Rationality and Society

USING COMPUTATIONAL METHODS TO PERFORM COUNTERFACTUAL ANALYSES OF FORMAL THEORIES

Andrew D. Martin and Kevin M. Quinn

The online version of this article can be found at:

Published by:

$SAGE

Additional services and information for *Rationality and Society* can be found at:

Email Alerts: http://rss.sagepub.com/cgi/alerts

Subscriptions: http://rss.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://rss.sagepub.com/content/8/3/295.refs.html

>> Version of Record - Aug 1, 1996

What is This?

USING COMPUTATIONAL METHODS TO PERFORM COUNTERFACTUAL ANALYSES OF FORMAL THEORIES

Andrew D. Martin and Kevin M. Quinn

ABSTRACT

Recently there has been an increase in the number of researchers who use rational choice models to explain single cases and rare events. Because of the small number of cases under study, these researchers must rely either explicitly or implicitly on counterfactual reasoning. This paper argues that computational methods provide a profitable means of carrying out rigorous counterfactual analysis. The authors advocate robustness analysis as one important part of the counterfactual analysis of formal theories. Specifically, they evaluate the robustness of the behavioral assumptions of two formal models using various heuristic search algorithms and Markov chains. They find that Kuran's (1989) threshold model of mass protest and Ingberman's (1985) model of direct-democracy referenda are robust to perturbations in their behavioral assumptions. These findings increase the plausibility of causal claims made by scholars who use these models to explain specific events.

KEY WORDS • counterfactuals • discrete optimization • genetic algorithms • robustness

Introduction

Empirical analysis of theoretically derived propositions is at the heart of modern social science. The validity of social scientific explanations depends on many different types of empirical support, ranging from large-sample statistical methods to well-chosen case studies. Regardless of the empirical methods employed, causal explanations must stand up to some type of counterfactual scrutiny. Perhaps it is never more important to explicitly discuss the counterfactual propositions implied by an explanation than when using formal rational choice models to explain single events and small numbers of cases. Because statistical methods cannot legitimately be used to lend *implicit* counterfactual support to such explanations, empirical support must come from the plausibility of *explicit* counterfactual propositions. Green and Shapiro, in their recent (1994) criticism of rational choice theory, claim that

rational choice theories are intrinsically untestable. However, this paper uses computational techniques to generate counterfactual claims about rational choice theories. These claims explicitly test the ability of rational choice models to withstand relaxed behavioral assumptions.

While the analytical rigor of a formal model can help the researcher assess the validity of counterfactual claims, this rigor can also make counterfactual reasoning more difficult. In this article we explore these difficulties and argue that modern computational techniques offer means to circumvent many of these potentially daunting problems. The next section discusses the relationship between counterfactuals and small n explanations derived from rational choice models. Section 2 outlines two models of political behavior on which we later perform computational experiments: Kuran's (1989) threshold model of mass protest and Ingberman's (1985) model of direct democracy referenda. In Section 3, we discuss computational methods which can be used to aid counterfactual reasoning. Section 4 reports the results of these experiments for the two models. We demonstrate that the behavioral assumptions of these models can be expanded so as to make the models applicable to a wider range of cases than would at first seem likely. The final section concludes.

Before moving on to the body of the paper, we define a few key concepts. By *model* we mean a set of definitions and assumptions which, taken together, produce either trivially true or trivially false conclusions. The reason for producing a model is not to make claims about the empirical world, but to explore conceptual questions (Hausman 1992, 77). Because a model, by our definition, is nothing more than definitions, axioms and trivially true conclusions, it constitutes a set of counterfactual worlds of which the modeler has near perfect knowledge. By *theory* we mean, 'a model plus a *general* theoretical hypothesis asserting that the assumptions of the model are true of some portion of the world' (Hausman 1992, 77). An *explanation* is a statement, or set of statements, which makes a causal claim about some phenomena or event, either by appealing to a theory or by making a claim that the event in question meets the assumptions of a model.

1. Counterfactual Analysis and Small n Rational Choice Explanations

In a perfect world, researchers would formulate theories of social phenomena from models which yield general, testable propositions. Researchers would then collect large amounts of data on the relevant

variables and conduct statistical tests to assess how much faith one should place in the theory. Often, this ideal is far from feasible. For instance, scholars in such fields as anthropology, comparative politics, international relations and sociology are often unable to collect data from a large number of cases or are simply interested in explaining a single case. The fact that a small number of cases cannot confirm a particular theory does not imply that a model cannot be used to explain either a specific case, or a small number of cases. As Fearon (1991) argues, counterfactual reasoning can be used to make causal claims about single cases and, in fact, *must* be used to make *any* causal claim.

As Davidson (1980) implies, and Elster (1986) argues explicitly, explanations derived from rational choice models are types of causal explanations. Consequently, the success of a rational choice explanation depends crucially upon counterfactual propositions. While the explicit nature of formal modeling aids the generation of counterfactual propositions by allowing the researcher to vary parameter values and to observe the logical consequences, the complexity of rational choice models and their reliance upon unobservable cognitive processes tends to hinder such propositions.

While explanations based on formal rational choice models increase the likelihood that the underlying logic of explanation is internally valid, they often decrease the plausibility of the connection between the model and the event in question. The reasons for this are three-fold. First, formal models make explicit assumptions which may not be entirely accurate representations of the empirical world. While all models must employ simplifying assumptions, many formal rational choice models make behavioral assumptions which most cognitive scientists would take to be inaccurate. This lack of initial plausibility tends to make all worlds implied by the model implausible, thus short-circuiting counterfactual reasoning before it can begin.

Second, rational choice models must take actor preferences as given. Consequently, accurate knowledge of actor preferences would appear to be a crucial part of any meaningful counterfactual test of a rational choice explanation. Even if actor preferences are assumed to be the same in the actual world and the counterfactual world in question, the validity of the counterfactual proposition depends upon an accurate knowledge of these preferences. For instance, take the stylized rational choice explanation: 'Y occurred because actors i and j had preferences satisfying requirements **I** and **II** and could choose strategies which were elements of Σ'. The empirical accuracy of this explanation may be supported by a counterfactual of the form: 'Y would not have occurred

had i and j been able to choose strategies which were elements of Σ''. Note that in both the original explanation and the counterfactual case, actor preferences are assumed to be known. Consequently, even counterfactual propositions which deal solely with the structural determinants of social outcomes depend crucially upon a more or less accurate understanding of actor preferences.

Third, because models are sets of counterfactual worlds, they imply a very large number of counterfactual propositions. If these propositions can be summarized with relatively simple comparative statics (e.g. 'If k increases, then Y'), this implies no difficulty for analysis and in fact greatly aids the assessment of counterfactual claims. However, the comparative statics derived from formal models may not yield simple linear relationships.[1] As a result, it is often not clear what a counterfactual claim of the form 'if k had been larger, then Y' means in the context of an explanation derived from a formal model.

The first two difficulties discussed so far are closely related. It is relatively easy to see that any departure from behavior predicted by a rational choice model can be explained by appealing to either a different cognitive model or a change in actor preferences. Methods to alleviate these two problems faced by rational choice practitioners have also taken a similar course. For instance, Satz and Ferejohn (1994) acknowledge the deficiencies of rational choice theory as a psychological theory; however, they argue that rational choice explanations in the social sciences are still extremely valuable because the success of these explanations depends upon observable structural constraints and institutional rules.[2] Similarly, Kuran (1995) acknowledges the difficulties encountered in attempts to obtain accurate measures of actor preferences, but goes on to argue that explanation is still possible because observable structural variables may provide information about preferences.

We agree with the positions of these authors. Nonetheless, to say that the results of a model depend more upon the structure of the interaction than upon the accuracy of either the cognitive assumptions or the assumed preference orderings is ultimately an empirical question. By examining the results derived from experimentally varied rational choice models, we demonstrate that the arguments of Satz and Ferejohn (1994) and Kuran (1995) do hold for the two models under study.

2. Models under Consideration

This article serves as both a methodological discussion regarding counterfactual reasoning and a test of the robustness of two formal

theories to perturbations in behavioral assumptions. Before we discuss the methodology and report the results, we outline the two models under consideration.

Kuran's (1989) Threshold Model[3]

One set of rational choice models often used to explain political protest is composed of models based loosely upon the work of Schelling (1978) and Granovetter (1978). Examples of such models include Petersen (1993), Karklins and Petersen (1993), and Kuran (1989, 1991). The basic intuition behind all such models is the following. An individual's decision to protest against his/her government or to support it is a function of how many others protest. These models assume that an agent's expected utility of any action is increasing in the number of others who take the same action. Furthermore, preferences in society are assumed to be heterogeneous. In other words, each actor is assumed to have a different threshold of fellow protestors, which must be reached before he/she will decide to protest. While all of these models share a similar structure, the Kuran model is the most formalized. Hence, it is the threshold model best suited for evaluation in this paper.

Kuran begins by assuming a political system characterized by two competing parties located at opposite ends of a unidimensional scale. One party currently controls the government while the other remains in opposition. It is assumed that neither party can alter its policy position. The primary actors are the members of society who are not publicly pre-committed to either party's position. These uncommitted members of society play a crucial role, since the party in power depends upon their (weighted) public support to maintain power. When deciding what preference to voice publicly, each member of society must weigh two distinct factors. On the one hand, it pays to be on the winning side. Those who are seen as being politically reliable fare better than those who are believed to be politically suspect. On the other hand, the individual pays a psychological cost for voicing a preference different from their private preference and compromising their integrity. By assuming a distribution of private preferences among the members of society, Kuran is able to show how slight changes in this distribution can lead to drastically different levels of protest behavior.

The Kuran model can be treated as a dynamic system with $n+1$ states. For notational purposes, let s_i denote the state in which i members of the population choose to voice a public preference equal to 1. Furthermore, let us call voicing a public preference equal to 1

protesting. Under full rationality conditions this model has two stable equilibria, s_0 and s_{1024}, as well as seven non-attracting equilibria, $\{s_{509}, s_{510}, \ldots, s_{515}\}$. Throughout the rest of the article this formulation of the original Kuran model will be referred to as the deterministic model. Even though there are only two possible actions in the Kuran model (*protest* and *not protest*) there are numerous strategies, owing to the conditional nature of each action. Since each action is conditional upon the previous actions of all n players, there are in fact $n+1$ strategies.

The reasons for focusing on the Kuran model are three-fold. First, it is a relatively simple model that lends itself to both computational and analytic methods of evaluation. Since the purpose of this artice is to provide a general example of several methods of robustness evaluation, the simplicity of the Kuran model was an obvious benefit. Second, it initially appears that this model depends heavily upon the assumption of perfect optimization. Since the model is based upon sequential moves, mistakes made by a few players early on could cause drastic departures from the predicted equilibrium. Finally, as noted earlier, models similar in structure to Kuran's are widely used to explain specific instances of mass protest (e.g. Kuran 1991; Petersen 1993; Karklins and Petersen 1993). Furthermore, Kuran himself (1995) has implicitly argued that explanations derived from his theory are valid because of their robustness. Consequently, the robustness of these types of models is of interest to a wide audience of researchers.

Ingberman's (1985) Setter Model[4]

The second model examined here is Ingberman's (1985) extension of Romer and Rosenthal's (1978) model of agenda-setter behavior in direct-democracy referenda. In the Romer and Rosenthal model, voters decide upon the amount of a publicly financed good to be allocated to their community. This amount is decided by a majority-rule referendum between two alternatives. One alternative is decided by an agenda setter, who is free to propose any level of expenditure. If a majority of voters do not favor the agenda-setter's proposal, the good is provided at what is known as the reversion level, which is assumed to be common knowledge to all. Furthermore, it is assumed that the agenda-setter's objective function is increasing in the amount of the good allocated. Consequently, the equilibrium proposal is the largest expenditure level which will be approved via majority-rule voting, conditional on the reversion level. The primary result of the Romer and Rosenthal model is that in institutional frameworks similar to those of the model, the

optimal proposal of the agenda setter entails substantially greater expenditures than would be expected under a competitive agenda-determining process.

Ingberman (1985) extends the earlier analysis of Romer and Rosenthal (1978) by incorporating a form of reversion rule known as permanent levy. Under this reversion rule the last expenditure level enacted before time t becomes the reversion at time t. In other words, the reversion is simply the status quo level of expenditure. Whereas earlier work focused on reversions which were exogenous to the agenda setter, Ingberman's model explicitly analyzes situations in which the agenda-setter's early proposals will dramatically affect later expenditure levels. Consequently, the agenda setter faces a much more difficult decision problem when confronted with a permanent levy reversion rule. To illustrate this complexity, consider the case of an agenda setter with a four-period tenure. Furthermore, assume that all feasible one-period expenditures lie in the set of integers from 0 to 1023. Under these assumptions, the agenda setter must find the global maximum in a four dimensional solution space composed of roughly 1.0995×10^{12} possible solutions in order to make the optimal series of proposals.

The Ingberman model was selected for evaluation for two reasons. First, owing to the complexity of the agenda-setter's objective function, the equilibrium results would initially appear to be heavily dependent upon the assumption of rationality. Second, these types of models are well suited to empirical testing. Therefore, performing robustness checks helps to determine the meaning of past tests and the validity of future tests.

3. Computational Methods and Counterfactual Analysis

As noted already, difficulties arise in three ways when attempting to derive counterfactual propositions from formal models. First, the narrowness of the assumptions employed may close off possibility before counterfactual propositions can be formed. Second, accurate measures of actor preferences are extremely difficult to obtain. Finally, the sheer number and complexity of relationships between the counterfactual worlds implied by a formal model can make counterfactual analysis very difficult. While this article concentrates on computational methods to alleviate the first two problems, it should be noted that computational methods can also be used to address the third problem as

well. For instance, Kollman (1995) argues for the use of numerical methods to derive conclusions where analytic techniques are not feasible. We prefer to concentrate on the question of how accurate behavioral assumptions must be, since this question is conceptually prior to the third, and because it presents a potentially more difficult problem to overcome.

The key to our argument is the idea of robustness. Robustness is here defined as the quality of a model which implies that small changes in its assumptions will not produce large changes in its conclusions. We use computational techniques to construct the following set of counter-factual cases: 'What would we observe given behavioral assumptions different from but approaching perfect rationality?' In other words, these techniques investigate counterfactual worlds where individuals are not perfect utility maximizers, but instead make choices according to a less stringent criterion. Robustness implies that the causal mechanisms of the model and the counterfactual propositions that they imply hold for cases in which the original assumptions are not quite met but are approached. As a result, a cognitively robust model derives its explana-tory power from very general cognitive assumptions and structural/ institutional factors, both of which are easily observable.

Even though the literature on model robustness is not well developed, robustness is important for models that are used to explain and understand social phenomena. By demonstrating that a model is robust to slight perturbations in its cognitive assumptions, one can claim that the structure of the model explains the phenomenon rather than the choice of assumptions. The strength of an explanation depends on the degree to which the model from which the explanation was constructed approaches the empirical world. By showing that the results from a cognitively robust model derive from extremely general cognitive assumptions and the model's empirically observable structure, robust-ness analysis increases the initial plausibility of a model.

Implementing Robustness Checks

To perform robustness checks, one needs to experimentally weaken the underlying cognitive assumptions of the model while keeping its logical structure intact. By collecting data generated from models with experi-mentally varied assumptions, a check can be made of any given model's ability to withstand such perturbations. The difficulties in evaluating model robustness have caused it to be neglected in much of the rational choice literature. Many of the models for which robustness is most

important are also the most difficult types of models to evaluate. However, recent advances in computing power have made computational investigations of model robustness possible where investigations based upon analytic techniques would be intractable.

Generally, an actor i in a rational choice model encounters optimization problems of the following type: choose a strategy $\sigma_i \in \arg_{a_i}\max$ $u_i(a_i/\sigma_j) \ \forall j \neq i$ s.t. $\sigma_i \in \Sigma$ where a_i is a feasible action, $u_i(\cdot)$ is i's utility function, and Σ is the strategy space. Instead of assuming perfect optimization, heuristic search algorithms can be used to simulate agents making choices in a near-optimal (but not perfectly optimal) manner. These algorithms can be used to relax the assumption of rationality. By observing results consistent with those obtained with models inhabited by perfectly rational actors, one can conclude that the model is robust to perturbations in this fundamental assumption. Thus the model can be further scrutinized empirically. We implement discrete optimization heuristics and models of probabilistic decision-making to serve as relaxations of the rationality assumption. The three widely used discrete global optimization heuristics outlined in this section are genetic algorithms, tabu search and simulated annealing. Probabilistic models of human behavior are also discussed as another check of robustness of these cognitive assumptions.

Methods for Checking Model Robustness

Genetic algorithms (GAs) were created by Holland (1992 [1975]) to study the mathematical underpinnings of adaptive behavior. While GAs were designed to simulate natural adaptive behavior, they have also proven to be powerful global optimization procedures. This twofold ability of GAs both to simulate adaptive behavior and to solve difficult optimization problems has proven quite useful to several social scientists. Our use of GAs to relax the assumption of perfect optimizing behavior falls in line with some recent work by cognitive scientists. For example, Clark has argued that '[b]iological reason ... is better conceived as an iterated process of adaptive response, made under extreme time-pressure, and exquisitely keyed to a variety of external structures and circumstances' (1995, 2). GAs choose courses of action under severe temporal and cognitive constraints, building on past experience and partial solutions. Furthermore, GAs process information in a manner which is implicitly parallel (Holland 1992 [1975]). GAs are well suited for the purposes of this paper, since their implicit parallelism and adaptive nature do seem to mimic some aspects of human learning

(Goldberg 1989, 217) and because they allow us to parameterize the amount of randomness in individual choice.

GAs work in roughly the following manner. First, a number of potential solutions to the optimization problem are selected from the solution space. The objective function is then evaluated at each of these potential solutions, which provides a measure of fitness for each potential solution. A series of genetic operators then modifies this initial set of potential solutions in a manner which generally increases the average fitness of the sample population from time t to time $t + 1$. By iteratively applying these genetic operators to each newly formed sample population, one can progress towards a global optimum. Unlike heuristics based on local search, GAs are not susceptible to traps of local optimality. Nonetheless, some types of objective functions are ill suited to GA-based optimization. Examples of social scientific work employing GAs include Axelrod (1987), Miller (1989), Andreoni and Miller (1990), and Kollman, Miller and Page (1992).[5] For a comprehensive description of GAs we refer the reader to Goldberg (1989).[6]

Tabu search, first proposed by Glover (1977), is a meta-heuristic used to solve both combinatorial and discrete optimization problems. Tabu search captures a different set of behavioral assumptions. Individuals making choices using tabu search consider options locally; i.e. they only consider alternative actions close to the action they are currently making. Within this set of feasible alternatives (the neighborhood) they choose the best option. Over time, individuals making choices based on tabu search will incrementally improve their solution to a local optimum. Tabu search also models memory, since the tabu list prevents the search from returning to inferior solutions.

For the purpose of discrete optimization problems, the heuristic used in the tabu search algorithm is a local improvement scheme, beginning with a good feasible solution. Local search starts from an initial solution σ_i^t and searches to find an improving solution σ_i^{t+1}. In other words, the search attempts to find an σ_i^{t+1} such that $u_i(\sigma_i^{t+1}) > u_i(\sigma_i^t)$. Local search therefore restricts the set of feasible alternatives and models for individuals, incrementally improving their strategy choice. Tabu search uses a memory structure (called the tabu list) that restricts the possible members of the user-defined neighborhood to which a search can progress. Thus, once a local optima is encountered, the search will not be able to revisit that area of the solution space. The tabu list must be small enough to allow the search to carefully scrutinize certain parts of the objective function, yet large enough to prevent a return to a previously generated solution. The tabu search meta-heuristic also uses

an aspiration criterion that defines a condition where the tabu status of a certain move can be over-ridden. Short-term memory functions are employed to intensify and diversify the search. Tabu search is allowed to run for a maximum number of iterations that is computationally practical. A comprehensive description of tabu search can be found in Glover and Laguana (1993).

Another meta-heuristic that relies on local search is called simulated annealing. Simulated annealing was first introduced by Kirkpatrick, Gelatt and Vecchi (1983) and Cerny (1985), and has roots in the work of Metropolis et al. (1953). Simulated annealing is analogous to the annealing process in physical chemistry, when liquid metals are heated and then left to cool into a steady, solid, organized state. Numerous successful applications of simulated annealing can be found in Collins, Eglese and Golden (1988). Simulated annealing algorithms use a probabilistic device called a cooling schedule to escape traps of local optimality. Simulated annealing also provides a different set of behavioral assumptions. Just as with tabu search, individuals making choices using simulated annealing also optimize locally. Simulated annealing also models the stochastic nature of human decision-making; when individuals randomly try a completely different strategy to assess its impact. Simulated annealing therefore captures randomness and spontaneity in decision making. For a comprehensive description and evaluation of many discrete optimization techniques, we refer the reader to Ackley (1987), who empirically assesses the success of each algorithm given different types of objective functions.

Probabilistic models of human behavior can also provide a satisfactory relaxation of the rationality assumption. In such models, one assumes that individuals act rationally (i.e. make the best choice) a certain percentage of the time, yet occasionally make mistakes. When using probabilistic models, a different set of behavioral assumptions are made. Instead of limiting possible search spaces, probabilistic models allow an individual to search over the entire space. However, to account for imperfect decision-making, a stochastic component is included to model cases when individuals make mistakes (and choose suboptimal strategies). This technique is appropriate when individuals make decisions over time; i.e. the use of probabilistic models can assess the path dependency of choices. By formulating a probabilistic model, we can determine if formal models are robust to such mistakes in human decision-making. A tool that is commonly used to model probabilistic processes is called a finite state Markov chain. For a discussion of Markov chains and their applications, we refer the reader to Hillier and

Lieberman (1990, 562–87). In the following section, we implement these analytic and computational techniques to assess the robustness of two rational choice models.

When performing robustness checks on the rationality axioms, it is important to choose an appropriate test. The computational methods presented earlier (genetic algorithms, tabu search and simulated annealing) are useful for nearly all situations, since optimization problems are at the heart of all rational choice models. The use of Markov chains, as well as other probabilistic simulations of behavior, are only appropriate when choices are made dynamically. Thus, when actors make meaningful choices at different points in time, analytic probabilistic models are appropriate. On the other hand, when actors choose a strategy at one point in time, this type of analysis is inappropriate. Therefore, the nature of the interaction must be kept in mind when determining what sort of check to perform. In the following section, robustness checks are performed on the two rational choice models presented earlier.

4. Results of Robustness Checks for Two Models

The basic process of evaluating a given model's robustness can be conceptualized in roughly the following manner. First, the parameters of the model under investigation are given actual values and the equilibria are found. These predictions serve as a baseline against which to compare the results derived from the experimentally varied models. Then, using the same parameter values as in the baseline model, results are derived either computationally or analytically from models with experimentally varied cognitive assumptions. These results are then compared to the results from the baseline model in order to gauge the degree to which the model's equilibrium predictions depend upon the assumptions being varied. In this section, we utilize computational techniques to gauge the robustness of the rationality assumptions used in Kuran's (1989) model of mass protest and Ingberman's (1985) extension of Romer and Rosenthal's (1978) setter model.

Robustness of Kuran's (1989) Threshold Model

In the deterministic model (assuming perfect rationality) each actor is assumed to know the number of fellow protesters needed to make protesting worthwhile. However, in the stochastic version (under relaxed behavioral assumptions) each individual began unaware of the optimal strategy. In these versions of the model each individual was

randomly given an initial strategy or a set of initial strategies. The heuristic method used then stochastically improved the quality of the initial strategy or strategies.

The first test compared the equilibrium predictions of the deterministic model to the results derived from a model inhabited by individuals who use a GA to search for improved strategies. In these runs, each individual was given a set of 20 initial strategies. Each strategy consisted of a ten-tuple bit vector. Each vector represented a threshold level in binary form. Genetic improvement occurred within individuals, as this is consistent with Kuran's assumption of methodological individualism. At each round, one strategy vector was chosen at random from each individual's set of potential strategies. After all individuals have chosen an action, the total number of protesters was recorded and strategy sets for the next round were computed using the GA. This process was repeated for 50 rounds, in 100 trials, and for two initial levels of protest: 400 and 500 expected protesters.[7] The expected accuracy of the GA was also set to three different levels for each of the two initial levels of protest. This was accomplished by varying the number of iterations of genetic improvement from 20 to 5 to 1.

The results of these trials are depicted in Figure 1. As can be seen from this figure, the long-run behavior of the GA modified model is very similar to that of the deterministic model. Although the qualitative movement of the time paths exhibits slightly different rates of change, both the deterministic and the genetic models approached the same stable equilibrium. Thus, it appears that the Kuran model is robust to imperfections in rationality modeled with a GA.

The second computational technique used to introduce imperfections in optimizing behavior was simple local search. Since the behavioral assumptions of local search capture the notion of incremental improvement, the use of local search (as opposed to computationally superior tabu search or simulated annealing) is appropriate. These tests were conducted in a manner similar to the GA tests aside from the limitation of each individual's strategy being set to a single random member, and the obvious difference of the heuristic method employed. The accuracy of the search was parameterized by the size of the neighborhood: from 200 to 50 to 2. Since the strategy space for this model is based on an integer ranging from 0 to 1023, the size of the neighborhood determines the width of the search area; i.e. for a size of the neighborhood equal to w, each iteration t of the local search examined the neighborhood bounded by the strategies $\sigma_i^{t-1} \pm w/2$. The results of the local search are presented in Figure 2.
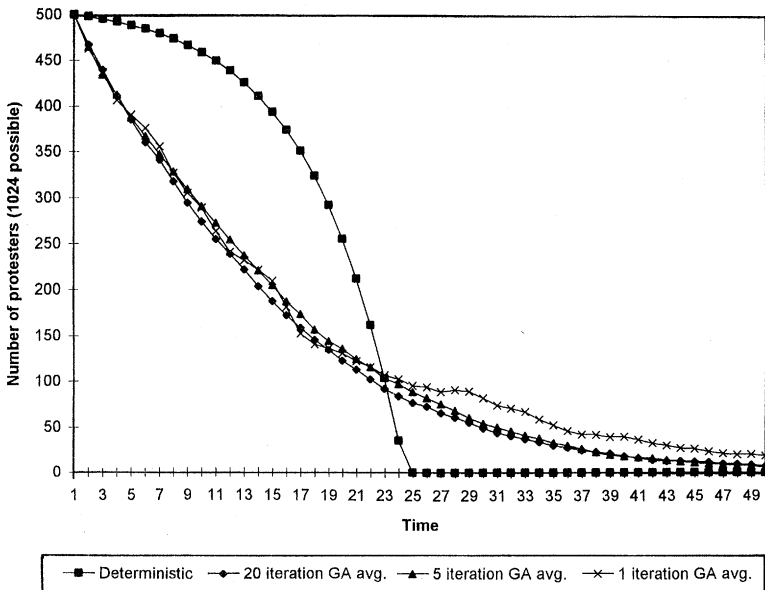
Figure 1. Deterministic results of Kuran (1989) vs GA results (500 initial protesters)

As can be seen in Figure 2, the results of the local search trials were also quite similar to those of the deterministic model. Although the qualitative behavior of the local search implementations was somewhat different from that of the deterministic model, each of the three trials did exhibit long-run behavior that approached the predicted stable equilibrium. This illustrates the robustness of Kuran's model to changes in the rationality assumption when these perturbations are modeled with a local search heuristic.

The final method used to evaluate the robustness of this model is analytic rather than computational. Owing to the relative simplicity of the interaction among individuals and the dynamic nature of the interaction, it is relatively straightforward to implement Kuran's model as a finite state Markov chain. Here we assumed that there is some fixed probability $\gamma \in (.5,1)$ with which each individual will take the optimal course of action. From this, the transition matrix is constructed using the normal approximation to the multinomial distribution. Owing to computational limitations, the number of states was reduced from 1025 to 129. The 129 state model is very similar in structure to the 1025 state model
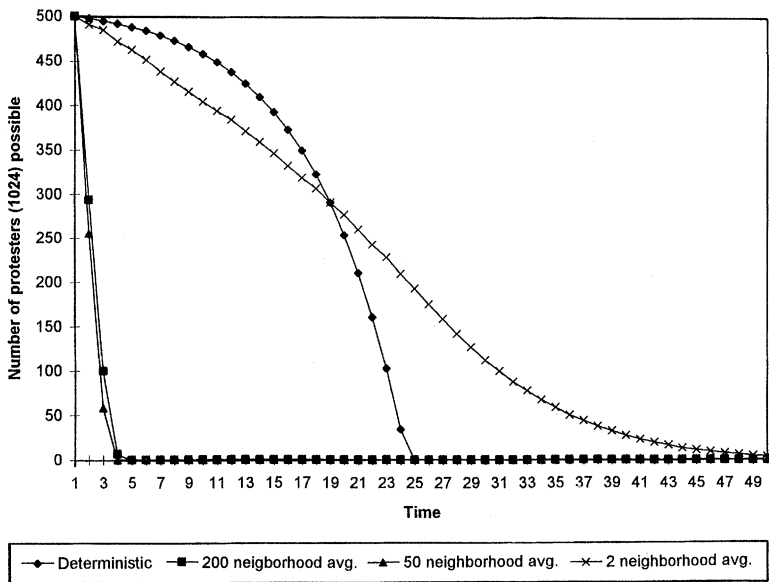
Figure 2. Deterministic results of Kuran (1989) vs local search results (500 initial protesters)

with stable equilibria at the two endpoints and seven non-attracting equilibria $\{s_{61}, s_{62}, \ldots, s_{67}\}$. Tests were performed in which γ was varied from 1.00 to 0.80 and the steady states were calculated. Figure 3 reports the steady state probability distributions for each of these five trials.

From Figure 3 it is apparent that even when the Kuran model is inhabited by individuals erring with relatively high probability, results very similar to the deterministic model are obtained. Although all of the non-deterministic distributions put some positive probability on states in the right-hand tail this is to be expected, since s_{128} is a stable equilibrium. This figure also illustrates that as individuals make fewer mistakes the steady state distribution approaches the deterministic distribution. This analysis indicates that even with imperfect optimizers, Kuran's equilibrium predictions hold.

If we look at the results obtained using three distinct methods of robustness evaluation, it is clear that even fairly sizable departures from perfect rationality do not degrade the model's original analytic predictions. In other words, the results of the Kuran model are not dependent on a specific behavioral assumption. Rather, they follow from the inter-
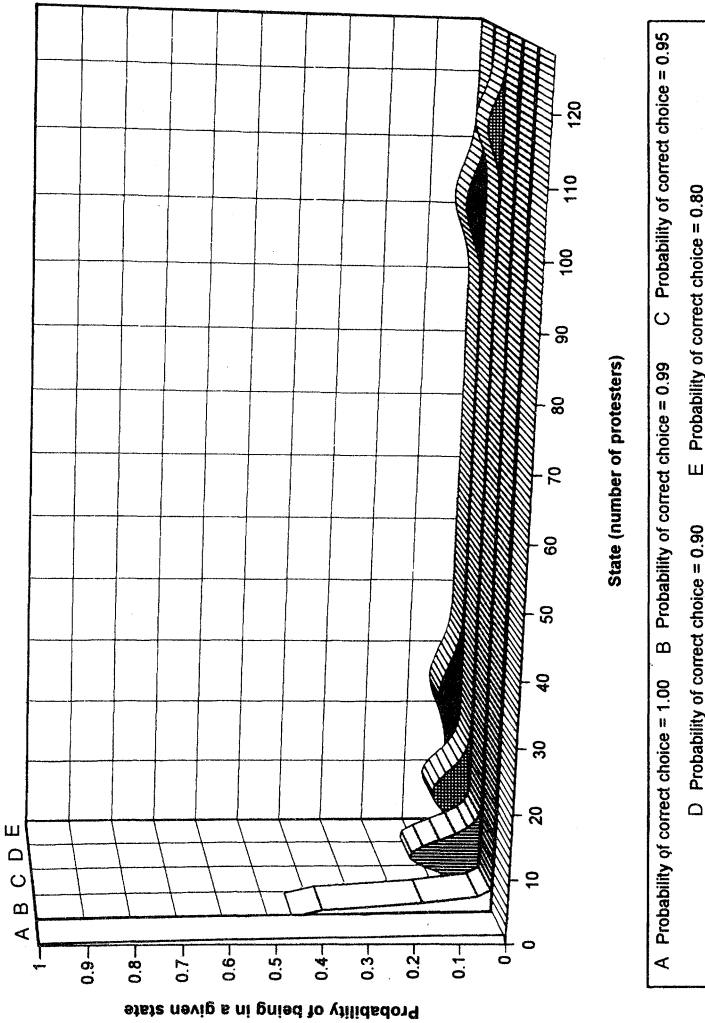
Figure 3. Steady state properties (with varying probabilities of error) of finite state Markov Chain implementation of Kuran (1989)

relation of the substantive parameters of interest. This is especially compelling when we recognize that this model has multiple equilibria.

In a recent symposium in the *American Journal of Sociology* (1995), Kuran's model of revolutionary protest is revisited in the face of the political revolutions in eastern Europe in the late 1980s. Here, Kuran attempts to explain the inability of his model to predict the revolutionary changes of the late 1980s. His argument is that any model of revolutionary change will fail to *predict* revolution owing to the problem of preference falsification; that individuals will hold different public and private preferences. Kuran's argument is that preference falsification disallows social science from ever truly predicting revolution. This is a compelling argument; however, Kuran leaves out a key assumption— namely, that the behavioral assumptions of his model are correct. The robustness experiments performed here show that Kuran's model *is* robust to perturbations in these behavioral assumptions. Thus, Kuran's argument about his model's predictive power is lent more credence. It is not the choice of assumptions that explains revolutions, it is rather the structure of the interaction. Thus, our results demonstrate the ability of the model to explain past instances of revolution.

Robustness of Ingberman's (1985) Setter Model

We also evaluated two versions of Ingberman's setter model.[8] These two versions of the model are the same as those presented in Ingberman (1985) as concrete illustrations of the analytic results. The robustness of these two versions of the model was tested using a GA as well as local search.[9] A Markov chain implementation of the model was not constructed, owing to the nature of the choice problem facing the agenda setter. Even though the Ingberman model takes place over time, the strategies are chosen only once, at the beginning of each four period round, thus making this technique inappropriate for this situation.

The first robustness test of the Ingberman (1985) model uses a GA to relax the assumption of perfect optimization. The GA implemented began with an initial random sample of 200 potential solutions drawn from $\mathbf{Z}^4_{1024}$ coded as bit vectors. The GA was iterated 5000 times to improve those potential solutions; 200 of these 5000 iteration runs were conducted using random initial samples from the search space. The general results from this test indicate that the version of the model in trial one is quite robust to changes in the optimization assumption. Figure 4 demonstrates how the distributions of each proposal changed as the GA progressed through its 5000 iterations.
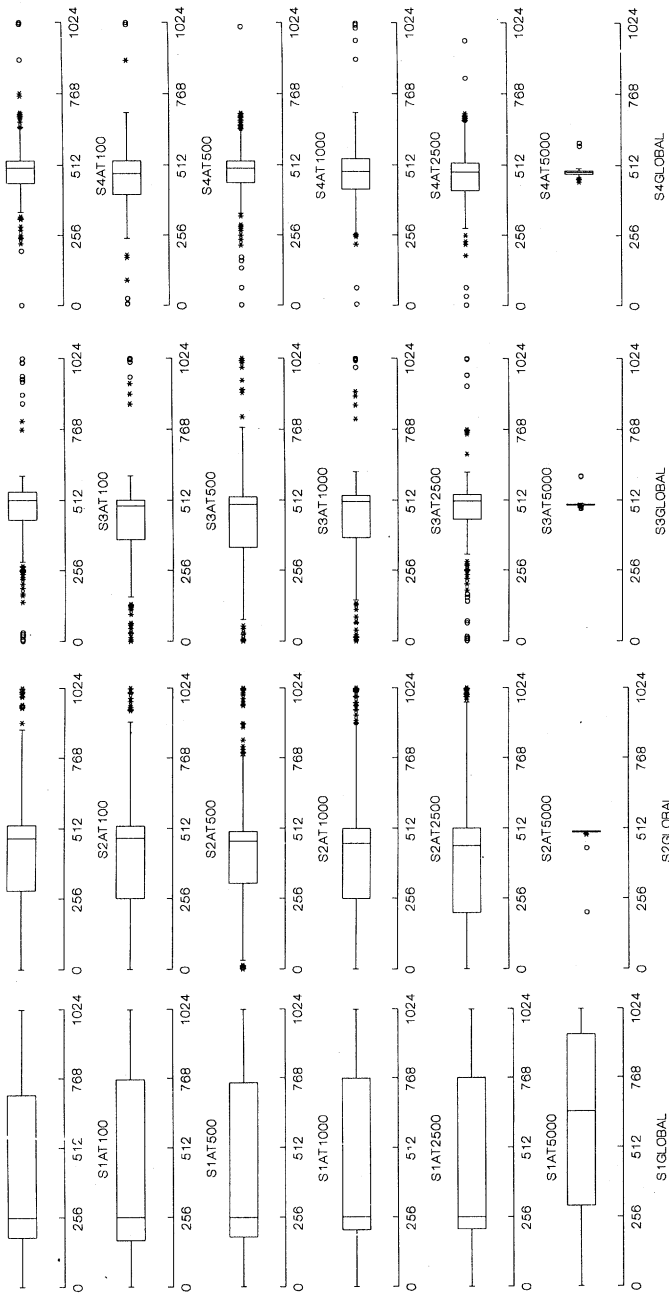
ANDREW D. MARTIN AND KEVIN M. QUINN

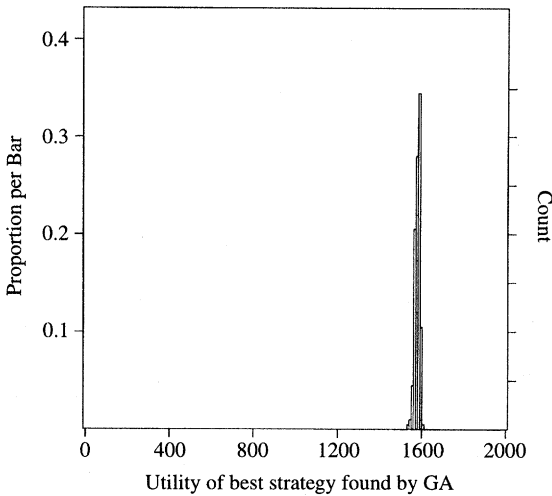**Figure 4.** Distributions of GA-produced strategies for Trial One as the algorithm progresses

Figure 5. Histogram of utility of best strategy found by GA for Trial One

As can be seen from Figures 4 and 6,[10] the variance of these distributions decreased over time. Furthermore, in trial one (Figure 4) the vector of proposals clearly moves towards the unique, optimal vector of proposals {100,500,500,500}. Note that the high mean and high variance of global best first-period proposal simply means that the agenda setter was willing to offer anything which would not be acceptable to a majority of the voters. This implies that the reversion of 100 was implemented. This logic also explains the high variance of the best second-period proposal in trial two. In trial two, the strategies are converging on a vector of proposals, even though there is no uniquely optimal strategy under these parameter values (Ingberman 1985). Figures 5 and 7 present the distribution of the agenda setter's utility derived from the best strategy found by the GA. As one would expect from the distributions of strategies in Figures 4 and 6, the distributions in Figures 5 and 7 are centered near the expected value and have very little variance. In fact, the best strategy found in trial two is superior to the strategy presented by Ingberman (1985, 35).

These results show that under genetic optimization the equilibrium strategy presented by Ingberman is obtained. In other words, even with an imperfect optimization routine, the researcher can expect to see agenda-setter behavior which very closely approximates that postulated by Ingberman (1985). Additionally, these results suggest that as agenda
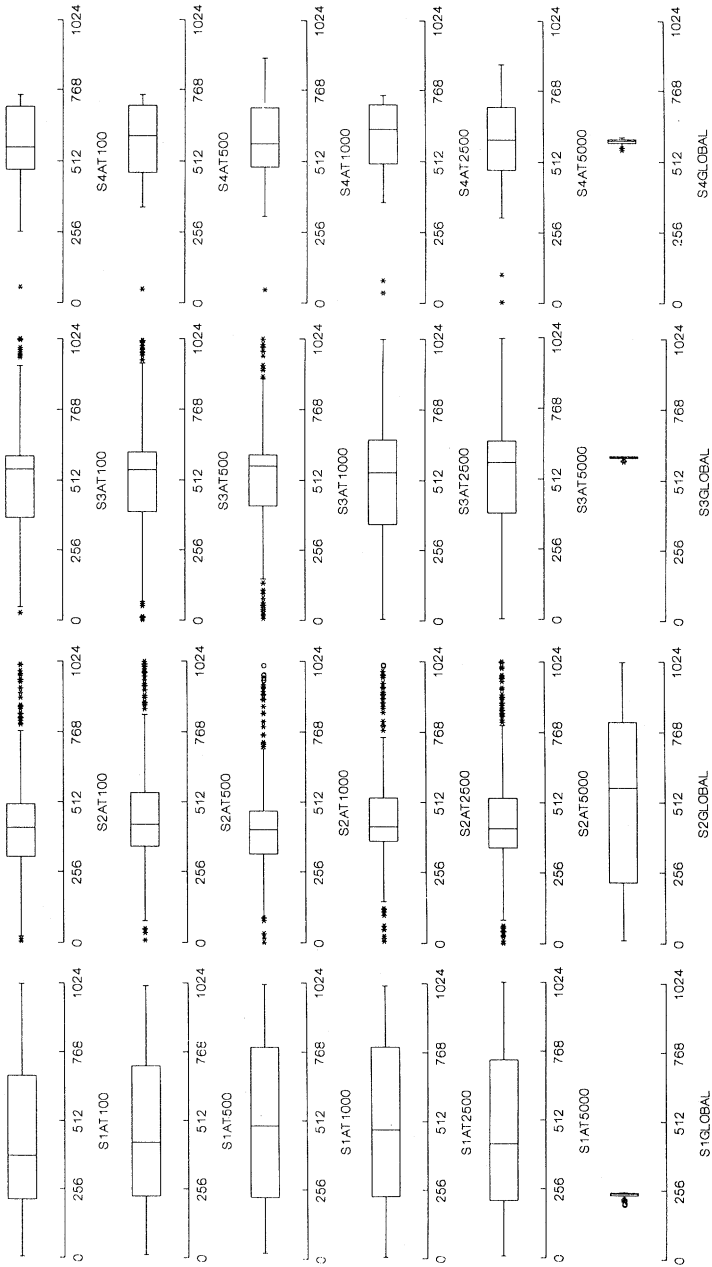
Figure 6. Distributions of GA-produced strategies for Trial Two as the algorithm progresses
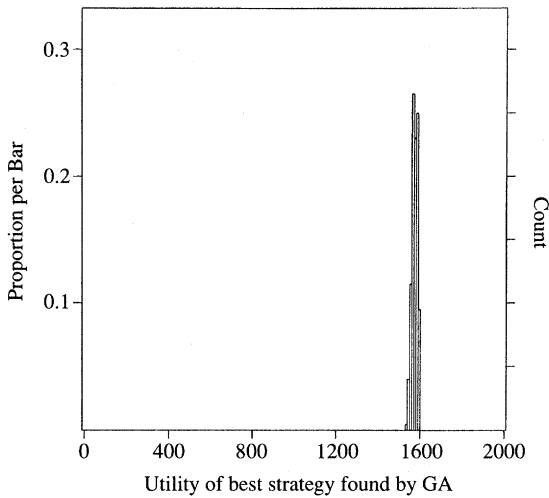
Figure 7. Histogram of utility of best strategy found by GA for Trial Two

setters gain experience with a given institutional setting their behavior will more closely approximate the ideal posited by Ingberman.

The second robustness check of the Ingberman agenda-setter model (1985) uses local search to relax the assumption of perfect optimization. The local search heuristic employed used a multi-started first improving local search that originated from 25 randomly selected members of the strategy space. Each of the 25 strategies was coded as a bit vector of length 40. We define the neighborhood by allowing one bit in the vector to flip (from 0 to 1 or from 1 to 0). This algorithm was then repeated 200 times to produce distributions over the strategies and objective function values obtained. As the search progressed, the objective function values obtained were recorded to compare results with those obtained from perfect optimizers. These results are reported in Figure 8 for trial one and Figure 9 for trial two.[11]

As the search progresses (i.e. the number of multi-starts which have taken place increases), the quality of the objective function values obtained increases. In both Figure 8 and Figure 9, the variances of the distributions are not only decreasing, but the mean is approaching the result obtained by perfect optimizers (Ingberman 1985). It should be noted that the distributions obtained over the four components of the strategies for both trials exhibit the same behavior as the GA results.

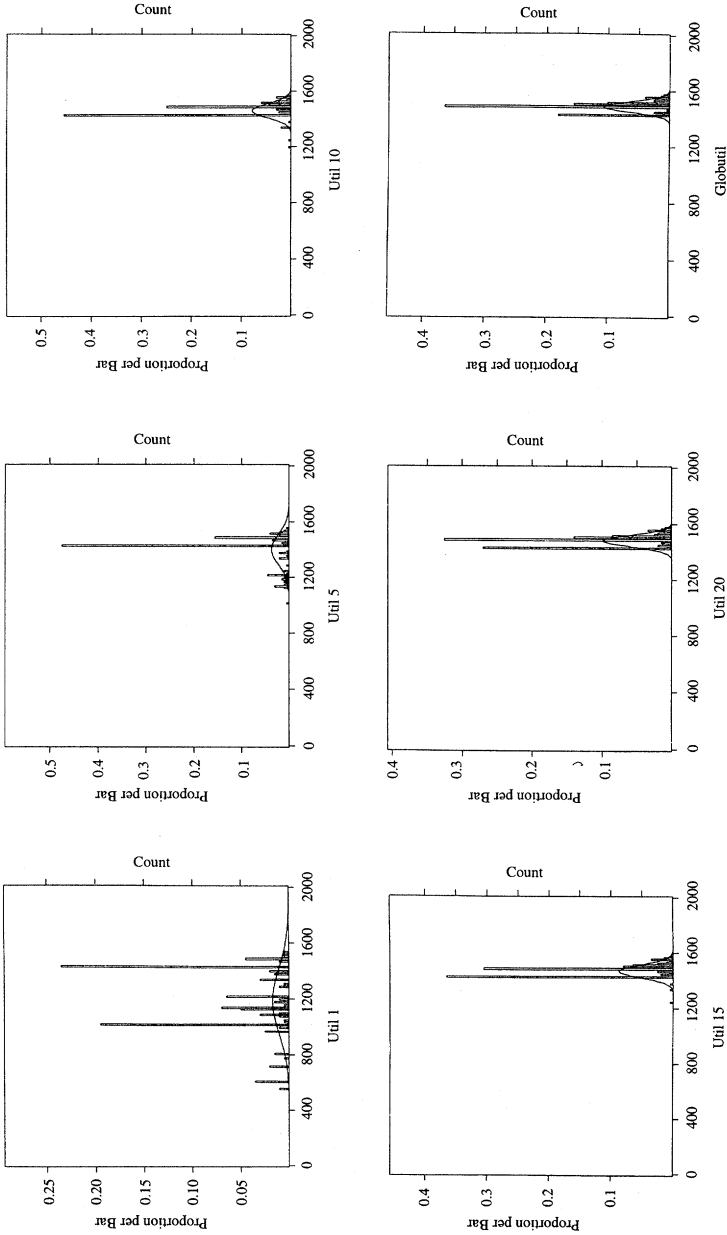The results show that even when agenda setters are not behaving in a

ANDREW D. MARTIN AND KEVIN M. QUINN

Figure 8. Local search utility values for Trial One

perfectly optimal manner, their behavior is practically indistinguishable from the stylized agenda setter posited by Ingberman (1985). From these results, we conclude that the Ingberman model is quite robust to perturbations in the rationality assumption. These results are replicated using both a genetic algorithm and an elementary local search heuristic.

5. Conclusion

Much recent criticism of the rational choice approach to the study of human behavior criticizes the use of strict rationality assumptions to generate empirical propositions about human behavior. Some, such as Green and Shapiro (1994), contend that strict rationality assumptions may not be accurate, and even if they were social scientists would be unable to test their validity. Similarly, cognitive science has yet to provide a complete explanation of human decision-making. Thus, if one accepts these two arguments, the application of rational choice approaches to study human interaction seems meaningless. Attempts to skirt this problem, such as that of Satz and Ferejohn (1994), place the emphasis not on individual cognitive processes but rather on empirically observable institutional and structural constraints. While this is a powerful line of reasoning, it rests on the empirical question of to what extent rational choice models rely on their cognitive assumptions. This paper provides computational techniques that can be used to judge the extent to which substantive results of a model depend on cognitive assumptions.

The results presented here demonstrate that Kuran's model of mass protest (1989) and Ingberman's (1985) setter model are robust to such perturbations. In other words, the explanations of mass protest and fiscal allocations are not driven by the strict assumptions of rationality, but are in fact driven by the structure of the models. The computational methods presented here provide a methodology other social scientists can employ to test their formal models. We argue that the rationality axioms used in rational choice models should be robust assumptions. If the counterfactual case demonstrates that substantive results are not robust to relaxations of the rationality assumptions, then little credence can be put in the explanatory and predictive force of a model. However, a successful robustness test (such as the two presented here) demonstrates that substantive parameters of interest drive explanations rather than the choice of specific behavioral assumptions.
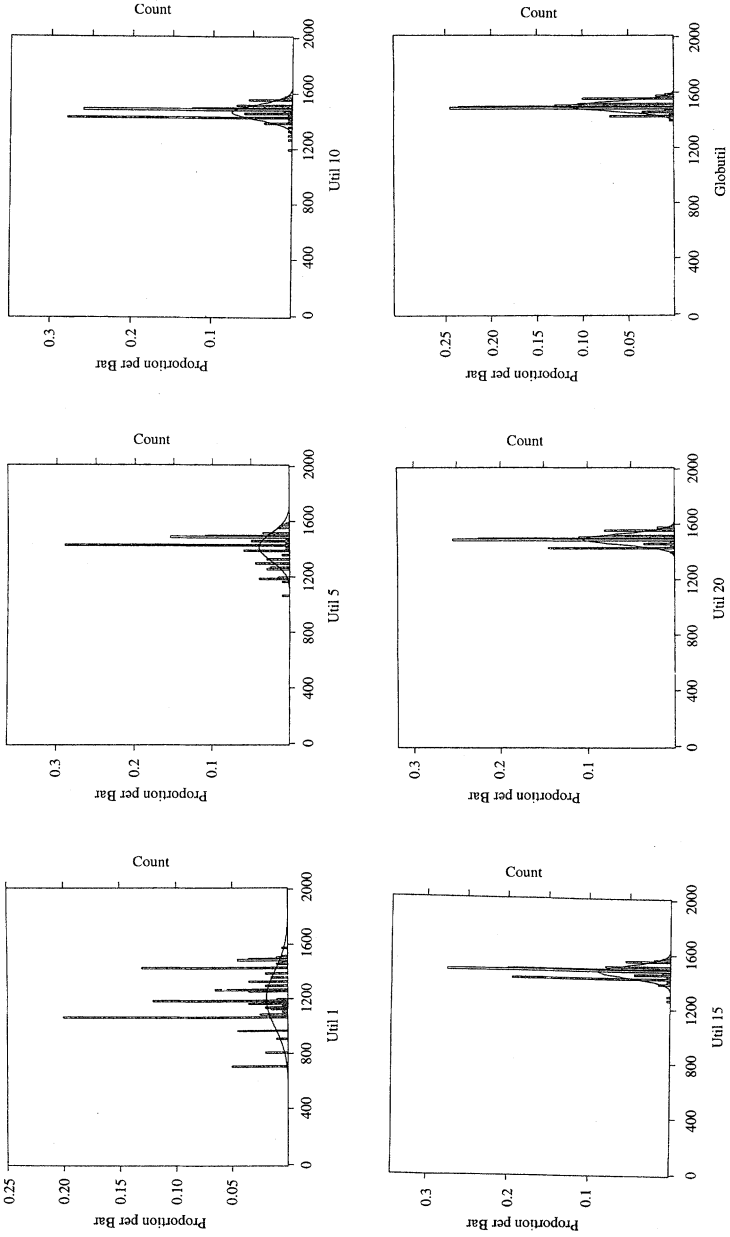
Figure 9. Local search utility values for Trial Two

APPENDIX A

1. Parameter Values Used To Test Kuran (1989)

We evaluated the robustness of a single realization of the model put forth in Kuran (1989), defined by the following parameter values. For a full explanation of these parameters we urge the reader to see Kuran (1989).

population size	$n = 1024$
weight of individual i's public preference	$w^i = \dfrac{1}{n}$, for all $i \in \{0, 1, \ldots, n - 1\}$
i's private preference	$x^i = \dfrac{i}{n - 1}$, for all $i \in \{0, 1, \ldots, n - 1\}$
utility i derives from reputation if i's public preference $= 0$	$f(s) = 0.815(s)$
utility i derives from reputation if i's public preference $= 1$	$F(S) = 0.815(S)$
i's utility derived from integrity	$N(\cdot) = 0.70(\cdot)$

2. Parameter Values Used To Test Ingberman (1985)

agenda-setter's discount rate	$\beta = 1$
curvature of voters indifference curves	$\theta = 1$
first period reversion	$R_1 = 100$
number of voters	$J = 5$
bliss points of voters at time period 1	$\mathbf{I} = \{100, 150, 200, 300, 400\}$
Changes in median income from time $t - 1$ to time t	$\lambda_{t=2} = 100$ $\lambda_{t=3} = 50$ $\lambda_{t=4} = 50$

Parameter Values for Trial One

agenda-setter's discount rate	$\beta = 1$
curvature of voters' indifference curves	$\theta = 1$
first period reversion	$R_1 = 150$
number of voters	$J = 5$
bliss points of voters at time period 1	$\mathbf{I} = \{100, 150, 200, 300, 400\}$
Changes in median income from time $t - 1$ to time t	$\lambda_{t=2} = 150$ $\lambda_{t=3} = 75$ $\lambda_{t=4} = 75$

Parameter Values for Trial Two

APPENDIX B

1. Details of the Genetic Algorithms Employed

The genetic algorithm used to test both the Kuran model and the Ingberman model used a remainder stochastic independent sampling (RSIS) selection function (Baker 1987). RSIS selection works in two stages. First, individuals are selected based on the integer values of their relative fitnesses. For instance, an individual with a relative fitness of 2.452 would produce two offspring in the first phase, while an individual with a relative fitness of 0.904 would produce no offspring in this stage. In the second stage of the selection process, the algorithm moves through the population, stochastically selecting offspring based on the fractional remainder of each individual's relative fitness. Using the example above, the first individual with the higher total relative fitness but lower remainder would be half as likely as the second individual to produce an offspring in the fractional phase. The crossover operator used was a simple crossover operator with replacement. The crossover rate in the Kuran trials was 0.500, while in the Ingberman trials it was set to 0.850. The mutation operator randomly flipped one element of each strategy vector with probability 0.025 in the Kuran trials and with probability 0.100 in the Ingberman trials. Sensitivity analyses indicate that our results are not dependent upon the exact values of these parameters.

NOTES

1. See Lippman, Mamer and McCardle (1987) for a discussion of the potential complexity of the comparative statics derived from game-theoretic models.
2. Even though Satz and Ferejohn (1994) argue against the internalist interpretation of rational choice theory offered by Davidson (1980), they nonetheless contend that rational choice explanations are causal explanations. Thus their position does not reduce the importance of counterfactuals for rational choice explanations.
3. See Appendix A for the specific parameter values used for our analysis.
4. See Appendix A for the specific parameter values used for our analysis.
5. It should be noted that our use of GAs in this paper is fairly different from most other applications using GAs or evolutionary models more generally. The work of Kollman, Miller and Page (1992); and Andreoni and Miller (1990) are notable exceptions. Whereas much of the existing evolutionary modelling literature is primarily concerned with formulating *new* models of social phenomena which incorporate evolutionary mechanisms, this paper uses heuristic search algorithms (some of which are based on evolutionary mechanisms) to assess the robustness of *existing* models. In other words, we are much more concerned with the testing of models than with the modeling of social phenomena.
6. See Appendix B for a discussion of the genetic algorithms used in this paper.
7. Owing to space limitations, only results from the model with 500 protesters are reported here. As expected, the results for the initial condition of four hundred initial protesters are similar. These results are reported in Quinn and Martin (1995).
8. For notational simplicity, we have chosen to call the first model trial one and the second model trial two.
9. As in the previous discussion of the Kuran (1989) model, results from simple local search are presented instead of results from tabu search and simulated annealing, owing to space constraints. Since both the tabu search and simulated annealing meta-heuristics are built on local search heuristics, this simplification only strengthens the following conclusions about the robustness of this model.
10. The variable names in Figure 4 are simply the letter s followed by the period of the proposal, the word *at*, and the iteration number at which the draw was taken. The global results obtained by the GA are simply the collection of best strategies encountered over each of the 5000 iterations. For all box-plots and histograms which follow $n = 200$.
11. Each of the histograms reported in Figures 8 and 9 report the objective function values obtained by searching for optimal strategies using this local search heuristic. At the bottom of each histogram, the *UTIL* indicates that utilities are being reported, and the appended number indicates the number of multi-starts which have already taken place. The final value, *GLOBUTIL* reports the best objective function value encountered during the search. For these histograms, $n = 200$.

REFERENCES

Ackley, David H. 1987. 'An Empirical Study of Bit Vector Optimization.' In *Genetic Algorithms and Simulated Annealing*, ed. Lawrence Davis. London: Pitman.

Andreoni, James and John H. Miller. 1990. 'Auctions With Artificial Adaptive Agents.' University of Wisconsin, Madison, mimeo.

Axelrod, Robert. 1987. 'The Evolution of Strategies in the Iterated Prisoner's Dilemma.' In *Genetic Algorithms and Simulated Annealing*, ed. Lawrence Davis. London: Pitman.

Baker, James Edward. 1987. 'Reducing Bias and Inefficiency in the Selection Algorithm.' In *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*, ed. John J. Grefenstette. Hillsdale, NJ: Lawrence Erlbaum.

Cerny, V. 1985. 'Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Program.' *Journal of Optimization Theory and Applications* 45: 41–51.

Clark, Andy. 1995. 'Economic Reason: The Interplay of Individual Learning and External Structure.' Paper presented at Conference in Honor of Douglass North, Washington University in St Louis.

Collins, N.E., R.W. Eglese and B.L. Golden. 1988. 'Simulated Annealing—An Annotated Bibliography.' *American Journal of Mathematical and Management Sciences* 8: 209–307.

Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Elster, Jon. 1986. 'Introduction.' In *Rational Choice*, ed. Jon Elster. New York: New York University Press.

Fearon, James D. 1991. 'Counterfactuals and Hypothesis Testing in Political Science.' *World Politics* 43: 169–95.

Glover, Fred. 1977. 'Heuristics for Integer Programming Using Surrogate Constraints.' *Decision Sciences* 8: 156–66.

Glover, Fred and Manual Laguana. 1993. 'Tabu Search.' In *Modern Heuristic Techniques for Combinatorial Problems*, ed. Colin R. Reeves. Oxford: Blackwell Scientific.

Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison Wesley.

Granovetter, Mark. 1978. 'Threshold Models of Collective Behavior.' *American Journal of Sociology* 83: 1420–43.

Green, Donald, and Ian Shapiro. 1994. *The Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.

Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hillier, Frederick S. and Gerald J. Lieberman. 1990. *Introduction to Operations Research: Fifth Edition*. New York: McGraw-Hill.

Holland, John H. 1992 [1975]. *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

Ingberman, Daniel E. 1985. 'Running Against the Status Quo: Institutions for Direct Democracy Referenda and Allocations Over Time.' *Public Choice* 46:19–43.

Karklins, Rasma and Roger Petersen. 1993. 'The Decision Calculus of Protesters and Regimes: Eastern Europe 1989.' *Journal of Politics* 55: 588–614.

Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi. 1983. 'Optimization by Simulated Annealing.' *Science* 220: 671–80.

Kollman, Ken. 1995. 'On The Use of Computers to Develop Political Theory.' *The Political Methodologist* 7: 7–10.

Kollman, Ken, John H. Miller and Scott Page. 1992. 'Adaptive Parties in Spatial Elections.' *American Political Science Review* 86: 929–37.

Kuran, Timur. 1989. 'Sparks and Prairie Fires: A Theory of Unanticipated Political Revolutions.' *Public Choice* 61: 41–74.

———. 1991. 'Now Out of Never: The Element of Surprise in the East European Revolutions of 1989.' *World Politics* 44: 7–48.

———. 1995. 'The Inevitability of Future Revolutionary Surprises.' *American Journal of Sociology* 100: 1528–51.

Lippman, Steven, John Mamer and Kevin McCardle. 1987. 'Comparative Statics in Non-cooperative Games via Transfinitely Iterated Play.' *Journal of Economic Theory* 2: 288–303.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. 1953. 'Equations of State Calculations by Fast Computing Machines.' *Journal of Chemical Physics* 21: 1087–92.

Miller, John. 1989. 'The Co-evolution of Automata in the Repeated Prisoners Dilemma.' Sante Fe Institute Economics Research Program Working Paper No. 89–003.

Petersen, Roger. 1993. 'A Community-based Theory of Rebellion.' *Archive Europeene Sociologie* 34: 41–78.

Quinn, Kevin M. and Andrew D. Martin. 1995. 'The Implications of Model Robustness for Empirical Tests of Formal Theories.' Washington University Political Science Paper Number 261.

Romer, Thomas and Howard Rosenthal. 1978. 'Political Resource Allocation, Controlled Agendas and the Status Quo.' *Public Choice* 33: 27–43.

Satz, Debra, and John Ferejohn. 1994. 'Rational Choice and Social Theory.' *The Journal of Philosophy* 91: 71–87.

Schelling, Thomas. 1978. *Micromotives and Macrobehavior*. New York: W.W. Norton.

ANDREW D. MARTIN received a AB in mathematics and government from The College of William and Mary in 1994. He is currently pursuing a PhD in political science at Washington University in St Louis. His academic interests include American politics, political communication, formal theory and quantitative methods.

ADDRESS: Department of Political Science, Washington University, Campus Box 1063, One Brookings Drive, St Louis, MO 63130, USA. [email: admartin@wugate.wustl.edu]

KEVIN M. QUINN received a BA in political science from The Johns Hopkins University in 1992 and an AM in political science from Washington University in 1994. He is presently a doctoral candidate in the department of political science at Washington University in St Louis. His academic interests include comparative politics, formal theory and quantitative methodology. His current research deals with the relationship between ethnicity and social action.

ADDRESS: Department of Political Science, Washington University, Campus Box 1063, One Brookings Drive, St Louis, MO 63130, USA. [email: kquinn@wugate.wustl.edu]