Getting disoriented by performance variability: a peek through the looking-glass separating average and extreme events

Fabrice Cavarretta

ESSEC Business School
Management Department
Avenue Bernard Hirsch B.P. 50105
95021 Cergy-Pontoise Cedex
FRANCE

INSEAD - OB Department
Boulevard de Constance
F-77305 Fontainebleau
Fabrice.Cavarretta@insead.edu
Tel. : +33 6 09 59 46 74

Abstract

The exploration-exploitation study (March, 1991) suggested that changes in performance variability may imply that more extremely high outcomes occur simultaneously with decreasing expected outcomes. That mean-variance tradeoff has wide theoretical consequences where ignoring variability effects can lead to wrong conclusions. I review the difficulties that limited focus on performance variability in organizational scholarship. In a constructive manner, I propose a generic approach to address it and illustrate it methodologically. This study contributes to the scholarship of extreme organizational outcomes by providing rationale to distinguish them from average outcomes.

Keywords : Performance, Risk, Methods

1. INTRODUCTION

Some of the most and least desirable outcomes for organizations are extreme. For example, large corporate fiascos as 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. 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 merits, one may wonder whether it is 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 those two outcomes are driven by different logic. Among Venture Capitalists (VCs), one encounters similar resistance when 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 intuitions find echoes in scholarship and appear most prominently in the exploration-exploitation study (March, 1991). It showed—using simulation—how changes in performance variability implies that increasing expected performance occurs simultaneously with decreasing extremely positive performance and therefore reduces survival. Few organizational studies explore the widespread theoretical consequences of this paradox (a notable exception appears in Denrell, 2003). Also, differentiating extreme outcomes from average outcomes has not become a common approach. To make things worse, classical methods have been designed to ignore or mute effects on variability by ways of "robust" statistical tools and removal of outliers. Overall, organizational researchers may have accumulated a body of prediction that is robust regarding average outcomes but could be systematically flawed concerning extreme outcomes.

Organizational scholarship could therefore benefit from a constructive approach to combining variability and average effects, if possible conceptualizing parsimoniously 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 whether such inversion occurs and how likely it is. Finally, one may be curious to identify specific organizational theories or contexts where such inversion matter 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, which I illustrate by a short simulated example. Third, I present the theoretical benefits of considering such approach for various perspectives.

2. LITERATURE REVIEW ON PERFORMANCE VARIABILITY

After a brief example to provide 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 account of variability appears in a few disjoint areas of organizational studies.

2.1. Why changes in variability matter: the intuition 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 L and H, which are each associated with a set stylized performance value: L leads to [0, 4, 8] and H to [4, 5, 6]. When considering the effect of the factor, classical organizational theory would only theorize a mean effect, and infer then that H is preferable because its expected value (5) is greater than the expected value for L (4).

This reasoning assumes the goal is to improve average outcome. By 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 example, if one seeks to reach at least 8, then L is more preferable, which is opposite to the conclusion than if one seeks to improve expected outcome. Typically, that factor has an effect on variability (negative from L to H) at the same time as an average effect (positive from L to H). When combined, those makes inferences change at a certain level of outcomes, intuitively somewhere between 5 and 6, i.e. H is preferable to reach any performance up to 5, but L is preferable to reach any performance above 6. This example shows that one cannot properly make inferences regarding extreme vs. average outcomes while neglecting effects on variability.

2.2. 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 posture, 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 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 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 average effect. This procedure, ingrained in the research community, is perfectly valid to estimate 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 takes 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 as 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, when studying the effects of culture on the reliability of firms' performances, 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.

2.3. Existing studies theorizing effects on variability

Variability is not absent from organizational studies and effects on that variable seem to occur in the field. 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), and that team experience diversity conditions “extreme outcomes” (Taylor & Greve, 2006). 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 no mechanism to balance mean effects against variability effects is proposed.

Reasoning on the consequences of variability also appear, more often in studies at the border with 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. That literature focuses on the relationship between the mean and variability of performance, which can be expressed symbolically as seeking a relationship between P and ΔP (Bromiley, Miller, & Rau, 2001). It differs in its objective from the current study that explores whether and how effects by 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 variability in outcomes plays a significant role in the evaluation of corporate opportunities. 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.

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 to 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, Denrell explored how, in the presence of variability effects, inference-making may be subject to a selection bias because of the disappearance of firms (2003).

However, the selection bias approach (Denrell, 2003) 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, such as IPOs, survive such outcomes and will be available when making inferences. Even extremely low outcomes do not imply disappearance. For instance, various firms file 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 variability because of the various contexts where "competition on extreme value" occurs. However, neither a general approach to deal with it nor the theoretical consequences of such neglect have been identified so far.

2.4. 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) claimed organizational studies should focus on exceptional organizational outcomes and praises the explicit analysis of outliers in an approach that requires reconsideration of assumptions about the distribution of outcomes. 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 normal distribution deserve attention such as fat-tail distributions (McKelvey & Andriani, 2005). 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 to model 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 distribution with fat tails. For example, effects on the volatility of financial assets (Black & Scholes, 1973) have been a cornerstone of financial theory for decades. In that field, even though suggestions to reconsider distributional assumptions appeared early (Mandelbrot, 1960), it has not yet produced even a fraction of the development attributed to theories studying effects on variability (Mandelbrot & Hudson, 2004). It would be awkward for organizational studies to move directly from (a) to (c) without first reaping all the theoretical benefits of considering 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 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.

3. WHEN VARIABILITY MAKES INFERENCES REGARDING EXTREME OUTCOMES THE INVERSE OF THOSE REGARDING 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 prediction 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.

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 neglect of performance variability has various theoretical consequences (next section) that can be avoided if properly taken into account (this section).

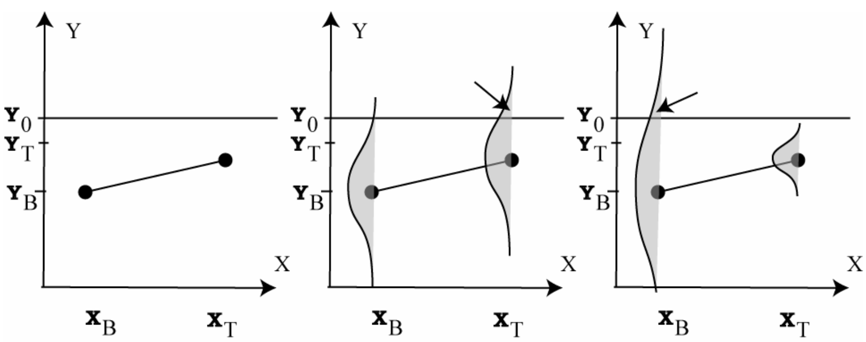
Regarding statistics and methodology, some details are provided in the appendix, but most of the reasoning is based on graphical arguments that only require basic statistical intuitions. On the issue of statistics, one should not confuse the objective of the current study with some estimation techniques that evoke thresholds on the dependent variable, for instance, when dealing with truncated data (Greene, 2003:chap 22), or discrete levels (Greene, 2003:chap 21). The issue here is not about such techniques to optimize estimation of first order effects but to identify the theoretical consequences of ignoring variability. For those interested in linking the approach to a statistical technique, the current study relates to quantile regressions (Greene, 2003:448).

3.1. Inference making is simple without changes in variability

I will now progressively introduce formal arguments showing 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 1). Imagine a factor X has the effects on a performance random variable Y (now assumed normal), as summarized by Figure 1(a): the average of Y is higher for the high values of the factor (X_H) than for its low values (X_L). 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 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 1(b)).

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 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.

Rule 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 at 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 X increases the chances to reach 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—as long as no variability effect exists.

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

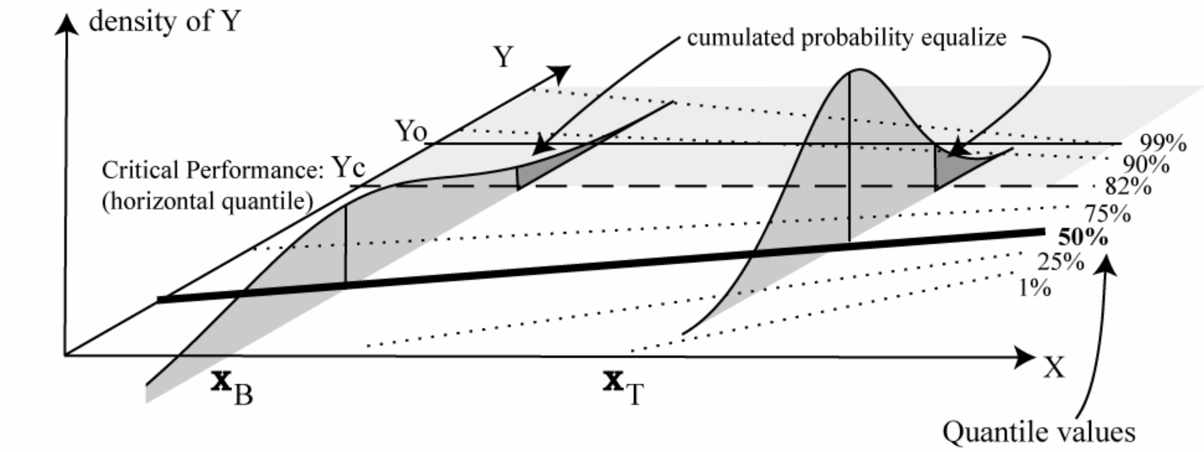
If performance variability changes, the generality of inferences does not hold anymore. Figure 1(c) matches Figure 1(a) for the mean effect but now assumes that X reduces the variability of Y, as represented by a bell curve that is more diffuse at X_L than at X_H. The probability of reaching the threshold is now higher at X_L (indicated by the arrow), even though the mean value is still higher at X_H. 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 fails to provide a simple mechanism of how inferences differ at various performance thresholds. No simple reasoning is available to balance 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 inferences change.

It still relies on a graphical argument, albeit slightly more sophisticated. 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 (actual reasoning in the appendix shows that it holds with more generality).

In Figure 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 roughly approximates the regression line. When accomplished for a few other values, it results in Figure 2—a more-detailed version of Figure 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 chance to reach the threshold.

Figure 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*.

Rule 2: The direction of an inference around a threshold of performance Y_0 correspond 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.

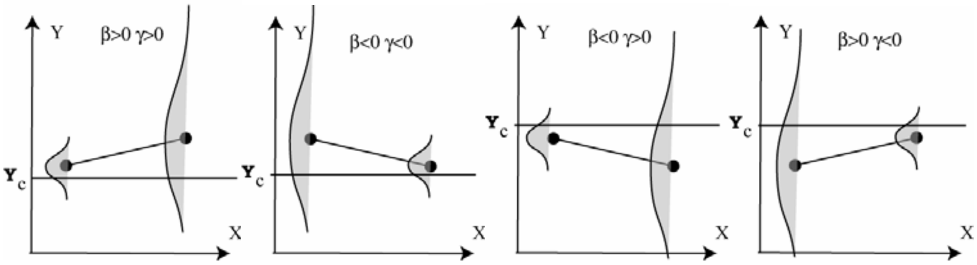
Under reasonable assumptions, the slopes of the quantile lines change direction once (see appendix for proof). In Figure 2, the quantile at 50% has a positive slope because X has a positive effect on the average of Y; when going up the diagram, the quantile line progressively loose slope because of the 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 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 decreases, so one would infer that *X* has a negative effect on the chances of reaching that threshold. This leads to the following rule:

Rule 3: Inference 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%, and it could be quite small if the mean effect is strong relative to the variability effect. At the extreme, if there is no effect on variability, the critical level disappears and all is driven by mean effect, as expressed in Rule 1. The position of the critical performance, relative to average performance, depends on both the mean and the variability effect (see Figure 3): 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.

Figure 3 Critical Performance Position

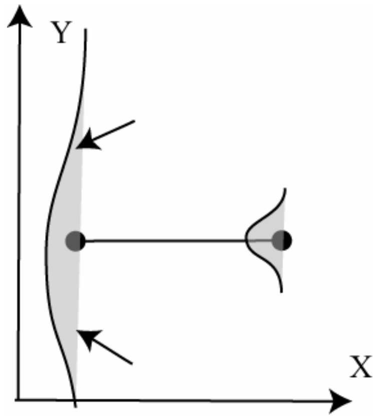


The mirroring of effects when crossing the critical performance level is the central mechanism justifying that one carefully takes into account variability and threshold when studying organizational outcomes. Overall, this approach provides a construct (threshold of performance) and a computable criterion (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 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. This leads to the final rule:

Rule 4: In the absence of average effect, variability effects allow inferences regarding both extremely low and high outcomes.

Figure 4 No mean effect with variability effect predicts both extreme low and high outcomes



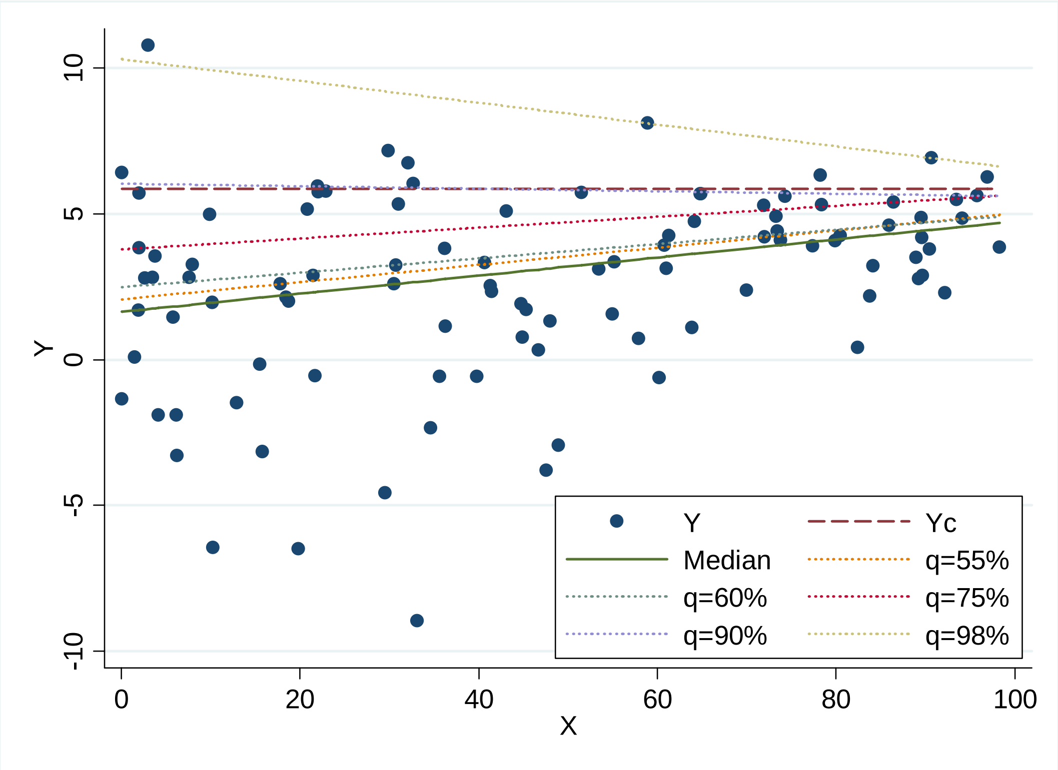
3.3. Illustration by a simulated example

A simulated example can illustrate the reasoning 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 (Equation 2) 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 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 5. The effects were made strong enough that the heteroskedasticity is visible, with a decrease in the dispersion of points around the regression line.

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



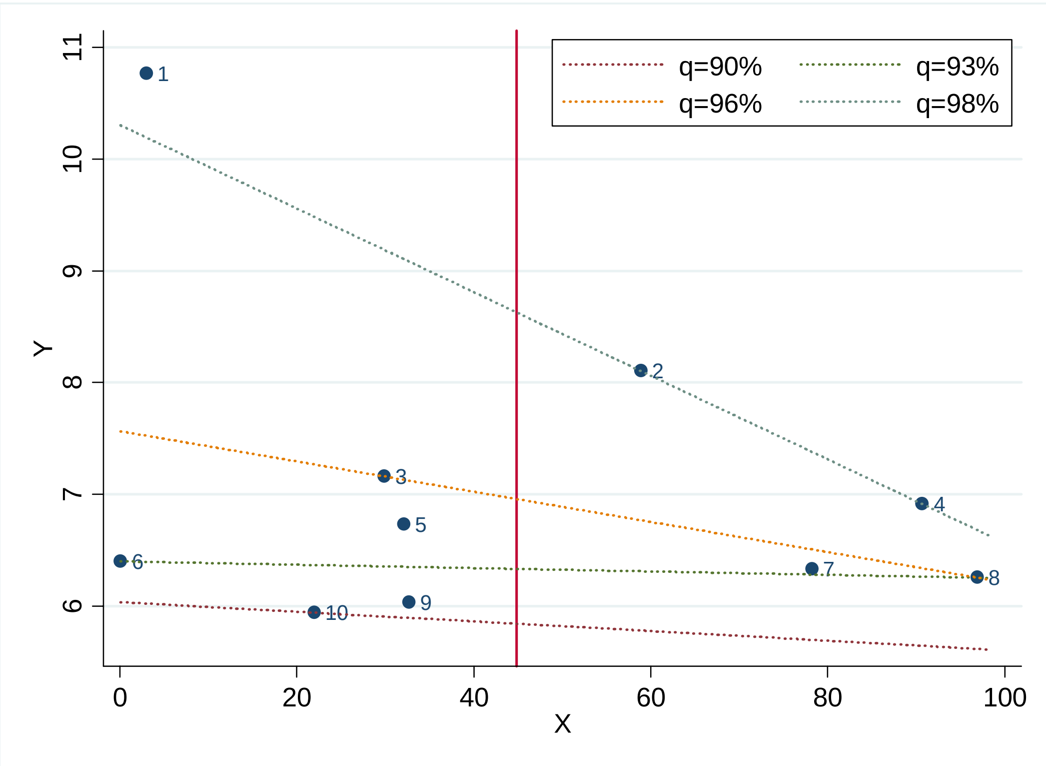
A first approach to determine the level of critical performance is to make successive approximations, plotting various quantile lines and figuring out the value at which the quantile line is horizontal. I computed and plotted in Figure 5 the quantile lines using the quantile regression Stata procedure (qreg), which suggests a value of the critical level around 6. In addition, I used the formal approach proposed in the appendix that takes as inputs the mean and variance effects. 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 4 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 can be determined by a graphical method as a first estimation (between 75% and 90% in Figure 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 Figure 2 was 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 6) illustrates the behavior beyond the critical performance level. I plotted only the range above the critical line (top 18%) and added a few additional quantile lines in that range. 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, I marked the points with their ranking in the performance range—number 1 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 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 on the basis of only 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[1]. More generally, if inferences are made by observing 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, if 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.

4. 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 to reach a given performance threshold and variability effects make that such inferences do not hold if considering another performance threshold. Typically, inferences regarding average outcomes apply to expected outcomes, but may be misdirected regarding extremely low or extremely high outcomes. Alternatively, inferences regarding extremely low or high outcomes may be misdirected regarding average outcomes. I will discuss below these four ideal cases, which will be summarized in summary Table 1.

In the study "Some Myths of Management", Denrell (2003) identified situations where managers get confused about the true effects of some factors due to sample bias because of bankruptcies and effects of variability. This section follows and extends this reasoning in two directions. First, it generalizes this logic to 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

[1] 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.

organizational scholars, whose research method or design determines the observation threshold, and for which the generality of predictions are endangered by variability effects. That being said, Denrell's study would fit in the fourth case (B.2. in Table 1) where one observes extremely low outcomes and be confused regarding average outcomes.

4.1. 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 based on averages which turn out to be misdirected regarding extremely high outcomes. 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 1). Applying a classical analysis (Mohr, 1982) to that context, one would find that the factor is beneficial. By contrast, taking into account effects on variability 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 IPO stage; the crowd of other entrants to that early market die after the market matures around the few that are properly funded by IPO. Inferences drawn from normal studies may therefore wrongly predict 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 chances to reach IPO.

By systematically clarifying the performance threshold (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 its conclusion on the nature of the outcome sought. This logic would apply to various other fields where exceptionally high outcome may be sought. For instance, given that the emergence of the Macintosh and iPod may be driven by variability effects more than accumulation of average effects, one may revisit innovation studies with a variability lens. This first ideal case is summarized in A.1. in Table 1.

4.2. Average-based inferences that do not apply to extremely low outcomes

The second ideal case occurs when one makes inferences based 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 1). Applying classical analysis, one would find that the factor is beneficial. By contrast, taking into account effects on variability 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, therefore signaling the existence of a low threshold of performance to avoid. In such contexts, relying on inferences based on average outcomes may be problematic. In simple words, 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, allowing distinguishing, for instance, the improvement of firm financial ratios 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 organization (HRO). This case is summarized in B.1 in Table 1.

4.3. 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 and are therefore potentially not applicable to *average* outcomes. The situation is less common and clear than for the first two cases. Since a large share of organizational studies uses regression, all those studies tend to predict average outcomes and therefore the applicability to extremes outcomes is easily endangered. However, because there is no single method or source to the problem as was the case with the regression, it is difficult to claim that all studies focusing on extreme outcomes are automatically endangered.

Another difficulty is that studies observing extreme outcomes often do not have information on a full scale of outcomes where such extremes fall; often, one only has a binary variable identifying whether a performance threshold has been attained. For instance, if studying promotion to CEO, often one has only that binary dependent variable and does not have access to an underlying scale where CEO promotion occurs at a certain defined level. Therefore, one may have difficulty distinguishing extreme cases vs. averages since the design may—by definition—only consider a binary outcome. Yet, the current study may provide motivation to find any (proxy) scale of performance 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 truncated scale, to determine whether there is a critical level where inferences invert, therefore signaling that what is good about becoming a CEO may be bad for other managers.

Once those restrictions are taken into account, I show that studies modeling the attainment of a single extreme threshold are potentially endangered by variability effect, except if one has already clarified the effect of variability and considered effects at different performance levels. Let us examine the last two ideal cases that illustrate that problem.

In the third ideal case, one makes inferences based on extremely high outcomes which turn out to be misdirected regarding 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 1). By observing attainment of the high threshold, one would find that the factor is detrimental. However, that could be due to an effect on variability which is masking the fact that the factor is beneficial on average. This distinction matters to 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 average performance of managers but decreases variability, it is therefore possible that it decreases the chances to become CEOs, 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). More generally, the issues apply to population ecology studies where very few survive, as was demonstrated in the exploration/exploitation study (March, 1991:chapter 3).

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).

4.4. Inferences about extremely low outcomes that do not apply to averages

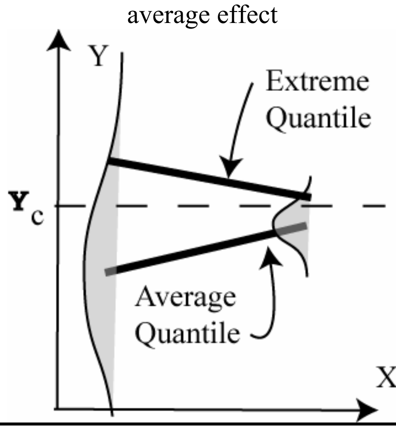
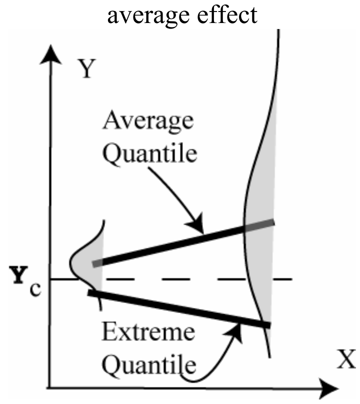
In the fourth ideal case, one makes inferences based on extremely *low* outcomes which turn out to be misdirected regarding average outcomes. This could occur when average and variability effects are in the same direction, for instance when a factor has a positive average effect and a positive variability effect (case B.2. in Table 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 which is masking the fact that the factor is beneficial on average. This distinction matters to all theoretical perspectives where average outcomes differ from extremely low outcomes.

For instance, population ecology relies on observing survival or death of organizations. In some contexts, death is a low outcome, i.e. it is rare enough 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 by Denrell in his study on undersampling of failure (Denrell, 2003): 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 suggests that various possible thresholds be considered, typically that a study predicting survival at a low threshold also consider effects on average. 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 when using a low selection threshold—may benefit taking into account effects of variability and the contingence of inferences to the level of performance observed.

Table 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, and both the average quantile (i.e. the regression line) and an 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: X is detrimental	On extremely low outcomes:
	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)

5. 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 inferences depend on the level of outcome 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 all perspectives studying 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 counter to what is predicted by averages effects. Such focus on variability could occur as free-standing 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, future research may find that various dimensions influence performance variability, finally providing an organizational dimension to the concept of organizational risk.

An alternative path of research may explore whether such reasoning would create myopias 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.

In particular, a qualitative approach may bring particular benefits to such program. One may explore the similarities between the extremely successful CEOs of large traded firms (currently, e.g., Steve Jobs) and the extremely unsuccessful ones (e.g., Jeff Skillings of Enron). Such qualitative approach may allow identifying if some of those characteristics which are similar in both extreme populations would appear different from those of the "average" manager (CEOs from a large traded firm whose performance are not exceptional in neither direction). For instance, one may find that some characteristics such as dishonstey or strong leadership style appear more often in CEOs with exceptional performance. A qualitative research would allow both to provide more richness to the approach advocated here as well as to exploit the intuition that the line separating success and failure is sometimes amazingly fine.

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

6. 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 $z_X(Y)=(Y-\mu_X) / \sigma_X$ and $\mu_X = \beta_1 X + \beta_0$ and $\sigma_X = \gamma_1 X + \gamma_0$.

This modeling accommodates many of the distributions used in organizational research. It generalizes the classic approach using a simple regression, $Y : N(\beta_1 X + \beta_0, \sigma_0)$, where N is the normal distribution and one assumes—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 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 2 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? For example, the exploration–exploitation study (March, 1991) explores the outcomes using a simulation but does not provide a constructive approach that is usable in other studies. Let us analytically explore the question of whether X improves the chance p(X) of reaching a threshold Y_0. The following reasoning builds on a logic exposed by Tsetlin, Gaba, and Winkler (2004). 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:

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

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 2. 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:

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

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.

7. REFERENCES

Argote, L. 1999. **Organisational Learning: Creating, Retaining and Transferring Knowledge**. Norwell, MA: Kluwer Academic Publishers.
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.
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.
Cabral, L. M. B. 2003. R&D Competition When Firms Choose Variance. **Journal of Economics & Management Strategy**, 12(1): 139-150.
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.
Denrell, J. 2003. Vicarious Learning, Undersampling of Failure, and the Myths of Management. **Organization Science**, 14(3): 227-243.
Fleming, L. 2004. Perfecting Cross-Pollination. **Harvard Business Review**, 82(9): 22-24.
Greene, W. H. 2003. **Econometric analysis** (5th ed.). Upper Saddle River, N.J.: Prentice Hall.
Hannan, M. T., & Freeman, J. 1984. Structural Inertia and Organizational Change. **American Sociological Review**, 49(2): 149-164.

Kahneman, D., & Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. **Econometrica**, 47: 263-291.
Khurana, R. 2002. The Curse of the Superstar CEO. **Harvard Business Review**, 80(9): 60.
Kogut, B. 1991. Joint Ventures and the Option to Expand and Acquire. **Management Science**, 37(1): 19-33.
Lieberson, S., & O'Connor, J. F. 1972. Leadership and Organizational Performance: A Study of Large Corporations. **American Sociological Review**, 37(2): 117-130.
Mandelbrot, B. 1960. The Pareto-Levy Law and the Distribution of Income. **International Economic Review**, 1(2): 79-106.
Mandelbrot, B. B., & Hudson, R. L. 2004. **The (Mis)Behavior of Markets: a Fractal View of Risk, Ruin, and Reward**. New York: Published by Basic Books.
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.
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.
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.
Pfeffer, J., & Salancik, G. R. 1978. **The External Control of Organizations: A Resource Dependence Perspective**. New York, NY: Harper and Row.
Sørensen, J. B. 2002. The Strength of Corporate Culture and the Reliability of Firm Performance. **Administrative Science Quarterly**, 47(1): 70.
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.
Sutton, R. I., & Callahan, A. L. 1987. The Stigma of Bankruptcy. **Academy of Management Journal**, 30(3): 405-436.
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.