INSEAD

The Business School for the World

Faculty & Research Working Paper

Better, Best or Worst Team? Linking Intra-Team Diversity to Extreme Team Performance

Fabrice CAVARRETTA

2007/56/OB

Better, Best Or Worst Team? Linking Intra-Team Diversity to Extreme Team Performance

Fabrice CAVARRETTA*

I am grateful to Martin Kilduff, Reinhard Angelmar and Ajay Mehra for sharing their dataset.
I thank my thesis adviser, Bruce Kogut, for his support, as well as Rodolphe Durand, Henrich Greve, Etty Jehn, Denrell Jerkell, Christoph Loch, Randall Peterson, and Maurizio Zollo
for their thoughtful comments on earlier drafts.

Working Paper: do not distribute or cite without author's authorization.

* Phd Student, Organisational Behaviour at INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, Fabrice.Cavarretta@insead.edu

Better, Best Or Worst Team? Linking Intra-Team Diversity to Extreme Team Performance

ABSTRACT

Intra-team diversity has contradictory effects, both improving information and hindering social integration, resulting in a neutral average effect on team performance. Combining those effects differently allows the prediction of a U-shaped relationship of diversity to performance variability, improving on past prediction of a simple increase. The reanalysis of a field study of 35 teams in a business simulation confirms such effect for three demographic and two cognitive diversity variables. An analysis combining both mean and variability effects leads to improved predictions, and the conclusions regarding the effect of age diversity drawn from a conventional mean analysis even reverse if aiming for an extreme performance goal.

INTRODUCTION

Most organizational research on intra-team diversity and team performance has sought to answer the question of whether diversity increases performance. However, the results have not been consistent across studies (O'Reilly, Williams, and Barsade, 1998) and have puzzled researchers for decades with two views, one optimistic and one pessimistic, that directly contradict each other (Mannix and Neale, 2005). On the one hand, intra-team diversity brings more information to the team (Nemeth, 1986), improving performance. On the other hand, diversity disrupts social integration (O'Reilly, Caldwell, and Barnett, 1989), impairing performance. Nemeth and Staw (1989) demonstrated that the effects of information availability and social integration can counter each other, which explains why most meta-analyses find a neutral effect (Williams and O'Reilly, 1998). Scholars seeking to reconcile these findings have either defined diversity with increasing subtlety (e.g., Bunderson and Sutcliffe, 2002; Harrison and Klein, 2007) or introduced variables to moderate the relationship between diversity and performance (e.g., Polzer, Milton, and Swann, 2002; van Knippenberg, De Dreu, and Homan, 2004).

However, there are contexts in which mean performance is not the performance measure of interest: in a "top-score competitive system" (Miner, Haunschild, and Schwab, 2003), one cares more about chances of a top ranking than about improving performance on average. For instance, in a winner-take-all contest, mean performance is uninteresting since one would only want to know what factors lead to the top position. For that purpose, performance variability—also called risk—may matter as much as mean performance in organizational research (March, 1991; Baum and McKelvey, 2006). For instance, an influencing factor may increase mean performance but decrease variability. If the effect on variability is strong enough, an increase in the factor may indeed decrease the probability of reaching the top rank in a competition, even though the factor improves performance on average. Concerning team

diversity, most research has explored its mean effect and does not adequately address the question of whether diversity increases the probability of finishing with either very high or very low performance. Following Fleming (2004), Taylor and Greve (2006), the current study focuses on the effect of diversity on performance variability.

Diversity is often considered as a variable whose mechanisms operate uniformly on its range. Yet teams with high diversity, with their potential for low integration and high knowledge, differ fundamentally from homogeneous teams, with their potential for low knowledge and high integration. In this paper I distinguish the mechanism that operates at high levels of diversity from the one at low levels of diversity, similar to theories that hypothesize curvilinear mean effects of diversity (e.g., Gibson and Vermeulen, 2003; Richard, et al., 2004; Uzzi and Spiro, 2005).

In a population of teams at high diversity levels, the lowest-performing teams are likely to suffer from conflicts; greater diversity then worsen their situation by further degrading social integration. At the same time, the best-performing teams are likely to have members who get along, so greater diversity increases knowledge and improves performance. It implies that some teams fail and some succeed, and the spread is exacerbated by greater diversity. This echoes previous claims (Taylor and Greve, 2006), except that I expect it to occur for only high levels of diversity rather than for the whole range of diversity.

In a population of teams at low diversity levels, the lowest-performing teams are likely to suffer from lack of proper information; greater diversity then leads to improved performance by improving available knowledge. At the same time, the best-performing teams likely enjoy a good knowledge-to-task fit, so greater diversity would not so much improve their information as introduce minority dissent, thus lowering performance. This implies that some teams fail and some succeed, but the spread is reduced by greater diversity. I therefore posit

that greater diversity reduces inter-team performance variability at low diversity levels and increases it at high diversity levels.

A field study of 35 teams engaged in a business simulation confirmed this effect for three demographic and two cognitive diversity variables. In addition, an innovative analysis explored whether taking into account effects of diversity on performance variability better predicts extreme performance. It found that although age diversity increases mean performance, it actually reduces the probability of achieving high performance by reducing performance variability. Overall, this study suggests that both extremely high and extremely low diversity may be preferable to middle-range diversity if the goal of reaching a high threshold of performance replaces the traditional goal of improving performance on average. Conversely, mid-range diversity would be preferable if the goal is to avoid low performance thresholds.

LITERATURE REVIEW

Research on the effects of intra-team diversity has identified two broad opposing perspectives, leading to the prediction that diversity has a neutral effect on average team performance (van Knippenberg and Schippers, 2007). I first review the literature on that relationship and the approaches used to disambiguate it. Then I present the alternative focus on the effects of diversity on performance variability, and explain why such effects matter.

Resolving a neutral mean effect with moderating factors or finer definitions of diversity

One perspective focuses on information availability and takes an "optimistic" view of diversity (Mannix and Neale, 2005:33), whereby intra-team diversity brings more information to the team, which can improve performance (Gruenfeld, et al., 1996; Dahlin, Weingart, and Hinds, 2005). For instance, exposure to minority thinking fosters a broader view of the issue at hand (Nemeth, 1986), and diverse teams exchange a wider range of information (Sommers, 2006). In contrast, a social integration perspective (O'Reilly, et al.,

1989) integrates the similarity-attraction (Festinger, 1954) and the self-categorization (Tajfel, 1982) approaches and takes a "pessimistic" view of diversity (Mannix and Neale, 2005:34). Here, diversity creates tensions in the team, which can be damaging to its performance. This view appears in Pfeffer's seminal study on demography (1983), suggesting a risk of conflict in organizations with discontinuous tenure distribution. It links heterogeneity to lower social integration (O'Reilly, et al., 1989), which has negative effects on various organizational outcomes such as turnover and performance.

These two perspectives suggest contradictory effects, so finding a simple main effect has proven elusive. Some authors have found results—linear or curvilinear—but starting with Nemeth and Staw (1989), various meta-analyses (Bowers, 2000; Webber and Donahue, 2001) and reviews have settled on an overall neutral effect of diversity (Milliken and Martins, 1996; Jackson, Joshi, and Erhardt, 2003). To address this ambiguity, at least two major streams of research have emerged (Mannix and Neale, 2005; van Knippenberg and Schippers, 2007). A first stream of research has focused on various moderating factors of the relationship between diversity and performance, such as team characteristics (e.g., entrepreneurial orientation in Richard, et al., 2004) or intra-team perceptions (e.g., interpersonal congruence in Polzer, et al., 2002) or context (e.g., people orientation of corporate culture in Kochan, et al., 2003:10). A second stream of research has narrowed the independent variable by adopting ever-finer definitions of diversity (e.g., Lau and Murnighan, 1998; Bunderson and Sutcliffe, 2002; Boone, Van Olffen, and Van Witteloostuijn, 2005; Cramton and Hinds, 2005). A wealth of definitions has led to a necessary theorization and classification of the various possible approaches to the diversity construct (Harrison and Klein, 2007).

Other ways to combine information and social integration effects

Those approaches attempted to balance the information availability and social integration effects by assuming they combine at the level of each team; yet, the summing of effects may

not be the only possible mechanism. For instance, addressing the issue of how various dimensions of diversity combine, van Knippenberg and Schippers proposed that "diversity's effects may be better understood if the influence of different dimensions of diversity is studied in interactions rather than as additive effects" (2007:519). Such logic underlies a stream of research that assumes that different types of diversity occur simultaneously, but their effects do not simply add up (Lau and Murnighan, 1998; Cramton and Hinds, 2005).

In the same spirit, the contradictory sub-effects of diversity on team performance could occur at the population level, with the sub-effects occurring across teams instead of adding up in each team. The categorization-elaboration model (van Knippenberg, et al., 2004) states that both task informational requirements and social integration moderate the relationship between diversity and performance. For instance, regarding moderation by social integration, it states that the effect of diversity differs depending on whether teams are in a good or bad social integration condition, improving performance in the first case, degrading it in the second. The theorizing below leverages the logic underlying such moderations: depending on conditions, diversity may simultaneously drive some teams to higher performance and other teams to lower performance.

Focusing on effects on performance variability

Van Knippenberg, De Dreu, and Homan suggested that "all dimensions of diversity may in principle elicit social categorization processes as well as information/decision-making processes" (2004:521), leading to the possibility of moving beyond "typologies of diversity" (2004:520). They even suggested that such a generic property of diversity applies to any "socially shared cognition" (2004:521). The current study follows that logic by defining team diversity as any dimension differentiating team members with the following two consequences: it leads to social categorization among team members, and it brings varied information.

Regarding the dependent variable, studies linking team composition to organizational performance have traditionally shared a common approach using linear regression to predict an effect on mean performance. However, performance variability, or how much the performance fluctuates around its expected value, sometimes matters to the organizational theorist more than mean performance. March's exploration-exploitation study (1991) seminally identified effects on performance variability as a better predictor of outcomes than mean effects. It led to a stream of literature on innovation and knowledge, where various factors, including diversity, might cause such variability (Sorensen, 2002; Denrell, 2003; Fleming, 2004). Accordingly, the current paper takes performance variability as its focal dependent variable.

Performance variability may sometimes better predict outcomes than a simple mean effect. In various settings, the context rewards teams performing a task only if they reach a performance threshold. For instance, the context may competitively rank the teams and reward only a top few, as is the case in the Olympic Games or a hot high-technology IPO market. Similarly, the management system at General Electric had the reputation to promote the top 25 percent of managers and dismiss the bottom 10 percent. In a study on variability, Miner, Haunschild, and Schwab (2003:803) defined such contexts as "competitions on extreme values." The common characteristic of all such settings is that one does not seek simply to increase the expected outcome—as predicted by the mean effect—but rather to reach a performance threshold.

Previous research explored how effects on variability would drive extreme outcomes. Cabral (2003) study the strategies of laggards to catch up with the leader of a race, and Tsetlin, Gaba and Winkler (2004) show how, in a contest, variability of outcomes may compensate for a handicap. To clarify the intuition on such mechanisms, let us reason on a few sketches (Figure 1). The sketch (a) illustrates a classical mean effect of X: on average, Y is higher at

X_T than at X_B (T for top values and B for bottom values). If the organizational goal is to reach a performance threshold Y_0, which values of X make it more likely?

---------- Insert Figure 1 roughly here ----------

Traditionally, one relies on the mean effect: since X increases the mean value of Y, one infers a greater probability of reaching the threshold at X_T simply by looking at (a). This assumes implicitly that the variability in Y is constant, as appears in (b) where the bell curves representing the distribution of Y have identical dispersion. The probability of reaching the threshold performance Y_0—measured by the tail of the distribution above the threshold—is driven only by the change in the mean, thereafter reaching Y_0 is more likely for X_T than for X_B.

The conclusion may change if X influences the variability of Y. The sketch (c) matches (a) for the mean effect, but assumes that X reduces the variability of Y, as represented by a more dispersed bell curve at X_B than at X_T. Now, the probability of reaching the threshold is higher at X_B (indicated by an arrow), even though the expected value is still higher at X_T. This simple reasoning shows that the variability effect can compensate for a mean effect if the aim is to reach a performance threshold. Such logic will be leveraged once the relationship between diversity and performance variability is specified.

THEORY

Differentiating effects in low vs. high levels of diversity

To propose that diversity influences not just mean performance but also performance variability, one must assume something more than simple additive effects of information availability and social integration. For instance, Taylor and Greve (2006:728) explained that an increase in experience diversity among team members increases team performance variability by the simultaneous possibility of low performance, because diversity increases

conflict, and high performance, because diversity increases the range of information available. Fleming (2004) used similar reasoning.

Those studies considered separately the sub-population of high performing teams from the sub-population of low-performing teams, and suggest that greater diversity influences the difference between those extremes. In other words, the neutral effect of diversity may represent an ecological fallacy (Robinson, 1950), in that greater diversity does not truly have a neutral effect for all sub-populations of teams. Instead, greater diversity makes some teams enjoy a greater positive outcome because of improved knowledge, while making others suffer a worse negative outcome because of degraded social integration.

The current paper follows Taylor and Greve (2006) in identifying the differentiated effect due to hazard on social integration, assuming that well performing teams tend to get along, and low performing team tend not to get along. However, difficult social integration should matter mainly for teams with diverse members. Therefore, it is necessary to distinguish the case of low diversity levels, where teams face not much the hazard of poor social integration but mainly the hazard of lack of knowledge. The method of distinguishing effects in low vs. high levels of diversity also appeared in studies linking diversity to performance by a curvilinear relationship (e.g., Gibson and Vermeulen, 2003; Richard, et al., 2004; Uzzi and Spiro, 2005). Those also distinguished the mechanism occurring in high diversity vs. the one occurring in low diversity. Below, I consider separately each diversity range (first high levels, then low levels), each with a differentiated hypothesis about the effects of diversity on performance variability.

Teams in high diversity: taking chances on integration

Let us first focus on teams in high diversity levels, characterized by a potential for high knowledge but low social integration, and consider the effect of greater diversity. In relatively high diversity, the main uncertainty is social integration, which can fluctuate

greatly, whereas knowledge is relatively available. On the one hand, the best teams are probably those whose members get along. Among them, greater diversity increases the knowledge available and used, thus improving performance. On the other hand, the worst teams are probably those whose members do not get along, where social integration is low. For them, greater diversity does not bring informational advantage, as team members do not get along; they can exploit only shared knowledge, which is reduced by diversity (Stasser and Titus, 1987), therefore hampering performance.

Such reasoning fits into the categorization-elaboration model (van Knippenberg, et al., 2004), which predicts that social integration positively moderates the relationship of diversity to performance (similar moderation also appeared in Van Der Vegt and Bunderson, 2005). Traditionally, such moderation took social integration as an exogenous factor. Here, since I study variability in the population of teams, I consider the differences in social integration that naturally occur inside the teams' population. The goal is not to predict the distribution of teams between good and bad integration. Rather, assuming random initial conditions exists, I focus on the effect of diversity on the spread of outcomes. Among the teams that are lucky to integrate properly, diversity increase performance, while it decrease it among those unfortunate to suffer from low integration. Overall, greater diversity is therefore associated with an increasing spread between the best- and worst-performing teams. This reasoning is similar to the one proposed by Taylor and Greve (2006), but restricted to the high levels of diversity. This lead to the following hypothesis:

Hypothesis 1: In a population of teams with high intra-team diversity, greater diversity leads to higher variability of team performance.

Figure 2 sketches the reasoning above. The x-axis represents a diversity variable, for example age diversity, and the y-axis represents team performance. The dashed line represents information availability sub-effect, with a positive slope since diversity improves

information. The dotted line represents the social integration sub-effect, with a negative slope since diversity reduces social integration. However, instead of assuming that those two simply add up, this theorizing suggests that the positive and the negative effects may occur in parallel. At high diversity levels (the right side of Figure 2), greater diversity leads to both higher and lower performance extremes, thus a greater spread of performance outcomes, which appear as two diverging continuous curves.

---------- Insert Figure 2 roughly here ----------

Teams in low diversity: taking chances on information

I now turn to discussing teams at relatively low diversity levels. For such homogeneous teams, social integration is expected but information is potentially seriously restricted; therefore, the main issue is whether each team possesses the knowledge to accomplish its task. For instance, consider a population of executive teams that are low in functional diversity. One team might be composed mainly of marketers and another one mainly of production specialists. If the task rewards marketing skill, the first team would perform strongly and the second one weakly. On an alternate task with different requirements, the results might be inverted. In low levels of diversity, whether the teams that are fitted to their task or not should therefore drive most of the variability. Now, how would diversity influence the spread between best and worst performing teams?

Past research has suggested taking into account task-knowledge fit when studying the link between diversity and performance. The categorization-elaboration model (van Knippenberg, et al., 2004:1012) predicts that task requirement moderates the relationship of diversity to performance. A similar moderation was also put forward by Jehn, Northcraft, and Neale (1999), whereby diversity would have a more positive effect on performance for complex task than for simple task. Traditionally, such moderation took task-knowledge fit as an exogenous factor. Here, since I study variability in a population of teams, the difference

occurs inside the team population: for some homogeneous teams a task will be simple (a marketing task for a marketing team), while complex for others (a marketing task for a production team). An unobserved heterogeneity occurs on that factor, and its moderation of the relationship from diversity to performance should lead to an effect on variability.

The worst-performing teams are likely to lack the proper information, and the task is therefore perceived as complex for them. For those teams, greater diversity improves their outcome by reducing the information penalty. The best-performing teams are likely to already possess the proper information, so greater diversity may not bring significant knowledge benefits; however, it implies the introduction of minority members, and therefore the reduction of social integration, which reduces performance. In the words of Jehn, Northcraft, and Neale, "diversity is more likely to increase workgroup performance when task are complex than routine" (1999:H6). Thereafter, greater diversity implies a more positive effect on performance for the worst teams (who gain knowledge) than for the best teams (who are penalized by reduced social integration), leading to a reduction of the spread of performance.

Hypothesis 2: In a population of teams with low intra-team diversity, greater diversity leads to lower variability of team performance.

The effect of greater diversity appears as a narrowing of the performance spread in the left-hand side of Figure 2. Overall, the diagram suggests greater performance variability when diversity goes toward its extremes. Although hypotheses 1 and 2 could also be expressed by stating that performance variability has a U-shaped relationship with diversity, separating them into two hypotheses provides three advantages. First, it reflects the distinct mechanisms operating at high and low levels of diversity. Second, separate hypotheses better match the empirical setting studied here, which allows testing only for an increase of variability at high diversity levels on some dimensions, or for a decrease at low diversity levels on other

dimensions. Finally, as explained in the next subsection, the study intends to compare variability vs. mean effects. Since this can be performed conveniently only on monotonic (i.e., linear) relationships, it is preferable to avoid curvilinear specification by identifying ranges where both variability and mean are simply linear.

One may question the compatibility of such curvilinear relationship with previous research stating that team diversity increases performance variability (Fleming, 2004; Taylor and Greve, 2006). The current study recognizes the possibility of a positive effect on variability, for high levels of diversity where social integration constitutes the main hazard. It adds up the task-knowledge fit as another hazard whose effect would be in the other direction, and perceived at low levels of diversity. The sole published study empirically linking team diversity with performance variability finds indeed only an increase of diversity (Taylor and Greve, 2006). A possible explanation is that the curvilinearity was present, but not detectable, especially given that the diversity was operationalized by counting the number of cartoon genres in which the members of cartoon creator teams had worked before. This variable was discrete (1, 2, 3, etc.), with an average of 2.56, so an effect at low levels of diversity might have been restricted by the limited number of possible values.

An alternative and possibly confounded explanation is that cartoon creator teams avoided projects into genres for which none of the members has any experience. It would imply that the cartoon creator teams were relatively immune the task-knowledge fit hazard because of a self-selection bias. Theoretically, this reasoning highlights that the current study puts the hazard of possessing the right information on an equal footing to the hazard of not getting along. However, in the field, the former might often be masked. On one hand, homogeneous teams are not likely to be assigned or choose tasks for which they are a priori not fitted; therefore that hazard is likely to be selected away. On the other hand, even though teams with diverse members can also mitigate the hazard of not getting along by assembling a priori

compatible team members, social integration of groups is a fluctuating property (Gersick, 1988). Thereafter, the later hazard is more difficult to avoid, and was therefore first documented. Empirically, this reasoning requires that the current study test its hypotheses in a setting without such possible selection bias.

Exploring the consequences of the variability effects of diversity

Previous studies predicting performance variability (e.g., Sorensen, 2002; Taylor and Greve, 2006) usually state such effect as a stand-alone conclusion. The current study intends to provide an additional conceptual step by contrasting the conclusion taking into account only the mean effect with the conclusion relying also on the variability effect. For instance, suppose that a diversity variable—such as age diversity—increased average performance. It implies that—on average—teams with high age diversity performed better than teams with low age diversity. Now, suppose also that the population of teams was characterized by low levels of age diversity. Then, hypothesis 2 predicts that the teams with lower diversity would exhibit greater performance variability. If the variability effect is strong enough—as in Figure 1.c—teams with relatively lower diversity have a higher probability of reaching a high performance threshold, although they have a lower expected performance.

When using a mean analysis only, the implications concern only the average outcome, with diversity being either beneficial or detrimental to the team. Such implication may change if diversity influences performance variability enough to change the probability of achieving a performance hurdle. For instance, diversity being beneficial on average may also decrease the high performance potential; diversity being detrimental on average may also increase the chances to remain above a low performance floor. The possibility that diversity has an opposite effect on average performance than on threshold achievement is theoretically important and will also be explored empirically below.

METHOD

The empirical analysis was designed to test the hypotheses on variability, as well as comparing mean vs. variability effects of diversity. The first objective would accommodate the classical approach of analyzing an original data set. However, the verification for the second objective becomes truly interesting if one can show how taking into account effects on variability adds up to—or even contradicts—the conclusions drawn from a classical mean analysis. To emphasize the counter-intuitive nature of the conclusions, I therefore chose to reanalyze an existing study of diversity, which also provided the benefit of reducing the presentation of the empirical setting and the mean analysis. Such parsimony in the first steps of the method section was welcome since the variability analysis and the innovative combination of effects to predict extreme outcomes implied non-standard methods whose presentation are therefore relatively lengthy.

Data

Kilduff, Angelmar, and Mehra (2000) analyzed the relationship between team performance and various cognitive and demographic diversity variables. Among its various conclusions, it appeared in that setting that diversity did not have much effect on performance, except maybe for age diversity that improved performance on average. The current paper builds on this study and complements the analysis of its data set. Below, I refer to the results of the original study (Kilduff, et al., 2000) as the *mean effect analysis* and the current analysis as the *variability analysis*.

The research setting was a MARKSTRAT business simulation, a game in which groups of players competed as management teams in a simulated market comprising five competitors. The game ran for three days during which teams made decisions on marketing, production, and R&D, and a computer determined competitive performance results measured by market share and profits. The sample consisted of 159 business executives divided into 35 teams,

which were the unit of analysis. Fourteen countries were represented, including at least 30 people from France, Germany, and Switzerland each. The managers occupied various roles in European firms, including more than 20 people each in marketing, R&D, manufacturing, and general management. The independent variables were measured early in the game, while the dependent variables were measured at the end of the game (see Kilduff, et al., 2000, for the full details of the empirical settings).

Variables

Independent variables. Demographic diversity was measured by age, nationality, and functional specialization. Age was a ratio scale, so diversity was computed through the coefficient of variation (standard deviations divided by the mean), where a low number indicated low diversity. The other measures were categorical, so their diversity was expressed by Blau's index of heterogeneity (1977), which takes values from 0 (perfect similarity of team members) to 1 (all team members different).

Team-level cognitive diversity was captured through items in a questionnaire administered early in the simulation. These items were derived from questions used by Zucker (1977) to measure cognitive variability in an institutionalization process. As in Zucker's study, participants were asked to indicate their perceptions of team processes. The diversity measures were calculated for each team, taking the coefficient of variation on the responses among members.

The current study focused on two variables, out of the six available in the original study, that best reflected cognitive diversity. Nevertheless, the analysis was also run on the other four as control, and the results section shows that those variables exhibited a pattern similar to the selected ones. The first cognitive variable, diversity in specialization perception, was measured by asking participants how specialized the team members were, on a scale from "no person has a specialized role to play" to "each person has a specialized role to play." The

second cognitive variable, diversity in power perception, was measured by asking participants how easy it would be to challenge the decision-making power of the dominant members, on a scale from "very easy to challenge the decision-making power of the dominant members" to "very hard to challenge the decision-making power of the dominant members."

The team-level coefficient of variation on these variables represented the measures of cognitive diversity. Therefore, for both the cognitive and demographic diversity independent variables, a low score showed low diversity, and a high score showed high diversity.

Determining, for each variable, the teams' population diversity level. The hypotheses stated that the effect of diversity on performance variability depends on whether the population of teams is characterized by a high or low levels of diversity. The empirical setting studied created the conditions of a natural experiment regarding the demographic diversity variables: in business education simulations, the organizers take pride in mixing people of various backgrounds, but rarely carry out this goal in a systematic way. Commonly, organizers try to mix participants based on their functional background, especially if the exercise simulates a general management team as in a MARKSTRAT simulation. Observation of the summary statistics (Table 1) suggests that functional diversity was attended to, as it was high for most teams. In contrast, teams were less varied on other demographic dimensions; since the executives' population had a limited nationality diversity with an overrepresentation of three countries, and given the consistent age of participants in most executive programs, teams had low diversity for nationality and age. In addition, the summary statistics show that the team sample was characterized by low level of cognitive diversity variables.

Therefore, the empirical setting presented a natural experiment condition on the demographic variables and an empirically clear situation on the cognitive variables. For each variable, the

sample of teams was either at low or high diversity levels, which allowed for testing one of the two linear hypotheses (either H1 or H2); this setting was convenient since the small sample size (N=35) would not allow testing of curvilinearity. The variables for which the teams were at low diversity levels—age, nationality, perception of power, and specialization—allowed testing of hypothesis 2; functional diversity, for which the teams were at high levels, allowed testing of hypothesis 1. The effect of diversity was assumed to be locally linear, respectively reducing or increasing performance variability.

Control variables. Since the current study builds on an existing mean effect analysis (Kilduff, et al., 2000), I kept the control variables of the previous study. Therefore, the size of the team and a measure of starting position—each team in the simulation did not start with the same market share—were included. Furthermore, all variables identified in the original study as relevant independent variables, but not used here as independent variables, appeared as control variables. These included the diversity in perception of ambiguity, decision difficulty, decision pressure, and effectiveness (see Kilduff, et al., 2000, for more details on those variables).

Dependent variables. Two performance measures—Final Market Share (FMS), expressed as a percentage, and Cumulated Net Marketing Contribution (CNMC)—reflected outcomes. These were the dependent variables of the mean analysis. Two different approaches were used to detect variability effect, and therefore variability was operationalized in two different ways accordingly, as exposed below.

Specifying the effects on performance variability

To test the hypotheses about the effects of diversity on performance variability, I first performed an Ordinary Least Square (OLS) multiple regression for both performance measures to determine the mean effects. One could object that since theory assumes a relationship between an independent variable (diversity) and the variability of a dependent

variable (performance), the implied heteroskedasticity could be problematic for OLS. I nevertheless chose to report these results instead of those drawn from a more robust method to preserve comparability with the original study (Kilduff, et al., 2000), which used OLS. Moreover, heteroskedasticity does not bias the directions of the conclusions or the value of the residual; it only reduces the efficiency of the estimates. In fact, Greene (2003:222) recommended using OLS for tests of heteroskedasticity since it provides, conveniently and without bias, the residuals of performance used in the various variance analysis methods.

White's heteroskedasticity test cannot help in hypothesis testing since it does not indicate the direction of the heteroskedasticity, only its existence. A first approach to detecting an effect of a variable on the variability is to consider the absolute value of the regression residual, and test for an effect of diversity using a regression (similar to Taylor and Greve, 2006:732). However, since a residual—the error term of a regression—should have a normal distribution centered at zero, its absolute value could not have a normal distribution. This introduced a misspecification that could bias the estimations, calling for further analysis. However, I reported the results of this intuitive method because they provided a first estimation of the direction and potential significance of the variability effect.[1]

The search for a test designed to determine the direction of heteroskedasticity, even with a small sample, led to the Goldfeld-Quandt test (the second heteroskedasticity test suggested by Greene, 2003). This groupwise test provided the statistical significance and direction of the heteroskedasticity without making any parametric assumption, and applied separately to each independent variable. It split the population of teams into two sub-samples, the bottom sub-sample (B) containing the n teams with the lowest value of the focal independent variable, and the top sub-sample (T) containing the n teams with the highest value. For each sub-

[1] A related unbiased approach avoid absolute value by parametrizing the dispersion of residual (e.g., Sorenson and Sorensen, 2001; Sorensen, 2002), and estimates it using a Maximum Likelihood Estimation (MLE). The small sample size (N=35) prevented its use here.

sample, I performed an OLS regression with all the independent variables. I then calculated the sum of the squares of the residuals for each sub-sample, providing a variance for each. The ratio of these two variances indicate the direction of the variability effect, with its significance indicated by an F-test. This approach is similar to an ANOVA, except that the sum of squares came from the residual of the regression in each subsample, rather than from the difference with the mean. Greene (2003:223) suggested using the following ratio:

$$r = \frac{e'_T \cdot e_T}{\mu_T{}^2 (n_T - K)} / \frac{e'_B \cdot e_B}{\mu_B{}^2 (n_B - K)},$$

where B notes the bottom sub-sample, T the top sub-sample, e the residual vector, μ_B and μ_T the means of Y in each sub-sample, n_B and n_T the number of teams in each sub-sample, N the total number of teams ($N > 2n_T = 2n_T$), and K the number of variables in the regression. I calculated the ratio for each independent variable, each calculation differing only by the different sorting of the teams into two sub-samples according to the values of that focal variable.

Here is a summary of the logic when testing, for example, for the effect of age diversity on performance variability. I sorted the 35 teams according to age diversity, with the bottom sub-sample (B) gathering the 17 teams with the lowest age diversity and the top sub-sample (T) the 17 teams with the highest age diversity[2]. For each sub-sample, I regressed performance on all variables, which provided a residual vector per sub-sample. The ratio r provided the direction of the effect, with $r > 1$ showing a larger variability in the top sub-sample, and $r < 1$ a larger variability in the bottom sub-sample. An F-test—on the ratio if $r > 1$ and on the inverse of the ratio if $r < 1$—measured significance.

[2] Greene (2003) suggested that the actual size of the sub-sample could be varied to search significance. Here, the greatest possible sub-samples (n=17) is preferable given the restricted size of the population (N=35).

Method to separate mean and variability effects

Once the effect of a diversity variable on performance variability was established, how could this change the predictions of the mean effect analysis? No simple criterion appears in the organizational theory literature to combine variability and mean effects to predict the attainment of threshold of performance. Here, I expose the logic for such a criterion, with the formal proofs and assumptions of such reasoning presented in the appendix.

Identifying and defining the criterion. Consider a situation where performance Y depends on factor X, and the context dictates to reach a performance threshold Y_0. The question to answer is whether X increases or decreases the probability of reaching this threshold. It complements the question answered by conventional regressions, whether X increases or decreases the mean value of Y. The theory section above provided intuition into why the mean effect alone might not predict the chances of reaching such a threshold.

To establish a criterion, let us use again the example represented by the sketches of Figure 1, where the factor X has a positive effect on the mean of performance Y and a negative effect on its variability. In the case of constant variability (Figure 1.b), the mean effect drives the chances of reaching the threshold Y_0. In the case where variability changes (Figure 1.c), estimating the chances of reaching the threshold is impossible without considering the distribution: the bell curve visually indicates the probability of reaching the threshold (i.e., the tail of the distribution above the threshold).

Now, imagine replacing the bell curves by drawing the lines that link—across different values of X—the points where performance is equally likely, those being traditionally called the quantile lines. An obvious line is the median line—at the 50 percent quantile—that roughly approximates the regression line. Adding a few other quantile lines results in Figure 3—a more detailed version of Figure 1.c. Now, one can simply read the probabilities—expressed by the quantile lines—of reaching the threshold. Take for example the position of

the threshold Y_0 relative to the 90 percent quantile line; it shows that reaching the threshold has more than a 10 percent chance of occurring for bottom values of the factor and less than a 10 percent chance for top values. One can therefore conclude that the factor decreases the chances to reach the threshold.

---------- Insert Figure 3 around here ----------

Determining the effect of the factor X on the probability of reaching the threshold amounts simply to reading the slope of the quantile line that crosses it. If the quantile lines that cross the threshold have a positive slope, the chances of reaching the threshold grow with X; if the quantile lines have a negative slope at the threshold, the chances of reaching the threshold diminish with X. Therefore, the direction of the effect corresponds to the slope of the quantile lines that cross it: a threshold crossed by a positive slope quantile line indicates a positive effect of the antecedent.

The diagram suggests that the slopes of the quantile lines change direction once. The quantile at 50 percent has a positive slope because X has a positive effect on the average of Y; when going up the diagram, the quantile lines progressively flatten because of the higher variability of performance for low values of X. At some point, the quantile line becomes horizontal—at a level of performance that is therefore equally likely for any value of X and that we will call the *critical performance level* (Y_c). Beyond that level, the quantile lines have a negative slope.

The important consequence is an inversion of the direction of the effect: for thresholds around average values of performance, the direction is dictated by the mean effect; for thresholds beyond the critical level, the threshold achievement effect is opposite the mean effect, driven in the other direction by the effect on variability. In the presence of a variability effect, one would therefore want to find at which level this inversion occurs, since it separates

the values of performance threshold where the factor has an effect dictated by mean analysis from those where its effect is in the opposite direction.

Computation of critical performance level in a groupwise setting. In the current setting, the limited sample size led to the computation of the variability effect using the Goldfeld-Quandt test. Since the estimation assumed that the population of teams was grouped into two sub-samples, these could be aggregated as two points, for which a mean and variability of performance could be computed.

For each sub-sample respectively, I denote the performance as the random variables Y_B and Y_T, with mean values μ_B and μ_T and standard deviation σ_B and σ_T. By definition, the critical performance occurs at level Y_c, where the cumulated probabilities are equal in the two sub-samples:

$$P[Y_B>Y_c] = P[Y_T>Y_c] \Longrightarrow F\left(\frac{\mu_B - Yc}{\sigma_B}\right) = F\left(\frac{\mu_T - Yc}{\sigma_T}\right)$$

Given that the cumulated probability function F is monotone, the equality implies that the ratios inside F have to be equal, which leads simply to the formula:

$$Y_c = (\frac{\mu_T}{\sigma_T} - \frac{\mu_B}{\sigma_B})/(\frac{1}{\sigma_T} - \frac{1}{\sigma_B})$$

Regarding the critical risk r_c attached to that critical performance level, its calculation takes the percentile of Y_c in the distribution of Y, with a condition. If the critical level appears at a low cumulated probability (below the median line), the probability of interest is the probability of a bad performance induced by variability, so the left tail of distribution. If the critical level appears at a high probability (above the median line), the probability of interest is the chance of a good performance, so the right tail of the distribution. In the example of Figure 3, the mean effect dominates in most of the diagram; the variability dominates only in a small fraction (the top 18 percent) of the distribution.

The hypotheses state that for low diversity levels, greater diversity decreases variability, and for high diversity levels, it increases variability. In the current empirical setting, for each variable, the teams' population was taken as entirely in either high or low diversity levels so the variability could be assumed linear. Hence, for each of the variables, I could estimate the mean and variability effect, which allowed calculation of the critical performance level beyond which the prediction of performance runs counter to the mean effect conclusion.

RESULTS

The current study builds on some of the conclusions of the mean effects analysis performed by Kilduff, Angelmar, and Mehra (2000). I first summarize the part of their results that interests us: it appears that age diversity significantly increased mean performance, whereas the other four diversity variables had no significant mean effect. I do not discuss the mean analysis here, instead focusing on the variability analysis, and on how it may complement the mean analysis.

The summary statistics and correlations for all variables appear in Table 1. To simplify cross-referencing with the original study, I present the variables in the same order. Therefore, I highlight our five independent variables in the table by putting their names in bold, while the other variables act as controls. Summary statistics identify diversity in nationality (mean=0.37, SD=0.28), age (mean=0.13, SD=0.07), perception of specialization (mean=0.43, SD=0.26), and perception of power (mean=0.38, SD=0.16) as the low-diversity variables. On the other hand, functional diversity (mean=0.63, SD=0.12) was a high-diversity variable. These results confirmed the design explained in the Method section, whereby for each of the variables, the teams' population is either in low or in high levels of diversity. Therefore, I could separately test hypothesis 1 on the low-diversity variables and hypothesis 2 on the high-diversity variable.

---------- Insert Table 1 roughly here ----------

Finding Variability Effects

The regressions and heteroskedasticity tests appear together in Table 2. Models 1a and 1b are the mean effect regressions, with the dependent variable being either Final Market Share (FMS) or Cumulative Net Marketing Contribution (CNMC). Models 2a and 2b are the regressions of the absolute value of the residual of the mean effect regressions for FMS and CNMC, respectively. In models 3a and 3b, I report the ratio calculated by the Goldfeld-Quandt method. If $r > 1$, it marks a growth of variability with the considered variable; and if $r < 1$, it marks a decrease. An F-test on this ratio of variances gives the significance of the effect.

---------- Insert Table 2 roughly here ----------

Hypothesis 1. This hypothesis predict that functional diversity should increase the variability of performance since the teams were at high levels of diversity. It should appear in model 2 as a positive coefficient, and in the Goldfeld-Quandt test as a ratio $r \geq 1$. The effect was in the right direction and significant for the regression of the absolute residual of FMS (deviation 0.09, $p<0.1$), and in the right direction but not significant for the Goldfeld-Quandt on FMS, and the absolute residual and the Goldfeld-Quandt of CNMC. Hypothesis 1 is supported.

Hypothesis 2. This hypothesis predicts that diversity in nationality, age, perception of specialization, and power should decrease variability since the teams' population was at low diversity levels for those variables. This result should appear in model 2 (regression of the absolute residual) as a negative coefficient, and in the Goldfeld-Quandt test (ratio of variances between two sub-samples, in columns 3a and 3b) as a ratio $r \leq 1$. The results showed a pattern of confirmation for diversity in nationality (for FMS: deviation 0.04, $p<0.1$; Goldfeld-Quandt 0.14, $p<0.05$. For CNMC: GQ 0.1, $p<0.01$), age (for FMS: deviation 0.18, $p<0.05$; for GQ: 0.28, $p<0.1$), specialization perception (for CNMC: –0.31, $p<0.1$; GQ 0.21,

p<0.05), and power perception (for FMS: GQ 0.22, p<0.1. For CNMC: GQ 0.17, p<0.05). Note that the other diversity variables we did not focus on here—but were nevertheless entered only as controls—exhibited the same pattern, with effects in the proper direction in the cases where they were significant. Overall, hypothesis 2 is supported.

The results supported the hypotheses linking diversity to performance variability, increasing it at high diversity levels and decreasing it at low diversity levels. One may nevertheless wonder how it matters.

Exploratory combination of mean and variability effects

Kilduff, Angelmar, and Mehra (2000) showed that age diversity increases performance, which appeared here too (FMS: 0.53, p<0.05; CNMC: 640, p<0.05). Classically, that seems to imply that higher diversity increases performance expectation, which is then interpreted as age diversity being beneficial. Would taking into account variability effects provide nuances to that conclusion, especially if the performance goal was not simply to improve mean performance, but to reach a threshold?

The Goldfeld-Quandt test split the population of teams into two sub-samples, grouping bottom and top values for each independent variable. Table 3 reports the mean and standard deviation of performance in the bottom and the top sub-samples, as well as the critical performance level and the attached critical risk (see the Method section for definitions and meanings of such computations).

---------- Insert Table 3 roughly here ----------

To interpret these results, I built graphs summarizing them, considering only the dependent variable FMS (Figure 4). Each of the five graphs synthesizes the results for one of the five independent diversity variables, and uses the representation introduced in Figure 3. For each independent variable, the two sub-samples partition the teams, those with low diversity in the bottom sub-sample (B) and those with high diversity in the top sub-sample (T). A circle

positions the mean performance level in each sub-sample. The line joining these two circles symbolizes the mean effect. The size of each circle represents the performance variability—as measured by standard deviation—in each sub-sample. The dotted line represents the critical performance level, where inversion of the effect of each variable occurs, legending it with the size of the associated critical risk.

---------- Insert Figure 4 roughly here ----------

I now detail the reasoning for one independent variable, age diversity, and for one performance measure, Final Market Share (FMS). Let us first consider mean effect. Teams in the top sub-sample reached, on average, 21 percent of FMS, while teams in the bottom sub-sample could expect only 19 percent of FMS, a positive mean effect of age diversity. The mean effect analysis therefore suggests that teams benefit from being in relatively high diversity.

A more nuanced conclusion appears if we consider performance variability. Diversity in age decreased performance variability, with a performance standard deviation of 5.2 percent FMS and 2.7 percent FMS in the bottom and the top sub-samples, respectively. The strength of the effect on variability puts the critical performance at 23 percent FMS, associated with a cumulated probability of 82 percent (critical risk is therefore the top 18 percent). A first interpretation implies that, when the goal of participants is to reach any performance beyond 23 percent FMS, teams in the bottom sub-sample have a higher probability of succeeding than teams in the top sub-sample. This conclusion held even though teams in the bottom sub-sample had a lower expected performance. It shows that, depending on the performance goal, the preferred team composition changed, and the normative implications of the mean analysis does not always hold.

To clarify the concept of critical risk, let us take another example of a high performance goal, this time expressed as a ranking: let us assume for instance that the goal was to belong to the top 10 percent of the teams, also expressed as being among the top three teams (since there were 35 teams). Given that the top 10 percent of performance was included in the top 18 percent (the critical risk zone), the analysis suggests that teams in the lower diversity condition would have more chances of reaching that goal. To clarify how that prediction relates to the empirics, Figure 5 presents a scatter-plot of all 35 teams, grouped in their respective sub-samples and showing their rank in the overall exercise, from 1 to 35. The lower sub-sample captures positions 2 and 3, illustrating the prediction that low age diversity is preferable in the presence of a performance threshold such as belonging to the top three teams[3]. If checking the goal of belonging to the top 5 (top 13%), the bottom sub-sample again dominates by having 3 in the top 5, while the top sub-sample has only 2.

---------- Figure 5 roughly here ----------

All those results contradict the mean analysis prediction on the superiority of higher diversity. One should notice that the bottom sub-sample also captures the worst performers—teams ranked 35,33, 32, 31, etc.—a conclusion over-determined by the combination of both mean and variability effects. Overall, lower diversity implies more extreme outcomes, both positive and negative, with the variability effect completely overcoming the mean effect in that setting.

Regarding the other variables, the mean effect analysis found no significant effect. The variability analysis showed that increasing diversity in nationality, power perception, and specialization perceptions (these factors being empirically low in this sample) reduced

[3] One should note that the first position was captured by the top sub-sample, this point appearing as an outlier of the variability analysis. Had the data been without such noise, the top three performers would have been in the lower diversity condition, reinforcing the idea that this condition leads to an extremely positive outcome even though the higher diversity condition leads to better outcome on average.

performance variability. Therefore, diversity on those factors simply decreased risk: greater diversity made both extremely good and bad outcomes less likely. Functional diversity (empirically high in this sample) increased variability but had no significant effect on mean, therefore simply increasing risk.

Overall, considering the effects of diversity on performance variability improves predictions compared with mean analysis by finding significant effects on variability for four variables for which no mean effect existed. Theorizing effects on variability therefore allows predictions of extreme outcomes in a situation where classic theory did not allow any prediction because there was simply no mean effect. For the fifth variable (age), and for a competitive goal such as belonging to any top fraction smaller than the top 18 percent, variability analysis contradicted mean analysis predictions that higher age diversity is preferable.

DISCUSSION

This study aims at two contributions. First, it shows that demographic and cognitive diversity have a U-shaped relationship to performance variability. This should help resolve contradictions in theories of diversity. Second, it shows, by using a quantifiable criterion, that variability effects lead to more nuanced conclusions about diversity than mean effect alone. This should inform the various organizational perspectives where extreme outcome, either high or low, play a particular role.

Two contradictions occur among theories of diversity. Those considering the effect of diversity on mean performance struggle between a pessimistic perspective—diversity decreases social integration, and therefore performance—and an optimistic perspective—diversity improves available knowledge, and therefore performance. Such a contradiction underlies the non-significant effect of team diversity on team performance. However, introducing performance variability allows to develop hypotheses that predict both

outstandingly good and bad outcomes even in the absence of mean effect. In the empirical analysis, four diversity variables had no mean effect on performance (Kilduff, et al., 2000), yet exhibited a significant effect on risk. A second contradiction appeared in previous studies considering the effect of diversity on performance variability. A knowledge perspective suggests that diversity creates risk (Fleming, 2001; Fleming, 2004; Taylor and Greve, 2006), while a socio-psychological perspective suggests that similarity creates risk (Schachter, et al., 1951; Janis, 1971; Luthans, 2002). I reconcile such views by proposing a curvilinear relationship, where risk appears at both extremes—high and low—of the diversity scales.

The variability analysis proposed above informs organizational theories perspectives studying contexts that either reward or punish extreme outcomes. Echoing the call to arms voiced by some scholars to expand our focus beyond effects on the mean (Daft and Lewin, 1990; Starbuck, 1993), it could generally be applied to any context where a variability effect occurs and the organizational goal can be expressed as reaching a threshold. In such contexts, cognitive and demographic diversity might then increase the chances of survival, even when these factors degrade mean performance.

Consider contexts where low outcomes have disastrous consequences, such as accidents in high-reliability organizations (HRO), or fiascos in governance and corporate social responsibility literature. This paper shows that although diversity might have no mean effect, it could still influence risk, and thus the chances of catastrophic outcomes. More important, it shows that one may find a mean effect of team diversity in one direction and an opposite effect on reaching an extreme threshold of performance. The possibility that the effect of team diversity on average performance may reverse regarding accidents should be emphasized, since it may require counterintuitive team composition. Regarding fiascos, this study suggest to revisit the concept of groupthink (1971; 1982), which was expected to predict catastrophes but for which little support was found linking concurrence seeking or

cohesion to performance (Nemeth and Staw, 1989; Esser, 1998). Here, I show how homogeneous teams are more likely to results in catastrophic outcomes not through a mean effect but through a variability effect.

Other contexts force organizations to take risks by rewarding those reaching high thresholds of performance, or by letting only a few winners of a contest survive. In entrepreneurial settings, many high and rare performance outcomes are salient, such as being the first firm to be profitable, or to benefit from network effects, or to complete an initial public offering (IPO), etc. Those lead to survival, suggesting that effects on variability may matter more than mean effect (March, 1991). Other organizational situations may reward particular positive outcomes, such as in the case of real options (Kogut, 1991) where the organization invests in projects with the hope of reaping high rewards. In all those cases, the prospects could be improved by manipulating team composition, with potentially greater chances of exceptionally positive outcomes for both highly homogeneous and highly diverse teams.

The long tradition of studying the effect of diversity on performance has been frustrated by its contradictory effects. In the meantime, practitioners could not wait for full conclusions, and pushed strongly to develop organizational diversity. In a world where risk—industrial, governance, entrepreneurial, etc.—increasingly matter to organizations, establishing that both extremes of team diversity scale entail increased chances of both positive and negative extreme outcomes warrants further study.

APPENDIX: EXISTENCE OF CRITICAL PERFORMANCE LEVEL

I assume that the performance Y is a function of X through a cumulated probability function F that depends simply on the z-score of Y, with its first two moments, mean $\mu(X)$ and standard deviation $\sigma(X)$, being linear on X:

Equation 1: $Y_X \sim F(z_X)$ with $z_X(Y)=(Y-\mu_X) / \sigma_X$ and $\mu_X = \beta_1 X + \beta_0$ and $\sigma_X = \gamma_1 X + \gamma_0$.

This modeling accommodates many of the distributions used in organizational research. It generalizes the classic approach using a simple regression, $Y : N(\beta_1 X + \beta_0, \sigma_0)$, where N is the normal distribution and one assumes homoskedasticity. Furthermore, such parameterization can be viewed as a linearization of more complex parameterization, modeled only by the first-order effects—on both the mean and the standard deviation—and ignoring all higher-order effects (quadratic, cubic, etc.). Overall, assuming the distribution is not too exotic (e.g., not power laws) and the model can be linearized (at least locally), the conclusion of the current study holds.

Estimation of the parameters in Equation 1 can be straightforward (e.g., Sorensen, 2002); however, the interpretation of signs is less clear. Traditionally, one seeks whether X increases Y, so the sign of β_1 is paramount. With a positive β_1, one assumes that X has a positive impact on performance.

When introducing the effect of the residual variability measured by γ_1, the problem becomes more complex. If the mean effect of X is positive but also reduces variability, what can we conclude? For example, the exploration–exploitation study (March, 1991) explored the outcomes using a simulation but did not provide a constructive approach usable in other studies. Let us analytically explore the question of whether X improves the chance p(X) of reaching a threshold Y_0. The following reasoning builds on a logic exposed by Tsetlin, Gaba, and Winkler (2004). Given the assumptions that the cumulative distribution function F depends only on the z-score, we have:

$$\text{Equation 2: } \mathrm{p(X)} = \mathrm{P}[\mathrm{Y_X} > \mathrm{Y_0}] = \mathrm{F}\left(\frac{\mu(X) - \mathrm{Y_0}}{\sigma(X)}\right) = \mathrm{F}\left(\frac{\beta_1 \mathrm{X} + \beta_0 - \mathrm{Y_0}}{\gamma_1 \mathrm{X} + \gamma_0}\right)$$

The quantile curves, linking points of equal probability, are defined by the equations taking p(X) as a constant. F is a cumulated distribution function and is therefore monotone; hence, quantile curves are defined by making the ratio inside F to be constant, leading to a linear

equation; therefore, those curves are simple lines, justifying the representation of Figure 3. To determine at which level such line is horizontal, one has simply to explore for which values of performance the ratio inside F is constant with X, and therefore has a null derivative. If the derivative is taken and made 0, it solves in Y_0, providing the critical value, Y_c:

$$\text{Equation 3: } Y_c = \beta_0 - \gamma_0 \frac{\beta_1}{\gamma_1}$$

This shows that the quantile lines change direction only once, at the critical performance level Y_c. Around average performance, all quantile lines have the same slope direction and invert on the other side of the critical level (inversion occurs in the grayed area of Figure 3).

REFERENCES

Baum, Joel A. C., and Bill McKelvey

2006 "Analysis of Extremes in Management Studies." In D. J. Ketchen, and D. D. Bergh (eds.), Research Methodology in strategy and management: 125-199.

Blau, Peter Michael

1977 Inequality and heterogeneity: a primitive theory of social structure. New York: Free Press.

Boone, Christophe, Woody Van Olffen, and Arjen Van Witteloostuijn

2005 "Team locus-of-control composition, leadership structure, information acquisition, and financial performance: a business simulation study." Academy of Management Journal, 48: 889-909.

Bowers, Clint A.

2000 "When Member Homogeneity Is Needed in Work Teams." Small Group Research, 31: 305.

Bunderson, J. Stuart, and Kathleen M. Sutcliffe

2002 "Comparing Alternative Conceptualizations of functional diversity in management teams: process and performance effects." Academy of Management Journal, 45: 875.

Cabral, Luis M. B.

2003 "R&D Competition When Firms Choose Variance." Journal of Economics & Management Strategy, 12: 139-150.

Cramton, Catherine Durnell, and Pamela J. Hinds

2005 "Subgroup dynamics in internationally distributed teams: Ethnocentrism or cross-national learning?" In B. M. Staw, and R. M. Kramer (eds.), Research in organizational behavior: 231-263: Elsevier Science/JAI Press.

Daft, Richard L., and Arie Y. Lewin

1990 "Can organizational studies begin to break out of the normal science straitjacket? An editorial essay." Organization Science, 1: 1.

Dahlin, Kristina B., Laurie R. Weingart, and Pamela J. Hinds

2005 "Team diversity and information use." Academy of Management Journal, 48: 1107-1123.

Denrell, Jerker

2003 "Vicarious learning, undersampling of failure, and the myths of management." Organization Science, 14: 227-243.

Esser, James K.

1998 "Alive and Well after 25 Years: A Review of Groupthink Research." Organizational Behavior & Human Decision Processes, 73: 116-141.

Festinger, Leon

1954 "A Theory of Social Comparison Processes." Human Relations, 7: 117-140.

Fleming, Lee

2001 "Recombinant uncertainty in technological search." Management Science, 47: 117-132.

2004 "Perfecting Cross-Pollination." Harvard Business Review, 82: 22-24.

Gersick, Connie J.G.

1988 "Time and Transition in Work Teams: Toward a New Model of Group Development." Academy of Management Journal, 31: 9-41.

Gibson, Cristina, and Freek Vermeulen

2003 "A Healthy Divide: Subgroups as a Stimulus for Team Learning Behavior." Administrative Science Quarterly: 202-239: Administrative Science Quarterly.

Greene, William H.

2003 Econometric analysis, 5th ed. Upper Saddle River, N.J.: Prentice Hall.

Gruenfeld, Deborah H., et al.

1996 "Group composition and decision making: How member familiarity and information distribution affect process and performance." Organizational Behavior & Human Decision Processes, 67: 1-15.

Harrison, David A., and Katherine J. Klein

2007 "What's the difference? Diversity constructs as separation, variety, or disparity in organizations." Academy of Management Review, 2007-Forthcoming.

Jackson, Susan E., Aparna Joshi, and Niclas L. Erhardt

2003 "Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications." Journal of Management, 29: 801.

Janis, Irving Lester

1971 "Groupthink." Psychology Today: 43-46.

1982 Groupthink: psychological studies of policy decisions and fiascoes, 2nd ed. Boston: Houghton Mifflin.

Jehn, Karen A., Gregory B. Northcraft, and Margaret A. Neale

1999 "Why differences make a difference: A field study of diversity, conflict, and performance in workgroups." Administrative Science Quarterly, 44: 741-763,.

Kilduff, Martin, Reinhard Angelmar, and Ajay Mehra

2000 "Top Management-Team Diversity and Firm Performance: Examining the Role of Cognitions." Organization Science: 21: INFORMS: Institute for Operations Research.

Kochan, Thomas, et al.

2003 "The Effects of Diversity on Business Performance: Report of the Diversity Research Network." Human Resource Management, 42: 3.

Kogut, Bruce

1991 "Joint ventures and the option to expand and acquire." Management Science, 37: 19-33.

Lau, Dora C., and J. Keith Murnighan

1998 "Demographic diversity and faultlines: the compositional dynamics of organizational groups." Academy of Management Review, 23: 325-340.

Luthans, Fred

2002 Organizational behavior, 9th ed. Boston: McGraw-Hill/Irwin.

Mannix, Elizabeth A., and Margaret A. Neale

2005 "What Differences Make a Difference?" Psychological Science in the Public Interest, 6: 31-55.

March, James G.

1991 "Exploration and Exploitation in Organizational Learning." Organization Science, 2: 71.

Milliken, Frances J., and Luis L. Martins

1996 "Searching for common threads: understanding the multiple effects of diversity in organizational groups." Academy of Management Review, 21: 402-433.

Miner, Anne S., Pamela R. Haunschild, and Andreas Schwab

2003 "Experience and convergence: curiosities and speculation." Industrial & Corporate Change, 12: 789-813.

Nemeth, Charlan J.

1986 "Differential contributions of majority and minority influence." Psychological Review, 93: 23-32.

Nemeth, Charlan Jeanne, and Barry M. Staw

1989 "The tradeoffs of social control and innovation in groups and organizations." In L. Berkowitz (ed.), Advances in experimental social psychology: 175-210: Academic Press, Inc.

O'Reilly, C.A., David F. Caldwell, and W.P. Barnett

1989 "Work Group Demography, Social Integration, and Turnover." Administrative Science Quarterly, 34: 21-37.

O'Reilly, Charles A. III, Katherine Y. Williams, and Sigal Barsade

1998 "Group demography and innovation: Does diversity help?" In D. H. Gruenfeld (ed.), Composition.: 183-207: Elsevier Science/JAI Press.

Pfeffer, J.

1983 "Organizational demography." Research in Organizational Behavior, 5: 299.

Polzer, Jeffrey T., Laurie P. Milton, and William B. Jr. Swann

2002 "Capitalizing on diversity: Interpersonal congruence in small work groups." Administrative Science Quarterly, 47: 296-324.

Richard, Orlando C., et al.

2004 "Cultural Diversity in Management, Firm Performance, and the Moderating Role of Entrepreneurial Orientation Dimensions." Academy of Management Journal, 47: 255.

Robinson, W. S.

1950 "Ecological Correlations and the Behavior of Individuals." American Sociological Review, 15: 351.

Schachter, Stanley, et al.

1951 "An experimental study of cohesiveness and productivity." Human Relations, 4: 229-238.

Sommers, Samuel R.

2006 "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." Journal of Personality and Social Psychology, 90: 597-612.

Sorensen, Jesper B.

2002 "The Strength of Corporate Culture and the Reliability of Firm Performance." Administrative Science Quarterly, 47: 70.

Sorenson, Olav, and Jesper B. Sorensen

2001 "Finding the right mix: franchising, organizational learning, and chain performance." Strategic Management Journal, 22: 713.

Starbuck, William H.

1993 "Keeping a butterfly and an elephant in a house of cards: the elements of exceptional success." Journal of Management Studies, 30: 885-921.

Stasser, G., and W. Titus

1987 "Effects of Information Load and Percentage of Shared Information on the Dissemination of Unshared Information During Group Discussion." Journal of Personality & Social Psychology, 53: 81-93.

Tajfel, Henri

1982 Social identity and intergroup relations. Cambridge: Cambridge University Press.

Taylor, Alva, and Henrich R. Greve

2006 "Superman or the fantastic four? Knowledge recombination and experience in innovative teams." Academy of Management Journal, 49: 723-740.

Tsetlin, Ilia, Anil Gaba, and Robert L. Winkler

2004 "Strategic Choice of Variability in Multiround Contests and Contests with Handicaps." Journal of Risk & Uncertainty, 29: 143-158.

Uzzi, Brian, and Jarrett Spiro

2005 "Collaboration and Creativity: The Small World Problem." American Journal of Sociology, 111: 447.

Van Der Vegt, Gerben S., and J. Stuart Bunderson

2005 "Learning and performance in multidisciplinary teams: the importance of collective team identification." Academy of Management Journal, 48: 532-547.

van Knippenberg, Daan, Carsten K. W. De Dreu, and Astrid C. Homan

2004 "Work Group Diversity and Group Performance: An Integrative Model and Research Agenda." Journal of Applied Psychology, 89: 1008-1022.

van Knippenberg, Daan, and Michaela C. Schippers

2007 "Work Group Diversity." Annual Review of Psychology, 58: 515-541.

Webber, Sheila Simsarian, and Lisa M. Donahue

2001 "Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis." Journal of Management, 27: 141.

Williams, Katherine Y., and Charles A. O'Reilly, III

1998 "Demography and Diversity in Organizations: A Review of 40 Years of Research." In B. M. Staw, and L. L. Cummings (eds.), Research in Organizational Behavior: 77-140. London: JAI Press Inc.

Zucker, Lynne G.

1977 "The role of institutionalization in cultural persistence." American Sociological Review, 42: 726-743.

Table 1: Summary Statistics and Correlations

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Market Share Change	0.00	0.07														
Early NMC (a) Change	1.84	0.35	-0.21													
Final Market Share	0.20	0.09	0.67***	0.38*												
CNMC (b)	267.36	137.40	0.37*	0.56***	0.88***											
Size	4.23	0.73	0.00	-0.21	-0.04	-0.15										
Starting Point	1.60	0.50	-0.20	0.69***	0.46**	0.65***	-0.06									
Demographic Diversity																
National	0.37	0.28	0.34*	-0.05	0.16	0.15	-0.40*	-0.15								
Functional	0.63	0.12	0.07	-0.31+	-0.01	-0.15	0.31+	-0.05	-0.24							
Age	0.13	0.07	0.30+	-0.17	0.25	0.16	-0.08	-0.14	0.21	-0.09						
Cognitive Diversity																
Specialization	0.43	0.26	0.19	-0.12	0.16	0.13	-0.06	0.08	0.23	-0.05	0.19					
Power	0.38	0.16	-0.01	-0.18	0.01	0.03	-0.10	-0.03	0.13	0.09	0.24	0.11				
Ambiguity	0.16	0.13	0.25	0.27	0.45**	0.50**	-0.22	0.30+	0.18	-0.28	-0.07	0.09	-0.26			
Decision Difficulty	0.25	0.13	0.05	-0.31+	-0.24	-0.33+	0.37*	-0.32+	-0.09	0.28	0.16	-0.13	0.07	-0.09		
Decision Pressure	0.36	0.21	0.19	-0.28	-0.12	-0.31+	0.16	-0.39*	0.13	0.13	0.20	-0.12	0.06	-0.10	0.36*	
Effectiveness	0.37	0.19	0.29+	-0.42*	0.06	0.00	0.03	-0.26	0.16	0.26	-0.04	0.03	0.31+	0.05	0.33+	0.11

N=35 firms; a: Net Marketing Contribution Changes; b: Cumulative Net Marketing Contribution

+: p<0.1 *: p<0.05 **: p<0.01 ***: p<0.001

Table 2: Analyses

Performance Variable:	Final Market Share (FMS)			Cumulative Net Marketing Contribution (CNMC)		
	1a	2a	3a	1b	2b	3b
Model:	Regression	Regression	Goldfeld-Quant	Regression	Regression	Goldfeld-Quant
Dependent Variable:	FMS(=Y)	abs(res(Y))		CNMC(=Y)	abs(res(Y)/Yhat)	
Size	0.03	0.01	1.10	24.69	0.09	1.92
	(0.02)	(0.01)		(26.60)	(0.09)	
Starting Position	0.08*	0.01	2.22	157.09**	-0.21+	2.13
	(0.03)	(0.02)		(40.02)	(0.13)	
National Diversity	0.05	-0.04+	0.14*	67.48	-0.18	0.10**
	(0.05)	(0.03)		(67.86)	(0.22)	
Functional Diversity	0.11	0.09+	2.70	19.76	0.12	1.23
	(0.12)	(0.06)		(151.54)	(0.48)	
Age Diversity	0.53**	0.18*	0.28+	640.32**	-0.56	0.47
	(0.21)	(0.10)		(255.13)	(0.82)	
Specialization Perc. Div.	-0.01	-0.03	0.36	-42.47	-0.31+	0.21*
	(0.05)	(0.03)		(66.13)	(0.21)	
Power Perc. Div.	-0.02	-0.06	0.22+	10.60	-0.22	0.17*
	(0.10)	(0.05)		(121.77)	(0.39)	
Ambiguity Perc. Div.	0.27*	0.02	0.77	361.32*	-0.10	0.39*
	(0.13)	(0.06)		(155.77)	(0.50)	
Decision Difficulty Perc. Div	-0.23*	-0.08	0.09*	-275.04*	-0.60	1.82
	(0.12)	(0.06)		(154.86)	(0.50)	
Decision Pressure Perc. Div.	0.01	-0.02	0.63	-62.82	0.38+	0.65
	(0.07)	(0.03)		(87.20)	(0.28)	
Effectiveness Perc. Div.	0.10	0.05	2.03	155.38+	-0.32	0.72
	(0.09)	(0.04)		(106.11)	(0.34)	

Note: N = 35; unstandardized betas (with standard errors in parentheses)
Note: res(Y) notes the residual of the main regression on Y; Yhat notes the predicted value of that regression
+: p <0.1 *: p <0.05 **: p <0.01

Table 3: Critical Performance and Probability

Performance Variable:	Final Market Share (FMS)						Cumulative Net Marketing Contribution (CNMC)					
	Bottom Bin		Top Bin		Critical		Bottom Bin		Top Bin		Critical	
	Mean	S.D.	Mean	S.D.	Perf.	Risk	Mean	S.D.	Mean	S.D.	Perf.	Risk
Size	22%	4.4%	19%	4.6%	86%	0%	307.68	54.24	236.19	75.13	493.30	0%
Starting Position	17%	3.0%	23%	4.9%	6%	0%	202.18	40.67	324.11	65.06	-1.08	0%
National Diversity	19%	7.0%	21%	2.7%	22%	33%	215.96	65.54	298.84	21.09	338.15	3%
Functional Diversity	20%	3.6%	19%	6.0%	22%	29%	288.19	63.78	254.16	70.87	594.58	0%
Age Diversity	19%	5.2%	21%	2.7%	23%	18%	252.10	63.86	272.38	43.63	316.11	16%
Specialization Perc. Div.	18%	4.7%	22%	2.8%	29%	2%	250.23	70.46	293.62	32.23	330.19	13%
Power Perc. Div.	21%	4.7%	19%	2.2%	18%	28%	271.65	58.55	263.14	24.00	257.23	40%
Ambiguity Perc. Div.	16%	4.2%	23%	3.7%	73%	0%	212.75	54.87	322.05	34.12	501.70	0%
Decision Difficulty Perc. Div	22%	4.9%	19%	1.5%	17%	17%	300.46	29.96	238.45	40.38	478.90	0%
Decision Pressure Perc. Div.	22%	4.2%	18%	3.4%	3%	0%	318.07	57.94	226.39	46.84	-160.41	0%
Effectiveness Perc. Div.	19%	3.4%	21%	4.9%	15%	14%	260.33	62.59	269.12	53.03	317.95	18%

Note: S.D.=Standard Deviation; N=35; computing the Goldfeld-Quandt requires making a choice on the size of the bins, here n=17 for maximum significance.

Figure 1: Reaching a Threshold Depends on the Variability Effect

(a) Effect of X on Y (b) With equal variability (c) With change of variability

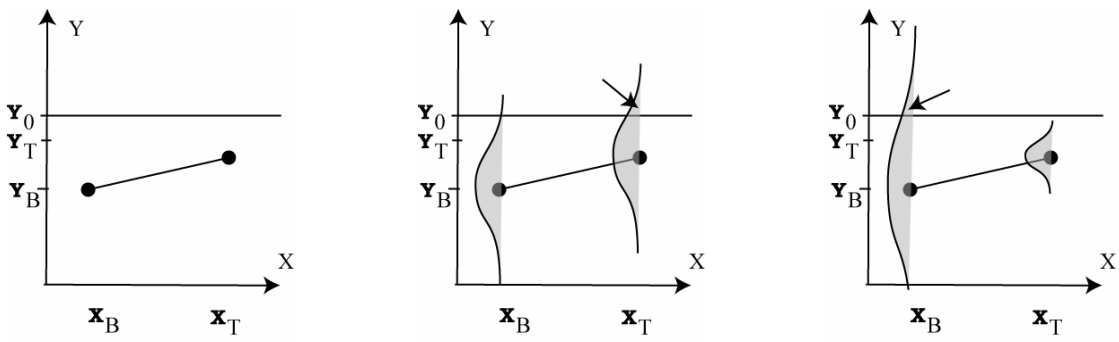


Figure 2: Combining the Influence of Information Availability and Social Integration

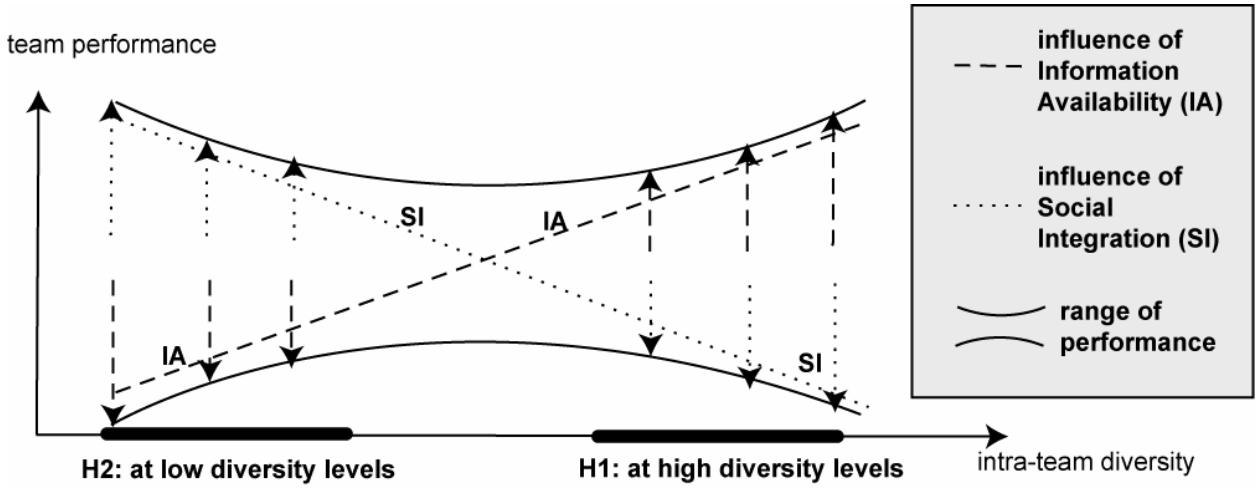


Figure 3: Mean and Variability effects represented by quantile lines, which change slope at critical performance level

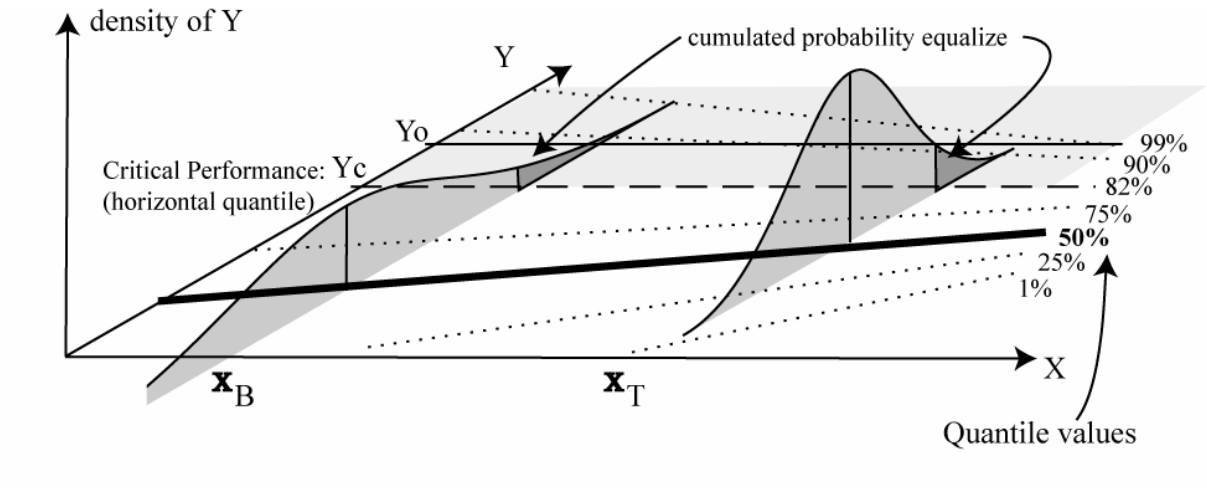


Figure 4: Separating where Mean Effect applies from where Variability Effect has Inversed Implications

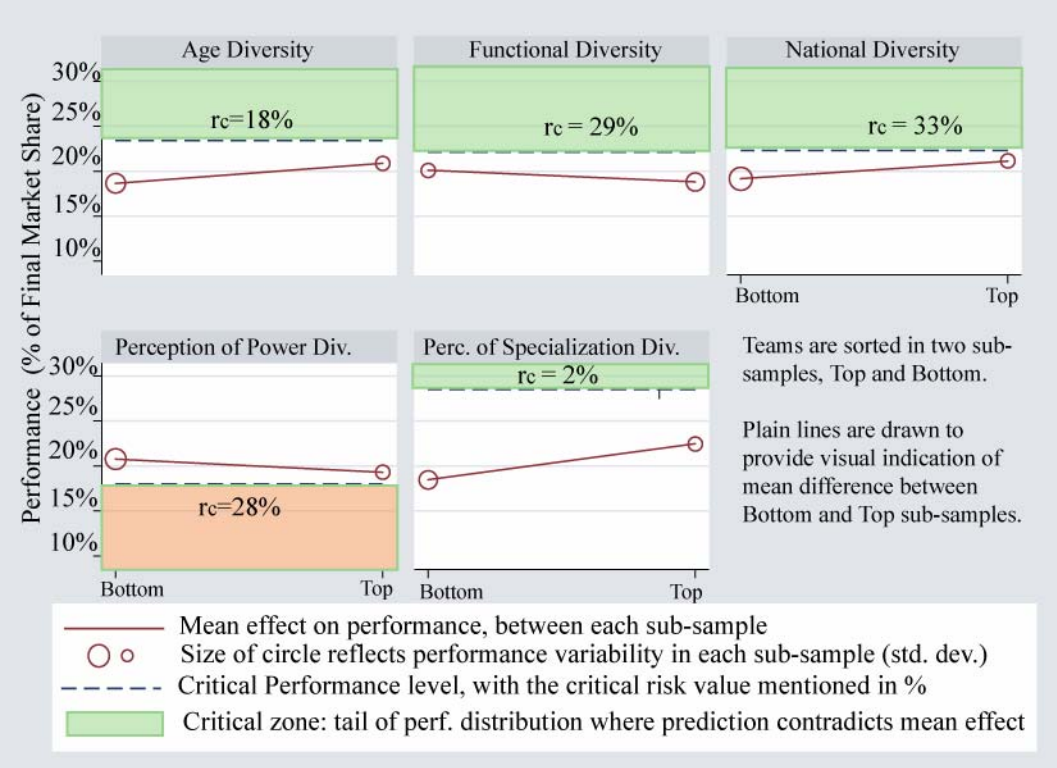
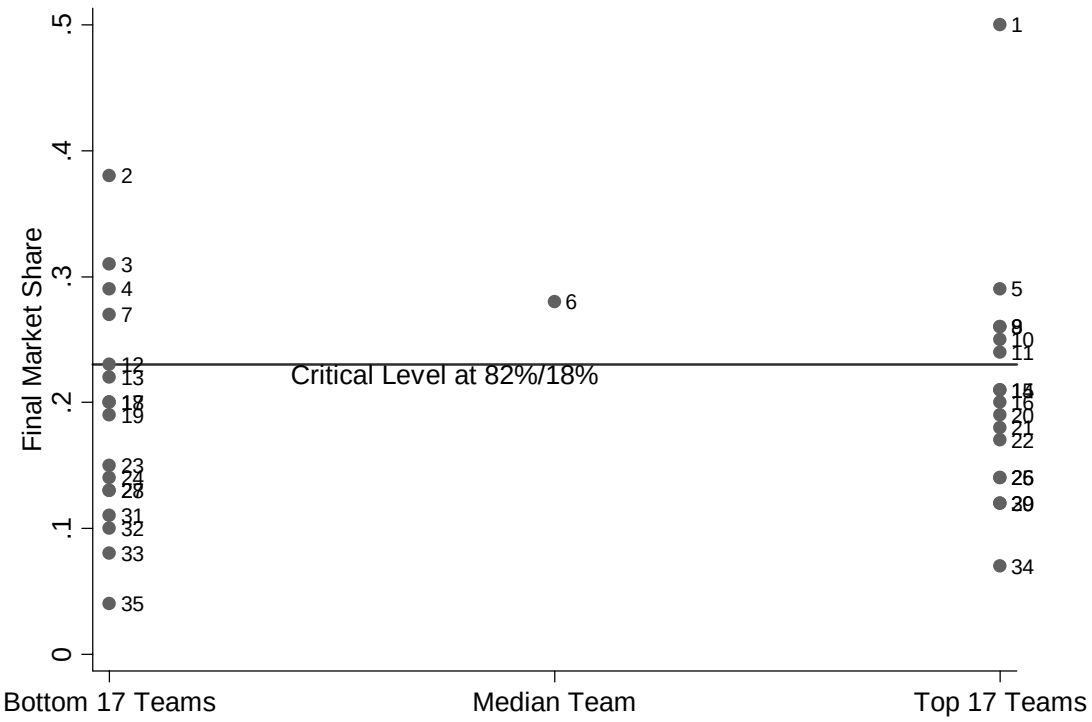


Figure 5: Lower age diversity imply more extremely high performance



Note: the team ranked number one is visibly an outlier in the variability analysis; worst results are also more likely for low age diversity.

Europe Campus
Boulevard de Constance,
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 6072 40 00
Fax: +33 (0)1 60 74 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

www.insead.edu

INSEAD

The Business School
for the World