CHAPTER 7

Learned versus optimizing behavior in simple situations

R. DUNCAN LUCE

By "simple situations" I mean those experimental designs commonly used in psychophysics, learning, and motivation, especially in studies intended to test one or another mathematical model. In fact, in this chapter only experiments for which the subject has two possible responses are discussed because mathematical theories are moderately well worked out only for these simple situations. Even there we still do not understand completely what is involved, as we shall see.

If we examine existing mathematical theories for these simple situations, we find two quite different classes of models, which, when we focus on their mathematical structure, may be described as stochastic and static decomposition models or, when we focus on their psychological interpretations, as adaptive and analytic models.

Current adaptive (or learning) models attempt to capture a subject's trial-by-trial choice behavior in terms of a stochastic process. These models apply to situations in which the same choice is repeatedly presented and the outcome to the subject on each trial depends both on his response and on the stimulus presented. The theoretical problem is to discover what remains invariant from trial to trial and, therefore, what constrains

This work was supported in part by National Science Foundation Grant NSF G-17637 to the University of Pennsylvania.

changes in the response probabilities. To be sure, there are many different and competing learning models: Markov chain models that arise from Estes' stimulus sampling theory versus operators that transform response probabilities; linear versus nonlinear operators; path-independent versus path dependent models; and so on, but their common features are rather more striking than their differences. All these models are restricted to highly repetitive situations in which the subject adjusts his behavior to his recent experience (outcomes) in an "adaptive" fashion.

The analytic or, if you will, cognative models are of a different character. First, as suggested by the term "static decomposition models," they are static in the sense that no mechanism for trial-by-trial change is provided. Since we know that response probabilities do change with experience, these models are surely incomplete. Generally, it is claimed that they deal only with the asymptotic behavior that follows the learning transient. Second, these models describe possible behavioral relations between two or more related choice situations, that is, they "decompose" or "analyze" one choice into a function of several others.

Perhaps the simplest example of an analytic proposition is the weak stochastic transitivity hypothesis: if a person chooses a over b at least half the time when a and b are presented and b over c at least half the time when b and c are presented, then he will choose a over c at least half the time when a and c are presented. A more sophisticated example of an analytic theory is the well-known subjective expected-utility hypothesis which says that, for each subject, numerical utilities can be assigned to outcomes and numerical subjective probabilities to events in such a way that the numerical ordering induced by the subjective expected utility of gambles formed from these outcomes and events is the same as the subject's preference ordering. Here the decomposition effected by the theory is completely apparent: choices among gambles are systematically decomposed into numerical statements about their two constituent elements, the outcomes and the events. A third example of an analytic theory is any of the predictions that we adduce from the logical structure of the alternatives. Specifically, suppose a subject is comparing the relative likelihood of two events; then, when one event logically implies the other, we usually assume that the first will not be judged more likely to occur than the second.

Whatever we may mean by optimization in human behavior—and I do not think that it is particularly easy to say—surely it has something to do with the decomposition of complex alternatives into more elementary ones and with comparisons among alternatives. In some normative theories, such as dynamic programming, the optimization may also take into account sequential aspects of the situation, but, as far as I am aware, in the descriptive theories of choice that involve optimization there are only static comparisons and decompositions. If that is so, it will cause no harm, and in some ways will make the argument simpler, to generalize the "optimizing" of the title of this chapter to "analytic," that is, to treat descriptive theories of optimizing behavior as special cases of analytic theories of behavior.

The adaptive-analytic dichotomy of choice theories constitutes a current dilemma of mathematical psychology: beyond a doubt, people exhibit both adaptive learning and analytic understanding, and any theory that fails to incorporate both aspects is surely going to be shown wrong some of the time. Moreover, many psychologists feel that subjects exhibit still another mode of behavior, namely, hypothesis testing, which is an aspect of learning not included in most of our models. People not only adapt to experience and analyze complex choices into simpler ones, but they generate and test hypotheses in much the way that statisticians do. Few "hypothesis-testing" models exist, and little that is convincing is known about the phenomenon experimentally, so we shall ignore it.

With this as background, let us examine some experiments with an eye to deciding whether learning or analytic models better describe what is going on. My purpose is to illustrate by means of two well-worked-over classes of experiments just how elusive the answer to this question can be.

THE DEPENDENCE OF p_∞ ON π IN GAMBLING EXPERIMENTS

Since analytic theories do not explicitly provide for trial-by-trial changes in behavior, we can evaluate the two classes of theories only in terms of asymptotic data. Actually, this restriction is quite convenient for the adaptive models as well, because on the whole their asymptotic properties are better understood than their transient ones.

Consider the following experiment:

$$\begin{array}{cc} & \text{Response} \\ & \begin{array}{cc} \text{I} & \text{II} \end{array} \\ \text{Event} \begin{array}{c} \alpha \\ \bar{\alpha} \end{array} & \begin{bmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{bmatrix}, \end{array}$$

where on each trial the subject chooses a column, j, a chance event α then chooses a row, i, and the subject receives the outcome o_{ij} so determined. Let p_∞ denote the asymptotic expected probability that the subject will select response I and let $\pi = Pr(\alpha)$ denote the probability that event α will occur, assuming its occurrences form a Bernoulli chain. Question: with the outcomes o_{ij} held fixed, how does p_∞ depend on π as we vary π over its range from 0 to 1?

In examining the literature, I have uncovered eight possible answers to this question, four of them provided by learning theories and the other four by analytic theories. Because some are special cases of others, the eight theories actually yield only five distinct proposals.

First, the learning theory predictions. A one-element stimulus sampling model (Suppes, 1961) and an experimenter-controlled linear operator model (Bush & Mosteller, 1955) both predict

$$(1) \qquad p_\infty = \frac{\pi}{\pi + (1-\pi)\theta_2/\theta_1}.$$

When $\theta_1 = \theta_2$, (1) reduces to the well-known probability matching prediction: $p_\infty = \pi$. Under certain conditions (Lamperti & Suppes, 1960) an experimenter-subject controlled (nonlinear) beta model (Luce, 1959) yields

$$(2) \qquad p_\infty = \frac{b_{12}\pi + b_{22}(1-\pi)}{(b_{11}-b_{12})\pi + (b_{21}-b_{22})(1-\pi)},$$

where the b_{ij} are logarithms of the multiplicative learning rate parameters β_{ij}. The fourth learning model arises from a threshold theory for psychophysical detection (Luce, 1963a), which we discuss in more detail later. Suffice it to say that the subject is assumed to act as if there were discriminative stimuli perfectly correlated with the chance events, whereas, of course, such stimuli do not exist. This forces the two stimulus parameters of the phychophysical model to be numerically equal, yielding,

$$(3) \qquad p_\infty = \begin{cases} \dfrac{\pi q}{\pi + (1-\pi)\theta_2/\theta_1}, & \text{if } \pi < \pi_0 \\[2ex] q + \dfrac{\pi(1-q)}{\pi + (1-\pi)\theta_2/\theta_1} & \pi > \pi_0. \end{cases}$$

Next, the predictions from analytic theories. Some attention has been given to models in which a "response strength" is attached to each response and the probability of a particular response occurring is simply its response strength normalized by the sum of the strengths of all the possible responses (Luce, 1959). Becker, DeGroot, and Marschak (1963) have suggested that for gambling situations the response strength of a response should be given as the expected utility of the several outcomes that may occur when that response is chosen. For the foregoing experiment this leads to

$$(4) \qquad p_\infty = \frac{v_{11}\pi + v_{21}(1-\pi)}{v_{11}\pi + v_{21}(1-\pi) + v_{12}\pi + v_{22}(1-\pi)}$$
$$= \frac{v_{11}\pi + v_{21}(1-\pi)}{(v_{11}+v_{12})\pi + (v_{21}+v_{22})(1-\pi)}.$$

Although (4) is formally the same as (2), which arose from the beta model, there is an important difference: all the v_{ij} must be positive, whereas some of the b_{ij} may be negative (corresponding to $\beta_{ij} < 1$). Siegel (1959, 1961) suggested a utility model in which the expected utility depends not only on the utilities of each of the outcomes but also on the variability of the responses, as measured by $p(1-p)$. Maximizing the resulting function leads to a linear relation between p_∞ and π, which is a special case of the beta model (2). Earlier, Edwards (1956) arrived at the same prediction by another route. His relative expected loss minimization (RELM) rule was based on Savage's loss matrix, that is, the matrix of differences between the best possible outcome for the event that occurred and the outcome corresponding to the response made. The expected loss is then calculated for each response. Edwards assumed that p_∞ is a linear function of the relative expected loss. Finally, in 1959 I showed that if the choice axiom[1] holds for the preference probabilities and for the judgments of event likelihood and these two classes of probabilities are statistically independent (a decomposition axiom), then, when $o_{11} > o_{21}$ and $o_{12} > o_{22}$, p_∞ must be a monotonic increasing step function of π.

The qualitative nature of these five predictions is shown in Figure 7-1 in which the several functions have been plotted for specific parameter values. The differences are sufficiently gross that we might hope to choose among them experimentally.

There are a variety of ways to do such an experiment, of which at least two have been realized. In the so-called probability prediction experiments the subject is told next to nothing about the events, but a long series of trials is run at each value of π. One advantage of this procedure is that, by plotting proportions of responses for blocks of trials, we can tell whether the behavior is approximately asymptotic. Although it is not inherent in the design, all published data of this type are for groups of subjects (of four to more than 100 members). From our point of view this is unfortunate because there is little reason to suppose that subjects all have the same parameter values, and averages of nonlinear functions with different parameters may yield a distorted picture of the family of functions involved.

Be that as it may, my best estimates of the asymptotic group probabilities are plotted in Figure 7-2 for the five experiments that I have found for which at least two different values of π were run, explicit payoff matrices were employed, and the payoffs were symmetric in the sense that $o_{11} = o_{22}$ and $o_{12} = o_{21}$. Because the data curves seem to have slopes greater than 1

[1] The choice axiom asserts that if x is a response, R, a set of possible responses including x, and S, another set including R, then the response probabilities of x occurring when S and R are presented are related by $p_R(x) = p_S(x \mid R)$ provided that the conditional probability exists.

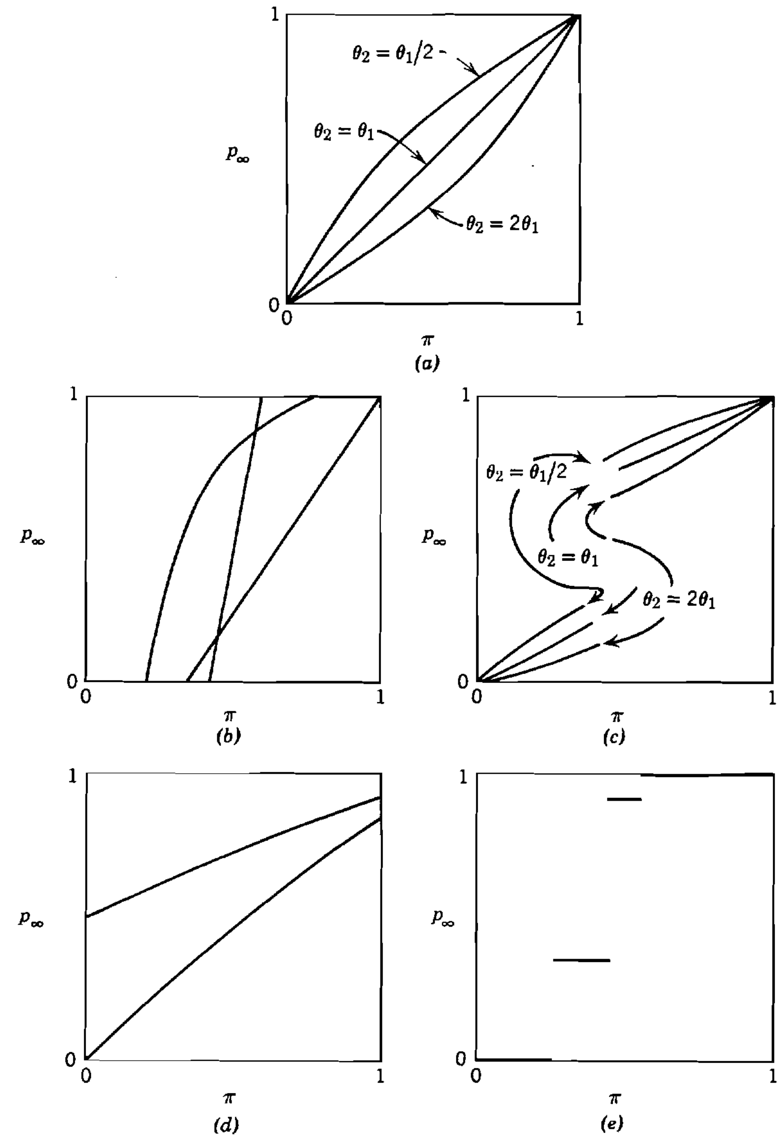

FIGURE 7-1. (*a*) Stimulus sampling and linear learning models. (*b*) Beta learning, utility of variability, and relm models. (*c*) Threshold model. (*d*) Strict expected utility model. (*e*) Decomposition model.

and to cut across the $p_{\infty} = \pi$ line, we can reject the one-element and linear learning models (1) as well as the combination of the choice model with the expected utility hypothesis (4). With so few data points—the most extensive study used only four values of π—and with group data it does not seem possible to choose among the other models.

Another way to do the experiment, which has not been popular among learning theorists but has been among utility theorists, involves giving the subject as much information as possible about the events, except, of course,

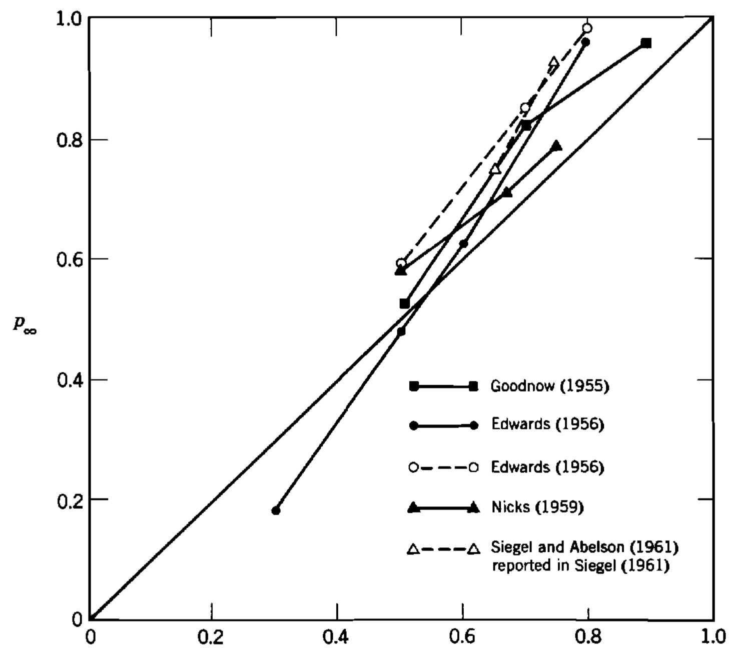

FIGURE 7-2

advance knowledge of the actual outcome, and to interleave many different presentations. For example, Luce and Shipley (1962) used 6 payoff matrices, 15 events (spanning a 0.2 probability range) per matrix, and 50 observations for each matrix-event pair; the subjects were told the mathematical probability of each event. The plots of p_{∞} versus π obtained in this way differ materially from those shown in Figure 7-2. The data for three of the five subjects are shown in Figure 7-3. Perhaps the transition from 0 to 1 for Subject 1 is comparable to that shown in Figure 7-2, but for Subjects 2 and 3, the transition is surely much more rapid. It is even more so for Subjects 4 and 5, who both exhibited simple discontinuities.

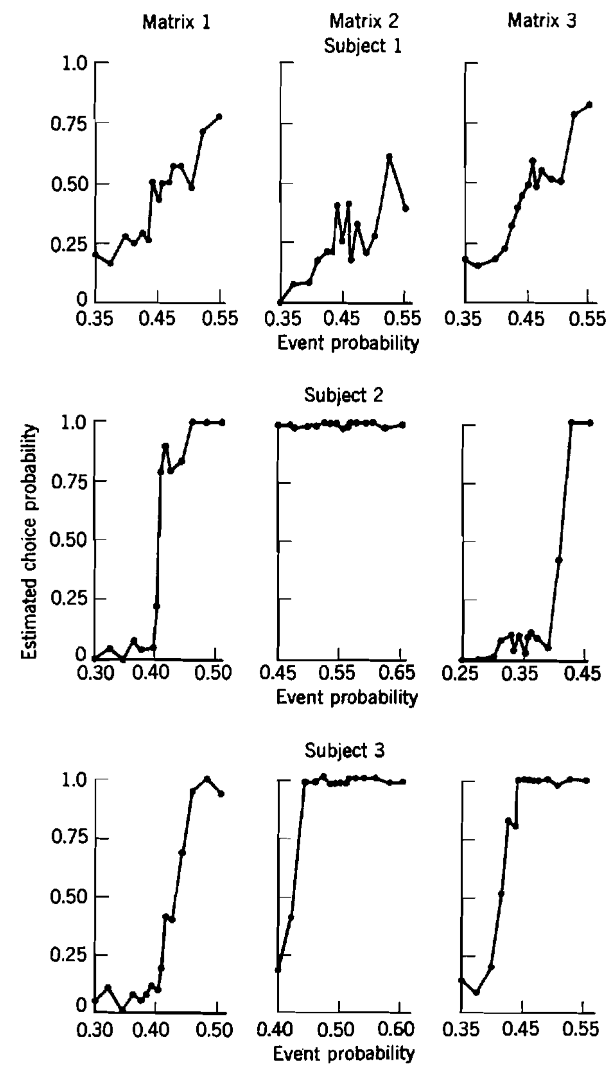

FIGURE 7-3

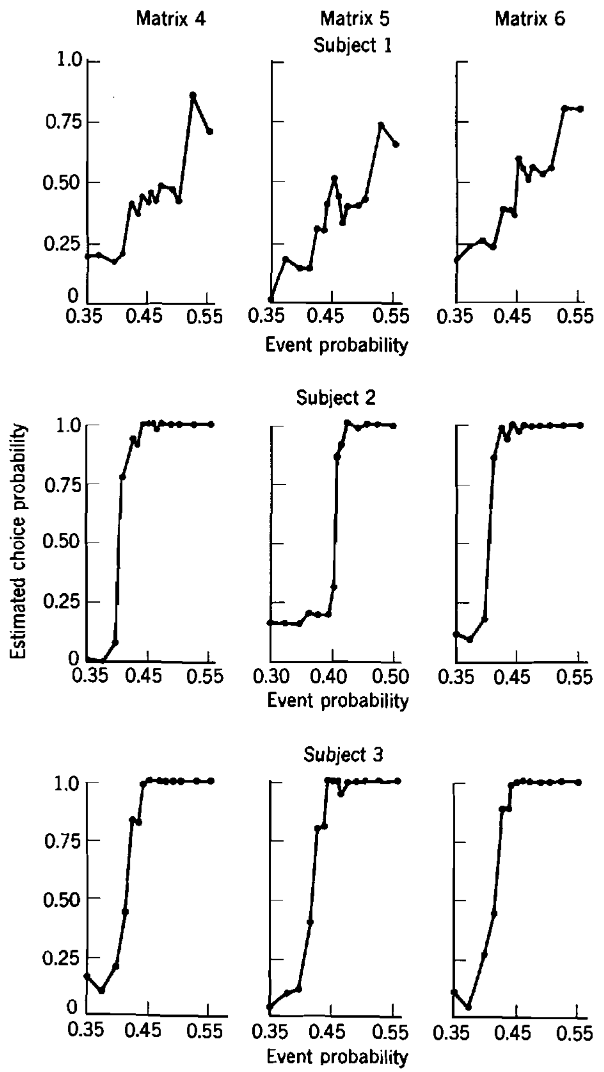

FIGURE 3 (*continued*)

As Luce and Shipley showed by a Monte Carlo calculation, the patterns of reversals in the estimated probabilities are extremely unlikely under the hypothesis of binomial variability superimposed on a smooth, strictly increasing function of the usual ogival type. Just what family of functions underlies these data is, however, another matter. The "plateaus" that seem to exist are consistent with the step-function prediction that arises from the choice axiom coupled with the decomposition axiom, but there exists enough rounding of the "discontinuities" to suggest that something more is involved. Undoubtedly learning effects were present, although they do not seem to be the primary phenomenon, and by the nature of the experiment we cannot be very confident that any of the plotted points really are estimated from asymptotic data.

This is just about all that is published about the plot of p_∞ versus π when explicit outcomes are employed. Surely these data are not sufficiently clear to decide whether the behavior is controlled primarily by an adaptive or an analytic process. Moreover, it is doubtful whether more experiments like these will help much, so we must consider alternatives. It seems to me that it would be worthwhile to carry out studies that possess the following virtues of both designs:

1. The subjects are informed fully about the probability structure of the mechanism generating the events.
2. Long runs are carried out at single π values so that it is possible to decide empirically what part of the data is approximately asymptotic.
3. A relatively large number (e.g., 15 to 30) of π values are used over the range in which the function is different from 0 and 1.
4. Data for individual subjects are reported.

DETECTION EXPERIMENTS

Current analyses of psychophysical detection experiments illustrate another difficulty in deciding whether a decision process is primarily adaptive or analytic, namely, that the observed behavior is assumed to result from a concatenation of two processes, neither of which can be observed separately.

Until recently, psychophysical experiments were analyzed as if only a sensory process were involved; however, with the advent of the ROC curve, or what I prefer to call the isosensitivity curve,[2] there is no longer

[2] The ROC (receiver operating characteristic) or isosensitivity curve is the exchange relation existing in any two stimulus (s_1 and s_2)-two response (r_1 and r_2) situation between the conditional response probabilities $p(r_1 \mid s_1)$ and $p(r_1 \mid s_2)$. It is substantially the same as a plot of the probabilities of the two types of errors in testing between two hypotheses. The nature of this exchange relation in a Yes-No detection experiment is discussed more fully later.

much doubt that a motivational process also controls the observed behavior. For example, in the simple Yes-No design a stimulus is presented in some proportion P of the trials and not in the remainder. On each trial the subject reports whether he thinks it was presented. Assuming asymptotic behavior, the data provide estimates of $p(Y \mid s)$, i.e., the probability of a Yes response when the stimulus is presented, and of $p(Y \mid n)$, i.e., the probability of a Yes response when it is not presented. When the stimulating conditions are fixed and either the payoff matrix or the presentation probability P is varied from run to run, the pairs of estimates of $p(Y \mid s)$ and $p(Y \mid n)$ appear to sweep out a smooth curve; see the data points in Figure 7-4 (the theoretical curves are discussed later). The brightness data

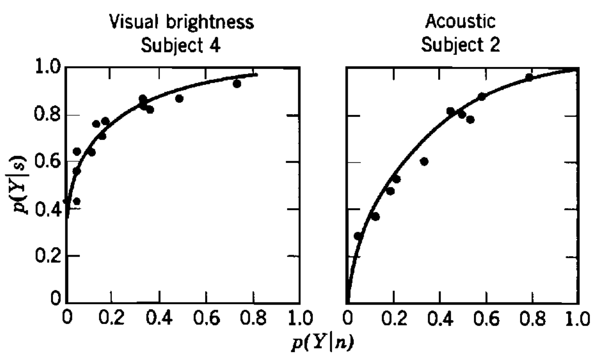

FIGURE 7-4

(Swets, Tanner, & Birdsall, 1961) were generated by varying the payoff matrix, and the acoustic data (Tanner, Swets, & Green, 1956), by varying the presentation probability P.

As already mentioned, current models propose that two distinct processes underlie these data: a sensory one that relates the stimuli to hypothetical internal states, which are interpreted as the possible representations of the stimuli, and a decision process that combines the internal state with other experimental information, such as the payoff matrix and the presentation probabilities, to arrive at a response. In most models the sensory process is assumed to depend only on the stimulating conditions and to be independent of the past experiences of the subject; the decision process is assumed to depend explicitly on the payoff aspects of the experiment, including in some cases the subject's experiences on previous trials.

To be more specific, in the signal detectability model (Green, 1960; Licklider, 1959; Luce, 1963b; Swets, Tanner, & Birdsall, 1961) the internal states are assumed to form a continuum, and the effect of a stimulus presentation is treated as a random variable distributed (normally)

over this continuum. When there are two presentations, as in the Yes-No experiment, the separation of the means of the two distributions, when normalized by the standard deviation of one of them, is treated as a basic stimulus parameter, d', of the model. The decision process involves the selection of a cut-point or criterion, c, such that the response is No to all presentations that result in an internal state less than c and Yes to those that result in a state larger than c.

For a given set of stimulating conditions, d' is assumed to remain fixed, and the isosensitivity curve is generated by varying the cut-point c over its possible values. Mathematically, eliminating c from the equations for $p(Y \mid s)$ and $p(Y \mid n)$ yields the functional relation between them with d' the single parameter of the resulting family of functions. The theoretical curves shown in Figure 7-4 were obtained this way.

As another example of a two-process model, consider the simplest sort of threshold model (Luce, 1963a) in which there are two internal states, D and $\bar{D}$, corresponding to whether or not a presentation exceeds the threshold. The conditional probabilities of going into state D given a stimulus and given no stimulus are denoted, respectively, by $q(s)$ and $q(n)$. These are the stimulus parameters of the model. The decision process is assumed to divide into two parts. Either the subject chooses to say Yes to all occurrences of the D state and to some proportion u of the occurrences of the $\bar{D}$ state, or says Yes only to some proportion t of the occurrences of the D state and No to the remainder of the D occurrences and to all of the $\bar{D}$ ones. As the bias parameter, u or t as the case may be, ranges over its possible values (the unit interval), the isosensitivity curve is generated for fixed stimulus conditions. This results in two straight-line segments, one from $\langle 0, 0\rangle$ to $\langle q(n), q(s)\rangle$ and the other from $\langle q(n), q(s)\rangle$ to $\langle 1, 1\rangle$, as shown in Figure 7-5. (The data points are for increments of brightness; they were reported by Swets, Tanner, & Birdsall, 1961.)

Given psychophysical models of these types, two new theoretical problems are presented. First, how do the stimulus parameters d' or $q(s)$ and $q(n)$ depend on physically measurable aspects of the stimulating conditions? Second how do the decision parameters c or u and t depend on other aspects of the design, such as the payoffs and the presentation probabilities that are believed to affect these parameters? We consider only the second problem here.

In signal detectability theory it is usually assumed that the subject chooses the cut-point c so that his expected money return is maximized. It turns out that the relevant function of payoffs and presentation probabilities is

$$\left(\frac{1-P}{P}\right)\left(\frac{o_{22}-o_{21}}{o_{11}-o_{12}}\right).$$

Green (1960) examined the relation between the predicted and actual cutoffs, assuming, of course, that the rest of the signal detectability model is correct, and he found that although they are monotonically related there is an appreciable and systematic discrepancy from perfect prediction. He argued that this failure of the model is due largely to the fact that the expected value function is nearly flat in the neighborhood of its maximum. It is also possible, however, that it stems from the fact that subjects use an entirely different decision mechanism, such as a quite simple adaptive

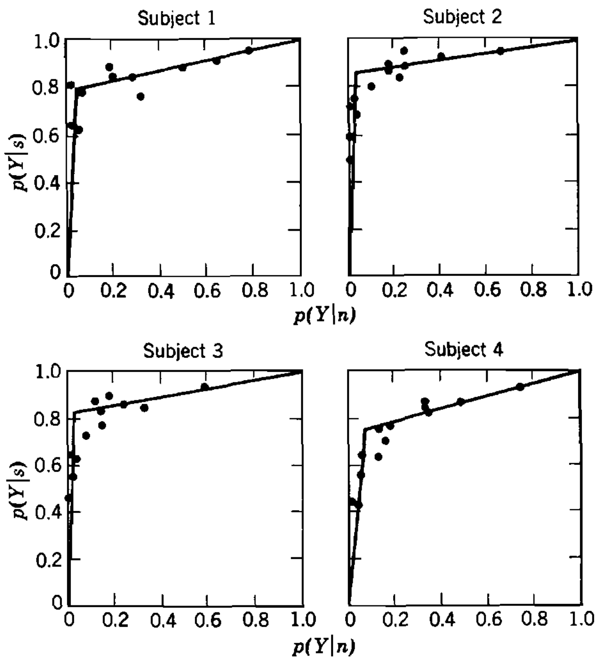

FIGURE 7-5

process. Because a continuum of states is involved, the mathematical problem of trying to formulate a simple learning model is considerable, and no formal test of this hypothesis has yet been attempted.

For the threshold model we have described, the maximization of expected payoffs results in a clearly incorrect prediction, namely, that the response probability pairs are either $\langle 0, 0\rangle$, $\langle q(n), q(s)\rangle$, or $\langle 1, 1\rangle$, which pair depending on the values of o_{ij}, P, $q(n)$, and $q(s)$. On the other hand, because there are only two internal states, it is not difficult to set up a simple experimenter-controlled linear learning model to describe the choice of the

bias parameters. This leads to asymptotic biases of the form

$$u_\infty = \frac{1 - q(s)}{1 - q(s) + [1 - q(n)]b} \quad \text{and} \quad t_\infty = \frac{q(s)}{q(s) + q(n)b},$$

where

$$b = \left(\frac{1 - P}{P}\right)\frac{\theta'}{\theta}$$

and θ and θ' are learning rate parameters. For the data shown in Figure 7-4 it is possible to select a simple relation between θ'/θ and the payoffs so that the predicted biases are in good accord with the observed ones (see Luce, 1963a, p. 73).

The important point about these examples is that any decision we make regarding the nature of the decision process is completely confounded with our decision about the nature of the sensory process. For example, if we were sure that the threshold model correctly described the sensory process, then there would be no question that subjects do not maximize their expected money returns. If, however, the threshold model were wrong, and the signal detectability model described the sensory mechanism, then this conclusion would not be so clearly warranted.

I might add that in some of our still unpublished experiments we find a suggestion that different decision processes may be involved at different times. It appears that subjects initially adjust their biases in much the way we would expect from a learning model, but after a considerable number of trials the "asymptotic" biases begin to drift toward one of the three "optimal" corners of the threshold isosensitivity curve.

CONCLUSION

It seems to me that the main conclusion to be drawn from these examples is that we are still not able to decide, even for the simplest experimental situations, whether a subject's asymptotic decision process is primarily adaptive learning or whether it is based on a more analytic evaluation of the situation. Our difficulties stem from at least two sources. In the gambling and probability prediction experiments they seem to come mainly from the way in which these experiments are carried out and from the presentation of only group data. In the psychophysical experiments, in which the data frequently seem to be in a more useful form, difficulties arise from the fact that a concatenation of at least two processes seems to govern the over-all behavior. A third possible source of trouble is that the decision process itself may be some mixture of adaptive and analytic processes. If so, then until better theories are devised, we can only hope to learn how to do experiments that primarily tap one or the other of the two processes.

REFERENCES

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Beh. Sci.*, **8**, 41–55.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.

Edwards, W. (1956). Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, **52**, 177–188.

Goodnow, Jacqueline J. (1955). Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, **68**, 106–116.

Green, D. M. (1960). Psychoacoustics and detection theory. *J. Acoust. Soc. Amer.*, **32**, 1189–1203. Reprinted in R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Readings in mathematical psychology*. New York: Wiley, 1963. Pp. 41–68.

Lamperti, J., & Suppes, P. (1960). Some asymptotic properties of Luce's beta learning model. *Psychometrika*, **25**, 233–241. Reprinted in R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Readings in mathematical psychology*, New York: Wiley, 1963. Pp. 404–412.

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: a study of a science*. New York: McGraw-Hill. Pp. 41–144.

Luce, R. D. (1959). *Individual choice behavior: a theoretical analysis*. New York: Wiley.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychol. Rev.*, **70**, 61–79. (a)

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1 New York: Wiley. Pp. 103–189. (b)

Luce, R. D., & Shipley Elizabeth F. (1962). Preference probability between gambles as a step function of event probability. *J. exp. Psychol.*, **63**, 42–49.

Nicks, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, **57**, 105–114.

Siegel, S. (1959). Theoretical models of choice and strategy behavior: stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, **24**, 303–316.

Siegel, S. (1961). Decision making and learning under varying conditions of reinforcement. *Ann. N.Y. Acad. Sci.*, **89**, 766–783.

Suppes, P. (1961). Behavioristic foundations of utility. *Econometrica*, **29**, 186–202.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychol. Rev.*, **68**, 301–340.

Tanner, W. P., Jr., Swets, J. A., & Green, D. M. (1956). *Some general properties of the hearing mechanism*. T.R. 30, Electronic Defense Group, Univer. Mich., Ann Arbor.