연세대학교 상경대학

**경제연구소**

Economic Research Institute

Yonsei University

서울시 서대문구 연세로 50

50 Yonsei-ro, Seodaemun-gu, Seoul, Korea

TEL: (+82-2) 2123-4065 FAX: (+82-2) 364-9149

E-mail: yeri4065@yonsei.ac.kr

http://yeri.yonsei.ac.kr/new

# Talent vs. Fit: Partner Selection and the Moderate's Trap

**Byung-Cheol Kim**

*University of Alabama*

**Jin Yeub Kim**

*Yonsei University*

**Hyunjun Cho**

*University of Pennsylvania*

February 2026

2026RWP-279

**Economic Research Institute**

**Yonsei University**

# Talent vs. Fit: Partner Selection and the Moderate's Trap*

Byung-Cheol Kim† Jin Yeub Kim‡ Hyunjun Cho§

February 11, 2026

**Abstract**

We study partner selection when higher productivity comes at the cost of weaker expected collaborative fit. When this tradeoff is strong, expected surplus is U-shaped: both high-productivity "difficult stars" and low-productivity "team players" generate higher expected value than moderately productive partners. This Moderate's Trap overturns the standard economic prediction that surplus increases monotonically with talent. In competitive labor markets, moderate workers are disadvantaged, and workers invest to specialize at the extremes, polarizing the productivity distribution. Our framework microfounds a demand-side force that operates against moderate performers, complementing existing technology-based explanations of modern labor market polarization and specialization.

*Keywords:* Partner selection, productivity, collaborative fit, difficult stars, wage polarization

*JEL Classification:* C78, D82, J31, L23, M51

*We thank Jon Eguia, Hülya Eraslan, Paan Jindapon, Ashwin Kambhampati, Jeongsik Lee, and Tigran Melkonyan for their helpful comments. Jin Yeub Kim acknowledges that this work was supported by the Yonsei University Research Grant of 2024.

†Department of Economics, Finance, and Legal Studies, Culverhouse College of Business, University of Alabama. Tuscaloosa, AL 35487, USA. E-mail: bkim34@ua.edu.

‡Corresponding author. School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea. E-mail: jinyeub@yonsei.ac.kr.

§Department of Economics, University of Pennsylvania, 133 South 36th Street, Philadelphia, PA 19104, USA. E-mail: hyn1652@sas.upenn.edu.

# 1 Introduction

Law firms promoting associates to partner, venture capital firms choosing startup partners, pharmaceutical companies forming joint ventures, and academics selecting coauthors face a common problem. They must choose partners before knowing their *collaborative fit*, that is, how well they will work together effectively. A potential partner's individual talent can often be evaluated ex ante using observable credentials, track records, and references. By contrast, collaborative fit is typically difficult to assess until the partnership begins. Facing this uncertainty, principals must base their selection on *expected* collaborative fit given observable characteristics. When higher talent systematically predicts weaker expected fit, this tradeoff shapes partner selection decisions and, in turn, affects organizational performance and labor-market outcomes.

To illustrate, consider a law firm deciding whom to promote to partner. One candidate is a highly productive "star" associate with exceptional billable hours and a strong book of business, but is widely regarded as difficult to work with. Another candidate is less productive by conventional metrics, but is expected to integrate smoothly into the partnership and facilitate collaboration within the firm. Between them is a third candidate with solid but unremarkable billables and no salient strengths or weaknesses. From a risk-minimization perspective, this intermediate option may appear safest, avoiding the potential frictions associated with the star while offering higher individual output than the least productive candidate. Our analysis shows that this reasoning can be exactly backwards.

The tension between individual talent and collaborative fit is pervasive in organizations. Management practitioners have long recognized that highly skilled, top-performing employees are sometimes difficult to work with, often labeled "brilliant jerks."[1] In popular discourse, the related notion of "difficult stars" is used to describe famous actors who are known for being hard to work with. We adopt this term, difficult stars, to refer to highly talented individuals with low expected collaborative fit in a partnership environment.

Despite the widespread recognition of difficult stars in practice, standard models of partner selection and matching predict that higher productivity is always preferred—surplus increases monotonically in talent. This prediction sits uneasily with observed patterns in team composition and organizational hiring. In professional services, up-or-out systems retain only exceptional performers while discarding the competent but unexceptional.[2] Technol-

[1] The term "brilliant jerks" was widely popularized by Netflix's 2009 culture deck, which explicitly advised against hiring such workers regardless of performance.

[2] Up-or-out promotion systems are prevalent in law firms, consulting, investment banking, and academia. See Galanter and Palay (1991) on law firms and Kahn and Huberman (1988) on academia. Galanter and Henderson (2008) document a subsequent shift toward "elastic tournaments" with multi-tiered structures

ogy firms often target either star individual contributors or strong culture-fit candidates, rejecting those who are average along both dimensions. Professional sports teams assemble rosters mixing superstars with cooperative role players, bypassing mid-tier talent.[3] These practices suggest that organizations systematically favor extremes over the middle, a pattern difficult to rationalize under standard monotonic models. At a broader level, a large literature documents wage and employment polarization across labor markets, with growth concentrated at the top and bottom of the skill distribution while the middle declines (Autor et al., 2003; Autor and Dorn, 2013; Goos and Manning, 2007). Standard explanations emphasize technological change that displaces routine middle-skill tasks. While these practices have been documented empirically and explained through setting-specific mechanisms (tournament incentives for up-or-out, culture arguments for bimodal hiring, coordination costs for team composition), no unified theoretical framework derives them from a common economic primitive. Our framework identifies a complementary demand-side mechanism operating within organizations: in partnership environments where higher productivity comes at the cost of weaker collaborative fit, expected surplus can be U-shaped, making moderate performers the least attractive.

Our model fills this gap by showing that bimodal selection, avoidance of moderate performers, and wage compression in the middle can all emerge from a single tradeoff between observable productivity and unobservable collaborative fit. This demand-side mechanism operates at the level of individual hiring and team formation decisions, providing a microfoundation for organizational preferences that favor extremes.

We study partner selection in a model with three central features. First, partners exert effort that combines through a constant elasticity of substitution (CES) production technology. We formalize individual talent as the productivity of a partner's effort and collaborative fit as the substitutability of partners' efforts, which reflects the flexibility of the production process. Second, productivity is observable prior to partnership formation, whereas collaborative fit is unobservable and revealed only through interaction. This information structure reflects a basic feature of the hiring process: a résumé can be evaluated ex ante, but the functioning of collaboration cannot be assessed beforehand. Third, we introduce the key assumption that *stars are difficult to work with*—more productive individuals tend to be less collaborative. We capture this tendency by a systematic negative relationship between productivity and expected collaborative fit. In this environment, a principal selects a partner from a pool of agents, each characterized by an observable level of productivity, without

that hollow out the middle of the partnership.

[3] Swaab et al. (2014) document a "too-much-talent" effect in professional sports, whereby teams with too many top performers underperform on tasks requiring coordination.

knowing the exact collaborative fit with the potential partner.

Our analysis characterizes how expected surplus varies with partner productivity when higher productivity comes at the cost of weaker collaborative fit. Depending on the intensity of this tradeoff, two regimes emerge. When the tradeoff is mild, so that highly productive agents are expected to retain a reasonable level of collaborative fit, expected surplus increases monotonically with productivity. In this regime, the most productive agent is unambiguously preferred, as productivity gains always outweigh losses from reduced collaborative fit. The conventional wisdom that 'talent is king' therefore applies.

This wisdom breaks down when the tradeoff is sufficiently strong, so that high productivity requires substantial sacrifices in collaborative fit. In this case, the structure of the partner selection problem changes fundamentally. Expected surplus becomes *U-shaped* in partner productivity, with both the most and the least productive agents generating relatively high expected values, while moderately productive agents generate the lowest expected values. We refer to this pattern as the *Moderate's Trap*. This finding overturns the central prediction of standard matching and superstar models, in which surplus increases monotonically with talent. In those frameworks, moderate performers occupy a natural middle ground between stars and low performers. The Moderate's Trap reverses this logic: the middle becomes the worst place to be.

The underlying mechanism is driven by the convexity of value creation, whereby losing collaborative fit is more costly for stronger collaborative fit. Low-productivity agents, whom we refer to as team players, generate value through strong collaborative fit and low frictions in the production process. For these team players, gains from stronger collaborative fit overpower the losses from lower individual productivity. High-productivity agents, whom we call difficult stars, generate value through exceptional individual talent. For these difficult stars, gains from exceptional individual productivity outweigh the frictions from weak collaborative fit. Agents with moderate productivity levels possess *neither* leverage. Their productivity is insufficient to create substantial value independently, while their collaborative fit is too weak to compensate for limited productivity gains. They are stuck in the middle.

The Moderate's Trap has sharp implications for optimal partner selection. Because expected surplus is U-shaped, the optimal partner selection strategy becomes *bimodal*: when making selection decisions, principals need only compare the least productive agent with the most productive agent in the candidate pool, while bypassing all intermediate types. When the tradeoff is sufficiently strong but not prohibitively so, the productivity advantage of the most productive agent dominates the collaborative-fit advantage of the least productive agent, making the most productive agent the optimal choice. As the tradeoff intensifies further, the collaborative-fit advantage of the least productive agent becomes increasingly pronounced, leading the optimal choice to switch to the least productive agent.

We embed this framework in a competitive labor market to trace the implications of the Moderate's Trap for wages and investment incentives. When firms compete for workers and earn equal profits in equilibrium, wage differences mirror differences in expected surplus. The U-shaped surplus profile therefore translates into a wage structure that penalizes moderate productivity. In isolation, the model predicts U-shaped wages in levels; in practice, when combined with other forces such as standard returns to skill, the mechanism compresses wages in the middle relative to what monotonic models would predict (Section 6.3). This provides a demand-side mechanism for wage polarization that operates through organizational preferences over talent and fit, complementing standard explanations based on technological displacement of routine tasks.

The trap can be self-reinforcing through endogenous investment. When market wages penalize intermediate productivity, workers respond by specializing at the extremes. High-talent agents invest fully to raise their productivity and become difficult stars, while low-talent agents do not invest and remain team players. The resulting productivity distribution is polarized, with the middle hollowed out. This amplification mechanism shows how the demand-side force we identify can interact with workers' human capital decisions to generate and sustain polarization in equilibrium.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the model and the key assumption of difficult stars. Section 4 characterizes the equilibrium. Section 5 establishes our main results on expected surplus and optimal partner selection. Section 6 embeds the model in a competitive labor market, characterizing equilibrium wages, endogenous investment, and the connection to labor market polarization. Section 7 concludes. Appendix A contains all proofs. Appendix B provides supplementary results on bargaining. Appendix C examines how overconfidence can distort partner selection.

# 2 Related Literature

This paper relates to several strands of literature. We make three contributions: (i) we show that when productivity trades off against collaborative fit, expected surplus can be U-shaped rather than monotonically increasing in talent; (ii) we characterize bimodal partner selection, whereby principals optimally choose from the extremes of the productivity distribution; and (iii) we embed this mechanism in a competitive labor market and show that it generates wage polarization and self-reinforcing investment patterns.

**Superstar economics.** Rosen (1981) develops a theory of superstars in which small talent differences generate large earnings differences, as scalability and convex demand allow more talented agents capture disproportionate market share. Gabaix and Landier (2008) extend this logic to CEO compensation. In these models, surplus and earnings are monotonic in talent. We depart from this framework by introducing collaborative fit as an additional source of value creation that trades off against productivity. When individual talent trades off with collaborative fit, high surplus can arise either from exceptionally talented but difficult stars or from less talented agents who work particularly well with others. Moderately talented agents leverage neither channel. Expected surplus therefore need not be monotonic in talent and can instead be U-shaped, giving rise to a "moderate penalty" absent from standard superstar models.

Empirical evidence further suggests that star performance is less portable than standard models assume. Groysberg et al. (2008) document sharp performance declines among star security analysts after switching firms, indicating dependence on firm-specific skills and environments. Azoulay et al. (2010) find that shocks to superstar scientists reduce collaborators' output, highlighting spillovers generated within teams. Oettl (2012) show that peer performance depends on the social dimension of stars, defined as helpfulness to peers, beyond their individual productivity. Together, these findings motivate a modeling approach in which the effective value of high productivity is mediated by collaboration. Psychological research on the "compensation effect" between perceived competence and warmth provides independent support for this tradeoff (Fiske et al., 2002).

**Matching theory.** The seminal work of Becker (1973) establishes positive assortative matching under supermodular production: matched partners have similar quality types. Subsequent work examines how this prediction is modified when frictions are introduced. Shimer and Smith (2000) show that search frictions can disrupt assortative matching even when production is supermodular, as agents weigh the opportunity cost of continued search against immediate matching. Legros and Newman (2007) demonstrate that when utility is imperfectly transferable and the degree of transferability varies with type, the standard prediction can reverse entirely—their "Beauty is a Beast" result shows that high types may match negatively assortatively when they cannot easily transfer surplus to partners.

We depart from this literature in a different direction. Rather than introducing frictions that disrupt otherwise-efficient matching, we model partner type as a bundle of productivity and collaborative fit that are negatively correlated in expectations. This embeds a technological tradeoff directly into the type space. While canonical matching models implicitly aggregate multiple attributes into a single index, such aggregation presumes that all relevant characteristics move in the same direction. Our formulation instead makes the tradeoff between distinct sources of value creation explicit. This perspective relates to work on multidimensional heterogeneity in matching: Lindenlaub (2017) shows that vector-valued types need not admit a scalar ordering, and Lise and Postel-Vinay (2020) find that interpersonal skills can bind even when cognitive ability is high. Unlike these studies, ours is not a two-sided matching model but one in which a principal selects a single partner from a heterogeneous pool. This asymmetry shifts the focus from assortative matching to optimal partner selection. Our contribution is to show that when productivity and *expected* collaborative fit are negatively correlated, expected surplus can be U-shaped in productivity, generating bimodal selection. The mechanism operates through this systematic tradeoff, with residual uncertainty about fit playing an attenuating rather than amplifying role.

**Team composition and selection.** A related literature examines optimal team composition and the foundations of team cooperation. Prat (2002) shows that whether a team should be homogeneous or heterogeneous depends on the sign of complementarity between jobs: positive complementarities favor similar types, while negative complementarities favor diversity. Mello and Ruckes (2006) study how uncertainty and the nature of tasks affect optimal team heterogeneity. Glover and Kim (2021) show how diversity can foster implicit team incentives through reputational concerns, while Imhof and Kräkel (2023) analyze how task-related heterogeneity—differences in team members' expected abilities—affects effort provision in promotion tournaments. Empirically, Hamilton et al. (2003) find that heterogeneous teams outperform homogeneous ones in a garment plant, attributing this to mutual learning and intrateam bargaining. On the behavioral side, Zeitoun et al. (2023) develop a theory of how team members sustain cooperation through virtual bargaining, a reasoning process that supports tacit commitments even absent explicit contracts. These studies focus on the composition of teams with multiple members, where diversity can provide informational or skill-based advantages, or on the behavioral mechanisms that sustain cooperation once teams form.

Our contribution is distinct in two respects. First, we study bilateral partnerships rather than multilateral teams, so diversity within a team is not the relevant margin. Second, our focus is on the tradeoff between observable productivity and uncertain collaborative fit within a single partner, rather than on combining partners with different observable characteristics. The Moderate's Trap arises from the tension between these two attributes of the same individual, a margin that is suppressed when partner types are one-dimensional.

**Team production.** The contract theory literature focuses on incentive design given a fixed team composition (Holmström, 1982; Che and Chung, 1999), while the team production literature studies optimal team composition abstracting from incentives. Recent work examines their interaction. Kambhampati and Segura-Rodriguez (2022) show that non-assortative team formation can arise when productivity is private information and effort is hidden. Kambhampati et al. (2024) study team formation with shared output and show that free-riding can prevent efficient diverse teams from forming.[4] We complement these works by showing that non-monotonic partner selection can arise even absent agency problems, driven purely by technological tradeoffs between productivity and collaborative fit.

Empirical studies of scientific collaboration show that prolific researchers maintain extensive coauthorship networks (Newman, 2004; Börner et al., 2010), and that team-authored papers receive more citations than solo work (Wuchty et al., 2007). While these findings underscore the importance of collaboration, they reflect outcomes of realized partnerships and do not speak directly to the ex ante likelihood that a high-ability individual will fit well with an arbitrary partner. Evidence more closely aligned with our framework comes from settings in which ex ante ability is measurable and team performance depends on coordination. Swaab et al. (2014) document a "too-much-talent" effect, whereby teams of highly talented individuals underperform when tasks are highly interdependent. This pattern is broadly in line with our mechanism, in which high individual productivity can be associated with weaker collaborative fit.

**Labor market polarization.** Several studies document wage and employment polarization, with growth concentrated at the top and bottom of the skill distribution (Autor et al., 2003; Autor and Dorn, 2013; Goos and Manning, 2007). Deming (2017) shows that growth has been strongest in occupations requiring social skills, a finding consistent with the increasing complementarity between cognitive and social skills (Weinberger, 2014; Heckman and Kautz, 2012). Standard explanations emphasize routine-biased technological change that displaces middle-skill tasks while complementing high-skill cognitive work and preserving demand for low-skill manual and service work (Autor and Dorn, 2013). We offer a complementary demand-side mechanism. Our framework does not predict that wages are U-shaped in practice, as real-world wages reflect many forces beyond the one we study. Rather, we microfound a force that operates against moderate performers in partnership environments, providing a theoretical rationale for why the middle of the productivity distribution can be systematically disadvantaged even absent technological displacement.

[4] See also Kaya and Vereshchagina (2022) who study sorting of workers who differ in their ability to acquire information.

# 3 Model

## 3.1 Environment

Two risk-neutral parties form a partnership. Partner 1 exerts effort $x \geq 0$ at cost $c(x)$ and Partner 2 exerts effort $y \geq 0$ at cost $c(y)$. The cost function $c : \mathbb{R}_+ \to \mathbb{R}_+$ is continuously differentiable, strictly increasing, and strictly convex, with $c(0) = 0$, $c'(0) = 0$, and $\lim_{e\to\infty} c'(e) = \infty$.

Output is produced according to the following symmetric CES technology:[5]

$$F(x, y; \alpha, \rho) = \left(\frac{x^\rho + (\alpha y)^\rho}{2}\right)^{1/\rho} \tag{1}$$

where $\rho \in (0, 1)$ is the substitution parameter[6] and $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, with $0 < \underline{\alpha} < 1 < \overline{\alpha}$, captures the *productivity* of Partner 2's effort per unit relative to that of Partner 1. When $\alpha > 1$, Partner 2 is relatively more productive than Partner 1, and vice versa when $\alpha < 1$.

Two partners' *collaborative fit* describes how well they are able to work together in the production process. We capture this notion through the substitution parameter $\rho$. A higher value of $\rho$ corresponds to a more flexible production process, in which efforts are closer substitutes and coordination frictions are relatively low; a lower value of $\rho$ corresponds to a more rigid production process, in which efforts are poorer substitutes and frictions are higher. Accordingly, we interpret a higher (resp. lower) $\rho$ as indicating stronger (resp. weaker) collaborative fit. Our notion of collaborative fit does not refer to technological complementarity of inputs in the sense of supermodularity; rather, it captures the flexibility of the production process.

Individual productivity, captured by $\alpha$, reflects how effective Partner 2's effort is in isolation and is observable prior to partnership formation. By contrast, collaborative fit is typically harder to assess ex ante. We represent this through the substitution parameter $\rho$, which is uncertain at the partner selection stage. As we formalize below, the principal's

[5]The standard CES aggregator allows asymmetric weights $(\omega, 1-\omega)$. We set $\omega = 1/2$ so that the production function is symmetric in partners' efforts when $\alpha = 1$. This normalization ensures that the bargaining problem is symmetric at equal productivity, yielding $s^* = 1/2$ (Proposition B.1), and that the surplus gradient $G(\alpha; \lambda)$ is continuous at $\alpha = 1$ (Theorem 1). Asymmetry between partners is captured entirely by the productivity parameter $\alpha$.

[6]We restrict the substitution parameter $\rho \in (0, 1)$. The upper bound comes from the fact that at $\rho = 1$, the aggregator is linear and collaborative fit is irrelevant. The lower bound serves two purposes: (i) it maintains the gross-substitutes interpretation of collaborative fit. For modestly negative $\rho$, the formal analysis extends. However, efforts become complements and the interpretation of $\rho$ as production flexibility no longer applies; (ii) it avoids the near-Leontief regime ($\rho \to -\infty$), where optimal efforts are driven toward corners under convex costs and $F$ loses smoothness.

selection decision depends on expected surplus, which is governed by the relationship between productivity and *expected* collaborative fit.

Once formed, the partnership operates under a linear sharing rule $s \in [0, 1]$, where Partner 1 receives a share $s$ of output and Partner 2 receives the remaining share $1 - s$. Payoffs are given by $U_1(s) = s \cdot F(x, y; \alpha, \rho) - c(x)$ and $U_2(s) = (1 - s) \cdot F(x, y; \alpha, \rho) - c(y)$, where the dependence of $U_i(s)$ on $x$, $y$, $\alpha$, and $\rho$ are suppressed for simplicity. Total surplus from the partnership is

$$S = U_1(s) + U_2(s) = F(x, y; \alpha, \rho) - c(x) - c(y).$$

## 3.2 Timing of the Game and Equilibrium

The partnership game proceeds in three stages specified below.

1. **Partner Selection:** A principal selects an agent from a pool of potential partners, each with an observable level of $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. The exact value of $\rho$ is not known, but the conditional distribution of $\rho$ given $\alpha$ is common knowledge. Upon selection, a partnership is formed between the principal (Partner 1) and the selected agent (Partner 2).

2. **Revelation and Bargaining:** The realization of $\rho$ is revealed to both partners. They then bargain over the sharing rule $s$, with full knowledge of $(\alpha, \rho)$.

3. **Effort and Payoffs:** Partners simultaneously choose efforts $x$ and $y$. Output is produced and payoffs are distributed according to the agreed sharing rule.

The key informational assumption is that productivity $\alpha$ is observable prior to partnership formation, whereas collaborative fit $\rho$ is revealed only after partnership formation and before bargaining.

**The equilibrium.** The solution concept is subgame perfection, with the bargaining stage resolved using the Nash bargaining solution. An equilibrium consists of $\{\alpha^*, s^*, (x^*, y^*)\}$, which together satisfy the conditions listed below.

(i) For every $\alpha^*$ and $s^*$ given any realized $\rho$, the effort profile $(x^*, y^*)$ constitutes a Nash equilibrium of the effort subgame.

(ii) For every $\alpha^*$ and given any realized $\rho$, the sharing rule $s^*$ is the Nash bargaining outcome of the bargaining subgame given the continuation payoffs generated by $(x^*, y^*)$.

(iii) The partner choice $\alpha^*$ maximizes the principal's expected continuation payoff under the prior beliefs about $\rho$, given the continuation strategies $s^*$ and $(x^*, y^*)$.

## 3.3 Difficult Stars Assumption and Information Structure

We give more structure to uncertainty about $\rho$ at the partner selection stage by introducing the key assumption of our model. A common empirical regularity across organizations is that highly talented, top-performing workers (whom we call "stars") are often less collaborative; that is, *stars are difficult to work with*. We focus on environments in which higher individual talent is systematically associated with weaker collaborative fit. In the model, this regularity is captured by the following assumption.

**Assumption DS** (Difficult Stars)**.** Collaborative fit is negatively related to productivity according to the following function that determines the substitution parameter:

$$\rho := \rho_0 - p(\alpha; \lambda), \qquad p(\alpha; \lambda) \equiv \lambda(\alpha - 1), \tag{2}$$

where $\rho_0 \in [\underline{\rho_0}, \overline{\rho_0}]$, with $0 < \underline{\rho_0} < \overline{\rho_0} < 1$, denotes the baseline substitution parameter at $\alpha = 1$, and $\lambda \geq 0$ measures the intensity of the tradeoff between productivity and collaborative fit.[7]

We hereafter refer to $\lambda$ simply as *the tradeoff intensity*. If $\lambda = 0$, productivity and collaborative fit are independent. If $\lambda > 0$, higher productivity is associated with a lower value of $\rho$, implying weaker collaborative fit. The linear specification in (2) is parsimonious and captures the idea that higher talent comes at the cost of reduced collaborative fit.

At Stage 1, the principal and agents in the pool know the conditional distribution of $\rho$ given $\alpha$, and all other aspects of the game are common knowledge, including the functional form in (2). In essence, they do not observe the baseline substitution parameter $\rho_0$ but only know its distribution. Let $\rho_0$ be a random variable supported on $[\underline{\rho_0}, \overline{\rho_0}]$ with cumulative distribution function $\mathcal{H}$, independent of both $\alpha$ and $\lambda$. Expected collaborative fit given productivity is then $\mathbb{E}[\rho|\alpha, \lambda] = \mathbb{E}[\rho_0] - \lambda(\alpha - 1)$, which is decreasing in $\alpha$ for $\lambda > 0$. The baseline parameter $\rho_0$ introduces residual variation in collaborative fit that is orthogonal to productivity. As we show in Section 5.4, greater dispersion in $\rho_0$ attenuates rather than amplifies the Moderate's Trap, confirming that the mechanism operates through the systematic component of the productivity-fit relationship captured by $\lambda$. To ensure that the realized $\rho$ lies in $(0, 1)$, we restrict attention to $\lambda \in [0, \lambda^{max})$, where

$$\lambda^{max} \equiv \min\left\{\frac{\underline{\rho_0}}{\overline{\alpha} - 1}, \frac{1 - \overline{\rho_0}}{1 - \underline{\alpha}}\right\}.$$

[7] Fiske et al. (2002) document a compensation effect between perceived competence and warmth in social cognition, providing independent support for a systematic tradeoff between productivity and collaborative fit.

For expositional simplicity, we suppress explicit references to $\lambda^{max}$ when stating results.

### 3.4 Discussion on Modeling Assumptions

We discuss several modeling assumptions and clarify their interpretations.

**The tradeoff intensity parameter.** The parameter $\lambda$ is assumed common across agents within a given environment. This reflects the idea that collaborative difficulty associated with high productivity is shaped largely by the task or organizational context rather than by idiosyncratic individual traits. Accordingly, $\lambda$ can be interpreted as varying across settings (e.g., occupations, firms, or industries) while remaining common knowledge within each setting. Importantly, $\lambda$ does not capture the overall importance of coordination in production. Environments where coordination is essential can still exhibit a low $\lambda$. Rather, $\lambda$ measures how strongly increases in individual productivity destabilize collaboration. Section 6 embeds the model in a competitive labor market and discusses how variation in $\lambda$ across environments generates cross-sectional predictions for wages and organizational practices.

**Timing and collaborative fit.** Collaborative fit $\rho$ is revealed only *after partnership formation* but prior to bargaining over the sharing rule. This information structure reflects a basic feature of our motivating applications, including law firm partnerships, joint ventures, firm co-founding, and research collaborations. In these environments, individual talent is often verifiable ex ante through credentials and track records, whereas collaborative fit can be learned only through interaction. In practice, fit often becomes apparent during an initial phase of partnership, such as a probationary period or early development stage, before partners finalize terms such as contracts, credit allocation, or equity splits.

Our notion of collaborative fit is deliberately narrow. The parameter $\rho$ captures the flexibility of the production process and the extent to which partners' efforts combine with low intrinsic frictions. It represents a structural, match-specific dimension of compatibility that is largely stable once the partnership operates and can be revealed through initial interaction. It is not meant to subsume all relationship-specific or evolving features of a partnership. We do not claim that all relevant features of a partnership are learned prior to contracting, nor that collaborative fit ceases to evolve. Many aspects of interaction, such as trust development or responses to unforeseen contingencies, are beyond the scope of the model. Our focus is on partner selection under uncertainty about a production-related dimension of fit that governs the substitutability of efforts.

Importantly, the Moderate's Trap is driven by the systematic tradeoff between productivity and expected collaborative fit, not by uncertainty about fit per se. As shown in Section 5.4, greater residual uncertainty in $\rho_0$ raises the threshold tradeoff intensity required for the trap to emerge and reduces its depth, confirming that uncertainty attenuates rather than amplifies the mechanism. Moreover, the assumption that $\rho$ is revealed prior to bargaining is not essential for the main results. Appendix B.2 considers an alternative timing in which partners bargain before collaborative fit is revealed and commit to the agreed sharing rule. The U-shaped expected surplus persists under this ex ante bargaining structure, albeit at a higher threshold tradeoff intensity.

**Nash bargaining.** To characterize behavior at the bargaining stage, we adopt the Nash bargaining solution with equal bargaining weights. This solution has well-established axiomatic foundations (Nash, 1950) and noncooperative foundations (Rubinstein, 1982; Binmore et al., 1986), and is standard in the relational contracting literature (e.g., Aghion et al., 1994; Miller et al., 2020). It provides a tractable reduced-form representation of efficient agreement and proportional surplus division without requiring explicit modeling of bargaining protocols. This approach allows us to isolate partner selection incentives in the presence of a productivity-collaborative fit tradeoff.

**Bilateral partnerships.** We focus on partnerships between two agents rather than larger teams. This restriction yields a tractable benchmark for studying partner selection under the productivity-fit tradeoff and isolates the core mechanism without the combinatorial complexity of multilateral team formation. In practice, many high-stakes collaborations are bilateral or can be approximated as such: law firm partnerships often operate through pairwise mentoring and case staffing, academic coauthorships frequently involve pairs, and joint ventures typically unite two firms. A fuller treatment allowing for richer team structures and pairwise interactions within teams remains for future work.

# 4 Equilibrium Analysis

We analyze the equilibrium of the continuation game at the bargaining and effort stages, and use the resulting outcomes to characterize expected surplus at the partner selection stage.

## 4.1 Effort Equilibrium

At Stage 3, after the sharing rule $s$ is determined for a given realization of $\rho$, partners simultaneously choose efforts. In equilibrium, given the known pair $(\alpha, \rho)$ and the sharing rule $s$, Partner 1 solves $\max_x \; sF(x, y; \alpha, \rho) - c(x)$ and Partner 2 solves $\max_y \; (1-s)F(x, y; \alpha, \rho) - c(y)$, where $F(x, y; \alpha, \rho)$ is given in (1). Boundary values $s = 0$ or $s = 1$ correspond to degenerate cases that eliminate one partner's incentives to exert effort; hence, we focus on equilibria with interior sharing rules $s \in (0,1)$. For tractability, we specialize to quadratic costs, $c(e) = e^2$, which yields a unique closed-form equilibrium, characterized below.[8]

**Proposition 1** (Effort Equilibrium)**.** *For every $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ and $s \in (0,1)$, given any realized $\rho \in (0,1)$, there exists a unique Nash equilibrium of the effort game, with efforts characterized by (3):*

$$x^*(s;\alpha,\rho) = \frac{s}{2^{1+1/\rho}} \left(1 + \left(\frac{\alpha}{r^*(s;\alpha,\rho)}\right)^{\rho}\right)^{\frac{1-\rho}{\rho}} \quad \textit{and} \quad y^*(s;\alpha,\rho) = \frac{x^*(s;\alpha,\rho)}{r^*(s;\alpha,\rho)}, \tag{3}$$

Let $S(s;\alpha,\rho)$ denote the (total) surplus induced by the effort equilibrium $(x^*, y^*)$ given $(\alpha,\rho)$ and $s$. Then $S(s;\alpha,\rho) = F(x^*,y^*) - (x^*)^2 - (y^*)^2$, where the dependence of $x^*$ and $y^*$ on $s$, $\alpha$, and $\rho$ is suppressed.

## 4.2 Nash Bargaining Outcome

At Stage 2, after the exact value of $\rho$ is revealed, partners bargain over the sharing rule. The equilibrium sharing rule is characterized using the Nash bargaining solution with given bargaining weights. We focus on equal bargaining power, a restriction that is immaterial for our substantive results. The disagreement point is $(0,0)$, reflecting the assumption that dissolution of the partnership precludes further joint production. In practice, partners may have positive autarky payoffs from working alone. We abstract from this possibility to focus on environments where the partnership is the sole productive arrangement, such as joint ventures, coauthorships, or case teams that require collaboration by design.[9]

For any sharing rule $s$, let $(x^*(s), y^*(s))$ denote the equilibrium efforts in the subsequent effort stage, and let $U_1(s)$ and $U_2(s)$ denote the continuation payoffs generated by $(x^*(s), y^*(s))$ given $(\alpha,\rho)$. The Nash bargaining problem chooses a sharing rule $s$ that maximizes the Nash product such that both partners receive at least their outside options:

$$s^*(\alpha,\rho) \in \arg\max_{s \in \mathcal{S}(\alpha,\rho)} U_1(s) \cdot U_2(s), \tag{4}$$

where $\mathcal{S}(\alpha,\rho) = \{s \in [0,1] : U_1(s) \geq 0,\ U_2(s) \geq 0\}$ is the set of feasible sharing rules.

[8] Our qualitative results do not depend on this specific functional form and extend to general cost functions satisfying the assumptions stated in Section 3.1.

[9] Incorporating positive outside options is straightforward: it shifts the disagreement point and reduces the bargaining surplus without altering the structure of the partner selection problem. In the limiting case where autarky payoffs are sufficiently high for the most productive agents, difficult stars may prefer to work alone rather than partner, generating voluntary non-participation at the top of the productivity distribution.

Standard results imply that a solution to the Nash problem exists in our setting. The feasible set for the sharing rule is compact and convex, and the Nash product is continuous in $s$, ensuring existence of a maximizer. While the Nash product attains a unique maximum value whenever both partners' payoffs are strictly positive, the set of sharing rules that achieve this maximum need not be singleton and may, in principle, include boundary points. For any interior solution, the first-order condition $\frac{1}{U_1(s^*)} \cdot \frac{dU_1}{ds} + \frac{1}{U_2(s^*)} \cdot \frac{dU_2}{ds} = 0$ characterizes the maximizer.

**Assumption 1** (Interior Bargaining Solution)**.** The Nash bargaining problem admits an interior solution; that is, there exists $s^* \in (0, 1)$ such that $U_1(s^*) > 0$ and $U_2(s^*) > 0$.

This assumption is not required for uniqueness per se but plays a technical role by ensuring that the equilibrium sharing rule can be characterized by first-order conditions and that symmetry arguments apply.

When productivity is symmetric ($\alpha = 1$), the bargaining problem is symmetric across partners. Any interior maximizer of the Nash product therefore comes in symmetric pairs around 1/2. If the interior maximizer is unique, symmetry implies $s^* = 1/2$; when multiple interior maximizers exist, we select $s^* = 1/2$ as a canonical symmetric selection. This case is formally established in Appendix B.1. When $\alpha \neq 1$, the bargaining problem is generally asymmetric, and the interior maximizer need not equal 1/2. In this case, Assumption 1 guarantees existence of at least one interior maximizer. When the set of maximizers is not singleton, we adopt a tie-breaking rule and select the smallest interior maximizer. This choice is purely technical and serves to ensure a well-defined and measurable selection.

Importantly, neither the assumption of equal bargaining power nor the exact characterization of $s^*$ affects the substantive results of the paper. All subsequent analysis depends only on the induced equilibrium surplus, not on the particular division of surplus between partners. Because all Nash bargaining solutions yield the same surplus, the choice of $s^*$ is payoff-irrelevant for expected surplus comparisons and partner selection. Nevertheless, we model the bargaining stage explicitly to capture the interactive nature of partnerships. The bargaining stage provides a parsimonious way to represent how anticipated negotiations shape equilibrium effort provision and surplus creation, without committing to a specific bargaining protocol or distributional mechanism.

### 4.3 Expected Surplus at Partner Selection Stage

At Stage 1, the principal selects a partner *before* knowing the exact value of $\rho$. Because surplus is divided through bargaining over the sharing rule $s$ after the exact value of $\rho$ is revealed, the partnership environment features transferable utility. As discussed previously, the Nash solution of sharing rule reallocates total surplus across partners without effecting its level, implying that individual payoffs differ only in distribution, not in efficiency. As a result, partner selection is governed by *expected surplus* generated by partnerships rather than individual payoffs.[10]

Formally, at Stage 1, the principal evaluates a potential partner based on expected surplus from the partnership with that partner. In equilibrium, the principal selects an agent with $\alpha$ to maximize expected surplus, where the expectation is taken over $\rho_0$. Let $S^*(\alpha, \rho) \equiv S(s^*(\alpha, \rho); \alpha, \rho)$ denote the surplus induced by the continuation equilibrium $(s^*, x^*, y^*)$ following the realization of $\rho$ given $\alpha$. Then the expected surplus from a partnership with an agent with $\alpha$, under the prior beliefs about $\rho_0$, is given by

$$\bar{S}(\alpha) \equiv \mathbb{E}_{\rho_0} \left[ S^*(\alpha, \rho_0 - p(\alpha; \lambda)) \right], \tag{5}$$

where $p(\alpha; \lambda)$ captures the endogenous shift in collaborative fit induced by productivity.

To characterize partner selection equilibrium, it is necessary to analyze how $\bar{S}(\alpha)$ varies with $\alpha$ when higher productivity comes at the cost of weaker expected collaborative fit.

# 5 Partner Selection

In this section, we derive the paper's central results regarding the shape of $\bar{S}(\alpha)$ and the principal's optimal partner selection at Stage 1.

## 5.1 Preliminaries

The principal's optimal partner choice maximizes expected surplus, given in (5). When there is no confusion, we simply write $\bar{S} = \mathbb{E}[S^*]$. Recall Assumption **DS**, under which $\rho = \rho_0 - p(\alpha; \lambda)$, where $\lambda \geq 0$ denotes the tradeoff intensity. We begin by establishing how $\bar{S}$ responds marginally to changes in productivity and collaborative fit.

**Lemma 1** (Expected Surplus Monotonicity)**.** *(i)* $\partial \bar{S}/\partial \alpha > 0$. *(ii)* $\partial \bar{S}/\partial p \leq 0$.

Lemma 1 shows that a more productive agent generates higher expected surplus holding the shift $p$ in collaborative fit fixed; and weaker collaborative fit (higher $p$) generates lower expected surplus holding productivity fixed.

[10]This criterion is standard in matching models with transferable utility (Becker, 1973; Shimer and Smith, 2000). Although individual payoffs depend on the division of surplus, total match value is invariant to the sharing rule under Nash bargaining. Using Partner 1's expected payoff as the selection criterion leads to the same characterization of the expected surplus function in Theorem 1.

By the chain rule for the total derivative of (5) with respect to $\alpha$, we have

$$\frac{d\bar{S}}{d\alpha} = \underbrace{\frac{\partial \bar{S}}{\partial \alpha}}_{\text{productivity effect } (+)} + \underbrace{\lambda \frac{\partial \bar{S}}{\partial p}}_{\text{collaborative-fit effect } (-)}. \tag{6}$$

By Lemma 1, the two effects operate in opposite directions. An increase in $\alpha$ raises expected surplus through greater individual talent (the productivity effect), but lowers expected surplus by worsening collaborative fit (the collaborative-fit effect).

To simplify notation, we define the *surplus gradient* as $G(\alpha; \lambda) \equiv d\bar{S}/d\alpha$. The sign of $G(\alpha; \lambda)$ determines whether expected surplus is locally increasing or decreasing in productivity. When $\lambda = 0$, $G(\alpha; 0) > 0$ everywhere, meaning expected surplus is strictly increasing in productivity. As $\lambda$ increases, the collaborative-fit effect strengthens relative to the productivity effect. The question is whether, and where, the collaborative-fit effect dominates.

We impose several assumptions on the behavior of $\bar{S}$ over the relevant parameter region to discipline how the two effects in (6) interact. Assumption 2 plays a central role in generating our main result; the remaining assumptions are made for analytical convenience, allowing us to isolate the key economic forces at work without sacrificing substantive generality.

**Assumption 2** (Expected Diminishing Sensitivity)**.** The expected surplus function satisfies $\partial^2 \bar{S}/\partial p^2 > 0$. That is, $|\partial \bar{S}/\partial p|$ is decreasing in $p$.

Unlike the monotonicity properties in Lemma 1, which follow analytically from the CES specification, Assumption 2 is not implied by the production function. Assumption 2 imposes a curvature restriction on the expected surplus function with respect to the shift parameter $p(\alpha; \lambda)$, requiring expected surplus to exhibit diminishing sensitivity to collaborative-fit losses. The assumption formalizes the idea that the marginal value of gaining collaborative fit is highest when collaborative fit is already strongest and declines as fit deteriorates. When partners already work together well, even small reductions in collaborative fit lead to large losses in expected surplus, whereas when collaborative fit is already poor, further losses in fit are less harmful at the margin.

This curvature condition finds support across several empirical settings. In professional sports, Swaab et al. (2014) document that basketball teams with initially strong coordination suffer disproportionately when adding players who disrupt team chemistry, while already-dysfunctional teams experience smaller marginal declines from additional coordination problems. Similarly, Housman and Minor (2015) find that high-performing workplace teams experience larger productivity losses from a single toxic worker than teams where collaboration has already deteriorated. The pattern parallels the intuition underlying Kremer

(1993)'s O-ring theory: in high-quality production processes, a single failure is catastrophic, whereas in low-quality processes, additional failures matter less at the margin.

**Assumption 3** (Boundary)**.**

(i) $G(\underline{\alpha}; \lambda^H) < 0$ for some $\lambda^H \in (0, \lambda^{max})$.

(ii) $G(\overline{\alpha}; \lambda) > 0$ for all $\lambda \in (0, \lambda^H)$.

Assumption 3(i) requires that, for some tradeoff intensity, the collaborative-fit effect dominates the productivity effect near the *lower* boundary of the productivity range. Intuitively, for agents with near-minimal talent, value creation relies primarily on strong collaborative fit, making marginal losses in fit more costly than marginal productivity gains. Assumption 3(ii) requires that, for relatively low tradeoff intensities, the productivity effect dominates the collaborative-fit effect near the *upper* boundary of the productivity range. Intuitively, agents with near-maximal talent continue to generate high expected surplus despite being difficult to work with, because their exceptional individual talent more than compensates for reduced collaborative fit. Assumption 3 does not require these dominances to hold for all tradeoff intensities, but only that such regimes are feasible.

**Assumption 4** (Regularity)**.** There exist finite constants $\kappa > 0$ and $\delta > 0$ such that:

(i) $-(\partial \bar{S}/\partial p)|_{\alpha=\underline{\alpha}} > (1 - \underline{\alpha})\delta$;

(ii) $\partial^2 \bar{S}/\partial p^2 \geq \kappa$, $|\partial^2 \bar{S}/\partial p \partial \alpha| \leq \delta$, and $\partial^2 \bar{S}/\partial \alpha^2 \geq -\delta$ for all relevant $(\alpha, \rho)$.

These regularity conditions serve two purposes. Condition (i) ensures that the threshold tradeoff intensity is well defined by guaranteeing that the collaborative-fit effect is sufficiently strong at the lower boundary. Condition (ii) imposes uniform curvature and bounded cross-effects on the expected surplus function, ensuring that it admits a unique interior minimizer. Note that $\partial^2 \bar{S}/\partial \alpha^2$ may fail to exist at $\alpha = 1$, but this non-differentiability does not invalidate our subsequent analysis (see the proof of Theorem 1 in Appendix A). Taken together, Assumption 4 guarantees a single reversal in the slope of $\bar{S}(\alpha)$, which underlies the U-shaped relationship between expected surplus and productivity derived below.

## 5.2 The Moderate's Trap

We state the paper's main result.

**Theorem 1** (Shape of Expected Surplus)**.** *Suppose Assumption **DS** and Assumptions 1–4 hold. Then there exists a unique threshold $\bar{\lambda} \in (0, \lambda^H)$ defined by $G(\underline{\alpha}; \bar{\lambda}) = 0$ such that:*

*(i) For mild tradeoff intensity* $(0 < \lambda \leq \bar{\lambda})$*:* $\bar{S}(\alpha)$ *is strictly increasing in* $\alpha$*.*

*(ii) For strong tradeoff intensity* $(\lambda > \bar{\lambda})$*:* $\bar{S}(\alpha)$ *is U-shaped in* $\alpha$*; that is, it is decreasing on* $[\underline{\alpha}, \alpha_{min})$ *and increasing on* $(\alpha_{min}, \overline{\alpha}]$*, with a unique interior minimizer* $\alpha_{min} \in (\underline{\alpha}, \overline{\alpha})$*.*

When the tradeoff intensity is mild $(\lambda \leq \bar{\lambda})$, the productivity effect dominates throughout the productivity range. Collaborative fit plays a limited role, so increases in productivity translate directly into higher surplus. This regime may describe environments in which individual productivity is closely tied to joint output and high performance cannot be realized without effective collaboration. In such settings, the conventional wisdom that more talent is better holds.

When the tradeoff intensity is relatively strong $(\lambda > \bar{\lambda})$, this wisdom no longer applies. The model admits a "stuck in the middle" phenomenon, which we call the *Moderate's Trap*. Low-productivity agents, referred to as team players, generate value through strong collaborative fit: a highly flexible production process with low coordination frictions overpowers limited individual talent. Highly productive agents, or difficult stars, generate value through exceptional individual talent, compensating for weak collaborative fit. Moderate-productivity agents possess neither advantage. Their productivity is insufficient to generate substantial value independently, while their collaborative fit is too weak to sustain low-friction production, leaving them trapped between the two effective modes of value creation.

The U-shaped pattern of expected surplus that defines the Moderate's Trap follows from diminishing sensitivity to collaborative-fit losses. The key economic content of this property is that collaborative fit matters most when it is already strong: when collaborative fit is strong, small reductions in fit are particularly costly, so moving from a team player with strong expected collaborative fit toward a moderate agent entails a large loss in surplus. When collaborative fit is already weak, further reductions have limited impact, so moving from a moderate agent toward a difficult star sacrifices relatively little. This asymmetry renders the middle of the productivity range the least attractive region and gives rise to the trap faced by moderately productive agents.

We adopt a CES production function for tractability, but the underlying economic mechanism is more general. Our main results follow from properties of the expected surplus function stated in Lemma 1 and Assumption 2. The CES specification does not mechanically deliver Assumption 2. Any production technology that yields expected surplus satisfying these conditions can generate the Moderate's Trap when the tradeoff intensity is sufficiently strong. The key economic content of Assumption 2 is that collaborative fit matters most when it is already strong—a property we view as empirically plausible across a broad class of collaborative environments. The trap thus reflects a fundamental feature of the environment rather than an artifact of functional form.

**Remark** (The Role of Uncertainty)**.** The Moderate's Trap arises from the tradeoff between productivity and expected collaborative fit, not from uncertainty about fit per se. The principal's selection problem depends on $\bar{S}(\alpha) = \mathbb{E}_{\rho_0}[S^*(\alpha, \rho)]$, where higher $\alpha$ predicts lower $\mathbb{E}[\rho|\alpha]$ under Assumption DS. The residual uncertainty in $\rho_0$ (variation in fit that is orthogonal to productivity) affects expected surplus but does not drive the U-shape. Indeed, Proposition 3 shows that greater dispersion in $\rho_0$ raises the threshold $\bar{\lambda}$ and reduces the depth of the trap, confirming that residual uncertainty attenuates rather than amplifies the mechanism.

## 5.3 Bimodal Selection

We now study the principal's optimal partner selection at Stage 1. Let $\alpha^*$ denote the productivity level of the optimal partner: $\alpha^* = \arg\max_\alpha \bar{S}(\alpha)$. We simply use $\alpha^*$ to refer to the optimally chosen agent with that productivity level.

When $\lambda = 0$ so that collaborative fit is independent of individual talent, the partner selection problem is trivial: $\bar{S}$ is strictly increasing in $\alpha$ by Lemma 1(i); hence the optimal partner is $\alpha^* = \overline{\alpha}$, the most productive agent in the pool. The following describes optimal partner selection in equilibrium when $\lambda > 0$.

**Corollary 1** (Optimal Partner Selection)**.**

*(i) For mild tradeoff intensity* $(0 < \lambda \leq \bar{\lambda})$*: The optimal partner choice is* $\alpha^* = \overline{\alpha}$.

*(ii) For strong tradeoff intensity* $(\lambda > \bar{\lambda})$*: The optimal partner choice lies at a boundary, that is,* $\alpha^* \in \{\underline{\alpha}, \overline{\alpha}\}$.

When highly productive agents are only mildly difficult to work with ($\lambda \leq \bar{\lambda}$), the same conclusion holds as in the case with $\lambda = 0$ by Theorem 1(i). Hence, the optimal choice is $\alpha^* = \overline{\alpha}$, the most productive agent in the pool.

When highly productive agents are sufficiently difficult to work with ($\lambda > \bar{\lambda}$), the U-shaped relationship between expected surplus and productivity fundamentally alters the selection problem. Because any interior productivity level yields strictly lower expected surplus than at least one boundary type, no interior agent $\alpha \in (\underline{\alpha}, \overline{\alpha})$ can be optimal. This establishes a pattern of *bimodal selection*: when making a partner selection decision, the principal need only compare the least productive agent ($\underline{\alpha}$) with the most productive agent ($\overline{\alpha}$) in the candidate pool, while strictly avoiding intermediate types. The implication for hiring decisions is stark: rather than seeking well-rounded agents in the middle of the candidate pool when ranked by observable credentials, firms should target the extremes.

By Corollary 1(ii), the principal's optimal partner choice reduces to a comparison of expected surplus at the two boundary productivity levels. Specifically, $\alpha^* = \overline{\alpha}$ if $\bar{S}(\overline{\alpha}) > \bar{S}(\underline{\alpha})$ and $\alpha^* = \underline{\alpha}$ if $\bar{S}(\overline{\alpha}) < \bar{S}(\underline{\alpha})$, with the principal randomizing between the two when $\bar{S}(\overline{\alpha}) = \bar{S}(\underline{\alpha})$. We label the $\underline{\alpha}$-agent the "best team player" and the $\overline{\alpha}$-agent the "superstar," reflecting their respective strengths in collaborative fit and individual talent. We now identify sufficient conditions under which the principal prefers to select the superstar over the best team player.

**Proposition 2** (Superstar vs. Best Team Player)**.** *Suppose $\lambda > \bar{\lambda}$. The principal selects the superstar ($\overline{\alpha}$) over the best team player ($\underline{\alpha}$) if (i) the tradeoff intensity $\lambda$ is relatively low, or (ii) the productivity range $[\underline{\alpha}, \overline{\alpha}]$ is relatively tilted toward higher productivity levels (i.e., a higher $\overline{\alpha}$ or a higher $\underline{\alpha}$).*

The comparison between the superstar and the best team player hinges on how productivity gains trade off against losses in collaborative fit.

First, consider the extensive margin, captured by the tradeoff intensity $\lambda$. When $\lambda$ is only moderately above $\bar{\lambda}$, stars are difficult to work with but not prohibitively so; and the productivity advantage of the superstar dominates the collaborative-fit advantage of the best team player, favoring selection of the superstar. As $\lambda$ increases further, collaborative-fit losses at high productivity become more severe, reducing expected surplus at the superstar boundary. At the same time, collaborative-fit gains at low productivity become more pronounced, increasing expected surplus at the best team player boundary. These opposing movements shift the relative advantage toward the best team player. Thus, changes in $\lambda$ can generate a switching pattern in optimal boundary selection.

Second, consider the intensive margin, captured by the productivity range $[\overline{\alpha}, \underline{\alpha}]$. An increase in the upper boundary $\overline{\alpha}$ expands the candidate pool to include substantially more productive agents, extending the right arm of the U-shape upward and making selection of the superstar more attractive. Conversely, a decrease in the lower boundary $\underline{\alpha}$ expands the pool to include candidates with considerably stronger expected collaborative fit, extending the left arm of the U-shape upward and making the best team player more favorable.

The extensive margin captures how variation in tradeoff intensity across environments affects optimal selection, whereas the intensive margin captures how, within a given environment, changes in the composition of the candidate pool shape the optimal choice.

The bimodal selection result admits a stronger interpretation. While Corollary 1(ii) establishes that the optimal partner lies at a boundary, some interior productivity levels may still generate expected surplus between the two boundary values. However, productivity levels sufficiently close to the interior minimizer $\alpha_{min}$ are strictly dominated by *both* boundaries. Formally, the set $\mathcal{D}(\lambda) \equiv \{\alpha \in [\underline{\alpha}, \overline{\alpha}] : \bar{S}(\alpha) < \min\{\bar{S}(\underline{\alpha}), \bar{S}(\overline{\alpha})\}$ is nonempty for $\lambda > \bar{\lambda}$.

Agents in this intermediate range are categorically unattractive; they would be passed over regardless of whether the principal ultimately prefers the superstar or the best team player.

## 5.4 Uncertainty about Collaborative Fit

The preceding analysis establishes that the Moderate's Trap arises when the tradeoff intensity $\lambda$ exceeds the threshold $\bar{\lambda}$. A natural question is how residual uncertainty about collaborative fit, captured by the dispersion of $\rho_0$, affects this threshold and the depth of the trap. This question helps clarify the mechanism: if the trap arose from uncertainty itself, greater uncertainty should amplify it; if instead the trap arises from the systematic productivity-fit tradeoff, greater residual uncertainty should attenuate it by diluting the informativeness of productivity as a signal of expected fit.

Recall that uncertainty enters through the baseline substitution parameter $\rho_0$, which follows a distribution $\mathcal{H}$ with mean $\mathbb{E}[\rho_0]$ and variance $\tau^2$. Our main results imposed no restriction on $\tau^2$. Introducing an explicit variance parameter makes clear that the threshold $\bar{\lambda}$ depends on $\tau^2$; we write $\bar{\lambda}(\tau^2)$ to denote this dependence. When $\lambda > \bar{\lambda}(\tau^2)$, define the *depth* of the Moderate's Trap as $\bar{S}(\underline{\alpha}) - \bar{S}(\alpha_{min})$, the surplus gap between the best team player and the interior minimum.

**Proposition 3** (Dispersion of $\rho_0$ and the Moderate's Trap)**.** *Suppose the assumptions of Theorem 1 hold. Then:*

*(i)* $\bar{\lambda}(\tau^2)$ *is strictly increasing in* $\tau^2$.

*(ii)* $\bar{S}(\underline{\alpha}) - \bar{S}(\alpha_{min})$ *is strictly decreasing in* $\tau^2$ *for any given* $\lambda > \bar{\lambda}(\tau^2)$.

Greater uncertainty about collaborative fit raises the threshold tradeoff intensity required for the trap to arise and attenuates its depth when it does arise. The intuition is straightforward. Productivity predicts collaborative fit: higher productivity signals weaker fit. When fit is relatively predictable (low $\tau^2$), this signal is informative, and the tradeoff translates into a sharply U-shaped surplus profile. Greater uncertainty weakens the informativeness of productivity, mitigating the impact of collaborative-fit losses on expected surplus. A stronger tradeoff intensity is then required for the trap to emerge, and the resulting U-shape is shallower. The trap is therefore most pronounced in environments where collaborative fit is relatively predictable at the partner selection stage.

This comparative static suggests that the Moderate's Trap should be more salient in settings where reputations for collaborative behavior are well established and observable. In mature industries with stable professional networks, such as law, investment banking, or Hollywood, a candidate's collaborative track record is often known through references, repeated interactions, and industry reputation. A partner at a law firm considering lateral hires can draw on decades of relationship history within the legal community; a film producer assembling a cast can consult well-established knowledge about which actors are "difficult." In such environments, productivity remains a reliable signal of expected collaborative fit, and the trap operates at full force.

By contrast, the trap should be attenuated in settings where collaborative fit is inherently harder to predict. Early-career hiring presents one such case: candidates lack the track record that would allow employers to infer fit from productivity. Cross-industry transitions present another: a star performer's collaborative reputation in one sector may not transfer to a new context with different norms and workflows. Emerging industries or novel team structures similarly lack the accumulated reputational information that sharpens the productivity-fit signal. In these high-uncertainty environments, principals face greater residual variance in collaborative fit conditional on productivity, weakening the case for bimodal selection and pushing partner choice back toward conventional productivity-based rankings.

Appendix C examines how overconfidence can distort partner selection, showing that productivity overconfidence and collaborative-fit overconfidence push selection in opposite directions and that correcting only one bias can induce mis-selection.

# 6 Labor Markets

The preceding analysis characterizes partner selection by a single principal. We now embed the model in a competitive labor market to study equilibrium wages and investment incentives. This extension addresses a natural question: if moderately productive agents generate lower expected surplus, how does this disadvantage manifest when all workers are employed in equilibrium? The answer is that the Moderate's Trap appears not as unemployment but as wage compression in the middle of the productivity range.

## 6.1 Competitive Equilibrium and Wages

Consider a labor market with a unit mass of principals (firms) and a unit mass of agents (workers). Workers are heterogeneous in productivity $\alpha$, distributed according to a continuous distribution $\mathcal{F}$ with full support on $[\underline{\alpha}, \overline{\alpha}]$. Each firm seeks to hire exactly one worker, and each worker can be hired by at most one firm. Workers have a common outside option normalized to zero.

Let $w(\alpha)$ denote the wage paid to a worker of type $\alpha$. A firm that hires a worker of type $\alpha$ earns profit $\Pi(\alpha) = \bar{S}(\alpha) - w(\alpha)$. In a competitive equilibrium, firms take wages as given and choose which worker type to hire, workers accept employment if wages weakly exceed their outside option, and markets clear.

**Proposition 4** (Equilibrium Wages)**.** *Suppose* $\bar{S}(\alpha) > 0$ *for all* $\alpha \in [\underline{\alpha}, \overline{\alpha}]$*. Then:*

*(i) A competitive equilibrium exists in which all worker types are employed. In any such equilibrium, firms earn equal profits across all worker types, and wages satisfy* $w(\alpha) - w(\alpha') = \bar{S}(\alpha) - \bar{S}(\alpha')$ *for all* $\alpha, \alpha' \in [\underline{\alpha}, \overline{\alpha}]$*.*

*(ii) For mild tradeoff intensity* $(0 < \lambda \leq \bar{\lambda})$*:* $w(\alpha)$ *is strictly increasing in* $\alpha$*.*

*(iii) For strong tradeoff intensity* $(\lambda > \bar{\lambda})$*:* $w(\alpha)$ *is U-shaped in* $\alpha$*, with a minimum at* $\alpha_{min}$*.*

The key insight is that equilibrium wage differences mirror differences in expected surplus. If some worker type were to generate strictly higher profits than another at prevailing wages, firms would compete for that type, bidding up its wage until profits equalize. The equilibrium wage function therefore inherits the shape of the expected surplus function: strictly increasing when the tradeoff intensity is mild, U-shaped when the tradeoff intensity is strong.

This result resolves an apparent tension with bimodal partner selection. In the single-principal model, when $\lambda > \bar{\lambda}$, only boundary types are ever selected. In the labor market, by contrast, all workers are employed. The difference is that wages adjust: moderately productive workers become sufficiently inexpensive that firms are willing to hire them despite their lower surplus contribution. The Moderate's Trap therefore manifests not as exclusion from employment but as a wage penalty for workers in the middle of the productivity range.

Proposition 4 characterizes wages in an environment where the Moderate's Trap is the sole determinant of surplus. Observed wage distributions reflect additional forces, including technological complementarities and returns to skill, that generate monotonically increasing baseline wages. Section 6.3 discusses how our mechanism interacts with these forces to produce relative wage compression in the middle rather than U-shaped wages in levels.

## 6.2 Endogenous Investment

We endogenize the productivity distribution by allowing agents to invest in productivity prior to entering the market. We first analyze investment incentives in the single-principal environment to establish the basic mechanism, then show how these incentives generate polarization in the competitive labor market.

### Single-Principal Environment

Extend the partnership game to include an initial investment stage. At Stage 0, each agent $k$ chooses an investment level $i_k \in [0, 1]$ without knowing the realization of collaborative fit. Investment is observable, costless, and raises effective productivity according to

$$\alpha(i_k; \theta_k) = \theta_k + \gamma i_k, \tag{7}$$

where $\theta_k \in [\underline{\theta}, \overline{\theta}]$ denotes agent $k$'s baseline talent and $\gamma > 0$ is the common return to investment. At Stage 1, the principal observes all agents' effective productivity levels and selects a partner. The game then proceeds as specified in Section 3.2.

Collaborative fit continues to be governed by Assumption **DS**, with the baseline substitution parameter $\rho_0$ unknown at both Stages 0 and 1. Investment affects expected collaborative fit only indirectly through productivity: agents who invest more become more productive but, in expectation, more difficult to work with.

Let $\bar{\bar{\lambda}} \in (\bar{\lambda}, \lambda^{max})$ satisfy $\bar{S}(\underline{\theta}) = \bar{S}(\overline{\theta}+\gamma)$. By Corollary 1 and Proposition 2, the principal selects $\alpha^* = \overline{\theta} + \gamma$ for $\lambda \in (\bar{\lambda}, \bar{\bar{\lambda}})$ and $\alpha^* = \underline{\theta}$ for $\lambda \in (\bar{\bar{\lambda}}, \lambda^{max})$.

**Proposition 5** (Optimal Investment)**.** *Suppose $\lambda > \bar{\lambda}$ and the investment return $\gamma$ is sufficiently large that $\bar{S}(\overline{\theta} + \gamma) > \bar{S}(\overline{\theta})$ and $\bar{S}(\underline{\theta}) > \bar{S}(\underline{\theta} + \gamma)$.*

*(i) For $\lambda < \bar{\bar{\lambda}}$, the agent with $\theta_k = \overline{\theta}$ chooses $i_k^* = 1$.*

*(ii) For $\lambda = \bar{\bar{\lambda}}$, the agent with $\theta_k = \overline{\theta}$ chooses $i_k^* = 1$; the agent with $\theta_k = \underline{\theta}$ chooses $i_k^* = 0$.*

*(iii) For $\lambda > \bar{\bar{\lambda}}$, the agent with $\theta_k = \underline{\theta}$ chooses $i_k^* = 0$.*

*In all cases, all other agents are indifferent over investment choices.*

The proposition reveals that only a single extreme agent—the one who will be selected in equilibrium—has strict investment incentives. When $\lambda < \bar{\bar{\lambda}}$, the principal selects the most productive agent, so the highest-talent agent invests fully to guarantee selection. When $\lambda > \bar{\bar{\lambda}}$, the principal selects the least productive agent, so the lowest-talent agent refrains from investment. All other agents are never selected regardless of their investment and are therefore indifferent. Endogenous investment thus reinforces the principal's extreme selection rather than smoothing the productivity distribution.

### Competitive Labor Market

The indifference of non-selected agents disappears in a competitive labor market. When all agents are employed and wages reflect expected surplus, every agent has strict investment incentives, and the Moderate's Trap generates a sharper prediction: bimodal investment that polarizes the productivity distribution.

**Proposition 6** (Bimodal Investment under Competitive Labor Market)**.** *Suppose $\lambda > \bar{\lambda}$ and the investment return $\gamma$ is sufficiently large.*

*(i) In equilibrium, all agents choose corner investment levels: $i_k^* \in \{0, 1\}$ for all $k$.*

*(ii) There exists a unique cutoff $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$ such that agents with $\theta_k > \hat{\theta}$ choose $i_k^* = 1$, while agents with $\theta_k < \hat{\theta}$ choose $i_k^* = 0$. An agent with $\theta_k = \hat{\theta}$ is indifferent.*

*(iii) The equilibrium productivity support is $[\underline{\theta}, \hat{\theta}] \cup [\hat{\theta} + \gamma, \overline{\theta} + \gamma]$.*

In a competitive labor market, wages allow agents to internalize surplus differences fully (Proposition 4). Given the U-shaped expected surplus, intermediate productivity levels are penalized while extremes are rewarded. Anticipating this payoff structure, agents avoid intermediate investment choices and specialize. Sorting across extremes follows comparative advantage: because productivity gains are more valuable at higher productivity levels, high-talent agents invest fully to become difficult stars, while low-talent agents do not invest and remain team players.

The resulting productivity distribution is polarized, with a gap in the middle. No agents optimally choose investment levels that generate productivity in $(\hat{\theta}, \hat{\theta} + \gamma)$—precisely the region where expected surplus is lowest. Rather than producing smooth sorting, investment incentives hollow out the middle of the productivity distribution. The Moderate's Trap is thus self-reinforcing: bimodal market rewards induce investment patterns that sustain the bimodal structure.

## 6.3 Relation to Labor Market Polarization

The mechanism we identify connects to a large literature documenting wage and employment polarization. Several studies find growth concentrated at the top and bottom of the skill distribution (Autor et al., 2003; Autor and Dorn, 2013; Goos and Manning, 2007). Standard explanations emphasize technological change that displaces routine middle-skill tasks while complementing high-skill cognitive work and preserving demand for low-skill manual and service work (Autor and Dorn, 2013).

Our model, in isolation, predicts U-shaped wages in levels: moderately productive workers earn less than both difficult stars and team players. Observed wage distributions do not exhibit this pattern; wages generally increase with measured productivity, even if growth rates differ across the distribution. This discrepancy reflects the fact that real-world wages are determined by many forces beyond the mechanism we study, including technological complementarities, institutional constraints, search frictions, and human capital accumulation. Our contribution is not to claim that wages are U-shaped in practice, but to microfound a demand-side force that operates against moderate performers. When this force is combined with other forces that generate monotonically increasing wages, such as standard returns to skill, the net effect compresses wages in the middle of the distribution relative to what standard monotonic models would predict.

Two channels can generate the observed pattern of wage polarization through our mechanism. First, within industries, the tradeoff intensity $\lambda$ may increase over time as work becomes more collaborative and knowledge-intensive. As $\lambda$ rises, the U-shape becomes more pronounced: wages at the extremes rise relative to wages in the middle, generating polarization in growth rates even if wages remain monotonic in levels. Second, to the extent that the economy shifts toward partnership-intensive sectors, the tradeoff becomes more globally relevant. In routine production tasks, individual productivity may dominate the collaborative fit (low $\lambda$), whereas in professional services and knowledge-intensive work, effective collaboration is essential and stars tend to be difficult (high $\lambda$). As partnership-intensive industries expand, the U-shaped wage modifier becomes more prevalent in the aggregate, depressing wages in the middle of the overall distribution relative to the extremes. Both channels provide a demand-side complement to explanations based on routine-biased technological change: polarization arises not only from task displacement but also from the expanding scope of environments where collaborative fit commands a premium.

Cross-sectional variation in $\lambda$ generates testable implications. In settings where individual productivity is highly scalable and tasks are largely modular, - such as law, consulting, or investment banking - a high performer can generate value relatively independently, and the premium on collaboration is correspondingly reduced. Increases in productivity are more likely associated with losses in collaborative fit, implying a higher $\lambda$, and wage compression in the middle should be more pronounced. By contrast, in settings where individual productivity is tightly linked to joint output and cannot be realized without effective collaboration - such as research teams or ensemble-based creative work - high performers remain dependent on working well with others. In such environments, $\lambda$ is lower, and wages should remain closer to monotonic in productivity.

## 6.4 Organizational Practices

The Moderate's Trap provides a unified rationale for several organizational practices that systematically favor workers at the extremes of the productivity distribution. We discuss three settings in which these patterns are particularly salient.

**Up-or-out and elastic tournaments.** In professional services such as law, consulting, and investment banking, firms commonly employ up-or-out rules that retain only exceptional performers while separating competent but unexceptional employees (Galanter and Palay, 1991; Kahn and Huberman, 1988). Galanter and Henderson (2008) document an evolution toward "elastic tournaments" in which the binary up-or-out structure gives way to a multi-tiered hierarchy: a thin equity core of rainmakers, a permanent non-equity tier of service partners and of-counsel, and de-equitization of those in between. The existing literature explains these practices through tournament incentives (Lazear and Rosen, 1981), human capital externalities, and signaling (Waldman, 1990). Our framework suggests an additional rationale grounded in surplus creation: when the productivity-fit tradeoff is strong, moderate performers generate lower expected surplus than either difficult stars or team players. The structure of modern professional firms—reserving equity ownership for star rainmakers at one end while leveraging juniors at the other, and confining moderates to the non-equity mantle—is consistent with bimodal selection under the Moderate's Trap.

**Bimodal hiring in technology firms.** Technology firms frequently target candidates at opposite ends of the spectrum: exceptional individual contributors or strong culture-fit candidates, often rejecting the middle (Bock, 2015). This aligns with the tradeoff between competence and likability: Casciaro and Lobo (2005) demonstrate that managers frequently select "lovable fools" (high fit, moderate competence) over "competent jerks," prioritizing network cohesion over raw individual productivity. Conversely, high-performance cultures like Netflix explicitly enforce bimodal selection by screening out adequate performers entirely to maintain talent density (Hastings and Meyer, 2020). Similarly, elite professional service firms often prioritize cultural matching over technical credentials for non-star roles (Rivera, 2015). These practices reflect the logic of the Moderate's Trap: when the production function depends on both individual output and collaborative fit, intermediate types generate less expected surplus than the boundary types who specialize in one dimension.

**Team composition in sports and entertainment.** In professional sports, teams with high concentrations of top performers suffer from diminishing returns due to coordination failures—the "too-much-talent" effect documented by Swaab et al. (2014). Instead, teams rely on productivity spillovers from role players who—despite lower individual stats—enhance the output of stars (Arcidiacono et al., 2017), while shared tenure creates the tacit knowledge necessary for collaborative advantage (Berman et al., 2002). This creates an incentive for teams to assemble rosters that mix superstars with cooperative role players rather than accumulating mid-tier talent. Parallel dynamics appear in entertainment: Elberse (2013) documents a "blockbuster strategy" where studios rely on marquee stars to anchor projects, hollowing out the middle tier. This is consistent with the "motley crew" property described by Caves (2000), where the risk that a moderately talented misfit derails a complex project leads decision-makers to prefer safe, high-fit collaborators. Across these settings, the non-monotonic valuation of talent suggests that moderate types contribute less expected value than specialists at either extreme.

# 7 Conclusion

This paper develops a theory of partner selection in environments where individual productivity and collaborative fit are negatively related. We characterize how expected surplus varies with partner productivity and identify two regimes. When the productivity-fit tradeoff is mild, conventional wisdom holds: higher productivity is always preferred. When the tradeoff is sufficiently strong, expected surplus becomes U-shaped in productivity, and moderate performers generate the lowest expected surplus. We refer to this pattern as the Moderate's Trap.

The trap arises because moderate performers lack both sources of value creation present at the extremes. Team players generate surplus through strong collaborative fit, while difficult stars generate surplus through exceptional individual talent. Moderate types benefit from neither mechanism. As a result, partner selection exhibits a bimodal pattern: firms need only compare the best team player with the superstar, bypassing the middle of the distribution.

We embed this mechanism in a competitive labor market. The U-shaped surplus profile translates into wages that compress in the middle, with moderate workers disadvantaged relative to both team players and difficult stars. Moreover, the trap can be self-reinforcing through endogenous investment: when market wages penalize intermediate productivity, workers invest to specialize at the extremes, hollowing out the middle. This demand-side force operates alongside technology-driven explanations and may help explain why polarization patterns persist across sectors with varying technological exposures. The framework also rationalizes organizational practices that systematically avoid moderate performers, including up-or-out promotion systems, bimodal hiring strategies, and team compositions that mix superstars with role players.

Several directions for future research merit attention. First, the difficult stars assumption that links higher productivity to weaker expected collaborative fit is central to our results but may not apply uniformly across settings. Identifying environments in which this tradeoff is empirically relevant remains an important task. Second, we focus on bilateral partnerships with a simple information structure in which productivity is observable while collaborative fit is uncertain. In practice, firms obtain partial signals about fit through interviews or references. Extending the model to richer information structures, including partial revelation of fit prior to matching, would connect our framework more directly to screening and signaling applications. Third, introducing heterogeneity across principals would connect our framework to the sorting literature. Because high-productivity agents bring both a productivity advantage and a collaborative-fit disadvantage, joint surplus may be neither globally supermodular nor submodular in principal and agent types, potentially generating non-standard sorting patterns. Fourth, dynamic extensions could incorporate learning about collaborative fit over time, relationship-specific investment, and partnership dissolution. In our static framework, partnerships form under uncertainty and outcomes are realized. In practice, early signals of poor fit may trigger renegotiation or separation, and the anticipation of such dynamics could alter initial selection decisions. Modeling these forces would yield richer predictions about career trajectories and the evolution of partnerships over time.

The core insight is general: when individual talent and collaborative fit are in tension, standard monotonic relationships between ability and value creation need not hold. By formalizing this tradeoff and tracing its implications for partner selection, wages, and investment, we provide a foundation for future empirical and theoretical work on team formation and the organization of work.

# References


Aghion, P., M. Dewatripont, and P. Rey (1994). Renegotiation design with unverifiable information. *Econometrica 62*(2), 257–282.

Arcidiacono, P., J. Kinsler, and J. Price (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics 35*(1), 191–225.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

Azoulay, P., J. Graff Zivin, and J. Wang (2010). Superstar extinction. *Quarterly Journal of Economics 125*(2), 549–589.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy 81*(4), 813–846.

Berman, S. L., J. Down, and C. W. Hill (2002). Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Academy of Management Journal 45*(1), 13–31.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The Nash bargaining solution in economic modelling. *RAND Journal of Economics 17*(2), 176–188.

Bock, L. (2015). *Work Rules!: Insights from Inside Google That Will Transform How You Live and Lead*. Twelve.

Börner, K., N. Contractor, H. J. Falk-Krzesinski, S. M. Fiore, K. L. Hall, J. Keyton, B. Spring, D. Stokols, W. Trochim, and B. Uzzi (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine 2*(49), 49cm24.

Casciaro, T. and M. S. Lobo (2005). Competent jerks, lovable fools, and the formation of social networks. *Harvard Business Review 83*(6), 92–99.

Caves, R. E. (2000). *Creative Industries: Contracts between art and commerce*. Harvard University Press.

Che, Y.-K. and T.-Y. Chung (1999). Contract damages and cooperative investments. *RAND Journal of Economics 30*(1), 84–105.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics 132*(4), 1593–1640.

Elberse, A. (2013). *Blockbusters: Hit-making, risk-taking, and the big business of entertainment*. Henry Holt and Company.

Fiske, S. T., A. J. C. Cuddy, P. Glick, and J. Xu (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology 82*(6), 878–902.

Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics 123*(1), 49–100.

Galanter, M. and W. D. Henderson (2008). The elastic tournament: A second transformation of the big law firm. *Stanford Law Review 60*, 1867–1929.

Galanter, M. and T. Palay (1991). *Tournament of Lawyers: The Transformation of the Big Law Firm*. University of Chicago Press.

Glover, J. and E. Kim (2021). Optimal team composition: Diversity to foster implicit team incentives. *Management Science 67*(9), 5800–5820.

Goos, M. and A. Manning (2007). Lousy and lovely jobs: The rising polarization of work in britain. *Review of Economics and Statistics 89*(1), 118–133.

Groysberg, B., L.-E. Lee, and A. Nanda (2008). Can they take it with them? The portability of star knowledge workers' performance. *Management Science 54*(7), 1213–1230.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy 111*(3), 465–497.

Hastings, R. and E. Meyer (2020). *No Rules Rules: Netflix and the Culture of Reinvention.* New York: Penguin Press.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464.

Holmström, B. (1982). Moral hazard in teams. *Bell Journal of Economics 13*(2), 324–340.

Housman, M. and D. Minor (2015). Toxic workers. Working Paper 16-057, Harvard Business School.

Imhof, L. A. and M. Kräkel (2023). Team diversity and incentives. *Management Science 69*(4), 2497–2516.

Kahn, C. and G. Huberman (1988). The economics of up-or-out rules in academia. *Journal of Labor Economics 6*(4), 461–480.

Kambhampati, A. and C. Segura-Rodriguez (2022). The optimal assortativity of teams inside the firm. *RAND Journal of Economics 53*(3), 484–515.

Kambhampati, A., C. Segura-Rodriguez, and P. Shao (2024). Why informationally diverse teams need not form, even when efficient. *Journal of Economic Behavior & Organization 226*, 106689.

Kaya, A. and G. Vereshchagina (2022). Sorting expertise. *Journal of Economic Theory 204*, 105497.

Kremer, M. (1993). The O-ring theory of economic development. *Quarterly Journal of Economics 108*(3), 551–575.

Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*(5), 841–864.

Legros, P. and A. F. Newman (2007). Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica 75*(4), 1073–1102.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *Review of Economic Studies 84*(2), 718–789.

Lise, J. and F. Postel-Vinay (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review 110*(8), 2328–2376.

Mello, A. S. and M. E. Ruckes (2006). Team composition. *Journal of Business 79*(3), 1019–1039.

Miller, D. A., T. E. Olsen, and J. Watson (2020). Relational contracting, negotiation, and external enforcement. *American Economic Review 110*(7), 2153–2197.

Nash, J. F. (1950). The bargaining problem. *Econometrica 18*(2), 155–162.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences 101*(suppl 1), 5200–5205.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science 58*(6), 1122–1140.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review 46*(7), 1187–1207.

Rivera, L. A. (2015). *Pedigree: How elite students get elite jobs.* Princeton University Press.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica 33*(3), 520–534.

Rosen, S. (1981). The economics of superstars. *American Economic Review 71*(5), 845–858.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica 50*(1), 97–109.

Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica 68*(2), 343–369.

Swaab, R. I., M. Schaerer, E. M. Anicich, R. Ronay, and A. D. Galinsky (2014). The too-much-talent effect: Team interdependence determines when more talent is too much or not enough. *Psychological Science 25*(8), 1581–1591.

Waldman, M. (1990). Up-or-out contracts: A signaling perspective. *Journal of Labor Economics 8*(2), 230–250.

Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics 96*(5), 849–861.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in production of knowledge. *Science 316*(5827), 1036–1039.

Zeitoun, H., T. Melkonyan, and N. Chater (2023). The social contract in miniature: How virtual bargaining supports team production. *Academy of Management Review 48*(3), 393–413.

# Appendices

# A Proofs

**Proof of Proposition 1 (Effort Equilibrium).**

Fix $(\alpha, \rho, s)$ with $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, $\rho \in (0,1)$, and $s \in (0,1)$.

*Existence and unique best responses.* For $\rho \in (0,1)$, the CES production function $F(x, y; \alpha, \rho)$ is strictly concave in $(x, y)$ on $\mathbb{R}^2_{++}$, and $c(e) = e^2$ is strictly convex. Hence each player's payoff $U_i$ is strictly concave in own effort, yielding a unique interior best response. The cost function $c(e) \to \infty$ ensures best responses are bounded. Existence of a pure-strategy equilibrium follows by standard arguments.

*Uniqueness.* Define the weighted potential

$$\Phi(x, y) \;=\; F(x, y; \alpha, \rho) \;-\; \frac{c(x)}{s} \;-\; \frac{c(y)}{1-s}.$$

Its gradient satisfies $\partial\Phi/\partial x = (1/s)\,\partial U_1/\partial x$ and $\partial\Phi/\partial y = (1/(1-s))\,\partial U_2/\partial y$, so the weighted pseudo-gradient of the game coincides with $\nabla\Phi$. Because $F$ is concave and $c/s$, $c/(1-s)$

are strictly convex, $\Phi$ is strictly concave on $\mathbb{R}^2_{++}$. By Theorem 2 of Rosen (1965), the game admits at most one Nash equilibrium. Combined with existence, the equilibrium is unique.

*Characterization.* With quadratic costs $c(e) = e^2$, the first-order conditions are $s \cdot F_x(x, y; \alpha, \rho) = 2x$ and $(1 - s) \cdot F_y(x, y; \alpha, \rho) = 2y$. For the CES production function in (1), the marginal products are given by

$$F_x(x, y; \alpha, \rho) = \frac{1}{2^{1/\rho}} \left(x^\rho + (\alpha y)^\rho\right)^{\frac{1-\rho}{\rho}} x^{\rho-1}, \tag{8}$$

$$F_y(x, y; \alpha, \rho) = \frac{\alpha^\rho}{2^{1/\rho}} \left(x^\rho + (\alpha y)^\rho\right)^{\frac{1-\rho}{\rho}} y^{\rho-1}. \tag{9}$$

Dividing the two first-order conditions yields $\frac{s}{1-s} \cdot \frac{F_x}{F_y} = \frac{x}{y}$. Using (8)–(9), we obtain:

$$\frac{F_x}{F_y} = \frac{x^{\rho-1}}{\alpha^\rho y^{\rho-1}} = \frac{1}{\alpha^\rho} \left(\frac{x}{y}\right)^{\rho-1}.$$

Letting $r \equiv x/y$, we obtain $\frac{s}{1-s} \cdot \frac{r^{\rho-1}}{\alpha^\rho} = r$, or equivalently $r^{2-\rho} = \frac{s}{(1-s)\alpha^\rho}$. Because $2 - \rho > 1$ for $\rho \in (0, 1)$, the mapping $r \mapsto r^{2-\rho}$ is strictly increasing on $\mathbb{R}_{++}$, and hence there exists a unique solution $r^* > 0$:

$$r^* = \left(\frac{s}{1-s}\right)^{\frac{1}{2-\rho}} \alpha^{-\frac{\rho}{2-\rho}}.$$

Given this unique ratio $r^*$, substitute $y = x/r$ into $F_x$. By the homogeneity of degree one of the CES aggregator,

$$F_x(x, x/r; \alpha, \rho) = \frac{1}{2^{1/\rho}} \left(1 + \left(\frac{\alpha}{r}\right)^\rho\right)^{\frac{1-\rho}{\rho}},$$

which is independent of the scale of efforts. Substituting into the first-order condition for Partner 1 yields the unique effort level $x^*$ in (3), and $y^* = x^*/r^*$ follows immediately.

**Proof of Lemma 1 (Expected Surplus Monotonicity).**

We begin by establishing the following lemma. Recall that $S(s; \alpha, \rho)$ is the (total) surplus induced by the effort equilibrium $(x^*, y^*)$ given $(\alpha, \rho)$ and $s$.

**Lemma A.1** (Surplus Monotonicity)**.**

*(i)* $\partial S/\partial \alpha > 0$*: Higher productivity of Partner 2 increases surplus, holding collaborative fit fixed.*

*(ii)* $\partial S/\partial \rho \geq 0$ *(with equality iff* $x^* = \alpha y^*$*): Stronger collaborative fit (higher* $\rho$*) increases surplus, holding productivity fixed.*

*Proof.* *Part(i)*: The partial derivative of $S(s;\alpha,\rho)$ with respect to $\alpha$ is given by

$$\frac{\partial S}{\partial \alpha} = \frac{\partial F}{\partial \alpha} + \frac{\partial F}{\partial x}\frac{dx^*}{d\alpha} + \frac{\partial F}{\partial y}\frac{dy^*}{d\alpha} - 2x^*\frac{dx^*}{d\alpha} - 2y^*\frac{dy^*}{d\alpha},$$

where all derivatives of $F$ are evaluated at $(x^*, y^*; \alpha, \rho)$. Rearranging terms yields:

$$\frac{\partial S}{\partial \alpha} = \frac{\partial F}{\partial \alpha} + \frac{dx^*}{d\alpha}\left(\frac{\partial F}{\partial x} - 2x^*\right) + \frac{dy^*}{d\alpha}\left(\frac{\partial F}{\partial y} - 2y^*\right).$$

Using the first-order conditions, $sF_x(x^*, y^*; \alpha, \rho) = 2x^*$ and $(1-s)F_y(x^*, y^*; \alpha, \rho) = 2y^*$, together with Euler's theorem, $F(x^*, y^*; \alpha, \rho) = F_x x^* + F_y y^*$, we obtain:

$$\frac{2(x^*)^2}{s} + \frac{2(y^*)^2}{1-s} = F(x^*, y^*; \alpha, \rho).$$

Differentiating this identity with respect to $\alpha$ shows that the terms involving $dx^*/d\alpha$ and $dy^*/d\alpha$ cancel exactly. Hence, $\partial S/\partial\alpha = \partial F/\partial\alpha|_{(x^*,y^*)}$. Because $\partial F/\partial\alpha > 0$ for all $y^* > 0$, it follows that $\partial S/\partial\alpha > 0$.

*Part (ii)*: By the same envelope argument as in part (i), we have $\partial S/\partial\rho = \partial F/\partial\rho|_{(x^*,y^*)}$. By the Power Mean Inequality of the CES production function, $\partial F/\partial\rho \geq 0$. It follows that $\partial S/\partial\rho \geq 0$ with equality if and only if $x^* = \alpha y^*$. □

Because $\rho_0$ is independent of $\alpha$ and $S^*(\alpha,\rho)$ is continuously differentiable in $(\alpha,\rho)$, differentiation and expectation can be interchanged. Hence, Lemma A.1(i) carries over to expected surplus, implying $\partial\bar{S}/\partial\alpha > 0$. For part (ii), because collaborative fit enters through the induced shift $p(\alpha;\lambda)$ and $\rho$ is decreasing in $p(\alpha;\lambda)$, it follows that $\partial S/\partial p \leq 0$, with equality if and only if $x^* = \alpha y^*$. This completes the proof.

**Proof of Theorem 1 (Shape of Expected Surplus).**

The proof proceeds in three steps.

**Step 1: Preliminary results.** Recall the expected surplus $\bar{S}(\alpha) = \mathbb{E}_{\rho_0}[S^*(\alpha,\rho)]$, where $\rho = \rho_0 - p(\alpha;\lambda)$ with $p(\alpha;\lambda) = \lambda(\alpha - 1)$, and the surplus gradient $G(\alpha;\lambda) = \partial\bar{S}/\partial\alpha + \lambda(\partial\bar{S}/\partial p)$. We establish two lemmas that are the key ingredients in the proof.

**Lemma A.2** (Existence and Uniqueness of $\bar{\lambda}$)**.** *Under Assumptions 2–4, there exists a unique $\bar{\lambda} \in (0, \lambda^H)$ such that $G(\underline{\alpha}; \bar{\lambda}) = 0$. Moreover, $G(\underline{\alpha};\lambda) > 0$ for $\lambda < \bar{\lambda}$, and $G(\underline{\alpha};\lambda) < 0$ for $\lambda > \bar{\lambda}$.*

*Proof.* When $\lambda = 0$, the surplus gradient $G(\underline{\alpha}; 0)$ is strictly positive by Lemma 1(i). By Assumption 3(i), there exists $\lambda^H \in (0, \lambda^{max})$ such that $G(\underline{\alpha}; \lambda^H) < 0$. Because $G(\underline{\alpha};\lambda)$ is continuous in $\lambda$, by the intermediate value theorem, there exists a $\bar{\lambda} \in (0, \lambda^H)$ that satisfies $G(\underline{\alpha}; \bar{\lambda}) = 0$. Next, we show that $G(\underline{\alpha}; \lambda)$ is a strictly decreasing function for all $\lambda \geq 0$. Taking the derivative of $G(\alpha; \lambda)$ with respect to $\lambda$, we have

$$\frac{dG(\alpha; \lambda)}{d\lambda} = (\alpha - 1)\left[\frac{\partial^2 \bar{S}}{\partial p \partial \alpha} + \lambda \frac{\partial^2 \bar{S}}{\partial p^2}\right] + \frac{\partial \bar{S}}{\partial p},$$

where $\partial \bar{S}/\partial p \leq 0$ (by Lemma 1(ii)), $|\partial^2 \bar{S}/\partial p \partial \alpha| \leq \delta$ and $\partial^2 \bar{S}/\partial p^2 \geq \kappa$ (by Assumption 4(ii)). Thus, the expression in the square bracket is bounded below by $\lambda\kappa - \delta$ where $\lambda\kappa \geq 0$. Because $(\partial \bar{S}/\partial p)|_{\alpha=\underline{\alpha}} < (\underline{\alpha} - 1)\delta$ (by Assumption 4(i)) and $\underline{\alpha} - 1 < 0$, the derivative $\frac{dG(\alpha;\lambda)}{d\lambda}$ evaluated at $\alpha = \underline{\alpha}$ is strictly negative; indicating that $G(\underline{\alpha}; \lambda)$ is strictly decreasing in $\lambda \geq 0$. This shows the uniqueness of $\bar{\lambda}$, and $G(\underline{\alpha}; \lambda) > 0$ for $\lambda < \bar{\lambda}$ and $G(\underline{\alpha}; \lambda) < 0$ for $\lambda > \bar{\lambda}$. Note that even without Assumption 4(i), $G(\alpha; \lambda)$ is a strictly decreasing function for $\alpha < 1$ and for a sufficiently large $\lambda$ (i.e., $\lambda \geq \delta/\kappa$). This is because, the restriction $\lambda \geq \delta/\kappa$ makes the expression in the bracket positive. But Assumption 4(i) is necessary to guarantee monotonicity even for $\lambda \in (0, \delta/\kappa)$. □

**Lemma A.3** (Gradient Monotonicity). *Under Assumptions 2–4, there exists $\underline{\lambda} \in (0, \lambda^{max})$ such that for all $\lambda \in (\underline{\lambda}, \lambda^{max})$, the surplus gradient $G(\alpha; \lambda)$ is strictly increasing in $\alpha$.*

*Proof.* Note that a technical issue arises at $\alpha = 1$, where $G(\alpha; \lambda)$ is continuous but not differentiable. This non-differentiability stems from the construction of $\rho$, where at $\alpha = 1$, $p(\alpha; \lambda) = 0$ and $\rho$ is independent of $\lambda$; whereas in any neighborhood of $\alpha = 1$, $\rho$ varies with $\lambda$. We therefore establish differentiability on the domain excluding $\alpha = 1$ and rely on continuity to complete the argument.

We first verify continuity at $\alpha = 1$, and then show that $G(\alpha; \lambda)$ has a strictly positive derivative for all $\alpha \neq 1$ given sufficiently large $\lambda$.

*Continuity at $\alpha = 1$:* As $\alpha \to 1$, symmetry implies that the Nash bargaining solution satisfies $s^*(\alpha, \rho) \to 1/2$ and the equilibrium effort ratio $r^*(\alpha) \to 1$. Consequently, $x^* \to y^*$, implying $\lim_{\alpha \to 1} \partial \bar{S}/\partial p = 0$. Hence,

$$\lim_{\alpha \to 1} G(\alpha; \lambda) = \lim_{\alpha \to 1} \left[\frac{\partial \bar{S}}{\partial \alpha} + \lambda \frac{\partial \bar{S}}{\partial p}\right] = \frac{\partial \bar{S}}{\partial \alpha}\Big|_{\alpha=1} = G(1; \lambda),$$

which establishes continuity of $G(\alpha; \lambda)$ at $\alpha = 1$, thus at all $\alpha$.

*Strict Monotonicity for $\alpha \neq 1$:* For $\alpha \neq 1$, $G(\alpha; \lambda)$ is differentiable, and its total derivative with respect to $\alpha$ is given by

$$\frac{dG}{d\alpha} = \frac{\partial^2 \bar{S}}{\partial \alpha^2} + 2\lambda \frac{\partial^2 \bar{S}}{\partial p \partial \alpha} + \lambda^2 \frac{\partial^2 \bar{S}}{\partial p^2}.$$

By Assumption 4(ii), it follows that

$$\frac{dG}{d\alpha} \geq -\delta - 2\lambda\delta + \lambda^2\kappa \equiv b(\lambda),$$

for $\lambda \geq 0$. The function $b(\lambda)$ is continuous, convex, satisfies $b(0) = -\delta < 0$, and diverges to $+\infty$ as $\lambda \to \infty$. Hence, there exists a smallest threshold $\underline{\lambda} = (\delta + \sqrt{\delta^2 + \kappa\delta})/\kappa > 0$ such that $b(\lambda) > 0$ for all $\lambda > \underline{\lambda}$. Therefore, for all $\lambda \in (\underline{\lambda}, \lambda^{max})$, we have $dG/d\alpha > 0$ for all $\alpha \neq 1$.

*Conclusion*: Fix $\lambda \in (\underline{\lambda}, \lambda^{max})$. Consider any $\alpha_1 < \alpha_2$. If the interval $[\alpha_1, \alpha_2]$ does not contain 1, by the Mean Value Theorem: $G(\alpha_2; \lambda) - G(\alpha_1; \lambda) = G_\alpha(c; \lambda)(\alpha_2 - \alpha_1)$ for some $c \in (\alpha_1, \alpha_2)$; and by the previous derivations, we have $G_\alpha(c; \lambda) > 0$ for $\lambda > \underline{\lambda}$ and $(\alpha_2 - \alpha_1) > 0$. Thus, $G(\alpha_2; \lambda) - G(\alpha_1; \lambda) > 0$. If the interval contains 1, we apply the Mean Value Theorem separately to $[\alpha_1, 1]$ and $[1, \alpha_2]$, using continuity of $G(\alpha; \lambda)$ at $\alpha = 1$: $G(\alpha_2; \lambda) - G(\alpha_1; \lambda) = [G(\alpha_2; \lambda) - G(1; \lambda)] + [G(1; \lambda) - G(\alpha_1; \lambda)]$, where both terms in the square bracket are positive because $dG/d\alpha > 0$ on the respective open intervals by the Mean Value Theorem. In both cases, $G(\alpha; \lambda)$ is strictly increasing in $\alpha$ for all $\alpha$. □

Lemma A.3 ensures that productivity and collaborative fit interact in a well-behaved manner, establishing a single-crossing property for the surplus gradient for some large enough $\lambda$. Although higher productivity increases friction, the marginal effect of productivity on expected surplus becomes increasingly favorable as $\alpha$ rises. Intuitively, when productivity is already high, further reductions in collaboration are less costly due to diminishing sensitivity, while productivity gains remain valuable. The lemma does not claim global monotonicity of the surplus gradient for all values of $\lambda$, but only for some large enough $\lambda$. This restriction is without loss for the subsequent proof.

The two thresholds in Lemmas A.2 and A.3 play distinct roles. The threshold $\bar{\lambda}$ has an economic interpretation: it is the point at which the surplus gradient at the lower boundary changes sign, which is necessary for the emergence of a U-shaped surplus. By contrast, $\underline{\lambda}$ is a technical threshold: it is the smallest value of $\lambda$ for which the surplus gradient satisfies a single-crossing property, guaranteeing a unique interior minimum. For expositional clarity, we impose $\bar{\lambda} > \underline{\lambda}$ in what follows. This restriction is purely technical and does not affect the economic content of our results. Our main conclusion concerns the behavior of the expected surplus for sufficiently large $\lambda$, where the U-shaped pattern arises. If $\bar{\lambda} < \underline{\lambda}$, then the expected surplus need not be monotone or U-shaped for values of $\lambda \in (\bar{\lambda}, \underline{\lambda})$,[11] but this region is irrelevant for the existence and characterization of the Moderate's Trap. We

[11]The shape of expected surplus may not be pinned down, or expected surplus may be decreasing, have multiple local extrema, display non-U-shaped patterns, or exhibit more irregular shapes, depending on how productivity and complementarity interact in this region.

therefore abstract from such cases to focus on economically interpretable regimes.

**Step 2: Characterization of $G(\alpha;\lambda)$.** When $\alpha = 1$, $G(1;\lambda)$ is constant across $\lambda$ and $G(1;\lambda) = G(1;0) > 0$ for all $\lambda$. For $\alpha \neq 1$, we consider the regimes separately.

*Part (i)*: $0 < \lambda \leq \bar{\lambda}$. By Lemma A.2, $G(\underline{\alpha};\lambda) > 0$ for $\lambda < \bar{\lambda}$, and $G(\underline{\alpha};\bar{\lambda}) = 0$. Assumption 3(ii) implies $G(\overline{\alpha};\lambda) > 0$ for $\lambda \in (0,\bar{\lambda}]$, where $\bar{\lambda} \in (0,\lambda^H)$. Because $G(\alpha;\lambda)$ is strictly increasing in $\alpha$ by Lemma A.3, it follows that $G(\alpha;\lambda) \geq 0$ for all $\alpha \in [\underline{\alpha},\overline{\alpha}]$, with equality if and only if $\lambda = \bar{\lambda}$ and $\alpha = \underline{\alpha}$.

*Part (ii)*: $\lambda > \bar{\lambda}$. By Lemma A.2, $G(\underline{\alpha};\lambda) < 0$ for $\lambda > \bar{\lambda}$. Assumption 3(ii) implies $G(\overline{\alpha},\lambda) > 0$ for $\lambda \in (\bar{\lambda},\lambda^H)$, where $\bar{\lambda} \in (0,\lambda^H)$. By Lemma A.3, $G(\alpha;\lambda)$ is strictly increasing in $\alpha$, and can have at most one zero on $[\underline{\alpha},\overline{\alpha}]$ (single crossing). Therefore, there exists a unique $\alpha_{min} \in (\underline{\alpha},\overline{\alpha})$ such that $G(\alpha_{min};\lambda) = 0$.

**Step 3: Characterization of $\bar{S}$.** Recall that $d\bar{S}/d\alpha = G(\alpha;\lambda)$.

*Part (i)*: $0 < \lambda \leq \bar{\lambda}$. For $\lambda < \bar{\lambda}$, $G(\alpha;\lambda) > 0$ for all $\alpha \in [\underline{\alpha},\overline{\alpha}]$, and hence $\bar{S}(\alpha)$ is strictly increasing in $\alpha$ over the entire productivity range. When $\lambda = \bar{\lambda}$, we have $G(\underline{\alpha};\bar{\lambda}) = 0$ and $G(\alpha;\bar{\lambda}) > 0$ for all $\alpha \in (\underline{\alpha},\overline{\alpha}]$. Because $G(\alpha;\bar{\lambda})$ is nonnegative everywhere and strictly positive except at a single boundary point, the integral of $G(\alpha;\bar{\lambda})$ over any nontrivial interval is strictly positive. It follows that $\bar{S}(\alpha)$ remains strictly increasing on $[\underline{\alpha},\overline{\alpha}]$.

*Part (ii)*: $\lambda > \bar{\lambda}$. In this case, $G(\alpha;\lambda) < 0$ for $\alpha \in [\underline{\alpha},\alpha_{min})$, $G(\alpha_{min};\lambda) = 0$, and $G(\alpha;\lambda) > 0$ for $\alpha \in (\alpha_{min},\overline{\alpha}]$. Consequently, $\bar{S}$ is strictly decreasing on $[\underline{\alpha},\alpha_{min})$ and strictly increasing in $\alpha$ on $(\alpha_{min},\overline{\alpha}]$, yielding a U-shaped relationship between expected surplus and productivity. At the unique interior minimum $\alpha_{min}$, strict monotonicity on each side implies $\bar{S}(\alpha_{min}) < \bar{S}(\underline{\alpha})$ and $\bar{S}(\alpha_{min}) < \bar{S}(\overline{\alpha})$.

**Proof of Corollary 1 (Optimal Partner Selection).**

*Part (i)*: Follows directly from Theorem 1(i). *Part (ii)*: The U-shaped surplus function on $[\underline{\alpha},\overline{\alpha}]$ achieves its maximum at a boundary. Because $\bar{S}$ is decreasing on $[\underline{\alpha},\alpha_{min})$ and increasing on $(\alpha_{min},\overline{\alpha}]$, we have $\max_{\alpha\in[\underline{\alpha},\overline{\alpha}]} \bar{S}(\alpha) = \max\{\bar{S}(\underline{\alpha}),\bar{S}(\overline{\alpha})\}$. Hence, $\alpha^* \in \{\underline{\alpha},\overline{\alpha}\}$. Note that, by continuity of $\bar{S}$, there exists a neighborhood of the interior minimum $\alpha_{min}$ such that for any $\alpha$ in this neighborhood, $\bar{S}(\alpha) < \min\{\bar{S}(\underline{\alpha}),\bar{S}(\overline{\alpha})\}$. Because $\bar{S}(\alpha_{min}) < \bar{S}(\underline{\alpha})$ and $\bar{S}(\alpha_{min}) < \bar{S}(\overline{\alpha})$ by Theorem 1, continuity implies that both boundary types strictly dominate interior productivity levels sufficiently close to $\alpha_{min}$. Importantly, this dominance need not extend to all interior productivity levels. Productivity levels whose surplus lies between the two boundary values are dominated by one boundary, but not by both.

**Proof of Proposition 2 (Superstar vs. Best Team Player).**

By Corollary 1(ii), for $\lambda > \bar{\lambda}$, the principal's optimal choice lies at a boundary, $\alpha^* \in \{\underline{\alpha}, \overline{\alpha}\}$. The principal prefers the superstar ($\overline{\alpha}$) over the best team player ($\underline{\alpha}$) if and only if $\Delta(\lambda, \underline{\alpha}, \overline{\alpha}) \equiv \bar{S}(\overline{\alpha}) - \bar{S}(\underline{\alpha}) \geq 0$. We prove the comparative statics of $\Delta(\lambda, \underline{\alpha}, \overline{\alpha})$.

*Part (i)*: Using $\rho = \rho_0 - p(\alpha; \lambda)$ with $p(\alpha; \lambda) = \lambda(\alpha - 1)$, we obtain

$$\frac{\partial \Delta}{\partial \lambda} = \frac{\partial \bar{S}(\overline{\alpha}; \lambda)}{\partial p}(\overline{\alpha} - 1) - \frac{\partial \bar{S}(\underline{\alpha}; \lambda)}{\partial p}(\underline{\alpha} - 1).$$

By Lemma 1(ii), $\partial \bar{S}/\partial p \leq 0$, with equality if and only if $x^* = \alpha y^*$, which is a knife-edge case. Moreover, $\overline{\alpha} > 1$ and $\underline{\alpha} < 1$. Hence, except in the degenerate equality case, we have $\partial \Delta/\partial \lambda < 0$. In particular, an increase in $\lambda$ reduces $\bar{S}(\overline{\alpha})$ but increases $\bar{S}(\underline{\alpha})$. Therefore, higher values of $\lambda$ reduce the relative advantage of the superstar. Equivalently, for $\lambda > \bar{\lambda}$, lower values of $\lambda$ strengthen the case for selecting $\overline{\alpha}$.

*Part (ii)*: The boundaries $\underline{\alpha}$ and $\overline{\alpha}$ are arbitrary parameters satisfying $0 < \underline{\alpha} < 1 < \overline{\alpha}$. From the proof of Theorem 1, we have $G(\alpha; \lambda) > 0$ for $\alpha \in (\alpha_{min}, \overline{\alpha}]$, so $\bar{S}(\alpha)$ is strictly increasing in $\alpha$ in this region. Hence, an increase in the upper boundary $\overline{\alpha}$ raises $\bar{S}(\overline{\alpha})$. In contrast, for $\alpha \in [\underline{\alpha}, \alpha_{min})$, we have $G(\alpha; \lambda) < 0$, so $\bar{S}(\alpha)$ is strictly decreasing in $\alpha$ on this region. Therefore, an increase in the lower boundary $\underline{\alpha}$ lowers $\bar{S}(\underline{\alpha})$. Either change increases $\Delta(\lambda, \underline{\alpha}, \overline{\alpha})$, strengthening the case for selection of $\overline{\alpha}$.

**Proof of Proposition 3 (Dispersion of $\rho_0$ and the Moderate's Trap).**

Under the assumptions of Theorem 1, and as discussed in Section 5.4, there exists a threshold $\bar{\lambda}(\tau^2)$ such that expected surplus is U-shaped in $\alpha$ for all $\lambda > \bar{\lambda}(\tau^2)$. This threshold is defined implicitly by

$$G(\underline{\alpha}; \bar{\lambda}(\tau^2)) \equiv \frac{\partial \bar{S}}{\partial \alpha} + \bar{\lambda}(\tau^2)\frac{\partial \bar{S}}{\partial p} = 0,$$

evaluated at $\alpha = \underline{\alpha}$.

*Part (i)*: By Lemma 1, $\partial \bar{S}/\partial \alpha > 0$ and $\partial \bar{S}/\partial p \leq 0$, with strict inequality holding generically. An increase in $\tau^2$ corresponds to a mean-preserving spread of the distribution of $\rho_0$. Under Assumption 2, the expected surplus function is convex in the shift parameter $p$, implying that the expected marginal effect of collaborative-fit losses, $\partial \bar{S}/\partial p$, becomes weaker in absolute value as dispersion increases. That is, greater uncertainty attenuates the expected marginal value of collaborative fit: $\partial \left|\partial \bar{S}/\partial p\right|/\partial \tau^2 < 0$. Holding $\partial \bar{S}/\partial \alpha$ fixed, a smaller absolute value of $\partial \bar{S}/\partial p$ requires a larger value of $\lambda$ to satisfy the threshold condition. Hence, $d\bar{\lambda}(\tau^2)/d\tau^2 > 0$.

*Part (ii)*: Fix $\lambda > \bar{\lambda}(\tau^2)$, so that expected surplus is U-shaped with a unique interior minimizer $\alpha_{min}$. Define the depth of the Moderate's Trap as $\Delta(\tau^2) \equiv \bar{S}(\underline{\alpha}) - \bar{S}(\alpha_{min})$. Under Assumption 2, the marginal value of collaborative fit is higher at lower values of the shift parameter $p(\alpha; \lambda)$. Because $p(\underline{\alpha}; \lambda) < p(a_{min}; \lambda)$, an increase in the dispersion of $\rho_0$ has a larger positive effect on expected surplus at $\underline{\alpha}$ than at $\alpha_{min}$. Consequently, the expected surplus gap between the boundary and the minimum of the U-shape shrinks as uncertainty increases. It follows that $d\Delta(\tau^2)/d\tau^2 < 0$. Greater uncertainty therefore attenuates the depth of the trap without eliminating it.

**Proof of Proposition 4 (Equilibrium Wages).**

*Part (i)*: Define the wage function $w(\alpha) = \bar{S}(\alpha) - \bar{\Pi}$, where $\bar{\Pi} \geq 0$ is a constant. Under this wage function, a principal who hires a worker of type $\alpha$ earns profit $\Pi(\alpha) = \bar{\Pi}$ regardless of the worker's productivity $\alpha$. Hence, principals are indifferent across all worker types and profit maximization is satisfied. Choosing $\bar{\Pi} = \min_\alpha \bar{S}(\alpha)$ guarantees that $w(\alpha) \geq 0$ for all $\alpha$, so that workers weakly prefer employment to their outside option. Because profits are equalized across worker types and the measure of principals equals the measure of workers, a competitive equilibrium exists in which every worker is matched with exactly one principal. By construction, wage differences satisfy $w(\alpha) - w(\alpha') = \bar{S}(\alpha) - \bar{S}(\alpha')$ for all $\alpha$, $\alpha'$.

*Part (ii)*: For $\lambda \leq \bar{\lambda}$, Theorem 1(i) establishes that $\bar{S}(\alpha)$ is strictly increasing in $\alpha$. Because the wage function $w(\alpha) = \bar{S}(\alpha) - \bar{\Pi}$ differs from expected surplus only by a constant, $w(\alpha)$ is also strictly increasing in $\alpha$.

*Part (iii)*: For $\lambda > \bar{\lambda}$, Theorem 1(ii) establishes that $\bar{S}(\alpha)$ is U-shaped with a unique interior minimizer $\alpha_{min}$. The wage function inherits this shape: it is strictly decreasing on $[\underline{\alpha}, \alpha_{min})$ and strictly increasing on $(\alpha_{min}, \overline{\alpha}]$.

**Proof of Proposition 5 (Optimal Investment).**

Take the continuation equilibrium at Stages 1–3 as given. At Stage 0, an agent's investment problem is governed by the expected surplus generated by a partnership in which the agent is selected. An agent who is not selected receives zero payoff regardless of investment. For $\lambda > \bar{\lambda}$, expected surplus is U-shaped in productivity. Let $\bar{\bar{\lambda}}$ satisfy $\bar{S}(\underline{\theta}) = \bar{S}(\overline{\theta} + \gamma)$. Corollary 1 and Proposition 2 together characterize the principal's continuation strategy $\alpha^*$.

*Part (i)*: Fix $\lambda < \bar{\bar{\lambda}}$. The principal's optimal choice is $\alpha^* = \overline{\alpha}$, the highest attainable productivity that is determined in equilibrium. Agent $k$ with $\theta_k = \overline{\theta}$ guarantees selection by choosing $i_k = 1$, yielding the highest attainable productivity $\overline{\theta} + \gamma$ among agents. By assumption, $\bar{S}(\overline{\theta} + \gamma) > \bar{S}(\overline{\theta})$, so any deviation to a lower investment level than $i_k = 1$ reduces productivity and therefore risks losing selection without any compensating gain. Hence, for $\lambda < \bar{\bar{\lambda}}$, the agent with $\theta_k = \overline{\theta}$ strictly prefers $i_k = 1$ to any other investment level. For any other agent $j \neq k$ with $\theta_j < \theta_k = \overline{\theta}$, even $i_j = 1$ yields productivity strictly below $\overline{\theta} + \gamma$. Such agents are never selected in equilibrium and therefore obtain zero payoff regardless of investment. Thus, they are indifferent over all investment choices. Given such investment choices, the upper bound on the equilibrium productivity support faced by the principal at Stage 1 is $\overline{\alpha} = \overline{\theta} + \gamma$, in turn rationalizing the principal's optimal choice $\alpha^* = \overline{\alpha}$.

*Part (iii)*: We prove this part before proving part (ii). Fix $\lambda > \bar{\bar{\lambda}}$. The principal's optimal choice is $\alpha^* = \underline{\alpha}$, the lowest attainable productivity in equilibrium. Agent $k$ with $\theta_k = \underline{\theta}$ guarantees selection by choosing $i_k = 0$, yielding the lowest attainable productivity $\underline{\theta}$. By assumption, $\bar{S}(\underline{\theta}) > \bar{S}(\underline{\theta} + \gamma)$, so $i_k > 0$ increases productivity and therefore may cause the principal to select another agent instead. Hence, for $\lambda > \bar{\bar{\lambda}}$, the agent with $\theta_k = \underline{\theta}$ strictly prefers $i_k = 0$ to any other investment level. As before, all other agents cannot affect the principal's choice through investment and are therefore indifferent. Given such investment choices, the lower bound on the equilibrium productivity support faced by the principal at Stage 1 is $\underline{\alpha} = \underline{\theta}$, again rationalizing the principal's optimal choice $\alpha^* = \underline{\alpha}$.

*Part (ii)*: At $\lambda = \bar{\bar{\lambda}}$, the principal is indifferent between the two extreme productivity levels, $\overline{\alpha}$ and $\underline{\alpha}$. The agent with $\theta_k = \overline{\theta}$ weakly prefers $i_k = 1$, while the agent with $\theta_k = \underline{\theta}$ weakly prefers $i_k = 0$. All other agents remain indifferent.

**Proof of Proposition 6 (Bimodal Investment under Competitive Labor Market).**

*Part (i)*: Fix an agent $k$ with baseline talent $\theta_k$. Given an investment choice $i_k \in [0, 1]$, the agent's expected payoff is $V(i_k; \theta_k) \equiv \bar{S}(\alpha(i_k; \theta_k)) = \bar{S}(\theta_k + \gamma i_k)$. When $\lambda > \bar{\lambda}$, Theorem 1 implies that $\bar{S}(\alpha)$ is U-shaped in $\alpha$ with a unique interior minimizer $\alpha_{min}$. Viewed as a function of the investment choice $i_k$, $V(i_k; \theta_k)$ is then either monotonic if the interval $[\theta_k, \theta_k + \gamma]$ lies entirely on one arm of the U-shape, or U-shaped with an interior minimum if the interval spans $\alpha_{min}$. In either case, there is no interior maximum, and the maximum of $V(i_k; \theta_k)$ over $i_k \in [0, 1]$ is attained at a boundary. Hence, the optimal investment choice satisfies $i_k^* \in \{0, 1\}$ for all agents.

*Part (ii)*: Define, for a generic baseline talent $\theta$, the payoff gain from full investment:

$$\Delta(\theta) \equiv V(1; \theta) - V(0; \theta) = \bar{S}(\theta + \gamma) - \bar{S}(\theta).$$

Because $\bar{S}$ is continuously differentiable, $\Delta(\theta)$ is continuous. Differentiating yields

$$\Delta'(\theta) = \bar{S}'(\theta + \gamma) - \bar{S}'(\theta) = G(\theta + \gamma; \lambda) - G(\theta; \lambda),$$

where $G(\alpha;\lambda) = d\bar{S}/d\alpha$ is the surplus gradient. By Lemma A.3, for $\lambda > \bar{\lambda}$, $G(\alpha;\lambda)$ is strictly increasing in $\alpha$. Because $\theta+\gamma > \theta$, it follows that $\Delta'(\theta) > 0$ for all $\theta$. Hence, $\Delta(\theta)$ is strictly increasing in baseline talent. By assumption, the investment return $\gamma$ is sufficiently large that $\Delta(\underline{\theta}) < 0$ and $\Delta(\overline{\theta}) > 0$. By the intermediate value theorem, there exists a unique cutoff $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$ such that $\Delta(\hat{\theta}) = 0$. Applying this cutoff to individual agents, an agent $k$ with baseline talent $\theta_k$ optimally chooses full investment $i_k^* = 1$ if $\theta_k > \hat{\theta}$, and no investment $i_k^* = 0$ if $\theta_k < \hat{\theta}$. An agent with $\theta_k = \hat{\theta}$ is indifferent and may randomize between the two.

*Part (iii)*: Follows from part (ii). Agents with $\theta_k > \hat{\theta}$ choose $i_k^* = 1$, generating effective productivity levels in $(\hat{\theta}+\gamma, \overline{\theta}+\gamma]$. Agents with $\theta_k < \hat{\theta}$ choose $i_k^* = 0$, generating effective productivity levels in $[\underline{\theta}, \hat{\theta})$. An agent with $\theta_k = \hat{\theta}$ is indifferent between $i_k = 0$ and $i_k = 1$. Because both investment choices are consistent with equilibrium, both boundary productivity levels $\hat{\theta}$ and $\hat{\theta}+\gamma$ may arise under equilibrium selection. Hence, the equilibrium productivity support is characterized as $[\underline{\theta}, \hat{\theta}] \cup [\hat{\theta}+\gamma, \overline{\theta}+\gamma]$.

**Proof of Lemma C.1 (Directional Effects of Overconfidence).**

Recall that the star premium is defined as $\Delta(\tilde{\beta}, \tilde{\lambda}) = \bar{S}(\overline{\alpha}; \tilde{\beta}, \tilde{\lambda}) - \bar{S}(\underline{\alpha}; \tilde{\beta}, \tilde{\lambda})$, where $\tilde{\beta}$ denotes the principal's perceived productivity and $\tilde{\lambda}$ the perceived tradeoff intensity.

*Part (i)*: Consider an increase in perceived productivity from $\tilde{\beta}$ to $\tilde{\beta}' > \tilde{\beta}$, holding $\tilde{\lambda}$ fixed. Differentiating the star premium with respect to $\tilde{\beta}$ yields

$$\frac{\partial \Delta}{\partial \tilde{\beta}} = \frac{\partial \bar{S}(\overline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\beta}} - \frac{\partial \bar{S}(\underline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\beta}}.$$

Because the principal's productivity and the agent's productivity are substitutes in joint production, the marginal contribution of the principal's productivity is decreasing in $\alpha$. Consequently, the marginal effect of $\tilde{\beta}$ on expected surplus is smaller when the agent is the superstar than when the agent is a best team player:

$$\frac{\partial \bar{S}(\overline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\beta}} < \frac{\partial \bar{S}(\underline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\beta}}.$$

It follows that $\partial \Delta / \partial \tilde{\beta} < 0$.

*Part (ii)*: Differentiating the star premium with respect to $\tilde{\lambda}$ yields

$$\frac{\partial \Delta}{\partial \tilde{\lambda}} = \frac{\partial \bar{S}(\overline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\lambda}} - \frac{\partial \bar{S}(\underline{\alpha}; \tilde{\beta}, \tilde{\lambda})}{\partial \tilde{\lambda}}.$$

Under Assumption **DS**, collaborative fit is given by $\rho = \rho_0 - p(\alpha; \lambda)$, so that the perceived tradeoff intensity operates through the shift parameter $p(\alpha;\tilde{\lambda}) = \tilde{\lambda}(\alpha-1)$. Therefore, differentiating expected surplus with respect to $\tilde{\lambda}$ yields

$$\frac{\partial \bar{S}}{\partial \tilde{\lambda}} = \frac{\partial \bar{S}}{\partial p}\frac{\partial p}{\partial \tilde{\lambda}} = \frac{\partial \bar{S}}{\partial p}(\alpha-1).$$

By Lemma 1(ii), $\partial \bar{S}/\partial p \leq 0$, with strict inequality holding generically. For the superstar, $\overline{\alpha} > 1$, we have $\partial \bar{S}/\partial \tilde{\lambda}\big|_{\overline{\alpha}} < 0$; that is, a higher perceived tradeoff intensity lowers expected surplus at the upper boundary. For the best team player, $\underline{\alpha} < 1$, we have $\partial \bar{S}/\partial \tilde{\lambda}\big|_{\underline{\alpha}} > 0$; that is, a higher perceived tradeoff intensity raises expected surplus at the lower boundary. It follows that $\partial \Delta/\partial \tilde{\lambda} < 0$. Equivalently, a reduction in perceived tradeoff intensity, i.e., collaborative-fit overconfidence with $\tilde{\lambda} = \lambda - \epsilon$, increases the star premium.

**Proof of Proposition C.1 (Partial Debiasing and Mis-selection).**

Suppose $\lambda > \bar{\lambda}$, so that expected surplus is U-shaped and partner selection is bimodal. The principal holds biased beliefs $(\tilde{\beta}, \lambda-\epsilon)$, with $\tilde{\beta} > 1$ and $\epsilon > 0$, while the true parameters are $(1,\lambda)$, under which the optimal choice is given by $\alpha^* \in \{\underline{\alpha}, \overline{\alpha}\}$. By assumption, the two biases offset so that the principal selects correctly; that is, the star premium under biased beliefs, $\Delta(\tilde{\beta}, \lambda-\epsilon)$, has the same sign as the star premium under the true parameters, $\Delta(1,\lambda)$.

First, suppose that under the true parameters $(1,\lambda)$, $\alpha^* = \underline{\alpha}$, so that $\Delta(1,\lambda) < 0$. Because the biases offset, we also have $\Delta(\tilde{\beta}, \lambda-\epsilon) < 0$. Now remove productivity overconfidence by setting $\tilde{\beta} = 1$, while holding $\tilde{\lambda} = \lambda - \epsilon$. By Lemma C.1(i), the star premium is strictly decreasing in $\tilde{\beta}$. Because $\tilde{\beta} > 1$, it follows that $\Delta(1, \lambda-\epsilon) > \Delta(\tilde{\beta}, \lambda-\epsilon)$, implying an upward shift in the star premium. Because $\Delta(\tilde{\beta}, \lambda-\epsilon) < 0$, this upward shift moves $\Delta(1, \lambda-\epsilon)$ closer to zero and may cause it to become positive. Hence, removing productivity overconfidence tends toward over-selection of superstars; that is, it is possible that $\Delta(1, \lambda-\epsilon) > 0 > \Delta(1,\lambda)$.

Next, suppose that under the true parameters $(1,\lambda)$, $\alpha^* = \overline{\alpha}$, so that $\Delta(1,\lambda) > 0$. Because the biases offset, we also have $\Delta(\tilde{\beta}, \lambda-\epsilon) > 0$. Now remove collaborative-fit overconfidence by setting $\epsilon = 0$, while holding $\tilde{\beta} > 1$. By Lemma C.1(ii), because $\epsilon > 0$, it follows that $\Delta(\tilde{\beta}, \lambda) < \Delta(\tilde{\beta}, \lambda-\epsilon)$, implying a downward shift in the star premium. Because $\Delta(\tilde{\beta}, \lambda-\epsilon) > 0$, this downward shift moves $\Delta(\tilde{\beta}, \lambda)$ closer to zero and may cause it to become negative. Hence, removing collaborative-fit overconfidence tends toward over-selection of best team players; that is, it is possible that $\Delta(\tilde{\beta}, \lambda) < 0 < \Delta(1,\lambda)$.

# B Supplementary Results: Bargaining

## B.1 Nash Bargaining Symmetry

We formally prove the symmetry property of the Nash bargaining outcome discussed in Section 4.2 by showing that, when productivity is symmetric ($\alpha = 1$), the Nash bargaining problem admits a symmetric set of interior maximizers around 1/2.

**Proposition B.1** (Symmetry of the Nash Bargaining Outcome)**.** *Suppose $\alpha = 1$ and Assumption 1 holds. Then the set of interior maximizers of the Nash product is symmetric around $1/2$. In particular, if $s^*$ is an interior maximizer, then $1 - s^*$ is also an interior maximizer. If the interior maximizer is unique, it must satisfy $s^* = 1/2$.*

*Proof.* Fix $\alpha = 1$. Then the CES production function and the cost functions are invariant to relabeling the two partners. Hence, the underlying environment is symmetric. Consider the transformation $(s, x, y) \mapsto (1 - s, y, x)$, which exchanges the partners' roles and reflects the sharing rule. By symmetry of the effort equilibrium when $\alpha = 1$, equilibrium efforts satisfy $x^*(s; 1, \rho) = y^*(1 - s; 1, \rho)$ and $y^*(s; 1, \rho) = x^*(1 - s; 1, \rho)$. It follows that the induced continuation payoffs satisfy $\big(U_1(s; 1, \rho), U_2(s; 1, \rho)\big) = \big(U_2(1-s; 1, \rho), U_1(1-s; 1, \rho)\big)$. Therefore, the Nash product $N(s) \equiv U_1(s; 1, \rho) \cdot U_2(s; 1, \rho)$ is symmetric around 1/2, that is, $N(s) = N(1-s)$ for all $s \in (0, 1)$. Hence, if $s^*$ is an interior maximizer of the Nash product, then $1 - s^*$ is also an interior maximizer. The set of interior maximizers is thus symmetric around 1/2. If the interior maximizer is unique, symmetry implies $s^* = 1/2$. □

## B.2 Ex Ante Nash Bargaining

In the partnership game specified in Section 3.2, partners bargain over the sharing rule after collaborative fit is revealed. We refer to this timing *ex post bargaining*. In this appendix, we consider an alternative timing in which partners bargain over the sharing rule *before* collaborative fit is revealed, which we refer to as *ex ante bargaining*. We show that the Moderate's Trap can arise under ex ante bargaining, and clarify that any apparent disappearance of the trap reflects feasibility constraints rather than a change in the underlying mechanism.

Under ex ante bargaining, the game proceeds as follows. At Stage 1, the principal selects an agent with an observable productivity level $\alpha$ and forms a partnership. At Stage 2, the two partners bargain over the sharing rule without knowing the exact value of $\rho$; in particular, they do not observe $\rho_0$ and only know its distribution. After the sharing rule is determined, $\rho_0$ is revealed and the game proceeds to Stage 3 as in the baseline model. We assume that partners commit to the sharing rule agreed upon at Stage 2, so that it cannot be renegotiated after the realization of $\rho_0$.

The Nash bargaining solution under ex ante bargaining is characterized by

$$\hat{s}(\alpha) \in \arg\max_{s\in[0,1]} \mathbb{E}_{\rho_0}\left[N\left(s;\alpha,\rho\right)\right], \tag{10}$$

where $\rho = \rho_0 - p(\alpha;\lambda)$, $N(s;\alpha,\rho) \equiv U_1(s,\alpha,\rho)\cdot U_2(s,\alpha,\rho)$ denotes the Nash product, and $U_i(s,\alpha,\rho)$, $i = 1,2$, denote the continuation payoffs induced by equilibrium effort choices in the subsequent stage given $s$, $\alpha$, and $\rho$.[12]

The following lemma provides a convenient characterization of the Nash bargaining solution under ex ante bargaining. We refer to this solution as the *ex ante sharing rule* and to the Nash bargaining solution under ex post bargaining as the *ex post sharing rule*.

**Lemma B.1** (Ex Ante Sharing Rule as Certainty Equivalent)**.** *The ex ante sharing rule $\hat{s}(\alpha)$ can be characterized as the ex post sharing rule if collaborative fit were known to equal a certainty-equivalent level $\hat{\rho}(\alpha)$. In general, $\hat{\rho}(\alpha) \neq \mathbb{E}[\rho|\alpha]$.*

*Proof.* Fix $\alpha$. For any given $\rho$, the Nash product $N(s;\alpha,\rho)$ is finite and continuously differentiable in $s$, and $\rho_0$ is independent of other given parameters. Differentiation and expectation can be interchanged; and so the ex ante sharing rule $\hat{s}(\alpha)$ satisfies the first-order condition $\partial\mathbb{E}_{\rho_0}[N(s;\alpha,\rho)]/\partial s\big|_{s=\hat{s}(\alpha)} = \mathbb{E}_{\rho_0}[\partial N(s;\alpha,\rho)/\partial s]\big|_{s=\hat{s}(\alpha)} = 0$. Define $\phi(s,\rho) \equiv \partial N(s;\alpha,\rho)/\partial s$. For interior values of $\rho$, the function $\phi$ is continuous in both arguments and satisfies $\phi(0,\rho) > 0$ and $\phi(1,\rho) < 0$, reflecting the requirement that both partners obtain positive payoffs. Hence, for each $s \in (0,1)$, the Intermediate Value Theorem implies the existence of a value $\tilde{\rho}(s)$ such that $\phi(s,\tilde{\rho}(s)) = 0$. Define $\hat{\rho}(\alpha) \equiv \tilde{\rho}(\hat{s}(\alpha))$. By construction, $\hat{s}(\alpha)$ coincides with the ex post sharing rule evaluated at $\rho = \hat{\rho}(\alpha)$. In general, however, $\hat{\rho}(\alpha) \neq \mathbb{E}[\rho|\alpha]$. The certainty-equivalent level depends on the curvature of $\phi$ in $\rho$, and Jensen's inequality implies that equality holds only in the special case where $\phi$ is linear in $\rho$. □

Let $\hat{N}(\alpha) \equiv \mathbb{E}_{\rho_0}\left[N(\hat{s}(\alpha);\alpha,\rho)\right]$ denote the maximized expected Nash product given $\alpha$ under ex ante bargaining, and let $\bar{N}(\alpha,\rho) \equiv N(s^*(\alpha,\rho);\alpha,\rho)$ denote the maximized Nash product given $\alpha$ and $\rho$ under ex post bargaining, where $s^*(\alpha,\rho)$ is the Nash bargaining solution defined in (4). Because ex post bargaining maximizes the Nash product pointwise in $\rho$, it follows immediately that $\mathbb{E}_{\rho_0}[\bar{N}(\alpha,\rho)] \geq \hat{N}(\alpha)$, with equality if and only if $s^*(\alpha,\rho)$ is independent of $\rho$. This ordering, however, does not generally extend to expected surplus, because Nash bargaining maximizes $U_1 \cdot U_2$ rather than $U_1 + U_2$.

[12] To ease the exposition, we suppress the dependence of $U_i$ and $N$ on $\lambda$.

Let $\hat{S}(\alpha) \equiv \mathbb{E}_{\rho_0}\left[S(\hat{s}(\alpha);\alpha,\rho)\right]$ denote the expected surplus under ex ante bargaining. Recall that $\bar{S}(\alpha) \equiv \mathbb{E}_{\rho_0}\left[S(s^*(\alpha,\rho);\alpha,\rho)\right]$ is the expected surplus under ex post bargaining. The following result establishes that the U-shaped relationship between expected surplus and productivity persists under ex ante bargaining for sufficiently strong tradeoff intensity.

**Proposition B.2** (U-Shaped Expected Surplus under Ex Ante Bargaining)**.** *Under the assumptions of Theorem 1, there exists a threshold $\hat{\lambda} > \bar{\lambda}$ such that: if $\hat{\lambda} < \lambda^{max}$, then for all $\lambda \in (\hat{\lambda}, \lambda^{max})$, $\hat{S}(\alpha)$ is U-shaped in $\alpha$ on $[\underline{\alpha}, \overline{\alpha}]$, with a unique interior minimizer $\hat{\alpha}_{min} \in (\underline{\alpha}, \overline{\alpha})$.*

*Proof.* The proof establishes that the surplus gradient under ex ante bargaining inherits the sign pattern required for a U-shaped expected surplus function. The argument proceeds by comparing the ex ante surplus gradient with its ex post counterpart and showing that, for sufficiently strong tradeoff intensity, the distortion induced by ex ante bargaining does not overturn the boundary behavior and monotonicity of the ex post gradient.

**Step 1: Distortion under ex ante bargaining.** Recall the surplus gradient under ex post bargaining: $G(\alpha;\lambda) \equiv d\bar{S}/d\alpha = \partial\bar{S}/\partial\alpha + \lambda(\partial\bar{S}/\partial p)$. Let $\hat{G}(\alpha;\lambda)$ denote the ex ante surplus gradient: $\hat{G}(\alpha;\lambda) \equiv d\hat{S}/d\alpha = \partial\hat{S}/\partial\alpha + \lambda(\partial\hat{S}/\partial p) + (\partial\hat{S}/\partial s)(d\hat{s}/d\alpha)$. Define the *distortion* term as $R(\alpha;\lambda) \equiv \hat{G}(\alpha;\lambda) - G(\alpha;\lambda)$. For large $\lambda$, both $G(\alpha;\lambda)$ and $\hat{G}(\alpha;\lambda)$ admit leading terms proportional to $\lambda$: $G(\alpha;\lambda) \approx \lambda \cdot g(\alpha)$ and $\hat{G}(\alpha;\lambda) \approx \lambda \cdot \hat{g}(\alpha)$, where $g(\alpha) \equiv \mathbb{E}_{\rho_0}[\partial S/\partial p|_{s=s^*(\alpha,\rho)}]$ and $\hat{g}(\alpha) \equiv \mathbb{E}_{\rho_0}[\partial S/\partial p|_{s=\hat{s}(\alpha)}]$. Accordingly, $R(\alpha;\lambda) \approx \lambda \cdot r(\alpha)$, where $r(\alpha) \equiv \hat{g}(\alpha) - g(\alpha)$. Because $\partial S/\partial p$ is continuous and evaluated on a compact set of sharing rules and $\rho$ realizations, the coefficients $g(\alpha)$, $\hat{g}(\alpha)$, and $r(\alpha)$ are bounded.

**Step 2: Boundary behavior of the ex post gradient.** By the proof of Lemma A.2, $G(\underline{\alpha};\lambda)$ is strictly decreasing in $\lambda$, with $G(\underline{\alpha};\bar{\lambda}) = 0$. Hence, for all $\lambda > \bar{\lambda}$, $G(\underline{\alpha};\lambda) < 0$. By Assumption 3(ii), $G(\overline{\alpha};\lambda) > 0$ for all $\lambda \in (0, \lambda^H)$. For sufficiently large $\lambda$, the term $(\overline{\alpha} - 1)\lambda(\partial^2\bar{S}/\partial p^2)$ dominates in $\partial G/\partial\lambda$. Because $\partial^2\bar{S}/\partial p^2 > 0$, it follows that $G(\overline{\alpha};\lambda)$ is increasing in $\lambda$. Hence, $g(\underline{\alpha}) < 0$ and $g(\overline{\alpha}) > 0$.

**Step 3: Existence of a threshold.** The ex ante gradient inherits the sign of the ex post gradient at each boundary whenever $|r(\alpha)| < |g(\alpha)|$. This condition holds when the ex post sharing rule $s^*(\alpha,\rho)$ varies little with $\rho$, or when the cross-partial $\partial^2 S/\partial p\partial s$ is small relative to $\partial S/\partial p$. Define $\hat{\lambda}$ as the smallest value of $\lambda > \bar{\lambda}$ such that $|g(\underline{\alpha})| > |r(\underline{\alpha})|$ and $|g(\overline{\alpha})| > |r(\overline{\alpha})|$. If $\hat{\lambda} < \lambda^{max}$, then for all $\lambda \in (\hat{\lambda}, \lambda^{max})$, $\hat{G}(\underline{\alpha};\lambda) = G(\underline{\alpha};\lambda) + R(\underline{\alpha};\lambda) < 0$ and $\hat{G}(\overline{\alpha};\lambda) = G(\overline{\alpha};\lambda) + R(\overline{\alpha};\lambda) > 0$.

**Step 4: Monotonicity of the ex ante surplus gradient.** By Lemma A.3, for large $\lambda$, $\partial G/\partial\alpha$ is dominated by $\lambda^2(\partial^2\bar{S}/\partial p^2)$, which is positive by Assumption 2. Differentiating $R(\alpha;\lambda)$ with respect to $\alpha$ and applying the chain rule yields terms involving $d\hat{s}/d\alpha$, which is bounded. Hence, $\partial R/\partial\alpha$ also grows at rate $O(\lambda^2)$, with coefficient $[\partial^2\hat{S}/\partial p^2 - \partial^2\bar{S}/\partial p^2]$. Because diminishing sensitivity is a property of the surplus function $S$ itself, $\partial^2 S/\partial p^2 > 0$ regardless of the sharing rule. Hence, the leading term in $\partial\hat{G}/\partial\alpha$ is $\lambda^2(\partial^2\hat{S}/\partial p^2) > 0$, implying that $\hat{G}(\alpha;\lambda)$ is strictly increasing in $\alpha$ for sufficiently large $\lambda$. Adjusting $\hat{\lambda}$ upward if necessary to ensure this monotonicity, we conclude that for all $\lambda \in (\hat{\lambda}, \lambda^{max})$, $\hat{G}(\alpha;\lambda)$ is negative at $\underline{\alpha}$, positive at $\overline{\alpha}$, and strictly increasing in $\alpha$.

**Step 5: Conclusion.** By the Intermediate Value Theorem, there exists a unique $\hat{\alpha}_{min} \in (\underline{\alpha}, \overline{\alpha})$ such that $\hat{G}(\hat{\alpha}_{min};\lambda) = 0$. Hence, $\hat{S}(\alpha)$ is strictly decreasing on $[\underline{\alpha}, \hat{\alpha}_{min})$ and strictly increasing on $(\hat{\alpha}_{min}, \overline{\alpha}]$, yielding a U-shaped relationship. □

The condition $\hat{\lambda} < \lambda^{max}$ holds when the sharing rule distortion is bounded. Specifically, for large $\lambda$, write $G(\alpha;\lambda) \approx \lambda \cdot g(\alpha)$ and $R(\alpha;\lambda) \approx \lambda \cdot r(\alpha)$, where $g(\alpha) \equiv \mathbb{E}_{\rho_0}\big[(\partial S/\partial p)\big|_{s=s^*(\alpha,\rho)}\big]$ and $r(\alpha) \equiv \mathbb{E}_{\rho_0}\big[(\partial S/\partial p)\big|_{s=\hat{s}(\alpha)}\big] - g(\alpha)$. If $|r(\alpha)| < |g(\alpha)|$ at both boundaries $\alpha \in \{\underline{\alpha}, \overline{\alpha}\}$, then $\hat{\lambda}$ is finite. This bound is satisfied when the ex post sharing rule $s^*(\alpha,\rho)$ varies little with $\rho$, or when the cross-partial $\partial^2 S/\partial p\partial s$ is small relative to $\partial S/\partial p$. The condition $\hat{\lambda} < \lambda^{max}$ then holds for a nonempty set of parameter configurations.

# C Overconfidence in Partner Selection

This appendix examines how the principal's overconfidence affects partner selection in the single-principal environment of Section 5. We consider two forms of bias, show that they distort selection in opposite directions, and demonstrate that correcting only one bias can induce mis-selection.

We consider two forms of bias, both reflecting the principal's tendency to overestimate her own capabilities.

(1) *Productivity overconfidence*: The principal believes her productivity is $\tilde{\beta} > \beta$, rather than the true value $\beta$.

(2) *Collaborative-fit overconfidence*: The principal perceives the tradeoff intensity as $\tilde{\lambda} = \lambda - \epsilon$ for some $\epsilon > 0$, rather than the true value $\lambda$.

Under productivity overconfidence, the principal overestimates her own productivity. Under collaborative-fit overconfidence, she underestimates the extent to which highly productive partners are difficult to work with, perhaps believing that her skills allow her to mitigate collaborative frictions more effectively than is objectively warranted. Although both biases stem from excessive self-confidence, they operate in distinct domains.

To formalize these biases, we evaluate expected surplus under the principal's perceived parameters. In the baseline model, the principal's productivity is normalized to one. Productivity overconfidence is captured by allowing the principal to believe her own productivity is $\beta > 1$, which scales her effort in production:

$$F(x, y; \beta, \alpha, \rho) = \left( \frac{(\beta x)^\rho + (\alpha y)^\rho}{2} \right)^{1/\rho}.$$

Let $\bar{S}(\alpha; \beta, \lambda)$ denote expected surplus evaluated under perceived parameters. Holding $\lambda$ fixed, a higher perceived $\beta$ reduces the perceived marginal contribution of an agent's productivity $\alpha$: when the principal believes she is more productive, additional productivity from an agent is viewed as less valuable.

**Definition 1** (Star Premium)**.** Given beliefs $(\tilde{\beta}, \tilde{\lambda})$, the *star premium* is the perceived surplus advantage of selecting the superstar over the best team player:

$$\Delta(\tilde{\beta}, \tilde{\lambda}) \equiv \bar{S}(\overline{\alpha}; \tilde{\beta}, \tilde{\lambda}) - \bar{S}(\underline{\alpha}; \tilde{\beta}, \tilde{\lambda}).$$

The principal prefers the superstar if $\Delta > 0$ and the best team player if $\Delta < 0$.

The following lemma characterizes how each bias affects the star premium.

**Lemma C.1** (Directional Effects of Overconfidence)**.**

*(i) Productivity overconfidence reduces the star premium: for $\tilde{\beta}' > \tilde{\beta}$, $\Delta(\tilde{\beta}', \tilde{\lambda}) < \Delta(\tilde{\beta}, \tilde{\lambda})$.*

*(ii) Collaborative-fit overconfidence increases the star premium: for $\epsilon > 0$, $\Delta(\tilde{\beta}, \lambda - \epsilon) > \Delta(\tilde{\beta}, \lambda)$.*

Part (i) reflects that a principal who believes she is more productive perceives less marginal value from a highly productive agent, tilting selection toward the best team player. Part (ii) reflects the asymmetric effect of perceived tradeoff intensity across the productivity range. A lower perceived $\lambda$ raises perceived surplus more for $\overline{\alpha} > 1$ than for $\underline{\alpha} < 1$, because the collaborative-fit penalty under Assumption **DS** is larger for highly productive agents. This tilts selection toward the superstar.

Because the two biases operate in opposite directions, they may partially offset. Recall from Corollary 1(ii) that, under the true parameters $\beta = 1$ and $\lambda$, the optimal partner choice is $\alpha^* \in \{\underline{\alpha}, \overline{\alpha}\}$. We say that the principal *selects correctly* if her partner choice under beliefs $(\tilde{\beta}, \tilde{\lambda})$ coincides with $\alpha^*$, and *mis-selects* otherwise.

**Proposition C.1** (Partial Debiasing and Mis-selection)**.** *Suppose $\lambda > \bar{\lambda}$ and both biases are present with beliefs $(\tilde{\beta}, \lambda - \epsilon)$, where $\tilde{\beta} > 1$ and $\epsilon > 0$. If the biases offset so that the* *principal selects correctly, removing only one bias can induce mis-selection. In particular, removing productivity overconfidence tends toward over-selection of superstars, while removing collaborative-fit overconfidence tends toward over-selection of team players.*

The Moderate's Trap amplifies the welfare consequences of biased beliefs. When expected surplus is U-shaped, partner selection is bimodal, so overconfidence can cause the principal to select from the wrong arm of the U-shape, generating welfare losses equal to the full surplus gap $|\bar{S}(\overline{\alpha}) - \bar{S}(\underline{\alpha})|$.

Proposition C.1 cautions that debiasing interventions may backfire when multiple biases coexist. A hiring manager who is overconfident about both her productive ability and her collaborative skills may nevertheless select correctly because the two biases offset. Removing only one bias leaves the other unopposed and may induce mis-selection. Organizations should therefore evaluate multiple dimensions of overconfidence jointly rather than addressing them in isolation.