Thinking inside the box: A participatory, computer-assisted approach to scenario discovery

Benjamin P. Bryant, Robert J. Lempert *

RAND, 1776 Main St., Santa Monica, CA, USA 90407

ARTICLE INFO

ABSTRACT

Scenarios provide a commonly used and intuitively appealing means to communicate and characterize uncertainty in many decision support applications, but can fall short of their potential especially when used in broad public debates among participants with diverse interests and values. This paper describes a new approach to participatory, computer-assisted scenario development that we call *scenario discovery*, which aims to address these challenges. The approach defines scenarios as a set of plausible future states of the world that represent vulnerabilities of proposed policies, that is, cases where a policy fails to meet its performance goals. Scenario discovery characterizes such sets by helping users to apply statistical or data-mining algorithms to databases of simulation-model-generated results in order to identify easy-to-interpret combinations of uncertain model input parameters that are highly predictive of these policy-relevant cases. The approach has already proved successful in several high impact policy studies. This paper systematically describes the scenario discovery concept and its implementation, presents statistical tests to evaluate the resulting scenarios, and demonstrates the approach on an example policy problem involving the efficacy of a proposed U.S. renewable energy standard. The paper also describes how scenario discovery appears to address several outstanding challenges faced when applying traditional scenario approaches in contentious public debates.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Information technology's growing power offers a potential revolution in new tools and methods to support and improve human decision-making. A new approach to participatory computer-assisted scenario development we call *scenario discovery* [1,2] assists policy-makers and analysts in identifying policy-relevant scenarios by interactively applying statistical and data-mining algorithms to large databases of simulation-model results. Traditional scenarios provide a commonly used and intuitively appealing means to communicate and characterize uncertainty in many decision support applications, but can fall short of their potential, especially when used in public sector applications with diverse audiences [3,4]. Initial applications of scenario discovery, in particular in two high-impact public policy studies, suggest the approach may help overcome some limitations of the purely qualitative approaches to choosing scenarios. This paper provides the first complete description of scenario discovery, introduces diagnostic tools to evaluate the statistical significance of the scenarios suggested by the algorithms, and suggests how the approach can address several outstanding challenges faced by traditional scenario approaches when applied in contentious public debates.

Numerous definitions exist for scenarios and a vast array of methodologies are used to create them (see for instance reviews in [5–7]). Key approaches derive from the school *La Prospective* developed in France by Gaston Berger and Michel Godet, the Probabilistic Modified Trends school originally developed at Ted Gordon and Olaf Helmer at RAND, and the intuitive logics or

Anglo-American school that originated at RAND in the 1960s now often associated with the scenario groups at Shell Oil and the Global Business Network [7]. Scenario discovery builds on the intuitive logics school, the most commonly used among scenario practitioners [5]. One intuitive logics school definition describes scenarios as "internally consistent and challenging descriptions of possible futures" [8]. Scenario discovery provides a set of analytic tools that help decision makers identify such scenarios by focusing on those futures most important to the design of and choice among candidate robust strategies.

Scenarios from the intuitive logics school can offer important benefits when used to support decisions under conditions of deep (or Knightian) uncertainty.[1] Such scenarios often describe uncertainty about the future in a way that decision makers find easy to understand, can reduce overconfidence, and encourage individuals and groups to reflect on a broader range of future possibilities than they would otherwise consider. If presented as possibilities rather than firm predictions, such scenarios can prove psychologically less threatening to those holding different worldviews and thus make it easier for decision makers to consider inconvenient or contentious futures [11].

A small evaluative literature[2] provides empirical evidence that while scenarios and the process of developing them can in some cases produce these claimed benefits, they often fail to do so, particularly when applied for groups with diverse interests and worldviews [4]. In many cases observers see the choice of scenarios as arbitrary or highly subject to the particular interests and values of those choosing them [3]. Observing an exercise by a government agency in the Netherlands, Van 't Klooster and van Asselt [12] noted three distinct and conflicting interpretations of the scenario axes developed by the group. The authors conclude that the diffuse and heterogeneous nature of public agencies' objectives and interests may make it impossible for them to come to consensus about the meaning of scenario axes. Comparative analyses also suggest that many scenario processes systematically exclude surprising or paradoxical developments (the so called "Black Swans") as inconsistent or logically impossible [13]. Van Notten et al. [14] compare twenty-two scenario studies, some using simulation models and others entirely qualitative, and find that none of the model-based exercises included discontinuities in system behavior. Finally, it often proves difficult to include probabilistic information in a traditional scenario analysis without contaminating the simplicity and sense of possibility, as opposed to prediction, that makes the scenarios useful in the first place [15].[3]

These difficulties flow in large part from the challenge of choosing a small number of scenarios to summarize the full breadth of uncertainty about the future. A set of scenarios cannot contain more than a handful of members and remain intelligible to decision makers, who nonetheless may face a multiplicity of potentially interesting, plausible, and important futures. Schwartz [16] provides the classic exposition of how the intuitive logics (also called scenario axis) approach aims to reduce many futures to a manageable few. Developers first identify the decision the scenarios are meant to inform. They next use their expertise to rank key factors in the external environment that may affect the decision according to two criteria – their importance to the success of the decision and the degree of uncertainty surrounding them. Developers select the most important driving forces as the axes of the scenarios that define a small set of (often four) scenarios. Developers craft a "scenario logic" for each case, weave a narrative around the logic, and check the narrative for internal consistency. This approach can be applied qualitatively, or with the support of computer simulation models. But even in the latter case, developers often choose the scenarios using their intuition and some variant of the scenario axis approach. Only after they have identified the scenarios will developers run the models to create projections that illuminate the scenarios' logic. The evaluative literature suggests that while this traditional scenario axis process often succeeds admirably with small groups of clients whose interests and concerns are known well by the scenario developers, it can fail to summarize a multiplicity of plausible futures in a way that generates consensus and common understanding among participants in broad public debates with diverse interests and values.

Scenario discovery aims to address this challenge by employing a key concept and an innovative assemblage of statistical tools for implementing the concept. The concept defines scenarios as a set of future states of the world that represent vulnerabilities of proposed policies. We interpret vulnerability broadly, referring to the states of the world where a proposed policy may fail to meet its performance goals as well as those where a policy's performance deviates significantly from the optimum policy in that state of the world.[4] Scenario discovery implements this concept by using statistical or data-mining algorithms to find easy-to-interpret, policy-relevant regions in the space of uncertain input parameters to computer simulation models. These simulations are run many (hundreds to millions) of times over a space defined by combinations of values for uncertain input parameters. Some policy-relevant criterion (e.g. the cost of the policy, whether some environmental goal is met) is used to distinguish a subset of the cases, typically by applying a threshold to the model's outputs. Statistical or data-mining algorithms applied to the resulting multi-dimensional database then find simple descriptions of the input space that best predict these cases of interest. These regions of input space can then be usefully considered as scenarios for decision analytic applications, and the input parameters used to define the regions become the key driving forces for these scenarios. Scenario discovery offers a quantitative implementation of the scenario axis approach that appears to address these difficulties and, in particular, improves the efficacy of scenarios for diverse audiences in public debates.

---

[1] We define deep uncertainty as the condition where parties to the decision do not know or do not agree on the system model relating actions to consequences or the prior probability distributions for input parameters to these system models [9,10].

[2] The term *evaluative scenario literature* describes studies that systematically attempt to evaluate scenario approaches with methods such as comparative case studies, ethnographic or other detailed observations of the impact of scenarios within organizations, laboratory studies of the cognitive impacts of scenarios, and structured comparisons of the performance of organizations that do and do not use scenarios. See [4] for a recent survey. This literature largely focuses on the intuitive logics school.

[3] Parson et al. [3] summarizes the debate in the climate change community about the efficacy of including probabilistic information with scenarios.

[4] The former represents an absolute performance measure and the latter a regret measure.

Two practical examples illuminate both scenario discovery and its potential benefits. One recent study [17] evaluated alternative policies considered by the United States Congress while debating reauthorization of the Terrorism Risk Insurance Act (TRIA). TRIA provides a federal guarantee to compensate insurers for losses due to very large terrorist attacks in return for insurers providing insurance against attacks of all sizes. Congress was particularly interested in the cost to taxpayers of alternative versions of the program. RAND used a simulation model to project these costs for various TRIA options for each of several thousand cases, each representing a different combination of 17 deeply uncertain assumptions about the type of terrorist attack, the factors influencing the pre-attack distribution of insurance coverage, and any post-attack compensation decisions by the Federal government. Scenario discovery demonstrated that the existing TRIA program would *always* cost the taxpayers the same or less than proposed alternatives except in cases where the losses from a terrorist attack exceed some very large monetary threshold (roughly $40 billion), largely independent of the value of the other uncertain parameters. The resulting scenario and the analysis based on it seem to have influenced a diverse, divisive public debate,[5] providing benefits that may have been difficult to achieve with traditional scenario or other decision analytic approaches. Using scenarios made it easier to consider a wide range of assumptions about difficult-to-predict events – in particular any post-attack compensative decisions of a future Congress – thereby enabling this study to reach different conclusions than those of the Congressional Budget Office and Treasury Department. The systematic, analytic, and reproducible scenario discovery approach limited any perceived arbitrariness in the choice of scenarios and made the resulting analysis more resilient to challenges in the public debate.[6]

A second scenario discovery analysis helped Southern California's Inland Empire Utilities Agency (IEUA) reduce the vulnerability of its long-range water management plans to potential climate change [19,20]. In 2005 the agency, which supplies water to a fast growing population in an arid region, completed a legally mandated plan for ensuring reliable water supplies for the next twenty-five years. This plan did not however consider the potential impacts of future climate change. RAND used a simulation model to project the present value cost of implementing IEUA's current plans, including any penalties for future shortages, in several hundred cases contingent on a wide range of assumptions about six parameters representing climate impacts, IEUA's ability to implement its plan, and the availability of imported water. A scenario discovery analysis identified three key factors – a 8% or larger decrease in precipitation, any drop larger than 4% in the rain captured as groundwater, and meeting or missing the plan's specific goals for recycled waster water – that, if they occurred simultaneously, would cause IEUA's overall plan to fail (defined as producing costs exceeding by 20% or more those envisioned in the baseline plan). These factors became the key driving forces defining what was called the *Dry, Flashy, Low-Recycling* scenario. Similarly to the TRIA example, this scenario and resulting analysis provided benefits difficult to achieve with other approaches, allowing IEUA's managers, constituents, and elected officials, who did not all agree on the likelihood of climate impacts, to understand in detail vulnerabilities to their plan and discuss ways to ameliorate those vulnerabilities.[7]

The analytic tools described here warrant the term "scenario" in their name because they support a collaborative process among people and computers designed to facilitate, rather than dictate, human creativity and problem solving. As widely noted, the *process* of developing scenarios often proves at least as important to decision-makers as the scenarios themselves (see, for example, discussions in [3,21]). Scenario discovery represents a participatory, computer-assisted process that supports Robust Decision Making (RDM) [2,9,10], a quantitative decision analytic method that uses available information (such as that contained in computer simulation models), not to improve predictions of a deeply uncertain future, but rather to help decision-makers craft strategies that can more effectively achieve their goals in the face of these uncertainties.

In brief, RDM first helps decision makers characterize the vulnerabilities of a series of candidate strategies and then helps these decision makers identify and choose among alternative means for ameliorating the vulnerabilities. Scenario discovery facilitates this first step, concisely summarizing a wide range of future states of the world in a way that helps decision makers more clearly understand the strengths and weaknesses of candidate strategies. In the TRIA example above, this process generated a scenario with a single key driving force – any terrorist attack with losses greater than $40 billion. In the IEUA example, the process identified a scenario defined by a combination of three driving forces – any future with a large drop in precipitation, even a small decrease in rain captured as ground water, and any failure to meet the recycling goals in the agency's plans. In both cases scenario discovery, exploiting cognitive mechanisms discussed in Section 4, helped to generate insights that facilitated the design of more effective strategies and the choice among them. Overall, scenario discovery and the robust decision making process of which it is often a part aims to harness the increasing power of information technology as a tool for facilitating human insight and imagination. Building on the exploratory modeling concepts originally described by Bankes [22], these tools help the computer move beyond its role as an elaborate calculator and serve instead as a "prosthesis of the imagination."

The next section of this paper describes the analytic underpinnings of scenario discovery, including measures of merit and diagnostic tools for evaluating scenario quality. The third section demonstrates the method on an example policy problem involving the efficacy of a proposed U.S. renewable energy requirement. The fourth section situates scenario discovery within the larger scenario literature, in particular describing how the method can successfully inform decisions by drawing on some of the cognitive benefits of traditional qualitative scenario methods while addressing some of these methods' shortcomings. The final section concludes by suggesting potential improvements to the tools currently used to implement scenario discovery.

---

[5] The results of this analysis were quoted on the floor of the U.S. Senate (Congressional Record, Sen. Christopher Dodd, Nov 16, 2007).

[6] One author who opposed the study's findings called its arguments "insidious" but was otherwise was unable to find fault with the analysis or its choice of scenario [18].

[7] See Congressional testimony by IEUA General manager Rich Atwater (Jan 29, 2008).

2. Analytic underpinnings of scenario discovery

Scenario discovery aims to summarize sets of plausible future states of the world that illuminate key vulnerabilities in proposed policies and to describe these scenarios in a manner useful for decisionmakers and stakeholders. Implementing scenario discovery involves four steps shown in Fig. 1. The following discussion emphasizes the analytic underpinnings needed by analysts to implement the method and evaluate the quality of the results. Most users will, however, be more interested in the results and the concepts behind them, rather than the details described here.

Step 1 involves specifying a sampling design for computational experiments and an output criterion that distinguishes the cases of interest. Step 2 involves applying one or more algorithms to the resulting database to identify candidate scenarios that provide a good description of these cases of interest. Step 3 involves evaluating these scenarios using various diagnostic tools. The process aims to be participatory and iterative. As will be described below, the algorithms present options and tradeoffs to users, who use this information to guide the choice of scenarios. Users may also revise their earlier choices based on statistical assessments of scenarios' quality. We next describe each step in more detail.

2.1. Generate data from a simulation model (Step 1)

Scenario discovery begins with one or more computer simulation models $y = f(s,\mathbf{x})$ that relate policy makers' actions s to consequences of interest y, contingent on a vector $\mathbf{x}$ representing a particular point in an M-dimensional space of uncertain model input parameters. For instance, in the TRIA example from the previous section, the simulation reports the proposed legislation's future savings or costs to taxpayers contingent on assumptions about future terrorist attacks and the response of Congress and the insurance industry. Congress was interested in the conditions under which the legislation would impose a net cost on taxpayers. In the water management example the simulation reported the costs of implementing IEUA's current plan contingent on assumptions about climate change, the availability of imported water, and the agency's ability to implement its plans. IEUA's leaders and constituents were interested in the conditions where the plan would impose higher than expected costs.

To implement scenario discovery we sample the system model of interest with an experimental design over the uncertain inputs $\mathbf{x}$ while holding constant a candidate strategy s. Using some policy-relevant criteria, we choose some threshold performance level Y^I that defines a set of cases of interest $I_s = \{\mathbf{x}^I | f(s,\mathbf{x}^I) \geq Y^I\}$ or $\{\mathbf{x}^I | f(s,\mathbf{x}^I) \leq Y^I\}$, contingent on that strategy. For instance, in the TRIA example I_s is the set of cases where the legislation imposes net costs on taxpayers and for IEUA I_s is the set of cases where the agency's costs exceed 20% or more of those assumed in the current plan.

A variety of alternative types of experimental designs, such as full factorial or Monte Carlo, can be used to construct the N-point experimental design resulting in a dataset $\{y_i, \mathbf{x}_i\}$, i = 1,...,N. In past work, we have found that the Latin Hypercube (LHS) provides a convenient experimental design for scenario discovery because it provides an efficient sample of a model's behavior over the input space. LHS is a randomized experimental design based on the higher dimensional generalization of a Latin Square. In traditional regression it proves to have superior variance reduction properties relative to other sampling designs [23]. Our experience to date suggests LHS is more useful for scenario discovery than other standard sampling methods.[8] Note that no probabilistic information is included at this stage of the analysis because the sample is only intended to explore the full range of the model's behavior. Probabilistic information can be incorporated at later stages by placing a joint probability density function $\rho(x)$ over the cases in the experimental design.

2.2. Algorithms Help Users Identify Candidate Scenarios (Step 2)

Scenario discovery next uses statistical or data-mining algorithms to identify the combinations of values of uncertain model input parameters that best predict the set of interesting cases I_s. In our current work, we characterize these regions of input parameter space with algorithms that generate multi-dimensional "boxes." Specifically, we seek to describe the set I_s using one or more sets of limiting constraints, $B_k = \{a_j \leq x_j \leq b_j, j \in L_k\}$, on the ranges of a subset of the input parameters $L_k \subseteq \{1, ..., M\}$. Input parameters not in L_k remain unconstrained for B_k. We call each individual set of simultaneous constraints B_k a box and a set of boxes B a box set.

We interpret each box as a scenario and the set of boxes as a set of scenarios. As described in Section 4, these boxes bring to decision support applications many of the same attractive features as do scenarios from the intuitive logics school.

Note that it often also proves useful to consider all those states not in any box as a scenario. For instance, a single box might represent a scenario where a policy has high costs. All the other states might represent the scenario where the policy has reasonable costs. In addition, the scenario discovery algorithms will in some cases yield boxes that overlap. To date, we have for convenience considered such boxes as distinct scenarios, though they might be more usefully viewed as a single scenario with a shape poorly described by a box. Section 5 will mention improvements to the current algorithms that could address such situations.

2.2.1. Measures of Merit for Scenario Quality

Choosing among box sets requires measures of the quality of any box and box set. The traditional scenario planning literature emphasizes the need to employ a small number of scenarios, each explained by a small number of "key driving forces," lest the scenario users become confused or overwhelmed by complexity [16]. In addition to this desired simplicity, the quantitative

---

[8] This discussion assumes continuous inputs, but sampling designs and resulting scenario definitions can accommodate input of mixed types.
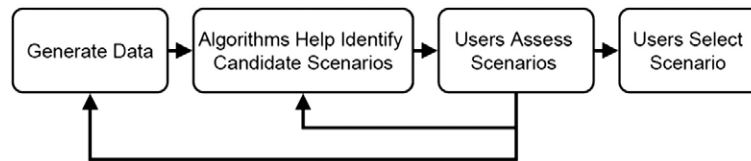
Fig. 1. Steps of Scenario Discovery.

algorithms employed here seek to maximize the explanatory power of the boxes, that is, their ability to accurately differentiate among the cases of interest and the other cases in the database.

These characteristics suggest three useful measures of merit for scenario discovery. To serve as a useful aid in decision-making, a box set should capture a high proportion of the total number of policy-relevant cases (high *coverage*), capture primarily policy-relevant cases (high *density*), and prove easy to understand (high *interpretability*). We define and justify these criteria as follows:

Coverage measures how completely the scenarios defined by box set B capture the cases of interest (I_s) and is analogous to the "sensitivity" or "recall" in the classification and information retrieval fields. With binary output, coverage is simply the ratio of the total number of cases of interest in the set of scenarios B to the total number of cases of interest, that is,[9]

$$Coverage = \sum_{\mathbf{x}_i \in B} y_i' / \sum_{\mathbf{x}_i \in \mathbf{x}^I} y_i' \tag{1}$$

where $y_i' = 1$ if $\mathbf{x}_i \in I_s$ and $y_i' = 0$ otherwise.

Decision makers should find this coverage measure important because they would like the scenarios to explain as many of the cases of interest as possible.

Density measures the purity of the scenarios and has analogues with "precision" or "positive predictive value" in other fields. With binary output, density can be expressed as the ratio of the total cases of interest in a scenario to the number of cases in that scenario, that is,

$$Density = \sum_{\mathbf{x}_i \in B} y_i' / \sum_{\mathbf{x}_i \in B} 1 \tag{2}$$

Decision makers should find this measure important because they would like each scenario to be a strong predictor of the cases of interest.

Interpretability measures the ease with which decision makers can understand a box set and use it to gain insight about their decision analytic application. This measure is thus highly subjective, but we can nonetheless approximate it quantitatively by reporting the number of boxes in a box set and the maximum number of model input parameters constrained by any box, equivalent to the size of the set L above. Based on the experience reported by the traditional scenario planning literature [16], a highly interpretable box set should consist of on the order of three or four boxes, each with on the order of two or three constrained parameters.

An ideal set of scenarios would combine high density, coverage, and interpretability. Unfortunately, these measures generally compete, so that increasing one typically comes at the expense of another. For instance, coverage and density are in tension, so that increasing coverage often decreases density. Increasing interpretability by constraining fewer parameters can increase coverage but typically decreases density. Achieving the same goal by reducing the number of boxes will likely decrease either density or coverage. For a given dataset, these three measures define a multi-dimensional efficiency frontier. The process in Fig. 1 envisions that users interactively employ a scenario-discovery algorithm to generate alternative box sets at different points along this frontier and then choose that set most useful for the decision analytic application.

2.2.2. Scenario Discovery Algorithms

To our knowledge, no existing algorithm performs tasks identical to that required for scenario-discovery. The interpretability measure poses requirements distinct from most other applications. In addition, while many algorithms seek to maximize coverage, which is equivalent to the success-oriented quantification of the Type II error rate, few consider density, which is related to but does not neatly correspond to the Type I error rate because the denominator in Eq. (2) refers to the set of scenarios rather than the overall dataset.

Among existing algorithms, the scenario-discovery task appears most similar to classification and bump-hunting approaches. For datasets with binary output, classification algorithms partition the input space into regions of high purity, that is, regions that contain predominantly one output class. Bump-hunting algorithms find regions of input space with a comparatively high mean output value. To generate candidate scenarios, we generally employ a version of the PRIM (Patient Rule Induction Method) bump-hunting algorithm originally developed by Friedman and Fisher [24]. PRIM is useful for scenario discovery because it is highly interactive, presents multiple options for the choice of scenarios, and provides visualizations that help users balance among the three measures of scenario quality: coverage, density, and interpretability. The Appendix describes PRIM's operation in more

---

[9] The definitions presented here assume a sufficiently dense and space-filling experimental design. Non-uniform sampling would require applying generalized definitions based on volume integrals.

detail. Lempert, Bryant, and Bankes [25] also tested the ability of the classification algorithm CART (Classification and Regression Tree) to perform scenario discovery. CART appears to generate similar results as PRIM, but with less user interactivity and more work required by the analyst to create box sets with high interpretability.

2.3. Users Assess Scenarios with Diagnostics (Step 3)

Our previous work [25] applied PRIM and CART to datasets with regions of known shape to test the algorithms' strengths and weaknesses for scenario discovery. These tests suggest that both algorithms can perform the scenario discovery task even for relatively complex shapes, but that under some conditions they make several types of errors. In particular, PRIM may needlessly slice off the end of a parameter's range, incorrectly suggesting that a proposed policy may prove sensitive to even a small variation in some parameter. The potential for such errors is troubling because a policy can be truly sensitive to small variations, as was the case, for instance, with IEUA where scenario discovery properly revealed that the agency's plan was very sensitive to any change in the amount of rain captured as groundwater. In addition, when applied to low-dimensional shapes in high-dimensional data PRIM may erroneously constrain extraneous parameters that do not in fact predict the cases of interest.

Such potential errors highlight the importance of using diagnostic tools to evaluate the statistical significance of the parameter constraints proposed by the scenario discovery algorithm. We propose that users employ a quasi p-value test and resampling test for this purpose. These techniques, commonly employed in the field of statistical learning to diagnose the quality of models fit to data, prove appropriate because the PRIM errors result from the finite and stochastic sampling of the LHC experimental design. Given this stochasticity, the scenario definitions can be considered statistical models with potentially nonzero bias and variance about the true model f. These two tests help detect the errors described above by estimating the probability that any particular parameter constraint is due to chance and by examining the extent to which the scenario definition varies over multiple samples of the original data. These tests are applied only to individual boxes, rather than on the entire box set.

2.3.1. Resampling test

This diagnostic tool evaluates a scenario definition by assessing how frequently the same definition arises from different samples of the same database. The resampling test runs the algorithm on multiple subsamples of the original dataset and notes which of the parameter constraints consistently emerge as important in the resulting scenario definitions. PRIM complicates automation of this technique because the algorithm is fundamentally interactive, requiring the user to select from a large number of options with different combinations of coverage and density. We thus generate two sets of "reproducibility statistics" – one in which the algorithm generates a scenario matching as closely as possible the coverage of the original box, and one in which it matches the density. These two criteria will often but not always generate identical results. Ideally for both criteria the parameters constrained in the initial scenario definition will also be constrained in 100 percent of the samples, while the unconstrained parameters will remain so in all the samples.

2.3.2. Quasi-p-value test

This diagnostic tool uses what is essentially a p-value test to estimate the likelihood that PRIM constrains some parameter purely by chance. Consider a single box β within box set B, defined by limiting constraints on parameters in the set L_β, and which contains H high-value ($y_i' = 1$) cases out of a total of T cases. To compute this quasi-p-value consider the box β_{-j} defined by constraints on all parameters in L_β except parameter $x_j \in L_\beta$. This box contains T_{-j} total cases and H_{-j} cases of interest, with $T_{-j} \geq T$ and $H_{-j} \geq H$. We then consider the null hypothesis that the cases of interest within β_{-j} are distributed among all cases in β_{-j} according to a binomial distribution with $p(1) = H_{-j}/T_{-j}$. The "qp-value" test thus answers the question: what is the probability that T points drawn from the above binomial distribution would have H or more high valued points? When the ratio H_{-j}/T_{-j} is close to H/T this number is high, the additional contribution of parameter r_j is low, and thus possibly due to chance. The opposite is the case when H/T is much larger than H_{-j}/T_{-j}.

We call this a quasi-p-value test, because we are aware that, contingent on sampling, it is not an entirely accurate model of the system, since it does not take into account spatial proximity and its interaction with whatever algorithm is defining the box. Nevertheless, the relative magnitudes of the qp-values provide useful information for comparing parameter relevance, as we illustrate in Section 3.

These diagnostic techniques, combined with the measures of coverage, density and interpretability, help users achieve a more complete understanding of the scenarios and their ability to characterize the cases of interest in the database.

3. Scenario discovery applied

This section now demonstrates scenario discovery on a real policy problem, in the process illustrating the utility of both the approach and the various tools developed to aid the task.

For this example, we consider the potential impacts of a renewable energy requirement in the United States. We use a simulation model developed by Toman, Griffin, and Lempert [26] [henceforth TGL] that examines the benefits and costs of a "25 by 25" policy requiring 25 percent of electricity and motor fuels to be derived from renewable sources by 2025. The model projects the greenhouse gas emissions and economic costs of such a policy contingent on a wide variety of assumptions about future cost and performance of various technologies, consumer behavior, and other factors. The model suggests that in many cases, the $25 \times x25$ policy would likely produce significant social benefits – including reducing emissions of greenhouse gases and U.S. dependence on foreign oil – but in some cases may be extremely costly. Griffin [27] conducts a full robust decision making (RDM)

analysis [2,9,10] with this model, employing scenario discovery to identify vulnerabilities of the 25×25 policy (e.g. high cost cases), then using this information to modify the policy to make it more robust. RDM analyses often employ scenario discovery methods in this manner to find the vulnerabilities of proposed strategies [1,2].

To demonstrate scenario discovery this paper focuses on assessing one particular vulnerability analyzed in TGL, which forms the starting point for Griffin's RDM analysis: Under what conditions does the 25×25 policy results in unacceptably high economic costs? In particular, this example demonstrates visualization and diagnostic tools not available for Griffin's and other previous work. Main question of analysis

3.1. Generate Dataset

The TGL simulation model has nine uncertain input parameters, as shown in Table 1 along with their allowed range of variation. As described in detail in TGL, these ranges were derived from high and low estimates for each parameter in the literature. For instance, the parameter *Low-cost biomass supply* represents the total amount of inexpensive biomass available in the United States in 2025 for production of biofuels. TGL used Aden *et al* [28] and Solomon *et al* [29] to provide the high and low estimates, respectively. These two studies represent the wide range of expert opinion relevant to the renewable requirements debate. We have assumed the input parameters in Table 1 vary independently across their ranges. If they were correlated we might use a transformed set of inputs, add a parameter expressing the degree of correlation, or address any correlations at a later stage when the analysis considers the relatively likelihood of various scenarios.

3.1.1. Latin Hypercube sample

We use a Latin Hypercube sample (LHS) to create an experimental design over the space defined by these nine uncertain model input parameters. Running this sample through the simulation creates a database that explores the implications all combinations of the full range of expert opinion about the values of the nine uncertain parameters in Table 1.

We generated a 1000-point LHS experimental design over this space, which required approximately 30 minutes of CPU time. The design yields a database of 882 cases, because the model fails to converge in 118 cases.[10] Each record in the database has ten entries, nine with a value for each of the uncertain input parameters and one with the resulting value of the model output of interest, that is, the cost to consumers of the 25×25 policy contingent on the values of the nine inputs. Fig. 2 shows the distribution of costs over these 882 cases.

3.1.2. Specify Outputs of Interest

We next characterize the output values in this database, differentiating between the cases of interest with unacceptably high costs for the 25×25 policy and the other cases with lower costs. Some policy studies offer a natural boundary between such cases dictated by some externally imposed standard, such as a budget constraint that policies do or do not meet. In other studies, stakeholders may have well-defined preferences that help distinguish the cases of interest. For instance, Griffin [27] uses data on consumers' willingness to pay to define high costs outputs from the TGL model. This present example lacks either a natural boundary or well-defined stakeholder preferences, so for convenience we define high cost to be any case that exceeds the 90th percentile cost, at approximately $78 billion. The resulting database thus assigns a value of unity for cases whose costs fall above the 90th percentile and zero otherwise. Users could, if desired, explore the sensitivity of the resulting scenarios to this choice of threshold.

3.2. Identify Scenarios

We now use the PRIM algorithm to concisely summarize the combinations of uncertain model input parameter values that best predict these high cost cases. As in our previous work, we employ the same basic algorithm as that introduced by Friedman and Fischer [24], enhanced with a newly implemented software toolkit that provides a variety of useful features for scenario discovery.[11] In particular, this toolkit directly reports the coverage measure in Eq. (1), replacing the support measure provided by Friedman and Fischer's original code. The new toolkit automatically calculates qp-values, the resampling statistics, and the visualizations shown below. In addition, the new toolkit provides the option to display coverage-density tradeoff curves calculated using both PRIM's peeling and pasting steps (described in the Appendix). The original PRIM software generated tradeoff curves using the peeling step, only adding pasting to candidate scenarios selected by the user.

Fig. 3 displays several coverage-density tradeoff curves generated by the scenario discovery toolkit from the database described in Section 3.1. The filled-in circles represent the coverage and density combinations identified by PRIM when it is free to use any number of parameters. The figure also shows density-coverage combinations achieved while holding constant the total number of parameters used. These latter points are found by dropping the least important parameters from more complicated boxes. They thus do not represent a complete or optimal search, but do serve to illuminate tradeoffs between the scenario quality measures of coverage, density and interpretability.

---

[10] The TGL simulation uses marginal cost curves with step functions and may fail to converge near the steps. Excluding these cases does not appear to bias the remaining samples.

[11] The scenario discovery toolkit is implemented in the free R statistical computing environment and is available at http://cran.r-project.org/web/packages/sdtoolkit/index.html.

Table 1
Range of values considered for uncertain model input parameters.

Uncertain Model Input Parameter	Low	High	Units
Biofuel Production Cost	67	134	$ per unit input
Low-cost biomass supply	450	1000	millions of tons
Feedstock supply distribution	0	1	pessimistic to optimistic
Biofuel yield	80	100	gallons per ton
Oil supply elasticity	0.2	0.6	
Transportation demand elasticity	-0.2	-0.8	
Electricity co-product	0	2	kWh per gallon
Shift in oil supply curve	-10	10	% change
Biomass backstop price	90	200	$ per ton

A box representing a perfect scenario would be defined by constraints on only one or two parameters and would lie in the upper right-hand corner with 100% coverage and 100% density, and thus capturing all the cases of interest and excluding all the other cases. Since such a box is not available, users must choose one with the combination of coverage, density, and interpretability that best supports their decision application. In general, dimensionality increases with density and decreases with coverage, and both decrease with interpretability. For the purposes of this example, we initially consider Scenario A, which uses four parameters to achieve 79% coverage and 73% density. After evaluating this scenario however, users may wish to modify this choice, possibly improving interpretability by dropping parameters deemed less important or choosing another scenario entirely.

Fig. 4 displays the combinations of parameter values PRIM uses to define Scenario A.[12] The scenario includes potential future states of the world where *transportation demand elasticity* and *biomass backstop price* are at the upper half of their ranges, *biofuel production costs* is at anywhere above the lowest end of its range, and the *supply of low-cost biomass* is at the lower half of its range. Overall, 73% of the cases in the dataset that meet these four conditions have high costs (i.e., 73% density). Of all the high-cost cases in the dataset, 79% meet these four conditions (i.e. 79% coverage).

As shown in Fig. 4, PRIM also reports each parameter's marginal contribution to explaining the high cost cases. With no parameters constraints the box would have 10% density and 100% coverage, since we have defined 10% of the cases in the database to have high-cost. Using only one parameter – *transportation demand elasticity* – provides a scenario that captures 97% of the high cost cases with a density of 25%. A two-parameter scenario that adds PRIM's constraint on the *biomass backstop price* increases the density from 25 to 46% with only a 13% drop in coverage. The fourth parameter yields diminishing returns. Adding PRIM's constraint on the *biofuel production cost* only gains an additional 11% (from 68% to 79%) density, compared to the jumps of over 20% for the previous two. This information, along with that provided below, can help users decide whether Scenario A represents an appropriate balance among coverage, density, and interpretability. If these prove unsatisfactory, users may also refer back to the parameter contours of Fig. 3 to assess whether more suitable balances might be achieved by starting with an alternative box.

Fig. 5a displays the cases in the database as a function of the first two parameters, *transport demand elasticity* and *biomass backstop price*. The black dots show the high cost cases, open circles show the other cases, and the boundaries of Scenario A shown in red. The figure does not show cases that fail to satisfy the constraints on the third and fourth parameters, *biofuel production costs* and the *supply of low-cost biomass*. Fig. 5b similarly displays the cases in the database as a function of the first and fourth parameter from Fig. 4, *transport demand elasticity* and *biofuel production cost*. Not surprisingly, both Fig. 5a and b show that the high-cost cases correlate strongly with the predicted ranges for the input parameters. Fig. 5a also suggests that Scenario A's lack of 100% coverage owes to a small number of high cost cases with high demand elasticity but relatively lower biomass backstop price.

Note that PRIM has generated a figure analogous to that of the standard scenario axis displays of intuitive-logics school. The axes in Fig. 5 represent key driving forces that define the scenario representing unacceptably high costs for the 25×25 policy. The scenario suggests that in any future where four conditions hold – *transportation demand elasticity* exceeds 0.43, *biomass backstop price* exceeds $148/ton, *biofuel production costs* exceed $73/unit and the *supply of low-cost biomass* is less than 770 megatons – then the 25×25 policy is likely to produce high costs for consumers, irrespective of the value of the other five uncertain parameters in the model. If these conditions fail to hold, then the policy is less likely to produce high costs. This description provides a concise summary of a potential vulnerability of the 25×25 policy.

3.3. Evaluate Scenarios

Based on the information from Section 3.2, Scenario A appears to provide a useful description of the vulnerabilities of the 25×25 policy. Before proceeding to employ this scenario, users may wish to acquire more information about the significance of the four parameters used to define it to ensure that each represents a real property of the TGL model and not a statistical artifact.

The Table 2 shows the frequency with which PRIM uses each model input parameter when run on ten different N/2 sized resamplings of the dataset. Note that the first three parameters defining Scenario A are all used in every resampling and that the

---

[12] Note that the scenario defined here is slightly different than that identified in Griffin [27] because the PRIM toolkit now uses more computational cycles to try to expand the coverage-density frontier. Overall, however, this study confirms the choice of scenarios used by Griffin.

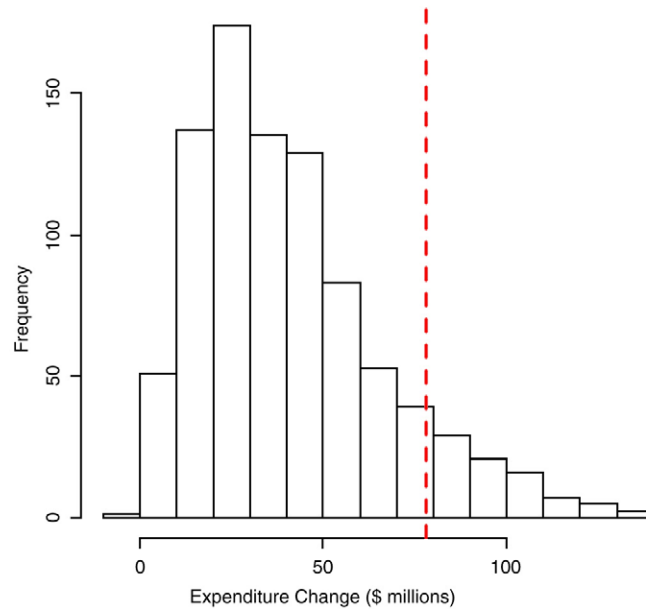**Histogram of Expenditure Change for Revenue Neutral Policy**

Fig. 2. Distribution of costs over the 882-point experimental design for the 25 × 25 policy. The dashed red line indicates the threshold for the worst 10% of cases.

biofuel production cost is constrained in 90% of the cases. Parameters not used to define Scenario A appear in 40% or less of the resampled cases. The frequency with which these other parameters appear is higher than would be likely in the full-sized dataset, because with fewer points more parameters are likely to be falsely constrained.

Table 2 also shows the qp-values associated with each of the four parameters used to define Scenario A. Note that each additional parameter is orders of magnitude less significant than its predecessor, though all but the last would qualify as highly significant. A standard threshold for significance would reject using *biofuel production cost* in the scenario definition because it has a significance level of only 0.19.
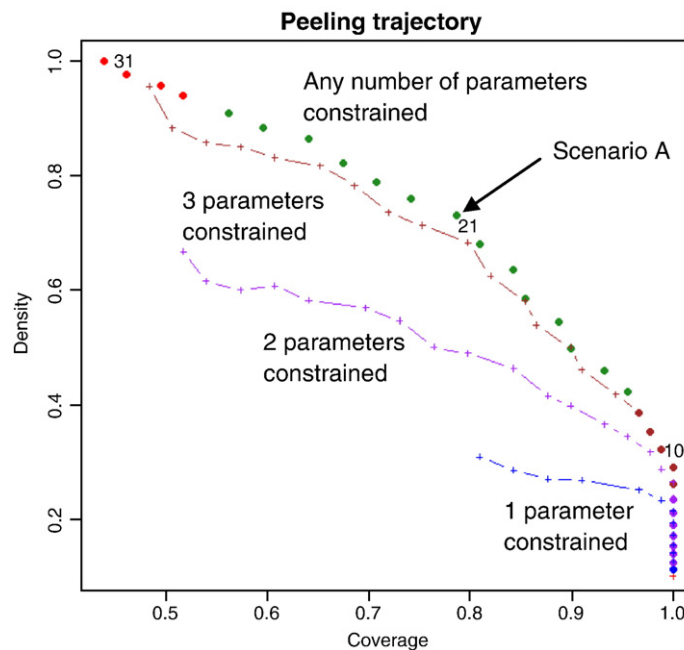
Peeling trajectory

Fig. 3. Coverage density tradeoff curves for scenarios that describe the high cost cases for the 25 × 25 policy using 1, 2, 3, and any number of parameters.

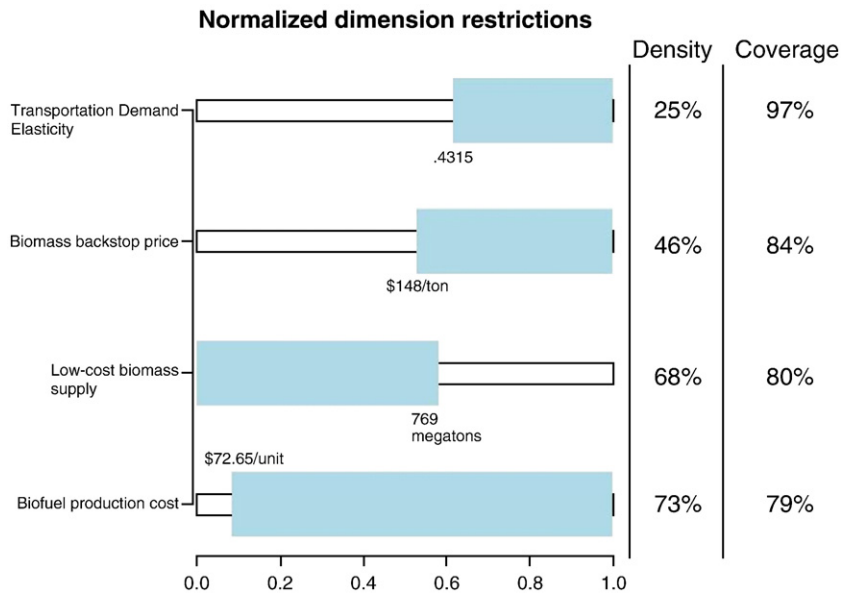**Normalized dimension restrictions**

Fig. 4. Parameters used to define Scenario A in Fig. 3. Also shown are the density and coverage for a box defined by each successive parameter and all those above it.

Together, these reproducibility statistics and qp-values provide high confidence that the first three parameters in Fig. 4 each play an important role in defining the high cost scenario. While the resampling statistics also support the importance of the fourth parameter, *biofuel production cost*, the qp-value test suggests the inclusion of this parameter may be due to chance.

3.4. Choosing a scenario

Users can now employ the information generated in Sections 3.2 and 3.3 to choose a scenario that best represents the potential vulnerabilities of the 25×25 renewable energy policy. Some users may choose to retain Scenario A despite the equivocal test results for the *biofuel production cost* parameter. Others might decide this fourth parameter was not worth the cost to interpretability and choose a scenario defined by only the first three parameters. Some users might return to Fig. 3 and examine one or more of the other scenarios offered there. Others, noting from Fig. 5b that the relatively sparse sampling near the lower scenario boundary may influence the qp-value test, might choose to generate a new database with a higher density of cases. In any of these situations, users would repeat the evaluation if they choose to explore an alternative to Scenario A.

Once users have chosen an initial scenario with less than 100% coverage, they may wish to identify a second scenario to characterize some or all of the remaining points. As described in the Appendix, PRIM can help identify a second scenario by removing all the points from the first scenario from the database, and then repeating the scenario discovery process on the remaining points. In this example however, we judged Scenario A has sufficiently high coverage to render a second scenario unnecessary.
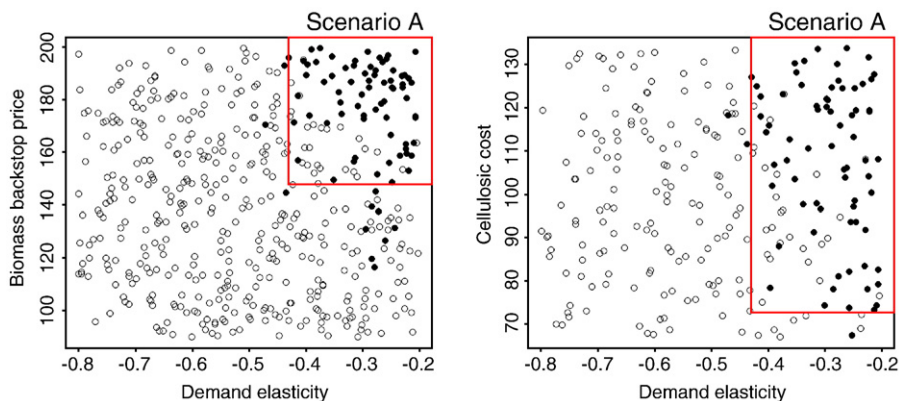

Fig. 5. Cases in database plotted as function of: a) first two parameters and b) first and fourth parameters shown in Fig. 4. Black and open dots show high-cost and lower cost cases, respectively. Red lines show parameters values corresponding to the boundaries of Scenario A.

Table 2

Results of qp-value and resampling tests for Scenario A. Highlighted parameters are used to define the scenario.

Model input parameter	Reproduce coverage	Reproduce density	qp-value
Transportation demand elasticity	100%	100%	1.2 E-16
Biomass backstop price	100%	100%	4.9 E-11
Low-cost biomass supply	100%	100%	1.4 E-05
Biofuel production cost	90%	90%	1.9 E-01
Biofuel yield	40%	30%	
Electricity co-product	40%	20%	
Shift in oil supply curve	10%	10%	
Feedstock supply distribution	0%	0%	
Oil supply elasticity	0%	0%	

4. Informing decisions with scenario discovery

Decision makers evaluating renewable energy requirements, and many other policies under conditions of deep uncertainty, face a multiplicity of alternative plausible futures [10]. Any process that aims to structure decisions and help evaluate the consequences of alternative choices must summarize this multiplicity of futures in a way that proves useful to the parties to the decision. Traditional risk management approaches assume precise probability distributions over future states of the world, and can thus summarize the future with a small set of numbers – a mean and a few moments – for each property of interest. Scenario methods assume the future is less knowable and characterize uncertainty by describing a set of alternative futures. But decision makers can only engage with a relatively small number of scenarios, and thus scenario methods must help their users determine which futures appear most worthy of attention. The scenario axis method of the intuitive logics school offers a qualitative approach for making such choices. The school *La Prospective* can use computer-assisted morphological analysis, as described in more detail below, to systematically lay out the full set of self-consistent possibilities and identify those most significant to the system's dynamics. Scenario discovery offers a new computer-assisted approach for summarizing the multiplicity of plausible futures by focusing on those most important to the design and choice among strategies in the face of deep uncertainty.

In many respects, the results of scenario discovery appear to provide similar cognitive benefits to those of the intuitive logics school. As noted above, Fig. 4 identifies four model input parameters analogous to what the intuitive logics school calls key driving forces. Like the intuitive logics drivers, these parameters rank most highly according to their uncertainty and importance to the users' decision. Fig. 5 displays how the combination of these driving forces define a region of particular importance to this decision. This region offers a clear "scenario logic" – with high cost for biomass production and inelastic consumer demand for transportation a 25×25 policy would cause fuel prices to rise significantly, while demand drops relatively little, leading to large costs paid by consumers. Like a scenario created through the qualitative intuitive logics approach, this region can be given a name and its logic used to build a narrative description that captures decision makers' imaginations. The region also provides a basis for deepening decision makers' understanding of the key forces that might affect the success of the 25×25 policy.

These features may help scenario discovery to successfully inform policy debates by helping decision makers to focus on key assumptions about the world which may reinforce or challenge their policy preferences, and helping to build consensus among parties to a debate with different expectations about the future. Schoemaker [11] describes how traditional scenarios, by using multiple plausible futures to describe uncertainty and by presenting each future as a possibility rather than a prediction with known likelihood, can become psychologically less threatening to those holding different worldviews. The same mechanism can help a set of scenarios gain common acceptance among parties to a decision who disagree with one another. "Precisely because scenarios do not aim to predict the future, but rather bound it," Schoemaker writes, "that a consensus building approach can work even if based with starkly different viewpoints."

Scenario discovery appears to provide both these cognitive benefits. The approach divides multiple runs of a simulation model into cases where a proposed policy performs well and where it performs poorly, and then provides a concise summary of the key assumptions about the future that lead to these different outcomes. Such information has helped decision makers consider and understand futures they might otherwise challenge because they prove inconvenient or threatening. Similarly, scenario discovery analyses have engaged parties with very different beliefs and expectations because the ensemble of model runs includes cases that reinforce the worldviews of all sides. For instance, the range of model input parameters in Table 1 include without prejudice estimates from studies that would tend to support and to undermine the 25×25 policy. In briefings and meetings with many parties involved with the renewables debate, visualizations similar to Fig. 5 allowed most individuals to find outcomes with which they felt comfortable. Similar visualizations played analogous roles in the scenario discovery work on Congress' Terrorism Risk Insurance Act and the climate change vulnerability and response options analysis for Southern California's Inland Empire Utilities Agency. Building on the resulting buy-in, the scenario discovery analyses in each case then helped parties with diverse views agree which combinations of key drivers suggest the proposed policy would succeed and which would cause it to fail.

Scenario discovery, similarly to scenarios from the intuitive logics school, may also help decision makers reduce overconfidence and expand the range of potentially surprising futures they consider. As noted above, traditional model-based scenario exercises often fail to include cases with discontinuities or surprise. To our knowledge, no study has demonstrated the reasons behind this

shortcoming, but one might presume a reluctance among modelers to run cases that deviate too far from those anchored in observations from the past. Scenario discovery may help address this challenge. By using simulation models to actively search for cases that would stress proposed policies, and by providing a framework to judge their importance, scenario discovery may facilitate more systematic consideration of surprise in quantitative analyses [10,30].

The scenario discovery process does offer important differences from that of the intuitive logics school. As a quantitative, analytic approach with quantified measures of merit for scenario quality, scenario discovery can provide a more concise and decision-focused summary of important futures for decision makers to consider. In part, this focus becomes possible because scenario discovery can also demonstrate what combinations of factors prove *less* important to a decision. For instance, note that Fig. 4 and Table 2 not only provide evidence of the most important driving forces, they also provides evidence that the other uncertain parameters are less important. For example, prior to the scenario discovery exercise one might have expected the response of world oil markets, captured by the *oil price elasticity* parameter, to play an important role in affecting the cost of the 25×25 policy. The fact that it does not suggests that disputes pertaining to this factor may prove unimportant to any evaluation of the policy. Similarly, the typical intuitive logics exercise creates four scenarios from two key driving forces, one for each combination of high/low excursions of the two drivers. In contrast scenario discovery may suggest fewer scenarios but with more drivers. For instance, Fig. 5 shows only two scenarios generated from four driving forces: Scenario A and a scenario representing all the other possible combinations of uncertain model input parameters. The analysis demonstrates that these two scenarios are sufficient, because none of the other fourteen[13] combinations of parameters consistently lead to high cost outcomes for the 25×25 policy.

Scenario discovery's analytic process and measures of merit may also help reduce ambiguity and disagreement about the meaning of the scenarios and their key driving forces when used in public debates among parties with diverse expectations and values. As reported by van 't Klooster and van Asselt [12] the intuitive logics approach may have difficulty producing consensus among diverse groups on the interpretation of the scenario axes. In large part, this difficulty may arise because participants attempt to use the small set of scenarios to provide a comprehensive summary of the multiplicity of plausible futures. Not surprisingly, this proves difficult because individuals may have strongly divergent views about what futures are most important, likely, hopeful, and interesting. In contrast, scenario discovery asks a narrower question, and one of inescapable general interest – in what futures might an organization's proposed strategies fail (or excel)? The scenario discovery process then provides transparent, reproducible answers to this question, along with quantifiable measures of merit that can be used to compare the quality of alternative answers.

The scenario literature often suggests decision makers use scenarios to develop strategies that are robust across a wide range of plausible futures [8]. The robust decision-making (RDM) method provides a systematic, quantitative process for implementing this concept [2,9,10]. As described above, scenario discovery supports a key RDM step by helping decision-makers to identify vulnerabilities of candidate robust strategies. This information can then be used to design and choose among more effective strategies. For example, Griffin [27] used results similar to Scenario A to suggest and examine three potential modifications to the 25×25 policy that could be implemented alone or in combination: allowing unrestrained corn ethanol, allowing efficiency investments to substitute for renewables, and implementing a safety valve that relaxes the renewable targets if costs exceed some threshold. He described the tradeoffs between the base 25×25 and augmented policies by comparing their performance in several scenarios identified by scenario discovery, and suggests that an approach combining all three modifications should prove most robust.

This RDM process also provides a means to include imprecise probabilistic information in the analysis without undercutting the cognitive benefit of the scenarios. As a purely technical matter, traditional scenarios generally represent points in a space of model input parameters, which makes it difficult to assign them probabilities in a meaningful way since any point should have zero likelihood of occurrence [31]. In contrast, scenario discovery defines scenarios as regions in model input parameter space, so probabilities can be more meaningfully assigned. More importantly, RDM analyses use the results of scenario discovery to identify the threshold probability for such scenarios that could cause a decision maker to prefer one strategy over another (see examples in Lempert et al. [2] and Groves and Lempert [1]).[14] For instance, one might identify the threshold likelihood of Scenario A that would suggest augmenting the base 25×25 policy with one or more of the policies mentioned above. This shift from asking the prediction question – "How likely is this scenario?" – to asking the decision impact question – "How likely would this scenario need to be to affect one's choice of strategy?" – allows the RDM process to include probabilistic information while preserving the sense of possibility rather than prediction that Schoemaker identified as so important to the cognitive benefit of scenarios.

The scenario literature also offers methods more quantitative than the scenario axis technique to explore the multiplicity of plausible futures and focus on the most important few. But these approaches generally employ different selection criteria than that used by scenario discovery. In contrast to traditional sensitivity analysis [23] scenario discovery not only indicates the most important inputs driving variation in model output, but also identifies the combinations of such parameters and their threshold values that most strongly predict outcomes of particular policy relevance. Tietje [32] proposes a quantitative method that identifies a small set of scenarios that best describe a large range of futures. Numerical methods generate a set of scenarios that meet criteria including: a small total number of scenarios, internal consistency of each scenario, significant differences among the scenarios, and a reproducible process for generating the scenarios (which Tietje calls reliability). While these criteria may prove useful for some situations, they focus on providing the best description of the future, the prediction question, rather than scenario

---

[13] In the traditional scenario axis approach four driving forces would yield 2^4 scenarios, fourteen more than considered here.

[14] This represents a special and often particularly useful case of placing a joint probability distribution over the cases in the database of simulation results.

discovery's policy impact question, which seeks a concise description of those futures most important to consider when designing and choosing among alternative decision options.

Morphological analysis [33,34] offers an approach to computer-assisted scenario development that aims to structure complex policy spaces by enumerating the full set of self-consistent possibilities. The method begins by defining the most important parameters relevant to a decision, lays out the full set of combinatoric configurations, and then eliminates, often with computer support, those that fail a self-consistency test. The resulting database of cases can represent the full set of future possibilities useful for a variety of purposes, such as threat assessment or fault analysis. The school *La Prospective* often employs morphological analysis for exploratory scenario building [35,36], combining it with structural analysis techniques to identify key parameters and with expert judgments about the most important or most likely factors to identify those combinations most worthy of consideration. In one sense, morphological analysis contrasts with scenario discovery because the former offers a highly quantitative approach to structuring a fundamentally qualitative system description (the beliefs of those involved in the scenario process), whereas the latter assumes that these beliefs have been largely quantified into a limited set of mathematical models. But more fundamentally, morphological analysis differs from scenario discovery because the latter seeks to provide a concise summary of the most decision-relevant futures and provides quantifiable measures for its success in that endeavor. In fact, morphological analysis and scenario discovery might prove complementary, with the former providing an alternative approach (perhaps without the use of a simulation model) for structuring the dataset used in the initial step of the latter, as shown in Fig. 1.

The benefits of scenario discovery do come with limitations and costs. First, as currently practiced, scenario discovery requires a quantitative simulation model to generate the initial database of results. Such models can be costly to produce and may restrict the range of phenomenon that can be considered. For instance, many social, cultural, political, and organizational factors may be important to a decision problem but difficult to meaningfully quantify in a simulation model. This constraint may be relaxed somewhat when specifically designing simulations to support decision-making under deep uncertainty, as opposed to developing simulations to predict future outcomes. The former case may allow simple parametric representations of difficult to quantify factors that can then be explored in a meaningful way alongside more easy-to-quantify state variables [10]. For instance, the IEUA work usefully combines information about future climate conditions obtained from extensive scientific modeling with rough estimates obtained from expert judgments of the agency's ability to implement its current strategy. Nonetheless the need for a simulation does raise the entry price for scenario discovery and currently restricts the range of phenomenon it can consider compared to traditional approaches that do not rely on quantitative models to inform the choice of scenarios.

Second, scenario discovery generates results contingent on a proposed policy action. While generally a benefit, in some situations this requirement may conflict with the charter or instructions of the group performing the analysis. For instance, U.S. intelligence services are legally forbidden to consider policy options and thus would have to conduct any scenario discovery analysis using policy options clearly generated by others. Similarly, as reported by [3], those developing the IPCC (Intergovernmental Panel on Climate Change) greenhouse gas emissions scenarios were instructed not to consider any scenarios with an explicit climate policy, and thus might have had to restrict any scenario discovery analysis to the business as usual case.

5. Summary

Methods to inform decisions under conditions of deep uncertainty face a key challenge in summarizing a multiplicity of plausible futures in a way that proves meaningful and useful to decision-makers. Scenario discovery addresses this challenge by defining scenarios as a set of plausible future states of the world that illuminates key vulnerabilities of a policy. It implements this concept by applying algorithms, modified from those in burgeoning fields of statistics and machine learning, to databases of simulation-model-generated results to identify the combinations of values of a small number of key uncertain model input parameters most predictive of the cases of interest. The concept draws from the intuitive logics school, whose process begins with the decision its scenarios aim to inform. However, this focus may prove hard to maintain in many qualitative exercises as even experts find it difficult to follow and compare a multiplicity of plausible paths into the future. In addition, many qualitative scenario exercises inexorably shift from illuminating decisions to an attempt to describe the future. Scenario discovery provides a systematic, quantitative implementation of some of the key ideas underlying qualitative scenario approaches in a way that may address some of their limitations, provide a firmer foundation for use in decision analysis, and prove more effective in broad public debates. Limitations and future work

While the current tools for implementing scenario discovery have often proved useful, they represent only an initial exploration of those potentially available to extend application of the approach into a wide range of policy areas. For instance, all scenario discovery applications to date have used datasets generated prior to applying PRIM or other algorithms, usually with only a Latin-Hypercube design. Sparse sampling density could easily limit the statistical significance of the resulting scenarios in applications with slower running models or higher dimensional space of model input parameters. A clear improvement would employ adaptive sampling methods to generate more simulation model runs near the boundaries of scenarios in order to improve confidence in the statistical significance of the results. In addition, the multidimensional boxes generated by PRIM may not always prove a good approximation of the shape of the underlying scenarios. New methods that transform the original dataset into a new coordinate system prior to applying PRIM, or algorithms that generate interpretable shapes other than boxes, might significantly enhance the quality of the generated scenarios as measured by coverage, density, and interpretability. Improvements in each of these areas can draw on a rich and growing body of the work in the related fields of statistics, machine learning and data-mining. In addition, the tools for evaluating scenarios could be significantly enhanced. For instance, algorithm and sampling-dependent estimates could improve the accuracy of the p-value test for scenario evaluation. In addition to these algorithmic advances, much

valuable work could be done in the evaluation scenario literature to deepen understanding of how decision-makers respond to scenario discovery, how the process could more effectively inform decisions, and how scenario discovery compares to other computer-assisted approaches to scenario development.

<mark>Scenario discovery applies a key concept – choosing scenarios that represent vulnerabilities of proposed policies– and an assemblage of statistical tools for implementing this concept.</mark> Together these can be used to better understand individual policies and the tradeoffs among them, helping to craft robust strategies for long-term and deeply uncertain problems. Despite significant potential for improvement, the initial methods proposed here should support a wide range of useful decision analytic applications and may help provide a new quantitative foundation for thinking about scenarios.

Acknowledgments

Appendix A. Description of PRIM Algorithm

The Patient Rule Induction Method (PRIM) [24] is technically a "bump-hunter" which seeks "high-density regions" by peeling away thin faces of the input space to generate smaller and smaller regions each containing a higher mean value. PRIM employs an iterative, two-step process to find regions within the input space where the mean of the output is significantly higher than the mean of the output over the entire dataset. The algorithm begins with the entire input space and none of the input dimensions constrained. It then creates a series of increasingly smaller and denser (higher mean) boxes. As shown in Fig. A1, PRIM finds each new box by removing a thin low density slice from whichever face of the current box will most increase the mean inside the new (remaining) box. PRIM's developers call the resulting series of boxes a "peeling trajectory."

At this stage, PRIM is designed to receive user input. The algorithm presents the user with a visualization of the peeling trajectory, plotting the density of each box against its coverage. (Friedman and Fisher's original PRIM algorithm uses a support measure, rather than coverage.) The user is then asked to choose the box that best balances the competing goals of high density and high coverage for their particular application.
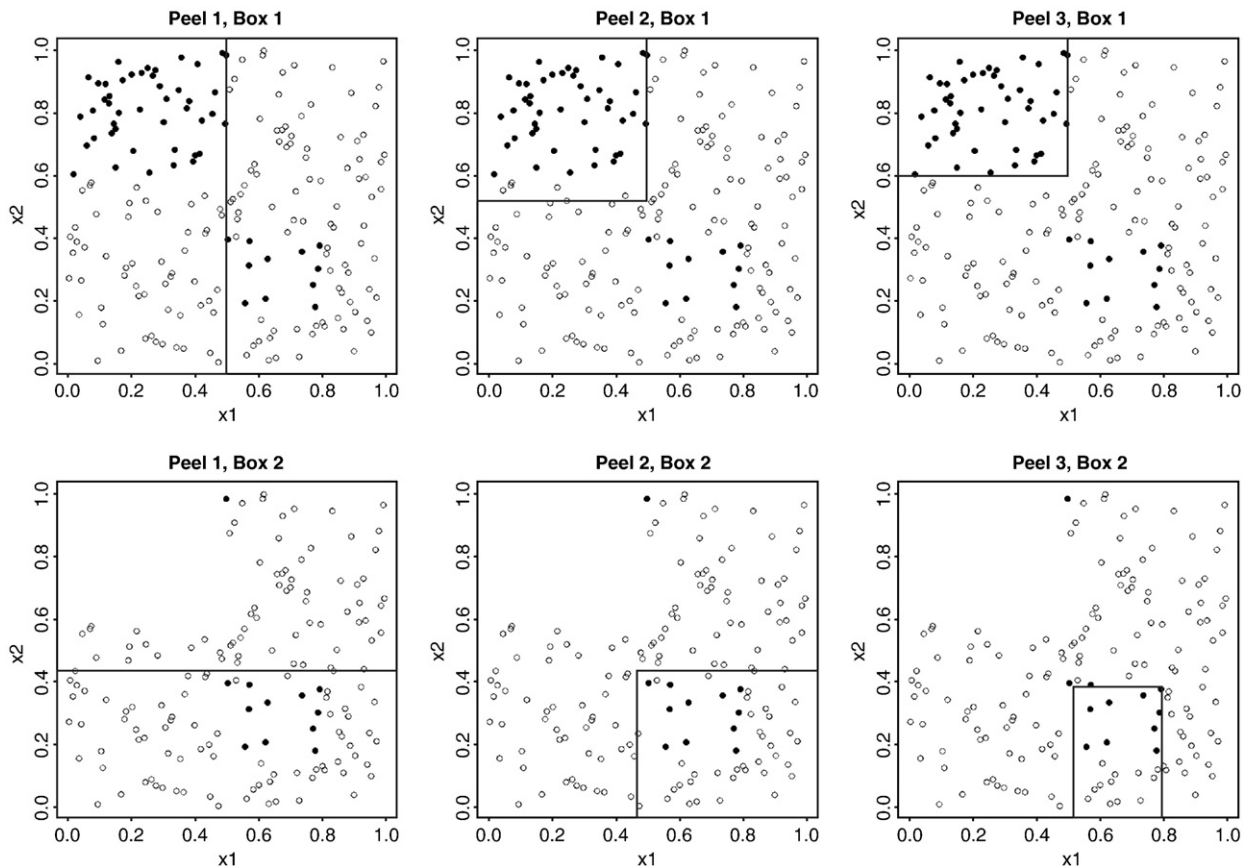

Fig. A1. Sequence of operations by the PRIM algorithm on a notional database.

These candidate boxes can then be expanded in a process called pasting, in which other dimensions that were perhaps unnecessarily constrained are allowed to expand. While conceptually optimal, PRIM's authors found that in general pasting plays little role in improving box quality. However, except in very complex data situations pasting is computationally short and feasible and can only improve box statistics, so our implementation of PRIM actually pastes prior to presenting the peeling trajectory to the user.

Once the user has chosen a box from the pasted candidates, he or she can iterate PRIM using a process called "covering." As shown in the second row of Fig. A1, the algorithm removes all the data points from the dataset inside the first box and replicates the peeling/pasting process with the remaining data. The user may repeat this process of box selection followed by covering until he or she determines the algorithm has exhausted its ability to generate useful boxes.

As one key feature, PRIM peels away only a small fraction of data at each step. This "patience" improves the algorithm's ability to determine the most important input parameters before it runs out of data and helps minimize the adverse consequences of any suboptimal step. PRIM's patience may, as shown in Lempert, Bryant, and Bankes [22], lead the algorithm to inappropriately clip the ends of boxes that would otherwise extend the length of some parameter range, motivating the development of the parameter significance tests described in Section 2.3.

References

[1] D.G. Groves, R.J. Lempert, A new analytic method for finding policy-relevant scenarios, Glob. Environ. Change 17 (1) (2007) 73–85.
[2] R.J. Lempert, D.G. Groves, S.W. Popper, S.C. Bankes, A general, analytic method for generating robust strategies and narrative scenarios, Manage. Sci. 52 (4) (2006) 514–528.
[3] E.A. Parson, V. Burkett, K. Fischer-Vanden, D. Keith, L. Mearns, H. Pitcher, C. Rosenweig, M. Webster, Global-change scenarios: their development and use, synthesis and assessment product 2.1b., US climate change science program, Washington, DC, 2007.
[4] European Environmental Agency, Looking back on looking forward: A review of evaluative scenario literature, Technical Report No 3/2009., European Environment Agency, Luxembourg, 2009.
[5] P. Bishop, A. Hines, T. Collins, The current state of scenario development: an overview of techniques, Foresight 9 (1) (2007) 5–25.
[6] L. Börjeson, M. Hojer, K.-H. Dreborg, T. Ekvall, G. Finnveden, Scenario types and techniques: Towards a user's guide, Futures 38 (7) (2006) 723–739.
[7] R. Bradfield, G. Wright, G. Burt, G. Cairns, K. van der Heijden, The origins and evolution of scenario techniques in long range business planning, Futures 37 (8) (2005) 795–812.
[8] K. Van der Heijden, Scenarios: The Art of Strategic Conversation, John Wiley and Sons, Chichester, UK, 1996.
[9] R.J. Lempert, M.T. Collins, Managing the risk of uncertain threshold response: Comparison of robust, optimum, and precautionary approaches, Risk Anal. 27 (4) (2007) 1009–1026.
[10] R.J. Lempert, S.W. Popper, S.C. Bankes, Shaping the next one hundred years: New methods for quantitative, long-term policy analysis, MR-1626-RPC, RAND Corporation, Santa Monica, California, 2003.
[11] P.J.H. Schoemaker, Multiple scenario development: Its conceptual and behavioral foundation, Strateg. Manage. J. 14 (3) (1993) 193–213.
[12] S.A. Van 't Klooster, M.B.A. van Asselt, Practicing the scenario-axis technique, Futures 38 (1) (2006) 15–30.
[13] T.J.B.M. Postma, F. Liebl, How to improve scenario analysis as a strategic management tool? Technol. Forecast Soc. Change 72 (2) (2005) 161–173.
[14] P. van Notten, A.M. Sleegers, M.B.A. van Asselt, The future shocks: on discontinuity and scenario development, Technol. Forecast Soc. Change 72 (2) (2005) 175–194.
[15] E. Best (Ed.), Probabilities. Help or hindrance in scenario planning? Deeper news: exploring future business environments, 2(4) (Summer 1991) Topic 154.
[16] P. Schwartz, The Art of the Long View, Doubleday, New York, New York, 1996.
[17] L. Dixon, R.J. Lempert, T. LaTourrette, R.T. Reville, The federal role in terrorism insurance: evaluating alternatives in an uncertain world, MG-679-CTRMP, RAND Corporation, Santa Monica, California, 2007.
[18] H. Jenkins, Terror insurance is here to stay, Wall St. J. Aug 8, 2007 A12.
[19] D.G. Groves, D. Knopman, R.J. Lempert, S.H. Berry, L. Wainfan, Presenting uncertainty about climate change to water resource managers, TR-505-NSF, RAND Corporation, Santa Monica, California, 2007.
[20] D.G. Groves, R.J. Lempert, D. Knopman, S.H. Berry, Preparing for an uncertain future climate in the Inland Empire: identifying robust water-management strategies, DB-550-NSF, RAND Corporation, Santa Monica, California, 2008.
[21] J. Coates, From My Perspective: Scenario Planning, Technol. Forecast Soc. Change 65 (1) (2000) 115–123.
[22] S.C. Bankes, Exploratory Modeling for Policy Analysis, Oper. Res. 41 (1993) 435–449 no. 3.
[23] A. Saltelli, K. Chan, E.M. Scott, Sensitivity analysis, Wiley, New York, NY, 2000.
[24] J.H. Friedman, N.I. Fisher, Bump hunting in high-dimensional data, Stat. Comput. 9 (1999) 123–143.
[25] R.J. Lempert, B.P. Bryant, S.C. Bankes, Comparing Algorithms for Scenario Discovery, WR-557-NSF, RAND Corporation, Santa Monica, California, 2008.
[26] M. Toman, J. Griffin, R.tJ. Lempert, Impacts on U.S. Energy Expenditures and Greenhouse-Gas Emissions of Increasing Renewable-Energy Use, TR-384-1-EFC, RAND Corporation, Santa Monica, California, 2008.
[27] J. Griffin, Improving cost-effectiveness and mitigating risks of renewable energy requirements, RGSD-236, Pardee RAND Graduate School Dissertation, Santa Monica, California, 2008, pp. 266.
[28] A. Aden, M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, B. Wallace, L. Montague, A. Slayton, J. Lukas, Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover, NREL/TP-510-32438, National Renewable Energy Laboratory,Golden, Colorado, (June) 2002, As of May 12, 2008: http://www.nrel.gov/docs/fy02osti/32438.pdf.
[29] B.D. Solomon, J.R. Barnes, K.E. Halvorsen, Grain and cellulosic ethanol: history, economics, and energy policy, Biomass Bioenergy 31 (6) (2007) 416–425.
[30] R.J. Lempert, Can scenarios help policymakers be both bold and careful? in: Fukuyama Francis (Ed.), Blindside: How to anticipate forcing events and wild cards in global politics, Brookings Institution Press, Washington, DC, 2007.
[31] M.G. Morgan, D.W. Keith, Improving the way we think about projecting future energy use and emissions of carbon dioxide, Clim. Change 90 (3) 2008 189–215.
[32] O. Tietje, Identification of a small reliable and efficient set of consistent scenarios, Eur. J. Oper. Res. 162 (2) (2005) 418–432.
[33] F. Zwicky, Discovery, invention, research-through the morphological approach, MacMillan, Toronto, Canada, 1969.
[34] T. Ritchey, Morphological analysis, Futures Research Methodology-V3.0, The Millenium Project, 2009.
[35] M. Godet, Creating Futures: Scenario Planning as a Strategic Management Tool, Economica, Paris, France, 2006.
[36] M. Godet, The Art of Scenarios and Strategic Planning: Tools and Pitfalls, Technol. Forecast Soc. Change 65 (1) (2000) 3–22.

Benjamin Bryant is a doctoral fellow in the Pardee RAND Graduate School where he works on techniques for decisionmaking under uncertainty and their application to issues in environment and development. His dissertation research uses robust decision-making methods to evaluate sustainable groundwater management policies in South Asia. He earned his Bachelors degree in mathematics from Harvey Mudd College.

Dr. Robert Lempert is a senior scientist at RAND and Director of RAND's Frederick S. Pardee Center for Longer Range Global Policy and the Future Human Condition. He earned his PhD in applied physics from Harvard University. Dr. Lempert is a Fellow of the American Physical Society, a member of the Council on Foreign Relations, a member of the U.S. National Academy of Sciences' Climate Research Committee, and an author of the book "Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis."