In recent years, economists have become increasingly involved in the design of labor market clearinghouses. The design of a market-clearing mechanism affects not only the ex post allocation but also individuals’ ex ante choices, such as those regarding the amounts and types of human capital to acquire.

In this paper, we provide an illustration of how the design of labor market-clearing mechanisms can affect incentives for human capital acquisition. Specifically, we show that the worker-optimal stable matching mechanism (Gale and Shapley, 1962; Kelso and Crawford, 1982; Hatfield and Milgrom, 2005) incentivizes workers to make (nearly) efficient human capital investments.

Our work extends the labor market matching model of Kelso and Crawford (1982) to incorporate the possibility that agents may invest in human capital before matching. To analyze the investment incentives under the worker-optimal stable matching mechanism, we employ general results from our other work (Hatfield, Kojima, and Kominers, 2014) characterizing the mechanisms that are fully efficient, in the sense that they both are ex post efficient and incentivize efficient ex ante investment. En route to our main result, we show that so long as the space of salaries is sufficiently rich, every stable outcome in the Kelso and Crawford (1982) setting is approximately efficient.

I. Investment Efficiency and Strategy-Proofness

First, we survey the results of our work (Hatfield, Kojima, and Kominers, 2014) characterizing the ex post efficient mechanisms that induce individuals to make efficient ex ante investments.

A. Underlying Framework

There is a finite set of agents $I$ and a finite set of alternatives $\Omega$. Each agent $i \in I$ has a valuation function $v^i : \Omega \rightarrow \mathbb{R}$; the space of possible valuations for $i$ is denoted $V^i$. The space of all valuation profiles is denoted by $V \equiv \times_{i \in I} V^i$. As we describe in Section I.B, agents’ valuation functions are determined endogenously, through ex ante investment.

A transfer vector $t \in \mathbb{R}^I$ specifies transfer payments for each agent $i \in I$. The (ex post) utility of agent $i \in I$ given an alternative $\omega \in \Omega$ and transfer vector $t \in \mathbb{R}^I$ is

$$u^i((\omega, t); v^i) \equiv v^i(\omega) - t^i.$$  

We call an alternative-transfer pair $(\omega, t)$ an outcome.

An allocation rule $\mu : V \rightarrow \Omega$ is a map from the space of valuations to the set of alternatives. A transfer rule $s : V \rightarrow \mathbb{R}^I$ is a map from the space of valuations to the set of transfer vectors. A mech-
anism $\mathcal{M} \equiv (\mu, s)$ consists of an allocation rule $\mu$ and a transfer rule $s$; we denote $\mathcal{M}(v) \equiv (\mu(v), s(v))$. We focus on direct revelation mechanisms, i.e., mechanisms that take agent valuations as input.

We define the (ex post) social welfare of an alternative $\omega$ as the sum of the agents’ valuations for that alternative:

$$V(\omega; v) \equiv \sum_{i} v^i(\omega).$$

We abuse notation slightly by, for a mechanism $\mathcal{M} = (\mu, s)$, writing $v^i(\mathcal{M}(v)) \equiv v^i(\mu(v))$ and $V(\mathcal{M}(v); v) \equiv V(\mu(v); v)$.

B. Ex Ante Investment

Before participating in a mechanism, each agent makes an investment decision that determines his valuation over alternatives. We model the investment decision of agent $i$ as an explicit choice of the valuation function $v^i$, with the cost of investment determined by a cost function $c^i : V^i \to \mathbb{R}$. Each agent $i$ invests so as to maximize his ex ante utility,

$$r^i((\omega, t); v^i) \equiv u^i((\omega, t); v^i) - c^i(v^i).$$

We define the ex ante social welfare of an outcome–investment pair $((\omega, t), v)$ as

$$\sum_{i \in I} (v^i(\omega) - c^i(v^i)) = V(\omega; v) - \sum_{i \in I} c^i(v^i).$$

C. Characterizing Fully Efficient Mechanisms

We focus on (ex post) efficient mechanisms, i.e., mechanisms that choose an alternative that maximizes social welfare with respect to the submitted valuation profile.

DEFINITION 1: A mechanism $\mathcal{M}$ is efficient if, for all $v \in V$,

$$V(\mathcal{M}(v); v) = \max_{\omega \in \Omega} V(\omega; v).$$

We say that a mechanism is strategy-proof for $i \in I$ if reporting truthfully is a dominant strategy for $i$.

DEFINITION 2: A mechanism $\mathcal{M}$ is strategy-proof for $i \in I$ if, for all $v \in V$ and $\bar{v}^i \in V^i$,

$$u^i(\mathcal{M}(v); v^i) \geq u^i(\mathcal{M}(\bar{v}^i, v^{-i}); v^i).$$

Finally, we say that a mechanism induces efficient investment for $i \in I$ if, for any valuation profile $v^{-i}$ for agents other than $i$ (assuming those agents report truthfully), the choice of $v^i \in V^i$ that maximizes the ex ante utility of $i$ also maximizes ex ante social welfare.

DEFINITION 3: A mechanism $\mathcal{M}$ induces efficient investment by $i \in I$ if, for all $v^{-i} \in V^{-i}$,

$$\arg\max_{\bar{v}^i \in V^i} \{u^i(\mathcal{M}(\bar{v}^i, v^{-i}); \bar{v}^i) - c^i(\bar{v}^i)\} = \arg\max_{\bar{v}^i \in V^i} \{V(\mathcal{M}(\bar{v}^i, v^{-i}); (\bar{v}^i, v^{-i})) - c^i(\bar{v}^i)\},$$

for any cost function $c^i$.

In other work (Hatfield, Kojima, and Kominers, 2014), we characterize the class of ex post efficient mechanisms that induce efficient investment.

THEOREM 1: Suppose that mechanism $\mathcal{M}$ is efficient and that $V^i$ is path-connected. Then, $\mathcal{M}$ is strategy-proof for $i$ if and only if $\mathcal{M}$ induces efficient investment by $i$.

The “only if” implication of Theorem 1 generalizes earlier work by Rogerson (1992) showing that the Vickrey-Clarke-Groves mechanism (which is efficient and strategy-proof for all agents) induces agents to make efficient ex ante investments.2

Using Theorem 1, we can show the existence of ex ante welfare-maximizing equilibrium outcomes of the game induced by a particular mechanism $\mathcal{M}$. Given cost functions $\{c^i\}_{i \in I}$, let the investment game induced by $\mathcal{M}$ be the game in which each agent $i \in I$ simultaneously chooses a valuation $v^i \in V^i$ and receives payoff $r^i(\mathcal{M}(v); v^i)$.3

2Bergemann and Välimäki (2002) use a similar insight to show that agents efficiently acquire information prior to participating in Vickrey-Clarke-Groves mechanisms.

3We could extend the game to allow each agent $i$ to report a valuation $\theta^i$, with the payoff of $i$ now given by...
COROLLARY 1: Suppose that mechanism $M$ is efficient and strategy-proof. Then, there exists a Nash equilibrium $v$ of the investment game induced by $M$ such that $v$ maximizes ex ante social welfare.\(^4\)

**D. Approximating Full Efficiency**

We now introduce approximate analogues of the exact properties described in Section I.C.

Our definition of approximate efficiency is standard.

**DEFINITION 4:** A mechanism $M$ is efficient within $\epsilon$ if, for all $v \in V$,

$$ V(M(v); v) + \epsilon \geq \max_{\omega \in \Omega} \{ V(\omega; v) \}. $$

Our definition of approximate strategy-proofness is also standard: A mechanism is approximately strategy-proof for $i$ if reporting truthfully provides $i$ with a utility close to the highest utility he can attain from any report.

**DEFINITION 5:** A mechanism $M$ is strategy-proof within $\epsilon$ for $i \in I$ if, for all $v \in V$ and $\bar{v} \in V^i$,

$$ u^i(M(v); v^i) + \epsilon \geq u^i(M(\bar{v}^i; v^{-i}); v^i). $$

Finally, we say that a mechanism approximately induces efficient investment by $i$ if, for any valuation profile $v^{-i}$ for agents other than $i$ (assuming those agents report truthfully), the choice of $v^i \in V^i$ that maximizes the ex ante utility of $i$ approximately maximizes ex ante social welfare.

**DEFINITION 6:** A mechanism $M$ induces efficient investment within $\epsilon$ by $i \in I$ if, for all $v^{-i} \in V^{-i}$, if

$$ \hat{v} \in \arg \max_{v^i \in V^i} \{ u^i(M(\hat{v}^i; v^{-i}); \hat{v}^i) - c^i(\hat{v}^i) \}, $$

then

$$ V(M(\hat{v}^i; v^{-i}); (\hat{v}^i, v^{-i})) - c^i(\hat{v}^i) + \epsilon \geq \sup_{\bar{v} \in V^i} \{ V(M(\bar{v}^i; v^{-i}); (\bar{v}^i, v^{-i})) - c^i(\bar{v}^i) \} $$

for all cost functions $c^i$.

In our other work (Hatfield, Kojima, and Kominers, 2014), we show that the results of Theorem 1 generalize to relate approximate strategy-proofness to approximately efficient investment incentives.\(^5\)

**THEOREM 2:** Suppose that mechanism $M$ induces efficient investment within $\epsilon$ by $i$ and is efficient within $\eta$. Then, $M$ is strategy-proof within $(\epsilon + \eta)$ for $i$.

**THEOREM 3:** Suppose that mechanism $M$ is strategy-proof within $\epsilon$ for $i$ and efficient within $\eta$, and that $V^i$ is path-connected. Then, $M$ induces efficient investment within $|\Omega^i|(\epsilon + \eta)$ by $i$, where

$$ \Omega^i \equiv \{ \Psi \subseteq \Omega : \forall \psi \in \Psi, \forall \omega \in \Omega, \forall v^i \in V^i, v^i(\psi) = v^i(\omega) \Leftrightarrow \omega \in \Psi \} $$

is the set of equivalence classes of alternatives for $i$.

II. Labor Market Matching

In this section, we show that the worker-optimal stable mechanism induces approximately efficient investment by workers in the job matching model (with discrete transfers) of Kelso and Crawford (1982).\(^6\)

A. Model

There are finite sets $W$ and $F$ of workers and firms; together, these sets comprise the set of agents $I = W \cup F$. A pairing $(w, f)$ specifies that a worker $w$ is employed by firm $f$. The set of pairings is given by $W \times F$.\(^7\) In this setting, an alternative is a set of employment pairings $\omega \subseteq W \times F$ such

\(^4\)Unfortunately, as an example of Hatfield, Kojima, and Kominers (2014) shows, it is not the case that every Nash equilibrium induced by an efficient and strategy-proof mechanism $M$ maximizes ex ante social welfare.

\(^5\)Here, we assume that $\epsilon \geq 0$ and $\eta \geq 0$.


\(^7\)Our focus on worker–firm pairs is consistent with Kelso and Crawford (1982). Just as in the Kelso and
that each worker is paired with at most one firm, i.e., a matching.

Each worker \(w\) has a valuation \(v^w: F \to \mathbb{R}\) over firms. This naturally induces a valuation over alternatives which, slightly abusing notation, we also denote by \(v^w\):

\[
v^w(\omega) = \begin{cases} v^w(f), & (w,f) \in \omega \\ 0 & \text{otherwise.} \end{cases}
\]

Similarly, each firm \(f\) has a valuation \(v^f: \wp(W) \to \mathbb{R}\) over sets of workers, where \(\wp(W)\) denotes the set of subsets of \(W\). This naturally induces a valuation over alternatives which we also denote by \(v^f\), taking \(v^f(\omega) = v^f(W')\) with

\[
W' = \{ w \in W : (w,f) \in \omega \}.
\]

The valuations \(v^i\) induce utility functions as in (1).

We supplement the model of Kelso and Crawford (1982) with the investment framework of Section I.B by allowing each agent \(i\) to choose a valuation \(v^i \in V^i\) at cost \(c^i(v^i)\); agents maximize ex ante utility as in (2). Thinking of workers’ valuations as negative—working requires effort—we can understand workers’ investments as lowering the effort required by employment. For instance, if a job requires drawing up contracts, legal training greatly reduces the effort required—but comes at significant expense(!).

There is a finite set of possible salaries \(P \subseteq \mathbb{R}\); we assume that \(0 \in P\). A valuation function \(v^f\) for a firm \(f\) induces a choice function \(C^f\), given a salary vector \(\pi \in \Pi \equiv P^{W \times F}\):

\[
C^f(\pi; v^f) = \arg \max_{W' \subseteq W} \left\{ v^f(W') - \sum_{w \in W'} \pi^w \right\}.
\]

We say that firm \(f\) (with valuation function \(v^f\)) has a substitutable choice function if \(\pi^f \leq \tilde{\pi}^f\), \(\tilde{\pi}^w \leq \pi^w\), and \(w \in C^f(\pi; v^f)\) together imply that \(w \in C^f(\tilde{\pi}; v^f)\). That is, the firm’s choice function is substitutable if a firm never desires to fire a worker \(w\) when the wages of other workers rise (while the wage of \(w\) is unchanged). We assume throughout that for each firm \(f \in F\), every valuation \(v^f \in V^f\) gives rise to a substitutable choice function.

A transfer vector \(t \in \mathbb{R}^f\) is \(\omega\)-compatible if

1) for each worker \(w \in W\), we have\(^9\)

\[
-t^w \in P
\]

and, if there does not exist a firm \(f \in F\) such that \((w,f) \in \omega\), then \(t^w = 0\);

2) for each firm \(f \in F\),

\[
t^f = - \sum_{w \in W' \setminus \{f\} \in \omega} t^w.
\]

That is, a transfer vector is \(\omega\)-compatible if it corresponds to each worker receiving a salary that is in \(P\), with workers who are unemployed in \(\omega\) receiving a salary of 0, while each firm pays an amount equal to the sum of the salaries received by workers it employs in \(\omega\).

A matching mechanism \(\mathcal{M}\) maps a vector of valuations to an alternative \(\omega\) and a \(\omega\)-compatible transfer vector \(t\).

An outcome \((\omega, t)\) is stable if it is

1) individually rational: \(u^i((\omega, t); v^i) \geq 0\) for all \(i \in I\), and

2) unblocked: there does not exist a firm \(f \in F\) and a set of workers \(W' \subseteq W\) such that, for the alternative \(\hat{\omega} = W' \times \{f\}\) there is some \(\hat{\omega}\)-compatible transfer vector \(\hat{t}\) such that

\[
u^i((\hat{\omega}, \hat{t}); v^i) \geq u^i((\omega, t); v^i)
\]

for all \(i \in W' \cup \{f\}\), with at least one inequality holding strictly.

Kelso and Crawford (1982) show that if all firms’ preferences are substitutable (as

\(^9\)Note that as we have defined \(u^w((\omega, t); v^w) = v^w(\omega) - t^w\), the salary paid to worker \(w\) takes the form of a negative transfer.

Crawford (1982) model, it is possible to augment the pairings with a finite set of (non-pecuniary) contractual terms \(E\) so that the set of possible contractual relationships is a subset of \(W \times F \times E\). See, for example, the setting of Hatfield et al. (2013).

\(^8\)We assume for simplicity that the optimal set of workers is unique for every salary vector \(\pi \in \Pi\) and valuation \(v^f \in V^f\).
we have assumed), then there exists a stable outcome. Furthermore, Hatfield and Milgrom (2005) show in a more general model that there exists a worker-optimal stable outcome, i.e., a stable outcome that all workers weakly prefer to any other stable outcome. They also show that the worker-optimal stable mechanism, i.e., the matching mechanism which selects the worker-optimal stable outcome, is strategy-proof for workers.

B. Analysis

Now, we analyze workers’ investment incentives in the Kelso and Crawford (1982) framework.

While any stable outcome is Pareto efficient in the ordinal sense, it need not be efficient in the sense of the present paper—that is, it need not maximize total welfare—as the set of possible salaries is discrete. This implies that our exact equivalence result for efficient mechanisms (Theorem 1) cannot be applied to study investment incentives under the worker-optimal stable mechanism.\textsuperscript{10}

However, there is a sense in which the worker-optimal stable mechanism is approximately efficient. We assume that the salary increment is at most $\epsilon > 0$, that is, $\bar{p} - p \leq \epsilon$, for any $p, \bar{p} \in P$ such that $\bar{p}$ is the smallest salary that is greater than $p$. We also assume that the domain of the salaries is sufficiently large.\textsuperscript{11} These conditions on the set of salaries imply approximate efficiency of stable outcomes.

THEOREM 4: Suppose that the salary increment is at most $\epsilon$ and that the domain of the salaries is sufficiently large. Then, for any stable outcome $(\omega, t)$, the alternative $\omega$ is efficient within $|I|\epsilon$, i.e.,

$$V(\omega) + |I|\epsilon \geq \max_{\psi \in \Omega} \{V(\psi)\}.$$ 

\textsuperscript{10}Mechanisms that are strategy-proof but inefficient do not necessarily induce efficient investment (Hatfield, Kojima, and Kominers, 2014).

\textsuperscript{11}More specifically, we assume that in any individually rational outcome no worker can be paid either $p_{\text{min}} \equiv \min_{p \in P} \{p\}$ or $p_{\text{max}} \equiv \max_{p \in P} \{p\}$. A sufficient condition for this is that both $-p_{\text{min}}$ and $p_{\text{max}}$ are larger than $\sup_{v \in V, \omega \in \Omega} \{V(\omega; v)\}$; weaker sufficient conditions exist, but are more cumbersome to formalize.

PROOF:

Suppose that an outcome $(\omega, t)$ is stable. Assume for contradiction that $\omega$ is not efficient within $|I|\epsilon$. Then there exists an alternative $\psi$ such that $V(\psi) > V(\omega) + |I|\epsilon$.

Consider a transfer vector $\hat{t}$ defined as follows: For each $w \in W$, if $(w, f) \in \psi$ for some $f \in F$, let $-\hat{t}^w$ be the smallest salary in $P$ such that $u^w(\psi, \hat{t}); v^w) \geq u^w(\omega, t); v^w)$; otherwise, let $\hat{t}^w = 0$. For each $f \in F$, let

$$\hat{t}^f = - \sum_{w \in \{w' \in W : (w', f) \in \psi\}} \hat{t}^w.$$ 

Because the salary increment is at most $\epsilon$, we have $u^w((\psi, \hat{t}); v^w) \leq u^w((\omega, t); v^w) + \epsilon$ for any $w \in W$.\textsuperscript{12} Because $V(\psi) > V(\omega) + |I|\epsilon$, there must exist $f \in F$ such that $u^f((\psi, \hat{t}); v^f) > u^f((\omega, t); v^f)$. Hence, $(\omega, t)$ is not stable, as it is blocked: Consider the alternative $\hat{\psi} = W' \times \{f\}$, where $W' = \{w \in W : (w, f) \in \psi\}$, and the $\hat{\psi}$-compatible transfer $\hat{t}$ that is the restriction of $\hat{t}$ to $W' \cup \{f\}$, i.e., $\hat{t}^i = \hat{t}^i$ for all $i \in [W' \cup \{f\}]$ and $\hat{t}^0 = 0$ for all $i \notin [W' \cup \{f\}]$.

Theorems 3 and 4 together imply that the worker-optimal stable mechanism induces workers to make approximately efficient investments.

THEOREM 5: Suppose that the salary increment is at most $\epsilon$ and that the domain of the salaries is sufficiently large. Furthermore, suppose that $V^w$ is path-connected. Then the worker-optimal stable mechanism induces efficient investment within

$$(|F| + 1)|I|\epsilon \leq |I|^2\epsilon$$

by worker $w \in W$.

PROOF:

By Theorem 11 of Hatfield and Milgrom (2005), the worker-optimal stable mechanism is strategy-proof for $w$, and, by Theorem 4, it is efficient within $|I|\epsilon$. Thus, Theorem 3 implies that the worker-optimal stable mechanism induces efficient investment

\textsuperscript{12}Note that as $(\omega, t)$ is stable, it is individually rational. It follows that $u^w((\omega, t); v^w) + \epsilon \geq u^w((\omega, t); v^w) \geq 0$; hence, if $(w, f') \notin \psi$ for any $f' \in F$, then we have $u^w((\psi, \hat{t}); v^w) = 0 \leq u^w((\omega, t); v^w) + \epsilon$. 

within \(|\Omega^w|(|I|e)\) by \(w\), where \(\Omega^w\) is the set of equivalence classes of alternatives for \(w\).

As the valuation of a worker \(w\) depends only on the identity of the firm to which he is paired (if any), we have \(|\Omega^w| \leq |F| + 1\); combining this with our preceding observations shows the theorem.

III. Conclusion

In this paper, we show that the worker-optimal stable mechanism induces workers to make approximately efficient investments in human capital. Gul and Stacchetti (1999) show that the Kelso and Crawford (1982) framework can be used to model the allocation of indivisible objects. Our work here can be adapted to show that, when consumers have unit demand, the mechanism suggested by Gul and Stacchetti (1999, 2000) incentivizes consumers to invest efficiently in assets complementary to the objects allocated by the mechanism.

Meanwhile, it is well-known that the worker-optimal stable mechanism is not strategy-proof for firms. Consequently, Theorem 2 implies that firms are generally not incentivized to make efficient pre-labor market investments.\(^{13}\) Understanding the investment behavior of firms will likely require new techniques.

REFERENCES


\(^{13}\)However, Cole et al. (2001) do obtain efficient investment by firms in the case that investment is unidimensional and there is a continuum of firms.