The Political Economy of Preference Falsification: Timur Kuran’s *Private Truths, Public Lies* *

BY ROBERT H. FRANK

Johnson Graduate School of Management and Department of Economics, Cornell University

When asked whether athletes are generally less intelligent than nonathletes, a majority of American university students will say yes (Gary Sailes 1993). Indeed, many of my own students seem to think it odd that anyone could ask such a stupid question. In fact, however, there is no evidence that athletes nationwide are any less intelligent, on average, than nonathletes, and some evidence that they may be more intelligent (Dov Nudel 1989). The reason for the prevailing misperception is that most universities relax their admissions standards when recruiting athletes. The result is that athletes—even though slightly more intelligent on the whole—tend to be academically less well qualified than their classmates within any given university.

In the early 1980s, I used to employ this example in an undergraduate economic policy class to make the point that an unintended consequence of affirmative action programs is to create, within each local environment, a misleadingly negative impression of preferentially admitted groups. This was one among several costs and benefits of affirmative action that I discussed in an attempt to present what I considered a balanced and evenhanded assessment of the program.

Several years later, I was thus distressed to discover in a conversation after class that several students had mistaken my willingness even to discuss this cost as a signal that I was unsympathetic to the interests of minorities. Although I had no reason to question my own motives, I never felt completely comfortable in discussing the athletes example thereafter. And when I stopped teaching the course a few years later, I was relieved to have to confront the issue no longer.

Who has not felt similar discomfort about articulating a position that others might misconstrue? Because being misunderstood involves costs—sometimes high costs—the mere possibility that a remark may be misinterpreted often provides reason enough for remaining silent. Of course, there are often other, more compelling, reasons for not speaking one’s mind. Citizens of the former communist bloc countries, for example, knew that to criticize their governments openly could derail their careers or land them in jail. And even in the absence of formal sanctions, there is the universal human desire for approval, which often prevents people from voicing minority opinions.

Timur Kuran’s bold new book, *Private Truths, Public Lies*, explores the causes and consequences of holding one set of opinions privately while expressing another in public.

What he calls people's "private preferences" are their true preferences—what they actually desire. Their public preferences are the ones they express openly—which, for the reasons noted, often do not coincide with their private preferences. In his carefully reasoned account, Kuran argues that the divergence between private and public preferences sheds light on a long list of important social phenomena: why, in hindsight, revolutions often seem inevitable, yet are rarely foreseen; why, contrary to the claims of structural theorists, specific leaders so often change the course of human history; why, despite being cruelly oppressed by it, even the victims of India's caste system often speak out in favor of it; and why, despite substantial and growing opposition, as reflected in confidential opinion polls, vocal support for affirmative action programs in the U.S. has remained strong. Not all of the arguments he offers in support of these claims are compelling. But even readers who remain unpersuaded will have to take them seriously.

I. How the Theory Works

Kuran explains many of his ideas in the context of a simple model in which there are two alternative positions on a public policy issue—say, full support for a communist regime or full opposition to it. People's private preferences over these alternatives are generated by the usual self-interest calculus: They favor whichever alternative will personally benefit them most. Whether one's public preference differs from the corresponding private preference will then depend on three forms of utility associated with speaking out.

First is what Kuran calls "intrinsic utility," which is the reward from speaking out in favor of the achievement of the privately preferred alternative. This may be an important factor in small groups, in which an influential individual might hope to tip the balance by persuading even a few others. But in most public policy settings, not many individuals can realistically hope to affect the final outcome, and in these cases the intrinsic utility of speaking out is essentially zero.

A second form of utility from speaking out Kuran calls "reputational utility." Here he has in mind the fact that voicing a majority opinion tends to enhance one's reputation, whereas voicing a minority opinion tends to diminish it.

The third and final form of utility from speaking out Kuran calls "expressive utility." Drawing on notions such as self-respect and peace of mind, Kuran argues that people might derive utility from voicing their opinions honestly even though they believe that there is no prospect of affecting the final outcome.

What Kuran calls "preference falsification" occurs when the sum of intrinsic, reputational, and expressive utility is highest for the alternative that the individual privately opposes. For example, people who support the regime publicly even though privately they would prefer to see it fall are preference falsifiers.

The interesting dynamics of Kuran's story emerge from the dependence of reputational utility on the distribution of people who favor the two alternatives. As more and more people favor one position publicly, the reputational cost of favoring the alternative position rises, and vice versa. Thus, if \( p \) denotes the proportion of the population that publicly opposes the regime, each individual may be seen as having a threshold value of \( p \) above which he, too, will oppose the regime. Under these circumstances, Kuran explains, it is no mystery that seemingly small changes can wreak social havoc.

Consider a population consisting of ten persons with the following thresholds for speaking out against the regime:

| Person | a b c d e f g h i j |
|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 0 .2 .2 .3 .4 .5 .6 .7 .8 .9 |

In this population, individual \( a \) so strongly opposes the regime that he will speak out even when no one else does so. Next strongest in their opposition are individuals \( b \) and \( c \), who will speak out only if at least 20 percent of the population speaks out. Individual \( j \), the person who most favors the regime, will speak against it only if at least 90 percent of the population does so. The equilibrium outcome here is that only \( a \) speaks out. The regime coasts along with the public support of 90 percent of the people.
Now imagine that \( b \) has an unpleasant encounter with government officials, the effect of which is to lower her threshold from .2 to .1. Seeing \( a \) speak out, \( b \)'s threshold is met, so she too speaks out. With 20 percent of the population now speaking out, \( c \)'s threshold becomes satisfied and he joins ranks. In short order, the regime's support topples like so many dominoes.

Kuran emphasizes that such dramatic changes are inherently difficult to predict because it is so hard to observe the discrepancies between private and public opinion from which they stem. This difficulty notwithstanding, he argues that relevant signals exist. In some countries, for example, it may be possible to conduct private opinion polls, and even when this is not possible, we can still make useful inferences. Just as Sherlock Holmes learned something from the dog that failed to bark in the night, the mere fact that East Germany prohibited scholars from conducting private opinion polls ought to have revealed the gap between private and public opinion, and hence ought to have alerted scholars to at least the possibility of chaotic social upheaval.

II. A Methodological Middle Ground

Describing the difference between sociology and economics, James Duesenberry once wrote that "economics is all about how people make choices; sociology is all about how people don't have any choices to make" (1960; quoted by Mark Granovetter 1990, p. 97). The battle continues between methodological individualism and competing theories that take groups as the unit of analysis. Kuran's model of preference falsification stakes out an attractive middle ground in this debate. In this respect it is like the models the late James Coleman discussed in his Foundations of Social Theory (1990).

In Kuran's approach, as in both the standard rational choice theory of economics and the theories articulated by Coleman, individuals do the best they can subject to the constraints they face. But whereas Coleman's models and other narrow versions of rational choice theory view the individual as a ruthlessly selfish monad, Kuran's theory adopts a much broader view: People need not only material resources but also the respect and affection of their peers; they have con- sciences; and they make systematic errors when trying to draw inferences from experience. Indeed, many of the novel results in Kuran's theory depend critically on these departures from the standard rational choice model.

Preference falsification itself, of course, can occur even within the confines of narrow rational choice models. These models have no trouble explaining why, for example, a repressive government might impose high penalties for speaking out; or why a selfishly rational person might find it prudent to remain silent in the face of such penalties.

What the narrow rational choice model cannot explain is the willingness to speak out at personal cost even when doing so will make no difference. From the rich descriptive literature on individual behavior in the weeks and months that precede revolutions, we know that people often speak out at enormous cost in the service of hopeless causes. Failure to explain such behavior must be reckoned as a serious shortcoming of narrow rational choice models.

Kuran's account of heroic behavior stresses the individual's need to behave in certain ways in order to maintain a coherent sense of self—to maintain identity. I have no quarrel with this argument, which rings true at a descriptive level and is consistent with a large social psychological literature.

Many economists, however, remain uncomfortable about assuming such motives. From their perspective, the best theory of human behavior is one that relies least on assumptions about broader human motives. These economists will be tempted to try to explain actions similar to the ones Kuran describes on the basis of narrow self-interest alone.

They may concede, for example, that speaking against the majority will often involve reputational costs and have little prospect of making any real difference, just as Kuran describes. They will note, however, that the same act of speech may also involve reputational benefits. Revolutionaries, after all, are members not just of society at large
but also of numerous local groups within it. Indeed, their most intimate ties are generally to others whose views are much like their own. Whereas speaking against the received wisdom will reduce one’s standing in the eyes of the majority, it will boost one’s standing among intimate friends. Conversely, failure to speak out may make it impossible to maintain the personal ties upon which people most strongly depend.

These observations, however, take us only so far. They cannot explain why someone would endure torture and even death in the service of an unpopular cause. People have some latitude in their choice of close associates, and it is difficult to see why a narrowly rational individual would choose a peer group likely to pressure him into self-destructive behavior. In the end, traditional economists must weigh the satisfaction they take in clinging to narrow versions of the rational choice model against the inability of that model to account for important categories of human behavior. Kuran’s book provides powerful reasons for such an assessment.

III. Is Democracy Efficient?

There is a tradition in the social sciences that views social evolution under democratic institutions as fundamentally efficient. Thus, in the Chicago School’s view of the evolution of common law, people who lose from an inefficient law argue more forcefully than those who gain from it; and in the process, laws evolve toward forms that maximize wealth (see, e.g., Richard Posner 1992). Indeed, as Kuran points out, this optimistic world view exists far beyond the academy—as evidenced by the popular slogans like “right will prevail” and “the triumph of evil is short lived” (p. 116). And without question, social institutions do evolve toward efficient forms much of the time.

But not always. Just as Kuran must depart from the narrow rational choice model to explain why people incur ruinous costs to promote social change, so too must he relax the model’s assumptions to explain why beneficial social change often fails to occur.

And here again he draws on both social and cognitive psychology to construct a richer portrait of the individual actor. He discusses first the human need to conform, citing the early work of social psychologists Solomon Asch and Stanley Milgram. In one especially illuminating experiment, Asch (1956) asked subjects to identify which of three lines had the same length as a given line. The task was transparently simple, and subjects who performed it in isolation had an error rate of less than one percent. When subjects were tested in groups, however, the results were startlingly different. All but one member of each group were Asch’s confederates who had been instructed to provide wrong judgments. Even though these judgments contradicted the clear impression of the subject’s own senses, subjects reported a judgment that conformed with the wrong view of the majority in 32 percent of the trials. If the desire to conform can distort judgment so strongly when the objective standard of proof is so clear, its effect will presumably be even more powerful with respect to social questions, for which the standards of proof are invariably less well defined.

Given this psychological tendency, it is easy to see how a false belief, once articulated forcefully by a group in power, might proliferate. And once it does so, Kuran argues, the learning environment is fundamentally transformed. Cognitive psychologists have demonstrated the power of simple heuristics in human judgment. When trying to assess the validity of a position, for example, people often tend to compare the number of reasons they can think of that favor the position with the number of reasons they can think of against it (Eldar Shafir, Itamar Simonson, and Amos Tversky 1993). This heuristic is obviously far from perfect. Thus, contrary to the heuristic, one should oppose a position if there were three weak reasons to favor it and only a single decisive reason to oppose it. Nonetheless, such heuristics economize on decision time and work well enough in many cases.

Useful though they may be, judgmental heuristics lead not only to random decision errors but also to systematic ones. Thus, as Tversky and Daniel Kahneman have explained, the number of reasons people can think of that favor a position depends not
only on how many such reasons actually exist, but also on their “availability”—that is, on how easily the various reasons can be summoned from memory (1981). Arguments that are repeated frequently tend to be more available than others.

This insight clears up important mysteries. For example, although it is easy to see how a determined, but ludicrously false, propaganda campaign might transform the opinions that people feel free to voice in public, it is much less clear how such a campaign might transform private opinion. If private opinion depends on the availability of reasons for holding various beliefs, however, such an effect can easily occur. And once people become sufficiently wedded to a certain view, they often fail even to consider possible alternatives. As Kuran puts it, “the unthinkable becomes the unthought.”

Once adjustments to private preferences begin to occur, they have a self-reinforcing quality. Thus, Kuran writes:

the instrument of positive feedback is adjustments in private preferences. These adjustments either strengthen the incumbent equilibrium or replace it with a more extreme alternative. (p. 193)

Distortions of this sort are probably most important in the context of politically oppressive regimes. But they would also be present even in the absence of government. Thus, citing John Stuart Mill’s view that government oppression is generally less burdensome than the tyranny of friends, neighbors, and fellow citizens, Kuran notes that the state is by no means the only force capable of coercing public opinion (p. 88). But state oppression is clearly important, and he recognizes that even democratic regimes are not above the fray. For instance, he writes that

Political theories in the neoclassical tradition . . . posit that parties compete by tailoring their agendas to fixed voter preferences. In reality, of course, parties also try to shape what voters want. (p. 175)

If Kuran’s point is that outcomes under democracy fall short of some full-information, frictionless ideal, he is surely right. His arguments provide ammunition for those who must do battle in the public policy realm with opponents who insist that democratic institutions tend always and everywhere toward efficiency.

But Kuran sometimes pushes his case too far. He makes much of the fact, for example, that reliance on “social proof” (testimonials and the like) is important even in democracy, and that it can lead to bad outcomes. Fair enough, but reliance on social proof also has its upside. As Kuran fully realizes, our cognitive capacities are limited, and without heavy reliance on social proof no one could manage even to get through the day.

Reliance on social proof also causes harm by foreclosing certain areas of discourse. Yet in many circumstances, the foreclosure of discourse is a manifestly beneficial outcome. For example, the world is surely a better place because social pressure deters proponents of human slavery from exercising their constitutional right to articulate their views in public. Yet, curiously, Kuran cites the fact that such pressure prevents Nazis from speaking freely as evidence democracy’s imperfection (p. 85).

In the political realm, as elsewhere, perfect outcomes are rarely attainable. Our choice is among flawed alternatives. Kuran’s arguments undercut those who make overblown claims on behalf of democracy, but do not challenge Churchill’s view that democracy is the worst form of government—except for the alternatives.

To my eye, the most intriguing aspect of Kuran’s discussion of the efficiency of democracy is one he does not confront—namely, the question of how even to define an efficiency standard when individual preferences are highly malleable. Kuran’s view on this issue seems to be that of the traditional economist—that efficiency is in the eyes of the beholder. Thus, he writes that “an inefficient decision will become efficient if enough people support it publicly for a sufficiently long time” (p. 193).

On this view, we would have to regard the institution of slavery as efficient if it could be shown that slaves grew to like their condition once they got accustomed to it. Those of us who reject such a conclusion implicitly embrace the proposition that one’s preferences
of the moment are not the final standard for welfare analysis. Given the power of circumstance to mold preferences, we must ask whether it would be better for preferences to be molded in some ways rather than others. This is a profoundly difficult question, for it requires the use of some standard other than individual preferences of the moment—the bedrock of traditional welfare economics. Some of the profession’s deepest thinkers—notably Amartya Sen—have grappled with this question (1985). But we remain far from a satisfactory answer to it.

IV. Affirmative Action

Kuran illustrates many of the points of his thesis with an in-depth case study of affirmative action programs in the United States. The essential argument, which has been made previously by Shelby Steele (1990), Glen Loury (1995), Robert Hughes (1993), and Stephen Carter (1991), among others, is that despite widespread private opposition to affirmative action programs, these programs have endured because whites remain silent out of fear of being branded as racists.

That public debate on racial matters is impoverished by this fear is a widely held view. Thus the New York Times praised Steele’s book The Content of Our Character for giving “eloquent voice to painful truths that are almost always left unspoken in the nation’s circumscribed public discourse on race.” Loury’s account of the dynamics of public discourse contains perhaps the richest description offered by any of these writers. According to Loury, people can be partitioned into two camps, those who are more sympathetic to the economic and social interests of disadvantaged minorities and those who are less sympathetic. Even though there are valid reasons that a member of the former camp might oppose affirmative action programs, the proportion of people who oppose these programs is significantly smaller among sympathizers than nonsympathizers. Thus, when a person expresses opposition to an affirmative action program, the rational listener’s estimate of the conditional probability that the speaker is a nonsympathizer goes up.

Most speakers—even those who are, in fact, unsympathetic to minority interests—will count this change in perception as a personal cost. And this cost will lead those who are only marginally opposed to affirmative action programs to remain silent. Because this marginal group contains disproportionately many who were sympathetic to minority interests, the remaining group of people who publicly oppose affirmative action will now consist even more strongly of people who are unsympathetic. The effect of speaking out against affirmative action now creates an even larger change in the listener’s perception and this leads still another set of former opponents to remain silent. In the end, even though most people may be strongly opposed to affirmative action programs, few may be willing to say so publicly.

Kuran argues in favor of this interpretation by describing a long list of negative consequences of affirmative action. Thus, he writes,

Americans scarcely recognize the costs of affirmative action. Many harbor misconceptions about its benefits and about its gainers and losers. Few realize how the emphasis on affirmative action has diverted attention from the root causes of the prevailing inequalities, to the detriment of truly disadvantaged Americans, black and nonblack. Only some appreciate how double standards are fueling tribalization and, ultimately, endangering the viability of the American nation. (p. 222)

A weakness of Kuran’s argument is that he discusses only the costs of affirmative action. What about its benefits? For example, the local TV news teams in most major cities now contain at least one minority and one female member, something that probably would not have happened as quickly, if at all, in the absence of affirmative action. These visible placements have undoubtedly helped kindle the aspirations of many young Americans. Affirmative action has also helped provide access of minorities and women to the elite social networks that channel talent to the top jobs in this country. In the absence of some sort of compensatory program, the opportunities facing white Americans with a given ability level would surely continue to dominate
those facing African Americans. Affirmative action programs have helped to level this imbalance.

Kuran cites the economic slide of poorest blacks as evidence that affirmative action failed in its goal to eliminate black poverty. The benefits of affirmative action may indeed have been concentrated on the those who were doing best to begin with. But this does not mean that the program has made things worse for the poorest blacks. Poor whites, after all, have also been losing ground economically in recent decades.

Nor is Kuran’s case convincing that preference falsification by itself can explain how a program opposed by an overwhelming majority could survive politically for more than 20 years. Kuran is right that speaking out against affirmative action entails political costs. (At any rate, this seemed to be true until the 1994 congressional elections.) Yet if most people really are deeply opposed to the policy, shouldn’t it have been a simple matter for politicians to use coded messages to curry favor with voters?

There are alternative explanations for the survival of affirmative action. One is that it persisted because both parties believed that blacks, women, and liberal males strongly favored it. If these three groups do not constitute a working majority, they come close. Rent-seeking is another plausible explanation. The benefits of affirmative action are relatively concentrated and easy to identify, whereas its costs are relatively diffuse. Thus the person who gets the job because of affirmative action knows who she is, whereas those rejected cannot know whether they would have been hired in the absence of affirmative action. Kuran objects that the rent-seeking “thesis is silent . . . on the deficiencies of private knowledge about the consequences of the incumbent racial agenda” (p. 240). But the free-rider hypothesis makes the same prediction as the preference falsification hypothesis, telling us to expect deficiencies of private knowledge concerning virtually any public policy issue.

Preference falsification may indeed cause us to underestimate the costs of affirmative action programs. But we still need to do the overall cost benefit analysis before we conclude that the program does more harm than good.

V. The Caste System

Another of Kuran’s case studies examines the survival of India’s caste system. Preference falsification, he argues, accounts not only for why members of the untouchable castes failed to speak out, despite being severely disadvantaged by the system, but also for why so many of them actively defended it. As Kuran recognizes, this is a hard case to make because the evidence most relevant for assessing it is the behavior of untouchables at the time the caste system was first implemented. That was several millennia ago, and the historical record is essentially silent on this issue.

Even after the caste system was firmly established, of course, there remains the question of why it has resisted repeated attempts by the government to eradicate it. Conventional rational choice models suggest that the untouchables should be actively involved in this effort, but even today many actively resist change. Our understanding of this behavior is enriched by Kuran’s account of social influences in belief formation.

As in the case of affirmative action, however, the persistence of the caste system remains open to alternative interpretations. A powerful group may establish a social order that works to its advantage by the use of force or other conventional sanctions. And once it consolidates its advantage, it is easy to see why rational members of the oppressed group would be reluctant to speak out, or indeed why they might even attempt to curry favor with the upper castes by defending the system. Kuran is thus able to show only that preference falsification could have contributed to the establishment and maintenance of the caste system. The extent to which it did so remains an open question.

VI. Scientific Revolution

Do scientists evaluate theories and evidence independently and disinterestedly, as they would like to believe, or do social influences intrude even in this hallowed domain?
Almost in passing, Kuran applies his model to this question, noting several ways in which it supports the thesis put forward by Thomas Kuhn in his *The Structure of Scientific Revolutions* (1970). According to Kuhn, scientific progress takes place less by the falsification of specific hypotheses, as Karl Popper believed it ought to, than by the replacement of one overarching theory by another. Anomalies in the data tend to be ignored, Kuhn argues, until a new theoretical framework emerges that can accommodate them.

Kuran’s story helps explain this pattern. As long as the received scientific wisdom favors a given theory, someone who speaks out against it risks being thought incompetent or ill-informed. The result can be the same kind of belief persistence that helped shore up the communist regimes of Eastern Europe. Kuran’s story also helps explain how a seemingly small discovery can, through the dynamics of the social influence process, snowball quickly into a major scientific revolution.

The power of social forces to enshrine manifestly false beliefs in science is not to be underestimated. The economics profession, for all its pretensions of being an objective science, is hardly immune. An example close to my own heart: In the face of overwhelming evidence to the contrary, our prevailing theories of welfare economics continue to be grounded on the view that satisfaction depends on absolute, not relative, income. These head-in-the-sand theories insist that higher taxes on the highest earners will diminish welfare by causing people to work too little. By contrast, a welfare theory in which satisfaction depends on both absolute and relative income implies that people, left to their own devices, will tend to work too much. Although a preponderance of evidence favors the latter formulation, the celebrated tradeoff between equity and efficiency remains a cornerstone of the conventional wisdom. This helps to explain why, even in the face of burgeoning income inequality, the U.S. has the least progressive tax structure of any industrial nation. If economic historians of the future are puzzled by the persistence with which most economists once clung to such a transparently deficient model, they need only consult Timur Kuran’s book.

VII. Concluding Remarks

Private Truths, Public Lies is a splendid book. It tackles a long list of interesting and important questions that have been discussed at length, and largely unsuccessfully, by scholars from each of the social sciences. The narrow rational choice model simply cannot answer many of these questions. Psychological theories by themselves cannot even address many of them. And sociological theories that take the group as the unit of analysis have made little progress. Kuran patiently and intelligently blends the insights of these disciplines into a behavioral model that moves the discussion forward on many fronts. If his is not the model that many traditional economists might have chosen, it is at least one that most can live with—and a good thing, too, because living with it is likely to prove easier than persevering with the traditional model.

REFERENCES


