-\frac{(M-1)\ell}{N}\right)$$

额外项 $-(M-1)\ell/N$ （相较于方程（13））之所以出现，是因为每位联盟成员都会将其自动化给其他 $M-1$ 成员造成的需求损失纳入考虑，而每位成员的收入都会减少 $\ell L/N$。令一阶条件等于零:

$$\alpha^M = \frac{s-M\ell/N}{k},$$

限制为 $[0,1]$。在 $M=1$时，这可简化为 $\alpha^{NE}=(s-\ell/N)/k$。在 $M=N$时，可简化为 $\alpha^{CO}=(s-\ell)/k$。

当 α^M 和 α^{CO} 均为内部解时，剩余楔子为

$$\alpha^M-\alpha^{CO} = \frac{s-M\ell/N}{k}-\frac{s-\ell}{k} = \frac{\ell(1-M/N)}{k},$$

对于 $M<N$，该值严格为正，且仅在 $M=N$时为零。 □

命题 *5*的证明（庇古式自动化税）。

（i）征税后，企业 i从（6）获得的利润变为

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s-\tau-\frac{\ell}{N}\right)-\frac{k}{2}\alpha_i^2-\frac{\ell}{N}\sum_{j\neq i}\alpha_j\right].$$

一阶条件为 $s-\tau-\ell/N-k\alpha_i=0$，得到 $\alpha^{NE}(\tau)=(s-\tau-\ell/N)/k$。令$\alpha^{NE}(\tau)=\alpha^{CO}=(s-\ell)/k$，可得 $\tau^*=\ell-\ell/N=\ell(1-1/N)$。

(ii) 在 $\tau = \tau^*$ 时，所有企业都选择 α^{CO}。单个企业利润为 $\pi^{\text{tax}} = \pi(\alpha^{CO})$ $-\tau^* L\alpha^{CO} = \pi^{CO} - \tau^* L\alpha^{CO}$。税收总收入为 $\tau^* LN\alpha^{CO}$；平均返还后，每家企业获得 $\tau^* L\alpha^{CO}$，利润恢复至 π^{CO}。

□

命题 *6*（人工智能生产率扩大了过度自动化楔子）。

¯（i）在对称策略组合 $\alpha_i = \alpha$ 下，若 $\phi > 1$，企业 i的收入为 收入$_i = DY_i/(NY)$。企业的一阶条件是自动化的边际收益等于其边际成本：∂ 收入$_i/\partial\alpha_i + sL - kL\alpha_i = 0$，其中 sL 表示用人工智能替代一名工人所带来的单位成本节约，$kL\alpha_i$ 表示边际整合摩擦。代入式（10）中的边际收入并整理后，对称的一阶条件为

$$k\alpha = s - \frac{\ell}{N} + \frac{D(\alpha)\,(\phi-1)(N-1)}{N^2[1+(\phi-1)\alpha]\,L}.$$

定义 LHS$(\alpha) = k\alpha$，并令 RHS(α) 表示右侧。左侧严格递增（斜率为 k）。右侧严格递减：市场份额项的分子与 $D(\alpha) = A + \lambda wLN - \ell LN\alpha$ 成正比（随 α 增大而递减），分母因子 $1+(\phi-1)\alpha$（随 α 增大而递增）。因此，LHS $=$ RHS 有唯一解。

为证明 $\alpha^{NE}(\phi) > \alpha^{NE}(1)$，在基准均衡 $\alpha = \alpha^{NE}(1) = (s - \ell/N)/k$ 处分别计算两边：左侧为 $= s - \ell/N$，而右侧为 $= s - \ell/N +$（正的市场份额项）$>$ 左侧。由于左侧递增、右侧递减，唯一交点必然位于 $\alpha^{NE}(\phi) > \alpha^{NE}(1)$。

（ii）合作规划者最大化 $\sum_i \pi_i = D - \sum_i C_i$。总收入等于总需求 D，无论产出如何在各企业之间分配：支出 D 由（2）确定，因此重新分配产出份额不会改变总收入。由此，规划者的一阶条件仅取决于成本：

$$\frac{\partial(\sum_i \pi_i)}{\partial \alpha_i} = -\ell L + sL - kL\alpha_i,$$

w其与 ϕ 无关。令其等于零可得 $\alpha^{CO}(\phi) = (s-\ell)/k = \alpha^{CO}(1)$.

对于广义规划者：$S(\mu) = \mu\mathcal{W} + (1-\mu)\,\mathcal{K}$。工人收入 $\mathcal{W} = wLN[1-(1-\eta)\bar{\alpha}]$不取决于 ϕ。在对称策略组合下，所有者剩余 $\mathcal{K} = D - \sum_i C_i$，且 $D(2)$ 和 $C_i(1)$ 都不取决于 ϕ。因此，在每个对称的 $\bar{\alpha}$ 下，$S(\mu)$ 都具有 ϕ 不变性，而 $\alpha^{SP}(\mu;\,\phi) =$

对所有 μ，$\alpha^{SP}(\mu; 1)$。

(iii) 结合 (i) 和 (ii)：$\alpha^{NE}(\phi) > \alpha^{NE}(1)$，而对于所有 μ，$\alpha^{SP}(\mu; \phi) = \alpha^{SP}(\mu; 1)$，因此楔子 $\alpha^{NE}(\phi) - \alpha^{SP}(\mu; \phi)$ 对每个 μ 都严格大于 $\alpha^{NE}(1) - \alpha^{SP}(\mu; 1)$。由于 $\alpha^{NE}(\phi)$ 随 ϕ 递增（采用相同的左边/右边论证，但市场份额项更大），楔子也随 ϕ严格递增。

□

命题7 的证明（内生进入、无摩擦基准）。我们假设 $k = 0$ （无摩擦情形），$\lambda = 1$ （完全回收），$0 < \kappa < A$ （进入需要成本，但市场至少能容纳一家企业），且 $\ell > s$ （因此 $N^* > 1$ 根据推论1）。

步骤1：利润函数在 $\mathbb{N}$上严格递减。令 $m = \lfloor N^* \rfloor$，则 $m \le N^* < m + 1$。对于 $N \le N^*$，推论1给出 $\alpha = 0$，因此由 (5) 和 $\lambda = 1$可得：$\Pi^*(N) = A/N$，它严格递减。对于 $N > N^*$，完全自动化是占优策略（推论1），且单个企业利润减少 $\Delta = L(\ell - s) > 0$，因此 $\Pi^*(N) = A/N - \Delta$，同样严格递减。在交点处：$\Pi^*(m) = A/m > A/(m+1) > A/(m+1) - \Delta = \Pi^*(m+1)$。因此 Π^* 在 $\mathbb{N}$上严格递减。

步骤 2：N^{FE}的存在性与唯一性。由于 $\kappa < A$, $\Pi^*(1) = A > \kappa$。由于 $\Delta > 0$,$\Pi^*(N) \to -\Delta < 0$作为 $N \to \infty$。因此，集合 $\mathcal{S} = \{N \in \mathbb{N} : \Pi^*(N) \ge \kappa\}$非空且有限。令 $N^{FE} = \max \mathcal{S}$。根据严格单调性，$\Pi^*(N^{FE}) \ge \kappa$且 $\Pi^*(N^{FE} + 1) < \kappa$，因此 N^{FE}满足（11），且具有唯一性。

步骤 3：按机制刻画。我们通过检查哪些整数属于 $\mathcal{S}$ 来确定 N^{FE}。

在无自动化分支（$N \le m$）上：$\Pi^*(N) = A/N \ge \kappa$ 当且仅当 $N \le A/\kappa$。因此，该分支上的可行整数为 $\{1, \ldots, \min(\lfloor A/\kappa \rfloor, m)\}$。

在完全自动化分支（$N \ge m + 1$）上：当且仅当 $N \le A/(\kappa + \Delta)$ 时，$\Pi^*(N) = A/N - \Delta \ge \kappa$。该分支上的最小整数为 $m + 1$，因此存在可行整数当且仅当 $m + 1 \le A/(\kappa + \Delta)$，即 $\kappa + \Delta \le A/(m + 1)$。当集合非空时，可行集合为 $\{m + 1, \ldots, \lfloor A/(\kappa + \Delta) \rfloor\}$。

由于 N^{FE} 是整体上最大的可行整数：

情形（*i*）：（低进入成本）$\kappa + \Delta \le A/(m + 1)$。完全自动化分支包含不超过 $\lfloor A/(\kappa + \Delta) \rfloor \ge m + 1$ 的可行整数。由于 $\kappa + \Delta > \kappa$，我们有 $A/(\kappa + \Delta) < A/\kappa$，因此完全自动化分支的最大值在绝对值上不超过无自动化分支的最大值，但超过 m（无自动化分支的上限）。因此 $N^{FE} = \lfloor A/(\kappa + \Delta) \rfloor \ge m + 1 > N^*$，且每家企业都完全自动化。见图 4（a）。

情形（*ii*）：（中等进入成本）$\kappa + \Delta > A/(m + 1)$ 且 $\kappa < A/m$。第一个条件意味着 $A/(m + 1) - \Delta < \kappa$，因此 $\Pi^*(m + 1) = A/(m + 1) - \Delta < \kappa$。由于 Π^* 递减，不存在更大的整数

以上的 N^* 是可行的。第二个条件给出 $A/\kappa > m$，因此 $\lfloor A/\kappa \rfloor \geq m$。因此，无自动化分支上最大的可行整数是 m，并且 $N^{FE} = m$。利润为 $\Pi^*(m) = A/m > \kappa$（严格成立，因为 $A/\kappa > m$意味着 $A/m > \kappa$）。在此情形下，由于 $N^{FE} = m < N^*$，没有企业会采用自动化。见图4（b）。

情形（*iii*）：（高进入成本） $\kappa \geq A/m$。那么 $A/\kappa \leq m$，所以 $\lfloor A/\kappa \rfloor \leq m$。由于 $\kappa + \Delta > \kappa \geq A/m > A/(m+1)$，完全自动化分支上没有整数可行（与情形（ii）的论证相同）。因此 $N^{FE} = \lfloor A/\kappa \rfloor$。由于 $N^{FE} = \lfloor A/\kappa \rfloor \leq m \leq N^*$，没有企业会自动化。见图4(c)。

这三种情形涵盖了所有 $\kappa \in$(0<style id="2">, A)。当 $\kappa > A$ 时，则 $N^{FE} = 0$。 □

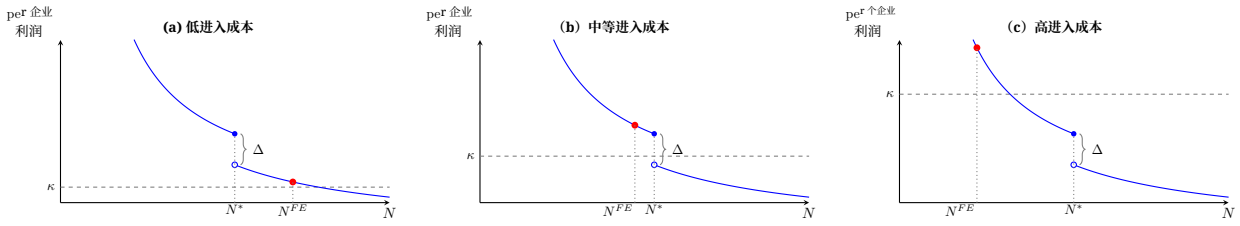


图4：无摩擦基准（$k = 0$，$\lambda = 1$）中单个企业利润随企业数量的变化。在每个面板中，利润在 $N \leq N^*$（不自动化）时遵循 A/N，并在完全自动化变为占优策略的 N^* 处按$\Delta = L\,(\ell - s)$ 离散下降。红点标记自由进入均衡 N^{FE}。（a）低进入成本： N^{FE}高于 N^*，所有企业都进行自动化。（b）中等进入成本： $N^{FE} = \lfloor N^* \rfloor$；自动化威胁遏制了边际进入者，使正利润得以维持，同时没有发生任何自动化。（c）高进入成本： N^{FE}远低于 N^*；仅进入成本就限制了竞争，自动化阈值始终未被接近。

命题 *8*（凸成本下的内生进入）。首先，我们证明单个企业利润随 N递减。根据方程（5），我们可以写成

$$\pi^{NE}(N) = A/N + C + g(N) \tag{19}$$

其中

$$C := (\lambda - 1)wL \tag{20}$$

$$g(N) := L\,\alpha^{NE}(N)\left[(s - \ell) - \tfrac{k}{2}\alpha^{NE}(N)\right] \tag{21}$$

由于 A/N 在 $\mathbb{N}$上严格递减，只需证明 g 在 $N \in \mathbb{N}$上单调不增。首先注意到， g(·)随 α^{NE}递减

$$\frac{dg}{d\alpha^{NE}} = (s - \ell) - k\,\alpha^{NE} \tag{22}$$

根据假设 $\ell > s$，因此导数始终为负。然后，根据命题 1，我们有

$$\alpha^{NE} = \begin{cases} 0 & \text{if } N < \frac{\ell}{s} \\ \frac{s-\ell/N}{k} & \text{if } \frac{\ell}{s} \leq N \leq \frac{\ell}{s-k} \\ 1 & \text{if } \frac{\ell}{s-k} < N \end{cases} \tag{23}$$

因此， α^{NE}关于 N单调不减。因此， $g(\alpha^{NE}(N))$关于 N单调不增。

由于 A/N关于 N递减，且 $g\left(\alpha^{NE}(N)\right)$ 关于 N单调不增，因此 $\pi^{NE}(N) = A/N + C + g(N)$ 关于 N严格递减，即 $\pi^{NE}(N) > \pi^{NE}(N+1)$。

现在，令

$$S \coloneqq \{N \in \mathbb{N} : \pi^{NE}(N) \geq \kappa\} \tag{24}$$

由于根据假设， $\pi^{NE}(1) = \Pi_0(1) > \kappa$ ，所以集合 S 非空。要证明 S 是有限的，我们需要证明 $\lim_{N\to\infty} \pi^{NE} < \kappa$。首先，

$$\alpha_\infty \coloneqq \lim_{N\to\infty} \alpha^{NE}(N) = \min\left(1, \frac{s}{k}\right) \tag{25}$$

So

$$\lim_{N\to\infty} g(N) = L\,\alpha_\infty \left[(s-\ell) - \tfrac{k}{2}\alpha_\infty\right] < 0 \tag{26}$$

因为 $s < \ell$ （根据假设）。因此

$$\lim_{N\to\infty} \pi^{NE}(N) = \lim_{N\to\infty} [A/N + C + g(N)] = (\lambda - 1)wL + \lim_{N\to\infty} g(N) < 0 \tag{27}$$

由于 $C = (\lambda - 1)wL < 0$。又因为 $\lim_{N\to\infty} \pi^{NE} < 0 < \kappa$，所以集合 S 是有限的。令

$$N^{FE} \coloneqq \max\left(N \in S\right) \tag{28}$$

那么 $\pi^{NE}(N^{FE}) \geq \kappa$和 $\pi^{NE}(N^{FE}+1) < \kappa$，因此 N^{FE}满足（11）。若 $N^{FE} > N^*$，则 $\alpha^{NE}(N^{FE}) = \min\left((s-\ell/N^{FE})/k, 1\right) > 0$；由于 $\ell > s$意味着 $\alpha^{CO} = \max\{0, (s-\ell)/k\} = 0$，我们有 $\alpha^{NE}(N^{FE}) > \alpha^{CO}$，因此过度自动化持续存在。 □

命题 *9* （内生工资）。对称均衡是一个不动点： $\bar{\alpha}$ 使得 $\bar{\alpha} = \alpha^{NE}$（$w$（$\bar{\alpha}$））。

（i）在对称率 $\alpha_i = \alpha$ 对所有 i，我们有 $\bar{\alpha} = \alpha$ 并且，由（5）可得，

$$\pi_i = \Pi_0 + L\left(s\,\alpha - \ell\,\alpha - \tfrac{k}{2}\alpha^2\right).$$

总利润为 $\sum_i \pi_i = N\,\pi_i$。对 α 求导并除以 NL，得到规划者的企业自动化边际收益：

$$g(\alpha) \coloneqq s(\alpha) - \ell(\alpha) - k\alpha = w(\alpha) - c - \lambda(1-\eta)w(\alpha) - k\alpha = w(\alpha)\big[1-\lambda(1-\eta)\big] - c - k\alpha,$$

其中我们用了 $s = w - c$ 和 $\ell = \lambda(1-\eta)w$。相比之下，根据（6），每家企业的利润仅通过其自身的收入份额 $\bar{\alpha}$取决于 ℓ/N。对 α_i 求导并除以 L，得到私人边际收益：

$$h(\alpha) \coloneqq s(\alpha) - \frac{\ell(\alpha)}{N} - k\alpha = w(\alpha)\left[1 - \frac{\lambda(1-\eta)}{N}\right] - c - k\alpha.$$

求导可得，$g'(\alpha) = w'(\alpha)[1-\lambda(1-\eta)] - k$；第一项非正，且 $k > 0$，因此$g' < 0$。对 h（其系数 $1-\lambda(1-\eta)/N > 0$ 对 $N \geq 2$）也可作同样的论证，因此两者都严格递减。注意，对所有 α，$g(\alpha) = h(\alpha) - \ell(\alpha)(1-1/N)$。在纳什均衡不动点处，根据定义，$h(\alpha^{NE}) = 0$，因此

$$g(\alpha^{NE}) = 0 - \ell(\alpha^{NE})(1-1/N) = -\ell(\alpha^{NE})(1-1/N) < 0.$$

由于 g 严格递减，且 $g(\alpha^{CO}) = 0$（规划者的最优性条件），不等式$g(\alpha^{NE}) < 0 = g(\alpha^{CO})$意味着 $\alpha^{CO} < \alpha^{NE}$。

(ii) 由（7）可知，阈值为 $N^*(w) = \lambda(1-\eta)w/(w-c)$。应用商法则：

$$\frac{\partial N^*}{\partial w} = \frac{\lambda(1-\eta)(w-c) - \lambda(1-\eta)w}{(w-c)^2} = \frac{-\lambda(1-\eta)c}{(w-c)^2} < 0,$$

因此，N^*关于 w 严格递减。由于根据假设，$w'(\bar{\alpha}) \leq 0$，对于所有 $\bar{\alpha} \geq 0$，$w(\bar{\alpha}) \leq w(0)$。结合可得：$N^*(w(\bar{\alpha})) \geq N^*(w(0))$；当 $w(\bar{\alpha}) < w(0)$ 时，不等式严格成立。 □

推论 4 (工资调整下的广义规划者)。 在命题 9的条件下，令 $\mu \in [0,1)$ 并定义 $g_\mu(\alpha) \coloneqq s(w(\alpha)) - \ell(w(\alpha))\big[1 + \mu/(\lambda(1-\mu))\big] - k\alpha$。如果 g_μ 严格递减，则 $\alpha^{NE} > \alpha^{SP}(\mu)$。

证明概要。 论证与命题 9第 (i) 部分相同，只需将合作规划者的边际收益 g 替换为 μ-规划者的边际收益。 μ-规划者最大化 $S(\mu) = \mu\,\mathcal{W} + (1-\mu)\,\mathcal{K}$，并在工资给定的情况下选择共同费率 α。根据命题2的证明，一阶条件为

$$(1-\mu)\big[s(\alpha) - \ell(\alpha) - k\alpha\big] = \frac{\mu\,\ell(\alpha)}{\lambda},$$

可将其改写为 $g_\mu(\alpha)=0$。该函数与私人边际收益相关 $h(\alpha)=s(\alpha)-\ell(\alpha)/N-k\alpha$，根据命题 9 的证明可得

$$g_\mu(\alpha)=h(\alpha)-\ell(\alpha)\left(1-\frac{1}{N}+\frac{\mu}{\lambda(1-\mu)}\right).$$

在纳什均衡处，$h(\alpha^{NE})=0$，因此

$$g_\mu(\alpha^{NE})=-\ell(\alpha^{NE})\left(1-\frac{1}{N}+\frac{\mu}{\lambda(1-\mu)}\right)<0,$$

因为 $\ell>0$、$N\geq 2$ 和 $\mu>0$。在规划者不动点处，根据定义，$g_\mu(\alpha^{SP}(\mu))=0$。如果 g_μ严格递减，则不等式 $g_\mu(\alpha^{NE})<0=g_\mu(\alpha^{SP})$ 意味着 $\alpha^{SP}(\mu)<\alpha^{NE}$，与 $\mu=0$ 情形完全相同。

接下来需要验证单调性。求导可得：

$$g'_\mu(\alpha)=w'(\alpha)\left[1-(1-\eta)\left(\lambda+\frac{\mu}{1-\mu}\right)\right]-k.$$

将方括号内的表达式定义为 $C_\mu:=1-(1-\eta)[\lambda+\mu/(1-\mu)]$。在 $\mu=0$处，$C_0=1-\lambda(1-\eta)\geq 0$，这正是命题 9 的证明中所用的系数；由于 $w'\leq 0$，$w'C_0\leq 0$ 与 $g'_0<0$的乘积可立即得出。对于 $\mu>0$，C_μ 递减。只要 $C_\mu\geq 0$，同样的论证便成立；这对所有 $\mu\leq\bar{\mu}$ 都成立：$=[1-(1-\eta)\lambda]/[2-\eta-(1-\eta)\lambda]$（在 $\lambda=0.5$、$\eta=0.30$处约为 0.48）。当 $C_\mu<0$时，乘积 $w'C_\mu\geq 0$，因此 $g'_\mu<0$ 要求 $k>|w'(\alpha)|\cdot|C_\mu|$：整合摩擦必须超过工资敏感度。

作为一个数值示例，考虑 $w(\bar{\alpha})=1-0.5\,\bar{\alpha}$，$c=0.30$，$\lambda=0.5$，$\eta=0.30$，$k=1$,$N=7$，以及 $\mu=0.3$。均衡率为 $\alpha^{NE}\approx 0.44$，$\alpha^{CO}\approx 0.26$，以及 $\alpha^{SP}(0.3)\approx 0.04$，证实了 $\alpha^{SP}<\alpha^{CO}<\alpha^{NE}$。分配溢价相当可观：$\mu$-规划者会将自动化降至接近零，远低于本就远低于纳什率的合作最优值。

□

命题 *10*（资本收入再循环）的证明。我们先对一般情形 $k\geq 0$推导结果；命题的两个部分由 $k=0$ 特例得出。

总需求为 $D=A+\lambda wLN-\ell LN\bar{\alpha}+\hat{}\eta\Pi$，其中总利润为 $\Pi=D-NL(w-s\bar{\alpha})-\frac{k}{2}L\sum_j\alpha_j^2$。代入并求解 D：

$$D(1-\hat{\eta})=A+(\lambda-\hat{\eta})wLN-\ell_{\hat{\eta}}\,LN\bar{\alpha}-\tfrac{\hat{\eta}k}{2}L\textstyle\sum_j\alpha_j^2,$$

其中 $\ell_{\hat\eta} = \ell - \hat\eta s$。当 $k = 0$ 时，得到 (12)。由于 $N\bar\alpha = \alpha_i + \sum_{j\neq i}\alpha_j$，求导得：

$$\frac{\partial D}{\partial \alpha_i} = -\frac{\ell_{\hat\eta} L}{1-\hat\eta} - \frac{\hat\eta k L}{1-\hat\eta}\,\alpha_i.$$

收入 $\text{Rev}_i = D/N$ 给出 $\partial \text{Rev}_i/\partial\alpha_i = -\ell_{\hat\eta}L/[N(1-\hat\eta)] - \hat\eta k L\alpha_i/[N(1-\hat\eta)]$。企业 i 的边际利润，包括直接成本节约 sL和边际摩擦 $kL\alpha_i$，为

$$\frac{\partial \pi_i}{\partial \alpha_i} = L\left(s - \frac{\ell_{\hat\eta}}{N(1-\hat\eta)} - k\alpha_i \cdot \frac{\hat N}{N(1-\hat\eta)}\right),$$

其中 $N := N(1-\hat\eta) + \hat\eta$。这只取决于 α_i，因此均衡采用严格占优策略。

第 (*i*) 部分： $k = 0$。边际利润可化简为 $L\left(s - \ell_{\hat\eta}/[N(1-\hat\eta)]\right)$，这是一个与 α_i 无关的常数。当且仅当 $sN(1-\hat\eta) > \ell_{\hat\eta}$ 时，该常数为正，即 $N > \frac{\ell_{\hat\eta}}{s(1-\hat\eta)} = N_{\hat\eta}$。因此，当 $N > N_{\hat\eta}$ 时，完全自动化是严格占优策略；当 $N < N_{\hat\eta}$时，不自动化是严格占优策略，这重现了推论1的结构。

(*ii*) $\ell_{\hat\eta} = \ell - \hat\eta s \le 0$ 当且仅当 $\hat\eta \ge \ell/s$。

推广到 $k > 0$。令一阶条件等于零：

$$\alpha^{NE} = \frac{N(1-\hat\eta)}{k\hat N}\left(s - \frac{\ell_{\hat\eta}}{N(1-\hat\eta)}\right) = \frac{s\hat N - \ell}{k\hat N} = \frac{s - \ell/\hat N}{k}.$$

当且仅当 $N > N^*$ 时，该值为正，这重现了命题 1，其中 N 被 N 取代。在对称策略组合（$\alpha_j = \alpha$ 对所有 j 均成立）下，总利润为 $\Pi = \frac{N}{1-\hat\eta}[\Pi_0 + L(s-\ell)\alpha - \frac{k}{2}L\alpha^2]$。该 $1/(1-\hat\eta)$ 乘数只会缩放目标函数，不会改变最优解，因此 $\alpha^{CO} =$ 最大值$\{0, (s-\ell)/k\}$，与命题1 相同。

□