Three Useful Tools For Goal Setting: Judgment Analysis, The Taylor Russell Diagram, and the System Dynamics Model

Elise A. Weaver, Ph.D., Worcester Polytechnic Institute

1 Introduction

I would like to introduce you to three useful tools for goal setting. Each of these tools was developed for a different purpose and adapted for policy formulation over time. The first, judgment analysis, can be used to design a safety indicator that is based on expert judgment. The method of judgment analysis and the theory surrounding it, Social Judgment Theory (SJT), is about 50 years old (Cooksey, 1996; Stewart, 1988; Hammond, 1955;Brunswik, 1955). SJT was recently reviewed by Meacham (in press) as a potentially useful approach to fire risk problems.

The second, the Taylor-Russell diagram, can be used to decide an appropriate threshold for that safety indicator. It dates back to signal detection theory (Swets et al., 1991; Green & Swets, 1966) and has been nicely applied to policy formulation by Hammond (1996).

The final tool, the system dynamics model, can be used to investigate a regulatory structure to allow for changes to the indicator threshold over time and across contexts. The system dynamics model is a computer simulation tool developed by an electrical engineer looking at business strategy problems (Forrester, 1961). Like the Taylor-Russell diagram, this tool has been applied outside its original domain to great effect; these applications have been reviewed in depth by Sterman (2000).

In this paper, I will consider each tool in turn, presenting the scenario for its application, the method, and the outcome or deliverable that would result. An overall chart of how to use these tools is shown in Figure 1.

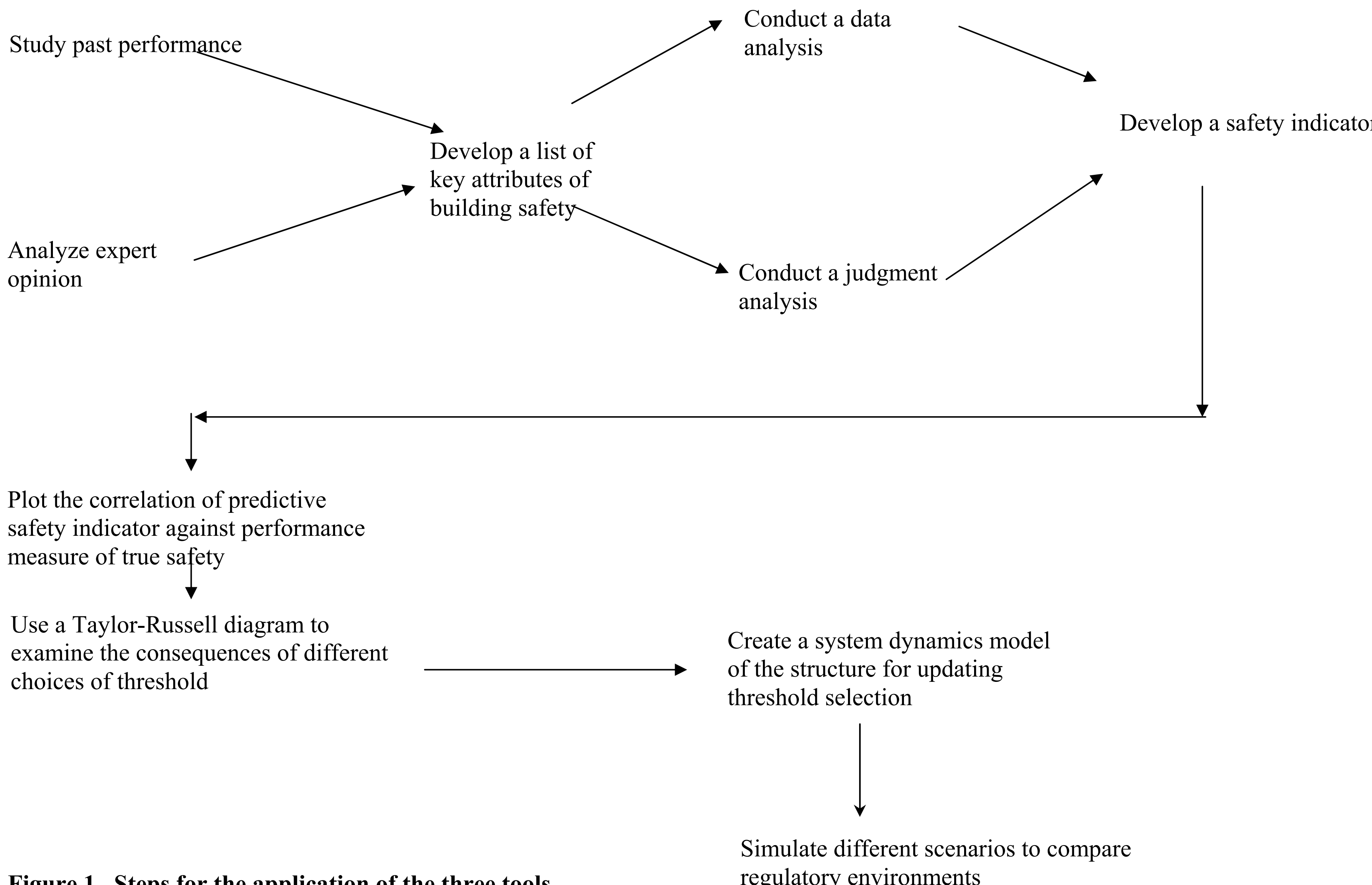

Figure 1. Steps for the application of the three tools.

2 Tool #1: Judgment Analysis

2.1 Scenario for Application

In the development of performance-based building codes, it might be helpful to have an index or indicator of safety. In order to construct an indicator for building safety, the first possible step is to select a list of safety-related attributes of a building and to learn how these have interacted and performed in the past.

If there were a way to represent the attributes of each building and how those buildings performed in a safety threatening situation, you could develop a safety indicator from simply conducting a regression analysis of performance with respect to these attributes. However, to use this approach, you would have to have adequate data. Without such data, it is possible to develop an indicator by analyzing the judgment of experts.

In this case, you would ask experts to judge building safety and extract their judgment policies. The resulting indicator would represent a weighted combination of factors important to expert judgments of safety, with the importance weights also determined from expert judgment.

If there were a consensus among experts, this might be a simple process. However, consider a scenario in which there was no clear expert consensus. Perhaps one expert thinks that egress during a fire is the most important factor and another thinks that implementation of fire prevention measures is the most important. Experts may differ on an attribute's importance, but agree on the technical implications of the different safety measures in addressing that attribute.

In this case, you would have a situation with a diversity of expert opinion, multiple attributes to consider, facts as to what different safety efforts will accomplish and values as to which of these measures is most important. The creation of such a safety indicator in such a context is the kind of problem suited to judgment analysis (Cooksey, 1996; Hammond, 1955; Brunswik, 1955).

2.2 Method

A judgment analysis comprises the following steps. First, you would interview experts about the attributes of the problem that constitute the cues to the judgment of safety, such as egress time, placement of sprinklers, etc. As mentioned earlier, while experts might differ on the importance of each of these, they may be able to agree on a list.

Second, you would sample from a set of real buildings to assemble a representative set of cases with attributes that occur together in plausible ways. (For more on representative design, consult Hammond & Stewart, 2001; Cooksey, 1996; Stewart, 1988; Brunswik, 1955).

Third, once you have a set of representative cases, you would have each expert rate each case on how safe he or she considers each building to be.

Finally, you would use regression analysis to model each expert's "policy" for making judgments of safety. Note that the reason for using regression analysis is that people are not always accurate about the judgment policies they hold or that they assume others hold. The use of statistics guarantees that you get actual expert judgments rather than the experts' guesses about how they make judgments. Where experts show differing judgment rules, it may be possible to see a small number of possible expert approaches. In this case, you might develop multiple indices representing different approaches or conduct research to assess the relative effectiveness of these differing judgment policies.

2.3 Outcome: An Experience-Based Safety Index

The outcome of such a judgment analysis would be insight into how diverse experts rate buildings on safety, as well as models of the judgment policies of clusters of experts. A policy maker could consciously select a compromise among such judgment policies to create an acceptable indicator of safety. This would be an experience-based safety index.

3 Tool #2: The Taylor Russell Diagram

Assume there is an indicator with a known success rate at predicting building safety, whether developed by analyzing past data or by analyzing expert judgment. The next step would be to select a safety threshold or cut-off point, such that buildings above the threshold would be considered "safe," and those below it would be considered "unsafe."

Unless the indicator is perfect, any threshold for a safe/not-safe decision will result in some buildings rated as safe when they are not (false positives) or some buildings being rated as unsafe when they are safe (false negatives). A Taylor-Russell diagram can be used to clarify the components of this situation (Hammond, 1996; Green & Swets, 1966).

A Taylor Russell diagram is presented in Figure 2. Along the horizontal axis are building safety indicator scores. Along the vertical axis are building safety performance scores. Each point represents a particular building. The quality of the indicator is shown by the spread of the points around a line angled at 45 degrees.

A lower threshold for an acceptable level of building safety may reduce costs, but there is a risk of constructing unsafe buildings (false positives). On the other hand, as you raise the threshold for the acceptable level of safety, you may impose unnecessary costs on the builder as you implement unnecessary safety measures (false negatives). Hammond (1996) describes this tradeoff between false positives and false negatives for any choice of threshold as the duality of error, made visual in a Taylor-Russell diagram.

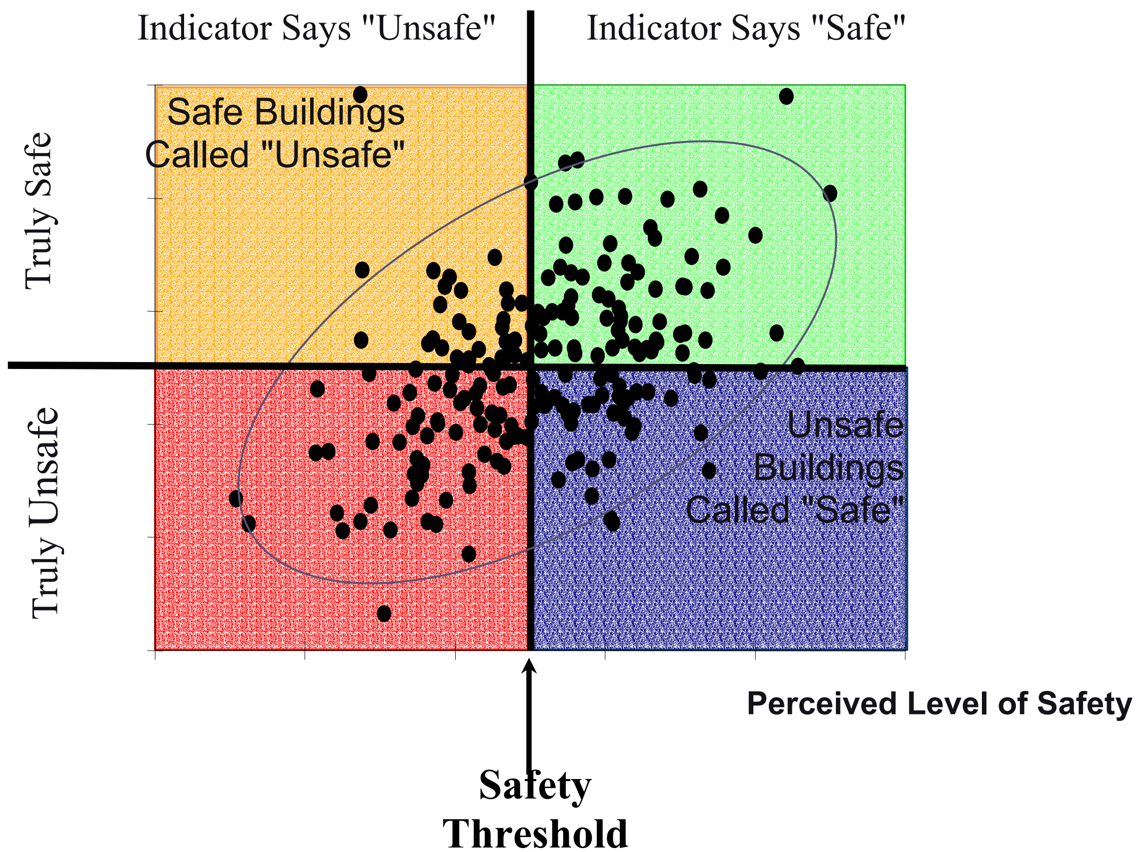

Figure 2. The Taylor-Russell Diagram

3.1 Method

To construct a Taylor Russell diagram, you plot individual buildings as points, with their predictive indicator score along the horizontal axis and their actual performance score along the vertical axis. Next, you set a threshold for true safety. Finally, you set a value-based threshold on your predictive indicator that results in consequences (false negatives and false positives) you are willing to accept.

By splitting the diagram into quadrants, you can see the number of true positives (i.e. "safe" on your indicator and safe in reality), true negatives (i.e. "unsafe" on your indicator and unsafe in reality), false positives (i.e. positive on your indicator for safety, but unsafe in reality), and false negatives (i.e. negative on your indicator for safety, but quite safe in reality) resulting from your selection of threshold, given the uncertainty in your indicator.

Note that the number of false positives and false negatives depends not only on the threshold chosen, but also on the degree of association between your indicator and the true safety rating. The correlation between these values represents the quality of the indicator. The indicator will show as much uncertainty as is currently present in the predictive science of safety assessment. Based on this diagram, individuals can discuss the numbers of false positives and/or false negatives they are willing to accept and, given the quality of the indicator, they can select an appropriate threshold.

3.2 Outcome

The Taylor Russell diagram is a way of envisioning simultaneously the connection between 1) the choice of threshold; 2) the effectiveness of an indicator; and 3) the resulting consequences.

4 Tool #3: The System Dynamics Model

4.1 Scenario for Application

Consider a situation in which you have an indicator, you have data on its performance, and you have chosen a threshold. At this point, you are hoping that the community at large will accept this threshold and that it will be stable. Unfortunately, you can't guarantee that your tradeoff of false positives and false negatives will be shared by others or that your particular threshold will always be appropriate.

In fact, if a highly salient event occurs for which there is a false positive, say, a building turns out unexpectedly to be unsafe, those constituents concerned with safety will pressure policy makers to move the threshold higher. As additional safety measures are implemented for buildings that were already safe enough, there will be diminishing returns on investment. The community concerned with the cost of buildings might then pressure the policy makers to lower the index.

Hammond (1996) described the potential for a threshold to oscillate as stakeholders respond to recent events and pressure policy makers to change a threshold. Weaver and Richardson (2002) designed a system dynamics simulation to analyze the systemic requirements for such an oscillation to occur.

In summary, as the selection of a threshold represents a value-based decision, it may become outmoded as societal values respond to recent events and as indicators improve. In order to have a responsive policy context that is protected from too wide an overreaction to recent events, it may be necessary to build in legal structures that regulate the threshold in an appropriately responsive manner. Rather than test out these structures in practice, it makes sense to simulate these phenomena in a computer simulation.

4.2 Method

First, create a model that includes the legal and political regulatory structures that affect the policy threshold (for an example, see Weaver and Richardson, 2002). This model would allow you to test different regulatory environments, including penalties for non-compliance, avenues for complaint, community values about outcomes, and the predictive quality of the safety indicator. Such a simulation might include stakeholder pressures to change the threshold, the quality of the index, the resulting false positives and false negatives and how the community would react to an unacceptable number of either of these.

A method that can be used to create such a computer simulation is to use system dynamics (Sterman, 2000). Such a simulation would include not only the information for which you have clear data, but also would embed the rich intuitions of experts, so that you could test what would happen under certain circumstances.

4.3 Outcome

The outcome of such a modeling effort would be a simulation that would allow policy makers to test the consequences of various threshold choices and a safety index that improves its predictive quality over time. In addition, it would allow them to set up and test a regulatory environment that would build in constraints against too sensitive a response to recent events, while guaranteeing the flexibility to update the model.

5 Summary

Three tools have been introduced, for use at different stages of the process of development of codes. The first, judgment analysis, could be used to develop an index of safety. The second, the Taylor Russell diagram, could be used to select an appropriate policy threshold for that index. The third, the system dynamics model, could be used to simulate a policy environment that would respond to unexpected events with appropriate adjustments.

6 References

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193-217.

Cooksey, R. (1996). *Judgment Analysis: Theory, Methods, and Applications*. NY: Academic Press.

Forrester, J. W. (1961). *Industrial Dynamics*. Cambridge MA: Productivity Press.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*(4), 255-262.

Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York: Oxford University Press.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The Essential Brunswik: Beginnings, Explications and Applications*. New York, NY: Oxford University Press.

Meacham, B. (in press). Decision making for fire risk problems: A review of challenges and tools.

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin McGraw-Hill.

Stewart, T. R. (1988). Judgment Analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human Judgment: The SJT View*. Amsterdam: North-Holland.

Swets, J. A., Getty, D. J., Pickett, R. M., D'Orsi, C. J., Seltzer, S. E., & McNeil, B. J. (1991). Enhancing and evaluating diagnostic accuracy. *Medical Decision Making, 11*(1), 9-18.

Weaver, E. A., & Richardson, G. P. (2002). *Threshold setting and the cycling of a decision threshold.* Paper presented at the System Dynamics '02, Palermo, Italy. http://www.wpi.edu/News/Transformations/2003Spring/pendulum.pdf