UNIVERSITE PARIS DAUPHINE

COMPOSITION DE L'EQUIPE ET VARIABILITE DE LA PERFORMANCE

THESE

pour l'obtention du titre de

DOCTEUR EN SCIENCES DE GESTION

(Arrêté du 7 Août 2006)

Présentée et soutenue publiquement par

Fabrice CAVARRETTA

JURY

Directeur de Thèse :

Mme Isabelle HUAULT
Professeur à l'université de Paris-Dauphine

Rapporteurs :

M. Rodolphe DURAND
Professeur à HEC

Mme Isabelle ROYER
Professeur à l'Université de Lille

Suffragants :

M. Christoph LOCH
Professeur à l'INSEAD

M. Raymond-Alain THIETART
Professeur des Universités Emérite, Université Paris-Dauphine

Le 8 septembre 2008

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions doivent être considérées comme propres à leurs auteurs.

RESUME

Dans cette thèse, j'explore l'influence de la diversité de l'équipe sur la variabilité de la performance. La théorie des organisations a relativement négligé les effets sur la variabilité alors que ceux-ci ont des conséquences organisationnelles importantes : avec l'augmentation de la variabilité, la performance extrême devient plus probable et, à partir d'un certain point, ces pointes de performances sont entièrement déterminées par la variabilité, rendant erronées les analyses basées seulement sur les effets moyens.

Ensuite, je fais l'hypothèse que la diversité de l'équipe a un effet curvilinéaire en U sur la variabilité de la performance, impliquant que des équipes très homogènes ou très diverses produisent une plus forte variabilité de performance que des équipes à diversité médiane. L'analyse complémentaire d'une simulation d'entreprise par 35 équipes de dirigeants confirme ces hypothèses.

Enfin, j'explore la variabilité de la performance au niveau de l'équipe, distinguant la variabilité longitudinale de la variabilité inter-tâche, faisant l'hypothèse que la diversité augmente l'une et diminue l'autre. Ceci est vérifié dans des archives de performance académique de 200 équipes d'étudiants dans une grande école de commerce.

ABSTRACT

This dissertation explores the effects of team diversity on performance variability and its organizational consequences. First, I show that organizational theory has mostly assumed variability to be detrimental to organizations, and that it is not a common dependent variable. However, effects on performance variability matter: with increasing performance variability, extreme performance becomes more likely, and—at some point—such extremes are entirely determined by variability, making predictions based on mean effect only to be flawed. I review the consequences of this logic on various organizational theories.

Second, I hypothesize that team diversity has a U-shaped effect on performance variability, with both very diverse and very homogeneous teams exhibiting high performance variability, and teams of middle-range diversity limited performance variability. The reanalysis of the performance of 35 teams engaged in a business simulation confirms those effects on performance variability measured at the team population level. It illustrates how effects on variability better predicts extreme outcomes—here finishing in the top 10% of the population—and how team diversity can have such an effect.

Third, I explore team performance variability at unit level, distinguishing along-time from across-tasks variability for each team. I hypothesize that diversity increases the former and decreases the latter. A longitudinal dataset of 200 student teams performing actual group-work confirms such contradictory effects. This study identifies the two hazards—social and informational—underlying the diversity-performance relationship.

This dissertation contributes to team composition literature by clarifying the various effects of diversity on performance variability. For practitioners, it identifies the low risk of mid-level diversity, from higher risk at the highest and the lowest levels of diversity, driven by organizational hazards.

REMERCIEMENTS

A ma femme,
A mes enfants,

J'exprime ma gratitude à Isabelle Huault pour avoir accepté d'encadrer ce travail à Paris Dauphine.

Je remercie également Rodolphe Durand, Christoph Loch, Isabelle Royer et Raymond-Alain Thiétart pour leur participation à mon jury de thèse.

ACKNOWLEDGMENTS

A dissertation is a very long journey, for which I have been blessed with great support and companionship. Several people deserve recognition for their positive roles in the current work.

My family has gracefully tolerated my going back to studies, good naturedly allowing themselves to believe that my getting deep into a second decade of university education makes any sense at all. More amazingly, my wife, Marion, has maintained her positive stance on this effort, even though, as the joke goes between us, she had married a young, promising, soon-to-be-powerful and increasingly well-paid executive, who quickly turned into a definitively-not-so-young, not-powerful-at-all, poorly-paid grad student father… My two boys, Marcel and Joseph, should also be thanked, because their arrival in my life helped me take a different view about what is truly important in life, which eventually led me back to academia. They also tolerate that daddy is a student like them, and were not too concerned when they could not figure out exactly what it was that I was doing. When I finally decided to tell them I was a researcher studying how to prevent airplanes from crashing, I really enjoyed their buying into that social construction, at the same time that my committee was buying into it—such agreement was a sign that I was onto something.

Many individuals in the academic community have helped me, and primarily, my chair, Bruce Kogut: even before I decided to apply to INSEAD, he had had discussions with me and, by signaling his interest in my profile, participated actively in my decision to make the big jump. Roughly two years later, I went to him again whilst trying to find a way to develop my dissertation project, and Bruce was one of the very first to see something in the unstructured and unconventional approach I had in mind. His accepting to supervise the work was a courageous decision, since, as one would expect, he then had to tolerate the labor pains involved in the making of the dissertation. For these three rounds of pushing me forward, he deserves the first position in my thank you list.

The other members of my committee have also been great supporters. Randall Peterson took a very genuine interest in my work very early on, proposing to share data, help with the writing and so forth, all of which was of immense practical and psychological help to get the project going. Also, his enthusiasm was crucial for a dissertation that had two legs and would have been crippled without him. One of the legs was in a quantitative and conceptual tradition ("performance variability matters") that Bruce Kogut could fully support, but the other leg was grounded in the more classical OB tradition (team and diversity) and therefore required an experienced OB committee member, which Randall provided to the fullest extent. Christoph Loch came on board because of his experience in the literature on risk. However, his down-to-earth, let's-do-it approach to everything he endeavors to do, combined with an amazing multidisciplinary mind, was crucial during the low points of that long process. If we made it a rule that every dissertation committee had to take on a successful junior-league soccer coach like Christoph, then maybe scientific progress would move both faster and smoother. Finally, Henrik Bresman brought his own fresh views and support on how to behave through the dissertation process, in addition to giving me most useful suggestions regarding additional recent studies to consider.

Beyond this team of formal supporters (a.k.a. the committee), various members of faculty have stepped up too, and some much more than once. First, there is Herminia Ibarra, who successfully dealt with her identity changes towards me, from being a professor to a friend (from HBS to her arrival in Paris), then going back to being a professor again (when I joined INSEAD). She made sincere and quality efforts to discourage me from joining academia, an approach that I now emulate when I advise other managers who are innocently dreaming about joining academia. She nevertheless allowed the experiment to go on, and maintained a discreet and arm's length relationship with me during those five years, always making me feel as though she was keeping a "good eye" on me, whatever that means. I hope our blend of friendship and professional relationship will go on forever. Maurizio Zollo also deserves a warm thank you for bringing enlightened and unconditional support to my research very early on and being the first one to sponsor the study that would eventually lead to my dissertation. Early in my first year, he accepted to supervise an early version of this research through an Instructed Research Project. Not only did he formally put his signature on my very rough ideas, but also he did not budge when my initial presentation of that work did not bring unanimously positive feedback from the rest of faculty. I also owe a lot to Martin Gargiulo, who made two important push during my PhD, even though he was not directly involved on the dissertation. First, he took the responsibility, when in charge of PhD recruitment, to let me into the program, not an easy decision when considering the mixed track record of "older" students in such program. Second, he accepted to play the role of advisor on my comprehensive examination research, which turned out a longer and more painful process that we all had hoped. Tom D'Aunno played a similar role—as my academic advisor during the pre-comps period. Eventually, Martin's and Tom's good will and advice in that laborious early phase of the program allowed it to unravel properly.

My research also owes a lot to Martin Kilduff, Reinhard Angelmar, and Ajay Mehra for not only allowing me to reanalyze their data, but also providing valuable advice. Martin Kilduff was one of the very few—out of the tens of scholars I contacted—to agree to share his data, simply responding to a cold email request from an unknown student. This data allowed me to get the very early positive results currently embodied in the chapter II. Our experience of my reanalysis of the data (that he had been good enough to share) also illustrates why such sharing does not occur often in our field: Martin first went through the agony of answering my many questions just for me to be able to run the analysis. Later, I was foolish enough to wonder whether there could be a problem with the original analysis (which it did not take long to eliminate as an hypothesis). Then, there was also the risk that my reinterpretation could put a shadow to their own study (which I can confirm is really not the spirit of my research). Finally, I even added insult to injury by letting—unwittingly—my first conference proceeding publication go out without proper acknowledgement of the source of that data. I have to thank Martin emphatically again for putting up with all that, and I sincerely hope he will do all this again with another scholar; may these pains be counted as his selfless contributions towards educating a junior (one contribution among so many others).

Michael Brimm was present early on, a fatherly figure that would bring the validation necessary when the implacable logic of a PhD program gets one to wonder about one's worth. In such initiation rituals, as in many human endeavors, the psychological aspects are at least as important all the other rational and technical aspects and I enjoyed every minute of my regular conversations with Michael. On these matters, I also benefited greatly from many counseling sessions with Jane Plimsoll, a psychologist that INSEAD has had the resourcefulness to make available to its PhD students. This brings me back to the Brimm family, since I understand that Linda Brimm played a crucial role in initiating and maintaining such support, so thank you also, Linda!

I thank all the professors who trained, advised, or evaluated me at INSEAD during those years: Miguel Brendl, Yianis Sarafidis, Ilia Tsetlin, Jean-Claude Thoenig, Hubert Gatignon, Vibha Gaba, Huy Quy, Yves Doz, Charlie Galunic, Allan Filipowicz, Ilian Mihov, Anca Metiu, John Weeks, Henrich Greve, Filipe Santos, Ha Hoang, Philip Anderson, Enrico Diecidue, Philippe Delquié, Marwan Sinaceur, Stephen Chick, Ayse Onculer, Neil Bearden, Bala Vissa, Morten Hansen. Among them, a special thought goes out to Erin Anderson who left us this year. She was an impressive professor that imprinted many INSEAD PhD cohorts with her positive views of research; she was also a very nice person. At my last meeting with her, we talked about my research, and reflecting about our current methods to deal with heteroskedasticity, she said: “Oh, that’s interesting, we remove outliers, and maybe we are wrong.” Obviously, she was not the first one to make that remark, neither am I the first one to work on that issue, but such frankness, and willingness to use the W-word is rare enough to be brought to her credit.

Université Paris Dauphine have an agreement that allows INSEAD PhD students to defend their dissertation jointly in the doctoral program of both institutions, and this work will be validated under that statute. I have to thank the faculty at the DMSP / Paris Dauphine—in particular my chair, Isabelle Huault, and the committee members Isabelle Royer and Raymond-Alain Thiétart—for bringing their share of supervision and advice, as well as patiently socializing me into the French academic system.

In the larger academic world, any friends and colleagues have supported me, and it will obviously be impossible to list them all. Before attempting to draw up a tentative list, I would like to thank those who took a particular interest in my academic endeavors. Cécile Dejoux immediately decided that I would well in academia, from the first day we met in a diner party before I even started at INSEAD. Unfortunately, we still need to wait for a few decades for her prediction to be fully verified, but in the meantime, she was always there to keep on repeating it and it really felt good. My old accomplice Armand Ajdari also kept thinking I could do it, so thank you Armand, and I still owe you a citation in my first published paper. Frédéric Dalsace deserves special recognition since he was the one to push me into sending an application, putting something of a “because I am worth it” cachet in applying to the INSEAD PhD. He also put me in touch with Nicola Dragonetti, who became something of a role model for me, as he is for many others, because of his encyclopedic knowledge of our field. Heidi Gardner has played a crucial role in the dissertation by helping significantly in my data collection efforts, and bringing her legendary mix of kindness and professionalism in a process that could otherwise have been ugly.

Also worth mentioning are the various researchers around the world who have helped and even encouraged me in this research path. I have to thank in particular Jerker Denrell, Etty Jehn, Scott Page, Stuart Bunderson, Rodolphe Durand, Tony Simons, Joel Baum, Michael Cohen, Monica Higgins, Bill McKelvey, Don Palmer, Karlene Roberts, Ray Reagans and Lee Fleming.

I will not risk trying to draw a line between colleagues and friends, nor will I try to have a narrative for every name. Most of the people below have been my companions during the PhD experience, and most of them reading early versions of my work, a substantive effort

that reflects interest, friendship, and often both. I thank you all, at INSEAD and elsewhere, for supporting me over those years: Anne-Claire Pache, Julie Battilana, Atalay Atasu, Prashant Deshpande, Aamir Khan, Metin Sengul, Dimo Ringov, Roxana Barbulescu, Gökhan Ertug, Manuela Giangrande, Yu Zhang, Olivier Chatain, Oliver Gottschalg, Jonghoon Bae, Konstantin Korotov, Govert Vroom, Michael Yaziji, Hakan Ener, Rahul Kapoor, Hajo Adam, Donal Crilly, Amit Jain, Otilia Stanciu, Jennifer Petriglieri, Imran Chowdhury, Hana Shepherd, Chris Rider, Spela Trefalt, Gergana Todorova, Denis Deschamps, Lionel Roure, Ioanna Tziri, Sam Garg, Patrick Ciarlet, Ben Hallen, Guillaume Chevillon, Olivier Saulpic, Maryse Dubouloy, Jean-Pierre Ponssard, Christophe Midler, Basit Chaudhry, Alexis Bonnet and Bernard Prieur-Smester.

A few people outside of academia also deserve recognition because they gracefully helped me when I was searching for a field setting. I thank Vincent Schoendoerffer and Pierre-Alexis Dezard at Altares / Dun & Bradstreet; General Georgelin, General Gilles, Colonel de Maisonneuve, and Commandant Combi in the French Armed Forces; Philippe Hellich at Danone; Grégoire Sentilhes at NetStage Ventures; Laurent Detrie at Sodexho, Jean de Vauxclair, Olivier Brousse and Philippe Huc at Veolia; Marc Villeneuve at Air France; Olivier Thorel at Shell; Pierre Hurstel at Ernst&Young; Jean-Laurent Poitou at Accenture; Philippe Compagnion and Pierre Mogenet at Egon Zehnder.

The PhD Office staff at INSEAD, in particular Alina Jacquet, Sabine Tognotti, and Valérie Hache also deserve my warm thanks. Alina has been so available and supportive of the PhD students for all my years in the program that I dare to use the word *blessing* for her presence in that position. I also warmly thank the various colleagues in the administrative staff at INSEAD in particular Muriel Moureaux, Alexandra Cappadoro, Sylvie Loisy, Miranda Helmes, Carole Lorusso, Stéphanie Paille, Ana de Sa and Laureen Sorreda. Their presence brought a significant amount of humanity, rationality, and practical sense to the craziness of graduate student life. At Paris Dauphine, I also really enjoyed the caring administrative support of Patricia Lenfant. Finally, this work was produced by a non-native English speaker, and therefore required, just to reach a reasonable state of readability, the help of nearly professional copy-editors. I thank in particular Angshu Chatterjee and Melissa Haveman for their careful comments on the current text.

Aside from the help of the above colleagues, friends and family, this research has been informed by the reviews and feedback from various conferences. Here is a list of the events where the essays underlying this dissertation were presented (or at least reviewed and accepted as of print date):

Chapter I

- Organization Science Winter Conference 2008 (Squaw Creek)
- AIMS 2008 (Sofia Antipolis)
- EGOS Conference 2008 (Amsterdam)

Chapter II

- LBS Transatlantic Conference 2006 (London, UK)
- Research on Managing Groups and Teams Conference 2007 (Cornell)
- INGroup Conference 2007 (Lansing, MI)
- Academy of Management Conference 2007 (Philadelphia; OMT division; Best Paper proceedings)
- EGOS Conference 2007 (Vienna)

Chapter III

- LBS Transatlantic Conference 2007 (London, UK)
- Academy of Management Conference 2008 (Anaheim; OMT division)

TABLE DES MATIERES

Remarques sur les langues employées : la langue de travail pour cette recherche a été l'anglais, et les essais sont donc rédigés dans cette langue. Néanmoins, une traduction est fournie dans le texte pour les parties suivantes : résumé, introduction et conclusion de la thèse, ainsi que le titre et le résumé de chaque chapitre.

Résumé iii

Remerciements v

Table des matières xii

Liste des tableaux xiv

Liste des figures xiv

Introduction Générale 1

Les questions de recherche 3
Plan 8

Introduction 10

Focal Questions of the Dissertation 11
Overview 14

Chapter I. Comment la variabilité inverse les effets entre la moyenne et la performance extrême 17

Introduction 18
Literature Review on Performance Variability 19
When Variability Makes Inferences about Extreme Outcomes the Inverse of those about Average Outcomes 25
Theoretical Consequences of Neglecting Variability 34
Future Directions 39

Chapter II. Mieux, meilleur, ou pire ? Diversité d'équipe et performance extrême 41

Introduction 42
Literature Review 43
Theory 47
Method 52
Results 60
Discussion 67

Chapter III. Influence de la diversité de l'équipe sur le risque organisationnel : contraste entre les effets sur la variabilité longitudinale et inter-tâches 71

Introduction 72
Literature Review 74
Distinguishing Along-Time vs. Cross-Task Performance Variability at Team Level 79
Method 81
Conclusion 89

Dissertation Closing Remarks 91

Contributions 91

Limitations 92
Next Steps 94
Conclusion 95

Discussion et conclusion 97

Contributions de l'étude 97
Limitations de l'étude 98
Prochaines étapes 100
Conclusion 102

Appendix 103

Formal Determination of Critical Performance Level 103

Références 105

LISTE DES TABLEAUX

Table 1 : Synthèse des différents chapitres 9
Table 2 : Summary of Dissertation Essays 15
Table I-1: Illustration of Theoretical Consequences of Mean-Variance Tradeoff 38
Table II-1: Summary Statistics and Correlations 61
Table II-2: Analyses 62
Table II-3: Computation of Critical Performance Level and Probability 63
Table III-1: Summary and Correlations 85
Table III-2: Effects on Performance Variability Along-Time 86
Table III-3: Effects on Performance Variability Across-Task 87
Table III-4: Contradictory Effects of Team Diversity on Performance Variability 89

LISTE DES FIGURES

Figure 1: L'effet d'un facteur de groupthink sur la variabilité de la performance (Bourgeois, 1985) 5
Figure 2: Pourquoi des effets sur la variabilité peuvent induire à la fois des performances très élevés et très basses 6
Figure 3 : Les groupes à culture mixte ont des performances plus variables que les groupes monoculture (Adler, 2002) 8
Figure I-1: Attainment of a Threshold Y_0 Depends on the Variability Effect 26
Figure I-2: Quantile Lines Represent Both Mean and Variability Effects 28
Figure I-3: Critical Performance Position 30
Figure I-4: No Mean Effect with Variability Effect Predicts both Extremely Low and Extremely High Outcomes 30
Figure I-5: Scatter Plot including Regression Line and Critical Level Surrounded by Quantile Lines of Converging Slope 31
Figure I-6: Zoom on Performance Range beyond the Critical Level 33
Figure II-1: Reaching a Threshold Depends on the Variability Effect 47
Figure II-2: Combining the Influence of Information Availability and Social Integration 49
Figure II-3: Mean and Variability Effects Represented by Quantile Lines, which Change Slope at Critical Performance Level 58
Figure II-4: Separating where Performance Range where Mean Effect applies from Range where Variability Effect has Inversed Implications 64
Figure II-5: Lower Age Diversity Imply More Extremely High performance 66
Figure II-6: Greater Functional Diversity Increases both Extremely High and Low Performance 67
Figure III-1: The Moderating Effects of social Integration and Task-Information Fit 76

INTRODUCTION GENERALE

L'équipe est un élément de base des organisations et sa composition un des leviers principaux de la construction organisationnelle. Comprendre la relation entre la composition de l'équipe, en particulier sa diversité, et sa performance est un point de recherche important de la théorie organisationnelle (Mannix et Neale, 2005). En outre, la vie organisationnelle contemporaine prend en compte des impératifs moraux et sociaux tendant à promouvoir la diversité, ce qui renforce la nécessité de comprendre son implication en terme de performance.

La plus grande attention a été portée à la relation entre la diversité de l'équipe et sa performance *en moyenne*. Traditionnellement, les chercheurs ont étudié si telle ou telle dimension de la diversité, par exemple la nationalité, augmentait la performance des équipes. L'expression '*en moyenne*' est mise en exergue car la question, dans ce cas, est de savoir si des équipes ayant une plus forte diversité nationale ont une performance *en moyenne* meilleure. Néanmoins, ceci ne recouvre pas la totalité des questions possibles autour de la performance de ces équipes. Par exemple, la diversité pourrait-elle augmenter la performance *en moyenne*, en même temps qu'on pourrait constater que de plus en plus d'équipe ont une performance *catastrophique* (un tel effet est possible si les autres ont une performance qui compense, dans la moyenne, le score des très mauvaises).

Se pose donc la question, normative, de savoir laquelle de ces deux conséquences – l'impact moyen vs. l'impact sur les catastrophes – est pertinente dans le contexte considéré. La logique de l'effet moyen est la norme dans la recherche organisationnelle. Ceci est justifié si l'objectif est littéralement la moyenne atteinte par les équipes en question. Par exemple, une entreprise qui aurait dix équipes de ventes pourrait ne se préoccuper que de la vente totale de ses équipes, et donc être intéressé par les effets moyens (vente totale = vente moyenne x 10).

Néanmoins, cette logique ne s'applique pas dans les nombreux contextes où la probabilité d'atteindre une performance extrême – soit extrêmement basse, soit extrêmement haute – est plus pertinente que le résultat en moyenne.

Par exemple, des groupes de travail peuvent être mis en place dans le but de concevoir une innovation radicale. Dans l'industrie pharmaceutique, des équipes conçoivent des médicaments, et ne sont retenus que ceux dont le chiffre d'affaires annuel peut dépasser 100 millions de dollars, ceux qu'on appelle les "blockbusters" et qui génèrent la plupart des bénéfices. Dans ce contexte, on se préoccupe plus d'augmenter la chance que quelques équipes découvrent un blockbuster que de la performance moyenne des équipes projet (Galambos & Sturchio, 1998). De même, dans le secteur du capital-risque, en particulier dans la haute technologie de la Silicon Valley, on attend des équipes qu'elles atteignent des succès exceptionnels, comme l'introduction en bourse (IPO). Ainsi, le capital-risqueur (Venture Capitalist, « VC ») espère que quelques-unes des entreprises de son portefeuille réussiront exceptionnellement bien afin de rembourser les investissements faits dans les entreprises dont les résultats s'avéreront mauvais ou seulement moyens. Améliorer leurs chances d'obtenir ces quelques succès rares et exceptionnellement rentables a plus d'importance pour certains capital-risqueurs que d'améliorer le sort de l'entreprise moyenne de leur portefeuille (Kenney & von Burg, 1999). Ainsi, dans ces deux industries, le management attend des équipes qu'elles atteignent des niveaux de haute performance qui sont à la fois extrêmes et rares, objectif qui peut ne pas être équivalent à l'amélioration de la moyenne des performances de toutes les équipes.

Symétriquement, les équipes peuvent avoir pour objectif d'éviter d'atteindre un niveau de performance extrêmement bas. Par exemple, dans une compagnie aérienne, un objectif fondamental de l'équipage est d'accomplir sa mission en évitant les accidents, avant toute considération de métriques de performance moyenne telles que la consommation de carburant ou les retards (Weick, 1990). Dans la gouvernance des grandes entreprises, le premier souci des régulateurs est d'éviter les résultats extrêmement faibles, par exemple les faillites comme celle d'Enron. De tels événements affectent la confiance des investisseurs et ont des coûts sociaux disproportionnés. Toutefois, la prévention de tels fiascos n'est souvent pas équivalente à l'optimisation de la performance moyenne (voir le débat autour du règlement Sarbanne Oxley comme un exemple). Dans ces deux exemples, la prévention des résultats extrêmement faibles semble plus importante que l'amélioration de la performance en moyenne.

Jusqu'à présent, la recherche sur la diversité dans les équipes a appliqué un paradigme classique préconisé par Mohr (1982), cherchant à prévoir seulement les effets sur la moyenne des résultats. En revanche, le lien entre la diversité de l'équipe et les très hautes ou basses performances n'a pas été exploré systématiquement. En particulier, peu de recherches étudient l'effet de la diversité de l'équipe sur la variabilité de la performance de l'équipe.

Cette thèse vise à couvrir trois classes de problèmes qui dérivent de cette observation :

- Dans les cas où la performance extrême est importante, que va apporter l'étude de la variabilité de la performance en tant que variable dépendante ?
- Quelles sont les méthodes qui pourraient être mises en œuvre pour étudier ces effets ?
- Quelle théorie, et donc quels mécanismes, pourraient lier la diversité de l'équipe et la variabilité de sa performance ?

Je présente ci-dessous comment les questions de recherche ont émergé et, dans la section suivante, je résume la structure du mémoire, et la façon dont ces questions sont adbordées.

LES QUESTIONS DE RECHERCHE

Traditionnellement, le chercheur cherche à expliquer l'effet d'un facteur, par exemple la diversité, sur une variable dépendante, par exemple la performance. Il constate alors si la variable indépendante augmente ou diminue la moyenne de la variable dépendante. Quel peut être l'intérêt de considérer comme variable dépendante, non plus la moyenne, mais la variabilité de la performance ? Si les effets sur la variabilité apparaissent dans divers domaines scientifiques, en jouant par exemple un rôle central dans la théorie financière (Black et Scholes, 1973), ils figurent rarement dans les études organisationnelles (par exemple, Sørensen, 2002 ; Taylor & Greve, 2006). L'étude de la variabilité n'est pas un angle naturel de la recherche dans notre domaine.

A l'origine, je me suis intéressé aux catastrophes organisationnelles telles que les fiascos d'Enron ou de Vivendi. Ces deux entreprises phares de la fin des années 90, aux Etats-Unis et en Europe, ont respectivement connu un succès foudroyant avant d'échouer de manière retentissante, Enron ayant même été conduite à la faillite. Dans les deux cas, nombre de commentateurs ont tenté d'analyser la cause du fiasco. Sans doute celui-ci est-il dû à de nombreuses causes, mais qu'il s'agisse de Vivendi ou d'Enron, l'équipe des dirigeants, et en premier lieu son leader, s'est vu attribuer la plus grosse part de responsabilité dans l'échec, avec pour conséquence le départ dans des circonstances peu glorieuses de Jean-Marie Messier, PDG de Vivendi, et la condamnation de Kenneth Lay, PDG d'Enron.

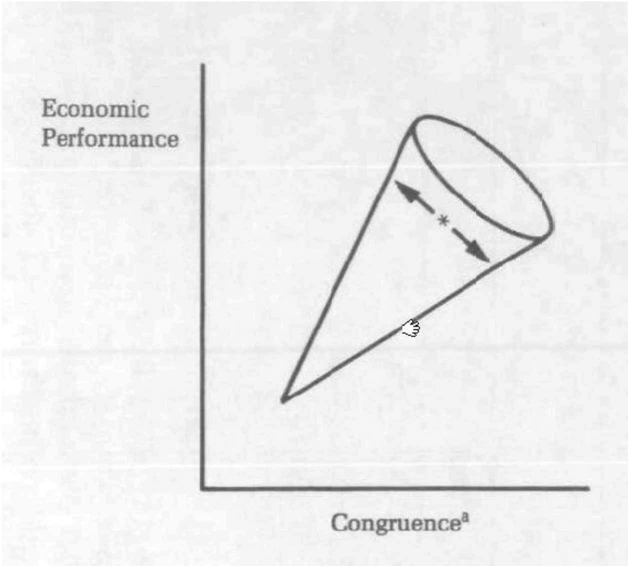
La recherche organisationnelle tend à analyser de telles catastrophes en se focalisant sur l'individu, cherchant chez le dirigeant des fautes ou des traits caractérisés, tels la préférence au risque (Tversky & Kahneman, 1981) ou des réponses comportementales risquées (Mars & Shapira, 1992).

Cette approche diffère notablement de celles utilisées pour analyser les accidents d'avion (Weick, 1990) ou les morts accidentelles (les pompiers chez Weick, 1993). Dans la littérature sur les industries à risques (High Reliability Organization, HRO), les catastrophes sont étudiées plus sous l'angle des risques systémiques ou des facteurs sociaux que par la recherche de fautes individuelles (Perrow, 1984). Par exemple, pour expliquer la catastrophe de la navette Challenger, les travaux ont moins cherché à pointer les erreurs d'individus isolés qu'à explorer la dynamique sociale qui a conduit à ignorer les avertissements au sujet de fissures finalement fatales (Vaughan, 1997). Cette approche n'ayant pas été largement utilisée pour analyser les fiascos d'organisations telles que Enron ou Vivendi, mon objectif de départ était donc de proposer une approche systémique pour expliquer de telles catastrophes dans des entreprises.

Le niveau d'analyse choisit fut l'équipe, car il est le plus abordable pour examiner la dynamique sociale. Comme c'est le plus courant dans la littérature organisationnelle, la composition de l'équipe est analysée à travers des métriques de diversité, principalement démographiques : âge, sexe, nationalité, origine ethnique, expérience, éducation etc. Les études sur le « groupthink » engagée par Janis (1971) avaient déjà étudié le lien entre la diversité et des catastrophes. Toutefois, elles n'avaient pas été en mesure de fournir des résultats concluants à l'intuition que « les équipes homogènes sont dangereuses». En effet, nombre d'études cherchent à démontrer que l'homogénéité dans l'équipe entraîne une diminution des performances en moyenne, mais les analyses selon cette logique ne sont finalement pas concluantes (voir Esser, 1998, pour un résumé sur la littérature du groupthink).

En réfléchissant à l'idée que «les équipes homogènes sont dangereuses », je croisais une étude réalisée par Bourgeois (1985) qui laissait entendre que cela pourrait être le cas s'il existait un effet sur la variabilité de la performance (voir Figure 1). Cette figure, qui était accompagnée de peu de commentaires dans le texte, suggérait que le facteur « congruence » augmente la gamme de performance possible, en plus d'augmenter la performance moyenne. Cet écart grandissant de performance pourrait jouer un rôle dans l'apparition de catastrophes, et il convenait donc d'explorer si la diversité intra-équipe pourrait avoir un tel effet.

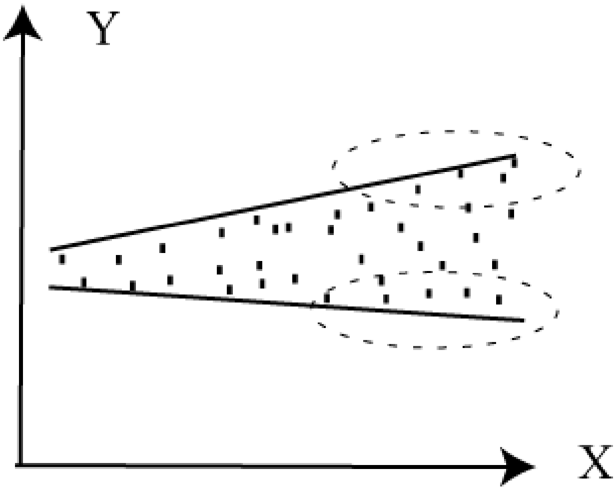
Figure 1: L'effet d'un facteur de groupthink sur la variabilité de la performance (Bourgeois, 1985)



Ce schéma montre la relation possible entre une caractéristique de l'équipe (la congruence, donc l'accord cognitif) et la performance économique de l'équipe. Le cône suggère que plus la congruence est grande, plus la gamme des performance obtenues augmente, en plus du fait que la congruence augmente la performance en moyenne.

Ceci mène à une première intuition, celle qu'un facteur augmentant la variabilité de la performance pourrait engendrer des résultats à la fois très élevés et très faibles (voir un des cas possibles en Figure 2). Toutefois, une telle possibilité doit être pondérée par l'effet moyen. Si un facteur augmente légèrement la variabilité, mais a un très fort effet positif sur la moyenne, il semble qu'il n'augmentera pas les chances de faire pire (comme c'est le cas dans la Figure 1). Alors qu'un fort effet sur la variabilité combiné à un effet négligeable sur moyenne augmenterait à la fois l'occurrence de résultats extrêmement élevés et extrêmement faibles. Dans la Figure 2 représentant la relation entre un facteur X hypothétique et une mesure de performance Y, une telle combinaison est représentée : quand X augmente, la moyenne de Y augmente légèrement, mais la dispersion aussi, à tel point qu'il y a à la fois de plus en plus de très bon scores et de très mauvais scores, ceux-ci étant repérés par un ovale en pointillé.

Figure 2: Pourquoi des effets sur la variabilité peuvent induire à la fois des performances très élevés et très basses



Ce schéma montre la relation possible entre un facteur X et la performance Y. De nouveau le facteur augmente la moyenne et la dispersion valeurs obtenues Néanmoins, cet effet de dispersion est assez fort pour qu'il y ait des valeurs de plus en plus basses (zone pointillée du bas), malgré le fait que la moyenne augmente.

La première question est donc de formaliser ce compromis moyenne-variance et de trouver un critère pour distinguer les cas où les résultats sont plus influencé par l'effet sur la moyenne plutôt que ceux où ils sont influencés par l'effet sur la variabilité. Cela conduit également à la question de savoir si la théorie des organisations a déjà traité des effets sur la variabilité de la performance. Il remarquait alors que cette variable dépendante est rarement étudiée, ce qui mène alors à s'interroger sur cette carence, et surtout à clarifier pourquoi l'étude des effets sur la variabilité de la performance pourrait contribuer à la théorie des organisations.

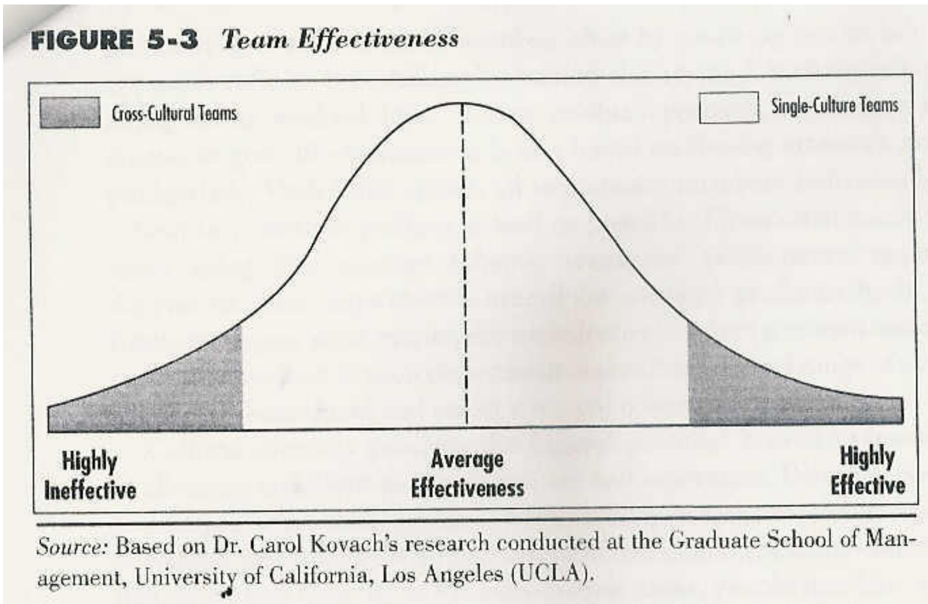
Il apparu aussi que différentes définitions de la variabilité sont possibles. Par exemple, si on étudie la performance d'équipes, on peut tout d'abord conceptualiser la « variabilité au niveau de la population », c'est-à-dire l'écart entre les meilleurs et les pires équipes dans une population d'équipes (exemple : l'écart entre les meilleures et les pires startups du portefeuille d'un capital-risqueur). Il est aussi possible d'imaginer une « variabilité individuelle de la performance de l'équipe», c'est-à-dire des variations de performance de chaque équipe. Dans ce deuxième cas, la performance peut fluctuer au fil du temps, ou entre deux tâches de nature différentes. Par exemple, si on étudie les équipes de direction d'entreprise, ceci pourrait être l'écart entre la performance financière et la performance environnementale. Les différentes définitions de la variabilité – au niveau de la population d'équipe vs. au fil du temps vs. entre diverses tâches – conditionnent le type d'impacts organisationnels que l'on peut étudier.

Après avoir clarifié les raisons d'étudier la variabilité de la performance (objet du chapitre I), examinons la question substantive de cette thèse, à savoir les effets de la diversité de l'équipe. Peu d'études lient ce construit à la variabilité des performances. La psychologie sociale classique fait allusion au fait que les équipes homogènes sont plus risquées (Schachter, Ellertson, McBride, & Gregory, 1951), ce qui avait été étudié de manière tangentielle dans la littérature sur la polarisation (Friedkin, 1999). Ces idées n'ont pas donné lieu à des formalisations complètes, suggérant ainsi le besoin de théoriser qu'une grande homogénéité implique une plus grande variabilité de la performance de l'équipe.

Au contraire, d'autres perspectives suggèrent que les équipes très diverses sont les plus risquées. Par exemple, il semble que des équipes de culture différente sont soit très performantes, soit très dysfonctionnelles (par exemple, voir la Figure 3). De plus, la littérature sur l'apprentissage organisationnel suggère aussi que l'augmentation de la diversité dans les équipes conduit à une plus grande variabilité de la performance (Fleming, 2004 ; Taylor & Greve, 2006). Ceci crée donc une apparente contradiction qui appelle à une exploration approfondie de la relation entre la diversité et la variabilité de la performance de l'équipe. [1]

[1] Une préoccupation supplémentaire, de nature épistémologique, apparut lorsque je présentais les premiers résultats de cette recherche liant diversité et variabilité de la performance. L'argument pour expliquer que la variabilité est influencé par la diversité s'appuie sur un argument contingent de ce type : « l'entente de l'équipe est aléatoire ; pour les équipes où les membres s'entendent, la diversité est bénéfique, pour les équipes où les membres ne s'entendent pas, la diversité est préjudiciable ; donc, la diversité augmente la variabilité de la performance ». Face à ces arguments, certains chercheurs s'attendaient à une vérification empirique de la modération par le facteur de contingence (dans cet exemple, l'entente de l'équipe). Toutefois, la mesure explicite de telles modérations non seulement contraindrait la vérification empirique, mais un tel angle de la recherche ne serait pas une contribution significative puisque des études ont déjà exploré ces modérations (par exemple Van der Vegt & Bunderson, 2005). Par conséquent, ma théorisation suppose de telles modérations, et mon approche empirique ne cherche pas à les prouver.

Figure 3 : Les groupes à culture mixte ont des performances plus variables que les groupes monoculture (Adler, 2002)



Ce diagramme illustre comment se distribuerait, en termes de performance, une population d'équipe dans laquelle la diversité culturelle varie. Les équipes homogènes culturellement auraient une performance moyenne alors que les équipe cross-culturelles seraient soit très efficaces soit très inefficaces.

PLAN

La thèse est structurée en trois chapitres portant sur trois ensembles de questions et utilisant trois approches empiriques distinctes. Ces chapitres ont été construits comme des essais indépendants, qui permettent de distinguer et de rendre plus lisibles les différents apports théoriques et empiriques.

Le premier chapitre ("Comment la variabilité inverse…") est un essai conceptuel explorant pourquoi et comment la variabilité de la performance est importante pour la théorie des organisations ; il est illustré par des données simulées. Il ne traite pas de théorie substantive, telle que celle relative à la diversité de l'équipe, et a vocation à être invoqué dans de nombreuses perspectives substantives de la théorie des organisations. Il fournit la motivation à étudier les effets sur la variabilité de la performance et identifie un critère essentiel : le niveau de performance critique.

Le deuxième chapitre ("Diversité d'équipe et performance extrême") est l'essai de base concernant la diversité des équipes, et explore l'effet de la diversité sur la variabilité de la performance. Il démontre que la variabilité des performances est plus grande à la fois pour les équipes très homogènes et très diverses, par rapport à la variabilité limitée des équipes moyennement diverses. L'un des objectifs est de démontrer empiriquement comment prendre en compte la variabilité pour parvenir à des conclusions différentes d'une approche ne prenant en compte que des effets moyens. Des données qui avaient déjà été étudiées et pour lesquelles un effet moyen avait été établi furent identifiées. Kilduff, Angelmar et Mehra eurent la bonté de me permettre la ré-analyse des données de leur étude sur la diversité démographique et cognitive (2000). Du point de vue méthodologique, ce chapitre utilise la logique développée dans le chapitre I, et démontre comment le niveau de performance critique peut être calculé dans le cas où la variable indépendante n'est pas continue.

Le troisième chapitre (“Contraste entre variabilités, longitudinale vs. inter-tâches”) explore les effets de la diversité sur différents types de variabilité, et fournit aussi la vérification – par une approche indirecte – des modérations qui sous-tendent mes inductions. Il est basé sur la même induction théorique que le chapitre II, mais il met davantage l'accent sur l'exploration de sources d’aléa, et finit donc par promouvoir les concepts d’aléa informationnel et d’aléa social de la vie de l’équipe.

La Table 1 résume les caractéristiques de chaque chapitre et identifie les liens entre eux.

Table 1 : Synthèse des différents chapitres

	Chapitre I : **“Comment la variabilité inverse…”**	**Chapitre II :** **“…Diversité d’équipe et performance extrême”**	**Chapitre III :** **“…contraste entre variabilités, longitudinale vs. inter-tâches”**
Contribution Théorique	Conceptuel: la variabilité est importante pour la théorie des organisations	• La diversité a une relation en U avec la variabilité • L’approche par la variabilité change les conclusions sur les effets de la diversité	• La diversité augmente la variabilité longitudinale mais diminue celle inter-tâche • Identifie les aléas social et informationnel de la composition d’équipe • Confirmation indirecte des modérations
Données empiriques	Un ensemble de données simulées	Ré-analyse d’une étude publiée sur 35 équipes de cadres supérieurs.	Données originales sur 4 promotions de 50 équipes d’étudiants
Méthode	Introduit le critère de « performance critique » (version *continue*)	Utilise la version *binaire* du critère	Utilise différentes définitions de la variabilité
Lien entre les chapitres		Utilise la méthode et les concepts exposés au chapitre I	Utilise une induction théorique similaire au chapitre II

INTRODUCTION

Teams are the basic building blocks of organizations, and their composition is one of the most obvious cornerstones of organizational engineering. Understanding the relationship between team composition—in particular team diversity—and team performance has been one of the perennial research problems of organizational theory (Mannix & Neale, 2005). In addition, modern organizational life recognizes the moral and social imperatives for diversity, thus reinforcing the necessity to understand its performance implications.

Most attention has thus far gone into exploring whether team diversity improves performance *on average*. Yet, such logic may not apply in the many contexts where the chances of reaching extreme performance—either extremely low or extremely high outcomes—matter more than expected outcome.

For instance, project teams may be assembled with the objective of producing a radical innovation. In the pharmaceutical industry, project teams are expected to work towards producing drugs whose yearly revenues might surpass 100 million dollars, the so-called "blockbuster drugs" that generate the most profits in the pharmaceutical industry. In this context, one cares more about the ability to increase the chances of a few project teams discovering a blockbuster drug than about the average performance of all project teams (Galambos & Sturchio, 1998). Similarly, in the venture capital industry, particularly in the Silicon Valley high-tech sector, venture teams are expected to reach the exceptional success of an Initial Public Offering (IPO). Accordingly, the Venture Capitalist (VC) firms rely on only a few firms in each portfolio to reach the exceptional outcomes that make the whole portfolio profitable. Improving the chances of the exceptionally profitable IPOs matters more to some VCs than improving the fate of the average venture (Kenney & von Burg, 1999). In those two exemplar industries, management expects teams to reach high performance levels that are both extreme and rare, which may not be equivalent to improving the average of the performance of all the teams.

Symmetrically, in some other contexts, teams are expected to avoid extremely low performance levels. For instance, in the context of an airline, the cockpit crews' first objective is to avoid crashes, above attending to any average performance metric such as fuel consumption or arrival delays (Weick, 1990). In the governance context, the top management teams of large traded firms and the market regulators are expected first to avoid extremely low outcomes, for instance corporate fiascos and bankruptcies, such as Enron's. Those events disproportionately affect the trust of investors and have disproportionate societal costs. However, the avoidance of fiascos is often not equivalent to optimizing the expected performance (see the debate around the Sarbanne Oxley regulation as an example). In these two examples, the avoidance of extremely low outcomes appears more important than the improvement of performance on average.

Team diversity research has so far followed a classical paradigm—advocated by Mohr (1982)—of predicting effects on average outcome. According to accepted wisdom, teams with diverse members should perform better because they can leverage better information. Therefore, the average performance of a population of teams made of diverse members is expected to be higher than that of a population of teams made of similar members. In contrast, the link between team diversity and very high or very low performance has not been explored systematically. In particular, little research links the effects of team diversity on team performance variability, a predictor of both extremely high and extremely low outcomes.

This dissertation aims to cover the three classes of issues that such exploration raises:

- the conceptual issue of why performance variability may be a dependent variable of interest,
- the methods required by an approach taking variability as a dependent variable,
- and a theory grounded on substantive mechanisms linking team diversity and team performance variability.

These three classes of issues have emerged in the development of this research program. I present the logic of this emergence here, and, in the next section, I summarize the structure of the dissertation, the questions it raises, and how they are answered.

FOCAL QUESTIONS OF THE DISSERTATION

To start from the beginning, one may ask why someone would consider performance variability as the dependent variable in an organizational theory research program. Even though effects on variability appear in various fields, playing, in particular, a central role in financial theory (Black & Scholes, 1973) and appearing sparingly in existing organizational studies (e.g. Sørensen, 2002; Taylor & Greve, 2006), studying performance variability is not a natural angle of research in our field. Rather, I was originally interested in studying organizational catastrophes such as the fiascos at Enron or Vivendi. I had been puzzled by the fact that organizational research tends to explain such catastrophes by focusing on individual level explanations, seeking to put the blame on individual agency or traits, for instance the risk preference (Tversky & Kahneman, 1981) or risk behavior (March & Shapira, 1992) of top managers.

This approach differs markedly from those used to analyze aircraft crashes (Weick, 1990) or accidental life losses in firefighter teams (Weick, 1993). In the high-reliability organization literature (HRO), catastrophes are studied more through the lens of systemic or social factors than by seeking individual faults (Perrow, 1984). For instance, when trying to explain the Challenger catastrophe, research sought less to characterize isolated individuals than to explore the social dynamics that lead to ignoring warnings about cracks in booster rings (Vaughan, 1997). Interestingly, this approach had not been used extensively to explain the corporate fiascos such as Enron or Vivendi, giving way rather to an obsession with the characteristics and intent of prominent managers. My initial goal was therefore to search a systemic lens to explain organizational extreme outcomes.

I narrowed my interest on teams, feeling that this construct could be the most tractable to consider social dynamics. For similar reasons, I narrowed onto team composition, in particular, team diversity (in age, gender, nationality, ethnicity, experience, education, etc.), because the groupthink literature initiated by Janis (1971) had already linked such constructs to catastrophic outcomes. However, the ensuing literature had not been able to deliver on its original premise, that homogeneous teams are "dangerous", by failing to demonstrate that team homogeneity leads to lower performance on average (see Esser, 1998 for the introduction of a modern revisit of the groupthink literature).

While playing with the idea that "homogeneous teams are dangerous", I encountered a study by Bourgeois (1985) hinting at a groupthink effect possibly occurring by influencing performance variability (Figure 1). It led me to wonder a) whether making performance more or less variable could explain catastrophic outcomes, and b) if intra-team diversity could have such effect.

Regarding that first intuition, it feels that a factor increasing variability in performance could both predict extremely high as well as extremely low performance (see Figure 2 for an illustration). However, such a possibility would have to be weighted by the average effect. If a factor slightly increases variability, but has a very strong positive effect on the average, it seems it will not increase the chances of doing worse (as is the case in Figure 1). Whereas, with a strong effect on performance variability but a negligible average effect, the factor increases occurrences of both extremely high and extremely low outcomes (circled in Figure 2).

The first question that emerged was therefore how to formalize this mean-variance tradeoff, and to find a criterion to distinguish when outcomes are more predicted by average effect, from when they are predicted by effect on variability. This also led to the question of whether organizational theory had already dealt with effects on performance variability. Since it appears that such a dependent variable is rarely studied, one may wonder about such neglect and why studying effects on performance variability could matter to organization theorists.

Finally, one may notice that various definitions of variability are possible. For instance, if considering team performance, one can distinguish "population variability," the spread of performance between the best and worst teams in the population, from "individual team variability" such as changes of performance of a focal team[2]. In that second category, one could imagine that performance fluctuates over time. Alternatively, team performance may fluctuate when accomplishing two different tasks concurrently. For instance, if studying the top-management-teams of firms, one could consider the spread between the financial performance and environmental performance. The definition of variability—across-population vs. along-time vs. across-tasks—matters since it conditions what type of organizational outcome each approach could explain.

Turning to the substantive issue of team diversity, little was available to link it to performance variability. Ancient social psychology literature seemed to hint that cohesive and homogenous teams are more risky (Schachter, Ellertson, McBride, & Gregory, 1951), which had been somewhat studied in the group polarization literature (Friedkin, 1999). Those ideas had not led to a formal statement, therefore calling for a theoretical formalization linking greater team homogeneity to greater team performance variability.

Contradictorily, other perspectives suggested the opposite intuition, that very diverse teams are the risk ones. For instance, cross-cultural researchers have hinted that culturally diverse groups could do either very well or very badly (e.g. the Figure 3 reproduced by Adler (2002)[3]). Also, recent learning literature suggests that increasing diversity in teams leads to greater variability (Fleming, 2004; Taylor & Greve, 2006). The apparent contradiction called for a more thorough exploration of the general relationship between team diversity and team performance variability.

---

[2] For those accustomed to the terminology of experimental social sciences, the research seeks to contrast *within-subject* variability to *between-subjects* variability (taking the team as the unit of analysis).

[3] Citing a study by Kovach difficult to trace in bibliographic databases

An epistemological concern appeared when I presented the first results of my research linking team diversity to performance variability. The argument to explain that performance variability is influenced by team diversity relied on a contingent argument along those lines: "for teams where members get along, diversity is beneficial; for teams where members do not get along, diversity is detrimental; there is some randomness regarding whether teams get along or not; therefore, diversity increases variability of performance". Faced with such arguments, some researchers expected an empirical verification of the moderation by the contingency factor (in this example, team integration). However, explicitly measuring or predicting such moderation would not only constrain the empirical verification, but such an angle of research would not be a significant contribution since studies have already explored those (e.g. Van Der Vegt & Bunderson, 2005). Therefore, my own theoretical inductions assume and leverage those moderations but do not seek to prove them.

Rather, I have endeavored to study whether the randomness of real teams that have to perform tasks that are meaningful to its members (i.e. not in experimental conditions) could create effects on performance variability, and whether those effects can be driven by team composition.

OVERVIEW

The dissertation is structured in three chapters that address three different sets of questions using three different empirical approaches. These chapters have been constructed as self-contained essays, which allow distinguishing between the various theoretical and empirical angles of the research program in blocks that are also more readable. The Table 2 summarizes the characteristics of each chapter and identifies the links between them.

Table 2 : Summary of Dissertation Essays

	Chapter I: "Getting Disoriented by Performance Variability …"	**Chapter II: "…Intra-Team Diversity and Extreme Team Performance"**	**Chapter III: "…Distinguishing Effects on Along-Time vs. Cross-Task Performance Variability"**
Theoretical Contribution	Conceptual: variability matters to organizational theory	• Diversity has a U-shaped relationship to variability • Variability approach may alter conclusions (including of existing studies)	• Diversity increases along-time but decreases across tasks variability • Identify social vs. informational hazard of team composition • Indirect confirmation of moderations
Empirics	A simulated dataset	Reanalysis of published study about 35 executive teams	Original archival data on 4 cohorts of 50 students teams
Methodology	Introduces "critical performance" criterion (*continuous* version)	Use *groupwise* version of the criterion	Uses different definitions of variability.
Linkage		Leverages method and concepts exposed in chapter I	Builds on similar theoretical induction as chapter II

The first chapter is a conceptual essay answering the question of "why and how variability matters for organizational theory" using a simulated dataset as an illustration. It does not address any specific substantive issue such as team diversity and could be leveraged in a variety of theoretical perspectives of organizational theory. It provides the motivation to study effects on performance variability and identifies a key criterion—the critical performance level—usable when studying mean-variance tradeoff in an organizational context.

The second chapter is the core essay regarding team diversity theory, and answers the substantive question "what is the effect of team diversity on performance variability." It demonstrates how variability of performance is expected to be large for both highly homogenous and highly diverse teams, as compared with limited variability for mid-range diverse groups. One of the objectives is to demonstrate empirically how taking variability into account may bring about different conclusions from an approach that considers only average effects. Therefore, a dataset that had already been studied and for which some average effect have already been identified was sought. Kilduff, Angelmar and Mehra were graceful enough to allow the re-analyzing of the data of their study (2000) on demographics and cognitive diversity. Methodologically, this chapter leverages the logic developed in chapter I, and demonstrates how the critical performance level can be computed in groupwise testing, when the independent variable is not continuous but contains only two separate bins.

The third chapter was developed with an initial objective of studying the effects of team diversity on different types of variability, as well as verifying—by an indirect approach—the moderations that underlie my research. Since it relies on the same theoretical induction used in chapter II, part of the demonstration is similar; yet, it puts more emphasis on exploring the sources of randomness in social integration and information-to-task fits, and therefore ends up promoting the concepts of informational and social hazards of team life.

CHAPTER I. COMMENT LA VARIABILITE INVERSE LES EFFETS ENTRE LA MOYENNE ET LA PERFORMANCE EXTREME

Quand un facteur augmente la moyenne de la performance, il peut aussi augmenter l'apparition de performances extrêmement basses, à condition que le facteur influence simultanément la variabilité de la performance. Ce compromis moyenne-variance – évoqué par March (1991) dans l'étude exploration-exploitation – a de nombreuses conséquences théoriques car l'ignorance des effets sur la variabilité mène à des conclusions erronées. Je résume les raisons ayant limité l'usage de la variabilité dans la théorie des organisations. De manière constructive, une nouvelle approche est proposée et illustrée avec une simulation numérique. Cette étude contribue aux différentes perspectives de la théorie des organisations telles que l'écologie des populations, l'entrepreneuriat, et les organisations à haute fiabilité (HRO) – où l'amélioration de la performance extrême peut être distinguée de l'amélioration de la performance moyenne.

GETTING DISORIENTED BY PERFORMANCE VARIABILITY: A PEEK THROUGH THE LOOKING-GLASS SEPARATING AVERAGE AND EXTREME EVENTS

When performance variability increases while average performance decreases, more extremely high outcomes can occur simultaneously with decreasing expected outcomes,. This mean-variance tradeoff was first demonstrated in the exploration-exploitation study (March, 1991) and could have wide theoretical consequences, because ignoring effects on performance variability can lead to wrong conclusions. I review the difficulties that have prevented the focusing on performance variability in organizational scholarship. In a constructive manner, a new conceptual approach to this problem is proposed and illustrated with a simulation. This study contributes to the various organizational theory perspectives—such as population ecology, entrepreneurship, and High-Reliability Organizations (HRO)—where extreme outcomes should be distinguished from outcomes on average.

INTRODUCTION

Some of the most and the least desirable outcomes for organizations are extreme. For example, large corporate fiascos such as what occurred at Enron or the outstanding success of the Initial Public Offering (IPO) of Google have disproportionate impacts on all the stakeholders of those organizations. Extreme outcomes may however present particular difficulties for scholars (Baum & McKelvey, 2006). The principal method of inquiry in organizational studies predicts the improvement of average outcome (Mohr, 1982) and one usually assumes equivalence with predicting extreme outcomes.

Even though this approach has its merits, it might be misleading. In the field, some practitioners signal an ambivalent relationship with the scholarly focus on improving expected outcomes (i.e. improving average performance), and appear to see extreme outcomes as distinct phenomena. In aviation, if one talks about cockpit crew performance, one may soon be stonewalled by the response that "there is no good pilot, only old pilots." It suggests that the necessity of survival is more important than average performance, and that different types of logic drive those two outcomes. Among Venture Capitalists (VCs), one encounters similar resistance if advocating that a factor might increase average outcomes by a few percentage points. In the VC world, firms doing slightly better than the average fall into the dreaded category of "zombies," the firms good enough to stay in business, but not good enough to make it big. The goal to perform outstandingly appears more important than average outcome, and those two seem distinguishable too.

These ideas find echoes in scholarship and appear most prominently in the exploration-exploitation study (March, 1991). This showed—using a simulation—how changes in performance variability imply that increasing expected performance occurs simultaneously with decreasing extremely positive performance and therefore reduces survival. Methodologically, this paradox has been studied as a mean-variance tradeoff but in ways that did not apply or were not tractable for organizational studies in general.

Therefore, few organizational scholars explored the widespread theoretical consequences of this paradox, with the notable exceptions of Denrell(Denrell, 2003) and Kalnins (2007). In addition, differentiating extreme outcomes from average outcomes has not become a common approach. Yet effects on variability are already theorized and verified in some organizational studies (e.g. Chatterjee & Hambrick, 2007; Sørensen, 2002; Taylor & Greve, 2006). To make things worse, classical methods have been designed to ignore or mute effects on variability by ways of "robust" statistical tools and the removal of outliers. Overall, organizational researchers may have accumulated a body of prediction that is robust regarding average outcomes but could be systematically flawed about extreme outcomes.

Organizational scholarship could therefore benefit from a constructive approach to combining variability and average effects, if possible parsimoniously conceptualizing the difference between extreme and average outcomes. In particular, when inferences regarding extreme outcomes differ from inferences regarding average outcomes, one may be curious to determine at which performance level the inversion occurs. One may also be curious as to whether such an inversion occurs, and how likely it is. Finally, one may be curious to identify specific organizational theories or contexts where such an inversion matters and should therefore be taken into account.

In this paper, I review how performance variability has been taken into account in previous organizational research and why organizational studies may have a bias towards averages and suppression of variability. Second, I propose a constructive approach to combining variability and average effects, illustrated by a short simulated example. Third, I present the theoretical benefits of considering such an approach for various perspectives.

LITERATURE REVIEW ON PERFORMANCE VARIABILITY

After a brief example to provide an explanation for why changes in performance variability matter, I briefly review organizational theory literature regarding effects on variability. It appears that, contrary to various other sciences like finance, engineering or natural evolution, variability has not gained status as a principal object of study. Yet studies demonstrate that such effects on variability exist in organizational studies and accounts of variability appear in a few disjointed areas of organizational studies.

Why Changes in Variability Matter: the Insight through an Example

The link between effects on outcome variability and extreme outcomes can be intuitive, as illustrated by the following simple example. Imagine a factor taking two values B and T, each of which are associated with a stylized set of performance values: B leads to the values (0, 4, 8) and T to (4, 5, 6) [4]. When considering the effect of the factor, classical organizational theory would only theorize a mean effect, and infer then that T is preferable because its expected value (5) is greater than the expected value for B (4).

[4] B stands for Bottom and T for Top; one could have imagined labeling those Low and High, but for consistency reason with later parts, noting B and T is preferable.

This reasoning assumes that the goal is to improve the average outcome. In contrast, one may seek to improve the chances of reaching a *threshold* of performance. In the field, it could be maintaining the positive value of a financial ratio to avoid bankruptcy or reaching a high value of a metric like revenues that gives access to an Initial Public Offering (IPO). This concept of threshold of performance echoes the one used in population ecology literature that identifies the level of performance where the salient selection outcome occurs (see for instance Barnett, Swanson, & Sorenson, 2003, where various thresholds are considered). In our stylized example above, assume that one seeks to reach at least 8; then, B is more preferable, which is opposite to the conclusion that is reached if one seeks to improve expected outcome. Typically, that factor has an effect on variability (negative from B to T) at the same time as an effect on the mean (positive from B to T). When combined, it makes inferences change at a certain level of outcomes, intuitively somewhere between 5 and 6; for any performance target up to 5, T is preferable, but for any performance target above 6, B is preferable. This example shows that one cannot properly make inferences regarding extreme vs. average outcomes while neglecting effects on variability.

Organizational Research Focuses on Averages

Traditionally, most organizational studies consider effects on performance by using average performance and rarely consider performance variability as a dependent variable. One possible explanation for this neglect is the focus on *explaining away* variance—a common operationalization of variability—on the dependent variable. As in most social sciences, the objective of organizational studies is to predict performance through a more or less sophisticated linear effect of factors on expected performance (Mohr, 1982). For instance, one seminal study explores whether variance in leadership can explain variance in organizational performance (Lieberson & O'Connor, 1972). This study proceeds by systematically eliminating sources of variance on the dependent variable and concluding that leadership accounts for less variance than various other factors such as industry characteristics. In a typical stance, this study attempts to predict what increases expected performance (an average effect) and therefore searches for factors that *eliminate variance* on the dependent variable. As the variability of the dependent variable and variance are related concepts, the focus on *explaining variance away* may cognitively block the use of variability as a dependent variable.

The negative image of heteroskedasticity also motivates the avoidance of performance variability. Heteroskedasticity occurs when an independent variable influences the residual of a regression (Greene, 2003, chapter 11). Most researchers remember heteroskedasticity as problematic (it makes the estimator inefficient, although unbiased); hence, they usually try to eliminate it. A typical procedure is to detect heteroskedasticity by an omnibus test such as White's test. If detected, one removes outliers until the heteroskedasticity seems eliminated, or uses a robust estimator so the variability effect does not disturb the estimation of the average effect. This procedure, ingrained in the research community, is perfectly valid for estimating average effect. However, it may have added to the relative neglect or confusion surrounding performance variability by suggesting it is not an interesting dependent variable.

Another reason to ignore performance variability is a rarely challenged assumption in organizational studies that organizations benefit from reliability. The arguments range from the need to buffer internal processes against uncertainty (Thompson, 1967) to the legitimacy derived from respecting institutional norms of consistency (Meyer & Rowan, 1977). For others, consistency improves organizational autonomy (Pfeffer & Salancik, 1978) and relationships with external stakeholders (Hannan & Freeman, 1984). Population Ecology accepts as a fundamental assumption that "selection in populations of organizations […] favors forms with high reliability of performance" (identified as assumption 1 by Péli, Masuch, Bruggeman, & Nualláin, 1994: table 1). Reliability even appears more important than efficiency in the structural inertia approach to population ecology (Hannan & Freeman, 1984). These traditions lead the few scholars who study performance variability as a dependent variable to assume variability as detrimental to organizational life. For instance, Sørensen, while studying the effects of culture on the reliability of firms' performance, writes about "reliability benefits" to summarize the assumption that increasing performance variability hampers organizations (2002:70).

Finally, variability is sometimes ignored because it is assumed to decrease naturally over time and disappear. Through a cycle of performance and adaptation (Cyert & March, 1963 [1992]), organizations narrow down to well-defined outcomes in which variability is eliminated. The concept of exploitation (March, 1991) embodies this idea of convergence to a narrow outcome. Argote even notes that most models of learning assume that variability diminishes while performance increases on average (1999). If this were true, it might justify a relative neglect of variability—assuming, in addition, that one cares only for the result of the convergence. However, Miner, Haunschild, and Schwab disagree with that general impression, stating that "rules and vicarious learning … may be engines of variability"

(2003:807). They illustrate, in the airline, movie, and biotech industries, cases in which variability grows with experience through various mechanisms.

Overall, the relative neglect of effects on variability occurs by the confluence of statistical tools that prime researchers to eliminate or camouflage variance on the one hand, and substantive organizational theories that assume natural convergence and normative pressure towards reliability on the other hand. Consequently, for many organizational scientists, performance variability does not seem a natural or even valid dependent variable with predictive powers of its own.

Existing Studies Theorizing Effects on Variability

Performance variability is not absent from organizational studies and effects on this construct variable appear in the field[5]. For instance, it has been demonstrated that the strength of corporate culture influences "reliability" of performance (Sørensen, 2002), that intra-team demographic diversity influences "risk" (Fleming, 2004), that team experience diversity conditions "extreme outcomes" (Taylor & Greve, 2006) and that firms with narcissistic CEOs have more extreme performance (Chatterjee & Hambrick, 2007). Hence, such relationships exist in organizational contexts, they just have not been extensively explored yet, and theoretical motivations are lacking. We should note that in those studies, little about extreme outcomes appears, and the balance between mean and variability effects is not studied.

Reasoning on the consequences of variability also appear, more often in studies bordering on other fields—sociology, economics, statistics, or finance. For instance, literature around Bowman's paradox (1980) has debated the nature of the relationship between the mean of returns and the variability of returns, theorizing either psychological (Kahneman & Tversky, 1979) or behavioral (March & Shapira, 1992) mechanisms. This literature focuses on the relationship between the mean and variability, which can be expressed as seeking a relationship between risk (ΔP) and performance (P) (Bromiley, Miller, & Rau, 2001). It differs in its objective from the current study that explores whether and how the effects of a factor (X) on performance variability (ΔP) could nuance conclusions regarding attaining some performance threshold.

Closer to that objective, Kogut (1991) explores how organizational projects such as joint ventures can be analyzed with the concept of financial options and suggests that

[5] The cases where variability is the independent variable are not reviewed here, but they also appear in the literature (e.g. Dineen, Noe, Shaw, Duffy, & Wiethoff, 2007).

variability in outcomes plays a significant role in the evaluation of corporate opportunities. In the conclusion, it even opens up to the generic possibility that "firms, if not organizations, may also profit from uncertainty" and notes that "in some cases they may even seek higher risk" (Kogut, 1991:32)[6]. Cabral explores how firms can favor effects on "variance" when competing in research and development tournaments (2003). Tsetlin et al. (2004) show that "variability can be an important strategic variable in a contest" and state also that—in competitive situations—variability may matter more than effects on the mean.

In addition, the behavioral theory of the firm perspective has acknowledged the potentially crucial role of performance variability. Seminally, the exploration–exploitation study demonstrated how crucial variance effects are in determining the outcomes in competitions where only a few survive (March, 1991). Miner, Haunschild, and Schwab (2003:803) echo that idea, identifying "competitions on extreme values" as situations in which only a few competitors out of many get rewarded and thus where one may benefit from increasing performance variability. In that spirit, a few studies explored how, in the presence of variability effects, inference-making may be subject to a selection bias because of the disappearance of firms (Denrell, 2003; Kalnins, 2007).

However, the selection bias approach does not address the issue of extreme outcomes in general. First, it relies on selective sampling, whereas many extreme outcomes do not imply such selective sampling. Firms successful at extremely positive outcomes, e.g. IPOs, often survive these outcomes and are therefore available to make inferences. Even extremely low outcomes do not imply disappearance. For instance, various firms file for bankruptcy protection at some point and emerge from it without being liquidated. Second, those studies (Denrell, 2003; March, 1991) rely on simulations to study the relationship between variability and average effect, but no constructive approach appears that could be usable in empirical settings.

Overall, literature recognizes that organizations may care about effects on performance variability; but, neither a general approach to deal with it nor the theoretical consequences of such neglect have been identified so far.

Organizational Research Looking Beyond Mean Performance and Towards Extremes

When it comes to the study of extreme organizational outcomes, it appears that some scholars already advocate expanding our focus beyond effects on the mean. Starbuck (1993)

[6] The words uncertainty and risk stand for the concept of variability discussed in the current study.

claimed that organizational studies should focus on exceptional organizational outcomes, and praises the explicit analysis of outliers in an approach that requires the reconsideration of the assumptions about outcomes distribution. Daft and Lewin even recommended the exploration of “heretical methods” in the opening paper of *Organization Science* (1990:6) by proposing the preliminary study of outliers as a potential way of renewing organizational studies. Recently, McKelvey summarized the spirit and intensity of the critique of average effects in organizational studies: “All of the cases used in M.B.A. classrooms are stories about good and bad examples—extremes, never averages … If one thinks of organization and management phenomena as appearing in all sorts of weird shapes, what happens in discipline research is that all these weird shapes are crammed into the square hole of Gaussian statistics” (2006:827).

This critique suggests that the study of extreme outcomes requires considering other distributions than the Gaussian (normal) distribution. Distributions more sophisticated than the normal distribution, for instance fat-tail distributions (McKelvey & Andriani, 2005), deserve attention. These distributions can powerfully model phenomena where the occurrence of extreme outcomes does not die quickly, at least as compared to normal distributions. Fat-tail distributions have particular characteristics such as unstable means and infinite variance. Other fields have proved their effectiveness in modeling phenomena such as earthquakes, traffic jams, and epidemics that follow non-standard distributions with fat tails (Baum & McKelvey, 2006:128).

Even though one should acknowledge the potential benefits of such advanced considerations, the study of extreme outcomes may still gain ground without invoking fat-tail distributions. Many scientific fields have moved progressively from (a) predicting average effect to (b) taking into account variability to (c) finally using non-normal distributions with fat tails. For example, the volatility of financial assets—a measure of variability—and their relationship to valuation have been the cornerstone of financial theory for decades (Black & Scholes, 1973). In that field, suggestions to reconsider distributional assumptions—toward fat tails distributions—appeared early (Mandelbrot, 1960). However, even though many contemporary scholars acknowledge its potential (Taleb, 2007), focus on exotic distributions has so far had a minor role compared to theories studying effects on variability. It would be awkward for organizational studies to move directly from (a) focusing on averages to (c) focusing on fat tails without first reaping all the theoretical benefits of (b) focusing on variability effects.

Overall, a gap appears in the study of extreme organizational outcomes. On the one hand, most organizational studies focus on average outcomes, ignoring extreme outcomes and the effects of variability. On the other hand, one emerging stream of research suggests focusing on extreme outcomes by overhauling the statistics we use. There appears a need for a mid-range approach focusing on variability effects, which would allow better predicting extreme outcomes without requiring drastic changes of our statistical models.

WHEN VARIABILITY MAKES INFERENCES ABOUT EXTREME OUTCOMES THE INVERSE OF THOSE ABOUT AVERAGE OUTCOMES

If classical mean analysis provides a simple framework to study expected performance (Mohr, 1982), no such clear approach is available to link variability analysis to extreme outcomes. In this section, I propose a constructive approach to the mean-variance tradeoff that is adapted to organizational theory. To allow differentiated predictions between extreme vs. average outcomes, it determines the separation between the two performance ranges where the causal factor influences performance in opposite directions. The literatures on economics (Cabral, 2003) and statistics (Tsetlin, Gaba, & Winkler, 2004) develop related ideas but without proposing a compact and simple approach that could be used in a large range of organizational studies. Currently, few organizational studies reason on a tradeoff between a mean and a variance effect. If average effects are supported by a conceptual approach—as advocated by Mohr (1982)—and embodied methodologically in the regression line, no equivalent conceptual approach is available in our field to study effects on variability.

The constructive approach proposed below does not intend to be a statistical treatise. Obviously, the reasoning relies on a strong methodological approach that could be reused in the future. However, the intended final contribution is to establish that effects on performance variability have various theoretical consequences (next section), which can be taken into account with a conceptual simplicity that comes close to the conceptual simplicity of the regression line (this section).

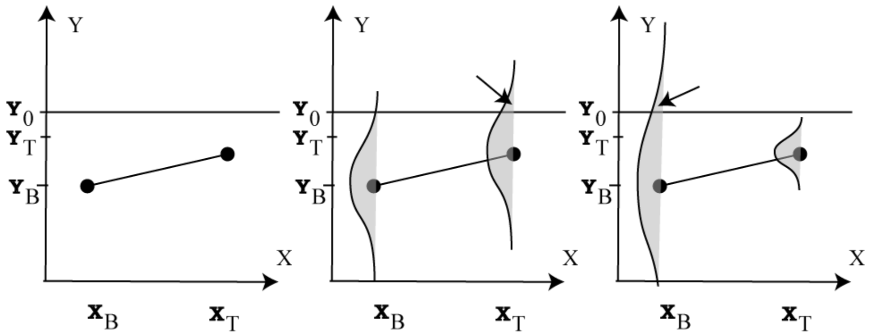
Regarding statistics and methodology, some details are provided in the appendix, but most of the reasoning is based on graphical arguments that only require basic statistical concepts. The current study shares the same methodological logic as studies on selection biases (Denrell, 2003; Kalnins, 2007) but aims to further generalize the theoretical consequences and provide a generic conceptual approach to those problems. For those interested in linking the approach to a statistical technique, the current study relates to quantile regressions (Greene, 2003:448)[7].

Inference Making is Simple without Changes in Variability

Let us now progressively discover how inferences are influenced by the presence of variability effect. Let us start by first revisiting the premise presented at the beginning with a few sketches (Figure I-1). Imagine a factor X has the effects on a performance random variable Y (now assumed normal), as summarized by Figure I-1 (a): the average of Y is higher for the top values of the factor (X_T) than for its bottom values (X_B). Now, imagine the goal is to reach a target performance threshold Y_0 above the average value of performance. The question of interest is: for which values of the factor is performance more likely to reach that threshold?

Figure I-1: Attainment of a Threshold Y_0 Depends on the Variability Effect

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



Traditionally, only the mean effect is considered. Here, X increases the mean value of Y, so one infers that X increases the chances of performance reaching the threshold. This implicitly assumes that the variability in Y does not change with X, as is the case in (b), with the bell curves sketching the distribution of Y having identical variability. The probability of reaching the threshold performance Y_0 corresponds to the area in the tail of the distribution above the threshold. With constant variability, the higher mean of Y at X_T implies a greater size of the tail above the threshold and therefore more chances to reach the threshold at X_T than at X_B (see the arrow in Figure I-1 (b)).

[7] On the issue of statistics, one should not confuse the objective of the current study with some existing technical issues of statistics. For instance, some estimation techniques evoke thresholds on the dependent variable, for instance, when dealing with censored data (Greene, 2003: Chap. 22), or discrete levels (Greene, 2003: Chap. 21); by contrast, even though I consider here threshold of performance, the issue is rather to show why various theoretical perspectives where threshold play a particular role should consider effects on variability. The issue here is nor about techniques to optimize estimation of first order effects when faced with effects on variability, usually referenced as heteroskedasticity (see Greene, 2003 for comprehensive treatment).

If no effect on variability exists, inferences around any threshold reach the same conclusion—as indicated by the average effect. In other words, making inferences around the average performance threshold is strictly equivalent to making inferences around any high or low threshold. To be convinced, one should notice in Figure I-1 (b) that—assuming a constant distribution of Y—the cumulated probability to reach any given threshold is always superior in X_T, as the distribution of outcomes is simply shifted upward when the factor increases.

Proposition 1: When no effects on performance variability exist, inferences around any threshold are all equivalent

In the absence of variability effects, conclusions drawn from regression analyses apply to any performance level. For instance, assuming no effect on variability, if one finds by a regression a positive coefficient of X on Y, not only does X increase the chances of reaching average performance, but it also increases the chances of reaching any threshold of performance. This conclusion concurs with the classical approach assuming mean effect predicts extremes—so long as no variability effect exists.

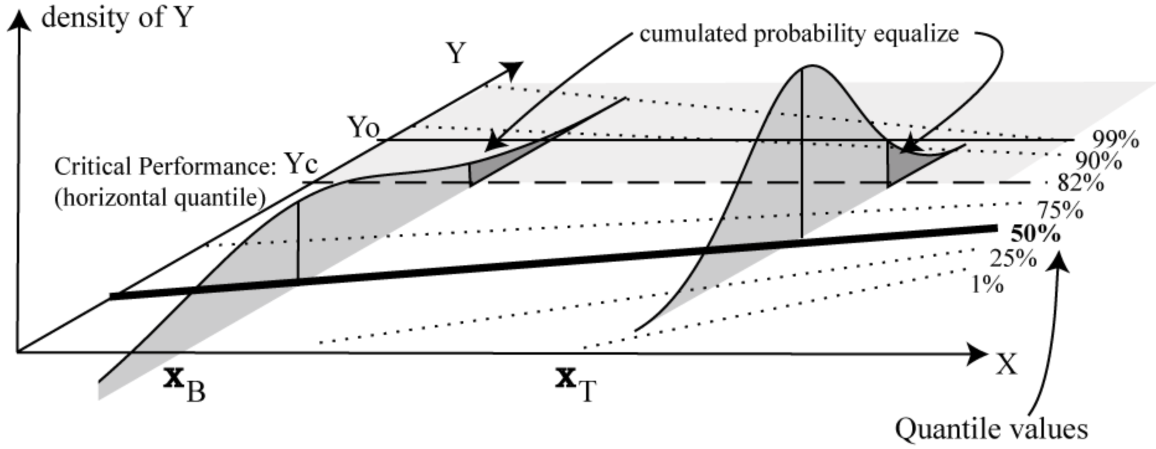
With Effects on Variability, Inferences Inverse at a Critical Performance Level

If performance variability changes, the generality of inferences does not hold anymore. Figure I-1 (c) matches Figure I-1 (a) for the mean effect but now assumes that X reduces the variability of Y, as represented by a bell curve that is more diffuse at X_B than at X_T. The probability of reaching the threshold is now higher at X_B (indicated by the arrow), even though the mean value is still higher at X_T. This is the same reasoning as the one used for the small numeric example proposed at the beginning, where a variability effect can compensate for a mean effect when trying to reach a performance threshold.

Current organizational literature does not conceptualize that inferences differ depending on the performance target. There is no simple reasoning available for balancing variability with average effect, something that would be similar to the slope of the regression line used when studying average effects (Mohr, 1982). This section proposes combining variability and mean by defining a criterion—the *critical performance* level—around which the inferences change. It still relies on a graphical argument, albeit slightly more sophisticated than the previous ones. The appendix presents the assumptions on the distributional properties of performance and presents one possible formula to compute critical performance level.

Let us assume that performance follows a normal distribution and is related to the factor by a simple linear relation (the reasoning in the appendix shows that this holds with more generality). In Figure I-1 (c), imagine replacing the sketchy bell curves by drawing a few of the lines that link the points where performance is equally likely—traditionally called the *quantile lines*. An obvious line is the median line (at the 50% quantile), which approximates the regression line. When accomplished for a few other values, it results in Figure I-2—a more-detailed version of Figure I-1 (c). Now, one can simply read the chances—expressed by the quantile lines—of reaching any threshold. In the diagram, observe the position of the threshold Y_0 relative to the 90% quantile line. It appears that for low values of X, there is less than a 90% chance that performance remains below the threshold (so more than 10% goes above), and for high values of X, there is more than a 90% chance it remains below the threshold (so less than 10% goes above). One can therefore conclude that the factor *decreases* the chances of reaching the threshold.

Figure I-2: Quantile Lines Represent Both Mean and Variability Effects



Hence, if one considers the quantile lines, an inference around a 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*.

Proposition 2: The direction of an inference around a threshold of performance Y_0 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 factor.

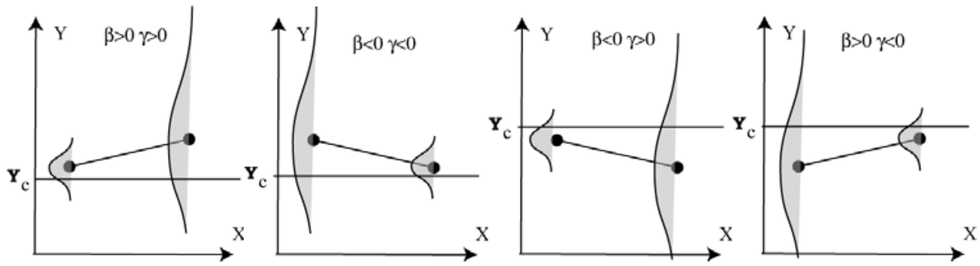
Under reasonable assumptions, the slopes of the quantile lines change direction once (see appendix for proof). In Figure I-2, the quantile at 50% has a positive slope because X has a positive effect on the average of Y; when going up the diagram, the quantile line progressively lose slope because of the higher performance variability for low values of X. At some level, 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 point, the quantile lines get a negative slope.

The important consequence is an inversion of the direction of the effect: for thresholds around average values of performance, all inferences go in the direction indicated by the mean effect; for thresholds beyond the critical level, all inferences indicate effect in the other direction. If the critical level is above average, the inversion occurs above it, if the critical level is below average, the inversion occurs below it. For instance, with the assumptions of the current examples and reading the assumed values of the quantile lines in Figure I-2, we can draw the following conclusions: if the threshold falls on the critical quantile (here 82%), the conclusion of the inference is neutral, as *X* has no effect on the chances of reaching that level. For a threshold around average values of *Y*, the quantile lines grow, so one would infer that *X* has a positive effect on the chances of reaching that threshold. For a threshold above the critical level, the quantile lines decrease, so one can infer that *X* has a negative effect on the chances of reaching that threshold:

Proposition 3: Inferences around thresholds close to average performance have a direction opposite to inferences around thresholds beyond the critical performance level.

The critical performance occurs on one side of the distribution of performance, bounding the part of the distribution that is driven more by performance variability than by its mean. By definition, that part of the distribution is a fraction that is lower than 50%. Its position of the critical performance, relative to average performance, depends on both the mean and the variability effect: critical performance occurs below the median line when mean and variability effects are in the same direction; otherwise, critical performance lies above the median line (illustration in Figure I-3).

Figure I-3: Critical Performance Position



The mirroring of effects when crossing the critical performance level is the central mechanism justifying that one carefully takes into account variability and threshold when studying organizational outcomes. Overall, this approach provides a construct (threshold of performance) and a computable criterion (critical level) to determine when inferences around extreme outcomes differ from inferences about average outcomes.

A particular case of that general situation should be mentioned because it illustrates the power of considering variability effects. Imagine a situation where a variability effect occurs—for instance negative—but no mean effect exists (simple illustration in Figure I-4). It implies then that the factor decreases occurrence of both extremely high and extremely low outcomes. This situation is remarkable since classical method would not allow any inference when the mean effect is insignificant. Now, two inferences are available and they paradoxically predict both beneficial (on increasing extremely high outcomes) and detrimental (on decreasing extremely low outcomes) effects at the same time:

Proposition 4: In the absence of average effect, increasing performance variability implies both more extremely low and more extremely high outcomes.

Figure I-4: No Mean Effect with Variability Effect Predicts both Extremely Low and Extremely High Outcomes

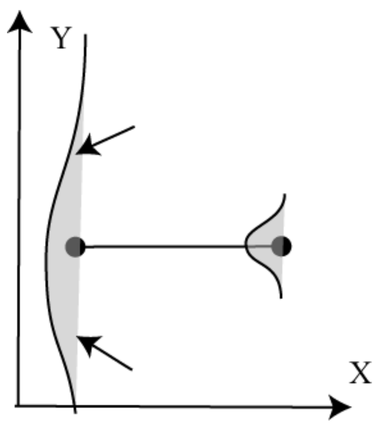


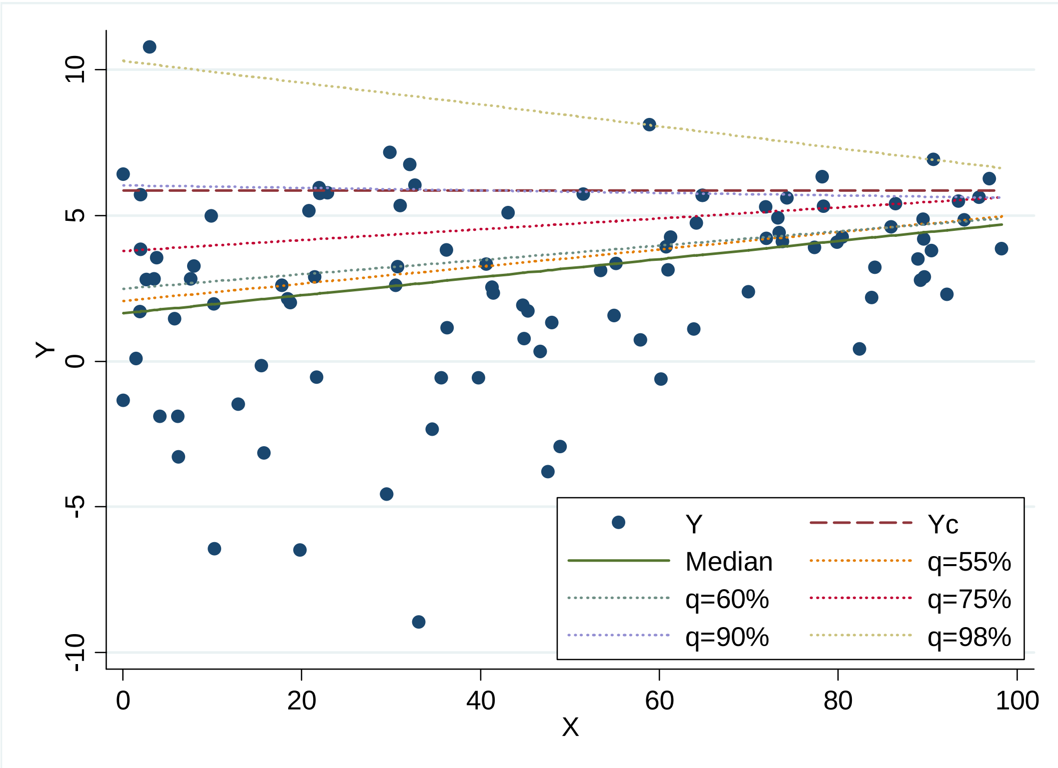
Illustration by a Simulated Example

A simulated example can illustrate the reasoning given above for the general case where both mean and variability effects occur. I generated a dataset linking an independent variable *X* with a performance variable *Y* by both a mean effect and a variability effect. The relation was modeled on a version of the generic parameterization exposed in the appendix, which was further simplified using the standard normal distribution to produce:

Equation 1 $Y: N(\mu,\sigma)\ ,\ \mu = \beta_1 X + \beta_0\ \sigma = \gamma_1 X + \gamma_0.$

I generated 100 points to match a classic order of magnitude in organizational studies, with X randomly uniformly distributed between values 0 and 100. With regard to the effects on the dependent variable *Y*, I picked a positive mean effect (β_1=0.05) and a negative variability effect (γ_1=—0.04). Various values and seeds of the random generator function were tested to build an example that would be illustrative. A scatter plot of the data, which includes the median line (an estimation of the regression line), appears in Figure I-5. The effects were made strong enough so that the heteroskedasticity is visible, with a decrease in the dispersion of points around the regression line.

Figure I-5: Scatter Plot including Regression Line and Critical Level Surrounded by Quantile Lines of Converging Slope



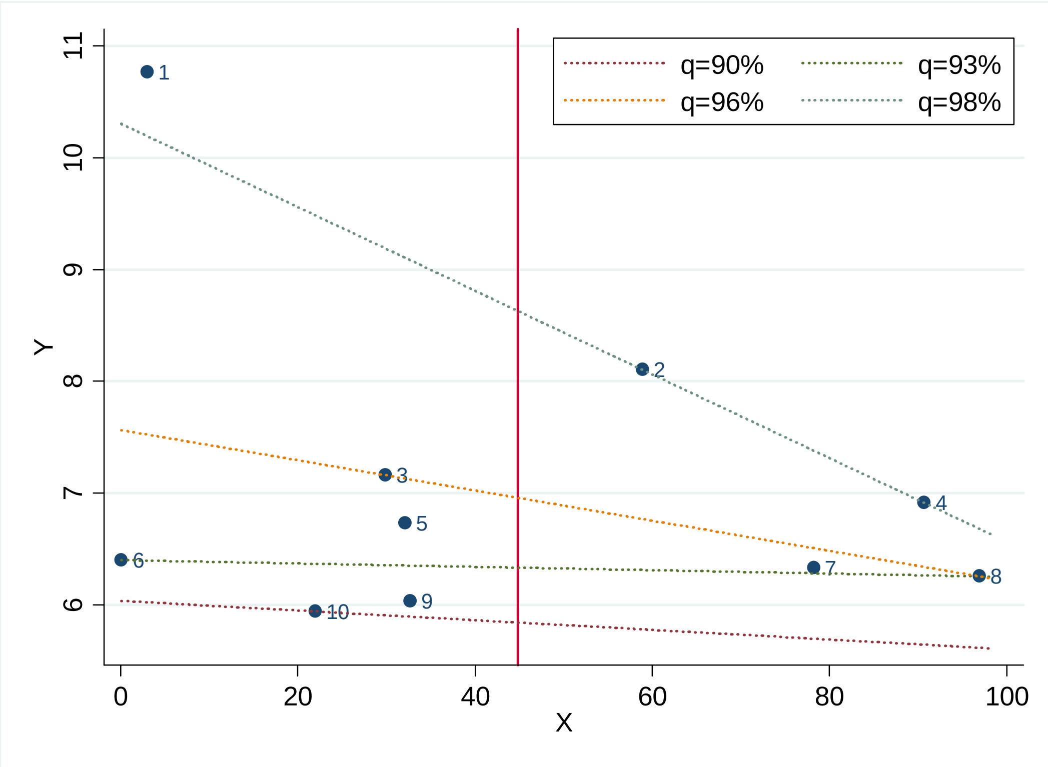
A first approach for determining the level of critical performance is to make successive approximations, plotting various quantile lines and figuring out the value at which the quantile line is horizontal. I computed and plotted the quantile lines using the quantile regression Stata procedure (qreg), which suggests a value of the critical level around 6. In addition, I used the formal approach proposed in the appendix that takes the mean and variance effects as inputs. A maximum likelihood estimation (MLE) on Equation 1 provides estimates of the parameters in the simulated data set (β_1=0.039 and γ_1=-0.037, consistent with the underlying actual values). Equation 2 of the appendix computes the critical level (Y_c=5.86).

Once the value of critical performance is established, the percentile at which it occurs may be determined by a graphical method as a first estimation (between 75% and 90% in Figure I-5). A formal determination method is also possible. Estimates of the mean and variance effects can be used to compute the mean and standard deviation of *Y* for any arbitrary value of *X*. I use the average of *X*, and assuming a normal distribution for *Y*, I took the inverse normal of the *z*-score of Y_c, which lead to the percentiles of Y_c at z_c=82%. By definition of Y_c, this percentile would be the same for any value of *X*, thus identifying the percentile of the critical performance.

To interpret it, notice that the example in Figure I-2 has been constructed for values similar to the results here, so the quantile plot here can be interpreted in the same way. The median line (the quantile at 50%) and all the quantiles around the average values have a positive slope, which is consistent with the positive effect of *X* on the average of *Y*. At the critical level—the 82% quantile—the expectation of performance does not depend on *X*. Beyond that value, the quantile lines take a negative slope.

A final graph (Figure I-1) illustrates the behavior beyond the critical performance level. Only the range above the critical line (Y_c = 5.86) is plotted, and a few additional quantile lines in that range were added. The quantile lines now have a clear negative slope in that range, which hints at a negative effect of *X* on the chances of reaching any threshold there. To give an example of such a phenomenon, the points are marked with their ranking in the performance range—number one being the point with the best performance. I separated by a vertical line the points for the low values of the factor X on the left, from the one with the high values of X on the right.

Figure I-6: Zoom on Performance Range beyond the Critical Level



For the points that fall in any top percentile smaller than 18%, performance is more likely to be superior for low values of *X* than for high values. For instance, if inferences are made only based on who finishes first (i.e. top 1%), low values of *X* are more favorable and, indeed, capture the top score. If inferences are made about being in one of the top three positions (i.e. top 3%), low values of X are again preferable and indeed capture two of the top three scores. If inferences are made about being in the top 10%, low values are still more favorable, and indeed capture 6 of the top 10 scores[8]. More generally, if inferences are made by observing the attainment of any performance threshold above the critical performance of Y_c (~5.9)—or, if expressed as a ranking, those who belong to any top percentile smaller than the critical risk (18%)—low values of *X* seem preferable to high values of *X*.

By comparison, while making inferences around average values—for instance, about the attainment of the average performance (~2.67)—one would conclude that high values of *X* are preferable. This illustrates clearly that inferences made about extreme outcomes can be opposed to the inferences made about average outcomes.

[8] Let us be clear that sample size should not play any role here. If necessary, imagine the sample was large enough that inferences are significant at one's level of comfort.

THEORETICAL CONSEQUENCES OF NEGLECTING VARIABILITY

The review showed that organizational studies have focused on average effects and ignored variability. Then, the approach suggested above clarified why and how neglecting effects on variability can be problematic, since inferences about average outcomes may contradict inferences regarding extreme outcomes. This potential contradiction calls for identifying clearly the consequences of such neglect.

Effects on variability have serious theoretical consequences in situations with such a pattern: one observes the effects of a factor on the chances of reaching a given performance threshold, and variability effects make it so that such inferences do not hold if considering another performance threshold. Typically, inferences regarding average outcomes apply to expected outcomes, but may be misdirected when it comes to extremely low or extremely high outcomes. Alternatively, inferences regarding extremely low or high outcomes may be misdirected regarding average outcomes. These four ideal cases will be discussed below, and summarized in summary Table I-1.

A few studies have already identified situations where one might get confused about the true effects of some factors due to sample bias (Denrell, 2003; Kalnins, 2007). This section follows and extends this reasoning in two directions. First, it generalizes this logic for situations where there is no sample selection bias and where the problem appears around various performance levels (low, high or average). Second, it applies mainly to the inference-making process of organizational scholars, whose research method or design determines the observation threshold, and for which variability effects endanger the generality of predictions. That having been said, sample selection bias studies would fit in the fourth case (B.2. in Table I-1) where one observes extremely low outcomes and is confused regarding average outcomes.

Average-based Inferences that Do Not Apply to Extremely High Outcomes

The first two cases deal with situations where one makes inferences based on *average* outcomes, inferences that are therefore potentially not applicable to *extreme* outcomes. In the first ideal case, one makes inferences regarding extremely *high* outcomes, based on average effects, inferences that turn out to be misdirected. This could occur when the average and variability effects are not in the same direction; for instance, if a factor has a positive average effect but a negative effect on variability (case A.1. in Table I-1). Applying a classical analysis (Mohr, 1982) to that context, one would find that the factor is beneficial. In contrast, taking effects on variability into account predicts a critical performance level beyond which the factor reduces the chances of *reaching extremely high outcomes*. This contradiction matters to all theoretical perspectives where extremely high outcomes play a role distinct from average outcomes.

For example, extremely high outcomes matter in entrepreneurship. Having a greater chance of ranking at the top of one's cohort, for instance when trying to reach IPO stages, may be more predictive of final success than improving performance on average. In the high-technology market, only a few players in each market reach the IPO stage; the crowd of other entrants to that early market dies out after the market matures around the few that are properly funded by IPOs. Hence, inferences drawn from normal studies may wrongly predict the attainment of any extremely high performance threshold beyond the critical level. In simple words, a factor may appear to increase performance on average while it actually reduces the chances of reaching IPO.

By systematically clarifying the performance thresholds (high, low, average) that apply to each context, and taking into account variability effects when making inferences, entrepreneurship scholarship may progress both in accuracy—by capturing the right direction of effects—and relevance—by nuancing conclusions on the nature of the outcome sought. This logic would apply to various other fields where exceptionally high outcome may be sought. For instance, one may feel that outstanding products like the Macintosh and iPod may be driven by variability effects more than accumulation of average effects, suggesting a revisit to innovation studies armed with a variability lens.

Average-based Inferences that do not Apply to Extremely Low Outcomes

The second ideal case occurs when one makes inferences on *average* again, but which turn out to be misdirected regarding *extremely low* outcomes. This could occur when the variability and average effects are in the same direction, for instance when a factor has a positive average effect and a positive variability effect (case B1 in Table I-1). Applying classical analysis, one would find that the factor is beneficial. By contrast, taking the effects on variability into account predicts a critical performance level below which the factor reduces the chances of *avoiding extremely low outcomes*. This contradiction matters to all theoretical perspectives where such extreme low outcomes play a role distinct from average outcomes.

For example, extremely low outcomes matter in the governance perspective. Business organizations, especially large traded firms, seek to avoid bankruptcy because of its disproportionate impact on stakeholders (Sutton & Callahan, 1987). Avoiding large bankruptcies, or fiascos in various social or environmental areas, becomes an important organizational goal, which therefore signals the existence of a low threshold of performance to avoid. In such contexts, relying on inferences based on average outcomes may be problematic. To put it simply, a factor may increase performance on average while it actually increases the chances of fiascos.

In contexts like governance, variability analysis brings benefits by making inferences contingent on the target performance threshold, which allows distinguishing between, for instance, the improvement of firm financial measures from the avoidance of bankruptcy. This logic would apply to various other fields where exceptionally low outcomes matter, such as corporate social responsibility (CSR) and the study of high reliability organizations (HRO).

Inferences about Extremely High Outcomes that do not Apply to Averages

The next two cases deal with situations where one makes inferences by observing *extreme* outcomes that are potentially misleading regarding *average* outcomes. The claim is that studies modeling the attainment of an extreme performance target can only be generalized to other performance levels—in particular average performance—if the effect of variability is clarified.

In the third ideal case, one makes inferences based on extremely *high* outcomes, which turn out to be misdirected when it comes to average outcomes. This could occur when average and variability effects are in opposite directions, for instance if a factor has a positive average effect but a negative effect on variability (case A.2. in Table I-1). By observing the attainment of the high threshold, one would find that the factor is detrimental. However, that could be due to an effect on variability that is masking the fact that the factor is beneficial on average. This distinction matters for all theoretical perspectives where average outcomes differ from extremely high outcomes.

For instance, following the leadership example above, one may be studying the factors that lead managers to become the CEO of a large firm. This outcome is by definition extremely high, at least relative of the population of managers. If a factor increases the average performance of managers but decreases variability, it is therefore possible that it decreases the chances of becoming CEO, even though it increases expected performance. Such logic could, for instance, provide mechanisms for literature that study the contradictions surrounding CEO selection (e.g. Khurana, 2002).

This logic considering variability and making inferences contingent to target performance level would benefit other perspectives that focus on high level outcomes, for instance population ecology (e.g. March, 1991 demonstrates how survival at a high threshold differs from average).

Inferences about Extremely Low Outcomes that Do Not Apply to Averages

In the fourth ideal case, one makes inferences based on extremely *low* outcomes, which turn out to be misdirected when it comes to average outcomes. This could occur when average and variability effects are in the same direction, for instance when a factor has a positive average effect and a positive variability effect (case B.2. in Table I-1). By observing attainment of a low threshold, one could find that the factor is detrimental. However, that could be due to an effect on variability that is masking the fact that the factor is beneficial on average. This distinction matters for all theoretical perspectives where average outcomes differ from extremely low outcomes.

For instance, population ecology relies on observing the survival or death of organizations. In some contexts, death is a low outcome, i.e. it is rare enough so that most firms survive. If a factor increases variability, it may however increase death rate, even though it increases performance on average. This reasoning is closely related to the argument and modeling developed in the studies on selection bias (Denrell, 2003; Kalnins, 2007): estimation of survival strongly depends on variability, somewhat independently of mean effect. The idea that population ecology could distinguish various threshold has been proposed by Barnett, Swanson and Sorenson (2003). However, the logic was that various processes may be at work (entry vs. exit) and that the threshold may be different. The current reasoning proposes that various possible thresholds be considered, typically that a study predicting survival at a low threshold also consider effects on average[9]. Effects at that level could be inverted if the selection threshold is beyond the critical level.

Overall, perspectives that study extremely low outcomes—such as evolution or population ecology mechanisms with low selection threshold—may benefit from taking into

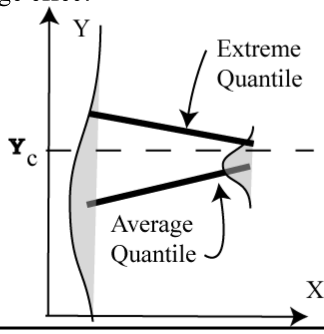
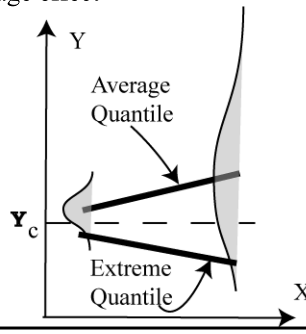
---

[9] Scholars might have not distinguished extreme performance vs. average performance because extreme outcomes are often available as a binary variable identifying whether a natural performance target has been attained, whereas no information is available on the full scale of outcomes where such extremes fall. For instance, if studying promotion to CEO, only that binary dependent variable is available and one does not have access to any underlying continuous performance scale of CEOs, into which promotion would occur at a defined level. Therefore, one may have difficulty distinguishing extreme cases vs. averages since the empirics only provide a binary outcome. Yet, the current study offers motivation to search for any scale of performance—potentially proxies such as salary—to check whether mean-variance tradeoff may endanger the generality of inferences. Typically, in the example of promotion to CEO (a binary variable), one could check the effects on salary, a more continuous and less censored scale, to determine whether there is a critical level where inferences invert, therefore signaling that what is good for becoming a CEO may be bad for other managers.

account the effects of variability and the contingence of inferences to the level of performance observed.

Table I-1: Illustration of Theoretical Consequences of Mean-Variance Tradeoff

The table considers a hypothetical situation where a factor X improves expected outcomes. By column it varies the effect on variability (A negative, B positive), and by rows varies whether (1) one makes inferences based on average but tries to apply those to extremes or (2) makes inferences around an extreme and tries to apply it to averages. For each diagram, the approximate position of the critical performance level where inference changes direction is indicated with a dotted line; both the average quantile (i.e. the regression line) and a representative extreme quantile on the other side of the critical level are drawn with thick lines.

		A. Variability effect opposite direction of average effect	B. Variability effect same direction as average effect
			
	Theories Impacted	When extremely *high* outcomes play a distinct role from average outcomes	When extremely *low* outcomes play a distinct role from average outcomes
1. Observing averages and applying to extremes	**Observed Effect**	Observing effect around average performance: X is beneficial	
	Contradiction Regarding	Extremely *high* outcomes	Extremely *low* outcomes
	Actual Effect	X is detrimental	
	Theoretical Example	E.g. **Entrepreneurship** “X appears to increase performance on average … but it actually reduces chances to reach IPO” Also: **Innovation**	E.g. : **Governance** “X appears to increase performance on average … but it actually increases fiascos” Also: **Corporate Social Responsibility, High-Reliability-Organization**
2. Observing extremes and applying to averages	**Observed Effect**	On extremely high outcomes:	On extremely low outcomes:
		X is detrimental	
	Contradiction Regarding	Average outcomes	
	Actual Effect	X is beneficial	
	Theoretical Example	E.g. **Leadership** “X appears to decreases chances to become CEO of a large firm … but it actually increases expected performance” E.g.: **Population Ecology** with high threshold, as evoked in (March, 1991)	E.g. : **Population Ecology** with low threshold “X appears to increases bankruptcy … but it actually improves expected performance” Similar to (Denrell, 2003; Kalnins, 2007)

FUTURE DIRECTIONS

Variability can have serious and widespread theoretical consequences for organizational studies. The obsession with expected performance and treating performance variability as an effect to be eliminated has hidden the possibility that conclusions could depend on the performance level considered. Thereafter, various theoretical perspectives may progress both in accuracy—by capturing the right direction of effects—and relevance—by nuancing its conclusion on the level of the outcome sought. This applies in particular to the various perspectives interested in extreme outcomes such as entrepreneurship, high-reliability organizations, governance, or leadership. It also applies to perspectives that typically study binary outcomes (survival, promotion) such as population ecology or leadership.

The current study suggests various future directions for research. The most obvious one is to start accumulating a body of knowledge on factors that influence performance variability. Until now, most organizational research has focused on average effects. Yet, the current study and a few empirical studies suggest that effects on variability may appear and have consequences contrary to what is predicted by averages effects. Such focus on variability could occur as freestanding research or simply appear more systematically as a complement of any research with an initial mean effect objective. As past research has explored risk as a trait of individuals or as a behavioral research, future research may find that many factors determine organizational risk by influencing performance variability.

An alternative path of research may explore whether such reasoning would create a state of myopia for managers in the field. For instance, one may study how the outstanding success of a few firms and entrepreneurs such as Bill Gates and Jeff Bezos may have influenced the larger population of managers and entrepreneurs in ways that may be interestingly dysfunctional, because such learning (inference making by managers) ignores variability effects and may be misdirected regarding the effect on the average manager.

Contemporary organizational life increasingly provides examples of extreme outcomes, with the mass media ensuring that such outcomes get a disproportionate share of attention. It would be paradoxical if organizational theory could neither caution about what to learn from extreme outcomes nor predict them differentially from average outcomes.

CHAPTER II. MIEUX, MEILLEUR, OU PIRE ? DIVERSITE D'EQUIPE ET PERFORMANCE EXTREME

La diversité intra-équipe a des effets contradictoires, améliorant à la fois l'information disponible et handicapant la socialisation. Il en résulte alors un effet moyen neutre sur la performance de l'équipe. Considérant la variabilité de la performance, je prédis un effet curvilinéaire (U) de la diversité. La ré-analyse d'une étude de terrain sur 35 équipes de dirigeants en formation confirme cet effet sur trois variables de diversité démographique et deux variables de diversité cognitive. Une analyse combinant l'effet sur la moyenne et sur la variabilité améliore les prédictions ; par exemple, concernant la diversité d'âge, la conclusion fournie par l'analyse conventionnelle s'inverse si l'on considère que le but est d'atteindre un très haut niveau de performance.

BETTER, BEST OR WORST TEAM: INTRA-TEAM DIVERSITY AND EXTREME TEAM PERFORMANCE

Intra-team diversity has been shown to have contradictory effects, both improving information and hindering social integration, resulting in a neutral average effect on team performance. I predict a U-shaped relationship of diversity to performance variability, improving on the past prediction of a simple increase. The reanalysis of a field study of 35 teams in a business simulation confirms such an effect for three demographic and two cognitive diversity variables. An analysis combining both mean and variability effects leads to improved predictions and to the conclusion that the effect of age diversity drawn from a conventional mean analysis reverses 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, & Barsade, 1998) and have puzzled researchers for decades with two views, one optimistic and one pessimistic, that directly contradict each other (Mannix & 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, & 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 & O'Reilly, 1998). Scholars seeking to reconcile these findings have either defined diversity with increasing subtlety (e.g., Bunderson & Sutcliffe, 2002; Harrison & Klein, 2007) or introduced variables to moderate the relationship between diversity and performance (e.g., Polzer, Milton, & Swann, 2002; van Knippenberg, De Dreu, & 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, & Schwab, 2003), one cares more about the chances of a top ranking than about improving performance on average. For instance, in a winner-takes-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 (Baum & McKelvey, 2006; March, 1991). 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. In the context of team diversity, most research has explored its mean effect and does not adequately address the question of whether diversity increases the probability of 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 as they have 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 & Vermeulen, 2003; Richard, Barnett, Dwyer, & Chadwick, 2004; Uzzi & Spiro, 2005).

In a population of teams at high diversity levels, the lowest-performing teams are likely to suffer from conflicts; greater diversity then worsens their situation by further degrading social integration. At the same time, the best-performing teams are likely to have members who are on good terms with one another, 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 & Greve, 2006), except that I expect it to occur for only high levels of diversity and not over the whole range.

In a population of teams at low diversity levels, the lowest-performing teams are likely to suffer from a 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 the effects of diversity on performance variability into account 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 & Schippers, 2007). I first review the literature on this relationship and the approaches used to make it less ambiguous. Then I present the alternative focus on the effects of diversity on performance variability, and explain why such effects matter.

Team Diversity Research Sorting out Good vs. Bad Outcomes

Researching a simple relationship between intra-team diversity and team performance lead to two competing perspectives. The first focuses on information availability and takes an "optimistic" view of diversity (Mannix & Neale, 2005:33): intra-team diversity brings more information to the team, which improves expected team performance (Dahlin, Weingart, & Hinds, 2005; Gruenfeld, Mannix, Williams, & Neale, 1996). For instance, exposure to minority thinking fosters a broader view of the issue at hand (Nemeth, 1986; Page, 2007), and diverse teams exchange a wider range of information (Sommers, 2006). By contrast, a social integration perspective (O'Reilly, Caldwell, & Barnett, 1989) combine the similarity-attraction (Festinger, 1954) and the self-categorization (Tajfel, 1982) approaches and takes a "pessimistic" view of diversity (Mannix & Neale, 2005:34): intra-team diversity creates tensions in the team, which damages expected team performance.

These two perspectives suggest contradictory effects, so finding a simple main effect has proven elusive. Some authors have found results—linear or U-shaped (Earley & Mosakowski, 2000) or inverted u-shaped (Dahlin, Weingart, & Hinds, 2005) or with even more complex shapes (Allmendinger & Hackman, 1995)—but starting with Nemeth and Staw (1989), various meta-analyses (Bowers, 2000; Webber & Donahue, 2001) and reviews have settled on an overall neutral effect of diversity (Jackson, Joshi, & Erhardt, 2003; Milliken & Martins, 1996). To address this ambiguity, at least three major streams of research have emerged (Mannix & Neale, 2005; van Knippenberg & Schippers, 2007). The first has focused on the independent variable, diversity, by adopting ever finer definitions (e.g., Boone, Van Olffen, & Van Witteloostuijn, 2005; Bunderson & Sutcliffe, 2002; Cramton & Hinds, 2005; Lau & Murnighan, 1998). A wealth of definitions has led to a necessary theorization and classification of the various possible approaches to the diversity construct (Harrison & Klein, 2007). A second stream of research, following Lawrence's recommendation to "open the black box of demography" (1997), explores the mediating processes, for instance distinguishing the effects of diversity on emotional vs. task conflicts, those leading to effects on performance (e.g. Pelled, Eisenhardt, & Xin, 1999). Finally, a 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, Barnett, Dwyer, & Chadwick, 2004), intra-team perceptions (e.g., interpersonal congruence in Polzer, Milton, & Swann, 2002) or context (e.g., people orientation of corporate culture in Kochan et al., 2003:10).

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 (Cramton & Hinds, 2005; Lau & Murnighan, 1998).

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, De Dreu, & Homan, 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 a higher level of performance and other teams to a lower level of 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 continues with 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.

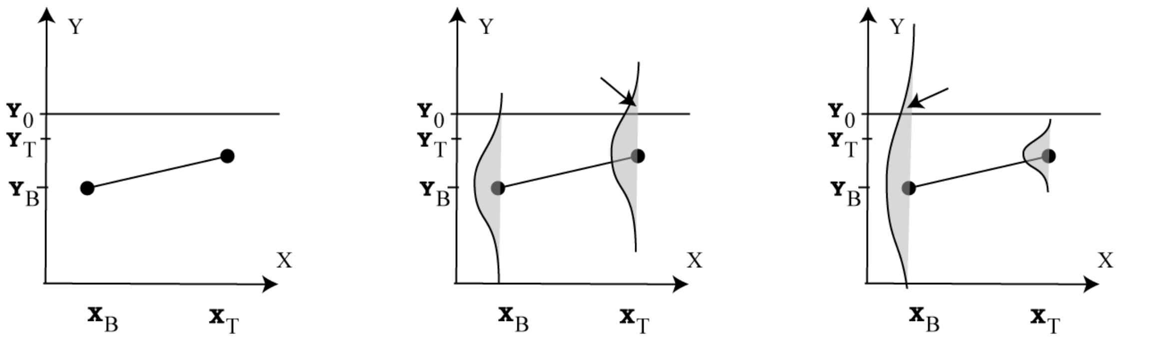
Regarding the dependent variable, studies linking team composition to organizational performance have traditionally shared a common approach using linear regression to predict an effect on the mean performance. However, performance variability—the fluctuations around its expected value—sometimes matters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

Performance variability may sometimes be a better predictor of outcomes than a simple mean effect. Various examples exist where team are rewarded only if they reach a performance threshold. For instance, teams may be competitively ranked and only a top few rewarded, as in the case of the Olympic Games or a high-technology IPO market. Similarly, the interests of many organizations tend only toward the both ends of performance spectrums (Zenger, 1992). For instance, General Electric promotes the top 25 percent of managers and dismisses the bottom 10 percent. In a study on variability, Miner, Haunschild, and Schwab (2003:803) define these instances as "competitions on extreme values." The common characteristic of these examples is that one does not seek simply to increase the expected outcome—as predicted by the mean effect—but rather to reach an extreme performance level. Therefore, in these cases, the effects on performance variability play a role at least as important as effects on the mean.

Previous research has explored how effects on variability could drive extreme outcomes. The study by Cabral (2003) explores the strategies used by laggards in trying 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 other handicaps. To clarify the mechanisms, let us consider some examples (Figure II-1). The first (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?

Figure II-1: Reaching a Threshold Depends on the Variability Effect

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



Traditionally, one relies on the mean effect: since X increases the mean value of Y, one may infer from (a) a greater probability of reaching the threshold at X_T. This assumes implicitly that the variability in Y is constant, as appears in (b) where the bell curves representing the distribution of Y have an 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. In this way, 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 the mean performance but also performance variability, one must assume something more than the 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 as both higher and lower performance is made possible simultaneously, because diversity both increases the amount of information available, and yet increases conflict. 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 the problem is not the hazard of poor social integration but the lack of information. 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 & Vermeulen, 2003; Richard, Barnett, Dwyer, & Chadwick, 2004; Uzzi & 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 & Titus, 1987), thereby hampering performance.

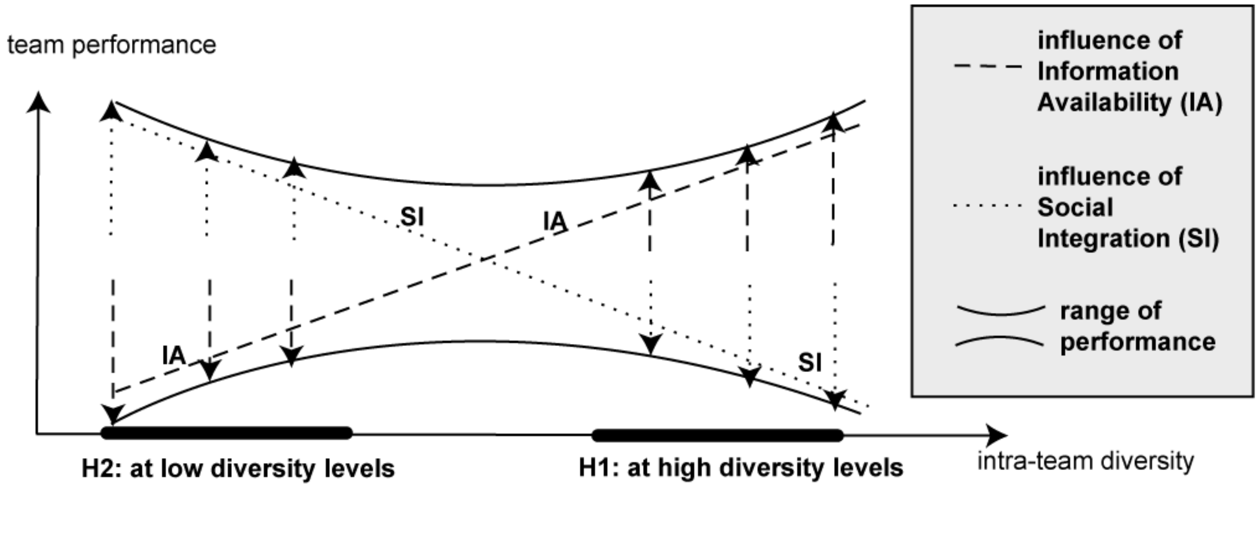
Such reasoning fits into the categorization-elaboration model (van Knippenberg, De Dreu, & Homan, 2004), which predicts that social integration positively moderates the relationship of diversity to performance (similar moderation also appeared in Van Der Vegt

& 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. Instead, assuming the existence of random initial conditions, I focus on the effect of diversity on the spread of outcomes. Among the teams that are lucky to integrate properly, diversity increases performance, while it decrease it among teams 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 II-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 these are additive, this theory suggests that the positive and the negative effects may occur in parallel. At high diversity levels (the right side of Figure II-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.

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



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. For an alternative task with different requirements, the results might be inverted. In low levels of diversity, whether the teams 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-information fit when studying the link between diversity and performance. The categorization-elaboration model (van Knippenberg, De Dreu, & Homan, 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-information 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 greater diversity improves their outcome by reducing the information penalty. The best-performing teams are likely to 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 tasks 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 II-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 & 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 includes task-information fit as an additional hazard whose effect would be in the other direction, and perceived at low levels of diversity. The only published study empirically linking team diversity with performance variability finds only an increase in diversity (Taylor & Greve, 2006). A possible explanation is that the curvilinearity was present, but not detectable, especially given that diversity was operationalized by counting the number of cartoon genres in which the members of cartoon creator teams had previously worked. 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 confounding explanation is that cartoon creator teams avoided projects into genres for which none of the members had any experience. It would imply that the cartoon creator teams avoided the task-information fit hazard due to a self-selection bias. Theoretically, this reasoning highlights the advantage in the current study of making the informational and the social hazards of equal importance. However, in the field, the former might often be masked. On one hand, homogeneous teams are not likely to be assigned or to choose tasks for which they are not suited a priori; therefore, the informational hazard is likely to be avoided. On the other hand, even though teams with diverse members can also mitigate the hazard of not getting along by assembling an already compatible team, 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 these hypotheses in a setting without selection bias.

Exploring the Consequences of the Variability Effects of Diversity

Previous studies predicting performance variability (e.g., Sørensen, 2002; Taylor & Greve, 2006) usually state such effects as a stand-alone conclusion. The current study intends to provide an additional conceptual step by contrasting the conclusion reached by considering only the mean effect with that reached by also relying on the variability effect. For instance, suppose that a diversity variable—such as age diversity—increased average performance. The implication is 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 teams with lower diversity would exhibit greater performance variability. If the variability effect is strong enough—as in Figure II-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 only using a mean analysis, the implications concern only the average outcome, with diversity being either beneficial or detrimental to the team. Such an implication may change if diversity influences performance variability enough to change the probability of achieving a performance hurdle. For instance, if diversity were beneficial on average, it may also decrease the high performance potential; if diversity were detrimental on average, it may also increase the chances of remaining 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 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 the effects of variability adds up to—or even contradicts—the conclusions drawn from a classical mean analysis. To emphasize the counter-intuitive nature of these conclusions, I therefore chose to reanalyze an existing study of diversity, which also had 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 conclusions, it appeared in that diversity did not have much effect on performance, except 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, Angelmar, & Mehra, 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; 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, Angelmar, & Mehra, 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 measured using 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 controls. The results section shows that those variables exhibited a pattern similar to those selected. 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 measured 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 the teams' population diversity level for each variable. The hypotheses stated that the effect of diversity on performance variability depends on whether the population of teams is characterized by a high or a low level 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 II-1) suggests that functional diversity was attended to, as it was high for most teams. In contrast, teams were less varied for other demographic dimensions: since the executives' population had a limited diversity of nationality 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 levels of cognitive diversity variables.

Therefore, the empirical setting presented a natural experiment condition for the demographic variables and an empirically clear situation for the cognitive variables. For each variable, the sample of teams was either at low or high diversity levels, which allowed for the testing of 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 the testing of hypothesis 2; functional diversity, for which the teams were at high levels, allowed the testing of hypothesis 1. The effect of diversity was assumed locally linear, respectively reducing or increasing performance variability.

Control variables. Since the current study builds on an existing mean effect analysis (Kilduff, Angelmar, & Mehra, 2000), I kept the control variables of the previous study. Therefore, the size of the team and a measure of the starting position—the teams in the simulation did not each start with the same market share—were included. Furthermore, all variables which were identified in the original study as relevant independent variables, but were 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, Angelmar, & Mehra, 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 shown 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, Angelmar, & Mehra, 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 & 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.[10]

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 assumptions, 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-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 between these two variances indicates 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:

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

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

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 ($N > 2\ n_T = 2\ n_T$), 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 used while 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[11]. For each sub-sample, I regressed the 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.

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 the insight into why the mean effect alone might not predict the chances of reaching such a threshold.

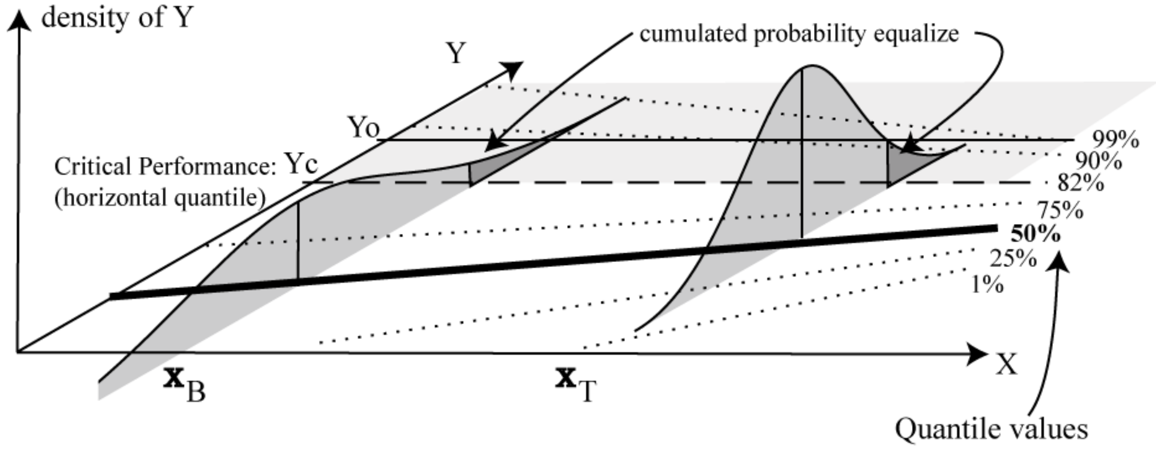
To establish a criterion, let us use again the example represented by the sketches of Figure II-1, where the factor X has a positive effect on the mean of performance Y and a negative effect on its variability. In the case of constant variability (Figure II-1.b), the mean effect drives the chances of reaching the threshold Y_0. In the case where variability changes

[11] 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).

(Figure II-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, these 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 II-3—a more detailed version of Figure II-1.c. Now, one can simply read the probabilities—expressed by the quantile lines—of reaching the threshold. For example, observe that the 90 percent quantile line crosses the threshold Y_0 with a negative slope; it shows that falling beyond 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 of reaching the threshold.

Figure II-3: Mean and Variability Effects Represented by Quantile Lines, which Change Slope at Critical Performance Level



Determining the effect of the factor X on the probability of reaching the threshold simply amounts 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 increase 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 the 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 the level at which 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:

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

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:

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

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 II-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 either entirely in 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 the 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, I focus on the variability analysis, and on how it may complement the mean analysis.

The summary statistics and correlations for all variables appear in Table II-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.

Table II-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

Finding Variability Effects

The regressions and heteroskedasticity tests appear together in Table II-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.

Table II-2: Analyses

Performance Variable:	Final Market Share (FMS)			Cumulative Net Marketing Contribution (CNMC)		
	1a	2a	3a	1b	2b	3b
Model:	Regression	Regression	Goldfeld-	Regression	Regression	Goldfeld-
Dependent Variable:	FMS(=Y)	abs(res(Y))	Quant	CNMC(=Y)	abs(res(Y)/Yhat)	Quant
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

Hypothesis 1. This hypothesis predicts that diversity increases performance variability for teams of relatively high diversity. In the sample, teams were at relatively high levels of functional diversity, so we should observe the positive effect of that variable, in model 2 by a positive coefficient, and in the Goldfeld-Quandt test by 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 decreases performance variability for teams of relatively low diversity. In the sample, teams were at relatively high levels of nationality, age, perception of specialization, and power diversity, so we should observe the negative effect of those variables, 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. Nevertheless, one might 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 variability effects into account 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 II-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).

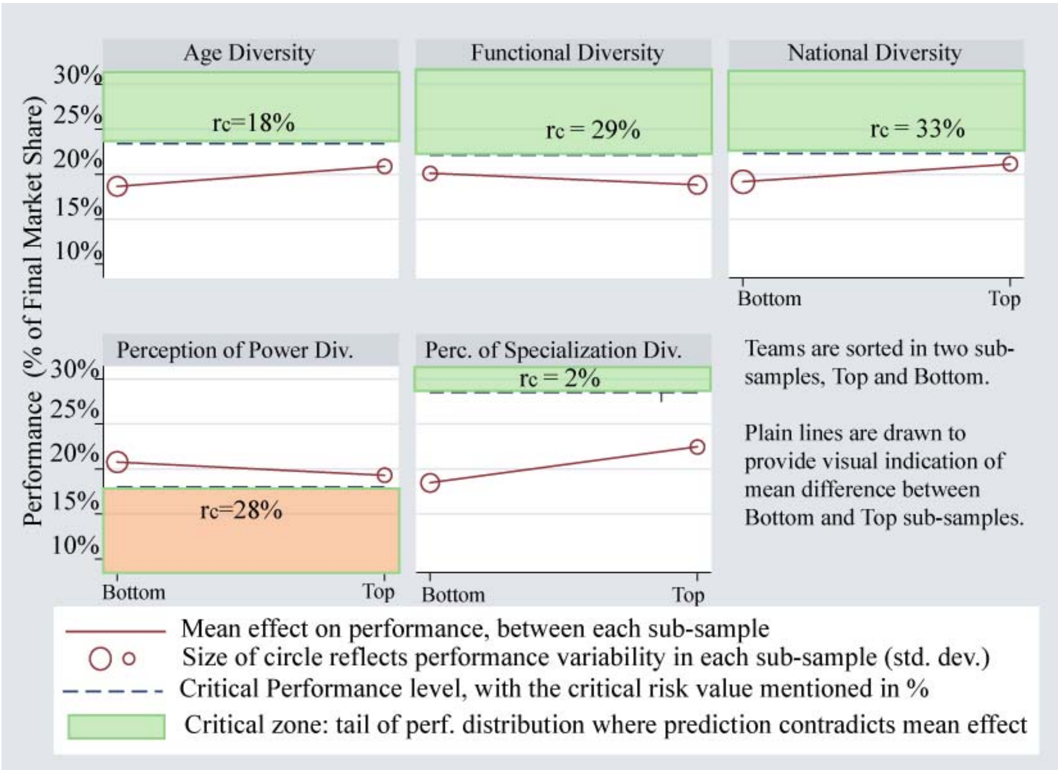
Table II-3: Computation of Critical Performance Level and Probability

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.

To interpret these results, I built graphs summarizing them, considering only the dependent variable FMS (Figure II-4). Each of the five graphs synthesizes the results for one of the five independent diversity variables, and uses the representation introduced in Figure II-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 depicts 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 the inversion of the effect of each variable occurs, labeling it with the size of the associated critical risk.

Figure II-4: Separating where Performance Range where Mean Effect applies from Range where Variability Effect has Inversed Implications



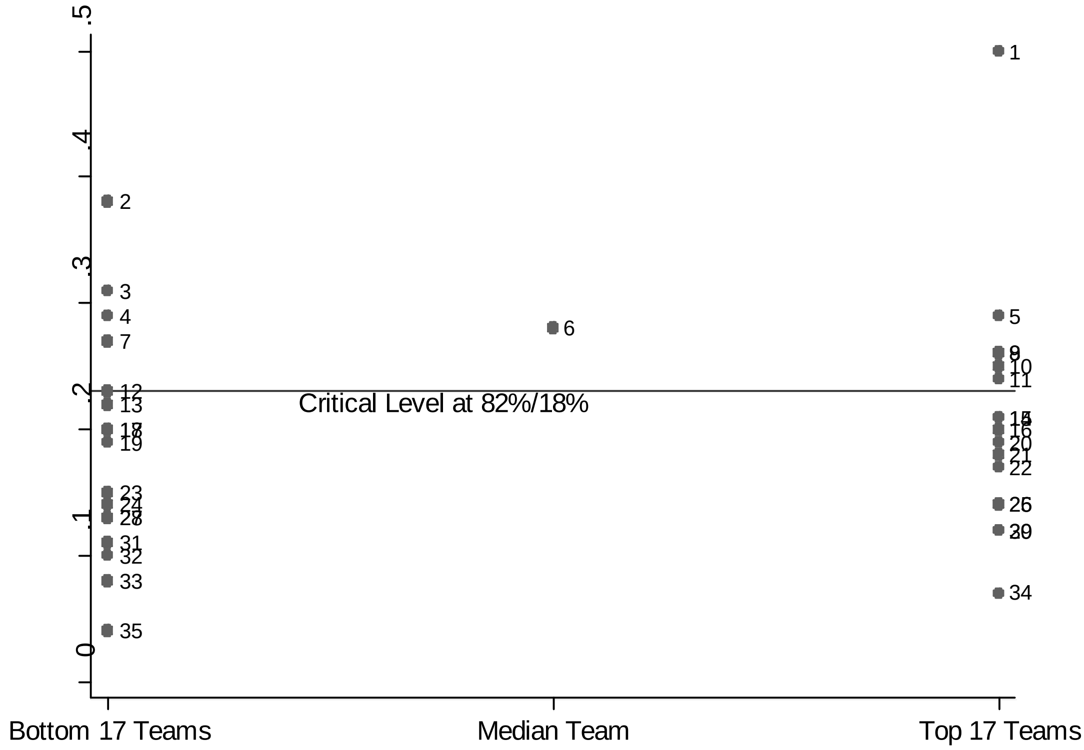
I now go into detail with the reasoning for one independent variable, age diversity, and for one performance measure, Final Market Share (FMS). Let us first consider mean effect. Teams in the top sub-sample reached, on average, 21 percent of FMS, while teams in the bottom sub-sample could expect only 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 do 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 II-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[12]. 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.

[12] One should note that the first position was captured by the top sub-sample, this point appearing as an outlier in this variability analysis. A graph below (Figure II-6) plot will show that team number one was in extremely high functional diversity (higher diversity in a sample of relatively high diversity), and therefore its extreme performance can be explained by that factor.

Figure II-5: Lower Age Diversity Imply More Extremely High performance[13]



All these 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.

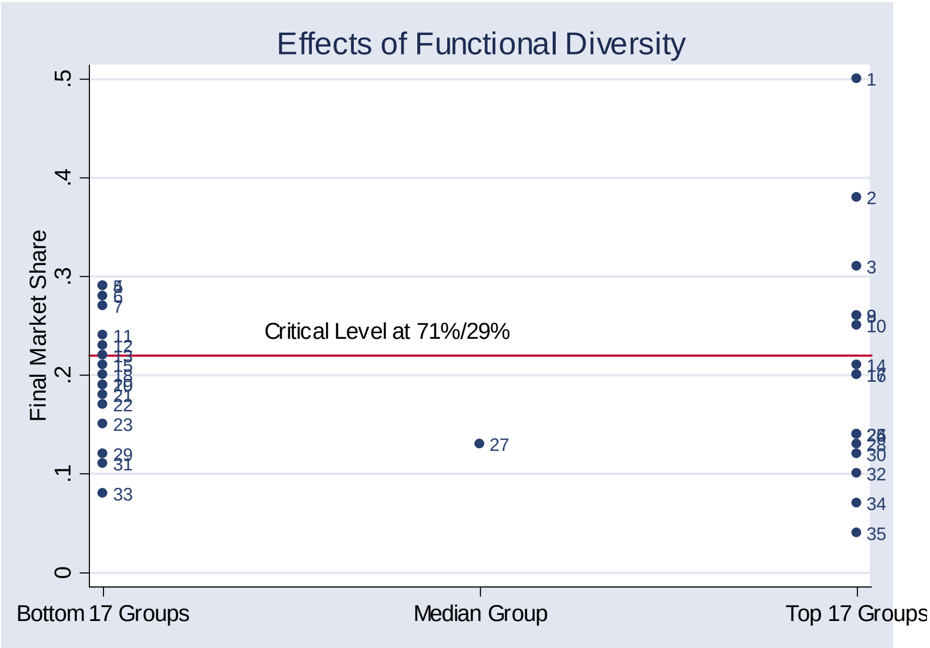
Regarding the other variables, the mean effect analysis found no significant effect. The variability analysis showed that increasing diversity in nationality, power perception, and specialization perceptions (these factors being empirically low in this sample) reduced performance variability. Therefore, 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. Figure II-6 illustrates that the variability effect appears clearly, and the greater chances of reaching both extremely low and high diversity being

---

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

illustrated by the higher diversity sub-sample (right side here) containing the more functionally diverse teams capturing both best and worst scores.

Figure II-6: Greater Functional Diversity Increases both Extremely High and Low Performance[14]



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 predictions 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 has a U-shaped relationship with 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 the mean effect

[14] This graphs also explain that team number one may not be an outlier as suggested by Figure II-5 but rather a team that may perform extremely well due to its functional diversity.

alone. This should inform the various organizational perspectives where extreme outcomes, either high or low, play a particular role.

Two contradictions occur within 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 the development of hypotheses that predict both outstandingly good and bad outcomes even in the absence of the mean effect. In the empirical analysis, four diversity variables had no mean effect on performance (Kilduff, Angelmar, & Mehra, 2000), yet exhibited a significant effect on risk. A second contradiction appeared in previous studies that considered the effect of diversity on performance variability. A knowledge perspective suggests that diversity creates risk (Fleming, 2001, 2004; Taylor & Greve, 2006), while a socio-psychological perspective suggests that similarity creates risk (Janis, 1971; Luthans, 2002; Schachter, Ellertson, McBride, & Gregory, 1951). 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 theory perspectives studying contexts that either reward or punish extreme outcomes. Echoing the call to arms of some scholars for expanding our focus beyond effects on the mean (Daft & 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 importantly, 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 suggests revisiting the concept of groupthink (Janis, 1971; 1982), where one expects homogeneity in teams to lead to catastrophes. Little support for that theory was found when trying to link concurrence seeking or cohesion to performance on average (Esser, 1998; Nemeth & Staw, 1989). Here, I show how homogeneous teams are more likely to result in catastrophic outcomes; however, I do so not through a mean effect but through a variability effect. This can potentially prove the spirit of Janis' statement, that homogeneous teams are dangerous.

Other contexts force organizations to take risks by rewarding those reaching high thresholds of performance, or by allowing only a few winners of a contest to survive. In entrepreneurial settings, many high and rare performance outcomes are salient, such as being the first firm to be profitable, to benefit from network effects, or to complete an initial public offering (IPO), etc. These lead to survival, suggesting that the effects on variability may matter more than the 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 have not waited for full conclusions, and have pushed strongly for developing organizational diversity. In a world where risk—industrial, governance, entrepreneurial, etc.—increasingly matters to organizations, establishing that both extremes of the team diversity scale entail increased chances of both positive and negative extreme outcomes warrants further study.

CHAPTER III. INFLUENCE DE LA DIVERSITE DE L'EQUIPE SUR LE RISQUE ORGANISATIONNEL : CONTRASTE ENTRE LES EFFETS SUR LA VARIABILITE LONGITUDINALE ET INTER-TACHES

Ce chapitre explore les effets contradictoires de la diversité de l'équipe sur différents types de variabilité de la performance. Quelques études (Cavarretta, 2007b; Taylor & Greve, 2006) relient déjà la diversité de l'équipe aux écarts de performance dans une population d'équipes, mais aucune ne considère la variabilité individuelle de chaque équipe. On prédit ici que la diversité augmente la variabilité longitudinale mais diminue la variabilité inter-tâche, et on le vérifie sur un riche ensemble de données de 191 équipes d'étudiants effectuant des exercices notés sur une période d'un an. Cela démontre que la composition de l'équipe influence deux sources d'incertitudes de la vie organisationnelle : un aléa social – selon que les membres de l'équipes s'entendent – et un aléa informationnel – selon l'adaptation des connaissances des membres de l'équipe à la tâche à accomplir.

INFLUENCE OF TEAM DIVERSITY ON ORGANIZATIONAL HAZARD: DISTINGUISHING EFFECTS ON ALONG-TIME VS. CROSS-TASK PERFORMANCE VARIABILITY

This paper explores the contradicting effects of team diversity on the different types of performance variability. A few studies already link team diversity to performance spread across a population of teams (Cavarretta, 2007b; Taylor & Greve, 2006), but none of them consider variability at the team level. The current study predicts a contradictory effect of team diversity, increasing performance variability over-time or decreasing variability across-tasks. It uses a rich archival dataset of 191 teams performing sanctioned business exercises over a period of a year. It demonstrates that team composition influences two uncertainties of team life: a social hazard—the uncertainty on team members' social integration—and an information hazard—the uncertainty on the match of team members' information and the tasks to accomplish.

INTRODUCTION

Team life is full of uncertainties, and the act of assembling a group of people to accomplish a task is a gamble by its very nature. The organizational actors that teams in the field often confront crucial issues of uncertainty: the team is designed to perform various tasks of which little is known ex-ante; its actual processes may be unobservable during the action; and finally, decisions about the composition will be practically impossible to adjust on the fly.

For instance, a board of directors can influence the composition of the top-management-team (TMT) of a firm, but it cannot be sure in advance of what kind of competitive environment that team will confront, or even assume that full information about the team's performance and processes will be available for correcting its course. Similar concerns plague those assembling teams for High-Reliability Organizations (HRO), such as firefighter teams, airplane cockpit crews, or nuclear power-plant control-room teams. The unfolding of team performance depends highly on various uncertainties that occur in the heat of action.

In trying to classify the uncertainties facing teams in the field, two large classes of factors may be identified. The first one concerns the information required for the tasks the team has to accomplish. One of the motivations for getting more than one individual to perform a task is to accumulate information—knowledge, experience, cognitive abilities, etc.—superior to what a single person can reasonably possess. Typically, the top management team of a firm aggregates differentiated backgrounds in order to manage the complexities of business organization. However, once assembled, it will be uncertain—contingent on the situation—whether a team actually possesses the required informational mix for all the upcoming challenges.

For instance, in the period preceding the Enron fiasco, the main issues facing the top-management-team (TMT) of that firm appeared to be those of value creation (Eisenhardt & Sull, 2001). However, the true challenges for that team turned out to be more about issues of ethics. This dimension had not been originally identified as crucial, and the financial bubble context of the late 90s created the conditions and the motivations for deleterious team behavior. The Enron TMT is an example of a team whose skills were matched to one of the dimensions of performance (financial creativity), but not another dimension (ethical strength) whose crucial relevance only appeared after the fact. If most research has focused on the ethical or legal dimensions of such fiascos (Palmer & Maher, 2004), sheer uncertainty and information overload may have also been at work (Gladwell, 2007).

The second class of uncertainty concerns whether the team members get along and, therefore, function properly as a team. Social integration (Nemeth & Staw, 1989) tends to fluctuate over time (Gersick, 1988), which creates uncertainty on team performance. For instance, sports teams have to perform regularly a well-defined task, yet their performance exhibits high volatility often driven by the changing dynamics of the teams. This uncertainty is so intrinsic to team life that it constitutes the major factor in many team sports championships, like in soccer where the fluctuating dynamics of each team feed most of the buzz surrounding that sport.

Since research on team diversity developed according to the classical variance-reduction approach (Mohr, 1982), it eliminated such uncertainties by determining whether factors increase or decrease expected performance (Mannix & Neale, 2005). Social integration or information appears either as mediating or moderating variables while attempting to figure out an average effect (e.g. van Knippenberg, De Dreu, & Homan, 2004). This approach can be problematic in the field since those two major contextual conditions—whether team members get along or not on one hand and whether they possess information matching the task to accomplish on the other hand—are often quite uncertain when making composition decision.

A gap therefore appears between scholarship proposing complex contingencies to predict an effect of team diversity onto expected team performance and practitioners facing real situations full of irreducible social and information uncertainties. That gap can be addressed by shifting the focus away from predicting the direction of the outcome (an average effect) to predicting the spread of the outcome (a variability effect). The performance variability induced by the uncertainties of team life then become the focal variable with both theoretical and practical consequences.

Following that logic, recent research on team diversity acknowledges the uncertainties facing practitioners in the field and views them as sources of performance variability, a dependent variable that can be influenced by team diversity. A first approach considers a population of teams and studies how team diversity influences the performance spread among the teams (Cavarretta, 2007b; Fleming, 2004; Taylor & Greve, 2006).

However, one may wonder about performance variability at the unit level, instead of the population level. For instance, in high reliability organizations, a repetitive task has to be accomplished regularly and consistently over time, like flying a plane full of passengers on a daily basis. Then, reliability of performance is an important measure of outcomes. Alternatively, in an organizational governance setting, different goals—like attending to financial as well as social and environmental performance—have to be attended to simultaneously (Cyert & March, 1963 [1992]). If performance is considered along time or across tasks, the question then becomes whether team diversity influences such performance variability.

Accordingly, this paper considers unit-level performance variability and distinguishes along-time vs. across-task variability. I hypothesize that team diversity decreases across-task performance variability but increases along-time performance variability. Empirically, this study leverages a field dataset gathering demographic and performance archival information about 191 student teams during their first year at a top business school. For most of their courses, the teams had to perform team tasks, which were graded, having both academic and reputational consequences. This population is representative of business executives in their early career, and anecdotal evidence suggests their team dynamics are similar to teams in many professional firms. Furthermore, the teams were not designed for, nor submitted to, an experiment, but assembled by school administration to accomplish academic learning goals with the intention of reaching the best possible performance, as reflected by grades.

A particular merit of that field context is that it provides performance measures for each team in three dimensions: across-level by providing performance scores both at team and individual levels; along-time when courses contain more than on team exercise on the same subject; and across-task when different courses contains team exercises in the same period (cross-task). This context matches field situations as much as possible while providing the required richness of information, which would not be available in most industrial situations.

LITERATURE REVIEW

Team Diversity Research Sorting out Good vs. Bad Outcomes

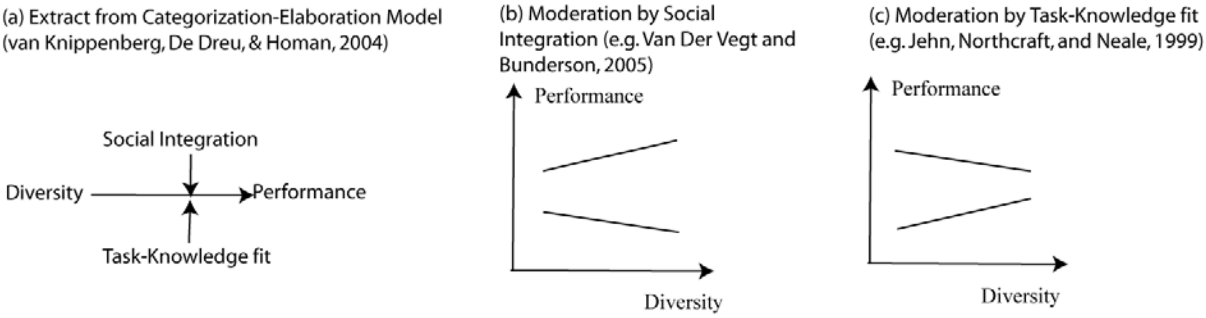
The research paradigm for team diversity follow Mohr's prescriptions (1982), attempting to find factors that reduce variance in outcomes. In other words, one observes variance in performance and hopes that under certain theoretical conditions, team diversity has an effect on the expected (i.e. mean) performance that allows reducing the unexplained variance. This respectable—and dominant—method of Organizational Behavior scholarship has allowed building a body of predictions regarding which diversity, and under what conditions, leads to better performance on average.

Researching a simple relationship between intra-team diversity and team performance lead to two competing perspectives. The first focuses on information availability and takes an “optimistic” view of diversity (Mannix & Neale, 2005:33): intra-team diversity brings more information to the team, which improves expected team performance (Dahlin, Weingart, & Hinds, 2005; Gruenfeld, Mannix, Williams, & Neale, 1996). For instance, exposure to minority thinking fosters a broader view of the issue at hand (Nemeth, 1986; Page, 2007), and diverse teams exchange a wider range of information (Sommers, 2006). By contrast, a social integration perspective (O'Reilly, Caldwell, & Barnett, 1989) combine the similarity-attraction (Festinger, 1954) and the self-categorization (Tajfel, 1982) approaches and takes a “pessimistic” view of diversity (Mannix & Neale, 2005:34): intra-team diversity creates tensions in the team, which damages expected team performance.

These two perspectives suggest contradictory effects, so finding a simple main effect has proven elusive. Some authors have found results—linear or U-shaped (Earley & Mosakowski, 2000) or inverted u-shaped (Dahlin, Weingart, & Hinds, 2005) or with even more complex shapes (Allmendinger & Hackman, 1995)—but starting with Nemeth and Staw (1989), various meta-analyses (Bowers, 2000; Webber & Donahue, 2001) and reviews have settled on an overall neutral effect of diversity (Jackson, Joshi, & Erhardt, 2003; Milliken & Martins, 1996). To address this ambiguity, at least three major streams of research have emerged (Mannix & Neale, 2005; van Knippenberg & Schippers, 2007). The first has focused on the independent variable, diversity, by adopting ever finer definitions (e.g., Boone, Van Olffen, & Van Witteloostuijn, 2005; Bunderson & Sutcliffe, 2002; Cramton & Hinds, 2005; Lau & Murnighan, 1998). A wealth of definitions has led to a necessary theorization and classification of the various possible approaches to the diversity construct (Harrison & Klein, 2007). A second stream of research, following Lawrence’s recommendation to “open the black box of demography” (1997), explores the mediating processes, for instance distinguishing the effects of diversity on emotional vs. task conflicts, those leading to effects on performance (e.g. Pelled, Eisenhardt, & Xin, 1999). Finally, a 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, Barnett, Dwyer, & Chadwick, 2004), intra-team perceptions (e.g., interpersonal congruence in Polzer, Milton, & Swann, 2002) or context (e.g., people orientation of corporate culture in Kochan et al., 2003:10).

Two moderations deserve particular attention. The categorization-elaboration model (van Knippenberg, De Dreu, & Homan, 2004) states that both informational requirements and social integration moderate the relationship between diversity and performance (Figure III-1.a). 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 (see Figure III-1.b), improving performance in the first case, degrading it in the second (e.g. Van Der Vegt & Bunderson, 2005). Regarding moderation by information, it states that the effect of diversity differs depending on whether the teams has the information fitted to the task or not (see Figure III-1.c), decreasing performance in the former case and increasing it in the later case (for example, see the moderation by task complexity in Jehn, Northcraft, & Neale, 1999).

Figure III-1: The Moderating Effects of social Integration and Task-Information Fit



Applicability to the Uncertainties Faced when Composing Teams in the Field

Unfortunately, the moderating conditions are often not actionable because one is often quite uncertain about those when making composition decisions. Some literature attempts to identify the characteristics of teams that could enable them to succeed in the long term in the complexity of organizational life (see for instance innovation performance of X-Teams in Ancona, Bresman, & Kaeufer, 2002). However, little theoretical literature has attempted to model the uncertainties of teams' organizational life. Regarding social integration, one might have the ex-ante information about personal relationships between members, but once a team is composed and starts performing, one has to accept that the social integration will fluctuate, often faster than one can change composition. The dynamics in teams vary over time (Gersick, 1988) and in ways that are often unpredictable and path dependent (Ancona, Goodman, Lawrence, & Tushman, 2001). The importance of such fluctuations is made particularly relevant since groups can engage into spirals where success feeds success and failure feeds failure, therefore possibly amplifying negligible initial differences to significant amounts (Hackman, 1990). The mechanisms for such spirals have been suspected to be a causality between conflicts and performance running both ways, from the social state to the outcome, as well as from the outcome to the social state (Peterson &

Behfar, 2003). Except in experimental conditions, even the best thought design cannot prevent significant fluctuations in the social integration of a team over time. One can imagine adjustments or interventions on the team but in many situations, e.g. for teams of astronauts, firefighters, or even top managers, practitioners accept some level of uncertainty about social integration during the life of the team. Attempting to improve that condition is a laudable goal, but it remains an intrinsic hazard in many practical situations.

Similarly, one can have ex-ante expectations about what skills the team will need and staff accordingly, however, once the action is started, the team can encounter issues that were difficult to predict and difficult to adjust on the fly. Typically, in a governance context, top management teams are assembled according to some expectation of the organization, with the hope that conditions—market, competition, technology, etc.—move in ways that are compatible with the skills available in the team (including the second-order learning skills necessary to adapt). In a well documented example, the top management team at Enron was designed for excellence in financial creativity (Eisenhardt & Sull, 2001), but its composition did not bring enough other soft skills to confront the dazzling market conditions of the late 90s' and their incentives for ethical lapses (Palmer & Maher, 2004). The Enron management team had a good composition for the task of financial innovation, but was lacking the skills for the task of managing the firm in a socially responsible way, with recent suggestions that such a fiasco could be explained through a cognitive overflow lens (Gladwell, 2007).

The social integration and information-to-task fit therefore constitute hazards that practitioners have to live with. Current research on team diversity rightfully addresses process questions (which factor will improve social integration and information-to-task fit?) and expected performance questions (what is the effect of a factor on performance given social condition or information-to-task fit condition). However, little research considers that those hazards are sources of uncertainty, hence inducing variability of performance that one may want to control.

Focusing on Effects on Performance Variability

Beyond the observation that composing a team implies uncertainties that are difficult to control in practice, why would we consider variability of performance as a dependent variable? Studies linking team composition to organizational performance have traditionally shared a common approach using linear regression to predict an effect on the mean performance. However, performance variability—the fluctuations around its expected value—sometimes matters to the organizational theorist more than mean performance does.

March's exploration-exploitation study (1991) seminally identified the effects on performance variability as a better predictor of outcomes than mean effects. It led to a stream of literature on innovation and learning, where various factors, including team diversity, might cause such variability (Denrell, 2003; Fleming, 2004; Sørensen, 2002). Accordingly, the current paper takes performance variability as its primary dependent variable.

Performance variability may sometimes be a better predictor of outcomes than a simple mean effect. Various examples exist where team are rewarded only if they reach a performance threshold. For instance, teams may be competitively ranked and only a top few rewarded, as in the case of the Olympic Games or a high-technology IPO market. Similarly, the interests of many organizations tend only toward the both ends of performance spectrums (Zenger, 1992). For instance, General Electric promotes the top 25 percent of managers and dismisses the bottom 10 percent. In a study on variability, Miner, Haunschild, and Schwab (2003:803) define these instances as "competitions on extreme values." The common characteristic of these examples is that one does not seek simply to increase the expected outcome—as predicted by the mean effect—but rather to reach an extreme performance level. Therefore, in these cases, the effects on performance variability play a role at least as important as effects on the mean.

More generally, performance reliability—the constraint of variability—has been identified as a common objective of organizational life. The arguments range from the need to buffer internal processes against uncertainty (Thompson, 1967) to the legitimacy derived from respecting institutional norms of consistency (Meyer & Rowan, 1977). For others, consistency improves organizational autonomy (Pfeffer & Salancik, 1978) and relationships with external stakeholders (Hannan & Freeman, 1984). Population ecology takes as a fundamental assumption that "selection in populations of organizations […] favors forms with high reliability of performance" (Péli, Masuch, Bruggeman, & Nualláin, 1994). Reliability even appears as more important than efficiency in the structural inertia approach to population ecology (Hannan & Freeman, 1984).

Accordingly, recent team diversity literature has considered effects on performance variability. A learning perspective focused on the uncertainty of social integration and proposed that team diversity increases performance variability (Fleming, 2004; Taylor & Greve, 2006). More generally, taking uncertainty into account both with social integration and information-to-task fit leads to a U-shaped effect of team diversity on performance variability (Cavarretta, 2007a). The logic is that performance is highly variable in the lower end of team diversity due to a high information hazard and in the higher end of diversity due to high social hazard, as compared to the middle range of team diversity where social and information hazards are limited.

DISTINGUISHING ALONG-TIME VS. CROSS-TASK PERFORMANCE VARIABILITY AT TEAM LEVEL

So far, the performance variability was considered at the population level, so the studies predicted cross-sectional performance variability *between-teams,* in other words a spread of performance between the best and worst teams. What happens then if we consider performance at team level? One may be more interested in the outcome at the team level rather than predictions concerning the spread between the best and the worst. Statistically, taking a unitary view of a team that wants to ignore distributional properties of the population may appear—at first—as problematic: how could one explore variability for a population of one?

This conundrum can be resolved if defining variability not so much by comparison to a population, but by comparison to the focal team itself. For instance, in various contexts, teams have to perform a task repetitively, and therefore performance can vary over time. The change in performance along time constitutes a variability of performance that can be measured at the team level. For instance, in financial theory and some branches of strategy, the longitudinal changes of firm performance—called volatility—matters. Another organizational context where one cares for longitudinal performance variability is the high reliability organization (HRO) context (Weick & Sutcliffe, 2001). This includes situations such as managing nuclear power plants, aircraft, firefighting, etc., where similar tasks have to be accomplished reliably along time, and the objective is to limit variability to avoid extremely low performance outcomes.

Once it is observed that interesting team-level variability can be constructed by measuring changes in performance along the time dimension, it is natural to consider differences of performance between tasks. In the field, teams evaluated on only one dimension of performance are the exception more than the rule. Most organizational situations require attending to multiple goals that all constitute different tasks (Cyert & March, 1963 [1992]; Simon, 1947 [1997]). In the field, a context where cross-task performance variability certainly matters is the field of governance (Fligstein, 1987), where one considers the various objectives that can be accomplished by a firm, and how these may be driven by the composition of its top management team (Hambrick & Mason, 1984). For instance, firms can be viewed as having to accomplish both short-term financial goals and long-term innovation goals, as well as some social and environmental goals. Stakeholders such as regulators (SEC, EPA, etc.) or different classes of shareholders may be concerned that firms will perform particularly well on one dimension, but fail miserably on the other (Barnett, 2007). Avoiding an asymmetric fiasco on one of those dimensions has become a general concern in modern business organizations (Palmer & Maher, 2004).

The current paper builds its hypotheses on those two ideal cases of within-team performance variability, contrasting *unit-level across-time* variability against *unit-level across-task* variability.

Effects of Team Diversity on Along-Time Performance variability

Let us first consider the influence of team diversity on unit-level performance variability when keeping task constant but varying time. Teams tend to evolve in their internal dynamics (e.g. Gersick, 1988; Harrison, Price, Gavin, & Florey, 2002), and therefore their social integration varies between two different points in time. However, since the task is constant, the information-to-task fit is constant for each team. Therefore, we can consider that the only uncertainty will be on social integration. That condition will fluctuate, and we can then consider how it moderates the diversity-performance relationship (Van Der Vegt & Bunderson, 2005; van Knippenberg, De Dreu, & Homan, 2004), as illustrated in Figure III-1.b.

Assuming some uncertainty (i.e. stochasticity) on social integration, the moderation relationship dictates that performance outcomes will become extreme with growing diversity. For properly integrated teams, higher diversity leads to higher performance because team diversity brings information that the team members can collectively leverage. For teams that do not integrate properly, higher diversity leads to lowered performance because teams with low social integration cannot make use of the different information available among their members. Overall, since teams fluctuate on their social integration, higher diversity amplifies that fluctuation, formalized as follows:

Hypothesis 1: intra-team diversity increases along-time variability of team performance

Effects of Team Diversity on Cross-Task Performance Variability

Now, consider the influence of diversity on performance variability at team level when simultaneously accomplishing tasks of different natures. One can assume that the social integration condition will be constant since the tasks are accomplished simultaneously. In this case, only the moderation of the diversity-performance relationship by task-information fit

(Jehn, Northcraft, & Neale, 1999; van Knippenberg, De Dreu, & Homan, 2004) applies, as illustrated by Figure III-1c.

Assuming some uncertainty (i.e. stochasticity) on task-information fit, the moderation relationship dictates that performance outcomes will become less extreme with growing diversity. For teams confronted with tasks of differing natures, having higher diversity implies a more balanced mix of information and therefore reduces the chances to perform very well on one task and perform badly on another task. For instance, consider top management teams that have to deal with various business tasks. Consider two homogenous teams, one composed only of engineers, and the other only of marketers, and one diverse team mixing engineers, marketers, and financiers. On a task of an engineering nature, the first team would do very well and the second one very badly, and the third one would do reasonably well. On a task of a different nature, for instance marketing, the first team would do particularly badly, the second particularly well, and the third one again reasonably well. Overall, all teams face uncertainty on the fit of information to the task to accomplish, but higher diversity teams experience reduced range of outcomes, formalized as follows:

Hypothesis 2: intra-team diversity decreases the across-task variability of team performance

METHOD

Setting and Sample

Verification of such theory required a field setting (the moderation having already been explored in experimental settings) offering rich data on tasks varying both with time and task. In addition, the effects are sought on performance variability (a residual of a mean effect), so are of the second order; hence, a controlled environment was preferable. I chose to study the archives of the MBA program of a top European business school that required extensive teamwork and collected full information about the teams' performance. In this context, students were assembled into stable teams of 6 to 7 members for the first year of their 2-year curriculum. During that whole year, all students took the same courses, most of which contained at least one graded team exercise and all courses included individual graded exercises. Both group and individual grades were combined into a final individual letter grade, but the archives contained both the individual scores (individual exams for instance) as well as group scores. The business school's administration provided access to the archives for four cohorts (2008-2005) with roughly 48 teams each, bringing the total to 191 teams.

Those teams have the characteristics of teams in many business organizations. Members did not choose each other. They had to stay together for an extended period. They cared about the outcome of teamwork, which affected both objective outcomes (grades) as well as reputational outcomes for both the team and each individual member. Finally, the teams went through various phases of group life, including serious and real conflicts that sometimes endangered teamwork.

Measures

Independent Variables. The theoretical model developed here draws on the following two effects of intra-team diversity: diversity leads to social categorization between team members and it brings varied information. I intend to verify the theory on any dimension that is expected to trigger both such effects. Using such a parsimonious and generic criteria for diversity dimensions is consistent with van Knippenberg, de Dreu and Homan's observation that "all dimensions of diversity may in principle elicit social categorization processes as well as information/decision making processes" (2004:521). The empirical verification presented here tests the theory using traditional demographic variables, such as age, nationality, and experiences, which typically both bring varied information as well as trigger social categorization.

The team-level diversity variables were coded using three different methods depending on the nature of the variable at the individual level. The first one applies to the variables that were on a numerical scale like age. It was also applied to binary variables like gender and the variable coding such as whether an individual's native language was English or not. In those cases, diversity was constructed through a simple coefficient of variation among team members, so a low value indicated low diversity.

The second method applied to categorical variables. *The ethnicity* of each individual had been coded by an independent coder based on pictures in the yearbook, with the possible values being Caucasian, African, Chinese, and Indian. In cases where the first coder was not sure on the individual's ethnicity, a second independent coder resolved it. I coded *Educational subject* background by attributing each subject to one of following broad categories: "Literary," "Science," "Business," and "Others" for entries too general to be easily matched. *Work industry* background was coded by attributing each industry to one of the following categories: "Consulting," "Media/Marketing," "High-Tech," "Manufacturing," "Finance," "Health," and "Others." *Work Function* background was coded by attributing each function to one of the following categories: "Marketing," "Finance," "General Management," "Consulting," "Engineering," "Legal," "Operations," and "Others." I coded education and work with the intent of creating meaningful categories out of extremely dispersed and non-standardized information. Overall, 493 educational subjects, 64 functions, and 83 industries appeared in the database, which were reduced into less than a dozen broad categories each. All those categorical variables (education subject, work industry, and work function) were aggregated at team level into diversity by using a Blau index (1977), with a low value indicating low diversity.

For *Nationality*, each individual was attributed the score of his or her nationality in the four dimensions of Hofstede's cultural dimensions scales (1990): power distance, individualism, masculinity, and uncertainty avoidance. The values were obtained from the most current list (Hofstede, 2007) and the few missing countries were matched to the closest country (based on private exchange with Hofstede). To compute a team-level diversity measure of nationality, no classical method exists to compute distance between a set of points (the individuals in a team) in a multi-dimensional space (the four dimensions of the Hofstede scales). By analogy to the concept of variance, I defined the diversity metric by summing the variance of the points when projected on each orthogonal dimension. That metric—the sum of the Euclidian distance of all points to the center of the team—captures how far apart the team members are on that four-dimension nationality space. It is also strongly related to the cultural distance metrics used by Kogut and Singh (Kogut & Singh, 1988), except that here the Euclidian distance is not to a reference point (it was USA in that paper) but to the mean of all the points in a group.

Controls. The analyses were controlled by the skill and characteristics of the team members. For that purpose, I used the GMAT aggregated as an average of team level as a proxy of the overall skill of the team. I also used the team average of age and years of work experience to control for both life and professional accumulated experience.

Dependent variables. In this academic context, the natural performance measures were scores in various academically sanctioned exercises/exams. The school archives have grade reports by each instructor with a level of detail such that, for each individual, one or more individual exercises scores were available, and one or more group scores; finally, all of this had been aggregated into an individual letter grade. I used the group scores as measures of group performance. Those were reported as numerical grades. The grade distributions were visually inspected to search for problematic grading patterns. It appeared in a few cases that the professor had negligible or problematic variability in its distribution, for instances appearing to give only 2 or 3 possible values, reflecting a limitation of the range. Those were eliminated. Finally, to make such performance measures compatible for comparison between courses or even along time, I normalized each grade—across each cohort—to an average of 0 and a mean of 1. The actual computation of the performance variability depended on the hypothesis tested (see analysis below).

Analyses

The hypotheses predict effects on *along-time* and *across-task variability*. These dimensions are respectively operationalized as team performance variability on *different yet simultaneous* tasks and on *longitudinal yet similar* tasks.

To test along-time variability (H1), I searched the archives for courses where two different grades would be available on the same subject, in order to compute the difference between grade 2 and grade 1. This would provide both a mean effect (improvement over time signaled by the sign of that difference) as well as a variability effect (the size of that difference). The actual equation defining across-time variability was therefore:

$$\Delta_{\text{time}}\, \mathrm{P} = \mathrm{P}_2 - \mathrm{P}_1$$

The Finance course provided two grades on different exercises for three cohorts, and the International Management course provided two different grades for two cohorts. The teams of one cohort appeared twice, once for with their Finance grades and once with their IM grade. It constituted a sample with 239 points. I use maximum likelihood estimation to estimate both mean and variability effect together. The equation tested in Stata using maximum likelihood estimation (MLE) procedure was:

$$(\text{mu: } \Delta_{\text{time}}\, \mathrm{P} = \beta_1\, \mathrm{X})\ (\text{sigma: } \gamma_1\, \mathrm{X})$$

To test cross-task variability (H2), I selected one period per year, and took all the available grades, but retained only one score (i.e. for courses with many group scores, I chose one only, usually the first score). That provides for each group a sample of performance measure where task is varied greatly, but time variation is somewhat limited (more on that nuance in the discussion of results). The archival information did not allow determining the precise time when each exercise was occurring, so the “time-constant” condition is approximated to “during the same period” condition. For that sample of scores at team level, the variability is then operationalized as the variance of those scores at team level, over all courses *i* in one period, taking the first score if more than one is available.

$$\Delta_{\text{task}}\, P = \text{variance}(P_i),$$

To estimate the effect, I regressed the variability against the independent variables and controls. To improve the significance, I moved from an OLS estimation to an MLE estimation. In addition, it appeared that including the most distant cohort (2005) was reducing the significance of results. I therefore reported the stronger version using only the most recent three years (2008-2006), which provide a sample of N=143.

Results

Table III-1 presents the summary and correlations for the variables. The values in both tables differ for a given variable because the sample used differs between analyses.

Table III-1: Summary and Correlations

Variable	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11
1 Variability (sum of deltas)	1.03	0.73	0.01	3.30	1.00										
2 GMAT	680.74	10.13	651.67	710.00	0.04	1.00									
					-0.56										
3 Experience	5.54	0.82	3.50	7.86	-0.09	-0.02	1.00								
					-0.15	-0.70									
4 Age	28.41	0.83	26.71	31.43	0.02	-0.08	0.52	1.00							
					-0.73	-0.21	0.00								
5 Gender Diversity	0.45	0.05	0.38	0.55	-0.01	0.22	0.03	-0.04	1						
					-0.82	0.00	-0.59	-0.56							
6 Language (Eng) Diversity	0.52	0.04	0.38	0.58	0.03	0.02	-0.01	-0.05	-0.12	1					
					-0.66	-0.74	-0.91	-0.47	-0.06						
7 Nationality Diversity	40.05	3.72	29.75	50.40	0.01	-0.16	0.13	0.16	-0.06	-0.02	1				
					-0.88	-0.02	-0.05	-0.01	-0.35	-0.76					
8 Ethnicity Diversity	0.49	0.12	0.00	0.69	-0.09	-0.03	-0.01	-0.05	-0.06	-0.22	0.07	1			
					-0.16	-0.60	-0.84	-0.47	-0.34	0	-0.3				
9 Education Diversity	0.54	0.11	0.00	0.72	0.01	0.11	0.00	0.07	-0.05	0.15	-0.02	-0.06	1		
					-0.84	-0.09	-0.95	-0.27	-0.48	-0.02	-0.8	-0.36			
10 GMAT Diversity	43.64	11.96	19.41	69.14	0.00	0.01	-0.05	-0.05	0.06	0.02	-0.15	-0.07	0.11	1	
					-0.97	-0.87	-0.46	-0.46	-0.32	-0.76	-0.02	-0.31	-0.08		
11 Industry Diversity	0.74	0.06	0.50	0.83	-0.05	-0.06	0.06	0.07	-0.04	-0.04	0.11	0.12	0.01	0.05	1
					-0.46	-0.37	-0.36	-0.26	-0.51	-0.53	-0.09	-0.07	-0.89	-0.48	
12 Function Diversity	0.71	0.08	0.28	0.86	0.03	-0.02	0.03	-0.02	-0.05	-0.08	0.03	0.04	-0.01	0.1	0.09
					-0.64	-0.80	-0.61	-0.77	-0.46	-0.24	-0.59	-0.57	-0.91	-0.13	-0.17

Variable	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Variability (sum of variances)	0.84	0.46	0.01	2.27	1.00												
2 GMAT	679.849	10.159	651.667	708.333	-0.04	1											
					-0.6												
3 GMAT Diversity	44.188	11.779	19.76	68.034	-0.06	0.08	1										
					-0.44	-0.32											
4 Experience	5.603	0.813	3.714	7.857	0.06	0.04	-0.01	1									
					-0.5	-0.67	-0.86										
5 Experience Diversity	1.947	0.693	0.447	3.971	0.15	0.1	-0.08	0.56	1								
					-0.07	-0.23	-0.32	0									
6 Age	28.411	0.845	26.714	31.429	-0.03	-0.05	-0.02	0.58	0.38	1							
					-0.76	-0.52	-0.79	0	0								
7 Age Diversity	2.221	0.695	0.69	4.082	0.06	-0.01	0.01	0.3	0.49	0.41	1						
					-0.49	-0.89	-0.87	0	0	0							
8 Gender Diversity	0.459	0.053	0.378	0.548	0.06	0.28	0.08	0.09	0.12	-0.04	0.08	1					
					-0.51	0	-0.37	-0.3	-0.16	-0.63	-0.33						
9 Language (Eng) Diversity	0.517	0.037	0.378	0.577	-0.13	0.03	0.1	-0.08	-0.11	-0.06	-0.04	-0.08	1				
					-0.13	-0.75	-0.24	-0.36	-0.19	-0.48	-0.6	-0.33					
10 Nationality Diversity	40.66	3.674	29.747	50.4	-0.14	-0.09	-0.11	0.04	0.07	0.09	0.08	-0.12	0.05	1			
					-0.1	-0.26	-0.18	-0.66	-0.42	-0.27	-0.32	-0.14	-0.59				
11 Ethnicity Diversity	0.492	0.123	0	0.694	0.02	-0.12	-0.11	-0.04	0.05	-0.07	0.06	-0.09	-0.13	-0.01	1		
					-0.82	-0.16	-0.2	-0.62	-0.54	-0.39	-0.5	-0.31	-0.12	-0.95			
12 Education Diversity	0.53	0.11	0	0.722	-0.05	0.08	0.17	0.01	0.08	0.09	0.03	0.01	0.16	0	-0.09	1	
					-0.59	-0.32	-0.04	-0.92	-0.36	-0.27	-0.72	-0.94	-0.06	-0.95	-0.26		
13 Industry Diversity	0.747	0.052	0.571	0.833	-0.09	-0.01	0.01	0.01	0	0.06	0.04	-0.06	-0.11	0.14	0.18	0.05	1
					-0.3	-0.93	-0.95	-0.93	-0.97	-0.51	-0.68	-0.48	-0.2	-0.11	-0.04	-0.56	
14 Function Diversity	0.712	0.082	0.278	0.833	-0.16	0.01	-0.02	-0.08	-0.11	-0.09	-0.16	-0.05	-0.09	0.01	0	-0.04	-0.01
					-0.06	-0.94	-0.83	-0.37	-0.19	-0.28	-0.05	-0.53	-0.29	-0.9	-0.98	-0.63	-0.87

Hypothesis 1 predicts that team diversity increases along-time performance variability. Table III-2 presents the results, showing that nationality diversity has a significant positive effect ($p<0.007$), therefore supporting hypothesis 1. Regarding the other variables, the effects are negligible (see below for discussion on non-significance).

Table III-2: Effects on Performance Variability Along-Time

Variables	MLE	
Mean effect (mu)		
GMAT	0.00	(0.587)
Experience	-0.13	(0.099)+
Age	0.09	(0.203)
Gender Diversity	-0.03	(0.970)
Language (Eng) Diversity	0.59	(0.672)
Nationality Diversity	0.00	(0.846)
Ethnicity Diversity	-0.54	(0.208)
Education Diversity	0.09	(0.845)
GMAT Diversity	0.00	(0.508)
Industry Diversity	-0.63	(0.484)
Function Diversity	0.16	(0.791)
Constant	-2.03	(0.606)
Variability effect (sigma)		
GMAT	0.00	(0.733)
Experience	-0.03	(0.636)
Experience Diversity	-0.03	(0.604)
Age Diversity	-0.05	(0.467)
Gender Diversity	0.09	(0.894)
Language (Eng) Diversity	-0.59	(0.638)
Nationality Diversity	0.02	(0.007)**
Ethnicity Diversity	-0.32	(0.363)
Education Diversity	0.32	(0.363)
GMAT Diversity	0.00	(0.643)
Industry Diversity	-0.87	(0.182)
Function Diversity	0.20	(0.644)
Constant	0.08	(0.973)

N=239; p values in parentheses; + significant at 10%; * at 5%; ** at 1%

The second prediction (H2) is that team diversity would decrease for cross-task variability. Table III-3 presents the results, showing significant negative effect for functional diversity ($p<0.049$). The other independent variables appear non-significant.

Table III-3: Effects on Performance Variability Across-Task

Variables	MLE	
GMAT	0.00	(0.363)
Experience	0.00	(0.969)
Age	-0.06	(0.333)
Age Diversity	0.00	(0.988)
Gender Diversity	0.15	(0.843)
Language (Eng) Diversity	-1.50	(0.149)
Nationality Diversity	-0.02	(0.106)
Ethnicity Diversity	-0.05	(0.862)
GMAT Diversity	0.00	(0.533)
Education Diversity	-0.08	(0.807)
Experience Diversity	0.11	(0.117)
Industry Diversity	-0.65	(0.367)
Function Diversity	-0.89	(0.049)*
Constant	7.18	(0.022)*

N=143; p values in parentheses; + significant at 10%; * at 5%; ** at 1%

Discussion on the Significance of Results

The pattern that appears in both analyses is that significance appears on one diversity variable, but various diversity variables do not appear to have any significant effect. Two possible explanations could account for that situation, and both are derived from the field nature of the empirical setting, a context where actual teams (students) perform actual tasks (of different academic natures).

First, the conditions may not have been as orthogonal as assumed by theory. The theory predicts contradictory effects on variability, which are assumed detectable on orthogonal conditions (across-task vs. along-time). In the context studied, taking different scores in the same course for the along-time condition vs. scores in a different course in the same period for the across-task condition was a good approximation. Unfortunately, in a normal academic setting in a business school, different exercises in the same course are normally strictly on the same task, so that the "along-time" condition also contained some "across-time" aspect. Similarly, considering the first team score across different courses in the same period could not ensure simultaneity of those tasks, because exercises are not given at a fixed date in the period and no archival information is available to control for the slight differences in time. Therefore, the "across-task" conditions also contained some "along-time" aspect. Overall, this setting—as most field settings—did not enforce the perfect orthogonality of conditions, which implied that each analysis contained a slight part of the contradictory effect, therefore reducing significance.

Second, as is common in non-experimental settings, some variance compression was probably occurring. The first source of variance compression is that the underlying population of individuals has a limited variance in some dimensions. MBA students in that top business school were selected with a remarkably limited dispersion in *age* (average of 28 with a standard deviation of roughly 2) and *years of experience* (average of 5 with standard deviation of less than a year). Accordingly, those dimensions are not expected to have enough variance and are included in the analysis as controls. The other potential cause of variance compression is that the teams were constituted with an educational objective to foster learning. Therefore, all teams were assembled with an identical logic, implying composition similarity on some dimensions. An interview was conducted with the current head of program administration to discover the logic used to assemble the teams. The expressed philosophy was: "The MBA office's philosophy in creating MBA study groups is to try to build a range of knowledge and experience into each team so that team members can support each other, especially during their core courses." The teams were assembled to obtain as high an intra-group diversity, with as little variability as possible across-groups. The process was performed manually, with roughly 8 to 10 iterations to obtain the best mix possible. The factors considered when assembling those teams were, in decreasing priority order: 1. nationality; 2. professional background/work experience; 3. gender; 4. educational background 5. GMAT; 6. native English speaker; 7. age.

Whether those dimensions would have enough variance to generate effects was an empirical question. The fact that some variables trigger significant effects proves that—because of the limited rationality of the team composition process—across-team variance in intra-team diversity existed and was not perceived by those assembling the groups. For some variables, one could have ex-ante expected some variance compression, in particular gender diversity. Because gender is such a salient dimension and an intrinsically binary variable, obtaining consistent diversity is a simple task, so the between-groups variance of gender diversity was expected to be minimal.

Despite the non-perfect orthogonality of the conditions and variance compression, significant effects appear on at least two dimensions of diversity: nationality and functional background. Furthermore, the effect appears nearly significant on the other condition for at least one variable: nationality (significantly positive on along-time condition and marginally negative on across-time condition) showing the possibility that a single diversity dimension triggers both social and informational hazard. Finally, one may notice that the different variables are conceptually fitted with their effects. Functional background is likely to have a stronger effect on information availability, especially in a business school context where the tasks are differentiated by such functions; nationality background is likely to have a stronger effect on social integration because it strongly triggers categorizations. Accordingly, the results show that functional diversity is the one amplifying the effect of informational hazard and national diversity is the one reducing the social hazard.

To get perfect significance specifically on each specific dimension, one would want to turn to experimental settings where those could be fully manipulated. However, the objective of the current study was to identify and demonstrate that the hazard hinted at by the moderations documented in team literature would have effects on the uncertainties met in the field. The results demonstrate those effects on a few variables, in the expected direction and in a configuration that makes sense.

CONCLUSION

This study contributes to research on team diversity by bringing nuance to the general U-shaped effect of team diversity on performance variability (Cavarretta, 2007b). Furthermore, it demonstrates in a field context, the contradictory effects of diversity if considering different definitions of variability and advocates the concepts of social vs. informational hazards of team composition.

Previous research hinted at two different underlying processes, driven in low diversity by a social integration hazard and in high diversity by information hazard. The current study uses panel data to disentangle those and show that each is relevant to specific situations, with team diversity increasing performance variability along time, but reducing it across tasks. The pattern that emerges is summarized in Table III-4. This pattern can be leveraged in the organizational situations where performance variability matters for either of the two ideal cases, either cross-task or along-time.

Table III-4: Contradictory Effects of Team Diversity on Performance Variability

Relevant Team Performance Variability:	**cross-tasks (e.g. governance)**	**along-time (e.g. airline, firefighting)**
Diverse Team	Low variability	High variability *"social hazard"*
Homogeneous Team	High variability *"informational hazard"*	Low variability

Such a distinction applies to various substantive organizational contexts where variability is to be avoided. For situations where cross-task performances matter, like governance, team diversity reduces variability. Therefore, homogeneity appears as a dangerous factor because it triggers an informational hazard. One could relate this conclusion to the group think literature (Janis, 1971, 1982) which was developed in the context of governance. It attempted to link negative events to various constructs like cohesion-seeking that are indeed related to team homogeneity. Interestingly, the research had difficulty in establishing that link, but the current study suggests a mechanism to explain the phenomena observed: the negative outcomes associated with homogeneous teams may be not a mean effect but rather a variability effect on team diversity limiting informational hazards.

Theoretically, introducing the concept of informational hazard vs. social hazard allows reinterpreting and reconciling previous studies. For instance, Westphal and Bednar (2005) take exception with Janis (1982) by showing that greater homogeneity is beneficial to avoid pluralistic ignorance, whereby Janis emphasized the detrimental effect of homogeneity. Those two points of view can be reconciled if considering what the underlying hazard is that the organization has to deal with. Janis focused on highly complex situations where cognitive limitations triggered catastrophes. In a situation dominated by an informational hazard, this study predicts that diversity increases chances of catastrophes. Westphal and Bednar focus on the ability of the board to challenge a situation, putting the emphasis on the social dynamics in particular relative to the ability to change over time. Accordingly, they find a beneficial effect of homogeneity, consistent with the conclusion of this study that highly diverse teams bring more catastrophes.

Another benefit of identifying those hazards is to explore the organizational situations rewarding performance variability (i.e. risk). Then, increasing the variability of outcomes, for instance by manipulating team diversity, may subtly improve the prospects. This paper shows that two different paths are possible for this purpose, which may depend on the context and suggests radically different tactics. It appears that in the ideal case where cross-task variability matters—e.g. a creative team inventing a new product with a lot of freedom on its attributes—one may benefit by betting on the informational hazard associated with homogeneous teams. When longitudinal variability matters—e.g. an advertising team trying to win competitions occasionally—one may benefit by betting on the social hazard associated with diverse teams.

The long tradition of studying the effect of diversity on performance has been frustrated by its contradictory effects. In a world where risk—industrial, governance, entrepreneurial, etc.—increasingly matters to organizations, establishing that team composition influences organizational risk by amplifying in opposite directions the social and informational hazard of team life warrants further study.

DISSERTATION CLOSING REMARKS

CONTRIBUTIONS

Inspired by the issue of organizational catastrophes, this dissertation aimed at exploring whether team composition could explain extreme team outcomes. It narrowed its focus onto the conceptual issue of relating extreme performance to a mean-variance tradeoff on one hand, and on the substantive relationship of intra-team diversity to team performance on the other hand. The three chapters cover that spectrum with the first one addressing the conceptual and methodological issue, the second one establishing a general U-shaped relationship between diversity and performance variability, and the third one refining issues of variability and attempting indirect verification of the underlying contingency mechanisms.

In addition to covering those objectives, each chapter opens additional venues of interest to organizational theory researchers. The first chapter enlarges the issue of the mean-variance tradeoff beyond the issue of predicting catastrophes. It shows that various organizational theory perspectives depend on attainment of performance levels that are not average, and therefore could all be impacted when effects on variability occurs. Furthermore, this concerns not only perspectives related to catastrophic low performance levels like in High Reliability Organization (HRO) or in Governance but also perspectives where extraordinary high performance levels matters such as in Entrepreneurship or in Innovation.

The second chapter demonstrates that existing research could gain from considering variability effects, whether applied substantively to team diversity theory or to any other perspective where effects on variability are expected. What is only a conceptual conjecture in chapter 1 turns into a example where the traditional analyses (as conducted by Kilduff, Angelmar, & Mehra, 2000) can be interestingly reinterpreted by considering extreme effects. In the context of that study (business executives in training), it is credible that the teams cared more about being among the very best teams more than being slightly above or below average. In that situation, the analysis shows that the effects of team diversity need then to be reassessed. In Figure I-1, low age diversity appears preferable even though the classical analysis suggest age diversity increases performance on average; in Figure II-6, variability analysis allows a conclusion—high functional diversity is better in order to belong to the top 3 teams—even though classical analysis is mute since no average effect is detectable.

The chapter III pushes the induction of chapter II further. The logic is that two factors (integration and information-task fit) moderate the relationship between diversity and performance. Since those factors can be considered stochastic in the field, I label them social and informational hazards and show that diversity changes how such hazards influence the performance variability depending if one consider population vs. cross-task vs. along-time variability. In a very large sense, it is encouraging to consider that the many moderation relationships already demonstrated in organizational theory could be reinterpreted in the same spirit (through a variability lens) to contribute to the scholarship of organizational risk.

LIMITATIONS

Various limitations appeared in this research, most notably regarding the empirical verifications. In the chapter II, not all effects are significant at the same time, and in chapter III, the effects do not appear on all dimensions. If such apparent limited significance may lead to questioning whether the theory is truly verified, one should however consider that those limitations might arise because both the dependent variable and the independent variables have particular characteristics.

Regarding the dependent variables, using performance variability and in some cases, predicting a curvilinear effect (chapter II) may require changing the expectations about significance. Statistically, variability is the second moment of the distribution and therefore a second-order quantity. It will hence be more subject to noise than the first order quantity (the first moment, the mean). Therefore, one could expect that the significance of effects on variability will be less significant than the mea effects usually analyzed by classical methods. Anticipating such issues, the second chapter therefore avoids stating a curvilinear hypothesis because testing for it would require an additional order of analysis, therefore leading to testing a third order effect which would be even less comparable to classical studies. The study leverages the natural-experiment setting, allowing testing for second order effects that are already difficult to establish. The third chapter takes advantage of a similar trick to avoid having to estimate the curvilinear effect on the residual.[15]

Regarding the independent variables—the team diversity measures—most were subjects to serious variance limitations in the field contexts that this research explored, whether the executive education or the MBA programs in business schools, or even in the alternatives to those settings that I considered (military teams, startup teams, etc.). A first mechanical issue is that, for categorical variables (e.g. ethnicity), the diversity will exhibit very few possible points. For instance, on a team of seven, there are only three possible

[15] Future research might explore—with the help of professional statisticians—improved estimation methods for linear and curvilinear effects on variability.

values if two ethnicities are present (the three possible combinations are: 1 individual of a category and 6 of the others; 2+5; and 3+4). A few more points appear with three distinct ethnicities (1+1+5;1+2+4;1+3+3;2+2+3). Those, plus the few cases for more than three diversity categories, and including the complete homogeneity still makes less than a dozen points. This mechanical restriction of range is even worse in many contexts where teams are small. For instance, most startup teams would contain three individuals, with a very few at four and even less beyond. This leads to the actual variance available on categorical measures (gender, ethnicity, etc.) being highly constrained, sometimes exhibiting no more than three values. Obviously, such cases hamper many analyses, in particular for testing curvilinearity.

A second variable limitation concerns the constraints on the demographic variety in the population of individuals out of which the teams are built. For instance, the executives in chapter II were from a relatively narrow pool of nationality, a situation typical of many executive teams in many industries. Similarly, the MBA students in Chapter III exhibit a narrow range of age. Therefore, the diversity range observed on a population of teams built on a pool of individual with narrow range can only be of narrow range of diversity.

The third issue of independent variable limitations is of social nature. Whatever the nature of the variable (categorical or continuous) and whatever the distribution of individuals, in the field context studied, there were norms regarding the proper construction of teams and therefore the observed range is limited. In Chapter III, teams are assembled following the norm that functional diversity should be high in management teams; therefore, all teams were in relatively high values of diversity, effectively limiting the range on that variable. This also occurred in Chapter III; the MBA students' teams were assembled with the strong intent to maximize the diversity on the few dimensions where some individual variance was available. For instance, nationalities in that program were relatively varied; however, the MBA administration made a point of maximizing the intra-team national diversity, effectively reducing the range of nationality diversity between-teams.

Those factors combine to a point where one wonders how much variance of each diversity independent variables is actually available for analysis. Most variables are not subject to at least one, if not a combination, of the three above issues: mechanical limitation on the number of points, limited underlying population, norms constraining construction of teams. Nevertheless, these limitations strongly match the limitations occurring in most field conditions, whether in an industrial context or in other organizational contexts such as in the military. In all contexts, the mechanical limitations occur by definition. In many field contexts, the underlying population has a limited variety, therefore constraining diversity at the team level (Pfeffer, 1997). In addition, in most field contexts, norms exist that make the teams have similar characteristics.

Overall, various conditions conspire to make the statistical significance of the results not as strong as could be desired. Yet, the significances that appear in Chapter II are strong enough to have material consequences, as illustrated by the Figure II-5 and Figure II-6. Furthermore, this research sheds light on the particularities of the diversity variables as they occur in the field.

NEXT STEPS

Future research will be inspired both by the positive thrust of these studies as well as by their current limitations. Obviously, a first step would be to explore further empirical situations in order to clarify the direction and significance of the results. Additional field data may allow for clearer results if the identified sources of problems are mitigated. In particular, one may explore contexts where teams are assembled with less determinism regarding the diversity of the members and pulling from a population of individuals that are truly diverse. In the extreme, this also suggests the possibility of running experiments where the population and the teams' composition could be manipulated. Such experimental approach would also allow either for measuring the stochastic factors (social integration and information-task fit) or for manipulating them. The former would strengthen the argument about stochasticity in team life; the later would strengthen the theoretical induction by directly manipulating the moderating factors and therefore settling the exact nature of those moderations.

An alternative approach to quantitative verification would study specifically top management teams (TMT) and firm performance. Its benefits are two-fold: first, TMT information and firm performance are collectable or already collected on a large scale, therefore bringing more statistical power to the analysis; second, linking TMT composition to variability of firm performance could be appealing to a larger research context beyond organizational theory. Variability is strongly related to measures both of risk (interesting to governance and audit scholarship) as well as volatility (interesting to financial theory), such study could build an interesting bridge between organizational scholars and other scholarships in business schools.

Beyond strengthening the current substantive developments, the research could be enlarged in various directions. A first direction would be to explore whether the relationship between team composition and performance variability has differential effects depending on the perception of risk in different contexts. Some fields may be driven by a serious aversion to risk: for instance, airline cockpit crews have a primary obligation to ensure safety and avoid the catastrophic outcome of accidents above the objective to improve average performance (e.g. fuel consumption). By contrast, some fields may be driven by an appetence to risk: for instance, startup teams in the portfolio of large Silicon Valley venture capital firms are expected to reach rare and extraordinary level of performance (e.g. IPO). In the middle, some fields may be neutral to risk: for instance, sales teams may be rewarded with a constant ratio of the sales, and therefore only motivated by the expected outcome (i.e. mean).

When considering those three ideal cases, a relationship between team composition and performance variability should have different effects in the different environments, and therefore it may appear that, by adaptation, those different fields institutionalize team composition differentially. For instance, airlines may decree that both extremely homogeneous teams and extremely diverse teams are "dangerous" and therefore seek to avoid them. By contrast, venture capitalist may be tolerant to extreme teams (very homogeneous or very diverse), and may even seek them at the same time that they consciously or unconsciously attempt to mitigate the hazards that such extremes entail. Finally, the linear-reward sales environment may be completely neutral to such concerns, or be focused only on composition issues that relates to mean effect (e.g. some diversity increases performance).

Those considerations suggest designing research that would study the processes and the qualitative aspects of team composition and life, across different fields. Finding different institutionalizations of team composition would not only bring further verification to theory, but also deliver practical knowledge about how different types of diversity can be managed, in relationship to organizational risk.

CONCLUSION

The dissertation was initiated on a simple intuition: the top management team of firms like Vivendi-Universal or Enron failed not because they were predictable underperformers, but more because some characteristics made them more likely to both perform extremely well, as well as fail dramatically. Assuming the firm outcomes reflect the performance of the top management team, it focused on linking the characteristics of teams, in particular their demographic diversity, to team performance variability. Extreme performance occurred for teams that had either the most homogeneous and most diverse compositions. I hope future research can link the social and informational hazards demonstrated here to examples of actual outstanding failures and successes in various organizational contexts, and that such an approach finds its place in the scholarship of organizational risk.

DISCUSSION ET CONCLUSION

CONTRIBUTIONS DE L'ETUDE

Inspiré par l'étude des catastrophes organisationnelles, cette étude explore l'hypothèse que la composition d'équipe influence la performance extrême des équipes. Son objectif conceptuel est focalisé sur le contraste entre les performances extrêmes et les performances moyennes d'une part, et sur la relation substantive entre la diversité intra-équipe et la performance de l'équipe d'autre part. Les trois chapitres traitent successivement des questions conceptuelles et méthodologiques relatives à la variabilité, de la relation générale (curvilinéaire en U) entre la diversité et la variabilité des performances, et finalement d'un affinement des définitions possibles de la variabilité. De plus, chaque chapitre ouvre des voies de recherches possibles en théorie des organisations.

Le premier chapitre développe le compromis moyenne-variance au-delà de la question de prévoir les catastrophes. Il montre que diverses perspectives théoriques dépendent du niveau de performance recherché, que celui-ci n'est pas toujours la moyenne, et que leurs conclusions pourraient donc être altérées lorsque des effets sur la variabilité sont pris en compte. Cela concerne non seulement les perspectives étudiant les faibles niveaux de performance comme dans l'étude des organisations à haute fiabilité (HRO) ou en gouvernance, mais aussi les perspectives étudiant les niveaux de performance élevés comme dans l'entrepreneuriat ou l'innovation.

Le deuxième chapitre démontre une relation curvilinéaire en U entre la diversité d'équipe et la variabilité de la performance : les équipes très homogènes et les équipes très diverses auront une performance plus variable que les équipes moyennement diverses. Au-delà d'établir cette relation générale entre diversité et variabilité pour les équipes, ce chapitre montre, d'un point de vue empirique, comment la recherche organisationnelle peut être améliorée en prenant en compte la variabilité des effets, que ce soit lorsqu'elle étudie la diversité des équipes ou toute autre question pour laquelle des effets sur la variabilité sont possibles.

Ce qui n'est qu'une conjecture conceptuelle dans le chapitre I se transforme au chapitre II en un exemple concret où des analyses traditionnelles à base d'effets moyens (telles qu'effectuées par Kilduff, Angelmar, & Mehra, 2000) peuvent être réinterprétés en prenant en compte les effets sur la variabilité. Dans le contexte de cette étude (des dirigeants en formation), il est crédible que les équipes ambitionnaient davantage d'être parmi les meilleures que d'être légèrement au-dessus ou en dessous de la moyenne. Dans ce cas, l'analyse montre que les effets de la diversité doivent être réévalués. Dans la Figure II-5, la faible diversité en âge apparaît préférable, même si l'analyse classique suggérait que la diversité en âge augmente la performance en moyenne ; dans la Figure II-6, l'analyse de la variabilité permet d'atteindre une conclusion (la diversité fonctionnelle est préférable si l'on désire faire partie des trois meilleures équipes) même si l'analyse classique est inopérante car aucun effet moyen n'est décelable.

Le chapitre III pousse l'induction du chapitre II plus loin. La logique est que deux facteurs (intégration sociale et correspondance entre la tâche et les connaissances disponibles) modèrent la relation entre la diversité et la performance. Étant donné que ces facteurs peuvent être considérés comme stochastiques, je les dénomme aléas sociaux et informationnels et montre que la diversité a un effet différent sur la variabilité selon qu'elle est considérée longitudinalement ou inter-tâche. De manière générale, il est encourageant d'envisager que les nombreuses relations où des modérations ont déjà été démontrées pourraient être réinterprétées dans le même esprit afin de contribuer au champ du risque organisationnel.

LIMITATIONS DE L'ETUDE

Diverses limitations apparaissent dans cette recherche, notamment en ce qui concerne les vérifications empiriques. Dans le chapitre II, tous les effets ne sont pas significatifs conjointement et, au chapitre III, les effets n'apparaissent pas sur toutes les dimensions. Si ces limitations semblent amener à douter de la vérification de la théorie, il faut cependant considérer que les variables étudiées, dépendantes et indépendantes, avaient des caractéristiques qui posaient des obstacles particuliers.

En ce qui concerne les variables dépendantes (variabilités de la performance) en général et l'estimation d'effets curvilinéaires au chapitre II en particulier, il convient de réexaminer quelles attentes sont raisonnables en terme de significativité statistique. La variabilité est le deuxième moment de la distribution de la performance et, par conséquent, une quantité d'ordre deux. Elle est donc plus sujet au bruit que la quantité d'ordre un (le premier moment, la moyenne). Par conséquent, on doit s'attendre à ce que le caractère significatif des effets sur la variabilité soit moindre que les effets généralement constatés dans des analyses traditionnelles. Pour anticiper ces questions, le deuxième chapitre évite donc d'énoncer une hypothèse curvilinéaire car cela exigerait un ordre de grandeur supplémentaire dans l'analyse, donc nécessitant la détection d'effet d'ordre trois qui seraient encore moins comparables à des études classiques (l'effet moyen est d'ordre un ; l'effet sur la variabilité d'ordre deux ; la curvilinéarité sur la variabilité est donc d'ordre trois). L'étude tire profit de la nature quasi expérimentale du contexte pour ne tester que des effets d'ordre deux, qui sont malheureusement déjà difficiles à établir. Le troisième chapitre tire parti de la même astuce pour éviter d'avoir à estimer des effets curvilinéaires sur des résidus.

En ce qui concerne les variables indépendantes (les mesures de diversité de l'équipe), la plupart étaient sujettes à de graves limitations empiriques, limitations qui seraient apparues aussi dans les alternatives empiriques qui ont été envisagées dans la période de sélection du terrain de recherche (ex. équipes militaires ou entrepreneuriales). Un premier problème, mécanique, est que pour les variables catégorielles (par exemple l'origine ethnique), très peu de valeurs sont possibles. Par exemple, sur une équipe de sept personnes, il n'existe que trois valeurs possibles si deux ethnies sont représentées (les trois combinaisons sont les suivantes : 1 personne d'une catégorie et 6 de l'autre ; 2 +5 et 3 +4). Un peu plus de possibilités apparaissent avec trois ethnies distinctes (1 +1 +5 ; 1 +2 +4 ; 1 +3 +3 ; 2 +2 +3). Dans les rares cas où plus d'ethnies sont représentées, et en prenant en compte l'homogénéité complète, cela ne donne que moins d'une dizaine de valeurs possibles à la variable de diversité. Cette limitation mécanique est encore plus forte dans les contextes où les équipes sont de petite taille. Par exemple, la plupart des équipes de startup sont formées de trois personnes maximum. Cela conduit à ce que le nombre de points réellement disponibles pour les variables catégorielles (sexe, origine ethnique, etc.) est très limité, ne présentant parfois pas plus de trois valeurs. De toute évidence, de tels cas gênent les analyses, en particulier si des effets curvilinéaires sont recherchés.

Une deuxième limitation sur les variables concerne les contraintes qui pèsent sur la diversité démographique dans la population des individus à partir de laquelle les équipes sont composées. Par exemple, au chapitre II, les dirigeants ont une diversité de nationalité relativement étroite, une situation typique de nombreuses équipes de direction. De même, au chapitre III, les étudiants MBA présentent une gamme d'âge étroite. Ceci a pour conséquence que la gamme de diversités observées sur une population d'équipes est réduite tout simplement car la population des individus composant les équipes manque de diversité.

La troisième limitation sur les variables indépendantes est de nature sociale. En effet, sur les terrains considérés, de fortes normes existaient concernant la composition des équipes, ce qui tendait à homogénéiser la population d'équipe et donc limitait la gamme des valeurs possibles. Au chapitre II, les équipes de dirigeants sont assemblées selon la norme que la diversité fonctionnelle doit être élevée dans les équipes de management et, par conséquent, toutes les équipes sont relativement diverses fonctionnellement, créant une forte limitation de gamme sur cette variable « fonction ». Il en est de même dans le chapitre III, les équipes d'étudiant en MBA ayant été assemblées avec la ferme intention de maximiser la diversité d'équipe sur les quelques aspects où de la diversité individuelle était disponible. Par exemple, les étudiants sont relativement divers en termes de nationalités dans ce programme, mais l'administration du MBA fait de gros efforts pour maximiser la diversité nationale au sein de chaque équipe, réduisant la gamme des valeurs possibles pour la diversité intra-équipes.

Ces facteurs se combinent au point que l'on peut se demander s'il reste assez de variance sur les variables indépendantes pour pouvoir mener des analyses. La plupart des variables peuvent être l'objet d'une – ou d'une combinaison – des trois limitations évoquées ci-dessus : 1) mécanique de limitation sur le nombre de points, 2) faible diversité de la population d'individus composant les équipes, et 3) existence de normes contraignants les compositions des équipes. Ces limitations correspondent à la réalité du terrain, que ce soit dans un contexte industriel ou dans d'autres contextes organisationnels - l'armée par exemple. En premier lieu, dans tous les contextes, la limitation mécanique va se produire, par définition. Ensuite, dans de nombreux contextes, les populations d'individus ont une diversité limitée, ce qui va donc limiter la diversité au niveau de l'équipe (Pfeffer, 1997). Enfin, dans la plupart des contextes, des normes existent qui poussent les équipes à avoir des caractéristiques similaires.

Globalement, diverses conditions concourent à faire que la significativité statistique des résultats soit limitée. Néanmoins, on remarquera que les résultats qui apparaissent dans le chapitre II sont suffisamment forts pour avoir des conséquences concrètes, comme le montrent la Figure II-5 et la Figure II-6. Et même dans les cas où la significativité est très réduite, cette étude a le mérite de démontrer les particularités des variables de diversité telles qu'elles se présentent sur le terrain.

PROCHAINES ETAPES

De futures recherches pourraient s'inspirer des résultats positifs et des limitations de ces études. De toute évidence, une première étape serait d'explorer de nouvelles situations empiriques afin de clarifier le sens et la signification des résultats. D'autres données de terrain permettraient de clarifier les résultats, si les sources de problèmes sont réduites. En particulier, on pourrait explorer les contextes où des équipes sont assemblées avec moins de déterminisme en ce qui concerne la diversité des membres, et partant d'une population d'individus qui soit vraiment diverse. Dans les cas extrêmes, cela suggère également d'utiliser des méthodes expérimentales où la population et les compositions des équipes seraient manipulées. Ceci permettrait soit de mesurer les facteurs stochastiques (intégration sociale et correspondance connaissance-tâche), soit de les manipuler. La première approche permettrait d'approfondir le concept d'aléa de la vie des équipes ; la seconde de renforcer l'induction théorique en manipulant directement les facteurs de modération, et donc d'établir leur effet exact.

Une autre approche quantitative pourrait étudier spécifiquement les équipes de direction et leur influence sur le fonctionnement des entreprises. Deux bénéfices seraient attendus : premièrement, les données sur des équipes de direction et la performance des entreprises sont collectables à grande échelle, ce qui apporterait donc plus de puissance statistique pour l'analyse. Deuxièmement, l'existence d'un lien entre la composition de l'équipe de direction et la variabilité de la performance de l'entreprise pourrait être attrayante pour une recherche au-delà de la théorie des organisations. La variabilité étant étroitement liée à des mesures de risque (qui intéressent la gouvernance et l'audit) et de volatilité (qui intéressent la théorie financière), cette approche pourrait constituer un pont intéressant entre la recherche organisationnelle et d'autres départements des écoles de gestion.

Cette recherche pourrait aussi être élargie par une approche qualitative. Une première orientation serait d'étudier si la relation entre composition de l'équipe et la variabilité des performances a des effets différents selon la perception du risque. Certains contextes font apparaître une forte aversion au risque : par exemple, les équipages d'avions de ligne ont une obligation d'assurer la sécurité et d'éviter toute catastrophe avant tout objectif d'amélioration de la performance moyenne (par exemple la consommation de carburant). En revanche, d'autres contextes font apparaître une grande appétence au risque : par exemple, les startups dans le portefeuille des grands fonds d'investissement en capital-risque doivent atteindre des niveaux de performance rare et extraordinaire (par exemple l'introduction en bourse). Enfin, dans d'autres contextes, il apparaît une certaine neutralité au risque : par exemple, des équipes de vente peuvent être récompensées au prorata des ventes, et donc seulement motivées par le résultat attendu (c'est-à-dire moyen).

En comparant ces trois cas typiques, la relation entre la composition de l'équipe et la variabilité de la performance devrait avoir des effets différents dans les différents environnements. En conséquence, il est possible que chaque contexte se soit adapté et ait institutionnalisé la composition des équipes différemment. Par exemple, les compagnies aériennes peuvent constater que les équipes très homogènes et très diverses sont «dangereuses» et, par conséquent, chercher à les éviter. En revanche, le capital-risqueur peut être tolérant aux équipes extrêmes (très homogènes ou très diverses), et même les rechercher, tout en tentant – consciemment ou non – d'atténuer les risques que de tels extrêmes entraînent. Enfin, l’environnement de vente à récompense linéaire pourrait faire apparaître une neutralité par rapport à la composition, ou se concentrer seulement sur les effets moyens de la composition (par exemple une certaine diversité augmente les performances).

Ces considérations suggèrent de concevoir des recherches qui étudieraient les processus et les aspects qualitatifs de la vie de l'équipe. Mettre en évidence différentes institutionnalisations de la composition de l'équipe pourrait non seulement apporter une confirmation de la théorie, mais aussi nous informer sur la façon dont différents types de diversité peuvent être gérés.

CONCLUSION

Ce travail fut initié sur une simple supputation: les équipes de direction d'entreprises comme Vivendi ou Enron pourraient avoir échoué non pas parce qu'elles étaient sous-performantes en moyenne, mais plus parce que certaines caractéristiques les avaient rendu plus susceptibles soit de réussir extraordinairement, soit d’échouer extraordinairement. En supposant que les résultats de l’entreprise reflètent la performance de l'équipe de direction, ce travail s’est focalisé sur le lien entre les caractéristiques des équipes, en particulier leur diversité démographique, et la variabilité de la performance de l'équipe. Il est apparu que l’on pouvait prédire des performances plus extrêmes à la fois pour les équipes les plus homogènes et les plus diverses. J'espère que des travaux futurs pourront lier les aléas sociaux et informationnels identifiés ici à d’autres exemples dans différents contextes de performance organisationnelle exceptionnelle, et que cette approche trouvera sa place dans les théories du risque organisationnel.

APPENDIX

FORMAL DETERMINATION 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 2: $Y_X \sim F(z_X)$ with $Z(Y_X) = \dfrac{Y - \mu_X}{\sigma_X}$ 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—or enforces—homoskedasticity. Furthermore, such parameterization can be viewed as a linearization of more complex models, modeling only the first-order effects—both on the mean and on the variance—and ignoring all higher-order effects (quadratic, 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. Sørensen, 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* impacts positively 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 at the same time reduces variability, what can we conclude?

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). We assume that the cumulative distribution function *F* depends only on the *z*-score. Then, for each X, the cumulated probability of Y being above Y_0 can be computed as follow:

Equation 3: $p_{Y_0}(X) = P[Y_X > Y_0] = F(z_{Y_0}) = F\left(\dfrac{\mu(X) - Y_0}{\sigma(X)}\right) = F\left(\dfrac{\beta_1 X + \beta_0 - Y_0}{\gamma_1 X + \gamma_0}\right)$

The quantile curves, linking points of equal probability, have therefore the equation where p(X) is a constant. *F* is a cumulated distribution function and is therefore monotone; hence, quantile curves are defined by making the ratio inside F constant, leading to a linear equation. Therefore, those curves are simple lines, justifying the representation of Figure I-1.

To determine at which level such line is horizontal, one has simply to explore when the ratio inside F has a null derivative. If the derivative is taken and made zero, it solves in Y_0, providing the critical value, Y_c:

Equation 4: $Y_c = \beta_0 - \gamma_0 \dfrac{\beta_1}{\gamma_1}$

The quantile lines therefore change direction and only once. Around average performance, all quantile lines have the same slope direction, which inverts beyond the critical level.

RÉFÉRENCES

Adler, N. J. 2002. **International dimensions of organizational behavior** (4th ed.). Cincinnati, Ohio: South-Western.

Allmendinger, & Hackman, J. R. 1995. The More the Better? A Four-Nation Study of the Inclusion of Women in Symphony Orchestras. **Social Forces**, 74(2): 423.

Ancona, D., Bresman, H., & Kaeufer, K. 2002. The Comparative Advantage of X-Teams. **MIT Sloan Management Review**, 43(3): 33.

Ancona, D. G., Goodman, P. S., Lawrence, B. S., & Tushman, M. L. 2001. Time: a new research lens. **Academy of Management Review**, 26(4): 645-563.

Argote, L. 1999. **Organisational Learning: Creating, Retaining and Transferring Knowledge**. Norwell, MA: Kluwer Academic Publishers.

Barnett, M. L. 2007. Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. **Academy of Management Review**, 32(3): 794-816.

Barnett, W. P., Swanson, A.-N., & Sorenson, O. 2003. Asymmetric Selection Among Organizations. **Industrial & Corporate Change**, 12(4): 673-695.

Baum, J. A. C., & McKelvey, B. 2006. Analysis of Extremes in Management Studies. In D. J. Ketchen & D. D. Bergh (Eds.), **Research Methodology in Strategy and Management**, Vol. 3: 125-199.

Black, F., & Scholes, M. 1973. The Pricing of Options and Corporate Liabilities. **Journal of Political Economy**, 81(3): 637.

Blau, P. M. 1977. **Inequality and Heterogeneity: a Primitive Theory of Social Structure**. New York: Free Press.

Boone, C., Van Olffen, W., & Van Witteloostuijn, A. 2005. Team locus-of-control composition, leadership structure, information acquisition, and financial performance: a business simulation study. **Academy of Management Journal**, 48(5): 889-909.

Bourgeois, L. J., III. 1985. Strategic Goals, Perceived Uncertainty, and Economic Performance in Volatile Environments., **Academy of Management Journal**, Vol. 28: 548: Academy of Management.

Bowers, C. A. 2000. When Member Homogeneity Is Needed in Work Teams. **Small Group Research**, 31(3): 305.

Bowman, E. H. 1980. A Risk/Return Paradox for Strategic Management. **Sloan Management Review**, 21(3): 17-31.

Bromiley, P., Miller, K. D., & Rau, D. 2001. Risk in Strategic Management Research. **Blackwell Handbook of Strategic Management**: 259.

Bunderson, J. S., & Sutcliffe, K. M. 2002. Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. **Academy of Management Journal**, 45(5): 875.

Cabral, L. M. B. 2003. R&D Competition When Firms Choose Variance. **Journal of Economics & Management Strategy**, 12(1): 139-150.

Cavarretta, F. 2007a. Better, best or worst teams? Linking intra-team diversity to extreme performance, **INSEAD Working Paper 2007/56/OB**. Fontainebleau.

Cavarretta, F. 2007b. Impacts of intrateam diversity on team performance variance: two ways to take chances. **Academy of Management Proceedings**: 1-6.

Chatterjee, A., & Hambrick, D. C. 2007. It's All about Me: Narcissistic Chief Executive Officers and Their Effects on Company Strategy and Performance. **Administrative Science Quarterly**, 52(3): 351-386.

Cramton, C. D., & Hinds, P. J. 2005. Subgroup dynamics in internationally distributed teams: ethnocentrism or cross-national learning? In B. M. Staw & R. M. Kramer (Eds.), **Research in Organizational Behavior**, Vol. 26: 231-263: Elsevier Science/JAI Press.

Cyert, R. M., & March, J. G. 1963 [1992]. **A Behavioral Theory of the Firm**. Malden, MA: Blackwell.

Daft, R. L., & Lewin, A. Y. 1990. Can Organizational Studies Begin to Break out of the Normal Science Straitjacket? An Editorial Essay. **Organization Science**, 1(1): 1.

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. 2005. Team diversity and information use. **Academy of Management Journal**, 48(6): 1107-1123.

Denrell, J. 2003. Vicarious Learning, Undersampling of Failure, and the Myths of

Management. **Organization Science**, 14(3): 227-243.

Dineen, B. R., Noe, R. A., Shaw, J. D., Duffy, M. K., & Wiethoff, C. 2007. Level and dispersion of satisfaction in teams: Using foci and social context to explain the satisfaction-absenteeism relationship. **Academy of Management Journal**, 50(3): 623-643.

Earley, P. C., & Mosakowski, E. 2000. Creating hybrid team cultures: an empirical test of transnational team functioning. **Academy of Management Journal**, 43(1): 26-49.

Eisenhardt, K. M., & Sull, D. N. 2001. Strategy as simple rules. **Harvard Business Review**, 79(1): 106-116.

Esser, J. K. 1998. Alive and Well after 25 Years: A Review of Groupthink Research. **Organizational Behavior & Human Decision Processes**, 73(2/3): 116-141.

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

Fleming, L. 2001. Recombinant uncertainty in technological search. **Management Science**, 47(1): 117-132.

Fleming, L. 2004. Perfecting Cross-Pollination. **Harvard Business Review**, 82(9): 22-24.

Fligstein, N. 1987. The Intraorganizational Power Struggle: Rise of Finance Personnel to Top Leadership in Large Corporations, 1919-1979. **American Sociological Review**, 52(1): 44-58.

Friedkin, N. E. 1999. Choice Shift and Group Polarization. **American Sociological Review**, 64(6): 856-875.

Galambos, L., & Sturchio, J. L. 1998. Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation. **Business History Review**, 72(2): 250.

Gersick, C. J. G. 1988. Time and Transition in Work Teams: Toward a New Model of Group Development. **Academy of Management Journal**, 31(1): 9-41.

Gibson, C., & Vermeulen, F. 2003. A Healthy Divide: Subgroups as a Stimulus for Team Learning Behavior., **Administrative Science Quarterly**, Vol. 48: 202-239: Administrative Science Quarterly.

Gladwell, M. 2007. Open Secrets, **The New Yorker**, Vol. Jan 8, 07: 44.

Greene, W. H. 2003. **Econometric analysis** (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. 1996. Group composition and decision making: How member familiarity and information distribution affect process and performance. **Organizational Behavior & Human Decision Processes**, 67(1): 1-15.

Hackman, J. R. 1990. **Groups that work (and those that don't): creating conditions for effective teamwork** (1st ed.). San Francisco: Jossey-Bass.

Hambrick, D. C., & Mason, P. A. 1984. Upper Echelons: The Organization as a Reflection of Its Top Managers. **Academy of Management Review**, 9(2): 193.

Hannan, M. T., & Freeman, J. 1984. Structural Inertia and Organizational Change. **American Sociological Review**, 49(2): 149-164.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. 2002. Time, teams, and task performance: changing effects of surface- and deep-level diversity on group functioning. **Academy of Management Journal**, 45(5): 1029-1045.

Harrison, D. A., & Klein, K. J. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. **Academy of Management Review**, 32(4): 1199-1228.

Hofstede, G., Neuijen, B., Ohayv, D. D., & Sanders, G. 1990. Measuring Organizational Cultures. **Administrative Science Quarterly**, 35(2): 286-316.

Hofstede, G.; Geert Hofstede™ Cultural Dimensions; http://www.geert-hofstede.com/hofstede_dimensions.php.

Jackson, S. E., Joshi, A., & Erhardt, N. L. 2003. Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications. **Journal of Management**, 29(6): 801.

Janis, I. L. 1971. Groupthink. **Psychology Today**: 43-46.

Janis, I. L. 1982. **Groupthink: psychological studies of policy decisions and fiascoes** (2nd ed.). Boston: Houghton Mifflin.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. 1999. Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. **Administrative Science Quarterly**, 44(4): 741-763,.

Kahneman, D., & Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. **Econometrica**, 47: 263-291.

Kalnins, A. 2007. Sample selection and theory development: implications of firms' varying abilities to appropriately select new ventures. **Academy of Management Review**, 32(4): 1246-1264.

Kenney, M., & von Burg, U. 1999. Technology, entrepreneurship and path dependence: industrial clustering in Silicon Valley and Route 128. **Industrial & Corporate Change**, 8(1): 67.

Khurana, R. 2002. The Curse of the Superstar CEO. **Harvard Business Review**, 80(9): 60.

Kilduff, M., Angelmar, R., & Mehra, A. 2000. Top Management-Team Diversity and Firm Performance: Examining the Role of Cognitions., **Organization Science**, Vol. 11: 21: INFORMS: Institute for Operations Research.

Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K. A., Leonard, J., Levine, D., & Thomas, D. 2003. The Effects of Diversity on Business Performance: Report of the Diversity Research Network. **Human Resource Management**, 42(1): 3.

Kogut, B., & Singh, H. 1988. The effect of national culture on the choice of entry mode. **Journal of International Business Studies**, 19(3): 411-432.

Kogut, B. 1991. Joint Ventures and the Option to Expand and Acquire. **Management Science**, 37(1): 19-33.

Lau, D. C., & Murnighan, J. K. 1998. Demographic diversity and faultlines: the compositional dynamics of organizational groups. **Academy of Management Review**, 23(2): 325-340.

Lawrence, B. S. 1997. The Black Box of Organizational Demography. **Organization Science**, 8(1): 1.

Lieberson, S., & O'Connor, J. F. 1972. Leadership and Organizational Performance: A Study of Large Corporations. **American Sociological Review**, 37(2): 117-130.

Luthans, F. 2002. **Organizational behavior** (9th ed.). Boston: McGraw-Hill/Irwin.

Mandelbrot, B. 1960. The Pareto-Levy Law and the Distribution of Income. **International Economic Review**, 1(2): 79-106.

Mannix, E. A., & Neale, M. A. 2005. What Differences Make a Difference? **Psychological Science in the Public Interest**, 6(2): 31-55.

March, J. G. 1991. Exploration and Exploitation in Organizational Learning. **Organization Science**, 2(1): 71.

March, J. G., & Shapira, Z. 1992. Variable Risk Preferences and the Focus of Attention. **Psychological Review**, 99: 172-183.

McKelvey, B., & Andriani, P. 2005. Why Gaussian Statistics are Mostly Wrong for Strategic Organization. **Strategic Organization**, 3(2): 219-228.

McKelvey, B. 2006. Van De Ven and Johnson's "Engaged Scholarship": Nice Try, But. **Academy of Management Review**, 31(4): 822.

Meyer, J. W., & Rowan, B. 1977. Institutionalized Organizations: Formal Structure as Myth and Ceremony. **American Journal of Sociology**, 83(2): 340-363.

Milliken, F. J., & Martins, L. L. 1996. Searching for common threads: understanding the multiple effects of diversity in organizational groups. **Academy of Management Review**, 21(2): 402-433.

Miner, A. S., Haunschild, P. R., & Schwab, A. 2003. Experience and Convergence: Curiosities and Speculation. **Industrial & Corporate Change**, 12(4): 789-813.

Mohr, L. B. 1982. **Explaining organizational behavior** (1st ed.). San Francisco: Jossey-Bass.

Nemeth, C. J. 1986. Differential contributions of majority and minority influence. **Psychological Review**, 93(1): 23-32.

Nemeth, C. J., & Staw, B. M. 1989. The tradeoffs of social control and innovation in groups and organizations. In L. Berkowitz (Ed.), **Advances in experimental social psychology**, Vol. 22: 175-210: Academic Press, Inc.

O'Reilly, C. A., Caldwell, D. F., & Barnett, W. P. 1989. Work Group Demography, Social Integration, and Turnover. **Administrative Science Quarterly**, 34: 21-37.

O'Reilly, C. A. I., Williams, K. Y., & Barsade, S. 1998. Group demography and innovation: Does diversity help? In D. H. Gruenfeld (Ed.), **Composition.**: 183-207: Elsevier Science/JAI Press.

Page, S. E. 2007. **The diversity effect : how individual difference creates collective benefits**. Princeton: Princeton University Press.

Palmer, D., & Maher, M. 2004. Developing the Process Model of Collective Organizational Wrongdoing, **Annual Meeting of the Western Academy of Management**. Alyeska, Alaska: AOM.

Péli, G., Masuch, M., Bruggeman, J., & Nualláin, B. ó. 1994. A Logical Approach to Organizational Ecology: Formalizing the Inertia-Fragment in First-Order Logic. **American Sociological Review**, 59(4): 571-593.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. 1999. Exploring the Black Box: An Analysis of Work Group Diversity, Conflict, and Performance., **Administrative Science Quarterly**, Vol. 44: 1: Administrative Science Quarterly.

Perrow, C. 1984. **Normal accidents: living with high-risk technologies**. New York: Basic Books.

Peterson, R. S., & Behfar, K. J. 2003. The dynamic relationship between performance feedback, trust, and conflict in groups: A longitudinal study., **Organizational Behavior & Human Decision Processes**, Vol. 92: 102.

Pfeffer, J., & Salancik, G. R. 1978. **The External Control of Organizations: A Resource Dependence Perspective**. New York, NY: Harper and Row.

Pfeffer, J. 1997. **New directions for organization theory: problems and prospects**. New York: Oxford University Press.

Polzer, J. T., Milton, L. P., & Swann, W. B. J. 2002. Capitalizing on diversity: Interpersonal congruence in small work groups. **Administrative Science Quarterly**, 47(2): 296-324.

Richard, O. C., Barnett, T., Dwyer, S., & Chadwick, K. 2004. Cultural Diversity in Management, Firm Performance, and the Moderating Role of Entrepreneurial Orientation Dimensions. **Academy of Management Journal**, 47(2): 255.

Robinson, W. S. 1950. Ecological Correlations and the Behavior of Individuals. **American**

Sociological Review, 15(3): 351.

Schachter, S., Ellertson, J., McBride, D., & Gregory, D. 1951. An experimental study of cohesiveness and productivity. **Human Relations**, 4: 229-238.

Simon, H. A. 1947 [1997]. **Administrative behavior: a study of decision-making processes in Administrative Organizations** (3rd ed. ed.). New York: Free Press ; London: Collier Macmillan.

Sommers, S. 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(4): 597-612.

Sørensen, J. B. 2002. The Strength of Corporate Culture and the Reliability of Firm Performance. **Administrative Science Quarterly**, 47(1): 70.

Sorenson, O., & Sørensen, J. B. 2001. Finding the right mix: franchising, organizational learning, and chain performance. **Strategic Management Journal**, 22(6/7): 713.

Starbuck, W. H. 1993. Keeping a Butterfly and an Elephant in a House of Cards: the Elements of Exceptional Success. **Journal of Management Studies**, 30(6): 885-921.

Stasser, G., & Titus, W. 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.

Sutton, R. I., & Callahan, A. L. 1987. The Stigma of Bankruptcy. **Academy of Management Journal**, 30(3): 405-436.

Tajfel, H. 1982. **Social identity and intergroup relations**. Cambridge: Cambridge University Press.

Taleb, N. 2007. **The black swan : the impact of the highly improbable** (1st ed.). New York: Random House.

Taylor, A., & Greve, H. R. 2006. Superman or the Fantastic Four? Knowledge Recombination and Experience in Innovative Teams. **Academy of Management Journal**, 49(4): 723-740.

Thompson, J. D. 1967. **Organizations in Action; Social Science Bases of Administrative Theory**. New York: McGraw-Hill.

Tsetlin, I., Gaba, A., & Winkler, R. L. 2004. Strategic Choice of Variability in Multiround Contests and Contests with Handicaps. **Journal of Risk & Uncertainty**, 29(2): 143-158.

Tversky, A., & Kahneman, D. 1981. The framing of decisions and the psychology of choice. **Science**, 211: 453-458.

Uzzi, B., & Spiro, J. 2005. Collaboration and Creativity: The Small World Problem. **American Journal of Sociology**, 111(2): 447.

Van Der Vegt, G. S., & Bunderson, J. S. 2005. Learning and performance in multidisciplinary teams: the importance of collective team identification. **Academy of Management Journal**, 48(3): 532-547.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. 2004. Work Group Diversity and Group Performance: An Integrative Model and Research Agenda. **Journal of Applied Psychology**, 89(6): 1008-1022.

van Knippenberg, D., & Schippers, M. C. 2007. Work Group Diversity. **Annual Review of Psychology**, 58(1): 515-541.

Vaughan, D. 1997. The Trickle-Down Effect: policy decisions, risky work and the challenger tragedy, **California Management Review**, Vol. 39: 80: California Management Review.

Webber, S. S., & Donahue, L. M. 2001. Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis. **Journal of Management**, 27(2): 141.

Weick, K. E. 1990. The vulnerable system: An analysis of the Tenerife air disaster. **Journal of Management**, 16(3): 571-593.

Weick, K. E. 1993. The collapse of sensemaking in organizations: The Mann Gulch disaster. **Administrative Science Quarterly**, 38(4): 628.

Weick, K. E., & Sutcliffe, K. M. 2001. **Managing the unexpected: assuring high performance in an age of complexity** (1st ed.). San Francisco: Jossey-Bass.

Westphal, J. D., & Bednar, M. K. 2005. Pluralistic Ignorance in Corporate Boards and Firms' Strategic Persistence in Response to Low Firm Performance. **Administrative Science Quarterly**, 50(2): 262-298.

Williams, K. Y., & O'Reilly, C. A., III. 1998. Demography and Diversity in Organizations: A Review of 40 Years of Research. In B. M. Staw & L. L. Cummings (Eds.), **Research in Organizational Behavior**, Vol. 20: 77-140. London: JAI Press Inc.

Zenger, T. R. 1992. Why Do Employers Only Reward Extreme Performance? Examining the Relationships among Performance. Pay, and Turnover. **Administrative Science Quarterly**, 37(2): 198-220.

Zucker, L. G. 1977. The role of institutionalization in cultural persistence. **American Sociological Review**, 42(5): 726-743.