DRMI (Defense Resources Management Institute)   Faculty and Researchers' Publications

2011

The Trouble With Risk Matrices

Wall, Kent D.

Monterey, California: Naval Postgraduate School.

https://hdl.handle.net/10945/32570

THE TROUBLE WITH RISK MATRICES

Kent D. Wall
Naval Postgraduate School (DRMI)

August 18, 2011

Abstract

Risk matrices are used in many organizations as a basis for risk management decisions. They are easy-to-use and have intuitive appeal. They are often defended as the only avenue of approach to risk management because quantitative information is "scarce or nonexistent". Unfortunately their theoretical basis is superficial and the validity of the qualitative information they employ is highly suspect. Assessments of the likelihood of occurrence and their impacts suffer all the shortcomings associated with subjective assessment. These all combine to produce a *less than useful* portayal of risk management information. *The true value of subject matter experts (SME) is in aiding the devlopemnt of quantitative models representing the inter-relations between variables that ultimately determine the outcome of interest for the decision maker.*

1 INTRODUCTION

Many organizations in public and private sector confront decision making in an environment fraught with significant uncertainty. As a result risk becomes an overriding concern and many decisions become ones of risk management. Risk management requires that: (1) risk is unambiguously defined; (2) the definition supports the measurement (quantification) of the risk; (3) the risk reduction produced by each prospective risk management alternative is quantifiable; and (4) the cost of each risk management alternative is known. Given this information the manager can then choose the "best" alternative (where best means that alternative that reduces the risk the most subject to the budget constraint). Often the information supporting the quantification or measurement of risk appears lacking or is hard to obtain. In such circumstances a "qualitative" approach is sought. Predominant among such approaches is the use of the "risk matrix".

The risk matrix is like a simple spreadsheet model composed of rows and columns. There usually are four or five rows and columns. The rows and columns define categories of likelihood (probability) and "impact" (severity). We assume throughout that the rows define the categories of probability (likelihood) while the columns define the categories of the outcome, often called the "impact", "severity" or "consequence". The cells within the matrix are assigned numbers called risk scores that purport to represent a quantitative assessment of the risk – the higher the score the higher the risk. Figure 1 displays a generic risk matrix.

	negligible	marginal	serious	catastrophic
almost certain				
highly likely				
likely				
unlikely				
almost never				

Figure 1: Example Risk Matrix

These matrices have intuitive appeal. Individuals easily can imagine risk as a function of the two dimensions portrayed by the matrix. For example, it seems self-evident that a highly likely event associated with catastrophic loss is riskier than an unlikely event associated with negligible loss. The cells of the matrix often are color-coded to communicate this (i.e., the cells in the upper right corner are red, the cells in the lower left corner are green, while the cell in between range from yellow to orange as we proceed along the diagonal from lower right to upper left). Green is "good" (none or very low risk), red is "bad" (very high risk), yellow and orange indicate increasing risk.

As a management tool the matrix offers a way to rank things like R&D projects in terms of their potential for cost overruns, schedule delays or substandard performance. Other examples of ranking objects by perceived risk are: military operations alternatives, national defense budgets, possible terrorist threats and deficiencies in accounting systems.

Risk matrices purport to serve risk management decisions. They are seen as useful tools by many practitioners. In reality their deficiencies render them less than useful. To understand why, we must first look behind the curtain and see how a risk matrix produces a ranking of risks. This is followed by a section reviewing the relevant theory, both prescriptive and descriptive, that forms the foundation of decision making under risk and risk management. Finally, we contrast the risk matrix assessment process with theoretical foundations using a example. This highlights the deficiency in the rationale behind the risk matrix. Our findings are summarized in the last section.

2   RISK ANALYSIS USING RISK MATRICES

The use of risk matrices to rank risks proceeds as follows. Each of the alternatives (projects, operational alternatives, budget allocations, etc.) become the objects of evaluation by subject matter experts (SME). The SME assign to each object a probability category and an impact category. This locates objects in the matrix using the two labels as row and column addresses. The result is a matrix "populated" with the alternatives. Numerical scales are assigned to the row and column categories and used to compute a risk score for each cell. Often the cells are color-coded in an attempt to highlight the meaning and interpretation of the scores. Figure 2 is an example from the U.S. Army (TRADOC 2009, Appendix I). Here the categories of likelihood are assigned "scores" from 1 (= almost never) through 5 (= almost certain), while the categories of consequence are assigned "scores" from 1 (= negligible) through 4 (= catastrophic). The reasons for this coding is not given. Sometimes the cell scores are obtained by "weighting" (e.g., consequence is weighted by 0.6 and likelihood by 0.4 before) multiplying the two. The justification for this is not explained. We find cells of the same color are not scored the same. For example, the red cells in the "Northeast" corner are not all equal. It appears that the risk corresponding to "almost certain and serious" is less than the risk associated with "highly likely and catastrophic". Evidently there is more to risk than what is communicated by color-coding.

The risk score calculation is described formally as follows. Let S_i denote the ith probability (row) category score and S_j denote the jth outcome (column) category score. Denote the *subjectively assessed* probability by $\widehat{p}$ and the *subjectively assessed* consequence by $\widehat{y}$. If the assessed probability is such that the SME declares it to fall within the ith row category, then this probability is denoted $\widehat{p}_i$ and is represented in subsequent calculations by its score value $S_i(\widehat{p}_i)$. If the assessed outcome is such that the SME declares it a member of the jth column category, this outcome is denoted $\widehat{y}_j$ and is represented in subsequent

	negligible	marginal	serious	catastrophic
almost certain	2.6 [Y]	3.2 [O]	3.8 [R]	4.4 [R]
highly likely	2.2 [G]	2.8 [Y]	3.4 [O]	4.0 [R]
likely	1.8 [G]	2.4 [Y]	3.0 [O]	3.6 [O]
unlikely	1.4 [G]	2.0 [G]	2.6 [Y]	3.2 [O]
almost never	1.0 [G]	1.6 [G]	2.2 [G]	2.8 [Y]

Figure 2: Example Risk Matrix with Color Code and Risk Score

calculations by its score value $S_j(\widehat{y}_j)$. The risk score assigned to the $(i, j)^{th}$ cell within the risk matrix is now represented by:

$$R_{ij} = R(S_i(\widehat{p}_i), S_j(\widehat{y}_j)). \tag{1}$$

R_{ij} exhibits monotonic behavior in its arguments:

$$R_{ij} \leq R_{i+1,j} \text{ and } R_{ij} \leq R_{i,j+1}.$$

There is no guidance as to the form of $R(S_i(\widehat{p}_i), S_j(\widehat{y}_j))$. Its definition appears to depend upon who is using the risk matrix. Most often we find R_{ij} computed as the product of its two arguments:

$$R_{ij} = S_i(\widehat{p}_i) \cdot S_j(\widehat{y}_j) \tag{2}$$

or a modification of this as in the U.S. Army TRADOC example in Figure 2:

$$R_{ij} = [0.4 \times S_i(\widehat{p}_i)] \cdot [0.6 \times S_j(\widehat{y}_j)].$$

3 SOME THEORY OF CHOICE UNDER RISK

The rationale for risk matrices apparently derives from the quantitative definition of risk proposed by Kaplan and Garrick (1981) They state that ".... risk consists of the answer to three questions: (1) What can go wrong?; (2) How likely is it to go wrong?; and (3) If it does go wrong, what is the outcome?" These authors represent the answers symbolically as a triple:

$$\{s_i, p_i, y_i\} \tag{3}$$

4

where s_i is the name of the scenario (event) representing "what can go wrong", p_i is the probability of occurrence (likelihood) of s_i and y_i is the measure (magnitude) of the outcome (damage) associated with the occurrence of s_i. Thus it appears as self-evident that risk is a function of two quantities: p_i and y_i. It is determined by the probability and the associated outcome of a scenario. A matrix representation is a natural way to describe this: let the rows represent the p_i and the columns represent the y_i. Thus Figure 1 appears to capture all that is relevant. Right?

Wrong? There is more to risk than just likelihood and outcome. In fact, this is implicitly acknowledged by those who use risk matrices, for they feel *compelled to add more information*. First, there is universal use of color-coding of the cells to convey something about the amount of risk to associate with a cell: red cells represent more risk than orange cells, orange cells represent more risk than yellow cells and green cells represent less risk than yellow cells. Second, there is an apparent need for more detailed information than the colors provide. Indeed, we find the cells contain a number representing either the *risk score* or the *risk rank*. These numbers are used to order the risks, even among cells with the same color. Evidently, all cells of the same color are not created equal! The use of colors and risk scoring are attempts to provide missing information – risk assessment requires more than just probability, p_i, and outcome, y_i.

The information omitted from the Kaplan and Garrick triple is *decision maker preference*[1]. To make a decision (e.g., NO, there is no significant risk; or YES, there is significant risk and something must be done to manage it), the decision maker must be able to express preference over the set of probable outcomes. That is, the decision maker must be able to express how he/she *feels* about the outcomes, $\{y_i; 1 \leq i \leq n\}$. Is there a range, or specific values, of y_i that the decision maker *prefers to avoid?* Is there a range, or particular values, of y_i that the decision maker prefers to have happen? Decision theory describes decision maker preference by a *payoff function* (sometimes called a value function). We denote this function by $v(y_i)$.

The Kaplan and Garrick "trio" needs to be expanded to a "quartet". Risk is really a combination of the answers to *four* questions: (1) "What can go wrong?"; (2) "How likely is it to go wrong?"; (3) "If it does go wrong, what is the outcome?"; and "How do you feel about it (the likely outcomes)?" The Kaplan and Garrick "trio" must be

---

[1] Kaplan and Garrick omit any reference to decision maker preferences because their primary focus is provivision of information necessary for *public* evaluation of risk. They leave it to the public through the democratic *political process* to establish preferences.

replaced by a "quartet":

$$\{s_i, p_i, y_i, v(y_i)\}. \tag{4}$$

The importance of decision maker preferences requires a brief review of a well-developed and widely accepted model of decision maker preference. This model will serve to highlight a major source of error in risk matrices: If the color-coding and scoring are not determined by the decision maker, then they do not represent risk and are meaningless!

3.1   Models of Preference

Decision maker preference in risky decision problems is represented by a payoff function, $v(Y)$. Its development has a long history beginning with Daniel Bernoulli (1738, published in translated form in 1954). His model of preference was later called a utility function by John von Neuman and Oscar Morgenstern (1944). Their work established the utility function as the gold-standard for the payoff function until the mid-1970's. Utility theory always had troubling deficiencies when it came to descriptive power – it is unable to explain observed decision behavior in important situations. This led Daniel Kahneman and Amos Tversky (1979) to develop the prospect function as a better payoff function for representing preferences in risky decisions. This is now the new gold-standard for decision theory (although its embrace by the economics community proceeds slowly).

The prospect function version of the decision maker's payoff has three salient characteristics: (i) a reference point for the outcome that separates gains from losses; (ii) loss aversion in which losses loom larger than gains; and (iii) risk aversion for gains and risk taking for losses. Much empirical research finds a power function form adequate to represent the payoff function:

$$v(Y) = v^+(Y) + v^-(Y) \tag{5}$$

where

$$
\begin{aligned}
v^+(Y) &= [Y - Y^R]^\beta \text{ if } Y - Y^R > 0 \\
v^-(Y) &= -\lambda |Y - Y^R|^\gamma \text{ if } Y - Y^R < 0
\end{aligned}
$$

when more Y is preferred to less Y. If preferences are reversed so less Y is preferred to more Y then we have

$$
\begin{aligned}
v^+(Y) &= -\lambda [Y - Y^R]^\gamma \text{ if } Y - Y^R > 0 \\
v^-(Y) &= |Y - Y^R|^\beta \text{ if } Y - Y^R < 0.
\end{aligned}
$$

Y^R denotes the reference point that distinguishes the Y representing gain from the Y representing loss. Loss aversion is represented by

the constant, $\lambda > 1$. Decreasing marginal returns for both losses and gains are captured by the powers $0 \le \beta < 1$ and $0 \le \gamma < 1$. The prospect functions is convex for losses an concave for gains. Empirical estimates, in terms of median values, are:

$$2 < \lambda < 5 \tag{6}$$

and

$$0.5 \le \beta \text{ and } \gamma \le 0.8 \tag{7}$$

for a large number of studies with a wide range of subjects. Figure 3 depicts a payoff prospect function in the case where more Y is preferred to less when $Y^R = 3.0$, $\beta = \gamma = 0.70$.
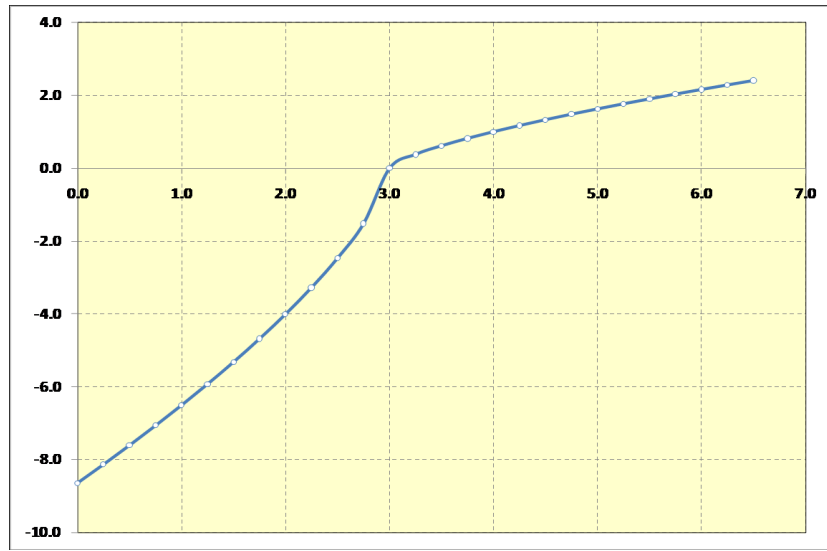

Figure 3: Payoff Prospect when "More is Better"

3.2 Models of Choice under Risk

Normative decision theory represents the choices made by a decision maker using a mathematical optimization construct: The alternative chosen is that which maximizes the decision maker's payoff function, $v(Y)$. In decisions involving uncertainty the theory replaces the payoff function with the expected payoff, $E\{v(Y)\}$ where the expectation is taken with respect to the distribution function of Y, P_Y. If $\{A_m, 1 \le m \le M\}$ denotes the set of alternatives, and $P_{Y|A_m}$ denotes the distribution function of Y conditioned on the alternative, then the decision problem is

$$\max_{A_m} E\{v(Y)|A_m\} = \max_{A_m} \int_{-\infty}^{\infty} v(Y)dF_{Y|A_m} \tag{8}$$

if Y is a continous-valued outcome and

$$\max_{A_m} E\{v(Y)|A_m\} = \max_{A_m} \sum_{k=1}^{k=N} v(y_k) \cdot p(Y = y_k|A_m) \qquad (9)$$

if Y is discrete-valued with N distinct values [2]. The solution is denoted as A^*. Since $v(Y)$ is negative for outcomes the decision maker prefers to avoid, the maximization operation means that A^* will be the risk minimizing alternative (i.e., the least negative among the set of alternatives).

This model is useful both normatively and descriptively. In the normative mode it informs the decision maker by inducing an ordering over the alternatives, from "best" to "worst". Descriptively it provides an explanation of why and how decisions are made. No matter what the mode of use, the model tells us what information the decision maker requires to make a rational informed decision: (i) the outcomes as a function of the alternatives, $Y(A_m)$; (ii) the conditional likelihood over the outcomes, $P_{Y|A_m}$; and (iii) the preferences (payoff function) of the decision maker over the outcomes, $v(Y)$.

3.3 Theoretical Implications for Risk Matrices in Practice

Risk matrices *cannot integrate all that matters using only two variables*: one defining the rows (probability or likelihood) and one defining the columns (outcome, consequence, impact or severity). A third quantity is required: payoff (preference). This describes how the decision maker values the probable outcomes. Risk cannot be quantified/measured/described until we know how much the decision maker desires to avoid the various outcomes representing loss. Thus, the need to add a third dimension to what is in reality a two-dimensional object. Risk matrices in practice attempt this by introducing color-coding and/or numerical-coding (scoring) to the cells.

Risk matrices facilitate decision making only if they faithfully depict the information the theory demonstrates is necessary for informed decisions. Risk matrices must accomplish this in situations where quantitative data appears unavailable, nonexistent or too difficult and costly to obtain. This means that risk matrix construction relies on subjective assessment. Subjective assessment is, however, fraught with problems.

4 AN ILUSTRATIVE EXAMPLE

We now contrast the qualitative risk matrix approach with the quantitative approach based on the theory of decision making under risk.

---

[2] Eq. (9) also applies when Y is continuous-valued but we use a relative frequency histogram approximation to the conditional probability ditribution of Y.

An illustrative example serves our purposes.

Suppose you want to know the risk inherent in doing nothing (the *status quo*) to mitigate infrastructure damage due to earthquakes. This is the first step in risk management: assessing the risk inherent in the "do nothing" alternative. If the answer is "NO", there is only insignificant risk, then it can be ignored. If the answer is "YES", there is significant risk, then you will be interested in assessing the risk associated with risk management alternatives. The process of deciding which of the proposals is "best" involves the same process of assessment as the original assessment, so we only consider this. Suppose further that the outcome of concern is the total program cost (implementation plus damage repair) and the time horizon is the next 20 years.

4.1 The Quantitative Approach

Given access to cost analysts, actuaries, earthquake engineers, simulation modelers and data you develop a simulation model that represents this knowledge. The uncertainty derives from the number, and strength, of earthquakes occurring over the next 20 years. Geologists say the number of annual earthquakes is modeled by a Poisson process with mean given by geological data. They say the size of the next earthquake (Richter scale) is modeled by an exponential distribution with mean given by geological data. The earthquake engineers and insurance industry provide a model that explains damages as a function of Richter scale value. You develop a spreadsheet model, use "add-in" simulation software, and produce a relative frequency histogram of total program cost (approximating the distribution, $P_{Y|A_0}$). This is depicted in Figure 4. This provides a *complete* picture of the
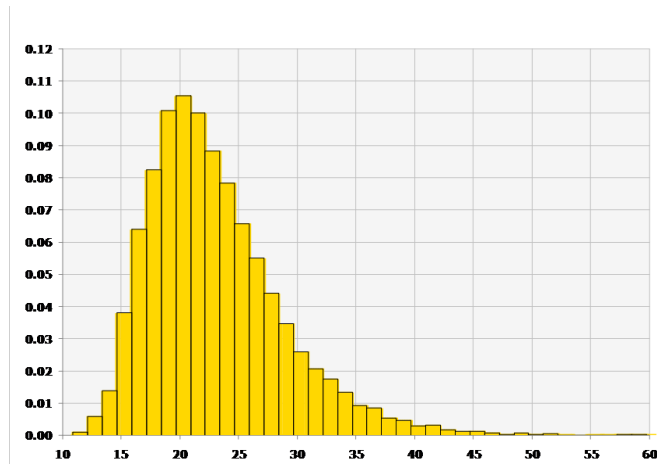

Figure 4: Example Total Cost Relative Frequency

uncertain situation confronting the decision maker. All that remains is to determine how the decision maker "feels about it".

We know from theory that this is captured by the computations in Eq. (9). The decision maker evaluates P_Y in the light of his/her preferences, $v(Y)$. The decision made depends on the value of $E\{v(Y)\}$: if $E\{v(Y)\} > 0$ there is an expected gain and insignificant risk; if $E\{v(Y)\} < 0$ then there is an expected loss and significant risk.

To illustrate, assume the decision maker has preferences over Y that give rise to the payoff function of Fig. 5 (here $\lambda = 5, \beta = 0.8, \gamma = 0.7$ and $Y^R = 20$.). A decision maker with this payoff function finds there
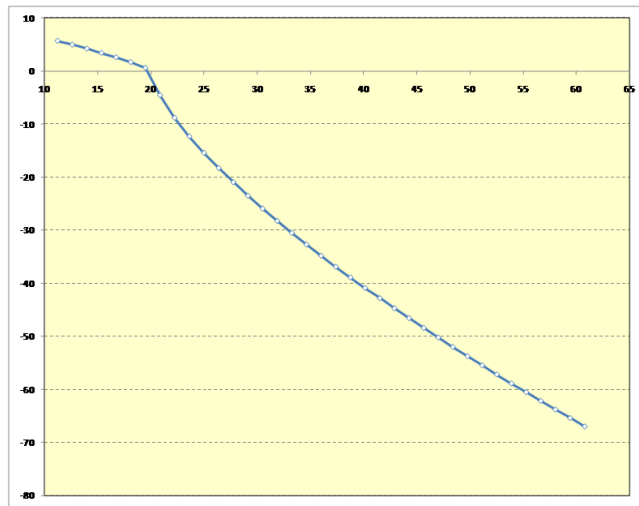

Figure 5: Decision Maker $v(y)$ for Earthquakes

is expectation of a loss; i.e., $E\{v(Y)\} = -8.431$. This person will conclude there is significant risk and seek management alternatives to reduce the risk. The dependence on Y^R is important. If $Y^R > 27.772$ then $E\{v(Y)\} > 0$ and the decision maker will conclude there is no significant risk. Of course, the actual decision is a function of all three parameters of the payoff function model.

The quantitative approach provides the decision maker the information required to make an informed decision. There is no more complete a characterization of the uncertainty in the outcome of interest, Y, than that provided by P_Y. Given this information (cf., Fig.4) the decision maker can determine its implication in terms of his/her payoff function for Y; i.e., $v(Y)$. Without knowledge of P_Y there is no hard evidence upon which to base a decision.

4.2 The Qualitative Approach

If you have no access to data and simulation modeling you rely on a qualitative approach using a risk matrix. You seek direct subjective assessment of likelihood and outcome by subject matter experts (SME), some or all of whom would be used by the quantitative approach in model construction. Let's assume you decide on four categories for this outcome: negligible (N), marginal (M), serious (S) and catastrophic (C). You also decide to use five categories for likelihood: almost never (AN); unlikely (U), likely (L), highly likely (HL) and almost certain (AC). Together, these define a risk matrix with four columns and five rows as in Figure 1 reproduced here.

Figure 6: Earthquake Risk Matrix

Once the risk matrix is defined the qualitative approach requires subjective assessment of the situation by the SME. This involves three steps.

4.2.1 Subjective assessment of probability and outcome are unreliable

First, and foremost, the SME must subjectively assess the existing risk in terms of the two dimensions of the matrix: the estimate of the likely outcome, $\widehat{y}_j$ and the estimate of its probability of occurrence, $\widehat{p}_i$. Unfortunately, all humans suffer substantial limitations in judgment under uncertainty (Kahneman, Slovic and Tversky, 1982). Three factors significantly affect the accuracy of subjective assessment: the representativeness heuristic; the availability heuristic; and anchoring with insufficient adjustment. Each lead to subjective assessments that exhibit bias and error. These limitations guarantee systematic errors in $\widehat{p}_i$ and $\widehat{y}_i$. Appendix A provides a brief summary of the details.

4.2.2 Interpretation of category labels is unreliable

Second, there are problems with interpretation of the subjective assessments in terms of the categories of the risk matrix. To illustrate, let the outcome categories be $\mathbb{Y}_j = \{Y_j^l \leq y < Y_j^u\}$ and probability categories be, $\mathbb{P}_i = \{P_i^l \leq p < P_i^u\}$. Clearly these are incomplete without specifying the upper and lower limits (signified by the superscripts l and u). Unfortunately there is no guidance how to do this. It is sometimes left to the SME for their definition. Other times definitions are provided by organizational doctrine, policy, or an analyst other than the SME.

4.2.3 Combining unreliable information compounds the problem

The last step requires the SME to reconcile the outputs of the first two steps. The SME must find a "home" for the assessed risk by placing it in one of the cells of the risk matrix. There are at least two possibilities here.

If no explicit (quantitative) definitions exist for the categories, then it is left to the SME to interpret the words defining them. There are significant difficulties in the interpretation, especially with respect to probabilities. Research finds that individuals attach very different meanings to the words defining the probability categories. Individuals interpret probabilities in the approximate range 0.0 - 0.05 as if they are indistinguishable from $P = 0$, while probabilities in the approximate range 0.92 - 0.99 as if they are indistinguishable from $P = 1$. Individuals over-weight the importance of probability when $0.05 < P < 0.4$ while they under-weight probabilities when $0.4 < P < 0.9$. This can result in an unlikely or highly unlikely risk receiving a likely classification, or a likely or highly likely risk receiving an almost certain classification. These difficulties also arise in the inverse. Given a numerical probability, different individuals will assign it to different rows in the matrix without explicit knowledge of the probability category definitions. On the other hand, given a probability identified by a verbal label, such as "highly likely", different individuals will identify it with different numerical probabilities and insert it into different rows. Again there is no guidance on how to do this. The same can be said for the outcome columns. Even when categories are explicitly defined in quantitative terms, individuals interpret these ranges using different words. Appendix B presents research findings that illustrate these problems.

Even if explicit category definitions exist for the SME, problems still arise. We illustrate using three different partitions for each outcome category:

- N: $\mathbb{Y}_1 = \{0 \leq y < 5\}; \{0 \leq y < 10\}; \{0 \leq y < 5\}$

- M: $\mathbb{Y}_2 = \{5 \leq y < 20\}; \{10 \leq y < 15\}; \{5 \leq y < 10\}$

- S: $\mathbb{Y}_3 = \{20 \leq y < 40\}; \{15 \leq y < 25\}; \{10 \leq y < 20\}$

- C: $\mathbb{Y}_4 = \{40 \leq y\}; \{25 \leq y\}; \{20 \leq y\}$

and three different partitions for each probability category:

- AN: $\mathbb{P}_1 = \{p \leq .10\}; \{p \leq .05\}; \{p \leq .01\}$

- U: $\mathbb{P}_2 = \{.10 < p \leq .33\}; \{.05 < p \leq .25\}; \{.01 < p \leq .20\}$

- L: $\mathbb{P}_3 = \{.33 < p \leq .90\}; \{.25 \leq p \leq .80\}; \{.20 \leq p \leq .70\}$

- HL: $\mathbb{P}_4 = \{.90 \leq p \leq .95\}; \{.80 \leq p \leq .95\}; \{.70 \leq p \leq .90\}$

- AC: $\mathbb{P}_5 = \{.99 \leq p\}; \{.95 \leq p\}; \{.90 \leq p\}$

Let us assume the SME *are in possession of the probability information* given in Figure 4. Even with all this information a single cell location is not possible. For example, denote by "X" the risk classification using the first partition of the $\{\mathbb{P}_i\}$ and $\{\mathbb{Y}_j\}$. Likewise, denote the classification corresponding to the second partition by "Y". Finally, denote the classification associated with the third partition "Z". None of the three category groups produce unambiguous risk classifications. as pictured in Figure 7. There is no unambiguous position

	Negligible	Marginal	Serious	Catastrophic
almost certain				
highly likely				
likely			**XYZ**	**YZ**
unlikely		**X**		
almost never				

Figure 7: Risk Matrix Alternative Locations

for "X" because $P\{Y_2\} = 0.33$ while $P\{Y_3\} = 0.66$. There is no unambiguous location for "Y" because $P\{Y_4\} = 0.66$ while $P\{Y_5\} = 0.30$. There is no unambiguous location for "Z" because $P\{Y_4\} = 0.33$

while $P\{Y_5\} = 0.66$. Even when the complete probabilistic picture is available to the SME subjectively judge must still be applied to force the situation into the matrix framework. The introduces error and may depend upon a proclivity to be pessimistic or optimistic when classifying the risk.

4.2.4 Risk scores are not informative

The final piece of information presented in the risk matrix is the risk score of Eq. (1). This score supposedly provides the decision maker a way to order the risks in the matrix from highest to lowest score. Unfortunately, the calculated score is based on subjective judgment and suffers from the same errors in judgment we find in the assessment of $\widehat{p}_i$ and $\widehat{y}_i$.

Developing a risk score requires specifying two components. First, there is the definition of the category scales that convert categories into numerical values. Second, there is the form of the function S_{ij}. Again there is no guidance provided to help decide what to do.

The current example requires five numerical values for the probability categories $\{S_i; 1 \leq i \leq 5\}$ and four numbers for the outcome categories, $\{S_j; 1 \leq j \leq 4\}$. The most common scaling uses a direct linear mapping:
$$S_i = i \text{ and } S_j = j.$$

This appears self-evident to most users of the qualitative approach. There are however, important implications for the eventual risk score calculation that may be unintended. For example, the outcome "serious" is three times more important than "negligible" and "catastrophic" is 1.25 more important than "serious". *There is no guarantee that the implied trade-offs correspond to the preferences of the decision maker.* In reality, the decision maker may consider the "catastrophic" outcome to be 5 times more important than the "serious" outcome. Similar observations apply to the probability category scaling. Appendix B presents more detail on the errors introduced by all category scaling.

The next step in computing a risk score is specification of S_{ij}. The only requirement is that it be monotonically increasing in each factor. If no other information is provided, then the simple product of the scales defined in Eq. (2) is used:
$$S_{ij} = S_i \cdot S_j = i \cdot j.$$

In particular we have $S_{ij} = S_i \cdot j$ which implies risk neutrality with respect to the outcome. This *contradicts the fact that decision makers are not risk neutral.* Using what appears to be an intuitively

attractive risk scoring function requires one to ignore the contradiction. The risk score computed this way has little to do with decision maker preferences and cannot be used to rank risks.

If the more general case of Eq.(1) is employed, the qualitative approach offers no guidance as to the preferred form of this function. The theory of decision making under risk offers guidance in terms of p_i and y_j but *not in terms of S_i and S_j*.

4.2.5   Summary of the qualitative approach

Every step of the qualitative approach involves subjective judgment. The need to *force all the information* contained in Figure 5 into a *single cell* of the matrix requires this. Unfortunately, all subjective judgment is imbued with bias and heuristics that limit its efficacy.

- Subjective judgment is required to extract from Figure 5 *one* representative $\widehat{y}_j$ and its corresponding $\widehat{p}_i$. We find this leads to non-unique classification.

- Subjective judgment is required in deciding the definition (partitioning) of each category. Different individuals use different partitions, even when explicit definitions exist.

- Subjective judgment is required in deciding the scaling of the categories, S_i and S_j. Scaling is arbitrary and ignores decision maker preferences.

- Subjective judgment is required in deciding on the form of the risk scoring function, S_{ij}. The most popular form, multiplication of S_i and S_j, combined with the most popular form of scaling, the direct linear map from categories to the positive integers, produces an inappropriate risk score.

5   COMPUTED SCORES DO NOT REPRESENT RISK

Users of risk matrices claim risk scores provide the information needed to rank risks. This claim has little or no foundation. Both the theory of decision making, and research results describing actual decisions, yield models that do not support the risk scoring in risk matrices. This, above all else, makes the information value of risk scores highly suspect. Let us see why.

Normative decision theory defines risk as the expected value of the

decision maker's payoff [3]:

$$E\{v(Y)\} = \sum_{k=1}^{k=N} p(Y = y_k) \cdot v(y_k) = R. \qquad (10)$$

(cf., Eq.9 with reference to the decision alternative, A_m, suppressed for notational convenience)[4]. Comparison of Eq.2 and Eq.10 shows some similarity: (i) both involve probability values; (ii) both involve the outcome values; and (iii) both involve combining this information via multiplication. Indeed, it may be argued that Eq.10 motivates the form of Eq.2. The differences are, however, significant. Both factors in Eq.2 are of dubious quality and content. The scales used to represent the $\widehat{p}_i$ category and the $\widehat{y}_j$ category of are arbitrary. Multiplying the two only compounds the errors. Unless $S_i(\widehat{p}_i) = \widehat{p}_i$ and $S_j(\widehat{y}_j) = v(\widehat{y}_j)$ we obtain meaningless results. The multiplication bears no resemblance to the expected pay-off of the decision maker. The $S_j(\widehat{y}_j)$ factor does not represent decision maker preferences and does not serve as a proxy for the decision maker pay-off function.

6 SUMMARY

Risk matrices have intuitive appeal and seem easy-to-use. They offer an approach to risk management that appears to avoid the need for quantitative information. As a result they have found widespread use. Unfortunately their true value is little more than a "placebo effect" (cf. Chpt. 7 in Hubbard, 2009).

First and foremost, they rely on a flawed process of subjective assessment. Any information obtained this way is fraught with bias and systematic errors. The use of SME does not eliminate this problem. SME are humans and are no less insulated from these problems than any other individual.

Second, the construction of the matrix itself is arbitrary. The number of rows and columns is arbitrary. Risk matrices can exhibit anywhere between three and six categories of outcome and probability. Increasing the number of categories does not improve accuracy or precision.

The definition of the categories is arbitrary. Category labels for probability and outcome use terms that vary from one application to the next. More importantly, the interpretation of the meaning of these labels is left entirely to the individual, whether SME or the decision

---

[3] Only the discrete form (c.f., Eq.9) of the normative theory is used to provide a straight forward comparison with Eq.10

[4] If the decision maker is *strictly loss averse* then $v(y_k)$ reduces to a simple switch where $v(y_k) = -1$ if y_k corresponds to a "loss" and is zero otherwise. In this case $E\{v(Y)\} = \sum p(Y = y_k) = R$ where the sum is only over those y_k that represent a loss.

maker. In fact, there is no assurance that the SME and decision maker employ identical interpretations of the categories. Experimental results show this creates errors understanding the information being communicated by the matrix. The term "high risk" that is given one combination of probability and outcome by an individual may be interpreted as "moderate risk" by another individual. These effects are exacerbated when SME attempt to force a veritable continuum of possibilities into a finite set of categories. Figure 7 illustrates this problem.

Third, category scaling arbitrary. Scales are necessary to convert categories to quantitative values for computing risk scores. No consideration, however, is given to the consequences of employing ordinal, interval or a ratio scales. Yet the choice of scale can have a substantial impact on resulting risk score and the decision as to how to allocate scare resources in an effort to manage risk. In reality we find scaling leads to range compression errors.

Fourth, and most important, the entire exercise of risk scoring is arbitrary. The theoretical foundations that should be employed in quantifying risk are ignored. The risk scoring functions that appear in the risk matrix literature are offered as a way to combine the numerically scaled values for probability and outcome but no basis is provided for their definition. They appear to be an effort to capture the belief that risk is a monotonic function of the severity (impact, magnitude or size) of the outcome of concern and the probability of its occurrence. Unfortunately, the most popular scoring function, the product of probability and outcome scales, is just plain wrong. Decision makers are risk averse and not risk neutral. More important is a total neglect of decision maker preferences in defining the risk scoring function. There is a veritable mountain of experimental results that shows that (decision maker) risk is determined by decision maker preferences over the likely values of the outcome. Moreover, the function that best expresses this risk is a prospect function. This function does not use the product of probability and outcome. It's form is dictated by four fundamental components: (a) a critical value by which the decision maker distinguishes those outcomes representing loss from those representing gain; (b) loss aversion by which losses are weighted more than gains (of the same absolute magnitude); (c) risk aversion for gains; and (d) risk taking behavior for losses. None of these components is contained in a function formed by the product of probability and outcome. All four components are determined by decision maker preferences, yet nowhere in the development of the risk matrix is there any attempt to solicit and incorporate decision maker preferences.

7 CONCLUSION

The principal reason for using the risk matrix is a perceived lack of quantitative information. Users claim it either does not exist or cannot be obtained without requiring significant time and resources. Apparently the time and resources expended to employ SME is viewed differently than those expended on quantitative analysis.

SME have valuable insights and experience that can and should be employed. Their expertise, however, is misapplied. It would be a far better use of SME experience and judgment if they are applied to the development of models describing the process that generates the outcome of concern. This includes the specification of variables, their interrelationships, ranges of likely values for parameters (confidence intervals), and information on possible probabilistic models of the uncertainties. Simulation models and the generation of data for a thorough quantitative risk analysis would follow in a matter of hours. It is hard to imagine the time required for this to be any more than the time required for SME to develop and complete a risk matrix!

8 EXAMPLES OF RISK MATRICES

The followiing section provides a small sample of risk matrices found in the public sector.

8.1 State of California

This risk management matrix uses risk ranking instead of risk scoring to order the risks in the matrix. A risk score (i.e., larger score = greater risk) is easily obtained. For example, replace the rank, r_{ij}, by $S_{ij} = 13 - r_{ij}$. The categories are explicitly defined, but no explanation or justification is given the cell ranking. One may ask what is the necessity of including the bottom row.

8.2 U.S. Navy

This risk matrix provides no explicit category definitions. The category scale is linear for both the row and the column. There is no explanation or justification for the risk scoring. The patterns of risk scores suggests a linear relationship over a subregion of the matrix.

8.3 Australia/New Zealand

This matrix does not have explicit definitions for categories and no risk scores. Risks can be ordered only by color (wording). The difficulties in using this matrix is illustrated by noting that an insignificant consequence is rated differently when it is almost certain *but*

	Neg. $0 - $20	Marg. $20-$40	Crit. $40-$80	Cata. $80-$100
0.7 – 1.0	9	5	2	1
0.4 - 0.7	11	8	4	3
0.0 – 0.4	12	10	7	6
0	0	0	0	0

Figure 8: Californina Department of Transportation Risk Management Matrix

	Low Impact 1	Mod. Impact 2	High Impact 3	Critical Impact 4
Very High Likelihood 4	4	5	6	6
High Likelihood 3	3	4	5	6
Moderate Likelihood 2	1	2	4	5
Low Likelihood 1	1	1	3	4

Figure 9: U.S. Navy Region SW Risk Management Process Matrix

still insignificant. If such a risk is insignificant then it must be so no matter how likely it is.

9   REFERENCES

1. Bernoulli, Daniel (1954), "Exposition of a New Theory on the Measurement of Risk" (original 1738), *Econometrica*, Vol.22, pp. 23-36.

2. Hubbard, D. W., *The Failure of Risk Management*, Hoboken, NJ: John Wiley & Sons, 2009.

3. Hubbard, D. W., *How to Measure Anything* (2nd), Hoboken, NJ: John Wiley & Sons, 2010.

4. Kahneman, D. and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Making under Risk", Econometrica,

LIKELIHOOD	CONSEQUENCES				
	Insignificant	Minor	Moderate	Major	Catastrophic
Almost certain	Medium	Medium	High	Extreme	Extreme
Likely	Low	Medium	High	High	Extreme
Possible	Low	Medium	Medium	High	High
Unlikely	Low	Low	Medium	Medium	Medium
Rare	Low	Low	Low	Low	Medium

Figure 10: Australia - New Zealand Standard for Risk Management

Vol.47, no. 2, pp. 263-291.

5. Kahneman, D., Paul Slovic and Amos Tversky (1982) *Judgment under uncertainty: Heuristics and biases*, Cambridge, UK: Cambridge University Press.

6. Kaplan, S. and B. John Garrick, "On the Quantitiative Definition of Risk", *Journal of Risk Analysis*, Vol.1, no.1, pp. 11-27.

7. Neumann John von and Oscar Morgenstern (1947), *The Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

8. Savage, Sam, *The Flaw of Averages*, Hoboken, NJ: John Wiley & Sons, 2009.

9. U. S. Army Training and Doctrine Command (TRADOC), "Capabilities Based Assessment (CBA) Guide Version 3.0, Sept. 2009.

A APPENDIX: SUBJECTIVE ASSESSMENT ERROR

Subjective assessment is subject to cognitive limitations that affect human ability to interpret and evaluate information. These produce errors and bias in the qualitative inputs to risk matrices. The most important of these, for our purposes, are those that affect the subjective assessment of: (1) the impact/severity/consequence; and (2) the likelihood/probability/frequency. Subject matter experts (SMEs) are no less insulated from these shortcomings than the man in the street – SME are human beings – enough said!

A.1 Representativeness Heuristic

Humans are susceptible to the "Representativeness" heuristic. Individuals are more apt to classify an object based on how representative the object is of a certain class or group of objects. This has serious implications for probability assessment. Questions like: "What is the

probability that object A belongs to class B?" "What is the probability that event A originates from process B?" "What is the probability that process B will generate event A?"are examples of where this heuristic leads to serious errors in assessment. Research has found six factors that give rise to the representativeness heuristic.

Ignoring the prior probabilities.   Individuals ignore facts related to the "base rate". In terms of conditional probability we know that $P(A \text{ and } B) = P(A|B) \cdot P(B)$. Research findings show that humans erroneously omit the prior probability or "base rate", $P(B)$, from this calculation.

Failure to account for sample size.   The variability in a sample statistic is a function of the sample size. The sample statistics (e.g., the mean value) computed from a small sample (e.g., 10 observations) will have more variability than the same sample statistic computed from a larger sample (e.g., 100 observations). Research shows that humans consistently fail to incorporate this fundamental result in probability statements.

Misconception of chance (effects of random behavior).   The classic example of this failure is the sequence of coin tosses: HTHTTH, HHHTTT, and HHHTH. Individuals consistently rate the first sequence as more indicative of a fair coin than the second sequence. Likewise the third sequence is rated as more indicative of a fair coin than the fourth sequence. Individuals infer the characteristics of a long sequence (global result) when judging the characteristics of a very short sequences. This erroneous inference is a demonstration of an effect called the "law of small numbers". Statistically sophisticated individuals commit this error more than one would think.

Insensitivity to predictability of the outcome.   Normative statistical theory maintains predictions are a function of prior information *and the expected predictive accuracy of the available evidence.* When predictability is nonexistent, prior probabilities dominate the process so all predictions should be very close to the base rate. If predictability is perfect, then the evidence dominates the process so predictions should be very close to actual outcomes. In reality, however, the ordering of the perceived likelihood of outcomes coincides with the ordering of outcomes by their representativeness.

Illusions of validity.   Subjects develop an unwarranted confidence in their predictions when there is a good fit between the predicted outcome and the input information. This confidence is the product of the representativeness of the input information and nothing else.

Failure to recognize the regression effect. Humans do not incorporate the regression effect or "regression toward the mean" when making subjective assessment . First, they do not expect this effect in many situations where it is bound to occur. Second, when they experience this effect, they invent spurious causal explanations for it. Failure to appreciate the regression effect leads to subjective assessment based on representativeness. This can produce pernicious effects. For example, experienced flight instructors have concluded that harsh criticism of students rather than verbal rewards is better in training, contrary to accepted psychological doctrine.

A.2   Availability Heuristic

Humans are susceptible to the "Availability" heuristic. Individuals assess the likelihood of an event by the ease with which instances or occurrences come to mind. This heuristic is useful in assessing likelihood by frequency but it is influenced by three factors that produce serious biases.

Retrievability. An event whose instances are easily retrieved from memory will appear more numerous than an event of equal frequency. Salience also effects retrievability. For example, the impact on frequency of actually witnessing an event will be greater than that from merely reading about it in the newspaper.

Imaginability. Sometimes one must evaluate the likelihood of an event not stored in memory. In such situations one usually constructs mental images of the event and evaluates the likelihood of the event by the ease with which the event can be constructed. Unfortunately, the ease with which an event can be constructed does not reflect their actual frequency. This bias is observed in many real-life situations.

Illusory correlation. Individuals significantly over-estimate the co-occurrence natural associates. Here the ease with which one associates the occurrence of two events seriously biases the likelihood estimate that these two events will co-occur in the future.

A.3   Anchoring and Adjustment

Many times individuals make estimates by starting from some initial value and then adjust it to obtain the final estimate. This is the "Anchoring and Adjustment" heuristic. The initial value may come from the problem formulation or it may be from some other partial computation. In either case, the adjustment is typically insufficient. There are three ways it fails.

Insufficient adjustment. Experiments reveal how significant is the anchoring effect on the final estimate of a quantity. Individuals given a low initial value are anchored so that their final estimate is consistently lower than the true value. Individuals given a high initial value produce final estimates consistently above the true value. Humans do not adjust sufficiently to offset the bias in the initial information.

Biases in the evaluation of conjunctive and disjunctive events. Studies show that individuals consistently underestimate the probabilities of disjunctive events and overestimate the probabilities of conjunctive events. This bias is the result of insufficient adjustment of prior probability estimation.

Anchoring in the assessment of subjective probability distributions. Both naive and sophisticated individuals fail to adequately adjust probability estimates from initially given values. This is revealed in many experiments where individuals consistently state overly narrow confidence intervals for the likelihood of some event. This is but one example of a trait exhibited by individuals when assessing random outcomes: overconfidence. Figure 11 presents an example of overconfidence in expert judgment. Seven experts (geotechnical engineers) were asked to estimate the failure height of an earthen wall. Each expert provided their estimate (blue square) along with its 25% and 50% percentiles (red triangle and red horizontal bar, respectively). The heavy black line depicts the actual failure value. Even SMEs are not well calibrated!
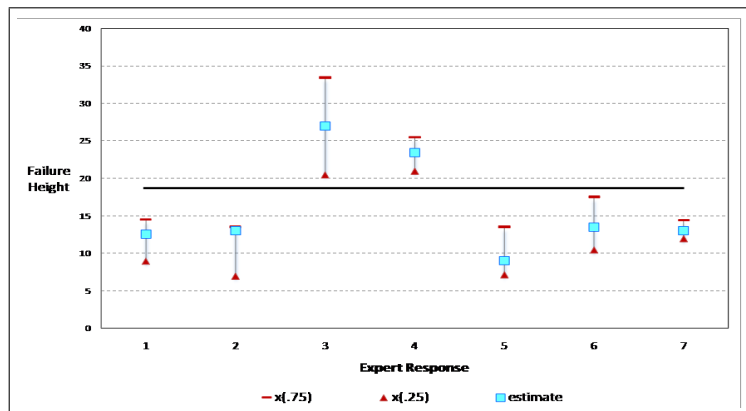

Figure 11: Expample of Expert Overconfidence

B APPENDIX: CATEGORY AND SCALE ERRORS

The use of categories for the row dimension and the column dimension introduce ambiguities that exacerbate errors. For example, individuals attach varying meanings to words like "high", "medium" and "low" in relation to assessing outcomes and probabilities. The use of more categories to produce a finer grid does nothing to alleviate problems and may even make them worse. The problems with the use of categories is very similar to the ambiguities of the χ^2- statistic in deciding goodness-of-fit: choosing different categories gives gives different values for the statistic. This can lead to the wrong decision.

Errors introduced by range compression. Forcing the conversion of an otherwise meaningful unambiguous quantity into a score with few (usually 3 to 5) values introduces error. Consider the Intergovernmental Panel on Climate Change (IPCC) example given below in Figure 12. The definition of the likelihood categories require both an 11% and 30% likelihood to be grouped together as "unlikely". Yet 11% corresponds to 1-in-9 odds while 30% represents almost 1-in-3 odds. Clearly this is a case of introducing "round-off" error. The same problems arise in relation to the outcome/impact axis of the risk matrix. For example, Hubbard (2009, pg. 131) describes an IT project risk matrix in which a return-on-investment (ROI) of between 1% and 299% is assigned a score of 1, while 300%<ROI<499% is assigned a score of 2. Thus a project with a ROI = 299% is only half as valuable as a project with a ROI = 300%.

Hubbard (2009, pg.132) also gives an example of "clustering". He shows how the five-point category assessment on the outcome (impact) scale for IT project evaluations in seven different companies produces a situation where the SMEs give an average score of 1 with probability = 0.01, an average score of 2 with probability = 0.12,an average score of 3 with probability = 0.30, an average score of 4 with probability = 0.48, .and an average score of 5 with probability = 0.09. Hubbard infers from this that SMEs suffer from bias: they tend to score more projects 4 and are biased away from the extremes of 1 or 5.

Errors with interpretation of the meaning of categories. The terminology used is ambiguous. An example from Hubbard (2009, pp.126-127) suffices. Research results from experiments by David Budescu at University of Illinois, Champagne-Urbana demonstrates how variable are individuals interpretation of terms used to define the likelihood of events. Figure 12 presents results from the experiments. The first column presents the phrases used to define category. The second column presents the official interpretation of the phrase ac-

cording to the guidelines of the Intergovernmental Panel on Climate Change (IPCC). The third and fourth columns summarize subjects responses and the fifth column presents the errors of the responses. It is most interesting that, when provided with the explicit official guidelines, subjects still gave answers at variance with the instructions!

Probability Phrase	IPCC Guidelines	Subject's Interpreted Meaning		Percent of Responses Violating Guidelines
		Minimum of Responses	Maximum of Responses	
Very likely	> 90%	43%	99%	58%
Likely	> 66%	45%	84%	46%
Unlikely	< 33%	8%	66%	43%
Very unlikely	< 10%	3%	76%	67%

Figure 12: Example of Interpretations of Probability

Problems with assigning numerical values (scales) to represent and quantify the categories. The arbitrary assignment of numerical scales is based on the presumption of regular or uniformly spaced intervals. Hubbard (2009, pp.133-134) presents the following example illustrating the errors introduced by this presumption. Figure 13 shows the relative value of the mitigating effect against IT project failure risk of the project sponsor rank in an organization. The scale assignment means that the influence of a CEO, CFO or CIO is three times as important as a VP. The influence of a VP is twice as important as that of a line manager. In reality, after acquiring historical data on the actual impact of sponsor rank on project success, we find the actual differences do not support the presumption of regular (uniformly spaced) intervals. While the rank ordering is preserved, the difference between C-level, senior VP and VP is quite small and that the influence of a manager is very much less that even the "ordinary" VP.

Problems with lack of independence between events. Scoring methods presume independence among factors and risks. They do not account for correlation between random variables or the effect of deterministic interdependencies. Two "medium-impact", "medium-likelihood" outcomes located in a risk matrix may represent a very high risk if they happen together (are correlated or interdependent). The only way such a situation can be taken into account is through the definition of a new outcome that represents the joint occurrence

	Scale Value	Scale	Actual
CEO, CFO, CIO	3	▬	☰
Senior VP	2	▬	
VP	1	▬	
Manager	0	▬	▬

Figure 13: Sponsor Rank and IT Sucess

of the two constituent outcomes. There is no evidence that this is re-definition of outcomes is practiced in the application of risk matrices.