Theoretical Economics 13 (2018), 377–414 1555-7561/20180377

Matching information

Hector Chade
Department of Economics, Arizona State University

Jan Eeckhout
Department of Economics, University College London and UPF-ICREA-GSE

We analyze the optimal allocation of experts to teams, where experts differ in the precision of their information, and study the assortative matching properties of the resulting assignment. The main insight is that in general it is optimal to diversify the composition of the teams, ruling out positive assortative matching. This diversification leads to negative assortative matching when teams consist of pairs of experts. And when experts' signals are conditionally independent, all teams have similar precision. We also show that if we allow experts to join multiple teams, then it is optimal to allocate them equally across all teams. Finally, we analyze how to endogenize the size of the teams, and we extend the model by introducing heterogeneous firms in which the teams operate.

Keywords. Assortative matching, teams, diversification, correlation.

JEL classification. C78, D83.

1. Introduction

The aggregation of decentralized information is a fundamental source of value creation within firms and organizations. Management heavily relies on the information and judgement of its employees. The influential work by Marshak and Radner (1972) has explored this topic in detail. In their basic formulation, a team is 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.

Hector Chade: hector.chade@asu.edu
Jan Eeckhout: j.eeckhout@ucl.ac.uk

We are grateful to three anonymous referees for their helpful comments. We have benefited from seminar audiences at UCL, Cambridge, Yale, UPF, Arizona State, Caltech, Wisconsin, Brigham Young, Indiana–IUPUI, Latin American Econometric Society Meetings, Society for Economic Dynamics, European Summer Symposium in Economic Theory at Gerzensee, Fundação Getulio Vargas Rio, Sidney, Edinburgh, Adelaide, Louvain, Essex, London Business School, Central European University, ITAM, Oxford, Rutgers, and Bocconi, as well as from comments by Federico Echenique, Ilse Lindenlaub, Alejandro Manelli, Giuseppe Moscarini, Meg Meyer, and Edward Schlee. Eeckhout gratefully acknowledges support from the ERC Grant 339186 and from MINECO-ECO2012-36200, and Chade is grateful to Universitat Pompeu Fabra for its hospitality where part of this research was conducted.

https://doi.org/10.3982/TE1820

In this paper, we embed such teams in a matching framework, and analyze optimal sorting patterns 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. A novel feature of our model is that the characteristic of an agent is the expertise to generate valuable information for a team. This is important in many applications such as research and development collaborations, teams of consultants, and the executive board of corporations. It also provides a rationale for the diversity of worker characteristics commonly observed within firms.

The model consists of a group of agents or experts that must be partitioned into fixed-size teams. Experts differ in the precision of their signals. Within a team, each agent draws a signal about an unknown parameter before making a joint decision. Building on the standard paradigm in team theory, we assume normal distributions and quadratic payoff functions. 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; alternatively, 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. While this precludes potential incentive issues, it allows us to zero in on the impact of information aggregation on matching, which is the novel question we address.

We first provide a closed form solution for a team's value function. It depends on experts' characteristics and correlation parameter through an index that summarizes the information contained in the experts' signals. The index reveals that negatively correlated signals are more informative 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 characteristic adds to the team's precision, this gain needs to be adjusted by how correlated the signal is with those of the other members.

We then analyze the optimal composition of teams. The main insight is that it is generally optimal to *diversify* the composition of the teams. Positive assortative matching (PAM), that is, allocating the best experts to one team, the second tier of best experts to another, and so on, is suboptimal in this setting. In most cases there cannot even be two teams where one contains uniformly better experts than another one. This is because for a wide range of correlation values, the value function is strictly submodular in experts' precision. As a result, there would be profitable swaps of experts between any two ordered teams that will improve efficiency.

When teams contain two experts, diversification leads to negative assortative matching (NAM), so the optimal team composition pairs 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, our model is akin to partition problems that are NP-hard (Garey and Johnson 1978, Vondrák 2007), and thus a succinct characterization of the composition of optimal teams is not possible. We can, however, derive further properties of the optimal matching when signals are conditionally independent, since in this case a team's precision is simply the sum of the precision of its members, and this permits us to obtain sharp results. We show that in this case it is optimal to build teams that are maximally balanced, i.e., team precision tends to be equalized across teams.

We also explore fractional assignment of 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). When signals are conditionally independent, the precision of *all* teams is equalized, which can be accomplished by allocating each expert to every team uniformly (i.e., divide the expert's time equally among all teams). This diversification result is a generalization of NAM, for the team value function is strictly submodular in agents' characteristics. Fractional assignment also affords a simple decentralization of the optimal matching as a Walrasian equilibrium, which reduces 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, which combine 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 can be diversification of expertise, which depends on how spread out firm productivity is: the higher is the spread, the higher is the difference between team precision across teams and, hence, the lower the diversity of expertise within teams.

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 our main assumptions and potential extensions that seem interesting to pursue. In particular, we show that many of the insights extend if we relax the assumption that groups are of equal size.

Related literature

The paper is related to several strands of literature.

Assortative matching Given our focus on sorting, the obvious reference is Becker (1973). The novel features are that agents differ in their signal informativeness, which is relevant in matching settings of economic interest, and the multi-agent nature of teams, unlike the standard pairwise paradigm in matching models with transferable utility. In Kremer (1993), identical firms hire multiple workers, which could be interpreted as teams. The team payoff in Kremer (1993) is multiplicatively separable in the workers' characteristics and, thus, it is strictly supermodular. This implies that PAM is optimal and that, given his large market assumption, each firm hires all agents with the same characteristic. Unlike his setting, in our model the team payoff function is in most cases submodular, and NAM or a generalization thereof ensues. Kelso and Crawford (1982) also analyze multi-agent matching and prove the existence of equilibrium for a class of such models. Here we focus on sorting patterns in a 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 for positive sorting. Similarly, Damiano et al. (2010) analyze group formation and assortative matching with peer effects. Our paper differs in many ways from theirs, e.g., our focus on information and the diversification property.

Teams We also build on the large literature on teams started by Radner (1962) and Marshak and Radner (1972). As in that literature, we abstract from incentive problems. But instead of analyzing a team in isolation with a given information structure, we study teams in a matching setting. Three recent related papers are Prat (2002), Lamberson and Page (2011), and Olszewski and Vohra (2012). Prat (2002) analyzes the optimal composition of a single team, and provides conditions for a homogeneous or a diversified team to be optimal. In his setting, the cost of an information structure for the team is exogenously given. Lamberson and Page (2011) analyze the optimal composition of a team of forecasters, who use weighted averages of the signals to estimate an unknown state, with weights chosen to minimize expected square error. Unlike these papers, we focus on sorting of agents into teams, 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 statics results with respect to the cost of hiring and productivity externalities. 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 there is no 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 the ability of team members. In our model, it also increases efficiency as it equalizes the precision of teams. Finally, Hong and Page (2001) study the optimal composition of a problem solving team, where agents have bounded abilities and apply heuristics to tackle a problem. Our teams do not engage in problem solving but aggregate their information in a Bayesian fashion.

Value of information Since each team runs a multivariate normal experiment, the paper 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 for the index of informativeness in our problem, and study the effects of correlation on the index. Also, we analyze the complementarity properties among signals, and this is related to Börgers et al. (2013), 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' characteristics). 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 characteristics 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.

2. The model

There is a finite set $\mathcal{I} = \{1, 2, \dots, m\}$ of agents. A mapping $x : \mathcal{I} \to [\underline{x}, \overline{x}]$ assigns a "level of expertise" (henceforth characteristic) $x(i) \equiv x_i \in [\underline{x}, \overline{x}]$, $0 < \underline{x} < \overline{x} < \infty$, to each agent $i = 1, 2, \dots, m$. Without loss of generality, we assume $x_1 \leq x_2 \leq \cdots \leq x_m$.

Each agent is allocated to one of N teams of fixed size k, and we assume that $m = kN$. In each team, agents solve a joint decision problem, indexed by an unknown state of nature s. The prior belief agents have about s is given by a density h that is normal with mean μ and precision (inverse of the variance) τ, so the random variable $\tilde{s}$ is distributed normal with mean μ and precision τ, that is, $\tilde{s} \sim \mathcal{N}(\mu, \tau^{-1})$.

Once in a team, an agent with characteristic x_i draws a signal $\tilde{\xi}_i$ from $f(\cdot \mid s, x_i)$ that is normal with mean s and precision x_i; that is, $\tilde{\xi}_i \sim \mathcal{N}(s, x_i^{-1})$. Better experts are those endowed with more precise signals, which are more informative in Blackwell's sense.[1] Conditional on the state s, signals can 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 covariance between any two agents x_i and x_j is given by $\rho(x_i x_j)^{-0.5}$, where $\rho \in (-(k-1)^{-1}, 1)$.[2] An important case that we analyze in detail is $\rho = 0$, conditionally independent signals, which is commonly assumed in normal models of information acquisition.

As in the classical theory of teams, 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 profit level that is reduced by the error in estimating the unknown state.[3] To ensure that a team's expected payoff is positive, we assume that $\pi > \tau^{-1}$.

Formally, a group with characteristics $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$ solves

$$V(\boldsymbol{x}) = \max_{a(\cdot)} \pi - \int \cdots \int \big(a(\boldsymbol{\xi}) - s\big)^2 f(\boldsymbol{\xi} \mid s, \boldsymbol{x}, \rho) h(s) \prod_{i=1}^{k} d\xi_i \, ds,$$

where $a : \mathbb{R}^k \to \mathbb{R}$ is a measurable function, $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_k)$, $f(\cdot \mid s, \boldsymbol{x}, \rho)$ is the joint density of the signals generated by the members of the group, and $V(\boldsymbol{x})$ is the maximum expected payoff for the team. The density $f(\cdot \mid s, \boldsymbol{x}, \rho)$ is multivariate normal with mean given by a $k \times 1$ vector with all coordinates equal to s, and a $k \times k$ covariance matrix Σ_k with diagonal elements $1/x_i$ and off-diagonal elements $\rho(x_i x_j)^{-0.5}$ for all $i \neq j$. The resulting (ex post) payoff is shared among team members via transfers.[4] Agents have linear preferences over consumption, and as a result, utility is fully transferable.

[1] A signal is more informative in Blackwell's sense than another one if the second is a "garbling" of the first. Formally, if $x_i > x_j$, then $\tilde{\xi}_i$ is more informative than $\tilde{\xi}_j$ since $\tilde{\xi}_j = \tilde{\xi}_i + \tilde{\varepsilon}$, where $\tilde{\varepsilon} \sim \mathcal{N}(0, x_j^{-1} - x_i^{-1})$ is independent of $\tilde{\xi}_i$ (Lehmann 1988, p. 522, and Goel and Ginebra 2003, p. 519).

[2] We have multiple random variables that can be pairwise negatively correlated. For a simple example, consider $\tilde{\xi}_1 = \tilde{s} + \tilde{z}_1 + \tilde{z}_2$, $\tilde{\xi}_2 = \tilde{s} - \tilde{z}_1 + \tilde{z}_3$, and $\tilde{\xi}_3 = \tilde{s} - \tilde{z}_2 - \tilde{z}_3$, where $\tilde{z}_i$, $i = 1, 2, 3$, are independent. The lower bound $-1/(k-1)$ ensures that the covariance matrix of any team is positive definite; it is clear that when k is large, we cannot have too much pairwise negative correlation.

[3] The results extend with minor changes to the case in which π depends on N.

[4] A team's ex post payoff can be negative, but this poses no problem as it is expected transfers that matter at the matching stage, and these can be chosen to be nonnegative since $V(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$.

Transferable utility implies that any optimal matching maximizes the sum of the teams' profits. Hence, an optimal matching is the partition of the set of agents into N teams of size k that maximizes $\sum_n V(\boldsymbol{x}_n)$.[5] Formally, let $\Upsilon = \{x_1, x_2, \ldots, x_{kN}\}$ be the multiset of characteristics and let $P(\Upsilon)$ be the set of all partitions of Υ into sub-multisets of size k.[6] The optimal matching problem is

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

The model subsumes several possible interpretations. It could be 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, by a competitive market, or by a chief executive officer if all the teams belong to a single firm. Regarding the state of nature s, it could either be the same for all groups or each team could obtain an independent draw of s from h (e.g., different teams perform different tasks). Finally, all teams are of equal size k, and k is given. This simply extends the pairwise assumption made in most of the matching literature. We later discuss relaxing this assumption.

3. Correlation, informativeness, and diversification

There are two stages in the model: the *team formation* stage, where agents sort into N teams each of size k, and the *information aggregation* stage, in which team members pool their information and choose an action. We proceed backward and analyze first the information aggregation problem and then the sorting of agents into teams.

3.1 *The team's decision problem and value function*

Consider a team with vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$. After observing the signal realizations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$, the team updates its beliefs about the state s. Since the prior distribution of s is normal and so is the distribution of the signals, it follows that the posterior distribution of s is also normal and is denoted by $h(\cdot \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho)$. Then the team solves

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

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

$$a^*(\boldsymbol{\xi}) = \int s h(s \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho)\, ds = \mathbb{E}[\tilde{s} \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho].$$

[5] For simplicity, in the text we use $\sum_n$ in place of $\sum_{n=1}^N$, $\sum_i$ in place of $\sum_{i=1}^k$, and so on.
[6] A multiset is a generalization of a set that allows for repetition of its members.

Inserting $a^*(\boldsymbol{\xi})$ into the team's objective function, we obtain

$$\begin{aligned} V(\boldsymbol{x}) &= \pi - \int \cdots \int \big(\mathbb{E}[\tilde{s} \mid \boldsymbol{\xi}, \boldsymbol{x}] - s\big)^2 f(\boldsymbol{\xi} \mid s, \boldsymbol{x}, \rho) h(s) \prod_{i=1}^{k} d\xi_i \, ds \\ &= \pi - \int \cdots \int \left(\int \big(s - \mathbb{E}[\tilde{s} \mid \boldsymbol{\xi}, \boldsymbol{x}]\big)^2 h(s \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho) \, ds \right) f(\boldsymbol{\xi} \mid \boldsymbol{x}, \rho) \prod_{i=1}^{k} d\xi_i \\ &= \pi - \int \cdots \int \operatorname{Var}(\tilde{s} \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho) f(\boldsymbol{\xi} \mid \boldsymbol{x}, \rho) \prod_{i=1}^{k} d\xi_i, \end{aligned}$$

where $f(\boldsymbol{\xi} \mid \boldsymbol{x}, \rho) \equiv \int f(\boldsymbol{\xi} \mid s, \boldsymbol{x}, \rho) h(s) \, ds$. The second equality uses $h(s \mid \boldsymbol{\xi}, \boldsymbol{x}, \rho) f(\boldsymbol{\xi} \mid \boldsymbol{x}, \rho) = h(s) f(\boldsymbol{\xi} \mid s, \boldsymbol{x}, \rho)$, and the third equality 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 well known that when signals are identically distributed normal and conditionally independent ($\rho = 0$), then the posterior precision is deterministic (it does not depend on $\boldsymbol{\xi}$) and it is equal to the sum of the prior's and the signals' precisions (see De Groot 1970, p. 167). The next proposition extends this result to our more general setting.

Proposition 1 (Value of a team). *The value of a team with characteristics $\boldsymbol{x}$ is*

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

where

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

The proofs of all the results are provided in the Appendix.[7] As illustrations of (1)–(2), notice that in the conditionally independent case, $\mathcal{B}(\boldsymbol{x}, 0) = \sum_i x_i$ and, thus,

$$V(\boldsymbol{x}) = \pi - \frac{1}{\tau + \sum_{i=1}^{k} x_i}.$$

Also, when $k = 2$, then $\mathcal{B}(\boldsymbol{x}, \rho) = (x_1 + x_2 - 2\rho(x_1 x_2)^{0.5})/(1 - \rho^2)$. And if $x_1 = x_2 = \cdots = x_k = x$, then $\mathcal{B}(\boldsymbol{x}, \rho) = kx/(1 + \rho(k-1))$.

[7] The proof proceeds by induction after obtaining the general functional form of the inverse of the covariance matrix. Another route is first to show that the optimal decision is a weighted average of the signals, then obtain the general form of the inverse of a covariance matrix, and finally appeal to Theorem 2 in Lamberson and Page (2011). A third route is simply to adapt formula 3.19 on p. 128 stated without proof in Figueiredo (2004) to our setup. We provide a simple proof for completeness.

3.2 *Correlation and informativeness*

The function $\mathcal{B}(\boldsymbol{x},\rho)$ is the index of informativeness of the vector of signals $\boldsymbol{\xi}$ drawn from a multivariate normal distribution centered at s and with covariance matrix Σ_k. The higher is the value of $\mathcal{B}(\boldsymbol{x},\rho)$, the more informative is $\boldsymbol{\xi}$ in Blackwell's sense.[8] An interesting question to explore is how correlation affects the informativeness of a team. The following proposition contains the most important properties of $\mathcal{B}$.

PROPOSITION 2 (Team precision). *The function $\mathcal{B}$ satisfies the following properties:*

(i) It is positive for all $(\boldsymbol{x},\rho)$.

(ii) The addition of a signal increases $\mathcal{B}$ for all $(\boldsymbol{x},\rho)$.

(iii) It is strictly convex in ρ for each $\boldsymbol{x}$, with $\arg\min_\rho \mathcal{B}(\boldsymbol{x},\rho)>0$ *for all* $\boldsymbol{x}$.

The first two properties are intuitive: since $\underline{x}>0$, the precision of any signal is positive and so is $\mathcal{B}$, which yields property (i). And adding a signal cannot decrease team precision (at worst it can be ignored); hence property (ii) should hold.

More interesting is the convexity of $\mathcal{B}$ in ρ and that its minimum is at a *positive* value of ρ. In particular, it implies that $\mathcal{B}(\boldsymbol{x},0)>\min_\rho \mathcal{B}(\boldsymbol{x},\rho)$ and thus conditionally independent signals are *less* informative than negatively correlated ones, but *more* informative than positively correlated ones in an interval of positive values of ρ.[9] Since this is an issue that has received some attention in the statistical literature in the context of *equal* precision normal signals (see Shaked and Tong 1990), we discuss it a bit further.

Suppose $k=2$ and $x_1=x_2=x$. Then $\mathcal{B}(\boldsymbol{x},\rho)=2x/(1+\rho)$, and thus $2x/(1+\rho)>2x$ if and only if $\rho<0$. Moreover, $\mathcal{B}(\boldsymbol{x},\rho)$ is decreasing in ρ. This is easiest to see in the extreme cases: if $\rho=1$, then observing the second signal is useless, while it is valuable under independence, and if $\rho=-1$, the second signal is infinitely informative, as it reveals the state. Consider $\rho\in(-1,1)$. The conditional distribution of $\tilde{\xi}_2$ given ξ_1 and s is $\mathcal{N}((1-\rho)s+\rho\xi_1,(1-\rho^2)/x)$. Both positive and negative correlation reduce the variance of the second signal. But negative correlation makes the mean more "sensitive" to s than positive correlation, making the second signal more informative about s.

Consider $k=2$ but now with an *arbitrary* $\boldsymbol{x}$, so $\mathcal{B}(\boldsymbol{x},\rho)=(x_1+x_2-2\rho(x_1x_2)^{0.5})/(1-\rho^2)$. Then $\mathcal{B}(\boldsymbol{x},\rho)>\mathcal{B}(\boldsymbol{x},0)$ if $\rho<0$, and this also holds when $\rho>0$ if and only if $\rho>\hat{\rho}=(x_1x_2)^{0.5}/((x_1+x_2)/2)$, where the right side is less than 1 by the arithmetic–geometric mean inequality (with equality if and only if $x_1=x_2$). Moreover, now $\mathcal{B}$ reaches a minimum at $\hat{\rho}$, displaying a U shape unlike the constant precision case. The explanation is that now 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. In the above example, the conditional distribution of the second signal is

[8] This follows from Theorem 2 in the survey by Goel and Ginebra (2003, p. 521), which is a celebrated result by Hansen and Torgersen (1974). See Appendix A.3 for details.

[9] For a simple illustration, let $\tilde{\theta}_1=\tilde{s}+\tilde{z}$ and $\tilde{\theta}_2=\tilde{s}-\tilde{z}$, with $\mathbb{E}[\tilde{z}]=0$. Conditional on s, they reveal the state as $\tilde{\theta}_1+\tilde{\theta}_2=s$. This would not happen if $\tilde{\theta}_2=\tilde{s}+\tilde{z}$ or with different and independent $\tilde{z}$s.

$\mathcal{N}((1-(x_1/x_2)^{0.5}\rho)s+\rho\xi_1,(1-\rho^2)/x_2)$, where the mean continues to depend on s even when $\rho=1$ and the variance goes to 0. Hence, one can perfectly infer the state with perfect positively correlated signals and heterogeneous experts, something that does not occur with $0<\rho<1$.

3.3 *Diversification of expertise across teams*

We now turn to the team formation stage and use the properties of the team value function (1) to shed light on optimal sorting. We begin with a crucial lemma.

LEMMA 1 (Value function properties). *Consider a team with characteristics $\boldsymbol{x}$:*

(i) If $\rho=0$, then V is strictly submodular in $\boldsymbol{x}$.

(ii) If $\rho<0$, then V is strictly submodular in $\boldsymbol{x}$ if $\tau\leq\mathcal{B}(\underline{\boldsymbol{x}})$ and $\overline{x}\leq 16\underline{x}$.

(iii) If $\rho>0$, then V is strictly submodular in $\boldsymbol{x}$ if $\rho\leq 1/((k-1)(\overline{x}/\underline{x})^{0.5}-(k-2))$.

Lemma 1 reveals that the value function is submodular in the characteristic of the team members in many cases of interest. Part (i) asserts that this is true in the conditional independent case, and by continuity it is so for a small amount of correlation. Regarding part (ii), it shows that under negative correlation, V is submodular when the value of experts' information is higher than the precision of the prior, which is arguably the most relevant case. It also contains a restriction on the domain of the precision of each expert (the proof provides a weaker bound for $\overline{x}$ that depends also on k, and that in the standard case in matching with $k=2$ is equal to infinity). Finally, part (iii) provides an upper bound on ρ under positive correlation below which V is strictly submodular.[10]

We will say that the optimal matching exhibits *diversification* if no two teams are ordered in the sense that all experts in one team have higher characteristics than those of the other team. In particular, the presence of diversification implies that positive assortative matching (PAM)—i.e., the best k experts are assigned to one team, the next best k to another, and so on, and hence *all* teams are ordered by expertise—cannot be the optimal sorting pattern. In the case of $k=2$, optimal diversification is equivalent to negative assortative matching (NAM).

PROPOSITION 3 (Optimality of diversification). *Under the conditions in Lemma 1, the optimal matching exhibits diversification and it is NAM if $k=2$.*

This result follows from Lemma 1. Given any partition with two *ordered* teams, one can find a swapping of two experts, one from each team, that strictly increases $\sum_n V(\boldsymbol{x}_n)$

[10]For an intuitive explanation of the bounds in (ii) and (iii), notice that V is concave in $\mathcal{B}$, which is a force toward submodularity. If $\rho<0$, then $\mathcal{B}$ is increasing and supermodular in $\boldsymbol{x}$, which is a force against it. The bound in (ii) ensures that the first effect prevails. If $\rho>0$, then $\mathcal{B}$ is submodular, and it is increasing if the bound in (iii) holds. In that case, the effects of V and $\mathcal{B}$ reinforce each other.

by strict submodularity. And if $k = 2$, it is well known (Becker 1973) that submodularity implies NAM. Hence, matching leads to a balanced expertise assignment across teams.

A natural question is whether, in the cases not covered by Lemma 1, V can be supermodular and thus optimal matching exhibits PAM. The answer is no when $\rho > 0$, for in this case V *cannot* be supermodular: there is always a subset of $[\underline{x}, \overline{x}]^k$ around the "diagonal" $x_1 = x_2 = \cdots = x_k$ where the team value function is strictly submodular. Indeed, in this case, V is strictly submodular in $\boldsymbol{x}$ if $\overline{x} - \underline{x}$ is sufficiently small, so there is not much heterogeneity among experts. If $\rho < 0$, however, then for any given value of ρ, the function V is supermodular in $\boldsymbol{x}$ if τ is large enough, in which case we obtain PAM. But we show in Appendix A.6 that τ must be unboundedly large as ρ converges to 0 or $-1/(k-1)$, and that for any $\rho \in (-1/(k-1), 0)$, a necessary condition for V to be supermodular is that τ be strictly larger than $10k\overline{x}$, i.e., more than 10 *times* the precision of the *best* team possible. That is, PAM can ensue in this case *only if* prior precision is much larger than that of any team, and thus experts are of little value.

3.4 *Properties of optimal teams*

We have focused on the optimality of diversification but have been careful not to assert that NAM is optimal, except for $k = 2$. The reason is that it is unclear how to define it when $k > 2$ (unlike PAM, which extends straightforwardly). Moreover, we will see below that computing the entire matching is intractable when $k > 2$. The best we can do in Proposition 3 is to assert that the optimal matching will *not* contain *ordered teams*, i.e., diversification, which rules out PAM. There are, however, important classes of problems for which we can say much more about the properties of optimal teams.

3.4.1 *Size-two teams and negative sorting* An important instance we can fully solve is the case most matching models focus on, namely, pairwise matching or $k = 2$. Then NAM is optimal whenever V is submodular (Becker 1973). That is, given any four agents with characteristics $x_1 > x_2 \geq x_3 > x_4$, total payoff is maximized if x_1 is matched with x_4 and x_2 with x_3, something that easily follows from submodularity. The computation of the optimal matching is straightforward in this case: match x_i with x_{2N-i+1} for each $i = 1, 2, \ldots, N$.

Moreover, the conditions in Lemma 1 are easy to obtain. Recall that in this case $\mathcal{B}(x_1, x_2) = (x_1 + x_2 - 2\rho(x_1x_2)^{0.5})/(1-\rho^2) = 2(\mathrm{AM}(x_1, x_2) - \rho\,\mathrm{GM}(x_1, x_2)/(1-\rho^2)$, where $\mathrm{AM}(x_1, x_2) = (x_1 + x_2)/2$ is the arithmetic mean, and $\mathrm{GM}(x_1, x_2) = (x_1x_2)^{0.5}$ is the geometric mean. Inserting this expression into $V(x_1, x_2) = \pi - (1/(\tau + \mathcal{B}(x_1, x_2))$ and differentiating, reveals that the sign of the cross-partial derivative of V is equal to the sign of the expression

$$3\rho\,\mathrm{AM}(x_1, x_2) - \left(2 + \rho^2\right)\mathrm{GM}(x_1, x_2) - 0.5\rho\left(1 - \rho^2\right)\tau. \tag{3}$$

If $\rho < 0$, then it is easy to check that if $\tau \leq 3\mathcal{B}(\underline{x}_1, \underline{x}_2)$ (and there is no restriction on the support), then the sign of (3) is negative, thereby weakening Lemma 1(ii).[11] In turn,

[11] To see this, simply replace τ by $3\mathcal{B}(x_1, x_2)$ in (3), which results in a negative expression.

if $\rho > 0$, then it suffices to ignore the last term in (3) and provide a bound on ρ such that $3\rho\,\mathrm{AM}(x_1, x_2) - (2+\rho^2)\,\mathrm{GM}(x_1, x_2) \leq 0$ (recall that AM $\geq$ GM with strict inequality unless $x_1 = x_2$). The worst case for a negative sum is when we evaluate AM and GM at $(\underline{x}, \overline{x})$. Simple algebra reveals that in this case NAM obtains if $\rho \leq (0.75(1+R^2) - 0.25(9-14R^2+9R^4)^{0.5})/R$, where $R = (\overline{x}/\underline{x})^{0.5}$. Alternatively, one can use a similar argument as in the case of $\rho < 0$ and show that NAM ensues if $\tau \geq 3\mathcal{B}(x_1, x_2)$ for all (x_1, x_2).

It is also well known that the optimal matching can be decentralized as the allocation of a competitive equilibrium (e.g., see Chade et al. 2017). The following procedure is a standard way to derive the wage function that supports μ as an equilibrium, where our endogenous V pins down its properties. The simplest way to describe it is to assume a large market with a measure m of agents and $N = m/2$ of teams. Let x be distributed with cumulative distribution function (cdf) G and continuous and positive density g. Then NAM can be described by a function μ that solves $G(\mu(x)) = 1 - G(x)$ for all x. Posit a wage function w that agents take as given when "choosing" a partner. Each x solves $\max_{x'} V(x, x') - w(x')$, with first-order condition $V_2(x, x') = w(x')$, where V_2 is the derivative with respect to the second argument. In equilibrium, this holds at μ, so $V_2(\mu(x), x) = w'(x)$ for all x, which yields

$$w(x) = w(\underline{x}) + \int_{\underline{x}}^{x} V_2\big(\mu(s), s\big)\, ds.$$

The pair (μ, w) is the competitive equilibrium with NAM. Since $V_2(\mu(x), x) > 0$ for all x when $\rho \leq 0$, it follows that wages are higher for experts with higher characteristics. If $\rho > 0$, however, $V_2(\mu(x), x) < 0$ for low values of x and positive for higher values. Under conditional independence, one can easily show that wages are linear if f is symmetric, and convex if it is decreasing in x (i.e., if experts with higher characteristics are scarcer).

3.4.2 *Conditionally independent signals and balanced teams* A standard assumption in applications with information acquisition is that signals are conditionally independent. We now describe the properties of optimal teams in this case.

The team value function is $V(\boldsymbol{x}) = \pi - (\tau + \sum_i x_i)^{-1}$, which is strictly submodular in $\boldsymbol{x}$, and it is strictly concave in $\sum_i x_i$. For clarity, we set $v(\sum_i x_i) \equiv V(\boldsymbol{x})$ and denote the precision of team n by $X_n \equiv \sum_i x_{in}$. Any partition of the experts yields a vector $(X_1, X_2, \ldots, X_N)$, and hence we seek the one that maximizes $\sum_n v(X_n)$.

In this case *all* partitions have the same sum $\sum_n X_n = X$, so the problem is akin to a welfare maximization problem where a planner allocates an "aggregate endowment" X among N identical "consumers," equally weighted.

If X could be continuously divided, then the problem would reduce to finding $(X_1, X_2, \ldots, X_N)$ to maximize $\sum_n v(X_n)$ subject to $\sum_n X_n = X$. And since the objective function is strictly concave, the solution would be $X_n = X/N$ for all n, so all teams would have the same precision.[12] Clearly, this depends on X being continuously divisible, but a similar insight obtains in the discrete case under independence.

[12] 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.

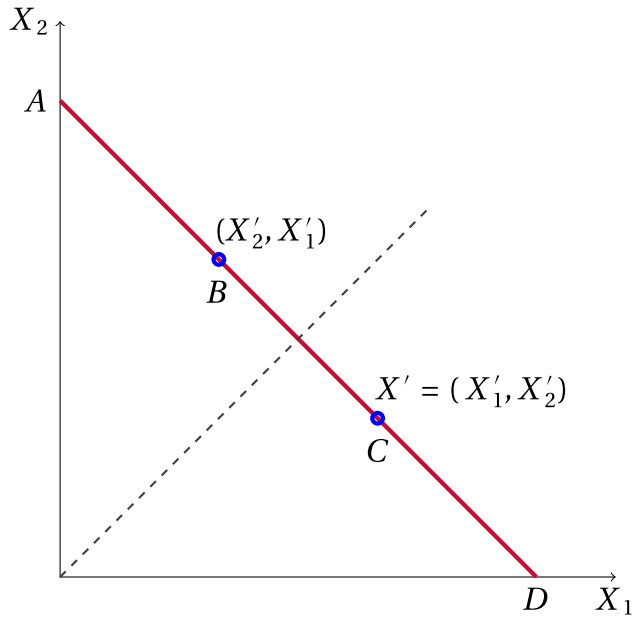

FIGURE 1. Geometry of majorization. The figure depicts vector $X' = (X_1', X_2')$ (and its mirror image (X_2', X_1')) whose sum equals that of any point along the line AD. Vectors on the segments AB and CD, that is, vectors that "pull away" from the 45 degree line, majorize X', and any vector on BC is majorized by X'. The closer a vector is to the 45 degree line, the higher the value is of any Schur concave function.

We need the following concepts. A vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ *majorizes* a vector $\boldsymbol{X}' = (X_1', X_2', \ldots, X_N')$ if $\sum_{\ell=1}^{m} X_{[\ell]} \geq \sum_{\ell=1}^{m} X'_{[\ell]}$ for $m = 1, 2, \ldots, N$, with $\sum_{\ell=1}^{N} X_{[\ell]} = \sum_{\ell=1}^{N} X'_{[\ell]}$, and where $X_{[\ell]}$ is the ℓth largest coordinate of $\boldsymbol{X}$ (see Figure 1 for an illustration of majorization). Also, a function $f : \mathbb{R}^N \to \mathbb{R}$ is *Schur concave* if $f(\boldsymbol{X}') \geq f(\boldsymbol{X})$ whenever $\boldsymbol{X}$ majorizes $\boldsymbol{X}'$, and it is *strictly Schur concave* if the inequality is strict (see Marshall et al. 2010).

Any partition generates a vector of team precisions $(X_1, X_2, \ldots, X_N)$. Let Γ be the set of such vectors partially ordered by majorization, which is a notion of how "spread out" a vector is. Then $\sum_n v(X_n)$ is strictly Schur concave on Γ, as it is the sum of strictly concave functions v (Marshall et al. 2010, Proposition C.1, p. 92).

It follows that if we compare the team precision vectors of two partitions of the experts and one majorizes the other, then the planner prefers the majorized one (by Schur concavity). Intuitively, since all partitions have the same "mean" and the majorized partition has less "spread," a planner with a "concave utility" function is better off with it. 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. This suggests that the optimal team structure *minimizes the spread* in the precision of the teams, thereby making them as balanced as possible.[13]

[13]That teams are diversified does not mean that we cannot find one with homogeneous members. For instance, consider six agents with characteristics 1, 4, 5, 5, 5, and 10, and let $k = 3$. Then the optimal

PROPOSITION 4 (Maximally balanced teams). *Assume conditional independence.*

(i) The optimal matching must be an element of the set of partitions whose team precision vectors $(X_1, X_2, \ldots, X_N)$ *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.

An implication of this result is that if there is a partition with $X_n = X/N$ for all n, then it is optimal, for the vector $(X/N, X/N, \ldots, X/N)$ is majorized by *all* partitions.

Although vectors majorized by every other vector 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 2) 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 ordered. 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 is that finding a polynomial-time algorithm to construct an optimal partition is a futile task except when $k = 2$. The problem requires finding a partition where team precisions are equalized as much as possible. When $N = 2$, this reduces to the 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).[14] When $N = 3$, this is like the three-partition problem, which is strong NP-complete (Garey and Johnson 1978, p. 504), and so on for $N > 3$.

A natural question is whether the optimal matching can be decentralized. As we saw, the answer is yes if $k = 2$, and in the next section on fractional assignment, we show that this is indeed the case as well. But we do not have an analogous result when $k > 2$ and each agent is assigned to only one group. Our setup is a special case of 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. We show in Appendix A.9, however, that our model fails their crucial gross substitutes (GS) condition, and thus we cannot appeal to their results.[15] We conjecture that in our setup where teams have identical payoff functions, the optimal matching can be decentralized. But this is a nontrivial task that we leave for future research.

partition is $\{1, 4, 10\}$ and $\{5, 5, 5\}$, and the second team contains homogeneous agents. But notice that (a) the two teams are not "ordered," and (b) generically, agents will have different characteristics.

[14] Garey and Johnson (1978, p. 499) provide an optimal dynamic programming algorithm.

[15] In words, GS asserts that if the wages of experts of different characteristics weakly increase, then a firm or partnership will still find it optimal to hire those experts to whom they made offers 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 byproduct 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 this is not true in our setting.

3.4.3 *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, and similarly for researchers. Applying this to our setting, we now assume that agents can be fractionally assigned to teams.[16]

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 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, her signals are independent in each team and each has half the precision it would have if she worked full time for one of them. To avoid information spillovers, we assume there is an independent draw of the state of nature in each team (e.g., each team works on a different task). Finally, we assume that $\rho = 0$, so signals are conditionally independent. These assumptions deliver clean results and make the model amenable to interesting extensions and variations.

It will be helpful to slightly change and abuse the notation: let $x(\mathcal{I}) = \{x_1, x_2, \ldots, x_J\}$ be the set of distinct characteristics of agents, and denote by m_j the number of agents of characteristic x_j, so that $\sum_j m_j = kN$ and $X = \sum_j m_j x_j$. Denote by $\mu_{jn} \geq 0$ the fractional assignment of characteristic-j agents to team n. Feasibility requires that $\sum_j \mu_{jn} = k$ for every n (i.e., the sum of the allocations of characteristics to team n must equal the fixed team size k), and $\sum_n \mu_{jn} = m_j$ for every j (i.e., the number of x_j agents allocated to all the teams must equal m_j, the total number of them in the population). Finally, let $X_n = \sum_j \mu_{jn} x_j$ be the precision of team n.

The fractional assignment problem is

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

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

As in the integer case, all partitions of agents have the same total sum X. But fractional assignment allows for continuous division of X, which yields the following result.

Proposition 5 (Perfect diversification). *Assume conditional independence. Then any optimal matching equalizes team precision across teams, i.e., $X_n = X/N$ for all n. An optimal solution is to allocate an equal fraction of each expert's characteristic to each team, i.e., $\mu_{jn} = m_j/N$ for all j, n.*

To gain some insight on this result, notice that a less constrained version of the fractional assignment problem is to maximize the objective function $\sum_n v(\sum_j \mu_{jn} x_j) = \sum_n v(X_n)$ subject only to $\sum_j \sum_n \mu_{jn} x_j = X$, which reduces to $\sum_n X_n = X$. Since v is

[16]This assumption is not without precedent in the literature on assignment games, as noninteger 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.

strictly concave, the planner wants to minimize the spread in precision across teams, and thus the unique solution is to set $X_n = X/N$ for all n. The additional constraints do not affect this property of the solution, and indeed it is easy to check that $\mu_{jn} = m_j/N$ for all j, n satisfies all the constraints and achieves the equal precision property of teams. This is actually the unique symmetric solution of the problem, which entails *perfect diversification* of expertise across teams, as they have exactly the same composition.[17]

Proposition 5 uses the conditional independence assumption in important ways, especially in the equal-precision property, which does not hold with correlation. The perfect diversification result, however, extends to the case with *negative* correlation. The details are given in the Appendix, but the logic is simple. When ρ is negative, $\mathcal{B}$ is concave in the fractions of each agent in the team, and since V is strictly increasing and strictly concave in $\mathcal{B}$, it follows that V is concave in the fractions as well. Given the constraints and the symmetry of the problem, it follows that distributing each agent in an equal fraction to each team solves the planner's problem. (If $\rho > 0$ the concavity of V in the fractions is lost and little can be said about the optimal fractional assignment.)

Allowing for fractional assignment makes the decentralization of the optimal matching straightforward under independence. Let there be competitive prices $(w_1, w_2, \dots, w_J)$ for different characteristics of experts. Consider the interpretation of identical firms hiring teams. Then each firm n solves the concave problem

$$\max_{\{\mu_{jn}\}_{j,n}} v\left(\sum_j \mu_{jn} x_j\right) - \sum_j \mu_{jn} w_j$$

$$\text{s.t.} \quad \sum_j \mu_{jn} = k, \qquad \mu_{jn} \geq 0 \quad \forall j.$$

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

$$v'\left(\sum_j \mu_{jn} x_j\right) x_j - w_j - \phi_n = 0 \quad \forall j, n,$$

where ϕ_n is the Langrange multiplier for firm n. These conditions are the same as the planner's once w_j is substituted by the planner's Langrange multiplier for this characteristic. It readily follows that the planner's and the decentralized allocations coincide.

Under the identical firms interpretation in a competitive market, we can also endogenize the size of the teams once we allow for free entry of firms.

Proposition 6 (Endogenous team size). *Assume $\rho = 0$ and fractional assignment, and let identical firms enter a competitive market by paying a fixed cost F, with*

[17] The solution in terms of team precision $X_n = X/N$ is unique. But unless there are only two characteristics, $\mu_{jn} = m_j/N$ 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 characteristic 1–4 and two agents with characteristic 5. Thus, $X = 20$. Both $\{\{1, 4, 5\}, \{2, 3, 5\}\}$ and $m_j/2$ yield $X_1 = X_2 = 10$. Multiplicity is due to the linearity of the constraints and because the experts' precision in a team are perfect substitutes. Also, the proof of Proposition 5 makes clear that fractional assignment allows for unequal size groups (not imposing the $\sum_j \mu_{jn} = k$ constraint), something that is hard to obtain with integer assignment.

$\pi - (1/\tau) < F < \pi$, which can hire workers at wages $(w_1, w_2, \dots, w_J)$ without a size constraint on teams. Then there is a unique equilibrium team size k^*, which is strictly increasing in F and strictly decreasing in τ and X.

To avoid the trivial cases with no firm or an infinite number of them, we assume $\pi - (1/\tau) < F < \pi$.[18] We show in Appendix A.12 that there is a unique equilibrium number of active $N^\star > 0$. As the number of experts m is fixed, we obtain a unique equilibrium team size $k^\star = m/N^\star$. A higher entry cost F implies that firms require higher post-entry profits. Thus, fewer firms enter the market, which leads to a larger team size 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. An increase in prior precision τ reduces wages and increases the revenue $v(X/N)$, which induces entry and and thus a lower equilibrium team size. A similar logic applies to an increase in aggregate precision X.

4. Heterogeneous firms, PAM, and diversification

One interpretation of the model is that of a market where identical firms compete to form teams of experts. In reality, firms are likely to be heterogeneous as well and solve problems of varying economic impact. For example, consulting firms that differ in their reputation consult for clients who are also heterogeneous. The value of expertise differs at those firms and so will their demand for experts, thus affecting sorting.

We now analyze such an extension. For simplicity, we assume conditional independent signals and fractional assignment, but the main insights hold more generally. There are N heterogeneous firms. Let y_i be the characteristic (e.g., productivity or technology) of firm i, with $0 < y_1 \le y_2 \le \cdots \le y_N$. Each firm matches with a team of size k. If a firm with characteristic y matches with a team with precision $\sum_i x_i$, then the expected payoff from the match is $yv(\sum_i x_i)$. The optimal matching problem is to partition the experts into k-size teams and assign them to the firms to maximize $\sum_n y_n v(\sum_i x_{in})$.

Since $yv(\sum_i x_i)$ is supermodular in $(y, \sum_i x_i)$, it is clear that it is optimal to match better firms with higher precision teams. And since it is submodular in experts' characteristics, there can still be diversification of expertise across teams. This insight is stated formally as follows.

Proposition 7 (Heterogeneous firms). *The optimal matching entails PAM between firm quality y_n and team precision X_n, $n = 1, 2, \dots, N$, with X_n given by*

$$X_n = \frac{y_n^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}} (\tau N + X) - \tau. \tag{4}$$

Moreover, there exists a fractional assignment rule $\{\mu_{jn}\}_{j,n}$ that yields (4) for all n.

[18] An alternative to free entry that yields the same insights is to assume that there is a cost of forming teams, c, that is strictly increasing and convex in N.

The optimal matching of firm quality and team precision solves the problem

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

The first-order conditions are $y_n v'(X_n) = y_{n'} v'(X_{n'})$ for all $n \neq n'$, which yields (4). For some intuition on its solution, we note that this problem is equivalent to a welfare maximization problem where a planner allocates an "aggregate endowment" X among N consumers, with weight y_n on consumer n.

A lot can be learned from (4) (see the details in Appendix A.13). First, unless $y_n = y$ for all n, in which case $X_n = X/N$, the optimal solution is increasing in n, so *higher characteristic firms match with higher precision teams.*[19] Second, 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. Third, an increase in prior precision τ *increases* X_n for high values of 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. Finally, if the y_ns become more "spread out" (in a precise sense related to majorization), then team precision *increases* for better ranked firms and *decreases* for lesser ranked ones. The intuition is similar to that given above for the planner's analogy and the changes in the marginal rate of substitution.

All these insights are based on the optimal team precision for the teams. To complete the analysis, we prove that there exists a fractional assignment rule $\{\mu_{jn}\}_{j,n}$ that implements (4). With heterogeneous firms, the assignment differs from perfect diversification.

For an illustrative example, consider two firms and four experts who match in pairs with the firms. Formally, $N = 2$, $k = 2$, $y_1 = 1$, $y_2 = y \geq 1$, there are two agents with characteristic $x_1 = 5$, and two with $x_2 = 20$. For simplicity, assume $\tau = 0$. From (4) we obtain $X_2 = (y^{0.5}/(y^{0.5} + 1))50$ and $X_1 = (1/(y^{0.5} + 1))50$, so $X_2 \geq X_1$. As y increases, the composition of teams ranges from NAM to PAM with different degrees of diversification. More precisely, if $y = 1$, then $X_2 = X_1$ and NAM is optimal, i.e., $\{5, 20\}$, $\{5, 20\}$; if $y \geq 16$, then $X_2 = 40$, $X_1 = 10$, and PAM is optimal, i.e., $\{5, 5\}$, $\{20, 20\}$; and if $1 \leq y < 16$, there is diversification within groups. The fractional assignment rule is $(\mu_{1n}, \mu_{2n}) = ((40 - X_n)/15, (X_n - 10)/15)$, $n = 1, 2$.

[19]The opposite would hold if match payoff were submodular in firm characteristic and team precision. An easy way to see this is to assume that it is $v(\sum_i x_i)/z$, $z > 0$, and replace y_i by $1/z_i$ in (4).

5. Discussion and concluding remarks

Many important economic applications entail the formation of teams composed of 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 calls for diversification of expertise: no team can have uniformly better experts than another team, which rules out PAM. In the case of pairwise matching, the optimal matching is NAM, while under conditional independence, diversification leads to teams that are maximally balanced, which takes the extreme form of equal-precision teams when experts can be fractionally assigned. We also explored the role of correlation on the informativeness of a team, the decentralization of the optimal matching, and endogenous team size. Finally, we analyzed the implications of adding another heterogeneous side of the market, i.e., firms that differ in their quality, and showed that the optimal sorting pattern exhibits a combination of PAM between firm quality and team precision with diversification within teams.

We close with some comments on robustness and describe some open problems.

Alternative information models

We build on a canonical model of information that plays a central role in statistical decision theory, i.e., the normal-prior–normal-signals model with quadratic payoff.[20] This setup features prominently in the economic literature on teams, networks, and global games, in part due to its tractability.[21] Another common 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, an agent's characteristic x_i is now 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 is concave in n (for example, the informative signals are drawn from a normal distribution centered at s with precision κ). If $k=2$, then the team value function is

$$V(x_1,x_2)=x_1x_2u(2)+\big(x_1(1-x_2)+x_2(1-x_1)\big)u(1)+(1-x_1)(1-x_2)u(0),$$

which is clearly strictly submodular in (x_1,x_2). Using Poisson's binomial distribution (Wang 1993), we show in Appendix A.14 that this is true for any k. Therefore, diversification is optimal in this alternative model as well.[22]

[20] It is easy to show that the results extends to a payoff function $\pi-(a-s)^n$ with n even, and we conjecture that they hold for a larger class of strictly decreasing and strictly concave functions of $a-s$.

[21] For a couple of representative contributions to networks and global games, see Ballester et al. (2006), Angeletos and Pavan (2007), and the references therein.

[22] As we note in Appendix A.14, 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.

Going beyond the analysis of canonical models is hard, as we lose tractability and, more importantly, we do not know much in general about curvature properties of the value of information. Indeed, an important feature of our model is that the value of a team is pinned down by the variance of the posterior beliefs, and that is concave in $\mathcal{B}$ and independent of the realization of the signals. This allows us obtain V in closed form. In turn, the Gaussian structure of the signals led to a tractable functional form for $\mathcal{B}$, whose complementarity properties are easy to derive. Similarly, in the variation above it comes in handy that u is independent of the experts' characteristics and probabilities are multiplicative in them. 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, and not much is known about complementarity properties of signals in such problems. Until there is more progress on this issue, a general analysis that covers a larger class of models will remain elusive.[23] As an illustration of the limits of the analysis, we construct two examples with PAM in Appendix A.15: one where utility is multiplicative in action and the state; the other with multidimensional actions.[24] In both cases the information structure and payoff function create complementarities among experts.

Different group sizes

We assume that all groups must be of size k. As mentioned, this is a generalization of the standard assumption in assignment games where agents match in pairs. We use this assumption 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. Most of the other propositions do not use this assumption in a crucial way. Clearly, Propositions 1 and 2 do not depend on it, as they apply to any given team. Proposition 4 is also independent of group size, since it is stated in terms of team precision, which equals the sum of the precision of its members. Its interpretation, however, can change if teams could have different group sizes. As an illustration, consider $\rho = 0$ and six agents, with characteristics 2, 2, 7, 7, 8, and 10. If $k = 3$, then the optimal partition is $\{2, 7, 10\}$ and $\{2, 7, 8\}$, with $X_1 = 19$ and $X_2 = 17$. If we allow the two groups to be of different size, then the optimal partition is $\{2, 2, 7, 7\}$ and $\{8, 10\}$, with $X_1 = X_2 = 18$, a strict improvement. The partition is consecutive in the sense that it puts agents with high characteristics in one team and agents with low characteristics in the other, and *diversification* occurs via the size of each group (the group with better characteristics is smaller). It would be interesting to know if the property of this example holds in general, although available results on partitioning problems with optimal consecutive partitions do not apply to our setup (see Hwang and Rothblum 2011). Notice, however, that this issue is irrelevant under fractional assignment, since perfect diversification yields $X_1 = X_2 = 18$. More

[23] If one restricts attention to a quadratic payoff and decisions that are weighted averages of the signals—which are optimal in our normal setting—then one can generalize the results to a large class of signal distributions. This follows from an application of Theorem 2 in Lamberson and Page (2011).

[24] We thank an anonymous referee for providing the multidimensional-action example.

generally, since there is a unique solution under fractional assignment in terms of teams' precision, there always exists an optimal assignment where all firms have the same size. As a result, Propositions 5 and 6 extend to variable group sizes.

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 it is clear from the analysis that what matters are the properties of $V(\mathcal{B}(\cdot))$ as a function of $\boldsymbol{x}$. Thus, the results also apply to a *class* of matching problems where the value of a team is strictly increasing and strictly concave in $\mathcal{B}$, and $\mathcal{B}$ exhibits the complementarity properties we have used in our results. 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.

Nontransferable utility

We make the standard assumption of transferable utility. This would not hold with risk aversion or moral hazard. If experts were risk averse, in addition to 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 turn the model into one with nontransferable utility (see Legros and Newman 2007). The analysis of sorting in these cases is a relevant open problem with several economic applications.

Appendix: Omitted proofs

A.1 *Preliminaries*

We first state a few facts about Bayesian updating and the normal distribution that we invoke in the proof. Recall that each signal $\tilde{\xi}_i \sim \mathcal{N}(s, x_i^{-1})$, $i = 1, 2, \ldots, k$, and the vector $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_k)$ is distributed $\tilde{\boldsymbol{\xi}} \sim \mathcal{N}(\boldsymbol{s}, \Sigma_k)$, where $\boldsymbol{s}$ 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 $1/x_i$, $i = 1, 2, \ldots, k$, and off-diagonal elements $\rho(x_i x_j)^{-0.5}$, $i \neq j$. 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 $i, j = 1, 2, \ldots, k$,

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

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 $\tilde{\xi}_i$, $i = 2, 3, \ldots, k$, given $(\xi_1, \xi_2, \ldots, \xi_{i-1})$ is

$$\tilde{\xi}_i|_{\xi_1,\xi_2,\ldots,\xi_{i-1}} \sim \mathcal{N}\left(a_i s + b_i, \frac{1}{q_{ii}}\right),$$

where a_i and b_i are given by the expressions

$$a_i = 1 + \frac{\sum_{j=1}^{i-1} q_{ij}}{q_{ii}}, \qquad b_i = -\frac{\sum_{j=1}^{i-1} q_{ij}\xi_j}{q_{ii}}.$$

This follows from normal distribution results (e.g., Section 5.4 in De Groot 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

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

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

FACT 4. Using $\tilde{\xi}_1 \sim \mathcal{N}(s, x_1^{-1})$ and Facts 2 and 3, we obtain

$$\tilde{s}|_{\xi_1} \sim \mathcal{N}\left(\frac{\tau\mu + \xi_1 x_1}{\tau + x_1}, \frac{1}{\tau + x_1}\right),$$
$$\tilde{s}|_{\xi_1,\xi_2,\ldots,\xi_i} \sim \mathcal{N}\left(\frac{\tau_{i-1}\mu_{i-1} + a_i(\xi_i - b_i)q_{ii}}{\tau_{i-1} + a_i^2 q_{ii}}, \frac{1}{\tau_{i-1} + a_i^2 q_{ii}}\right), \quad i = 2, 3, \ldots, k.$$

This follows immediately from the two facts mentioned.

A.2 *Proof of Proposition 1*

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

We proceed by induction. This is true for $k = 1$, as (2) 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

$$\begin{aligned}\tau_k &= \tau_{k-1} + a_k^2 q_{kk}\\ &= \tau_{k-1} + \left(1 - \frac{\rho}{\big(1+(k-2)\rho\big)}\frac{\sum_{j=1}^{k-1} x_j^{0.5}}{x_k^{0.5}}\right)^2 \left(\frac{x_k\big(1+(k-2)\rho\big)}{(1-\rho)\big(1+(k-1)\rho\big)}\right)\\ &= \tau_{k-1} + \frac{\left(x_k^{0.5}\big(1+(k-2)\rho\big) - \rho\sum_{j=1}^{k-1} x_j^{0.5}\right)^2}{(1-\rho)\big(1+(k-2)\rho\big)\big(1+(k-1)\rho\big)}.\end{aligned} \tag{5}$$

By the induction hypothesis, $\tau_{k-1} = \tau + \mathcal{B}(x_1, x_2, \ldots, x_{k-1}, \rho)$ or, equivalently (using (2)),

$$\tau_{k-1} = \tau + \frac{\big(1+(k-3)\rho\big)\sum_{i=1}^{k-1} x_i - 2\rho\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}(x_i x_j)^{0.5}}{(1-\rho)\big(1+(k-2)\rho\big)}. \tag{6}$$

Combining (5) and (6) yields, after long but straightforward algebra,

$$\begin{aligned}\tau_k = \tau + \Bigg(&\frac{\big(\big(1+(k-1)\rho\big)\big(1+(k-3)\rho\big)+\rho^2\big)\sum_{i=1}^{k-1} x_i}{1+(k-2)\rho}\\ &- 2\rho\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5} + x_k\big(1+(k-2)\rho\big)\Bigg) \Big/ \big((1-\rho)\big(1+(k-1)\rho\big)\big).\end{aligned} \tag{7}$$

Since $(1+(k-1)\rho)(1+(k-3)\rho)+\rho^2 = (1+(k-2)\rho)^2$, (7) can be written as

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

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

A.3 $\mathcal{B}(\boldsymbol{x}, \rho)$ *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 $X = (X_1, \ldots, X_n) \sim \mathcal{N}(A\beta, \Sigma_X)$ and

$Y = Y_1, \ldots, Y_m \sim \mathcal{N}(B\beta, \Sigma_Y)$, where $\beta = (\beta_1, \ldots, \beta_l)'$ 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.

We apply this result to our setting. Consider two teams with compositions $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. Let the vector β be simply the scalar s, and let the matrices A and B be the $k \times 1$ unit vector I_k. Then $\boldsymbol{\xi}$ is more informative than $\boldsymbol{\xi}'$ 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 $\mathcal{B}(\boldsymbol{x}, \rho) \geq \mathcal{B}(\boldsymbol{x}', \rho)$, thereby showing that $\mathcal{B}$ indexes the informativeness of the team signals.

A.4 *Proof of Proposition 2*

To simplify the notation, let $A \equiv \sum_{i=1}^{k} x_i$ and $C \equiv \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5}$. Below we will also use the notation A_k and $C_{k-1,k}$ when we want to highlight the limits of the sums.

We first show a result that we invoke below,

$$(k-1)A - 2C = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\left(x_i^{0.5} - x_j^{0.5}\right)^2 \geq 0,$$

with strict inequality unless $x_i = x_j = x$ for all i, j. To see this, expand the right side,

$$\begin{aligned}
\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\left(x_i^{0.5} - x_j^{0.5}\right)^2 &= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\left(x_i + x_j - 2x_i^{0.5}x_j^{0.5}\right) \\
&= \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i + x_j) - 2C \\
&= \left(\sum_{i=1}^{k-1} x_i(k-i) + \sum_{i=2}^{k-1} x_i(i-1)\right) - 2C \\
&= (k-1)\sum_{i=1}^{k} x_i - 2C \\
&= (k-1)A - 2C,
\end{aligned}$$

where the third and fourth equalities follow by expansion of the sums.

(i) To prove that $\mathcal{B}$ is positive for all $(\boldsymbol{x}, \rho)$, note that

$$\begin{aligned}
\mathcal{B}(\boldsymbol{x}, \rho) &= \frac{\big(1 + (k-2)\rho\big)A - 2\rho C}{(1-\rho)\big(1 + (k-1)\rho\big)} \\
&\geq 2C\frac{\dfrac{\big(1 + (k-2)\rho\big)}{k-1} - \rho}{(1-\rho)\big(1 + (k-1)\rho\big)}
\end{aligned}$$

$$= \frac{2C}{(k-1)\big(1+(k-1)\rho\big)}$$
$$> 0,$$

where the first inequality follows from $(k-1)A \geq 2C$.

(ii) Let $\boldsymbol{x}_k$ be a vector of size k. We must show that $\mathcal{B}(\boldsymbol{x}_k, x_{k+1}, \rho) \geq \mathcal{B}(\boldsymbol{x}_k, \rho)$ for all $(\boldsymbol{x}_k, \rho)$ or, equivalently, that

$$\frac{\big(1+(k-1)\rho\big)A_{k+1} - 2\rho C_{k,k+1}}{(1-\rho)(1+k\rho)} \geq \frac{\big(1+(k-2)\rho\big)A_k - 2\rho C_{k-1,k}}{(1-\rho)\big(1+(k-1)\rho\big)}. \tag{8}$$

By cross-multiplying and using $A_{k+1} = A_k + x_{k+1}$, $C_{k,k+1} = C_{k-1,k} + x_{k+1}^{0.5}\sum_{i=1}^{k} x_i^{0.5}$, $(1+(k-1)\rho)^2 - (1+k\rho)(1+(k-2)\rho) = \rho^2$, and $1+(k-1)\rho-(1+k\rho) = -\rho$, we can rewrite (8) as

$$\rho^2(A_k + 2C_{k-1,k}) + \big(1+(k-1)\rho\big)^2 x_{k+1} - 2\rho\big(1+(k-1)\rho\big)x_{k+1}^{0.5}\sum_{i=1}^{k} x_i^{0.5} \geq 0, \tag{9}$$

which clearly holds if $\rho \leq 0$. Assume then that $\rho > 0$. Notice that in this case (9) is strictly convex in x_{k+1}. Thus, we are done if it is nonnegative when this expression is minimized with respect to x_{k+1}. The first-order condition for an interior solution yields

$$\big(1+(k-1)\rho\big)^2 - \rho\big(1+(k-1)\rho\big)x_{k+1}^{-0.5}\sum_{i=1}^{k} x_i^{0.5} = 0$$
$$\Rightarrow \quad \hat{x}_{k+1} = \frac{\rho^2}{\big(1+(k-1)\rho\big)^2}\left(\sum_{i=1}^{k} x_i^{0.5}\right)^2.$$

By inserting $\hat{x}_{k+1}$ into (9), one can verify that the resulting expression is zero. And since it is strictly convex in x_{k+1}, it is positive for any other value $x_{k+1} \neq \hat{x}_{k+1}$. And if $\hat{x}_{k+1} > \overline{x}$, then the solution of the minimization problem is at $\overline{x}$ and (9) at $x_{k+1} = \overline{x}$ is positive. Thus, we have shown that for all $(\boldsymbol{x}, \rho)$, $\mathcal{B}(\boldsymbol{x}_k, x_{k+1}, \rho) \geq \mathcal{B}(\boldsymbol{x}_k, \rho)$.

(iii) Since $\boldsymbol{x}$ is fixed, we can rewrite $\mathcal{B}$ as $\mathcal{B}(\boldsymbol{x}, \rho) = Cz(\rho, A/C)$, where

$$z\left(\rho, \frac{A}{C}\right) = \frac{\big(1+(k-2)\rho\big)\dfrac{A}{C} - 2\rho}{(1-\rho)\big(1+(k-1)\rho\big)}.$$

The result follows if $z(\cdot, A/C)$ is strictly convex in ρ. Differentiating z twice with respect to ρ and simplifying yields

$$z_{\rho\rho}\left(\rho, k, \frac{A}{C}\right) = \frac{2}{k}\left(\frac{(k-1)\dfrac{A}{C} - 2}{(1-\rho)^3} + \frac{\left(2+\dfrac{A}{C}\right)(k-1)^2}{\big(1+(k-1)\rho\big)^3}\right),$$

which is positive since, as was shown above, $(k-1)A \geq 2C$, proving convexity. To show that the minimum is achieved at a positive ρ, one can verify that the sign of the first derivative of $\mathcal{B}$ with respect to ρ is determined by the sign of

$$A(k-1)\rho\big(2+(k-2)\rho\big) - 2C\big(1+(k-1)\rho^2\big),$$

which is negative at $\rho = 0$, and thus the minimum is achieved at a positive ρ.

A.5 *Proof of Lemma 1*

Recall that the team value function is

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

Since this function is $\mathcal{C}^2$, it follows that V is submodular (supermodular) in $\boldsymbol{x}$ 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 expression

$$\begin{aligned}
&\big(1+(k-2)\rho\big)\left(4\rho\left(x_l^{0.5}\sum_{j\neq m} x_j^{0.5} + x_m^{0.5}\sum_{j\neq l} x_j^{0.5}\right)\right.\\
&\left. - \rho\sum_{i=1}^{k} x_i - \big(1+(k-2)\rho\big)4(x_l x_m)^{0.5}\right) - \tau\rho(1-\rho)\big(1+(k-1)\rho\big) \qquad (10)\\
&+ 2\rho^2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5} - 4\rho^2\sum_{j\neq m} x_j^{0.5}\sum_{j\neq l} x_j^{0.5}.
\end{aligned}$$

(i) Expression (10) at $\rho = 0$ equals $-4x_l^{0.5}x_m^{0.5} < 0$ and thus $V_{lm}|_{\rho=0} < 0$. Since this holds for any l and m, and any values of x_l and x_m, V is strictly submodular in $\boldsymbol{x}$.[25]

(ii) Notice that when $\rho < 0$, $\mathcal{B}(\underline{\boldsymbol{x}}) \leq \mathcal{B}(\boldsymbol{x})$ for all $\boldsymbol{x}$. Hence, the premise implies that $\tau < \mathcal{B}(\boldsymbol{x})$. Replacing τ by $\mathcal{B}(\boldsymbol{x})$ and using (2), we obtain that (10) is smaller than

$$\begin{aligned}
&4\rho\big(1+(k-2)\rho\big)\left(x_l^{0.5}\sum_{j\neq m} x_j^{0.5} + x_m^{0.5}\sum_{j\neq l} x_j^{0.5}\right) - \big(1+(k-2)\rho\big)^2 4(x_l x_m)^{0.5}\\
&- 2\rho\big(1+(k-2)\rho\big)\sum_{i} x_i - 4\rho^2\left(\sum_{j\neq m} x_j^{0.5}\sum_{j\neq l} x_j^{0.5} - \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(x_i x_j)^{0.5}\right).
\end{aligned}$$

[25]Notice that (10) is continuous in ρ, so there is an open interval around $\rho = 0$, which can be made dependent only on τ, $\underline{x}$, and $\overline{x}$, on which V is strictly submodular.

Since $\sum_{j\neq m} x_j^{0.5} \sum_{j\neq l} x_j^{0.5} > \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (x_i x_j)^{0.5}$, it follows that (10) is smaller than

$$\big(1+(k-2)\rho\big)\Bigg[2\rho\Bigg(2\Bigg(x_l^{0.5}\sum_{j\neq m} x_j^{0.5} + x_m^{0.5}\sum_{j\neq l} x_j^{0.5}\Bigg) - \sum_i x_i\Bigg) - \big(1+(k-2)\rho\big)4(x_l x_m)^{0.5}\Bigg].$$

A sufficient condition for this expression to be negative (and thus V strictly submodular) is that the term multiplying 2ρ inside the square bracket is nonnegative for all $\boldsymbol{x}$. This would follow if and only if the minimum of this expression is nonnegative. Rewrite it as

$$2\Bigg(x_l^{0.5}\sum_{j\neq m} x_j^{0.5} + x_m^{0.5}\sum_{j\neq l} x_j^{0.5}\Bigg) - \sum_i x_i = x_l + x_m + 2\big(x_l^{0.5} + x_m^{0.5}\big)\sum_{j\neq l,m} x_j^{0.5} - \sum_{j\neq l,m} x_j.$$

Consider the problem

$$\min_{\underline{x}\leq x_1,\ldots,x_k\leq \overline{x}} x_l + x_m + 2\big(x_l^{0.5} + x_m^{0.5}\big)\sum_{j\neq l,m} x_j^{0.5} - \sum_{j\neq l,m} x_j.$$

It is easy to see that at the optimum $x_l = x_m = \underline{x}$ and all the x_j, $j\neq l,m$, are equal and either all $\underline{x}$ or all $\overline{x}$. If they are all $\underline{x}$, then the value of the objective at the minimum is $(2+3(k-2))\underline{x} > 0$ and then V is strictly submodular. If they are all $\overline{x}$, then the value of the objective is $2\underline{x} + (k-2)\overline{x}^{0.5}(4\underline{x}^{0.5} - \overline{x}^{0.5})$, which is positive if $\overline{x} \leq 16\underline{x}$.

(iii) From (1), we obtain that $V_{lm} < 0$ if and only if $\mathcal{B}_{lm} - (2\mathcal{B}_l\mathcal{B}_m/(\tau+\mathcal{B})) < 0$. When $\rho > 0$, $\mathcal{B}_{lm} < 0$. Hence, it suffices that B_i, $i = 1,2,\ldots,k$, be nonnegative. The sign of $\mathcal{B}_i$ reduces to that of

$$1+(k-2)\rho - \rho x_i^{-0.5}\sum_{j\neq i} x_j^{0.5} \geq 1+(k-2)\rho - \rho(k-1)\left(\frac{\overline{x}}{\underline{x}}\right)^{0.5},$$

which is nonnegative if $\rho \leq ((k-1)(\overline{x}/\underline{x})^{0.5} - (k-2))^{-1}$.

One can improve the bound given in (ii) by checking when $2\underline{x} + (k-2)\overline{x}^{0.5}(4\underline{x}^{0.5} - \overline{x}^{0.5}) \geq 0$. It is easy to show that this is the case if $\overline{x} \leq \alpha(k)\underline{x}$, where $\alpha(k) = (7 - 7.5k + 2k^2 - 2(k-2)^{\frac{2}{3}}(k-1)^{0.5})^{-1} > 16$ for all $k \geq 2$, and $\lim_{k\to 2}\alpha(k) = \infty$.

A.6 *Supermodularity of V and prior precision*

Assume $\rho < 0$ as in Lemma 1(ii). We assert in Section 3.2 that in this case supermodularity of V in $\boldsymbol{x}$ requires $\tau > 9.89k\overline{x}$. We now justify this assertion.

Using (11), it follows that a necessary condition for V supermodular in $\boldsymbol{x}$ when correlation is ρ is that prior precision be larger than the bound

$$\tau > \frac{4(1-\rho)+k\rho}{-k\rho\big(1+(k-1)\rho\big)}k\overline{x}.$$

Notice that the right-hand side goes to infinity as ρ goes to 0 or $-1/(k-1)$. More generally, consider the problem

$$J(k) = \min_{\rho\in(-\frac{1}{1-k},0)} \frac{4(1-\rho)+k\rho}{-k\rho\big(1+(k-1)\rho\big)}.$$

One can show that the minimum is achieved at $-(2(2(k-1)-\sqrt{3k(k-1)})/(4-5k+k^2)$ and that $J(k)=(-4+7k+4\sqrt{3k(k-1)})/k$, which is increasing in k. Since $k\geq 2$, it follows that $J(k)\geq J(2)=9.898979$ for all $k\geq 2$. Therefore, τ must be at least $J(2)k\overline{x}$ for V to be supermodular in $\boldsymbol{x}$ at ρ.

A.7 *Failure of supermodularity of V around the diagonal*

Assume $\rho>0$. We asserted in the text that V cannot be supermodular. To show it, let $x\in[\underline{x},\overline{x}]$ and $x_1=x_2=\cdots=x_k=x$. Evaluating the expression for the cross-partial in (10) at this vector yields

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

After algebraic manipulation, this expression can be written as

$$-\tau\rho\big(1+(k-1)\rho\big)-x\big(4(1-\rho)+k\rho\big)\leq-\tau\rho\big(1+(k-1)\rho\big)-\underline{x}\big(4(1-\rho)+k\rho\big)<0. \quad (11)$$

Hence, $V_{lm}|_{x_1=\cdots=x_k=x}<0$ 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 $\boldsymbol{x}$ belong to that interval, then $V_{l,m}<0$ for all l and m. Thus, for any $\rho>0$, V cannot be supermodular on $[\underline{x},\overline{x}]^k$. Finally, 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. Hence, diversification ensues if heterogeneity of expertise is small.

A.8 *Proof of Proposition 4*

Let $M\subseteq\Gamma$ consist of all the vectors (generated by partitions) that are majorized by all the remaining vectors in M^c (the complement of M in Γ).

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

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

A.9 *Failure of the Gross substitutes property*

Recall that the model can be reinterpreted as one of matching groups of experts with identical firms. A decentralized version of the problem would have each firm face a vector of characteristic-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

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

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 $T(A^*) = \{i \in A^* \mid w(x_i) = w'(x_i)\}$.

The following example shows that GS fails in our model.

The firm faces three experts, 1, 2, and 3, with $x_1 = 1$, $x_2 = 2$, 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 $v(x_1) - w(x_1) = (5/12) + \varepsilon$, $v(x_2) - w(x_2) = 7/12$, $v(x_3) - w(x_3) = 7/12$, $v(x_1 + x_2) - w(x_1) - w(x_2) = (7/12) + \varepsilon$, $v(x_1 + x_3) - w(x_1) - w(x_3) = (33/60) + \varepsilon$, $v(x_2 + x_3) - w(x_2) - w(x_3) = 7/12$, and $v(x_1 + x_2 + x_3) - w(x_1) - w(x_2) - w(x_3) = (44/84) + \varepsilon$. 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 $v(x_1) - w(x_1) = (5/12) + \varepsilon$, $v(x_2) - w(x_2) = 1/2$, $v(x_3) - w(x_3) = 7/12$, $v(x_1 + x_2) - w(x_1) - w(x_2) = (1/2) + \varepsilon$, $v(x_1 + x_3) - w(x_1) - w(x_3) = (33/60) + \varepsilon$, $v(x_2 + x_3) - w(x_2) - w(x_3) = 1/2$, and $v(x_1 + x_2 + x_3) - w(x_1) - w(x_2) - w(x_3) = (37/84) + \varepsilon$. Thus, if $\varepsilon < 1/30$, then the optimal choice is unique and given by $B^* = \{3\}$.

Since in this case $T(A^*) = \{1\} \not\subseteq B^*$, 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.

A.10 *Proof of Proposition 5*

Let $X_n \equiv \sum_j \mu_{jn} x_j$; multiply both sides of $\sum_n \mu_{jn} = m_j$ by x_j and sum with respect to j to obtain $\sum_n X_n = X$. If we ignore the other constraints in the problem, we obtain the doubly relaxed problem of finding $(X_1, \ldots, X_N)$ to maximize $\sum_n v(X_n)$ subject to $\sum_n X_n = X$, whose unique solution (by strict concavity of v) is $X_n = X/N$ for all n. Thus, any vector of μ_{jn}s such that $\sum_j \mu_{jn} x_j = X/N$ and $\mu_{jn} \geq 0$ for all j, n solves the doubly relaxed problem. If, in addition, such a vector satisfies $\sum_n \mu_{jn} = m_j$ for all j, then it solves the relaxed problem where only the size constraint is omitted.[26] Finally, if it also satisfies $\sum_j \mu_{jn} = k$ for all n, then it solves the fractional assignment problem.

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.

Since there is at least one solution in terms of μ_{jn}s that achieves the value $Nv(X/N)$ (the value of the less constrained problem above), it follows that any optimal matching must have the property that $X_n = X/N$ for all n.[27]

[26] Unlike the case with integer assignment, now we could easily solve the problem with unequal group sizes, since the main requirement is that the equal-team precision condition is satisfied.

[27] An alternative proof of this property is to show that if it does not hold at the optimum of the fractional assignment problem, then there are two teams n'' and n' such that $X_{n''} < X_{n'}$, and two characteristics $x_\ell < x_p$ with $\mu_{\ell n''} > 0$ and $\mu_{pn'} > 0$. Then an $\varepsilon > 0$ reallocation of these types reduces $X_{n'}$ and increases $X_{n''}$, which increases the value of the objective function. Thus, $X_n = X/N$ for all n.

A.11 *Fractional assignment with $\rho \leq 0$*

Recall that there are m agents with $x_1 \leq x_2 \leq \cdots \leq x_m$. Assume that $\rho \in (-1/(m-1), 0]$. It will be convenient to reinterpret μ_{in} as the fraction of agent i, between 0 and 1, allocated to team n. As in Section 3.4.3, we assume that if i allocates a fraction $0 \leq \mu_{in} \leq 1$ to team n, then she contributes with the realization of a signal whose precision is $\mu_{in}x_i$. For clarity, denote the informativeness of team n given $\{\mu_{in}\}_{i=1}^m$ by

$$\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho) = \frac{\big(1+(m-2)\rho\big)\sum_{i=1}^{m}\mu_{in}x_i - 2\rho\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(\mu_{in}x_i\mu_{jn}x_j)^{0.5}}{(1-\rho)\big(1+(m-1)\rho\big)}. \tag{12}$$

Notice that $\mathcal{B}$ is concave in $\boldsymbol{\mu}_n$ when $-1/(m-1) < \rho \leq 0$, as it is a sum of concave functions of the μ_{in}s. Denote the value of team n given $(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho)$ by $v(\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho)) = \pi - (1/(\tau + \mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho)))$. Since $v(\mathcal{B}(\boldsymbol{x}, \cdot, \rho))$ is a strictly increasing and concave transformation of $\mathcal{B}(\boldsymbol{x}, \cdot, \rho))$, it follows that $v(\mathcal{B}(\boldsymbol{x}, \cdot, \rho))$ is concave in $\boldsymbol{\mu_n}$.

We will assume that all agents are present in all teams in some fraction, which can be zero for some teams. The planner then solves the problem

$$\max_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_N} \sum_{n=1}^{N} v\big(\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho)\big)$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\mu_{in} = k, \quad n = 1, 2, \ldots, N, \tag{13}$$

$$\mu_{in} \geq 0, \quad i = 1, \ldots, m, n = 1, 2, \ldots, N,$$

$$\mu_{in} \leq 1, \quad i = 1, \ldots, m, n = 1, 2, \ldots, N, \tag{14}$$

$$\sum_{n=1}^{N}\mu_{in} = 1 \quad i = 1, 2, \ldots, m. \tag{15}$$

Proposition 8. *Under the assumptions made, the perfect diversification assignment $\mu_{in} = 1/N$ for all i, n solves the planner's problem.*

Proof. The objective function is continuous in $\boldsymbol{\mu}_n$ and the constraints set is compact and consists of linear constraints. Thus, a solution exists, and since the objective function is concave in $\boldsymbol{\mu}_n$, any critical point that satisfies the Kuhn–Tucker conditions is a global optimum. Concavity and symmetry make $\mu_{in} = 1/N$ for all i, n a natural candidate solution.

Notice that $\mu_{in} = 1/N$ for all i, n is feasible, for it satisfies all the constraints (the first one because $k = m/M$). We now show that it is indeed an *optimal solution*.

Let λ_n, ξ_{in}, and ν_i, $i = 1, 2, \ldots, m$, $n = 1, 2, \ldots, N$, be the multipliers of constraints (13), (14), and (15), respectively. The Kuhn–Tucker conditions are, for all i, n,

$$v'\big(\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho)\big)\mathcal{B}_{\mu_{in}}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho) - \lambda_n - \xi_{in} - \nu_i \leq 0, \quad \mu_{in} \geq 0$$

with complementary slackness.

Consider, as part of a candidate solution, $\mu_{in} = 1/N$ for all i, n. Since this is positive and strictly less than 1, it follows that $\xi_{in} = 0$ for all i, n and

$$v'\left(\mathcal{B}\left(\boldsymbol{x}, \frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}, \rho\right)\right)\mathcal{B}_{\mu_{in}}\left(\boldsymbol{x}, \frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}, \rho\right) = \lambda_n + \nu_i. \tag{16}$$

Notice that (16) is a system of $i \times n$ equations. Since the compositions of all teams are equal, it follows that $\mathcal{B}_{\mu_{in}}$ is independent of n; denote it then by $\mathcal{B}_{\mu_i}$. But this implies that the right-hand side of (16) is independent of n and as a result $\lambda_n = \lambda$ for all n. Thus, there exists a $\lambda \geq 0$ and $\nu_i = v'\mathcal{B}_{\mu_i} - \lambda \geq 0$ for all i, thereby completing the description of a critical point with $\mu_{in} = 1/N$ for all i, n. Since the planner's problem is a concave programming problem, $\mu_{in} = 1/N$ for all i, n is an optimal solution. □

We have implicitly assumed in the proof that all experts are present in each team even if the fraction of the agent allocated to the team is zero. That is, we kept m fixed in $\mathcal{B}$ even when some fractions were zero. This is irrelevant in the conditionally independent case since $\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, 0) = \sum_{i=1}^{m} \mu_{in} x_i$ and thus if $\mu_{in} = 0$ for some n, then the number of characteristics present in a team is automatically reduced by one.

Although we have assumed that we keep zero precision members as part of the team, the perfect diversification result under negative correlation still obtains if we reduce the number of experts present in a team by one when the fraction of an expert is zero. To see this, notice that the sums in (12) do not change if we replace m by $m-1$ when one of the μ_{in}s is zero. So the difference is in the first term in the numerator (which is a function of m) and the denominator. To simplify the notation, let $A = \sum_{i=1}^{m} \mu_{in} x_i$ and $B = -2\rho \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (\mu_{in} x_i \mu_{jn} x_j)^{0.5}$. Then

$$\mathcal{B}(\boldsymbol{x}, \boldsymbol{\mu}_n, \rho) = \frac{\big(1 + (m-2)\rho\big)A + B}{(1-\rho)\big(1 + (m-1)\rho\big)}.$$

It is straightforward to show that $\operatorname{sign}(\partial\mathcal{B}/\partial m) = \operatorname{sign}(\rho^2 A - \rho B) > 0$. Hence $\mathcal{B}$ increases in m or, equivalently, the informativeness of a team is lower if a zero fraction of an expert reduces m by one. As a result, the value of the team is lower, and hence if $\mu_{in} = 1/N$ for all i, n solves the problem under our assumption, then it also solves it if a zero fraction of an expert reduces the number of fractionally allocated experts in a team by one.

A.12 *Proof of Proposition 6*

Firms choose the μ_{jn}s to maximize $v(\sum_j \mu_{jn} x_j) - \sum_j \mu_{jn} w_j$. From the first-order conditions, we obtain $X_n = \sum_j \mu_{jn} x_j = X/N$ for all n, and thus $\mu_{jn} = m_j/N$ is the unique symmetric equilibrium allocation. Inserting the solution into the zero profit constraint $v(\sum_j \mu_{jn} x_j) - \sum_j \mu_{jn} w_j - F = 0$, taking $w_j = v'(X/N)x_j$ from the first-order condition, and using $X = \sum_j m_j x_j$ yields the following equilibrium condition for N:

$$\pi - F = N(N\tau + X)^{-1} + NX(N\tau + X)^{-2}.$$

Rewrite the equilibrium condition as

$$\pi - F = \frac{N^2\tau + 2NX}{(N\tau + X)^2}. \tag{17}$$

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 (17).[28]

The comparative statics of N^* with respect to F, τ, and X are as follows. Rewrite (17) 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.

A.13 *Proof of Proposition 7*

Since $v'(X_i) = (\tau + X_i)^{-2}$, the first-order condition $y_n v'(X_n) = y_m v'(X_m)$ yields

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

Fix n and sum both sides for all m. Using $\sum_m X_m = X$, we obtain

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

which is (4). By the strict concavity of v, 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.

To prove the second part of the proposition, notice that the optimal precision vector $(X_1, X_2, \ldots, X_N)$ can be implemented by any fractional assignment $\{\mu_{jn}\}_{j,n}$ that satisfies

$$\sum_{j=1}^{J} \mu_{jn} x_j = X_n \quad \forall n,$$

$$\sum_{n=1}^{N} \mu_{jn} = m_j \quad \forall n,$$

$$\mu_{jn} \geq 0 \quad \forall j, n,$$

where we have omitted from the system the equations $\sum_j \mu_{jn} = k$ for all n since, as the proof of Proposition 5 shows, one can relax the equal-size group restriction in this case (what is important is that the precision of all teams is pinned down uniquely).

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

We now show that there exists a nonnegative vector $\{\mu_{jn}\}_{j,n}$ that satisfies these equations. Rewrite the above system as

$$A\mu = b, \quad \mu \geq 0, \tag{18}$$

where $A = \left[\begin{smallmatrix} A_1 & A_2 & \cdots & A_N \\ I & I & \cdots & I \end{smallmatrix}\right]$ is a $(N+J) \times JN$ matrix; each A_n is a $N \times J$ matrix with $x_1, x_2, \ldots, x_J$ in row n and the rest zeroes, and $\mathbb{I}$ is a $J \times J$ identity matrix; μ is a $JN \times 1$ vector whose entries are $\mu_{11}, \ldots, \mu_{J1}, \mu_{12}, \ldots, \mu_{J2}, \ldots, \mu_{1N}, \ldots, \mu_{JN}$; and $b = [X_1, \ldots, X_N, m_1, \ldots, m_J]$ is a $(N+J) \times 1$ vector.

By Farkas' lemma, (18) has a solution if and only if there is no $1 \times (N+J)$ vector z that solves the system

$$zA \geq 0, \qquad zb < 0. \tag{19}$$

Assume there is a solution to (19). Notice that $zA \geq 0$ consists of

$$z_i x_j + z_{N+j} \geq 0, \quad i = 1, 2, \ldots, N, j = 1, 2, \ldots, N.$$

Multiplying by m_j and summing for all j yields

$$z_i \sum_{j=1}^{J} m_j x_j + \sum_{j=1}^{J} z_{N+j} m_j \geq 0, \quad i = 1, 2, \ldots, N.$$

Since $X = \sum_j m_j x_j$ and the inequality holds for all z_i, $i = 1, 2, \ldots, N$, it follows that it also holds for $\min_i z_i$, so

$$\Big(\min_i z_i\Big) X + \sum_{j=1}^{J} z_{N+j} m_j \geq 0.$$

But $(\min_i z_i)X = (\min_i z_i) \sum_i X_i \leq \sum_i z_i X_i$ and thus

$$\sum_i z_i X_i + \sum_{j=1}^{J} z_{N+j} m_j \geq 0,$$

and this contradicts $zb < 0$, which is equal to

$$\sum_i z_i X_i + \sum_{j=1}^{J} z_{N+j} m_j < 0.$$

Hence, system (19) does not have a solution; by Farkas' lemma, there is a solution to (18), which proves the existence of a fractional assignment of agents into teams with precision X_n given by (4) for all n.

We now prove the comparative statics properties of X_n, $n = 1, 2, \ldots, N$, asserted in the text. The derivative of (4) with respect to X is $y_n^{0.5} / \sum_m y_m^{0.5} > 0$, and thus an increase in X increases X_n for all n.

The derivative of (4) with respect to τ is $(y_n^{0.5}N/\sum_m y_m^{0.5}) - 1$, and this is positive if and only if $y_n^{0.5} > \sum_m y_m^{0.5}/N$. 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

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

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

We now show that increasing the spread of the vector $\boldsymbol{y}$ increases the team precision for better teams and decreases it for worse ones. We do so for a class of vectors ordered by majorization that in addition satisfy a single crossing property. Consider $\boldsymbol{y'} = (y'_1, y'_2, \ldots, y'_N)$ such that $y'_n = y_n - \Delta_n$ for all $n \leq \hat{n}$, with $\Delta_1 \geq \Delta_2 \geq \cdots \geq \Delta_{\hat{n}} \geq 0$, and $y'_n = y_n + \Delta_n$, $\Delta_n \geq 0$, for all $n \geq \hat{n}+1$, and such that $\sum_{n=1}^{\hat{n}} \Delta_n = \sum_{n=\hat{n}+1}^{N} \Delta_n$. Notice that $\boldsymbol{y'}$ majorizes $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$. We will show that there exists an n^* such that X_n increases for all $n \geq n^*$ and decreases otherwise when $\boldsymbol{y}$ is replaced by $\boldsymbol{y'}$.

Since $\sum_n y_n^{0.5}$ is the sum of concave functions in one variable $y_n^{0.5}$, it follows that it is Schur concave in $(y_1, y_2, \ldots, y_N)$. Hence, $\sum_n (y'_n)^{0.5} \leq \sum_n y_n^{0.5}$ since $\boldsymbol{y'}$ majorizes $\boldsymbol{y}$. It is now immediate that X_n increases for all $n \geq \hat{n}+1$, for in this case $(y'_n)^{0.5} > y_n^{0.5}$ and thus $(y'_n)^{0.5}/\sum_n (y'_n)^{0.5} > y_n^{0.5}/\sum_n y_n^{0.5}$, thereby increasing X_n (see (4)).

Consider now teams $1, 2, \ldots, \hat{n}$. If X_n decreases for any $n \leq \hat{n}$ when $\boldsymbol{y'}$ replaces $\boldsymbol{y}$, then it must decrease for all teams $1, \ldots, n$, for X_n decreases if and only if

$$\frac{(y_n - \Delta_n)^{0.5}}{\sum_{n=1}^{N} (y'_n)^{0.5}} \leq \frac{y_n^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}} \quad \Leftrightarrow \quad \left(1 - \frac{\Delta_n}{y_n}\right)^{0.5} \leq \frac{\sum_{n=1}^{N} (y'_n)^{0.5}}{\sum_{n=1}^{N} y_n^{0.5}}.$$

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

Finally, we complete the analysis of the example in the text. When $J = 2$, the following vector for each $n = 1, 2, \ldots, N$ solves the fractional assignment problem:

$$(\mu_{1n}, \mu_{2n}) = \left(\frac{kx_2 - X_n}{x_2 - x_1}, \frac{X_n - kx_1}{x_2 - x_1}\right). \tag{20}$$

Intuitively, μ_{1n} decreases in n while μ_{2n} increases in n. When $N = k = 2$, a simple rewriting of (20) yields the formulas used in the example in the text.

A.14 *Alternative information model*

We now prove the assertion made in Section 5 that the team value function is strictly submodular when an agent's characteristic 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 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, $i=1,\dots,k$, the number of informative signals in a k-size team is a random variable with Poisson's binomial distribution (Wang 1993). Let $m=0,1,\dots,k$ and define $\mathcal{F}_m \equiv \{B : B\subseteq\{1,2,\dots,k\}, \mid B \mid = m\}$.

The probability of m informative signals out of k in a team with characteristics $\boldsymbol{x}$ is

$$\sum_{B\in\mathcal{F}_m}\Big(\prod_{\ell\in B}x_\ell\Big)\Big(\prod_{p\notin B}(1-x_p)\Big).$$

Thus, the team value function is

$$\begin{aligned}
V(\boldsymbol{x}) &= \sum_{m=0}^{k} u(m)\Big(\sum_{B\in\mathcal{F}_m}\Big(\prod_{\ell\in B}x_\ell\Big)\Big(\prod_{p\notin B}(1-x_p)\Big)\Big)\\
&= \sum_{R\subseteq\{1,2,\dots,k\}} u\big(|R|\big)\Big(\prod_{\ell\in R}x_\ell\Big)\Big(\prod_{p\notin R}(1-x_p)\Big)\\
&= \sum_{R\subseteq\{1,2,\dots,k\}\setminus\{i,j\}}\Big(\prod_{\ell\in R}x_\ell\Big)\Big(\prod_{p\notin R}(1-x_p)\Big)\big(x_ix_ju\big(|R|+2\big)\\
&\quad + x_i(1-x_j)u\big(|R|+1\big)+x_j(1-x_i)u\big(|R|+1\big)+(1-x_i)(1-x_j)u\big(|R|\big),
\end{aligned} \tag{21}$$

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 et al. 2011).

Differentiating (21) with respect to x_i and x_j yields

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

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

A.15 *Examples of PAM under alternative assumptions*

Our analysis shows that when standard models of information acquisition are embedded in a matching setting, optimal sorting entails diversification of expertise.

As we mention in the concluding remarks, it is difficult to provide more general results given the current knowledge of curvature and complementarity properties in the value of information as a function of the informativeness of the signals.

We now present two examples that illustrate that, for carefully chosen primitives (outside the canonical models), one can construct cases where PAM is optimal.

The first example assumes that $s \in \{0, 1\}$ with prior $\mathbb{P}[s = 1] = \mu \in (0, 1)$; the action set is $A = \mathbb{R}_+$; the signal of each agent i is $\tilde{\xi}_i \sim N(s, x_i)$, i.e., normally distributed with mean s and precision $x_i > 0$; signals are conditionally independent; $k = 2$; and the payoff function a team maximizes is $2a - sa^2$.[29]

Consider any given team. The posterior belief after observing (ξ_i, ξ_j) is

$$\mu(\xi_i, \xi_j, x_i, x_j) = \frac{\mu f(\xi_i \mid 1, x_i) f(\xi_j \mid 1, x_j)}{\mu f(\xi_i \mid 1, x_i) f(\xi_j \mid 1, x_j) + (1 - \mu) f(\xi_i \mid 0, x_i) f(\xi_j \mid 0, x_j)},$$

where

$$f(\xi \mid s, x) = \frac{1}{\sqrt{2\pi x^{-1}}} e^{-\frac{x}{2}(\xi - s)^2}.$$

After observing the signal realizations and updating beliefs, the team solves

$$U(\xi_i, \xi_j, x_i, x_j) = \max_a \mu(\xi_i, \xi_j, x_i, x_j)\big(2a - a^2\big) + \big(1 - \mu(\xi_i, \xi_j, x_i, x_j)\big)2a.$$

Simple algebra reveals that $a^*(\xi_i, \xi_j) = 1/\mu(\xi_i, \xi_j, x_i, x_j) = U(\xi_i, \xi_j, x_i, x_j)$.

The marginal density of the signals is

$$f(\xi_i, \xi_j, x_i, x_j) = \mu f(\xi_i \mid 1, x_i) f(\xi_j \mid 1, x_j) + (1 - \mu) f(\xi_i \mid 0, x_i) f(\xi_j \mid 0, x_j),$$

which we denote by $\mu f_{1i} f_{1j} + (1 - \mu) f_{0i} f_{0j}$. Then the value of the team is given by

$$\begin{aligned}
V(x_i, x_j) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\mu(\xi_i, \xi_j, x_i, x_j)} \big(\mu f_{1i} f_{1j} + (1 - \mu) f_{0i} f_{0j}\big) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\big(\mu f_{1i} f_{1j} + (1 - \mu) f_{0i} f_{0j}\big)^2}{f_{1i} f_{1j} \mu} \\
&= 2 - \mu + \frac{(1 - \mu)^2}{\mu} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f_{0i}^2}{f_{1i}} \frac{f_{0j}^2}{f_{1j}} \\
&= 2 - \mu + \frac{(1 - \mu)^2}{\mu} \left(\int_{-\infty}^{\infty} \frac{f_{0i}^2}{f_{1i}} \right) \left(\int_{-\infty}^{\infty} \frac{f_{0j}^2}{f_{1j}} \right) \\
&= 2 - \mu + \frac{(1 - \mu)^2}{\mu} e^{x_i} e^{x_j},
\end{aligned}$$

where the second equality follows by replacing $\mu(\xi_i, \xi_j, x_i, x_j)$, the third follows by straightforward algebra, the fourth follows by independence of $\tilde{\xi}_i$ and $\tilde{\xi}_j$, and the last follows from ($\ell = i, j$)

$$\int_{-\infty}^{\infty} \frac{f_{0\ell}^2}{f_{1\ell}} = e^{x_\ell} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi x_\ell^{-1}}} e^{-\frac{x_\ell}{2}(y+1)^2} = e^{x_\ell},$$

[29]The example generalizes to multiple experts a single-agent example in Chade and Schlee (2002).

where we use that the integrand in the middle expression is the density of a random variable that is normal with mean -1 and precision x_i, and hence integrates to 1.

Notice that $V(x_i, x_j) = 2 - \mu + ((1-\mu)^2/\mu)e^{x_i}e^{x_j}$ is *strictly supermodular* in (x_i, x_j). Hence, in this setting the optimal matching is PAM.

In the second example, the state is a pair (s_1, s_2), $s_i \in \{0, 1\}$, independently distributed with $\mathbb{P}[s_i = 1] = 0.5$, $i = 1, 2$. There are four agents: 1 and 2 observe uninformative signals about the states; 3 observes a perfectly informative signal about s_1 and an uninformative signal about s_2, and the opposite is true for agent 4. Agents match pairwise and each teams maximizes the expected value of $-\max\{(a_1 - s_1)^2, (a_2 - s_2)^2\}$ with respect to a_1, a_2, with $a_i \in \mathbb{R}$, $i = 1, 2$. We will show that PAM is optimal.

Under NAM, 1 matches with 3 and 2 matches with 4. Consider team $\{1, 3\}$: since 3 receives a signal that reveals state s_1, his action will match the state perfectly and hence the value of this team, $V(1, 3)$, is

$$V(1,3) = 0.5 \max_{a_2} \left(-(a_2 - 1)^2 - a_2^2\right) = -0.25,$$

where the second inequality follows by replacing the optimal action $a_2 = 0.5$. Similarly, $V(2, 4) = -0.25$, and the overall payoff for the planner under NAM is -0.5.

Under PAM, 1 matches with 2 and 3 matches with 4. Clearly, $V(3, 4) = 0$ as each agent has perfect information about one state. Regarding the other team, $V(1, 2)$ is given by

$$\begin{aligned} V(1,2) = 0.25 \max_{a_1, a_2} \big(&- \max\left\{(a_1 - 1)^2, (a_2 - 1)^2\right\} - \max\left\{a_1^2, (a_2 - 1)^2\right\} \\ &- \max\left\{(a_1 - 1)^2, a_2^2\right\} - \max\left\{a_1^2, a_2^2\right\}\big) \\ = -0.25, \end{aligned}$$

where the second equality follows by replacing the optimal actions $a_1 = a_2 = 0.5$. Hence, the planner's payoff under PAM is $-0.25 > -0.5$, thereby proving that *PAM is optimal*.

References

Angeletos, George-Marios and Alessandro Pavan (2007), "Efficient use of information and social value of information." *Econometrica*, 75, 1103–1142. [394]

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006), "Who's who in networks. Wanted: The key player." *Econometrica*, 74, 1403–1417. [394]

Becker, Gary S. (1973), "A theory of marriage: Part I." *Journal of Political Economy*, 81, 813–846. [378, 379, 386]

Börgers, Tilman, Angel Hernando-Veciana, and Daniel Krähmer (2013), "When are signals complements or substitutes?" *Journal of Economic Theory*, 148, 165–195. [380]

Calinescu, Gruia, Chandra Chekuri, Martin Pál, and Jan Vondrák (2011), "Maximizing a monotone submodular function subject to a matroid constraint." *SIAM Journal on Computing*, 40, 1740–1766. [410]

Chade, Hector, Jan Eeckhout, and Lones Smith (2017), "Sorting through search and matching models in economics." *Journal of Economic Literature*, 55, 1–52. [387]

Chade, Hector and Edward Schlee (2002), "Another look at the Radner–Stiglitz nonconcavity in the value of information." *Journal of Economic Theory*, 107, 421–452. [395, 411]

Damiano, Ettore, Hao Li, and Wing Suen (2010), "First in village or second in Rome?" *International Economic Review*, 51, 263–288. [379]

De Groot, Morris H. (1970), *Optimal Statistical Decisions*. McGraw-Hill, New York. [383, 397]

Figueiredo, Mário A. T. (2004), "Lecture notes on Bayesian estimation and classification." Unpublished paper, Instituto de Telecomunicações. [383]

Garey, M. R. and D. S. Johnson (1978), "Strong NP-completeness results: Motivation, examples, and implications." *Journal of the Association for Computing Machinery*, 25, 499–508. [378, 389]

Goel, Prem and Josep Ginebra (2003), "When is one experiment 'always better than' another?" *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52, 515–537. [381, 384, 398]

Hansen, Ole Havard and Erik N. Torgersen (1974), "Comparison of linear normal experiments." *The Annals of Statistics*, 2, 367–373. [380, 384, 398]

Hong, Lu and Scott E. Page (2001), "Problem solving by heterogeneous agents." *Journal of Economic Theory*, 97, 123–163. [380]

Hwang, Frank and Uriel G. Rothblum (2011), *Partitions: Optimality and Clustering Volume I: Single-Parameter*. Series on Applied Mathematics. World Scientific, Singapore. [380, 395]

Kelso, Alexander S. and Vincent P. Crawford (1982), "Job matching, coalition formation, and gross substitutes." *Econometrica*, 50, 1483–1504. [379, 389, 404]

Kremer, Michael (1993), "The O-ring theory of economic development." *Quarterly Journal of Economics*, 108, 551–575. [379]

Lamberson, P. J. and Scott E. Page (2011), "Optimal forecasting groups." *Management Science*, 58, 805–810. [380, 383, 395]

Legros, Patrick and Andrew F. Newman (2007), "Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities." *Econometrica*, 75, 1073–1102. [396]

Lehmann, E. L. (1988), "Comparing location experiments." *The Annals of Statistics*, 16, 521–533. [381]

Marshak, Jakob and Roy Radner (1972), *Economic Theory of Teams*. Yale University Press, New Haven. [377, 380]

Marshall, Albert, Ingram Olkin, and Barry Arnold (2010), *Inequalities: Theory of Majorization and Its Applications*, Second edition. Springer Series in Statistics. Springer, New York. [388]

Mertens, Stephan (2006), "The easiest hard problem: Number partitioning." In *Computational Complexity and Statistical Physics* (Allon Percus, Gabriel Istrate, and Cristopher Moore, eds.), 125–139, Oxford University Press, New York. [389]

Meyer, Margaret A. (1994), "The dynamics of learning for team production: Implications for task assignments." *The Quarterly Journal of Economics*, 109, 1157–1184. [379, 380]

Moscarini, Giuseppe and Lones Smith (2002), "The law of large demand for information." *Econometrica*, 70, 2351–2366. [395]

Murota, Kazuo (2003), *Discrete Convex Analysis*. Discrete Mathematics and Applications. Siam, Philadelphia. [389]

Olszewski, Wojciech and Rakesh Vohra (2012), "Team selection problem." Unpublished paper, Northwestern University. [380]

Prat, Andrea (2002), "Should a team be homogeneous?" *European Economic Review*, 46, 1187–1207. [380]

Pycia, Marek (2012), "Stability and preference alignment in matching and coalition formation." *Econometrica*, 80, 323–362. [379]

Radner, Roy (1962), "Team decision problems." *The Annals of Mathematical Statistics*, 33, 857–881. [380]

Shaked, Moshe and Y. L. Tong (1990), "Comparison of experiments for a class of positively dependent random variables." *The Canadian Journal of Statistics*, 18, 79–86. [380, 384]

Vondrák, Jan (2007), *Submodularity in Combinatorial Optimization*. Doctoral Thesis, Department of Applied Mathematics, Charles University. [378]

Wang, Y. H. (1993), "On the number of success in independent trials." *Statistica Sinica*, 3, 295–312. [394, 410]

Weitzman, Martin L. (1998), "The Noah's ark problem." *Econometrica*, 66, 1279–1298. [394, 410]

Williams, David (1991), *Probability With Martingales*. Cambridge University Press. [397]

Co-editor George J. Mailath handled this manuscript.

Manuscript received 7 April, 2014; final version accepted 7 February, 2017; available online 14 March, 2017.