*Matching Information**

Hector Chade† Jan Eeckhout‡

October, 2012

Abstract

We analyze the allocation of heterogeneous experts to teams of fixed size. Experts differ in the precision of their information about an unknown state and make a joint decision, aggregating their potentially correlated information. In this setting, we study the assortative matching properties (Becker (1973)) of teams.

The main insight is that it is optimal to diversify the composition of each team. This diversification is a generalization of negative assortative matching (NAM) to large teams, as it arises when the team payoff is submodular in their members's characteristics. When experts' signals are conditionally independent, diversification within teams is maximal, and the precision tends to be equalized across teams. We extend the model to introduce heterogeneity of firms in which the teams operate, and also analyze how to endogeneize the size of the teams.

Keywords. Assortative Matching, Teams, Diversification, Correlation.

*We have benefited from seminar audiences at UCL, Cambridge, UPF, Arizona State, Caltech, Brigham Young, Indiana–IUPUI, 2011 Latin American Econometric Society Meetings, 2012 Society for Economic Dynamics, 2012 European Summer Symposium in Economic Theory at Gerzensee, Sidney, Edinburgh, Adelaide, Louvain, Essex, London Business School, and Central European University, as well as from comments by Federico Echenique, Ilse Lindenlaub, Alejandro Manelli, Meg Meyer, and Edward Schlee. Eeckhout gratefully acknowledges support from the ERC grant 208068, and Chade is grateful to Universitat Pompeu Fabra for its hospitality in the winter of 2011 when part of this research was conducted.

†Arizona State University, `hector.chade@asu.edu`

‡University College London and Barcelona GSE-UPF, `j.eeckhout@ucl.ac.uk`.

1 Introduction

The aggregation of decentralized information is one of the fundamental sources of value creation within firms and organizations. Management heavily relies on the information and judgement of its employees. The team theory of Marshak and Radner (1972) that formalizes this insight has had a profound influence on our understanding of the role of information aggregation in organizations. In its basic formulation, a team consists of a group of agents with a common objective who take actions based on their information. In most economic settings, however, a team does not work in isolation but is embedded in a market or a larger organization with multiple teams that compete for their members. This makes the composition and information structure of teams endogenous. In this paper we analyze the allocation of differentially informed agents across teams and ask how competition for informed agents shapes the organization within and between teams.

To this end, we incorporate teams in a matching framework, and analyze the sorting properties that emerge, in the tradition of Becker (1973). Matching models shed light on how competition shapes the allocation of heterogeneous agents, such as partners in marriage, business, or firms and workers. But information as a sorting variable has not received much attention, even though it is important in many applications where experts match, such as R&D groups, teams of consultants, co-authorship, etc. Nor do we have many results on sorting properties when agents are matched in groups of size larger than two. While pairwise matching is relevant for marriage, to really understand sorting in the context of firms we need to study matching in larger teams.

The model consists of a group of agents or experts that must be partitioned into fixed-size teams. Within a team, each agent draws a signal about an unknown parameter before making a joint decision. Experts differ in the precision of their signals. Building on the standard paradigm in team theory, we assume normal distributions and quadratic payoff function. Unlike most of the literature, however, we allow for a general correlation structure. Conditional on the unknown state, experts' signals may be positively correlated, as is the case if they have access to common resources or received similar training; or they may be negatively correlated. We also assume that team members can transfer utility and observe each other's signal realizations before making a decision, allowing us to abstract from incentive issues and focus on the impact of information aggregation.

As in Marshak and Radner (1972), each team in our model uses information to make a decision in pursuance of a common goal. But unlike team theory, the composition of each team is determined endogenously, and so is the cost of its information structure, which is determined by the opportunity cost of assigning each member to another team. As in Becker (1973), the optimal matching maximizes the total payoff of all teams and can be decentralized. But unlike most of this literature, the sorting characteristic is an agent's information, and we do not assume pairwise matches. In this set up, we ask the standard assortative matching question: Who matches with whom? Will better experts match with similar ones to form superstar teams? Or will they be matched with less informed ones? The main robust insight is that optimal teams consist of diverse experts. That is, information as a sorting characteristic leads to *diversification* of expertise.

The following is a more detailed summary of our contributions. We first analyze the information aggregation problem of a team with a *given* composition of experts. Although Bayesian updating under normality in combination with a quadratic payoff function yield closed forms when there is only one signal, the presence of multiple signals that can be correlated considerably complicates the analysis. We nonetheless provide a closed form solution for the team value function that plays a crucial role in the analysis and is of independent interest. It depends on experts' types and correlation parameter through a function that can be formally interpreted as an index of informativeness of the team that summarizes the information contained in the experts' signals. Using this index, we show that negatively correlated signals are more informative (in Blackwell's sense) than conditionally independent ones. The opposite is true when correlation is positive but small. The intuition relies on the marginal value of adding a signal. While an expert's type adds to the team's precision, this gain needs to be adjusted by how correlated the signal is with those of the other team members.

With the team value function in place, we then analyze the optimal team composition, i.e., the matching stage. The main insight that emerges from the properties of the team value function is that it is typically optimal to diversify the composition of the teams. Optimal teams do not consist of members of the same types or, equivalent, positive assortative matching (PAM) is suboptimal in this setting. If a team structure were to exhibit PAM, there would always be profitable swaps of experts between any two teams that will improve efficiency. This is because for a wide range of correlation values, the team value function is submodular in experts' types.

When teams are of size two, diversification is simply NAM. In this case, it is easy to build the optimal team composition: pair the best expert with the worst, the second best with the second worst, and so on until depleting the pool of experts. But for larger teams, such a sweeping description of the exact composition of optimal teams is not possible when there is correlation in signals due to its complicated effects on a team's value of information. We can, however, fully pin down the optimal allocation when signals are conditionally independent. The remainder of the paper therefore focuses on this canonical case, also because it is the most commonly used assumption in models of information acquisition. Now the precision of a team is equal to the sum of the precision of the signals of its members, and this permits us to derive sharp results.

Using basic concepts of majorization theory, we show that the optimal matching problem amounts to maximizing a strictly Schur concave function of the precision of the teams subject to a total precision constraint. As a result, it is optimal to build teams that maximize the spread in precision *within* teams, and at the same time the precision tends to be equalized *across* teams. This calls for an extreme form of diversification of expertise: if possible, all teams should have the same composition. A by-product of this result is that the underlying combinatorial problem embedded in our model is akin to partition problems that are NP-hard (Garey and Johnson (1978), Vondrak (2007)).

We also explore the possibility of fractionally assigning experts across teams. In reality, team members often spend a fraction of their time on different tasks and thus are members of more than one team (e.g., Meyer (1994)). Similarly, researchers typically work on different projects in parallel with varying coauthors. Now the precision of all teams is equalized, and this can be accomplished by perfectly diversifying the experts across teams. That is, it is optimal to allocate each expert to every team in a uniform fashion (i.e., divide the expert's time equally among all teams). These sharp diversification results are a natural generalization of NAM to teams of any given size, for the team value function is strictly submodular in agents' types. Fractional assignment also affords a simple decentralization of the optimal matching as a Walrasian equilibrium, which boils down to comparing first order conditions.

We then extend the model by adding heterogeneous firms that match with teams of experts, and find that there are two dimensions to the optimal sorting pattern, mixing both PAM and diversification. More productive firms match with teams of higher precision, so there is PAM between firm quality and team precision. Yet within each team, there is diversification of expertise. Moreover, such diversification depends on how spread out firm productivity is: the higher the spread, the higher the difference between team precision across teams, and hence the lower the diversity of expertise within teams. These insights are important for economic applications, since they shed light on skilled work force composition in each firm and across firms. Although in reality firms differ in productivity and size, the distribution of skills within firms — whether they are large or small, highly productive or not — is typically diverse.

There is a lot of structure in the model that permits the derivation of all the results. At the end of the paper, we provide a thorough discussion of the main assumptions of the model and of potential extensions that seem interesting to pursue.

Related Literature. The paper is related to several strands of literature.

Assortative Matching. Since the focus of the paper is on assortative matching, the obvious point of departure is Becker (1973). The novel features are: the analysis of sorting based on an agent's signal informativeness, to be sure a characteristic that is relevant in many matching settings of economic interest; and the multi-agent nature of teams, which goes beyond the pairwise paradigm that is standard in matching models with transferable utility. Multi-agent matching and equilibrium existence is also analyzed in Kelso and Crawford (1982). Here we focus on the sorting patterns that emerge in the team setting. Our model also relates to the literature on matching and peer effects. In the presence of correlation, each agent's signal imposes an 'externality' on the group via its effect on aggregate precision. Pycia (2012) provides a comprehensive analysis of this topic, and gives conditions under which positive sorting emerges. Similarly, Damiano, Hao, and Suen (2012) analyze group formation and assortative matching with peer effects. Our paper differs in many ways from theirs, in particular in our focus on information as the sorting variable and the diversification pattern that emerges.

Teams. We also build on the large literature on teams started by the seminal contributions of Radner (1962) and Marshak and Radner (1972). Like that literature, we abstract from incentive problems. But instead of analyzing a team in isolation with an exogenously given information structure, we study teams in competition for the expertise (information) of its members. Two recent papers that are related are Prat (2002) and Olszewski and Vohra (2012). To the best of our knowledge, Prat (2002) is the first to analyze the optimal composition of a single team, and to provide conditions on the payoff function that makes a homogeneous or a diversified team optimal. In his setting, the cost of an information structure for the team is exogenously given, and he analyzes the team in isolation. Instead, we derive our team value function from first principles, and analyze optimal teams in a matching framework, where the cost of endowing a team with an information structure is the opportunity cost of matching the experts with another team. Olszewski and Vohra (2012) analyze the optimal selection of a team where members' productivities are interdependent in an additive way. They provide a polynomial time algorithm to construct the optimal set and comparative static results with respect to the cost of hiring and productivity externalities. And when agents in a team can form sub-teams under equal sharing of the output, they show that there is a unique stable coalition structure. Unlike their paper, we assume transferable utility and derive our match payoff function from the information aggregation of experts' signals, which does not fit their model. As a result, our sorting analysis is quite different, and due to its nature we cannot obtain a polynomial-time algorithm to select the optimal teams. Although less related, Meyer (1994) shows that fractional assignment can increase the efficiency of promotions, for it may enhance learning about ability of team members. In our model, it also increases efficiency as it equalizes the precision of teams.

Value of Information. A crucial step in our analysis of assortative matching is the derivation of the team value function, which summarizes the value of the information contained in the experts' signals. Since each team in our model runs a multivariate normal experiment, it is related to the literature on comparison of such experiments, e.g., Hansen and Torgersen (1974) and Shaked and Tong (1990). We provide a closed form solution that appears to be new for the index of informativeness of a normal experiment, and shed light on the effect of the correlation among signals on their informativeness. Also, we analyze the complementarity properties among signals, and this is related to Borgers, Hernando-Veciana, and Krahmer (forthcoming), who provide a characterization result for two (discrete) signals to be complements or substitutes. Unlike that paper, we study a normal framework and embed it in a matching setting.

Partition Problems. There is a significant discrete optimization literature on partition problems, nicely summarized in Hwang and Rothblum (2011). Ours is a partition problem, and in the conditionally independent case we actually solve a sum-partition problem (each team is indexed by the sum of its members' types). Most of the related results in the literature are derived for maximization of Schur convex objective functions, in which case one can find optimal consecutive partitions (i.e., constructed starting from the highest types downward). We instead deal with the maximization of a Schur concave objective function, and thus cannot appeal to off-the-shelf results. Moreover, we also shed light on many other properties of the solution.

The next section presents the model. Section 3 derives the team value function and the optimality of diversification. Section 4 focuses on conditionally independent signals, in which case diversification is extreme. Section 5 introduces heterogeneous firms. Section 6 concludes. The Appendix contains all the proofs.

2 The Model

There is a finite set $\mathcal{I} = \{1, 2, ..., m\}$ of agents, where $m = kN$. Each agent is assigned a 'level of expertise' (henceforth the agent's type) from a set $[\underline{x}, \overline{x}]$, with $\underline{x} > 0$ and $\overline{x} < \infty$, via the mapping $x : \mathcal{I} \to [\underline{x}, \overline{x}]$, where $x(i) \equiv x_i$, $i = 1, 2, ..., kN$.[1] Without loss of generality, we shall assume that $x_1 \leq x_2 \leq \cdots \leq x_{kN}$.

Each agent is assigned to one of N teams of fixed size k. Once in a team, each agent uses his or her expertise to cooperate in the solution of a problem that the team faces. More precisely, the team operates in an uncertain environment indexed by an unknown state s. For example, s could be the return of an asset, or the performance of a technology adopted, or a parameter of the demand for a product produced by the team, etc. The prior belief about the state s in each team is given by a density $h(\cdot)$ that is normal with mean μ and precision (inverse of the variance) τ, i.e., $\tilde{s} \sim \mathcal{N}(\mu, \tau^{-1})$.

Agents bring to the team their expertise in assessing the unknown state. Once in a team, an agent of type x_i draws a signal $\tilde{\sigma}_i$ drawn from $f(\cdot|s, x_i)$ that is normal with mean s and precision x_i, i.e., $\tilde{\sigma}_i \sim \mathcal{N}(s, x_i^{-1})$. In other words, an agent's type is the precision of her signal about s: better experts are those endowed with more precise signals, i.e., more informative in Blackwell's sense.[2] Conditional on the state s, we allow signals to be correlated across team members. For instance, it could be the case that agents use similar technologies to estimate the state, or they have acquired their training in similar places, etc. The pairwise correlation between any two agents x_i and x_j is given by $\rho(x_i x_j)^{-0.5}$, where $\rho \in (-(k-1)^{-1}, 1)$ (the lower bound $-1/(k-1)$ ensures that the covariance matrix of a team is positive definite). An important special case that we analyze in detail is that of $\rho = 0$, i.e., conditionally independent signals, arguably the most commonly made assumption in normal models of information acquisition.

There is no conflict of interest among team members. After observing the signal realizations of every member they jointly take an action $a \in \mathbb{R}$ to maximize the expected value of $\pi - (a - s)^2$, where $\pi > 0$ is an exogenous level of profit that is reduced by the error in estimating the unknown state.[3] We assume that $\pi > \tau^{-1}$: we shall see later

[1] Although convenient, strictly speaking it is an abuse to call x_i the type of agent i as there is no private information in the model. Notice also that agents i and i' can have the same type, i.e., $x_i = x_{i'}$.

[2] A signal is Blackwell-more-informative than another one if the second is a 'garbling' of the first. Formally, if $x_i > x_j$, then $\tilde{\sigma}_i$ is more informative than $\tilde{\sigma}_j$ since $\tilde{\sigma}_j = \tilde{\sigma}_i + \tilde{\varepsilon}$, where $\tilde{\varepsilon} \sim \mathcal{N}(0, x_j^{-1} - x_i^{-1})$ is independent of $\tilde{\sigma}_i$ (Lehmann (1988) p. 522, and Goel and Ginebra (2003) p. 519). Recent contributions with economic applications are Persico (2000), Jewitt (2007), and Quah and Strulovici (2009).

[3] The exogenous profit π could depend on the number of teams N. Since this adds notation without affecting the main results of the paper, we will assume throughout that it is a constant.

that this implies that the expected payoff of a team is always positive.

Formally, a group with types $\vec{x} = (x_1, x_2, ..., x_k)$ solves

$$V(\vec{x}) = \max_{a(\cdot)} \quad \pi - \int \cdots \int (a(\vec{\sigma}) - s)^2 f(\vec{\sigma}|s, \vec{x}, \rho) h(s) \prod_{i=1}^{k} d\sigma_i ds, \tag{1}$$

where $a : \mathbb{R}^k \to \mathbb{R}$ is a measurable function, $\vec{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_k)$, and (with some abuse of notation) $f(\vec{\sigma}|s, \vec{x}, \rho)$ is the joint density of the signals generated by the members of the group, which is multivariate normal with mean s and covariance matrix Σ_k, with diagonal elements $1/x_i$ and off-diagonal elements $\rho(x_i x_j)^{-0.5}$ for all $i \neq j$. Denote by $V(\vec{x})$ the maximum expected payoff of a team whose composition is $\vec{x}$. The resulting (ex-post) payoff is shared among team members via suitable transfers.[4] Agents have linear preferences over consumption, and as a result utility is fully transferable.

Before any information is generated, agents are sorted into N teams. Transferable utility implies that the optimal assignment maximizes the sum of the teams' profits. Hence, the optimal matching is the partition of the set of agents that maximizes $\sum_{n=1}^{N} V(\vec{x}_n)$. Formally, let $\Upsilon = \{x_1, x_2, ..., x_{kN}\}$ be the multiset of types (a multiset is a generalization of a set that allows for repetition of its members) and let $P(\Upsilon)$ be the set of all partitions of Υ into sub-multisets of size k. Then the optimal matching problem is

$$\max_{\mathcal{P} \in P(\Upsilon)} \sum_{S \in \mathcal{P}} V(\vec{x}_S). \tag{2}$$

Notice that in our optimal matching problem, which we cast in terms of a group of agents that is partitioned into teams (partnerships) of fixed size, we have been silent on who conducts the assignment. The model subsumes several possible interpretations. We can think of the problem as a many-to-one matching problem between experts and identical firms of fixed size. Alternatively, we could think of these teams as different divisions within the same firm. The assignment can be accomplished by a social planner, or by a decentralized competitive market, or by a CEO if all the teams belong to a single firm.

Also, the state of nature s can accommodate more than one interpretation. For instance, s could be the same for all groups; alternatively, there could be an independent draw of s in each group (e.g., different teams perform different tasks). Finally, all teams

[4]Notice that a team ex-post payoff can be negative and thus agents may have to share losses. This poses no problem as it is expected transfers what matters at the matching stage, and these can always be chosen to be nonnegative since $V(\cdot)$ is positive for all $\vec{x}$.

are of equal size k, and k is given. This is just a straightforward extension of the standard pairwise assumption made in most of the matching literature. We discuss how to endogeneize k in Section 4 and the relaxation of equal-group size in Section 6.

3 Correlation, Informativeness, and Diversification

There are two stages in the model. The first one is the *team formation* stage, where agents sort into N teams each of size k. The second stage is the *information aggregation* stage, in which team members pool their information and choose an action. We proceed backwards by first analyzing the main properties of the information aggregation stage, and then solving for the sorting pattern that emerges in the team formation stage.

3.1 The Team's Decision Problem and Value Function

Consider a team (one out of the N teams formed) with types $\vec{x} = (x_1, x_2, ..., x_k)$. After observing the realizations of the signals of all team members, i.e., $\vec{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_k)$, the team updates its beliefs about the state s. Since the prior belief is normal and signals are drawn from a multivariate normal distribution, the posterior distribution of s is also normal and denoted by $h(\cdot|\vec{\sigma}, \vec{x}, \rho)$. Then the team solves:

$$\max_{a \in \mathbb{R}} \quad \pi - \int (a - s)^2 h(s|\vec{\sigma}, \vec{x}, \rho)\, ds.$$

It follows from the first-order condition that the optimal decision function is

$$a^*(\vec{\sigma}) = \int s \; h(s|\vec{\sigma}, \vec{x}, \rho)\, ds = \mathbb{E}\left[\tilde{s}|\vec{\sigma}, \vec{x}, \rho\right].$$

Inserting $a^*(\vec{\sigma})$ into the team's objective function we obtain, after algebra,

$$\begin{aligned} V(\vec{x}) &= \pi - \int \cdots \int \left(\mathbb{E}\left[\tilde{s}|\vec{\sigma}, \vec{x}\right] - s\right)^2 f(\vec{\sigma}|s, \vec{x}, \rho) h(s) \prod_{i=1}^{k} d\sigma_i ds \\ &= \pi - \int \cdots \int \left(\int \left(s - \mathbb{E}\left[\tilde{s}|\vec{\sigma}, \vec{x}\right]\right)^2 h(s|\vec{\sigma}, \vec{x}, \rho) ds \right) f(\vec{\sigma}|\vec{x}, \rho) \prod_{i=1}^{k} d\sigma_i \\ &= \pi - \int \cdots \int \mathrm{Var}\left(\tilde{s}|\vec{\sigma}, \vec{x}, \rho\right) f(\vec{\sigma}|\vec{x}, \rho) \prod_{i=1}^{k} d\sigma_i, \end{aligned}$$

where $f(\vec{\sigma}|\vec{x},\rho) \equiv \int f(\vec{\sigma}|s,\vec{x},\rho)h(s)ds$. The second equality uses $h(s|\vec{\sigma},\vec{x},\rho)f(\vec{\sigma}|\vec{x},\rho) = h(s)f(\vec{\sigma}|s,\vec{x},\rho)$, and the third follows from replacing the expression for the variance of posterior density. The team value function thus depends on the information conveyed by the signals only through the variance of the posterior density of the unknown state.

It is easy to compute the value function in the conditionally independent case, i.e., when $\rho = 0$. For in this case after observing $\vec{\sigma}$ the posterior density is normal with mean $(\mu\tau + \sum_{i=1}^{k} \sigma_i x_i)/(\tau + \sum_{i=1}^{k} x_i)$ and variance $(\tau + \sum_{i=1}^{k} x_i)^{-1}$ (this follows from an adaptation of DeGroot (1970) p. 167). Since the posterior variance is independent of the signal realizations, it follows that

$$V(\vec{x}) = \pi - \left(\frac{1}{\tau + \sum_{i=1}^{k} x_i}\right). \tag{3}$$

More generally, for any ρ, we have the following result:

Proposition 1 (Team Value Function) *The value function of a team with types $\vec{x}$ is*

$$V(\vec{x}) = \pi - \left(\frac{1}{\tau + \mathcal{B}(\vec{x},\rho)}\right) \tag{4}$$

where

$$\mathcal{B}(\vec{x},\rho) = \frac{(1 + (k-2)\rho)\sum_{i=1}^{k} x_i - 2\rho \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (x_i x_j)^{0.5}}{(1-\rho)(1+(k-1)\rho)}. \tag{5}$$

The proofs of all the results are in the Appendix. Observe that the functional form is independent of $\vec{\sigma}$, as it depends on the variance of the posterior belief which is $(\tau + \mathcal{B}(\vec{x},\rho))^{-1}$. Since the proof of this result is somewhat involved, we sketch the argument here. We proceed by induction after obtaining the general functional form of the inverse of the covariance matrix. When the number of agents in each team is $k = 1$, $\tilde{s} \sim \mathcal{N}(\mu, 1/\tau^s)$, $\tilde{\sigma}_1 \sim \mathcal{N}(s, 1/x_1)$, and $\tilde{s}|_{\sigma_1} \sim \mathcal{N}((\mu\tau + \sigma_1 x_1)/(\tau + x_1), 1/(\tau + x_1))$, and (4) trivially holds. We then assume it is true for $k = n-1$ and find $\tilde{\sigma}_n|_{\sigma_1,\ldots,\sigma_{n-1},s}$ so as to obtain $\tilde{s}|_{\sigma_1,\ldots,\sigma_n}$. Finally, we check that the posterior variance has the above functional form, and thus (4) is true also for $k = n$. As illustrations of (4)–(5), notice that $\mathcal{B}(\vec{x},0) = \sum_{i=1}^{k} x_i$ and we obtain (3). Also, when $k = 2$, then $\mathcal{B}(\vec{x},\rho) = (x_1 + x_2 - 2\rho(x_1x_2)^{0.5})/(1-\rho^2)$. And if $x_1 = x_2 = ... = x_k = x$, then $\mathcal{B}(\vec{x},\rho) = kx/(1+\rho(k-1))$.

3.2 Correlation and Informativeness

The function $\mathcal{B}(\vec{x}, \rho)$ is the index of informativeness of the vector of signals $\vec{\sigma}$ drawn from a multivariate normal distribution centered at s and with covariance matrix Σ_k. The higher the value of $\mathcal{B}(\vec{x}, \rho)$ is, the more informative $\vec{\sigma}$ is in Blackwell's sense.[5] An interesting question to explore is how correlation affects the informativeness of a team. In particular, are correlated signals more informative than conditional independent ones? This issue has received some attention in the statistical literature (e.g., Shaked and Tong (1990)), and we provide a clear answer in our setting.

To motivate it, suppose $k = 2$ and $x_1 = x_2 = x$. Then $\mathcal{B}(\vec{x}, \rho) = 2x/(1 + \rho)$, and it is immediate that $2x/(1 + \rho) > 2x$ if and only if $\rho < 0$. That is, observing a vector of negatively (positively) correlated signals is more (less) informative than observing a vector of conditionally independent ones. Moreover, $\mathcal{B}(\vec{x}, \rho)$ is decreasing in ρ. The intuition underlying this example is as follows. Consider the extreme cases of perfect correlation: if $\rho = 1$, then observing the second signal is useless, while this is not the case under independence; and if $\rho = -1$, the second signal is infinitely informative, as it reveals the state. What if $\rho \in (-1, 1)$? Consider the conditional distribution of σ_2 given σ_1 and s, which, by standard bivariate normal distribution results (e.g., Bickel and Doksum (1977) p. 22) is distributed $\mathcal{N}\left((1 - \rho)s + \rho\sigma_1, \frac{1-\rho^2}{x}\right)$. Both positive and negative correlation reduce the variance of the second signal. But the effect on the mean differs depending on the sign of ρ: negative correlation makes it more 'sensitive' to s than positive correlation, thus making the former more informative than the latter. Moreover, information about s disappears from the second signal in the limit when $\rho = 1$.

The above example assumed that $k = 2$ and that the precision was the same for every member of the team. In the general case, we have the following result:

Proposition 2 (Correlation and Team Precision)
(i) If $\rho < 0$, then signals $\vec{\sigma}$ are more informative than if $\rho = 0$ ($\mathcal{B}(\vec{x}, \rho) > \mathcal{B}(\vec{x}, 0)\ \forall \vec{x}$);
(ii) There is a $\hat{\rho}$ such that if $0 < \rho < \hat{\rho}$, then signals $\vec{\sigma}$ are less informative than when $\rho = 0$ ($\mathcal{B}(\vec{x}, \rho) < \mathcal{B}(\vec{x}, 0)\ \forall \vec{x}$).

When $k = 2$, the proof of part (i) is immediate by inspection of (5), and in the general case it follows easily as well. Part (ii) is less intuitive in light of the example discussed above: in the general case, positively correlated signals are less informative than conditionally independent ones *only* for values of ρ below a threshold.

[5]This follows from Theorem 2 in Goel and Ginebra (2003) p. 521. See the Appendix A.2 for details.

Moreover, $\mathcal{B}$ is not monotonic in ρ. The explanation is that when precision is not the same for all the team's members, then positively correlated signals continue to be informative even in the limit, and the reduction in the variance provided by correlation outweighs the lower sensitivity of the mean with respect to s when ρ is large enough. But when it is low, the opposite happens and part (ii) ensues. The failure of monotonicity follows along the same lines. This can be grasped by allowing for $x_1 \neq x_2$ in the above example . Now the conditional distribution of the second signal is $\mathcal{N}\left((1-(x_1/x_2)^{0.5}\rho)s + \rho\sigma_1, \frac{1-\rho^2}{x_2}\right)$, where the mean continues to depend on s even when $\rho = 1$. Moreover, $\mathcal{B}(\vec{x},\rho)$ becomes nonmonotonic in this case, first decreasing in ρ and then increasing after a positive threshold value. These properties are illustrated in Figure 1, which depicts $\mathcal{B}(\vec{x},\rho)$ and $\mathcal{B}(\vec{x},0)$ for $k = 2$.

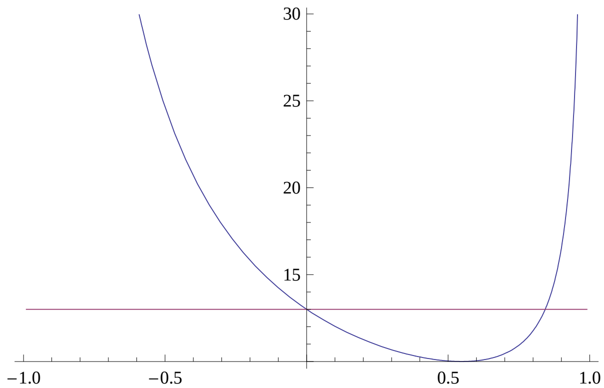

Figure 1: **Correlation and Informativeness.** For $k = 2$, the figure depicts both $\mathcal{B}(\vec{x},\rho) = (x_1 + x_2 - 2\rho(x_1x_2)^{0.5})/(1-\rho^2)$ and $\mathcal{B}(\vec{x},0) = x_1 + x_2$ as a function of ρ (with $x_1 = 10$ and $x_2 = 3$). Notice that $\mathcal{B}(\vec{x},\rho)$ is greater than $\mathcal{B}(\vec{x},0)$ when $\rho < 0$, less than $\mathcal{B}(\vec{x},0)$ when $0 < \rho < 0.84$, and greater beyond $\rho = 0.84$. Also, $\mathcal{B}(\vec{x},\rho)$ is not monotonic in ρ.

3.3 Diversification of Expertise across Teams

We now turn to the team formation stage of the problem. We will use the properties of the team value function (4) to shed light on the optimal sorting patterns. The following result is crucial for understanding the composition of the optimal teams that form.

Lemma 1 (Value Function Properties) *Consider a team with types $\vec{x}$:*
(i) There is a $\tilde{\rho}$ such that, if $-\tilde{\rho} < \rho < \tilde{\rho}$, then $V(\cdot)$ is strictly submodular in $\vec{x}$;
(ii) If $\rho > \tilde{\rho}$, then $V(\cdot)$ cannot be supermodular in $\vec{x}$, and it is strictly submodular if τ is sufficiently large;
(iii) If $\rho < -\tilde{\rho}$, then $V(\vec{x})$ cannot be supermodular in $\vec{x}$ unless τ is sufficiently large.

What Lemma 1 reveals is that the value function 'tends to be' submodular in types. Around zero correlation submodularity is strict. When the correlation parameter ρ is positive and large enough, the value function cannot be supermodular. Under negative correlation the extent to which supermodularity can arise is qualified. In particular, it depends on the exact values of both the prior precision τ about the state s and the precisions of the agents captured by $\vec{x}$. That is, for PAM to be optimal, the prior about the state has to be very precise to begin with, in which case the experts' information does not add much value to the team. Recall that only under supermodularity does an allocation exhibit PAM, i.e., there is specialization rather than diversification. The immediate implication of Lemma 1 is that the optimal matching exhibits *diversification* in all cases of interest. The next result now readily follows from this discussion.

Proposition 3 (Optimality of Diversification)
(i) Diversification within teams is always optimal for values of ρ in a neighborhood of 0.
(ii) Diversification is generically optimal when ρ is positive, and it is always optimal if τ is large enough.[6]
(iii) Diversification is optimal when ρ is negative, so long as τ is not too large.

Part (i), which subsumes the conditional independent case, implies that PAM — which assigns the best k types to one team, the second best k types to another team, and so on — can *never* be optimal. Given a partition that entails PAM, one can pick any two teams, swap the best expert in the better team with the worst in the other team and strictly increase $\sum_{n=1}^{N} V(\vec{x}_n)$ by strict submodularity (Lemma 1 (i)). That is, homogeneous teams can never be part of an optimal assignment. *Therefore, the optimal matching in this case always entails diversification of experts across teams.*

The logic underlying part (ii) is as follows. The proof of Lemma 1 (ii) shows that there is always a set $C \subseteq [\underline{x}, \overline{x}]^k$ around the 'diagonal' $x_1 = x_2 = ... = x_k$ where the team value function is strictly submodular when restricted to that set. If one starts from a PAM assignment, then one can do some profitable swaps if there are types that belong to the set C defined above. In other words, *diversification sometimes occurs in this case, and it does occur if the prior's precision is large enough.*[7]

For an 'intuition' of part (iii), notice that any partition of the agents into N teams generates a vector $(\mathcal{B}_1, ..., \mathcal{B}_N)$, $\mathcal{B}_i \equiv \mathcal{B}(\vec{x}^i, \rho)$ for all i, which has a 'mean' and a 'spread.' Differentiation of (5) reveals that $\mathcal{B}(\vec{x}, \rho)$ is supermodular in $\vec{x}$ when $\rho < 0$. Hence, PAM maximizes $\sum \mathcal{B}_n$. Obviously, this is not the same as maximizing $\sum V(\vec{x}_n)$, since V is strictly concave in $\mathcal{B}_n$. Now there is a force towards supermodularity in $\mathcal{B}_n$ and

[6] By 'generically' we mean that the optimality of diversification holds on a set of vectors of types (distributions of types) that has positive Lebesgue measure.

[7] The proof in the Appendix also reveals that $V(\cdot)$ is strictly submodular in $\vec{x}$ if $|\overline{x} - \underline{x}|$ is sufficiently small, i.e., diversification ensues if there is not much heterogeneity in agents' expertise.

one towards submodularity from the concavity of V. When τ grows large, the degree of concavity of the function $V = \pi - (1/(\tau + \mathcal{B}))$, i.e., $-V_{\mathcal{B}\mathcal{B}}/V_{\mathcal{B}}$, goes to zero.[8] Thus, when τ is large enough, the planner behaves as if maximizing $\sum \mathcal{B}_n$, thereby making PAM the optimal sorting pattern. In short, *when $\rho < -\tilde{\rho}$, diversification is optimal unless the prior precision is large enough.* How large should τ be for PAM? We show in Appendix A.5 that for the team value function to be supermodular τ must be be strictly larger than $8k\overline{x}$, i.e., more than *eight times* the precision of the *best* team possible.

Notice that we have focused on the optimality of diversification but have been careful not to assert that NAM is optimal. The reason is that it is unclear how to define it when $k > 2$. Also, we have not pinned down exactly what the composition of the optimal teams is. Clearly, we can say much more about optimal sorting when $k = 2$, for in this case $V(\cdot)$ submodular *implies* that NAM is optimal. That is, given any four types $x_1 > x_2 \geq x_3 > x_4$, the total payoff is maximized if x_1 is matched with x_4 and x_2 with x_3. The intuition is that x_4 can outbid x_2 and x_3 when competing for x_1. Moreover, the computation of the optimal matching is straightforward: match x_1 with x_{2N}, x_2 with x_{2N-1}, ..., and x_N with x_{N+1}. (Unlike NAM, there is no problem with the definition and computation of the optimal matching under PAM for *any* k, as when τ is large enough so that $V(\cdot)$ is supermodular (case (iii)).)

What can be said about the optimal composition of teams when $k > 2$? In order to address this question we will henceforth focus on the canonical case of *conditional independent signals*, where we can provide a complete answer.

4 Conditional Independence

Under conditional independence the team value function is $V(\vec{x}) = \pi - \left(\tau + \sum_{i=1}^{k} x_i\right)^{-1}$. By Lemma 1, it is strictly submodular in $\vec{x}$, and it is strictly concave in $\sum_{i=1}^{k} x_i$. Since $V(\cdot)$ depends on $\vec{x}$ only through its sum, for clarity we define $v(\sum_i x_i) \equiv V(\vec{x})$ and denote the precision of team n by $X_n = \sum_i x_i^n$. Any partition of the agents into N groups of size k generates a vector of team precisions $(X_1, X_2, ..., X_N)$. Hence, in the conditional independent case we seek a partition of the agents $x_1, x_2, ..., x_{kN}$ into N teams such that $\sum_{n=1}^{N} v(X_n)$ is maximum. Note that all partitions have the same sum $\sum_n X_n = X$. If X could be continuously divided, then the optimal matching problem would boil down to

[8]When $k = 2$, the same argument sheds light on the second part of (ii): when $\rho > 0$, $\mathcal{B}_n$ is submodular and thus if τ is large enough, then NAM is optimal since it maximizes $\sum \mathcal{B}_n$.

finding $(X_1, ..., X_N)$ to maximize $\sum_n v(X_n)$ subject to $\sum_n X_n = X$. This optimization problem is formally equivalent to a welfare maximization problem where a social planner allocates an 'aggregate endowment' X among N identical 'consumers,' equally weighted by the planner. And since the objective function is strictly concave, it follows that the optimal partition would be $X_n = X/N$ for all n, as optimality calls for the equalization of team precision.[9] The incentives to diversify lead to an extreme configuration of teams, all of them with equal precision. In a sense to be made precise below, expertise in each team is *maximally diversified*.

Clearly, the above argument is predicated on the assumption that X can be divided continuously, but a similar insight obtains more generally under independence. We next formalize it in two ways: first, we provide an analog of the above analysis for our discrete optimization problem; second, we explore the economically meaningful possibility of allocating each expert to multiple teams. Along the way, we shed light on the complexity of the problem, decentralization, and ways to endogeneize k.

4.1 Maximally Diversified Teams are Optimal

As mentioned, any partition of the experts into k-size teams generates a vector of team precisions $(X_1, X_2, ..., X_N)$. Consider the set Γ of such vectors generated by all partitions, and partially order Γ by majorization, which is a notion of how 'spread out' a vector is.[10] Then the objective function $\sum_n v(X_n)$ is strictly Schur concave on this partially ordered set, as it is the sum of single-variable strictly concave functions $v(\cdot)$.[11]

This property of the objective function is the key to both understanding the main characteristic of the optimal solution and providing a recipe for finding it. Notice that if we compare the team precision vectors of two partitions of the agents into k-sized teams and one majorizes the other, then the planner prefers the majorized one (by Schur concavity of the objective function). Continuing this way, each time a partition is 'improved' by decreasing the spread of its associated team precision vector in the sense of majorization, the objective function increases. Hence, the optimal solution to the

[9]From the first-order conditions, we obtain that $v'(X_n) = v'(X_m)$ for any $n \neq m$, i.e., the marginal benefit of team precision must be equalized across all teams.

[10]An N-vector X majorizes a N-vector X' if $\sum_{\ell=1}^{m} X_{[\ell]} \geq \sum_{\ell=1}^{m} X'_{[\ell]}$ for $m = 1, 2, ..., N$, with $\sum_{\ell=1}^{N} X_{[\ell]} = \sum_{\ell=1}^{N} X'_{[\ell]}$, and where $X_{[\ell]}$ is the ℓ-th largest coordinate of the vector X. For this and related concepts, see Marshall, Olkin, and Arnold (2009).

[11]$f : \mathbb{R}^N \to \mathbb{R}$ is Schur concave $f(X') \geq f(X)$ whenever X majorizes X', and it is strictly Schur concave if the inequality is strict.

matching problem must belong to the subset M of Γ consisting of all the vectors (generated by partitions) that are majorized by *all* the remaining ones in M^c (the complement of M in Γ). And if there exists a partition whose precision vector is majorized by *all* the precision vectors generated by the other feasible partitions, i.e., there exists a *minorizing vector*, then its associated partition is the *optimal* solution to the problem. In short, this suggests that *the optimal team structure is the one that minimizes the spread in the precision of the teams*, and this requires that teams be as diversified as possible. We call this property *maximal diversification* of expertise.

Proposition 4 (Maximal Diversification) *Let signals be conditionally independent. (i) The optimal matching must be an element of the set of partitions whose team precision vectors $(X_1, X_2, ..., X_N)$ are majorized by those generated by all the remaining partitions. (ii) If a team precision vector is majorized by the precision vectors of all the feasible partitions of the agents, then its associated partition is the optimal matching.*

Clearly, if there is a partition with $X_n = X/N$ for all n, then this is the optimal solution, for the vector $(X/N, ..., X/N)$ is a minorizing vector and Proposition 4 (ii) applies. Although minorizing vectors need not exist — and thus one needs to use Proposition 4 (i) instead —, two important cases where they do are $k = 2$ (teams of size two) and $N = 2$ (two-team partitions). The first case has already been discussed, and barring ties, the partition identified by the construction of the NAM partition is the optimal one. And if $N = 2$, then the precision vectors generated by partitions are *completely* ordered by majorization. This follows from $X_1 = X - X_2$ for all partitions, and thus any two precision vectors are comparable. Hence, a minorizing vector (X_1, X_2) exists and its associated partition of the set of agents into two teams solves the problem.

Another implication of Proposition 4 is that finding a polynomial-time algorithm seems a futile task except in the $k = 2$ case. For some perspective, notice that finding an optimum is tantamount to finding a partition where team precisions are equalized 'as much as possible.' When $N = 2$, this is reduces to so-called number partitioning problem with a constraint on the size of the two sets in the partition, which is well-known to be NP-hard (e.g., Garey and Johnson (1978), Mertens (2006)).[12] When $N = 3$, the task involved is as in the 3-partition problem, which has been shown to be strong NP-complete (Garey and Johnson (1978) p. 504), and so on for $N > 3$.

[12] Garey and Johnson (1978) p. 499 provides an optimal dynamic programming algorithm.

Thus far we have focused on the optimal matching problem of partitioning the set of agents into size-k teams, i.e., the planner's problem. A natural question is whether this can be decentralized. In the next section on fractional assignment we show that this is indeed the case. But we do not have an analogous result when each agent is assigned to only one group. In fact, we know that our set up is subsumed by the general framework in Kelso and Crawford (1982) – where firms hire (or partnerships consist of) groups of heterogeneous agents and the match payoff depends on their composition. If our model satisfied their sufficient conditions, then we could apply their proof of decentralization of the optimal allocation. We show in Appendix A.7, however, that our model fails their crucial gross substitutes condition (GS), which plays a fundamental role in their analysis, and thus we cannot appeal to their results.[13] We conjecture that a suitable modification of their analysis would yield a positive decentralization result even without gross substitutes. But since tackling this nontrivial task will lead us too far astray from our main results, we leave it as an interesting open problem.

4.2 Fractional Assignment and Perfect Diversification

Many assignment problems involve fractional time dedication. For instance, management consultants at McKinsey or partners in a law firm dedicate time to different projects that run in parallel. Similarly, researchers collaborate on different projects with different co-authors at the same time. In these examples experts are effectively working simultaneously in several teams. Applying this insight to our setting, we now assume that agents can be fractionally assigned to teams. An intuitive way to think about it is that each agent has a time budget and she can allocate it across teams.[14] Besides being of independent interest, the tractability of the solution under this assumption will be helpful below when we explore some variations of the model.

Clearly, to make this operational, we need an assumption about an expert's contribution to a team when working fractionally. We assume that the precision of the signal

[13] In words, GS asserts that if the wages of experts of different types weakly increase, then a firm or partnership will still find it optimal to hire those experts, whom they made offers to at the previous wages, whose wages did not change. Most of the general equilibrium with indivisibilities literature relies on this property. Moreover, if GS held, a by product would be that the planner's objective function would satisfy it as well, and a greedy algorithm would then find the optimal groups (Murota (2003) chapter 11, section 3). But we know that is not true in our setting.

[14] This assumption is not without precedent in the literature on assignment games, as non-integer assignment of agents is usually permitted in its formulation (although not used in equilibrium). Other combinatorial optimization problems (e.g., knapsack) also explore versions with fractional solutions.

the expert contributes to a team is *proportional* to her time dedication to that team. For example, if an agent works part-time for two teams, the signal observed in each team will have half the precision it would have if she worked full-time for one of them. To avoid information spillovers across teams, we assume there is an independent draw of the state of nature in each team (e.g., each team works on a different task).

It will be convenient to slightly change (and abuse) the notation: let $x(\mathcal{I}) = \{x_1, ..., x_J\}$ be the set of *distinct* types of agents, and denote by m_j the number of agents of type x_j, so that $\sum_{j=1}^{J} m_j = kN$ and $X = \sum_{j=1}^{J} m_j x_j$. Denote by $\mu_{jn} \geq 0$ the fractional assignment of type-j agents to team n. Feasibility requires that $\sum_{j=1}^{J} \mu_{jn} = k$ for every n (i.e., the sum of the allocations of types to team n must equal the fixed team-size k), and $\sum_{n=1}^{N} \mu_{jn} = m_j$ for every j (i.e., the fraction of type-j agents allocated to all the teams must equal m_j, the total number of them in the population).

The fractional assignment problem is:

$$\max_{\{\mu_{jn}\}_{j,n}} \sum_{n=1}^{N} v\left(\sum_{j=1}^{J} \mu_{jn} x_j\right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \mu_{jn} = m_j \ \forall j, \quad \sum_{j=1}^{J} \mu_{jn} = k \ \forall n, \quad \mu_{jn} \geq 0 \ \forall j, n. \tag{6}$$

Since fractional assignment allows for continuous division of X across teams, we have:

Proposition 5 (Perfect Diversification) *Any optimal matching entails the equalization of team precision across teams, i.e., $X_n = X/N$ for all n. It can be implemented by allocating an equal fraction of each expert's type to each team, i.e., $\mu_{jn} = \frac{m_j}{N}$ for all j, n.*

That is, under fractional assignment teams all have the same precision, and the unique symmetric solution that implements it distributes the total number of agents of a given type equally among all teams. We call this assignment *perfect diversification* of expertise across teams.[15] Once again, one can interpret the extreme diversification that ensues as a generalization of NAM to our setting when k is bigger than two.

Allowing for fractional assignment makes the *decentralization* of the optimal matching straightforward. To see this, let there be competitive prices $(w_1, w_2, ..., w_J)$ for

[15]The solution in terms of team precision $X_n = X/N$ is unique. Yet, unless there are only two types, the exact allocation of agents m_j/N is not the unique solution. For an example, let $k = 2$, $N = 6$, and $x(\mathcal{I}) = \{1, 2, 3, 4, 5\}$, with one agent of each type 1 through 4 and two with type 5. Thus, $X = 20$. Both $\{\{1, 4, 5\}, \{2, 3, 5\}\}$ and $m_j/2$ yield $X_1 = X_2 = 10$. Multiplicity derives from the linearity of the constraints and the fact that the precision of different experts in a team are perfect substitutes.

different types of experts. Then each firm n solves the following problem:

$$\begin{aligned} \max_{\{\mu_{jn}\}_{j,n}} \quad & v\left(\sum_j \mu_{jn} x_j\right) - \sum_j \mu_{jn} w_j \\ \text{s.t.} \quad & \sum_j \mu_{jn} = k, \quad \mu_{jn} \geq 0 \ \ \forall j. \end{aligned}$$

There are J first-order conditions for each of the N firms:

$$v'\left(\sum_j \mu_{jn} x_j\right) x_i - w_i + \phi_n = 0, \forall i, n.$$

where ϕ_n is the Langrange multiplier for firm n. The set of first-order conditions and constraints is the same as the planner's once w_i is substituted by the Langrange multiplier from the planner's problem for this type. It readily follows that the decentralized equilibrium allocation coincides with the planner's solution.

The tractability of fractional assignment also allows us to explore ways to *endogeneize the size of the teams*, which was thus far assumed fixed and equal to k. One alternative is to assume that there is a cost of forming teams, $c(\cdot)$, that is strictly increasing and convex in N. Since the optimal matching under fractional assignment equalizes the precision of the teams, i.e., $X_n = X/N$ for all n, the planner chooses N to maximize $v(X/N)N - c(N)$. Treating N as a continuous variable, the optimal N thus solves

$$\pi - c'(N) = N(N\tau + X)^{-1} + NX(N\tau + X)^{-2}. \tag{7}$$

Another way to endogeneize N is to envision a market with free entry of identical firms that hire groups of experts. By incurring an entry cost $F > 0$, identical entrepreneurs simultaneously enter the market. They hire workers of type x_j at a competitive wage rate w_j. Denote by N the number of firms that decide to enter the market. To avoid the trivial cases where either no firm or an infinite number of them enter the market, we assume that the entry cost satisfies $\pi - (1/\tau) < F < \pi$.

Each firm chooses the employment configuration μ_{jn} that maximizes its profits $v\left(\sum_j \mu_{jn} x_j\right) - \sum_j \mu_{jn} w_j$. From the first-order conditions, we obtain that $X_n = \sum_j \mu_{jn} x_j$ is equalized across all firms, i.e., $X_n = X/N$. Thus, $\mu_{jn} = m_j/N$ is the unique symmetric allocation of experts to firms.[16] Inserting the solution into the zero profit constraint

[16]Even if there are multiple solutions, $\sum_j \mu_{jn} w_j$ is constant since there is uniqueness in payoffs. By

$v\left(\sum_j \mu_{jn} x_j\right) - \sum_j \mu_{jn} w_j - F = 0$ yields the following equilibrium condition for N:

$$\pi - F = \frac{N}{N\tau + X} + \frac{NX}{(N\tau + X)^2}. \tag{8}$$

Either way of endogeneizing N yields the following insights about the equilibrium firm/group size (see Appendix A.9). First, there is a *unique* $N^\star$ that solves (7) or (8).[17] As the total number of experts m is fixed, this pins down a *unique* team-size $k^\star = m/N^\star$.

Second, a higher entry cost F *increases* $k^\star$. Intuitively, with $F > 0$, firms will compete away any surplus from entry into the market. A higher entry cost implies that firms require higher post-entry profits. Thus, fewer firms enter the market thereby increasing the size of the teams in each firm. (Since firm size is larger, this lowers the marginal product of experts, which leads to lower wages and higher post-entry profits.)

Finally, $k^\star$ *decreases* in both the precision of the prior τ and in the aggregate precision of the signals X. Prior precision increases the marginal product of additional workers at a given firm size (or the marginal benefit of adding teams in equation (7)). As a result, firms make higher profits, generating new entry, and thus lower firm size. The same logic applies to an increase in the aggregate precision X.

5 Heterogeneous Firms, PAM, and Diversification

One interpretation of the model is that of a market where identical firms compete to form a team of experts. In reality, those firms are likely to be heterogeneous as well and solve problems of varying economic impact. For example, consulting firms that differ in their market reputation consult for clients (firms) that are also heterogeneous in their attributes, such as size or market value. The value of expertise at each of those firms will differ, and as a result so will the demand for informed agents and the optimal/equilibrium matching pattern.

To avoid algebraic detours, we will assume conditional independent signals in this section. Although the analytical arguments heavily rely on this assumption, the main insights are more general. There are N heterogeneous firms. Let y_i be the type of firm i, e.g., the firm's productivity, say due to its capital, or technology, and assume that $y_1 \leq y_2 \leq \cdots \leq y_N$. As before, each firm matches with teams of size k. If a firm with

Proposition 5, X_n is constant across different allocations, and so is the total wage bill.

[17] Strictly speaking, the number of groups/firms will be the integer part of $N^\star$.

type y matches with a team with precision $\sum_{i=1}^{k} x_i$, then the expected payoff from the match is $y \cdot v\left(\sum_{i=1}^{k} x_i\right)$. The optimal matching problem is to partition the experts into k teams and assign them to the firms to maximize $\sum_{n=1}^{N} y_n v\left(\sum_{i=1}^{k} x_{in}\right)$.

Since the match payoff is supermodular in $\left(y, \sum_{i=1}^{k} x_i\right)$, it follows that it is optimal to match better firms with better teams (measured by their precision). That is, optimal sorting will exhibit more precise teams working for higher quality firms. Moreover, since match payoff is submodular in experts' types, there will still be diversification of expertise across teams. A natural conjecture is that groups will tend to be more homogeneous within firms as firm heterogeneity becomes more 'spread out.'

We can derive these insights in an elementary fashion by allowing for fractional assignment. The optimal matching of firm quality and team precision solves:

$$\max_{\{X_n\}_{n=1}^{N}} \sum_{n=1}^{N} y_n v\left(X_n\right)$$
$$\text{s.t. } \sum_{n} X_n = X.$$

This optimization problem is formally equivalent to a welfare maximization problem where a social planner allocates an 'aggregate endowment' X among N identical 'consumers,' with each consumer n 'weighted' by y_n.

The first-order conditions for this problem are $y_n v'\left(X_n\right) = y_{n'} v'\left(X_{n'}\right)$ for all $n \neq n'$. The unique optimal solution is, for all n, given by

$$X_n = \frac{y_n^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}} \left(\tau N + X\right) - \tau. \tag{9}$$

Notice that unless $y_n = y$ for all n, in which case we obtain $X_n = X/N$, the optimal solution is increasing in n, i.e., *higher-type firms match with higher precision teams.* This insight may be instructive for understanding differences of skilled workforce across firms. We observe a great diversity of skills within the firm even if the demand for skills across firms differs. This is consistent with this model, which predicts that better firms will on average have a more skilled workforce, yet still with diversification within them.

A straightforward comparative static result emerging from (9) is that an increase in aggregate precision X *increases* the precision of all teams. Tracing a welfare maximization parallel, if the 'aggregate endowment' increases, each consumer gets a higher level of 'consumption' X_n (i.e., team precision is a 'normal good' for the planner). Moreover, the *difference* between team precision at consecutive firms, $X_n - X_{n-1}$, increases.

A bit of work reveals that an increase in prior precision favors teams at better firms, that is, X_n *increases* for high n and *decreases* otherwise. In the planner's analogy, an increase in τ affects the planner's marginal rate of substitution between X_i and X_j for any $i \neq j$ in the direction of the consumer with the larger weight between the two. The *difference* between team precision at consecutive firms, $X_n - X_{n-1}$, also increases.

Less apparent from (9) is that if the y_n's become more 'spread out' (in a precise sense related to majorization), then this also favors teams at *better* firms. Indeed, we show that when the spread of firm quality *increases*, then team precision *increases* for better ranked firms and *decreases* for lesser ranked ones. The intuition is similar to the aforementioned changes in the planner's marginal rate of substitution. This result is important as it suggests that *in markets where firm heterogeneity is more spread out, we should also observe a more spread out skilled workforce across firms.*

Given the optimal team precision (9), we prove that there exists a fractional assignment rule $\{\mu_{jn}\}_{j,n}$ that implements it.[18] In some cases as in the example below, this can take the form of a suitable deviation from perfect diversification.

We summarize the insights of this section in the following proposition:

Proposition 6 (Heterogeneous Firms) .
(i) The optimal matching entails PAM between firm quality y_n and team precision X_n, given by equation (9);
(ii) An increase in aggregate precision X increases the precision of every team, X_n, and also $X_n - X_{n-1}$ for all n;
(iii) An increase in prior precision τ increases (decreases) the precision of teams whose index is above (below) a threshold $1 \leq n^* \leq N$, and increases $X_n - X_{n-1}$ for all n;
(iv) An increase in the spread of $(y_1, y_2, ..., y_N)$ increases (decreases) the precision of teams whose index is above (below) a threshold $1 \leq \hat{n} \leq N$;
(v) There exists a fractional assignment rule $\{\mu_{jn}\}_{j,n}$ that implements the unique solution to the optimal team precision given by (9).

As an illustration, consider the following example with two firms and four experts who match in pairs with the firms. Formally, $N = 2$, $y_1 = 1$, $y_2 = y \geq 1$, $\tau = 0$ (for simplicity), and there are two agents with type $x_1 = 5$ and two with type $x_2 = 20$.

[18] With identical firms, we showed existence by checking that perfect diversification solved the problem. This does not work with heterogeneous firms, and we use Farkas' Lemma to prove existence.

Using the above formulas, we obtain $X_2 = (y^{0.5}/(y^{0.5}+1))50$ and $X_1 = (1/(y^{0.5}+1))50$. That is, the better firm, namely firm 2, matches with a higher precision team than firm 1 does. Regarding the optimal matching pattern, it is easy to show that it spans the whole spectrum as y increases, ranging from NAM to PAM with different degrees of diversification in between. More precisely, the optimal matching satisfies the following properties: (a) if $y = 1$ then $X_2 = X_1$ and NAM is optimal, i.e. $\{5, 20\}, \{5, 20\}$; (b) if $y \geq 16$ then $X_2 = 40$, $X_1 = 10$, and PAM is optimal, i.e., $\{5, 5\}, \{20, 20\}$; (c) if $1 \leq y < 16$ there is some diversification within groups, which can be accomplished by the fractional assignment rule $(\mu_{12}, \mu_{22}) = (1 - (1/15)(X_2 - 25), 1 + (1/15)(X_2 - 25))$ for team 2 and $(\mu_{11}, \mu_{21}) = (1 - (1/15)(X_1 - 25), 1 + (1/15)(X_1 - 25))$ for team 1.

6 Discussion and Concluding Remarks

Many important economic applications entail the formation of teams composed by members of varying expertise. We have analyzed such a matching problem in a highly structured model of information, where the notion of a team, the expertise of its members, and the aggregate informativeness of their signals are easy to interpret. In this setting, we have derived several insights regarding the sorting patterns that emerge given the properties of the match payoff function of a team. In particular, we have shown that in most cases of interest the optimal formation of teams require diversification of expertise; i.e., it is optimal to have experts of different skills within each teams. Under conditional independence, this diversification takes an extreme form, essentially equating the precision of all teams. We have also explored the role of correlation and how it affects the informativeness of a team, as well as the decentralization of the optimal team configuration and endogenous team size. Finally, we analyzed the implications of adding another heterogeneous side of the market, namely, firms that differ in their quality, and showed that the optimal sorting pattern entails a peculiar combination of PAM between firm quality and team precision and diversification within teams.

We close with some comments on the model, robustness, and open problems.

Alternative Information Models. We build on a canonical model of information that is central to the statistical decision theory literature, i.e., the normal prior-normal signals model with quadratic payoff. This set up features prominently in the economic literature on, e.g., teams, networks, and global games, especially because it is tractable.[19] Despite the complications introduced by non-identically distributed signals and correlation, we were able to derive the team value function in closed form, a crucial step in our analysis. It is natural to wonder about the robustness of our insights when we venture beyond the confines of the model. For instance, another popular way to model information acquisition in economic applications is to assume that an agent receives an informative signal with some probability, and otherwise receives pure noise. In our setting, this can be modelled as follows: an agent's type x_i is her probability of receiving an informative signal; signals are conditionally independent; and the team's payoff if n informative signals are observed is $u(n)$, where $u(\cdot)$ is 'concave' in n (e.g., informative signals are drawn from a normal distribution centered at s with precision κ). If $k = 2$, then the team value function is

$$V(x_1, x_2) = x_1 x_2 u(2) + (x_1(1 - x_2) + x_2(1 - x_1))u(1) + (1 - x_1)(1 - x_2)u(0),$$

which is clearly strictly submodular in (x_1, x_2). Using Poisson's binomial distribution (Wang (1993)), we show in Appendix A.11 that this is true for any k. Therefore, diversification is optimal in this alternative model as well.[20]

Going beyond the analysis of canonical models is hard, as (*a*) we lose tractability and, more importantly, (*b*) we do not know much in general about curvature properties of the value of information. For instance, we have exploited the concavity in $\sum x_i$ of the value function and similarly in the alternative model just outlined. But it is well-known (e.g., Chade and Schlee (2002), Moscarini and Smith (2002)) that nonconcavities are hard to rule out in models with information acquisition. Until there is more progress on this issue, a general analysis of our matching problem will remain elusive.

Class of Matching Problems. We cast our model as an information aggregation problem that naturally occurs in economic contexts where teams form. Besides economic relevance, the model also provides a microfoundation for the value function of each team based on the informativeness of its members' signals. But as the conditionally independent case makes clear, all that matters are the properties of the value function $v\left(\sum x_i\right)$ and that the sum of the precision of the teams is constant across all partitions of the set of agents. That is, the results also apply to a *class* of matching problems

[19] For a couple of representative contributions to networks and global games, see Ballester, Calvó-Armengol, and Zenou (2006), Angeletos and Pavan (2007), and the references therein.

[20] As we note in Appendix A.11, it turns out that the team value function in this case is very similar to the expected diversity function in Weitzman (1998) in a different setting.

where the value of a team is strictly increasing and strictly concave in $\sum x_i$. Then the planner's objective function is Schur concave and the sorting properties derived in the conditionally independent case hold. Since submodular maximization problems are in general NP-hard, this constitutes a subset of such problems where a lot can be said about their solution. It would be interesting to know if one could extend this class further.

Nontransferable Utility. We have assumed that utility is transferable, which is a standard assumption in matching models. One could, however, envision applications where this assumption is less palatable, e.g., if experts were risk averse. Then besides the information aggregation motive they would be interested in sharing the risky payoff efficiently. Similarly, if moral hazard were added to the problem, e.g., agents exert unobservable effort to affect signal precision, then the incentive constraints would impose limits on transferability. These variations would (in most cases) turn the model into one with nontransferable utility (see Legros and Newman (2007)). Exploring the sorting patterns in this case is a relevant open problem to pursue.

Different Group Sizes. We have assumed that all groups must be of size k. As mentioned, this is a generalization of the standard assumption made in most assignment games such as Becker (1973), where it is exogenously imposed that agents match in pairs. We have used this assumption, e.g., in Proposition 3 when checking for profitable swaps of experts, for it was important to have teams' value functions defined on the same domain. We view the extension of the analysis to groups of different sizes as an important open problem. As an illustration, consider $\rho = 0$ and six agents, with types 2, 2, 7, 7, 8, and 10, respectively. If $k = 3$, then the optimal partition is $\{2, 7, 10\}, \{2, 7, 8\}$, with $X_1 = 19$ and $X_2 = 17$. If we allow the two groups to be of different size, then the optimal partition is $\{2, 2, 7, 7\}, \{8, 10\}$, with $X_1 = X_2 = 18$, a strict improvement. The partition is consecutive in the sense that it puts high types in one team and low ones in the other, and *diversification* occurs via the size of each group (the group with better types is smaller). Notice that this issue is irrelevant under fractional assignment, since perfect diversification yields $X_1 = X_2 = 18$. More generally, observe that under fractional assignment, since there is a unique solution in terms of teams' precision, there always exists an optimal assignment where all firms have the same size.

Approximation Algorithms. Although no efficient algorithm is known for this problem, there are approximation algorithms derived for similar ones. In particular, our matching problem is a variant of the so-called submodular welfare problem (Vondrak (2007) chapter 4). To see this, rewrite problem (2) as one that maximizes $\sum_{n=1}^{N} \hat{v}(S_n)$

over all partitions of the set of agents $\mathcal{I}$ with the additional constraint $|S_n| = k$ for all n, where $\hat{v}(S_n) = v\left(\sum_{i \in S_n} x_i\right)$. Ignoring the constraint, this is a submodular welfare problem with identical 'players,' and it is known to be NP-hard. Vondrak (2007) provides an approximation algorithm that captures a fraction $(1 - 1/e)$ of the optimal value. It would be interesting to explore this issue further in our setting.

STOCHASTIC SORTING. In our model, an agent's type is a scalar that indexes a family of conditional densities ordered by informativeness. That is, one can interpret the model as one of matching *distributions*. Besides information aggregation, it is easy to think of applications where agents' types index probability distributions. Consider for instance a marriage model a la Becker (1973) where an agent's type is his or her current attribute (e.g., education), but the payoff-relevant attribute (e.g., income) is drawn after the match from a distribution conditional on the current type.[21] Depending on the order imposed on those distributions (e.g., first-order stochastic dominance or mean-preserving spread), the conditions on the match payoff function that determine PAM or NAM based on current types will differ. In current work in progress we have found that this reinterpretation of an agent's type can be illuminating in matching problems (both with transferable and nontransferable utility) under uncertainty.

[21] For a variation thereof, see the sorting model with signals in Hoppe, Moldovanu, and Sela (2009), or with noise in Chade (2006). For the broader subject of matching with informational frictions, see for example Liu, Mailath, Postlewaite, and Samuelson (2012) and Chakraborty, Citanna, and Ostrovsky (2010) amongst others.

A Appendix: Omitted Proofs

A.1 Proof of Proposition 1

We first state a few facts about Bayesian updating and the normal distribution that we invoke in the proof. Recall that each signal $\sigma_i \sim \mathcal{N}(s, x_i^{-1})$, $i = 1, 2, ..., k$, and the vector $\vec{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_k)$ is distributed $\vec{\sigma} \sim \mathcal{N}(\vec{s}, \Sigma_k)$, where $\vec{s}$ is a $k \times 1$ vector with s in all entries, and Σ_k is a $k \times k$ symmetric positive definite matrix with diagonal elements $1/x_i$, $i = 1, 2, ..., k$, and off-diagonal ones $\rho(x_i x_j)^{-0.5}$, $i \neq j$. Also, $\tilde{s} \sim \mathcal{N}(\mu, \tau^{-1})$.

Fact 1. The inverse of Σ_k is the $k \times k$ matrix $\Sigma_k^{-1} = [q_{ij}]$, where for all $i, j = 1, 2, ..., k$

$$q_{ij} = \frac{-\rho(x_i x_j)^{0.5}}{(1-\rho)(1+(k-1)\rho)} \quad \forall i \neq j \quad q_{ii} = \frac{x_i(1+(k-2)\rho)}{(1-\rho)(1+(k-1)\rho)}.$$

To prove it, algebra shows that $\Sigma_k^{-1}\Sigma_k = I_k$, where I_k is the $k \times k$ identity matrix.

Fact 2. The conditional distribution of σ_i, $i = 2, 3, ..., k$, given $(\sigma_1, \sigma_2, ..., \sigma_{i-1})$ is

$$\sigma_i|_{\sigma_1,\sigma_2,...,\sigma_{i-1}} \sim \mathcal{N}\left(a_i s + b_i, \frac{1}{q_{ii}}\right),$$

where a_i and b_i are given by the following expressions:

$$a_i = 1 + \frac{\sum_{j=1}^{i-1} q_{ij}}{q_{ii}} \quad b_i = -\frac{\sum_{j=1}^{i-1} q_{ij}\sigma_j}{q_{ii}}.$$

This follows from normal distribution results (e.g., Section 5.4 in DeGroot (1970)).

Fact 3. Given random variables $\tilde{\theta} \sim \mathcal{N}(m, t^{-1})$ and $\tilde{y} \sim \mathcal{N}(b + a\theta, x^{-1})$, then

$$\tilde{\theta}|_y \sim \mathcal{N}\left(\frac{tm + xa(y-b)}{t + a^2 x}, \frac{1}{t+x}\right).$$

This is a standard result; e.g., see Williams (1991), section 15.7.

Fact 4. Using $\sigma_1 \sim \mathcal{N}(s, x_1^{-1})$ and Facts 2–3, we obtain:

$$\tilde{s}|_{\sigma_1} \sim \mathcal{N}\left(\frac{\tau\mu + \sigma_1 x_1}{\tau + x_1}, \frac{1}{\tau + x_1}\right) \tag{10}$$

$$\tilde{s}|_{\sigma_1,\sigma_2,...,\sigma_i} \sim \mathcal{N}\left(\frac{\tau_{i-1}\mu_{i-1} + a_i(\sigma_i - b_i)q_{ii}}{\tau_{i-1} + a_i^2 q_{ii}}, \frac{1}{\tau_{i-1} + a_i^2 q_{ii}}\right) \quad i = 2, 3, ..., k. \tag{11}$$

This follows immediately from the two facts mentioned.

Proof of Proposition 1. The derivation of $V(\vec{x})$ shows that $V(\vec{x}) = \pi - \mathbb{E}[\mathrm{Var}(\tilde{s}|\vec{\sigma}, \vec{x}, \rho)]$, where the expectation is taken with respect to the distribution of $\vec{\sigma}$. Thus, we must show that $\mathrm{Var}(\tilde{s}|\vec{\sigma}, \vec{x}, \rho) = 1/(\tau + \mathcal{B}(\vec{x}, \rho))$, which is independent of $\vec{\sigma}$. We will prove this result in terms of precision, i.e., we will show that $\tau_k = \tau + \mathcal{B}(\vec{x}, \rho)$.

We proceed by induction. This is true for $k = 1$, as (5) collapses to $\mathcal{B}(x_1, \rho) = x_1$ and thus $\tau_1 = \tau + x_1$. Assume it is true for $k - 1$. We will show it is true for k as well. Using (in this order) Facts 4, 2, and 1, we can write τ_k as follows:

$$\begin{aligned}
\tau_k &= \tau_{k-1} + a_k^2 q_{kk} \\
&= \tau_{k-1} + \left(1 - \frac{\rho}{(1 + (k-2)\rho)} \frac{\sum_{j=1}^{k-1} x_j^{0.5}}{x_k^{0.5}}\right)^2 \left(\frac{x_k(1 + (k-2)\rho)}{(1-\rho)(1 + (k-1)\rho)}\right) \\
&= \tau_{k-1} + \frac{\left(x_k^{0.5}(1 + (k-2)\rho) - \rho \sum_{j=1}^{k-1} x_j^{0.5}\right)^2}{(1-\rho)(1 + (k-2)\rho)(1 + (k-1)\rho)}.
\end{aligned} \tag{12}$$

By the induction hypothesis, $\tau_{k-1} = \tau + \mathcal{B}(x_1, x_2, ..., x_{k-1}, \rho)$ or, equivalently (using (5)),

$$\tau_{k-1} = \tau + \frac{(1 + (k-3)\rho) \sum_{i=1}^{k-1} x_i - 2\rho \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} (x_i x_j)^{0.5}}{(1-\rho)(1 + (k-2)\rho)}. \tag{13}$$

Combining (12) and (13) yields, after long but straightforward algebra,

$$\tau_k = \tau + \frac{\left(\frac{((1+(k-1)\rho)(1+(k-3)\rho)+\rho^2) \sum_{i=1}^{k-1} x_i}{1+(k-2)\rho} - 2\rho \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (x_i x_j)^{0.5} + x_k(1 + (k-2)\rho)\right)}{(1-\rho)(1 + (k-1)\rho)}. \tag{14}$$

Since $(1 + (k-1)\rho)(1 + (k-3)\rho) + \rho^2 = (1 + (k-2)\rho)^2$, (14) can be written as

$$\tau_k = \tau + \frac{(1 + (k-2)\rho) \sum_{i=1}^{k} x_i - 2\rho \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (x_i x_j)^{0.5}}{(1-\rho)(1 + (k-1)\rho)} = \tau + \mathcal{B}(\vec{x}, \rho).$$

Hence, the formula is true for k as well, and the induction proof is complete. □

A.2 $\mathcal{B}(\vec{x}, \rho)$ and Blackwell More Informative Signals

The following result, which follows from a theorem in Hansen and Torgersen (1974), is stated in Goel and Ginebra (2003) p. 521: Let $X = (X_1, ..., X_n) \sim \mathcal{N}(A\beta, \Sigma_X)$ and

$Y = Y_1, ..., Y_m \sim \mathcal{N}(B\beta, \Sigma_Y)$, where $\beta = (\beta_1, ..., \beta_l)'$ is a vector of unknown parameters, A is a known $n \times l$ matrix, B is a known $m \times l$ matrix, and Σ_X and Σ_Y are positive definite covariance matrices. Then X is more informative than Y if and only if $A'\Sigma_X^{-1}A - B'\Sigma_Y^{-1}B$ is a nonnegative definite matrix.

Let us apply this result to our setting. Consider two teams with composition $\vec{x}$ and $\vec{x}'$, respectively. Let the vector β be simply the scalar s, and the matrices A and B are the $k \times 1$ unit vector I_k. Then $\vec{\sigma}$ is more informative than $\vec{\sigma}'$ if and only if (the scalar) $I_k'\Sigma_k^{-1}I_k - I_k'\Sigma_k'^{-1}I_k \geq 0$. Tedious algebra using the inverse of the covariance matrix given in the proof of Proposition 1 reveals that this is equivalent to $\mathcal{B}(\vec{x}, \rho) \geq \mathcal{B}(\vec{x}', \rho)$, thereby showing that $\mathcal{B}$ indexes the informativeness of the team signals. □

A.3 Proof of Proposition 2

We need to compare $\mathcal{B}(\vec{x}, \rho)$ and $\mathcal{B}(\vec{x}, 0)$, or

$$\frac{(1 + (k-2)\rho)\sum_{i=1}^{k} x_i - 2\rho\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5}}{(1-\rho)(1+(k-1)\rho)} \gtrless \sum_{i=1}^{k} x_i. \tag{15}$$

Straightforward algebra reveals that (15) is equivalent to

$$\rho^2(k-1)\sum_{i=1}^{k} x_i - 2\rho\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5} \gtrless 0. \tag{16}$$

(i) It is immediate that (16) is positive when $\rho < 0$. Hence, negatively correlated signals are more informative than conditionally independent ones.

(ii) Assume $\rho > 0$ and rewrite (16) as follows:

$$\rho\left(\rho(k-1)\sum_{i=1}^{k} x_i - 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5}\right) \gtrless 0. \tag{17}$$

When ρ is equal to one (17) is positive, as $(k-1)\sum x_i - 2\sum\sum(x_i x_j)^{0.5} \geq \sum x_i - 2\sum\sum(x_i x_j)^{0.5} = (\sum x_i^{0.5})^2 > 0$. And when ρ is close to zero the expression in parenthesis is negative. Moreover, the expression in parenthesis is strictly increasing in ρ. Hence, for any $\vec{x}$ there is a threshold $0 < \tilde{\rho}(\vec{x}) = (2\sum\sum(x_i x_j)^{0.5})/((k-1)\sum x_i) \leq 1$ such that (17) is negative if and only if $0 < \rho < \tilde{\rho}(\vec{x})$.

Since the threshold must work uniformly for all $\vec{x}$, define

$$\hat{\rho} = \frac{2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\underline{x}^2)^{0.5}}{(k-1)\sum_{i=1}^{k}\overline{x}} = \frac{2\frac{k(k-1)}{2}\underline{x}}{(k-1)k\overline{x}} = \frac{\underline{x}}{\overline{x}}.$$

If $0 < \rho < \hat{\rho}$, then (17) is negative, and thus positively correlated signals are less informative than conditionally independent ones. □

A.4 Proof of Lemma 1

Recall that the team value function is:

$$V(\vec{x}) = \pi - \left(\frac{1}{\tau + \frac{(1+(k-2)\rho)\sum_{i=1}^{k} x_i - 2\rho\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5}}{(1-\rho)(1+(k-1)\rho)}} \right).$$

Since this function is $\mathcal{C}^2$, it follows that $V(\cdot)$ is submodular (supermodular) in $\vec{x}$ if and only if $V_{lm} = \partial^2 V/\partial x_l \partial x_m \leq (\geq)\ 0$ for all $1 \leq l \neq m \leq k$. Simple yet long algebra reveals that the sign of V_{lm} is equal to the sign of the following expression:

$$(1+(k-2)\rho)\left(4\rho\left(x_l^{0.5}\sum_{j\neq m} x_j^{0.5} + x_m^{0.5}\sum_{j\neq l} x_j^{0.5}\right) - \rho\sum_{i=1}^{k} x_i - (1+(k-2)\rho)4(x_l x_m)^{0.5}\right)$$
$$-\tau\rho(1-\rho)(1+(k-1)\rho) + 2\rho^2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5} - 4\rho^2\sum_{j\neq m} x_j^{0.5}\sum_{j\neq l} x_j^{0.5}. \quad (18)$$

(i) Notice that expression (18) evaluated at $\rho = 0$ is equal to $-4x_l^{0.5}x_m^{0.5} < 0$ and thus $V_{lm}|_{\rho=0} < 0$. Since this holds for any l, m and any values of x_l, x_m, and since the expression is continuous in ρ, it follows that there exists a $\tilde{\rho} > 0$ — that holds for any $\vec{x} \in [\underline{x}, \overline{x}]^k$ — such that, if $\rho \in (-\tilde{\rho}, \tilde{\rho})$, then $V(\cdot)$ is strictly submodular in $\vec{x}$.[22]

(ii) Let $x \in [\underline{x}, \overline{x}]$ and $x_1 = x_2 = ... = x_k = x$. Evaluating (18) at this vector yields:

$$(1+(k-2)\rho)\left(8\rho(k-1)x - \rho k x - (1+(k-2)\rho)4x\right) - \tau\rho(1-\rho)(1+(k-1)\rho) + \rho^2(k-1)x(k-4(k-1)).$$

[22] To see that $\tilde{\rho}$ can be chosen to hold for any $\vec{x}$, assume that $\rho > 0$ ($\rho < 0$) and replace any x in a positive (negative) term in (18) by $\overline{x}$ ($\underline{x}$) and then do the same continuity argument as above. The resulting $\tilde{\rho}$ only depends on $\underline{x}$ and $\overline{x}$. Clearly, this is not the largest $\tilde{\rho}$ that can be constructed.

After algebraic manipulation, this expression can be written as follows:

$$-\tau\rho(1-\rho)(1+(k-1)\rho) - x(4(1-\rho)^2 + k\rho(1-\rho)) < 0. \tag{19}$$

Hence, $V_{lm}|_{x_1=\ldots=x_k=x} < 0$ for any $x \in [\underline{x}, \overline{x}]$ and any $l \neq m$. By continuity, there is an interval around x such that if the components of $\vec{x}$ belong to that interval, then $V_{l,m} < 0$ for all l, m. Thus, for any $\rho \in (\tilde{\rho}, 1)$, $V(\cdot)$ cannot be supermodular on $[\underline{x}, \overline{x}]^k$.[23] And since the term containing τ in (18) is negative when $\rho > 0$, it follows that for τ large enough the team value function is strictly submodular.[24]

(*iii*) Notice that (19) is negative for τ sufficiently small, as the parenthesis in the second term is positive for all $\rho \in (-(k-1)^{-1}, 0)$. Thus, the team value function cannot be supermodular by a similar argument as in part (*ii*). And since the term containing τ in (18) is negative when $\rho < 0$, for τ large enough it is strictly supermodular. $\square$

A.5 Supermodularity of $V(\cdot)$ and Prior Precision

Assume $\rho < \tilde{\rho}$ as in Lemma 1 (*iii*) . We assert in Section 3.2 that in this case supermodularity of $V(\cdot)$ in $\vec{x}$ requires $\tau > 8k\overline{x}$. We now justify this assertion. Using (19), it follows that a necessary condition for $V(\cdot)$ supermodular in $\vec{x}$ is that prior precision be larger than the following bound:

$$\tau > \frac{\frac{4(1-\rho)}{k} + \rho}{-\rho(1+(k-1)\rho)}\ k\overline{x} > \frac{4(k-1)\min\left\{4, \frac{3k}{k-1}\right\}}{k}\ k\overline{x} \geq 8\ k\overline{x},$$

where the second inequality follows by maximizing the denominator and minimizing the numerator separately with respect to $\rho \in (-(k-1)^{-1}, 0)$.

A.6 Proof of Proposition 4

(*i*) Towards a contradiction, assume that the optimal partition has a precision vector $(X_1, .X_2, ..., X_N)$ that does not belong to M (this set is defined in the text). Since any element of M is majorized by $(X_1, X_2, ..., X_N)$ and the objective function is Schur concave, an improvement is possible, thereby contradicting the optimality of $(X_1, X_2, ..., X_N)$.

[23] Since the result holds for any $x \in [\underline{x}, \overline{x}]$, it follows that for $|\overline{x} - \underline{x}|$ sufficiently small, i.e., equal to the aforementioned neighborhood around x, the team value function is strictly submodular. That is, for any $\rho \in (\tilde{\rho}, 1)$, diversification ensues if heterogeneity of expertise is small.

[24] The threshold on τ can be made independent of $\vec{x}$ by a similar argument than that in footnote 22.

(ii) This follows from (i) and the singleton property of M. $\square$

A.7 Failure of the Gross Substitutes Property

Recall our interpretation of matching groups of experts with identical firms. A decentralized version of the problem would have each firm face a vector of type-dependent wages w at which it can hire them. Ignore the size-k restriction in what follows (it is easy to introduce it). The firm solves:

$$\max_{A \subseteq \mathcal{I}} v\left(\sum_{i \in A} x_i\right) - \sum_{i \in A} w(x_i)$$

Let $D(w)$ be the set of solutions to this problem. The crucial property in Kelso and Crawford (1982) is the gross substitutes condition (GS): If $A^* \in D(w)$ and $w' \geq w$, then there is a $B^* \in D(w')$ such that $T(A^*) \subseteq B^*$, where $T(A^*) = \{i \in A^* | w(x_i) = w'(x_i)\}$.

The following example shows that GS fails in our model.

The firm faces three experts, 1, 2, and 3, with types $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$. Make the innocuous assumption that $\pi = \tau = 1$ (so hiring nobody yields zero profits).

Let $w = ((1/12) - \varepsilon, 1/12, 1/6)$. Then it is easy to verify that $v(x_1) - w(x_1) = (5/12) + \varepsilon$, $v(x_2) - w(x_2) = 7/12$, $v(x_3) - w(x_3) = 7/12$, $v(x_1 + x_2) - w(x_1) - w(x_2) = (7/12) + \varepsilon$, $v(x_1 + x_3) - w(x_1) - w(x_3) = (33/60) + \varepsilon$, $v(x_2 + x_3) - w(x_2) - w(x_3) = 7/12$, $v(x_1 + x_2 + x_3) - w(x_1) - w(x_2) - w(x_3) = (44/84) + \varepsilon$. Thus, the optimal choice is unique and given by $A^* = \{1, 2\}$.

Suppose now that $w = ((1/12) - \varepsilon, 1/6, 1/6)$, so that only the wage of expert 2 has increased. Profits from each subset of experts are $v(x_1) - w(x_1) = (5/12) + \varepsilon$, $v(x_2) - w(x_2) = 1/2$, $v(x_3) - w(x_3) = 7/12$, $v(x_1 + x_2) - w(x_1) - w(x_2) = (1/2) + \varepsilon$, $v(x_1 + x_3) - w(x_1) - w(x_3) = (33/60) + \varepsilon$, $v(x_2 + x_3) - w(x_2) - w(x_3) = 1/2$, $v(x_1 + x_2 + x_3) - w(x_1) - w(x_2) - w(x_3) = (37/84) + \varepsilon$. Thus, if $\varepsilon < 1/30$, then the optimal choice is unique and given by $B^* = \{3\}$.

Since in this case $T(A^*) = \{1\} \nsubseteq B^*$, it follows that GS does not hold.

Notice that the same example shows that if the firm were constrained to hire at most two experts, GS would still fail. $\square$

A.8 Proof of Proposition 5

Let $X_n \equiv \sum_{j=1}^{J} \mu_{jn} x_j$, and notice that if we multiply both sides of the first constraint in (6) by x_j and sum with respect to j, we obtain $\sum_{n=1}^{N} X_n = X$. If we ignore the other two constraints in (6), we obtain the 'relaxed problem' of finding $(X_1, ..., X_N)$ to maximize $\sum_{n=1}^{N} v(X_n)$ subject to $\sum_{n=1}^{N} X_n = X$, whose unique solution is $X_n = X/N$ for all n. Thus, the optimal matching equalizes the precision of all teams.

If the unique solution to the relaxed problem solves the original fractional assignment problem, then it is a solution to that problem. The task is to find values for the μ_{jn}'s such that the equal precision rule X/N for all n satisfies all the feasibility constraints. Let $\mu_{jn} = m_j/N$ for all j, n. Then $X_n = \sum_j \mu_{jn} x_j = \sum_j (m_j/N) x_j = X/N$ for all n. Moreover, $\sum_n \mu_{jn} = \sum_n m_j/N = N m_j/N = m_j$ and $\sum_j \mu_{jn} = \sum_j m_j/N = kN/N = k$. Thus, $\mu_{jn} = m_j/N$ for all j, n solves the fractional assignment problem. $\square$

A.9 Endogenous Team Size

We now prove the assertions made in Section 4.2. We focus on (8) since the other case is analogous. Rewrite the equilibrium condition as follows:

$$\pi - F = \frac{N^2 \tau + 2NX}{(N\tau + X)^2}. \tag{20}$$

The left-hand side is a positive constant. The right-hand side is zero at $N = 0$, it is strictly increasing in N, and converges to $1/\tau$ as N goes to infinity. Since $\pi - F < 1/\tau$, there is a unique N^* (and hence k^*) that solves (20).

The comparative statics of N^* with respect to F, τ, and X are as follows. Rewrite (20) as $\pi - F = z(N^*, \tau, X)$. It is easy to verify that the right-hand side is strictly decreasing in τ and also in X. Thus, $\partial N^*/\partial F = -1/z_N < 0$, $\partial N^*/\partial \tau = -z_\tau/z_N > 0$, and $\partial N^*/\partial \tau = -z_X/z_N > 0$. Hence, k^* increases in F, and it decreases in τ and X. $\square$

A.10 Proof of Proposition 6

(*i*) Since $v'(X_i) = (\tau + X_i)^{-2}$, the first-order condition $y_n v'(X_n) = y_m v'(X_m)$ can be written as follows:

$$\tau + X_m = \frac{y_m^{0.5}}{y_n^{0.5}} (\tau + X_n).$$

Fix n and sum both sides for all m. Using $\sum_{m=1}^{N} X_m = X$, we obtain

$$N\tau + X = \frac{\sum_{m=1}^{N} y_m^{0.5}}{y_n^{0.5}}(\tau + X_n) \Rightarrow X_n = \frac{y_n^{0.5}}{\sum_{m=1}^{N} y_m^{0.5}}\ (\tau N + X) - \tau,$$

which yields (9). Since the first-order conditions are necessary and sufficient for optimality, this is the unique solution to the optimization problem. Notice that X_n is increasing in n, thus showing the PAM property between firm quality and team precision.

(ii)–*(iii)* The derivative of (9) with respect to X is $y_n^{0.5}/\sum y_n^{0.5} > 0$, and thus an increase in X increases X_n for all n.

The derivative of (9) with respect to τ is $(y_n^{0.5}N/\sum y_n^{0.5}) - 1$, and this is positive if and only if $y_n^{0.5} > \sum y_n^{0.5}/N$. Since y_n is increasing in n, it follows that there is an n^* such that X_n increases in τ for $n \geq n^*$ and decreases otherwise.

The difference $X_n - X_{n-1}$ is given by

$$X_n - X_{n-1} = \frac{(y_n^{0.5} - y_{n-1}^{0.5})}{\sum_{n=1}^{N} y_n^{0.5}}\ (\tau N + X)\,.$$

Since $y_n \geq y_{n-1}$, an increase in X or in τ increases $X_n - X_{n-1}$.

(iv) Let N be odd and $\Delta > 0$. Consider $y' = (y'_1, y'_2, ..., y'_N)$ such that $y'_n = y_n - \Delta$ for all $n < (N+1)/2$, $y'_n = y_n - \Delta$ for all $n > (N+1)/2$, and $y'_{\frac{N+1}{2}} = y_{\frac{N+1}{2}}$. Notice that y' majorizes $y = (y_1, y_2, ..., y_N)$. We will show that there exists an n^* such that X_n increases for all $n \geq n^*$ and decreases otherwise when y is replaced by y'.

Since $\sum (y_n)^{0.5}$ is the sum of concave functions in one variable $y_n^{0.5}$, it follows that it is Schur concave in $(y_1, y_2, ..., y_N)$. Hence, $\sum (y'_n)^{0.5} \leq \sum (y_n)^{0.5}$ as y' majorizes y. It is now immediate that X_n increases for all $n \geq (N+1)/2$, for in this case $(y'_n)^{0.5} > (y_n)^{0.5}$ and thus $(y'_n)^{0.5}/\sum (y'_n)^{0.5} > (y_n)^{0.5}/\sum (y_n)^{0.5}$, thereby increasing X_n (see equation (9)).

Consider now teams with index in the set $\{1, 2, ..., (N+1)/2\}$. We will show that if X_n decreases for any n in this set, then it must decrease for all teams $1, ..., n$. This will prove that there is an n^* with the aforementioned properties. Let there be an n in $\{1, 2, ..., (N+1)/2\}$ such that X_n decreases when y' replaces y. This holds if and only if

$$\frac{(y_n - \Delta)^{0.5}}{\sum_{n=1}^{N} y_n'^{0.5}} \leq \frac{y_n^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}} \;\; \Leftrightarrow \;\; \left(1 - \frac{\Delta}{y_n}\right)^{0.5} \leq \frac{\sum_{n=1}^{N} y_n'^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}}.$$

Since the last term is a constant and $(1 - (\Delta/y_n))^{0.5}$ decreases when y_n is replaced by a lower value y_{n-i}, it follows that if X_n decreases for such an n, it must decrease for all teams with a lower index, thus proving the claim.

We have focused on N odd. If N is even, then let $y'_n = y_n - \Delta$ for all $n \leq N/2$, $y'_n = y_n - \Delta$ for all $n > N/2$, and proceed exactly as before.

(v) The optimal $(X_1, X_2, ..., X_N)$ can be implemented by any fractional assignment vector $\{\mu_{jn}\}_{j,n}$ that satisfies the following system:

$$\begin{aligned} \sum_{j=1}^{J} \mu_{jn} x_j &= X_n \quad \forall\, n \\ \sum_{j=1}^{J} \mu_{jn} &= k \qquad \forall\, n \\ \mu_{jn} &\geq 0 \qquad \forall\, j, n, \end{aligned}$$

where we have omitted from the system the equations $\sum_{n=1}^{N} \mu_{jn} = m_j$ for all j since they are implied by the other ones.[25] Since there are only $2N$ equations and JN unknowns, there is an infinite number of solutions if $J > 2$; any one that satisfies the nonnegativity constraints will implement the optimal precision vector $(X_1, X_2, ..., X_N)$.

We now show that such a vector always exists. Rewrite the above system as

$$A\mu = b, \quad \mu \geq 0 \tag{21}$$

where:

$A = [A_1, ..., A_N]$ is a $2N \times JN$ matrix; each A_n is a $2N \times J$ matrix with zeroes in all rows except for row n, which is $[x_1, x_2, ..., x_J]$, and row $n + N$, which is $[1, 1, ..., 1]$;

μ is a $JN \times 1$ vector whose entries are μ_{jn} for all j, n; and

$b = [b_1, b_2]$ is a $2N \times 1$ vector, with $b_1 = [X_1, X_2, ...X_N]$ and $b_2 = [k, k, ..., k]$.

By Farkas' Lemma, (21) has a solution if and only if there is no $1 \times 2N$ vector y that solves the following system:

$$yA \geq 0 \quad yb < 0 \tag{22}$$

Assume there is a solution to (22). From $yb < 0$, there is a $j \in \{1, 2, ..., 2N\}$ such that $y_j < 0$. Suppose first that $j \in \{1, 2, ..., N\}$. From $yA \geq 0$, we obtain that $y_j x_J + y_{j+N} \geq 0$

[25] Notice that $\sum_n \sum_j \mu_{jn} x_j = \sum_j (\sum_n \mu_{jn}) x_j = \sum_n X_n = X = \sum_j m_j x_j$; hence, $\sum_n \mu_{jn} = m_j \;\; \forall j$.

(this follows from multiplying y and the J-th column of A_j). Since $kx_J \geq X_n$ for all n,

$$ky_{j+N} \geq -y_j kx_J \geq -y_j X_j \Rightarrow y_j X_j + ky_{j+N} \geq 0.$$

But then for any $y_j < 0$, $j \in \{1, ..., N\}$, the term $y_j X_j$ in yb will be dominated by ky_{j+N}, and hence we cannot have $yb < 0$, contradiction.

Assume now that $j \in \{N+1, N+2, ..., 2N\}$. Then, to satisfy $yb < 0$, either $\sum_{n=N+1}^{2N} y_n < 0$ or, if this is nonnegative, there is a $j' \in \{1, 2, ..., N\}$ such that $y_{j'} < 0$ (otherwise $yb < 0$ could not hold). If the latter, proceed as above to reach a contradiction. If the former, assume first that only $y_j < 0$ among those y_n with $n \geq N+1$. Notice that $yb < 0$ implies

$$\sum_{n=1}^{N} y_n X_n < -k \sum_{n=N+1}^{2N} y_n \leq -ky_j. \tag{23}$$

Now, from $yA \geq 0$, we obtain $y_{j-N} x_1 + y_j \geq 0$ (this follows from multiplying y and the first column of A_{j-N}). But then $y_{j-N} kx_1 \geq -ky_j$, and since $X_n \geq kx_1$ for all n, $y_{j-N} kx_1 \leq \sum_{n=1}^{N} y_n X_n$, and thus $\sum_{n=1}^{N} y_n X_n \geq -ky_j$, which is a contradiction.

If besides $y_j < 0$ there are other $y_n < 0$, $n \geq N+1$, let $B = \{n \geq N+1 | y_n < 0\}$ and replace the rightmost expression in (23) by $-k \sum_{n \in B} y_n$, and use the same argument as in the previous paragraph for each $N - n$, $n \in B$ to reach the desired contradiction.

Hence, system (22) does not have a solution; by Farkas' Lemma, there is a solution to (21), which proves existence of a fractional assignment of agents into k-size teams with precision X_n given by (9) for all n. $\square$

When $J = 2$, it is easy to verify that the unique solution to the fractional assignment problem is given by the following vector for each $n = 1, 2, ..., N$:

$$(\mu_{1n}, \mu_{2n}) = \left(\frac{kx_2 - X_n}{x_2 - x_1}, \frac{X_n - kx_1}{x_2 - x_1} \right). \tag{24}$$

Intuitively, μ_{1n} decreases in n while μ_{2n} increases in n. When $N = k = 2$, a simple rewriting of (24) yields the formulas used in the example in the text.

A.11 Alternative Information Model

We now prove the assertion made in Section 6 that the team value function is strictly submodular when an agent's type is the probability of receiving an informative signal. We could have appealed to Weitzman (1998) Theorem 2, since the value function is similar to the submodular expected diversity function considered in that paper. To make the paper self-contained, we include a proof that is slightly different.

Denote by $u(n)$ the payoff to the team when there are n informative signals out of the k realizations. We assume that $u(\cdot)$ is strictly increasing in n and satisfies strictly decreasing differences in n, i.e., $u(n) - u(n-1)$ is strictly decreasing in n. Since each signal is informative with probability x_i, $i = 1, ..., k$, the number of informative signals in a k-size team is a random variable with Poisson's binomial distribution (Wang (1993)). Let $m = 0, 1, ..., k$ and define $\mathcal{F}_m \equiv \{B : B \subseteq \{1, 2, ..., k\}, |B| = m\}$.

The probability of m informative signals out of k in a team with types $\vec{x}$ is given by

$$\sum_{B \in \mathcal{F}_m} \left(\prod_{i \in B} x_i \right) \left(\prod_{j \notin B} (1 - x_j) \right).$$

Thus, the team value function is

$$\begin{aligned}
V(\vec{x}) &= \sum_{m=0}^{k} u(m) \left(\sum_{B \in \mathcal{F}_m} \left(\prod_{i \in B} x_i \right) \left(\prod_{j \notin B} (1 - x_j) \right) \right) \\
&= \sum_{R \subseteq \{1,2,...,k\}} u(|R|) \left(\prod_{i \in R} x_i \right) \left(\prod_{j \notin R} (1 - x_j) \right) \\
&= \sum_{R \subseteq \{1,2,...,k\} \setminus \{i,j\}} \left(\prod_{i \in R} x_i \right) \left(\prod_{j \notin R} (1 - x_j) \right) \ (x_i x_j u(|R|+2) + x_i(1-x_j) u(|R|+1) \\
&\quad + x_j(1-x_i) u(|R|+1) + (1-x_i)(1-x_j) u(|R|), \qquad (25)
\end{aligned}$$

where the second equality follows from the fact that summing over all sets is the same as summing first over all sets of a given cardinality and then over all feasible set sizes, and the third equality follows from a straightforward decomposition of the sum (see Lemma 3 in Calinescu, Chekuri, Pal, and Vondrak (2007)).

Differentiating (25) with respect to x_i and x_j yields

$$\operatorname{sgn}\left(\frac{\partial^2 V(\vec{x})}{\partial x_i \partial x_j}\right) = \operatorname{sgn}\left((u(|R|+2) - u(|R|+1)) - (u(|R|+1) - u(|R|))\right) < 0,$$

where the inequality follows from the strictly decreasing difference property of $u(\cdot)$ in n. Since i and j were arbitrary, $V(\cdot)$ is strictly submodular in $\vec{x}$. $\square$

References

Angeletos, G.-M., and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75(4), 1103–1142.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 74(5), 1403–1417.

Becker, G. (1973): "A Theory of Marriage I," *Journal of Political Economy*, 81, 813–846.

Bickel, P., and K. Doksum (1977): *Mathematical Statistics*. Holden-Day, Inc.

Borgers, T., A. Hernando-Veciana, and D. Krahmer (forthcoming): "When are Signals Complements or Substitutes?," *Journal of Economic Theory*.

Calinescu, G., C. Chekuri, M. Pal, and J. Vondrak (2007): "Maximizing a Monotone Submodular Function subject to a Matroid Constraint," *Proceedings of 12th IPCO*, pp. 182–196.

Chade, H. (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1), 81–113.

Chade, H., and E. Schlee (2002): "Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information," *Journal of Economic Theory*, 107, 421–452.

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): "Two-sided matching with interdependent values.," *Journal of Economic Theory*, 145(1), 85–105.

Damiano, E., L. Hao, and W. Suen (2012): "First in Village or Second in Rome?," *International Economic Review*, 51, 263–288.

DeGroot, M. (1970): *Optimal Statistical Decisions*. McGraw-Hill.

Garey, M., and D. Johnson (1978): "Strong NP-Completeness Results: Motivation, Examples, and Implications," *Journal of the Association for Computing Machinery*, 25, 499–508.

Goel, P., and J. Ginebra (2003): "When Is One Experiment 'Always Better Than' Another One?," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52, 515–537.

HANSEN, O. H., AND E. TORGERSEN (1974): "Comparison of Linear Normal Experiments," *The Annals of Statistics*, 2, 367–373.

HOPPE, H. C., B. MOLDOVANU, AND A. SELA (2009): "The Theory of Assortative Matching Based on Costly Signals," *Review of Economic Studies*, 76(1), 253–281.

HWANG, F., AND U. ROTHBLUM (2011): *Partitions: Optimality and Clustering*. World Scientific.

JEWITT, I. (2007): "Information Order in Decision and Agency Problems," Oxford University Working Paper.

KELSO, ALEXANDER S, J., AND V. P. CRAWFORD (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50, 1483–1504.

LEGROS, P., AND A. NEWMAN (2007): "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102.

LEHMANN, E. L. (1988): "Comparing Location Experiments," *The Annals of Statistics*, 16, 521–533.

LIU, Q., G. J. MAILATH, A. POSTLEWAITE, AND L. SAMUELSON (2012): "Matching with Incomplete Information," .

MARSHAK, J., AND R. RADNER (1972): *Economic Theory of Teams*. Yale University Press, New Haven.

MARSHALL, A., I. OLKIN, AND B. ARNOLD (2009): *Inequalities: Theory of Majorization and its Applications*. Springer, New York.

MERTENS, S. (2006): "The Easiest Hard Problem: Number Partitioning," in *Computational Complexity and Statistical Physics*, ed. by A. Percus, G. Istrate, and C. Moore, pp. 125–139. Oxford University Press, New York.

MEYER, M. (1994): "The Dynamics of Learning for Team Production: Implications for Task Assignments," *The Quarterly Journal of Economics*, 109, 1157–1184.

MOSCARINI, G., AND L. SMITH (2002): "The Law of Large Demand for Information," *Econometrica*, 70(6), 2351–2366.

Murota, K. (2003): *Discrete Convex Analysis*. Siam, Philadelphia.

Olszewski, W., and R. Vohra (2012): "Team Selection Problem," Northwestern University Working Paper.

Persico, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68(1), 135–148.

Prat, A. (2002): "Should a Team be Homogeneous?," *European Economic Review*, 46, 1187–1207.

Pycia, M. (2012): "Stability and Preference Alignment in Matching and Coalition Formation," *Econometrica*, 80, 323–362.

Quah, J. K.-H., and B. Strulovici (2009): "Comparative Statics, Informativeness, and the Interval Dominance Order," *Econometrica*, 77(6), 1949–1992.

Radner, R. (1962): "Team Decision Problems," *The Annals of Mathematical Statistics*, 33, 857–881.

Shaked, M., and Y. Tong (1990): "Comparison of Experiments for a Class of Positively Dependent Random Variables," *The Canadian Journal of Statistics*, 18, 79–86.

Vondrak, J. (2007): "Submodularity in Combinatorial Optimization," Doctoral Thesis, Department of Applied Mathematics, Charles University.

Wang, Y. H. (1993): "On the Number of Success in Independent Trials," *Statistica Sinica*, 3, 295–312.

Weitzman, M. (1998): "The Noah's Ark Problem," *Econometrica*, 66, 1279–1298.

Williams, D. (1991): *Probability with Martingales*. Cambridge.