

Kets, Willemien; Sandroni, Alvaro

Working Paper

Challenging conformity: A case for diversity

Discussion Paper, No. 1585

Provided in Cooperation with:

Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Kets, Willemien; Sandroni, Alvaro (2015) : Challenging conformity: A case for diversity, Discussion Paper, No. 1585, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/146988>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



CMS-EMS
Center for Mathematical Studies in Economics
And Management Science

Discussion Paper #1585

“Challenging Conformity: A Case for Diversity”

Willemien Kets and Alvaro Sandroni

Northwestern University

November 16, 2015

Challenging Conformity: A Case for Diversity*

Willemien Kets[†]

Alvaro Sandroni[‡]

November 16, 2015

Abstract

Why do diverse groups outperform homogeneous groups in some settings, but not in others? We show that while diverse groups experience more frictions than homogeneous ones, they are also less conformist. Homogeneous groups minimize the risk of miscoordination, but they may get stuck in an inefficient equilibrium. Diverse groups may fail to coordinate, but if they do, they tend to attain efficiency. This fundamental tradeoff determines how the optimal level of diversity varies with social and economic factors. When it is vitally important to avoid miscoordination, homogeneous groups are optimal. However, when it is critical to implement new and efficient practices, diverse groups perform better.

*Part of the material incorporated here was previously in a paper entitled “A belief-based theory of homophily” by the same authors ([Kets and Sandroni, 2015a](#)). We thank David Ahn, Larbi Alaoui, Sandeep Baliga, Vincent Crawford, Vessela Daskalova, Georgy Egorov, Tim Feddersen, Matthew Jackson, Wouter Kager, Rachel Kranton, George Mailath, Niko Matouschek, Friederike Mengel, Rosemarie Nagel, Alessandro Pavan, Antonio Penta, Nicola Persico, Debraj Ray, Yuval Salant, Larry Samuelson, Paola Sapienza, Rajiv Sethi, Eran Shmaya, Andy Skrzypacz, Jakub Steiner, Colin Stewart, Jeroen Swinkels, and numerous seminar audiences and conference participants for helpful comments and stimulating discussions.

[†]Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu

[‡]Kellogg School of Management, Northwestern University. E-mail: sandroni@kellogg.northwestern.edu

1. Introduction

Interacting with people like ourselves allows us to stay in our comfort zone. People with a similar background pick up on the same subtle social cues, have a similar communication style, and have similar experiences and beliefs (Gudykunst, 2004). Accordingly, homogeneous societies tend to be more cohesive, experience less conflict, and have fewer coordination problems than more diverse ones (Jackson and Joshi, 2011). Yet, it pays for organizations to get out of their comfort zone. Using 15 years of data on a large and comprehensive sample of public U.S. corporations, Dezső and Ross (2012) show that, for innovation-focused firms, diversity leads to better performance. Similarly, for innovation-focused banks, increases in racial diversity are related to enhanced financial performance (Richard et al., 2003).¹ However, in other settings, diverse teams can experience severe communication problems, trumping all potential benefits of diversity (De Dreu, 2006). Infamously, in January 1990 Avianca Flight 52 crashed, killing 73 of the 158 people on board, in large part due to poor communication between U.S. air traffic control and the Colombian crew (Cushman, 1990).

We analyze the effects of diversity on performance in a formal model. We consider a setting where a manager chooses the composition of teams to maximize performance, that is, total payoffs. Players are matched with other players in their team to play a game. Players belong to different groups. If most of the team members belong to the same group, the team is homogeneous. If the distribution of team members over groups is more even, the team is diverse. Group membership is exogenous and observable; one can think of groups as types, such as race, gender, socioeconomic background, and so on.

Empirical evidence suggests that diversity can affect performance even if group membership does not affect payoffs. For example, mergers and acquisitions often fail to meet expectations due to incompatibilities in culture, even if there are obvious economic benefits (Cartwright and Cooper, 1993). Indeed, if a player’s background and experiences influence his strategic reasoning, then diversity can affect performance even if it has no direct effect on incentives. As noted by Schelling (1960, p. 57, pp. 96–98), which equilibrium is played in a game with multiple equilibria may depend on “who the parties are and what they know about each other.” Players with different backgrounds may thus select different strategies, even if all are fully rational and face the same incentives.

While intuitive, this is difficult to formalize using traditional game-theoretic methods. Clas-

¹Also see Alesina et al. (2013) for cross-country evidence on the positive impact of diversity of skilled immigration on economic development; Peri (2012) for state-level evidence that there is a positive association between immigration and TFP; and Ottaviano and Peri (2006) for evidence that people that live in U.S. cities that have become more culturally diverse experienced a significant increase in their wages. See Alesina and La Ferrara (2005) for a survey of the effects of ethnic diversity on economic policies and outcomes.

sical game theory lacks a formal language to describe how a player’s identity can affect his reasoning. To capture this, we enrich the standard game-theoretic framework by building on the dual process account of Theory of Mind in psychology. This theory posits that an individual has an initial, instinctive reaction, and then adapt his views by reasoning about what he would do if he were in the opponent’s position.² In our model, players’ instinctive reactions are modeled by impulses, that is, payoff-irrelevant signals that direct players to a course of action. Each player then introspects on his impulse. That is, each player uses his own impulse to form a conjecture about how his opponent might behave. In particular, a player who views his opponent as someone who is similar to himself may think that his opponent is likely to have a similar impulse. Players then formulate a best response to their conjecture. However, players recognize that their opponent may have gone through a similar process. This may lead them to revise their conjecture, and to formulate a best response to this revised conjecture. This process continues to higher orders and the limit of this process is an *introspective equilibrium*.

In an introspective equilibrium, players may fail to coordinate on a pure Nash equilibrium, yet achieve higher rates of coordination than in a mixed Nash equilibrium. This is consistent with experimental evidence (Mehta et al., 1994). While introspective equilibrium is based on psychology and allows for non-Nash behavior, this should not be conflated with irrationality. As we show, the behavior predicted by introspective equilibrium is always consistent with common knowledge of rationality.

Unlike in standard models, identity matters in an introspective equilibrium even if it is not directly payoff relevant. This is the case if players with different backgrounds have different impulses. For example, a major issue after the Sprint-Nextel merger in 2005 was that employees who were used to the rigid rules at Sprint approached a situation differently than employees accustomed to the more freewheeling culture of Nextel. Likewise, people who grew up in collectivist cultures tend to react differently than people from individualistic cultures (Hofstede, 2001).

Cultural differences can be difficult to predict. The French, for example, generally communicate in a more indirect way than Americans, yet they are more direct in providing criticism (Meyer, 2014). A Frenchman may know how to interpret criticism from a compatriot, while an American may be taken aback. In general, it is easier to understand the perspective of members of one’s own group, as opposed to outsiders.³

²See Epley and Waytz (2010) for a survey. The dual process account of Theory of Mind relies on a rapid instinctive process and a slower cognitive process. As such, it is related to the two-systems account of decision-making under uncertainty, popularized by Kahneman (2011), the foundations of which go back to the work of the psychologist William James (1890/1983). See Section 2 for more discussion.

³For evidence from psychology and neuroscience that supports this hypothesis, see the meta-analysis by Elfenbein and Ambady (2002) and the survey of de Vignemont and Singer (2006), respectively. For experimental

Accordingly, members of diverse teams may find it difficult to anticipate others’ instincts. As a result, diverse teams may experience more frictions than homogeneous teams. However, the very clash of perspectives that causes mishaps and misunderstandings in diverse teams may also stimulate people to implement new solutions rather than sticking with established routines. The case of Pixar illustrates this point. In 2000, Pixar had just released three blockbuster movies. Instead of continuing to work with the same team, Pixar hired Brad Bird, a director who had just come off a movie that had been a financial failure. Bird engaged the “black sheep” at Pixar – the artists “who have another way of doing things that nobody’s listening to.” Pixar thus intentionally chose a diverse team. Indeed, it hired Bird “to come shake things up” (Rao et al., 2008).

Pixar is renowned for stimulating diversity of thought, but the idea of bringing in outsiders to provide a fresh perspective is of course hardly new.⁴ To model the effects of diversity, we focus on coordination games.⁵ Examples of coordination problems in organizations abound. Coordination on the same communication style (e.g., direct or indirect) is central to communication (Cr  mer, 1993). Coordination on common meaning facilitates trade (Lazear, 1999a) and makes it possible for an organization to use implicit contracts (Camerer and Knez, 2002) and to deal with unforeseen contingencies (Kreps, 1990).

In coordination games, players face two types of difficulties: how to avoid miscoordination and how to avoid inefficient coordination. When players can accurately predict other players’ reactions, as in homogeneous teams, they can avoid miscoordination by conforming to what they expect others will do, even if that means coordinating on an inefficient option. By contrast, in diverse teams, players’ impulses are not very informative of others’ reactions, and choices are driven primarily by payoff considerations.

We use this insight to characterize the optimal team composition in different economic environments (i.e., payoff distributions). A manager (or planner) chooses the team composition before payoffs are realized. His goal is to maximize expected total payoffs. After payoffs are realized, players observe the game and select an action using introspection. In coordination games where no option is clearly superior, a player has an incentive to follow his impulse if he expects other players to follow theirs. The more homogeneous the team, the more aligned the evidence in economics, see Jackson and Xing (2014) and Le Coq et al. (2015).

⁴For example, organizations routinely hire outside consultants to work with their employees to implement novel practices, and include representatives from various groups when they set up transition management teams. Other notable examples include the design firm IDEO; a key feature of their organization is that “they throw a bunch of people with different backgrounds together in a room [...] Doctors, opera singers and anthropologists for example, and get them to brainstorm” (CBS 60 Minutes, 2013).

⁵We thus abstract away from incentive problems. This allows us to isolate the effect of identity and reasoning on team performance.

impulses are, and the lower the risk of miscoordination. It follows that in stable environments, where it is unlikely that new, Pareto improving, options will become available, it is optimal to have homogeneous teams.

Now consider a more changeable environment where superior alternatives are likely to arise. Coordinating on a new option gives higher payoffs than following established routines. In such environments, homogeneity has a downside: the ability to accurately predict others' impulses makes it difficult to break away from inefficient routines. If a player's initial impulse is to follow the inefficient practice, then it is likely that his opponent has a similar impulse. So, he has a strong incentive to conform. By contrast, if a player cannot anticipate the reaction of his opponent, then his choice tends to be guided by payoffs. Hence, if diverse teams manage to coordinate, they are more likely to coordinate on the efficient option. It follows that in changeable environments, diversity is optimal.

This is true even if there are no direct effect of diversity on payoffs. In our model, diversity is not valuable or detrimental in itself. Hence, the beneficial effects of diversity that we identify are above and beyond any direct, exogenously given ones. Our results demonstrate that diversity can improve performance in a much broader context than previously assumed: diverse teams may outperform homogeneous teams even in the absence of skill complementarities or differences in information.

The basic advantage of diversity that our model captures has been observed empirically. Homogeneous societies with a strong culture often find it more difficult to break out of inefficient equilibria than more diverse, open-minded societies ([Mokyr, 1990](#)). Likewise, organizations that foster an open, adaptive atmosphere are less likely to become mired in ineffective practices ([Pfeffer and Sutton, 1999](#)).

We next ask how the likelihood of efficient coordination depends on both economic conditions (i.e., payoff distribution) and social factors, such as group identity (i.e., correlation of impulses within a group). Since the introspective equilibrium is (essentially) unique in our games, we can obtain unambiguous comparative statics. In stable environments, a stronger group identity is conducive to coordination. In changeable environments, groups with a strong identity may be *less* likely to implement the efficient option. This is true even if they face a lower risk of miscoordination overall. These results reveal an interplay between cultural and economic factors, which would be missed if identity and introspection are not taken into account.

The driving forces behind our analysis are intuitive, yet they are difficult to capture with standard game-theoretic approaches. First, identity must be incorporated into the theory.⁶

⁶Existing work on identity in economics incorporates identity into the payoffs ([Akerlof and Kranton, 2000](#)). It does not seem straightforward to capture the idea that diversity can reduce the pressure to conform in such

Moreover, the theory must allow for miscoordination in some games and select an equilibrium in others. When players try to coordinate their actions, they encounter the problem of miscoordination as well as the problem of ensuring coordination on the efficient action. The former is about playing a Nash equilibrium versus not playing a Nash equilibrium. The latter is about selecting the “right” Nash equilibrium. Existing theories mostly focus on one or the other. For example, level- k models can model nonequilibrium behavior, while introducing payoff perturbations or modeling players’ learning process can be used to select a unique Nash equilibrium. Existing approaches typically cannot do both.⁷ Building on findings from psychology, we can tackle both issues simultaneously. This allows us to explain why players fail to coordinate on a Nash equilibrium in some settings, while selecting a particular Nash equilibrium in others. This, in turn, provides clear, intuitive, and unambiguous comparative statics and novel testable predictions.

The outline of this paper is as follows. We introduce the basic model in Section 2. Section 3 studies the optimal team composition in different economic environments. Section 4 provides comparative statics. Section 5 discusses the related literature, and Section 6 concludes. Proofs are in the appendix.

2. Model

2.1. Coordination and introspection

There are two groups, A and B , each consisting of a unit mass of players. Members of these groups are called A -players and B -players, respectively. Group membership is observable.⁸ Players are matched to play a coordination game G , with payoffs given by:

$$\begin{array}{cc} & s^1 & s^2 \\ \begin{array}{c} s^1 \\ s^2 \end{array} & \begin{array}{|c|c|} \hline v^*, v^* & z, y \\ \hline y, z & v, v \\ \hline \end{array} \end{array}, \quad v^* \geq v > z, v^* > y.$$

This game has two strict Nash equilibria: one in which both players choose s^1 , and one in which both players choose s^2 . Coordinating on s^1 is (Pareto) efficient, but it may also be risky.

models; see Section 5 for a discussion.

⁷A notable exception is quantal response equilibrium, or QRE (McKelvey and Palfrey, 1995). QRE does not deal with issues of diversity and identity.

⁸This is appropriate when groups differ in their demographic attributes, such as gender or race, or when people can signal their identity using markers (e.g., distinctive clothing or tattoos). All our results extend qualitatively to settings where group membership is imperfectly observable or even unobservable (or, equivalently, where players cannot condition their behavior on the group of their opponent, perhaps for legal reasons), as when groups are defined by socioeconomic or educational background, sexual orientation, or religion.

It is thus not clear a priori how to play this game. However, the context of the game may give players some insight on how to play. The dual process account of Theory of Mind in psychology suggests how players can use introspection to anticipate others' actions. According to the dual process account, people have impulses, and through introspection (i.e., by observing their own impulse) players can learn about the impulses of others and thus form a conjecture about their behavior. This may lead them to consider a different action than the one suggested by their impulse; realizing that their opponent may likewise adjust their behavior, they may revise their initial conjecture (see [Epley and Waytz, 2010](#), for a survey).^{9,10}

A person's instinctive reaction to a strategic situation is shaped by his background ([Triandis and Suh, 2002](#)). People with a similar background have a shared history or have similar experiences. They can thus be expected to have a similar instinctive reaction to a given context. People of different backgrounds, on the other hand, do not share the same history and may respond differently. Moreover, a shared history makes it easier to anticipate someone's instinctive response. Accordingly, players find it easier to predict the impulse of someone who is similar to them.¹¹

We formalize the dual process account of Theory of Mind as follows. Each player has an initial impulse to take an action. A player's impulse is payoff-irrelevant. It is influenced by his background (i.e., group) in the following way. Nature draws a (payoff-irrelevant) state $\theta_C = 1, 2$ for each group $C = A, B$. Each state $\theta_C = 1, 2$ is equally likely ex ante for each group C . The states are (positively) correlated across groups: conditional on the state θ_A being $m = 1, 2$, we have $\theta_B = m$ with probability $\lambda \in (\frac{1}{2}, 1)$, and likewise with the group labels interchanged. If $\theta_A = 1$ then the initial impulse of an A -player is to take action s^1 with probability $q \in (\frac{1}{2}, 1)$, independently across players; likewise, if $\theta_A = 2$, then an A -player has an impulse to choose action s^2 with probability q . Analogous statements hold for B -players. If q is close to 1, then group membership strongly influences impulses; if q is close to $\frac{1}{2}$, the effect of group membership on impulses is weak. We define the strength of players' *group identity* as the probability Q_{in} that two players from the same group receive the same impulse. Lemma

⁹These ideas have a long history in philosophy. According to [Locke \(1690/1975\)](#) people have a faculty of "Perception of the Operation of our own Mind" and [Mill \(1872/1974\)](#) writes that understanding others' mental states first requires understanding "my own case." [Russell \(1948\)](#) observes that "[t]he behavior of other people is in many ways analogous to our own, and we suppose that it must have analogous causes."

¹⁰[Robalino and Robson \(2015\)](#) interpret Theory of Mind as the ability to learn other players' payoffs, and shows that this confers an evolutionary benefit in volatile environments.

¹¹For experimental evidence from psychology and neuroscience that shows that it is easier to predict the behavior or expectations of similar people, see [Elfenbein and Ambady \(2002\)](#) and [de Vignemont and Singer \(2006\)](#), respectively. For experimental evidence in economics, see, e.g., [Jackson and Xing \(2014\)](#) and [Le Coq et al. \(2015\)](#).

[A.2](#) in the appendix shows that Q_{in} lies strictly between $\frac{1}{2}$ and 1, and is increasing in q . So, impulses are correlated within a group; and if q is close to 1, then group identity strongly guides impulses. The probability that players from different groups have the same impulse is denoted by Q_{out} . By Lemma [A.2](#), Q_{out} lies strictly between $\frac{1}{2}$ and Q_{in} . Thus, impulses are more strongly correlated within groups than across groups, reflecting the idea that players find it easier to anticipate the impulses of members with a similar background.

Players are matched in pairs to play the game. A player's first instinct is to follow his impulse, without any strategic considerations. We refer to this initial stage as level 0. That is, a level-0 strategy σ_j^0 for player j maps his impulse into an action $s^i = s^1, s^2$. At higher levels, players realize that if their opponent belongs to the same group, then they are likely to have a similar impulse. So, through introspection (i.e., by observing their own impulses), players obtain an informative signal about what their opponents will do. At level 1, a player formulates a best response to the belief that his opponent will follow her impulse. This defines a level-1 strategy σ_j^1 for each player j that maps his impulse and the identity of the opponent into an action. This introspective process continues to higher orders: at level $k > 1$, players formulate a best response to their beliefs about their opponents' action at level $k - 1$; this defines a level- k strategy σ_j^k for each player j . The levels do not represent actual behavior; they are merely constructs in a player's mind. We are interested in the limit of this process as the level k goes to infinity. If such a limit strategy σ_j exists for each player j , then the profile $\sigma = (\sigma_j)_j$ is an *introspective equilibrium*.

Proposition 2.1. [Existence and Uniqueness Introspective Equilibrium] *Every coordination game G has an introspective equilibrium $\sigma^G = (\sigma_j^G)_j$, and, it is essentially unique.*¹²

The proof follows from Proposition [A.3](#) in the appendix, which provides a complete characterization of the equilibrium for every combination of parameters.¹³ Moreover, Proposition [A.4](#) demonstrates that every introspective equilibrium is a correlated equilibrium. So, behavior in an introspective equilibrium is consistent with common knowledge of rationality ([Aumann, 1987](#)). However, players need not follow their impulse in an introspective equilibrium, unlike with action recommendations in correlated equilibrium or sunspots. This proves to be important for the relative performance of homogeneous and diverse teams as we discuss in Section [3](#). Another critical distinction is that the introspective process selects an (essentially) unique prediction, while games often have many correlated equilibria. This delivers a powerful tool. It allows us to characterize the optimal team composition and to obtain comparative statics

¹²That is, the range of parameters for which the introspective equilibrium is not unique has Lebesgue measure 0; see Appendix [A](#).

¹³[Kets and Sandroni \(2015b\)](#) show a similar result in games where identity plays no role.

regarding the likelihood of efficient coordination as a function of cultural and economic factors; see Section 3 and 4, respectively.

2.2. Teams

Players interact in teams. A manager assigns players to one of two *teams*, labeled T_1 and T_2 . Each team consists of a unit mass of players. Teams can be more or less diverse. For example, if all A -players are assigned to team T_1 (say), and all B -players to team T_2 , teams are completely homogeneous. If half of the players of each group are assigned to each of the teams, teams are (maximally) diverse. Each player is matched to play the coordination game with a member of the same team. That is, members of team T_1 are matched with members of T_1 , and members of T_2 are matched with members of team T_2 . Matchings within a team are uniform and independent across players.

A manager chooses the team assignments to maximize team performance. In the model, team performance is measured by the total payoffs. Suppose player j is matched with some player k and follows a strategy $\sigma_j(i_j, k)$ which maps his impulse $i_j = 1, 2$ and the identity of his opponent (viz., k) into an action $s = s^1, s^2$. If players j and k have impulses i_j and i_k , respectively, the payoff of player j is

$$\pi_j(\sigma_j(i_j, k), \sigma_k(i_k, j)),$$

where $\pi_j(s, s')$ is the payoff in the coordination game for a player if he chooses action s and his partner chooses action s' (e.g., $\pi_j(\sigma_j(i_j, k), \sigma_k(i_k, j)) = v^*$ if $\sigma_j(i_j, k) = \sigma_k(i_k, j) = s^1$). Taking the expectation over the distribution of impulses and the random matching (given the team assignments α), and summing over all players gives the total payoff $C^G(\sigma; \alpha)$, where G denotes the game (i.e., payoff realizations).

At the time a manager assigns players to teams, he does not know the payoffs in the game G . He chooses the team assignment α to maximize the total payoffs given the economic environment, i.e., the distribution of the payoffs. After payoffs are realized, players observe the payoffs and play the coordination game, taking the team assignment as given, following the introspective process described earlier. So, if players play according to the introspective equilibrium σ^G , then the manager's optimization problem is

$$\max_{\alpha} \mathbb{E}[C^G(\sigma^G; \alpha)], \tag{2.1}$$

where the expectation is taken over the possible payoff realizations. By Proposition 2.1, the manager's maximization problem (2.1) is well defined. We assume throughout that the payoff distribution is smooth in an appropriate sense; see Section 3.

Solving the maximization problem (2.1) is challenging because there are many possible team assignments α . However, by Lemma 2.2 below, it suffices to consider the team composition, that is, the unevenness with which the two groups are distributed across teams. For a given team assignment α , define the *team composition* d as:

$$d = \frac{1}{2} \cdot \left| \frac{\text{share of } A\text{-players assigned to } T_1}{\text{total measure of } A\text{-players}} - \frac{\text{share of } B\text{-players assigned to } T_1}{\text{total measure of } B\text{-players}} \right| + \frac{1}{2} \cdot \left| \frac{\text{share of } A\text{-players assigned to } T_2}{\text{total measure of } A\text{-players}} - \frac{\text{share of } B\text{-players assigned to } T_2}{\text{total measure of } B\text{-players}} \right|.$$

The team composition d measures the share of players that need to be reassigned in order to attain an even distribution of groups across teams (Duncan and Duncan, 1955). The team composition d lies between 0 and 1. If d is close to 1, then teams are homogeneous; if d is close to 0, then teams are (maximally) diverse. The next result shows that the total payoff depends only on the team composition, not on the exact team assignment.

Lemma 2.2. *Suppose team assignments α, α' lead to the same team composition d . Then, the expected total payoff in the introspective equilibrium is the same under both team assignments (i.e., $\mathbb{E}[C^G(\sigma^G; \alpha)] = \mathbb{E}[C^G(\sigma^G; \alpha')]$, with σ^G the introspective equilibrium for the game G).*

With some abuse of notation, we write $\mathbb{E}[C^G(\sigma^G; d)]$ for $\mathbb{E}[C^G(\sigma^G; \alpha)]$ when the team assignment α gives rise to the team composition d . Lemma 2.2 implies that we can simplify the manager's optimization problem significantly: his optimization problem becomes

$$\max_{d \in [0,1]} \mathbb{E}[C^G(\sigma^G; d)]. \quad (2.2)$$

The *optimal team composition* d^* is the team composition that maximizes the total payoff (i.e., d^* solves (2.2)). If the optimal team composition d^* is close to 0, then it is optimal to have diverse teams; if the optimal team composition d^* is close to 1, then it is optimal to have homogeneous teams. In the next section, we consider the optimal team composition for different economic environments.

3. Challenging conformity

We characterize how the benefits of diversity vary with the economic environment (i.e., the payoff distribution). We consider two extreme cases that differ in the likelihood of a substantial innovation. To fix ideas, consider the following game:

$$\begin{array}{cc} & \begin{matrix} s^1 & s^2 \end{matrix} \\ \begin{matrix} s^1 \\ s^2 \end{matrix} & \begin{array}{|cc|} \hline v^*, v^* & 0, 0 \\ \hline 0, 0 & 1, 1 \\ \hline \end{array} \end{array}, \quad v^* \geq 1.$$

If v^* is equal to 1, then coordination on either action gives the same payoff. If v^* is greater than 1, then coordinating on s^1 is better. In this example, the ratio v^*/v gives the value of coordinating on s^1 relative to s^2 . In general, the off-diagonal payoffs y and z need not be 0 and need to be taken into account. So, define the risk-adjusted ratio of payoffs by

$$R := \frac{v^* - y}{v - z}.$$

The risk-adjusted ratio R is high (i.e., $R \gg 1$) if the payoff v^* to coordinating on action s^1 or the payoff z to playing s^1 when the other player chooses s^2 is high relative to the payoffs v and y to action s^2 . If the payoffs to action s^1 and s^2 are the same, then $R = 1$.

From the viewpoint of the manager, the game payoffs are uncertain, that is, the payoffs are random variables.¹⁴ To keep notation simple, we take the payoffs to action s^2 to be fixed, while the payoffs to action s^1 are random.¹⁵ The payoff distribution has a well defined density $f(v^*, z)$. We restrict attention to economic environments where the higher-order moments of the payoffs are finite, that is,

$$\mathbb{E}[|v^*|^{1+\eta}], \mathbb{E}[|z|^{2+2\eta}], \mathbb{E}[|v^* \cdot z|^{1+\eta}] < \infty \quad (3.1)$$

for some $\eta > 0$. This includes the case where payoffs are bounded and many other cases.

3.1. Stable environments

We first consider economic environments that are stable in the sense that will affect the risk-adjusted payoff ratio significantly. That is, the risk-adjusted payoff ratio R is likely to be close to 1. In particular, any Pareto improvement is likely to be small.

Formally, for $\delta < 1$, two actions are δ -equally strong if the payoffs are in the $(1 - \delta)$ -neighborhood $U_{1-\delta}^{R=1}$ of the event that $R = 1$.¹⁶ The environment is δ -stable if the probability that the two actions are δ -equally strong is greater than δ . If δ is close to 1, then the risk-adjusted payoff ratio is likely to be close to 1.

The next result shows that in stable economic environments, homogeneity is optimal.

Theorem 3.1. [Stable Environments: Homogeneous Teams] *In stable economic environments, it is optimal to have homogeneous teams. For every $\varepsilon > 0$, there is $\delta < 1$ such that*

¹⁴With some abuse of notation, we use the same symbol (e.g., v^*) for both the random variable and its realization in the main text.

¹⁵Our results do not depend on this.

¹⁶So, the open neighborhood $U_{1-\delta}^{R=1}$ of the event that $R = 1$ contains all points (v^*, z) such that the distance between (v^*, z) and a point (\tilde{v}^*, \tilde{z}) for which $\tilde{R} = (\tilde{v}^* - y)/(v - \tilde{z}) = 1$ is less than $1 - \delta$ (in the usual Euclidean topology on \mathbb{R}^2). Since the risk-adjusted payoff ratio R is continuous, the risk-adjusted payoff ratio is close to 1 on $U_{1-\delta}^{R=1}$.

if the economic environment is δ -stable, then the optimal team composition d^* is greater than $1 - \varepsilon$.

Intuitively, if both actions are equally strong, total payoffs are maximized when the potential for miscoordination is minimized. As members of the same group can more accurately predict each others' reactions than the responses of members of the other group, the risk of miscoordinating is minimized when teams are homogeneous.

Theorem 3.1 is consistent with experimental evidence that shows that subjects are more successful at coordinating when they interact with their own group (Weber and Camerer, 2003; Chen and Chen, 2011; Jackson and Xing, 2014). It is also consistent with empirical evidence that demonstrates that conflict is minimized in homogeneous teams with congruent expectations and values; see Jackson and Joshi (2011) for a survey. For example, Reagans and Zuckerman (2001) find that diversity has a negative impact on communication. Indeed, homogeneous teams perform better than diverse teams on simple tasks that require ample coordination (Bowers et al., 2000).

Thus, if the primary aim is for players to coordinate, and it does not matter much which action they coordinate on, then it is optimal to have homogeneous teams to reduce the risk of miscoordination. While intuitive, standard approaches seem to be unable to deliver this result. For example, suppose the realized payoffs are as follows:

	s^1	s^2
s^1	1,1	0,0
s^2	0,0	1,1

Experimental evidence shows that in this game, subjects often fail to coordinate on one of the pure Nash equilibria. However, they manage to coordinate at a significantly higher rate than in the mixed Nash equilibrium (Mehta et al., 1994). This is consistent with our results. The game has a unique introspective equilibrium, and in this introspective equilibrium, all players follow their impulse (Proposition A.3(c)). Since impulses are correlated, the probability that players coordinate and choose the same action is (strictly) higher than in the mixed Nash equilibrium. On the other hand, since the correlation is imperfect, players do not play according to a pure Nash equilibrium. Standard approaches cannot capture this: equilibrium selection criteria either pick out one of the pure Nash equilibria (like Pareto efficiency), select the mixed Nash equilibrium (e.g., Harsanyi and Selten, 1988), or have no bite in this game (e.g., global games). More fundamentally, standard game-theoretic approaches are unable to model how identity influences behavior,¹⁷ and thus cannot explain how team composition can affect performance.

¹⁷An exception is the work of Akerlof and Kranton (2000, 2005). However, in their work and the literature that builds on it, identity affects payoffs and not reasoning; see Section 5 for a discussion.

Modeling the introspective process explicitly delivers intuitive results that are consistent with experimental evidence.

3.2. Changeable environments

In many environments of interest, an innovation can significantly increase the payoff to one of the actions. It is convenient to define

$$\tilde{R} := \frac{Q_{out}}{1 - Q_{out}}. \quad (3.2)$$

Note that $\tilde{R} > 1$. Action s^1 is δ -strong (for $\delta < 1$) if the payoffs are in the $(1 - \delta)$ -neighborhood $U_{1-\delta}^{R \geq \tilde{R}}$ of the event that the risk-adjusted payoff ratio is at least \tilde{R} . The environment is δ -changeable if (1) the probability that action s^1 is δ -strong is greater than δ and (2) the joint density $f(v^*, z)$ is positive whenever $R \geq \tilde{R}$. If δ is close to 1, then it is likely that the payoff ratio R is greater than 1.

So, both in stable and in changeable environments, coordinating on action s^1 (weakly) Pareto-dominates coordinating on s^2 . The critical difference is that is that the gain of coordinating on s^1 (relative to s^2) is likely to be limited in stable environments, but can be substantial in changeable environments.

The next result shows that in changeable environments, diversity is optimal.

Theorem 3.2. [Changeable Environments: Diversity] *In changeable economic environments, it is optimal to have diverse teams. For every $\varepsilon > 0$, there is $\delta < 1$ such that if the economic environment is δ -changeable, then the optimal team composition d^* is less than ε .*

Theorem 3.2 is consistent with empirical evidence that shows a positive effect of diversity on financial performance for organizations focused on innovation, but not for other types of organizations (e.g., [Richard et al., 2003](#); [Dezső and Ross, 2012](#)). Moreover, diverse management teams are better able to adapt to changing conditions ([Wiersema and Bantel, 1992](#)).

The mechanism through which diversity can improve performance in changeable environments can be illustrated with a simple game. Suppose realized payoffs are as follows:

	s^1	s^2
s^1	5,5	-1,0
s^2	0,-1	1,1

Suppose that groups have a strong identity (i.e., Q_{in} close to 1), and consider a player who is matched with a member of his own group. Since group identity is strong, the player's

impulse is highly informative of the impulse of his opponent. If the player has an impulse to play action s^2 , then, through introspection, he realizes that his opponent is likely to have a similar impulse. At level 1, the expected payoff of choosing action s^2 is thus close to 1, while the expected payoff of action s^1 is close to -1. Accordingly, it is optimal for the player to follow his impulse at level 1. A simple inductive argument shows that the same is true at higher levels. So, players who are matched with their own group may coordinate on the inefficient Nash equilibrium (i.e., (s^2, s^2)).¹⁸

Now suppose the player is matched with a member of the other group. Since impulses are less strongly correlated across groups (i.e., $Q_{out} < Q_{in}$), it is difficult for the player to put himself into his opponent's shoes. In the extreme case where impulses are minimally correlated across groups (i.e., Q_{out} close to $\frac{1}{2}$), a player's impulse is almost completely uninformative of his opponent's impulse. Consequently, the player assigns about equal probability to his opponent playing each action at level 1. His expected payoff from playing action s^2 is thus close to $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}$ at level 1, while the expected payoff of playing action s^1 is close to $\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot (-1) = 2$. Accordingly, a player who is matched with a player from the other group selects action s^1 at level 1 in this case, even if he has an impulse to play the other action. The same is true at higher levels. In effect, players' inability to anticipate the impulses of members of the other group reduces the pressure to conform.¹⁹ Team performance is thus maximized by maximizing the fraction of cross-group interactions.

Diversity can thus improve team performance even if groups do not have different information and there are no skill complementarities across groups. It does so by reducing conformity and inertia, allowing teams to coordinate on the Pareto-dominant equilibrium. There is ample evidence that groups have a tendency to follow familiar routines even in situations where they are no longer appropriate. While this reduces the risk of miscoordination, it also limits the group's ability to adapt to changing circumstances (Gersick and Hackman, 1990). Strategic complementarities often lead to inertia. For example, in their study of young high technology firms, Baron et al. (1996) show that choices made at the time a firm was founded strongly affects the time that new practices are adopted. Even if a new alternative presents a clear Pareto improvement, it is not always adopted. Individuals may choose an inefficient action because they believe that this is the norm for the group (Bicchieri and Fukui, 1999). Pfe-

¹⁸Van Huyck et al. (1991) report experimental evidence that subjects may indeed coordinate on inefficient alternatives even if the losses of miscoordination are the same across actions and if there is no historical precedent.

¹⁹This explains the form of the threshold \tilde{R} in (3.2): if the impulses of players from the other group are easy to predict (i.e., Q_{out} close to Q_{in}), then the threshold is high (i.e., $\tilde{R} \gg 1$), and players face considerable pressure to conform, even when interacting with the other group. When the impulses from the other group are difficult to anticipate (i.e., Q_{out} close to $\frac{1}{2}$), then the pressure is mitigated, and \tilde{R} is close to 1.

fer and Sutton (1999) present evidence that in many industries, organizations fail to adopt work practices that are commonly known to be superior. According to Pfeffer and Sutton, the critical difficulty for organizations lies in the implementation of efficient practices, not in identifying new solutions to complex problems. So, consistent with our approach, conformity and strategic complementarities can greatly affect performance.

The mechanism through which diversity improves performance in changeable environments does not rely on the premise that members of diverse teams have complementary skills or information, as is the focus of much of the existing literature on diversity (e.g., Hong and Page, 2001). Instead, diversity improves performance because it increases strategic uncertainty. A lack of congruent expectations makes it harder to anticipate the impulses of team members and so, discourages persistent practices. Eisenhardt (1989) shows that considering different perspectives can allow groups to avoid an escalation of commitment to any one option and to be more open to changing course if that leads to better outcomes. A number of papers have established a link between the diversity of teams and their willingness to consider alternative options and to deviate from established routines. Simons et al. (1999) show that diverse management teams are less inclined to focus on a single alternative and are more open to considering different courses of action. Phillips et al. (2006) present experimental evidence that diversity can stimulate team members to consider the issue from multiple perspectives, even if different groups have exactly the same information.²⁰

The introspective process plays a critical role in deriving this result. While a player’s impulse suggests a course of action, he need not follow his impulse in an introspective equilibrium (unlike in a correlated equilibrium or with sunspots). Schelling (1960, pp. 112–113) recognized that focal points can be inherently unstable, in which case they merely shape mutual expectations. The instability of focal points in diverse teams makes it possible for these teams to break out of Pareto-dominated Nash equilibria. This is hard to capture using existing approaches. For example, in global games, the risk dominant Nash equilibrium is always selected, independent of other factors. So, the global games approach cannot account for how diversity affects the ability of teams to adapt to changing circumstances.²¹

In effect, teams face two types of coordination problems. The first is basic: how to coordinate in the first place (on either alternative). The second is no less important: how to ensure coordination on the efficient Nash equilibrium? It is difficult to solve both coordination

²⁰There is also ample anecdotal evidence. President Franklin D. Roosevelt, for example, is known for assembling teams of clashing personalities, and for trying to avoid working with agencies with homogeneous, entrenched cultures (Greenstein, 1988, p. 28–30).

²¹Indeed, introspection does not always lead players to play the risk dominant Nash equilibrium. When identity affects the likelihood that players coordinate on the efficient equilibrium, choosing the team composition optimally becomes important.

problems simultaneously. If agents can accurately predict the impulses of others (as in homogeneous teams), then they can successfully coordinate on one of the alternatives. But, this comes at the risk of being stuck in the Pareto-dominated Nash equilibrium. On the other hand, if there is a lack of congruent expectations (as in diverse teams), then teams can break out of the Pareto-dominated Nash equilibrium and coordinate on the efficient outcome. However, it may also lead to miscoordination. This fundamental tradeoff is resolved differently in different economic environment, leading the manager to opt for diverse teams in changeable environments, and homogeneous teams in more stable ones. The next section studies the comparative statics of the likelihood that the efficient Nash equilibrium is played when the team composition is chosen optimally.

4. Implementation rate

Having characterized the optimal team composition, we can ask how behavior changes when cultural and economic factors vary. We focus on the likelihood that the efficient alternative is implemented. We take an *ex ante* perspective: for a given economic environment (i.e., distribution over payoffs), we ask how the probability that players coordinate on the efficient action in the introspective equilibrium varies when cultural and economic factors change.

For a given economic environment $f(v^*, z)$, the implementation rate $\mathcal{I}_f(Q_{in})$ is the probability that a randomly chosen pair of players coordinates on the Pareto-dominant Nash equilibrium (s^1, s^1) in the introspective equilibrium when the team composition is chosen optimally. That is, if we denote by $\text{Prob}_{\sigma^G, d}(s^1, s^1)$ the probability that two randomly matched players both choose action s^1 in the introspective equilibrium σ^G (given the game G) when the team composition is d^* , then the implementation rate is

$$\mathcal{I}_f(Q_{in}) := \mathbb{E}[\text{Prob}_{\sigma^G, d^*(Q_{in})}(s^1, s^1)],$$

where the expectation is taken over payoffs, as before. For simplicity, we focus on environments that satisfy somewhat stronger conditions than the ones in Section 3. The economic environment is *strongly stable* if it is δ -stable for some $\delta < 1$ and the risk-adjusted payoff ratio R lies between $1/\tilde{R}$ and \tilde{R} (with probability 1). The economic environment is *strongly changeable* if it is δ -changeable for some $\delta < 1$ and R is at least \tilde{R} (with probability 1). The results extend qualitatively to more general environments.

We first consider how the implementation rate varies when the economic environment changes. The first result shows that players are more likely to coordinate on the efficient outcome when its relative payoff increases. Say that the economic environment $f(v^*, z)$ strongly dominates the economic environment $f'(v^*, z)$ if the distribution F_R of the risk-adjusted payoff

ratio R induced by $f(v^*, z)$ strictly first-order stochastically dominates the distribution F'_R of R under $f'(v^*, z)$ (i.e., $F_R(r) < F'_R(r)$ for all r in the interior of the support of F_R and F'_R).

Proposition 4.1. [Implementation Rate Increases with Payoffs] *If the economic environments f and f' are both strongly stable or both strongly changeable and $f(v^*, z)$ dominates $f'(v^*, z)$, then $\mathcal{I}_f(Q_{in}) \geq \mathcal{I}_{f'}(Q_{in})$, with strict inequality when $f(v^*, z)$ and $f'(v^*, z)$ are strongly changeable.*

Intuitively, when the efficient outcome (s^1, s^1) becomes more attractive, the payoff structure of the game provides more guidance. In this case, the efficient outcome becomes a natural focal point for coordination. This leads to more coordination on the efficient action. This is consistent with empirical evidence. [Van Huyck et al. \(1991\)](#) show that subjects choose the efficient option less often when other concerns (such as history or risk) become more important. However, standard game-theoretic approaches seem to be unable to deliver this result.

We next turn to the effect of group identity. In stable environments, groups with a strong identity are more successful at coordinating on the efficient outcome.

Proposition 4.2. [Implementation Rate Increases with Identity if Stable] *In strongly stable environments, the implementation rate increases when group identity is strengthened. If the economic environment $f(v^*, z)$ is strongly stable and $Q_{in} > Q'_{in}$, then $\mathcal{I}_f(Q_{in}) > \mathcal{I}_f(Q'_{in})$.*

The intuition is straightforward: if a group has a stronger identity, players can accurately predict the impulses of the members of their own group. This allows them to coordinate more effectively, increasing the implementation rate.

However, the next result suggests that this intuition is incomplete: in changeable environments, groups with a strong identity are *less* successful at coordinating on the efficient outcome. Say that the environment is strongly* changeable (for Q_{in}) if it is strongly changeable and the density of the risk-adjusted payoff ratio R is weakly increasing in R for $R \leq \frac{Q_{in}}{1-Q_{in}}$.

Proposition 4.3. [Implementation Rate Decreases with Identity if Changeable] *In strongly* changeable environments, the implementation rate decreases when group identity is strengthened. If the economic environment $f(v^*, z)$ is strongly* changeable for Q_{in} , then the implementation rate strictly decreases in Q_{in} .*

So, strengthening group identity *reduces* the probability that the Pareto-dominant Nash equilibrium is played when the environment is changeable. This result may seem surprising at first sight. As noted above, it is easier for group members to anticipate each others' impulses when the group's identity is strong. However, there is also a counteracting effect. Because group members can more accurately predict each others' impulses when group identity is strong, they

have a strong incentive to follow their impulse. Players thus feel a stronger pressure to conform. In strongly* changeable environments, the latter effect dominates the former: an increase in the pressure to conform leads to a significant decrease in likelihood of coordinating on the efficient outcome.

On the other hand, if group identity is weak, players are less conformist and more open to selecting the efficient option. Proposition 4.3 may thus shed light on why more diverse and open-minded societies are more likely to abandon outdated practices than homogeneous ones with a strong culture (Mokyr, 1990). Likewise, organizations with an open culture are less likely to stick with inefficient practices (Pfeffer and Sutton, 1999).

Our approach allows us to consider the interplay of cultural and economic factors, in particular group identity and the payoff distribution. The results shed light on why groups with a weak identity may be more successful at implementing the efficient course of action in some environments, but not in others.²² Standard equilibrium analysis does not produce these results. Equilibrium selection typically does not depend on cultural factors. Existing models of learning and evolution in games are also unable to capture the interplay of cultural and economic factors highlighted here. Some models in the literature on learning and evolution predict that a particular Nash equilibrium (such as the risk-dominant Nash equilibrium) is selected in the limit (e.g., Kandori et al., 1993; Young, 1993). However, predictions do not vary with social or cultural factors such as diversity and group identity.²³

More broadly, players face two types of problems when trying to coordinate their actions, as noted earlier. First, players need to avoid miscoordination. Second, and no less important, they need to decide what to coordinate on. While it is desirable to implement the efficient option, this may become impossible once a group has developed routines. This leads to a fundamental tension. On the one hand, players can avoid miscoordination when there are strong group norms. On the other hand, while norms can reduce the risk of miscoordination, they may also make it difficult for players to switch to superior alternatives when these become available. That is, reducing the risk of miscoordination may come at the cost of increasing inertia and conformism. Capturing this fundamental tension requires a solution concept that allows miscoordination in some cases and the selection of Nash equilibria in other cases. The introspective equilibrium introduced here does exactly that: when the payoff structure gives little guidance, players may fail to coordinate; but if one option is far superior, players coordinate efficiently.

²²We could also do comparative statics on more traditional measures, such as team payoffs. Similar results obtain.

²³When players interact in a network, the structure of local interactions may determine whether the risk-dominant equilibrium spreads to the entire population when players are myopic (e.g., Blume, 1995; Morris, 2000); see Section 5.

5. Related literature

An emerging literature in economics studies the effect of identity and culture on economic outcomes. In the seminal work of [Akerlof and Kranton \(2000\)](#) and much of the subsequent literature, an agent’s identity enters his utility function; see, for instance, the work of [Akerlof and Kranton \(2005\)](#) on the role of identity in organizations. By contrast, in our work, a player’s identity affects his reasoning about others. [Kets and Sandroni \(2015a\)](#) use the approach developed here to show how identity shapes social interaction patterns. [Bisin and Verdier \(2001\)](#), [Kuran and Sandholm \(2008\)](#), and [Bisin et al. \(2015\)](#) develop models of cultural transmission of preferences and cultural integration. A number of authors have investigated the effects of diversity on economic outcomes such as public good provision both theoretically and empirically; see [Alesina and La Ferrara \(2005\)](#) for a survey.

The nature, design, and performance of organizations has been widely studied in economics; see [Gibbons and Roberts \(2013\)](#) for a recent overview. Within this literature, our work is most closely related to the papers that study the costs and benefits of diversity. Important contributions to this literature include [Lazear \(1999a,b\)](#), [Hong and Page \(2001\)](#), [Page \(2007\)](#), [Prat \(2002\)](#), and [Van den Steen \(2010\)](#). [Hong and Page \(2001\)](#), [Page \(2007\)](#), [Prat \(2002\)](#), and [Van den Steen \(2010\)](#) focus on the problem of identifying the optimal solution in complex environments, but do not consider the question how to implement the efficient outcome, as we do. These works thus abstract away from the strategic dimension and thus cannot explain why organizations may fail to adopt work practices that are commonly known to be superior, as is the case in many industries ([Pfeffer and Sutton, 1999](#)). [Lazear \(1999b\)](#) shows that diverse teams may outperform homogeneous teams if different groups have complementary skills. We show that diversity can have economic benefits even if this conditions is not met, by reducing conformity, thus uncovering a novel source of benefits of diversity in organizations.

The literature has also considered the effect of corporate culture on organizational performance; see [Kreps \(1990\)](#) and [Cr  mer \(1993\)](#) for seminal contributions, and see [Hermalin \(2013\)](#) for a survey. Rather than focusing on corporate culture, we consider the effect of cultural differences between the different groups that make up an organization. [Kreps \(1990\)](#) noted that cultural rules can act as focal principles in organizations and reduce coordination problems. We provide a formal mechanism through which identity and culture can aid in equilibrium selection, and use this to derive comparative statics.

[B  nabou \(2013\)](#) provides an economic analysis of groupthink, which may lead to unwarranted optimism about a new course of action. By contrast, we are interested in conformity to outdated practices and organizational inertia. [Bernheim \(1994\)](#) presents a model of conformity which is different from ours, and his paper focuses on different questions.

Focal points have been studied formally in a number of papers (e.g., Crawford and Haller, 1990; Sugden, 1995; Bacharach and Stahl, 2000; Janssen, 2001). This literature focuses on how players describe their options and under what conditions an alternative can become salient.

The process we consider bears some resemblance with level- k and cognitive hierarchy models which successfully predict behavior in a wide range of games (Crawford et al., 2013). In these models, players formulate a best response to the conjecture that other players are of lower level, and level-0 players are nonstrategic. A key difference is that we are interested in equilibrium behavior, while the level- k literature focuses on non-equilibrium outcomes. In addition, the level- k literature does not consider payoff-irrelevant signals such as impulses, which are critical in our setting.²⁴

Our model is also very different from global games (Carlsson and van Damme, 1993). There is no payoff uncertainty in our model, and, most importantly, the economic implications differ significantly. While a global-games approach always selects the risk-dominant equilibrium, this need not be the case in our setting even if there is a unique strict risk-dominant solution that is also payoff dominant.²⁵ This allows us to provide a novel rationale for diversity in organizations, and to provide novel comparative statics.

Equilibrium selection has also been studied in the literature on learning and evolution. Most of the models in this literature either select a particular Nash equilibrium (such as the risk-dominant one) or predict a distribution over Nash equilibria that is independent of social and cultural factors such as group identity and diversity (Mailath, 1998). When players interact only with their neighbors in a network, the evolution of play may depend on the network structure (e.g., Blume, 1995; Morris, 2000). The questions that these papers focus on are different from the ones that concern us here.

The methodological contribution of this paper is that it can account both for the failure to play according to a Nash equilibrium (like level- k models) in certain settings as well as players' ability to select a unique Nash equilibrium (like global games or learning models) in others. Whether players may fail to coordinate or select a unique Nash equilibrium depends in a natural way on economic incentives. The theory can do so without departing from the as-

²⁴The introspective process bears some resemblance to the tracing procedure (Harsanyi and Selten, 1988). This procedure involves an axiomatic determination of players' common priors and the construction of fictitious games. The fictitious games are used to select a Nash equilibrium of the original game. Our approach is inspired by evidence from psychology, both in its definition of the reasoning process and the initial beliefs, does not require the construction of auxiliary games, and gives different predictions. For example, the tracing procedure selects the risk-dominant equilibrium in coordination games, but this need not be the case for our process.

²⁵Grout et al. (2014) study the relative performance of homogeneous and diverse teams in a beauty contest game. The setting they consider is fundamentally different from ours. In particular, there is no Pareto superior option in their setting. The results are also different.

sumption that players are perfectly rational: in all cases, behavior is consistent with correlated equilibrium. However, while a game can have many correlated equilibria, the theory selects an (essentially) unique equilibrium, allowing us to characterize the optimal team composition in different environments and to provide clear and intuitive comparative statics.

6. Conclusions

This paper shows that there is a clear and compelling economic rationale for diversity even in the absence of skill complementarities or differences in information: diversity challenges conformity and stimulates the adoption of efficient practices. However, the very clash of perspectives that can make diverse teams successful may also enhance conflict and miscoordination. So, when the primary goal is to align behavior, diversity can be harmful. Consistent with empirical evidence, diversity improves performance in innovation-based industries, but not in others ([Jackson and Joshi, 2011](#)).

At the heart of our contribution is our model of introspection. Building on findings from psychology, we develop an explicit, and fully rational, model of introspection and reasoning that delivers a formal account of the effects of identity and diversity on economic outcomes. The model predicts non-Nash equilibrium outcomes in some games and Nash-equilibrium selection in other games, while returning a unique prediction in each of our games. The model provides intuitive comparative statics, and reveals a rich interplay between cultural and economic factors. The comparative statics are consistent with empirical evidence, but are difficult to produce using standard approaches.

While we focused on teams and organizations, our model can be used more broadly to study which societies are more likely to thrive and innovate ([Mokyr, 1990](#)), to design diversity policies ([Kets and Sandroni, 2015a](#)), and to identify optimal organizational cultures ([Hermalin, 2013](#)). Exploring the economic implications of culture and identity promises to be an exciting research agenda.

Appendix A Preliminary results

A.1 Team composition

Fix a team T . The majority group of team T is the group (say, A) such that at least half the players in the team belong to that group. We refer to the other group (e.g., B) as the minority group. If A is the majority group in team T_1 , then B is the majority group in T_2 and vice versa.

We can characterize the team composition in terms of the share of the team members that belongs to the majority group.

Lemma A.1. *Let $p_1, p_2 \geq \frac{1}{2}$ be the share of the majority group in teams T_1 and T_2 , respectively. Then, $p_1 = p_2 =: p$, and the team composition is $d = 2p - 1$.*

Proof. By symmetry, if a share $p_1 \geq \frac{1}{2}$ of the members of team T_1 belong to group A (say), then the share of A -players in team T_2 is $1 - p_1$. It follows that the share of B -players in team T_2 is $p_2 = 1 - (1 - p_1)$. In other words, if the majority in one team makes up a share p of the team, then the majority in the other team also makes up a share p of the team. We can thus write the team composition d as

$$\frac{1}{2} \cdot |p_1 - (1 - p_1)| + \frac{1}{2} \cdot |(1 - p_2) - p_2| = 2p - 1,$$

where the last line uses that $p_2 = p_1 =: p$. □

A.2 Impulses

It will be helpful to characterize the probability that two players have the same impulse. Recall that, conditional on $\theta_A = 1$, an A -player has an impulse to play action s^1 with probability $q \in (\frac{1}{2}, 1)$. Likewise, conditional on $\theta_A = 2$, an A -player has an impulse to play action s^2 with probability q . Analogous statements apply to B -players. Conditional on $\theta_A = m$, we have $\theta_B = m$ with probability $\lambda \in (\frac{1}{2}, 1)$, where $1, 2$. The following result characterizes the probability that two players have the same impulse.

Lemma A.2. *Let $q \in (\frac{1}{2}, 1)$ be the probability that a player of group A has the impulse to choose s^1 conditional on $\theta_A = 1$, and analogously for group B . Then:*

- (a) *the probability that two distinct A -players have the same impulse is $Q_{in} := q^2 + (1 - q)^2 \in (\frac{1}{2}, 1)$;*

- (b) the probability that two distinct A -players have the impulse to play s^1 is equal to $\frac{1}{2}Q_{in}$;
- (c) the conditional probability that an A -player j has the impulse to play action s^1 given that another A -player j' has the impulse to play action s^1 is equal to Q_{in} ;
- (d) the probability that an A -player and a B -player have the same impulse is $Q_{out} := \lambda \cdot Q_{in} + (1 - \lambda) \cdot (1 - Q_{in}) \in (\frac{1}{2}, Q_{in})$;
- (e) the probability that an A -player and a B -player have the impulse to play s^1 is equal to $\frac{1}{2}Q_{out}$;
- (f) the conditional probability that an A -player j has the impulse to play action s^1 given that a B -player j' has the impulse to play s^1 is equal to Q_{out} ;

Proof. We denote the probability measure over impulses and θ_A, θ_B by \mathbb{P} . For example, the probability that $\theta_A = 1$ is $\mathbb{P}(\theta_A = 1)$, and the conditional probability that an A -player j_A has an impulse to play action s^2 conditional on $\theta_A = 1$ is $\mathbb{P}(j_A = 2 \mid \theta_A = 1)$. Also, the probability that two A -players j_A and j'_A have an impulse to play action s^1 and s^2 , respectively, is denoted $\mathbb{P}(j_A = 1, j'_A = 2)$, the probability that j_A has an impulse to play s^1 given that j'_A has an impulse to play s^2 is $\mathbb{P}(j_A = 1 \mid j'_A = 2)$, and the probability that an A -player j_A and a B -player j_B have an impulse to play s^1 is $\mathbb{P}(j_A = 1, j_B = 1)$.

(a) Consider two A -players j_A and j'_A , $j_A \neq j'_A$. The probability that j_A and j'_A have the same impulse is

$$\mathbb{P}(j_A = 1, j'_A = 1) + \mathbb{P}(j_A = 2, j'_A = 2).$$

By symmetry, $\mathbb{P}(j_A = 1, j'_A = 1) = \mathbb{P}(j_A = 2, j'_A = 2)$, so it suffices to compute $\mathbb{P}(j_A = 1, j'_A = 1)$. We have

$$\begin{aligned}
\mathbb{P}(j_A = 1, j'_A = 1) &= \mathbb{P}(j_A = 1, j'_A = 1 \mid \theta_A = 1) \cdot \mathbb{P}(\theta_A = 1) + \\
&\quad \mathbb{P}(j_A = 1, j'_A = 1 \mid \theta_A = 2) \cdot \mathbb{P}(\theta_A = 2) \\
&= \mathbb{P}(j_A = 1 \mid \theta_A = 1) \cdot \mathbb{P}(j'_A = 1 \mid \theta_A = 1) \cdot \mathbb{P}(\theta_A = 1) + \\
&\quad \mathbb{P}(j_A = 1 \mid \theta_A = 2) \cdot \mathbb{P}(j'_A = 1 \mid \theta_A = 2) \cdot \mathbb{P}(\theta_A = 2) \\
&= \frac{1}{2} \cdot q^2 + \frac{1}{2} \cdot (1 - q)^2, \tag{A.1}
\end{aligned}$$

where the second line uses that impulses of A -players are conditionally independent given θ_A , and the last line follows by definition. The probability that two distinct A -players have the same impulse is thus

$$2 \cdot \left(\frac{1}{2} \cdot q^2 + \frac{1}{2} \cdot (1 - q)^2 \right) =: Q_{in}.$$

(b) By (A.1), the probability that two distinct A -players have an impulse to play action s^1 is $\frac{1}{2} \cdot q^2 + \frac{1}{2} \cdot (1 - q)^2 = \frac{1}{2} \cdot Q_{in}$.

(c) The conditional probability that an A -player j_A has the impulse to play action s^1 given that another A -player j'_A has the impulse to play action s^1 is

$$\begin{aligned} \mathbb{P}(j_A = 1 \mid j'_A = 1) &= \frac{\mathbb{P}(j_A = 1, j'_A = 1)}{\mathbb{P}(j'_A = 1)} \\ &= \frac{\frac{1}{2} \cdot q^2 + \frac{1}{2} \cdot (1 - q)^2}{\frac{1}{2}} \\ &= Q_{in} \end{aligned}$$

where we have used (A.1) again, and where we have used that the ex ante probability that a player j has an impulse to play action s^1 is $\frac{1}{2}$.

(d) Consider an A -player j_A and a B -player j_B . The probability that j_A and j_B have the same impulse is

$$\mathbb{P}(j_A = 1, j_B = 1) + \mathbb{P}(j_A = 2, j_B = 2).$$

As before, $\mathbb{P}(j_A = 1, j_B = 1) = \mathbb{P}(j_A = 2, j_B = 2)$, by symmetry. It thus suffices to compute $\mathbb{P}(j_A = 1, j_B = 1)$. We have

$$\begin{aligned} \mathbb{P}(j_A = 1, j_B = 1) &= \mathbb{P}(j_A = 1, j_B = 1 \mid \theta_A = 1, \theta_B = 1) \cdot \mathbb{P}(\theta_A = 1, \theta_B = 1) + \\ &\quad \mathbb{P}(j_A = 1, j_B = 1 \mid \theta_A = 2, \theta_B = 1) \cdot \mathbb{P}(\theta_A = 2, \theta_B = 1) + \\ &\quad \mathbb{P}(j_A = 1, j_B = 1 \mid \theta_A = 1, \theta_B = 2) \cdot \mathbb{P}(\theta_A = 1, \theta_B = 2) + \\ &\quad \mathbb{P}(j_A = 1, j_B = 1 \mid \theta_A = 2, \theta_B = 2) \cdot \mathbb{P}(\theta_A = 2, \theta_B = 2) \\ &= \frac{1}{2} \cdot [\lambda q^2 + 2(1 - \lambda) \cdot q \cdot (1 - q) + \lambda \cdot (1 - q)^2] \\ &= \frac{1}{2} \cdot [\lambda Q_{in} + (1 - \lambda) \cdot (1 - Q_{in})]. \end{aligned} \tag{A.2}$$

The probability that an A -player and a B -player have the same impulse is thus

$$2 \cdot \mathbb{P}(j_A = 1, j_B = 1) = \lambda Q_{in} + (1 - \lambda) \cdot (1 - Q_{in}) =: Q_{out}.$$

In the limit where λ approaches $\frac{1}{2}$ (i.e., θ_A and θ_B are uncorrelated), we have $Q_{out} = \frac{1}{2}$ (i.e., no correlation in impulses across groups). In the limit where λ approaches 1 (θ_A and θ_B are perfectly correlated), we have $Q_{out} = Q_{in}$ (i.e., correlation in impulses across groups equals the correlation in impulses within groups).

(e) This result is immediate from (A.2).

(f) The conditional probability that an A -player j_A has the impulse to play action s^1 given that a B -player j_B has the impulse to play s^1 is

$$\mathbb{P}(j_A = 1 \mid j_B = 1) = \frac{\mathbb{P}(j_A = 1, j_B = 1)}{j_B = 1} = Q_{out}.$$

□

A.3 Characterization introspective equilibrium

We characterize the introspective equilibria for symmetric (2×2) coordination games. Recall that payoffs are given by:

	s^1	s^2
s^1	v^*, v^*	z, y
s^2	y, z	v, v

By assumption, $v^* \geq v$, $v^* > y$, and $z < v$. So, the game has two strict Nash equilibria, viz., (s^1, s^1) and (s^2, s^2) , and the first is (weakly) payoff dominant. The ratio

$$R := \frac{v^* - y}{v - z}$$

is the ratio of risk-adjusted payoffs in the coordination game. The pure Nash equilibrium of the coordination game in which both players choose action s^1 is risk dominant in the sense of [Harsanyi and Selten \(1988\)](#) if $R \geq 1$, and the pure Nash equilibrium in which both players choose s^2 is risk dominant if $R \leq 1$. By Lemma [A.1](#), the team composition can be written as $d = 2p - 1$, where $p \geq \frac{1}{2}$ is the fraction of players in each team that belongs to the majority group, and $1 - p$ is the fraction of players that belongs to the minority group (see Lemma [A.1](#) for definitions).

The following result provides a complete characterization of introspective equilibria in symmetric (2×2) coordination games. Equilibrium behavior depends on the realization of the payoff parameters; there are different regions, separated by the following thresholds:

$$\begin{aligned} U_{in}(Q_{in}) &= \frac{Q_{in}}{1 - Q_{in}}; \\ U_{out}(Q_{in}) &= \frac{Q_{out}}{1 - Q_{out}}; \\ L_{out}(Q_{in}) &= \frac{1 - Q_{out}}{Q_{out}}; \\ L_{in}(Q_{in}) &= \frac{1 - Q_{in}}{Q_{in}}. \end{aligned}$$

Since $Q_{in} > Q_{out} > \frac{1}{2}$, we have $U_{in}(Q_{in}) > U_{out}(Q_{in}) > 1 > L_{out}(Q_{in}) > L_{in}(Q_{in})$.

Proposition A.3. [Introspective Equilibrium Symmetric Coordination Games] *Consider a symmetric coordination game and let $R = \frac{v^* - y}{v - z}$ be the risk-adjusted payoff ratio. Then:*

(a) *If*

$$R > U_{in}(Q_{in}),$$

then there is a unique introspective equilibrium. In this unique introspective equilibrium, all players choose action s^1 , regardless of their impulse or whom they are matched with;

(b) *If*

$$R \in (U_{out}(Q_{in}), U_{in}(Q_{in})),$$

then there is a unique introspective equilibrium. In this unique introspective equilibrium, players follow their impulse when they are matched with a member of their own group, and choose s^1 (regardless of their impulse) otherwise;

(c) *If*

$$R \in (L_{out}(Q_{in}), U_{out}(Q_{in})),$$

then there is a unique introspective equilibrium. In this unique introspective equilibrium, all players follow their impulse, regardless of whom they are matched with.

(d) *If*

$$R \in (L_{in}(Q_{in}), L_{out}(Q_{in})),$$

then there is a unique introspective equilibrium. In this unique introspective equilibrium, players follow their impulse when they are matched with a member of their own group, and choose s^2 (regardless of their impulse) otherwise;

(e) *If*

$$R < L_{in}(Q_{in})$$

then there is a unique introspective equilibrium. In this unique introspective equilibrium, all players choose action s^2 , regardless of their impulse or whom they are matched with.

(f) *If R is equal to $U_{in}(Q_{in})$, $U_{out}(Q_{in})$, $L_{out}(Q_{in})$, or $L_{in}(Q_{in})$, then at least some players are indifferent between playing a fixed action and following their impulses in some of their matches.*

Proof. By assumption, all players follow their impulse at level 0. For level $k > 0$, suppose that all players follow their impulse at any level $\ell < k$. Since group membership is observable, players can condition their action on which group their opponent belongs to. We thus consider two cases: the case where players are matched with members of their own group, and the case where players are matched with members of the other group.

Case 1: Within-group matchings. Consider an A -player who is matched with another A -player. If this player has an impulse to play action s^2 , then ignoring his impulse and choosing s^1 at level k is the unique best response if and only if

$$v^* \cdot (1 - Q_{in}) + z \cdot Q_{in} > v \cdot Q_{in} + y \cdot (1 - Q_{in})$$

or, equivalently, if and only if

$$R > \frac{Q_{in}}{1 - Q_{in}} =: U_{in}(Q_{in}),$$

where we have used Lemma A.2 for the conditional probability that players from the same group have an impulse to play action s given that player j has an impulse to play action s' . Conversely, if $R < U_{in}(Q_{in})$, the unique best response for an A -player at level k is to follow his impulse when he is matched with a player from his own group. When $R = U_{in}(Q_{in})$, the player is indifferent between following his impulse and choosing action s^1 . Next suppose that the player (who is matched with a player from his own group) has an impulse to play action s^1 . Ignoring his impulse and choosing action s^2 is the unique best response at level k if and only if

$$v \cdot (1 - Q_{in}) + y \cdot Q_{in} > v^* \cdot Q_{in} + z \cdot (1 - Q_{in}),$$

or, equivalently, if and only if

$$R < \frac{1 - Q_{in}}{Q_{in}} := L_{in}(Q_{in}).$$

Conversely, if $R > L_{in}(Q_{in})$, the unique best response for the player (who is matched with his own group and has an impulse to choose s^1) at level k is to follow his impulse and choose action s^1 . The same statements, of course, apply to B -players.

Case 2: Across-group matchings. Consider an A -player who is matched with a B -player and who has an impulse to play action s^2 . In this case, ignoring the impulse and playing action s^1 is the unique best response at level k if and only if

$$v^* \cdot (1 - Q_{out}) + z \cdot Q_{out} > v \cdot Q_{out} + y \cdot (1 - Q_{out})$$

or, equivalently, if and only if

$$R > \frac{Q_{out}}{1 - Q_{out}} =: U_{out}(Q_{in}),$$

where we have used Lemma A.2 for the conditional probability that players from different groups have an impulse to play action s given that player j has an impulse to play action s' .

Conversely, if $R < U_{out}(Q_{in})$, then it is a unique best response at level k for the player (who is matched with a player from the other group and has an impulse to play s^2) to follow his impulse. If $R = U_{out}(Q_{in})$, then the player is indifferent between following his impulse and playing action s^1 (i.e., ignoring his impulse). Next suppose the player has an impulse to play action s^1 . Ignoring his impulse and choosing action s^2 is the unique best response if and only if

$$R < \frac{1 - Q_{out}}{Q_{out}} := L_{out}(Q_{in}).$$

Conversely, if $R > L_{out}(Q_{in})$, then the unique best response at level k is to follow his impulse in this case. When $R = L_{out}(Q_{in})$, the player is indifferent between following his impulse and ignoring it. Analogous statements apply to B -players who are matched with a member of group A .

The above characterizes the best responses at level k . Next consider level $k + 1$, taking the best responses at level k as given. It is straightforward to check that the best responses at level k are also best responses at level $k + 1$. To see this, first suppose that $R > U_{in}(Q_{in})$. In this case, all players choose s^1 at level k , regardless of their impulse or whom they are matched with. Clearly, the unique best response at level $k + 1$ is to choose s^1 , independent of one's impulse, group, or group of the opponent. Next suppose $R \in (U_{out}(Q_{in}), U_{in}(Q_{in}))$. At level k , players ignore their impulse and play action s^1 when they are matched with a player of the other group, and follow their impulse when they are matched with the own group. At level $k + 1$, it is thus the unique best response for a player who is matched with a player of the other group to ignore his impulse and choose action s^1 . Also, for a player who is matched with the own group, following his impulse is the unique best response at level $k + 1$ (given that following one's impulse is the unique best response at level k against the conjecture that players of the own group follow their impulse). If $R \in (L_{out}(Q_{in}), U_{out}(Q_{in}))$, all players follow their impulse at level k , and this is the unique best response to the belief that all players follow their impulse; so, following one's impulse is the unique best response in this case. If $R \in (L_{in}(Q_{in}), L_{out}(Q_{in}))$, it is the unique best response for players at level k to choose s^2 when matched with a player of the other group and to follow their impulse otherwise; by a similar argument as above, this is the unique best response at level $k + 1$ also. Finally, if $R < L_{in}(Q_{in})$, then all players choose s^2 at level k ; clearly, the unique best response at level $k + 1$ is to choose action s^2 . One can repeat this argument to show that the best responses at level k are also best responses for any level $\ell \geq k$. So, if R is not equal to one of the thresholds (i.e., $R \neq L_{in}(Q_{in}), L_{out}(Q_{in}), U_{out}(Q_{in}), U_{in}(Q_{in})$), we have a unique introspective equilibrium. If R falls on one of the thresholds (i.e., $R = L_{in}(Q_{in}), L_{out}(Q_{in}), U_{out}(Q_{in}), U_{in}(Q_{in})$), there are multiple introspective equilibria where players that are indifferent between following their

impulse and playing a fixed action mix between their actions. \square

We next show that the introspective equilibrium characterized in Proposition A.3 is a correlated equilibrium. We take an interim perspective where we condition on a match. Let $i_C, i_{C'}$ be a randomly selected pair of (distinct) players, with i_C and $i_{C'}$ belonging to group C and C' , respectively. Define a probability distribution $\{\rho_{s,\tilde{s}}\}_{s,\tilde{s}=s^1,s^2}$ over action profiles for i_C and $i_{C'}$: $\rho_{s,\tilde{s}} \geq 0$ for $s, \tilde{s} = s^1, s^2$, and $\rho_{s^1,s^1} + \rho_{s^1,s^2} + \rho_{s^2,s^1} + \rho_{s^2,s^2} = 1$. In an introspective equilibrium σ^G , players i_C and $i_{C'}$ follow the strategies σ_{i_C} and $\sigma_{i_{C'}}$, respectively. For simplicity, we write $\sigma_{i_C}^G(\theta_{i_C})$ for the strategy $\sigma_{i_C}^G(\theta_{i_C}, i_{C'})$ of player i_C when his impulse is θ_{i_C} and he is matched with $i_{C'}$, and likewise for the strategy of $i_{C'}$. Following Aumann (1987), the strategy profile $(\sigma_{i_C}^G, \sigma_{i_{C'}}^G)$ is a correlated equilibrium if it generates a probability distribution $\{\rho_{s,\tilde{s}}^G\}_{s,\tilde{s}=s^1,s^2}$ over action profiles that satisfies the following conditions:

$$\begin{aligned} \rho_{s^1,s^1}^G \pi_{i_C}(s^1, s^1) + \rho_{s^1,s^2}^G \pi_{i_C}(s^1, s^2) &\geq \rho_{s^1,s^1}^G \pi_{i_C}(s^2, s^1) + \rho_{s^1,s^2}^G \pi_{i_C}(s^2, s^2); \\ \rho_{s^2,s^1}^G \pi_{i_C}(s^2, s^1) + \rho_{s^2,s^2}^G \pi_{i_C}(s^2, s^2) &\geq \rho_{s^2,s^1}^G \pi_{i_C}(s^1, s^1) + \rho_{s^2,s^2}^G \pi_{i_C}(s^1, s^2); \\ \rho_{s^1,s^1}^G \pi_{i_{C'}}(s^1, s^1) + \rho_{s^2,s^1}^G \pi_{i_{C'}}(s^1, s^2) &\geq \rho_{s^1,s^1}^G \pi_{i_{C'}}(s^2, s^1) + \rho_{s^2,s^1}^G \pi_{i_{C'}}(s^2, s^2); \\ \rho_{s^1,s^2}^G \pi_{i_{C'}}(s^2, s^1) + \rho_{s^2,s^2}^G \pi_{i_{C'}}(s^2, s^2) &\geq \rho_{s^1,s^2}^G \pi_{i_{C'}}(s^1, s^1) + \rho_{s^2,s^2}^G \pi_{i_{C'}}(s^1, s^2); \end{aligned} \tag{A.3}$$

where we recall that $\rho_{s,\tilde{s}}$ is the probability that i_C chooses s and $i_{C'}$ chooses \tilde{s} .²⁶

Proposition A.4. [Any introspective equilibrium is a correlated equilibrium] *If σ^G is an introspective equilibrium, then for any pair of (distinct) players $i_C, i_{C'}$ (belonging to group C and C' , respectively), the strategy profile $(\sigma_{i_C}^G, \sigma_{i_{C'}}^G)$ generates a probability distribution $\{\rho_{s,\tilde{s}}^G\}_{s,\tilde{s}=s^1,s^2}$ over action profiles that satisfies the conditions in (A.3).*

Proof. We check the various cases in Proposition A.4. First suppose that $R \geq \frac{Q_{in}}{1-Q_{in}}$. By Proposition A.4, there is an introspective equilibrium in which all players choose s^1 , regardless of their impulse (and if R is strictly greater than $\frac{Q_{in}}{1-Q_{in}}$, then this is the unique introspective equilibrium). So, $\rho_{s^1,s^1}^G = 1$, and (A.3) reduces to $v^* \geq y$, and this holds given that (s^1, s^1) is a Nash equilibrium.

Next suppose that $R \in [\frac{Q_{out}}{1-Q_{out}}, \frac{Q_{in}}{1-Q_{in}}]$. By Proposition A.4, there is an introspective equilibrium in which players follow their impulse when matched with the own group, and play s^1 otherwise (and this is the unique introspective equilibrium in the interior of the interval). First suppose that $C = C'$, so that players i_C and $i_{C'}$ belong to the same group. Then, by Lemma

²⁶Alternatively, we could write the equilibrium conditions in terms of conditional probabilities; this would give the same results. The present approach is the one taken by Aumann (1987) and has the advantage we do not have to be concerned about conditioning on probability-0 events.

A.2, $\rho_{s^1, s^1}^G = \frac{1}{2}Q_{in}$, $\rho_{s^1, s^2}^G = \rho_{s^2, s^1}^G = \frac{1}{2} \cdot (1 - Q_{in})$, and $\rho_{s^2, s^2}^G = \frac{1}{2}Q_{in}$, and (A.3) becomes

$$\begin{aligned} \frac{1}{2}Q_{in} \cdot v^* + \frac{1}{2} \cdot (1 - Q_{in}) \cdot z &\geq \frac{1}{2}Q_{in} \cdot y + \frac{1}{2} \cdot (1 - Q_{in}) \cdot v; \\ \frac{1}{2}Q_{in} \cdot v + \frac{1}{2} \cdot (1 - Q_{in}) \cdot y &\geq \frac{1}{2}Q_{in} \cdot z + \frac{1}{2} \cdot (1 - Q_{in}) \cdot v^*. \end{aligned} \quad (\text{A.4})$$

These inequalities are satisfied for $R \in [\frac{Q_{out}}{1-Q_{out}}, \frac{Q_{in}}{1-Q_{in}}]$. Next suppose that i_C and $i_{C'}$ belong to different groups, that is, $C' \neq C$. Then, $\rho_{s^1, s^1}^G = 1$, and (A.3) again reduces to $v^* \geq y$.

The next case we consider is $R \in [\frac{1-Q_{out}}{Q_{out}}, \frac{Q_{out}}{1-Q_{out}}]$. By Proposition A.4, there is an introspective equilibrium in which all players follow their impulse, regardless of whom they are matched with (and this is the unique introspective equilibrium in the interior of the interval). If both players belong to the same group (i.e., $C = C'$), then (A.3) again reduces to (A.4), and these inequalities are satisfied for this range of R . So suppose the players belong to different groups (i.e., $C \neq C'$). Then, (A.3) becomes

$$\begin{aligned} \frac{1}{2}Q_{out} \cdot v^* + \frac{1}{2} \cdot (1 - Q_{out}) \cdot z &\geq \frac{1}{2}Q_{out} \cdot y + \frac{1}{2} \cdot (1 - Q_{out}) \cdot v; \\ \frac{1}{2}Q_{out} \cdot v + \frac{1}{2} \cdot (1 - Q_{out}) \cdot y &\geq \frac{1}{2}Q_{out} \cdot z + \frac{1}{2} \cdot (1 - Q_{out}) \cdot v^*. \end{aligned}$$

where we have again used Lemma A.2. Again, these inequalities are satisfied for $R \in [\frac{1-Q_{out}}{Q_{out}}, \frac{Q_{out}}{1-Q_{out}}]$.

Next suppose $R \in [\frac{1-Q_{in}}{Q_{in}}, \frac{1-Q_{out}}{Q_{out}}]$. By Proposition A.4, there is an introspective equilibrium in which players follow their impulse when matched with the own group, and play s^2 otherwise (and this is the unique introspective equilibrium in the interior of the interval). If i_C and $i_{C'}$ belong to different groups (i.e., $C' \neq C$), then $\rho_{s^2, s^2}^G = 1$ and (A.3) reduces to $v \geq z$. This holds because (s^2, s^2) is a Nash equilibrium. So suppose that i_C and $i_{C'}$ belong to the same group. Then (A.3) again reduces to (A.4), and these inequalities are satisfied for this range for R .

Finally, suppose $R \leq \frac{1-Q_{in}}{Q_{in}}$. By Proposition A.3, there is an introspective equilibrium in which all players choose s^2 , regardless of their impulse or whom they are matched with (and if the inequality is strict, this introspective equilibrium is unique). Then, (A.3) reduces to $v \geq z$. Again, this inequality is satisfied given that (s^2, s^2) is a correlated equilibrium.

As noted in Proposition A.3(f), some players are indifferent among their actions when R takes on “threshold” values. In those cases, some players are indifferent among actions, and there may be introspective equilibria in which different players from a given group take different actions even if they are matched with the same group and have the same impulse. For example, some players from group C''' follow their impulse when matched to a member of group C''' while other players from group C''' choose a fixed action (e.g., s^1) regardless of their impulse when matched with a member of group C''' . In all cases, players’ behavior satisfies (A.3). \square

So, every introspective equilibrium is a correlated equilibrium. It follows from the results in [Aumann \(1987\)](#) that players' behavior in a correlated equilibrium is consistent with rationality and common knowledge of rationality. However, a game may have many correlated equilibrium, while the introspective equilibrium is essentially unique.

A few aspects of the characterization in Proposition [A.3](#) are worth noting.

First, while players sometimes coordinate on the risk dominant Nash equilibrium, it is not always played. For example, if $R \geq 1$, (s^1, s^1) is a risk dominant Nash equilibrium, but players only coordinate on s^1 (with probability 1) if $R > U_{in}(Q_{in}) > 1$. By contrast, if $R \in [1, U_{out}(Q_{in}))$, players follow their impulse in all their matches and may potentially coordinate on action s^2 or miscoordinate. In the intermediate range (i.e., $R \in (U_{out}(Q_{in}), U_{in}(Q_{in}))$), players choose s^1 (regardless of their impulse) when they are matched with the other group, but follow their impulse otherwise.

Second, players may ignore their impulse in the introspective equilibrium. In this case, players do not follow the “action recommendation” from the impulse, unlike in correlated equilibrium.

Third, the introspective process ends after a few rounds in this class of games. So, even if players are bounded in their reasoning about others, they will play according to the introspective equilibrium. In other games, players adjust their action at each level; see [Kets and Sandroni \(2015a\)](#) and [Kets and Sandroni \(2015b\)](#) for examples.

Fourth, the result is robust to relaxing the assumption that group membership is perfectly observable: it largely goes through if players can observe their opponent's group only imperfectly. Even if players cannot observe their opponent's group at all, the introspective equilibrium is very similar. It is again (essentially) unique; if R is sufficiently high (low), all players choose s^1 (respectively, s^2); and for intermediate values of R , some players follow their impulse, while others choose a fixed action (s^1 or s^2 depending on whether R is high or low).

Finally, the introspective equilibrium is essentially unique: it is unique for almost all values of the parameters, that is, the parameters for which the introspective equilibrium is unique has measure 1 under the Lebesgue measure or, in fact, under any payoff distribution that admits a density (i.e., is absolutely continuous with respect to Lebesgue measure). It will be useful to adopt a tie-breaking rule to handle the case where players are indifferent, as discussed in the next remark.

Remark 1. For the remainder of the paper, we will assume that if players are indifferent between following their impulse and choosing a fixed action s (regardless of their impulse) when matched with members of a certain group, they will follow their impulse. The choice of tie-breaking rule does not affect our results in any way. This is because the values of the payoff parameters for which (some subset of) players are indifferent has measure 0 in the economic

environments that we consider. So, the manager would choose the same team composition regardless of the tie-breaking rule that is used.

Fixing a tie-breaking rule allows us to simplify notation: we can fix a particular introspective equilibrium for values of the parameters where players are indifferent, instead of working with equivalence classes of introspective equilibria (which differ only on a set of parameters of measure zero). Henceforth, when we calculate expected payoff (for given majority share p and group identity Q_{in}), we will do so assuming that players play according to the introspective equilibrium characterized in Proposition A.3 whenever no player is indifferent, and follow the above tie-breaking rule otherwise; we will denote this introspective equilibrium by σ^G . \triangleleft

A.4 Equilibrium payoffs

We characterize the total payoff in the (essentially) unique introspective equilibrium. Fix the majority share p and the strength Q_{in} of players' group identity (where the majority share p is defined in Lemma A.1). By Proposition A.3 (and using the tie-breaking rule discussed in Remark 1), we have the following:

1. If $R > U_{in}(Q_{in})$, then in the introspective equilibrium, all players choose action s^1 , regardless of their impulse;
2. If $U_{out} < R \leq U_{in}(Q_{in})$, then in the introspective equilibrium, players choose action s^1 , regardless of their impulse, when matched with a player of the other group and follow their impulse otherwise;
3. If $L_{out}(Q_{in}) \leq R \leq U_{out}(Q_{in})$, then in the introspective equilibrium, all players follow their impulse;
4. If $L_{in}(Q_{in}) \leq R < L_{out}(Q_{in})$, then in the introspective equilibrium, players choose action s^2 , regardless of their impulse, when matched with a player of the other group and follow their impulse otherwise;
5. For any $R < L_{in}(Q_{in})$, then in the introspective equilibrium, all players choose action s^2 , regardless of their impulse.

We calculate the team payoff in the introspective equilibrium for each of these regions. As before, let $p \geq \frac{1}{2}$ be the share of the majority group in the team. The game G can be summarized by the payoff parameters v^* and z , so we will write $G = (v^*, z)$ here.

It will be useful to calculate the team payoff in the introspective equilibrium (i.e., the sum of payoffs accruing to members of a given team), as opposed to the total payoff (i.e., the sum of

payoffs accruing to all players). By symmetry, the total payoff in the introspective equilibrium is simply twice the payoff of an individual team (in equilibrium).

We need to consider different ranges of parameters. First, suppose that the payoff parameters $G = (v^*, z)$ and group identity Q_{in} are such that $R > U_{in}(p, Q_{in})$. In that case, all players choose s^1 regardless of their impulse. Then, a team's payoff in the introspective equilibrium σ^G is

$$\gamma_1(p, Q_{in}, v^*, z) := v^*.$$

Next, suppose that $U_{out}(p, Q_{in}) < R \leq U_{in}(p, Q_{in})$ when the game is $G = (v^*, z)$. Then, in the introspective equilibrium σ^G , players choose action s^1 , regardless of their impulse, when matched with a player of the other group and follow their impulse otherwise. So, a team's payoff in the introspective equilibrium σ^G is

$$\begin{aligned} \gamma_2(p, Q_{in}, v^*, z) := & p \cdot \left[p \cdot \left(\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}) \right) + (1 - p) \cdot v^* \right] + \\ & (1 - p) \cdot \left[(1 - p) \cdot \left(\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}) \right) + p \cdot v^* \right], \end{aligned}$$

where we have used Lemma A.2 for the probability that two distinct players have an impulse to a given action. Again, the team payoff is the sum of the payoffs accruing to the two groups. When players interact with members of their own group, their expected payoff is $\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in})$; if they interact with members of the other group, it is v^* .

Suppose $L_{min}(p, Q_{in}) \leq R \leq U_{out}(p, Q_{in})$ when the game is $G = (v^*, z)$, so that all players follow their signal in the introspective equilibrium. The team payoff in the introspective equilibrium σ^G is then

$$\begin{aligned} \gamma_3(p, Q_{in}, v^*, z) := & \frac{1}{2} \cdot \left[(p^2 + (1 - p)^2) \cdot Q_{in} + 2p \cdot (1 - p) \cdot Q_{out} \right] \cdot (v^* + v) + \\ & \frac{1}{2} \cdot \left[(p^2 + (1 - p)^2) \cdot (1 - Q_{in}) + 2p \cdot (1 - p) \cdot (1 - Q_{out}) \right] \cdot (y + z), \end{aligned}$$

where we have again used Lemma A.2.

If $L_{maj}(p, Q_{in}) \leq R < L_{min}(p, Q_{in})$ when the game is $G = (v^*, z)$, then in the introspective equilibrium σ^G , players choose action s^2 when matched with the other group, and follow their impulse otherwise. The team payoff in this introspective equilibrium is

$$\begin{aligned} \gamma_4(p, Q_{in}, v^*, z) := & p \cdot \left[p \cdot \left(\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}) \right) + (1 - p) \cdot v \right] + \\ & (1 - p) \cdot \left[(1 - p) \cdot \left(\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}) \right) + p \cdot v \right] \end{aligned}$$

Finally, suppose that $R < L_{maj}(p, Q_{in})$ when the game is $G = (v^*, z)$. In that case, all players choose action s^2 , and the team payoff in the introspective equilibrium σ^G is

$$\gamma_5(p, Q_{in}, v^*, z) := v.$$

The team payoff is thus given by

$$\mathcal{C}(p, Q_{in}, v^*, z) := \begin{cases} \gamma_1(p, Q_{in}, v^*, z) & \text{if } R > U_{in}(Q_{in}); \\ \gamma_2(p, Q_{in}, v^*, z) & \text{if } U_{out}(Q_{in}) < R \leq U_{in}(Q_{in}); \\ \gamma_3(p, Q_{in}, v^*, z) & \text{if } L_{out}(Q_{in}) \leq R \leq U_{out}(Q_{in}); \\ \gamma_4(p, Q_{in}, v^*, z) & \text{if } L_{in}(Q_{in}) \leq R < L_{out}(Q_{in}); \\ \gamma_5(p, Q_{in}, v^*, z) & \text{if } R < L_{in}(Q_{in}). \end{cases}$$

As there are two teams, the total payoff in the introspective equilibrium is $C^G(\sigma^G; d) = 2\mathcal{C}(p, Q_{in}, v^*, z)$ when the game is $G = (v^*, z)$, where we recall that the team composition d is equal to $2p - 1$ when the share of the majority is $p \geq \frac{1}{2}$ (Lemma A.1). While the functions $\gamma_1, \gamma_2, \gamma_3, \gamma_4$, and γ_5 are all continuous in the payoff parameters v^*, z , the share p of the majority, and group identity Q_{in} , the function $\mathcal{C}(p, Q_{in}, v^*, z)$ is generally discontinuous at the thresholds $R = U_{in}(p, Q_{in}), U_{out}(p, Q_{in}), L_{min}(p, Q_{in}), L_{maj}(p, Q_{in})$.

Appendix B Proofs

B.1 Proof of Lemma 2.2

Fix a game G . By Proposition A.3 in Appendix A, for generic values of the parameters, there is a unique introspective equilibrium $\sigma^G = (\sigma_j^G)_j$. In this introspective equilibrium, the strategy σ_j^G of a player depends only on his own group and the group of the player he is matched with. That is, if players j and k belong to the same group C and are matched with players j' and k' , respectively, who both belong to group C' , then $\sigma_j^G = \sigma_k^G$.²⁷ So, for each pair of groups $C, C' = A, B$ and a given team $T = T_1, T_2$, there is a strategy $\sigma_{C, C', T}^G$ that maps impulses into actions such that for each C -player j in team T with impulse i_j that is matched with a C' -player c' , we have $\sigma_j^G(i_j, c') = \sigma_{C, C', T}^G(i_j)$. Since the payoff distribution has a well-defined density, the values of the payoff parameters for which the introspective equilibrium is not unique have zero probability for a given Q_{in} . So, the expected total payoff in the introspective equilibrium does not depend on the individual identity of players (i.e., who is matched with whom), only on the shares of players of each group assigned to different teams, that is, on the team composition (Lemma A.1). \square .

²⁷Of course, depending on the realization of their impulses, players j and k may end up taking different actions even if they belong to the same group and are both matched to the same group (and neither is indifferent between his actions): if player j 's impulse i_j is different from player k 's impulse i_k , then we could have $\sigma_j^G(i_j, k) = s$ and $\sigma_k^G(i_k, j) = s'$ for $s, s' = s^1, s^2$ and $s \neq s'$.

B.2 Proof of Theorem 3.1

We use the results and notation from Appendix A. It will be convenient to work with random variables rather than with densities. For the proofs, we make the distinction between a random variable and its realization: we write \mathbf{v}^* for the random variable that gives the coordination payoff for s^1 , and v^* for its realization. Likewise, we write \mathbf{z} for the random variable that gives the payoff to a player who chooses s^1 when his opponent chooses s^2 , and z for its realization. So, $(\mathbf{v}^*, \mathbf{z})$ is a random vector (taking values in \mathbb{R}^2), and $f(v^*, z)$ is its joint density. The joint density $f(v^*, z)$ corresponds to a distribution function $F(v^*, z)$ (i.e., $F(v^*, z)$ is the probability that \mathbf{v}^* is at most v^* and \mathbf{z} is at most z). We also define the random variable \mathbf{R} by

$$\mathbf{R} := \frac{\mathbf{v}^* - y}{v - \mathbf{z}}.$$

The collection of δ -stable environments thus corresponds to a set of random variables. For $\delta < 1$, let $\Pi^{S,\delta}$ be the set of random vectors $(\mathbf{v}^*, \mathbf{z})$ (taking values in \mathbb{R}^2) that satisfy the moment conditions (3.1) as well as the following conditions:

- S_δ -I the probability that \mathbf{v}^* and \mathbf{z} are in $U_{1-\delta}^{\mathbf{R}=1}$ is greater than δ ;
- S_δ -II the payoffs have a well-defined joint density;
- S_δ -III the event that $\mathbf{v}^* \geq v$, $\mathbf{v}^* > y$, and $v > \mathbf{z}$ has probability 1.

With some abuse of terminology we will refer to the elements of $\Pi^{S,\delta}$ as δ -stable economic environments.

Let $\Pi^{R=1}$ be the set of random vectors (taking values in \mathbb{R}^2) that (1) are such that $\mathbf{R} = 1$ with probability 1; and that (2) satisfy the moment condition (3.1) and condition S_δ -III. Note that the elements of $\Pi^{R=1}$ do not represent an economic environment, since they do not have a well-defined density (the event that $\mathbf{R} = 1$ is an event that has Lebesgue measure zero). However, it will be a useful benchmark. In this benchmark case, it is optimal to have maximally homogeneous teams.

Lemma B.1. *Fix Q_{in} . For any random vector in $\Pi^{R=1}$, the unique optimal team composition is $d^* = 1$.*

Proof. Fix Q_{in} and $(\mathbf{v}^*, \mathbf{z}) \in \Pi^{R=1}$. Since $\mathbf{R} = 1$ with probability 1, the expected team payoff is $\mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})] = \mathbb{E}[\gamma_3(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]$. By (3.1), we can “pass to the limit,” that is, we have

$$\frac{d\mathbb{E}[\gamma_3(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]}{dp} = \mathbb{E}\left[\frac{d\gamma_3(p, Q_{in}, \mathbf{v}^*, \mathbf{z})}{dp}\right].$$

Since

$$\frac{d\gamma_3(p, Q_{in}, v^*, z)}{dp} = (2p - 1) \cdot (Q_{in} - Q_{out}) \cdot (v^* + v - y - z),$$

we have $\frac{d\gamma_3(p, Q_{in}, v^*, z)}{dp} \geq 0$ for any realization (v^*, z) of $(\mathbf{v}^*, \mathbf{z})$, with strict inequality if $p > \frac{1}{2}$ (given that $p \geq \frac{1}{2}$ and given S_δ -III). Hence, the expected team payoff is maximized at $p^* = 1$. By Lemma A.1, the unique optimal team composition in this environment is $d^* = 2p^* - 1 = 1$. \square

If the environment is δ -stable, then it is close to the case where $\mathbf{R} = 1$ (with probability 1). The next step is to show that the expected team payoff is continuous. Define Π to be the set of random variables that take values in \mathbb{R}^2 and that satisfy conditions (3.1), S_δ -II and S_δ -III, and let $\Pi^S := \Pi \cup \Pi^{R=1}$. Clearly, $\Pi^{S,\delta} \subset \Pi^S$ for every $\delta < 1$.

The law $\mu_{(\mathbf{v}^*, \mathbf{z})}$ of a random vector $(\mathbf{v}^*, \mathbf{z})$ is the distribution of $(\mathbf{v}^*, \mathbf{z})$: for every measurable subset A of \mathbb{R}^2 , $\mu_{(\mathbf{v}^*, \mathbf{z})}(A)$ is the probability that $(\mathbf{v}^*, \mathbf{z}) \in A$. We identify the collection Π^S with the subset Δ^S of the set of Borel probability measures on \mathbb{R}^2 such that $\mu \in \Delta^S$ if and only if μ is the law of some $(\mathbf{v}^*, \mathbf{z}) \in \Pi^S$; we can likewise define the set $\Delta^{R=1}$ as the set of laws of random vectors in $\Pi^{R=1}$. We endow Δ^S with the relative topology (with the set of Borel probability measures endowed with its usual weak convergence topology). A sequence $\{(\mathbf{v}_n^*, \mathbf{z}_n)\}_n$ of random vectors in Π^S converges to a random vector $(\mathbf{v}^*, \mathbf{z})$ if the laws of $\{(\mathbf{v}_n^*, \mathbf{z}_n)\}_n$ converge to the law of $(\mathbf{v}^*, \mathbf{z})$ in the weak topology.

Define the function $g^S : [\frac{1}{2}, 1] \times (\frac{1}{2}, 1) \times \Pi^S \rightarrow \mathbb{R}$ by $g^S(p, Q_{in}, \mathbf{v}^*, \mathbf{z}) := \mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]$. This function is continuous:

Lemma B.2. *For every Q_{in} , the function $g^S(\cdot, Q_{in}, \cdot, \cdot)$ is jointly continuous in its arguments.*

Proof. Fix $(\mathbf{v}^*, \mathbf{z}) \in \Pi^S$. Since \mathbf{v}^* and \mathbf{z} are random variables, the team payoff $\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$ is also a random variable (for any p and Q_{in}). Denote by M the collection of random variables $\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$ for which $(\mathbf{v}^*, \mathbf{z}) \in \Pi^S$. Then, since the elements of Π^S satisfy the moment conditions (3.1), the set M is uniformly integrable (Billingsley, 1968).

Consider a sequence $\{(p_n, \mathbf{v}_n^*, \mathbf{z}_n)\}_{n=1,2,\dots}$, where for every n , $p_n \in [\frac{1}{2}, 1]$ and $(\mathbf{v}_n^*, \mathbf{z}_n) \in \Pi^S$. Suppose the sequence $\{(p_n, \mathbf{v}_n^*, \mathbf{z}_n)\}_{n=1,2,\dots}$ converges to $(p, \mathbf{v}^*, \mathbf{z})$ for $p \in [\frac{1}{2}, 1]$ and $(\mathbf{v}^*, \mathbf{z}) \in \Pi^S$. We claim that the sequence $\{\mathcal{C}(p_n, \mathbf{v}_n^*, \mathbf{z}_n)\}_{n=1,2,\dots}$ of random variables converges weakly to the random variable $\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$. The first step is to show the set of discontinuities of the measurable function $\mathcal{C}(\cdot, \cdot, \mathbf{v}^*, \mathbf{z})$ has measure 0 under the distribution whenever $(\mathbf{v}^*, \mathbf{z}) \in \Pi^S$. To see this, suppose that $(\mathbf{v}^*, \mathbf{z}) \in \Pi \subset \Pi^S$. Then, the distribution of $(\mathbf{v}^*, \mathbf{z})$ has a well-defined density (condition S_δ -II) and thus assigns zero probability to the event that $\mathbf{R} = U_{in}(Q_{in}), U_{out}(Q_{in}), L_{out}(Q_{in}), L_{in}(Q_{in})$, which are the potential points of discontinuity of \mathcal{C} (Appendix A.4). Alternatively, if $(\mathbf{v}^*, \mathbf{z}) \in \Pi^{R=1} \subset \Pi^S$, then the function \mathcal{C} is equal to the

function γ_3 with probability 1. The result then follows because γ_3 does not have any discontinuities. So, the set of discontinuities of $\mathcal{C}(\cdot, \cdot, \mathbf{v}^*, \mathbf{z})$ has zero measure. We can then apply the mapping theorem (Billingsley, 1968, Thm. 5.1) to show that the sequence $\{\mathcal{C}(p_n, \mathbf{v}_n^*, \mathbf{z}_n)\}_{n=1,2,\dots}$ of random variables converges weakly to the random variable $\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$. By uniform integrability, it follows that $\mathbb{E}[\mathcal{C}(p_n, Q_{in}, \mathbf{v}_n^*, \mathbf{z}_n)]$ converges to $\mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]$ (Billingsley, 1968, Thm 5.4). That is, $g^S(p_n, Q_{in}, \mathbf{v}_n^*, \mathbf{z}_n) \rightarrow g^S(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$. \square

Since the team payoff \mathcal{C} is neither continuous in its arguments nor bounded, some work is required that the expected team payoff is continuous. The argument in the proof of Lemma B.2 has two main steps. First, the assumption that the payoff parameters have a continuous distribution (condition \mathcal{S}_δ -II) implies that the set of discontinuities of \mathcal{C} has measure 0, and this ensures that the measurable function \mathcal{C} converges weakly when its arguments converge in the appropriate sense. Second, the moment restrictions (3.1) ensure that the expectation of \mathcal{C} converges when the measurable function \mathcal{C} converges. So, even if the function \mathcal{C} is not continuous or bounded, its expectation is continuous and finite.²⁸

We can now apply standard tools to show that when we are close to the benchmark case where $\mathbf{R} = 1$ with probability 1, then the optimal team composition will be close to homogeneous. Define the “value function” $v_{Q_{in}}^S : \Delta^S \rightarrow \mathbb{R}$ for Q_{in} by

$$v_{Q_{in}}^S(\mu(\mathbf{v}^*, \mathbf{z})) := \max_{\tilde{p} \in [\frac{1}{2}, 1]} g^S(\tilde{p}, Q_{in}, \mathbf{v}^*, \mathbf{z}).$$

Define the associated correspondence of maximizers by

$$P_{Q_{in}}^*(\mu(\mathbf{v}^*, \mathbf{z})) := \{p \in [\frac{1}{2}, 1] : g^S(p, Q_{in}, \mathbf{v}^*, \mathbf{z}) = v_{Q_{in}}^S(\mu(\mathbf{v}^*, \mathbf{z}))\}.$$

That is, $v_{Q_{in}}^S(p, \mu(\mathbf{v}^*, \mathbf{z}))$ is the expected team payoff when the majority share p is chosen to maximize the team payoff for given Q_{in} and $(\mathbf{v}^*, \mathbf{z})$, and the set $P_{Q_{in}}^*(\mu(\mathbf{v}^*, \mathbf{z}))$ (possibly empty) contains the majority shares at which this maximum is attained. For example, for $(\mathbf{v}^*, \mathbf{z}) \in \Pi^{R=1}$, we have $P_{Q_{in}}^*(\mu(\mathbf{v}^*, \mathbf{z})) = \{1\}$ by Lemma B.1.

By the Berge maximum theorem (e.g., Aliprantis and Border, 2006, Thm. 17.31), the value function $v_{Q_{in}}^S$ is continuous, and the correspondence $P_{Q_{in}}^*$ has nonempty compact values and is upper hemicontinuous. Fix $\varepsilon > 0$. Since the correspondence $P_{Q_{in}}^*(\cdot)$ is upper hemicontinuous and nonempty, the set

$$V_{\varepsilon, Q_{in}}^S := \{\mu \in \Delta^S : P_{Q_{in}}^*(\mu) \subset (1 - \frac{\varepsilon}{2}, 1]\}$$

²⁸It is clear from the proof of Lemma B.2 that we could alternatively have shown that the function $g^S(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$ is jointly continuous in p , \mathbf{v}^* , \mathbf{z} , and Q_{in} ; the present result, however, is what we need for our proofs.

is open in Δ^S and nonempty; in fact, it contains the closed set $\Delta^{R=1}$. For $\delta < 1$, define

$$W_{1-\delta}^S := \{\mu \in \Delta^S : \mu(U_{1-\delta}^{R=1}) > \delta\},$$

where we recall that $U_{1-\delta}^{R=1}$ is the (open) neighborhood of the event that $R = 1$. Clearly, the sets $W_{1-\delta}^S$ shrink with δ , that is, if $\delta < \delta' < 1$, then $W_{1-\delta'}^S \subsetneq W_{1-\delta}^S$. Also, $W_{1-\delta}^S$ is open in Δ^S (Billingsley, 1968) and contains the closed set $\Delta^{R=1}$. Moreover, $\Delta^{S,\delta} \subset W_{1-\delta}^S$ for any $\delta < 1$. So, by choosing δ sufficiently close to 1, we have

$$\Delta^{S,\delta} \subset W_{1-\delta}^S \subset V_{\varepsilon, Q_{in}}^S.$$

That is, for any $\varepsilon > 0$, if the environment is δ -stable for δ sufficiently close to 1 (for a given ε), then for any $p_{Q_{in}}^*(\mu_{(\mathbf{v}^*, \mathbf{z})}) \in P_{Q_{in}}^*(\mu_{(\mathbf{v}^*, \mathbf{z})})$, we have $p_{Q_{in}}^*(\mu_{(\mathbf{v}^*, \mathbf{z})}) > 1 - \varepsilon/2$, and thus $d^*(Q_{in}) > 1 - \varepsilon$ (Lemma A.1). In words, if the manager assigns high probability to the alternatives being roughly equally strong, then it is optimal to have almost homogeneous teams. \square

B.3 Proof of Theorem 3.2

We follow a similar approach as in the proof of Theorem 3.1: we characterize the optimal team composition for a benchmark case, and show that the optimal team composition is close to the optimal team composition for the benchmark case whenever the environment is close to the benchmark case.

As in the proof of Theorem 3.1, it will be convenient to work with and their laws rather than with densities. We use the notation introduced there.

So, as before, the collection of δ -changeable environments corresponds to a set of random variables. For $\delta < 1$, let $\Pi^{D,\delta}$ be the set of random vectors $(\mathbf{v}^*, \mathbf{z})$ (taking values in \mathbb{R}^2) that satisfy the moment conditions (3.1) as well as the following conditions:

D_δ -I the probability that \mathbf{v}^* and \mathbf{z} are in $U_{1-\delta}^{R \geq \tilde{R}}$ is greater than δ ;

D_δ -II the payoffs have a well-defined joint density $f(v^*, z)$;

D_δ -III $f(v^*, z) > 0$ for every v^*, z such that $R > \tilde{R}$;

D_δ -IV the event that $\mathbf{v}^* \geq v$, $\mathbf{v}^* > y$, and $v > \mathbf{z}$ has probability 1.

We note that $\tilde{R} = U_{out}(Q_{in})$ in the notation of Appendix A, and we will henceforth use this notation in the proof.

Consider the set Π^D of random vectors (on \mathbb{R}^2) such that $R \geq U_{out}(Q_{in})$ with probability 1 and that satisfy conditions D_δ -II– D_δ -IV as well as condition (3.1). Clearly, $\Pi^D \subset \Pi^{D,\delta}$ for every $\delta < 1$. In this benchmark case, it is optimal to have maximally diverse teams.

Lemma B.3. Fix Q_{in} . For any $(\mathbf{v}^*, \mathbf{z}) \in \Pi^D$, the unique optimal team composition is $d^* = 0$.

Proof. Fix Q_{in} and $(\mathbf{v}^*, \mathbf{z}) \in \Pi^D$. The expected team payoff when the majority share is p is

$$\mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})] = \mathbb{E}[\mathbb{I}_{\mathbf{R} \in (U_{out}(Q_{in}), U_{in}(Q_{in}))} \cdot \gamma_2(p, Q_{in}, \mathbf{v}^*, \mathbf{z})] + \mathbb{E}[\mathbb{I}_{\mathbf{R} > U_{in}(Q_{in})} \cdot \gamma_1(p, Q_{in}, \mathbf{v}^*, \mathbf{z})],$$

where \mathbb{I}_E is the indicator function for the event E , and where we have used the notation in Appendix A.4.²⁹ By (3.1), we can bring the derivative inside the expectation operator when calculating the derivative, as before. Using that the thresholds (viz., $U_{out}(Q_{in})$ and $U_{in}(Q_{in})$) and the function $\gamma_1(p, Q_{in}, \mathbf{v}^*, \mathbf{z})$ do not depend on p , we thus have

$$\frac{d\mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]}{dp} = \mathbb{E} \left[\mathbb{I}_{R \in (U_{out}(Q_{in}), U_{in}(Q_{in}))} \cdot \frac{\gamma_2(p, Q_{in}, \mathbf{v}^*, \mathbf{z})}{dp} \right].$$

If the function $\gamma_2(\cdot, Q_{in}, v^*, z)$ has a unique maximum $p^* = p^*(Q_{in})$ that is independent of the realizations v^* and z of \mathbf{v}^* and \mathbf{z} , respectively, such that $v^* \geq v$, $v^* > y$, and $v > z$ (consistent with condition D_δ -IV), then this is the unique maximum for the expected team payoff $\mathbb{E}[\mathcal{C}(\cdot, Q_{in}, \mathbf{v}^*, \mathbf{z})]$ (for a given Q_{in}). (Note that condition D_δ -III rules out that the manager is indifferent among team compositions.)

So, it remains to show that the function $\gamma_2(\cdot, Q_{in}, v^*, z)$ is maximized when $p = \frac{1}{2}$ assuming that the payoff realizations v^* and z satisfy $v^* \geq v$, $v^* > y$, and $v > z$. For any $(v^*, z) \in \mathbb{R}^2$,

$$\frac{d\gamma_2(p, Q_{in}, v^*, z)}{dp} = (4p - 2) \cdot \left[\frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}) - v^* \right].$$

It thus suffices to show that

$$v^* > \frac{1}{2} \cdot (v^* + v) \cdot Q_{in} + \frac{1}{2} \cdot (y + z) \cdot (1 - Q_{in}).$$

We can rewrite this inequality as

$$\frac{1}{2} \cdot (1 - Q_{in}) \cdot (v^* + v - y - z) + \frac{1}{2} \cdot (v^* - v) > 0.$$

This inequality is satisfied if $v^* \geq v$, $v^* > y$, and $v > z$ (given that $Q_{in} < 1$). So, under conditions D_δ -III and D_δ -IV, the expectation of γ_2 is maximized uniquely (for given Q_{in}) at $p^*(Q_{in}) = \frac{1}{2}$. By Lemma A.1, the unique optimal team composition in this environment is $d^*(Q_{in}) = 2p^*(Q_{in}) - 1 = 0$ for any group identity Q_{in} . \square

²⁹Note that the event that $\mathbf{R} = U_{out}(Q_{in})$ has zero probability (by condition D_δ -II), so that $\mathbb{E}[\mathbb{I}_{R=U_{out}(Q_{in})} \cdot \gamma_3(p, Q_{in}, \mathbf{v}^*, \mathbf{z})] = 0$. We can thus omit this term.

We can now apply a continuity argument, as in the proof of Theorem 3.2. As in the proof of Theorem 3.2, let Π be the set of laws of random variables on \mathbb{R}^2 that satisfy conditions (3.1), D_δ -II, and D_δ -IV.³⁰ As before, we can identify Π with the set Δ of laws of random vectors in Π , and we can endow it with the relative topology induced by the weak topology. Clearly, $\Pi^{D,\delta} \subset \Pi$ for every δ , so that Π^D is also a subset of Π . Define the function $g^D : [\frac{1}{2}, 1] \times (\frac{1}{2}, 1) \times \Pi \rightarrow \mathbb{R}$ by $g^D(p, Q_{in}, \mathbf{v}^*, \mathbf{z}) := \mathbb{E}[\mathcal{C}(p, Q_{in}, \mathbf{v}^*, \mathbf{z})]$. The following is a direct corollary of Lemma B.2.

Corollary B.4. *For every Q_{in} , the function $g^D(\cdot, Q_{in}, \cdot, \cdot)$ is jointly continuous in its arguments.*

We can again apply the Berge maximum theorem to show that when we are close to the benchmark case, then the optimal team composition is almost maximally diverse. For $Q_{in} \in (\frac{1}{2}, 1)$, define the value function $v_{Q_{in}}^D : \Delta \rightarrow \mathbb{R}$ by

$$v_{Q_{in}}^D(\mu(\mathbf{v}^*, \mathbf{z})) := \max_{\tilde{p} \in [\frac{1}{2}, 1]} g^D(\tilde{p}, Q_{in}, \mathbf{v}^*, \mathbf{z}).$$

Also, define the correspondence of maximizers by

$$P_{Q_{in}}^*(\mu(\mathbf{v}^*, \mathbf{z})) := \{p \in [\frac{1}{2}, 1] : g^D(p, Q_{in}, \mathbf{v}^*, \mathbf{z}) = v_{Q_{in}}^D(\mu(\mathbf{v}^*, \mathbf{z}))\}.$$

As before, $v_{Q_{in}}^D(\mu(\mathbf{v}^*, \mathbf{z}))$ is the expected team payoff when the majority share p is chosen so as to maximize the team payoff (for given Q_{in} and \mathbf{v}^*, \mathbf{z}), and this maximum is attained at the majority shares that belong to the set $P_{Q_{in}}^*(\mu(\mathbf{v}^*, \mathbf{z}))$.

Again, it follows from the Berge maximum theorem (e.g., Aliprantis and Border, 2006, Thm. 17.31) that the value function $v_{Q_{in}}^D$ is continuous. Moreover, the correspondence $P_{Q_{in}}^*$ has nonempty compact values and is upper hemicontinuous. Let $\varepsilon > 0$. Since the correspondence $P_{Q_{in}}^*(\cdot)$ is upper hemicontinuous and nonempty, the set

$$V_{\varepsilon, Q_{in}}^D := \{\mu \in \Pi : P_{Q_{in}}^*(\mu) \subset [\frac{1}{2}, \frac{1}{2} + \frac{1}{2}\varepsilon]\}$$

is open in Π and nonempty. Clearly, it contains the closed set Π^D . For $\delta < 1$, define

$$W_{1-\delta}^D := \{\mu \in \Pi : \mu(U_{1-\delta}^{R \geq \tilde{R}}) > \delta\},$$

where we recall that $U_{1-\delta}^{R \geq \tilde{R}}$ is the (open) neighborhood of the event that $R \geq \tilde{R} = U_{out}(Q_{in})$. As before, $W_{1-\delta}^D$ decreases in δ and is open in Δ (Billingsley, 1968). It also contains the closed set Δ^D . In addition, $\Delta^{D,\delta} \subset W_{1-\delta}^D$ for every δ . So, if δ is sufficiently close to 1,

$$\Delta^{D,\delta} \subset W_{1-\delta}^D \subset V_{\varepsilon, Q_{in}}^D.$$

³⁰It is immediate that the set Π as defined here is identical to the set Π in the proof of Theorem 3.1. Likewise for the set Δ .

In words: if the environment is δ -changeable for δ sufficiently close to 1 (for a fixed $\varepsilon > 0$), then the majority share that maximizes the expected team payoff is less than $\frac{1}{2} + \frac{1}{2}\varepsilon$. By Lemma A.1, we thus have $d^*(Q_{in}) < \varepsilon$. So, if it is likely that an innovation significantly improves the payoffs of action s^1 , then it is optimal to have diverse teams. \square

Remark 2. Some comments on the role of the moment conditions (3.1) are in order. The conditions are used in two ways. First, it ensures that we can “pass to the limit” when taking derivatives. Second, it ensures that the optimal team composition is continuous in an appropriate sense in the proofs of Theorems 3.1 and 3.2. While condition 3.1 is much weaker than assuming that payoffs are bounded, it is stronger than what is needed; the condition could be relaxed at the expense of complicating notation and proofs. Finally, the condition on the moment of \mathbf{z} in (3.1) is stronger than the other conditions (viz., we require $\mathbb{E}[|\mathbf{z}|^{2+2\eta}]$ to be finite for some $\eta > 0$, while we only require $\mathbb{E}[|\mathbf{v}^*|^{1+\eta}]$ to be finite). This is because the relevant equations include higher powers of \mathbf{z} than of \mathbf{v}^* . \triangleleft

B.4 Proof of Proposition 4.1

As in the proofs of Theorem 3.1 and 3.2, we distinguish between a random variable \mathbf{x} and its realization x . Denote the distribution of the risk-adjusted payoff under $f(v^*, z)$ and $f'(v^*, z)$ by $F_R(r)$ and $F'_R(r)$, respectively. So, $F_R(r)$ is the probability of the event that $\mathbf{R} = \frac{\mathbf{v}^* - \mathbf{y}}{v - \mathbf{z}} \leq r$ when the economic environment is $f(v^*, z)$; $F'_R(r)$ is defined analogously.

The probability that the Pareto-dominant Nash equilibrium (s^1, s^1) is played can easily be derived from the characterization of the team payoffs in Appendix A. Fix the majority share p and group identity Q_{in} . Then, when the risk-adjusted payoff ratio is equal to R , the probability that the payoff dominant Nash equilibrium (s^1, s^1) is played is given by

$$\nu(R; p, Q_{in}) := \begin{cases} \nu_1(p, Q_{in}) & \text{if } R > U_{in}(Q_{in}); \\ \nu_2(p, Q_{in}) & \text{if } U_{out}(Q_{in}) < R \leq U_{in}(Q_{in}); \\ \nu_3(p, Q_{in}) & \text{if } L_{out}(Q_{in}) \leq R \leq U_{out}(Q_{in}); \\ \nu_4(p, Q_{in}) & \text{if } L_{in}(Q_{in}) \leq R < L_{out}(Q_{in}); \\ \nu_5(p, Q_{in}) & \text{if } R < L_{in}(Q_{in}); \end{cases}$$

where

$$\begin{aligned}
\nu_1(p, Q_{in}) &:= 1 \\
\nu_2(p, Q_{in}) &:= (p^2 + (1-p)^2) \cdot \frac{1}{2} Q_{in} + 2p \cdot (1-p) \\
\nu_3(p, Q_{in}) &:= (p^2 + (1-p)^2) \cdot \frac{1}{2} Q_{in} + 2p \cdot (1-p) \cdot \frac{1}{2} Q_{out} \\
\nu_4(p, Q_{in}) &:= (p^2 + (1-p)^2) \cdot \frac{1}{2} \cdot Q_{in} \\
\nu_5(p, Q_{in}) &:= 0.
\end{aligned}$$

It is easy to check that for every $p < 1$ and Q_{in} ,

$$\nu_1(p, Q_{in}) > \nu_2(p, Q_{in}) > \nu_3(p, Q_{in}) > \nu_4(p, Q_{in}) > \nu_5(p, Q_{in}),$$

so that for every p and Q_{in} , the function $\nu(R; p, Q_{in})$ is increasing in R .

Fix Q_{in} and a random vector $(\mathbf{v}^*, \mathbf{z})$ with joint density $\tilde{f}(v^*, z)$ (which represents the economic environment, as before) and associated distribution $\tilde{F}_R(r)$ over the risk-adjusted payoff ratio. If there is a unique majority share $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in})$ that maximizes the expected team payoff in the introspective equilibrium, then the implementation rate is given by

$$\begin{aligned}
\mathcal{I}_{\tilde{f}}(Q_{in}) &:= \mathbb{E}[\nu(R; p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in})] \\
&= \tilde{F}_R(L_{in}(Q_{in})) \cdot \nu_5(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) + \\
&\quad (\tilde{F}_R(L_{out}(Q_{in})) - \tilde{F}_R(L_{in}(Q_{in}))) \cdot \nu_4(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) + \\
&\quad (\tilde{F}_R(U_{out}(Q_{in})) - \tilde{F}_R(L_{out}(Q_{in}))) \cdot \nu_3(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) + \\
&\quad (\tilde{F}_R(U_{in}(Q_{in})) - \tilde{F}_R(U_{out}(Q_{in}))) \cdot \nu_2(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) + \\
&\quad (1 - \tilde{F}_R(U_{in}(Q_{in}))) \cdot \nu_1(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}).
\end{aligned}$$

We consider strongly stable and strongly changeable economic environments in turn.

Strongly stable environments A minor modification of the proof of Lemma B.1 shows that it is optimal to have maximally homogeneous teams when the environment \tilde{f} is strongly stable: $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}) = 1$. Moreover, when the environment is strongly stable, it assigns probability 1 to the event that $\mathbf{R} \in [L_{out}(Q_{in}), U_{out}(Q_{in})] = [1/\tilde{R}, \tilde{R}]$. So, the implementation rate can be written as

$$\mathcal{I}_{\tilde{f}}(Q_{in}) = \nu_3(1, Q_{in}).$$

This is true for any strongly stable environment $\tilde{f}(v^*, z)$. So, $\mathcal{I}_f(Q_{in}) = \mathcal{I}_{f'}(Q_{in})$ for any strongly stable environments $f(v^*, z)$, $f'(v^*, z)$.

Strongly changeable environments When the environment $\tilde{f}(v^*, z)$ is strongly changeable, Lemma B.3 shows that the unique optimal team composition is $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}) = \frac{1}{2}$. Moreover, when the environment is strongly stable, we have that $\mathbf{R} \geq U_{out}(Q_{in}) = \tilde{R}$ with probability 1. We can thus simplify the expression for the implementation rate to

$$\mathcal{I}_{\tilde{f}}(Q_{in}) = \tilde{F}_R(U_{in}(Q_{in})) \cdot \nu_2(\frac{1}{2}, Q_{in}) + (1 - \tilde{F}_R(U_{in}(Q_{in}))) \cdot \nu_1(\frac{1}{2}, Q_{in})$$

As f dominates f' , we have $F_R(r) < F'_R(r)$ for all $r \in (\tilde{R}, \infty)$. In particular,

$$F_R(U_{in}(Q_{in})) < F'_R(U_{in}(Q_{in})),$$

and the result follows. \square

B.5 Proof of Proposition 4.2

As before, we distinguish between a random variable \mathbf{x} and its realization x . As noted in the proof of Proposition 4.1, it is optimal to have homogeneous teams when the environment $f(v^*, z)$ is strongly stable, that is, $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}) = 1$. Moreover, as in the proof of Proposition 4.1, the implementation rate is simply

$$\mathcal{I}_f(Q_{in}) = \nu_3(1, Q_{in}) = \frac{1}{2}Q_{in}.$$

This function is clearly strictly increasing in the strength Q_{in} of players' identity. \square

This result extends to more general environments. We can use a similar continuity argument as in the proof of Theorem 3.1 and then apply the implicit function theorem to account for the fact that the optimal team composition may change when group identity changes.

B.6 Proof of Proposition 4.3

Fix Q_{in} . As before, if there is a unique majority share $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in})$ that maximizes the expected team payoff in the introspective equilibrium (given the economic environment $f(v^*, z)$), then the implementation rate is given by

$$\mathcal{I}_f(Q_{in}) = F_R(U_{in}(Q_{in})) \cdot \nu_2(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) + (1 - F_R(U_{in}(Q_{in}))) \cdot \nu_1(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}),$$

where $F_R(r)$ is the distribution of the risk-adjusted payoff ratio induced by $f(v^*, z)$. Differentiating this expression with respect to Q_{in} gives:

$$\begin{aligned} \frac{d\mathcal{I}_f(Q_{in})}{dQ_{in}} &= F_R(U_{in}(Q_{in})) \cdot \frac{d\nu_2(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in})}{dQ_{in}} - \\ &\quad - f_R(U_{in}(Q_{in})) \cdot \frac{dU_{in}(Q_{in})}{dQ_{in}} \cdot (\nu_1(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in}) - \nu_2(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in})), \quad (\text{B.1}) \end{aligned}$$

where f_R is the density of R induced by F_R . The first term in (B.1) is the increase in the implementation rate due to the enhanced coordination when players follow their impulses in at least some of their matches (i.e., when $\mathbf{R} \in (U_{out}(Q_{in}), U_{in}(Q_{in}))$). We encountered a similar effect in stable environments; cf. the proof of Proposition 4.2. The second term is the decrease in the implementation rate due to the stronger pressure to conform: when Q_{in} increases, there is a smaller range of R for which all players choose action s^1 (i.e., $U_{in}(Q_{in})$ increases). For a given economic environment, the derivative of ν_2 at the optimal team composition p^* is given by

$$\begin{aligned} \frac{d\nu_2}{dQ_{in}} &= \frac{\partial \nu_2}{\partial Q_{in}} + \frac{\partial \nu_2}{\partial p^*} \cdot \frac{dp^*}{dQ_{in}} \\ &= \frac{1}{2} \cdot ((p^*)^2 + (1 - p^*)^2) + p^* \cdot (1 - p^*) \cdot (\lambda - (1 - \lambda)) + \\ &\quad (2p^* - 1) \cdot (Q_{in} - Q_{out}) \cdot \frac{dp^*}{dQ_{in}} \end{aligned}$$

where we have used the expression for Q_{out} in Lemma A.2. The existence of $\frac{dp^*}{dQ_{in}}$ follows from the implicit function theorem by standard arguments. We also have

$$\begin{aligned} \frac{dU_{in}(Q_{in})}{dQ_{in}} &= \frac{1}{(1 - Q_{in})^2} > 0; \\ \nu_1(\tfrac{1}{2}, Q_{in}) - \nu_2(\tfrac{1}{2}, Q_{in}) &= \tfrac{1}{2} \cdot (1 - \tfrac{1}{2}Q_{in}) > 0. \end{aligned}$$

Finally, by Lemma B.3, it is optimal to have maximally diverse teams when the environment $(\mathbf{v}^*, \mathbf{z})$ is strongly* changeable, that is, $p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}) = \tfrac{1}{2}$. This gives

$$\frac{d\nu_2(p_{(\mathbf{v}^*, \mathbf{z})}^*(Q_{in}), Q_{in})}{dQ_{in}} = 2\lambda.$$

Using these results, we see that the implementation rate strictly decreases in Q_{in} if and only if

$$f_R(U_{in}(Q_{in})) \cdot \frac{2 - Q_{in}}{(1 - Q_{in})^2} > 2\lambda f_R(U_{in}(Q_{in})).$$

Using that in strongly* changeable environments, the density f_R is weakly increasing in R for $R \leq U_{in}(Q_{in})$, we have

$$F_R(U_{in}(Q_{in})) = \int_{U_{out}(Q_{in})}^{U_{in}(Q_{in})} f_R(r) dr \leq f_R(U_{in}(Q_{in})) \cdot (U_{in}(Q_{in}) - U_{out}(Q_{in})).$$

So, the implementation rate strictly decreases with group identity whenever

$$f_R(U_{in}(Q_{in})) \cdot \frac{2 - Q_{in}}{(1 - Q_{in})^2} > 2\lambda f_R(U_{in}(Q_{in})) \cdot \left(\frac{Q_{in}}{1 - Q_{in}} - \frac{Q_{out}}{1 - Q_{out}} \right).$$

Given that $f_R(U_{in}(Q_{in})) > 0$ by definition (see (2) in the definition of δ -changeable environments), this inequality holds if and only if

$$\frac{2 - Q_{in}}{(1 - Q_{in})^2} > 2\lambda \left(\frac{Q_{in}}{1 - Q_{in}} - \frac{Q_{out}}{1 - Q_{out}} \right).$$

Applying Lemma A.2 and rewriting gives that this holds if and only if

$$(2 - Q_{in}) \cdot (1 - \lambda Q_{in} - (1 - \lambda) \cdot Q_{in}) > 2\lambda(1 - \lambda) \cdot (2Q_{in} - 1) \cdot (1 - Q_{in}).$$

It can be checked that this inequality holds for any $Q_{in} \in (\frac{1}{2}, 1)$ and $\lambda \in (\frac{1}{2}, 1)$. \square

As is the case for Proposition 4.2, this result extends to more general environments. As in the proof of Theorem 3.2, we can use a continuity argument to show that similar results hold for environments that are almost strongly changeable.

References

- Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics* 115, 715–753.
- Akerlof, G. A. and R. E. Kranton (2005). Identity and the economics of organizations. *Journal of Economic Perspectives* 19, 9–32.
- Alesina, A., J. Harnoss, and H. Rapoport (2013). Birthplace diversity and economic prosperity. Working Paper 18699, NBER.
- Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature* 43, 762–800.
- Aliprantis, C. and K. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker’s Guide* (3rd ed.). Springer.
- Aumann, R. J. (1987). Correlated equilibria as an expression of Bayesian rationality. *Econometrica* 55, 1–18.
- Bacharach, M. and D. O. Stahl (2000). Variable-frame level- n theory. *Games and Economic Behavior* 32, 220–246.
- Baron, J. N., M. D. Burton, and M. T. Hannan (1996). The road taken: Origins and evolution of employment systems in emerging companies. *Industrial and Corporate Change* 5, 239–275.

- Bénabou, R. (2013). Groupthink: Collective delusions in organizations and markets. *Review of Economic Studies* 80, 429–462.
- Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy* 102, 841–877.
- Bicchieri, C. and Y. Fukui (1999). The great illusion: Ignorance, informational cascades, and the persistence of unpopular norms. *Business Ethics Quarterly* 9, 127–155.
- Billingsley, P. (1968). *Convergence of Probability Measures* (1st ed.). Wiley.
- Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2015). Bend it like Beckham: Ethnic identity and integration. Working paper.
- Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory* 97, 298–319.
- Blume, L. E. (1995). The statistical mechanics of best-response strategy revision. *Games and Economic Behavior* 11, 111–145.
- Bowers, C. A., J. A. Pharmer, and E. Salas (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research* 31, 305–327.
- Camerer, C. F. and M. Knez (2002). Coordination in organizations: A game-theoretic perspective. In Z. Shapira (Ed.), *Organizational Decision Making*, Chapter 8. Cambridge University Press.
- Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica* 61, 989–1018.
- Cartwright, S. and C. L. Cooper (1993). The role of culture compatibility in successful organizational marriage. *Academy of Management Executive* 7, 57–70.
- CBS 60 Minutes (2013, January 7). How to design breakthrough inventions?
- Chen, R. and Y. Chen (2011). The potential of social identity for equilibrium selection. *American Economic Review* 101, 2562–2589.
- Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51, 5–62.
- Crawford, V. P. and H. Haller (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica* 58, 571–595.

- Cr  mer, J. (1993). Corporate culture and shared knowledge. *Industrial and Corporate Change* 2, 351–386.
- Cushman, J. (1990). Avianca flight 52: The delays that ended in disaster. *New York Times*. February 5.
- De Dreu, C. K. W. (2006). When too little or too much hurts: Evidence for a curvilinear relationship between task conflict and innovation in teams. *Journal of Management* 32, 83–107.
- de Vignemont, F. and T. Singer (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences* 10(10), 435–441.
- Dezs  , C. L. and D. G. Ross (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal* 33, 1072–1089.
- Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indices. *American Sociological Review* 20, 200–217.
- Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal* 32, 543–576.
- Elfenbein, H. A. and N. Ambady (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128, 243–249.
- Epley, N. and A. Waytz (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of Social Psychology* (Fifth ed.), Volume 1, Chapter 14. Wiley.
- Gersick, C. J. and J. R. Hackman (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes* 47, 65–97.
- Gibbons, R. and J. Roberts (Eds.) (2013). *Handbook of Organizational Economics*. Princeton University Press.
- Greenstein, F. I. (1988). *Leadership in the Modern Presidency*. Harvard University Press.
- Grout, P. A., S. Mitraile, and S. Sonderegger (2014). The costs and benefits of coordinating with a different group. Working paper, University of Nottingham.
- Gudykunst, W. B. (2004). *Bridging Differences: Effective Intergroup Communication*. SAGE.
- Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. MIT Press.

- Hermalin, B. E. (2013). Leadership and corporate culture. In R. Gibbons and J. Roberts (Eds.), *Handbook of Organizational Economics*, pp. 432–478. Princeton University Press.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE.
- Hong, L. and S. E. Page (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory* 97, 123–163.
- Jackson, M. O. and Y. Xing (2014). Culture-dependent strategies in coordination games. *Proceedings of the National Academy of Sciences* 111, 10889–10896.
- Jackson, S. E. and A. Joshi (2011). Work team diversity. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology*, Volume 1, Chapter 20. American Psychological Association.
- James, W. (1890/1983). *Principles of Psychology*. Harvard University Press.
- Janssen, M. C. W. (2001). Rationalizing focal points. *Theory and Decision* 50, 119–148.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
- Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56.
- Kets, W. and A. Sandroni (2015a). A belief-based theory of homophily. Working paper, Northwestern University.
- Kets, W. and A. Sandroni (2015b). A rational approach to coordination. Mimeo, Northwestern University.
- Kreps, D. M. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (Eds.), *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge University Press.
- Kuran, T. and W. Sandholm (2008). Cultural integration and its discontents. *Review of Economic Studies* 75, 201–228.
- Lazear, E. P. (1999a). Culture and language. *Journal of Political Economy* 107(S6), S95–S126.
- Lazear, E. P. (1999b). Globalisation and the market for team-mates. *Economic Journal* 109, C15–C40.

- Le Coq, C., J. Tremewan, and A. K. Wagner (2015). On the effects of group identity in strategic environments. *European Economic Review* 76, 239–252.
- Locke, J. (1690/1975). *An essay concerning human understanding*. Oxford University Press.
- Mailath, G. J. (1998). Do people play Nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature* XXXVI, 1347–1374.
- McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior* 10, 6–38.
- Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review* 84, 658–673.
- Meyer, E. (2014). *The Culture Map: Breaking Through the Invisible Boundaries of Global Business*. PublicAffairs.
- Mill, J. S. (1872/1974). *A system of logic, ratiocinative and inductive*, Volume 7 of *Collected works of John Stuart Mill*. University of Toronto Press.
- Mokyr, J. (1990). *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press.
- Morris, S. (2000). Contagion. *Review of Economic Studies* 67, 57–78.
- Ottaviano, G. I. P. and G. Peri (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography* 6, 9–44.
- Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- Peri, G. (2012). The effect of immigration on productivity: Evidence from U.S. states. *Review of Economics and Statistics* 94, 348–358.
- Pfeffer, J. and R. I. Sutton (1999). Knowing “what” to do is not enough: Turning knowledge into action. *California Management Review* 42, 83–108.
- Phillips, K. W., G. B. Northcraft, and M. A. Neale (2006). Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group Processes & Intergroup Relations* 9, 467–482.
- Prat, A. (2002). Should a team be homogenous? *European Economic Review* 46, 1187–1207.

- Rao, H., R. Sutton, and A. P. Webb (2008). Innovation lessons from Pixar: An interview with Oscar-winning director Brad Bird. *McKinsey Quarterly*. April.
- Reagans, R. and E. W. Zuckerman (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science* 12, 502–517.
- Richard, O., A. McMillan, K. Chadwick, and S. Dwyer (2003). Employing an innovation strategy in racially diverse workforces: Effects on firm performance. *Group & Organization Management* 28, 107–126.
- Robalino, N. and A. J. Robson (2015). The evolution of strategic sophistication. *American Economic Review*. forthcoming.
- Russell, B. (1948). *Human knowledge: Its scope and limits*. George Allen & Unwin.
- Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press.
- Simons, T., L. H. Pelled, and K. A. Smith (1999). Making use of difference: Diversity, debate, and decision comprehensiveness in top management teams. *Academy of Management Journal* 42, 662–673.
- Sugden, R. (1995). A theory of focal points. *Economic Journal* 105, 533–550.
- Triandis, H. C. and E. M. Suh (2002). Cultural influences on personality. *Annual Review of Psychology* 53, 133–160.
- Van den Steen, E. (2010). Culture clash: The costs and benefits of homogeneity. *Management Science* 56, 1718–1738.
- Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1991). Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. *Quarterly Journal of Economics* 106, 885–910.
- Weber, R. A. and C. F. Camerer (2003). Cultural conflict and merger failure: An experimental approach. *Management Science* 49, 400–415.
- Wiersema, W. F. and K. A. Bantel (1992). Top management team demography and corporate strategic change. *Academy of Management Journal* 35, 91–121.
- Young, H. P. (1993). The evolution of conventions. *Econometrica* 61, 57–84.