

Hattori, Keisuke

Working Paper

The Hidden Value of Human Followers

Suggested Citation: Hattori, Keisuke (2026) : The Hidden Value of Human Followers, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/344041>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Hidden Value of Human Followers^{*}

Keisuke Hattori[†]

School of Business, Aoyama Gakuin University

September 29, 2026

Abstract

Replacing a human follower with AI that supplies more input can reduce team output. With complementary effort, a leader works harder because a human follower observes this effort and responds with greater effort. AI that matches the follower’s contribution removes this reason to work hard, so the leader works less and team output falls. In larger teams, the remaining human followers also work less. Profit-maximizing firms may nevertheless adopt AI that contributes more while the team produces less. Falling AI costs can trigger an output drop at adoption, followed by recovery. Replacing followers can mean discouraging leaders.

Keywords: Artificial intelligence; Team production; Leadership; Strategic complementarity; Automation.

JEL classification: D23; J23; M54; O33.

1 Introduction

Can an AI system replace a human team member without reducing team output if it supplies exactly the same contribution? Experiments show that even a superior AI teammate can lower team performance and human effort (Dell’Acqua et al., 2025). We study the effects of replacing a human follower with AI in a team where efforts are complementary and the follower chooses effort after observing the leader’s.

With a human follower, the leader has both an incentive to contribute directly to team output and an incentive to elicit additional follower effort. Replacing the follower with AI that contributes the same amount removes the latter incentive and puts the human leader in the position of choosing effort in response to the AI’s contribution. In this sense, the roles of the party that commits first and the party that responds are reversed. Leader effort and team output consequently fall. Offsetting this loss requires AI to contribute more than the displaced follower. When the two members are identical apart from their positions as leader and follower, the AI contribution that exactly preserves output equals the leader’s contribution before replacement.

A growing literature examines the effects of replacing human labor with AI and other forms of automation. Within this literature, studies of team production highlight how replacement affects the

^{*}This research was supported by JSPS KAKENHI Grant Number 24K04912.

[†]E-mail: hattori@busi.aoyama.ac.jp

incentives of remaining workers through interactions within the team. Dogan and Yildirim (2022) show that automation can remove incentive mechanisms sustained by interactions among workers. More closely, Cheng et al. (2025) show how AI replacement disrupts peer monitoring in sequential teams. Both studies design compensation to sustain effort. We examine how effort and output adjust under output sharing, identifying the AI contribution needed to preserve output and showing that firms may profitably adopt AI even when team output falls.

A quadratic example endogenizes AI capacity. The firm can profitably choose an AI contribution exceeding the displaced worker's while reducing total output. As AI becomes cheaper, adoption may initially lower output; further cost reductions subsequently restore it. Appendix A shows that replacing one follower reduces every remaining human's effort in a larger team. Appendix B provides a numerical example of output-decreasing adoption with separately optimized output shares. Appendix C shows that equal-contribution replacement causes a larger output loss in teams with a more able leader.

2 Team production

Output, also revenue at a unit price, is $Y = G(e_L, e_F)$, where $e_L, e_F \geq 0$ are the leader's and follower's efforts, respectively. The smooth technology satisfies $G_L, G_F > 0$ and $G_{LF} > 0$, so the members' efforts are complementary. Human member $i \in \{L, F\}$ receives the fixed share $b \in (0, 1/2)$, has effort cost $C_i(e_i)$, and has payoff

$$u_i = bG(e_L, e_F) - C_i(e_i). \quad (1)$$

Costs are smooth, increasing, and strictly convex. Effort is noncontractible. The leader first commits to an effort level; the follower observes this effort and then chooses its own. Assume finite, interior optima, $C_i'' - bG_{ii} > 0$ for each human, and strict concavity of the leader's payoff along the follower's response. These conditions ensure the unique equilibria used below.

Let $R_F(e_L)$ denote the human follower's best response. Its first-order condition and slope are

$$C'_F(R_F) = bG_F(e_L, R_F), \quad R'_F(e_L) = \frac{bG_{LF}}{C''_F - bG_{FF}} > 0. \quad (2)$$

The superscript H denotes the all-human team. Write its equilibrium as (e_L^H, e_F^H) , where $e_F^H = R_F(e_L^H)$, and $Y^H = G(e_L^H, e_F^H)$. The leader's condition is

$$C'_L(e_L^H) = b[G_L(e_L^H, e_F^H) + G_F(e_L^H, e_F^H)R'_F(e_L^H)]. \quad (3)$$

AI replaces the follower's effort with a purchased input $a \geq 0$, installed before the leader commits to effort and maintained during production. The leader's share and technology remain unchanged. The superscript A denotes the team after AI replacement. Let $R_L(a)$ maximize the leader's payoff $u_L^A(e_L; a) \equiv bG(e_L, a) - C_L(e_L)$. Given installed input a , write the equilibrium leader effort as $e_L^A = R_L(a)$ and output as $Y^A(a) = G(R_L(a), a)$. Thus

$$C'_L(R_L(a)) = bG_L(R_L(a), a), \quad R'_L(a) = \frac{bG_{LF}}{C''_L - bG_{LL}} > 0. \quad (4)$$

Technological complementarity is preserved: a larger AI input raises the marginal product of leader

effort through $G_{LF} > 0$.¹

3 Replacing a responsive follower

Proposition 1. *Replacing the human follower with AI input $a = e_F^H$ strictly reduces leader effort and team output: $e_L^A = R_L(e_F^H) < e_L^H$ and $Y^A(e_F^H) < Y^H$.*

Proof. At the former leader effort, equations (3) and (4) imply

$$\left. \frac{\partial u_L^A}{\partial e_L} \right|_{e_L=e_L^H, a=e_F^H} = -bG_F(e_L^H, e_F^H)R'_F(e_L^H) < 0. \quad (5)$$

Strict concavity implies $R_L(e_F^H) < e_L^H$. Since the follower input is unchanged and $G_L > 0$, output falls. $\square$

AI matching the follower's contribution removes the leader's incentive to elicit additional follower effort. The leader best responds to the AI input with lower effort, reducing output. The roles of commitment and response are thus reversed. By continuity, this loss can persist even with greater AI input. In a horizontal team with simultaneous effort choices, equal-contribution replacement leaves the other member's effort unchanged.

Corollary 1. *If $G(x, y) = G(y, x)$ for all $x, y \geq 0$ and $C_L = C_F$, the unique AI contribution preserving human-team output is $\hat{a} = e_L^H > e_F^H$.*

Proof. Symmetry and the follower's first-order condition imply $R_L(e_L^H) = e_F^H$. Hence $Y^A(e_L^H) = G(e_F^H, e_L^H) = Y^H$. Moreover, $(Y^A)' = G_L R'_L + G_F > 0$. Proposition 1 then implies uniqueness and $e_L^H > e_F^H$. $\square$

Under symmetry, setting AI input equal to the former leader's effort makes the human choose the former follower's effort, exchanging their contributions and preserving output. The threshold concerns preserving output, not the firm's adoption decision. It identifies the contribution needed to compensate for the lost human response without specifying a particular production function.

Two extensions under a quadratic specification examine effort-cost differences and larger teams. Appendix C shows that teams with a more able leader, whose effort is less costly, incur larger output losses from equal-contribution replacement, both absolutely and proportionally. Equal-contribution replacement also lowers output in teams with multiple followers: Appendix A shows that replacing one follower reduces leader effort, prompting every remaining human follower to work less.

¹A programmed rule reproducing R_F could preserve the leader's incentive to elicit follower effort. Section 5 discusses why such a rule may not be credible when AI settings remain adjustable.

4 Profitable automation with lower output

A firm, distinct from the leader, now chooses whether to retain the follower or install AI and select a . Reward shares remain fixed. Specialize the technology, effort costs, and AI costs to

$$G(e_L, e_F) = e_L + e_F + \beta e_L e_F, \quad C_i(e) = \frac{c}{2}e^2, \quad K(a) = F + \frac{k}{2}a^2, \quad (6)$$

where $\beta, c, k > 0$ and $F \geq 0$ is paid only upon adoption. Define $\lambda = b\beta/c \in (0, 1/\sqrt{2})$. Backward induction gives

$$e_L^H = \frac{b(1+2\lambda)}{c(1-2\lambda^2)}, \quad e_F^H = \frac{b(1+\lambda)}{c(1-2\lambda^2)}, \quad Y^H = e_L^H + e_F^H + \beta e_L^H e_F^H. \quad (7)$$

The leader's reduced payoff has second derivative $-c(1-2\lambda^2) < 0$. At $a = e_F^H$, leader effort falls by λe_F^H . Corollary 1 implies the required contribution premium $(\hat{a} - e_F^H)/e_F^H = \lambda/(1+\lambda)$, increasing in complementarity. For $b = 1/3$ and $c = \beta = 1$, equal-contribution replacement reduces output by 17.7%; preserving output requires 25% more AI input.

Human-team profit is $\Pi^H = (1-2b)Y^H$. Under AI,

$$Y^A(a) = a + \frac{b}{c}(1+\beta a)^2, \quad \Pi^A(a) = (1-b)Y^A(a) - F - \frac{k}{2}a^2. \quad (8)$$

To ensure a finite optimal AI input, assume $k > k_0 \equiv 2b(1-b)\beta^2/c$, making profit strictly concave in a . Conditional on adoption,

$$a^* = \frac{(1-b)(1+2\lambda)}{k-k_0}, \quad \Pi^{A*} = \frac{b(1-b)}{c} - F + \frac{(1-b)^2(1+2\lambda)^2}{2(k-k_0)}. \quad (9)$$

The resulting leader effort is $e_L^A = R_L(a^*)$. The human team's profit advantage over the leader alone is $\Delta \equiv \Pi^H - b(1-b)/c$. We assume $\Delta > 0$, so the firm retains the human follower when AI is sufficiently costly.

Proposition 2. *AI adoption raises profit and lowers output if and only if $k_Y < k < k_A(F)$, where*

$$k_Y = \frac{c(1-b)}{b}, \quad k_A(F) = k_0 + \frac{(1-b)^2(1+2\lambda)^2}{2(\Delta + F)}. \quad (10)$$

This interval is nonempty if and only if

$$F < \bar{F}_Y \equiv bY^H - \frac{c(1-b)(e_L^H)^2}{2b}. \quad (11)$$

Proof. Equation (9) gives $\Pi^{A*} > \Pi^H$ if and only if $k < k_A(F)$. Since Y^A increases in a , Corollary 1 gives $Y^A(a^*) < Y^H$ exactly when $a^* < e_L^H$. Substitution yields $k > k_Y$, with $k_0 < k_Y$. At k_Y , AI profit minus human profit is $\bar{F}_Y - F$. AI profit strictly decreases in k , proving the interval condition. $\square$

The saving on the follower's pay can make adoption profitable before AI contributes enough to preserve output. When $F < \bar{F}_Y$, falling AI costs therefore trigger adoption at k_A while output remains below the human-team level. Output drops at adoption, but further cost reductions increase both AI

input and leader effort, restoring output at k_Y .

For $b = 1/3$, $c = \beta = 1$, and $F = 0.01$, the thresholds are $k_Y = 2$ and $k_A \simeq 2.196$. At $k = 2.1$, human efforts (0.714, 0.571) become leader effort 0.557 and AI input 0.671. AI contributes 17.4% more than the displaced worker, yet output falls from 1.694 to 1.602 (5.4%). The higher AI contribution does not compensate for the leader’s reduced effort. Profit rises from 0.565 to 0.585 (3.6%). Cost savings exceed the revenue loss. Strict inequalities ensure this example persists under small parameter changes. Appendix B gives a numerical example with the same three inequalities when the firm optimizes each human’s output share separately.

Figure 1 separates the adoption threshold from the output-recovery threshold. Higher fixed cost lowers k_A without changing optimal capacity conditional on adoption; sufficiently high fixed cost eliminates output-reducing adoption. Lower effort can benefit the leader through cost savings; an output loss alone does not establish a welfare loss.

5 Conclusion

Human followers contribute both their own input and, through their responsiveness, an incentive for leaders to work harder. Equal-contribution replacement reduces output under general complementary technologies; under symmetry, preserving output requires AI to match the former leader’s contribution. Profit-maximizing firms may nevertheless adopt AI that contributes more than the displaced follower while the team produces less. With multiple followers, replacing one can reduce every remaining human’s effort.

Even if AI can be programmed to reproduce a human follower’s response, the rule need not be credible: if settings remain adjustable, those controlling the AI must also have an incentive to uphold the rule. With no marginal cost of using installed capacity, a leader controlling AI prefers unconditional full utilization, so greater effort cannot elicit more AI input. A firm may promise conditional utilization to motivate the leader but would prefer full utilization once the leader’s effort is sunk. If the leader anticipates an override, the rule loses its incentive effect. The hidden value of human followers thus lies in their autonomy: they respond according to their own interests. Since neither the leader nor the firm can simply dictate this response, the leader must work harder to make greater follower effort worthwhile.

References

- Cheng, X., Dogan, M., Yildirim, P., 2025. Artificial intelligence in team dynamics: Who gets replaced and why? NBER Working Paper 34259. <https://www.nber.org/papers/w34259>.
- Dell’Acqua, F., Kogut, B., Perkowski, P., 2025. Super Mario meets AI: Experimental effects of automation and skills on team performance and coordination. *Review of Economics and Statistics* 107, 951–966. doi:10.1162/rest_a.01328.
- Dogan, M., Yildirim, P., 2022. Managing automation in teams. *Journal of Economics & Management Strategy* 31, 146–170. doi:10.1111/jems.12456.

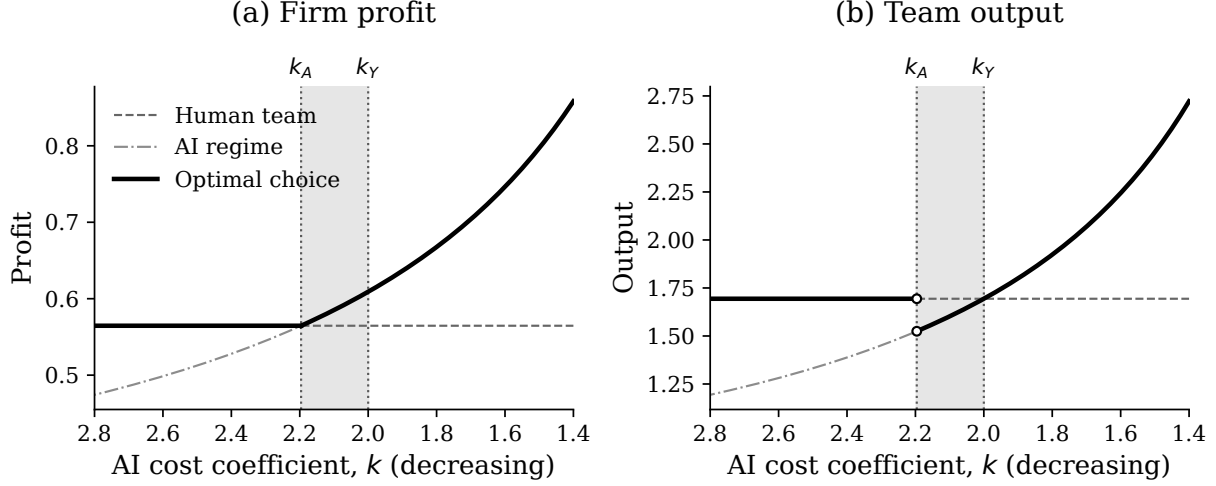


Figure 1: Cheaper AI, endogenous adoption, and team output.

Note: $b = 1/3$, $c = \beta = 1$, $F = 0.01$. Moving right means lower AI cost. Dotted lines mark adoption (k_A) and output recovery (k_Y); the shaded region features profitable adoption with lower output. At k_A the firm is indifferent.

Appendices

The appendices extend the analysis to teams with multiple followers (Appendix A), endogenous output shares (Appendix B), and heterogeneous effort costs (Appendix C).

A One leader and simultaneous followers

This appendix extends the quadratic technology to one leader and $n \geq 2$ follower positions. It shows that replacing one follower with a fixed AI input equal to that follower's original contribution reduces the efforts of the leader and every remaining human follower.

A.1 Technology and follower responses

For each follower position $j = 1, \dots, n$, let x_j denote its input. For a human follower, $x_j = e_{F,j}$, where $e_{F,j}$ is the effort of follower j ; for a position occupied by AI, $x_j = a$. The leader remains human and supplies effort e_L . Output is

$$Y = e_L + \sum_{j=1}^n x_j + \beta \left(e_L \sum_{j=1}^n x_j + \sum_{j < m} x_j x_m \right), \quad \beta > 0. \quad (\text{A.1})$$

The leader's payoff is $bY - ce_L^2/2$, and human follower j has payoff $bY - ce_{F,j}^2/2$, with $b, c > 0$. If shares are financed from output, impose $(n+1)b \leq 1$. The leader chooses observable effort; human followers then move simultaneously. An AI input is known and fixed before the leader acts. All human reward shares are unchanged by replacement.

Set $\lambda = b\beta/c$. Let $h \in \{n-1, n\}$ denote the number of human followers and define

$$d_h = 1 - \lambda(h-1), \quad \eta_h = 1 - \frac{2h\lambda^2}{d_h} - \frac{h(h-1)\lambda^3}{d_h^2}. \quad (\text{A.2})$$

Assume $d_n > 0$ and $\eta_n, \eta_{n-1} > 0$. The first condition makes each follower subgame a contraction in the sup norm and yields a unique positive symmetric equilibrium. The second pair gives strict concavity of the leader's reduced payoff in both regimes. The effort solutions are finite and positive.

With all followers human, their equilibrium effort at a given leader effort is

$$R_{F,n}^H(e_L) = \frac{b/c + \lambda e_L}{d_n}, \quad (R_{F,n}^H)'(e_L) = \frac{\lambda}{d_n}. \quad (\text{A.3})$$

With one AI input a and $h = n-1$ human followers, the remaining humans choose

$$R_{F,n}^A(e_L; a) = \frac{b/c + \lambda(e_L + a)}{d_h}, \quad \frac{\partial R_{F,n}^A}{\partial e_L} = \frac{\lambda}{d_h}. \quad (\text{A.4})$$

These formulas follow by solving the first-order conditions $ce_{F,j} = b[1 + \beta(e_L + \sum_{m \neq j} x_m)]$ simultaneously for all human followers. All terms are positive, so nonnegativity constraints do not bind.

Let $e_{L,n}^H$ denote the leader's all-human equilibrium effort and $e_{F,n}^H$ the common equilibrium effort of each follower. In these equilibrium symbols, the subscript n indicates the total number of follower positions, not an individual follower's index. By contrast, the subscript j in $e_{F,j}$ identifies an individual human follower. Backward induction yields

$$e_{L,n}^H = \frac{b[1 + 2\lambda - (n-1)\lambda^2]}{cd_n^2\eta_n}, \quad e_{F,n}^H = \frac{b(1 + \lambda)}{cd_n\eta_n}. \quad (\text{A.5})$$

The second derivative of the leader's reduced payoff with h human followers is $-c\eta_h$, irrespective of the fixed AI input. Let $e_{L,n}^A$ and $e_{F,n}^A$ denote the equilibrium efforts of the leader and each remaining human follower after one human is replaced by AI input $a = e_{F,n}^H$. The index n continues to count all follower positions, including AI.

Proposition A.1. *Replacing one follower with $a = e_{F,n}^H$ strictly reduces every remaining human's effort and total output. Writing $\Delta e_{L,n} = e_{L,n}^H - e_{L,n}^A$ and $\Delta e_{F,n} = e_{F,n}^H - e_{F,n}^A$,*

$$\Delta e_{L,n} = \frac{e_{F,n}^H \lambda (1 + \lambda)}{d_n d_{n-1} \eta_{n-1}} > 0, \quad (\text{A.6})$$

$$\Delta e_{F,n} = \frac{\lambda}{d_{n-1}} \Delta e_{L,n} > 0. \quad (\text{A.7})$$

Consequently, the total effort decrease among the remaining humans is

$$\Delta E_H = \Delta e_{L,n} + (n-1)\Delta e_{F,n} = \frac{1 + \lambda}{d_{n-1}} \Delta e_{L,n}. \quad (\text{A.8})$$

Proof. At $e_L = e_{L,n}^H$ and $a = e_{F,n}^H$, equation (A.4) gives $R_{F,n}^A(e_{L,n}^H; e_{F,n}^H) = e_{F,n}^H$. Thus every input and every marginal product initially equals its all-human value. Let M_n be a follower's marginal product at this point. The follower's first-order condition implies $bM_n = ce_{F,n}^H$.

Writing $Y_L = \partial Y / \partial e_L$ for the leader's marginal product at this point, the leader's marginal payoff in the all-human regime is

$$b \left[Y_L + M_n \frac{n\lambda}{d_n} \right] - ce_{L,n}^H = 0.$$

After replacement, at that same effort profile it becomes

$$\begin{aligned} & b \left[Y_L + M_n \frac{(n-1)\lambda}{d_{n-1}} \right] - ce_{L,n}^H \\ &= -ce_{F,n}^H \lambda \left[\frac{n}{d_n} - \frac{n-1}{d_{n-1}} \right] = -\frac{ce_{F,n}^H \lambda (1+\lambda)}{d_n d_{n-1}} < 0. \end{aligned} \quad (\text{A.9})$$

The mixed-regime leader payoff is quadratic with second derivative $-c\eta_{n-1}$. Moving its zero marginal payoff leftward therefore gives (A.6). Equation (A.4) gives (A.7); summing yields (A.8). The mixed-regime marginal payoff at zero effort is positive, so its optimum remains interior. Since output is strictly increasing in each nonnegative input, and the AI input stays at $e_{F,n}^H$, the reductions in all human efforts strictly reduce output. $\square$

The proof also identifies how the effort reduction spreads. If the leader maintained the original effort $e_{L,n}^H$, the remaining human followers would still choose $e_{F,n}^H$: AI supplies exactly the displaced follower's contribution, preserving all inputs and marginal products at the former allocation. What changes is the response to a change in leader effort. The aggregate follower-response slope falls from $n\lambda/d_n$ to $(n-1)\lambda/d_{n-1}$. Replacement removes one responder and weakens each remaining follower's equilibrium response, since the displaced worker no longer adjusts effort to reinforce the others' responses.

Anticipating this weaker aggregate response, the leader reduces effort. The remaining human followers then face a lower return to effort and reduce their own, as equation (A.7) shows. When at least two human followers remain, complementarity among their efforts further reinforces these reductions within the simultaneous follower subgame. The response coefficient λ/d_{n-1} incorporates this interaction.

A single replacement can therefore reduce effort throughout the human part of the team, even though AI fully reproduces the displaced worker's contribution. A human follower supports colleagues' effort indirectly by sustaining the leader's incentive to elicit effort. This propagation to the other followers is the additional effect relative to the two-person baseline. For $n = 1$, interpreting the absent human-follower terms as zero recovers the main-text identity $\Delta e_{L,1} = \lambda e_{F,1}^H$.

For comparison, consider a horizontal team with the same technology, costs, and reward shares, in which all members choose effort simultaneously. Replacing one member with AI that matches that member's contribution at the horizontal team's own equilibrium leaves every remaining human's effort unchanged: each still faces the same inputs from the others, and the equilibrium is unique under the maintained parameter restrictions. The effort reductions among the remaining followers in the sequential team therefore arise through the leader's adjustment.

A.2 Team size and the magnitude of the loss

Equation (A.8) isolates the propagation of a given leader-effort reduction. For $n > 1$, its multiplier exceeds one. It does not imply that the percentage output loss must increase with n : team size also changes baseline output and the leader's initial response. Moreover, fixing individual shares and dividing a fixed reward pool across more humans are different comparisons.

Table A1 fixes $b = 0.1$, $c = \beta = 1$ across team sizes. Each experiment replaces one follower with that team's own all-human follower contribution. Let Y_n^H and Y_n^A denote equilibrium team output before and after replacement, respectively. All displayed cases satisfy the existence and concavity conditions and have total human shares below one.

The absolute output loss rises in these examples, but the percentage loss initially falls. The robust additional result is the reduction in every remaining human's effort.

B Endogenous output shares

This numerical check lets the firm choose each human's linear output share before efforts are chosen. Set $c = \beta = 1$. In the human team, payoffs are $b_i Y - e_i^2/2$, and the firm chooses $b_L, b_F \geq 0$ with $b_L + b_F \leq 1$. Backward induction gives

$$e_L^H = \frac{b_L(1 + 2b_F)}{1 - 2b_L b_F}, \quad e_F^H = b_F(1 + e_L^H), \quad \Pi^H = (1 - b_L - b_F)(e_L^H + e_F^H + e_L^H e_F^H). \quad (\text{B.1})$$

The denominator is positive throughout the feasible set. Under AI, the firm chooses $b_L \in [0, 1]$ and $a \geq 0$. Given these choices, the leader supplies $e_L^A = R_L(a) = b_L(1 + a)$. For $k > 1/2$, optimal capacity conditional on b_L and the resulting profit are

$$a^*(b_L) = \frac{(1 - b_L)(1 + 2b_L)}{k - 2b_L(1 - b_L)}, \quad (\text{B.2})$$

$$\Pi^{A*}(b_L) = b_L(1 - b_L) - F + \frac{(1 - b_L)^2(1 + 2b_L)^2}{2[k - 2b_L(1 - b_L)]}. \quad (\text{B.3})$$

We maximize profits over feasible output shares by comparing stationary points and boundaries. For $F = 0.01$ and $k = 2.05$, the optimal shares are $(b_L, b_F) = (0.3542, 0.3692)$ in the human team and $b_L = 0.3792$ in the AI team. Table B1 reports the resulting allocations. AI contributes 2.1% more, yet output falls 14.4% and profit rises 5.0%. Numerically, these three inequalities hold throughout $1.80 \leq k \leq 2.08$ when $F = 0.01$. Thus, within linear output-sharing contracts, fixing reward shares is not necessary for this outcome.

C Leader ability and asymmetric effort costs

Return to one leader and one follower, with technology $G(e_L, e_F) = e_L + e_F + \beta e_L e_F$ and a common fixed output share $b \in (0, 1/2)$. Allow their quadratic effort costs to differ:

$$C_L(e_L) = \frac{c_L}{2} e_L^2, \quad C_F(e_F) = \frac{c_F}{2} e_F^2, \quad c_L, c_F, \beta > 0. \quad (\text{C.1})$$

A lower c_L represents greater leader ability. Hold b, c_F, β fixed and restrict attention to $D \equiv c_L c_F - 2b^2 \beta^2 > 0$. The follower's response is $R_F(e_L) = b(1 + \beta e_L)/c_F$, and the leader's reduced payoff has second derivative $-D/c_F < 0$. Backward induction gives the unique human equilibrium

$$e_L^H = \frac{b(c_F + 2b\beta)}{D}, \quad e_F^H = \frac{b(c_L + b\beta)}{D}, \quad Y^H = e_L^H + e_F^H + \beta e_L^H e_F^H. \quad (\text{C.2})$$

Table A1: Replacing one follower with an equal-contribution AI

Team size	Human team			Mixed team				Loss (%)
	$e_{L,n}^H$	$e_{F,n}^H$	Y_n^H	$e_{L,n}^A$	$e_{F,n}^A$	a	Y_n^A	
2	0.122	0.112	0.248	0.111	—	0.112	0.236	5.03
3	0.154	0.128	0.467	0.138	0.127	0.128	0.444	4.74
4	0.201	0.150	0.810	0.177	0.147	0.150	0.767	5.28
5	0.277	0.182	1.409	0.238	0.178	0.182	1.315	6.66

Note: $b = 0.1$ and $c = \beta = 1$. Team size is $n + 1$, including the leader and all follower positions, one of which is occupied by AI in the mixed team. Superscripts H and A denote the human and mixed teams, respectively. $e_{L,n}$ is leader effort, $e_{F,n}$ is effort per human follower, a is AI input, and Y_n is team output. In each row, $a = e_{F,n}^H$. A dash indicates that no human follower remains. Loss (%) is $100(Y_n^H - Y_n^A)/Y_n^H$, calculated before rounding.

Table B1: Separately optimized output shares

	Human team	AI team
Leader effort	0.834	0.641
Follower or AI input	0.677	0.691
Team output	2.076	1.776
Firm profit	0.574	0.603

For each leader type, replace the follower with AI input $a = e_F^H$ evaluated at that type's human equilibrium. Thus AI contributions are matched within each team, not held constant across leader types. The AI-team leader's best response is $R_L(a) = b(1 + \beta a)/c_L$, so the replacement yields $e_L^A = R_L(e_F^H)$.

Proposition C.1. *Equal-contribution replacement reduces leader effort and output by*

$$\Delta e_L \equiv e_L^H - e_L^A = \frac{b\beta}{c_L} e_F^H > 0, \quad \Delta Y \equiv Y^H - Y^A(e_F^H) = \frac{b\beta}{c_L} e_F^H (1 + \beta e_F^H) > 0. \quad (\text{C.3})$$

Holding b, c_F, β fixed, both ΔY and $\Delta Y/Y^H$ strictly increase as c_L falls within the admissible region.

Proof. The human leader's first-order condition is $c_L e_L^H = b(1 + \beta e_F^H) + b\beta e_F^H$. Subtracting the AI leader's condition gives the first identity in (C.3). Since the follower input remains e_F^H , the output difference is $(1 + \beta e_F^H)\Delta e_L$, giving the second identity. Moreover,

$$\frac{\partial e_F^H}{\partial c_L} = -\frac{b^2\beta(c_F + 2b\beta)}{D^2} < 0.$$

Hence both e_F^H and $b\beta/c_L$ increase as c_L falls, so the absolute output loss strictly increases.

For the proportional loss, set $t = b\beta/c_L$ and $u = b\beta/c_F$, with $1 - 2tu > 0$. Write

$$z \equiv \beta e_L^H = \frac{t(1 + 2u)}{1 - 2tu}, \quad w \equiv \beta e_F^H = \frac{u(1 + t)}{1 - 2tu}, \quad \delta \equiv \frac{\Delta e_L}{e_L^H} = \frac{u(1 + t)}{1 + 2u}. \quad (\text{C.4})$$

Then

$$\frac{\Delta Y}{Y^H} = \frac{\delta z(1 + w)}{z + w + zw} = \frac{\delta}{1 + (w/z)/(1 + w)}. \quad (\text{C.5})$$

As t rises, δ and w strictly increase, while $w/z = u(1+t)/[t(1+2u)]$ strictly decreases. The numerator of the last expression therefore increases and its denominator decreases. Since lower c_L means higher t , the proportional output loss strictly increases with leader ability. $\square$

A more able leader induces a larger contribution from a human follower and adjusts effort more strongly when that follower's response disappears. The resulting loss is larger even relative to the team's higher initial output. Table C1 illustrates both comparisons; its symmetric-cost row reproduces the main-text example.

For otherwise identical independent teams differing only in c_L , an organization replacing exactly one follower with an equal-contribution AI minimizes its total output loss by choosing the team with the less able leader. Human responsiveness is therefore more valuable alongside a more able leader. This ranks replacements by output, not profit: the follower payment saved, bY^H , and the cost of supplying e_F^H can also vary across teams.

Cost asymmetry preserves the replacement loss but generally changes the output-preserving threshold. Since $Y^A(a) = a + (b/c_L)(1 + \beta a)^2$ is strictly increasing and unbounded, the unique threshold $\hat{a} > e_F^H$ solves $Y^A(\hat{a}) = Y^H$. The identity $\hat{a} = e_L^H$ is recovered when $c_L = c_F$.

Table C1: Leader ability and the output loss from equal-contribution replacement

Leader cost c_L	Human output	AI-team output	Output loss	Loss (%)
1.50	1.121	0.964	0.157	14.02
1.00	1.694	1.395	0.299	17.67
0.75	2.457	1.945	0.512	20.84
0.50	5.000	3.667	1.333	26.67

Note: $b = 1/3$ and $c_F = \beta = 1$. In each row, AI supplies that team's original human follower contribution e_F^H .