Coping with Complexity: The Adaptive Value of Changing Utility

By MICHAEL D. COHEN AND ROBERT AXELROD*

In everyday life we frequently act on the basis of beliefs that are incomplete or, to some degree, false. Yet academic research on human choice behavior pays surprisingly little attention to the difficulties facing a decision maker whose current model of the world usually has a substantial probability of being misspecified. Most research can be viewed as providing potential elements of an improved future system of beliefs, one that will better match the world. But that very fact implies that today's decisions and those made earlier may well have been based on wrong beliefs. How should a decision maker behave under such conditions? When it is too costly to finance adequate investigation or to postpone decisions while awaiting its results, how might a reasonable, but not omniscient, person proceed using beliefs about the world that may be partially incorrect?

Existing literature gives only limited help on this question. One major research tradition deals with the updating of beliefs in view of new experience. This tradition includes both econometrics and Bayesian inference. These techniques can be very helpful in updating beliefs where new information can be incorporated into a model that is correctly specified. But they offer less help when the "improved" parameter estimates are being obtained within an incorrect model of how things work. Among the many difficulties that can occur with a misspecified model, one of the most trying is the possibility that such a model can make correct predictions over some range of policies. The apparent confirmation in this "misspecification trap" can prevent the decision maker from pursuing an improved model that would reveal superior alternatives remote from current policy.[1]

Our proposal for a response to the problems posed by wrong beliefs is a decision process that incorporates a controlled form of preference change. It sounds paradoxical to say that individuals may perform better as a result of moving away from their current goals. But we will show that a properly structured adaptive utility process can indeed help people make better decisions.

I. Misspecification and Preference Change

This paper offers a model which demonstrates that preference change can provide a course of action that is adaptive in the presence of misspecification. By "adaptive," we mean that such a process can improve performance. The preference change is driven by experience, but is not necessarily conscious.

Our model bears on both normative and descriptive issues. On the normative side, it

*Associate Professor and Professor of Political Science and Public Policy, respectively, Institute of Public Policy Studies, 1516 Rackham Building, University of Michigan, Ann Arbor, MI 48109. We thank Joel Cohen, Jon Bender, John Ferejohn, Aanund Hylland, Lewis Kornhauser, James March, Robert A. Pollak, Richard Quandt, and John Scholz for their helpful suggestions. We are especially grateful to Carl Simon who generously devoted many hours to refining an earlier version of our model. Earlier versions of this paper were presented at the Conference on the Multiple Self, Maison des Sciences de l'Homme, Paris, and the Carnegie-Mellon Symposium of Information Processing in Organizations. This work was supported by a Sloan Foundation grant to The University of Michigan Institute of Public Policy Studies, the Center for Advanced Study in the Behavioral Sciences, and NSF grant SOC-8023556 to Axelrod, and NSF grant SES-7726026 to Cohen.

[1] The other major research tradition dealing with the problem of inaccurate beliefs, investigates how people of limited rationality can still cope with their environment (for example, James March and Herbert Simon, 1958; Richard Cyert and March, 1963). Concepts such as satisficing and dynamic aspiration levels do suggest that decision mechanisms as well as beliefs are adaptable. Yet even in the research tradition of limited rationality, the underlying preferences are taken as given.

provides an approach to the problem of improving your performance when you don't completely understand what you are doing. On the descriptive side, the approach seeks to account for some of the ways in which preferences do in fact change as a function of experience. We seek not only to model adaptive preference change, but to do so using assumptions that are behaviorally plausible.

Economists interested in how tastes change have taken two basic approaches. The sophisticated behavior approach of Robert Strotz (1955–56) assumes that individuals know that their present choices influence their future preferences. This gives rise to a variety of problems of consistency, existence and stability of plans and choices over time (Carl Christian von Weizsacker, 1971; Bazalel Peleg and Menahem Yaari, 1973; Robert Pollak, 1968; Peter Hammond, 1976; and Gordon Winston, 1980). Besides these technical problems, there is also the practical question of whether a decision maker is likely to know much about how a current choice would affect future preference.

In contrast, the naive behavior approach assumes that the decision maker does not know how current choice would influence his or her future preferences. This approach has been used in empirical demand analysis. The work has explored four ways in which tastes change over time: habit formation, interdependence between people, advertising, and price signals. (For a review, see Pollak, 1978.) For our purposes, the most relevant strand of this literature is habit formation, the idea that future preferences depend directly upon prior choices. Habit formation is typically operationalized by providing a functional relationship between the amount of some current (or past) activity and the parameters of the next period's utility function.[2] This captures the concept of a habit developing as a function of action, but it leaves out the fact that beliefs may also have influenced the shaping of preferences.

The model of preference change which we propose remedies this omission by explicitly introducing the interaction between beliefs and actions in the shaping of future preferences. An action taken on the basis of beliefs can yield surprise. We define "surprise" as the difference between the utility experienced as the result of an action and the utility expected to result from that action. We then model an individual as coming to like the things that yield pleasant surprises and coming to dislike the things that yield unpleasant surprises.

The model forges a connection between the literature on changes in preferences and the literature on rules of thumb and bounded rationality. It does this by viewing preference change as being driven by the surprises that will inevitably occur if the world is too complex and dynamic for the decision maker to develop a correctly specified model of the environment. Because of such complexity and the resulting misspecification of belief, we do not want to assume that the decision maker can anticipate his or her own future preference changes. Instead, we are studying the situation in which utility change represents a largely unconscious adaptation to an environment which is not completely understood.

To illustrate our basic approach, consider the game of chess. With a correctly specified model, one could choose the best play at every turn, but the explosive combinatorics of chess have so far (and for the foreseeable future) prohibited the development of a correct model. Play must be heuristic, based on principles known to be imperfect. The goal, of course, remains the capture of the opponent's king. But at the beginning and the middle of the game, the player cannot see just how to accomplish this goal. So the

[2] George Stigler and Gary Becker (1977) have proposed a human capital theory approach to account for certain self-reinforcing patterns of behavior. But their theory accounts for the changes by alterations in "technology" (for example, the ability to make finer discriminations), rather than modifications in the underlying utility itself. This approach has been criticized for being both a doubtful ideology and having no practical relevance for normative prescriptions of choice (March, 1978, p. 597). The issue has also been tackled by Cyert and Morris DeGroot (1975; 1980), but their approach is to study how experience alters a person's *knowledge* of what he or she likes, rather than how it alters underlying utility.

player pursues other goals as well, hoping that will lead toward improved performance even on the criterion of winning the game. A chess player is typically taught to evaluate features of board positions in particular ways. For example, a rook is valued at five points, a knight at three points, and a pawn at one point. These valuations lead directly to useful policy advice. For instance, they suggest that a player should be willing to give up a knight to capture a rook, but should not be willing to make the reverse trade. These values for the pieces are not specified in the rules of the game. But neither are they arbitrary. They are the result of centuries of experience. The valuation of the pieces and of other aspects of board positions by successful players indicates that one can do better by *not* concentrating exclusively on the capture of the king, but by learning also to pursue goals that eventually make the ultimate goal more accessible.

Our model is actually a direct descendant of an artificial intelligence program by A. L. Samuel (1959) for playing checkers. The success of the Samuel Checker Player could be evaluated by an outsider using a simple performance measure: whether or not it wins games. The program itself was not given the set of values to be used in the evaluation of board positions. Instead it learned for itself what values to follow, beyond the basic one of piece advantage. The program's learning process was driven by surprise. When things were going surprisingly well (or badly), it would note what was correlated with surprise, and would adjust the values of its various goals accordingly. To take the chess analogy again, suppose a person valued a rook and a knight equally. Then when the player gave up a rook to capture a knight, the player might notice that a few moves later things were going surprisingly badly. If unpleasant surprises frequently occurred after giving up a rook to get a knight, the learning process would gradually raise the relative value imputed to rooks. Thus surprises can drive an adaptive change in values. This is just how Samuel's checker playing program learned to play good checkers.[3] In fact, the program was able to defeat a former state champion (Edward Feigenbaum and Julian Feldman, 1963, pp. 103–104).

This paper seeks to generalize the principles that make the Samuel checker playing program so successful. To demonstrate that the principles are applicable beyond checkers, we want to show their successful operation in another task domain. To maximize the clarity of the principles, we have used as simple a setting as we can without making the choice trivial.

II. A Dynamic Model of Preference Change

An extended example will illustrate both the nature of the problem we have in mind and the application of the principles that we believe to be useful in responding to the problem. We introduce some notation in order to make the inferences in the presentation precise, but will make a number of simplifying assumptions so that the mathematical details do not obscure the issues at hand.

Consider a factory manager with a fixed number of labor hours available for the upcoming period. The manager wants to maximize output and has to allocate labor hours between production and maintenance.

The manager knows that too much labor devoted to production would lead to suboptimal output because of inadequate maintenance. The manager also knows that too little labor devoted to production will lead to suboptimal output. Therefore the manager's problem is to choose the level of labor devoted to production, x, which maximizes output, y. We will assume that the relationship between x and y is *believed* to be of the following form:

$$\hat{y}_t = -x_t^2 + \hat{b}_{t-1}x_t, \tag{1}$$

where $\hat{y}_t$ is the expected output, and $\hat{b}_{t-1}$ is a parameter that can be estimated from the previous choice of x and the observed output in the previous time period.[4] This functional

[3] Chess programs typically have fixed evaluation functions, and hence do not learn. The learning in computer chess takes place in the programmer's mind and is embodied in the next version of the program.

[4] For example, this relationship could have come from an analysis of labor productivity. The manager

form is typical of a broad class of problems where increasing the level of activity is beneficial at first, but then becomes detrimental.

With a choice of x_t, and an observed level of output, y_t, the manager can estimate the unknown parameter in equation (1) by solving for $\hat{b}_t$. This gives (for x_t not equal to zero):

$$\hat{b}_t = y_t/x_t + x_t. \tag{2}$$

With this estimate of the unknown parameter, the manager can select the next policy choice to maximize expected output. A little calculus shows that the x_{t+1} which maximizes $\hat{y}_{t+1}$ is

$$x_{t+1} = \hat{b}_t/2. \tag{3}$$

Now we are in a position to introduce the problem that we want to study, namely misspecification.[5] Equation (1) gives the *believed* production relationship. Our interest is in the case in which this belief is inaccurate. So we will suppose that there is an unknown source of lost output in the factory due to pilferage. We will denote the number of units being lost each period as c ($c < 0$) and make the *true* output relationship

$$y_t = -x_t^2 + bx_t + c. \tag{4}$$

We will assume that this simple misspecification will not be discovered by the manager. The reason for this assumption is to allow us to embody the important principle that the world is so complex that there is always *something* that beliefs do not model adequately. We could easily endow the manager with a more powerful belief system capable of discovering the missing variable, and then increase the complexity of the reality in order to keep the belief system inadequate. Beliefs would still fall short of reality, but the mismatch would occur at a much higher level of complexity. This might add an appearance of realism, but at the cost of burdening our exposition with details that would obscure the structure of the argument without strengthening its fundamental logic.

The manager's method for choosing an optimal level of x is no longer ideal in light of this misspecification, but it will be used since the loss of c units each period is not known. The level of x chosen according to equation (3) will therefore no longer be optimal. If the manager sticks to the choice procedure, there will be a brief discrepancy between expected output and observed output. After that, the policy choice will settle down to

$$x_s^* = \left(b + (b^2 + 8c)^{1/2}\right)/4. \tag{5}$$

We will refer to the model of the manager's decision making that we have just sketched as the *standard model* so that we may contrast it with a dynamic model that we will develop below. The allocation x_s^* will be called the *standard result*.[6] It does not depend on what initial policy was tried to provide the first estimate $\hat{b}_1$, and it is suboptimal since the standard result will be less than $b/2$ whenever $c < 0$. Furthermore, the standard result has the devilish property that it produces a misspecification trap: output experience will exactly match expectations based on incorrect beliefs.

Figure 1 shows the geometry of the situation the unfortunate manager is in. The two

might believe that $\hat{Y}_t = kp_t x_t$, where k is a constant of proportionality, $p_t = q + r(L - x_t)$ where q is the (unknown) minimum productivity of a work hour that would occur if all the available labor were allocated to production, r is the (known) rate at which a labor hour invested in maintenance improves productivity, and L is the (known) total labor available. Then if we rescale the output by letting $\hat{y}_t = Y_t/kr$, we have equation (1) where $\hat{b} = L + q/r$. Note that $\hat{b}$ is unknown because q is unknown.

[5] The mechanisms for updating beliefs and choosing new policy incorporated in (2) and (3) are perhaps a bit more precise than one might expect from real decision makers. Our decision to make new beliefs perfectly consistent with the most recent experience, and new policy optimal with respect to those beliefs, rests on a desire to provide the most stringent test of the contribution of dynamic preferences to decision-making quality. With these forms for (2) and (3), we can be more confident that the performance improvement obtained is not merely a correction for faulty methods of updating beliefs or choosing policy.

[6] Equation (5) is derived by noting that at stability equations (2), (3), and (4) imply that $x = (b + c/x)/2$. The right-hand side of (5) is just the largest root of this quadratic. To avoid an imaginary term we need $b^2 + 8c$ to be nonnegative. We restrict our discussion to cases satisfying this constraint.

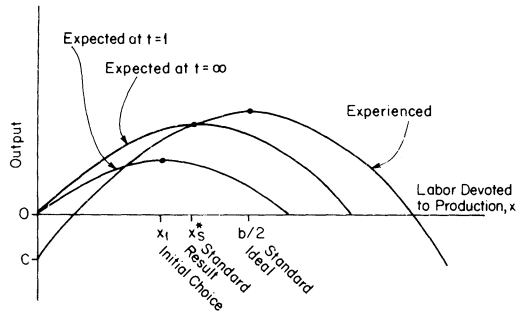

FIGURE 1. A MISSPECIFICATION PROBLEM

parabolas passing through the origin represent what is believed about the system initially and when the adjustment of beliefs to experience has settled down. The choice that would maximize output, given the final beliefs, is x_s^*. The true output relationship is shown by the other parabola. At the standard result it has the same value as the believed parabola and the incorrect model is therefore confirmed by experience.

To develop the dynamic model, we need an explicit expression for the manager's revealed utility. We will suppose it is the simple sum of output, about which the manager has explicit beliefs, and an additional factor that we call the *intrinsic utility* of x, the extent to which the manager regards production labor as a good (or bad) thing in itself, as a final value rather than as an instrument in the production of valued output. Intrinsic utility need not be conscious. When the manager places positive intrinsic value on x, this amounts, of course, to reducing the relative value of output in determining the choice of x. As we will show, however, the properly controlled evolution of the intrinsic value for x can yield an *increase* in output—the manager does better in output terms by caring less about it. The *expected utility* which the manager's choice will be maximizing is

$$\hat{U}_t = \hat{y}_t + w_t x_t, \tag{6}$$

where w_t is the intrinsic utility of a unit of x. The manager's satisfaction with observed results, *experienced utility*, is output y_t plus intrinsic utility $w_t x_t$, or

$$U_t = -x_t^2 + bx_t + c + w_t x_t. \tag{7}$$

For $w_t = 0$, (6) and (7) correspond to the standard model. Expected utility is maximized at $(\hat{b}_t + w_t)/2$. When w is a function of experience, we have our dynamic model. The dynamic choice process is like the standard process just presented in every respect except that in the dynamic case intrinsic utility evolves as a function of the discrepancy between expected and experienced utility, a quantity that we will call *surprise* and label D_t.

$$\text{Surprise} = D_t = U_t - \hat{U}_t = \left(b - \hat{b}_{t-1}\right)x_t + c. \tag{9}$$

Our proposed model is in essence a kind of learning process through which the manager comes to ascribe additional value to the assignment of labor to production if such assignments are associated with pleasant surprises, and comes to assign less value to the level of x when it is associated with negative surprises.[7]

We must specify a functional form governing the change in w and there are many possibilities. The field is narrowed sharply, however, by two constraints. First, the change in w should be proportional to the magnitude and direction of recent change in x. This can be represented as $(x_t - x_{t-1})/|x_t|$. The divisor is necessary to scale the expression so that it will depend only on the relative magnitude of the change. Second, the change in w should be proportional to the (scaled) magnitude and direction of recent surprise. This is just D_t/x_t. There are several ways these two expressions could be combined. Our choice is the relatively conservative position that change in w should be large only when surprise and policy change are

[7]Our decision to make surprise a fundamental quantity in our model follows Samuel. It is supported by recent findings that surprise is a genuine psychophysical state with reliably measurable corresponding brainwave patterns (Connie Duncan-Johnson and Emanual Donchin, 1977). Moreover, this measurable surprise can apparently derive from both conscious and unconscious expectations in correspondence to our distinction between instrumental and intrinsic utility. In cognitive psychology, an approach in the same spirit has been made by George Mandler (1981), who argues that changes in value are driven by a form of cognitive discrepancy or "schema incongruity."

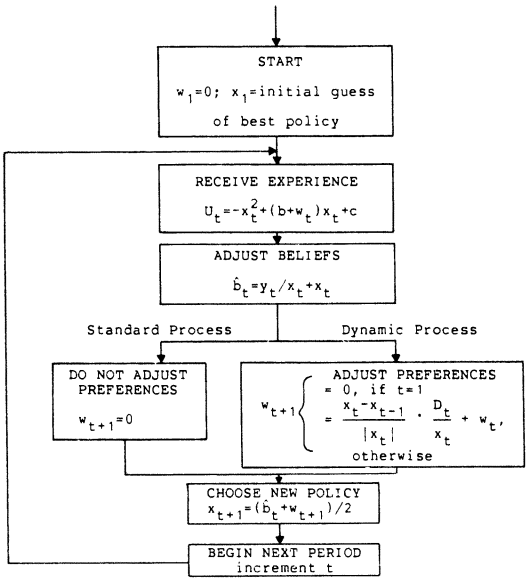

FIGURE 2. FLOWCHART FOR STANDARD AND DYNAMIC PROCESSES

both large.[8] This gives us an overall expression for the change in w as the product of these two expressions:

$$(10) \qquad w_{t+1} - w_t = \frac{x_t - x_{t-1}}{|x_t|} \cdot \frac{D_t}{x_t}.$$

This functional form captures three important properties that we believe to be essential for the satisfactory operation of the dynamic process: surprise as the main driver of the adaptation of preferences; drag on the adaptation process to keep it from runaway pathologies; and implicit context sensitivity, which keys the changes in preferences to the conditions associated with the surprise. In our simple setting, these are appropriate analogues of the procedures employed in the Samuel Checker Player.

The flow chart of Figure 2 presents the complete cycle that is followed by our dynamic choice process, and includes the standard choice process as a special case in which w_t is always zero. We have not been able to obtain a closed-form expression for the *dynamic result*, the value of x (denoted x_d^*) that is obtained from running the dynamic process under various initial conditions. It is easy to compute this value, however, for a great variety of initial values, x_1, and for many levels of c, the specification error. We now turn to the results of such computations.

III. Assessment of Performance

Thus far we have delineated a pair of processes that a decision maker could use when seeking to maximize utility in a world that is not completely understood. The dynamic model and the standard model both have a structure of beliefs about the world built from the same (mis)specification of reality. Both have a system for updating those beliefs as new experience comes in. Both choose new policies so as to maximize expected utility given current beliefs. Both fail to choose optimal policies because of misspecification. The utility functions being maximized are, however, not identical, since the dynamic model has a method of changing its utility function that is inoperative in the standard model. This is, in fact, the only difference between the two models. It does mean, however, that the objective function being pursued by the dynamic model will, in general, be different from that being pursued by the standard model.

This last fact raises the question of how we are to assess performance. The standard model may be regarded as embodying the process that would occur if no preference change were allowed. The dynamic process begins with the same utility function as that used throughout the standard process. However, the dynamic model becomes, in effect, a different person. If we use the standard model's utility function to assess the quality of the policy chosen by the dynamic model, or if we use the dynamic function to assess standard model policy choices, we will be making an intertemporal comparison of utilities. The difficulties this raises are very similar to those present in the more common case of interpersonal comparison. We resolve the problem as has been done so often in the

[8] Since two policy choices and observed outcomes are needed to scale the surprise, the adjustment in w does not start until $t = 2$.

more familiar setting, by invoking the standard proposed by Pareto: if one outcome is preferred to another on *both* utility functions, we will call it the better of the two. This criterion is restrictive in many ways. We do not mean to propose that normative advice that fails to satisfy the criterion (i.e., that leads to outcomes not preferred on both the current and the future utility functions) is necessarily bad advice. But we do think that advice which leads to preferred outcomes on both utility measures will definitely be more acceptable to the potential taker—who, after all, has the standard utility function at the moment and will have the dynamic function later.[9] Therefore, we wish to determine the conditions under which the dynamic model is better than the standard model in this Pareto sense.

To this end we introduce some additional definitions. The *standard ideal* is the policy at which the utility function of the standard model would be maximized. In our model, its value is $b/2$. The standard ideal might also be characterized as the "instrumental ideal," since it is the level of the activity x that maximizes utility when x is only an instrument and has no value of its own for the decision maker. Except when $c = 0$, the standard ideal differs from the standard result of (5), the policy normally reached by the standard model.

The *utility of the standard result* (U_s^*) is the value of equation (7) with $w = 0$ and $x = x_s^*$. A comparable quantity is the *utility of the dynamic result* (U_d^*) obtained from equation (7) with $w = 0$ and $x = x_d^*$.

When the inequality

$$U_d^*/U_s^* > 1 \tag{11}$$

is satisfied, the net output at the dynamic result is greater than at the standard result. Since the standard utility function values only output, we have a situation in which the dynamic process has come closer to the standard ideal than has the standard process. In this case, the standard utility function will have a higher value at the dynamic result than at the standard result, establishing the difficult part of the Pareto criterion. The other part, as one might expect for a rationalizing system, follows immediately, since a result that is better on the standard utility function is better on the dynamic utility function by an addition of the term for intrinsic utility, $w_t x_t$. (In the calculations we have performed, $w_t x_t$ is always positive if U_d^* is at least as large as U_s^*.)

Thus we have isolated the crucial measure to be used in assessing the performance of the dynamic model. When $U_d^* > U_s^*$, the dynamic result may be said to be better than the standard result in the Pareto sense. In terms of the chess example we have employed, this is equivalent to observing that a player who comes to care about objectives other than capturing the king may not only be happy in terms of the new objectives being pursued, but also may end up capturing the king more often—although the king's capture is no longer the sole objective.

We are now ready to turn to the results of the computations. Figure 3 is a topographic map, showing the performance of the dynamic model relative to the performance of the standard model. The results are given for particular values of the misspecification c, and the initial choice x_1. The computations were made with x_1 from just above zero through a starting value four times the standard ideal, and with c from just below zero to $-b^2/8$, the minimum for which the standard process converges. The unhatched portion of the surface corresponds to the set of parameter values for which the utility ratio is greater than one. These are the points where the dynamic process achieves more output than the standard process. It is the overwhelming majority of the surface.

The possible values of the initial belief x_1 should not be imagined to be equiprobable, however. For that reason we have shown what might be considered the most probable area of the figure by adding two vertical dotted lines at one-half and twice the standard ideal. A manager beginning a sequence of choices with the benefit of reasonable

[9] It is interesting to note that a common form of preference change—imitating those one already likes—also satisfies this Pareto logic. The imitator already likes the behavior of the imitated and can expect to like it more as the movement continues toward the imitated's values.

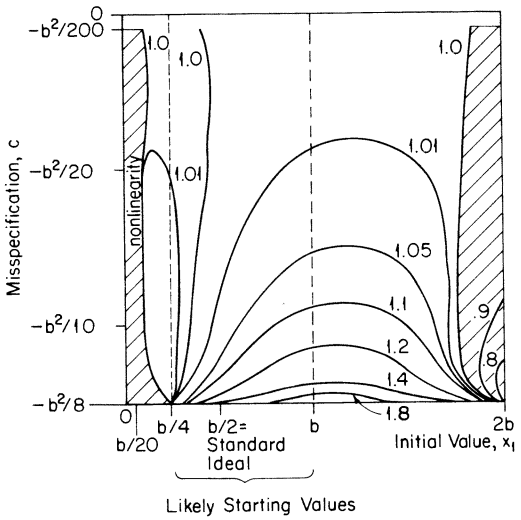

FIGURE 3. RATIO OF PERFORMANCE: DYNAMIC/STANDARD

prior beliefs derived from imitation or analysis would be quite likely to start at a point in this central area. In this region of plausible initial beliefs, the dynamic process completely dominates the standard process. Thus, with any initial belief that differs from the standard ideal by even as much as a factor of two, the dynamic process does better than the standard process.

When the pilferage rate c is high, the results are near the bottom of the figure, where the dynamic model does up to twice as well as the standard. For values of x_1 very close to zero, there are sharp nonlinearities in the behavior of the dynamic model, but they are confined to a small region well distant from plausible starting values for the dynamic process. Some points here do satisfy inequality (11), but others nearby do not.

Figure 3 verifies that the final policy reached from likely starting conditions is always better in the dynamic case. Our computations actually give the stronger result that from all likely initial conditions, the cumulative output of the dynamic model exceeds that of the standard model at every period. The dynamic model is better not only when stability is reached, but throughout the entire process of adaptation.

The consistent dominance of the dynamic result over the standard result displayed in the portion of the figure between the dotted lines demonstrates an important point. Under a wide range of initial conditions, a decision maker in this model would be well-advised to let his or her values drift away from their current configuration in a properly controlled adaptation to experience. Over a wide range of parameter values, a process that is essentially a controlled form of rationalization leads to results that will be preferable both as measured by the current utility function and as measured by the utility function at the conclusion of the dynamic process.

Once the dominance of the dynamic model over the static model has been established, a further question asserts itself. How robust is this superior performance of the dynamic model in misspecified environments? There is an enormous number of model variations that might be run to investigate robustness, but one line of such variations seems particularly important. These are versions of the models that test whether the strong results obtained depend on the particular form of misspecification used. The mismatch between reality (equation (4)) and belief (equation (1)) in the case that was studied was the omission of a constant ($c < 0$) from the belief system.

A variation on the experiment that suggests itself immediately is the symmetric misspecification in which the omitted term (c) would be greater than zero. We have repeated our calculations with all the variables covering the same absolute ranges as before and c positive. The results are essentially the same. None of our qualitative conclusions are affected. In particular, it is worth noting that this demonstrates the full symmetry of the dynamic model's performance: its adaptive contribution occurs whether the standard result and initial policy are above or below the best policy, and whether the ultimate value of w is positive or negative.

To assess robustness further, we subjected the dynamic and standard models to more radical misspecifications in three additional experiments. In each case we substituted a new reality for equation (4) of our models. All other aspects remained unchanged. The beliefs about the production relationship

continued the assumption that it was a parabola passing through the origin. The method of updating beliefs based on experience operated as before. The method of choosing a new policy expected to be optimal in a quadratic world was retained. In the dynamic model, the preference change process driven by surprise was identical. All that was different was the nature and magnitude of the mismatch between belief and underlying reality.

In the three robustness experiments, the new realities were

$$(4') \quad y_t = -|x_t - b| + c, \quad \text{(absolute value)};$$

$$(4'') \quad y_t = -\sin(cx_t + b), \quad \text{(sine)};$$

$$(4''') \quad y_t = x_t^3 - x_t^2 + bx_t + c, \quad \text{(cubic)}.$$

All three were scaled with respect to the variables x_1 (initial policy) and c (the omitted parameter) so as to be comparable to the original case. All three have only a single maximum in the range studied, but the curvatures in the range are all quite different from the original reality function and therefore from the parabola incorporated in the belief systems of both models. As a result, the models behaved quite differently in the three additional cases. However, the fundamental qualitative conclusion of the original case was sustained in all three of the new studies: over the same wide range of variation in initial policy (x_1) and in magnitude of the omitted parameter (c), the dynamic model outperformed the standard model. Its cumulative performance on both measures was better at every step along the way.

Thus the robust superior performance of the dynamic model demonstrates that the simple principles underlying the success of the Samuel Checker Player can be transferred with powerful effect to other task domains.

IV. A Management Example

Consider a newly hired middle manager who begins with a concern only to win the approval of his or her superiors. Proceeding from that orientation, such a manager might initially join them in opposing an expensive package of fringe benefits for the workers in the manager's unit. Over time, however, the operation of the dynamic process we described might well produce a growing sensitivity to winning the approval of subordinates if there is some regular but poorly understood relationship between subordinate approval and performance by the unit that satisfies top management. In turn, such a development can find the manager coming eventually to support the fringe benefit package that the workers want, getting better performance out of the unit and therefore winning more net approval of superiors. Such a result is consistent with a pattern frequently observed in organizations. Consider, for example, Peter Blau's report (1956) of a study by Daniel Katz et al. (1950): "...superiors who were primarily concerned with maintaining a high level of production, interestingly enough, were less successful in meeting this goal than those supervisors who were more interested in the welfare of their subordinates than in sheer production; in the latter case, productivity was generally higher" (p. 70).

All this can occur in the dynamic process without the manager ever having a full understanding of the causal connections producing the effects. It seems to us consistent with the observation that managers (and other regulators) are sometimes "captured" by the interests of those they manage. It also suggests that in a complex environment, a suitably controlled capture process may have some genuine virtues.

V. The Interpretations of Our Model

This example can be used to illustrate two different interpretations of our model. The conventional interpretation is that the manager only cares about production, and treats the welfare of the subordinates as a means to that end. Under this interpretation, the manager's utility is simply a function of production. The conventional interpretation would then regard the mechanism for using surprise as a "rule of thumb" which shapes a

pseudo-utility function.[10] The interpretation that we prefer drops the concept of a pseudo-utility function, and regards the manager as having undergone a genuine change in preferences. Certainly, this would be the natural interpretation from the perspective of revealed preference theory: if the manager is observed to be willing to give up some expected production to get some expected employee welfare, then this choice can be taken as reflecting a utility function which includes both production and welfare.

An individual need not realize that he or she is coming to like things that are associated with pleasant surprises—but the resulting change in the utility function can still be adaptive. As long as there is misspecification, there is need for more than the usual Bayesian approach by which new information is incorporated into the decision process through the revision of beliefs (Sanford Grossman, Richard Kihlstrom, and Leonard Mirman, 1977; and John Hey, 1981). Under our preferred interpretation, it would be reasonable to say that new information is used not only to update beliefs about how choices map into outcomes, but that it is also used to modify the very utility function that is being maximized. Indeed, we regard the demonstrated potential for adaptive utility change as the fundamental point of our model.

VI. Learning to Adapt in a Strategic Interaction

The theory of adaptive utility change has so far been discussed in the context of a single decision maker in a passive environment. The theory can also help to explain behavior in a setting of mutual interaction. Consider for example the iterated Prisoner's Dilemma in which a player chooses in each turn whether to cooperate or defect. A simple strategy in the game is Tit for Tat: cooperate on the first move and then do what the other player did on the previous move. This strategy has been shown by Stuart Oskamp (1971) and Warner Wilson (1971) to be highly successful when playing directly with people. It has also been shown by Axelrod (1980; 1981) to be highly successful when playing with a wide variety of more or less sophisticated decision rules. However, even a naive player without a clear knowledge of the other's strategy can realize that there is instrumental utility to defecting. After all, the logic of Prisoner's Dilemma guarantees that on a given move the payoff from defecting will always be greater than the payoff from cooperating, no matter what the other player chooses. But there is a pleasant surprise in store for the naive player who experiments while interacting with Tit for Tat. In the move after the naive player cooperates, the player gets a higher payoff than in the move after a defection is chosen. According to the model of adaptive utility developed here, this pleasant surprise will actually cause the naive player to place some intrinsic utility on the choice of cooperation after the other player cooperates. When enough intrinsic utility has been assigned to this conditional cooperation, the player will overcome the instrumental utility of defection, and come to cooperate when appropriate.

The trick in the Prisoner's Dilemma is to cooperate only with those who will reciprocate. This is already built into Tit for Tat. It is a trick that someone employing an adaptive utility process can learn from scratch.

VII. Conclusions

Beliefs are virtually always misspecified to a greater or lesser extent. Our results show that when beliefs are misspecified, controlled preference change can actually be adaptive. By allowing pleasant and unpleasant surprises to guide changes in utility, a decision maker can actually achieve better performance on both the original and revised utility functions. This possibility has now been demonstrated not only in the complex task environment of checkers, but also in four variations of a simple model of a management task.

[10] Rules of thumb have been investigated by William Baumol and Richard Quandt (1964) in the context of optimally imperfect decisions, namely decisions where the marginal cost of additional information or calculation equal the marginal expected yield. In our model, the amount of information is fixed, but the problem remains of how best to use it.

The significant novelty in our approach is the way in which preferences change as a function of experience. We believe the success we have had rests substantially on incorporating—as did Samuel in the Checker Player—three principles of preference change: 1) surprise as the main driver of adaptation; 2) drag to retard excess value change; and 3) implicit context sensitivity.

The first of these is essentially the principle that the utility function adapts not to experience itself, but to the difference between experienced and expected utility. In effect, the adaptation of the function is to the errors of the cognitive system: today's intrinsic utilities are in part the result of yesterday's misunderstandings of the world. The second characteristic of our treatment of utility dynamics is the use of drag on the process of bringing utilities into alignment with experience. If this process is allowed to occur too rapidly, there is a serious risk of a "runaway" utility function that dictates extremely high or extremely low levels of the activities in question. We regard this model pathology as akin to the development of an addiction or phobia. While a too-rapid adjustment is to be avoided, so also is too much drag. At the extreme, excessive drag would leave a dynamic process nearly indistinguishable from the standard one, and the benefits available would have been forfeited to caution. In his stimulating 1978 paper, March observed that the a posteriori adjustment of preferences toward consistency with outcomes ("rationalization" in common parlance) might not be maladaptive in every case. He lamented the lack of a formal framework in which one could pursue the issue with precision. Our model provides the first formal treatment of the question, so far as we are aware.

The third characteristic of our treatment of value change is an effort to arrange the adaptive process so that it implicitly incorporates as much sensitivity to context as possible. In the dynamic model, this occurs because the sign of the change in intrinsic utility depends on whether or not recent changes in policy (and, implicitly, recent changes in intrinsic utility) have been positive or negative. In the Samuel Checker Player, a considerably more elaborate system accomplished a similar result.

An issue addressed in only its simplest form by our model is the attribution problem: to which of the many things going on now or recently should the system give the credit (or blame) for the surprise it has experienced? Some portion of this dilemma is resolved by whatever cognitive mechanisms are available for updating the system's beliefs. But it is unlikely that these mechanisms will be entirely adequate, so that discrepancies will probably continue to occur. The intrinsic values that rise or fall in such circumstances need to have a better than random chance of being genuinely related to the experienced discrepancies if the learning is to be anything but superstition.[11] Samuel was able to specify a method, based on feature correlations, which handled the attribution problem in the special case of checkers, but we are not certain what the correct generalization of his methods will be.[12]

The success of the three simple principles of preference change that we have followed is certainly encouraging. They led us, almost without a false step, to the model we have presented, they seem to be consistent with the procedures used in Samuel's remarkable checker program, and they are generally consistent with results of psychological research. We can, at this point, only claim success for the approach in the special case presented here. But this special case does serve as an

[11]As John Anderson notes (1980), the most successful models of learning and cognition have postulated a fundamental capability of responding to features that are correlated in the environment. This process need not be conscious. Indeed, Anderson et al. (1979) have shown that subjects often cannot verbalize the bases of their learned (and correct) experimental responses. They sometimes even exhibit behaviorally powerful unconscious expectations which are directly at odds with their conscious models of the experimental situation (Daniel Kahneman and Amos Tversky, 1982).

[12]One possible solution may lie in employing an analogue of the adaptive process used by a pool of genes to become increasingly more fit in a complex environment. A promising effort to convert the main characteristics of this process to an heuristic algorithm is given by John Holland (1975). This algorithm has had some striking preliminary success in the heuristic exploration of arbitrary high dimensionality nonlinear functions.

existence proof that utility changes which are guided by surprise can be adaptive. If this principle turns out to be broadly applicable, it would have important consequences for fundamental questions in economics, political science, and organization theory.

REFERENCES

Anderson, John, *Cognitive Psychology and Its Implications*, San Francisco: Freeman, 1980.

_____, **Kline, Paul and Beasley, Charles,** "A General Learning Theory and Schema Abstraction," in G. Bower, ed., *The Psychology of Learning and Motivation: Advances in Theory and Research*, Vol. 13, New York: Academic Press, 1979.

Axelrod, Robert, "More Effective Choice in the Prisoner's Dilemma," *Journal of Conflict Resolution*, September 1980, *24*, 379–403.

_____, "The Emergence of Cooperation Among Egoists," *American Political Science Review*, June 1981, *75*, 306–18.

Baumol, William J. and Quandt, Richard E., "Rules of Thumb and Optimally Imperfect Decisions," *American Economic Review*, March 1964, *54*, 23–46.

Blau, Peter M., *Bureaucracy in Modern Society*, New York: Random House, 1956.

Cyert, Richard M. and DeGroot, Morris H., "Adaptive Utility," in R. H. Day and T. Groves, eds., *Adaptive Economic Models*, New York: Academic Press, 1975, 223–46.

_____ **and** _____, "Learning Applied to Utility Functions," in Arnold Zellner, ed., *Bayesian Analysis in Econometrics and Statistics*, New York: North-Holland, 1980, 159–68.

_____ **and March, James G.,** *A Behavioral Theory of the Firm*, Englewood Cliffs: Prentice-Hall, 1963.

Duncan-Johnson, Connie C. and Donchin, Emanual, "On Quantifying Surprise: The Variation of Event Related Potentials with Subjective Probability," *Psychophysiology*, September 1977, *14*, 456–67.

Elster, Jon, *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press, 1979.

Feigenbaum, Edward A. and Feldman, Julian, *Computers and Thought*, New York: McGraw-Hill, 1963.

Grossman, Sanford J., Kihlstrom, Richard E. and Mirman, Leonard J., "A Bayesian Approach to the Production of Information and Learning by Doing," *Review of Economic Studies*, October 1977, *44*, 533–47.

Hammond, Peter J., "Changing Tastes and Coherent Dynamic Choice," *Review of Economic Studies*, February 1976, *43*, 159–73.

Hey, John D., "Are Optimal Search Rules Reasonable? And Vice Versa?," *Journal of Economic Behavior and Organization*, March 1981, *2*, 47–70.

Holland, John H., *Adaptation in Natural and Artificial Systems*, Ann Arbor: University of Michigan Press, 1975.

Kahneman, Daniel and Tversky, Amos, "Variants of Uncertainty," in D. Kahneman et al., eds., *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press, 1982.

Katz, Daniel, MacCoby, Nathan and Morse, Nancy C., "Productivity Supervision, and Morale in an Office Situation," Institute for Social Research, University of Michigan, 1950.

Mandler, George, "The Structure of Value: Accounting for Taste," Center for Human Information Processing Report 101, University of California-San Diego, May 1981.

March, James G., "Bounded Rationality, Ambiguity, and the Engineering of Choice," *Bell Journal of Economics*, Autumn 1978, *9*, 587–608.

_____ **and Simon, Herbert A.,** *Organizations*, New York: Wiley & Sons, 1958.

Oskamp, Stuart, "Effects of Programmed Strategies on Cooperation in the Prisoner's Dilemma and Other Mixed-Motive Games," *Journal of Conflict Resolution*, June 1971, *15*, 225–59.

Peleg, Bazalel and Yaari, Menahem E., "On the Existence of a Consistent Course of Action when Tastes are Changing," *Review of Economic Studies*, January 1973, *40*, 391–401.

Pollak, Robert A., "Consistent Planning," *Review of Economic Studies*, April 1968, *35*,

201–08.

_____, "Habit Formation and Long-Run Utility Formation," *Journal of Economic Theory*, October 1976, *13*, 272–97.

_____, "Endogenous Tastes in Demand and Welfare Analysis," *American Economic Review Proceedings*, May 1978, *68*, 374–79.

Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, July 1959, *3*, 211–29; reprinted in E. A. Feigenbaum and J. Feldman, eds., *Computers and Thought*, New York: McGraw-Hill, 1963, 71–108.

Stigler, George J. and Becker, Gary S., De Gustibus Non Est Disputandum," *American Economic Review*, March 1977, *67*, 77–90.

Strotz, Robert H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, Winter 1955–56, *23*, 165–80.

von Weizsacker, Carl Christian, "Notes on Endogenous Changes of Tastes," *Journal of Economic Theory*, December 1971, *3*, 345–72.

Wilson, Warner, "Reciprocation and Other Techniques for Inducing Cooperation in the Prisoner's Dilemma Game," *Journal of Conflict Resolution*, June 1971, *15*, 167–95.

Winston, Gordon C., "Addiction and Backsliding," *Journal of Economic Behavior and Organization*, December 1980, *15*, 295–324.

Yaari, Menahem E., "Endogenous Changes in Tastes: A Philosophical Discussion," in Hans W. Gottinger and Werner Leinfellner, eds., *Decision Theory and Social Ethics, Issues in Social Choice*, Dordrecht: D. Reidd, 1977, 59–98.