Coarse Agents and Intergroup Phenomena[*]

Giacomo Weber [†]

August 27, 2025

Find the latest version here

Abstract

This paper proposes a framework for analyzing intergroup phenomena. A set of heterogeneous players is divided into two groups. Players are exogenously matched, and each pair plays a simultaneous-move game under complete information. When players are matched within the same group ("ingroup"), they form correct expectations about their opponent's behavior in equilibrium. Conversely, players form coarse expectations when matched with a member of the "outgroup." In equilibrium, such coarse expectations must coincide with the aggregate behavior of the outgroup. The concept is illustrated in coordination games and applied to an organizational setting where the groups represent subdivisions, and each game corresponds to a team task. These tasks are identical and exhibit strategic complementarities. An omniscient designer sorts players into pairs to maximize the overall probability of task success. The analysis of the optimal assignment emphasizes the role of coarse expectations: by pairing efficient players within their group and less efficient players across groups, the designer can induce the latter to overexert effort in equilibrium.

---

[*] I thank audiences at TOM and PSI seminars at PSE, Organizational Economics Lunch at MIT Sloan, Microeconomics Working Group at EUI, Transatlantic Theory Workshop 2024, TSE–PSE Workshop 2025. I am indebted to Philippe Jehiel, Rachel Kranton, Ran Spiegler, Evan Friedman, Bernard Caillaud, Bob Gibbons, Larry Samuelson, Frederic Koessler, Roberto Corrao, Giacomo Lanzani, Laura Carlotta Hoenig, and Manpreet Singh for useful suggestions and discussions. This work has benefited from the funding of the European Research Council (grant 742816) and the Agence Nationale de la Recherche through the program Investissements d'Avenir ANR-17-EURE-0001. This work has also been funded by European Union – Next Generation EU, Mission 4 Component 2, CUP J53D23004180006.

[†] PSE - EHESS, 48 boulevard Jourdan, 75014 Paris, France and University of Bologna, Piazza Scaravilli 2, Bologna, Italy; giacomo.weber@psemail.eu

1 Introduction

Beliefs regarding behaviors across various social groups wield significant influence over societal interactions and economic outcomes. Interactions across social groups are a core element of many phenomena studied in economics and psychology, such as stereotypes (Hilton and von Hippel (1996); Bordalo et al. (2016)), statistical discrimination (Phelps (1972); Arrow (1973)), and identity (Tajfel (1974); Tajfel and Turner (1979); Akerlof and Kranton (2000)). These phenomena have been shown to have a strong impact on labor markets, consumer choices, and organizations. Thus, studying intergroup behavior is crucial for improving our understanding of social interactions. Intergroup behavior is typically defined as a situation where *"individuals belonging to one group interact, collectively or individually, with another group or its members in terms of their group identification"* (Sherif, 1966, p. 12).

Psychological research on intergroup phenomena primarily highlights ingroup favoritism and the outgroup homogeneity effect. Ingroup favoritism refers to the tendency to prefer and act more positively toward members of one's own group (Tajfel et al. (1971)). The outgroup homogeneity effect, on the other hand, describes the perception that outgroup members are more similar to one another than ingroup members. The social psychology literature reveals a broad consensus about the existence of an outgroup homogeneity effect, but various theories have been proposed regarding the underlying cognitive mechanisms.

Linville et al. (1989) argue that familiarity must play a role. They link perceived variability to a memory retrieval process: because ingroup members are more familiar with each other, they are able to retrieve more exemplars from their memory in ingroup interactions compared to outgroup interactions. In this paper, familiarity with the ingroup is captured by the assumption that players in the ingroup are more likely to have informal interactions with one another. This is a realistic assumption when groups divide players based on some notion of proximity (e.g., neighborhoods or geographical proximity in a society; plants or subdivisions in organizations; etc.).

Judd and Park (1988) and Park et al. (1991) argue that familiarity alone cannot explain the evidence on the outgroup homogeneity effect in minimal groups.[1]

---

[1] The minimal group paradigm involves dividing experimental subjects into arbitrary groups (e.g., by flipping a coin). Intergroup phenomena have been observed in these settings as well. Although, Ostrom et al. (1993) points out that results of outgroup homogeneity seem to be less robust in settings with minimal groups.

They posit that encounters with both ingroup and outgroup members are stored in memory via categories and that more categories are dedicated to ingroup interactions. Finally, Ostrom et al. (1993) propose that the processes underlying ingroup and outgroup interactions are fundamentally different. Ingroup interactions involve storing encounters in memory via person-categories, whereas information about the outgroup is structured as an attribute-based category. The approach in this paper is closer to the latter view.

In this paper, intergroup phenomena are analyzed by positing that players form expectations based on the outgroup homogeneity effect. Specifically, the framework consists of a set of heterogeneous players, denoted by $\mathcal{I}$, where heterogeneity is captured by payoff-relevant individual characteristics, and players choose actions from a common action space A. It is assumed that players can accurately select the optimal action based on their expectations. The set $\mathcal{I}$ is partitioned into two subsets, each representing a group. For a player i belonging to group G, the other players in G are referred to as the *ingroup*, while those not in G constitute the *outgroup*.

Players are exogenously sorted into pairs according to a deterministic function ϕ. Each pair engages in a two-player simultaneous-move game, where individual payoffs depend on the characteristics of both players. The individual characteristics of all players, as well as the matching ϕ, are assumed to be commonly known.

Players can be matched either within the ingroup or with the outgroup. In the case of ingroup matches, it is assumed that players form expectations that are finely tuned to the behavior of their opponents. In equilibrium, these expectations must coincide with the strategy of the opponent to whom they are matched, ensuring that the equilibrium behavior corresponds to the strategies prescribed by Nash equilibrium for these matches. Conversely, when players are matched with the outgroup, they are assumed to form *coarse* expectations, relying on aggregate statistics of behaviors in the outgroup. In equilibrium, these expectations are required to coincide with the average distribution of strategies among all outgroup players (including those interacting within the ingroup). Clearly, in these matches, the equilibrium behavior may differ from the predictions obtained through more standard approaches.

The assumption of coarse expectations for outgroup interactions captures a stark version of the outgroup homogeneity effect. This is not far from the cognitive process proposed by Ostrom et al. (1993): players form finely tuned expectations when interacting ingroup ("person-category") and form coarse expectations when interacting outgroup (a prototype in the "attribute-category").

From another perspective, the difference in expectations for ingroup and outgroup interactions is consistent with a situation where players are exposed to two different datasets. For ingroup interactions, each data point consists of both the behavior and the identity of members in the ingroup, whereas data is anonymized for outgroup interactions. Players are then forced to rely on aggregate statistics to make sense of the latter dataset. Interestingly, this view is somewhat consistent with findings in psychology: Linville (1982) finds that subjects are able to retrieve more complex information from memory about the ingroup than the outgroup.

For a given matching ϕ, the equilibrium concept used in this paper can be viewed as a special case of the analogy-based expectations equilibrium (ABEE) proposed by Jehiel (2005), which posits that players bundle situations into categories and form expectations for each category relying on aggregate statistics (see Jehiel (2022) for a survey of this literature). The framework proposed in this paper can be interpreted as an analogy-based expectations framework where each player partitions other players in the following way: each ingroup member is placed in a singleton cell of the partition, while the outgroup players are all bundled together.[2]

A novel aspect, compared to the literature on ABEE, is that, in the current setup, players are sorted according to a function ϕ. The sorting governs the pairwise interactions and allows for the study of societies that are matched assortatively, anti-assortatively, or independently of players' individual characteristics.[3] Players are assumed to know the function ϕ. In particular, each player observes the individual characteristics of their match. However, players do not make inferences from ϕ when forming coarse expectations. That is, players do not make strategic considerations upon observing the individual characteristics of their match but simply rely on data about outgroup behavior to form their expectations. This requires players to treat

---

[2] Some recent papers on ABEE aim to endogenize the partitions (Jehiel and Weber (2025), Jehiel and Mohlin (2023), Goursat et al. (2025)), which were originally introduced as a primitive of the model. In this paper, the partitions are given exogenously because the scope of the paper is to study the effect of interactions across social groups, rather than the formation of social groups.

[3] ϕ expresses formal pairwise interactions, while informal interactions are not modeled in the framework (see the discussion on familiarity above).

the data as independent of the sorting and the individual characteristics of players in the outgroup, and it can be viewed as a form of correlation neglect.[4]

Next, the consequences of sorting and the outgroup homogeneity effect are analyzed in two economic scenarios.

First, an illustration is provided in section 3 through coordination games à la Carlsson and van Damme (1993), where players choose whether to invest or not. Groups are symmetric, and payoffs depend on the individual characteristics of both players, who are matched according to ϕ. In this setting, there are multiple Nash equilibria, and the literature typically focuses on equilibrium selection.

In the context of this illustration, it is established that when one match is such that both players find it strictly dominant to invest, the outgroup homogeneity effect induces all players to invest, provided that all are matched outgroup and in an assortative fashion.

The main goal of this illustration is to clarify the difference between the current framework under complete information and the more standard approach of Bayes–Nash equilibrium in games of incomplete information. To this end, a Bayesian game is introduced in the context of coordination games, where the matching is unobserved but there is a commonly known distribution over matchings. It is then shown that no distribution can uniquely induce all players to invest in a Bayes–Nash equilibrium. In particular, multiple Bayes–Nash equilibria arise for any mixture over matchings. The key insight of this illustration is that uniqueness can be achieved via the outgroup homogeneity effect when this is not feasible under standard approaches.

In section 4, an application to organizational contexts is proposed. Organizations typically comprise several distinct "groups" (e.g., departments, offices, floors, or plants), with possibly limited –informal– interactions among agents from different groups due to the organizational structure. More specifically, the application considers a setting where the groups represent subdivisions, and each match corresponds to a team. Each match involves a simultaneous-move game exhibiting strategic complementarities in a private value setting.

A designer is required to maximize the overall probability of success of the tasks

---

[4] Eyster and Rabin (2005) study correlation neglect in strategic environments of incomplete information and draw a link between fully-cursed and ABEE with coarsest analogy partitions.

in a pure moral hazard setting with risk-neutral agents and limited liability. The probability of success of each task depends linearly on the effort exerted by players in that match and is supermodular in the efforts. The designer can only affect the equilibrium efforts via the matching ϕ.[5] For general groups, some necessary conditions for the optimal matching are provided. In particular, given the subsets of agents that are matched ingroup/outgroup, the optimal ϕ should match the agents assortatively within their subset.

The attention is then restricted to organizations with a large number of agents. It is established that when groups are symmetric, the optimal matching takes the form of a threshold: all agents with costs above the threshold are assortatively matched outgroup, while the rest are assortatively matched ingroup. Finally, an iterative procedure is proposed to find this optimal matching. This consists of starting by matching all agents ingroup assortatively and progressively rematch outgroup those agents who exert effort below average, where the average increases at each rematch. This can be interpreted as the designer trying various matchings and improving the objective function over time, thereby reducing the tension in the model between an omniscient designer and coarse agents.

Section 5 discusses various extensions of the model. First, the case of decentralized matching is examined in the context of large organizations, and it is established that the threshold property arises under this alternative approach as well. However, the designer's optimal solution additionally requires that outgroup matches be assortative. The discussion then turns to how the current modeling approach can be adapted to capture alternative effects such as ingroup homogeneity (Brewer (1993)). Finally, it is noted that the framework can be extended by augmenting the game with a network, where players form correct expectations about the behavior of those to whom they are linked, and coarse expectations otherwise. Groups are derived from a specific network structure in which the vertices are partitioned into subsets and each vertex is linked to all (and only) others within its subset. From this perspective, the current notion can be viewed as a refinement of Peer-confirming Equilibrium (Lipnowski and Sadler (2019)).

---

[5] The more typical channel of bonus schemes is shut down by assuming a fixed bonus scheme.

1.1 Related Literature

This paper is part of the body of research in behavioral game theory that aims to address various forms of cognitive limitations and misperceptions (see, for instance, Osborne and Rubinstein (1998) on sampling equilibrium or Jehiel and Samet (2007) on coarse evaluation). The notion of equilibrium introduced in this paper can be interpreted as an analogy-based expectation equilibrium (Jehiel (2005)) with multiple players and a specific assumption on the analogy partition. As such, it can also be formulated under the general frameworks of Berk-Nash equilibrium (Esponda and Pouzo (2016)) and conjectural equilibrium (Battigalli (1987)). Spiegler (2016) explores misperceptions of causal relationships between relevant variables for a decision maker via the language of Directed Acyclic Graphs, and Spiegler (2020) offers an interpretation of coarseness in terms of omitted variables through this approach. This fits with the interpretation of coarse expectations as arising due to anonymized data in this paper. One novelty with respect to the aforementioned approaches is the formal introduction of the matching function, which affects coarse expectations.

In a related approach, Eyster and Rabin (2005) examine misperceptions regarding the impact of private information on behavior. They propose contrasting the standard concept of Bayes-Nash equilibrium with the case where a player neglects the correlation between the strategy and the type of the opponents ("fully-cursed"). In the latter case, expectations are correct on average, but each opponent is mistakenly perceived as playing according to the same distribution, regardless of her type.[6] They posit that players best-respond to a convex combination of the expectations of a Bayesian player and a fully-cursed player, where the weight can be seen as the degree of cursedness. In a similar fashion, the current paper takes a standard concept at one extreme (Nash equilibrium for ingroup matches) and the coarsest (outgroup) analogy partition at the other extreme (for outgroup matches). However, here the coarse expectations are determined only by actual equilibrium behaviors aggregated across players in the outgroup.

This paper is also related to recent literature that provides novel perspectives on intergroup phenomena through the lens of bounded rationality. In the context of social identity, Liqui Lung (2022) points out that if decision makers have a noisy

---

[6] The case with fully-cursed players coincides with the coarsest analogy partition in ABEE under incomplete information as defined in Jehiel and Koessler (2008).

perception of their own ability, they can improve their decision making on average by relying on, possibly irrelevant, aggregate statistics of their social group. However, this can lead to persistent self-driven asymmetries across social groups. From a statistical discrimination perspective, Chauvin (2020) argues that the Fundamental Attribution Error can lead to a self-enforcing cycle where agents mistakenly attribute outcome gaps to trait differences, which leads to discrimination, further fostering the outcome gaps. In this strand of literature, the most related paper is Frick et al. (2022). They propose a model where social interactions generate sample of behaviors for those involved in the interactions and these occur in as assortative fashion. Neglecting the assortativity, each player treats her sample of behaviors as representative of the whole population. This misperception leads to outcomes of socioeconomic disparities. Similarly to Frick et al. (2022), in this paper players neglect the effect of the matches on the distribution of behaviors. The main difference is that, unlike Frick et al. (2022), in this paper players are exposed to the aggregate distribution of behaviors over the whole outgroup.

2 Framework

Strategic Environment. Consider a set $\mathcal{I}$ containing an even number of players, who interact in two-player normal-form games under complete information. The pairwise interactions are governed by ϕ, which is a complete one-to-one matching in $\mathcal{I}$. The set of all such matchings is denoted by Φ, and $\phi(i)$ denotes the player $j \in \mathcal{I}$ that is matched with i according to $\phi \in \Phi$. The matching is given exogenously.

Each player i has payoff-relevant individual characteristics $x_i \in X$ and available actions $a_i \in A_i$. The action space is assumed to be common to all players: for each i, $A_i = A$. I denote by $\sigma_i \in \Delta A$ a mixed strategy of player i, and $\sigma = (\sigma_i)_{i \in \mathcal{I}}$ is a profile of strategies.

The payoff of player i matched with j is described by a Von Neumann-Morgenstern utility function, where $u_i(a_i, a_j; x_i, x_j)$ is the utility obtained by player i, with characteristics x_i, when she plays action a_i, given that player j, with characteristics x_j, plays a_j.

Given the matching ϕ, the match $(i, j) \in \phi$, characteristics x_i and x_j, and strategies σ_i and σ_j, the expected utility of player i is given by:

$$U_i(\sigma_i, \sigma_j; x_i, x_j) = \sum_{a_i, a_j} \sigma_i(a_i)\sigma_j(a_j)u_i(a_i, a_j; x_i, x_j).$$

Players are assumed to observe the matching ϕ and the characteristics x_i and x_j when choosing their actions.

Given ϕ, a strategy profile σ is a Nash equilibrium if, for every $i \in \mathcal{I}$, and $\sigma_i' \in \Delta A$,

$$U_i(\sigma_i, \sigma_{\phi(i)}; x_i, x_{\phi(i)}) \geq U_i(\sigma_i', \sigma_{\phi(i)}; x_i, x_{\phi(i)}).$$

Groups. I assume that players are divided into two groups. Players within a group are viewed as being familiar with one another, so that they are able to know or learn the strategy of their opponent whenever the opponent is from the same group. Therefore, for these types of interactions, the equilibrium play coincides with what is prescribed by Nash equilibrium.

Instead, players are assumed to learn only the aggregate behavior of players from different groups, where the aggregation is over all the strategies of players in a group.

The familiarity between members of the same group is not captured directly by the model, and it can be interpreted in different ways depending on what the groups represent.[7] For instance, one can think of groups as neighborhoods, with familiarity arising as a byproduct of frequent *informal* interactions among neighbors.[8] Another interpretation of familiarity is in terms of culture. One could view groups as, say, nationalities, and familiarity could derive from the fact that players share traditions and customs, or alternatively, groups could be interpreted as organizations (or subdivisions within an organization), each with its own culture (or subculture).

More formally, the set of players $\mathcal{I}$ is assumed to be partitioned into two groups, namely G_1 and G_2. I assume that in each group players are ordered so characteristics

---

[7] The idea of familiarity in relation to the outgroup homogeneity effect is discussed in Judd and Park (1988).

[8] A different, but somewhat similar, interpretation in terms of neighborhoods can be found in Frick et al. (2022), where agents are assumed to sample behaviors more frequently from their neighbors, but they treat the sample as representative of the whole population.

are weakly increasing. The notation $G(i)$ is used to refer to the group G that contains player i.

The matching ϕ determines whether player i interacts with a player from the same group or not. I say that player i is matched *ingroup*, according to ϕ, if $\phi(i) \in G(i)$, and player i is said to be matched *outgroup* otherwise.

The set of players in G that are matched ingroup according to ϕ is denoted by I_G^ϕ, while O_G^ϕ denotes the set of players in G that are matched outgroup according to ϕ.

Given ϕ, player i forms expectations about the behavior of her opponent according to $\beta_i(\sigma) \in \Delta A$, which maps from the set of strategy profiles $(\Delta A)^{|\mathcal{I}|}$ to ΔA. A profile of expectations is denoted by $\beta \equiv (\beta_i(\sigma))_{i \in \mathcal{I}}$.

A profile of expectations β is said to be *consistent* with σ, whenever, given ϕ, for all $i \in \mathcal{I}$,

$$\beta_i(\sigma) = \begin{cases} \sigma_{\phi(i)} & \text{if } \phi(i) \in G(i) \\ \sum_{j' \in G(\phi(i))} \frac{\sigma_{j'}}{|G(\phi(i))|} & \text{if } \phi(i) \notin G(i) \end{cases}$$

In other words, consistency means that the expectations are finely tuned to individual behaviors for ingroup interactions, while they correctly represent the aggregate behaviors in the outgroup when the play is governed by σ. The modeling of expectations for outgroup interactions as coarse captures a stark version of the outgroup homogeneity effect. It should be noted that the aggregate behavior in a group also includes the strategies of players interacting within their group.

A strategy σ_i is said to be a *best-response* to $\beta_i(\sigma)$. given ϕ, whenever for all $a_i \in A$,

$$U_i(\sigma_i, \beta_i(\sigma); x_i, x_{\phi(i)}) \geq U_i(a_i, \beta_i(\sigma); x_i, x_{\phi(i)})$$

In other words, players are assumed to be able to correctly perform the utility maximization problem, given their expectations. In this sense, player i best responds as if $\phi(i)$ played $\beta_i(\sigma)$.

Definition 1 *Given the strategic environment, groups (G_1, G_2), and matching ϕ, a profile of strategies σ is an equilibrium if and only if there exists a profile of expectations β such that:*

1. *For each player i, σ_i is a best response to β_i*

2. β is consistent with σ

Some comments are in order here.

Nash Equilibrium as a special case. The equilibrium concept introduced above has Nash Equilibrium as a special case when one group is empty.

More precisely, when the groups are $G = \mathcal{I}$ and $G' = \emptyset$, all agents are necessarily paired *ingroup*, and so definition 1 corresponds to Nash equilibrium. This is because, by consistency, each player i has correct expectations about the strategy of $\phi(i)$, and then she chooses σ_i so as to maximize $U_i(\sigma_i, \sigma_{\phi(i)}; x_i, x_{\phi(i)})$. More generally, the two definitions coincide whenever the matching ϕ is such that all agents are paired *ingroup*. Clearly, the equilibrium actions prescribed by definition 1 can differ from Nash equilibrium for general groups and matchings.

Related notions of bounded rationality. The equilibrium defined in (1) can be thought of as a specific formulation of several notions of equilibrium in the bounded rationality literature.

The closest one is the analogy-based expectation equilibrium (ABEE), first introduced for settings with multiple stages by Jehiel (2005), and extended to model games of incomplete information by Jehiel and Koessler (2008) (see Jehiel (2022) for a survey). The notion presented in this paper can be viewed as an ABEE where each agent i partitions the members of the *ingroup* finely (each agent in $G(i)$ belongs to a singleton cell in the analogy partition of i), while the members of the *outgroup* are all bundled together into one cell of the analogy partition.[9] A benefit of thinking about the framework of this paper in terms of ABEE is that one could allow for less coarse categorizations of the outgroup (say, a finite number K of cells for the members of the outgroup, rather than $K = 1$ as it is currently assumed), thereby capturing less stark formalizations of the outgroup homogeneity effect.

From another perspective, the equilibrium presented in definition 1 can be interpreted through the language of Directly Acyclic Graphs (DAG) as proposed by Spiegler (2016), where agents are modeled as misperceiving the causal links across the parameters and variables in the model, thereby making decisions based on a subjective model (as opposed to the true DAG). In this paper, when agents are paired

---

[9] This also implies that definition 1 is a special case of Berk-Nash equilibrium (Esponda and Pouzo (2016)).

outgroup, they disregard the causal links that indirectly determine the final outcome (the expected payoff)—specifically, the characteristics of their opponent. For a given match, the individual characteristics of the agents are the sole parameters that truly determine the actions chosen by the players. However, when i is paired outgroup, she disregards the type of $\phi(i)$ and instead focuses on $\phi(i)$ belonging to the outgroup.

Outgroup Homogeneity Effect. The outgroup homogeneity effect is most commonly defined as the tendency to perceive ingroups as relatively more variable than outgroups (Judd and Park, 1988). We propose a stylized modeling of this effect through the consistency of expectations, where agents are required to hold the same expectations for all members of the outgroup. It should be noted that our approach relies on the perception of homogeneous behaviors within the outgroup, rather than the perception of homogeneous traits (or characteristics). This distinction allows us to disentangle the perception of individual traits from the perception of behaviors. The first instance could be modeled using standard tools of incomplete information, as is often done in the statistical discrimination literature, while the second focuses on the role of group "membership".

To illustrate the difference between incomplete information and the outgroup homogeneity effect, a version of global games is analyzed in the next section.

3 An Illustration with Global Games

In this section, coordination games in the vein of Rubinstein (1989) and Carlsson and van Damme (1993) are analyzed. This illustration primarily aims to highlight the different insights provided by the current framework in comparison to the more standard cases of incomplete information, in particular the literature originated by the latter paper.

Let the set of players be $\mathcal{I} = (b_1, \ldots, b_n, d_1, \ldots, d_n)$, where $x_{b_k} = x_{d_k} \equiv x_k = \frac{k}{n}(1 + \varepsilon)$ for $k = 1, \ldots, n$, and $0 < \varepsilon < \frac{1}{4n^2-1}$. Consider symmetric groups $G_z = \{z_1, \ldots, z_n\}$ for $z = b, d$.

Each agent must decide whether to Invest (I) or not (NI). Whenever i and j are matched according to ϕ, the payoffs are given by the following matrix:

	I	NI
I	$\frac{x_i+x_j}{2}, \frac{x_i+x_j}{2}$	$\frac{x_i+x_j}{2} - 1, 0$
NI	$0, \frac{x_i+x_j}{2} - 1$	$0, 0$

Note that player i's best response is to choose I if and only if player i's expectations are such that the probability of Invest being played is strictly greater than $1 - \frac{x_i+x_j}{2}$.

In this game, there are multiple Nash Equilibria. More specifically, there exists at most one pair that necessarily plays (I, I), namely the pair (b_n, d_n). For all other possible pairs (i, j), three (Nash) equilibrium plays can be sustained: (I, I), (NI, NI), or the mixed equilibrium where I is played with probability $1 - \frac{x_i+x_j}{2}$.

In contrast, under the equilibrium notion defined in (1), there exists a matching ϕ such that there is a unique equilibrium in which everybody invests.

Proposition 1 *Given groups G_b, G_d, let ϕ be such that $\phi(b_k) = d_k$, for $k = 1, \ldots, n$. Then, all players Invest (play I) in the unique equilibrium.*

The result is driven by the pair (b_n, d_n) that finds it strictly dominant to play Invest, and the fact that all agents are matched outgroup in an assortative fashion.

This can be seen as follows. Since all agents are matched outgroup and players with characteristics x_n play I, every player i expects the opponent to play Invest with probability weakly above $\frac{1}{n}$: for each $i \in \mathcal{I}$,

$$\beta_i(\sigma)[I] \geq \frac{1}{n}.$$

Then, both players in the match (b_{n-1}, d_{n-1}) must choose I. This is optimal for both players in such a pair whenever they believe I to be played by their opponent with probability greater than $1 - \frac{x_{n-1}+x_{n-1}}{2} = \frac{1}{n} - \frac{n-1}{n}\varepsilon$, which is strictly smaller than $\frac{1}{n}$. This implies that, for each i,

$$\beta_i(\sigma)[I] \geq \frac{2}{n}.$$

Reiterating the argument yields that, at each k between 1 and $n-2$, the last $n-k$ agents must be playing I, which induces the pair (b_{k-1}, d_{k-1}) to also play I. Thereby obtaining a unique equilibrium where everybody plays I (see the Appendix for the full argument).[10]

The definition of groups G_b and G_d as symmetric and with equally-spaced players allows to illustrate starkly the logic of the argument in proposition 1, but it is not necessary for the result to hold. In fact, this can be relaxed considering the values $x_k = \frac{k}{n}(1+\varepsilon)$ for $k = 1, \ldots, n$ as thresholds over the real line and by ensuring that: in each group there is a sufficient number of players between each interval of consecutive thresholds; and there is a sufficient number of players with characteristics above 1.

In order to state such conditions explicitly, I introduce the sequence of thresholds $(\hat{x}_k)_{k=1}^{N+1}$ where

$$\hat{x}_k \equiv \frac{k}{N}(1+\varepsilon),\ k = 1, \ldots, N-1,\ \hat{x}_N = 1,\ \text{and } \hat{x}_{N+1} = 1+\varepsilon \tag{1}$$

where now $0 < \varepsilon < \frac{1}{4N^2-1}$. Let the set of players be $\mathcal{I}' = \{b'_j, d'_j\}_{j=1}^n$, where for each j, $x_{b'_j}$ and $x_{d'_j}$ are in $(\hat{x}_1, \hat{x}_{N+1})$, and $x_{b'_j}$ is possibly different $x_{d'_j}$. Consider now possibly asymmetric ordered groups, where $G_z = (z_j)_{j=1}^n$, for $z = b', d'$. Provided that at least one player in $\mathcal{I}'$ has characteristics strictly above one, the following is established:

Proposition 2 *Let $x_i > 1$ for some $i \in \mathcal{I}'$. Consider ordered groups $G_{b'}, G_{d'}$, each with $n \geq 1$ players. Let ϕ be such that $\phi(b_j) = d_j$, for $j = 1, \ldots, n$. If, for some $N \geq 1$, the sequence of thresholds $(\hat{x}_k)_{k=1}^{N+1}$ defined as in (1) is such that, for each group, the number of players with characteristic in $[\hat{x}_k, \hat{x}_{k+1})$ is non decreasing as k increases from 1 to N, then all players Invest (play I) in the unique equilibrium.*

Proposition 2 extends the insights of proposition 1 to asymmetric groups where the characteristics of the players follow less specific assumptions, following the same logic. Still, a lot of structure is imposed on the set of characteristics of the players. In particular, for each group, the number of players with characteristics in $[1, 1+\varepsilon)$ –which are induced to play I due to the assortative matching– should be no

---

[10] The logic of the argument in Proposition 1 is reminiscent of Winter (2004), where a designer can uniquely implement the desired strategy profile by choosing a bonus that makes an action strictly dominant for one agent, so that progressively lower bonuses are needed to ensure that all agents' best responses are single-valued, given the bonus scheme and the actions of the other players in the organization.

smaller than the number of those with characteristics in $[\frac{N-1}{N}(1+\varepsilon), 1)$. Although demanding, the condition ensures that the results of proposition 1 hold for non-generic cases. The condition becomes less demanding as the set characteristics becomes "denser". The argument (detailed in Appendix) follows the same logic as the proof of the previous proposition.

The literature stemming from Carlsson and van Damme (1993) is typically concerned with equilibrium selection under incomplete information. Carlsson and van Damme (1993) show that the risk-dominant Nash equilibrium is selected in coordination games when payoffs are perturbed. In this paper, players perfectly observe the payoffs of the game, but the outgroup homogeneity plays a similar role to the noise. Proposition 1 selects the payoff-dominant Nash equilibrium, rather than the risk-dominant one.[11] However, this is driven by the fact that the characteristics of the players are skewed towards positive values.

To illustrate this, the following remark establishes that if characteristics are skewed towards negative values, the equilibrium where no player invests would be selected.

Remark 1 *For* $z \in \{b, d\}$, *let* $G_z = (z_1, \ldots, z_n)$ *where* $x_{z_n} = -\varepsilon$ *and* $x_{z_k} = \frac{k}{n}(1-\varepsilon)$, $k = 1, \ldots, n-1$. *If* ϕ *is such that* $\phi(b_k) = d_k$, *for* $k = 1, \ldots, n$. *Then, all players not investing (play* NI*) in the unique equilibrium.*

The argument is analogous to the argument in proposition 1: pair (b_n, d_n) finds it dominant to play NI. Then $\beta_i(\sigma) \leq \frac{n-1}{n}$ and so the pair (b_1, d_1) must play NI as well. This in turn induces (b_2, d_2) to play NI and so on and so forth. The remark clarifies that the key insight of this illustration is the uniqueness of the equilibrium, rather than the specific equilibrium that is selected.

The analysis of this illustration is somewhat similar to Steiner and Stewart (2008). They study long-run outcomes of learning in a (large) family of games under complete information, where players' beliefs in one game are formed by extrapolation from similar games. In the context of coordination games, they show that a unique equilibrium play is selected in games where multiple equilibria could arise. This occurs through propagation from games where there is a unique equilibrium. This is not too far from the procedure that establishes uniqueness in Proposition 1.

---

[11] In this illustration, the Nash equilibrium (NI, NI) risk-dominates (I, I) when $x_i + x_j \leq 1$.

The pair (b_n, d_n) has a strictly dominant strategy, which affects the expectations of the pair (b_{n-1}, d_{n-1}), rendering their best response single-valued, and so on and so forth. The key modeling difference is that, in this paper, expectations are formed based on the aggregate behavior of the outgroup, rather than extrapolation across similar games.

Unobservable matching and Bayes-Nash equilibrium. It is worth stressing that the uniqueness result cannot be obtained under the standard paradigm, even when there is a random matching and its realization is not observed by the players.

Let $\mu \in \Delta\Phi$ be a probability distribution over the matchings in Φ. The probability that player $i \in \mathcal{I}$ is matched with player $j \in \mathcal{I}$ is denoted by

$$\mu_{i,j} \equiv \sum_{\phi \in \Phi} \mathbf{1}_{\{\phi(i)=j\}} \mu(\phi).$$

Clearly, no player can be matched with themselves, i.e., $\mu_{i,i} = 0$, and all players must be matched: for all i, $\sum_{j \in \mathcal{I}} \mu_{i,j} = 1$. Also, $\mu_{i,j} = \mu_{j,i}$. It is assumed that μ is common knowledge, but players do not observe the realization of μ.

The Bayesian game can be seen as follows. Nature picks the matching ϕ according to μ. The probability distribution μ and the characteristics $(x_i)_{i \in \mathcal{I}}$ are common knowledge. No player observes the move by Nature, nor does any player receive a signal about it.[12]

As usual, let $-i$ be a shortcut notation for all players other than i. Given $\sigma \in (\Delta A)^{\mathcal{I}}$ and $x \in X^{\mathcal{I}}$, the expected utility of player i in the Bayesian game is defined as

$$\begin{aligned}\bar{U}_i(\sigma_i, \sigma_{-i}; x_i, x_{-i}) &\equiv \sum_{j \in \mathcal{I}} \mu_{i,j} U_i\left(\sigma_i, \sigma_j; x_i, x_j\right) \\ &= U_i\left(\sigma_i, \sum_{j \in \mathcal{I}} \mu_{i,j} \sigma_j; x_i, \sum_{j \in \mathcal{I}} \mu_{i,j} x_j\right)\end{aligned}$$

---

[12] An alternative modeling approach entails assuming that agents receive a fully revealing signal when they are matched ingroup and no signal otherwise. For the current illustration, this alternative way of modeling the Bayesian game would not change qualitatively the results: the scope is to show that a random unobservable matching cannot induce a unique equilibrium. Under the alternative approach, one would then have to restrict attention to outgroup matches only, because, conditional on a pair receiving a fully revealing signal, multiple strategy profiles can be sustained in equilibrium for that pair.

where the equality holds due to the linearity of U_i in σ_j and x_j.

Note that the strategy of player i cannot be conditioned on the type of player j she faces. This effectively captures the notion that player i does not observe the type of the opponent, while knowing the distribution over types. From this perspective, μ can be thought of as the distribution over the finite set of types X.

The strategy profile σ is a Bayes-Nash equilibrium (BNE) if and only if, for each i,

$$\bar{U}_i(\sigma_i, \sigma_{-i}; x_i, x_{-i}) \geq \bar{U}_i(\sigma'_i, \sigma_{-i}; x_i, x_{-i}), \quad \forall \sigma'_i \in \Delta A.$$

The following proposition establishes that there always exists a Nash equilibrium of the Bayesian game where some agent is not playing Invest.

Proposition 3 *For all $\mu \in \Delta\Phi$, there always exist multiple Bayes-Nash equilibria. In particular, there is always a BNE where everyone plays I and a BNE where some player is playing Not Invest. That is, σ is a BNE and $\sigma_i = \{NI\}$ for some $i \in \mathcal{I}$.*

The result that σ is a BNE and $\sigma_i = \{NI\}$ for some $i \in \mathcal{I}$ can be understood in the following way. First, one can verify that the strategy profile σ where all players choose NI is a BNE under some conditions on μ. More specifically, if the probability that b_n is matched with d_n is weakly smaller than $1 - \frac{1}{2n}$, then such a σ is a BNE. Instead, whenever b_n and d_n are matched with probability above this threshold, it is necessarily the case that all other players are matched with b_n and d_n with relatively small probability. Because of this, an equilibrium can be sustained where all players choose NI, except for the players b_n and d_n, who play I (for the full argument for this equilibrium in the Appendix).

The multiplicity of the equilibria can then be easily established by observing that, not only is there a BNE where some player chooses NI, but there is also a BNE where all players invest, for all μ. This is readily verified by checking that I is a best response, given that all players are playing I: for each i, I is best response, given $\sigma_{-i} = (\{I\})_{j \in \mathcal{I} \setminus \{i\}}$ and μ, if and only if

$$1 = \sum_{j \in \mathcal{I} \setminus \{i\}} \mu_{i,j} \sigma_j(I) \geq 1 - \frac{x_i + \sum_{j \in \mathcal{I}} \mu_{i,j} x_j}{2}$$

which is always true because $\frac{x_i + \sum_{j \in \mathcal{I}} \mu_{i,j} x_j}{2}$ is positive for all μ. Thus, $\sigma_i = \{I\}$ for all i is a BNE.

Therefore, there is an equilibrium where all players invest in the Bayesian game that has been introduced above, where the matching is unobservable and the probability distribution over matchings is common knowledge. However, no matching nor mixture over matchings can generate a unique (Nash) equilibrium for such a Bayesian game. This is in contrast with the findings of Proposition (1), which shows that, under the outgroup homogeneity effect, all players investing is the unique equilibrium when the right matching is picked.

One of the main reasons for this difference in equilibrium is that, under outgroup homogeneity, players' expectations on the opponents are disentangled from the characteristics of the opponents. Players observe the characteristics of their match, but make no inference on her behavior based on such characteristics. More specifically, under the outgroup homogeneity effect, a player i mistakenly believes the behavior of the opponent to correspond to the aggregate behavior in the outgroup, despite the correct assessment of $u_i(\cdot, \cdot; x_i, x_{\phi(i)})$, where $\phi(i) \notin G(i)$. This allows to induce b_n and d_n to play Invest uniquely by matching them together, while affecting every player's expectations by matching all players outgroup. Whereas, in the Bayesian game, whenever b_n and d_n are matched with probability one, their behavior cannot affect the beliefs of other players.

Extensions. This illustration is intended to clarify the difference between a Bayesian game with unobservable matching and the outgroup homogeneity effect, as introduced in Section 2. The current analysis relies on specific assumptions about the payoff-relevant characteristics of the players, which allow to make the point is an immediate way. However, the illustration also draws a parallel between the information structure in coordination games and the outgroup homogeneity effect. Carlsson and van Damme (1993) relies on noise that follows a continuous distribution. To explore the differences more deeply with the literature on global games, the current illustration could be extended to allow for a continuum of players with individual characteristics distributed over some subset of the real line, including agents with characteristics both below 0 and above 1.

4 An Application to Organizations

This section considers a simple setting of (pure) moral hazard in teams (Holmstrom (1982)). A designer, who knows the agents' costs, matches agents into teams. Agents are risk-neutral, and there is limited liability. The usual channel for incentivizing effort is through a bonus scheme. To focus on the optimal matching that arises when the outgroup homogeneity effect plays a role, this channel is shut down by assuming a fixed bonus scheme.

Agents are divided into groups, which can be thought of as subdivisions of the organization. The coarse expectations in this application can be interpreted as the shared view of agents in one subdivision about the performance of those in another subdivision. The idea that culture can be identified as a shared view within the organization is proposed in Pettigrew (1979) and discussed more recently in Gibbons and Prusak (2020). Gibbons et al. (2021) study the effect of shared cognitive frames in interactions and relate this to culture in organizations. From this point of view, coarse expectations toward other subdivisions can be interpreted as a byproduct of the organization's culture (a subculture for each subdivision, with each subculture arising through the same process).

Strategic environment. Consider a setting with n identical two-person tasks characterized by $t \in (0,1)$, a parameter that captures the intensity of strategic complementarity in the tasks, and a finite set of heterogeneous agents, $\mathcal{I} = \{1, \dots, 2n\}$. Agent i exerts effort $e_i \in [0, \bar{e}]$ at a cost $c_i(e_i) = x_i \frac{e_i^2}{2}$, where $x_i \in [\underline{x}, \bar{x}] \subset \mathbb{R}_+$, so agents differ in their cost of effort. Agents are ordered based on their characteristics: $x_1 \leq x_2 \leq \cdots \leq x_{2n}$. A profile of efforts is defined as $e = (e_1, e_2, \dots, e_{2n})$.

When two agents are paired, each is rewarded through a fixed bonus scheme: each is paid 1 if the task is successful and 0 otherwise. Therefore, when agent i is matched with agent j, her utility is given by:

$$u_i(e_i, e_j; x_i, x_j) = \mathbb{E}[b_i] - c_i(e_i)$$

where b_i is the agent's bonus. As a consequence of the structure of the bonus scheme, the expected bonus $\mathbb{E}[b_i]$ is simply equal to the probability of success of the task. The probability of success is assumed to be determined by the function

$$p(e_i, e_j) \equiv \frac{e_i + e_j + te_ie_j}{2\bar{e} + \bar{e}^2}$$

which is linear in the effort of each agent and is normalized to ensure that it lies between 0 and 1. The cross derivative of $p()$ with respect to the efforts of the agents is equal to t, thereby implying that t represents the degree of strategic complementarity by construction.[13]

Designer's Problem. A designer knows the agents' characteristics and faces the problem of finding the matching $\phi \in \Phi$ that maximizes the sum of the probabilities of success of the tasks. More formally, the designer's problem is the following: given $t \in (0, 1)$,

$$\max_{\phi\in\Phi} \sum_{(i,j)\in\phi} p(e_i, e_j) \tag{2}$$

where e_i and e_j are the efforts exerted by agents i and j when they are paired.

The characteristics of the agents and the intensity parameter t are common knowledge. The intended sequence of events is as follows: the designer chooses the matching; agents observe the matching and choose effort simultaneously; the outcomes of the tasks are realized and observed; payments are made.

4.1 Groups and equilibrium notion

Let the set of agents $\mathcal{I}$ be partitioned into two groups, G_1 and G_2. As before, given a match (i, j), agent i's expectations regarding effort are denoted by $\beta_i(e)$. Since the utility of agent i is linear in the effort of the opponent, agent i only cares about the mean expected effort, $\mathbb{E}[\beta_i(e)] \in [0, \bar{e}]$. With slight abuse of notation, $\beta_i(e)$ is used to denote the mean expected effort. Given a matching ϕ, a profile of expectations is denoted as $\beta = (\beta_i(e))_{i\in\mathcal{I}}$.

As before, β is said to be *consistent* with e, given ϕ, whenever for all i,

$$\beta_i(e) = \begin{cases} e_{\phi(i)} & \phi(i) \in G(i) \\ \sum_{j'\in G(\phi(i))} \frac{e_{j'}}{|G(\phi(i))|} & \phi(i) \notin G(i) \end{cases}$$

---

[13] When U_i is twice differentiable, the degree of strategic complementarity is given by $\frac{\partial^2 U_i}{\partial e_i \partial e_j}$ (see Vives (1999) for a textbook definition).

where $\phi(i)$ is the agent paired with i under ϕ.

Whenever agents are matched with outgroup partners, they form expectations by aggregating the behaviors of all agents outside their own group $G(i)$. Consistency then implies that all agents in group G who are paired with someone in G' share the same belief about the effort of their task-mate. Thus, agents in group G share a common view of the behavior of agents in group G'. These shared views can be interpreted as capturing the reputations of the various groups within the organization. In particular, the mean action serves as the simplest representation of aggregate behavior within a given group.

The description of agents as coarse within the organization might appear to be in tension with the notion of an omniscient designer who can foresee the effects of various matchings in order to identify the optimal one—particularly so when the organization comprises many agents. However, the designer, being aware of the underlying coarseness in the organization, can be interpreted as experimenting with different matchings over time, receiving feedback on their effects, and gradually improving the objective function. This process eventually leads to the optimal matching. An iterative procedure that the designer might follow to implement this approach in the context of symmetric groups with a large number of agents is discussed at the end of subsection 4.3.

4.2 Analysis

This section first establishes the existence and uniqueness of equilibrium. Then, it investigates the designer's problem.

In the equilibrium analysis, the matching ϕ is held fixed. Given ϕ, agents interact in pairs in a simultaneous game with complete information.

The Nash equilibrium can be derived as follows. The assumptions $\underline{x} \geq 1$ and $\bar{e} \geq 1$ ensure that, for each i, the best response lies in the interior of the action space. This is because the utility function of agent i is strictly concave in e_i, given $e_{\phi(i)} \in [0, \bar{e}]$. The best response is thus:

$$e_i = \frac{1 + te_{\phi(i)}}{c_i}$$

where $c_i \equiv x_i(2\bar{e} + \bar{e}^2)$, for convenience of notation.

The resulting unique Nash equilibrium requires that, given ϕ, for each i,

$$e_i^{NE}(\phi) = \frac{c_{\phi(i)} + t}{c_i c_{\phi(i)} - t^2}$$

This is also the equilibrium attained under Definition 1 when ϕ matches all agents ingroup. More generally, the effort exerted in equilibrium by an agent $i \in I_G^\phi$ coincides with the effort prescribed by the Nash equilibrium. The aggregate effort exerted by all agents interacting ingroup in G can thus be written compactly as

$$e(I_G^\phi) \equiv \sum_{i \in I_G^\phi} \frac{c_{\phi(i)} + t}{c_i c_{\phi(i)} - t^2}. \tag{3}$$

The following result establishes that, for G_1 and G_2, there exists a unique equilibrium under the same assumptions used in the Nash equilibrium analysis.

Proposition 4 *Consider the strategic environment, groups (G_1, G_2), and a matching ϕ. If $\underline{x} \geq 1$ and $\bar{e} \geq 1$, there exists a unique equilibrium as defined in 1, where for each G, and for each $i \in G$, the equilibrium efforts are given by*

$$e_i(\phi) = \begin{cases} \frac{c_{\phi(i)}+t}{c_i c_{\phi(i)} - t^2} & \text{if } i \in I_G^\phi \\ \frac{1 + t\eta_{G'}(\phi)}{c_i} & \text{otherwise} \end{cases}$$

where $\eta_{G'}(\phi)$ denotes the average effort in the group $G' \neq G$, defined as

$$\eta_{G'}(\phi) = \frac{|G| \left(e(I_{G'}^\phi) + \sum_{i \in O_{G'}^\phi} \frac{1}{c_i} \right) + t \left(\sum_{i \in O_{G'}^\phi} \frac{1}{c_i} \right) \left(e(I_G^\phi) + \sum_{i \in O_G^\phi} \frac{1}{c_i} \right)}{|G||G'| - t^2 \left(\sum_{i \in O_G^\phi} \frac{1}{c_i} \right) \left(\sum_{i \in O_{G'}^\phi} \frac{1}{c_i} \right)} \tag{4}$$

The equilibrium efforts and expectations are derived using standard techniques (see the Appendix for details), while uniqueness follows from the fact that the best response is single-valued due to the strict concavity of the utility function and the assumptions $\underline{x} \geq 1$ and $\bar{e} \geq 1$, which are assumed to hold for the remainder of the paper.

Notably, the average effort in group G is unaffected by how agents are paired outgroup. That is, $\eta_G(\phi') = \eta_G(\phi)$ whenever the ingroup pairings are unchanged.

More formally, $\eta_G(\phi') = \eta_G(\phi)$ if, for each $G' \in \{G_1, G_2\}$, (i) $I^{\phi'}_{G'} = I^{\phi}_{G'}$, and (ii) $i \in I^{\phi'}_{G'}$ implies that $\phi'(i) = \phi(i)$.

This property reflects the fact that the utility of agent i does not depend directly on the characteristics of the opponent, namely $x_{\phi(i)}$. This can thus be interpreted as a framework with private values.

This contrasts with the global games framework discussed in Section 3. The distinction matters: in a private values framework, whenever all agents are matched outgroup, the equilibrium defined in Definition 1 can be replicated in a Bayesian game in which any agent in group G is uniformly randomly matched with agents in group G', and the realizations of the random matching are not observed.[14] However, the equilibrium may still differ when some players are matched ingroup and others outgroup, because the behavior of ingroup pairs in G affects the aggregate behavior in G and thus the expectations of all agents matched outgroup in G'. This is illustrated in the following example.

Example 1 *Consider the set $\mathcal{I} = \{b_1, d_1, b_2, d_2, b_3, d_3\}$, with $c_{b_k} = c_{d_k} = c_k$ for $k = 1, 2, 3$ and $c_3 > c_2 = c_1 + \xi$, where $\xi > 0$ is arbitrarily small. Note that $\frac{1}{c_1 - t}$ is an upper bound on the level of effort that can be sustained in any equilibrium, with ϕ and G, G' varying. Similarly, $\frac{1}{c_3 - t}$ is a lower bound on the level of effort that can be exerted in equilibrium, letting ϕ and G, G' vary.*

Let the groups be given by $G_z = \{z_1, z_2, z_3\}$ for $z = b, d$. Consider the matching ϕ, where in each group the agent with the highest cost is matched outgroup, and the other agents are matched ingroup. Specifically, $O^{\phi}_{G_z} = \{z_3\}$ and $I^{\phi}_{G_z} = \{z_1, z_2\}$ for $z = b, d$.

In this example, such a matching induces the agents matched ingroup to exert the maximal effort achievable (as $\xi \to 0$), and it induces the agents matched outgroup to overexert effort.

More precisely, the equilibrium efforts of the agents matched ingroup are approximately $\frac{1}{c_1 - t}$, since ξ is close to zero. This is the upper bound on the equilibrium efforts.

In contrast, the agents matched outgroup best respond to the average effort in the outgroup. Following Proposition 4, the average effort in group G can be found

---

[14] In Section 3, it is shown that in a setting without private values, the equilibrium notions can differ even when all players are matched outgroup.

by applying equation (4), which yields

$$\eta_G(\phi) \approx \frac{2\frac{1}{c_1-t}+\frac{1}{c_3}}{3-t\frac{1}{c_3}} = \frac{2c_3+c_1-t}{(c_1-t)(3c_3-t)}.$$

The average effort is the same in both groups and is greater than the effort exerted in equilibrium by agent $i \in \{b_3, d_3\}$. *This is because* $e_i(\phi) < \eta_G(\phi)$ *if and only if* $\eta_G(\phi) > \frac{1}{c_i-t}$, *where* $i \in \{b_3, d_3\}$. *The latter inequality always holds because* $\frac{1}{c_3-t}$ *is a lower bound on the equilibrium efforts.*

Thus, by choosing ϕ, *the designer can induce the agents with the highest costs to overexert effort (compared to the actual effort chosen by their match), while the other agents exert effort levels that are approximately at the upper bound of efforts that could be sustained in any equilibrium.* ■

This example shows that the designer can induce overexertion of effort by matching low-cost agents within the ingroup and high-cost agents within the outgroup. Such a level of overexertion cannot be replicated in a Bayesian game where the matching is unobserved. This is because inducing high effort from efficient agents requires them to believe they are likely matched within the ingroup, which, in equilibrium, implies that the least efficient agents know they are matched (in the outgroup) with another inefficient agent with high probability. Therefore, the fact that groups in the organization rely on aggregate statistics about the outgroup—as captured by coarse expectations—enables the designer to achieve a level of overexertion that cannot be replicated simply by not disclosing information about the matching under the standard paradigm.

The analysis now turns to the problem of choosing the matching that would maximize the overall probability of success of the tasks. Given $t \in (0,1)$, the designer's problem is given by

$$\max_{\phi\in\Phi} \sum_{(i,j)\in\phi} p(e_i(\phi), e_j(\phi)) \tag{5}$$

where $e_i(\phi)$ denotes the effort exerted by agent i according to Proposition 4. Whenever a matching is a solution to the maximization problem in (5), it is referred to as an *optimal matching*.

The example above illustrated that the designer can induce overexertion of effort, one of the key forces shaping the designer's problem. Another important force is the supermodularity of the objective function. While this property will be discussed in detail later, it can be roughly interpreted as the designer's preference for matching agents across the organization so that efforts are aligned in a positively assortative fashion.

The following definitions formally introduce the concept of assortative matching within this framework. Since matchings can occur either within the ingroup or the outgroup, we distinguish between two types of assortativeness: one-sided assortativeness, which refers to assortativeness within the ingroup, and two-sided assortativeness, which considers assortativeness across both groups.

Definition 2 *Let $I \subseteq \mathcal{I}$. The matching $\phi \in \Phi$ is said to be* ***one-sided assortative in*** *I if, for each $i \in I$,*

1. *$\phi(i) \in I$,*
2. *there is no $j \in I$ such that $\min\{c_i, c_{\phi(i)}\} < c_j < \max\{c_i, c_{\phi(i)}\}$.*

Definition 2 states that ϕ is one-sided assortative in some subset I if all agents in I are paired with another agent in I, and they are paired in order of their costs of effort.

Definition 3 *Let I_1, I_2 be ordered subsets of $\mathcal{I}$ with the same cardinality and let i denote the i-th element in I, for $I \in \{I_1, I_2\}$. The matching $\phi \in \Phi$ is* ***two-sided assortative in*** *$I_1 \times I_2$ if,*

1. *for each $i \in I$, $\phi(i) \in I'$, for $I' \neq I \in \{I_1, I_2\}$,*
2. *for each $(i_1, i_2) \in \phi \cap (I_1 \times I_2)$, $i_1 = i_2$.*

Definition 3 states that a matching is two-sided assortative in $I_1 \times I_2$ if for all $1 \leq m \leq |I_1|$, the agent in the m-th position in I_1 is paired with the agent in the m-th position in I_2.

Some necessary conditions for the optimal matching can now be established (note that existence is guaranteed by the finiteness of the set Φ).

Proposition 5 *Consider the strategic environment and groups (G_1, G_2). Let ϕ be an optimal matching. Then:*

1. *for each G, if I_G^ϕ is non-empty, then ϕ is one-sided assortative in I_G^ϕ;*
2. *if $O_{G_1}^\phi$ is non-empty, then ϕ is two-sided assortative in $O_{G_1}^\phi \times O_{G_2}^\phi$.*

Conditions (1) and (2) in the proposition suggest a preference for assortativeness in the optimal solution. Specifically, the matching is one-sided assortative in each set of agents paired ingroup and two-sided assortative in the sets of agents paired outgroup. This preference for assortative matchings depends on both supermodularity of the objective function in the efforts of the agents and strategic complementarity. More precisely, condition (1) is mainly driven by the strategic complementarity of the best response, whereas for condition (2) the key property that plays a role is the supermodularity of the objective function (as each player matched outgroup best-replies to the average effort in the outgroup, regardless of the cost-type of their match).

by the supermodularity of the objective function in the cost of efforts of the pairs (see the Appendix for details), together with the strategic complementarity of best responses. Conditions (1) and (2) specify how agents are matched, keeping fixed the sets of agents matched ingroup and outgroup.

The next proposition provides (partial) conditions indicating whether agents should be paired outgroup or ingroup in the optimal matching.

Proposition 6 *Consider the strategic environment and groups (G_1, G_2). Let $b, b' \in G_B$ and $d, d' \in G_D$, for $B \neq D \in \{1,2\}$. Let ϕ be an optimal matching. Then:*

3. *There are no pairs $(b,d), (b',d') \in \phi$, where $c_b \neq c_{b'}$ or $c_d \neq c_{d'}$, such that*
    - $\min\{c_d, c_{d'}\} \geq \max\{c_b, c_{b'}\}$, *and*
    - $\min\{e_d(\phi), e_{d'}(\phi)\} \geq \max\{\eta_{G_B}(\phi), \eta_{G_D}(\phi)\}$;
4. *There are no pairs $(b,b'), (d,d') \in \phi$, where $c_b \neq c_{b'}$ or $c_d \neq c_{d'}$, such that*
    - $\min\{c_{b'}, c_{d'}\} \geq \max\{c_b, c_d\}$, *and*
    - $\max\{e_b(\phi), e_d(\phi)\} \leq \min\{\eta_{G_B}(\phi), \eta_{G_D}(\phi)\}$.

Condition (3) states that there cannot be two outgroup pairs where all agents are exerting efforts above average, provided that the costs of the agents in one group are greater than those in the other group. In such a case, the designer could improve the objective by forming ingroup pairs. Analogously, condition (4) states that there cannot be ingroup pairs where all agents are exerting efforts below average, provided that one agent per group has a lower cost than the remaining two agents. In this case, the designer could improve the objective by forming outgroup matches.

The main difficulty in fully characterizing the optimal matching arises from a potential trade-off between two forces: (i) the supermodularity of the function $p(\cdot,\cdot)$, which favors assortative matching on efforts, regardless of grouping, and (ii) the overexertion of effort induced by the outgroup homogeneity effect in equilibrium, which favors pairing lower-cost agents ingroup while matching higher-cost agents outgroup. Conditions (3) and (4) describe situations where these forces are aligned.

If the channel of the outgroup homogeneity effect is shut down, the optimal matching is readily found. This is established in the following corollary.

Corollary 1 *Given the strategic environment and groups $G = \mathcal{I}$ and $G' = \emptyset$, ϕ is a solution to (5) if and only if ϕ is one-sided assortative in $\mathcal{I}$.*

The "only if" direction of the statement is a special case of condition (1) in Proposition 5, applied to the set $\mathcal{I}$. This establishes that, given the groups specified in the corollary, any solution to the designer's problem (if it exists) must be one-sided assortative in $\mathcal{I}$.

The "if" direction follows by noting that a solution must exist because Φ is finite, and if more than one matching is one-sided assortative in $\mathcal{I}$, they all yield the same value for the designer.[15]

The result in Corollary 1 provides a solution to the designer's problem in the standard case (Nash equilibrium), and is reminiscent of the positive assortativeness in marriage markets with complementarities established in Becker (1973), with two caveats: in this framework, efforts are endogenously determined, and the matching occurs between agents within the same group.

The result in the corollary may be used as a benchmark, since the value of the designer's objective function in this case corresponds to what would be obtained if agents always formed correct expectations.

---

[15] The only case where two matchings ϕ and ϕ' can both be one-sided assortative in some set I is when there are $i, j \in I$ with $c_i = c_j$ such that $\phi(i) = \phi'(j)$ and $\phi(j) = \phi'(i)$.

The following remark clarifies that there is scope for improving upon the benchmark just established.

Remark 2 *If $n > 1$ and $c_i \neq c_j$ for some $i, j \in \mathcal{I}$, then there exist groups G_1, G_2 and a matching ϕ such that the designer's objective function attains a value strictly greater than in the benchmark in Corollary 1.*

This is readily verified by taking the groups $G_1 = \{|\mathcal{I}|\}$ and $G_2 = \mathcal{I} \setminus G_1$. Let ϕ be the matching where every agent in G_2 is paired assortatively within G_2, except for the agent with the second-highest cost in $\mathcal{I}$, namely $c_{|\mathcal{I}|-1}$, who is paired outgroup with the agent with the highest cost, $c_{|\mathcal{I}|}$. Then ϕ yields a higher value than the benchmark. This is because all pairs other than $(|\mathcal{I}|-1, |\mathcal{I}|)$ exert the same efforts in both cases, while agent $|\mathcal{I}|$ necessarily exerts higher effort as she best-replies to the average in G_2, inducing a higher effort from agent $|\mathcal{I}|-1$, thereby ensuring a higher value for the designer's objective function compared to the benchmark. Remark 2 shows that the outgroup homogeneity effect can, at times, be exploited by the designer.

The discussion above considers only cases in which the trade-off is shut down. Either the outgroup homogeneity effect cannot arise (when one group is empty), or the choice on the number of outgroup matches is forced (when one group is a singleton). What follows considers another special case in which the groups are identical. This case, however, allows the trade-off faced by the designer to be analyzed.

Groups G and G' are said to be *symmetric* if $G = G'$.

In this special case, for each agent in G_1, there is an identical agent in G_2, and in order to fully exploit the supermodularity of $p(\cdot,\cdot)$, the matching should pair identical agents together. However, under this setting, it is impossible to fully exploit supermodularity without forming some outgroup matches. In this sense, symmetric groups provide a simpler environment that is still rich enough to capture the essence of the trade-off.

To illustrate this intuition, consider the setting presented in example 1, where the set of players is $\mathcal{I} = \{b_1, d_1, b_2, d_2, b_3, d_3\}$, with $c_{b_k} = c_{d_k} = c_k$ for $k = 1, 2, 3$, and $c_3 > c_2 = c_1 + \xi$, with $\xi > 0$, and the groups are symmetric. That is, $G_z = \{z_1, z_2, z_3\}$ for $z = b, d$.

Letting ξ vary between 0 and $c_3 - c_1$, the full range of values for c_2 is covered. There exists a threshold $0 < \hat{\xi} \leq c_3 - c_1$ such that the optimal matching consists of matching the c_1 and c_2 agents ingroup in each group and the c_3 agents outgroup when $\xi < \hat{\xi}$, while for $\xi \geq \hat{\xi}$, the optimal matching requires that all agents be matched outgroup assortatively.

To clarify the underlying mechanism, suppose c_3 is very large, so that the equilibrium effort of any agent with cost c_3 is close to zero, regardless of the matching. If ξ is close to c_3, then agents with cost c_2 also exert effort close to zero regardless of the matching. In this case, if the more efficient agents with cost c_1 are matched ingroup, they best-respond to an effort close to zero by choosing an effort close to $\frac{1}{c_1}$. By contrast, if they are matched outgroup, they best-respond to the average effort, which is approximately $\frac{1}{3c_1}$, and thus overexert effort.

Note that this is a setting with private information. When the designer's optimal solution requires matching all agents outgroup, the resulting equilibrium behaviors can be replicated in a Bayesian game in which the matching is undisclosed and agents are uniformly randomly matched outgroup. However, when the optimal solution involves matching the most efficient agents ingroup and the rest outgroup, this structure cannot be implemented in a Bayesian game with undisclosed matching. The reason is that the ingroup matches raise the average effort, thereby inducing overexertion from outgroup matches. In order to elicit high effort levels from the ingroup matches, the matching must be disclosed to those agents. Yet such disclosure prevents the generation of overexertion in a Bayesian setting, as outgroup matches no longer factor in the ingroup when forming expectations about the effort of their match. From this perspective, the optimal solution in this setting yields a value for the designer's objective function that cannot be replicated in a standard setting in which the designer can choose the (possibly mixed) matching and whether to disclose it to each pair.

The structure found to be optimal in the setting of example 1 is conjectured to extend to symmetric groups with $n > 3$ agents. That is, there exists a threshold such that all agents with costs above (below) the threshold are matched outgroup (ingroup) assortatively. The next section analyzes the continuum case, viewed as an approximation of a large population, in order to verify this conjecture in that setting.

4.3 Large Organization

The case with a large number of finitely many agents in the organization is approximated by taking a continuum of agents for each group. This requires defining a matching over a continuum of agents. Following Gretsky et al. (1992), Gretsky et al. (1999), Nöldeke and Samuelson (2018), and Greinecker and Kah (2021), a distributional approach is used to model large organizations (as introduced by Hart et al. (1974) for large economies in the context of general equilibrium), where agents are formally identified by their characteristics and treated as perfectly divisible.

More precisely, a group G is defined as a continuum over the interval of costs $[\underline{c}, \bar{c}]$ with positive and finite Borel measure ν_G. Two groups are considered, G_1 and G_2, with measures ν_{G_1} and ν_{G_2}, respectively. The measure of agents in the whole organization is given by the Borel measure $\nu = \nu_{G_1} + \nu_{G_2}$, which means that for any Borel set $B \subset [\underline{c}, \bar{c}]$, then $\nu(B) = \nu_{G_1}(B) + \nu_{G_2}(B)$. In terms of interpretation, the specific values of the measures are inconsequential, while if the ratio $\nu_{G_1}(G_1)/\nu_{G_2}(G_2) > 1$, the interpretation is that group G_1 is larger than group G_2.

Since both ingroup and outgroup matches are allowed, ϕ is referred to as a *matching structure* $(In_1, Out_1, In_2, Out_2; \phi_O, \phi_1, \phi_2)$, which is defined in two steps.

First, the matching structure ϕ separates group G_k, $k = 1, 2$, into two Polish subspaces, namely In_k—for ingroup matches—and Out_k—for outgroup matches, with Borel measures v_{In_k} and ν_{Out_k} respectively, such that:

(i) $\nu_{In_k}(In_k) \equiv \nu_{G_k}(In_k) \geq 0$,

(ii) $\nu_{Out_k}(Out_k) \equiv \nu_{G_k}(Out_k) \geq 0$,

(iii) $\nu_{In_k}(In_k) + \nu_{Out_k}(Out_k) = \nu_{G_k}(G_k)$.

Also, the mass of players matched outgroup must be the same in the two groups: $\nu_{Out_1}(Out_1) = \nu_{Out_2}(Out_2)$. In accordance with the rest of the paper, all agents are assumed to be matched. Recall also that the bonus scheme is fixed, so this is effectively a setup of non-transferable utility.

Secondly, matchings (ϕ_O, ϕ_1, ϕ_2) are introduced, where the first refers to outgroup matches, while the other two refer to ingroup matches for each group.

(i) ϕ_O is a Borel measure on $Out_1 \times Out_2$ with marginals $\phi_O(I \times Out_2) = \nu_{Out_1}(I)$ for all measurable $I \subseteq Out_1$, and $\phi_O(Out_1 \times I) = \nu_{Out_2}(I)$ for all measurable $I \subseteq Out_2$.

(ii) For $k = 1, 2$, ϕ_k is a Borel measure on $In_k \times In_k$ with marginals $\phi_k(I \times In_k) = \nu_{In_k}(I)$ for all measurable $I \subseteq In_k$, and $\phi_k(In_k \times I) = \nu_{In_k}(I)$ for all measurable $I \subseteq In_k$.

The following continuity assumption is imposed on each of the matchings ϕ_k, for $k = O, 1, 2$: let λ be a Borel measure on $A \times B$; if $A' \subseteq A$ (resp. $B' \subseteq B$) is an open interval with positive measure, then there exists some open $B' \subseteq B$ (resp. $A' \subseteq A$) such that $\lambda(A' \times B') > 0$.[16]

For a given matching structure $\phi = (In_1, Out_1, In_2, Out_2, \phi_O, \phi_1, \phi_2)$, the equilibrium efforts are summarized by the following continuous functions: for $(c, c') \in In_k \times In_k$,

$$e(c, c') = \frac{c' + t}{cc' - t^2};$$

and for $(c, c') \in Out_1 \times Out_2$, we have

$$e(c, c_2^*) = \frac{1 + t\frac{1}{c_2^* - t}}{c}, \quad \text{and} \quad e(c', c_1^*) = \frac{1 + t\frac{1}{c_1^* - t}}{c'} \tag{6}$$

where c_k^* is the value in $[\underline{c}, \bar{c}]$ (which exists and is unique) that solves $\frac{1}{c-t} = \eta_k(\phi)$, where $\eta_k(\phi)$, the average in group $k = 1, 2$, is determined by

$$\eta_k(\phi) = \frac{\nu_{G_k}\left(c_{In_{k'}} + c_{Out_{k'}}\right) + tc_{Out_{k'}}\left(c_{In_k} + c_{Out_k}\right)}{\nu_{G_k}\nu_{G_{k'}} - t^2 c_{Out_k} c_{Out_{k'}}} \tag{7}$$

where

$$c_{In_k} \equiv \frac{1}{2}\int_{In_k \times In_k} \frac{c' + t}{cc' - t^2}\, d\phi_k(c, c'), \quad c_{Out_k} \equiv \int_{Out_k} \frac{1}{c}\, d\nu_{Out_k}(c).$$

---

[16] This assumption is called continuity based on the concept of continuity for functions, rather than continuity of measures.

The designer's maximization problem is then defined as

$$\max_{\phi} \int_{Out_1 \times Out_2} p(e(c, c_2^*), e(c', c_1^*))\, d\phi_O(c, c') + \frac{1}{2} \sum_{k=1}^{2} \int_{In_k \times In_k} p(e(c, c'), e(c', c))\, d\phi_k(c, c')$$

$$\text{st. } \phi = (In_1, Out_1, In_2, Out_2, \phi_O, \phi_1, \phi_2) \text{ is a matching structure} \tag{8}$$

To verify that, in the setting with a continuum of agents, the optimal matching ϕ retains the assortativity properties established in Proposition 5, the following definitions are introduced.

Definition 4 *Given the matching structure ϕ, let In_k have positive measure $\nu_{In_k} > 0$ for some k. The matching ϕ_{In_k} is said to be* one-sided assortative *in In_k if, for any open set $I \subseteq In_k$ with $\nu_{In_k}(I) > 0$, it holds that $\phi_{In_k}(I \times I) = \nu_{In_k}(I)$.*

This definition requires that agents are matched among themselves within any open subset of the ingroup with positive measure.

The matching structure ϕ cannot be optimal unless it is one-sided assortative in both In_1 and In_2, as a local improvement would otherwise be possible.

Lemma 1 *Suppose that, for some ϕ and some $k \in \{1, 2\}$, the matching ϕ_k is not one-sided assortative in In_k, where $\nu_{In_k} > 0$. Then ϕ cannot be optimal.*

The argument is provided in the Appendix. The necessity of one-sided assortativeness arises from strategic complementarities. When ϕ_k is one-sided assortative in In_k, for $k = 1, 2$, cost-type c agents are matched with others of the same cost-type c, and the equilibrium effort within the ingroup is given by the function $e(c, c) = \frac{1}{c-t}$.

To define outgroup assortativity, it is useful to introduce a form of quantile distribution over characteristics for the agents matched outgroup in each group k, denoted Q_k^{ϕ}. Let $Q_k^{\phi}(c)$ be the measure of the set $Out_k \cap [c, \bar{c}]$. Note that $Q_k^{\phi}(c)$ increases (weakly) as c decreases, and that $Q_k^{\phi}(\underline{c}) = \nu(Out_k)$.

Definition 5 *Given ϕ, suppose Out_1 has positive measure. The matching ϕ_O is said to be* two-sided assortative *in $Out_1 \times Out_2$ if, for any interval $(c - \varepsilon, c + \varepsilon)$ with $\nu_{Out_1}((c - \varepsilon, c + \varepsilon)) > 0$, there exists an interval $(c' - \varepsilon', c' + \varepsilon')$ with $\nu_{Out_2}((c' - \varepsilon', c' + \varepsilon')) > 0$ such that $\phi_O((c - \varepsilon, c + \varepsilon) \times (c' - \varepsilon', c' + \varepsilon')) > 0$ and $Q_1^{\phi}(c) = Q_2^{\phi}(c')$.*

The next lemma establishes that a matching structure ϕ cannot be optimal if it is not two-sided assortative in $Out_1 \times Out_2$.

Lemma 2 *Suppose that for some ϕ, ϕ_O is not two-sided assortative in $Out_1 \times Out_2$, where $\nu_{Out_1} > 0$. Then ϕ cannot be optimal.*

The argument (in Appendix) relies solely on the supermodularity of the objective function of the designer, while strategic complementarity plays no direct role because types that are matched outgroup best reply to the average effort in the outgroup. Thus, the type of the match affects the effort exerted by type c only insofar as it alters the average in the outgroup. However, for fixed sets Out_1 and Out_2, the specific distribution induced by ϕ_O over $Out_1 \times Out_2$ does not affect the average effort (as evident from equation (7)).

Lemmas 1 and 2 ensure that the optimal matching structure must be both one-sided and two-sided assortative.

A matching structure $\phi = (In_1, Out_1, In_2, Out_2, \phi_O, \phi_1, \phi_2)$ is said to be *assortative* if ϕ_O is two-sided assortative in $Out_1 \times Out_2$ and ϕ_k is one-sided assortative in In_k, for $k = 1, 2$.

For what follows, symmetric groups are considered, assuming that the sets G_1 and G_2 are identically distributed ($\nu_{G_1} = \nu_{G_2}$). Under symmetric groups, the optimal matching can be identified by solving a threshold problem: cost-types below the threshold are matched ingroup, while those above are matched outgroup.

The next proposition identifies two thresholds on cost-types and establishes that an optimal assignment must match almost all types with cost below the lower threshold ingroup, and with cost above the upper threshold outgroup.

Proposition 7 *Let groups be symmetric. Consider some assortative matching structure ϕ and let $\eta_1(\phi) \geq \eta_2(\phi)$, and denote by $c_k^* = \frac{1+t\eta_k(\phi)}{\eta_k(\phi)}$ (so that $c_1^* \leq c_2^*$). Consider the following conditions:*

1. *For each $k = 1, 2$, there is a measure-zero set of types matched outgroup with costs below c_1^*, or $\nu_{Out_k}((\underline{c}, c_1^*)) = 0$.*

2. *For each $k = 1, 2$, there is a measure-zero set of types matched ingroup with costs above c_2^*, or $\nu_{In_k}((c_2^*, \bar{c})) = 0$.*

If condition 1 or 2 is not satisfied, then ϕ cannot be an optimal matching structure.

The argument is detailed in the Appendix. Proposition 7 can be interpreted as the continuum counterpart of the necessary conditions established in Proposition 6 for a finite setting.

The next proposition shows that optimality requires the thresholds c_1^* and c_2^* to coincide. That is, the optimal assignment must satisfy a unique-threshold property.

Proposition 8 *Let groups be symmetric. Consider a matching structure ϕ with $\eta_1(\phi) \geq \eta_2(\phi)$. If $\eta_1(\phi) \neq \eta_2(\phi)$, then ϕ cannot be optimal.*

Proof. Assume by contradiction that ϕ is optimal and $\eta_1(\phi) > \eta_2(\phi)$, so that $c_1^* < c_2^*$.

By Lemmas 1 and 2, ϕ is assortative. By Proposition 7, almost all types with $c < c_1^*$ are matched ingroup, while almost all types with $c > c_2^*$ are matched outgroup in both groups.

The equilibrium effort for ingroup matches of cost-type c in group k is given by $e_{In_k}(c,c) = \frac{1}{c-t}$, while for outgroup matches it is given by

$$e_{Out_k}(c, c_{k'}^*) = \frac{1 + t\frac{1}{c_{k'}^* - t}}{c} = \frac{c_{k'}^*}{c} \frac{1}{c_{k'}^* - t}.$$

Then for any $c \in (c_1^*, c_2^*)$, we observe:

(i) If $c \in In_1$ and $c \in In_2$, then $e_{In_1}(c,c) = e_{In_2}(c,c) = \frac{1}{c-t}$.

(ii) If $c \in In_1$ and $c \in Out_2$, $c < c_1^* \implies e_{In_1}(c,c) = \frac{1}{c-t} < \frac{c_1^*}{c}\frac{1}{c_1^*-t} = e_{Out_2}(c, c_1^*)$.

(iii) If $c \in Out_1$ and $c \in In_2$, $c > c_2^* \implies e_{Out_1}(c, c_2^*) = \frac{c_2^*}{c}\frac{1}{c_2^*-t} < \frac{1}{c-t} = e_{In_2}(c,c)$.

(iv) If $c \in Out_1$ and $c \in In_2$, then $c_1^* > c_2^* \implies e_{Out_1}(c, c_2^*) < e_{Out_2}(c, c_1^*)$.

Properties (i)–(iv) together imply that the effort for any $c \in (c_1^*, c_2^*)$ is weakly smaller in group 1 than in group 2. Property (iv) extends this to types with $c > c_2^*$, so efforts are strictly higher in group 2 on $(c_1^*, \bar{c}]$. Property (i) also holds for $c < c_1^*$, and thus on the full region $[\underline{c}, c_2^*)$.

Therefore, for all c, effort is weakly higher in group 2 than in group 1—and strictly so on a set of positive measure. This contradicts the assumption that $\eta_1(\phi) > \eta_2(\phi)$. ■

Propositions 7 and 8 confirm the threshold property for the optimal matching among symmetric groups: average effort is the same in both groups, $\eta_1(\phi) = \eta_2(\phi) = \eta(\phi)$, and there exists a threshold c^* such that

$$\eta(\phi) = \frac{1}{c^* - t},$$

with almost all types with cost below (respectively, above) c^* matched ingroup (respectively, outgroup).

By imposing the additional condition that the total mass in each group is 1, the distribution within each group is described by a probability density function f, with cumulative distribution function F. When f is absolutely continuous over $[\underline{c}, \bar{c}]$, the designer's maximization problem reduces to choosing c^* that solves:

$$\max_{\hat{c}} \int_{\underline{c}}^{\hat{c}} \frac{f(c)}{c - t}\, dc + \int_{\hat{c}}^{\bar{c}} \frac{\hat{c}}{c} \frac{f(c)}{\hat{c} - t}\, dc.$$

This formulation implies that types matched ingroup exert effort strictly above the average, while those matched outgroup exert effort strictly below. Effort for types at the threshold c^* coincides with the average, i.e., $e_{In}(c^*, c^*) = e_{Out}(c^*, c^*) = \eta(\phi)$.

This property also rules out threshold values at the extremes of the support $[\underline{c}, \bar{c}]$. If all types are matched ingroup ($c^* = \bar{c}$), the average effort becomes $\mathbb{E}\left[\frac{1}{c-t}\right]$, which is strictly greater than $\frac{1}{\bar{c}-t}$. Therefore, the threshold must lie strictly below $\bar{c}$, and the designer can strictly improve upon the equilibrium outcome with full ingroup matching.

Similarly, if all types are matched outgroup ($c^* = \underline{c}$), the average effort becomes $\frac{\mathbb{E}\left[\frac{1}{c}\right]}{1-t\mathbb{E}\left[\frac{1}{c}\right]}$, which is strictly smaller than $\frac{1}{\underline{c}-t}$ (since $\underline{c} > 1$). Hence, the threshold must also be strictly above $\underline{c}$. Note that under this private-value framework, the equilibrium induced by full outgroup matching coincides with the outcome of a Bayesian game where types are uniformly randomly matched. Thus, $c^* > \underline{c}$ guarantees that the designer can strictly improve upon the corresponding Bayes-Nash equilibrium.

Finally, the symmetry condition established in Proposition 8, together with Condition (1) in Proposition 7, suggests a natural iterative procedure to identify the optimal threshold. This procedure assumes throughout that ϕ is symmetric and always both one-sided and two-sided assortative, in line with Lemmas 1 and 2.

The next proposition introduces this procedure and shows that: (i) the sequence must converge, and (ii) at any steady state, the resulting efforts, mean efforts, and threshold coincide with those characterized in the framework above.

Proposition 9 *Consider the following procedure:*

- *Iteration 0. All agents are matched ingroup (threshold $\tau_0 = \bar{c}$), and the matching ϕ_0 is one-sided assortative in In_G^0 for both G. The average effort, namely $\eta(\tau_0)$, is equal to $\mathbb{E}[\frac{1}{c-t}]$.*
- *Iteration k. Given τ_{k-1}, ϕ_{k-1}, and $\eta(\tau_{k-1})$, the efforts exerted by the agents are summarized by the function:*

$$e_{\phi_{k-1}}(c; \tau_{k-1}, \eta(\tau_{k-1})) = \begin{cases} \frac{1}{c-t} & \text{if } c \leq \tau_{k-1}, \\ \frac{1+t\eta(\tau_{k-1})}{c} & \text{if } c > \tau_{k-1}. \end{cases}$$

  - *Step 1. Find τ_k such that $\frac{1}{\tau_k - t} = \eta(\tau_{k-1})$.*
  - *Step 2. Consider the (assortative) matching ϕ_k that matches outgroup all agents with cost $c > \tau_k$, thereby obtaining, for fixed average $\eta(\tau_{k-1})$, the effort function:*

$$e_{\phi_k}(c; \tau_k, \eta(\tau_{k-1})) = \begin{cases} \frac{1}{c-t} & \text{if } c \leq \tau_k, \\ \frac{1+t\eta(\tau_{k-1})}{c} & \text{if } c > \tau_k. \end{cases}$$

  - *Step 3. Compute the average $\eta(\tau_k) = \int_{\underline{c}}^{\bar{c}} e_{\phi_k}(c; \tau_k, \eta(\tau_{k-1})) f_G(c)\, dc$ and the effort function $e_{\phi_k}(c; \tau_k, \eta(\tau_k))$.*

Consider the steady state $\tau_k = \tau_{k-1} = \tau$, $\phi_k = \phi_{k-1} = \phi$, and $\eta(\tau_k) = \eta(\tau_{k-1}) = \eta(\tau)$ for some k. Then, for each G, $\eta_G(\phi) = \eta(\tau)$, for each type c, the effort exerted is given by $e_\phi(c; \tau, \eta(\tau))$, and $c^* = \tau$.

This procedure must converge.

The procedure allows us to find the optimal matching through the iteration of a few simple steps. In particular, it simplifies the problem because at each iteration, first a threshold is identified for fixed efforts and fixed mean efforts. Then, the efforts are recomputed for a fixed threshold and fixed mean efforts. Finally, the mean efforts are computed for given efforts and threshold. Throughout the procedure, the effort function does not necessarily correspond to actual equilibrium efforts. However, this must be the case in the steady state.

The convergence is established thanks to the fact that the procedure above constitutes a Lyapunov system (strategic complementarity is necessary for this): the threshold is weakly decreasing at each iteration, and it is bounded below by $\underline{c}$, or alternatively, the mean effort is weakly increasing at each iteration, and it is bounded above by $\frac{1}{\underline{c}-t}$ (see the Appendix for details). Furthermore, as observed before, the threshold c^* must necessarily be below $\underline{c}$, so the procedure must stop before reaching the lower bound for the threshold.

As mentioned before, this procedure can be seen as a guideline the designer might follow to improve over time his objective function and get closer to the optimal matching.

5 Further Modeling Discussions

In this section I compare the results obtained through the current modeling approach with alternative approaches.

5.1 Endogenous Matching

Throughout the paper, the matching function ϕ is treated as exogenous from players' perspective. Even in the application to organizations, the designer chooses the matching function, but players treat it as given. It seems natural to analyze the predictions that would be obtained were players allowed to match in a decentralized fashion.

A general definition of pairwise stability in a setting with a continuum of agents and matches that can be both ingroup and outgroup is beyond the scope of the paper. In line with Greinecker and Kah (2021), I propose a notion of stability, where a matching structure ϕ is said to be stable if, given ϕ and efforts profiles e^{ϕ}, there is a positive measure of types c that would like to rematch among themselves. Similarly to the main set-up, I do not consider individual rationality constraints, implicitly assuming that all types prefer to be matched rather than not be matched.

Let the set of players have measure $\nu > 0$ and let groups G_1 and G_2, with measures $\nu_{G_1} = v(G_1)$ and $\nu_{G_2} = v(G_2)$, be a partition of the set of players, and the matching function ϕ be defined in the continuum as in subsection 4.3. Given $\phi = (In_1, Out_1, In_2, Out_2, \phi_O, \phi_1, \phi_2)$, I denote by e^{ϕ} the efforts exerted by any type c matched with a type c' according to ϕ_k, $k = O, 1, 2$, and I denote by c_{ϕ_k} the type c' matched with c in ϕ_k. If type c is matched ingroup in G_k, $e^{\phi_k}(c, c_{\phi_k}) = \frac{c_{\phi_k}+t}{cc_{\phi_k}-t^2}$, $k = 1, 2$, if types are matched outgroup and type c is in G_k, $e^{\phi_O^k}(c, c_{\phi_O^k}) = \frac{1+t\eta_{k'}(\phi)}{c}$, for $k \neq k' = 1, 2$. Then the effort profile e^{ϕ} is a list $(e^{\phi_O^1}, e^{\phi_O^2}, e^{\phi_1}, e^{\phi_2})$, where $e^{\hat{\phi}} = (e^{\hat{\phi}}(c, c_{\hat{\phi}}))_{c \in [\underline{c}, \bar{c}]}$, for $\hat{\phi} = \phi_O^1, \phi_O^2, \phi_1, \phi_2$.

In order for a type c to consider to form a re-match (analog to a blocking pair), I require agents to have access to the following data: given a matching structure $\phi = (In_1, Out_1, In_2, Out_2, \phi_O, \phi_1, \phi_2)$ and a profile of efforts $e^{\phi} = (e^{\phi_O^1}, e^{\phi_O^2}, e^{\phi_1}, e^{\phi_2})$, a type-$c$ agent in group G_k has access to the effort exerted by the types in her own group, both the efforts in the ingroup matches, e^{ϕ_k}, and the efforts exerted in the outgroup matches, $e^{\phi_O^k}$. However, in parallel with the rest of the paper, I consider that they have access only to the average effort in the outgroup, namely $\eta_{k'}(\phi)$.

Definition 6 *Given ϕ and e^{ϕ}, I say that a currently matched ingroup type c in G_k is willing to engage in a rematch with*

(i) type $c' \in G_k$ if $u(e^{\phi_k}(c, c_{\phi_k}), e^{\hat{\phi}'}(c', c'_{\hat{\phi}'}); c) > u(e^{\phi_k}(c, c_{\phi_k}), e^{\phi_k}(c_{\phi_k}, c); c)$, where $\hat{\phi}' = \phi_k$ (resp. $\hat{\phi}' = \phi_O^k$) for type c' matched ingroup (resp. outgroup) in ϕ;

(ii) type $c' \in G_{k'}$ if $u(e^{\phi_k}(c, c_{\phi_k}), \eta_{k'}(\phi); c) > u(e^{\phi_k}(c, c_{\phi_k}), e^{\phi_k}(c_{\phi_k}, c); c)$.

Note that players, when considering a re-match, treat as fixed the data on the efforts of the ingroup, including their own effort.[17] This implicitly assumes that players do not engage in strategic reasoning (similarly to the rest of the paper).

In a similar fashion,

Definition 7 *given ϕ and e^{ϕ}, I say that a currently matched outgroup type c in G_k is willing to engage in a rematch with*

(i) type $c' \in G_k$ if $u(e^{\phi_O^k}(c, c_{\phi_O^k}), e^{\hat{\phi}'}(c', c'_{\hat{\phi}'}); c) > u(e^{\phi_k}(c, c_{\phi_k}), \eta_{k'}(\phi); c)$, where $\hat{\phi}' = \phi_k$ (resp. $\hat{\phi}' = \phi_O^k$) for type c' matched ingroup (resp. outgroup) in ϕ;

(ii) type $c' \in G_{k'}$ if $u(e^{\phi_O^k}(c, c_{\phi_O^k}), \eta_{k'}(\phi); c) > u(e^{\phi_k}(c, c_{\phi_k}), \eta_{k'}(\phi); c)$.

---

[17] In the same vein, Goursat (2024) (Chapter 2) proposes a heuristical approach to matching in marriage markets, where agents evaluate the possibility of forming a blocking pair based on the current realized match utilities.

Note that condition (ii) will never occur by construction as an type that is currently matched outgroup is indifferent about the type of her match, as she evaluates the effort through the average in the outgroup.

I identify the instability set $I \subseteq G_1 \cup G_2$, with measure $\nu(I)$, as the set containing all types such that for each type $c \in I$, currently matched with a type c_ϕ according to ϕ, there is a type $c' \in I$ such that c is willing to engage in a rematch with c' and c' is willing to engage in a rematch with c.

I say that a matching structure ϕ is stable if the instability set I has measure zero ($\nu(I) = 0$).

Remark 3 *A stable matching ϕ is necessarily one-sided assortative. However, it does not need to two-sided assortative.*

The first claim of the remark can be shown with an argument along the lines of lemma 1. As a matter of fact, the strategic complementarity drives the necessity of one-sided assortativeness in lemma 1. Strategic complementarity derives from the fact that the utility is increasing in the effort of the opponent (as is the case), which in enough to obtain the desired property.

The second claim of the remark is obtained by observing that there can never be a currently-matched-outgroup type who is willing to engage in a rematch with another outgroup type (as condition (ii) in definition 7 can never be satisfied). This observation implies that the stability requirements do not impose any structure on the distribution over the outgroup matches. That is, given Out_1 and Out_2, any $\phi_O(Out_1 \times Out_2)$ does not contribute to the instability set.

For symmetric groups, it is established that the matching structure ϕ that is optimal for the designer subsection 4.3 of the application to organizations is also stable. However, the two-sided assortativeness is not needed here. That is,

Remark 4 *Let groups be symmetric. Consider a matching structure ϕ and a threshold c^* such that: (i) in each group, every type $c \leq c^*$ is matched ingroup and assortatively; and (ii) every type $c > c^*$ is matched outgroup.*

If c^* solves $\eta_k(\phi) = \frac{1}{c^* - t}$ for $k = 1, 2$, then ϕ is a stable matching structure.

This can be seen with the following observations: consider matching ϕ as defined in the remark and let e^ϕ be the list of effort profiles. For any $c \leq c^*$, the effort

$e^{\phi}(c, c) = \frac{1}{c-t}$ is greater that the effort in the whole list e^{ϕ} of any other type c' such that $c > c'$ and also weakly greater than the average $\eta(\phi) = \frac{1}{c^*-t}$. Then, any type $c \leq c^*$ is matched ingroup and is willing to engage in a rematch only with types-costs c' in the same group, where c' is strictly greater than c . Clearly the type-cost c' will not reciprocate the willingness to engage in a rematch with type c, as c is smaller than c'. So the set of types $c \leq c^*$ that are willing to engage in any rematch is empty.

As far as the set of types $c > c^*$ is concerned, no type can engage in a rematch with any type above c^*, as the latter will not be willing to engage in a rematch with c. So any potential rematch must occur between types $c > c^*$, which are all currently matched outgroup. They will not engage in a rematch with some type from the same group and $c' < c^*$, because type c' is exerting effort below the average $\eta(\phi)$. They will not engage in a rematch with some type $c' < c^*$ from the outgroup either as condition (ii) in definition 7 is never satisfied. Hence, the instability set is empty and the matching structure ϕ is stable.

The discussion of decentralized matching shows that, in the context of the organization, the optimal solution of the designer's problem can be interpreted as arising from an endogenous matching where types are allowed to choose how to match. However, if the designer were to allow agents to choose how to match, he would be able to achieve overexertion of effort from the outgroup matches, but he would not be able to guarantee assortativeness within the outgroup matches.

5.2 Ingroup Homogeneity Effect

Along with the outgroup homogeneity effect, the psychological literature also finds evidence of ingroup homogeneity. Brewer (1993) observes that, while there is ample evidence of the tendency to perceive outgroups as homogeneous, other studies find that members of a group also attribute greater similarity of attitudes within the ingroup (as compared to the outgroup). In their meta-analysis, Mullen and Hu (1989) find that the perceived homogeneity of outgroups is almost constant across studies, whereas the ingroup is perceived as relatively more or less homogeneous, possibly depending on the social context.

From this perspective, the approach used in this paper is relevant in some contexts, while in others it might be appropriate to allow for ingroup homogeneity as well. Note that the framework is flexible enough to accommodate other specifications of ingroup and outgroup homogeneity.

To see this, recall that the current notion of equilibrium can be viewed in terms of the analogy-based expectation equilibrium (ABEE), by positing that players categorize all outgroup members together in one analogy class, while each ingroup member is categorized separately. However, this is only one specific way of categorizing the set of players. Through this lens, the approach in this paper can be extended by allowing members in each group to devote K_{In} categories to the ingroup and K_{Out} to the outgroup, thereby capturing the relative degree of perceived homogeneity in the attitudes of ingroups and outgroups through the cardinality of the respective categorizations.[18]

Building on this observation, one interesting avenue for future research is to allow for categories to be endogenously determined, using approaches similar to those introduced in Jehiel and Weber (2025) or Goursat et al. (2025), to shed light on the various degrees of ingroup and outgroup homogeneity through the lens of expectations in strategic environments. The focus on expectations can serve as complement to the literature spanning from Akerlof and Kranton (2000) on identity, which typically analyzes intergroup phenomena through the primitive of the utility function.

5.3 Groups as Networks: Peer-Confirming Equilibrium

Peer-Confirming Equilibrium (PCE) is an equilibrium notion proposed by Lipnowski and Sadler (2019) to capture how players' social networks can affect their expectations about others' behavior. Specifically, they augment a game with a network and posit that players form correct expectations about those to whom they are linked, while forming rationalizable conjectures about the behavior of players to whom they are not linked. The set of viable conjectures is restricted by the (correct) expectations over their links.

In this paper, the notion of groups captures the ability of players to form expec-

---

[18] If player-specific cardinalities for the categorization were also allowed, as is typically done in the ABEE literature, the approach could also relate to groupy and non-groupy behaviors as recently discussed in Kranton and Sanders (2017), where a non-groupy individual would be identified as a player with categorizations involving a large number of classes (e.g., as many categories as the there are potential opponents).

tations about opponents' behaviors. From this perspective, groups play a role that is similar in spirit to the network in Lipnowski and Sadler (2019). In particular, in both papers the additional primitive aims to capture the effect of sociality (social groups or social networks) on the perception of others' behavior.

The current framework can also be extended by augmenting the game with a network, allowing players to form finely tuned expectations when interacting with agents to whom they are linked, and coarse expectations over the behavior of players to whom they are not linked. Social groups are then simply subsets of players who are all linked to each other (and only to each other) in the network.

From this point of view, the notion of equilibrium introduced in this paper can be seen as a refinement of the set of peer-confirming equilibria. More broadly, the connection between this paper and peer-confirming equilibrium is analogous to the connection between analogy-based expectations equilibrium (ABEE) and conjectural or self-confirming equilibrium (see Jehiel (2022) for a discussion on the connection between ABEE and SCE; see Lipnowski and Sadler (2019) for a discussion on the connection between PCE and conjectural equilibrium).

6 Conclusion

In this paper, I have introduced a novel framework for analyzing intergroup phenomena by distinguishing between finely tuned expectations in ingroup interactions and coarse, aggregate expectations in outgroup interactions. The model formalizes the outgroup homogeneity effect and demonstrates how these differing expectations can lead to distinct equilibrium outcomes.

The analysis of coordination games illustrates that, under the influence of coarse expectations, a unique equilibrium can be achieved even in settings where multiple equilibria might exist under standard assumptions. In the organizational application, I illustrate how a designer can exploit these behavioral nuances to obtain overexertion of effort by some agents, when he can only choose how to match agents as a way to incentivize effort.

Appendix

Proof of Proposition 1

It is shown by induction that in the global games illustration if $\phi(b_i) = d_i$ for $i = 1, \dots, n$, then there is a unique equilibrium where all players play I.
Recall that player $i \in \mathcal{I}$, matched with $j \in \mathcal{I}$, prefers playing I to NI whenever i expects I to be played with probability greater than $1 - \frac{x_i + x_j}{2}$.

Step 0. Consider the pair (b_n, d_n). Let $\beta_i(\sigma)[I]$ denote the expectations of player i over the probability of I being played according to σ. By definition, $\beta_i(\sigma)[I] \geq 0$. Since $1 - x_n = -\varepsilon < 0$, it is strictly dominant for b_n and d_n to play I, regardless of β. Because every player is matched outgroup according to ϕ, then for each $i \in \mathcal{I}$, $\beta_i(\sigma)[I] \geq \frac{1}{n}$ by consistency.

Step k. Assume by the induction hypothesis that all pairs $(b_n, d_n), \dots, (b_{n-(k-1)}, d_{n-(k-1)})$ strictly prefer playing I rather than NI, so that for each i, $\beta_i(\sigma)[I] \geq \frac{k}{n}$. Then, the players in the match (b_{n-k}, d_{n-k}) strictly prefer to play I too: this is because, for each i,

$$1 - x_{n-k} = \frac{k}{n} - \frac{n-k}{n}\varepsilon < \frac{k}{n} \leq \beta_i(\sigma)[I]$$

Since all players are matched outgroup, this also implies that, for each $i \in \mathcal{I}$, $\beta_i(\sigma)[I] \geq \frac{k+1}{n}$ by consistency.

The induction argument can be simply seen in terms of expectations: we have shown that $\beta_i(\sigma)[I] \geq \frac{1}{n}$ for all i, and $\beta_i(\sigma)[I] \geq \frac{k}{n}, \ \forall i \implies \beta_i(\sigma)[I] \geq \frac{k+1}{n}, \ \forall i$ where $k = 1, \dots, n-1$, thereby implying $\beta_i(\sigma)[I] = 1$. By consistency, $\sigma_j = \{I\}$, for all $j \in \mathcal{I}$. ■

Proof of Proposition 2

The argument follows the same logic by induction as proposition 1, only moving downwards along the intervals generated by the various thresholds $(\hat{x}_k)_{k=1}^{N+1}$. All players are matched outgroup in an assortative fashion. Once established that the players with x above 1 play I, then all players with x between $[\frac{N-1}{N}(1+\varepsilon), 1)$ must play I (provided that are not more players within the latter interval as compared to the former).

Let $z(k)$ denote the number of players in group $z \in \{b', d'\}$ with characteristics in $[\hat{x}_k, \hat{x}_{k+1})$ for $k = 1, \dots, N$. Assume that the number of players within each interval is monotone in the sense that, for each z and k, $z(k+1) \geq z(k) \geq 0$. Observe that this property implies that in both groups, the fraction of players within the last interval, namely $\frac{b'(N)}{n}$ and $\frac{d'(N)}{n}$, is weakly grater than $\frac{1}{N}$, and more generally

$$\sum_{k'=k}^{N} \frac{z_{k'}}{n} \geq \frac{N+1-k}{N}, \text{ for } k' = 1, \dots, N, \text{ and } z = b', d'. \tag{9}$$

Let $m_k \equiv \min\{\sum_{k'=k}^{N} b'(k'), \sum_{k'=k}^{N} d'(k)\}$ (note that $m_1 = n$).

Because ϕ matches players outgroup and assortatively, for each k, there are m_k players in group $G_{b'}$ with characteristics weakly above $\hat{x}_k$ that are matched with some player in $G_{d'}$ with characteristics weakly above $\hat{x}_k$. Thereby implying that, for $k = N$, there are at least m_N players in each group playing I because for all such players $\frac{x_i + x_{\phi(i)}}{2} \geq 1$ and the inequality is strict for at least one pair, given that $x_i > 1$ for at least one player in $\mathcal{I}'$ (which induces all other $m_N - 1$ pairs to play I). Since ϕ matches all players outgroup, for each player $i \in \mathcal{I}'$, $\beta_i(\sigma)[I] \geq \frac{m_N}{n} \geq \frac{1}{N}$.

Say that at some step $k+1$, $\beta_i(\sigma)[I] \geq \frac{m_{k+1}}{n}$ by induction hypothesis.

Then it is easily verified that $\beta_i(\sigma)[I] \geq \frac{m_k}{n}$: note that $\frac{m_{k+1}}{n} \geq \frac{N-k}{N}$ by (9).

Consider the last m_k players in each group. For each such player i, $\frac{x_i + x_{\phi(i)}}{2} \geq \hat{x}_k = \frac{k}{N}(1+\varepsilon)$. Then, $1 - \frac{x_i + x_{\phi(i)}}{2} \leq 1 - \hat{x}_k = \frac{N-k}{N} - \varepsilon\frac{k}{N}$. Then, each such player i plays I because $\beta_i(\sigma)[I] > \frac{N-k}{N} - \varepsilon\frac{k}{N}$. Therefore, m_k players are playing I in each group, thereby obtaining, by consistency, $\beta_i(\sigma)[I] \geq \frac{m_k}{n}$. ■

Proof of Proposition 3

The following claim is established: for any $\mu \in \Delta(\Phi)$, there is some σ that is a BNE, where $\sigma_i = \{NI\}$ for some $i \in \mathcal{I}$.

The result is shown by finding the conditions on $\mu \in \Delta(\Phi)$ such that a strategy profile, where some player is choosing NI, is a Bayes-Nash equilibrium (BNE). The proof involves two steps. First, we find conditions on μ_{b_n,d_n} that suffice to ensure that $\sigma_i = \{NI\}$ for all i constitutes a BNE. After that, we find conditions on μ_{z_{n-1},z'_n} for $z, z' \in \{b, d\}$ that are sufficient for σ to be a BNE, where $\sigma_{b_n} = \sigma_{d_n} = \{I\}$ and $\sigma_i = \{NI\}$ for all $i \neq b_n, d_n$. Finally, the result is obtained by verifying that if the conditions in the first step are not met, then the conditions in the second step hold necessarily.

STEP 1: let $\mu \in \Delta(\Phi)$, $\mu_n \equiv \mu_{b_n,d_n}$, and let σ be such that, for each i, $\sigma_i = \{NI\}$. Given μ and σ, NI is a best response for player i iff $0 \leq 1 - \frac{x_i + \sum_{j \in \mathcal{I}} \mu_{i,j} x_j}{2}$. Note that, for each $i \in \mathcal{I}$, $x_i + \sum_{j\in\mathcal{I}} \mu_{i,j} x_j \leq x_n + \mu_n x_n + (1-\mu_n)x_{n-1}$. Thus, to ensure that NI is best response for all players it suffices to impose that

$$\begin{aligned} 0 \leq 1 - \frac{x_n + \mu_n x_n + (1-\mu_n)x_{n-1}}{2} &= \frac{2 - (1+\mu_n)(1+\varepsilon) - (1-\mu_n)\frac{n-1}{n}(1+\varepsilon)}{2} \\ &= \frac{1 - \mu_n - \varepsilon(2n - 1 + \mu_n)}{2n} \end{aligned}$$

Moving $\mu_n = \mu_{b_n,d_n}$ on the left-hand side yields:

$$\mu_{b_n,d_n} \leq \frac{1-(2n-1)\varepsilon}{1+\varepsilon}$$

where the term on the right-hand side tends to 1 as $\varepsilon \to 0$, and it tends to $1 - \frac{1}{2n}$ as $\varepsilon \to \frac{1}{4n^2-1}$. Therefore, for all μ such that $\mu_{b_n,d_n} \leq 1 - \frac{1}{2n}$, the strategy profile where $\sigma_i = \{NI\}$ for all i is a BNE.

Step 2: let $\mu \in \Delta(\Phi)$, $\mu_k \equiv \mu_{b_k,d_k}$, and $\mu_{i,x_k} \equiv \mu_{i,b_k} + \mu_{i,d_k}$ for $i \in \mathcal{I}$. Based on the result in step 1, assume that $\mu_n > 1 - \frac{1}{2n}$. Consider the strategy profile σ where $\sigma_{b_n} = \sigma_{d_n} = \{I\}$ and $\sigma_i = \{NI\}$ for all $i \neq b_n, d_n$. Given μ and σ, NI is a best response for player i iff

$$\mu_{i,x_n} = \sum_{j\in\mathcal{I}} \mu_{i,j}\sigma_j(I) \leq 1 - \frac{x_i + \sum_{j\in\mathcal{I}} \mu_{i,j}x_j}{2}$$

For $i \in \{b_n, d_n\}$, it is readily verified that I is a best response, given σ and μ. Whereas, for each $k < n$ and $i_k \in \{b_k, d_k\}$, we have that

$$x_{i_k} + \sum_{j\in\mathcal{I}} \mu_{i_k,j}x_j \leq x_{n-1} + \mu_k x_{n-1} + \mu_{i_k,x_n}x_n + (1 - \mu_k - \mu_{i_k,x_n})x_{n-2}.$$

Thus, to ensure that NI is best-response for all players other than b_n and d_n it is sufficient to require that

$$\begin{aligned} &\mu_{i_k,x_n} \leq 1 - \frac{x_{n-1} + \mu_k x_{n-1} + \mu_{i_k,x_n}x_n + (1-\mu_k-\mu_{i_k,x_n})x_{n-2}}{2} \iff \\ &\mu_{i_k,x_n}\left(1 + \frac{x_n - x_{n-2}}{2}\right) \leq 1 - \frac{x_{n-1}+x_{n-2}}{2} - \mu_k\frac{x_{n-1}-x_{n-2}}{2} \iff \\ &\mu_{i_k,x_n}\left(\frac{2(n+1+\varepsilon)}{2n}\right) \leq \frac{(3-\mu_k)(1+\varepsilon) - 2n\varepsilon}{2n} \iff \mu_{i_k,x_n} \leq \frac{(3-\mu_k)(1+\varepsilon)-2n\varepsilon}{2(n+1+\varepsilon)} \end{aligned}$$

where the third inequality is found by substituting x_k with $(1+\varepsilon)\frac{k}{n}$ and then simplifying.

Note that the right-hand side in the last inequality is minimized for $\mu_k \to 1$ [19] and $\varepsilon \to \frac{1}{4n^2-1}$, which yields $\frac{4n-1}{4n^2+4n-1}$. Therefore, we have that $\mu_{i_k,x_n} \leq 1-\mu_{b_n,d_n} < \frac{1}{2n} < \frac{4n-1}{4n^2+4n-1}$ for all $i_k \in \{b_k, d_k\}$ and $k = 1, \ldots, n-1$, thereby obtaining that for all μ such that $\mu_{b_n,d_n} > 1-\frac{1}{2n}$, the strategy profile with $\sigma_{b_n} = \sigma_{d_n} = \{I\}$ and $\sigma_i = \{NI\}$, for all $i \neq b_n, d_n$, constitutes a BNE. ■

Proof of Proposition 4

Let $\eta_G(\phi)$ denote the average effort in group G, given ϕ. By taking players' best responses, we have that

$$\begin{aligned}
\eta_G(\phi) =& \frac{1}{|G|}\left(e(I_G^\phi) + \sum_{i\in O_G^\phi} \frac{1+t\eta_{G'}(\phi)}{c_i}\right)\\
=& \frac{1}{|G|}\left\{e(I_G^\phi) + \sum_{i\in O_G^\phi} \frac{1}{c_i}\left[1+\frac{t}{|G'|}\left(e(I_{G'}^\phi) + \sum_{i\in O_{G'}^\phi} \frac{1+t\eta_G(\phi)}{c_i}\right)\right]\right\}
\end{aligned}$$

where $\beta_i(e) = \eta_{G'}(\phi)$ for $i \in O_G^\phi$ by consistency, and $e(I_G^\phi) \equiv \sum_{i\in I_G^\phi} \frac{c_{\phi(i)}+t}{c_i c_{\phi(i)}-t^2}$. Solving for $\eta_G(\phi)$ yields

$$\eta_G(\phi) = \frac{|G'|\left(e(I_G^\phi) + \sum_{i\in O_G^\phi}\frac{1}{c_i}\right) + t\left(\sum_{i\in O_G^\phi}\frac{1}{c_i}\right)\left(e(I_{G'}^\phi) + \sum_{i\in O_{G'}^\phi}\frac{1}{c_i}\right)}{|G||G'| - t^2\left(\sum_{i\in O_G^\phi}\frac{1}{c_i}\right)\left(\sum_{i\in O_{G'}^\phi}\frac{1}{c_i}\right)}.$$

Therefore, whenever $i \in O_G^\phi$, her equilibrium effort is given by $\frac{1+t\eta_{G'}(\phi)}{c_i}$. Whereas, for $i \in I_G^\phi$, the equilibrium effort of agent i is equal to $\frac{c_{\phi(i)}+t}{c_i c_{\phi(i)}-t^2}$.

As far as uniqueness is concerned, it suffices to note that the best-response of player i is single-valued for any $\beta_i(e) \in [0, \bar{e}]$ because of the strict concavity of the utility function, and the assumptions $\underline{x} \geq 1$ and $\bar{e} \geq 1$. Given that $\eta_G(\phi) \in [0, \bar{e}]$, there is a unique equilibrium (in pure strategies). ■

Proof of Proposition 5

---

[19] Since we are simply looking for a lower bound, we can ignore the fact that setting $\mu_k \to 1$ would have the effect that $\mu_{i_k,x_n} \to 0$. Whenever μ_{i_k,x_n} is below the lower bound, the inequality is satisfied for any value that μ_k might actually take between 0 and 1.

1. The proof relies of techniques similar to those used for proving the optimal sorting in Becker (1973). We prove the first statement in two steps. In step 1, we establish that the part of the designer's objective function which is obtained via ingroup pairs is supermodular in the costs of effort. In step 2, we make use of supermodularity to show that the optimal matching must be one-sided assortative in I_G^ϕ.

   Step 1. Let $P(c_i, c_j) \equiv p(e_i(c_i, c_j), e_j(c_j, c_i))$, where $e_i(c_i, c_j) = \frac{c_j+t}{c_i c_j - t^2}$. We first show that the cross derivative of $P(c_i, c_j)$ is positive by use of the chain rule. That is,

$$\begin{aligned}\frac{\partial^2 p(e_i(c_i,c_j), e_j(c_j,c_i))}{\partial c_j \partial c_j} &= t\frac{\partial e_j}{\partial c_j}\frac{\partial e_i}{\partial c_i} + (1+te_j)\frac{\partial^2 e_i}{\partial c_j \partial c_i} \\ &+ t\frac{\partial e_i}{\partial c_j}\frac{\partial e_j}{\partial c_i} + (1+te_j)\frac{\partial^2 e_j}{\partial c_j \partial c_i} > 0\end{aligned}$$

   where the inequality holds because all first order derivatives of efforts in the equation are negative and the cross derivatives are positive.

   We can conclude that P is supermodular in the costs: $\frac{\partial^2 P}{\partial c_j \partial c_i} > 0$ implies that, if $c_i < c_i'$ and $c_j < c_j'$, then

$$P(c_i', c_j) - P(c_i, c_j) < P(c_i', c_j') - P(c_i, c_j') \tag{10}$$

   Step 2: By contradiction, assume that ϕ solves (5) and it is not one-sided assortative in I_G^ϕ. Relabel the agents so that $I_G^\phi = \{1, 2, \ldots, k\}$ for some $k \leq |G|$, where $c_i < c_{i+1}$ for $i = 1, \ldots, k-1$. Without loss, let $i < j$ for any $(i, j) \in \phi$. Since ϕ is assumed not to be one-sided assortative in I_G^ϕ, there exists some pair $(i, j) \in \phi \cap (I_G^\phi \times I_G^\phi)$ such that $c_j > c_{i+1}$. Take the first such pair, we have:

$$\begin{aligned}P(c_i, c_j) + P(c_{i+1}, c_{\phi(i+1)}) &= P(c_i, c_j) + P(c_{\phi(i+1)}, c_{i+1}) \\ &< P(c_i, c_{i+1}) + P(c_{\phi(i+1)}, c_j)\end{aligned}$$

   by (10) because $c_i < c_{\phi(i+1)}$ and $c_{i+1} < c_j$.

   Furthermore, the rematching also increases the average effort in group G, thereby inducing greater effort levels for all the agents in G' matched outgroup (if any), which in turn induces a weakly greater average effort in G', so that also the effort of all agents in G increases weakly.

Overall, the objective function under the alternative matching is strictly greater. Hence, a contradiction.

2. Assume $\{(i_1, j_2), (i_2, j_1)\} \subseteq \phi$, with $c_{i_1} < c_{j_1}$ and $c_{i_2} < c_{j_2}$, where $i_1, j_1 \in O^{\phi}_{G_1}$ and $i_2, j_2 \in O^{\phi}_{G_2}$.

   Consider a matching ϕ' identical to ϕ except for the fact that ϕ' has the matches (i_1, i_2) and (j_1, j_2). Since in both matchings these pairs are outgroup, then $\eta_G(\phi') = \eta_G(\phi)$, for all G, and $e_i(\phi') = e_i(\phi)$, for all i. Given $e_{i_1}(\phi) > e_{j_1}(\phi)$ and $e_{i_2}(\phi) > e_{j_2}(\phi)$, it must be the case that

$$p\left(e_{i_1}(\phi), e_{i_2}(\phi)\right) + p\left(e_{j_1}(\phi), e_{j_2}(\phi)\right) > p\left(e_{i_1}(\phi), e_{j_2}(\phi)\right) + p\left(e_{j_1}(\phi), e_{i_2}(\phi)\right)$$

   where the difference between the LHS and RHS is the difference in the objective function of the designer under ϕ' as compared to ϕ. Hence, ϕ cannot be a solution to (5). ■

Proof of Proposition 6

3. We prove the statement by contradiction, showing that there is an improvement to the matching ϕ, when $(b, d), (b', d') \in \phi$.

   Let $P(c_i, c_j) \equiv p\left(\frac{c_j+t}{c_i c_j - t^2}, \frac{c_i+t}{c_i c_j - t^2}\right)$ as in part (1) of the proof of Proposition 5. Since, for all four agents, the effort exerted under ϕ is greater than the average in both groups, we have $\frac{c_{\phi(i)}+t}{c_i c_{\phi(i)} - t^2} \geq e_i(\phi)$ for $i = b, b', d, d'$, and then:

$$\begin{aligned} p\left(e_b(\phi), e_d(\phi)\right) + p\left(e_{b'}(\phi), e_{d'}(\phi)\right) &\leq P(c_b, c_d) + P(c_{b'}, c_{d'}) \\ &< P(c_b, c_{b'}) + P(c_d, c_{d'}) \end{aligned}$$

   where the second inequality holds for the supermodularity of $P()$, because $\min\{c_d, c_{d'}\} \geq \max\{c_b, c_{b'}\}$, and $c_b \neq c_{b'}$ or $c_d \neq c_{d'}$.

   Consider the matching ϕ' identical to ϕ with the exception of the pairs $(b, b'), (d, d') \in \phi'$. We have that $P(c_b, c_{b'}) + P(c_d, c_{d'}) = p\left(e_b(\phi'), e_{b'}(\phi')\right) + p\left(e_d(\phi'), e_{d'}(\phi')\right)$.

Moreover, under a matching ϕ', the average effort would increase for both groups, thereby implying that all other pairs' efforts would be weakly greater than under ϕ. Hence, ϕ cannot be optimal.

4. We prove the statement by contradiction. Let $(b, b'), (d, d') \in \phi$, where ϕ is an optimal matching.

   Since the effort of all four agents is weakly smaller than the average efforts in both groups, we have

   $$f(i) \equiv \frac{1 + t \min\{\eta_{G_B}(\phi), \eta_{G_D}(\phi)\}}{c_i} \geq \frac{1 + t e_{\phi(i)}(\phi)}{c_i} = e_i(\phi),$$

   for $i \in \{b, b', d, d'\}$, where the inequality holds by strategic complementarity. Given the $p(\cdot, \cdot)$ is supermodular in the efforts, we have:

   $$\begin{aligned} p\left(e_b(\phi), e_{b'}(\phi)\right) + p\left(e_d(\phi), e_{d'}(\phi)\right) &\leq p\left(f(b), f(b')\right) + p\left(f(d), f(d')\right) \\ &< p\left(f(b), f(d)\right) + p\left(f(b'), f(d')\right) \end{aligned}$$

   where the second inequality holds because $\min\{c_b, c_d\} \leq \max\{c_{b'}, c_{d'}\}$, and $c_b \neq c_{b'}$ or $c_d \neq c_{d'}$.

   Consider the matching ϕ' identical to ϕ except for the pairs (b, d) and (b', d') in ϕ'. Note that for each $i \in \{b, b', d, d'\}$, the fact that $f(i) \geq e_i(\phi)$ together with the assumption that all but the two aforementioned matches are identical in ϕ' and ϕ implies that $e_i(\phi') \geq f(i)$ and $\eta_G(\phi') \geq \eta_G(\phi)$, for all G. This is because the four agents are paired outgroup in ϕ' and so they best respond to the average effort in their respective outgroup. Hence, their best-reply cannot be lower than the effort exerted before ($e_i(\phi') \geq e_i(\phi)$ for $i \in \{b, b', d, d'\}$) and therefore for each G, $\eta_G(\phi') \geq \min\{\eta_{G_B}(\phi), \eta_{G_D}(\phi)\}$. This being established, we have that:

   $$\begin{aligned} p\left(e_b(\phi'), e_d(\phi')\right) + p\left(e_{b'}(\phi'), e_{d'}(\phi')\right) &\geq p\left(f(b), f(d)\right) + p\left(f(b'), f(d')\right) \\ &> p\left(e_b(\phi), e_{b'}(\phi)\right) + p\left(e_d(\phi), e_{d'}(\phi)\right) \end{aligned}$$

   Furthermore, $\eta_G(\phi') \geq \eta_G(\phi)$, for all G, also implies that $e_i(\phi') \geq e_i(\phi)$ for all $i \in \mathcal{I}$ so that the value of the designer's objective function is strictly greater under ϕ' as compared to ϕ. Hence, a contradiction. ■

Proof of Lemma 1.
Denote by $I(c,\varepsilon)$ an interval $(c-\varepsilon, c+\varepsilon)$.

Let ϕ be such that $\nu_{In_k} > 0$ and for some open $I \subseteq In_k$, $\phi_k(I \times I) < \nu_{In_k}(I)$. Then, for some ε small enough there must be some disjoint $I(c,\varepsilon) \subseteq I$ and $I(c',\varepsilon) \not\subseteq I$ both with positive measure and $\phi_k(I(c,\varepsilon) \times I(c',\varepsilon)) \equiv \kappa > 0$ (by continuity of ϕ_k).

The effort of each agent in $I(c',\varepsilon)$ matched with some agent-type in $I(c,\varepsilon)$ is approximated by $\frac{c+t}{cc'-t^2}$. This implies that the probability of success of the tasks over all the agents in $I(c,\varepsilon)$ and their respective matches in $I(c,\varepsilon)$ is approximately

$$\kappa\, p\left(\frac{c'+t}{cc'-t^2}, \frac{c+t}{cc'-t^2}\right) \tag{11}$$

We show that there is a matching that constitutes an improvement to ϕ. Consider the matching $\tilde{\phi}$ identical to ϕ, except for the fact that $\phi_k(I(c,\varepsilon) \times I(c',\varepsilon)) = 0$, while $\tilde{\phi}_k(I(c,\varepsilon) \times I(c,\varepsilon)) \geq \frac{\kappa}{2}$ and $\tilde{\phi}_k(I(c',\varepsilon) \times I(c',\varepsilon)) \geq \frac{\kappa}{2}$

Under $\tilde{\phi}$, the equilibrium effort is approximately $\frac{1}{c-t}$ in $I(c,\varepsilon)$ and $\frac{1}{c'-t}$ in $I(c',\varepsilon)$. This implies that the probability of success of the tasks in $I(c,\varepsilon)$ and in $I(c',\varepsilon)$ is no smaller than

$$\frac{\kappa}{2}p\left(\frac{1}{c-t}, \frac{1}{c-t}\right) + \frac{\kappa}{2}p\left(\frac{1}{c'-t}, \frac{1}{c'-t}\right). \tag{12}$$

Assume W.L.O.G. that $c < c'$ so that $\frac{1}{c'-t} < \frac{c+t}{cc'-t^2} < \frac{c'+t}{cc'-t^2} < \frac{1}{c-t}$. By the supermodularity of $p(\cdot,\cdot)$, we have that (12) is strictly greater than (11).

Furthermore, the rematching in $\tilde{\phi}$ strictly increases the average effort in group G. This induces higher effort levels for all the agents in G' that are matched outgroup (if any), thereby increasing (weakly) the average effort in G', so that the effort of all agents matched outgroup in G (if any) increases (weakly).

Overall, the objective function of the designer is necessarily strictly greater under $\tilde{\phi}$ as compared to ϕ. Hence, ϕ cannot be optimal. ■

Proof of Lemma 2.
Denote by $I(c,\varepsilon)$ an interval $(c-\varepsilon, c+\varepsilon)$.

Let ϕ be such that $\nu_{Out_1}(I(c_1,\varepsilon_1)) > 0$, but there is no c_2' and ε_2' such that, with $\phi_O(I(c_1,\varepsilon_1) \times I(c_2',\varepsilon_2')) = \nu_{Out_2}(I(c_2',\varepsilon_2')) > 0$ and $Q_1^{\phi}(c_1) = Q_2^{\phi}(c_2')$. By the assumption of continuity for the matching structure ϕ, one can always find an interval $I(c_2', \varepsilon_2')$ satisfying the conditions over ϕ_O. So it must be the case that $Q_1^\phi(c_1) \neq Q_2^\phi(c_2')$. Assume without loss that $Q_1^\phi(c_1) > Q_2^\phi(c_2')$.

Because the sets Out_1 and Out_2 must have the same mass, there exists also some $c_1' > c_2'$ and some neighborhood $I(c_1', \varepsilon_1')$ such that I is a subset of $I(c_1', \varepsilon_1')$ with measure $\phi_O(I \times Out_2) = \nu_{Out_1}(I) > 0$ and some measurable $I' \subseteq I(c_2, \varepsilon_2) \subseteq Out_2$ such that $\phi_O(I \times I') > 0$, for some c_2 such that $Q_2^\phi(c_1') < Q_2^\phi(c_2)$.

For the various ε terms small enough, the equilibrium efforts in set $I(c_1, \varepsilon_1)$, (resp. $, I(c_2', \varepsilon_2')$, I, and I') can be approximated by $e(c_1, c_2^*)$ (resp. $e(c_2', c_1^*)$, $e(c_1', c_2^*)$, and $e(c_2, c_1^*)$). Then, that the probability of success of the tasks within these matches is approximately

$$\phi_O(I(c_1, \varepsilon) \times I(c_2', \varepsilon'))\, p\left(\frac{1 + t\eta_2(\phi)}{c_1}, \frac{1 + t\eta_1(\phi)}{c_2'}\right) + \phi_O(I \times I')\, p\left(\frac{1 + t\eta_2(\phi)}{c_2}, \frac{1 + t\eta_1(\phi)}{c_1'}\right) \tag{13}$$

We show that there is a matching that constitutes an improvement to ϕ. Let $\lambda \equiv \min\{\phi_O(I(c_1, \varepsilon) \times I(c_2', \varepsilon')), \phi_O(I \times I')\}$. Consider the matching $\tilde{\phi}$ that is identical to ϕ, except for the following two type of rematches. First, a mass λ of agents in I is matched with a mass λ of agents in $I(c_1, \varepsilon)$. Second, the respective matches under ϕ are now matched together so that a mass λ of agents in $I(c_2', \varepsilon_2')$ is matched with mass λ of agents in I, under $\tilde{\phi}$.

The equilibrium efforts are not affected by the rematches, but under $\tilde{\phi}$ the probability of success of the tasks over all the re-matched agents in I and their respective matches, and over all re-matched agents in I' and their respective matches is approximately

$$\lambda\, p\left(\frac{1 + t\eta_2(\phi)}{c_1}, \frac{1 + t\eta_1(\phi)}{c_1'}\right) + \lambda\, p\left(\frac{1 + t\eta_2(\phi)}{c_2}, \frac{1 + t\eta_1(\phi)}{c_2'}\right) \tag{14}$$

Expressions (14) and (13) show that $\tilde{\phi}$ improves strictly over ϕ due to the supermodularity of $p(\cdot, \cdot)$ because $c_1 > c_2$ and $c_1' > c_2'$. ■

Proof of Proposition 7

I show that if either Condition 1 or 2 does not hold for some matching structure ϕ, then there exists some $\tilde{\phi}$ that is a local improvement of ϕ.

For $I \in \{In_k, Out_k\}$, for some k, a set $I' \subseteq I$ is referred to as the set with the smallest (*respectively*, largest) c in I when I' is a Borel set with $\nu_I(I') > 0$ such that, for some $\hat{c}$, $\hat{c} = \inf_c I'$ and $\sup_c\{c \mid \nu(I \cap (\underline{c}, c)) = 0\} = \hat{c}$ (*respectively*, $\hat{c} = \sup_c I'$ and $\inf_c\{c \mid \nu(I \cap (c, \bar{c})) = 0\} = \hat{c}$).

The function $e(c, c') \equiv \frac{1+t\frac{1}{c'-t}}{c}$ describes the equilibrium effort for all agents. If $c_k \in In_k$, then $e(c_k, c_k) = \frac{1}{c_k - t}$ by assortativity of ϕ, while if $c_k \in Out_k$, then $e(c_k, c^*_{k'}) = \frac{1}{c_k}\left(1 + \frac{1}{c^*_{k'} - t}\right)$ by the definition of c^*.

Condition 1. Assume by contradiction that, for some $k \in \{1, 2\}$, there exists a Borel set $A_k \subseteq Out_k \cap (\underline{c}, c^*_1)$. We focus on the set A_k with the smallest c. That is, A_k is such that $Q^{\phi}_k(\hat{c}) = \nu(Out_k)$, for $\hat{c} = \inf_c A_k$.

By the assumption on continuity of ϕ, there exists an open interval $A_{k'} \subseteq Out_{k'}$, for $k' \neq k$, such that $\phi_O(A_1 \times A_2) = \nu_{Out_{k'}}(A_{k'}) > 0$.

Note that for each type-cost c_k in A_k, the effort according to ϕ is given by the function $e(c_k, c^*_{k'})$, which is strictly smaller than $e(c_k, c_k) = \frac{1}{c_k - t}$, the effort exerted by a type-cost c_k if it were matched ingroup according to some assortative matching structure.

Therefore, if $\nu_{Out_{k'}}(A_{k'} \cap (\underline{c}, c^*_1)) > 0$, we can improve on ϕ by re-matching ingroup (assortatively) all type-costs in $A_{k'}$ and re-matching ingroup (assortatively) a subset $B \subseteq A_k$ with equal mass to $A_{k'}$ (according to the population measure ν) such that $\phi_O(B \times A_{k'}) = \nu_{Out_k}(B)$ (where B excludes those type-costs in A_k that are not matched with $A_{k'}$). Let $\tilde{\phi}$ be an assortative matching structure identical to ϕ, except that $A_{k'} \subseteq \widetilde{In}_{k'}$ and $B \subseteq \widetilde{In}_k$. Note that all c-types re-matched in $\tilde{\phi}$ exert strictly greater effort, as $e(c, c) > e(c, c^*_1)$; all type-costs matched ingroup in both ϕ and $\tilde{\phi}$ exert the same effort; and all outgroup matches only best respond to the average. This ensures that the average effort in each group is strictly greater under $\tilde{\phi}$, and thus the designer's objective function is strictly better under $\tilde{\phi}$ than under ϕ.

If instead $\nu_{Out_{k'}}(A_{k'} \cap (\underline{c}, c^*_1)) = 0$, we can find a local improvement by noting that there must be a Borel set I of ingroup matches with cost-types greater than a positive measure of cost-types in A_k (this follows from the symmetry of the groups, the assortativity of ϕ, and the fact that $Q^{\phi}_k(\inf_c A_k) = \nu(Out_k)$).

Let I be a set with the largest c in In_k, and denote the largest c in In_k by $\hat{c}$. Take a measurable subset $I' \subseteq I \subseteq In_k$ of mass $\lambda > 0$ with cost-types near $\hat{c}$. Consider a subset $A' \subseteq A_k \subseteq Out_k$ of mass λ with cost-type agents such that $\hat{c} > \sup_c A'$.

Let $\tilde{\phi}$ be a matching structure identical to ϕ, except that $\tilde{In}_k = In_k \cup A' \setminus I'$, with $\tilde{\phi}_O(I' \times A_{k'}) = \nu_{Out_k}(I')$ and $\tilde{\phi}_k(A' \times A') = \nu_{In_k}(A')$. To see that the matching structure $\tilde{\phi}$ improves upon ϕ, consider the following cases.

Case 1: $\hat{c} > c_2^*$, or, $c_1^* < \hat{c} < c_2^*$ and $k = 2$. In both cases, we have that $c_i \in I' \subseteq In_k$ implies $\frac{1}{c_i - t} < \frac{1}{c_{k'}^* - t}$. Then, for any $c_i \in I'$, $e^{\phi}(c_i, c_i) < e(c_i, c_{k'}^*)$. Also, for $c_a \in A'$, $e^{\phi}(c_a, c_{k'}^*) < e^{\tilde{\phi}}(c_a, c_a)$. The direction of the two inequalities implies that $\eta_k(\tilde{\phi}) > \eta_k(\phi)$ and $\tilde{c}_k^* < c_k^*$ (and $e(c_i, c_{k'}^*) < e^{\tilde{\phi}}(c_i, \tilde{c}_{k'}^*)$). For all $c_{k'} \in Out_{k'}$, including $c_{k'} \in A_{k'}$, $e^{\tilde{\phi}}(c_{k'}, \tilde{c}_k^*) > e^{\phi}(c_{k'}, c_k^*)$. For all $c \in In_{k'}$, $e^{\tilde{\phi}}(c, c) = e^{\phi}(c, c)$. Therefore, $\eta_{k'}(\tilde{\phi}) > \eta_{k'}(\phi)$. The designer's objective function thus increases strictly for outgroup matches in both groups, while it is unchanged for those ingroup matches that coincide in ϕ and $\tilde{\phi}$. Finally, taking λ small enough, I can approximate the change in the objective function within the re-matched cost-types for $c_i \in I'$, $c_a \in A'$, and $c_{k'} \in A_{k'}$,

$$
\begin{aligned}
&\lambda p(e^{\tilde{\phi}}(c_i, c_i), e^{\tilde{\phi}}(c_i, c_i)) + \lambda p(e^{\tilde{\phi}}(c_a, \tilde{c}_{k'}^*), e^{\tilde{\phi}}(c_{k'}, \tilde{c}_k^*)) > \\
&\lambda p(e^{\phi}(c_a, c_a), e^{\phi}(c_a, c_a)) + \lambda p(e^{\phi}(c_i, c_{k'}^*), e^{\phi}(c_{k'}, c_k^*))
\end{aligned}
$$

Case 2: $c_1^* < \hat{c} < c_2^*$ and $k = 1$. In this case, for $c_i \in I'$, $\frac{1}{c_2^* - t} < \frac{1}{c_i - t} < \frac{1}{c_1^* - t}$, and so re-matching cost-types in I' outgroup is not effort-enhancing within I'—for a fixed average $\eta_2(\phi)$—as $e^{\phi}(c_i, c_i) > e(c_i, c_2^*)$. However, the re-matching in $\tilde{\phi}$ re-matches a mass λ in I' outgroup, while re-matching a mass λ in A' ingroup, which is effort-enhancing, as for $c_a \in A'$, $e(c_a, c_a) > e^{\phi}(c_a, c_1^*)$. It can be easily verified that such a gain in effort outweighs the cost by noting that, for $c > c'$, the difference $e(c, c) - e(c, c')$ is increasing in c, and $c_a > c_i > c_1^*$ for any $c_a \in A'$ and $c_i \in I'$. As such, the overall average in group 2 must be larger after the re-matches. That is, $\eta_2(\tilde{\phi}) > \eta_2(\phi)$ for any positive mass of re-matches λ. This increases the effort of all outgroup matches in G_1 too, so that $\eta_1(\tilde{\phi}) > \eta_1(\phi)$.

While the change in the objective function for the mass λ of re-matched cost-types is not guaranteed to be positive (call it a loss), there is a strictly positive gain under $\tilde{\phi}$ compared to ϕ for all outgroup matches. Recall that λ is taken arbitrarily such that $0 < \lambda < \nu_{Out_1}(I)$. Note that decreasing the mass λ of re-matches has three effects: the difference in the averages under ϕ and $\tilde{\phi}$ becomes smaller; the (potential) loss incurred in the objective function occurs for a smaller mass of cost-types; and the mass of cost-types for which there is a strict gain increases ($\nu(Out_1) - \lambda \geq \nu_{Out_1}(I) - \lambda > 0$ and increasing as λ gets smaller). By taking $\lambda > 0$ small enough, $\eta_k(\tilde{\phi}) > \eta_k(\phi)$ for each k is small but still positive, and the mass λ for which a loss might occur is negligible when compared to the gain occurring for the mass $\nu(Out_1) - \lambda$ of outgroup matches. This concludes the argument for showing that Condition 1 is necessary for optimality of the matching structure.

Condition 2. Assume by contradiction that, for some $k \in \{1, 2\}$, there is a Borel set $A_k \subseteq In_k \cap (c_2^*, \bar{c})$. We focus on the set A_k with the largest c. By the assumption that groups have the same mass and outgroups must also have the same mass for the matching to be balanced, there must be a positive mass matched ingroup in group k' as well. We focus on the Borel set $A_{k'} \subseteq In_{k'}$ with the largest c in $In_{k'}$.

If $\nu_{In_{k'}}(An_{k'} \cap (c_2^*, \bar{c})) > 0$, it is easily verified that re-matching outgroup the cost-types in A_k above c_2^* with the cost-types in $A_{k'}$ above c_2^* is both effort-enhancing and improves the designer's objective function.

If $\nu_{In_{k'}}(An_{k'} \cap (c_2^*, \bar{c})) = 0$, all Borel sets of cost-types above $\sup_c A_{k'}$ are matched outgroup, and $\sup_c A_{k'} > c_2^*$. This is because $A_{k'}$ is the set with the largest c in $In_{k'}$. Suppose that there is no mass of cost-types in A_k with cost weakly above c_1^*. Then there must be a positive measure of outgroup matches in group k' with cost-types above c_1^*. Then Condition 1 is violated and ϕ cannot be optimal.

Therefore, there must exist a subset $A' \subseteq A_{k'}$ with positive mass such that for all $c \in A'$, $c > c_1^*$. I propose to re-match outgroup the cost-types in A' with cost-types in A_k. If $k = 2$, then for each $c_1 \in A'$, $c_1 > c_2^*$, thus $e^{\phi}(c_1, c_1) < e(c_1, c_2^*)$, and similarly, for $c_2 \in A_2$, $c_2 > c_1^*$, thus $e^{\phi}(c_2, c_2) < e(c_2, c_1^*)$. Hence, outgroup rematches among these sets would be effort-enhancing and would increase the designer's objective function, so that ϕ cannot be optimal.

If $k = 1$, the proposed re-matching would not work, as the cost-types in $A' \subseteq A_2$ exert more effort as ingroup: for each $c_2 \in A'$, $c_1^* < c_2 < c_2^*$, thus $e^{\phi}(c_2, c_2) > e(c_2, c_1^*)$. However, take the Borel set $I \subseteq Out_1$ with the smallest c in Out_1. Note that there cannot be a positive measure of cost-types in I below c_1^*, otherwise Condition 1 is violated. Also, there must be a subset $I' \subseteq I$ with positive mass $\lambda > 0$ such that all cost-types in I' are above c_2^*. This is because almost all type-costs below those in A_2 are matched outgroup and $c_2 \in A_2$ is above c_2^*, and $I \subseteq Out_1$ is picked so that almost all types in G_1 with costs larger than those in I are matched ingroup. Since the groups have the same mass and distribution, $c_2 < c_1^*$ implies that $c_i > c_1^*$ for any $c_i \in I'$. Then, re-matching types in I' ingroup in $\tilde{In}_1$ under assortative $\tilde{\phi}$, while re-matching outgroup types in A' with those types that were matched with I' under ϕ, namely $\phi_O(I' \times Out_2)$, leads to an effort-enhancing rematch and the average effort is then greater under $\tilde{\phi}$. That is, $\eta_1(\tilde{\phi}) > \eta_1(\phi)$ and $\eta_2(\tilde{\phi}) > \eta_2(\phi)$. By picking $\tilde{\phi}$ so that the mass $\lambda > 0$ of re-matches is small enough, $\tilde{\phi}$ improves upon ϕ for the designer (as there is a positive gain for outgroup matches with mass $\nu(Out_1) - \lambda = \nu(Out_2) - \lambda \geq \nu_{Out_1}(I) - \lambda > 0$. ■

Proof of Proposition 9 (convergence)

Note that $\eta(\tau_k)$ is bounded above by $\frac{1}{\underline{c}-t}$. If $\eta(\tau_k)$ is weakly increasing at each iteration k, then the procedure must converge (given that the action space is closed and bounded.

We show that the threshold is weakly decreasing at each iteration by induction.

In Iteration 1, $\tau_0 = \bar{c}$, ϕ_0 matches all agents ingroup and assortatively. Since $\eta(\tau_0) = \mathbb{E}[\frac{1}{c-t}]$, then there exists $\tau_1 < \bar{c}$ such that $\frac{1}{\tau_1 - t} = \eta(\tau_0)$ in step 1. Moreover, since $\frac{1}{\tau_1 - t} = \eta(\tau_0)$ the function obtained in step 2, namely $e_{\phi_1}(c, \tau_1, \eta(\tau_0))$, is continuous at $c = \tau_1$ (and hence continuous everywhere). In step 3, the average is computed and it is readily verified that $\eta(\tau_1) > \eta(\tau_0)$. This can be seen by the fact that

$$\begin{aligned}\eta(\tau_1) &= \int_{\underline{c}}^{\tau_1} \frac{1}{c-t} f(c)dc + \int_{\tau_1}^{\bar{c}} \frac{1 + t\eta(\tau_0)}{c} f(c)dc \\ &> \int_{\underline{c}}^{\tau_1} \frac{1}{c-t} f(c)dc + \int_{\tau_1}^{\bar{c}} \frac{1}{c-t} f(c)dc = \eta(\tau_0)\end{aligned}$$

where the first terms of the summation is the same on both sides of the inequality and the second term is strictly greater on the LHS because $\eta(\tau_0) > \frac{1}{c-t}$ for all $c \in (\tau_1, \bar{c}]$.

By the induction hypothesis, assume that at iteration $k-1$ it is verified that $\tau_{k-1} \leq \tau_{k-2}$ and that $\eta(\tau_{k-1}) \geq \eta(\tau_{k-2})$. We want to show that then we have also $\tau_k \leq \tau_{k-1}$ and $\eta(\tau_k) \geq \eta(\tau_{k-1})$.

For what concerns step 1., note that $\eta(\tau_{k-1}) \in [\frac{1}{\bar{c}-t}, \frac{1}{\underline{c}-t}]$, so that there must exists τ_k such that $\frac{1}{\tau_k - t} = \eta(\tau_{k-1})$. Furthermore, since $\eta(\tau_{k-1}) \geq \eta(\tau_{k-2})$ and $\frac{1}{x-t}$ is decreasing in x (for $x \geq 1$ and $0 < t < 1$), we must have that $\tau_k \leq \tau_{k-1}$.

Then, in step 3. the expression average effort can be written as:

$$\eta(\tau_k) = \int_{\underline{c}}^{\tau_k} \frac{1}{c-t} f(c)dc + \int_{\tau_k}^{\tau_{k-1}} \frac{1+t\eta(\tau_{k-1})}{c} f(c)dc + \sum_{h=0}^{k-2} \int_{\tau_h}^{\tau_{h+1}} \frac{1+t\eta(\tau_{k-1})}{c} f(c)dc$$

$$> \int_{\underline{c}}^{\tau_1} \frac{1}{c-t} f(c)dc + \int_{\tau_k}^{\tau_{k-1}} \frac{1}{c-t} f(c)dc + \sum_{h=0}^{k-2} \int_{\tau_h}^{\tau_{h+1}} \frac{1+t\eta(\tau_{k-2})}{c} f(c)dc = \eta(\tau_{k-1})$$

where the first terms of the summation is the same on each side of the inequality, while the second and third terms are weakly greater on the LHS. The second terms is weakly greater because by construction $\eta(\tau_{k-1}) > \frac{1}{c-t}$ for all $c \in (\tau_k, \tau_{k-1})$ (the term is strictly whenever $\tau_{k-1} < \tau_k$). Whereas for the third term is suffices to observe that $\eta(\tau_{k-1}) \leq \eta(\tau_{k-2})$ by the induction hypothesis. ■

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Arrow, K. J. (1973). The theory of discrimination. In *Discrimination in labor markets* (pp. 3–33). Princeton University Press.

Battigalli, P. (1987). Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali. *unpublished undergraduate dissertation, Bocconi University, Milano.*

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy*, *81*(4), 813–846.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794.

Brewer, M. B. (1993). Social identity, distinctiveness, and in-group homogeneity. *Social cognition*, *11*(1), 150–164.

Carlsson, H., & van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, *61*(5), 989–1018.

Chauvin, K. (2020). *A misattribution theory of discrimination* [mimeo].

Esponda, I., & Pouzo, D. (2016). Berk-nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, *84*(3), 1093–1130.

Eyster, E., & Rabin, M. (2005). Cursed equilibrium. *Econometrica*, *73*, 1623–1672.

Frick, M., Iijima, R., & Ishii, Y. (2022). Dispersed behavior and perceptions in assortative societies. *American Economic Review*, *112*(9), 3063–3105.

Gibbons, R., LiCalzi, M., & Warglien, M. (2021). What situation is this? shared frames and collective performance. *Strategy Science*, *6*.

Gibbons, R., & Prusak, L. (2020). Knowledge, stories, and culture in organizations. *AEA Papers and Proceedings*, *110*, 187–92.

Goursat, L. (2024, January). *Essays on matching with limited information and bounded rationality* (Publication No. 2024ENPC0017) [Theses]. École des Ponts ParisTech.

Goursat, L., Jehiel, P., & Weber, G. (2025). Evolutionarily stable analogy-based expectation equilibrium. *Working Paper*.

Greinecker, M., & Kah, C. (2021). Pairwise stable matching in large economies. *Econometrica*, *89*(6), 2929–2974.

Gretsky, N. E., Ostroy, J. M., & Zame, W. R. (1992). The nonatomic assignment model. *Economic Theory*, *2*, 103–127.

Gretsky, N. E., Ostroy, J. M., & Zame, W. R. (1999). Perfect competition in the continuous assignment model. *Journal of Economic Theory*, *88*(1), 60–118.

Hart, S., Hildenbrand, W., & Kohlberg, E. (1974). On equilibrium allocations as distributions on the commodity space. *Journal of Mathematical Economics*, *1*(2), 159–166.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*(Volume 47, 1996), 237–271.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, *13*(2), 324–340.

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, *123*(2), 81–104.

Jehiel, P. (2022, July). *Analogy-Based Expectation Equilibrium and Related Concepts:Theory, Applications, and Beyond* (Working Papers No. halshs-03735680). HAL.

Jehiel, P., & Koessler, F. (2008). Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior*, *62*(2), 533–557.

Jehiel, P., & Mohlin, E. (2023). Categorization in games: A bias-variance perspective.

Jehiel, P., & Samet, D. (2007). Valuation equilibrium. *Theoretical Economics*, *2*(2), 163–185.

Jehiel, P., & Weber, G. (2025). Endogenous clustering and analogy-based expectation equilibrium. *arXiv preprint arXiv:2505.13022*.

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of personality and social psychology*, *54*(5), 778.

Kranton, R. E., & Sanders, S. G. (2017). Groupy versus non-groupy social preferences: Personality, region, and political party. *American Economic Review*, *107*(5), 65–69.

Linville, P. W. (1982). The complexity–extremity effect and age-based stereotyping. *Journal of personality and social psychology*, *42*(2), 193.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of personality and social psychology*, *57*(2), 165.

Lipnowski, E., & Sadler, E. (2019). Peer-confirming equilibrium. *Econometrica*, *87*(2), 567–591.

Liqui Lung, C. W. (2022, December). *Optimal Self-Screening and the Persistence of Identity-Driven Choices* (Janeway Institute Working Papers No. 2232). Faculty of Economics, University of Cambridge.

Mullen, B., & Hu, L.-T. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology*, *10*(3), 233–252.

Nöldeke, G., & Samuelson, L. (2018). The implementation duality. *Econometrica*, *86*(4), 1283–1324.

Osborne, M. J., & Rubinstein, A. (1998). Games with procedurally rational players. *American Economic Review*, 834–847.

Ostrom, T. M., Carpenter, S. L., Sedikides, C., & Li, F. (1993). Differential processing of in-group and out-group information. *Journal of Personality and Social Psychology*, *64*(1), 21.

Park, B., Judd, C., & Ryan, C. (1991). Social categorization and the representation of variability information. *European Review of Social Psychology - EUR REV SOC PSYCHOL*, *2*, 211–245.

Pettigrew, A. M. (1979). On studying organizational cultures. *Administrative Science Quarterly*, *24*(4), 570–581.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659–661.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under" almost common knowledge". *The American Economic Review*, 385–391.

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation.* Houghton Mifflin comp.

Spiegler, R. (2016). Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, *131*(3), 1243–1290.

Spiegler, R. (2020). Behavioral implications of causal misperceptions. *Annual Review of Economics*, *12*(Volume 12, 2020), 81–106.

Steiner, J., & Stewart, C. (2008). Contagion through learning. *Theoretical Economics*, *3*(4), 431–458.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social science information*, *13*(2), 65–93.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations/Brooks/Cole.*

Vives, X. (1999). *Oligopoly pricing: Old ideas and new tools.* MIT Press (MA).

Winter, E. (2004). Incentives and discrimination. *American Economic Review*, *94*(3), 764–773.