*Risk Analysis, Vol. 31, No. 2, 2011* DOI: 10.1111/j.1539-6924.2010.01492.x

# How Probabilistic Risk Assessment Can Mislead Terrorism Risk Analysts

**Gerald G. Brown[1] and Louis Anthony (Tony) Cox, Jr.[2,*]**

Traditional probabilistic risk assessment (PRA), of the type originally developed for engineered systems, is still proposed for terrorism risk analysis. We show that such PRA applications are unjustified in general. The capacity of terrorists to seek and use information and to actively research different attack options before deciding what to do raises unique features of terrorism risk assessment that are not adequately addressed by conventional PRA for natural and engineered systems—in part because decisions based on such PRA estimates do not adequately hedge against the different probabilities that attackers may eventually act upon. These probabilities may differ from the defender's (even if the defender's experts are thoroughly trained, well calibrated, unbiased probability assessors) because they may be conditioned on different information. We illustrate the fundamental differences between PRA and terrorism risk analysis, and suggest use of robust decision analysis for risk management when attackers may know more about some attack options than we do.

**KEY WORDS:** Decision analysis; expert elicitation; terrorism risk analysis

## 1. INTRODUCTION

Following continued practice by the Department of Homeland Security (DHS), Ezell *et al.*[1] recently asserted that "a useful first-order indicator of terrorism risk is the expected consequences (loss of lives, economic losses, psychological impacts, etc.)" and claimed that: "We take it for granted that all probabilities are conditional on our current state of knowledge . . . [T]here is no fundamental difference in this type of conditioning compared to conditioning probability judgments in the case of natural or engineered systems." This important claim—that the same type of conditional probability assessment applies as well to terrorism risk analysis as to probabilistic risk assessment (PRA) of natural hazards and engineered systems—is presented without proof. We believe that this is importantly incorrect, and that PRA calculations based on this idea can be highly misleading, rather than useful, for terrorism risk analysis. In particular, applying conventional PRA to terrorism may result in recommendations that *increase* the risk of attacks, or that fail to reduce them as much as possible for resources spent. This article seeks to clarify why conditioning risk estimates on knowledge or beliefs about the future actions of others, who in turn may condition their preferences for alternative actions on what they know about our risk estimates, leads to new problems in terrorism risk analysis that cannot be solved well, if at all, by traditional PRA.

The particular formalism advocated by Ezell *et al.*, based on the threat–vulnerability–consequence (TVC) formula "*Risk = Probability of attack × Probability that attack succeeds, given that it occurs × Consequence of a successful attack*," has previously been criticized on other grounds, such as its failure to optimally diversify protective investments, or to

[1]Operations Research Department, Naval Postgraduate School, Monterey, CA, USA.
[2]Cox Associates, Denver, CO, USA.
*Address correspondence to Louis Anthony (Tony) Cox, Jr., Cox Associates, 503 Franklin Street, Denver, CO 80218, USA; tel: 303-388-1778; fax: 303-388-0609; TCoxDenver@aol.com.

 

account for correlations and dependencies among the factors on its right-hand side, or to reveal the costs and benefits of alternative risk management decisions, or to use well-defined concepts that demonstrably improve, rather than confuse and obscure, decision making.[2,3] Here, we focus on a different issue: considering why a belief that there is no fundamental difference in conditional probability calculations for systems with and without reasoning agents can provide a dangerously misleading foundation for terrorism risk analysis. More constructively, we suggest that a different approach, based on explicit recognition that attack probabilities may depend on information that an attacker has but that we do not have, can be used to make more robust and useful risk management decisions—demonstrably superior to those from PRA based only on our own information—by enabling a defender to allocate defensive resources to hedge against what he does not know about the attacker's information.

## 2. ATTACK RISKS MAY DEPEND ON THE DEFENDER'S RISK ANALYSIS RESULTS

The theory of "common knowledge" in information economics[4] explains why conditioning probability judgments on the actions, beliefs, or statements of other reasoning agents differs in crucial respects—some of them startling—from conditioning probability judgments on information about a system without reasoning agents. One fundamental difference is that the behavior of reasoning agents may depend on what they know (or believe or infer) about what we know . . . including what we know they know, what they know we know they know, and so forth. This can lead to risks very different from those in systems without reasoning agents. For example, an attacker who uses the defender's allocation of defensive resources to help decide where to attack (e.g., by reasoning that the defender will give priority to protecting what he values most) poses a different kind of threat from an earthquake or a tornado, which strikes at random. Traditional PRA (e.g., based on event trees, F-N curves, probabilistic simulation models, etc.) is appropriate for the second type of hazard, but not the first.

### 2.1. Example: PRA Estimates that Inform an Adversary May be Self-Defeating

Following the rationale of Ezell *et al.* (p. 578),[1] suppose that our terrorism experts compare the relative probabilities of several different possible attacks, based on our judgments of the "relative technical difficulties of executing these attacks." Specifically, suppose that our experts rank five possible attacks on this basis from most likely to least likely. Consider an attacker who understands our risk assessment and who uses this decision rule: Do not attempt either of the two attacks that we (the Defender) rank as most likely (because we may be prepared for those), nor either of the two that are ranked least likely (because they may be too unlikely to succeed to be attractive). Instead, undertake the attack that is midway between these extremes. Then, any PRA ranking that our experts generate will be self-defeating, in that the attacks that it ranks as most likely will actually have zero probability, whereas the middle-ranked attack will actually have a higher probability. This is because the attacker cares about, and uses, the results of our PRA to decide what to do. Natural and engineered systems do not act this way.

### 2.2. Example: PRA Estimates that Inform Enemy Actions May be Self-Fulfilling

Conversely, suppose that an attacker is very uncertain about whether an attack will succeed if attempted. He uses the decision rule: Attempt whatever attack our (the Defender's) PRA identifies as most likely to succeed (after any defensive measures have been taken). In this case, whatever potential attack our PRA ranks at the top becomes the one that is actually attempted. In this context, using a random number generator to rank-order attacks would be just as accurate as expert elicitation, or any other ranking method. Moreover, although our experts might assess identical PRA risk estimates in this example and the previous one, it is clear that *the true risk of any specific attack depends on what decision rule the attacker uses*, and on what our own PRA concludes. This is very different from the behaviors of any natural or engineered system that does not use decision rules and that does not respond to PRA results.

### 2.3. Example: Risk Depends on Attacker Patience and Choices, Not Random Variables

Ezell *et al.* advocate using the formula "*Risk = Probability of attack* × *Probability that attack succeeds, given that it occurs* × *Consequence of a successful attack*." But the phrase "given that it occurs" glosses over crucial information about *why*

(i.e., based on what decision rule) the attacker attacks. Without this information, the proposed formula for risk is ambiguous. For example, suppose the three factors on the right side vary from day to day. Suppose an attacker uses the decision rule "Attack if and only if the probability that the attack will succeed if it is attempted today is at least $p$," where $p$ is a number between 0 and 1. A patient attacker may wait for $p$ to be close to 1; a less patient attacker will set $p$ lower. The "*Probability that attack succeeds, given that it occurs*" depends on the attacker's patience parameter, $p$, because this (and the successive realizations of the random variables) determines when the conditions for an attack are triggered. More patient attackers will have higher success probabilities, on average, than less patient attackers, but will wait longer. An assessment of risk based in part on the factor "*Probability that attack succeeds, given that it occurs*," but without specifying $p$, is underdetermined. On the other hand, an assessment of risk based on specifying the above factors may be self-defeating. For example, suppose an attacker sets $p$ by dividing the attacker's own estimate of vulnerability by 2 (as would be appropriate if the attacker is a Bayesian with a uniform prior for this probability and if he interprets the defender's vulnerability estimate as an upper bound on the true but unknown value of this success probability). In this case, the defender's PRA estimate of vulnerability will always be twice the true (realized) value that triggers an attack. In short, the attacker's choice of decision rule, which makes no explicit appearance in the formula "*Risk = Probability of attack × Probability that attack succeeds, given that it occurs × Consequence of a successful attack,*" determines the true risk. This may be very different from the formula-predicted risk.

## 3. PRA FOR TERRORIST ATTACKS MAY RECOMMEND POOR RISK MANAGEMENT DECISIONS

The preceding examples all have the attacker exploit information about the defender's PRA results. Even if this is impossible (e.g., because the PRA results are kept secret), PRA can recommend poor risk management decisions. The example presented in this section shows that basing defender resource allocations on a PRA, without considering that the attacker may undertake research that will leave him better informed, with different probabilities than the defender, may lead to ineffective allocations of resources to defending potential targets. The key insight is that attack probabilities depend on what the attacker knows or believes, rather than on what the defender knows or believes. In contrast to risk analysis for defense against random events, risk analysis for reasoning attackers should consider how what the attacker discovers in the future may affect his future decisions. Failing to do so may lead a risk manager who relies on PRA to allocate resources based on a "best bet," given what is currently known (to the defender), without adequately hedging against what the attacker may know at the time of the attack.

### 3.1. Example: Traditional PRA Misallocates Defensive Resources

#### *3.1.1. Setting*

Consider an attacker who must choose between two possible attack plans, A and B. He can afford to undertake either one, but not both. Assume the following:

(1) It is initially common knowledge between the attacker and the defender that attack plan A has a 30% probability of succeeding if attempted, whereas attack plan B has a 60% probability of succeeding if attempted.
(2) The defender has enough budget to either (a) defend heavily against either one of the possible attacks, A or B (but not both), thus reducing the consequence of a successful attack that has been defended against from 1 million lives lost to 0.1 million; or else (b) partly protect against both possible attacks, thus cutting the consequences of a successful attack (whether A or B) from 1 million lives lost to 0.2 million. (For example, if A and B represent two different port cities that might be targeted for an attack, and if the available defensive budget can be spent on additional detection and early warning, preparation for damage control, and consequence mitigation countermeasures, then these countermeasures might be concentrated at one location, or spread more thinly across both.)
(3) The attacker will do some research on each option, A and B, and will then choose the one that gives the higher probability of success. The attacker's only goal is to complete a successful attack; thus, he always chooses the available option with the greatest success

**Table I.** Expected Loss (in Millions of Lives) for Each Possible Combination of Attacker Decision (Columns) and Defender Decision (Rows): Which Decision Row Should Defender Choose to Minimize Expected Loss?

| | Attack A | Attack B |
|---|---|---|
| Defend A heavily | 0.1 | 1 |
| Defend B heavily | 1 | 0.1 |
| Defend both lightly | 0.2 | 0.2 |
| Success probability for attack: | 0.3 for A | 0.6 for B |

probability, as this maximizes his expected utility.

Table I summarizes the above facts.

### *3.1.2. Problem*

Given the above information, what should the defender do? How should he allocate his available budget to best reduce the expected loss of life from an attack? What probabilities should he assess for attacks A and B?

### *3.1.3. Traditional PRA Analysis*

Expected-value and PRA analyses recommend that the attacker should choose attack B (because doing so maximizes his expected utility, as calculated from the success probabilities of 0.6 for B and 0.3 for A). The defender should therefore allocate his resources to defend against attack B. Thus, the predicted probability of attack A is 0 and the probability of attack B is 1.

### *3.1.4. Analysis Based on Attacker's Possible Information Sets*

We propose that the defender's best move should in reality depend crucially on the *unmodeled details* of part (3) above, i.e., on *what research opportunities are available to the attacker* to reduce uncertainties before attacking. For example, suppose that attack option B, but not option A, can be researched before attempting it. If such research will reveal whether B will succeed if it is attempted (for which the common knowledge prior probability, before doing the research is 0.6), then the attacker's best (expected utility-maximizing) choice *after* doing the research is a random variable with respect to the information available *before* doing the research. Specifically, with respect to the preresearch information (available to both the defender and the attacker), there is a probability of 0.6 that the attacker will choose attack option B (because this is the prior probability that the research will reveal that B will succeed), *but there is a probability of 0.4 that the attacker will choose attack option A* (because this is the prior probability that the research will reveal that B will fail, leaving A as the only viable attack option). In this case, defending only against attack B, as a naïve PRA analysis might recommend, would be expected to save 540,000 lives (= 60% attack probability for B × 900,000 lives saved if attack B is made). Splitting defenses across A and B would be expected to save 576,000 lives (= 0.6 × 800,000 lives saved if attack B is used + 0.4 × 0.3 × 800,000 lives saved if attack B is found to be nonviable, so that attack A is tried instead, and then succeeds). Thus, simply defending against the attacker's predicted "expected utility-maximizing" choice (with expected utilities calculated using the information available to a defender who does not know the attacker's research results) would yield a poor risk management recommendation.

By contrast, if research by the attacker cannot change the prior probabilities for success of attack options A and B by very much (e.g., resolving the 0.6 success probability for B into either a 0.8 if favorable, or a 0.4 if unfavorable, with equal prior probabilities; and resolving the 0.3 success probability for A into either a 0.25 or a 0.35, with equal probabilities), then the prediction that the attacker will choose attack B (and not A), and the resulting prescription that all resources should therefore be allocated to defending against attack B, would be correct.

This example illustrates the following general points, which do not depend on the many oversimplifications made for purposes of a simple illustration (e.g., common-knowledge priors, known research opportunities and conditional probabilities, known objectives and strategies determining actions as a function of information).

(1) First, *the probabilities of alternative attacker actions (such as A vs. B) assessed by the defender should depend on what research opportunities are available to the attacker*. PRA and event tree analysis are not developed or intended for systems that can actively perform their own research before deciding what to do. In the example, the probabilities of attacks A and B can be any of the following pairs,

depending on the research opportunities available to the attacker:

(a) (0.4, 0.6) (meaning Pr(A) = 0.4 and Pr(B) = 0.6), if the attacker can perform highly informative research on the success of B (but not A) before attempting it. This case is analyzed in detail above.
(b) (0.0, 1.0), if no such highly informative research is possible. In this case, as just discussed, the attacker's best bet is to select B and hope for success.
(c) (0.3, 0.42), if highly informative research is possible for both A and B, and if the attacker selects A rather than B if he finds that both attacks can succeed. (In this case, the attack probability for A is 0.3 = Pr(A will succeed). The probability for attack option B = Pr(A will not succeed & B will succeed) = (1 – 0.3) ∗ 0.6 = 0.42.

Thus the "threat" (i.e., probability of attack) for A can be any of 0.3, 0.4, or 1, depending on what information the attacker can collect before deciding what to do. Such a set of possible probabilities, all of which are fully consistent with the constraints imposed by what the defender knows, is called an *uncertainty set.*[5]

(2) If the research options available to the attacker are unknown to the defender, then *the probabilities of different attacks (based on the attacker's information) are uncertain and are not uniquely predictable by the defender* (unless and until he learns what the attacker knows): they can be any combination in the uncertainty set. An important decision-analytic tradition teaches that unique subjective probabilities can and should be assessed or elicited for any event, e.g., by assessing willingness to bet upon various contingencies. The logic of this traditional approach seems impeccable for a single decisionmaker, at least if the foundational problem of "small worlds" (that the probabilities and utilities of specific acts and outcomes can vary with the amount of details included in their description) can be ignored.[6–8] However, in the context of a decisionmaker being advised by experts, we recommend recognizing that unique correct attacker probabilities cannot necessarily be determined by the defender. Robust optimization[5]—acting to maximize the minimum possible expected utility, or, equivalently, to minimize the maximum possible expected loss, when attacker probabilities are only known to lie within some "uncertainty set" of possible probabilities—offers a constructive approach for decisionmaking with such unknown probabilities Its key ideas are that, even if there is no objective basis for quantifying a specific joint probability distribution for uncertain quantities, it may be possible to identify an uncertainty set of alternative possible probability distributions (e.g., corresponding to alternative observations that the attacker may have made), and to choose acts to minimize maximum expected loss, or maximize minimum expected utility, over all distributions in the uncertainty set. Such a conservative strategy for coping with uncertain probabilities is implied by normative axioms (which specialize to yield expected utility theory when the uncertainly set contains only one probability distribution). Moreover, for many uncertainty sets, solving the robust optimization problem is computationally easier than solving the corresponding expected utility maximization problem.

(3) *Additional research may have zero information value to the attacker and defender*. Resolving the high-level success probabilities of 0.6 for attack option B and 0.3 for attack option A into better-informed estimates, of either 0.4 or 0.8 for A, and either 0.25 or 0.35 for B, would have zero value of information (VOI), as these refined estimates would not change attacker behavior (and hence the defender's best decision). More generally, information that cannot change decisions has no value (VOI = 0), even if it enables more accurate predictions. When further refinements in information and models make relatively small differences in the current uncertainty set-based predictions and robust decision recommendations, then the value of the additional information is also small (or zero), and current decision recommendations are unlikely to change even if a more detailed, better-informed model is constructed. In this sense, a partial understanding of attack options may be good enough to establish the best strategies for attacker and defender, despite remaining uncertainties.

(4) *Better information for attackers may reduce threats*. The lowest risk to the defender, in this example, occurs when the attacker can obtain highly informative research on both attack options, A and B. Then, some bets that might have seemed worth taking with less information can be seen by the attacker to be not worth taking after all. Only in this case is the probability of *some* attack (A or B) less than 1. This feature of the example illustrates a new way of thinking about reducing threats: we should be allocating resources not only to make attacks less likely to succeed, if attempted (by reducing our vulnerabilities), but also to degrade the *assessed value of attack options to attackers*. From this perspective, an opportunity to improve our defenses arises from not having attackers believe that they can exploit superior information about attack options to identify attractive options that we have overlooked, due to our failure to research, identify, and forestall such options. The possibility of reducing or eliminating threats by revealing credible information about our research and countermeasures does not apply to random failures in engineered or natural systems. Yet, it may be valuable in terrorism risk analysis. Similarly, deterrence plays no role in PRA for safety systems, but is important in security risk assessment.

The problem illustrated here arises, not because of the use of expert judgments of probabilities *per se*, but because the assessed probabilities (e.g., 0.6 for the probability that A succeeds if attempted, and 0.3 for the probability that B succeeds if attempted) do not represent crucial aspects of the attacker's decision process—specifically, the future information on which the attacker will act. Failing to model what the attacker may learn that will influence his future decision makes the assessed probabilities misleading for use in deciding how best to defend against the attacker's future choice. We do not contend that modeling of the attacker's future information is impossible (e.g., via decision trees or game trees), but rather that it is important to do such modeling, instead of truncating the modeling of attacker behavior by assigning probabilities to events or actions based on our own current information. Modeling the attacker's decision tree in useful detail may even suggest opportunities for deterrence or risk reduction, which use of judged probabilities for actions (T) and successes (V) might not.

**Table II.** A Hypothetical Example of Historical Data that Might be Used to Inform the Defender's Risk Estimates

| | Attack A | Attack B |
|---|---|---|
| Attack succeeded | 505 | 110 |
| Attack failed | 595 | 901 |
| Total attacks of each type | 1,100 | 1,011 |
| Success fractions | 0.46 | 0.11 |

## 4. THE IRRELEVANCE OF DEFENDER INFORMATION TO PREDICTING HOW DEFENSES AFFECT RISK

Let us now put ourselves in the shoes of one of the defender's experts. Suppose we have collected historical data on different types of attacks, and their success rates, as shown in Table II. Moreover, suppose it is common knowledge the attackers will strike again as soon as they are ready, and that they will use the same decision processes (e.g., exploiting local knowledge of attack opportunities and costs, resources, constraints, and preparations) that generated the data in Table II. The defender can afford to improve defenses against either a type A attack or a type B attack (but not both). He seeks guidance from us (and the TVC formula) on which attack to defend against.

Armed with the knowledge in Table II, we might conclude that the probability that the next attack will be of type A is approximately $1{,}100/(1{,}100 + 1{,}011) = 0.52$; the probability that it will be of type B is approximately 0.48; and the probability that an attack of type A will succeed if attempted is 0.46, whereas the probability that an attack of type B will succeed if attempted is 0.11. Thus, we would assess vulnerability to type A attacks as unambiguously greater than vulnerability to type B attacks (0.46 vs. 0.11). Increasing defenses against type A attacks, even if it diverts some attacks to type B attacks, might be expected to significantly reduce the fraction of successful attacks.

Now let us examine the same situation from the attacker's point of view. Suppose the attacker knows something the defender does not: that some attacks are planned or carried out by relatively well-trained ("strong" or "elite") members, whereas the rest are carried out by less proficient ("weak" or "ordinary" members). Tables III and IV show the data for these

**Table III.** Example Data for Strong (Elite) Attackers

| | Attack A | Attack B |
|---|---|---|
| Attack succeeded | 500 | 10 |
| Attack failed | 500 | 1 |
| Total attacks of each type | 1,000 | 11 |
| Success fractions | 0.50 | 0.91 |

**Table IV.** Hypothetical Example Data for Weak (Ordinary) Attackers

| | Attack A | Attack B |
|---|---|---|
| Attack succeeded | 5 | 100 |
| Attack failed | 95 | 900 |
| Total attacks of each type | 100 | 1,000 |
| Success fractions | 0.05 | 0.10 |

two subgroups. Tables III and IV sum, cell by cell, to give Table II.

From the attacker's point of view, it is unequivocally clear that the defender is already much more vulnerable to type B attacks than to type A attacks (by either type of attacker)—precisely the opposite of what the defender's expert concluded based on Table II. If the defender now invests in further reducing vulnerability to type A attacks, displacing the attacker's allocation of attack resources toward more type B attacks, then instead of reducing the fraction of successful attacks (by approximately fourfold, as might be expected from Table II), the success fraction for attacks will approximately double (as revealed by Tables III and IV), for each type of attacker (ordinary and elite).

This example illustrates that *knowledge of threat and vulnerability data, such as those in Table II (together with consequence data), does not allow us to predict how alternative risk management interventions will affect risk.* Other factors are essential, such as what types of attacker resources (e.g., elite vs. ordinary) produce how much damage when deployed in alternative ways (e.g., to type A vs. type B attacks). To predict how risk will change when the attacker reallocates resources to adapt to changes made by the defender, one needs to consider the information that the attacker has (such as that type B attacks are about twice as likely to succeed as type A attacks, for each type of attack resource). That the defender's measure of vulnerability, as "the probability that an attack succeeds, given that it occurs" happens to be approximately four times greater for type A than for type B attacks is *irrelevant* to the attacker because it merely reflects past allocations of resources (elite vs. ordinary) to attack opportunities, but reveals nothing about the relative ease of successfully completing type A vs. type B attacks. For the same reason, this measure of vulnerability *should also be viewed as irrelevant by the defender*, rather than being made a central component of TVC "risk" calculations.

The larger point is that the TVC product is not necessarily a useful measure of risk, nor a useful guide for allocating defensive resources.[2,3] Because risk depends on the attacker's resource allocation decisions, which in turn may depend on information not included in the TVC data (such as the damage inflicted in this example by different allocations of attacker resources to attack opportunities), the TVC formula in general does not provide information needed to predict risk.

## 5. DISCUSSION: PRACTICAL IMPLICATIONS FOR U.S. TERRORISM RISK MANAGEMENT

Infrastructure operators have always had to contend with disruptions from accidents, failures, and Mother Nature, but our current critical infrastructure systems, such as power grids, roads and bridges, aqueducts, fuel pipelines, and medical services, were built when the threat of malicious adversaries was of scant concern. Because private industry seeks to deliver goods and services at minimum cost to make a profit, decades of competition have led to infrastructure that is very lean and very fragile.

The attacks of 9/11 showed that terrorist threats are different from risks from industrial accidents or natural disasters. Terrorists act with purpose. They observe defensive preparations, evaluate alternate plans, and choose to act at a time and place where they can apply their limited resources to maximum effect. Ignoring this reality in PRA analyses is a recipe for ineffective risk management, as illustrated in the preceding simplified examples.

In the wake of 9/11, our government integrated 22 federal agencies into a single DHS to protect against a broad range of risks. From the outset, Congress urged DHS to work with our national laboratories to assess newly recognized risks. The laboratories proposed PRA, whose origins trace to assessing risks of nuclear reactor accidents. The claim that actions of terrorists can usefully be modeled as random, and characterized in essentially the same way as natural disasters or industrial accidents, makes PRA

a cornerstone of DHS critical infrastructure protection.

We believe there are two fundamental problems with this claim. First, *terrorists do not act randomly*, but seek and use information to exploit weaknesses in defenses and to increase the impact of attacks. Assessing terrorism risk via PRA requires someone who may have different information from the terrorists to *guess* the probability that a terrorist will attack some component, the probability that the attack will succeed, and the consequence of the attack. Although subject matter experts are in vogue in Washington, their elicited probabilities do not necessarily agree, and it is impossible to reproduce their assessments independently. This is not science. Worse, the guesses of our experts may be irrelevant (e.g., uncorrelated with, negatively correlated with, or uninformative about what terrorists will actually do) because terrorists act on the basis of *their* information, not ours. Although we have made these points here using simple hypothetical examples, empirical research also abundantly confirms the inability of our best experts to usefully predict what other nations, combatants, or political leaders will actually do: *expert probability judgments for such events tend to be slightly less useful than purely random guesses.*[9] Incorporating expert judgments into PRA assessments of terrorism risks has never been established as an empirically valid method for predicting these risks.

The second problem is that, even if PRA worked for a single target, it is completely inadequate (and can be dangerously misleading) for interconnected infrastructures consisting of thousands of targets *because PRA, as currently practiced, does not represent the function of the infrastructure at all*. For example, when considering our electric power grid, what matters is not whether a terrorist attacks an electric power substation but whether that attack could lead to a blackout similar to the one that occurred by accident during the 2003 Northeast power outage. To understand this, one must consider the system as a whole, not just its individual components.

A third problem is that PRA results may themselves serve to encourage or deter potential attackers who learn the results. There is no mathematical or practical reason to expect that attacker responses to the information in a PRA will in general confirm, rather than invalidate, the PRA estimates. Thus, PRA estimates that are acted on by intelligent adversaries may be self-defeating, whereas PRA estimates do not affect the behaviors of reliability systems or of natural hazards.

The 2007 update to the National Strategy for Homeland Security marked a shift in emphasis from "managing risk" to "*increasing resilience*," stating clearly that improving resilience—that is, designing systems so that they can withstand (continue to function) or recover (resume functioning) quickly from a wide range of disruptions—mitigates worst-case terrorist attacks, as well as natural disasters and accidents. Yet, DHS continues to promote a form of PRA that seeks to replace modeling of attacker decision rules (mapping what attackers know to what they do) with expert judgments of T, V, and C. DHS requires "risk matrices" summarizing threat versus consequence for grant proposals, even though any insight they might provide is unrelated to resilience (or, as far as we can tell, to sound or useful risk management[2]).

Alternatives to PRA, used by the U.S. military, evaluate infrastructure resilience by considering how infrastructure operators will respond to severe disruptions, including worst-case attacks based on adversary *capabilities*. Deciding what redundancy to place where to increase resilience requires rigorous analysis, and making these decisions robustly requires considering what an adversary *might* do based on his information, not guessing what we think he is likely to do, based on our information. The previous sections illustrate why this distinction matters.

The United States has a long way to go to increase the resilience of our critical infrastructures. Weaning industry from no-fault terrorism insurance (adopted after 9/11), and instead providing clear economic incentives to improve resiliency of their systems might reduce the attractiveness to terrorists of some obvious targets in our current infrastructure by credibly signaling that our infrastructure owners and operators have the incentive to perform defensive risk research and close holes that might otherwise repay terrorist research efforts. Conversely, the Terrorism Risk Insurance Act, recently extended to 2014, indemnifies private infrastructure owners from loss *after* a terrorist attack, thus reducing incentives to preemptively invest in enhanced resilience. This moral hazard may inadvertently increase the number of attractive targets, and make attacks more likely—the kind of predictable adaptive response that PRA does not illuminate. Shifting DHS attention from guessing where risks are highest to calculating where targeted investments would most increase infrastructure resilience might do much to reduce risks from intelligent, resource-constrained adversaries.

## 6. CONCLUSIONS

The examples in this article have illustrated fundamental differences between risk analysis for systems that do not perform their own risk research (or consider ours) before deciding what to do, and risks from terrorists, who may do both. Key differences are that attack probabilities for terrorists may be impossible to estimate accurately (if the estimates themselves affect the attack probabilities), and that probabilities estimated based only on what we know, rather than on what the attacker might know, can lead to poor risk management decisions, compared to those that would be made by considering what the attacker might know. No analogous limitation applies to natural hazards or engineered systems. We therefore recommend making robust risk management decisions that acknowledge that the attacker may know things we do not. Doing so can change risk management recommendations from protecting against attack probabilities implied by our own expert elicited probabilities to hedging against possible attacks based on what the attacker might know. We recommend shifting the emphasis of risk management from using experts to guess where risk might be greatest (e.g., using the formula "*Risk = Probability of attack × Probability that attack succeeds, given that it occurs × Consequence of a successful attack*") to calculating where targeted investments will most improve the resilience of critical infrastructures. The distinction between conditioning T, V, and C estimates on our own information and considering (and hedging against) the alternative possible information sets that an attacker might have is a fundamental difference between PRA as developed for natural and engineered systems and risk analysis that is useful for terrorism risks.

## REFERENCES

1. Ezell B, Bennett S, von Winterfeldt D, Sokolowski J, Collins A. Probabilistic risk analysis and terrorism risk. Risk Analysis, 2010; 30(4):575–589.
2. Cox A. What's wrong with hazard-ranking systems? An expository note. Risk Analysis, 2009; 29(7):940–948.
3. Cox A. Some limitations of "risk = threat × vulnerability × consequence" for risk analysis of terrorist attacks. Risk Analysis, 2008; 28(6):1749–1761.
4. Aumann R. Agreeing to disagree. Annals of Statistics, 1976; 4(6):1236–1239.
5. Bertsimas D, Brown D, Caramanis C. Theory and applications of robust optimization, 2007. Available at: http://users.ece.utexas.edu/~cmcaram/pubs/RobustOptimizationSV.pdf, Accessed on June 27, 2010.
6. Bordley R, Hazen G. Nonlinear utility models arising from unmodelled small world intercorrelatons. Management Science, 1992; 38(7):1010–1017.
7. Joyce J. The Foundations of Causal Decision Theory. Cambridge, UK: Cambridge University Press, 1999.
8. Laskey K, Lehner P. Metareasoning and the problem of small worlds. IEEE Transactions on Systems, Man and Cybernetics, 1994; 24(11):1643–1652.
9. Tetlock P. Expert Political Judgement: How Good Is It? How Can We Know? Princeton, NJ: Princeton University Press, 2005. Available at: www.newyorker.com/archive/2005/12/05/051205crbo_books1, Accessed on June 27, 2010.