

# WILEY



---

Preference Falsification, Policy Continuity and Collective Conservatism

Author(s): Timur Kuran

Source: *The Economic Journal*, Vol. 97, No. 387 (Sep., 1987), pp. 642-665

Published by: [Wiley](#) on behalf of the [Royal Economic Society](#)

Stable URL: <http://www.jstor.org/stable/2232928>

Accessed: 13/06/2013 01:47

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley and Royal Economic Society are collaborating with JSTOR to digitize, preserve and extend access to *The Economic Journal*.

<http://www.jstor.org>

## PREFERENCE FALSIFICATION, POLICY CONTINUITY AND COLLECTIVE CONSERVATISM\*

Timur Kuran

In a moment of frustration over his country's economic woes, Leonid Brezhnev once complained of the tendency of Soviet enterprise managers to shy away from innovation 'as the devil shies away from incense'.<sup>1</sup> Implicit in this lamentation are two distinct attributions. There is the claim that the Soviet economy's lacklustre performance is caused by the failure of enterprises to alter their routines. And there is the allegation that this failure reflects an attachment to the status quo.

This paper seeks to explain why collective decisions are path-dependent – why, as Brezhnev's complaint implies, policy decisions bear the influence of previous policy decisions. The issue to be explained is not simply that one observes *policy continuity* over time. It is that this continuity is generated, at least partly, by society's attachment to its past choices, an attachment we may label as *collective conservatism*.

Of the terms introduced, the first has only a descriptive meaning, the latter a causal one. Policy continuity describes a congruence between choices in consecutive periods. Collective conservatism, on the other hand, refers to a causal process – a process analogous to hysteresis in physics. Thus, the fact that a particular economic prohibition of yesterday is still in effect today constitutes an instance of continuity; conservatism is involved only to the extent that this prohibition exists today *because* it existed yesterday.

The theory of collective conservatism to be developed is based on the observation that because of group pressures, the policy preferences people express in public often differ from those they hold privately. We shall see that under numerous plausible conditions the possibility of *preference falsification* yields multiple equilibrium distributions of public preferences, each associated with a different policy, and each attainable, given the right initial expectations, through a bandwagon process. What gives rise to collective conservatism is that the distribution in one period affects individuals' expected payoffs in the next. Among the contributions of the paper is a measure of collective conservatism, designed to quantify the status quo's influence.

An essential feature of the argument is the observation that people rely on the prevailing climate of opinion in developing the personal belief systems that

\* For their fruitful suggestions, I am grateful to Reuven Brenner, Mark Granovetter, Vibhooti Shukla, Jacques Silber, Karol Soltan, Ulrich Witt, and two anonymous referees of this JOURNAL. I also benefited from the comments of numerous others during presentations at the University of Southern California, Virginia Polytechnic Institute, the University of Montreal, California State University at Fullerton, and the March 1986 convention of the Public Choice Society, held in Baltimore. Under grant no. SES-8509234, my work was supported in part by the National Science Foundation of the United States.

<sup>1</sup> *Pravda*, March 31, 1971. As quoted by Berliner (1976, p. 437).

underlie their private policy preferences. With this climate being formed by the justifications offered for preferences expressed publicly, it turns out that a by-product of preference falsification may be a shift in the distribution of private preferences in favour of the status quo.

Through a unified framework, the paper thus illuminates not only why a policy that few people support privately may command an overwhelming public endorsement, but also why, once this policy is in place, the degree of private opposition will diminish. The argument is consistent both with the claim that societies retain many policies they would rather change and with the claim that policies they inherit from the past shape their members' views and wants.

The theory provides an individual-choice based interpretation for an array of phenomena on which other works that deal systematically with collective conservatism are silent.<sup>2</sup> The new insights afforded are illustrated by means of an inquiry into India's caste system, which, throughout the social sciences, is routinely offered as the quintessential example of the past's reign over the present. Among the particular facts explained are that caste councils reach decisions through open rather than secret voting; and that the so-called untouchables, situated at the very bottom of the caste hierarchy, tend to be among the system's most committed supporters. The latter issue is especially significant in view of the fact that social thinkers have long been intrigued by people's propensity to accept servitude.

#### I. A MODEL OF COLLECTIVE CHOICE

The framework for the analysis is a simplified version of a model developed in a companion paper (Kuran, 1987*a*), where a fuller exposition may be found.

##### *Basic Features*

Consider a society faced with the task of selecting a collective policy,  $p$ , from within the unit interval. Its members, who are in agreement as to the unidimensional nature of the issue, fall into two categories: activists and non-activists. The activists are split into two pressure groups, one that advocates  $p = 0$  and another that advocates  $p = 1$ .<sup>3</sup> The non-activists, far greater in number, are not pre-committed to any particular policy; and regardless of what they privately prefer, they can be pressured into supporting one of these positions. The activists will remain in the background throughout most of the discussion that follows, so unless indicated otherwise, the term 'individual' will stand for 'non-activist'.

<sup>2</sup> For a critical survey of these works, see Kuran (1987*b*). Several imaginative and instructive contributions, including Hirschman (1970), Goldberg (1976), and Olson (1982), attribute collective conservatism to various social phenomena. Two others, Heiner (1983) and Brenner (1983), invoke personal conservatism, the individual's personal attachment to his own or society's past choices.

<sup>3</sup> The establishment of new pressure groups is ruled out. Although this simplification suppresses an important aspect of the problem at hand, it is justifiable in some contexts by the fact, elucidated by Olson (1982), that it is cheaper to maintain an existing coalition than to negotiate a new one.

The pressure groups vie for individuals' endorsements because society's decision depends negatively on  $s_0$ , the share of the individuals who support  $p = 0$ , and positively on  $s_1$ , the share who support  $p = 1$ . Naturally,  $s_0 + s_1 \leq 1$ . For expositional convenience, I assume that the decision function has the form

$$p = \frac{1}{2}(1 - s_0 + s_1), \quad (1)$$

according to which  $p$  necessarily lies in the interval  $[0, 1]$ . Note that this function accords no influence to preferences favouring intermediate policies. The justification for this feature is that the costs of preference aggregation lead societies to ignore preferences held by unorganised groups, especially if they lack numerical strength. Indeed, the administrative and legislative branches of government generally pay little attention to policy proposals that appear to lack the organised support of a sizeable group.<sup>4</sup>

#### *The Individual's Preference Declaration Decision*

Each individual decides what position to advocate publicly on the basis of three considerations.

First is his utility stemming directly from society's policy choice. This is measured by his 'direct benefit' function,  $B^i(p)$ , where  $i$  indexes the individual. The form of this function, which for the time being is taken to be predetermined, reflects the preference ordering he would provide in a secret ballot. I assume that there exists a unique policy that tops this ordering. This is  $x^i$ , his privately held preference, or simply, his *private preference*. Given that he constitutes an infinitesimal segment of society, the individual justifiably expects his own wish to have no perceptible influence on society's choice. This does not mean that he disregards his private preference. As we shall see, he accords it an important, though roundabout, role in his preference declaration decision.

The second consideration that enters the individual's decision process is the utility associated with his publicly declared preference, which I shall denote by  $y^i$ , and refer to as his *public preference*. This utility is captured by his 'reputation' function, which is given by

$$R(y^i) = \begin{cases} f_0(s_0) & \text{if } y^i = 0, \\ 0 & \text{if } 0 < y^i < 1, \\ f_1(s_1) & \text{if } y^i = 1, \end{cases} \quad (2)$$

where  $f_0(0) > 0$ ,  $f_1(0) > 0$ ,  $df_0/ds_0 > 0$ , and  $df_1/ds_1 > 0$ . Thus, the individual can obtain positive reputational utility only by backing an organised group; he cannot gain such utility merely by coming close to the group's position.<sup>5</sup>

<sup>4</sup> This observation conflicts, of course, with the rhetoric of contemporary constitutional democracy, according to which members of society have equal *political* rights. What this means in practice is that everyone enjoys equal *organisational* rights. Those who do not use their organisational rights have less say in the political process than those who do. In according zero influence to preferences held by no organised block, the function (1) encapsulates an extreme version of the observation.

<sup>5</sup> This feature is not essential to the argument. For reasons discussed in Kuran (1987*a*, sect. 3), all that is necessary is that the function have sufficiently sharp discontinuities at 0 and 1.

Another significant feature is that the utility conferred by a given pressure group increases with the size of this group's following. The rationale for this specification is that people fortify their reputation as supporter of a given cause by rewarding other supporters and by withdrawing favours from opponents.<sup>6</sup>

The final consideration in the individual's preference declaration decision is the utility he derives from integrity. To the extent that  $y^i$  differs from  $x^i$ , he compromises his integrity and thereby incurs a utility loss. His utility from integrity is represented by

$$N(x^i, y^i) = N(1 - |x^i - y^i|), \quad (3)$$

which is increasing in its argument,  $1 - |x^i - y^i|$ . This argument, which measures the proximity of the individual's public and private preferences, assumes a value between 0 and 1. The nearer they are, the closer the value to 1.

Given the multitude of non-activists, the individual non-activist has reason to expect his personal influence on the outcome of the collective choice process to be negligible. Thus, he effectively maximises a function such as

$$V^i(x^i, y^i) = R(y^i) + N(x^i, y^i), \quad (4)$$

which incorporates the reputational and integrity components of his utility function, but not the direct benefit component. Two points need to be recognised. First, the absence of  $B^i(p)$  from the right-hand side of (4) does not imply that the individual's private preference ordering has no impact on his public declaration; it does have an impact, since  $x^i$ , which depends on the shape of  $B^i(p)$ , appears in the argument of  $N(\cdot)$ . Second, although by assumption the non-activists have the same reputation and integrity functions, they may well have different private preferences and, hence, different maximands.

The share variables  $s_0$  and  $s_1$ , which enter (4) through the reputation function, are not necessarily known with precision. Let us assume, for simplicity, that everyone employs the same point estimates,  $\hat{s}_0$  and  $\hat{s}_1$ . Using these, each individual computes (4) under three alternatives: supporting  $p = 0$ , supporting  $p = 1$ , and revealing his private preference.<sup>7</sup> The corresponding utility levels can be denoted by  $V_0^i$ ,  $V_1^i$ , and  $V_*^i$ .

To present the argument clearly, I focus on the case where, for all  $i$ ,

$$\max(V_0^i, V_1^i) > V_*^i, \quad (5)$$

which is to say that supporting either  $p = 0$  or  $p = 1$  constitutes the dominant option for all the non-activists.<sup>8</sup> In this setting, expectations can be assumed to satisfy the condition

$$\hat{s}_0 + \hat{s}_1 = 1, \quad (6)$$

<sup>6</sup> This rationale is developed at length in Kuran (1987a, sect. 3). The gist is that an individual claiming to support a cause comes across as insincere unless his words are reflected in his behaviour. Thus, to be perceived as an opponent of apartheid, it is not sufficient to pay lip service to the principle of racial equality. Words must be buttressed with concrete actions, such as applauding an anti-apartheid speaker or demonstrating against a company doing business in South Africa.

<sup>7</sup> All other options are dominated by correct preference revelation, because  $R(y^i) = 0$  for  $0 < y^i < 1$ , and because  $N(x^i, y^i)$  is maximised when  $y^i = x^i$ .

<sup>8</sup> For a more general analysis, see Kuran (1987a).

under which  $V_0^i$  and  $V_1^i$  become<sup>9</sup>

$$V_0^i = f_0(1 - \hat{s}_1) + N(1 - x^i), \tag{7}$$

$$V_1^i = f_1(\hat{s}_1) + N(x^i). \tag{8}$$

Equating (7) with (8), we can compute, for each  $\hat{s}_1$ , the value that  $x^i$  must assume to make the individual just indifferent between supporting  $p = 0$  and supporting  $p = 1$ . This exercise amounts to defining a downward-sloping function  $x(\hat{s}_1)$ , such that, for any given  $\hat{s}_1$ ,  $V_0^i \cong V_1^i$  according as  $x^i \cong x(\hat{s}_1)$ .<sup>10</sup> The function is the same for all non-activists, because they all have the same reputation and integrity functions. It is illustrated in Fig. 1, according to which the individual supports  $p = 0$  if the point  $[\hat{s}_1, x^i]$  lies below the curve  $x(\hat{s}_1)$ , and  $p = 1$  if the point lies above it. Note that  $x(\hat{s}_1)$  crosses the  $\hat{s}_1$  axis at 0.8, which means that for  $\hat{s}_1 > 0.8$ , the individual supports  $p = 1$  regardless of where his private preference lies.

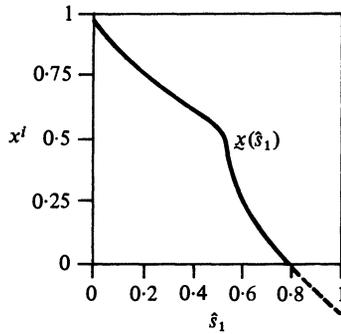


Fig. 1

*Equilibrium*

Now let  $\Phi(x)$  be the *cumulative density function* of the non-activists' private preferences. This function, depicted in Fig. 2 by a light curve, is defined to provide, for any  $x$ , the share of non-activists with a private preference greater than this  $x$ . The heavy curve, which I shall call the *threshold function* and denote by  $x(\hat{s}_1)$ , separates, for each  $\hat{s}_1$ , the private preferences for which supporting  $p = 0$  is optimal from those for which supporting  $p = 1$  is optimal. Its downward-sloping segment is precisely the function  $x(\hat{s}_1)$ , as depicted in Fig. 1. It also has a horizontal segment, which serves to indicate that if  $\hat{s}_1$  ever rises above 0.8,  $p = 1$  will enjoy unanimous support. (A horizontal segment along the top axis would signal that if  $\hat{s}_1$  ever became sufficiently low, all would support  $p = 0$ .) The reason for working with  $x(\hat{s}_1)$ , rather than  $x(\hat{s}_1)$ , is that this facilitates geometric interpretation.

<sup>9</sup> Insert (2) and (3) into (4), and then substitute  $\hat{s}_1$  for  $s_1$  and  $1 - \hat{s}_1$  for  $s_0$ . When evaluated at  $y^i = 0$  and  $y^i = 1$ , the resulting expression yields (7) and (8), respectively.

<sup>10</sup> To verify the slope of  $x(\hat{s}_1)$ , implicitly differentiate the equation  $V_0^i - V_1^i = 0$  to obtain  $dx/d\hat{s}_1 = - [Df_0(1 - \hat{s}_1) + Df_1(\hat{s}_1)] / [DN(1 - x) + DN(x)]$ , where D denotes the differential operator. By construction,  $Df_0(\cdot)$ ,  $Df_1(\cdot)$ , and  $DN(\cdot)$  are all positive.

Given  $\hat{s}_1$ , the actual shares supporting  $p = 1$  and  $p = 0$  turn out to be

$$s_1 = \Phi[\underline{x}(\hat{s}_1)], \tag{9}$$

$$s_0 = 1 - \Phi[\underline{x}(\hat{s}_1)]. \tag{10}$$

If  $s_1 \neq \hat{s}_1$ , both expectations are falsified, and the system is in disequilibrium. In Fig. 2, such is the case when  $\hat{s}_1 = 0.4$ : the dotted arrows indicate that the corresponding  $s_1$  is 0.25. In the event of disequilibrium, I assume,  $\hat{s}_1$  is revised in the direction of  $s_1$  until an equilibrium is attained – in which, by definition, expectations are self-confirming. Later, we will distinguish between temporary and permanent equilibria, but for now the distinction is immaterial.

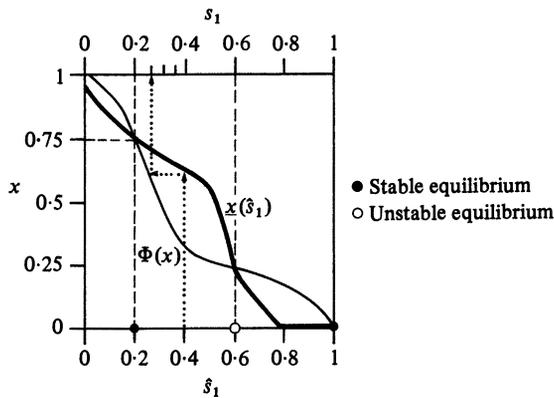


Fig. 2

The assumptions of the model guarantee the existence of an equilibrium, and there may be many.<sup>11</sup> The case shown in Fig. 2, for instance, features three equilibria. One of these is  $\hat{s}_1 = 0.2$ ; if everyone believes that exactly 20% will support  $p = 1$ , then the 20% whose private preferences exceed 0.75 will actually do so, while the 80% whose private preferences lie below 0.75 will support  $p = 0$ . It is stable, since an expectational displacement in either direction would generate further revisions resulting in its re-establishment. By analogous reasoning, one can show that  $\hat{s}_1 = 1$  also constitutes a stable equilibrium. The third equilibrium,  $\hat{s}_1 = 0.6$ , is unstable.

Which equilibrium is attained depends on individuals' initial share expectations, since these help determine how much pressure on behalf of each group is ultimately exerted. Not that the pressure groups' initial strengths are imma-

<sup>11</sup> To verify existence, let us consider in turn all possibilities as to the location of  $\underline{x}(\hat{s}_1)$  in the  $[\hat{s}_1, x]$  plane. (1)  $\underline{x}(\hat{s}_1)$  coincides with the axis  $x = 1$  or with  $x = 0$ : given that  $\Phi(x)$  begins at  $[0, 1]$  and descends toward  $[1, 0]$ , there is a unique equilibrium in each case, at  $\hat{s}_1 = 0$  in the former and at  $\hat{s}_1 = 1$  in the latter. (2)  $\underline{x}(\hat{s}_1)$  has a downward-sloping segment, and at least one horizontal segment at  $x = 1$  or  $x = 0$ : either  $\hat{s}_1 = 0$  or  $\hat{s}_1 = 1$  is an equilibrium, and, as in Fig. 2, there may be additional equilibria. (3)  $\underline{x}(\hat{s}_1)$  lies entirely below  $x = 1$  and entirely above  $x = 0$ : since both  $\underline{x}(\hat{s}_1)$  and  $\Phi(x)$  traverse the entire length of the  $\hat{s}_1$  range, and since  $\Phi(x)$  traverses the entire length of the  $x$  range as well, there is at least one equilibrium. Incidentally, the possibility of multiple equilibria is a feature shared by all models in which individuals' choices are interdependent. Other such models have been developed by Schelling (1978), Granovetter (1978), Akerlof (1980), and Arthur (1985).

terial. But the group that is relatively weaker at the outset can see its policy position gain approval if enough people support it during the crucial first stage. To put this concretely in terms of Fig. 2 and the notation of the reputation function, the fact that  $f_0(o) > f_1(o)$  would not preclude attainment of the rightmost equilibrium,  $s_1 = 1$ .

## II. POLICY CONTINUITY AND COLLECTIVE CONSERVATISM

Up to this point, we have examined the individual's preference declaration decision and the formation of self-confirming expectations. Continuing to treat private preferences as predetermined, we can turn now to our central concerns, policy continuity and collective conservatism. The conceptual arguments in this section will be quantified in Section III, after which, in Section IV, the model will be extended to make private preferences, too, endogenous.

Recall that to every pair  $[s_0, s_1]$  of actual shares corresponds a particular policy, given by the decision function (1). Since  $s_0 + s_1 = 1$ , this function reduces to

$$p = s_1. \quad (11)$$

Thus, in Fig. 2 the set of policies associated with equilibria is  $\{0.2, 0.6, 1\}$ .

Suppose, as a start, that the leftmost equilibrium has been attained and that society has adopted the corresponding policy, 0.2. If the threshold and cumulative density functions remain fixed, this policy will exhibit complete continuity: period after period, society will choose  $p = 0.2$ . Is this an indication of collective conservatism? Yes, in the sense that if past realisations of the underlying shares were suddenly forgotten,  $p = 0.2$  would not necessarily be retained. In particular, any initial  $s_1 \geq 0.6$  would lead to the adoption of  $p = 0.6$  or  $p = 1$ . But since an individual's utility depends on the shares, he would not just forget their realisations – least of all recent ones that have confirmed his prior expectations. With everyone fitting into this mould, self-confirming expectations will be retained indefinitely, and the other two viable policies,  $p = 0.6$  and  $p = 1$ , will not even get tested. What if, though, society had originally adopted either  $p = 0.6$  or  $p = 1$ , instead of  $p = 0.2$ ? By the same logic, it would hold onto this particular policy, thereby exhibiting collective conservatism.

Returning for a moment to the case discussed, I ought to point out that the 80% who support  $p = 0$  in one period do so in the next as well, and that the same goes for the 20% who support  $p = 1$ . This does not imply that an individual is personally attached to his own earlier choices. Free of *personal conservatism*, he does not give special consideration to past choices in maximising his utility.

Let us consider now the consequence of disturbing the system depicted in Fig. 2. A disturbance could entail a shift in either the cumulative density function of private preferences, or the threshold function, or both. The density function could shift because of a technological discovery, a change in the physical environment, a price shock, the emergence of a new externality, the influx of ideas from other societies, a transformation in a pressure group's efficiency in transmitting information, or population growth, among other

possibilities.<sup>12</sup> The threshold function could shift because, for instance, a group becomes more efficient at punishing its opponents, or the non-activists come to derive greater utility from integrity on the issue in question.<sup>13</sup>

The set of equilibria is not necessarily altered by such functional shifts. But even when it is, the observed shares of support may remain unaffected. Moreover, if the shares do change, they do so under the influence of the past.

Suppose, to illustrate the argument, that in Fig. 2 the established equilibrium is  $\hat{s}_1 = 1$ . Due to exogenous shocks, everyone eventually comes to perceive a greater direct benefit from  $p = 0$  than from  $p = 1$ , causing the cumulative density function to take the form given in Fig. 3. All individuals now have

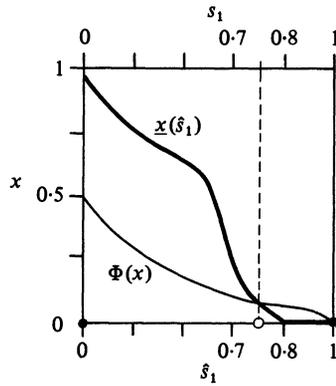


Fig. 3

private preferences between 0 and 0.5, and if a secret vote were taken, the average would turn out well below 0.5. Observe, however, that the pre-disturbance equilibrium,  $\hat{s}_1 = 1$ , is contained in the post-disturbance set of equilibria,  $\{0, 0.7, 1\}$ . This means that no one will announce a change of heart, unless a sufficient number of others do so first. Individuals most eager to switch – those with private preferences at 0 – will do so only if they believe that 20% have already defected. And even this belief would not lead to the establishment of a new equilibrium. Only if people somehow became convinced that over 30% were supporting  $p = 0$  would enough people switch to put in motion a bandwagon process toward the other stable equilibrium,  $\hat{s}_1 = 0$ . But this expectation is unlikely to emerge and spread, because nobody is willing to take the lead in publicising his opposition to the status quo. Society is stuck, therefore, at the equilibrium  $\hat{s}_1 = 1$ , and  $p = 1$  is retained.

In this example, many non-activists come to feel enchained by the existing policy. Period after period, however, they vote for their chains by keeping the shift in their private preferences *private*. It is critical to recognise that the

<sup>12</sup> The last source presents no analytical difficulties, because shares, not absolute numbers, enter individuals' maximands and society's policy function.

<sup>13</sup> A pressure group can help bring out the latter type of change by spreading the message that people who conceal their private preferences are in some sense inferior. Foreign trade is an issue conducive to this sort of symbolism. Protectionist groups sometimes succeed in convincing people that foreign raids on domestic markets constitute an issue that patriotic, dignified, honourable people do not keep quiet about.

enchainment felt by individual non-activists is not attributable entirely to the activists advocating  $p = 1$ . The pressure to support  $p = 1$  comes partly from the non-activists themselves, as each chooses, in a rational and voluntary effort to establish an advantageous reputation, to help make life difficult for those who fail to support  $p = 1$ . An important implication is that once a group of activists has obtained widespread support for its agenda, it may no longer itself have to punish its opponents. The non-activists may ensure that few people, if any, proclaim a desire for change. In fact, the original group could wither away after a while, without altering the non-activists' incentives to support the status quo. Witness how Sunni Islam, at one time tightly controlled by a centralised hierarchy, but for many years leaderless, remains an immense conservative factor in diverse policy contexts in a region stretching from Morocco to Indonesia.

In the case just analysed, the population was fixed. In an alternative scenario, a portion of the population dies in each period, to be replaced by a new generation. Under this interpretation, the shift in the distribution of private preferences is caused not by individual change but by generational replacement – the replacement of older cohorts by younger, different-minded cohorts. After a while, those originally responsible for selecting  $p = 1$  would all be dead. Their legacy would live on, however, moulding the public preferences of successive new generations. For an illustration of the legacy of dead generations, consider the Soviet agricultural system. Few of those who helped establish this system, either as committed instigators or as cowed supporters, are still alive. Yet a choice they made in the 1930s continues to guide Soviet agricultural policy in the 1980s.

A second case where the pre-disturbance equilibrium,  $\hat{s}_1 = 1$ , is contained in the post-disturbance set of equilibria is portrayed in Fig. 4. Here the density function is the same as in Fig. 2, but the threshold function has shifted to the right, because the group advocating  $p = 0$  has become more efficient at delivering reputational utility to its supporters. Again, no one shifts to  $p = 0$ , and society remains in equilibrium at  $\hat{s}_1 = 1$ . All would support  $p = 0$  if enough others were to do so, but this does not become known.

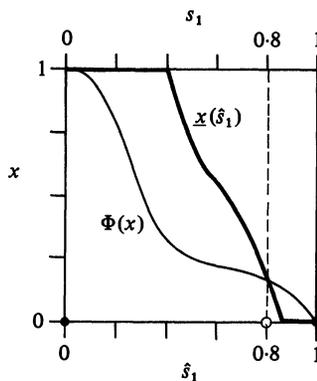


Fig. 4

Let us take up, finally, a case where the pre-disturbance outcome is not contained in the post-disturbance set of equilibria. Such a case is provided by Fig. 5, where, relative to Fig. 2, the cumulative density function has shifted in such a way as to move the leftmost equilibrium rightward. So suppose society was previously in equilibrium at  $\hat{s}_1 = 0.2$ . After the shift, this expectation is no longer sustainable, since, as the dotted arrows in Fig. 5 indicate, it causes the actual share to be 0.3. In accordance with the adjustment pattern outlined in Section I,  $\hat{s}_1$  will rise toward 0.3. But the new expectation will also prove to be unsustainable, and further upward revisions will be necessary until  $\hat{s}_1$  reaches 0.4. At this point, a new equilibrium is in place. Note that  $\hat{s}_1 = 0.4$  is the post-disturbance equilibrium closest to the pre-disturbance equilibrium,  $\hat{s}_1 = 0.2$ . This outcome, a reflection of the fact that the pre-disturbance equilibrium determines the initial post-disturbance expectation, ensures that the policy response is the smallest possible response.<sup>14</sup>

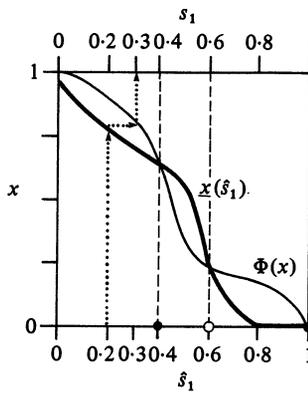


Fig. 5

### III. MEASURES

In each of the cases covered, society's policy choice is influenced by the status quo, which means, by definition, that it exhibits collective conservatism. But the argument developed thus far does not allow us to say whether collective conservatism is greater in one case than another. Nor does it allow us to quantify policy continuity.

Quantification requires the introduction of time subscripts as well as additional notation. So let  $p_t$  and  $p_{t-1}$  denote the policies actually in place at  $t$  and  $t-1$ ; and  $\hat{p}_t$  the policy society would have been expected to choose at  $t$  had it forgotten  $p_{t-1}$  and the underlying distribution of public preferences. Also, let  $d(p, p')$  represent the Euclidean distance between policies  $p$  and  $p'$ ; and  $D(p)$  the greatest distance between  $p$  and any other policy in the unit interval.

<sup>14</sup> According to (11), the set of viable policies is  $\{0.4, 0.6, 1\}$ . The member selected,  $p = 0.4$ , is the one closest to the old policy,  $p = 0.2$ .

*Policy Continuity*

In terms of this notation, a simple measure of policy continuity is

$$\gamma_t = \frac{D(p_{t-1}) - d(p_{t-1}, p_t)}{D(p_{t-1})}. \quad (12)$$

When  $p_t = p_{t-1}$ , the measure equals 1. When  $p_t$  differs from  $p_{t-1}$ , but by less than the maximum possible, it lies between 0 and 1. Continuity is *complete* in the former instance, *partial* in the latter. The measure is bounded below by 0.

It can easily be checked that in all the cases discussed in Section II, except the last, the degree of policy continuity is 1. In the final case, where the pre- and post-disturbance policies are 0.2 and 0.4, respectively, we find

$$\gamma_t = \frac{0.8 - 0.2}{0.8} = 0.75.$$

This result has a simple interpretation: the observed policy change is one fourth as large as the greatest possible change, which is to say that 75% of the possible change has not materialised. Note that the measure carries no causal connotation. It provides only descriptive information.

*Collective Conservatism*

A meaningful measure of collective conservatism is

$$\sigma_t = \frac{d(p_{t-1}, \tilde{p}_t) - d(p_{t-1}, p_t)}{D(p_{t-1})}. \quad (13)$$

According to this measure, the degree of collective conservatism is positive if, while no policy change is observed, some change would have been expected had realisations in  $t-1$  of the pertinent variables been forgotten. It is zero, on the other hand, if  $p_t = \tilde{p}_t$ , that is, if the policy actually chosen at  $t$  would also have been the expected choice under historical amnesia. The measure's range is  $-1 \leq \sigma_t \leq 1$ , although only the non-negative segment is relevant here. Note that  $t-1$  and  $t$  refer to points in time at which society is in equilibrium.<sup>15</sup> This means that disturbances and expectational adjustments, if any, take place in subperiods lying between  $t-1$  and  $t$ . For the time being, these subperiods will not concern us.

It is necessary to specify how  $\tilde{p}_t$  is found. But before doing this, let us introduce a variant of (13). Setting  $p_t = p_{t-1}$ , one obtains

$$\sigma_t^I = \frac{d(p_t, \tilde{p}_t)}{D(p_t)}, \quad (14)$$

<sup>15</sup> Like  $\gamma_t$ ,  $\sigma_t$  is sensitive to the unit of time: by altering it, one may affect  $p_{t-1}$ , which enters both measures. As an example, consider affirmative action policy in the United States. Both measures will be substantially higher if the unit period is a year than if it is a half a century. The inevitable arbitrariness in these measures calls for caution in interpretation.

a measure of *instantaneous* collective conservatism. When  $\gamma_t = 1$ ,  $\sigma_t = \sigma_t^I$ . The two measures may differ, however, when  $\gamma_t < 1$ . In such contexts, as we shall see shortly, the two measures provide different information.<sup>16</sup>

The hypothetical variable  $\tilde{p}_t$  is the weighted arithmetic average of the policies that could be selected under historical amnesia, where the weight associated with each policy is the probability of initially forming an expectation  $\hat{s}_1$  that would generate this policy. In determining this probability, it is appropriate, without a justification for viewing some initial expectations as more likely than others, to accord every possible initial expectation an equal likelihood.<sup>17</sup> Accordingly, I shall suppose that under a veil of historical ignorance the initial  $\hat{s}_1$  is distributed uniformly between 0 and 1.

### Illustrations

Equipped with measures, and with a specification for  $\tilde{p}_t$ , we can turn back now to Figs. 2-5 and quantify the collective conservatism implied by each of the portrayed cases. With respect to Fig. 2, observe the following: an initial expectation such that  $0 \leq \hat{s}_1 < 0.6$  leads to the policy  $p = 0.2$ ; one of  $\hat{s}_1 = 0.6$  to  $p = 0.6$ ; and one such that  $0.6 < \hat{s}_1 \leq 1$  to  $p = 1$ . Given the uniform distribution assumption, it follows that under historical amnesia the expected policy choice would be

$$\tilde{p}_t = d(0, 0.6) (0.2) + d(0.6, 0.6) (0.6) + d(0.6, 1) (1) = 0.52.$$

Making the appropriate substitutions into (14), we find that for the leftmost equilibrium the degree of instantaneous collective conservatism is

$$\sigma_t^I = \frac{|0.52 - 0.2|}{\max(|0.2 - 0|, |0.2 - 1.0|)} = 0.4.$$

Table 1 lists  $\sigma_t^I$  for each of the equilibria in Figs 2-5. The essential point to note is that the degree of instantaneous collective conservatism associated with an equilibrium is higher the smaller the probability that under historical amnesia it would have been attained. In Fig. 2, for instance,  $\sigma_t^I$  is higher for 1 than for 0.2, because if all initial expectations are equally probable, 1 is less likely to be selected than 0.2.

The table also illustrates how functional shifts affect the degree of instantaneous collective conservatism. Take the situation where the  $\sigma_t^I$  associated with  $p = 1$  rises from 0.48 in Fig. 2 to 0.7 in Fig. 3. This has a simple interpretation: as individuals' private preferences move away from  $p = 1$ , the range of initial expectations generating this policy narrows, implying that its retention comes to depend more heavily on the pull of the past.

<sup>16</sup> By construction, each of the three measures depends on chosen policies, and not on distributions of public preferences. But these distributions do play a role, since, according to the decision function (1), they determine each period's policy. If society's decision function were different, the measures themselves would remain the same, although they could take on different values.

<sup>17</sup> My argument here is analogous to Borda's (1781) celebrated argument that in a rank-order voting procedure distances should be equal in the absence of a convincing reason for introducing differences. While the equality assumption is arbitrary too, it has the virtue of making clear what is being assumed.

Table 1  
*Degrees of Instantaneous Collective Conservatism: Equilibria in Figs. 2-5*

Figure	$p_t$	$\hat{p}_t$	$\sigma_t^I$
2	$\left. \begin{array}{l} 0.2 \\ 0.6 \\ 1 \end{array} \right\}$	0.52	$\left\{ \begin{array}{l} 0.4 \\ 0.13 \\ 0.48 \end{array} \right.$
3	$\left. \begin{array}{l} 0 \\ 0.7 \\ 1 \end{array} \right\}$	0.3	$\left\{ \begin{array}{l} 0.3 \\ 0.57 \\ 0.7 \end{array} \right.$
4	$\left. \begin{array}{l} 0 \\ 0.8 \\ 1 \end{array} \right\}$	0.2	$\left\{ \begin{array}{l} 0.2 \\ 0.75 \\ 0.8 \end{array} \right.$
5	$\left. \begin{array}{l} 0.4 \\ 0.6 \\ 1 \end{array} \right\}$	0.64	$\left\{ \begin{array}{l} 0.4 \\ 0.1 \\ 0.36 \end{array} \right.$

Another instructive illustration is provided by the last case analysed in Section II, the one where the equilibrium  $\hat{s}_1 = 0.2$  in Fig. 2 gives way, following a disturbance, to  $\hat{s}_1 = 0.4$  in Fig. 5. This means, according to Table 1, that  $\sigma_{t-1}^I = 0.4 = \sigma_t^I$ . The reason for the identity is that the range of  $\hat{s}_1$  leading to  $p = 1$  is precisely  $0.6 < \hat{s}_1 \leq 1$  in both the pre- and post-disturbance situations. What about  $\sigma_t$  in this particular case? From (13) we find

$$\sigma_t = \frac{|0.64 - 0.2| - |0.4 - 0.2|}{\max(|0.2 - 0|, |0.2 - 1.0|)} = 0.3,$$

which, as one would expect, is less than both  $\sigma_{t-1}^I$  and  $\sigma_t^I$ . Recalling that (13) and (14) are equivalent in the absence of policy change, we can now suppose, to extend the illustration, that  $p_{t-2} = 0.2$  and  $p_{t+1} = 0.4$ . It follows that  $\sigma_{t-1} = 0.4$ ,  $\sigma_t = 0.3$ , and  $\sigma_{t+1} = 0.4$ . The dip at  $t$  reflects the policy change that has just occurred.

#### IV. ADAPTATION OF PRIVATE PREFERENCES

Up to this point, people's private preferences have been predetermined. This has meant that it would take an exogenous shock to shatter an established equilibrium and alter the degree of collective conservatism.

Removing the assumption that private preferences are exogenous to the system, I shall now develop the paradoxical argument that the persistence of preference falsification can cause the degree of conservatism to fall. The crux of this argument is that following a policy's adoption, an individual privately opposed to it might come to support it both publicly and privately. The argument rests on the observation that the evolution of an individual's private preference is guided by justifications others give for their public preferences.

The claim that chosen policies shape people's wants is not new. It has been

advanced by scores of renowned thinkers, among them Bagehot (1884/1956, chs. 3, 5), Veblen (1919, ch. 1), Freud (1921/59, chs. 1, 2, 9), and Keynes (1936, pp. 383–4); and it constitutes a major proposition of modern sociology. The contribution of this section lies, then, not in the claim itself, but rather, in the particular argument it furnishes in support of the claim.

My argument relies on the fact that people's cognitive abilities are bounded. Other arguments that rely on cognitive factors to explain why people come to favour established policies they once disliked have been advanced by Festinger (1957), Hirschman (1967), von Weizsäcker (1971), and Elster (1983, ch. 3). In them, preference adaptations are caused by 'tricks of the mind' aimed at mitigating cognitive inconsistencies. Here, we shall see, they stem from people's need to draw information from each other's declared beliefs.

### *Belief Systems*

Recall that the individual's private preference ordering is represented by a direct benefit function,  $B^i(p)$ , with a unique maximum at  $x^i$ , his private preference. On the basis of extensive research by psychologists, I postulate now that  $B^i(p)$  is governed by a personal belief system – a mental model – which maps each policy alternative into relevant consequences.<sup>18</sup> Through this belief system the individual combines available information and predicts the effects of each alternative – effects he can order according to their attractiveness. To a given issue, individuals with identical innate dispositions may bring different belief systems that generate different private preferences.

Because of his cognitive limitations, a person is able to formulate educated belief systems concerning a minute portion of the phenomena that bear on his happiness. Out of biological necessity, he must rely largely on beliefs conveyed by others. This dependence is the focus of a vast segment of the psychology literature. From our standpoint, the significant finding is that for any given opinion, frequency of exposure serves as a major criterion of validity.<sup>19</sup> The number of repetitions an individual hears of an opinion is likely to depend on the number of people conveying it. One would expect, in particular, that the greater the number of people who appear to hold a given opinion, the more validity it will assume.<sup>20</sup>

It is necessary at this point to distinguish between an individual's *private belief system* and his *public belief system*. Embodying his true convictions, his private belief system is what underlies his direct benefit function. It may never become

<sup>18</sup> A survey of this research is provided by Markus and Zajonc (1985).

<sup>19</sup> See Hasher *et al.* (1977), Montmollin (1977), and Schwartz (1982). One explanation for this finding is that in their deliberations, individuals rely on what Tversky and Kahneman (1973) call the availability heuristic, a mental shortcut whereby subjective familiarity gets equated with validity. Another is that they believe, as implied by the saying 'four eyes see better than two', that people are unlikely to fall into identical errors.

<sup>20</sup> This was recognised by James Madison (1787–8/1961, p. 349), a founding father of the United States. He wrote: '[T]he strength of opinion in each individual, and its practical influence on his conduct, depend much on the number which he supposes to have entertained the same opinion. The reason of man, like man himself, is timid and cautious when left alone, and acquires firmness and confidence in proportion to the number with which it is associated.'

known to others. His public belief system, on the other hand, embodies the convictions he conveys in public. Like his public preference, it is influenced by reputational considerations.

The distinction just drawn implies that the individual engages in *belief falsification*. It is reasonable to assume that this mirrors his preference falsification, since in many contexts appropriate argumentation is a precondition for meaningful preference declaration. Indeed, to make his preference declaration convincing, a person will generally need to back up his endorsement with appropriate substantive arguments. He must provide reasons, that is, as to why he expects the policy he ostensibly supports to fulfil his objectives most closely; his public belief system must be geared toward making his public preference look reasonable. An example may help clarify what the assumption involves. Someone who says he favours an import quota for textiles, but who gives the impression that he subscribes unequivocally to the doctrine of free trade, will fail to convince his audience that he really favours a quota. To be convincing, he must offer a reason, like the high costs of resource reallocation, as to why a departure from this doctrine is in this instance desirable.

Four new assumptions have been introduced, which, as we turn to their implications for collective conservatism, bear reiteration. First, individuals rely on each others' beliefs. Second, the relative influence of a particular belief depends on the share of society that asserts it. Third, an individual disguises a private belief when he expects thereby to benefit. And fourth, his belief falsification mirrors his preference falsification.

### *Formalisation*

To proceed, we need slightly more elaborate notation. So, for time  $t$ , let  $x_t^i$  be the individual's private preference;  $\bar{y}_t$  the arithmetic average of all public preferences; and finally,  $s_{0,t}$  and  $s_{1,t}$  the shares supporting  $p = 0$  and  $p = 1$ . Given that all individuals support either  $p = 0$  or  $p = 1$ , it follows that

$$\bar{y}_t = (0) s_{0,t} + (1) s_{1,t} = s_{1,t}. \quad (15)$$

On the basis of the first two of our four new assumptions, we can postulate that the individual's private preference evolves according to the relationship

$$x_{t+1}^i = x_t^i h^i(\bar{y}_t - x_t^i), \quad (16)$$

where the function  $h^i(\cdot)$  is subject to the restrictions  $dh^i/d(\bar{y}_t - x_t^i) > 0$ ,  $h^i(0) = 1$ ,  $0 \leq h^i(-x_t^i)$ , and  $h^i(1 - x_t^i) \leq 1/x_t^i$ . The latter two restrictions serve to keep the individual's private preference within the unit interval. The former two imply that  $x_{t+1}^i \cong x_t^i$  according as  $x_t^i \cong \bar{y}_t$ . Thus, individuals with private preferences below  $\bar{y}_t$  adjust them upward, while those with private preferences above  $\bar{y}_t$  adjust them downward.

For expositional clarity, it is useful to specify a temporal scheme for the various adjustments under consideration. So suppose that the unit period contains two subperiods. In the first of these, which is of length  $\epsilon < 1$ , individuals' private belief systems and, hence, their private preferences adapt to

the most recent distribution of public opinions. In the second, which is of length  $1 - \epsilon$ , the distribution of private preferences is fixed, and individuals' public preferences adapt and readapt until an equilibrium is attained. The postulated temporal scheme is depicted in Fig. 6 for two full periods. At the end of each full period, the figure indicates, public preferences are in *temporary* equilibrium. The equilibrium is not necessarily *permanent*, for it may be destroyed once private preferences adapt to it.

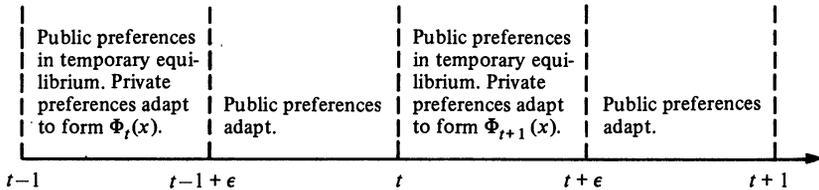


Fig. 6

We are prepared now to explore the evolution of private preferences and the implications of this evolution for society's policy choice and the degree of collective conservatism. It is instructive to begin with the case of corner equilibria. Later, we shall take up the more challenging case of interior equilibria.

*Corner Equilibria*

Let us turn back to Fig. 5 and suppose that the equilibrium  $\hat{s}_1 = 1$  is in place. According to (15) and (16), all individuals' private preferences will gravitate over time to  $x = 1$ . The cumulative density function will shift upward and to the right, therefore, as shown in Fig. 7. Ultimately, everyone's private preference will equal 1, and the function will take the form of an inverted L.

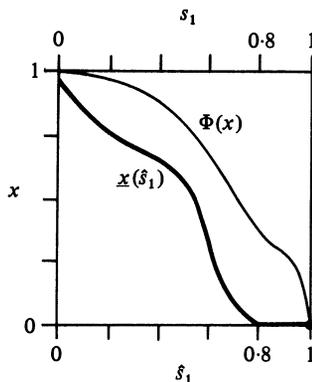


Fig. 7

What becomes of the degree of collective conservatism? After elimination of the interior equilibria, the set of viable policies contains a single member,  $p = 1$ . From (13) it follows that beyond this point in time,  $\sigma_t = 0$ . But this result

does not generalise: although the degree of collective conservatism always converges to some lower bound, this bound is not necessarily zero. Before explaining why, though, I shall interpret the argument thus far.

### *Interpretation*

It is paradoxical that when a policy becomes fixed, the degree of collective conservatism can fall to zero. But the explanation is simple. If the only self-confirming expectation is that associated with the status quo, the status quo would be retained even if people were to forget that  $p = 1$  commands unanimous support. The observed policy continuity owes nothing, therefore, to the pull of the past.

Let us be clear about the sequence of events involved. Once a public consensus forms in favour of  $p = 1$ , debate becomes one-sided, featuring arguments only in support of 1. This climate of opinion causes those privately opposed to amend their convictions. Eventually all come to see  $p = 1$  as the most desirable policy. At this point, society has lost touch, in effect, with the fact that its destiny could be different. It sees the status quo as self-explanatory – not an artifact of social experimentation, but inherent in nature itself.

The activists advocating  $p = 1$  are major beneficiaries of the described capture of the non-activists' minds. As we have seen, however, the key role in this capture is played not by them but by the mass of non-activists, who, by withholding their personal convictions from one another, distort the climate of opinion. This point is significant in view of the fact that policies benefiting special interest groups – sectoral subsidies, trade barriers – often receive the sympathy of a large majority. The argument just advanced links this puzzling phenomenon to the process by which members of the majority form their views of the world. This is not to say that the special interest groups play no role. They most certainly do, if only by manipulating the non-activists' information base. But such manipulation is possible precisely because the non-activists' cognitive limitations make their minds capturable.

The discussion should not be taken to imply that if  $p = 1$  ever gets established, it will be maintained forever. The whole exercise abstracts from factors pulling people's private preferences apart, such as changes in environmental conditions, the inevitable diversity of people's experiences, and opportunities for forming new pressure groups. In practice, the gravitation of preferences toward 1 could be arrested by a shock that propels private preferences away from 1.

Nothing has been said concerning the speed of the process by which private preferences respond to the climate of opinion. There is reason to believe, though, that this process can be slow: studies indicate that people exhibit resistance to information inconsistent with their belief systems.<sup>21</sup> Accordingly, the discussed sequence of events could play itself out over many decades or centuries, with shifts in the distribution of private preferences taking place primarily across generations.

<sup>21</sup> See, for instance, Edwards (1968).

*Interior Equilibria*

Turning once again to Fig. 5, we can now explore the transformation of an interior equilibrium. Suppose that at  $t = T$  the equilibrium in place is  $\hat{s}_{1, T} = 0.4$ . By (15), this means that  $\bar{y}_T = 0.4$ . It follows, by (16), that all private preferences below 0.4 will rise at  $T + 1$  and that all those below 0.4 will fall. How will the equilibrium be affected? The answer is illustrated in Fig. 8, where  $\underline{x}(\hat{s}_1)$  and  $\Phi_T(x)$  are as in Fig. 5. The latter function rotates counterclockwise around the point  $a$  (whose abscissa is 0.4). This yields  $\Phi_{T+1}(x)$ , which generates the equilibrium  $\hat{s}_{1, T+1} = 0.3$ .

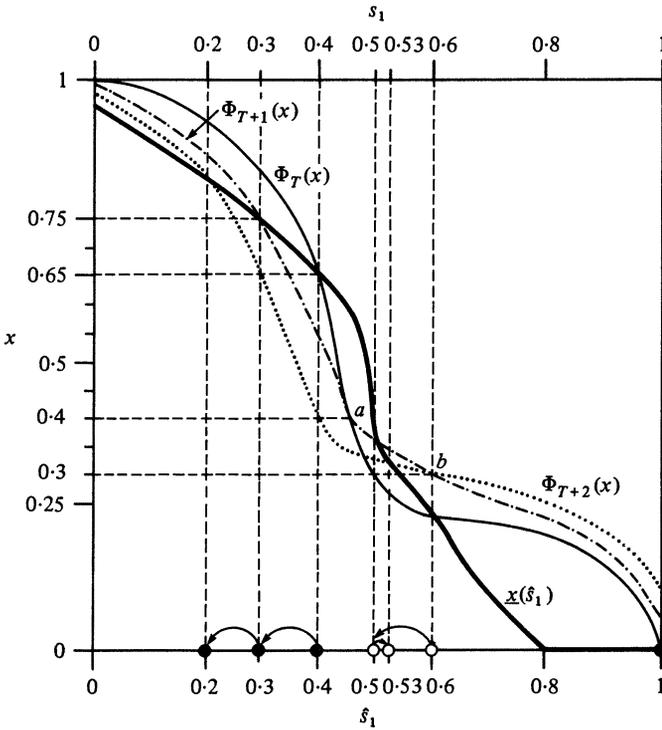


Fig. 8

The reason the equilibrium at  $T + 1$  lies to the left of that at  $T$  is that the intersection defining  $\hat{s}_{1, T} = 0.4$  is located on the portion of  $\Phi_T(x)$  that moves to the left at  $T + 1$  – the portion between  $x = 0.4$  and  $x = 1$ . The intersection defining  $\hat{s}_{1, T+1} = 0.3$  lies on the portion of  $\Phi_{T+1}(x)$  between  $x = 0.3$  and  $x = 1$ , so, by the same logic,  $\hat{s}_{1, T+1}$  must give way at  $T + 2$  to an equilibrium further to the left. Fig. 8 bears this out: a counterclockwise rotation of  $\Phi_{T+1}$  around the point  $b$  (whose abscissa is 0.3) yields  $\Phi_{T+2}(x)$  and the equilibrium  $\hat{s}_{1, T+2} = 0.2$ . Examination of the figure indicates that the equilibrium will continue to fall over time until, in some period  $T + N > T + 2$ , it reaches 0.

We see that the interior equilibrium  $s_{1,T} = 0.4$  gives way over a number of periods to a corner equilibrium,  $\hat{s}_{1,T+N} = 0$ . My earlier argument on corner equilibria suggests that from period  $T+N+1$  onwards, all private preferences will gravitate towards 0. Ultimately, the cumulative distribution function will become L-shaped, as shown in Fig. 9. The point to observe here is that in contrast to Fig. 7, there are equilibria other than the attained corner equilibrium. This is so because beyond  $\hat{s}_1 = 0.8$  the threshold function coincides with the  $x = 0$  axis. Even after all private preferences fall to 0,  $p = 1$  would receive unanimous support if  $\hat{s}_1$  were somehow to rise above 0.8.

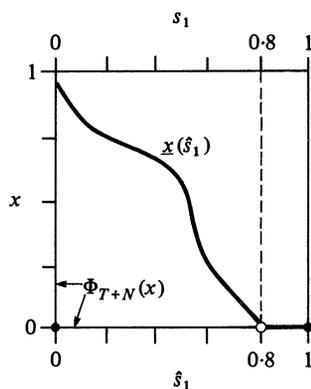


Fig. 9

It should now be apparent why the degree of collective conservatism associated with a corner equilibrium need not converge over time to 0. If the threshold function has a horizontal segment encompassing the *opposite* corner, it will converge to a positive number. In the present case, for instance, the degree of collective conservatism associated with the equilibrium  $\hat{s}_1 = 0$  converges to 0.2.<sup>22</sup>

This is not so say that from period  $T$  onwards the degree of collective conservatism must decrease *monotonically*. While it must ultimately fall to 0.2, it could rise for a while, depending on how shifts of the cumulative distribution function affect the other equilibria in the system. Table 2 shows that in the example depicted in Figs. 8 and 9, a rise does occur at  $T+2$ . It also shows that the degree of instantaneous collective conservatism rises at  $T+1$ , before falling at  $T+2$ . These outcomes reflect the fact that our measures depend on all policies supported by self-sustaining expectations, not just those society adopts. In our example, not only does the leftmost member of the set of equilibria change over time, but the middle member changes as well.

Relying as it does on a particular example, this analysis lacks generality. The issue deserves to be taken up in a setting equipped to isolate the effect on the

<sup>22</sup> In Fig. 9 an  $\hat{s}_1$  below 0.8 generates  $p = 0$ , and one above 0.8 generates  $p = 1$ . Therefore,  $\bar{p}_{T+N} = 0.2$ . Given that  $p_{T+N} = p_{T+N+1} = 0$  and that  $D(p_{T+N}) = 1$ , it follows from (13) that  $\sigma_{T+N+1} = 0.2$ .

Table 2

*Evolution of the Degree of Instantaneous Collective Conservatism and the Degree of Collective Conservatism: Figs. 8 and 9*

Period	Set of equilibria	$p_t$	$\bar{p}_t$	$\sigma_t^I$	$\sigma_t$
$T$	{0.4, 0.6, 1}	0.4	0.64	0.4	
$T+1$	{0.3, 0.5, 1}	0.3	0.65	0.5	0.25
$T+2$	{0.2, 0.53, 1}	0.2	0.576	0.47	0.251
$\vdots$					
$T+N+1$	{0, 0.8, 1}	0	0.2	0.2	0.2

measures of each specific change in the set of equilibria and to handle systematically variations in the number of equilibria. But an important point has been made, which is that even in the absence of exogenous shocks, both the degree of collective conservatism and the degree of instantaneous collective conservatism may reverse direction.

In our example, the prevailing equilibrium moved leftward. Under certain conditions, however, it would move rightward or stay fixed. The following relationship covers the possible cases:

$$\text{If } \bar{y}_t \leq x(\hat{s}_{1,t}), \text{ then, for } \hat{s}_{1,t} \text{ stable (unstable), } \hat{s}_{1,t+1} \leq (\geq) \hat{s}_{1,t}. \quad (17)$$

Note first that Fig. 8 is consistent with (17). Since the equilibrium  $\hat{s}_{1,T} = 0.4$  is stable, and since  $0.4 = \bar{y}_T < x(0.4) = 0.65$ , it must be the case, according to (17), that  $\hat{s}_{1,T+1} < \hat{s}_{1,T}$ . This is what we found. We also found that  $\hat{s}_{1,T+2} < \hat{s}_{1,T+1}$ , which follows from the fact that the equilibrium  $\hat{s}_{1,T+1} = 0.3$  is stable and that  $0.3 = \bar{y}_{T+1} < x(0.3) = 0.75$ . The second point to observe in (17) is that the status quo will not be disturbed if  $\bar{y}_t = \hat{s}_{1,t}$ . Even in this case, though, the distribution of private preferences will evolve, in that all private preferences will gravitate toward  $\bar{y}_t$ .<sup>23</sup>

#### *Another Implication*

The preceding discussion suggests that except in special circumstances, an interior equilibrium will give way to a corner equilibrium. The significance of this finding lies in the fact that a corner equilibrium has permanence: it is immune to endogenous private-preference adaptations. Once a corner equilibrium is established, in other words, there is no return. Not only does debate cease, but as a result, the possibility of future debate diminishes. I ought to re-emphasise that I have deliberately ignored processes that pull people's beliefs and preferences apart. In practice these interfere with the homogenisation process just analysed.

<sup>23</sup> Interestingly, this gravitation makes the equilibrium unstable – vulnerable, that is, to a disturbance. To verify this geometrically, observe that the cumulative distribution function becomes horizontal at  $x = \bar{y}_t$ , in which position it necessarily cuts the threshold function from below.

## V. THE THEORY APPLIED: INDIA'S CASTE SYSTEM

The value of a social theory lies in its ability to illuminate social phenomena. To demonstrate the merits of this one, I shall consider India's caste system, which has puzzled countless historians, anthropologists, sociologists, political scientists, and economists. My concern is with the process by which the system has for millennia been maintained.<sup>24</sup> It is a fact that defenders of the system have included the lower castes, even the so-called untouchables. I take it as a challenge to explain this intriguing fact.

A good starting-point is a theory developed by Akerlof (1976, 1980). It rests on two observations: first, that castes are economically interdependent; and second, that Indian society penalises the owners, operators, and consumers of firms that assign their low-caste employees to the lucrative tasks reserved for high castes. Everyone, according to the argument, recognises that the cost of production would decline if producers substituted low- for high-caste labour. Groups of firms, workers, and consumers wish, therefore, that they could collude to break the system. But the formation of anti-caste coalitions is hindered by the free-rider problem: given that the individual member of an economically viable coalition would be ostracised by the larger society from the moment he joined, his expected private gain from joining could be negative, even if he knew for sure that the coalition would form. With all potential members of a coalition fearing that it is doomed to failure, failure becomes a self-fulfilling prophecy, and the system survives unscathed.

Although enlightening, this theory overlooks the fact that social sanctions are aimed not just at actions against the system but also at expressions of disagreement. The evidence indicates that traditional Indian society discourages inquiries into the rationale for the caste system and that it aims to conceal disagreements, whether over general characteristics of the system or over its particular manifestations. In caste and village assemblies, protests and assertions of difference are discouraged, apparently in order to foster the image of a harmonious society. Conflicts over caste matters are often settled by caste leaders through deals made behind the scenes. Also significant is the fact that in meetings, voting takes place by a show of hands, not by secret ballot.<sup>25</sup>

The theory developed in Sections I and II sheds light on these realities. Caste leaders, who have a stake in the system, must expect to gain from the appearance of harmony, as this would lower would-be reformists' estimates of the potential opposition. The prevalence of open voting is attributable to the fact that opponents of the status quo can be coerced into falsifying their preferences in an open vote, but not in a closed vote that accords them anonymity. As for

<sup>24</sup> Students of the caste system have attributed its origin to factors ranging from labour shortages (Lal, 1985, chs. 2, 3) to ethnic differences in immunity to disease (McNeill, 1976, pp. 83-4). A survey of the competing theories is provided by Cox (1948, ch. 7). In this century a variety of groups have strenuously fought the system, and, consequently, its essential features are now illegal (Anant, 1972, ch. 2). As a matter of practice, however, the system is still strongly in place (Dumont, 1966/80, ch. 11).

<sup>25</sup> On these observations, see Dumont (1966/80, ch. 8) and Cox (1948, ch. 6).

the discouragement of inquiry, it betrays a fear that questions will be interpreted as a sign of dissatisfaction with the status quo.<sup>26</sup>

These arguments link the stability of the caste system to the fact that its potential opponents do not air their opposition and doubters of its wisdom do not publicise their doubts. To the extent that these factors do come into play, the existence of economically viable anti-caste coalitions will remain a secret.

A related problem with Akerlof's theory is that it assumes, in effect, that most Indians feel shackled by the prevailing system. But in reality, even the untouchables tend not to consider themselves oppressed. Regarded as 'polluted', they are barred from living in the village proper, from drawing water from the village well, and from entering Hindu temples. Yet they tend to consider these restrictions neither exploitative nor offensive.<sup>27</sup> Many an untouchable apparently believes, in accordance with the doctrine of Karma, that his inferiority is the result of mistakes he committed in his former lives, and that if he accepts his present station and patiently fulfils his duties, he will move into a higher caste in his next life. Having imbibed the Hindu teachings about reincarnation and the inter-caste mobility of souls, he genuinely feels that he is best off working with the system, not fighting it.<sup>28</sup> In terms of the model of this paper, his direct benefit is maximised by retention of the caste restrictions.

Why have the untouchables continued, generation after generation, to accept a set of beliefs that sanctifies their subjection and degradation? Akerlof's theory provides no answer. Such beliefs do not have a place in it, except as an exogenous factor influencing people's payoffs. A possible explanation is offered, however, by the argument in Section IV.

To get started, let us travel back a couple of millennia, to a time when the system was still in formation. We know that at first various groups fought the restrictions placed on them.<sup>29</sup> Evidently, alternative systems were openly being considered, which suggests that under the right expectations concerning public preferences, some other system might have been adopted and retained. If this inference is valid, it follows that the degree of collective conservatism associated with the caste system's retention was once large.

Moving forward in time, we begin to observe that punishments are meted out to those openly proclaiming their opposition to the system, even those simply questioning its wisdom. These punishments ensure that most opponents keep their private preferences and beliefs to themselves. As a result, new generations grow up hearing much in favour of the system and almost nothing against it. Their thought processes vitiated by the climate of opinion, they come to see the desirability of the inherited order as self-evident. Reaching the modern

<sup>26</sup> A person does not ordinarily raise questions about matters he regards as fully settled. He does not inquire into the wisdom of the implicit policy of allowing people to have a roof over their heads, because it never enters his mind that the policy might be undesirable.

<sup>27</sup> The restrictions are outlined by Dumont (1966/1980, pp. 46-9, *passim*).

<sup>28</sup> Extensive evidence concerning the untouchables' preferences is provided by Moore (1978, pp. 55-64). See also Cox (1948, chs. 1, 2).

<sup>29</sup> See Lal (1985, ch. 2).

era, we find that most Indians genuinely support the system, and correspondingly, that the degree of collective conservatism is nearly zero.<sup>30</sup>

This explanation for the caste system's persistence differs fundamentally from some theories that enjoy great popularity in India. In these theories, some of which bear the influence of contemporary Marxism, the stability of the caste system and of the beliefs associated with it are attributed simply to the power of the dominant castes.<sup>31</sup> Here, in contrast, the focus is on processes by which all the castes jointly strengthen the system. The lower castes, it is argued, play a vital role in preserving the status quo, by holding back their opposition and, in the process, shackling their own minds.

#### VI. FURTHER REMARKS

The foregoing arguments rest on distinctions between private and public preferences and between private and public beliefs. The first distinction explains why societies retain policies they might have abandoned if not for the pull of the past. The two together explain why adopted policies condition people's perceptions and wants.

The model uncovers a tendency for beliefs and preferences to become homogenised. Such outcomes may be sought intentionally by some. But they are ultimately caused by multitudes of individual decisions made without an awareness of where they will lead. Outcomes are not necessarily socially optimal. In contrast to popular approaches that attribute social optimality to every outcome, this one explicitly allows for suboptimality.

In recognising that people depend on each other for their beliefs about how the world works, the model confers to the process of belief formation an important role in the collective decision process. It does so without compromising the principle that social phenomena are to be explained by individual choices. Using the methodology of individualism, the basis of modern economics, it makes endogenous a variable that economics has traditionally treated as exogenous.

*University of Southern California*

*Date of receipt of final typescript: February 1987*

#### REFERENCES

- Akerlof, G. A. (1976). 'The economics of caste and of the rat race and other woeful tales.' *Quarterly Journal of Economics*, vol. 90 (November), pp. 599-618.
- (1980). 'A theory of social custom, of which unemployment may be one consequence.' *Quarterly Journal of Economics*, vol. 94 (June), pp. 749-75.
- Anant, S. S. (1972). *The Changing Concept of Caste in India*. Delhi: Vikas Publications.

<sup>30</sup> The argument is not invalidated by the fact that certain groups are now openly fighting the system. The formation of these groups has to do with exposure to Western egalitarian ideologies, a factor I have entirely ignored. In a more complete model than the one space limitations have allowed me to provide here, the West would appear as an anti-caste pressure group, which, partly for technological reasons, gains strength only in the modern era.

<sup>31</sup> See Anant (1972, p. 20) and Srinivas (1966, p. 104).

- Arthur, B. (1985). 'Competing technologies and lock-in by historical small events: the dynamics of allocation under increasing returns.' *Center for Economic Policy Research Working Paper*, Stanford University.
- Bagehot, W. (1884/1956). *Physics and Politics*. Boston: Beacon Press.
- Berliner, J. S. (1976). 'Prospects for technological progress.' In *Soviet Economy in a New Perspective* (Joint Economic Committee, Congress of the United States). Washington, D.C.: U.S. Government Printing Office.
- Borda, J.-C. de (1781). 'Mémoire sur les élections au scrutin.' In *Mémoires de l'Académie Royale des Sciences*. Paris.
- Brenner, R. (1983). *History – The Human Gamble*. Chicago: University of Chicago Press.
- Cox, O. C. (1948). *Caste, Class, and Race*. Garden City, N.Y.: Doubleday.
- Dumont, L. (1966/80). *Homo Hierarchicus: The Caste System and its Implications*, rev. ed. Chicago: University of Chicago Press.
- Edwards, W. (1968). 'Conservatism in human information processing.' In *Formal Representation of Human Judgment* (ed. B. Kleinmuntz). New York: Wiley.
- Elster, J. (1983). *Sour Grapes*. Cambridge: Cambridge University Press.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.
- Freud, S. (1921/59). *Group Psychology and the Analysis of the Ego* (trans. and ed. by J. Strachey). New York: Norton.
- Goldberg, V. P. (1976). 'Regulation and administered contracts.' *Bell Journal of Economics*, vol. 7 (autumn), pp. 426–48.
- Granovetter, M. (1978). 'Threshold models of collective behaviour.' *American Journal of Sociology*, vol. 83 (May), pp. 1420–43.
- Hasher, L., Goldstein, L. and Toppino, T. (1977). 'Frequency and the conference of referential validity.' *Journal of Verbal Learning and Verbal Behavior*, vol. 16 (February), pp. 107–12.
- Heiner, R. A. (1983). 'The origin of predictable behavior.' *American Economic Review*, vol. 73 (September), pp. 560–95.
- Hirschman, A. O. (1967). *Development Projects Observed*. Washington, D.C.: The Brookings Institution.
- (1970). *Exit, Voice, and Loyalty*. Cambridge: Harvard University Press.
- Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. London: Macmillan.
- Kuran, T. (1987a). 'Chameleon voters and public choice.' *Public Choice*, vol. 53, pp. 53–78.
- (1987b). 'The tenacious past: Theories of personal and collective conservatism.' Unpublished manuscript, University of Southern California.
- Lal, D. (1985). The Hindu Equilibrium. Unpublished manuscript, The World Bank.
- McNeill, W. H. (1976). *Plagues and People*. Garden City, N.Y.: Doubleday.
- Madison, J. (1787–8/1961). 'The Federalist No. 39.' In *The Federalist* (ed. B. F. Wright), pp. 347–51. Cambridge, MA: Harvard University Press.
- Markus, H. and Zajonc, R. B. (1985). 'The cognitive perspective in social psychology.' In *The Handbook of Social Psychology*, 3rd. ed., vol. 1 (ed. G. Lindzey and E. Aronson), pp. 137–230. New York: Random House.
- Montmollin, G. de (1977). *L'influence Sociale*. Paris: Presses Universitaires de France.
- Moore, B., Jr. (1978). *Injustice: The Social Bases of Obedience and Revolt*. White Plains, N.Y.: M. E. Sharpe.
- Olson, M. (1982). *The Rise and Decline of Nations*. New Haven: Yale University Press.
- Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: Norton.
- Schwartz, M. (1982). 'Repetition and rated truth value of statements.' *American Journal of Psychology*, vol. 95 (autumn), pp. 393–407.
- Srinivas, M. N. (1966). *Social Change in Modern India*. Los Angeles: University of California Press.
- Tversky, A. and Kahneman, D. (1973). 'Availability: a heuristic for judging frequency and probability.' *Cognitive Psychology*, vol. 5 (September), pp. 207–32.
- Veblen, T. (1919). *The Vested Interests*. New York: B. W. Heusch.
- von Weizsäcker, C. C. (1971). 'Notes on endogenous change of tastes.' *Journal of Economic Theory*, vol. 3 (December), pp. 345–72.