TOO GOOD TO FIRE: NON-ASSORTATIVE MATCHING TO PLAY A DYNAMIC GAME

BENJAMIN SPERISEN AND THOMAS WISEMAN

Abstract. We study stable outcomes in a two-sided, one-to-one matching market with firms and workers. The model provides a way of endogenizing how transferable utility is within a match: when a firm-worker pair is matched, they play an infinite-horizon discounted dynamic game with one-sided, observable effort. Partners' types are complementary in determining the maximal feasible payoffs. In our setting, the classic result that with complementary types stable matchings are positively assortative does not hold. Increasing the quality of a match harms dynamic incentives, because a firm cannot credibly threaten to fire a worker who is productive even with low effort. Thus, firms may prefer lower-type workers who will exert more effort. If wages are endogenous, then committing to pay a high wage can improve effort incentives indirectly by making the firm more willing to walk away.

1. Introduction

We consider frictionless matching markets for firms and workers. When a firm-worker pair is matched, they play an infinite-horizon dynamic game. In each period, the worker chooses how much effort to exert. The firm observes the effort level and decides whether to fire the worker or to continue the relationship for the next period. The potential output from a match increases both with the type of the worker (skill or ability) and the type of the firm (technology or capital stock). Further, firm and worker types are complementary – the maximal feasible payoffs are a supermodular function of the partners' types. The actual payoffs that the partners achieve are

Date: June 27, 2016.

PRELIMINARY AND INCOMPLETE – Please do not cite.

We thank V. Bhaskar, Dan Hamermesh, John Hatfield, Ayça Kaya, Leslie Marx, and seminar participants for helpful comments.

Sperisen: Department of Economics, Tulane University, bsperise@tulane.edu. Wiseman: Department of Economics, University of Texas at Austin, wiseman@austin.utexas.edu.

determined in equilibrium. Classic results from the literature on two-sided matching (starting with Becker, 1973) show that when payoffs are increasing and supermodular, then stable matchings are positively assortative: high-type workers match with high-type firms. In our setting, that result does not hold. The reason is that increasing the quality of a match harms effort incentives by reducing the firm's willingness to fire. That effect dominates the direct positive effect of complementarity in types, so that some firms prefer lower-type workers who will exert more effort. Adding endogenous wages may restore the stability of positively assortative matching, but the effect depends in a subtle way on the complementarity between effort and worker ability.

The feature that distinguishes our model from the standard matching environment is that here payoffs from a match are determined endogenously. The matching literature typically specifies those payoffs as an exogenous function of types. There are two main branches of that literature. In transferable utility (TU) models, partners can split the payoff generated from their match however they like. In a setting with non-transferable utility (NTU), in contrast, a match generates a specific payoff for each partner. Becker (1973) shows that 1) in the NTU case, stable matchings are positively assortative if each agent's payoff is increasing in his partner's type, and 2) in the TU case, supermodularity of the match value is sufficient for positively assortative matching. Legros and Newman (2007) consider an intermediate case where the degree of transferability depends exogenously on types. In our model, transferability is endogenous. Once they are matched, a firm and worker can get any payoffs in the equilibrium set of the repeated game. In our baseline setting, the wage is fixed and independent of types, as might be the case in labor markets where the wage level is set either by formal collective bargaining contracts (notably some professional sports leagues and government employment) or by informal industry standards (as in the market for new assistant professors). We also consider a setting with endogenous wages, where a matched firm and worker can commit to a per-period wage to be paid to the worker as long as the firm does not fire him.

Formally, let the stage-game payoff in a match between a firm (F) of type x and a worker (W) of type y be given by $U^r(e; x, y)$, where $r \in \{F, W\}$ denotes the role of

the agent and $e \in [0,1]$ denotes the level of effort exerted by the worker. Each payoff $U^r(e; x, y)$ is increasing and supermodular in the types x and y. Effort is costly for the worker but beneficial for the firm, so $U^F(e; x, y)$ is increasing in e and $U^W(e; x, y)$ is decreasing in e. If the firm fires the worker, then both get continuation payoffs of 0. A key assumption in the model is that for low values of x and y, the firm's payoff $U^F(e; x, y)$ is positive if and only if effort e is high enough, while for high values of x and y, the firm's payoff is strictly positive even at zero effort. The consequence is that in matches between high types, the only subgame perfect equilibrium is that the worker exerts zero effort in every period and the firm never fires him: the firm's minimum payoff in the stage game is higher than the payoff from firing. On the other hand, in matches with lower types, the firm prefers to fire the worker if it expects low effort in the future. In that case, firing is a credible threat and effort can be sustained in equilibrium.

That discontinuity in the equilibrium payoff set has implications for stable matching. In evaluating potential partners, a firm faces a potential trade-off between a high-type worker who will not exert effort and a worker with a lower type who will. In particular, the most preferred worker for a firm of type x is either 1) the worker with the highest type y, or 2) the highest-type worker who will exert effort – that is, whose type $y^e(x)$ satisfies $U^F(e; x, y^e(x)) = 0$. As a result, a firm's preferences over workers depend on the firm's type x, even though $U^F(e; x, y)$ is increasing in y for all x and e.

Our findings are as follows. Suppose that the market is thick, in the sense that the distance between an agent's type and the type of the next closest agent is small, and that effort is efficient, in the sense that firms are willing to trade partner type for effort at a higher rate than workers are. In our baseline setting of a fixed wage, we show that positively assortative matching is not stable and that a stable, non-assortative matching exists. With endogenous wages, the only stable matchings are again non-assortative if effort and worker type are substitutes in the firm's payoff function. Effort and ability are likely to be substitutes in industries with a standardized product,

where a more-skilled worker can produce a unit of output with less effort than a less-skilled worker. On the other hand, in industries that rely on innovation effort and ability may be complements: both are required to generate a breakthrough.

In Section 3, we construct a stable outcome in a parametrized example where the marginal utility of effort is independent of types for both firms and workers. Because the payoff functions $U^F(e;x,y)$ and $U^W(e;x,y)$ are supermodular in types, the marginal benefit from matching with a better type of worker increases with the firm's own type: high-type firms care more about their partners' types and less about effort relative to low-type firms. Roughly speaking, then, firms at the top of the type distribution prefer high-type partners, while firms in the middle want to match with lower types to get more effort. The stable matching follows that pattern: positively assortative matching at the top and negatively assortative matching in the middle, where a type-x firm matches with a worker of type $y^e(x)$ (the highest-type worker who will exert effort when matched with the type-x firm). At the bottom, a firm with a very low type x will be able to get effort from any partner, since $U^F(0;x,y)$ will be negative, so it prefers high-type partners, and matching will again be positively assortative.

Thus, the matching between firms and workers is positively assortative at the top and bottom of the type distribution and negatively assortative in the middle. A complication arises, however. Because the function $y^e(x)$ is generically nonlinear, matching firms of type x with workers of type $y^e(x)$ is not feasible – the mass of firms matched would differ from the mass of workers. Instead, it turns out that a stable matching in the middle is very discontinuous. Once firms with types near x "run out" of their most-preferred workers $y^e(x)$ to match with, some of those must match with their "second favorites," whose types are discretely lower.

Adding endogenous wages changes which matchings are stable by indirectly affecting effort incentives: committing to paying a wage lowers the firm's payoff for any effort level and so increases its willingness to fire the worker. In the absence of wages, if the payoff from effort 0 to a type-x firm matched with a type-y worker, $U^F(0;x,y)$, is positive, then the firm cannot credibly threaten to fire. If the firm commits to a wage of at least $U^F(0;x,y)$, though, then the firm will be willing to fire, and that

threat can be used to enforce effort in equilibrium. In that case, the firm's payoff net of the wage is $U^F(e;x,y) - U^F(0;x,y)$. In the example, the marginal benefit of effort is independent of worker type, so the firm is indifferent over partners, and there are multiple stable matchings, including the positively assortative matching. The wage that workers receive is higher than the minimum level that would induce effort under the threat of being fired. The motivation behind these "efficiency wages" is not directly to motivate workers by increasing the difference between the continuation wage if the worker exerts and effort and the outside option if he shirks and is fired. Instead, the higher wage improves the firm's incentive to punish the worker, and so indirectly improves the worker's effort incentives.

If we relax the knife's edge assumption that the marginal utility of effort does not vary with the worker's type, then the firm's net payoff $U^F(e;x,y) - U^F(0;x,y)$ and thus its preferences over workers depend on whether effort and type are substitutes or complements. If they are substitutes, then the firm prefers lower-type workers because $U^F(e;x,y) - U^F(0;x,y)$ is decreasing in y. In that case, positively assortative matching is not stable.

Related literature. As mentioned above, Legros and Newman (2007) examine a matching environment where the extent to which partners can transfer utility to each other is an exogenous function of types. Their sufficient condition for positively assortative matching to be stable fails in our setting. (See Footnote 1 below.)

Two other closely related papers consider matching where the transferability of utility is endogenous. Both Citanna and Chakraborty (2005) and Kaya and Vereshchagina (2015) study settings where partners derive payoffs from a repeated interaction with two-sided moral hazard. Citanna and Chakraborty (2005) find that pairing high-wealth agents with low-wealth partners improves effort incentives. In Kaya and Vereshchagina's (2015) model, effort is unobservable, and an agent's type affects the probability that each period's publicly observed output is high. They find that in some cases negatively assortative matching is stable. In both papers, an agent's type directly affects the ability to provide incentives. Serfes (2005, 2008) and Wright (2004) consider similar issues in modls of one-sided moral hazard where a worker's type is his level of risk aversion and the firm's type is the riskiness of its project.

The structure of the rest of the paper is as follows: Section 2 presents the model. In Sections 3 and 4we analyze a parametrized model in detail, both with and without endogenous wages. In Section 5 we show how the results from the example generalize, and in Section 6 we conclude.

2. Model

We consider a one-to-one, two-sided matching market. There are N firms and N workers. Agents are heterogeneous, with types lying in $\left[\underline{\theta}, \bar{\theta}\right] \subseteq \mathbb{R}$. Firm n's type is denoted x_n, where $x_1 < x_2 < \ldots < x_N$. Similarly, worker n's type is y_n, where $y_1 < y_2 < \ldots < y_N$. The sets of agent types are denoted by

$$\mathcal{X}_N \equiv \{x_1, ...x_N\}, \mathcal{Y}_N \equiv \{y_1, ...y_N\}.$$

Some of our results concern the case where the number of agents N grows large. In that case, we focus on sequences $(\mathcal{X}_N, \mathcal{Y}_N)_N$ with the feature that the type distributions for both firms and workers converge weakly to the continuous uniform distribution over $\left[\underline{\theta}, \bar{\theta}\right]$.

Types are publicly observed. When a type-x firm is matched with a type-y worker, they play the following infinite-horizon dynamic game: in each period, the firm decides whether to continue the game or to end it by firing the worker. If it fires the worker, then the continuation payoff for each player is 0. If it continues the game, then the worker chooses an effort level $e \in [0,1]$, which is publicly observed. The resulting instantaneous payoffs are $U(e; x, y) = \left(U^F(e; x, y), U^W(e; x, y)\right)$, where $U^F(e; x, y)$ is the firm's payoff and $U^W(e; x, y)$ is the worker's. Payoff functions are continuous and twice continuously differentiable, and they have the following properties:

Assumption. *For any $e \in [0,1]$, $x' > x$, and $y' > y$,*

(1) $U^F_e(e; x, y) > 0, U^F_{ee}(e; x, y) \leq 0,$
$U^W_e(e; x, y) < 0, U^W_{ee}(e; x, y) \leq 0,$ *and*
$U^F_e(e; x, y) + U^W_e(e; x, y) > 0.$

(2) $U^r(e; x, y') > U^r(e; x, y),$
$U^r(e; x', y) > U^r(e; x, y),$ *and*
$U^r(e; x, y) + U^r(e; x', y') > U^r(e; x', y) + U^r(e; x, y'), r \in \{F, W\}.$

(3) $U^W(1; x_1, y_1) > 0$,
$U^F(0; x_1, y_1) < 0 < U^F(0; x_N, y_N)$, *and*
$U^F(1; x_1, y_1) > 0$.

Assumption 1 says that effort is costly for the worker and beneficial for the firm, that the benefit exceeds the cost, and that the marginal cost weakly increases and the marginal benefit weakly decreases with the effort level. Assumption 2 says that payoffs are increasing and supermodular in types. Assumption 3 relates stage-game payoffs to the continuation payoff of 0 if the worker is fired. Workers get a strictly positive payoff in the stage game for any effort level and pair of types, as do firms with high enough types. Low-type firms, on the other hand, get a negative payoff when matched with a low-type worker who exerts low effort.

Firms and workers discount the future at the common rate $\delta \in (0, 1)$ per period, where

$$\delta > \max_{x,y} \frac{U^W(0; x, y) - U^W(1; x, y)}{U^W(0; x, y)} \in (0, 1). \tag{2.1}$$

That is, the threat of firing is enough to induce full effort from a worker regardless of match quality. The total payoff from the dynamic game if the worker is fired after T periods and exerts effort e_t in period $t \leq T$ is

$$(1 - \delta) \sum_{t=1}^{T} \delta^{t-1} U(e_t; x, y).$$

A public randomization device is available.

Let $E^\delta(x, y) \in \mathbb{R}^2$ denote the set of subgame perfect equilibrium (SPE) payoffs of the dynamic game with a firm of type x and a worker of type y. Given the restriction on the value of the discount factor δ, the equilibrium payoff set has the form described in Lemma 1 below. Define $\underline{e}(x, y)$ as the minimum level of effort that gives the firm a payoff at least as good as its outside option:

$$\underline{e}(x, y) \equiv \text{argmin}_{e \in [0,1]} \left\{ U^F(e; x, y) \geq 0 \right\}.$$

Let $\hat{V}(x, y) \equiv \{U(e; x, y) : e \geq \underline{e}(x, y)\}$ denote the set of feasible stage-game payoffs that give the firm a payoff of at least 0.

Lemma 1. *The equilibrium set* $E^\delta(x, y)$ *depends on the value of* $U^F(0; x, y)$*:*

- *If* $U^F(0;x,y) > 0$*, then* $E^\delta(x,y) = \{U(0;x,y)\}$.
- *If* $U^F(0;x,y) \leq 0$*, then* $E^\delta(x,y) = co\left\{\hat{V}(x,y),(0,0)\right\}$.

If the firm's payoff when the worker exerts zero effort is strictly higher than the payoff from firing (if $U^F(0;x,y) > 0$), then in equilibrium the firm will never fire the worker. In that case, the only SPE has the worker never exert effort and the firm never fire. If, on the other hand, $U^F(0;x,y) \leq 0$, then there is a SPE where the worker's strategy is to always exert zero effort and the firm's is to always fire. The threat of reverting to that equilibrium can be used to support any effort level in SPE.

This discontinuity in the equilibrium payoff set will generate non-assortative matching. Although the feasible payoffs exhibit complementarity in types, equilibrium payoffs need not.[1]

2.1. **Stability.** In our setting, it is necessary to describe both which firm gets matched with which worker and the play in the dynamic game resulting from each match. An *outcome* consists of a *matching* and an *equilibrium selection rule*. A matching $\mu : \mathcal{X}_N \to \mathcal{Y}_N$ specifies for each type x of firm the type $\mu(x)$ of worker that the firm matches with. An equilibrium selection rule γ maps each matched pair of types $(x,\mu(x))$ to an SPE payoff in $E^\delta(x,\mu(x))$.

We say that an outcome (μ,γ) is *stable* if there is no blocking pair. A firm of type x and a worker of type y form a blocking pair if 1) $y \neq \mu(x)$, so that the firm and worker are not matched; and 2) there is a payoff in $E^\delta(x,y)$ that both firm and worker strictly prefer to the payoffs in their current outcomes. The idea is that a firm (or worker) can ask a worker (or firm) to leave his current partner and join it instead, and in making that request can propose an equilibrium to play in the new match. Crucially, the proposal cannot specify an effort level that is not part of an SPE. That is, an outcome includes a fixed SPE for each match, but a potential blocking pair can deviate to any SPE.

[1] As a result, Legros and Newman's (2007) sufficient condition for stable matchings to be positively assortative, *generalized increasing differences*, fails here. The generalized increasing differences condition requires that the transferability of utility is increasing in types.

We say that a matching μ is stable if there exists an equilibrium selection rule γ such that the outcome (μ, γ) is stable. We denote the positively assortative matching by μ^+: $\mu^+(x_n) = y_n$ for all n.

Note that a stricter definition of stability would also require that a matched pair not be able to deviate to a Pareto superior SPE. Under that definition, the equilibrium selection rule in a stable outcome would specify for each match an SPE with a payoff on the Pareto frontier of $E^\delta(x, y)$. Our results would still hold under that stricter notion.

2.2. **Adding wages.** We will also consider the case where a matched firm and worker can agree on a nonnegative wage w to be paid to the worker in every period until he is fired. The stage-game payoffs are then $U^F(e; x, y) - w$ and $U^W(e; x, y) + w$ for the firm and the worker, respectively. Continuation payoffs after the firm fires the worker are still 0.

The set of SPE payoffs for a firm of type x matched with a worker of type y now depends on the wage w. The firm's payoff from zero effort is now $U^F(0; x, y) - w$. The firm is willing to fire if and only if that payoff is no greater than 0. Let $\underline{e}(x, y; w)$ be the minimum level of effort that gives the firm a payoff at least as good as its outside option; such an effort level exists if the wage is not too high. In particular,

$$\underline{e}(x, y; w) \equiv \text{argmin}_{e \in [0,1]} \left\{ U^F(e; x, y) - w \geq 0 \right\}$$

if $U^F(1; x, y) \geq w$. Let

$$\hat{V}(x, y; w) \equiv \left\{ \left(U^F(e; x, y) - w, U^W(e; x, y) + w \right) : e \geq \underline{e}(x, y; w) \right\}$$

denote the set of feasible stage-game payoffs that give the firm a payoff of at least 0. Then, paralleling the analysis for the baseline case without wages, the set of SPE payoffs $E^\delta(x, y; w)$ has the following structure. (The proof of Lemma 2 mirrors the proof of Lemma 1 and is omitted.)

Lemma 2. *In the setting with wages, the equilibrium set $E^\delta(x, y; w)$ depends on the values of $U^F(0; x, y)$, $U^F(1; x, y)$, and w:*

- *If $U^F(0; x, y) - w > 0$, then $E^\delta(x, y; w) = \left\{ \left(U^F(0; x, y) - w, U^W(0; x, y) + w \right) \right\}$.*
- *If $U^F(0; x, y) - w \leq 0 \leq U^F(1; x, y) - w$, then $E^\delta(x, y) = co \left\{ \hat{V}(x, y; w), (0, 0) \right\}$.*

- *If* $U^F(1;x,y) - w < 0$, *then* $E^\delta(x,y;w) = \{(0,0)\}$.

Note that paying a wage can in some cases increase the highest SPE payoff for the firm. In particular, if $U^F(0;x,y) > 0$, then with no wage the only SPE gives the firm a payoff of $U^F(0;x,y)$. At wage $w = U^F(0;x,y)$, though, there is an SPE where the worker always exerts full effort, giving the firm $U^F(1;x,y) - U^F(0;x,y)$. Thus, if $U^F(1;x,y) > 2U^F(0;x,y) > 0$, then the firm may be better off paying a positive wage.

In this setting, an outcome $(\mu, \gamma, \mathbf{w})$ specifies for each matched pair a per-period wage $\mathbf{w}(x, \mu(x))$ and a payoff consistent with some SPE given the wage. That is, $\gamma(x, \mu(x))$ must lie in $E^\delta(x, y; \mathbf{w}(x, \mu(x)))$. An outcome $(\mu, \gamma, \mathbf{w})$ is stable if there is no firm of type x and worker of type y such that 1) $y \neq \mu(x)$; and 2) there exists a wage $w \geq 0$ and a payoff γ in $E^\delta(x,y;w)$ such that both firm and worker strictly prefer γ to the payoffs in their current outcomes. Finally, we say that a matching μ is stable if there exist an equilibrium selection rule γ and a wage rule $\mathbf{w}$ such that the outcome $(\mu, \gamma, \mathbf{w})$ is stable.

3. Example

In this section, we examine a parameterized example to illustrate our results on stable outcomes. In the example, firm and worker types are distributed across the interval $[0,2]$. We will focus on the case where N grows large, so that the type distribution approaches a continuous uniform distribution over $[0,2]$. For expositional purposes, we will sometimes work directly with that limiting distribution.

The stage-game payoff functions are the following:

$$U^F(e;x,y) = 2e - 1 + xy$$

and

$$U^W(e;x,y) = 2 - e + xy.$$

When the match quality is high enough – specifically, when $xy > 1$ – then the firm will never choose to fire the worker. If $xy \leq 1$, on the other hand, then any effort level below $\underline{e}(x,y) = (1 - xy)/2$ gives the firm a payoff below zero. From 1, the SPE

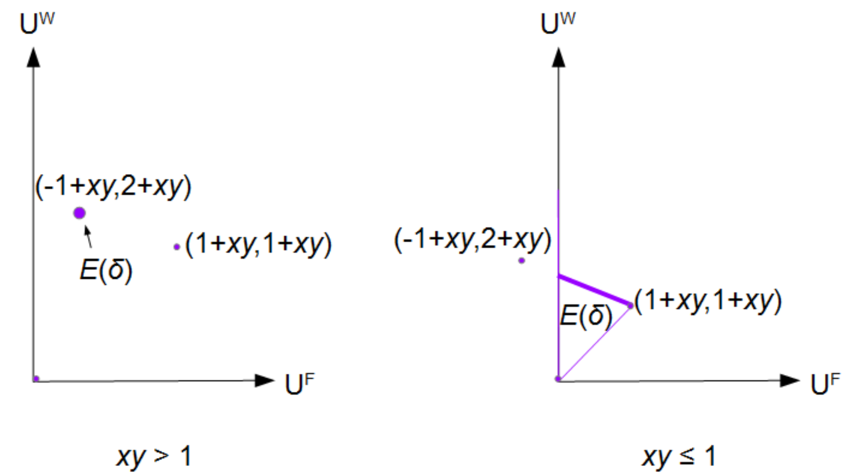

FIGURE 3.1. SPE payoffs as a function of match quality

payoff set $E^\delta(x, y)$ is given by

$$E^\delta(x,y) = \begin{cases} \text{co}\left\{(0,0),(1+xy,1+xy),(0,\frac{3}{2}(1+xy)\right\} & \text{if } xy \leq 1 \\ (-1+xy,2+xy) & \text{if } xy > 1. \end{cases}$$

(Recall that the discount factor δ is high enough – in this case, $\delta > \frac{1}{2}$ – that in any match the threat of being fired is sufficient to induce high effort.) The equilibrium set is illustrated in Figure 3.1.

3.1. **Positively assortative matching is not stable.** In this example, the positively assortative matching μ^+ is not stable when N is large enough – that is, when types are close enough together. Under μ^+, positive effort is sustainable in equilibrium when $x_n y_n \leq 1$. The problem arises at the transition where effort is no longer sustainable. For simplicity, consider the special case where $\mathcal{X}_N = \mathcal{Y}_N$ for all N, so that $x_n = y_n$. Let $x_{n^*(N)}$ be the lowest type of firm strictly higher than 1, and let $y_{n^{**}(N)} < y_{n^*(N)}$ be the highest type of worker such that $x_{n^*(N)} y_{n^{**}(N)} \leq 1$. Then the firm of type-$x_{n^*(N)}$ and the worker of type $y_{n^{**}(N)}$ form a blocking pair as long as N

is high enough. Since $x_{n^*(N)}y_{n^*(N)} > 1$, the type-$x_{n^*(N)}$ firm's match partner will not exert effort in equilibrium, and so the firm's payoff will be $-1 + x_{n^*(N)}y_{n^*(N)}$. The firm would prefer to match with the type-$y_{n^{**}(N)}$ worker if that worker exerted effort e^* strictly above $\underline{e} \equiv \frac{1}{2}x_{n^*(N)}(y_{n^*(N)} - y_{n^{**}(N)})$:

$$2e^* - 1 + x_{n^*(N)}y_{n^{**}(N)} > -1 + x_{n^*(N)}y_{n^*(N)}.$$

The type-$y_{n^{**}(N)}$ worker's current payoff is at most $2+x_{n^{**}(N)}y_{n^{**}(N)}$, so he also prefers to match with the type-$x_{n^*(N)}$ firm and exert effort e^* as long as e^* is strictly below $\bar{e} \equiv y_{n^{**}(N)}(x_{n^*(N)} - x_{n^{**}(N)})$:

$$2 - e^* + x_{n^*(N)}y_{n^{**}(N)} > 2 + x_{n^{**}(N)}y_{n^{**}(N)}.$$

For large N, $\underline{e} < 1$. Further, because $x_{n^*(N)} - x_{n^{**}(N)} = y_{n^*(N)} - y_{n^{**}(N)}$, we have $\bar{e} > \underline{e}$. Since $x_{n^*(N)}y_{n^{**}(N)} \leq 1$, any effort level is sustainable in equilibrium when the type-$x_{n^*(N)}$ firm and the type-$y_{n^{**}(N)}$ worker are matched. Thus, they form a blocking pair.

Intuitively, the reason that the switch benefits both is that firms put a higher utility weight on effort relative to match quality than do workers. The fact that payoffs are complementary in types acts in the opposite direction: the (high) type-$x_{n^*(N)}$ firm cares more about the type of its partner than does the (low) type-$y_{n^{**}(N)}$ worker. If the difference in types is small, though (that is, if N is large), then the first effect dominates: the type-$x_{n^*(N)}$ firm is willing to accept a slightly lower-quality partner in exchange for positive effort, while the type-$y_{n^{**}(N)}$ worker is willing to supply that increased effort for a higher-quality partner.

3.2. **A stable outcome.** Here we focus on the preferences of firms over match partners to derive the properties of stable outcomes.

The firm with the highest type, $x = 2$, prefers to match with the highest type worker, $y = 2$, and get no effort rather than match with the highest type of worker who would exert effort, $y = \frac{1}{2}$. That is, $-1 + 0 + 2 \cdot 2 > -1 + 2 + 2 \cdot \frac{1}{2}$. Since firm $x = 2$ and 0 effort is the most-preferred match result for the worker with type 2, that pairing must be part of any stable outcome. By the same argument, the next highest firm type, $x = 2 - \frac{1}{N}$, will match with the next highest worker type, $y = 2 - \frac{1}{N}$, who

will exert no effort (as long as $N > 3$). That logic extends for every firm with a type above $x = \sqrt{3} \approx 1.73$: because $-1 + x^2 > 2$, those firms prefer zero effort from a worker of their own type to full effort from the best potential partner willing to work when matched with them, $y = 1/x$. Thus, every stable outcome features positively assortative matching at the top.

Below $x = \sqrt{3}$, firms' preferences depend on equilibrium effort levels. For a start, we look for a *within-match firm-optimal* outcome $(\tilde{\mu}, \gamma^{\max})$, where the equilibrium selection rule $\gamma^{\max}$ specifies the highest effort level consistent with equilibrium for each match in μ. That is, if $xy \leq 1$, then the worker exerts effort 1, yielding payoffs $(1 + xy, 1 + xy)$. If $xy > 1$, then the worker puts in 0 effort, and payoffs are $(-1 + xy, 2 + xy)$.

In that case, a firm with a type just below $x = \sqrt{3}$ prefers being matched with the highest type of worker who will exert effort, $y = 1/x$, rather than getting no effort from the best worker type remaining, $y = \sqrt{3}$. The same argument applies for firm types down to $x = 1/\sqrt{3} \approx 0.58$. In that range there is *negatively* assortative matching as a firm of type x matches with a worker of type $1/x$. Below $x = 1/\sqrt{3}$, a firm's type is low enough that effort will be attainable ($xy < 1$) when the firm is matched with any type of worker remaining, and so matching is positively assortative again at the bottom. Figure 3.2 illustrates that matching.

The problem with the matching just described, however, is that the continuous limit (that is, as N grows without bound) of the matching for middle types – where a type x firm matches with a type $1/x$ worker – is not measure preserving. It specifies, for instance, that firms with types between 1 and $\sqrt{3}$ (measure $\frac{1}{2}(\sqrt{3}-1) \approx 0.37$) are matched to the workers with types between 1 and $1/\sqrt{3}$ (measure $\frac{1}{2}(1-1/\sqrt{3}) \approx 0.21$). Too many firms are matched to too few workers. Similarly, the relatively few firms with types between 1 and $1/\sqrt{3}$ are supposed to match with the relatively many workers with types between 1 and $\sqrt{3}$. For a large but fixed N, the problem is that for a firm of type x, its ideal type $y = 1/x$ of worker may not be an element of $\mathcal{Y}_N$. In that case, the firm's most preferred existing worker will be

$$\max\left\{y \in \mathcal{Y}_N : y \leq \frac{1}{x}\right\}.$$

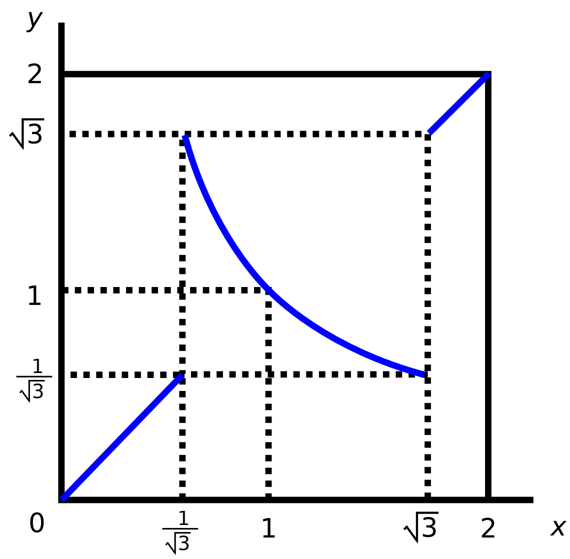

FIGURE 3.2. A non-measure preserving matching

But then multiple firms will have the same highest worker type who will exert effort for them.[2] Thus, it is not feasible to match each firm with the highest worker type who will exert effort.

The solution is that in the region where the firms outnumber their most preferred workers (that is, firm types in $[1, \sqrt{3}]$), some of the firms must match with their second choices instead. In this "high middle" region, first match the firms with their preferred worker types $1/x$ to the extent possible. There will be a leftover mass of $\frac{1}{2}(\sqrt{3}-1) - \frac{1}{2}(1-1/\sqrt{3}) \approx 0.15$ firms still unmatched. Those firms must match with their second-favorite workers. For example, consider a firm whose type x is just below $\sqrt{3}$. If the firm cannot get its preferred worker (among those workers remaining after firms with higher types have matched), type $1/x$ just above $1/\sqrt{3}$, then its next choice will be either a slightly lower type or a much higher type. (Recall that worker types between $y = 1$ and $y = \sqrt{3}$ are still available, as are types below $1/\sqrt{3}$.) The firm would never chose a type y *slightly* higher than $1/x$ – that worker would not exert effort, and conditional on no effort the firm would rather match with an even higher

[2]For example, suppose that types lie in $\{0.1, 0.2, \ldots, 1.9, 2\}$. Then for a firm of type 1.3, the largest y such that $1.3y \leq 1$ is $y = 0.7$. But $y = 0.7$ is also the largest y such that $1.4y \leq 1$.

type. Suppose that the firm prefers the slightly lower type. Then it will match with a worker of type just below $1/\sqrt{3}$; call that type $\hat{y}(x)$.

Next, a firm with a slightly lower type x' faces a similar choice. Suppose that it also, rather than matching with a much higher type, prefers to match with the highest remaining worker who will exert effort, which now is the type just below $\hat{y}(x)$, $\hat{y}(x')$. Suppose that all the leftover firms with types from $x = \sqrt{3}$ down to $x = \sqrt{3} - \Delta$ for some small $\Delta > 0$ match in that fashion. Then overall, the measure $\frac{1}{2}\Delta$ of firms with types in that interval (including those who matched in the "first round" with worker type $1/x$) match with the workers of types below $1/(\sqrt{3} - \Delta)$ and above $\hat{y}(\sqrt{3} - \Delta)$, because $1/x$ is decreasing in x and $\hat{y}(x)$, by construction, is increasing in x. For the measure of workers to match the measure of firms, it must be that $1/(\sqrt{3} - \Delta) - \hat{y}(\sqrt{3} - \Delta) = \Delta$, so $\hat{y}(\sqrt{3} - \Delta) = 1/(\sqrt{3} - \Delta) - \Delta$.

This construction does in fact lead to a stable matching. A firm with type $x \in [1, \sqrt{3}]$ matches either with a worker of type $1/x$ or with one of type

$$\hat{y}(x) \equiv \frac{1}{x} - \left(\sqrt{3} - x\right) = \frac{1}{x} + x - \sqrt{3}.$$

The need to match equal measures of firms and workers pins down the probability of each of a firm's two potential matches. If a firm of type x matches with a worker of type $1/x$ with probability $1/x^2$, then the total measure of firms in those matches equals the measure of available workers: for any positive $\Delta \leq \sqrt{3} - 1$,

$$\int_{\sqrt{3}-\Delta}^{\sqrt{3}} \frac{1}{x^2} \cdot \frac{1}{2} dx = \frac{1}{2}\left(\frac{1}{\sqrt{3} - \Delta} - \frac{1}{\sqrt{3}}\right) = \int_{\frac{1}{\sqrt{3}}}^{\frac{1}{\sqrt{3}-\Delta}} \frac{1}{2} dy.$$

With the complementary probability $1 - 1/x^2$, a firm of type x matches with a worker of type $\hat{y}(x)$.

It remains to verify that each firm of type $x \in [1, \sqrt{3})$ prefers high effort from a worker of type $\hat{y}(x)$ to no effort from worker type $y = \sqrt{3}$. Since

$$-1 + x\sqrt{3} < 1 + x\left(\frac{1}{x} + x - \sqrt{3}\right)$$

in that range, $\hat{y}(x)$ is in fact the firm's second choice after $y = 1/x$. Thus, in the proposed matching firms with types in $[1, \sqrt{3}]$ match with the workers with types in $[2 - \sqrt{3}, 1]$.

For the "low middle" region for firm types, $x \in \left[1/\sqrt{3}, 1\right]$, the opposite problem arises: the range of firms with such a type x is smaller than the range $[1, \sqrt{3}]$ of their preferred worker types $1/x$. After matching those firms, there will be a leftover mass of workers equal to $\frac{1}{2}(\sqrt{3} - 1) - \frac{1}{2}(1 - 1/\sqrt{3}) \approx 0.15$. Those workers then match with the next available firms, those with types from $x = 2 - \sqrt{3} \approx 0.27$ to $x = 1/\sqrt{3} \approx 0.58$, in positively assortative fashion. The matching probabilities mirror those in the high middle region, switching the roles of firms and workers: a worker of type $y \in [1, \sqrt{3}]$ matches with a firm of type $1/y$ with probability $1/y^2$, and with a firm of type $\frac{1}{y} + y - \sqrt{3}$ with probability $1 - 1/y^2$. The firms with types in $[2 - \sqrt{3}, 1]$ match with the workers with types in $[1, \sqrt{3}]$.

The matching $\tilde{\mu}$ just constructed is summarized here:

$$\tilde{\mu}(x) = \begin{cases} x & \text{if } x \geq \sqrt{3} \\ \frac{1}{x} \text{ w/prob. } \frac{1}{x^2}, \hat{y}(x) \text{ w/prob. } 1 - \frac{1}{x^2} & \text{if } 1 \leq x < \sqrt{3} \\ \frac{1}{x} & \text{if } \frac{1}{\sqrt{3}} \leq x < 1 \\ \hat{y}^{-1}(x) & \text{if } 2 - \sqrt{3} \leq x < \frac{1}{\sqrt{3}} \\ x & \text{if } x < 2 - \sqrt{3} \end{cases}$$

Recall that $\hat{y}(x) \equiv \frac{1}{x} + x - \sqrt{3}$, so that for values of x between 1 and $\sqrt{3}$, $\hat{y}$ is strictly increasing and maps to $(2 - \sqrt{3}, 1/\sqrt{3})$. Thus, its inverse,

$$\hat{y}^{-1}(x) \equiv \frac{1}{2}\left(x + \sqrt{3} + (x^2 + 2x\sqrt{3} - 1)^{\frac{1}{2}}\right),$$

is well-defined. Figure 3.3 illustrates the matching $\tilde{\mu}$. The thickness of the matching curve represents the probability of the match.

3.2.1. *Effort levels.* Note, however, that the outcome $(\tilde{\mu}, \gamma^{\max})$ is not stable. The reason is that a firm inside the high middle region, with a type $x \in (1, \sqrt{3})$, strictly prefers one of his possible partners ($y = 1/x$) to the other ($y = \hat{y}(x)$) if both exert full effort, as the the equilibrium selection rule $\gamma^{\max}$ specifies. But then a type-x firm matched with a type-$\hat{y}(x)$ worker and a type-$(1/x)$ worker matched with a different

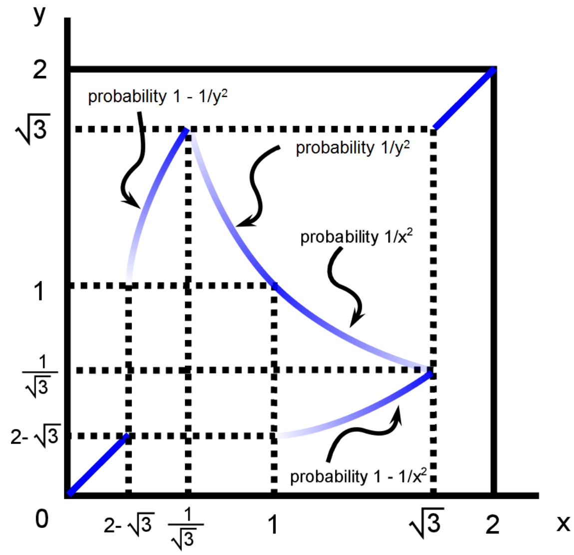

FIGURE 3.3.

type-x firm form a blocking pair. If those two match with each other and the worker exerts effort e' just below 1, then both do strictly better than in their assigned match and equilibrium. To get stability, we adjust γ so that the firm is indifferent between its two possible matches. In particular, a firm of type x gets effort 1 if it is matched with a worker of type $\hat{y}(x)$, and effort $\hat{e}(x) \equiv 1 - \frac{1}{2}x(\sqrt{3} - x)$ if it is matched with a worker of type $1/x$; the firm's payoff is $2 - x(\sqrt{3} - x)$ in either case.

In the low middle region of firms, where the firms are outnumbered by their preferred workers, specifying full effort does not cause a problem for stability. A worker with a type $x \in (1, \sqrt{3})$ who has matched with a firm of type $x = \hat{y}(y)$ would strictly prefer to match with his other possible partner ($x = 1/y$). But the firms of that type are already getting effort 1 from another worker of type y, so the worker cannot entice such a firm to switch.

One last issue involves the "left over" workers with types between 1 and $\sqrt{3}$: the match between a firm of type x' just above $2 - \sqrt{3}$ and a worker of type y' just above 1 must specify effort strictly less than 1. Otherwise, the type-y' worker could offer effort just below 1 to a firm of type x just below $2 - \sqrt{3}$ (matched with a worker of type $y = x < x'$ and getting effort 1) and form a blocking pair. In particular, the

type-y' worker exerts effort

$$\check{e}(y) \equiv \min\left\{1, \frac{13}{4} - 2\sqrt{3} - \frac{3}{2}\ln(y') + \frac{1}{4}(y')^2 + \frac{\sqrt{3}}{2}y'\right\}.$$

Thus, the matching $\tilde{\mu}$ is stable if the equilibrium level of effort $\tilde{e}$ in each match is as follows:

$$\tilde{e}(x, \tilde{\mu}(x)) = \begin{cases} 0 & \text{if } x \geq \sqrt{3} \\ \check{e}\left(\hat{y}^{-1}(x)\right) & \text{if } 2 - \sqrt{3} \leq x < \frac{1}{\sqrt{3}} \\ \hat{e}(x) & \text{if } 1 \leq x < \sqrt{3} \text{ and } \tilde{\mu}(x) = \frac{1}{x} \\ 1 & \text{otherwise} \end{cases}$$

Call the resulting equilibrium payoffs $\tilde{\gamma}$:

$$\tilde{\gamma}(x, \tilde{\mu}(x)) = \begin{cases} (-1 + x^2, 2 + x^2) & \text{if } x \geq \sqrt{3} \\ \left(2 - x(\sqrt{3} - x), 2 + \frac{1}{2}x(\sqrt{3} - x)\right) & \text{if } 1 \leq x < \sqrt{3} \text{ and } \tilde{\mu}(x) = \frac{1}{x} \\ \left(2 - x(\sqrt{3} - x), 2 - x(\sqrt{3} - x)\right) & \text{if } 1 \leq x < \sqrt{3} \text{ and } \tilde{\mu}(x) = \hat{y}(x) \\ (2, 2) & \text{if } \frac{1}{\sqrt{3}} \leq x < 1 \\ (-1 + 2\check{e}\left(\hat{y}^{-1}(x)\right) + x\hat{y}^{-1}(x), & \text{if } 2 - \sqrt{3} \leq x < \frac{1}{\sqrt{3}} \\ \qquad 2 - \check{e}\left(\hat{y}^{-1}(x)\right) + x\hat{y}^{-1}(x)) & \\ (1 + x^2, 1 + x^2) & \text{if } x < 2 - \sqrt{3} \end{cases}$$

Then the equilibrium selection rule $\tilde{\gamma}$ together with the matching $\tilde{\mu}$ represents a stable outcome in the continuous limit, in the following sense: there exists a sequence of sets of agent types $(\mathcal{X}_N, \mathcal{Y}_N)_N$ such that 1) the type distributions for both firms and workers converge weakly to the continuous uniform distribution over $\left[\underline{\theta}, \bar{\theta}\right]$, and 2) the outcome $(\tilde{\mu}, \tilde{\gamma})$ is stable for every N. The appendix gives a proof of that claim.

A firm's payoff as a function of its type x is graphed in Figure 3.4. Note in particular that the payoff is not monotonic in the firm's type: if x is just below 1, then the firm gets a payoff of 2, while a firm with a type just above 1 gets payoff $3 - \sqrt{3} \approx 1.27$. Figure 3.5, similarly, shows a worker's payoff as a function of its type y. A worker with a type y between 1 and $\sqrt{3}$ gets one of two possible payoffs, depending on whether he matches with a firm of type $x = 1/y$ or with a firm of type $\hat{y}(y)$. The first match

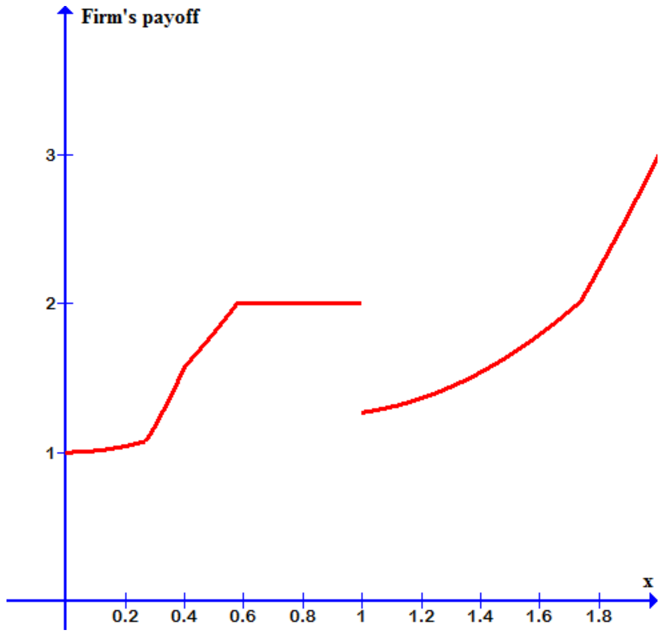

FIGURE 3.4. Firms' payoffs under the stable outcome $(\tilde{\mu}, \tilde{\gamma})$

occurs with probability $1/y^2$ and yields payoff 2; the second occurs with probability $1 - 1/y^2$ and yields payoff $2 - \check{e}\,(y) + y \cdot \hat{y}(y)$. The dotted line in Figure 3.5 shows the expected payoff. As with the firms, a worker's payoff is higher if his type is just below 1 than if it is just above. Intuitively, these nonmonotonicities arise from the fact that the function mapping a firm's type x to the type of the worker that it is just willing to fire, $1/x$, is not measure-preserving. For $x > 1$, the slope of that function is less than 1 in absolute value, so workers are the "short side" of the market. As described above, the resulting competition among relatively many firms for relatively few workers pushes down the firms' payoffs. For $x < 1$, the absolute value of the slope is greater than 1, and it is the firms who are in short supply.

Crawford (1991) and Ashlagi *et al.* (forthcoming) consider the effects on stable matching and payoffs of having more agents on one side of the market than on the other. Here, the imbalances arise endogenously in a setting where the two sides of the market are symmetric.

The stable outcome $(\tilde{\mu}, \tilde{\gamma})$ constructed above for the continuous limit (as N grows without bound) features "fractional matching" (Roth *et al.*, 1993): an agent of a given

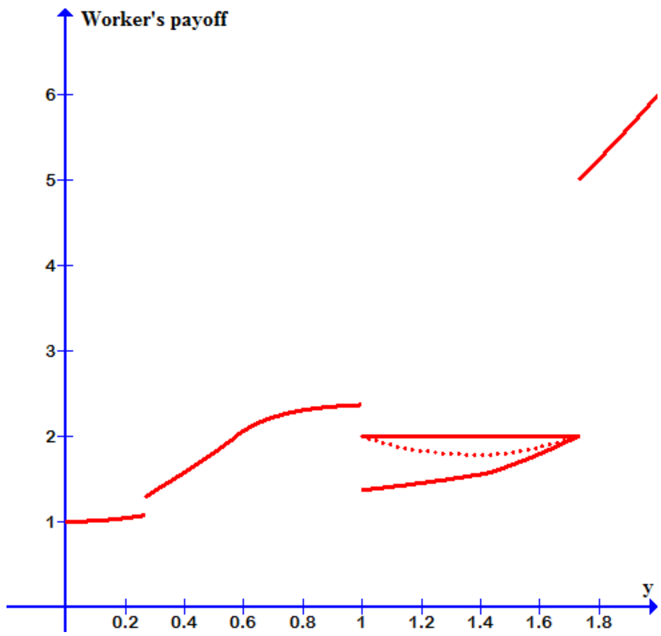

FIGURE 3.5. Workers' payoffs under the stable outcome $(\tilde{\mu}, \tilde{\gamma})$

type matches with two distinct types of partner with positive probability. For a large but fixed N, each agent type matches with a single partner type, but agents whose types are very close may match with very different partner types. Figure 3.6 shows an example.

4. Example with Wages

When a firm of type x pays a wage of $w \geq 0$ to a worker of type y who exerts effort e, the resulting stage-game payoffs are

$$U^W(e; x, y, w) = 2 - e + xy + w$$

and

$$U^F(e; x, y, w) = -1 + 2e + xy - w.$$

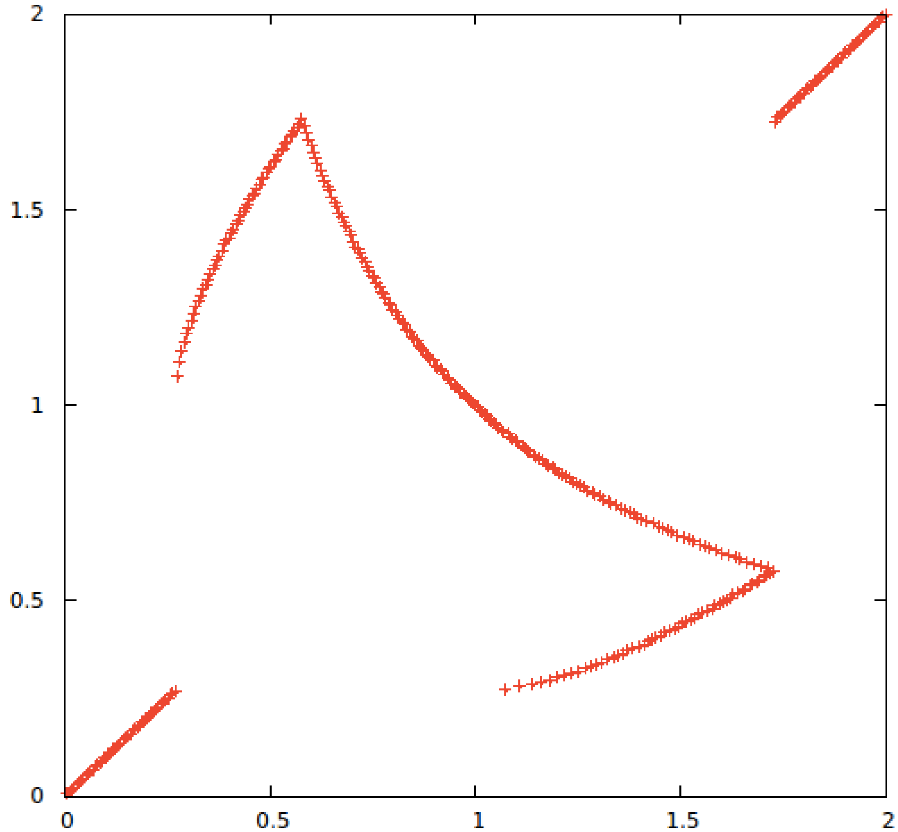

FIGURE 3.6. A discrete version of the stable matching $\tilde{\mu}$

The equilibrium payoff set of the stage game $E^\delta(x, y, w)$ is given by

$$E^\delta(x,y,w) = \begin{cases} (0,0) & \text{if } -1 \leq xy - w \\ \text{co}\left\{ \begin{array}{c} (0,0), (1+xy-w, 1+xy+w), \\ (0, \frac{3}{2}(1+xy) + \frac{1}{2}w \end{array} \right\} & \text{if } -1 \leq xy - w \leq 1 \\ (-1+xy-w, 2+xy+w) & \text{if } xy - w > 1. \end{cases}$$

For a firm type x and a worker type y such that $xy \geq 1$, define $\underline{w}(x,y) \equiv xy - 1 \geq 0$ as the lowest wage that makes effort sustainable. At that wage, the firm's payoff from zero effort is 0:

$$U^F(0; x, y, \underline{w}) = -1 + 2 \cdot 0 + xy - (xy - 1) = 0,$$

As a result, the firm is willing to fire the worker if it expects zero effort, and that threat can sustain any positive effort level in SPE. (Recall that when the wage is 0, effort is sustainable if and only if $xy \leq 1$.) At wage $\underline{w}(x, y)$ and effort $e = 1$, the payoff to the firm is

$$U^F(e; x, y, \underline{w}) = -1 + 2 + xy - (xy - 1) = 2.$$

Thus, the firm prefers paying $\underline{w}(x, y)$ and getting effort 1 over paying no wage and getting no effort if $2 > -1 + xy$; that is, if $xy < 3$.

Paying a wage in that situation improves the incentives for effort, but the effect is indirect. The worker would be willing to work to avoid being fired even if the wage were zero. Instead, committing to a positive wage increases the firm's willingness to discipline a worker who deviates. That is, the wage makes effort sustainable not because it makes the threat of being fired more costly for the worker, but because it makes that threat credible for the firm.

The fact that paying a wage can make effort sustainable even when the quality of a match is high (when $xy > 1$) will restore the stability of positively assortative matching in the example, as we show next.

4.1. **Positively assortative matching is stable.** We will again look for a stable outcome where a matched firm and worker choose the wage and effort level that give the firm the highest possible SPE payoff for that match. In particular, the wage will be either $\underline{w}(x, y)$ or 0, and effort will be 1 if it is sustainable.

As before, there will be positively assortative matching at the top. A firm with a type $x > \sqrt{3}$ prefers no wage and zero effort from a worker of its own type (yielding payoff $-1 + x^2$) to the payoff of 2 that it would get from full effort from either 1) the best potential partner willing to work wage 0 (type $y = 1/x$), or 2) a higher type of partner getting wage $\underline{w}(x, y)$.

Below that level, we can construct a stable outcome as follows: each firm with a type x between 1 and $\sqrt{3}$ matches with the worker of the same type $y = x$, pays wage $\underline{w}(x, x)$, and gets effort 1, yielding payoff 2. The outcome is stable in that region because no firm can get a payoff above 2 in SPE when matched with any remaining worker at any wage: the firm would have to pay a worker with type y greater than

$1/x$ a wage of at least $\underline{w}(x, y)$ to get effort, and matching with a lower type of worker gives a payoff below 2 even at effort $e = 1$ and wage $w = 0$. The firm might get a payoff above 2 by matching with a worker of type greater than $\sqrt{3}$ at $e = 0$, but such a blocking pair is ruled out by the supermodularity in types together with the efficiency of effort.

Finally, a firm with a type x below 1 does not need to pay a wage to induce effort: each such firm matches with the worker of the same type $y = x$, pays wage 0, and gets effort 1. To summarize, the constructed stable outcome $(\mu^+, \gamma^*, \mathbf{w}^*)$ is characterized by

$$\mathbf{w}^*(x, x) = \begin{cases} 0 & \text{if } x > \sqrt{3} \\ \underline{w}(x, x) & \text{if } 1 \leq x \leq \sqrt{3} \\ 0 & \text{if } x < 1 \end{cases}$$

and

$$\gamma^*(x, x) = \begin{cases} (-1 + x^2, 2 + x^2) & \text{if } x > \sqrt{3} \\ (2, 2x^2) & \text{if } 1 \leq x \leq \sqrt{3} \\ (1 + x^2, 1 + x^2) & \text{if } x < 1. \end{cases}$$

That outcome is illustrated in Figure 4.1.

In contrast to the stable outcome constructed in the setting without wages, here the payoff to a firm is (weakly) increasing in its type. The payoff to a worker is increasing in his type except at $y = \sqrt{3}$: a worker with a type just below that level gets payoff close to 6, while the worker just above gets only 5. The reason is that the total payoff in the match is higher just below $\sqrt{3}$, where there is full effort, than just above, where effort is 0. The endogenous wage allows some of that increased total to be passed on to the worker. Similarly, note that for types between 1 and $\sqrt{3}$, the firm's payoff is constant while the worker's is strictly increasing: all of the increase in match value goes to the worker.

4.2. **Negatively assortative matching is also stable.** For firms with types between 1 and $\sqrt{3}$, the argument above for why positively assortative matching is stable also applies to negatively assortative matching. Suppose that a firm with type

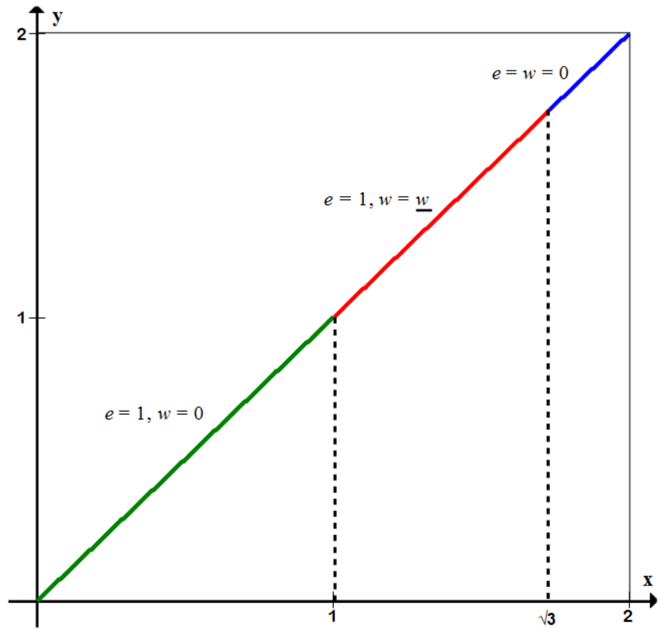

FIGURE 4.1. A stable, positively assortative outcome with wages

$x \in [1, \sqrt{3})$ matches with the worker of type

$$y^{neg}(x) \equiv 1 + \sqrt{3} - x,$$

pays wage $\underline{w}(x, y^{neg}(x))$, and gets effort 1, yielding payoff 2.[3] The outcome is stable in that region because, as before, no firm can get a payoff above 2 in SPE when matched with any remaining worker at any wage. Thus, the outcome $(\mu^{**}, \gamma^{**}, \mathbf{w}^{**})$, characterized by

$$\mu^{**}(x) = \begin{cases} x & \text{if } x > \sqrt{3} \\ y^{neg}(x) & \text{if } 1 \leq x \leq \sqrt{3} \\ x & \text{if } x < 1, \end{cases}$$

[3]Note that this matching is measure-preserving and thus feasible.

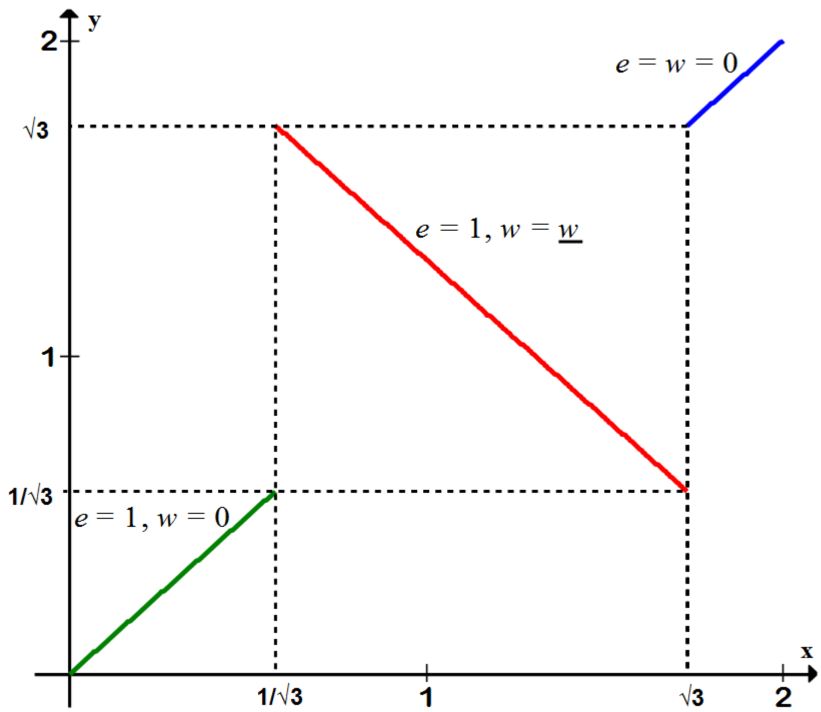

FIGURE 4.2. A stable, non-positively assortative outcome with wages

$$\mathbf{w}^{**}(x,x) = \begin{cases} 0 & \text{if } x > \sqrt{3} \\ \underline{w}(x, y^{neg}(x)) & \text{if } 1 \le x \le \sqrt{3} \\ 0 & \text{if } x < 1, \end{cases}$$

and

$$\gamma^{**}(x,x) = \begin{cases} (-1+x^2, 2+x^2) & \text{if } x > \sqrt{3} \\ (2, 2x(1+\sqrt{3}-x)) & \text{if } 1 \le x \le \sqrt{3} \\ (1+x^2, 1+x^2) & \text{if } x < 1, \end{cases}$$

is stable. Figure 4.2 illustrates that outcome.

In fact, *any* way of matching the firms and workers with types between 1 and $\sqrt{3}$ will yield a stable outcome. The firms cannot do better than the resulting payoff of 2. The reason for this multiplicity is that the payoff functions are additively separable

in effort e and match quality xy. If the firm pays the minimum effort-inducing wage $\underline{w}(x, y)$, then its net payoff is $2e$ – it is perfectly indifferent over any partner type $y \geq 1/x$. That indifference is fragile. We show next that if we relax the assumption of additive separability, then in the setting with wages the stability of positively assortative matching fails when effort and match quality are substitutes.

4.3. **Interaction between effort and match quality.** To introduce a small interaction between effort and match, suppose that now the stage-game payoff to a type-x firm from being matched with a type-y worker who exerts effort e is

$$U^F(e; x, y, w) = -1 + 2e + xy - \alpha exy - w,$$

where $\alpha > 0$ is a scalar. Thus, effort are worker quality are substitutes for the firm. The worker's payoff function is the same as before. Assume that $|\alpha|$ is close to 0, so that payoffs are still increasing and supermodular in types. As before, if $xy \geq 1$, then the smallest wage that will make the firm willing to fire is $\underline{w}(x, y) \equiv xy - 1$.

The payoff to a firm of type x paying $\underline{w}(x, y) + \Delta$, $\Delta \geq 0$, to a type-$y > 1/x$ worker for effort 1 is

$$U^F(e; x, y, \underline{w}(x, y)) = -1 + 2 + xy - \alpha xy - (xy - 1) - \Delta = 2 - \Delta - \alpha xy;$$

the payoff to the worker is

$$U^W(e; x, y, \underline{w}(x, y)) = 2 - 1 + xy + xy - 1 + \Delta = 2xy + \Delta.$$

Thus, conditional on effort 1, wage $\underline{w}(x, y) + \Delta$, and $y \geq 1/x$, the worker always prefers a higher type of firm, but the firm strictly prefers lower-type workers.

As a result, positively assortative matching is not stable for large N. The intuition is the following: suppose that there were a stable outcome in which each firm with type x above 1 and below $\sqrt{3}$, matched with the firm of the same type $y = x$, gets effort 1 and pays wage $\underline{w}(x, x) + \Delta$ for some Δ independent of x. (When N is large, competition between firms with very similar types ensures that that effort and wage pattern must hold.) But then a type-x worker and a type-$(x + \epsilon)$ firm form a blocking

pair with effort 1 and wage $\tilde{w} \equiv \underline{w}(x+\epsilon, x) + \Delta$:

$$\begin{aligned} U^F(e; x+\epsilon, x, \tilde{w}) &= 2 - \Delta - \alpha e x(x+\epsilon) \\ &> 2 - \Delta - \alpha e(x+\epsilon)^2 \\ &= U^F(e; x+\epsilon, x+\epsilon, \underline{w}(x+\epsilon, x+\epsilon) + \Delta) \end{aligned}$$

and

$$\begin{aligned} U^W(e; x+\epsilon, x, \tilde{w}) &= (2-\alpha)x(x+\epsilon) + \Delta \\ &> (2-\alpha e)x^2 + \Delta \\ &= U^W(e; x, x, \underline{w}(x,x) + \Delta). \end{aligned}$$

When effort and match quality are strict substitutes, then positively assortative matching is not stable if types are close together. The other outcome constructed above, $(\mu^{**}, \gamma^{**}, \mathbf{w}^{**})$, in which matching was positively assortative at the top and bottom and negatively assortative in the middle, is stable (once the cutoff types are adjusted to account for the introduction of α in the firm's payoff function). Relaxing the assumption that effort and match quality are additively separable in the stage-game payoff functions acts to reduce the multiplicity of stable matchings in the setting with wages.

5. General Results

In this section, we show how the results from the example generalize. First, consider the case without wages. A key step in the argument that positively assortative matching is not stable in the example is the following: if a firm of type x is matched with a worker of type x, and a type-$(x+\epsilon)$ firm with a worker of type $x+\epsilon$, and effort is 0 in both matches, then there is an effort level $e' > 0$ such that both the type-$(x+\epsilon)$ firm and the type-x worker benefit from matching with each other at effort e'. That is, at low effort levels firms are willing to trade quality for effort at a higher rate than are workers. Generalizing, we say that *firms care more about effort than workers do* if for any type $\theta \in \left[\underline{\theta}, \bar{\theta}\right]$ and any effort level $e \in [0,1]$,

$$\frac{U_e^F(0;\theta,\theta)}{U_y^F(0;\theta,\theta)} > -\frac{U_e^W(0;\theta,\theta)}{U_x^W(0;\theta,\theta)}. \tag{5.1}$$

If, all else equal, the marginal benefit of low effort is large enough or the marginal cost is small enough, then firms will care more about effort than workers do.

Let $(\mathcal{X}_N, \mathcal{Y}_N)_N$ be a sequence of sets of agent types such that the type distributions for both firms and workers converge weakly to the continuous uniform distribution over $\left[\underline{\theta}, \bar{\theta}\right]$.

Theorem 1. *Consider the setting without wages. If firms care more about effort than workers do, then for large enough N positively assortative matching is not stable.*

The argument for why positively assortative matching is unstable if firms care strictly more about effort is basically identical to the argument for the example in Section 3.1. Kaneko (1982) establishes the existence of a stable matching in this environment. That finding, together with Theorem 1, immediately implies the following result:

Corollary 1. *Consider the setting without wages. If firms care more about effort than workers do, then for large enough N there exists a stable matching that is not positively assortative.*

Thus, as long as firms care more about effort than workers do, the results from the example for the case without wages generalize. The same is true in the setting with wages. We say that *effort and match quality are substitutes* if for any types $x, y \in \left[\underline{\theta}, \bar{\theta}\right]$ and any effort level $e \in [0, 1]$,

$$U^F_{ey}(e; x, y) < 0. \tag{5.2}$$

The general result for the case where wages are endogenous is the following:

Theorem 2. *Consider the setting with wages. If effort and match quality are substitutes, then for large enough N positively assortative matching is not stable.*

Don't need firms care more - effort increasing total payoff is enough.

Kaneko's (1982) result implies Corollary 2.

Corollary 2. *Consider the setting with wages. If effort and match quality are substitutes, then for large enough N there exists a stable matching that is not positively assortative.*

6. Discussion

In this paper, we endogenize the value of a match between a firm and a worker as the outcome of a dynamic game: in each period the worker chooses his effort level and the firm decides whether or not to continue the relationship. We find that positively assortative matching may not be stable even though payoffs are increasing and supermodular in types. Allowing match partners to choose and commit to wages may restore positively assortative matching, but only if effort and worker ability are weak complements. In that case, we identify a new interpretation of efficiency wages: by paying workers above the level required to induce effort, firms increase their willingness to fire high-quality workers.

In the context of a parametrized example, we construct a stable outcome that features positively assortative matching at the top and bottom of the type distribution, and negatively assortative matching in the middle. Although the distributions of firms and workers are symmetric, the nonlinear negatively assortative matching in the middle generates endogenous imbalances. As a result, the matching is extremely discontinuous.

Our model specifies that the continuation value to an agent after a firing (0) is independent of the agent's type, but we note that our results do not require that condition. Our arguments go through as long as that outside option increases in type slower than the value of the stage-game payoff, as would be the case if the outside option is a fixed fraction of the stage-game payoff. That assumption would be satisfied if, for example, the outside option was the match value multiplied by (1 minus the expected discounted time to find a new partner). Endogenizing the post-match continuation value within a model of matching with search frictions is a subject for future research.

Appendix A. Proofs

A.1. **Proving Lemma 1.**

Proof. If $U^F(0; x, y) > 0$, then for any strategy of the worker, the firm's payoff from firing is strictly lower than its payoff from continuing. Thus, in any SPE the firm will continue after every history. The worker's best response is to always choose effort 0.

If $U^F(0; x, y) \leq 0$, then the strategy profile $\underline{\sigma}$ where after any history the firm chooses to fire and the worker chooses effort 0 is an SPE; $\underline{\sigma}$ yields payoffs $(0, 0)$. For any effort level $\hat{e} \geq \underline{e}(x, y)$, Condition 2.1 ensures that the threat of reversion to $\underline{\sigma}$ supports the outcome where the firm does not fire and the worker chooses effort level $\hat{e}$ in each period. Thus, $\text{co}\left\{\hat{V}(x, y), (0, 0)\right\} \subseteq E^\delta(x, y)$. Any feasible payoff outside $\text{co}\left\{\hat{V}(x, y), (0, 0)\right\}$ gives the firm less than its minmax payoff 0 and so cannot lie in $E^\delta(x, y)$. We conclude that $E^\delta(x, y) = \text{co}\left\{\hat{V}(x, y), (0, 0)\right\}$. $\square$

A.2. **Proving that $(\tilde{\mu}, \tilde{\gamma})$ is stable in the example.**

Proof. We will show that in the continuous limit, under $(\tilde{\mu}, \tilde{\gamma})$ there are no blocking pairs. It follows immediately that for any $(\mathcal{X}_N, \mathcal{Y}_N)$ such that $\tilde{\mu}(x_n) \in \mathcal{Y}_N$ for every $x_n \in \mathcal{X}_N$, $(\tilde{\mu}, \tilde{\gamma})$ is stable with respect to $(\mathcal{X}_N, \mathcal{Y}_N)$. If we choose a sequence $(\mathcal{X}_N)_N$ such that the distribution of firm types converges weakly to the continuous uniform distribution over $\left[\underline{\theta}, \bar{\theta}\right]$, and set $\mathcal{Y}_N = \tilde{\mu}(\mathcal{X}_N)$, then the result follows.

First, we note that for any matched pairs (x, y) and (x', y') with effort levels e and e', respectively, the firm of type x and the worker of type y' cannot form a blocking pair if

$$y' \cdot (x' - x) - \frac{1}{2} x \left[y' - y\right] \geq e' - e. \tag{A.1}$$

The reason is that they form a blocking pair only if there is an effort level $\hat{e} \in [0, 1]$ such that

$$-1 + 2\hat{e} + xy' > -1 + 2e + xy$$

and

$$2 - \hat{e} + xy' > 2 - e' + x'y'.$$

If Condition A.1 holds, then no such $\hat{e}$ exists.

Case 1. $x, y \in \left[\sqrt{3}, 1\right]$. Each firm with a type x in this region can get a higher payoff only by switching to a worker with a higher type, but each worker in this region also would only switch to a firm with a higher type. Thus, no firm or worker with a type in this region can be involved in any blocking pair, and they can be ignored for the rest of the analysis.

Case 2. $x \in [1, \sqrt{3}]$. By construction, each firm with a type x in this region is indifferent between its two possible matches, $1/x$ and $\hat{y}(x)$. The firm cannot form a blocking pair with any worker on the $\hat{y}$ branch (that is, $y \in [2 - \sqrt{3}, 1/\sqrt{3}]$): the firm would switch only to a higher y in that range (since it is getting effort 1 from the worker of type $\hat{y}(x)$), and for any $x' > x$ in that range straightforward algebra gives

$$\begin{aligned} \hat{y}(x')(x'-x) - \tfrac{1}{2}x\left[\hat{y}(x') - \hat{y}(x)\right] &= \tfrac{(x'-x)}{x'}\tfrac{1}{2}\left((x')^2 2 - \sqrt{3}x' + 3\right) \\ &\geq 0 \\ &= 1 - 1 = e(x') - e(x). \end{aligned}$$

Thus, Condition A.1 holds, so no such blocking pair exists.

The firm also cannot form a blocking pair with any worker whose type y is below $2 - \sqrt{3}$, again because the firm is already getting 1 from a worker of type $\hat{y}(x) \geq 2 - \sqrt{3}$. For $y \in \left(1, \sqrt{3}\right)$, then effort is not sustainable when the firm matches with a type-y worker, and simple algebra shows that the firm prefers effort 1 from a type-$\hat{y}(x)$ worker to effort 0 from a worker of type$\sqrt{3}$ (the best y in that range). Finally, for $y \in \left[1/\sqrt{3}, 1\right]$, similar reasoning means that we need only check for blocking pairs where the firm matches with a worker whose type is lower than $1/x$: matching with a higher type would lead to zero effort. For any $y < 1/x$ in this range, more algebra shows that Condition A.1 holds: there is no effort level $\hat{e}$ such that both the firm and the type-y worker prefer to match with each other at $\hat{e}$ over what they're getting currently.

Case 3. $x \in \left[1/\sqrt{3}, 1\right]$. Again, we need only check for blocking pairs where a firm with a type x in this range matches with a worker whose type y is lower than $1/x$ – effort is not sustainable when the firm is matches with a

worker whose type is above that level, and the firm prefers effort 1 from its currnet partner (type $1/x$) to effort 0 from a worker of type (the highest y potentially available). The firm would only prefer a worker with a lower y only if it could get higher effort, but it is already getting effort 1. Thus, firms with a type in this range are not part of any blocking pairs.

Case 4. $x \in \left[2-\sqrt{3}, 1/\sqrt{3}\right]$. A firm with a type x in this range gets a higher payoff from its current outcome to the payoff from getting effort 1 when matched with a worker of type $y < 2-\sqrt{3}$. Thus, it cannot form a blocking pair with such a worker. For each worker type $y \in \left[2-\sqrt{3}, \sqrt{3}\right]$, straightforward but somewhat tedious algebra establishes that Condition A.1 holds, so the firm cannot be part of a blocking pair.

Case 5. $x \in \left[0, 2-\sqrt{3}\right]$. A firm with a type x in this range is matched with a worker of the same type and gets effort 1, so it cannot form a blocking pair with a worker of type $y < x$. For each worker type $y \in \left[2-\sqrt{3}, \sqrt{3}\right]$, more algebra establishes that Condition A.1 holds, so the firm cannot be part of a blocking pair.

□

A.3. **Proving Theorem 1.**

Proof. Under the positively assortative matching μ^+, positive effort is sustainable in equilibrium when $U^F(0; x_n, y_n) \leq 0$. As N increases, the limit of μ^+ is the identity matching μ^I, where $\mu^I(x) = x$. Under μ^I, the condition for effort to be sustainable is $U^F(0; x, x) \leq 0$. Define $\hat{\theta} \in \left(\underline{\theta}, \bar{\theta}\right)$ as the type such that $U^F(0; \hat{\theta}, \hat{\theta}) = 0$.

Because $U^F(0; \hat{\theta}, \hat{\theta}) = 0$, for any sequence of strictly positive ϵ_k converging to 0, we can choose $x(k), y(k)$ such that $x(k) \in \left(\hat{\theta}, \hat{\theta} + \epsilon_k\right)$, $y(k) \in \left(\hat{\theta} - \epsilon_k, \hat{\theta}\right)$, and $U^F(0; x(k), y(k)) < 0$. We will show that under μ^I, in the continuous limit $x(k)$ and $y(k)$ form a blocking pair for large enough k. Then we will argue that for large N, there must be a similar blocking pair under μ^+.

To see that $x(k)$ and $y(k)$ are a blocking pair, note that since $U^F(0; x(k), y(k)) > 0$, the type-x_k firm's match partner will not exert effort in equilibrium, and so the firm's payoff will be $U^F(0; x(k), y(k))$. The firm would prefer to match with the type-$y(k)$

worker if that worker exerted effort e^* strictly above $\underline{e}(k)$, defined implicitly as

$$U^F(\underline{e}(k); x(k), y(k)) = U^F(0; x(k), x(k)).$$

The type-$y(k)$ worker's current payoff is at most $U^F(0; y(k), y(k))$, so he also prefers to match with the type-$x(k)$ firm and exert effort e^* as long as e^* is strictly below $\bar{e}(k)$, defined by

$$U^F(\bar{e}(k); w(k), y(k)) = U^F(0; y(k), y(k)).$$

Taking a first-order approximation, for large k, we get

$$\underline{e}(k) \approx (x(k) - y(k)) \frac{U_y^F(0; \hat{\theta}, \hat{\theta})}{U_e^F(0; \hat{\theta}, \hat{\theta})}$$

and

$$\bar{e}(k) \approx (x(k) - y(k)) \frac{U_x^W(0; \hat{\theta}, \hat{\theta})}{-U_e^W(0; \hat{\theta}, \hat{\theta})}.$$

Since $\hat{\theta} - y(k)$ shrinks to 0, $\underline{e}(k) < 1$ for large k. Taking the ratio, we get that for large enough k,

$$\frac{\bar{e}(k)}{\underline{e}(k)} \approx \frac{\frac{U_x^W(0; \hat{\theta}, \hat{\theta})}{-U_e^W(0; \hat{\theta}, \hat{\theta})}}{\frac{U_y^F(0; \hat{\theta}, \hat{\theta})}{U_e^F(0; \hat{\theta}, \hat{\theta})}}.$$

Condition 5.1 then implies that $\bar{e}(k) > \underline{e}(k)$ for large k. Because $U^F(0; x(k), y(k)) < 0$, any effort level $e^* \in [\underline{e}(k), \bar{e}(k)]$ is sustainable in equilibrium when the type-$x(k)$ firm and the type-$y(k)$ worker are matched. Thus, they form a blocking pair.

For a given N and k, let $x_{n^*(N,k)}$ be the closest type of firm in $\mathcal{X}_N$ to $x(k)$, and let $y_{n^{**}(N,k)}$ be the closest type of worker in $\mathcal{Y}_N$ to $y(k)$. As N increases, $x_{n^*(N,k)}$ and $y_{n^*(N,k)}$ converge to $x(k)$, and $y_{n^{**}(N,k)}$ and $x_{n^{**}(N,k)}$ converge to $y(k)$. Therefore, when k and N are large enough, under μ^+ $x_{n^*(N,k)}$ and $y_{n^{**}(N,k)}$ form a blocking pair, and positively assortative matching is not stable. $\square$

A.4. **Proving Theorem 2.**

Proof. For any match (x, y) such that $U^F(0; x, y) \geq 0$, define $\underline{w}(x, y) \equiv U^F(0; x, y)$ as the lowest wage that makes effort sustainable. Define $\hat{\Theta} \subseteq \left[\underline{\theta}, \bar{\theta}\right]$ as the set of firm types x such that

$$U^F(1; x, x) - U^F(0; x, x) > U^F(0; x, x).$$

When matched with a worker of the same type, a firm of type $x \in \hat{\Theta}$ prefers getting effort 1 and paying wage $\underline{w}(x,x)$ to zero effort and zero wage. Since $U^F(0;\underline{\theta},\underline{\theta}) < 0 < U^F(1;\underline{\theta},\underline{\theta})$, $\hat{\Theta}$ is nonempty and has nonempty interior.

First, we show that in the limit as N grows, in any stable outcome with the positively assortative matching μ^+, it must be that 1) a firm of type $x \in \hat{\Theta}$ gets effort 1, and 2) there exists a scalar $\Delta \geq 0$ such that each firm with a type $x \in \hat{\Theta}$ pays wage $\underline{w}(x,x) + \Delta$. To establish the first part of the claim, suppose that there is a firm of type $x_n^N \in \mathcal{X}_N$ such that the firm gets effort less than 1 and $x^N \in \text{int}\hat{\Theta}$. For large N, the adjacent type x_{n+1}^N also lies in $\text{int}\hat{\Theta}$. As N grows, x_n^N, x_{n+1}^N, and their matches y_n^N and y_{n+1}^N all become arbitrarily close to each other. Straightforward algebra then shows that either firm x_n^N and worker y_{n+1}^N form a blocking pair with effort 1, or firm x_{n+1}^N and worker y_n^N do. For example, suppose that effort is zero in both match $\left(x_n^N, y_n^N\right)$ and match$\left(x_{n+1}^N, y_{n+1}^N\right)$, and that wages are w and w', respectively. Then firm x_n^N and worker y_{n+1}^N form a blocking pair with effort 1 and wage $\underline{w}\left(x_n^N, y_{n+1}^N\right) + \hat{w}$ if there is a $\hat{w} \geq 0$ such that

$$U^F\left(1; x_n^N, y_{n+1}^N\right) - U^F\left(0; x_n^N, y_{n+1}^N\right) - \hat{w} > U^F\left(0; x_n^N, y_n^N\right) - w$$

and

$$U^W\left(1; x_n^N, y_{n+1}^N\right) + U^F\left(0; x_n^N, y_{n+1}^N\right) + \hat{w} > U^W\left(0; x_{n+1}^N, y_{n+1}^N\right) + w'.$$

As N increases, such a $\hat{w}$ exists unless

$$w' - w \geq U^F\left(1; x_n^N, x_n^N\right) + U^W\left(1; x_n^N, x_n^N\right) - \left[U^F\left(0; x_n^N, x_n^N\right) + U^W\left(0; x_n^N, x_n^N\right)\right];$$

the right-hand side is strictly positive because effort increases total payoff. Analogously, though, firm x_{n+1}^N and worker y_n^N form a blocking pair with effort 1 unless $w > w'$. Thus, there must be a blocking pair. The other cases are similar.

To establish the second part of the claim, observe that if firm x_n^N and firm x_{n+1}^N pay wages $\underline{w}\left(x_n^N, y_n^N\right) + \Delta$ and $\underline{w}\left(x_{n+1}^N, y_{n+1}^N\right) + \Delta'$, respectively, for effort 1, and $\Delta \neq \Delta'$, then there must be a blocking pair. Without loss of generality, suppose that $\Delta' > \Delta$. Then for large N, firm x_{n+1}^N can offer wage $(\Delta + \Delta')/2$ to worker y_n^N and form a blocking pair.

Thus, if there is a stable outcome with the positively assortative matching μ^+ in the limit as N grows, there is a $\Delta \geq 0$ such that each firm with a type $x \in \hat{\Theta}$ pays

wage $\underline{w}(x,x)+\Delta$ and gets effort. Such an outcome cannot be stable, however. In that case, the firm's payoff is

$$U^F(1;x,x)-U^F(0;x,x)-\Delta,$$

and the worker's is

$$U^W(1;x,x)+U^F(0;x,x)+\Delta.$$

But then a firm of type $x'>x$, $x'\in \mathrm{int}\hat{\Theta}$, forms a blocking pair with the worker of type x, with effort 1 and wage $\underline{w}(x,x)+\Delta$:

$$U^F(1;x',x)-U^F(0;x',x)-\Delta>U^F(1;x',x')-U^F(0;x',x')-\Delta$$

by Condition 5.2, and

$$U^W(1;x',x)+U^F(0;x',x)+\Delta>U^W(1;x,x)+U^F(0;x,x)+\Delta$$

because both U^F and U^W are strictly increasing in the firm's type. Thus, positively assortative matching is not stable for large enough N. □

References

[1] Ashlagi, I., Kanoria, Y., and Leshno, J. (Forthcoming.) "Unbalanced Random Matching Markets: the Stark Effect of Competition," *Journal of Political Economy.*

[2] Becker, G. (1973). "A Theory of Marriage: Part I," *Journal of Political Economy*, **81**, pp. 813-846.

[3] Citanna, A., and Chakraborty, A. (2005). "Occupational choice, incentives and wealth distribution," *Journal of Economic Theory*, **122**, pp. 206-224.

[4] Crawford, V. (1991). "Comparative statics in matching markets," *Journal of Economic Theory*, **54**, pp. 389-400.

[5] Kaya, A., and Vereshchagina, G. (2015). "Moral hazard and sorting in a market for partnerships," *Economic Theory*, **60**, pp. 73-122.

[6] Kaneko, M. (1982). "The Central Assignment Game and the Assignment Markets," *Journal of Mathematical Economics*, **10**, pp. 205–232.

[7] Legros, P., and Newman, A. (2007). "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, **75**, pp. 1073-1102.

[8] Roth, A., Rothblum, U., and Vande Vate, J. (1993) ."Stable Matchings, Optimal Assignments, and Linear Programming," *Mathematics of Operations Research,* **18**, pp. 803–828.

[9] Serfes, K. (2005). "Risk sharing and incentives: Contract design under two-sided heterogeneity," *Economics Letters*, **88**, pp. 343-349.

[10] Serfes, K. (2008). "Endogenous matching in a market with heterogeneous principals and agents," *International Journal of Game Theory*, **36**, pp. 587-619.

[11] Wright, D. (2004). "The Risk and Incentives Trade-off in the Presence of Heterogeneous Managers," *Journal of Economics*, **88**, pp. 209-223.