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

BENJAMIN SPERISEN AND THOMAS WISEMAN

EXTENDED ABSTRACT

We study a simple two-sided, one-to-one matching market with firms and workers. When a firm-worker pair is matched, they play an infinite-horizon discounted dynamic game. The range of feasible payoffs of the dynamic game is increasing in the players’ types, and their types are complementary — that is, maximal payoffs are a supermodular function of types. Classic results from the two-sided matching literature show that when types are complementary, then stable matchings are positively assortative: high-type workers match with high-type firms. In our setting, that result does not hold. There is positively assortative matching at the top and bottom ends of the market, but not in the middle. Intuitively, in this middle region increasing the quality of a match harms cooperative incentives. That effect dominates the direct positive effect of complementarity in types, so that higher-type firms prefer lower-type workers who will exert more effort. The key feature that distinguishes our model from the standard matching environment is that here payoffs from a match are determined endogenously rather than fixed as an exogenous function of types.

In the model, there are a continuum of firms and a continuum of workers. Each firm $i$’s type $X_i$ is drawn independently from a uniform distribution on $[0, 2]$, as is the type $Y_j$ of each worker $j$. (More precisely, we consider the limit of discrete approximations of those continuous uniform distributions.) 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 she
fires the worker, then the continuation payoff for each player is 0. If she continues, then the worker chooses an effort level $e \in [0, 1]$, which is publicly observed. The resulting instantaneous payoffs are

$$U_F(e; x, y) = 2e - 1 + xy$$

and

$$U_W(e; x, y) = 2 - e + xy.$$ 

That is, the worker’s payoff is decreasing in effort. The firm’s payoff is increasing in effort, as is the sum of payoffs. Holding effort fixed, both players’ payoffs are increasing in both types, and the marginal benefit of one player’s type is increasing in the type of the other player. Overall payoffs are the discounted sum (using the common discount factor $\delta \in (0, 1)$) of stage-game payoffs before the (possibly infinite) time when the worker is fired.

A key feature is that when the match quality is high enough (specifically, when $xy > 1$), then the firm will never choose to fire the worker. Even if she expects the worker to exert zero effort, the stage-game payoff of $xy - 1$ exceeds the zero payoff that she would get from firing. Thus, there is a unique subgame perfect equilibrium (SPE) in which the worker never works and the firm never fires him. If $xy \leq 1$, on the other hand, then any effort level below $e^*(x, y) \equiv (1 - xy)/2$ gives the firm a payoff below zero. In the limit as the players become patient, then, any effort level between $e^*(x, y)$ and 1 is enforceable in SPE.

To summarize, the limiting SPE payoff set as a function of $x$ and $y$, $E(x, y)$, is given by

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

where the first element is the firm’s payoff and the second the worker’s.

It is this discontinuity in the equilibrium payoff set that generates non-assortative matching. Although the feasible payoffs exhibit complementarity in types, equilibrium payoffs need not.
Outcomes. To investigate stable matchings, we first note that an outcome specifies both a matching and an equilibrium selection rule. A matching $\mu : [0, 2] \to [0, 2]$ is a measure-preserving correspondence that specifies for each type $x$ of firm the type $\mu(x)$ of worker that the firm matches with. An equilibrium selection rule $\gamma$ maps each pair of types $(x, y)$ to a SPE payoff in $E(x, y)$.

We say that an outcome is stable if there is no blocking pair. That is, for any types $x$ and $y$ of firm and worker such that $y \notin \mu(x)$, there is no payoff in $E(x, y)$ 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 her 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.

We claim that there is a stable outcome with the following properties, and (more tentatively) that it is the unique stable outcome:

- Firms with very high types prefer to be matched with a worker of the same type, even though in equilibrium there will be no effort exerted. The match quality effects dominates the effort effect. Thus, at the top there is positively assortative matching and no effort.
- For a firm with a middle type $x$, the type of her matched worker is less important (because of the complementarity in types), and so she prefers to be matched with the highest-type worker whom she is willing to fire (and thus can get effort from in equilibrium), $1/x$, rather than a firm of her own type $x$ (who would never exert effort in equilibrium). Thus, in the middle matching is negatively assortative and effort is exerted.
- Firms with very low types is willing to fire any type of worker, and so there is no effort effect. Thus, at the bottom there is positively assortative matching (as at the top) and positive effort (as in the middle).

A full description of the stable outcome is somewhat more complicated than the outline above, because the proposed matching in the middle ($\mu(x) = 1/x$) is not
measure preserving. To deal with that problem, the stable outcome involves discontinuous matching and adjustment of effort levels to make a firm indifferent between her match partner and the (very different) match partners of firms with nearby types.