Risk Analysis, Vol. 31, No. 2, 2011 DOI: 10.1111/j.1539-6924.2010.01563.x

Response

Making Terrorism Risk Analysis Less Harmful and More Useful: Another Try

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

Although Ezell claims that we only "repackaged in a new article the limitations identified by Ezell *et al.* of PRA in terrorism risk analysis," he neither addresses nor refutes any of our substantive technical points and examples, and his comments reflect a fundamental lack of understanding of our main ideas. We are therefore grateful for this opportunity to clarify our reasoning in light of his comments, as follows.

1. *Intelligence analysts cannot condition on knowledge that they do not have.* Ezell asserts that we, and the National Research Council, have overlooked what he calls the "obvious fact" that "intelligence analysts consider terrorist knowledge ... when assessing probabilities of attack." However, a major point of our article is that analysts *cannot* consider what terrorists know, when the terrorists know more than the analysts. The example in our Tables III and IV shows why attempting to have our "intelligence analysts consider terrorist knowledge ... when assessing probabilities of attack" may be worse than useless when the attacker knows things our analysts do not. (More formally, if what the attacker knows is represented by a finer information partition than what we know, then the attacker's action probabilities can be mathematically unmeasurable with respect to our information partition.) We believe that useful risk assessment requires acknowledging that attackers often have knowledge that we lack, and that we cannot produce valid probabilistic estimates for attacker actions that are informed by the knowledge we lack. Our best bet is not to pretend that there is some way ("obvious" or not) to accomplish this impossibility, but rather to proceed under the more realistic assumption that our experts cannot always form predictively useful attack probabilities when they have insufficient information.
2. *Simple examples suffice for proofs by contradiction.* Ezell's complaint that "simple toy examples offered by the authors lack proof that the methods would scale to do terrorism risk analysis" misunderstands the nature of these proofs. Showing via a small toy example that a general claim is mistaken suffices to disprove it. For example, Ezell *et al.*'s repeated claim that *Risk* = *Threat* × *Vulnerability* × *Consequence* (the "TVC" framework) provides a useful way to allocate defensive resources and priorities to combat terrorism is readily disproved by showing that the correlations among terms (which the formula omits) can reverse the rankings provided by the formula. A simple example suffices to make this point (Cox 2008), but the point is entirely general.
3. *The poor performance of expert judgments about future political and conflict events is well established by empirical studies.* Ezell writes that "The authors' implication that the U.S. intelligence community's judgment on our adversaries is less useful than purely random guesses is presented without proof and is aloof ...". In fact, we *do* prove (constructively, via our first example) that *any*

[1]Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA.
[2]Cox Associates, 503 Franklin Street, Denver, CO 80218, USA; TCoxDenver@aol.com.
*Address correspondence to Gerald G. Brown, Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA; ggbrown@nps.navy.mil.

judgment of an adversary's attack probabilities can be self-defeating, and hence strictly less accurate than a purely random guess, if the adversary uses the judgment to decide where to attack. More importantly, we stated that: "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* [Tetlock 2005]" (emphasis in original). The references we cite support our claims. Ezell objects that, for DARPA's Integrated Crisis Early Warning System (ICEWS), "the standards for accuracy are high—80% accuracy and 70% precision." But, our original point stands. The following warning, from an evaluation of the actual (not desired) performance of the ICEWS system, shows why the "high standards" to which Ezell refers do not translate to high predictive accuracy in practice.

> We early on discovered that we could come close to achieving our benchmark performance metrics [80% accuracy and 70% precision] using naïve models, which included lagged values of the EoI [Events of Interest] dependent variable, and a small number of policy-irrelevant correlates like size of population, presence or absence of mountainous terrain, and the like. Though such a naïve model may retrospectively achieve acceptable levels of overall performance, *it is useless for real world applications* . . . [M]odels that rely on dependent variable lags, as seen above, *provide only an illusion of high performance or goodness of fit*. A naïve model containing only lags of the dependent variable may score well on indicating the presence of some EoIs, but will miss every new onset and cessation of conflict, literally by definition. The illusory good performance metrics also operate as a disincentive to continue the search for more insightful, actionable crisis antecedents. (O'Brien 2010, emphases added)

The public derives no benefit from "the illusion of high performance." We suggest that, rather than retrospectively overfitting regression models to past data, and then misleadingly advertising that "the standards for accuracy are high—80% accuracy and 70% precision," it is more useful to recognize that such models do *not* perform well prospectively; that they are too often "useless for real world applications" (or nearly so) in reducing terrorism risks; and that we need to deal with this fact.

4. *Poor risk analysis threatens us all.* Ezell's passionate defense that "[m]any of these intelligence analysts risk their lives collecting data and making these difficult estimates" provides no guarantee that the resulting estimates are valid or useful. We believe they are not because the TVC framework does not ask the right questions or elicit relevant information for predicting risks, as explained in our paper. (It is also far from clear just what is life-threatening about making up numerical estimates for threat, vulnerability, or consequence numbers—a task frequently assigned to junior staff in various organizations competing for DHS dollars.) Of course, using a framework that is incapable of predicting how what we do will affect risk—as the discussion of Tables III and IV of our paper shows is the case for the TVC framework—may put the lives of *other* citizens at risk, by allocating defensive resources where they do little or no good, or even do harm, as shown in our examples. But the occupational hazards of guessing at "TVC" numbers are not self-evidently life-threatening, or even career-limiting, for those involved.
5. *Better risk analysis is easy.* We have tried, by exposition and example, to show that correctly applying existing techniques of applied probability, modeling, and optimization can provide useful insights for guiding effective allocation of limited defensive resources. Unfortunately, feeding expert judgments into the $Risk = Threat \times Vulnerability \times Consequence$ framework is *not* how to do it—for example, because the framework omits crucial information needed to predict and manage risks (such as correlations among terms, or bang-for-the-buck information about risk reductions achieved by implementing different subsets of possible actions); because its key terms are not well defined (Cox 2008); because our experts often lack the information needed to provide useful estimates, even if the concepts made sense; and because the framework has never been shown to produce good results (e.g., better than random). Better risk analysis is easy, but requires replacing the TVC approach with more useful analyses.

Ezell concludes with a brief account of the "substantial resources" that the U.S. government has

been investing in "research to develop and test new theories and approaches," and of a five-year debate "about PRA good vs. PRA bad." This misses the fact that whether PRA is good or bad—or, rather, useful or useless—*depends on how it is done*. Rather than continuing to devote "considerable resources" to creating simplistic, unvalidated, and low-performing "new theories and approaches" such as the TVC framework, we believe that the United States would be served far better by having competent risk analysts apply well-established techniques from operations research and risk analysis to model uncertainty and to robustly improve our infrastructure resilience and defensive resource allocations. Techniques of reliability analysis, causal modeling, simulation-optimization, robust and hierarchical optimization, and diversification and hedging of investment portfolios against uncertainties can demonstrably do far more than expert judgments and TVC calculations, for a fraction of the cost, to make our infrastructures more secure and resilient to natural and manmade attacks. To get there, however, we must stop denying that the TVC framework has fundamental logical and practical flaws that make it invalid for risk assessment. Ezell prefaces his paper with the aphorism: "All models are wrong, but some are useful." However, there is no guarantee that TVC models are useful in general, or usually, for correctly assessing attack risks or setting priorities, for reasons discussed in our paper and its references (e.g., that TVC typically omits the information that attackers use to decide when and where to attack). While the TVC framework offers simplicity, it confines analysis within a framework that does not best serve the operators and planners who are working diligently to protect our people and infrastructure. *It is time to adopt more useful analytics.* We must also stop pretending that our experts can produce predictively useful probabilities as inputs to the framework when they lack adequate information; stop pouring money into "research" to develop simplistic and flawed "new approaches" that do not address the fundamental limitations of the current approach; and start replacing it with sound predictive and prescriptive techniques, such as those listed above.

Only by vigilantly identifying, discussing, acknowledging, and rejecting flaws in approaches put forth under the name of "risk analysis" can professional risk analysts protect the long-term credibility and value of their profession. In our opinion, the TVC framework is a prime example of a currently fashionable approach that should not be used, and competent risk analysts should inform their clients of its deep technical flaws and use better analytics instead. Our reasons are fully explained in our paper and its references. We appreciate this opportunity to summarize some of them.

REFERENCES

Brown G, Cox A. How probabilistic risk assessment can mislead terrorism analysts. Risk Analysis, 2010; 31(2): 196–204.

Cox LA Jr. Some limitations of "*Risk = Threat × Vulnerability × Consequence*" for risk analysis of terrorist attacks. Risk Analysis, 2008; 28(6): 1749–1762.

Ezell B, Bennett S, von Winterfeldt D, Sokolowski J, Collins A. Probalistic risk analysis and terrorism risk. Risk Analysis, 2010; 30(4): 575–589.

Ezell B, Collins A. Letter to editor in response to Brown and Cox, "How probalistic risk assessment can mislead terrorism analysts." Risk Analysis, 2010; 31(2): 192.

O'Brien SP. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. International Studies Review, 2010; 12(1): 87–104. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2486.2009.00914.x/pdf.

Tetlock P. Expert Political Judgement: How Good Is It? How Can We Know? Princeton, NJ: Princeton University Press, 2005 (For an amusing and substantive review, see: www.newyorker.com/archive/2005/12/05/051205crbo_books1, accessed 27 June 2010.)