

What's Wrong with Risk Matrices?

Louis Anthony (Tony) Cox, Jr.*

Risk matrices—tables mapping “frequency” and “severity” ratings to corresponding risk priority levels—are popular in applications as diverse as terrorism risk analysis, highway construction project management, office building risk analysis, climate change risk management, and enterprise risk management (ERM). National and international standards (e.g., Military Standard 882C and AS/NZS 4360:1999) have stimulated adoption of risk matrices by many organizations and risk consultants. However, little research rigorously validates their performance in actually improving risk management decisions. This article examines some mathematical properties of risk matrices and shows that they have the following limitations. (a) *Poor Resolution*. Typical risk matrices can correctly and unambiguously compare only a small fraction (e.g., less than 10%) of randomly selected pairs of hazards. They can assign identical ratings to quantitatively very different risks (“range compression”). (b) *Errors*. Risk matrices can mistakenly assign higher qualitative ratings to quantitatively smaller risks. For risks with negatively correlated frequencies and severities, they can be “worse than useless,” leading to worse-than-random decisions. (c) *Suboptimal Resource Allocation*. Effective allocation of resources to risk-reducing countermeasures cannot be based on the categories provided by risk matrices. (d) *Ambiguous Inputs and Outputs*. Categorizations of severity cannot be made objectively for uncertain consequences. Inputs to risk matrices (e.g., frequency and severity categorizations) and resulting outputs (i.e., risk ratings) require subjective interpretation, and different users may obtain opposite ratings of the same quantitative risks. These limitations suggest that risk matrices should be used with caution, and only with careful explanations of embedded judgments.

KEY WORDS: AS/NZS 4360; decision analysis; enterprise risk management; Military Standard 882C; qualitative risk assessment; risk matrix; semiquantitative risk assessment; worse-than-useless information

1. INTRODUCTION

A *risk matrix* is a table that has several categories of “probability,” “likelihood,” or “frequency” for its rows (or columns) and several categories of “severity,” “impact,” or “consequences” for its columns (or rows, respectively). It associates a recommended level of risk, urgency, priority, or management action with each row-column pair, that is, with each cell. Table I shows an example of a standard 5×5 risk matrix developed by the Federal Highway Administration for

assessing risks and setting priorities in addressing issues as diverse as unexpected geotechnical problems at bridge piers and unwillingness of landowners to sell land near critical road junctions.

The green, yellow, and red cells indicate low, medium, and high or urgent risk levels based on ratings of probability (vertical axis) and impact (horizontal axis) ranging from “VL” (very low) to “VH” (very high).

Table II shows a similar example of a 5×5 risk matrix from a 2007 Federal Aviation Administration (FAA) Advisory Circular (AC) introducing the concept of a safety management system for airport operators. The accompanying explanation states: “Hazards are ranked according to the severity and the likeli-

* Address correspondence to Louis Anthony (Tony) Cox; Cox Associates and University of Colorado, 503 Franklin St., Denver, CO 80218; tel: 303-388-1778; fax: 303-388-0609; tcoxdenver@aol.com.

Table I. Standard 5 × 5 Risk Matrix for Federal Highway Administration

Probability\ Impact	VL	L	M	H	VH
VH	Green	Yellow	Red	Red	Red
H	Green	Yellow	Red	Red	Red
M	Green	Green	Yellow	Red	Red
L	Green	Green	Yellow	Red	Red
VL	Green	Green	Green	Yellow	Red

Source: Federal Highway Administration, 2006
http://international.fhwa.dot.gov/riskassess/images/figure_12.htm.

hood of their risk, which is illustrated by where they fall on the risk matrix. Hazards with high risk receive higher priority for treatment and mitigation.” Many similar examples can be found for regulatory agencies, regulated industries, and public- and private-sector organizations. Training courses and software tools, such as MITRE’s *Risk Matrix* tool for program risk management (MITRE, 1999–2007) help to automate risk matrix creation, application, and documentation.

The use of such risk matrices to set priorities and guide resource allocations has also been recommended in national and international standards. It has spread through many areas of applied risk management consulting and practice, including enterprise risk management (ERM) and corporate governance (partly under the influence of the Sarbanes Oxley Act and international standards such as AUS/NZ 4360:1999); highway construction project risk management (Table I); airport safety (Table II); homeland security; and risk assessment of potential threats to office buildings, ranging from hurricanes to terrorist attacks (Renfroe & Smith, 2007).

Risk matrices have been widely praised and adopted as simple, effective approaches to risk management. They provide a clear framework for systematic review of individual risks and portfolios of risks; convenient documentation for the rationale of risk rankings and priority setting; relatively simple-appearing inputs and outputs, often with attractively colored grids; opportunities for many stakeholders to participate in customizing category definitions and action levels; and opportunities for consultants to train different parts of organizations on “risk culture” concepts at different levels of detail, from simply positioning different hazards within a predefined matrix to helping thought leaders try to define risk categories and express “risk appetite” preferences in the color coding of the cells. As many risk matrix practitioners and advocates have pointed out, constructing, using,

Table II. Example of a Predictive Risk Matrix for the Federal Aviation Administration

Severity \ Likelihood	No Safety Effect	Minor	Major	Hazardous	Catastrophic
Frequent					
Probable					
Remote					
Extremely Remote					
Extremely Improbable					

HIGH RISK
MEDIUM RISK
LOW RISK

Source: Federal Aviation Administration, 2007
www.faa.gov/airports_airtraffic/airports/resources/advisory_circulars/media/150-5200-37/150_5200_37.doc.

and socializing risk matrices within an organization requires no special expertise in quantitative risk assessment methods or data analysis.

Yet, despite these advantages and their wide acceptance and use, there has been very little rigorous empirical or theoretical study of how well risk matrices succeed in actually leading to improved risk management decisions. Very little prior technical literature specifically addresses logical and mathematical limitations of risk matrices (but see Cox *et al.*, 2005). Risk matrices are different enough from other topics (such as multivariate classification, clustering, and learning with correct classes provided as training data) to require separate investigation of their properties, in part because “risk” is not a measured attribute, but is derived from frequency and severity inputs through *a priori* specified formulas such as $Risk = Frequency \times Severity$. This article explores fundamental mathematical and logical limitations of risk matrices as sources of information for risk management decision making and priority setting.

2. A NORMATIVE DECISION-ANALYTIC FRAMEWORK

Many decisionmakers and consultants believe that, while risk matrices may be only rough

Table III. A 2 × 2 Risk Matrix

Consequence Probability	Low	High
High	Medium	High
Low	Low	Medium

approximate tools for risk analysis, they are very useful for distinguishing qualitatively between the most urgent and least urgent risks in many settings and are certainly much better than nothing, for example, than purely random decision making. This section examines these beliefs from the standpoint of optimal statistical decision making in a simple framework for which it is possible to obtain exact results.

The simplest possible risk matrix is a 2 × 2 table that results from dichotomizing each of the two axes, referred to here as “probability” and “consequence.” (Many other axes such as “frequency” and “severity” or “likelihood” and “magnitude” are also used, but changing the names does not affect the logic.) Table III shows such a matrix. Now, consider using it to categorize quantitative risks. For simplicity, suppose that the two attributes, *Probability* and *Consequence* have quantitative values between 0 and 1, inclusive (where 0 = minimal or zero adverse consequence and 1 = maximum adverse consequence). Define the quantitative risk for any (*Probability*, *Consequence*) pair to be their product, $Risk = Probability \times Consequence$, as advocated in many risk matrix methodology documents. The risk matrix designer can choose where to draw the boundaries between *low* and *high* values on each axis. Let the boundary between *low* and *high* consequence corresponds to a numerical value x between 0 and 1; and let the boundary between *low* and *high* probability correspond to a value y between 0 and 1.

To assess the performance of the risk matrix in supporting effective risk management decisions, consider the following specific decision problem. The decisionmaker must choose which of two risks, *A* and *B*, to eliminate. (She can only afford to eliminate one of them.) The quantitative values of *Probability* and *Consequence* are *a priori* independently and uniformly distributed between 0 and 1 for each of *A* and *B*. The only information that the decisionmaker has is knowledge of which cell of the risk matrix each risk falls in. (Thus, the risk matrix provides statistical information about the true but unknown quantitative risk; it is a lossy information channel.) *How well can the information provided by the risk matrix be used to*

identify the quantitatively greater risk? Equivalently, how well can the categorizations of quantitative risks provided by the matrix be used to identify the decision that maximizes expected utility (minimizes expected loss)?

The answer depends on how the risk matrix is designed and on the joint probability distribution of *Probability* and *Consequence* values. In general, the two risks can be ranked with no error if one risk falls in the *high* (red) cell in the upper right of Table I and the other falls in the *low* (green) cell in the lower left (since every risk in the *high* cell is quantitatively as well as qualitatively greater than any risk in the *low* cell). The probability of this event is $2 \times (1 - x)(1 - y)xy$. This symmetric function is maximized by choosing $x = y = 0.5$. (Otherwise, if the two risks have the same qualitative rating, then there is no way to choose among them based on the risk matrix, and we can assume that there is a 50-50 chance of making the right choice, that is, 50% error probability. If one of the two ratings is medium and the other is not, then the error probability from choosing the risk with the higher rating is positive, since some points in the cell with the higher qualitative rating have smaller quantitative risk values than some points in the cell with the lower qualitative rating; see Lemma 1 in the next section.)

The probability that two risks can be unambiguously ranked (i.e., with zero error probability) using the risk matrix with $x = y = 0.5$ is $(1/2) \times (1/4) = 0.125$ (i.e., it is the probability that one of them falls in one cell of the “*high/low*” diagonal and the other falls in the other cell of that diagonal). The probability that the two risks cannot be compared using the matrix with better than random accuracy (50% error probability) is the probability that both risks receive the same qualitative rating; this is $0.375 = (1/4) \times [(1/2) + (1/4) + (1/2) + (1/4)]$ (considering the four cells clockwise, starting with the upper left). The probability that the two risks can be compared using the matrix with error probability greater than zero but less than 50% is $1 - 0.125 - 0.375 = 0.5$.

Next, suppose that the risk matrix is constructed with $x = y = 0.5$, but that it is applied in decision settings where the joint probability distribution of *Probability* and *Consequence* is uncertain. Now, how well the matrix can identify which of two risks is greater depends completely on the joint probability distribution of (*Probability*, *Consequence*) pairs. For example, if *Probability* and *Consequence* values are uniformly distributed along the diagonal from (0, 0) to (1, 1), then there is a 50% probability that the two risks can

be classified with zero error probability (if one of them is in the *high* cell and the other is in the *low* cell); otherwise, the error probability is 50% (if both are in the same cell). Thus, under these very favorable conditions of perfect positive correlation, the error probability is $0.5 \times 0.5 = 0.25$. Conversely, if *Probability* and *Consequence* values are perfectly negatively correlated and are concentrated along the diagonal from (0, 1) to (1, 0), then all risks will be assigned a risk rating of “Medium” (although their numerical values range from 0 at the ends of the upper-left to lower-right diagonal to 0.25 in the middle), and the risk matrix will provide no useful information for discriminating between greater and lesser risks. Under these less favorable conditions, the decisionmaker using the risk matrix can do no better than random decision making, and the error probability increases to 50%.

Finally, if *Probability* and *Consequence* values are negatively correlated and concentrated along the line $Probability = 0.75 - Consequence$ (for *Consequence* values between 0 and 0.75), then all points on this line in the medium cells (i.e., for *Consequence* values between 0 and 0.25 or between 0.5 and 0.75) have *smaller* quantitative risks than any points in the *low* cell (i.e., for *Consequence* values between 0.25 and 0.5). For example, the pair (0.1, 0.65) would be classified as a medium risk (although its quantitative risk value is $0.1 \times 0.65 = 0.065$), while the pair (0.37, 0.38) would be classified as a *low* risk, even though its quantitative risk value is more than twice as great, $0.37 \times 0.38 = 0.14$. (More generally, such counterexamples can be constructed by noting that each iso-risk contour $Probability \times Consequence = constant$ is convex, so that a straight line passing through the two points where such a contour intersects the edges of a cell of the matrix will lie above the contour within the cell but below it outside the cell.)

For this unfavorable joint distribution of (*Probability*, *Consequence*) pairs, the information provided by the risk matrix is *worse than useless* (Cox & Popken, 2007) in the sense that, whenever it discriminates between two risks (by labeling one *medium* and the other *low*), it reverses the correct (quantitative) risk ranking by assigning the higher qualitative risk category to the quantitatively smaller risk. Thus, a decisionmaker who uses the risk matrix to make decisions would have a lower expected utility in this case than one who ignores the risk matrix information and makes decisions randomly, for example, by tossing a fair coin. (Similar examples can be constructed for the *high* risk cell in the upper right corner of Ta-

ble III. For example, the (*Probability*, *Consequence*) pair (0.6, 0.6) is rated as *high* and the pair (0.48, 1) is rated as *medium*, even though the latter has a higher quantitative risk (0.48) than the former (0.36).)

The question of how risk matrices ideally should be constructed to improve risk management decisions has no simple answer, both because risk matrices are typically used as only one component in informing eventual risk management decisions and also because their performance depends on the joint distribution of the two attributes, *Probability* and *Consequence*, as illustrated in the above examples. Since risk matrices are commonly used when quantitative data are limited or unavailable, this joint distribution is typically unknown or very uncertain. This knowledge gap implies that the actual performance of a risk matrix and whether it is helpful, no better than random, or worse than useless may be unknown. It also prevents easy application of traditional decision-analytic, statistical, artificial intelligence, and engineering methods for similar problems (e.g., for optimal classification and for discretization of multivariate relations) that require the joint distribution of the attributes as an input.

However, the simplest case of a 2×2 risk matrix does suggest two important related conclusions. First, it is not necessarily true that risk matrices provide qualitatively useful information for setting risk priorities and for identifying risks that are high enough to worry about and risks that are low enough to be neglected or postponed. (As just discussed, the information they provide can be worse than useless when probability and consequence are negatively correlated.) Second, use of a risk matrix to categorize risks is not always better than—or even as good as—purely random decision making. Thus, the common assumption that risk matrices, although imprecise, do some good in helping to focus attention on the most serious problems and in screening out less serious problems is not necessarily justified. Although risk matrices can indeed be very useful if probability and consequence values are positively correlated, they can be worse than useless when probability and consequence values are negatively correlated. Unfortunately, negative correlation may be common in practice, for example, when the risks of concern include a mix of low-probability, high-consequence and higher-probability, low-consequence events.

Although this section has been restricted to 2×2 risk matrices, the nature of the counterexamples in which the optimal statistical decision is to ignore risk matrix information (e.g., examples with joint

distributions of probability-consequence pairs concentrated on negatively sloped lines that intersect with convex iso-risk contours where they cross cell boundaries) implies that simply changing the position or number of grid lines cannot eliminate the problem. A similar construction can be carried out no matter how many cells a matrix has and no matter where the cell boundaries are located. Generalizing the decision problem to that of selecting a subset of risks to remediate, from among a larger set of many risks (rather than only deciding which of two risks is greater) also does not change the main conclusion. For some joint distributions of probability and consequence values, normative decision theory would require *not* using the qualitative risk rating information provided by a risk matrix, as it reverses the correct (quantitative) risk ratings that would be obtained using perfect information.

What can be salvaged? Several directions for advancing research on risk matrices appear promising. One is to consider applications in which there are sufficient data to draw some inferences about the statistical distribution of (*Probability*, *Consequence*) pairs. If data are sufficiently plentiful, then statistical and artificial intelligence tools such as classification trees (Chen *et al.*, 2006), rough sets (Dreiseitl *et al.*, 1999), and vector quantization (Lloyd *et al.*, 2007) can potentially be applied to help design risk matrices that give efficient or optimal (according to various criteria) discrete approximations to the quantitative distribution of risks. In such data-rich settings, it might be possible to use risk matrices when they are useful (e.g., if probability and consequence are strongly positively correlated) and to avoid them when they are not (e.g., if probability and consequence are strongly negatively correlated).

A different approach is to consider normative properties or axioms that risk matrix designers might ideally want their matrices to satisfy, and then to identify whether such matrices exist (and, if so, whether they are unique). This normative axiomatic approach, explored in the following section, can be used even when sufficient data are not available to estimate the joint distribution of probability and consequence values.

3. LOGICAL COMPATIBILITY OF RISK MATRICES WITH QUANTITATIVE RISKS

What does a risk matrix mean? One natural intuitive interpretation is that it provides a rough discrete (ordered categorical) approximation to a

more detailed—but not readily available—underlying quantitative relation. At least in principle, the underlying relation is described by a risk formula such as one of the following:

$$\begin{aligned} \text{Risk} = & \text{probability} \times \text{consequence (or frequency)} \\ & \times \text{severity or likelihood} \times \text{impact or threat} \\ & \times (\text{vulnerability} \times \text{consequence}), \text{ etc.} \end{aligned}$$

(We will use “frequency” or “probability” and “severity” or “consequence” as the default names of the two axes, and “risk” as the name for their product, but the analysis applies to any similar mathematical structure, regardless of the names.) For example, it might be supposed that the division of the probability axis into five ordered qualitative categories (e.g., from very rare to almost certain) corresponds roughly to a partitioning of a quantitative probability axis into the intervals [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), and [0.8, 1] (where square brackets indicate that the corresponding end point is included in an interval and parentheses indicate that it is not). Similarly, the five ordered categories for the severity axis might naturally be interpreted as corresponding to numerical intervals, [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), and [0.8, 1], on a quantitative value scale (e.g., a von Neumann-Morgenstern utility scale) normalized to run from 0 to 1, where 0 = no adverse impact, 1 = worst possible adverse outcome considered, and values between 0 and 1 represent adverse impacts or consequences with values intermediate between no adverse impact and worst possible adverse impact.

However, such an intuitive interpretation of the risk matrix as an approximation to an underlying quantitative model can only be sustained if the risk matrix satisfies certain constraints. To be most useful, a risk matrix should, at a minimum, discriminate reliably between very high and very low risks, so that it can be used as an effective screening tool to focus risk management attention and resources. This requirement can be expressed more formally as the following principle of *weak consistency* between the ordered categorization of risks provided by the matrix and the ranking of risks by an underlying quantitative formula, such as one of those above.

DEFINITION OF WEAK CONSISTENCY: *A risk matrix with more than one “color” (level of risk priority) for its cells satisfies weak consistency with a quantitative risk interpretation if points in its top risk category represent higher quantitative risks than points in its bottom category.*

Here, “quantitative risk” is defined as the product of a point’s coordinates when the axes are interpreted quantitatively, for example, $frequency \times severity$. If weak consistency holds, then all risks in the top qualitative category are quantitatively larger than all risks in the lowest qualitative category. In this case, the risk matrix can discriminate reliably between at least some risks, even though it does not require quantifying the probability and consequence attributes. It may then serve as a useful screening tool, which is one of the main practical uses of risk matrices. But if weak consistency does not hold, then risks that are screened out as being relatively small according to the matrix may in fact be larger than some of those that the matrix classifies as top priority, thus leading to a misallocation of risk management resources. It is therefore desirable to construct risk matrices that satisfy weak consistency, if possible.

Weak consistency is not an arbitrary axiom. It is implied by the hypothesis that *some* quantitative interpretation of the risk categories in a matrix exists, at least in principle (i.e., that there is some underlying quantitative risk scale such that the consecutive ordinal risk categories of the matrix correspond, at least approximately, to consecutive intervals on the quantitative scale), even if this scale is unknown, imprecise, or undefined in practice. If it does not hold, then a risk matrix does not mean what many users might expect it to mean, that is, that risks rated in the top category (red) are larger than those rated in the bottom category (green). Thus, transparency of interpretation provides another incentive for designing risk matrices to satisfy weak consistency.

3.1. Discussion of Weak Consistency

More generally, a risk matrix partitions alternatives (typically representing different threats, hazards, risk reduction or investment opportunities, risk management actions, etc.) into distinct categories corresponding to the different priority levels or “colors” of the matrix cells. Weak consistency implies that this partitioning assigns the highest qualitative level (e.g., red) to the alternatives that actually do have higher quantitative risk values than those assigned the lowest qualitative level (e.g., green). If weak consistency holds, the qualitative classification given by the matrix is, in this sense, at least roughly consistent with what a quantitative analysis would show. Red cells do represent unambiguously higher risks than green cells, where we use “red” to denote the highest urgency

Table IV. A 5×5 Matrix Compatible with $Risk = Probability \times Consequence$

Prob\Consequence	0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1
0.8–1	Green	Green	Yellow	Red	Red
0.6–0.8	Green	Green	Yellow	Yellow	Red
0.4–0.6	Green	Green	Green	Yellow	Yellow
0.2–0.4	Green	Green	Green	Green	Green
0–0.2	Green	Green	Green	Green	Green

level (that of the upper right-most cell, if the matrix axes are oriented to represent increasing probability or frequency on one axis and increasing severity of consequences on the other) and we use “green” to denote the lowest urgency level (that of the lowest left-most cell in such a table). This provides a logical basis for screening risks into “larger” (red) and “smaller” (green) categories.

Table IV shows an assignment of risk levels that satisfies weak consistency for a 5×5 matrix in which the rows and columns are interpreted as equal partitions of two numerical scales, each normalized to run from 0 to 1. Any point in a red cell has a quantitative value (calculated as the product of the horizontal and vertical coordinates) of at least 0.48, while no point in any green cell has a value greater than 0.40.

3.2. Logical Implications of Weak Consistency

Weak consistency is more restrictive than might be expected. For example, neither of the colorings in Tables I and II satisfies weak consistency. See Lemma 2.) Indeed, it implies some important constraints on possible colorings of risk matrices.

LEMMA 1. *If a risk matrix satisfies weak consistency, then no red cell can share an edge with a green cell.*

Proof: Suppose that, to the contrary, a red cell and a green cell do share an edge. The iso-risk contour (i.e., the locus of all frequency-severity combinations having the same value of the product $frequency \times severity$) passing through the midpoint of the common edge is a curve with negative slope. (It is a segment of a rectangular hyperbola, running from northwest to southeast.) Thus, it divides both cells into regions above and below this contour curve. Points that lie above this contour in the green cell have higher quantitative risk values than points lying below it in the red cell, contradicting weak consistency. Therefore, in a

risk matrix satisfying weak consistency, red and green cells cannot share an edge. QED

Comment: It is sufficient for this proof that iso-risk contours exist and have negative slopes. Thus, risk could be any smooth increasing function of frequency and severity (or whatever attributes the two axes of the matrix represent), not necessarily their product. However, the product of the coordinates is often used in practice in discussions of the concept of quantitative risk that accompany risk matrices, and we will use it as the default definition for quantitative risk in numerical examples.

LEMMA 2: *If a risk matrix satisfies weak consistency and has at least two colors (“green” in the lower left cell and “red” in the upper right cell, if axes are oriented to show increasing frequency and severity), then no red cell can occur in the left column or in the bottom row of the risk matrix.*

Proof: Contours for all sufficiently small risk values (namely, values of all risk contours below and to the left of the one passing through the upper right corner of the lower left-most cell) pass through all cells in the left-most column and in the bottom row of a risk matrix. If any of these cells is red, then all points below one of these contours in the red cell will have lower quantitative risk levels than points above it in the green lower left-most cell of the table. This would contradict weak consistency; thus, no such red cell can exist. QED

An implication of Lemmas 1 and 2 is that any risk matrix that satisfies weak consistency and that does not assign identical priorities to all cells must have at least three colors: for example, red for the upper right-most cell; green for the lower left-most cell; and at least one other color (i.e., priority rating), which we will call yellow, to separate red and green cells.

3.3. The Betweenness Axiom: Motivation and Implications

The hypothesis that a risk matrix provides an approximate qualitative representation of underlying quantitative risks also implies that arbitrarily small increases in frequency and severity should not create discontinuous jumps in risk categorization from lowest priority (“green”) to top priority (“red”) without going through any intermediate levels (“yellow”). (Notice that this condition is violated in Tables I–III, but holds in Table IV.) Indeed, if the successive risk categories in a risk matrix represent (at least ap-

proximately) successive intervals on some underlying quantitative risk scale, then continuously increasing quantitative risk from 0 to 1 should cause the corresponding qualitative rating to pass through increasingly severe categorical values. A weaker condition is that the qualitative risk should pass through at least one intermediate value between green and red as the quantitative risk increases continuously from 0 to 1. Otherwise, a risk matrix does not mean what users might intuitively expect: that intermediate risk categories describe risks between the highest (red) and lowest (green) ones. These considerations motivate the following axiom.

DEFINITION OF BETWEENNESS: *A risk matrix satisfies the axiom of betweenness if every positively sloped line segment that lies in a green cell at its lower (left) end and in a red cell at its upper (right) end passes through at least one intermediate cell (meaning one that is neither green nor red) between them.*

Comment: Tables I and II both have red cells in Row 2 and violate betweenness, that is, in each an arbitrarily small increase in frequency and severity can cause a risk to be reclassified as red instead of green, without going through yellow. A 2×2 table such as Table III lacks sufficient resolution to allow betweenness, since there are no cells between the green lower left cell and the red upper right cell. Thus, betweenness can only be required for 3×3 and larger risk matrices.

Only some risk matrices satisfy both weak consistency and betweenness. Among all 3×3 matrices having more than one color, only one coloring of the cells satisfies both axioms. Using our conventional coloring scheme (green for lowest risk, red for highest risk, yellow for intermediate risk), this is the matrix with red in the upper right cell, green throughout the left column and bottom row, and yellow in all other cells.

3.4. Consistent Coloring

The final normative axiom considered in this article is motivated by the idea that equal quantitative risks should ideally have the same qualitative risk rating (color). Although this condition is impossible to achieve exactly in a discrete risk matrix, for the reason shown in the proof of Lemma 1 (essentially, horizontal and vertical grid lines cannot reproduce negatively sloped iso-risk contours), one rough approximation might be to enforce it for at least the two most extreme risk categories, red and green, while accepting some inconsistencies for intermediate colors. Accordingly,

we will consider a requirement that all cells that contain red contours (meaning iso-risk contours that pass through other red cells) should themselves be red, unless the low resolution of the risk matrix causes them to also contain green contours. (A cell that contains both red and green contours has insufficient resolution to separate top-priority and bottom-priority risks and will not be required *a priori* to have either color.) Conversely, cells that contain green contours but no red ones should themselves be green. This motivates the following axiom of *consistent coloring*.

DEFINITION OF CONSISTENT COLORING. (1) A cell is red if it contains points with quantitative risks at least as high as those in other red cells (and does not contain points with quantitative risk as small as those in any green cell). (2) A cell is colored green if it contains some points with risks at least as small as those in other green cells (and does not contain points with quantitative risks as high as those in any red cell). (3) A cell is colored an intermediate color (neither red nor green) only if either (a) it lies between a red cell and a green cell; or (b) it contains points with quantitative risks higher than those in some red cells and also points with quantitative risks lower than those in some green cells.

Intuitively, one might think of an iso-risk contour as being colored green if it passes through one or more green cells but not through any red cells; as being colored red if it passes through one or more red cells but not through any green cells; and as being colored yellow (or some other intermediate color) if it passes through both red and green cells (or through neither red nor green cells). Then, the consistent coloring principle implies that any cell that contains green contours but no red contours must itself be green, while any cell that contains red contours but no green ones must itself be red. This is admittedly only one possibility for trying to capture the intuitive idea that all sufficiently high risks should have the same color (“red”) and all sufficiently low risks should have the same color (“green”). Other normative axioms could perhaps be formulated, but this article will only use the three already defined.

3.5. Implications of the Three Axioms

THEOREM 1: *In a risk matrix satisfying weak consistency, betweenness, and consistent coloring: (a) all cells in the left-most column and in the bottom row are green (lowest-priority); and (b) all the cells in the second col-*

umn from the left and in the second row from the bottom are nonred.

Proof: See the Appendix.

COROLLARY: *A 3×3 or a 4×4 risk matrix satisfying weak consistency, betweenness, and consistent coloring (and having more than one color) has a unique coloring, as follows. The left column and bottom row are green; the top right cell (for a 3×3 matrix) or the 4 top right cells (for a 4×4 matrix) are red; and all other cells are yellow.*

Proof: Theorem 1 implies that the left column and bottom row are green. Assuming that the upper right cell is red (since there is more than one color and this is the most severe cell), consistent coloring implies that the two cells in a 4×4 matrix that share edges with it must also be red and that the cell that both of these share edges with (diagonally below and to the left of the upper right cell) must also be red. Betweenness then implies that all other cells in a 3×3 or 4×4 matrix must be yellow. QED.

This result shows that it is possible to construct 3×3 and 4×4 matrices (although not 2×2 matrices) satisfying all three of the normative axioms proposed in this section. There is only one way to do so, however: any other colorings violate one or more of the axioms. For larger matrices, there is greater flexibility, as illustrated next.

3.5.1. Example: The Two Possible Colorings of a Standard 5×5 Risk Matrix

Table V shows two possible colorings of a 5×5 risk matrix that are consistent with the axioms of weak consistency, betweenness, and consistent coloring and also with a fully quantitative interpretation of the two axes, whose product gives a quantitative measure of risk (e.g., *risk* = *frequency* \times *severity*; *expected utility* = *success probability* \times *utility of success*; *reduction in perceived risk* = *perceived reduction in expected annual frequency of adverse events* \times *perceived average severity per event*; and so forth). The axes are normalized to run from (0, 0) at the lower left corner of the matrix to (1, 1) at the upper right corner, and the grid lines partition the axes into equal quantitative intervals.

In these tables, a “green contour” (with numerical value of 0.18) extends from the upper left cell to the lower right cell of the matrix (both of which are green, by Theorem 1), passing through a total of 9 cells. (All cells containing this contour are green, as

Table V. Two Possible Colorings of a Standard 5 × 5 Risk Matrix

	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
0.8-1	0.18, 1	0.21, 0.86	Yellow	Red	Red
0.6-0.8	Green	0.24, 0.75	Yellow	Yellow	Red
0.4-0.6	Green	0.36, 0.5	0.42, 0.42	Yellow	Yellow
0.2-0.4	Green	Green	0.5, 0.36	0.75, 0.24	0.86, 0.21
0-0.2	Green	Green	Green	Green	1, 0.18

	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
0.8-1	0.18, 1	0.21, 0.86	Green	Yellow	Red
0.6-0.8	Green	0.24, 0.75	Green	Yellow	Yellow
0.4-0.6	Green	0.36, 0.5	0.42, 0.42	Green	Green
0.2-0.4	Green	Green	0.5, 0.36	0.75, 0.24	0.86, 0.21
0-0.2	Green	Green	Green	Green	1, 0.18

are all cells below and to the left of it, by consistent coloring.) The upper right-most cell is defined to be red (top risk priority). The cell to its left and the cell below it each contain points with higher quantitative risks than those of points in this top priority cell's lower left corner; therefore, they must also be red (by consistent coloring) unless adjacent green cells make them yellow. The other yellow cells are implied by betweenness.

4. RISK MATRICES WITH TOO MANY COLORS GIVE SPURIOUS RESOLUTION

The foregoing analysis implies that, for a 5 × 5 risk matrix to be consistent with a fully quantitative interpretation as in Table IV, it must have exactly three colors. This is violated in many practical applications. For example, Table VI shows a default risk matrix used in some commercial risk management software tools designed to help support risk analysis standards and recommendations. Such a four-color matrix is inconsistent with the assumption that the colors represent relative sizes of underlying quantitative risks as in Table IV. For example, if the horizontal and vertical axes of Table VI are interpreted quantitatively as in Table IV, then Table VI assigns a higher rating

Table VI. Default 5 × 5 Risk Matrix Used in a Risk Management Software System

Likelihood\Consequence	Insignificant	Minor	Moderate	Major	Catastrophic
Almost certain	Blue	Orange	Red	Red	Red
Likely	Light green	Blue	Orange	Red	Red
Possible	Light green	Blue	Blue	Orange	Red
Unlikely	Green	Light green	Blue	Blue	Orange
Rare	Green	Green	Light green	Light green	Blue

Source: Adapted from www.incom.com.au/risk.asp?ID=471.

to (0.81, 0.21) than to (0.79, 0.39), even though the former has a product of 0.17 and the latter a product of 0.31.

4.1. Example: A 4 × 4 Matrix for Project Risk Analysis

The use of risk matrices for risk analysis of projects has been described as follows by the California Division of the Federal Highway Administration.

Risk is computed as the probability of occurrence multiplied by the consequence of the outcome. Probability is between 0 [minimal] and 1 [certain]. Consequence is expressed in terms of dollars, features, or schedule. Multiplying probability of occurrence and consequence [impact analysis] together gives a risk assessment value between 0 [no risk] and 1 [definite and catastrophic]. . . Below is an example of the matrix used for such an evaluation. The numbers are the order in which the risks are to be considered. Anything that is in the box labeled "1" is the highest priority.

	Likely 0.7-1.0	Probable 0.4 to 0.7	Improbable 0.0 to 0.4	Impossible 0
Catastrophic 0.9 to 1.0	1	3	6	
Critical 0.7 to 0.9	2	4	8	
Marginal 0.4 to 0.7	5	7	10	
Negligible 0 to 0.4	9	11	12	

Source: California Department of Transportation, 2007 www.fhwa.dot.gov/cadiv/segb/views/document/Sections/Section3/3.19.4.htm.

Table VII presents this risk matrix with its horizontal and vertical axes exchanged and oriented to be increasing, consistent with the conventions in previous examples.

The matrix has 13 priority levels as possible outputs, far greater than the three levels needed for a

Probability\Consequence	Negligible 0 to 0.4	Marginal 0.4 to 0.7	Critical 0.7 to 0.9	Catastrophic 0.9 to 1.0
Likely 0.7-1.0	9	5	2	1
Probable 0.4-0.7	11	8	4	3
Improbable 0.0- 0.4	12	10	7	6
Impossible 0	0	0	0	0

Table VII. Example Risk Matrix for Airport Projects

quantitative risk interpretation consistent with our axioms. The excess levels make it inconsistent with a coherent quantitative interpretation. For example, it assigns a priority rating of 8 to a quantitative risk of 0.42 (from a probability = 0.65 of a loss of relative severity 0.65 on a scale from 0 = no loss to 1 = worst catastrophic loss considered), but it assigns a much higher priority rating of 3 to a lower quantitative risk of 0.37 (probability = 0.41, consequence = 0.91). (Recall that output levels in the cells are numbered so that 1 = top priority.) Similarly, a loss of 0.6 with probability 1 receives a lower priority level than a quantitative loss of 0.8 with probability 0.5 (5 vs. 4), even though the former has a quantitative risk greater than the latter (0.6 vs. 0.4). A priority level of 12 is assigned to a probability 0.33 of consequence 0.33, but a priority level of 6 is assigned to a numerically identical risk consisting of a probability 0.11 of consequence 0.99. Thus, as expected, the priority ratings implied by the 13 distinct priority levels in this matrix do not successfully represent the relative sizes of these quantitative risks. (That the qualitative ratings reverse the quantitative ratings in such examples cannot be justified by risk aversion, since the consequence axis is explicitly assumed to have been already transformed, scaled, or defined in such a way that the product of the two coordinate axes, probability and consequence, is the measure of quantitative risk that the qualitative matrix attempts to represent.)

The upper left-most cell of the risk matrix in Table VII illustrates *range compression*: discrete categorization lumps together very dissimilar risks, such as an adverse consequence of severity 0 occurring with probability 1 and an adverse consequence of severity 0.39 occurring with probability 1.

The two possible 5 × 5 risk matrices in Table V have very limited resolution. They assign a green rating to all risks less than 0.24, and a red rating to all risks greater than 0.64 (on a scale normalized to run from 0 to 1). Attempts to use more colors or risk rating levels to improve resolution, as in the preceding example, necessarily create more ranking-reversal errors, in which quantitatively smaller risks are assigned

qualitatively higher rating levels than some quantitatively larger risks.

As a rough measure of the degree to which these limitations might affect practical work, suppose that the cases being classified by a risk matrix have their two components independently and uniformly distributed between 0 and 1. Then the probability that a randomly selected pair of points can be correctly and unambiguously rank-ordered by a matrix such as the one in Table IVa (i.e., the probability that one point falls in a red cell and the other in a green cell) would be only (3/25 red fraction) × (17/25 green fraction) = 8.2%. Thus, over 90% of the time, the matrix will not be able to rank-order the two points correctly with certainty.

5. RISK RATINGS DO NOT NECESSARILY SUPPORT GOOD RESOURCE ALLOCATION DECISIONS

How well can the information provided by a risk matrix guide risk management resource allocation decisions? This section examines some limitations that hold even if the risk matrix provides qualitative ratings that perfectly represent underlying quantitative risks.

5.1. Example: Priorities Based on Risk Matrices Violate Translation Invariance

Suppose that a risk manager can afford to eliminate all but one of the following three risks: (A) lose \$95 with certainty; (B) lose \$75 with certainty; (C) lose \$95 with probability 50% (else lose nothing). Which one should she keep to minimize risk (here defined as expected loss)? According to the priority ranking in Table VII (and interpreting the normalized consequence axis running from 0 to 1 as corresponding dollar losses running from \$0 to \$100), the answer is (C). (This has the lowest rating, 3, compared to ratings of 1 for A and 2 for B. Recall that in Table VII, lower numbers in the cell indicate higher priority.)

Now, suppose that all potential losses are reduced by \$15, so that the new alternatives are: (A') lose \$80 with certainty; (B') lose \$60 with certainty; (C') lose \$80 with probability 50% (else lose nothing). According to Table VII, one should now choose to keep (B') (rating = 5, compared to ratings of 2 and 4 for the A' and B', respectively). Thus, simply reducing the potential loss by the same amount for all three risks changes the prescribed priority ordering among them. This violates the principle of translation invariance for coherent risk measures (Artzner *et al.*, 1999). Moreover, keeping (B') instead of (C') is inconsistent with minimizing risk (defined as expected loss in this example). Thus, the risk matrix in Table VII does not necessarily support effective risk management decision making.

Similarly in Table VI, if a risk manager can eliminate exactly two out of four risks, corresponding to the four lower left-most cells in the table, and if ties are broken at random, then the probability that the risk in the second column and the bottom row will be eliminated is one-third (since the risk in the higher-rated cell to its northeast will certainly be selected, followed by any one of the remaining three tied risks). Translating all consequences one cell to the right (by adding the same incremental consequence value to each of them) increases the probability to one-half (since this alternative will now tie with one other for second place). But a second translation by one step to the right reduces the selection probability to zero (since now the two blue cells in the second row dominate the two cells in the first row). Finally, one more rightward shift of the four alternatives increases the probability that this one will be selected to one-half again.

In Table IV, if only one of four risks in the four upper left cells (e.g., with respective (probability, consequence) values of (0.9, 0.1), (0.9, 0.3), (0.7, 0.1), and (0.7, 0.3)) can be selected to eliminate, and if ties are broken at random, then the probability that the numerically greatest of these risks, namely, (0.9, 0.3), would be selected for elimination is only one-fourth. Translating all four consequences rightward by the same amount, 0.4, would increase this selection probability to 1. Translating them further rightward by an additional 0.2 would reduce the selection probability to one-third (since the three red cells would then be tied). Thus, the probability of assigning top priority to the numerically greatest risk does not satisfy translation invariance. (This same pattern also occurs for successive rightward translations of the four lower left-most cells in Table I.)

5.2. Example: Priority Ranking Does Not Necessarily Support Good Decisions

Setting: A risk manager has identified the following three risk reduction opportunities:

- Act A reduces risk from 100 to 80. It costs \$30.
- Act B reduces risk from 50 to 10. It costs \$40.
- Act C reduces risk from 25 to 0. It costs \$20.

(This example can also be constructed so that all three acts start from the same base level of risk, say 50, and A, B, and C reduce risk by 20, 40, and 25, respectively. Using different base levels allows for the possibility that the different options A, B, and C being compared protect different subpopulations.) The risk manager's goal is to purchase the largest possible total risk reduction for the available budget.

To assist risk-management decision making, suppose that a risk matrix is used to categorize opportunities A, B, and C. Resources will then be allocated first to the top-rated alternatives, working down the priority order provided by the risk matrix until no further opportunities can be funded.

Problem: How should a risk matrix categorize A, B, and C to support the goal of achieving the largest risk reduction from allocation of limited funds?

Solution: The answer depends on the budget. For a budget of \$40, the largest feasible risk reduction is achieved by funding B, so the best priority order puts B first. If the budget is \$50, then funding A and C achieves the greatest risk reduction, so B should be ranked last. At \$60, the best investment is to fund B and C, so now A should be ranked last. In short, *no categorization or rank-ordering of A, B, and C optimizes resource allocation independent of the budget.* No possible priority order (or partial order, if some ratings are tied) is optimal for budgets of both \$49 and \$50. This illustrates a limitation on the type of output information—ordered categorical classification—provided to decisionmakers by risk matrices. Such information is in general not sufficient to support effective allocation of risk-reducing resources because solutions to such resource allocation optimization problems cannot in general be expressed as priority lists or categories that should be funded from the top down until no further items can be afforded (Bertsimas & Nino-Mora, 1996).

Thus, the input information going into a risk matrix (ordinal ratings of event frequencies and severities) is simply not sufficient to optimize risk

management resource allocations, or even to avoid very poor allocations, as in the above example. Calculating optimal risk management resource allocations requires quantitative information beyond what a risk matrix provides, for example, about budget constraints and about interactions among countermeasures. In general, risk rankings calculated from frequency and severity do not suffice to guide effective risk management resource allocation decisions.

5.3. Categorization of Uncertain Consequences is Inherently Subjective

To use a risk matrix, it is necessary to be able to categorize the alternatives being compared into the cells of the matrix. However, decision analysis principles imply that there is no objective way to categorize severity ratings for events with uncertain consequences. Subjective risk attitudes play an essential (but seldom articulated) role in categorizing severity for such events. Thus, the information in a risk matrix represents a mixture of factual (probability and consequence) information about the risk and (usually unstated) psychological information about the risk attitude of the person or people performing the risk categorization. Since the risk attitudes of the builders are seldom documented, it can be impossible to determine how consequence severity classifications should be changed when someone else views or uses the matrix.

5.4. Example: Severity Ratings Depend on Subjective Risk Attitudes

For a decisionmaker with an exponential utility function, the certainty equivalent (CE) value of a prospect with normally distributed consequences is $CE(X) = E(X) - k \times \text{Var}(X)$, where k is a parameter reflecting subjective risk aversion ($k = 0.5 \times$ coefficient of risk aversion); $E(X)$ is the mean of prospect X ; $\text{Var}(X)$ is its variance; and $CE(X)$ is its certainty-equivalent value (i.e., the deterministic value that is considered equivalent in value to the uncertain prospect) (Infanger, 2006, p. 208). Consider three events, A, B, and C, with identical probabilities or frequencies and having normally distributed consequences (on some outcome scale) with respective means of 1, 2, and 3 and respective variances of 0, 1, and 2. The certainty equivalents of prospects A, B, and C are:

$$CE(A) = 1$$

$$CE(B) = 2 - k$$

$$CE(C) = 3 - 2k.$$

For a risk-neutral decisionmaker (for whom $k = 0$), the ordering of the prospects from largest to smallest certainty equivalent value is therefore: $C > B > A$. For a risk-averse decisionmaker with $k = 1$, all three prospects have the same certainty equivalent value of 1. For a more risk-averse decisionmaker with $k = 2$, the ordering of the prospects is: $A > B > C$. Thus, the certainty equivalents of the severities of the prospects are oppositely ordered by decisionmakers with different degrees of risk aversion. There is no objectively correct ordering of prospect severity certainty equivalents independent of subjective attitudes toward risk. But risk matrices typically do not specify or record the risk attitudes of those who use them. Users with different risk attitudes might have opposite orderings, as in this example. Neither is objectively (independent of subjective risk attitude) more correct than the other. As a result there is no objective way to classify the relative severities of such prospects with uncertain consequences.

5.5. Example: Pragmatic Limitations of Guidance from Standards

In practice, various standards provide written guidance on how to classify severities for use in risk matrices. For example, Table VIII shows the severity ratings suggested in a 1998 General Accounting Office report on "Combating Terrorism," based on the widely cited Military Standard 882C (https://cra.army.mil/guidance/system_safety/882C.pdf). As that standard notes: "These hazard severity categories provide guidance to a wide variety of programs. However, adaptation to a particular program is generally required to provide a mutual understanding . . . as to the meaning of the terms used in the category definitions. The adaptation must define what constitutes system loss, major or minor system or environmental damage, and severe and minor injury and occupational illness." Even with these caveats, the guidance in Table VIII does not resolve the type of ambiguity in the previous example. For example, it offers no guidance on how to rate a consequence that is zero with probability 90% but catastrophic otherwise (perhaps depending on wind direction or crowding of a facility or of evacuation routes at the time of a terrorist attack). Moreover,

Table VIII. Severity Levels of Undesired Event Consequences for Combating Terrorism

Severity Level	Characteristics
I Catastrophic	Death, system loss, or severe environmental damage
II Critical	Severe injury, severe occupational illness, major system or environmental damage
III Marginal	Minor injury, minor occupational illness, or minor system or environmental damage
IV Negligible	Less than minor injury, occupational illness, or less than minor system or environmental damage

Source: GAO (1998).

it introduces other ambiguities. For example, how should one rate the severity of a consequence that consists of 1 death and 1 severe injury compared to that of a consequence of 0 deaths but 50 severe injuries? The answer is not obvious from Table VIII.

The discrete qualitative categories provided in guidance such as Table VIII are also inconsistent with the continuous quantitative nature of many physical hazards. For example, should a condition that causes “negligible” environmental damage on each occurrence (e.g., leaking 1 ounce of jet fuel per occurrence) but that causes a high frequency of these small events (e.g., averaging 5 events per hour) truly have a lower severity rating than a second condition that causes more damage per occurrence (e.g., leaking 10 pounds of jet fuel per occurrence) but that causes less frequent occurrences (e.g., once per week)? (Both would be assigned the highest possible frequency rating by Military Standard 882C.) If so, then the risk matrix analysis could give lower priority to eliminating a threat of leaking 52.5 pounds per week (= 5 ounces per hour × 24 hours/day × 7 days per week) than to eliminating a threat of leaking only 10 pounds per week, due to the greater “severity” of 10 pounds than 1 ounce and the equal “frequency” rating of common events (an example of range compression). In such cases, the idea of rating severity independently from frequency appears flawed.

Focusing on applying qualitative rating criteria, rather than on more quantitative comparisons of risks, can create irrational risk management priorities. The following example illustrates how uncritical application of risk matrix guidance might promote misperceptions and misrankings of the relative risks of different strategic investment opportunities.

5.6. Example: Inappropriate Risk Ratings in Enterprise Risk Management (ERM)

Suppose that a company must choose between the following two risky investment strategies for

responding to major and pervasive uncertainties, such as climate change risks.

- Strategy A has probability 0.001 of leading to a small growth rate that barely meets shareholder expectations; otherwise (probability 99.9%) shareholder value and growth will increase by a negligible amount (e.g., < 0.00001%), disappointing shareholders and failing to meet their expectations.
- Strategy B has probability 50% of causing rapid and sustained growth that greatly exceeds shareholder expectations; otherwise (e.g., if the outcome of a crucial R&D project is unsuccessful), shareholder value and growth will not grow (growth rate = 0%).

Which strategy, A or B, better matches a responsible company’s preferences (or “risk appetite”) for risky strategic investments?

Commonsense might suggest that Strategy B is obviously better than Strategy A, as it offers a 50% probability of greatly exceeding expectations instead of a 0.1% probability of barely meeting them, with no significant difference in downside risk. However, uncritical application of risk matrices suggested as examples for enterprise risk management (ERM) systems could rate B as more risky than A. For example, Australia published a risk management “guide for business and government . . . [that] is consistent with the Australian and New Zealand Standard for Risk Management, AS/NZS 4360:2004, which is widely used in the public and private sectors to guide strategic, operational and other forms of risk management. The Guide describes how the routine application of the Standard can be extended to include the risks generated by climate change impacts” (Australian Government, 2006). The illustrative risk matrix and category definitions for a commercial business (Tables 10–12 of the Guide) could be used to assign a “medium” risk priority to Strategy A but a “high”

risk to strategy B, making B appear to be less attractive than A. (For A, the likelihood of the adverse consequence, 99.9%, is classified as “almost certain.” The consequence is described as “Growth would be achieved but it would fail to meet expectations,” which is classified as a “minor” consequence. The risk matrix example in Figure 12 of the Guide categorizes the likelihood-consequence pair (*almost certain, minor consequence*) as a “medium” risk. For B, the likelihood of the adverse consequence is classified as “likely,” the consequence is described as “There would be no growth,” and this is classified as a “moderate” consequence. The combination (*likely, moderate consequence*) is categorized as a “high” risk.) Thus, a tight focus on implementing the discrete categorization criteria in the guidance could distract attention from the fact that most shareholders would gladly trade a negligible increase in adverse consequences for a large increase in the probability of a much better outcome. In the terminology of multicriteria decision making, the discrete categorization of consequences and probabilities inherent in risk matrices can produce noncompensatory decision rules that do not reflect the risk trade-off preferences of real decisionmakers and stakeholders.

Quantitative risk assessment was developed in part to help prevent the types of paradoxes illustrated in these examples. Even if the quantities in the fuel leaking example were quite uncertain (e.g., an average of 1–10 ounces every few minutes in the first case and 0–100 pounds every few months in the second), a rough quantitative calculation would reveal that the first threat is much more severe than the second. Similarly, even a rough quantitative comparison of strategies A and B in the enterprise risk management example would show that B is much more attractive than A. By contrast, qualitative or semiquantitative risk assessments based on ordered categories do not necessarily prevent rating reversals and misallocations of resources, as in these examples—and may even unintentionally encourage them, by directing risk management effort and attention away from the key quantitative comparisons involved and toward the (often inherently subjective) task of categorizing frequency and severity components.

6. DISCUSSION AND CONCLUSIONS

The theoretical results in this article demonstrate that, in general, quantitative and semiquantitative risk matrices have limited ability to correctly reproduce the risk ratings implied by quantitative models, es-

pecially if the two components of risk (e.g., frequency and severity) are negatively correlated. Moreover, effective risk management decisions cannot in general be based on mapping ordered categorical ratings of frequency and severity into recommended risk management decisions or priorities, as optimal resource allocation may depend crucially on other quantitative information, such as the costs of different countermeasures, the risk reductions that they achieve, budget constraints, and possible interactions among risks or countermeasures (such as when fixing a leak protects against multiple subsequent adverse events).

Categorizing severity may require inherently subjective judgments (e.g., reflecting the rater’s personal degree of risk aversion, if severity is modeled as a random variable) and/or arbitrary decisions about how far to aggregate multiple small and frequent events into fewer and less frequent but more severe events. The need for such judgments, and the potential for inconsistencies in how they are made by different people, implies that there may be no objectively correct way to fill out a risk matrix.

Conversely, the meaning of a risk matrix may be far from transparent, despite its simple appearance. In general, there is no unique way to interpret the comparisons in a risk matrix that does not require explanations—seldom or never provided in practice—about the risk attitude and subjective judgments used by those who constructed it. In particular, if some consequence severities are random variables with sufficiently large variances, then there may be no guarantee that risks that receive higher risk ratings in a risk matrix are actually greater than risks that receive lower ratings.

In summary, the results and examples in this article suggest a need for caution in using risk matrices. Risk matrices do not necessarily support good (e.g., better-than-random) risk management decisions and effective allocations of limited management attention and resources. Yet, the use of risk matrices is too widespread (and convenient) to make cessation of use an attractive option. Therefore, research is urgently needed to better characterize conditions under which they are most likely to be helpful or harmful in risk management decision making (e.g., when frequencies and severities are positively or negatively correlated, respectively) and that develops methods for designing them to maximize potential decision benefits and limit potential harm from using them. A potentially promising research direction may be to focus on placing the grid lines in a risk matrix to minimize the maximum loss from misclassified risks.

We hope to present some positive results from this optimization-based approach soon.

APPENDIX: PROOF OF THEOREM 1

By definition, the lower left-most cell is green. Consistent coloring implies that any contour must be green if it lies below/to the left of the one passing through the upper right corner of this lower left-most cell (i.e., the contour through the points (0.04, 1), (0.2, 0.2), (1, 0.04) in the numerical example in Table IV), since (a) it passes through the lower left-most cell (which is green by definition); and (b) none of the cells that it passes through is red (by Lemma 2). By construction, such a green contour passes through all cells in the left-most column and in the bottom row.

Now, consider the cell directly above the lower left-most cell (i.e., the cell containing the point (0.1, 0.3) in Table IV). Suppose that, contrary to the claimed result, this cell is not green. It cannot be red, by Lemma 2. For it to be an intermediate color (not green), it must contain at least one red contour (by color consistency and the fact that a green contour passes through it). This cell cannot be “between” a red and a green cell, since it is on an edge of the matrix, so it cannot acquire an intermediate color that way. This red color neither comes from the cell above it in the left-most column (which is nonred, by Lemma 2), nor from any cell in the bottom row (again by Lemma 2). Since contours are downward-sloping, the only remaining possibility is for the cell to its right (the cell containing (0.3, 0.3) in Table IV) to be red. But this would violate betweenness (at the point (0.2, 0.2) in Table IV). Therefore, the assumption that the cell directly above the lower left-most cell is not green leads to a contradiction. Hence, it must be green. By a symmetrical argument, the cell directly to the right of the lower left-most cell (the cell containing (0.3, 0.1) in Table IV) must also be green.

Next, suppose that the third cell in the left-most column (the one containing (0.1, 0.5) in Table IV) is not green. Since green contours pass through it (as it is in the left-most column), it can only be nongreen if some red contour also passes through it (by color consistency and the fact that it is an edge cell). This red contour could not come from a red cell below it in the left-most column, or in the bottom row (by Lemma 2), nor from the cell directly to its southeast (containing (0.3, 0.3) in Table IV) (since if that were red, it would violate Lemma 1 and betweenness for the cells so far proved to be green). The only remaining possibility is that the cell to its right (the one containing (0.3, 0.5)

in Table IV) is red. But this would violate betweenness (with the second cell in the left-most column, the cell containing (0.1, 0.3) in Table IV, which we have proved above must be green). Hence, the assumption that the third cell in the left-most column is not green implies a contradiction. So, it must be green. Symmetrically, the third cell in the bottom row must be green. This construction (showing that a cell directly above a green cell in the first column, with only nonred cells to its southeast, must itself be green) can be iterated for all remaining cells in the left-most column, thus establishing that they all must be green; symmetrically, all remaining cells in the bottom row must be green. This proves part (a). Part (b) is then an immediate consequence of part (a) and Lemma (2). QED

Comment: This proof does not depend on the number of rows or columns in the table. Therefore, its conclusion (that the left-most column and bottom row consist entirely of green cells) holds for risk matrices of any size, under the stated conditions of weak consistency, betweenness, and consistent coloring.

REFERENCES

- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.
- Australian Government. (2006). Australian Greenhouse Office, in the Department of the Environment and Heritage. *Climate Change Impacts & Risk Management: A Guide for Business and Government*. Canberra, Australia: Commonwealth of Australia. Available at <http://www.greenhouse.gov.au/impacts/publications/pubs/risk-management.pdf>. (Last accessed 8-19-2007.)
- Bertsimas, D., & Nino-Mora, J. (1996). 1986. Conservation laws, extended polymatroids and multiarmed bandit problems: Polyhedral approach to indexable systems. *Mathematics of Operations Research*, 21(2), 257–306.
- California Department of Transportation, Federal Highway Administration, California Division. (2007). *Systems Engineering Guidebook for ITS Version 2.0*. Available at www.fhwa.dot.gov/cadiv/segb/views/document/Sections/Section3/3_9_4.htm.
- Chen, J. J., Tsai, C. A., Moon, H., Ahn, H., Young, J. J., & Chen, C. H. (2006). Decision threshold adjustment in class prediction. *SAR QSAR Environmental Research*, 17(3), 337–352.
- Cox, L. A. Jr., Babayev, D., & Huber, W. (2005). Some limitations of qualitative risk rating systems. *Risk Analysis*, 25(3), 651–662.
- Cox, L. A. Jr., & Popken, D. A. (2007). Some limitations of aggregate exposure metrics. *Risk Analysis*, 27(2), 439–445.
- Dreiseitl, S., Ohno-Machado, L., & Vinterbo, S. (1999). Evaluating variable selection methods for diagnosis of myocardial infarction. *Proc AMIA Symposium*, 246–250.
- Federal Aviation Administration. (2007). *Introduction to Safety Management Systems (SMA) for Airport Operators (Advisory Circular)*, February 28, 2007. Washington, DC: U.S. Department of Transportation. Available at www.faa.gov/airports_airtraffic/airports/resources/advisory_circulars/media/150-5200-37/150_5200_37.doc. (Last accessed 8-19-2007.)
- Federal Highway Administration of the U.S. Department of Transportation. (2006). *Risk Assessment and Allocation for High-*

- way *Construction Management*. Washington, DC: U.S. Department of Transportation. Available at <http://international.fhwa.dot.gov/riskassess/index.htm>. (Last accessed 8-19-2007.)
- GAO. (1998). *Combating Terrorism: Threat and Risk Assessments Can Help Prioritize and Target Program Investments*. Washington, DC: U.S. Government Accounting Office. Available at <http://www.gao.gov/archive/1998/ns98074.pdf> (Last accessed 8-19-2007.)
- Infanger, G. (2006). Dynamic asset allocation strategies using a stochastic dynamic programming approach. In S. A. Zenios & W. T. Ziemba (Eds.), *Handbook of Assets and Liability Management* (Vol. 1, Ch. 5). New York: North Holland.
- Lloyd, G. R., Brereton, R. G., Faria, R., & Duncan, J. C. (2007). Learning vector quantization for multiclass classification: Application to characterization of plastics. *Journal of Chemical Information and Modeling*, 47(4), 1553–1563.
- MITRE Risk Management Toolkit. (1999–2007). Available at <http://www.mitre.org/work/sepo/toolkits/risk/ToolsTechniques/RiskMatrix.html>. (Last accessed 11-19-2007.)
- Renfro, N. A., & Smith, J. L. (2007). *Whole Building Design Guide: Threat/Vulnerability Assessments and Risk Analysis*. Washington, DC: National Institute of Building Sciences. Available at <http://www.wbdg.org/design/riskanalysis.php>. (Last accessed 8-19-2007.)

Copyright of *Risk Analysis: An International Journal* is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.