COMPLEXITY AVERSION: SIMPLIFICATION IN THE HERRNSTEIN AND ALLAIS BEHAVIORS[1]

James Stodder
Rensselaer Polytechnic Institute

INTRODUCTION

Every economics teacher knows how easily students confuse average with marginal values. Managerial decision-makers, even if they understand the difference between average and marginal products, seem to be more like economics students than economics teachers in this respect [Faulhaber and Baumol, 1988]. Does this imply that it is inefficient to try for perfect efficiency? This essay argues that the "sub-optimal" rules of students and managers are efficient if, like most of us, they are liable to make computational errors and are averse to the risk of so doing.

In the first part, I review two famous examples of individual substitution of average for marginal quantities: "matching" behavior, as studied by the late Harvard psychologist, Richard J. Herrnstein, and the lottery "paradox" discovered by Maurice Allais. In the second part, I show that a risk-averse individual who makes costly or imperfect calculations will rationally decide when to cease further estimates of expected utility. In the third part of the paper, I argue that such sub-optimizing is an *evolutionary stable strategy*, defined by Maynard-Smith [1982] as a behavior robust to genetic competition. I conjecture that the evolutionary stability of sub-optimizing behavior results not from the costs of calculation, but from the costs of acquiring data and communicating solutions to other (non-optimizing) individuals. I conclude with a summary and suggestions for future research.

TWO EXAMPLES OF SUB-OPTIMAL BEHAVIOR

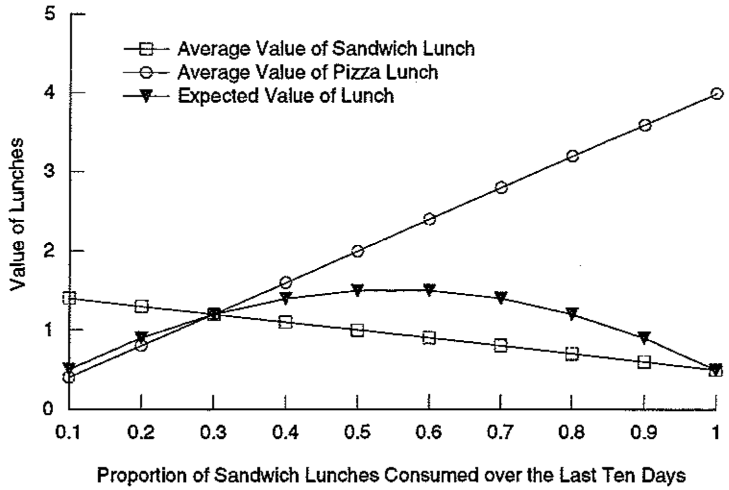
Herrnstein's Matching

The behavioral psychologist R. J. Herrnstein [Herrnstein and Prelec, 1991] provided the following homely illustration of sub-optimality. Let the average subjective value of eating a sandwich (s) or pizza (p) for lunch, be given as V_s or V_p. These are dollar values to the consumer, which depend on the proportion of sandwich (π_s) and pizza (π_p) lunches eaten over the last 10 days. V_s and V_p for a particular consumer are:

$$V_s = 1.50 - \pi_s, \quad V_p = 4(1 - \pi_p) = 4\pi_s, \quad \text{since } \pi_s = 1 - \pi_p.$$

If these are the only choices, then the values for each lunch can be shown in terms of π_s, the proportion of sandwiches consumed, as in Figure 1 below.

FIGURE 1
Value of Pizza and Sandwich Lunches



Note: Title is after Herrnstein and Prelec [1991]

This consumer is evidently one for whom sandwiches hold a fairly steady value as a lunch staple, whereas pizzas give high value as occasional treats, but quickly lose their appeal if eaten too often.

In such a situation, most people, even economists, respond sub-optimally. Assume that the two lunches are equally-priced. Upon entering the lunch room, our economist briefly considers which lunch gives him a greater subjective value, and buys that item. If both are equal, he is indifferent. This means equalizing the values, $V_s = V_p$, implying $\pi_s = 0.3$ and $\pi_p = 0.7$. His value from each meal is thus $1.20. If a meal costs $1, this would leave him with a consumer surplus of $0.20.

This equating of values is sub-optimal. Maximizing requires equating the marginal change in values, not the values themselves. Here the value (V) over lunches of both types is

(1) $$V = \pi_s V_s + \pi_p V_p = \pi_s(1.50 - \pi_s) + (1-\pi_s)(4\pi_s).$$

Setting the derivative in terms of π_s equal to zero, the optimum occurs at $\pi_s = 0.55$, where the expected value is now $1.51, yielding an average consumer surplus of $0.51 per lunch, two-and-a-half times the sub-optimal $0.20. At $\pi_s = 0.55$, our optimal consumer's value from pizza is $2.20, whereas he only gets $0.95 in value from a sandwich. He must resist increasing his rate of pizza consumption, however, to achieve optimal satisfaction. This degree of self-control may seem unrealistic, even for an economist. In fact, there is overwhelming experimental evidence that most people consistently fail to achieve such optimal control [Williams, 1988].

Herrnstein's "matching" applies in situations when the return from each alternative (i) depends on the frequency, π_j, with which every alternative — including i — is chosen; $j = (1, 2, ... i, ..., n)$.

Generalizing equation (1), the individual's total value can then be written:

(2) $$V(\pi_1, \pi_2, ..., \pi_n) = \sum_j^n \pi_j V_j(\pi_1, \pi_2, ..., \pi_n).$$

Matching means equating the average or expected values of each alternative i and j:

(3) $$V_i / P_i = V_j / P_j,$$

where P_i and P_j are the prices of alternatives i and j, as in our lunch example where $P_i = P_j$. In the true optimality condition, the ratios of *marginal* value (MV) to price are equated:

(4) $$MV_i / P_i = MV_j / P_j.$$

From equation (2), the marginal value of each alternative i is:

(5) $$MV_i \frac{\partial V}{\partial \pi_i} = V_i + \pi_i \sum_j^n \frac{\partial V_j}{\partial \pi_i}.$$

Matching as in equation (3) instead of (4) ignores the interdependencies between i and all the other j alternatives, as represented by the second term in equation (5). This term may be extremely difficult to calculate in practice. Herrnstein and Prelec [1991, 147] comment that this failure to coordinate among interdependent payoffs is formally analogous to a multi-agent externality problem. Consider, for example, commuters who must choose whether to get to work by car or subway. If the cost of each is the same, each commuter optimizes by adopting the quicker mode. This becomes an example of "matching" — commuters will crowd onto the faster alternative until the time spent on either is the same. While it is individually rational, it is socially sub-optimal — every commuter does not consider the congestion her choice imposes on others. Dixit and Nalebuff [1991, 229] have a simple exposition of this well-known problem.

This multi-person analogy can be pushed further. Consider the optimal consumption of sandwiches in our story as a form of subsidized "mass transit." Recall that the optimal consumption of sandwiches yielded a subjective value of only $0.95 per sandwich — less than the sandwich costs! To maximize total value, the high subjective return on pizzas should be used to "subsidize" the consumption of sandwiches — more than the sandwiches themselves would merit. This is like taxing people who keep cars in the city in order to subsidize the ridership of the subway.

When asked how his matching law differed from these classic problems of the prisoner's dilemma, Herrnstein once replied, "With this dilemma, you only need one prisoner" [Speech to the Society for the Advancement of Socio-Economics, New York City, March 1993]. In this sense we are all prisoners of our own desires. Failing to coordinate and control our desires, we wind up leaving each urge free to maximize "its own" return. It's not so bad to eat too much pizza, perhaps, but matching is now used to model all sorts of serious addictions [Herrnstein and Prelec, 1992; Heyman, forthcoming]. Someone may want to quit cocaine, a sexual obsession, or a life of crime, but finds the behavior too rewarding to resist in the short run. With initially high but steeply declining marginal returns, such addictions can clearly make one their prisoner.

There is a suggestive analogy here between the optimal provisioning of public goods by the state, and the optimal disciplining of desires by the superego. The superego is "super" in the sense that it can have a "meta-preference" (a preference about one's preferences) to not be addicted. The social and religious support for such a superego is a key aim of successful anti-addiction groups like Alcoholics Anonymous [Bartold and Hochman, 1988].

But not all matching behavior is "addictive." I will define addiction broadly as any behavior that is directly self-injurious (it may also be subject to social sanctions), but which the self-injuring person is nonetheless unable to control. Matching would not be so common if it were always self-injurious. But as we will see, it characterizes all known animal species. If not strictly "optimal," when does matching lead to some improvement? In formal terms, when does indulging in behavior i with the highest average value (AV_i) also produce the highest marginal value (MV_i)? In our sandwich-pizza problem:

$$(6) \qquad V_s = 1.50 - \pi_s, \qquad V_p = 4(1-\pi_p) = 4\pi_s, \qquad \text{since } \pi_s = 1 - \pi_p,$$

$$MV_s = 1.50 - 2\pi_s, \qquad MV_p = 4 - 8\pi_p = 8\pi_s - 4.$$

It is straightforward then that $V_s > V_p$ implies $\pi_s < 0.30$, and thus $MV_s > MV_p$. Then $MV_s < MV_p$ must imply $V_s < V_p$, and the former will be true whenever $\pi_s > 0.55$. In terms of Figure 1, this means that whenever sandwiches give a average payoff higher than pizzas (i.e., whenever $\pi_s < 0.30$, the point where the sandwich and pizza lines cross) we should eat more sandwiches. If, on the other hand, pizza is giving a higher average value, and in addition we are to the right of the average value curve's peak (i.e., whenever $\pi_s > 0.55$) then we should eat more pizza.

This means that choosing the lunch with highest average value winds up producing the right marginal allocation except when π_s is between 0.30 and 0.55. If π_s is uniformly distributed on the unit interval, it falls into this narrower interval only 25 percent of the time. Otherwise, the average and marginal indicators point in the same direction. This shifts our attention from *optimization*, requiring the equality of all marginal values, and focuses it on the concept of *melioration* [Herrnstein and Prelec, 1991] — requiring only that marginal values be brought somewhat closer together.

The Allais Paradox

Another violation of the expected utility axiom is the "paradox" discovered by Maurice Allais, the Nobel laureate. To illustrate, imagine that as Chief Financial Officer (CFO), you must instruct your lieutenants how to invest $100 million of your company's daily cash-flow. You tell them that you regard a certain return of one percent, or $1 million on this $100 million, as better than most of the "lotteries" offered by the market. Given a choice between lotteries A and A*, for example:

A :	The certainty of $1 million.
A* :	1/100 chance of No Profit ($0). 89/100 chance of $1 Million. 10/100 chance of $5 Million,

you instruct them to always choose A over A*, which you write A ≻ A*.

"On the other hand," you continue, "there are times when it is worth taking a small risk for a much greater profit." Thus, if faced with lotteries B and B*:

B :	89/100 chance of No Profit ($0). 11/100 chance of $1 Million.
B* :	90/100 chance of No Profit ($0). 10/100 chance of $5 Million.

you instruct them to always choose B* over B, or B* ≻ B.

This kind of choice has been staged by many experimenters [Machina, 1998], both as a thought-experiment and with modest amounts of real money. The most common of choices are the ones our CFO just made, A ≻ A* and B* ≻ B. But, as I shall now show, these choices are inconsistent under expected utility theory.

This inconsistency does not depend on any assumptions about risk aversion, or equivalently, the shape of the subject's utility function. Thus, for any "shape" of utility function U(·), we can write:

$$\text{(7a) } A \succ A^*: \quad U(\$1m) \succ (1/100)\cdot U(\$0) + (89/100)\cdot U(\$1m) + (10/100)\cdot U(\$5m)$$
$$\Rightarrow \quad (11/100)\cdot U(\$1m) \succ (1/100)\cdot U(\$0) + (10/100)\cdot U(\$5m),$$

while

$$\text{(7b) } B^* \succ B: (89/100)\cdot U(\$0) + (11/100)\cdot U(\$1m) \prec (90/100)\cdot U(\$0) + (10/100)\cdot U(\$5m)$$
$$\Rightarrow \quad (11/100)\cdot U(\$1m) \prec (1/100)\cdot U(\$0) + (10/100)\cdot U(\$5m).$$

Implications (7a) and (7b) are obviously contradictory. (We write these as the preference relation " >" rather than as the numerical inequality " >", since they clearly violate the algebra of expected utility.)

The violation of expected utility has the further problem that it is "dynamically inconsistent"— our CFO has given instructions that are self-contradictory in terms of decision trees. "Decision trees" are commonly used to formalize decisions in business, public policy, and computer programing. A decision tree is the "extensive form" of a game, as opposed to its "normal" or matrix form, if the information and payoffs available to its players are well-defined [Gardner, 1995, 28-32]. The extensive form includes information on the sequencing of moves, so the same normal form game may sometimes be decomposed into several different extensive forms. An extensive form game or decision tree that is equivalent to the choice between lotteries A and A*, and B and B*, can be shown as Figures 2a and 2b, respectively, following Machina [1989].

The square nodes in each tree are points at which a decision must be made, while round nodes represent stochastic moves by "nature," with probabilities summing to one. If one commits to a particular course of action at each decision node, then it can be seen that the *product* of the probabilities along all the non-excluded paths must sum to one across paths. At the decision node in Figure 2a for example, assume one commits to path A with certainty — eliminating path A* from consideration. Then at the root of this tree, before nature moves at the first node, there are two ways to arrive at the same final outcome of $1 m., with a probability summing to 100/100. Similarly, if one commits to A* rather than A, this means three final outcomes are possible, with associated probabilities: 10/100 ($5m), 1/100($0), and 89/100($1m). Thus it can be seen that committing to path A is equivalent to choosing lottery A, while committing to path A* is equivalent to choosing lottery A* in our original Allais problem. Figure 2b is identical to 2a except for the lowest branch outcome, which has fallen to $0. This makes a commitment to path B is equivalent to lottery B, whereas path B* is equivalent to lottery B*.

Thus we have the same two lotteries at this choice node (A = B and A* = B*), but embedded within two different larger lotteries. This is known as a *compound lottery*. The independence axiom of expected utility theory says that one's choice over simple lotteries should not depend on the particular compound lottery in which this choice is embedded. The Allais paradox clearly violates independence. Our CFO's underlings might well find themselves confused by this, since their boss's instructions to choose A *and* B* — rather than the pairs (A, B) or (A*, B*) — commits them to making the opposite choice over the same final options. Machina [1989, 1634-38] notes that this dynamic inconsistency makes one prey to what he calls "the money pump."

To illustrate, suppose that a manager follows her CEO's instructions in Figure 2a by committing to a futures contract of A rather than A*. The economic forecast then changes, and Figure 2b is now her best available information. If this manager wants to stick to her boss's instructions, she will now have to pay something to be released from her old commitment of A (now B), so that she can pledge herself to B* (formerly A*). A change of projections back to Figure 2a could send her scurrying back to A, and paying another penalty to be released from her previous contract. If such a "money pump" were to continue, her firm could obviously be sucked dry.

FIGURE 2a
Dynamic Choice Problem Generating Lotteries A and A*

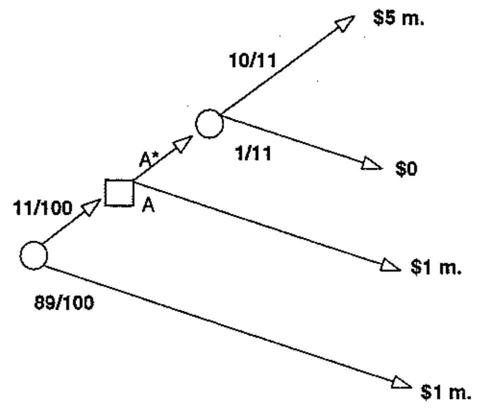
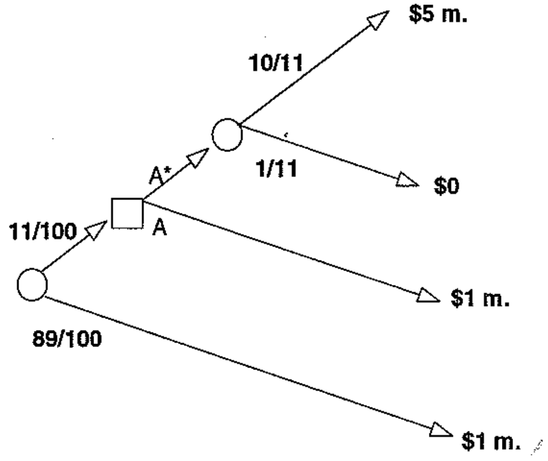


FIGURE 2b
Dynamic Choice Problem Generating Lotteries B and B*



Note that if the manager actually reaches the decision node in either Figure 2a or 2b, the tree's lowest branched payoff will at that point have been eliminated. It will have simply never happened, and so merit no further consideration. This is the meaning of the independence axiom — one's choice over a sub-lottery should be independent of the larger lottery in which it is nested. When confronted with the Allais results, Raiffa [1968] suggested that subjects be taught how to decompose the problem into compound lotteries, as in Figures 2a and 2b. If they really understood that the choice of sub-lotteries is identical in both figures, Raiffa conjectured, subjects would then wish to act in accord with the independence axiom. This would be like our manager wishing that her boss's instructions had been consistent. Raiffa's conjecture has been confirmed experimentally by Conlisk [1989]. The Allais paradox can thus be seen as the tendency to "oversimplify" a complex lottery, as will now be shown.

RISK-AVERSION AND "COMPLEXITY AVERSION"

Observed Preferences and Complexity

Expected utility maximization is an ideal norm of self-interested rationality. The Herrnstein and Allais versions of sub-optimality show that most people are, in fact, not terribly good at making simple calculations of expected utility. This means that in such risky-choice problems, people face an additional risk — the risk of making a mistake. But the Herrnstein and Allais behaviors can be also be seen as attempts to minimize that risk through "efficient simplifications."

To see this in the Allais case, note that in (7a) and (7b), U($0) can be normalized to zero, since Von Neuman-Morgenstern expected utility is only unique up to a linear transformation. The last line of (7b) is then (10/100)·U($5m) >(11/100)·U($1m), and substituting this into the top line of (7a) implies

$$\text{(8)} \qquad U(\$1m) > (89/100)\cdot U(\$1m) + (11/100)\cdot U(\$1m).$$

By saying that a simple lottery is preferred to its component sub-lotteries, (8) is clearly a violation of the "independence axiom," and so cannot be explained by simple risk-aversion. It does make sense, however, for individuals who are not only averse to risk, but to the risks of computational errors.

Consider a generalization of (8). By iterated substitution of sub-lotteries, (8) becomes

$$\text{(9a)} \qquad V(x) > \sum_j^m \pi_j \cdot V(x) > \sum_k^n \pi_k \cdot V(x), \text{ where } 2 \leqslant m < n.$$

Then it follows algebraically, dividing by $V(x)$, that

$$\text{(9b)} \qquad 1 \geqslant \sum_j^m \pi_j \geqslant \sum_k^n \pi_k .$$

Equations (9a) and (9b) say that the greater the number of partitions of expected value in a *compound* lottery, or "lottery of lotteries," the smaller their subjective sum. This can be interpreted as a preference for less complexity, *ceteris paribus*. Here it is important to note that the other inconsistent choice in the Allais problem, (A^*, B), would imply a preference for *more* complexity, since it reverses the preference relations in (7a, b) and therefore in (9a, b). However, this symmetric Allais-paradox choice is significantly less common than the choice (A, B^*), and may be due simply to random information processing errors [Conlisk, 1989].

The probabilities in (9b) are what Kahneman and Tversky [1979] call "subjective" and clearly in error, since they sum to less than one. This was also found by Ellsberg

[1961] in his famous experiment on unknown proportions. In this experiment, most subjects preferred to bet on the color of a ball pulled at random from an urn filled with an equal number of red and black balls, rather than bet on a second urn with unknown proportions of red and black. Ellsberg's subjects were "distributional pessimists" because, whatever color they might guess for the unknown-proportion urn, they acted as if its probability of being drawn was less than 50 percent, the known probability for the other urn. This leads to a contradiction, since the probability of picking *either* a red *or* a black ball from the unknown-proportion urn would then be less than one.

The complexity-aversion of behavioral inequality (9a) can be seen as predicting Ellsberg's result, since the unknown-proportion urn is really a complex compound lottery, one with as many different sub-lotteries as there are possible proportions of black balls, $\pi_b \in [0, 1]$. All these lotteries, or values of π_b, are equally likely. It is apparently not obvious to most people that this unknown-proportion urn has the same *ex-ante* probabilities as the known-proportion urn. They may be convinced by an intuitive argument, but proving it takes a bit of integral calculus.[1]

Imagine scanning a complicated compound lottery such as (9a) and making an initial guess that it holds with equality, rather than inequality. You would then be indifferent among all three lotteries of eaution (9a), only if you were sure that

a. you never made mistakes in calculating compound lotteries, or
b. you made imperfect but unbiased calculations and were risk-neutral.

But (a) is impossible, and (b) extremely unlikely for most human subjects. In fact, few people would ever have enough information to have any basis for saying whether or not (b) were true. Since behavioral inequality (9a) violates the independence axiom of Von-Neuman-Morgenstern utilities, I will say that such preferences exhibit *complexity-aversion*, rather than simple risk-aversion.

Recall Raiffa's [1968] conjecture that subjects shown how to build up compound lotteries would not violate the independence axiom. To test this conjecture, Conlisk [1989] guided subjects through the choices leading up to a compound lottery — using trees like Figures 2a and 2b. This "tutorial" reduces the incidence of both inconsistent choices, (A, B^*) and (A^*, B). Interestingly, this tutorial also eliminated the previous systematic bias — that (A, B^*), which implies complexity aversion, was much more likely than (A^*, B), which does not. These inconsistencies now appear as random rather than systematic errors. Conlisk argues that the usefulness of expected utility theory does not require human infallibility, but only that subjects' errors do not systematically violate its axioms.

Now return to the Herrnstein matching problem. Note that in allocating extra resources to average as opposed to marginal payoffs, the subject is in effect declaring by equations (3) to (5) that

$$(10) \qquad V_i > V_i + \pi_i \sum_j^n \frac{\partial V_j}{\partial \pi_i}, \forall i.$$

These Herrnstein preferences are similar to the Allais preferences in (9a) — the more complex the representation of a choice problem, the less desirable that choice becomes. If calculation is costly and error-prone, this is not obviously irrational. Often there will not be substantial preference interdependencies, so that the summation of terms in (10) is not worth calculating. Even if they do not net out to zero, one may well decide that:

a. perfect calculations are unlikely, and even if they are perfect,
b. by the time calculation is finished, the initial information is likely to be out of date.

A Simple Model of Complexity Aversion

Both the Herrnstein and Allais types of complexity aversion can be modeled as aversion to the risk of errors. Let an individual have a *certainty equivalent* income, x^{CE}, reflecting expected utility over a lottery of the form:

$$(11) \qquad V(x^{CE}) \equiv \sum_i^n \pi_i V(x_i).$$

Defining $\sum_i \pi_i x_i = \boldsymbol{x}$, assume that the agent knows $\boldsymbol{x}$, but must estimate x^{CE}. Let us assume simple risk aversion, or the concavity of V(x). This implies $\boldsymbol{x}^{CE}$, and that

$$(12) \qquad V(\boldsymbol{x}) - V'(\boldsymbol{x})(\boldsymbol{x} - x^{CE}) \geqslant V(x^{CE}) \geqslant V(\boldsymbol{x}) - V'(x^{CE})(\boldsymbol{x} - x^{CE}),$$

where $V'(x)$ indicates a first derivative with respect to x.

Inequality (12) is a simplification, because if the individual does not have perfect computational ability, x^{CE} in (11) will not be known with certainty, only estimated. Call $est^t(x^{CE})$ the *estimated* certainty equivalent, i.e. the estimate of the *true* certainty equivalent at time t. If every attempt to estimate x^{CE} gets more accurate, and if that additional accuracy can actually be comprehended, then $est^{t+1}(x^{CE}) > est^t(x^{CE})$. This follows from the concavity of $V(x)$; if two estimates of x^{CE} are unbiased around the same midpoint, the estimate with the narrower range of values must have a higher expected utility. If calculations were costless, then further estimations would continue until perfect accuracy was achieved.

There are two sorts of costs that can limit this process, however: the costs of *generating* better estimates, and the cost of assimilating or *processing* these estimates. It is almost a truism of the computer age that as the costs of generating more accurate information fall, our human ability to process information may only get more overwhelmed. Thus more could be less, and we might be less able to accurately *use* a more detailed estimate — even if that estimate itself were known to be more accurate. It is not clear which cost constraints, generating or processing, will be more binding in most choice problems.

First, consider the direct resource cost of estimation. One may just be making increasingly accurate estimates of a steadily shrinking pie. This resource cost can be expressed as a lowering of the average income derived from x, net of the effort of calculation. If resources are spent at every estimation effort t, then comparing estimates at times t and t+1, $\boldsymbol{x}^{t+1} < \boldsymbol{x}^t$. Assume this decay in real resources is known and measurable by the agent.

Next, consider the processing costs shown by behavioral inequalities (9a) and (10), which limit an agent's ability to comprehend estimates that have become too complex. These inequalities imply that even if resources have been spent to derive a more accurate estimate, so that $\boldsymbol{x}^{t+1} < \boldsymbol{x}^t$, we may still find that $est^{t+1}(x^{CE}) < est^t(x^{CE})$. Thus the limits to reasonable estimation may derive not just from the cost of generating further accuracy, as reflected in the fall of $\boldsymbol{x}^t$, but also in the costs of processing those estimates, as reflected in a fall in $est^t(x^{CE})$. I will try to model both sorts of limits.[2]

Assume that the "distributional pessimism" of complexity aversion means that the *estimated* certainty equivalent is never greater than the *true* certainty equivalent which a perfectly calculating agent would presumably know, so that $est^t[x^{CE}] \leqslant (x^{CE})^t$. To reflect the costs in both generating and processing further estimates, I will now rewrite the lower bound in (12) in terms of $est(x^{CE})$ instead of x^{CE}, and put a time superscript on all terms to indicate that this is an iterated estimate.

$$(12a) \quad V(\boldsymbol{x}^t) - V'(\boldsymbol{x}^t)(\boldsymbol{x}^t - (x^{CE})^t) \geqslant V(x^{CE})^t \gtreqless V(\boldsymbol{x}^t) - V'(est^t[x^{CE}])(\boldsymbol{x}^t - est^t[x^{CE}])$$

The upper bound in (12a) is just the limit imposed by simple risk-aversion, in the absence of any issues of complexity aversion. The lower bound can be thought of as the limit implied by complexity aversion; i.e., the rate at which further estimates are actually felt to be utility-enhancing. We can use (12a) to show a hierarchy in the costs of complexity — that a change in the *estimated* certainty equivalent is more important for expected utility than a change in the actual certainty equivalent.

> ***Proposition:*** The resource costs of estimation alone cannot decrease the lower bound of (12a). Both the upper and lower bounds of (12a) fall, however, if in addition to these resource costs, the estimated certainty equivalent falls.

The basic idea of the proof is not hard to see. First, assume an expenditure of resources, $\boldsymbol{x}^t > \boldsymbol{x}^{t+1}$, but with no loss of the *estimated* certainty equivalent. This means $(x^{CE})^t$ must fall with each iteration, but $est^t[x^{CE}]$ does not. Then the upper bound of (12a) must fall — $\boldsymbol{x}^t$ falls, its derivative therefore rises, and the final term $(\boldsymbol{x}^t - (x^{CE})^t)$ increases by simple risk aversion. The lower bound, however, would actually rise. Its derivative term $V'(est^t[x^{CE}])$ is greater than $V'(\boldsymbol{x}^t)$, so the drop in $\boldsymbol{x}^t$

times $V'(est^t\,[x^{CE}])$ must fall by more than $V(\boldsymbol{x}^t)$ itself falls. (Recall that we are here assuming $est^t\,[x^{CE}]$ does not change.)

If the estimated certainty equivalent $est^t\,[x^{CE}]$ actually rises, as one would hope, then this lower bound of (12a) will increase by even more, *a fortiori*. So using up resources has an ambiguous effect if it does not decrease the estimated certainty equivalent. Any rise in this estimated certainty equivalent, however, although it does use resources, at least improves the outlook on the worst case.

Now assume the opposite — a fall in the estimated certainty equivalent, $est^{t+1}\,[x^{CE}] < est^t\,[x^{CE}]$, but with no fall in resources or the true certainty equivalent: $\boldsymbol{x}^t = \boldsymbol{x}^{t+1} \Leftrightarrow (x^{CE})^t = (x^{CE})^{t+1}$. The fall in $est^t\,[x^{CE}]$ forces the derivative term and its multiplicand $(\boldsymbol{x}^t - est^t\,[x^{CE}])$ in the lower bound to rise, so the lower bound falls. Thus added complexity would be costly even if it did not use added resources. If $\boldsymbol{x}^t$ and $(x^{CE})^t$ fall, as they must whenever estimates use resources, then as already shown the upper bound must fall, and the lower bound would fall by even more.

This proposition shows a hierarchy of constraints, so that the final processing of more accurate estimates is more important than the raw costs of estimation itself. This may be related to Herrnstein's [1991] finding that there is a large dispersion of choices among individuals. De Palma et al. [1994] model this dispersion with different computational abilities. In the present model, such differences operate only on the level of resource costs, and efficiency on this level can be trumped by the costs of processing complexity. Even without any individual differences in estimation costs, there could still be significant differences in complexity aversion, leading to different final choices.

AN EVOLUTIONARILY STABLE STRATEGY

The Allais results have not to my knowledge been tested on nonhuman subjects, but there is little reason to expect other species to do better than humans over compound lotteries. Herrnstein's matching is certainly not encouraging in this regard. A review by Williams summarizes:

> The generality of the matching relation has been confirmed by a large number of different experiments. Such studies have shown matching, at least to a first approximation, with different species (pigeons, human, monkeys, rats), different responses (keypecking, lever pressing, eye movements, verbal responses), and different reinforcers (food, brain stimulation, money, cocaine, verbal approval). Apparently, the matching relation is a general law of choice. [1988, 178]

This robust behavior is precisely what evolutionary biologist John Maynard-Smith [1982] terms an *evolutionarily stable strategy* (ESS). An ESS is a non-cooperative (Nash) equilibrium which, once set by a large population, cannot be invaded by behavioral mutations in a few individuals. Since they are non-cooperative equilibria, ESS are in general neither unique nor Pareto-efficient. Complexity aversion is therefore probably only a local optimum, given the existing genetic limitations of most species.

A formal model of complexity-aversion as an ESS is not undertaken here. But a brief conjecture on the reasons for this apparent stability may be in order. It does not seem likely that raw computational power poses the main obstacle. A computationally capable mutation would not face severe physical constraints, unlike say, a human being born with wings. After all, generations of economics students have learned to solve problems like Herrnstein's sandwich-pizza choice, at least on tests. The mental ability required to do so without special training has almost certainly occurred in many mutations. If computational power is not the main problem, this reinforces the hierarchy of constraints in equation (12a).

Rather than raw computability, the binding constraints on the success of such mutations may be (a) the cost of acquiring data, and (b) the cost of communicating the results to other agents who are themselves not optimal processors. An example from business practice may illustrate the point. Few managers use marginal analysis, even if they understand the principle perfectly well [Baumol, 1977, 34-35]. Baumol conjectures that this is because marginal calculations require information the firm may not have. Salvatore [1993, 300-07] has an illuminating discussion of the difficulty in estimating marginal cost functions in practice. Unless a firm wants to experiment with radical changes in its production levels and input mix, it will usually not have data outside of a narrow range of experience. Past data, if they exist, are often not comparable due to changes in technology and product.

Calculating averages, by contrast, requires only a knowledge of the firm's current operating level. If the environment is changing rapidly enough, then "melioration", not optimization, may be the only feasible goal. By the time one collected data, estimated costs, calculated the equality of marginal payoffs, and then moved to equalize them, the true marginal payoffs would have changed.

SUMMARY AND EXTENSIONS

Herrnstein's matching and the Allais paradox both demonstrate a preference for simplifications, driven by the increased risk of mistakes in more complex calculations. This simplification may be efficient given the costs of generating more complex estimations, or in trying to understand them after they are generated. One empirical prediction is that an increased complexity of nested sub-lotteries, even without any change in objective probabilities, will be treated as having lower expected utility. Another is that differences in complexity aversion lead to variability in decisions, even for individuals with the same computational ability.

This behavior's extraordinary evolutionary stability should be investigated. If the real barrier to marginal calculation is a lack of usable data, then better models and simulations can help. If it is basic information processing costs, then these are certain to fall. But if it is assimilating this new information, there is a problem of design and presentation. Whatever the source, our inability to optimize can be greatly reduced by information technology, with large potential gains in efficiency. There may also be role for basic economic-psychological education here — at least to convince people that they really do need such support, even for apparently simple and everyday calculations.

NOTES

Thanks to Ralph Noble, my psychologist colleague, and to the editor and two annonymous referees for their extensive comments. I also wish to express gratitude to the late R. J. Herrnstein, who seemed pleased to enter into long arguments with undergraduates like me. The usual caveats apply with unusual force.

1. Integrating over all possible probabilities for the unknown-proportion urn shows that π_b and $(1-\pi_b)$, the probability of the Black (B) and Red (R) outcomes respectively, must each be ½:

$$\int_{\pi=0}^{\pi=1} \pi_b(B) + (1-\pi_b)(R)d\pi_b = (1/2)B + (1/2)R.$$

2. There is another sort of limit not modeled here which only strengthens our case. Apart from any processing costs, more detailed estimates may be often be, in fact, less accurate than "cruder" estimates. A recent study by Hlawitschka [1994], "The Empirical Nature of Taylor-Series Approximations to Expected Utility," uses parameterized utility functions to simulate selection of stock portfolios by expected utility. He shows that second-order Taylor approximations of expected utility are often more accurate than much higher-order expansions — even for series that converge asymptotically. After all, he notes, such convergence is only a property of the limit.

REFERENCES

Barthold, T. A., and Hochman, H. M. Addiction as Extreme-Seeking. *Economic Inquiry*, January 1988, 89-106.

Baumol, W. J. *Economic Theory and Operations Analysis*. Englewood, NJ: Prentice-Hall, 1977.

Conlisk, J. Three Variants on the Allais Example. *American Economic Review*, June 1989, 392-407.

De Palma, A., Meyers G. M. , and Papageorgeiou, Y. Y. Rational Choice Under an Imperfect Ability to Choose. *American Economic Review*, June 1994, 419-40.

Dixit, A. K. And Nalebuff, B. J. *Thinking Strategically: The Competitive Edge in Business, Politics and Everyday Life*. New York: W.W. Norton, 1991.

Ellsberg, D. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 1961, 643-69.

Faulhaber, G. R. and Baumol, W. J. Economists and Innovators: Practical Products of Theoretical Research. *Journal of Economic Literature*, June 1988, 577-600.

Gardner, R. *Games for Business and Economics*, New York: John Wiley & Sons, Inc., 1995.

Herrnstein, R. J. Experiments on Stable Suboptimality in Individual Behavior. *American Economic Review*, May 1991, 360-64.

Herrnstein, Richard J. And Draen Prelec. A Theory of Addiction, in *Choice Over Time*, edited by G. Loewenstein and J. Elster. New York: Russell Sage Foundation, 1992, 331-60.

__________. Melioration: A Theory of Distributed Choice. *Journal of Economic* Perspectives. Summer 1991, 137-56.

Heyman, G. M. Resolving the Contradictions of Addiction. *Brain and Behavioral Sciences*, forthcoming.

Hlawitschka, W. The Empirical Nature of Taylor-Series Approximations to Expected Utility. *American Economic Review*, June 1994, 713-19.

Kahneman, D., and Tversky, A. Prospect Theory: A Theory of Decision Under Risk. *Econometrica*, March 1979, 263-91.

Machina, M. J. Dynamic Inconsistency and Non-Expected Utility Models of Choice Under Uncertainty. *Journal of Economic Literature*, December 1989, 1622-68.

Maynard-Smith, J. *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press, 1982.

Raiffa, H. *Decision Analysis*. Reading, MA: Addison-Wesley, 1968.

Salvatore, D. *Managerial Economics in a Global Economy*, 2nd. Ed., New York: McGraw-Hill, 1993.

Williams, B. A. Reinforcement, Choice, and Response Strength, in *Stevens' Handbook of Experimental Psychology*, Vol. 2, edited by R.C.Luce et. al. New York: Wiley, 1988, 167-244.