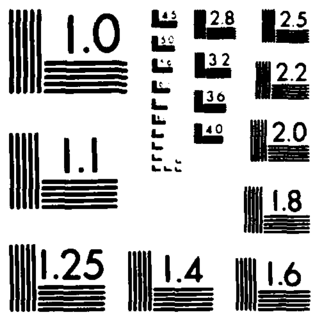
AD-A124 233 TOWARD A PROCEDURAL THEORY OF JUDGMENT(U) WISCONSIN HUMAN INFORMATION-PROCESSING PROGRAM MADISON L L LOPES DEC 82 WHIPP-17 N00014-81-C-0069 1/1

UNCLASSIFIED F/G 5/10 NL

END
DATE
FILMED
4-83
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

TOWARD A PROCEDURAL THEORY OF JUDGMENT

Lola L. Lopes

WHIPP 17 December 1982

Final Report

This research was supported by the Engineering Psychology Programs, Office of Naval Research, ONR Contract N00014-81-C-0069 Work Unit NR197-068.

WHIPP

WISCONSIN HUMAN INFORMATION PROCESSING PROGRAM

DEPARTMENT OF PSYCHOLOGY ■ UNIVERSITY OF WISCONSIN ■ MADISON, WISCONSIN 53706

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER WHIPP 17	2. GOVT ACCESSION NO	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Toward a Procedural Theory of Judgment		5. TYPE OF REPORT & PERIOD COVERED Final Report 1-Oct-82 - 31-Dec-82
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Lola L. Lopes		8. CONTRACT OR GRANT NUMBER(s) N00014-81-C-0069
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Wisconsin Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR197-068
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217		12. REPORT DATE December 1982
		13. NUMBER OF PAGES 56
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Judgment
Algebraic models
Bayesian inference
Averaging models
Multiplying models
Relative ratio models

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A procedural theory of judgment is described in which judgment is viewed as a serial "anchoring and adjustment" process. The process is described as comprising scanning, anchoring, and adjusting operations, the latter of which is applied iteratively (with order of adjustment steps usually determined by relative importance) until the judge deems that sufficient information has been integrated and outputs a final response. The paper has three major sections. First, the procedural theory is described and

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) contrasted with the implicit characterization of process that might be drawn from algebraic models of judgment. Next, three common forms of judgment are described (averaging rules, relative ratio rules, and multiplying rules) and discussed in terms of the factors that predispose subjects to "use" those rules for particular tasks. Finally, several common phenomena of judgment (primacy and recency effects, initial impression effects, differential weighting effects and violations of additivity) are interpreted within the serial adjustment model.

Toward a Procedural Theory of Judgment

Prologue

A few years ago I spent several months trying to interest a new graduate student in research on judgment, but I could not seem to arouse any enthusiasm. Finally, in desperation, I sent him off to the laboratory to run some subjects on a pilot project. A few weeks later he was back with a sheaf of computer output. "I want to show you something interesting," he said. He pointed to a sheet on which a single subject's data were laid out as a row-by-column array with the responses from the several replications listed one after another within the cells. "Look here," he said, "everytime you give the subject the same stimulus he gives almost the identical answer. And almost all the subjects do the same."

This student's revelation had, for me, a bittersweet quality, for he had "discovered" what all of us who have studied judgment seriously take as given: human beings, in making quantitative judgments about complex stimuli, do so in a manner that is systematic and replicable. Furthermore, their judgments often reveal algebraic patterns that suggest the operation of some internal mechanism for "computing" averages, products, relative ratios and so forth. Subjects, however, are not conscious of these computations nor are they aware of the algebraic form of their judgments. Where do the judgments come from? What psychological processes give rise to this algebra-less algebra? These are questions whose interest has eluded not only that particular graduate student, but also a number of professional researchers who have perfunctorily critized algebraic judgment models for their "as if" status and then moved on to study other topics.

Strictly speaking, the present paper is a final report on a research project that has attempted to show why subjects in a particular task -- the Bayesian inference task -- produce judgments that are more like averages than like inferences. The project, however, was based on a general procedural theory of judgment that applies not only to averaging, but to other algebraic forms as well. Thus, the broader purpose of the present report is to lay out the rudiments of this procedural theory more generally than has been possible in the previously published experimental reports (Lopes, 1981, 1982a,b).

The report will be divided into three sections. First, I will describe the procedural theory and contrast it with the implicit characterization of process that might be drawn from algebraic models of judgment. Second, I will describe three common forms of judgment (averaging rules, relative ratio rules, and multiplying rules) and will discuss the factors that predispose subjects to use these rules for particular tasks. Third and finally, I will show how several different phenomena of judgment (primacy and recency effects, initial impression effects, differential weighting effects, and violations of additivity) can be interpreted within the procedural theory.

A Procedural Theory of Judgment

In the present view, judgments are produced via a serial adjustment process in which an initial or "anchor" quantity is "adjusted" one or more times in accordance with other available information. The idea that judgment is a process of serially adjusting an internal quantity was first suggested by Slovic (1967) under the name "polarization and adjustment" and later named "anchoring and adjustment" by Tversky and Kahneman (1974). More recently, the process has been hypothesized to account for averaging-like results in Bayesian inference (Lopes, 1981, 1982a,b; Wallsten, 1976b), diagnostic inference (Einhorn & Hogarth, 1982), and similarity judgment (Lopes & Johnson, 1982; Lopes & Oden, 1980). The present paper extends these earlier treatments by showing how serial adjustment processes mediate the algebraic form of the judgments produced.

A flow diagram of the proposed process is given in Figure 1. The process is described as comprising a set of operations in which information is scanned, items are selected for processing, scale values are assessed, and adjustments are made (at least after the first step) to an interim quantity that summarizes the results from already-processed information. These various operations are described in detail below.

Scanning. In the scanning operation the judge assesses what information has been presented for judgment. The details of this stage depend, of course, on how the information is presented to the judge. Three cases can be distinguished. In *sequential presentation* the items of information are presented one at a time, with or without an overt response from the judge after each item. Obviously, the scanning stage is rudimentary in sequential presentation since at any moment there is only one item to scan.

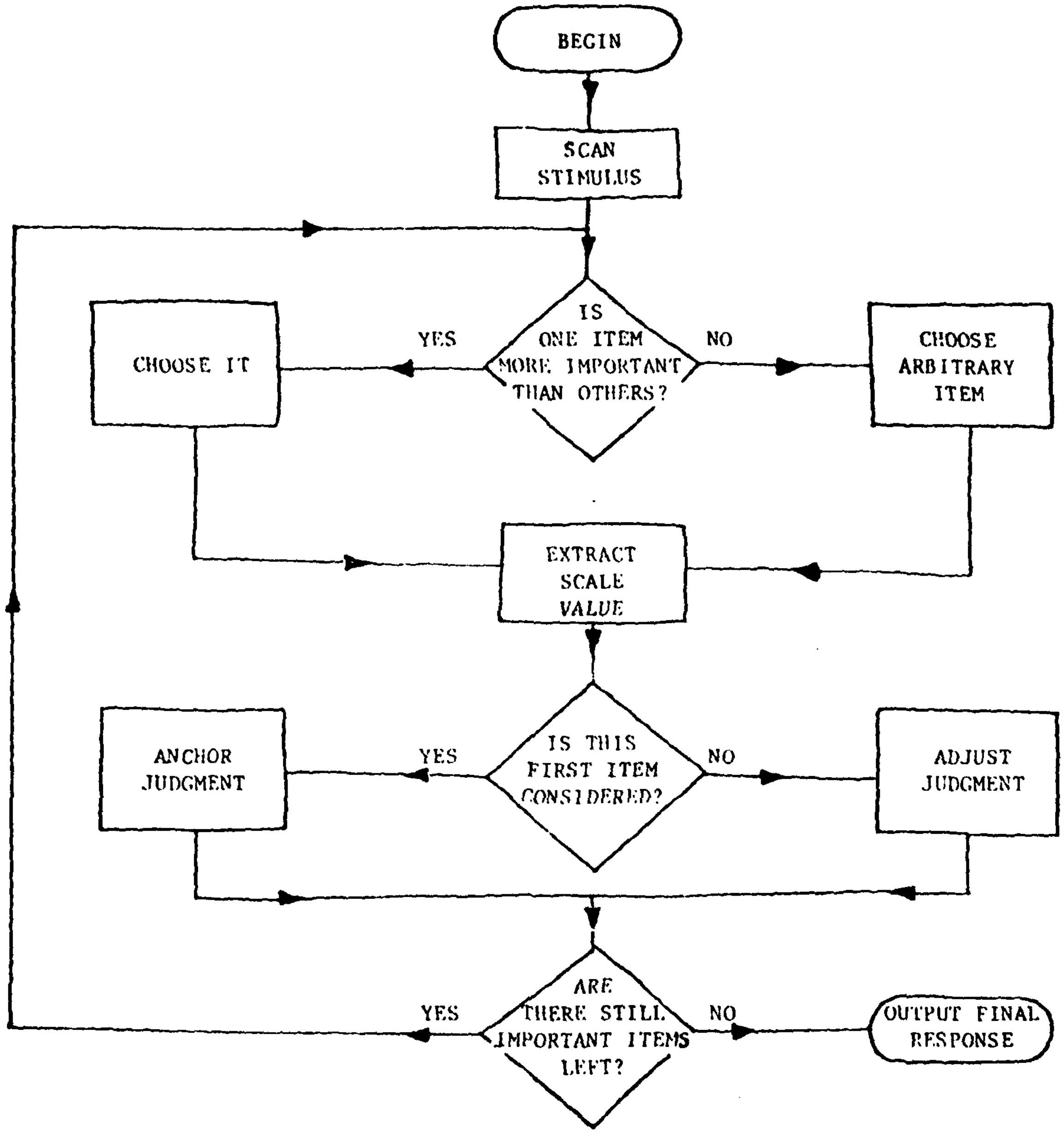


Figure 1. Flow diagram of serial adjustment judgment model.

In undifferentiated simultaneous presentation several items of information are presented at the same time and there is no a priori reason to suppose that any particular item is more important than any other. In such a case, if the number of items presented is small, scanning will probably include all available items, with the order of scanning determined by the physical format of the stimulus; i.e., if the information is presented as a list, scanning will probably proceed from top to bottom. If the number of items of information is large, however, scanning may terminate before all available items have been scanned.

In differentiated simultaneous presentation several items of information are presented at once but items differ from one another in terms of the kind of information they contain or in terms of their relevance for the judgment task. One example might be an application for graduate school containing an applicant's GPA, GRE scores, undergraduate college, letters of reference, personal history, and reasons for wanting to pursue advanced study. Another example might be an advertisement for an apartment giving rent, locale, number of rooms, and so forth. With complicated materials such as these the scanning process will undoubtedly reflect the judge's perception of the relative importance or usefulness of the different types of information. Some items may reliably be scanned before other items (e.g., GPA and GRE scores may be scanned first in a graduate school application) and some items may not be scanned at all, or at least not on the initial pass (e.g., hobbies or reasons for wanting to attend graduate school).

It is assumed that the primary function of the scanning operation is to orient the judge to the information available in the stimulus. Although some preliminary evaluation and integration of the stimulus material may occur during the stage, this would give only a rough impression of the value of the stimulus. Such a rough impression would ordinarily not be the final response, since that would be developed as the result of later and, usually, more deliberate passes through the stimulus information.

There may, of course, be exceptions to this rule. For example, in scanning applications for graduate school, a judge may come across one that is deficient in every respect. If a very crude scale of judgment is being used (e.g., assigning the applicant to one of five categories) the judge may assign the application to the lowest category without further processing. If, however, the judge is using a more continuous scale of judgment (e.g., rating the applicant's probability of success in graduate school) and if, as is usually the case in experimental tasks, the judge has been cautioned against making hasty "end responses," then further processing would be required in order to position the stimulus reliably on the judgment scale.

Anchoring. After the stimulus information has been scanned, the judge chooses some item of information to serve as a quantitative "anchor" (Tversky & Kahneman, 1974) for the judgment process. If the information is presented sequentially, of course, only one item is available at a time so the judge must use the first item presented as the anchor. If the information is presented simultaneously, the judge has several items to choose from. For items that are relatively undifferentiated (e.g., a list of trait adjectives in an impression formation task) the judge may simply anchor on whatever adjective happens to be on top of the list. For items that are differentiated, however, the judge will ordinarily select an anchor on the basis of the *a priori* importance of the information type. For instance, GRE scores or GPA might be chosen to anchor judgments of graduate school applicants. Alternatively, an item might be chosen because of the unusual diagnosticity of its particular information. For example, prior work experience would probably not be considered very important for judging graduate school applicants. But experience as a lion tamer might be considered interesting enough to use as an anchor for someone applying for study in animal behavior (or risky decision making).

As will be seen, the anchoring operation can have important consequences for the judgment process. As Tversky and Kahneman (1974) pointed out, adjustments to the anchor quantity are often "insufficient" in the sense that final judgments often lie nearer to the anchor than they would if the process were independent of which stimulus served as the anchor. Thus, anchoring and adjustment can lead to judgmental primacy which, in turn, can affect the algebraic pattern of the judgments produced. These anchoring effects will be discussed more fully in the sections on primacy/recency and on differential weighting.

Valuation. Once an anchor stimulus has been chosen, it must be evaluated relative to the scale of judgment. In some cases this valuation operation may yield a quantity that serves directly as the anchor value. For example, in a Bayesian inference task, subjects may simply anchor their judgment at the scale position that is proportional to the number of "target" items in the initial sample (Lopes, 1981). In other cases, however, the initial judgment may be less extreme than the scale value of the anchor stimulus. In these cases the initial judgment appears to be a compromise between the value of the stimulus information and some internal, presumably more neutral "initial impression" (Anderson, 1967).

As can be seen from the figure, stimulus valuation occurs not only in the course of anchoring the judgment, but also in subsequent stages as other stimulus information is brought into the judgment. This valuation process can be simple or complex depending on the nature of the judgment task. In some cases, the stimulus information seems to "contain" the pertinent value information fairly directly. This is seen most easily in the impression formation task in which the given personality trait adjectives seem to contain the required "likeableness" values as part of their meaning. In other cases, however, inference must be done to determine the scale value of the stimulus information. This would be true, for example, in a task in which subjects were asked to use aptitude information to judge the proficiency of people in various occupations (e.g., Anderson & Lopes, 1974).

In many judgment tasks the valuation process is complete when the scale value of the currently attended stimulus has been obtained. In some situations, however, the obtained scale value is processed further by comparing it with some expectancy or with the values of previous stimuli. For example, Lopes (1972) showed how some striking contrast effects obtained by Jones, Worchel, Goethals, and Grumet (1971) in an attitude attribution task could be explained without recourse to notions of stimulus interaction. It was proposed that subjects evaluated the content of an essay purportedly written by the target person against the sort of essay such a target person might have been expected to write. The discrepancy between the actual and the obtained essay was then integrated into the judgment along with the value of the essay itself.

In a similar vein, some subjects in a goal-setting task (Lopes, 1976) seemed to set their goal for the next trial in a maze-running problem in accordance both with their scores from the previous few trials and with the discrepancy between their most recent score and their next-most-recent score. For these subjects, it was as though they evaluated their performance on each trial not only in terms of the score itself, but also in terms of whether or not they showed improvement on the task.

Adjustment. Once anchoring has been accomplished, the judge must decide whether there are still important items unprocessed. If so, the process essentially reiterates, with the judge choosing which of the remaining items to consider next. As can be seen in the figure, the considerations at this point are exactly what they were at the time of choosing the anchor: if one of the remaining items is clearly more important than the others, the subject chooses it, otherwise an item is chosen arbitrarily and the scale value of the chosen item is then determined.

The next step in the judgment process is adjustment of the initial value according to the new information. It is this step that is most crucial in determining whether the algebraic form of the judgment is an average, product, relative ratio, or other combination rule. Detailed discussion of how the adjustment process is related to the algebraic form of the judgment will be deferred to the section on algebraic rules. For the moment it suffices to say that the adjustment process has two important subprocesses: choice of the <u>direction</u> in which adjustment will occur and choice of the <u>magnitude</u> of the adjustment. These subprocesses will reflect, in large part, the judge's internal representation of the judgment task and the judgment scale.

The Psychological Nature of Algebraic Processes

Evidence supporting the algebraic nature of human judgments has been accumulating for more than 20 years (cf. Anderson, 1981). However, as Graesser and Anderson (1974, p. 697) pointed out, establishing an algebraic model for judgment data is "only a first step in the analysis of the judgment process." Beyond this lies a clear need to reach an understanding of the cognitive mechanisms that are involved when people produce averages, products, or other algebraic forms.

Algebraic models are ordinarily expressed in what I will call <u>declarative</u> form. By this I mean that all the variables that enter the judgment are expressed along with the algebraic rule that defines the relationship among these variables in the judgment. No indication is given, however, of the <u>procedures</u> through which the computation is performed. Thus, the model applies to the final output of the judgment process rather than to the stages through which the judgment is generated.

For illustration, consider the forms in which averaging models are usually expressed. Equation 1 gives the averaging rule in <u>absolute weight</u> form:

$$R_N = \frac{\sum_{i=1}^{N} (w_i s_i)}{\sum_{i=1}^{N} w_i} \qquad (1)$$

In this form, the response given to N stimuli, R_N, is seen as the quotient of two sums: in the numerator, the sum of the various stimulus scale values (s_i) each multiplied by their respective weights (w_i), and in the denominator, the sum of the weights. (For simplicity I have ignored the possibility of an initial impression. If there were one, it would be assigned weight w_0 and scale value s_0 and the summation would run from i = 0 to N.)

Equation 2 gives the same averaging rule in relative weight form:

$$R_N = \sum_{i=1}^{N} \left[\frac{w_i}{\sum_{i=1}^{N} w_1} s_i \right] \qquad (2)$$

This expression differs from the absolute weight form primarily in that it suggests that the various scale values are each multiplied not by their absolute weights, but by their relative weights. These, in turn, are simply ratios of the individual absolute weights to the total absolute weight. (Sometimes the relative weight form of the averaging rule is written without showing the normalization of the absolute weights, i.e., the division by the sum of the weights. This is done by noting outside the equation that the sum of the weights is 1.)

Equations 1 and 2 are intended to describe the output of the judgment process, but not the process itself. In particular, the sequence of operations that one would use to literally instantiate computation of an average via either of these equations is not intended as a description of the sequence of operations that judges perform during the judgment process. Instead, the equations are intended to show (a) what quantities must be supposed to be functional in the judge's representation of the stimulus information and (b) how these quantities are related to one another mathematically in the description of the final judgment.

A major assumption of the present procedural theory is that the judgment process operates serially by producing "running" values as stimuli are sequentially selected and integrated into the judgment. Equation 3 shows how an averaging model can be written to emphasize this serial integration:

$$R_n = \frac{w_n}{\sum_{i=1}^{n}}(s_n) + \frac{\sum_{i=1}^{n-1} w_i}{\sum_{i=1}^{n}}(R_{n-1}) \qquad (3)$$

In the equation, R_n is the response (either internal or external) after the first n stimuli have been integrated. It is shown as a weighted average of the scale value of the new stimulus item (s_n) and the old response, i.e., the response value after n-1 stimuli.

Equation 3 captures the serial aspect of judgment, but it does not convey the sense that the old response is transformed to a new response by an adjustment process that increases or decreases the value of the old judgment in accord with the relationship between the value of the old judgment and the value of the new information. Equation 3 can be rewritten, however, to a change form that captures the notion of adjustment fairly well:

$$R_n = w_n'(s_n - R_{n-1}) + R_{n-1} \quad (4)$$

In this form, the variable w_n' stands for the relative weight of the information currently being integrated. Note that the new judgment, R_n, is simply the old judgment plus or minus some proportion of the difference between the value of the new information and the value of the old response.

Equation 4 comes much closer than the previous equations to capturing in algebraic terms the various psychological operations that are hypothesized to operate when judges produce averages. There are, however, certain respects in which even this equation misrepresents important psychological features of the judgment process. To begin with, Equation 4 is an equation, which is to say that it shows how certain numerical (i.e., digital) quantities are to be operated on arithmetically to produce other numerical quantities. In contrast, the procedural theory outlined above need not have its various values instantiated numerically (although they could be). Likewise, there would need be no literal subtraction of R_{n-1} from s_n, no literal multiplication of the difference by w_n', and no literal addition of the quantity to R_{n-1}.

In the second place, Equation 4 does not reveal how the characteristics of the various stimuli to be integrated influence the order in which they are processed or how, in turn, the order of processing affects the algebraic form of the judgment. This is particularly true with respect to the important ordering role played by weights in the procedural theory. Information integration theory has traditionally emphasized the theoretical necessity of construing stimuli in integration tasks to have both weights and scale values (cf. Anderson, 1981, p. 334-355), but it has not specified the different functions that weights and scale values play in judgment except to point out that weights measure, more or less, the salience or amount of information in a stimulus whereas scale values measure the position of a stimulus on a particular scale of judgment. The present procedural theory amplifies this important distinction between weights and scale values by showing their functional differences in the judgment process.

Finally, Equation 4 does not specify how the internal judgment is related to the external response. Traditionally, algebraic theories have treated the integration stage (in which the algebraic mechanism operates) as quite separate from the response stage in which the judge interprets his internal judgment in the explicit terms required by the experiment. Anderson (cf. 1981, p. 4-5), for example, refers to a "psychomotor law" in which a response function operates to map an implicit response into an observable response. This separation of internal from external response, in turn, has led some algebraic theorists to distrust the surface patterns of judgment data and to suppose that the internal algebraic process may be quite different from that suggested by the data (cf. Birnbaum, in press). In the prevent view, internal and external response processes are brought closer together. Although it is obvious that thoughts differ fundamentally from their observable manifestations in judgment tasks (e.g., marks made on paper, movements made on a potentiometer, etc.), it will be seen that there is often a close connection between internal or implicit responses and some of the more superficial characteristics of the external response scale.

Adjustment Processes and Algebraic Form

In this section I will show how adjustment processes can be related to three algebraic patterns (averaging rules, relative ratio rules, and multiplying rules) that are commonly observed in judgment data. Before proceeding, however, it is useful to point out some psychologically important differences between the sort of analogical procedures that are assumed by the present model and more ordinary numerical procedures for computing algebraic results.

For sake of exposition, suppose that a subject is asked to average 73 and 42. One way to proceed would be to add 73 and 42 to yield 115 and then divide the result by 2 to yield 57.5. Numerical processes such as this are characterized by three properties: (1) The numerals that are manipulated in the arithmetic are symbols that have been assigned to quantities arbitrarily. (2) The operations that are applied to the numerals are rewrite rules that specify the mapping between the ordered strings of numerals in the problem statement and the ordered string of numerals in the answer. For example, in adding 73 and 42, we know that the "3" and the "2" in the ones places of the two addends combine to yield a "5" in the ones place of the answer. (3) Quantitative interpretation of the answer string requires information that is not in the string itself. That is, since the answer string does not represent quantity directly, the interpreter or decoder of the string must apply outside knowledge about the mapping between strings of numerals and the quantities they represent.

Another way to average the numbers would be analogically. This could be done, for example, by locating the value 73 on some quantitative continuum, then finding the location of 42 on the same continuum, and finally moving halfway between 73 and 42. Note that none of these operations would need involve numerals since analogical processes use quantities (e.g., spatial position, intensity, etc.) directly to represent other quantities (e.g., the average of 73 and 42). For example, if the continuum were printed as a line on a piece of paper, the scale positions corresponding to 73 and 42 could be located and folded to meet one another. The position of the fold would then mark the answer of 57.5.

There are several good reasons to doubt that the cognitive processes that mediate algebraic judgment are numerical. The first is that for many judgment tasks, neither the stimuli nor the responses may ever be represented numerically, as would be necessary for numerical computation. In impression formation, for example, subjects are typically asked to read a set of personality trait adjectives and then rate (often on an unmarked continuous scale) the likeableness of a person having the traits. In verbal tasks such as these it is extremely unlikely that subjects would recode the stimulus information into numerical form and then operate on that. Instead, it appears that human judgment processes can operate as readily on quantities that are defined verbally or vaguely as it can on quantities that happen to be expressed numerically.

A second reason is that judgmental difficulty does not seem to increase as quickly as a function of problem size as would be expected if the judgment process involved the sorts of symbol manipulation that occur in numerical computation. For example, Shanteau (1974) had subjects rate the worth of single gambles (e.g., fairly likely to win sandals) and double gambles (e.g., fairly likely to win sandals and unlikely to win bicycle). Even though the double gambles were obviously more complex than the single gambles, there was no tendency for the single gambles to produce "better" data. This result contrasts strongly with results from studies of mental arithmetic (e.g., Dansereau & Gregg, 1968; Hitch, 1978) in which subjects perform numerical computations without pencil and paper. In these tasks, both errors and latency to solution tend to increase as problem difficulty increases.

A third and final reason is that it is rare for subjects in judgment tasks to know intellectually that their judgments conform to an algebraic rule. Since numerical computation requires explicit and rule-based manipulation of symbols, it seems exceedingly unlikely that subjects could be using even a rudimentary form of numerical computation without having at least some awareness of the basis for their judgments.

Analogical processes, on the other hand, can provide a plausible account of how algebraic judgments are produced. As will be detailed in the sections immediately following, they can operate via a variety of serial adjustment procedures to produce averages, relative ratios, products, and other algebraic results from stimulus quantities that are expressed either exactly or inexactly. In addition, they are computationally efficient, putting few demands on short term memory and allowing problems of varying size to be computed iteratively simply by recycling through the processes of stimulus selection, valuation, and adjustment until all (or sufficient) stimulus information has been integrated.

Averaging

Of the several algebraic models that have been proposed to describe human judgment processes, the model that is empirically most pervasive and theoretically best understood is averaging. Subjects appear to average in a variety of judgment situations ranging from number averaging in which averaging is the requested response to impression formation in which averaging is neither demanded nor disallowed logically to Bayesian inference in which averaging deviates both quantitatively and qualitatively from what is normative.

The most commonly obtained form of averaging is what is called constant weighted averaging. The term "constant weighting" refers to the fact that, for composite stimuli generated by factorial combination of stimulus elements, the weights associated with all stimuli within a given factor are equal. Weights can, however, vary between factors.

In constructing a procedural theory for constant weighted averaging, two important empirical findings must be explained. The first is the basic averaging result which, simply put, refers to the fact that averaging one quantity with another always results in a new quantity that lies between the two original quantities. Thus, if neutral information is averaged with information of high value, the neutral information lowers the average. On the other hand, if the same neutral information is averaged with information of low value it raises the average. The second finding is the parallelism result which refers to the fact that if judgments generated by a constant weighted averaging process are plotted as a factorial graph, parallelism will obtain between the various rows.

Both averaging and parallelism are easily explained within the proposed serial adjustment theory. Sticking for the moment to the number averaging example described above, suppose that a subject is asked to average the following set of numbers:

73
42
63
67

The first step would be to anchor the judgment at 73 (since it is first on the list) and then adjust downward to a position lying between 73 and 42. This between-ness property of the interim response would, by definition, constitute a generalized averaging result. If in addition the magnitude of the adjustment was one-half the distance between 73 and 42 (i.e., 57.5), the obtained average would be the simple arithmetic average of the two quantities and, since arithmetic averaging constitutes a special case of constant weighted averaging, parallelism would also hold in a factorially-defined set of such stimulus pairs.

The next step in the averaging procedure would be adjustment of the interim response to account for the value 63. For arithmetic averaging, the magnitude of this adjustment would have to be one-third the distance between 57.5 and 63 giving a new interim response of 59.33. The final adjustment step would follow the same pattern with an upward adjustment one-fourth the distance between 59.33 and 67. Note that for arithmetic averaging the adjustment rule is, for the nth stimulus in the average, to adjust 1/n of the distance between the scale value of the nth stimulus and the interim response to the first n-1 stimuli.

Arithmetic averaging is special in the sense that even when it is performed analogically, there still seems to be a strong numerical base to the process both in the representation of the stimulus values and in the fractional specification of the adjustment constants. These numerical overtones are, however, irrelevant in the more general definition of the averaging process. Constant weighted averaging requires only two things: first, that the adjustment process, however it occurs, yields a value that lies between the value of the stimulus being integrated and the value of the "old" response; and second, that the adjustment constant for a particular serial position be independent of the content of the stimulus information at that serial position.

For illustration, Figure 2 shows how a person might integrate a set of personality traits in an impression formation task without resorting to numerical

GREGARIOUS

DISORGANIZED

IMAGINATIVE

	Thought	Stage	Values
1.	*Let's see...gregarious, disorganized, imaginative...*	Scanning	Impression = neutral
2.	*I'll start here with gregarious.*	Anchor selection	
3.	*OK. I really like gregarious people...*	Valuation of anchor	Gregarious = very good
4.	*So that sounds like he'd be pretty good.*	Anchoring	Impression = good +
5.	*OK. What's next.*	Stimulus selection	
6.	*Uh oh. Disorganized isn't too good.*	Valuation of second item	Disorganized = not too good
7.	*That lowers the person quite a bit.*	Adjustment for the second item	Impression = above average
8.	*OK. Now imaginative...*	Stimulus selection	
9.	*That's good usually...*	Valuation of the third item	Imaginative = good
10.	*So it helps.*	Adjustment for the third item	Impression = fairly good

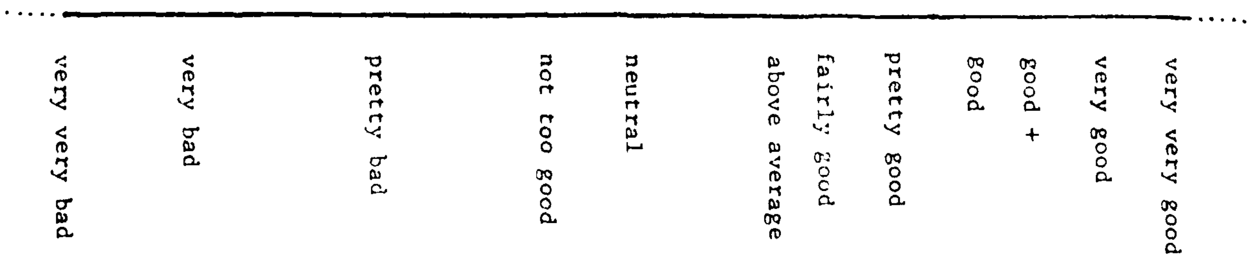


Figure 2. Example of how an analog judgment process might operate on non-numerical input.

processes. The stimulus set is shown at the top of the figure. The central portion of the figure shows (a) in the left column, fictional thoughts occurring during judgment, (b) in the middle column, a designation of the current judgment stage in terms of the procedural model of Figure 1, and (c) in the right column, a verbal expression of the running impression. At the bottom of the figure is given a hypothesized evaluative continuum with occasional verbal labels. This continuum is intended to represent a possible internal schema for evaluative terms. It is not intended to be taken literally as a spatial scale, nor is it intended that the verbal labels shown exhaust the verbal descriptions that a subject could give of positions on the continuum.

In Step 1, the judge scans the stimulus set. Since the items are fairly undifferentiated, the judge opts (in Step 2) for processing them in the order listed. In Steps 3 and 4, then, the first item is evaluated and used to anchor the judgment. Note that in this example, the judge began with a neutral initial impression. Thus, the impression value generated at Step 4 is somewhat less extreme than the value of the anchor adjective taken alone. In Step 5, the subject selects the next item for processing and, in Step 6, evaluates it. Step 7 represents the adjustment of the current impression toward the new stimulus item. Note that for constant weighted averaging to hold, the phrase "quite a bit" must be interpreted as a relational operator referring to a constant proportion of the difference between the new and the old value. In Step 8, the judge selects the next (and final) stimulus item. This is then evaluated in Step 9 and adjusted for in Step 10 to produce a final impression that is then mapped by the subject onto whatever overt response scale has been provided. Note that in this example I have shown the second adjustment (Step 9) to involve a smaller proportion of the difference between the old impression and the new item than was shown in Step 7. For anything approaching arithmetic averaging to hold, there would need to be successively smaller adjustment constants at each step with values of about a half, a third, and so forth. Ordinarily, however, averaging tasks display primacy, which implies adjustment constants that are even smaller than this. (More will be said about primacy and recency effects in a later section of the paper.)

As the example illustrates, averaging procedures seem intuitive and natural for the impression formation task, as they would for many other integration tasks. But averaging results are also found in some tasks where averaging seems irrational. A case in point is the Bayesian inference task in which subjects make judgments concerning the relative probability of two hypotheses (H1 and H2) based on one or more data samples (D1, D2, etc.).

According to Bayes' theorem, the probability of H1 given data sample D1 is as follows:

$$p(H1|D1) = \frac{p(D1|H1)p(H1)}{p(D1|H1)p(H1) + p(D1|H2)p(H2)} \tag{5}$$

where p(H1) and p(H2) are the prior probabilities of the two hypotheses and p(H1|D1) is the posterior probability of H1. If a second sample is introduced, the process simply reiterates with the values for p(H1|D1) and p(H2|D1) replacing the old prior probabilities. That is,

$$p(H1|D1\&D2) = \frac{p(D2|H1)p(H1|D1)}{p(D2|H1)p(H1|D1) + p(D2|H2)p(H2|D1)} \tag{6}$$

Looked at in terms of adjustment processes, Bayes' theorem constrains the direction in which adjustments can occur. This is easiest to see by simplifying and then rewriting Equations 5 and 6. First, let p(D1|H1)p(H1) be replaced by the value $\underline{a}$ and p(D1|H2)p(H2) by the value $\underline{b}$:

$$p(H1|D1) = \frac{a}{a+b} \tag{7}$$

Note that the equation reveals a simple relative ratio form. Now, we divide both numerator and denominator by $\underline{a}$ to give:

$$p(H1|D1) = \frac{1}{1+b/a} \tag{8}$$

In the same way, p(H2|D1) = b/(a+b) = 1/(1+a/b).

Consider how a new sample, D2, would affect the outcome. Let $\underline{a}'$ stand for p(D2|H1) and $\underline{b}'$ stand for p(D2|H2). Substituting into Equation 6 we obtain

$$p(H1|D1\&D2) = \frac{a'\frac{a}{a+b}}{a'\frac{a}{a+b} + b'\frac{b}{a+b}} = \frac{aa'}{aa'+bb'} \tag{9}$$

This can then be simplified as follows:

$$\frac{1}{1+(b/a)(b'/a')} \tag{10}$$

Focusing, for illustration, on H1, notice that if sample D2 favors H1, i.e., $b'/a' < 1$, then the value of $p(H1|D1\&D2)$ must exceed the value of $p(H1|D1)$ since the denominator term becomes smaller. Thus, anytime a new sample favors H1, adjustments must be toward increased support for H1. Under averaging, however, this directional constraint need not hold. In particular, if the current degree of support for H1 is more extreme than the degree of support offered by the new evidence, averaging-type adjustment will operate to lower the degree of support for H1.

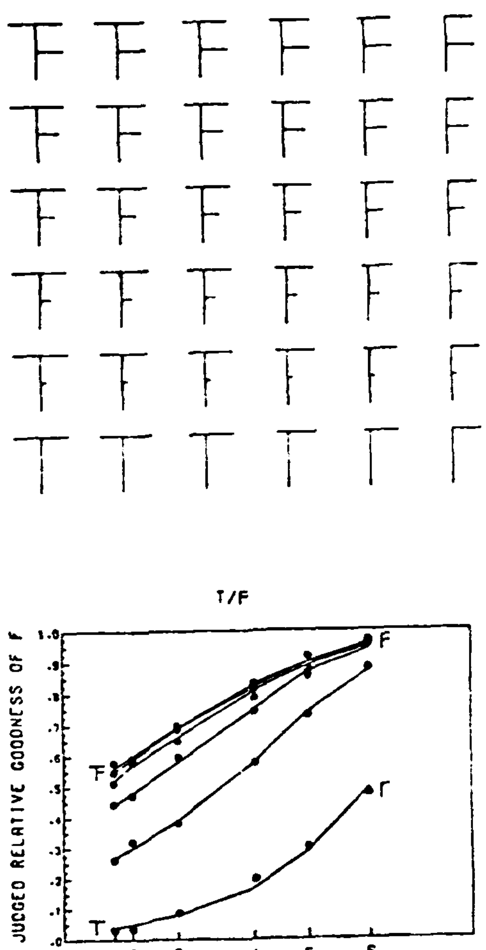
A considerable body of evidence supports the notion that subjects in Bayesian inference tasks produce data that are more like averages than they are like the relative ratios demanded by the normative theory (Beach, Wise, & Barclay, 1970; Marks & Clarkson, 1972, 1973; Shanteau, 1970, 1972). In addition, the current research program has demonstrated quite clearly that these averaging-like results can actually involve directionally inappropriate adjustments (Lopes, 1981, 1982a). In other words, subjects who are given new data that support the currently held hypothesis more than the alternative sometimes actually lower their confidence in the supported hypothesis and increase their confidence in the alternative.

If subjects always produced averages and never produced relative ratios, there would be little problem. One could suppose that the psychological pressure to make adjustments toward the scale value of the new information was simply so great that logical factors such as those influencing normative theories could not compete. But subjects sometimes do make ratio-like judgments even in tasks that are formally quite similar to Bayesian tasks. The question then arises as to the determinants of averaging-like and ratio-like processing. This is an important question, but it is best deferred until relative ratio rules have been discussed.

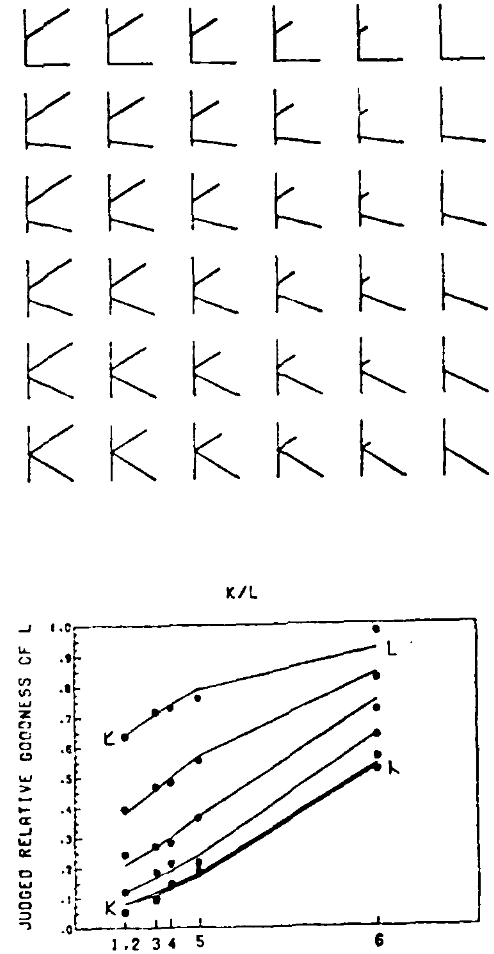
Relative Ratios

Ratio tasks are tasks of relative judgment. In the simplest case, there are two polar alternatives and a set of stimuli. The judge's task is to locate the values of the various stimuli relative to the alternatives. In many experimental tasks, the stimuli are generated by a factorial combination of features that vary in the degree to which they support the two alternatives.

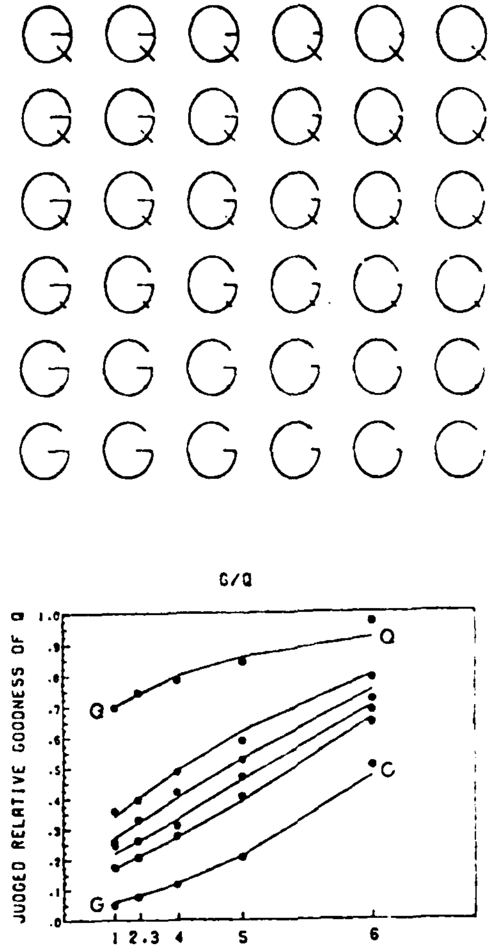
As noted above, the Bayesian inference task is a relative ratio task in which the "features" that are combined factorially are probabilities associated with different samples. That is, if we look, for example, at the three terms comprising



Mean judged relative goodness of *F* and predictions of the fuzzy logical model for the *T/F* matrix. (Note that the column spacing is proportional to the marginal means and that the order of the rows and of the columns is the same as for the stimuli.)



Mean judged relative goodness of *L* and predictions of the fuzzy logical model for the *K/L* matrix. (Note that the column spacing is proportional to the marginal means and that the order of the rows and of the columns is the same as for the stimuli.)



Mean judged relative goodness of *Q* and predictions for the fuzzy logical model for the *G/Q* matrix. (Note that the column spacing is proportional to the marginal means and that the order of the rows and of the columns is the same as for the stimuli.)

Figure 3. Example of task in which subjects generate relative ratio data. From Oden, 1979.

Equation 5, we see that each of them is separately the joint probability of the data and one or the other of the two hypotheses.

Relative ratio rules need not, however, involve probabilities. Oden (1979), for example, has used a relative ratio rule to model the degree to which people judge various character-like stimuli to be instances of one letter or another. His model is a fuzzy logical model for integrating the continuous truth values of various semantic propositions describing the candidate patterns.

Figure 3 shows both the data and the stimuli from Oden's experiment: Note that each stimulus matrix (upper panel) involves two features that vary continuously across a factor. For example, in the T/F matrix at the left, the row feature is the degree to which the left-hand portion of the T-bar is present or absent and the column feature is the degree to which the middle F-bar is present or absent. In the matrix, perfect examples of F and T are located in the upper right and lower left cells, respectively. The other stimuli are, to varying degrees, less than perfect and the subject's task is to say, for each stimulus, the degree to which it is a T or an F.

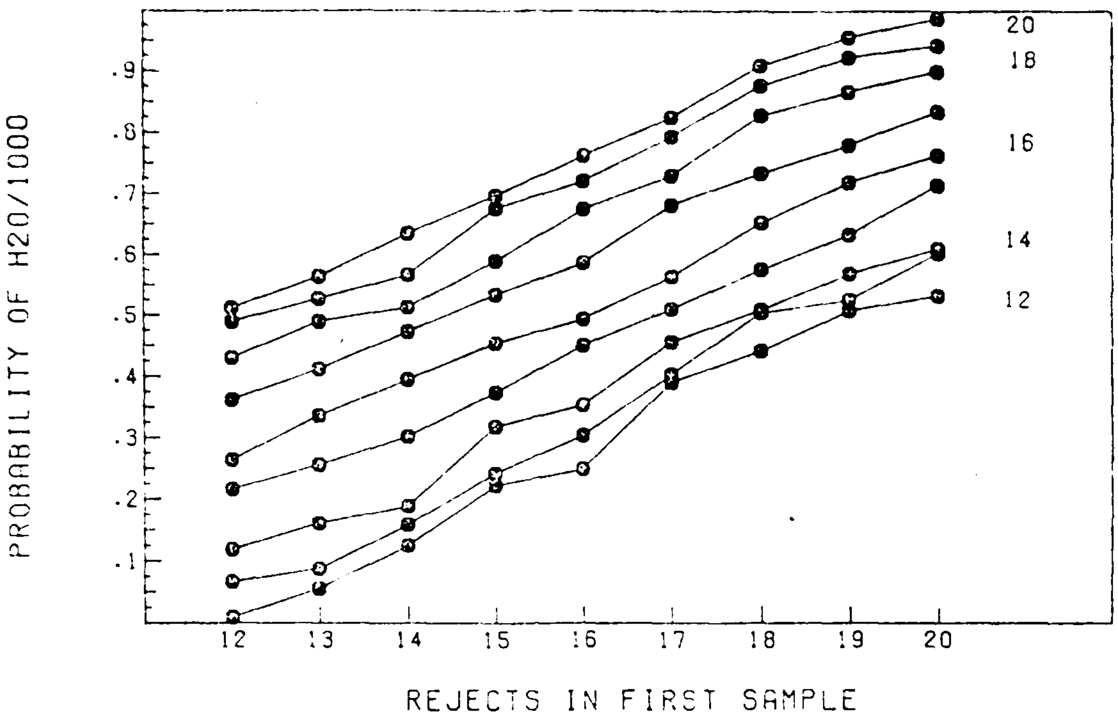
Oden's model can be paraphrased as follows: if we suppose that the subject is asked to rate the degree to which a given stimulus is a T or an F (i.e., the stimulus "T" would be rated 1 and the stimulus "F" would be rated 0), then,

$$\text{Relative T-ness of stimulus} = \frac{\dfrac{\text{T-ness of feature 1}}{\text{F-ness of feature 1}}}{\dfrac{\text{T-ness of feature 1}}{\text{F-ness of feature 1}} + \dfrac{\text{F-ness of feature 2}}{\text{T-ness of feature 2}}} \qquad (11)$$

A relative ratio rule of this sort is expected, for stimuli such as Oden used, to produce data that have a "barrel-shape" when they are graphed. As can be seen in the lower panels of Figure 3, Oden's data do, indeed, have such a barrel-shape for all three stimulus matrices and, thus, verify that people have access to ratio-like judgment processes.

The upper panel of Figure 4, however, shows the data pattern that is produced by naive judges in the Bayesian task. These particular data were taken from an experiment in which subjects had to judge the probability that a machine was in need of repair. The stimuli were pairs of estimates of the rate at which the machine produced rejected parts (Lopes, 1981a). As is evident graphically, these data are essentially parallel -- indicative of an averaging process -- rather than barrel-shaped as Bayes' theorem requires.

CONTROL SUBJECTS



TRAINED SUBJECTS

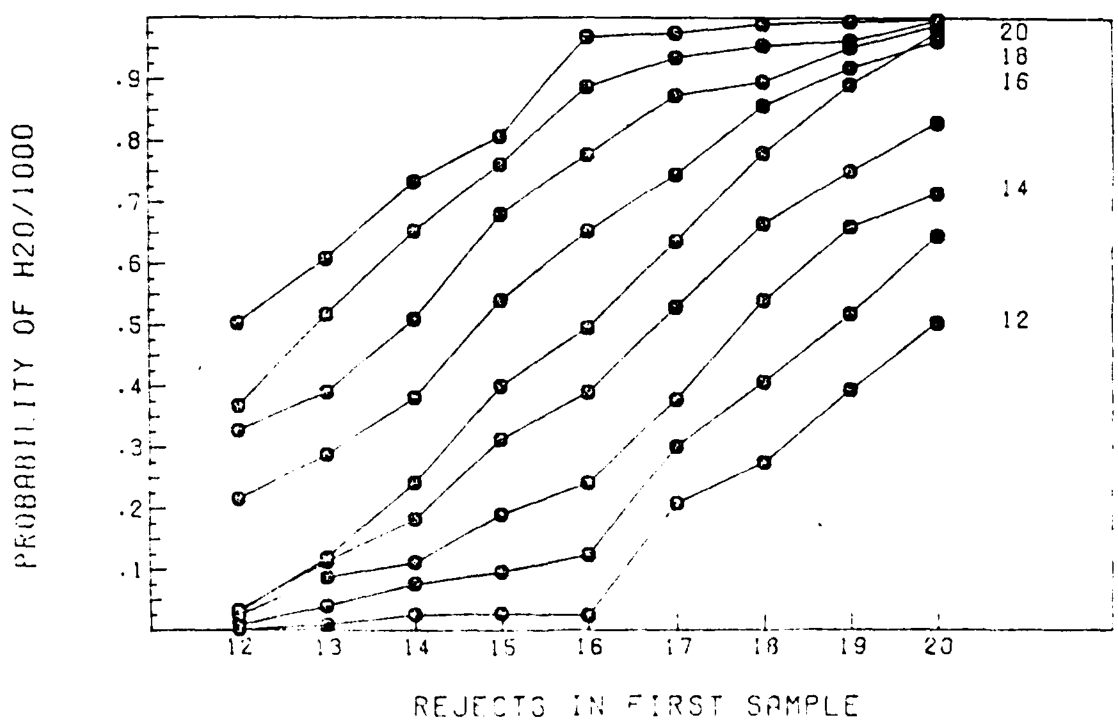


Figure 4. Examples of data produced by naive judges in Bayesian task (upper panel) and trained judges (lower panel). From Lopes, 1982a.

If people can produce relative ratios, why don't they do so in the Bayesian task? The answer seems to lie in the relation between the procedures underlying ratio-like responding and the subject's representation of the stimulus information and the response scale. There appear to be two fundamental differences between relative ratio procedures and averaging procedures. The first of these concerns the direction in which adjustments are made when both new and old information favor the same hypothesis. As was discussed above, adjustments in relative ratio processes must always be made in the direction of the hypothesis that is favored by the current information. In contrast, adjustments in averaging can be made either towards or away from the hypothesis favored by the current information depending on whether the current information is more or less extreme than the value being adjusted.

The major purpose of the Bayesian inference experiments described above was to find out whether naive subjects could be made to be better Bayesians by teaching them better adjustment strategies. In the first experiment, two groups of subjects were given sample information about the rate of rejected parts in a sequential-presentation format and were asked to judge the probability that the machine that produced the samples was broken. Half the subjects were given no further instructions; these were the control group. The other subjects were given instruction concerning the prevalence of directional adjustment errors and were taught the proper direction of adjustment for pairs in which a "weak" sample favoring a given hypothesis follows a "strong" sample favoring the same hypothesis. The results indicated that the instructions did, indeed, greatly reduce the occurrence of directional errors, but the data produced were still much more like averages than like relative ratios.

The largest discrepancies appeared to involve stimulus pairs in which one sample favored one hypothesis and the other sample favored the other hypothesis. For these pairs, subjects appeared to give the samples within serial position relatively equal weight regardless of whether they were extreme or not. In other words, the subjects were behaving as required by the parallelism principle described in the section on constant weighted averaging. Bayesian inference, on the other hand, requires a "differential weighting" strategy: Samples that are extreme in one direction or another are more diagnostic than nonextreme samples and must be accorded higher weight in the final judgment.

Experiment 2 (the results of which are shown in Figure 4) attempted to teach subjects a procedure that would facilitate differential weighting in a simultaneous-presentation task. Basically, this new procedure retained the instructions concerning the proper direction of adjustment and added only the idea that the subject should always anchor the judgment at the value given by the more extreme stimulus and then adjust for the value of the less extreme stimulus. The reasoning was that, since primacy is a commonly observed finding in judgment, processing stimuli in order of importance might be one way of producing differential weighting.

As can be seen in the lower panel of Figure 4, the training procedure was quite effective. In contrast to a group of control subjects (top panel), the trained subjects produced data that are clearly more barrel-shaped. In fact, these data were very close to the optimal values computed from Bayes' theorem.

It appears, then, that whether or not subjects produce ratio-like data or averaging-like data involves both their intuitions about the direction in which judgments should be made and their intuitions about the magnitude of the adjustment constants (or, equivalently, about the order in which stimulus features should be processed). To understand why their intuitions are appropriate in some tasks but not others, it is useful to compare Oden's task situation with the typical Bayesian task situation.

In Oden's task, the stimulus features seem to have a bipolar character. For example, presence of a moderately sized middle F-bar in a particular stimulus seems to be simultaneously evidence for F-ness and evidence against T-ness. Psychologically speaking, we might suppose that such features are on a bipolar scale than runs from positive to negative. In contrast, stimuli in Bayesian tasks often do not have this bipolar character. For example, suppose that the two hypotheses under test involve whether the rejection rate for a machine is 10 parts per 1000 (H10/1000) or 20 parts per 1000 (H20/1000). A sample with an estimated rate of 16 parts per thousand intuitively favors H20/1000 at least slightly. However, it does not seem particularly to deny the possibility of H10/1000. Thus, Bayesian stimuli seem to lie on a unipolar scale that runs from zero to positive.

Whether or not a subject evaluates stimuli on a bipolar or unipolar scale seems, intuitively, to be an important determinant of whether or not directional errors occur in adjustment. If, for example, a subject has coded a moderately-sized F-bar as "F-like and not T-like," it seems unlikely that the subject would ever adjust toward T-ness. On the other hand, if such a mid-range value were to be coded relative to a neutral point, as appears to be the case when subjects judge mid-range samples in the Bayesian task to be "a little bit on the H20/1000 side of neutral", adjustment might easily be toward the middle or neutral range of the response scale, as would be required for averaging.

The issue of order of processing in the two tasks also seems to be quite different. In Oden's task, one seems automatically to pick out the more diagnostic feature. Hence it would be natural to use it as an anchor and adjust for the weaker feature. For example, the stimulus

would be described by most people as a T with a blot on the vertical bar, suggesting that the full length left T-bar is more important or salient than the small middle F-bar. In judging the relative T-ness of the stimulus, they would consequently anchor on the value given by the T-bar and adjust for the value given by the F-bar. Bayesian stimuli, in contrast, are typically perceptually undifferentiated, as for example, in the stimulus pair below:

14 rejects
11 rejects

Although the second of the two samples is more diagnostic (given the hypotheses described above), the eye is not drawn to it and there is little reason to suppose that a naive subject would have any particular tendency to consider it more important or to process it first. Thus, Oden's task may tend to predispose people to process the stimulus components in an order that would facilitate the differential weighting necessary for a relative ratio rule. In contrast, the typical Bayesian task would tend to predispose people to process the stimuli in an order more amenable to constant weighted averaging.

Finally, a third factor that may differentiate between Oden's task and the typical Bayesian task may be that, for naive subjects, the response scale in Bayesian tasks is often confusable with an estimation scale. As is evident in the literature (Beach, Wise, & Barclay, 1970; Marks & Clarkson, 1972, 1973; Shanteau, 1970, 1972), judgments in Bayesian tasks are not only more often like averages than like relative ratios, they are also more like averages of estimates than they are like averages of inferences. For example, in Lopes (1981), it appeared that subjects in a serial-presentation version of the machine task simply treated the response scale as running linearly from 10 rejects (on the left) to 20 rejects (on the right) rather than from p(H20/1000) = 0 to p(H20/1000) = 1. In particular, their anchor judgments were proportional to the number of rejects in the anchor sample rather than being numerically at all like the normatively expected inferences.

Multiplying

There are several good examples of tasks in which naive subjects produce judgments that look as if they had been generated by multiplying. Prominent among these are the risky decision task in which subjects judge the worth of gambles (Anderson & Shanteau, 1970; Shanteau, 1974; Tversky, 1967) and the joint probability task in which subjects judge the likelihood of joint events (Beach & Peterson, 1966; Lopes, 1976; Shuford, 1959).

Figure 5 gives some data from a joint probability task in which experienced poker players were asked to judge the likelihood that they could beat pairs of opposing stud poker hands with a pair of sevens. The stimulus pairs were composed using an 11 x 3 factorial design in which the levels of the two factors were poker hands, each described by four up-cards, a bet amount, and the playing style of the opponent. For present purposes there is no need to get into the particular coding used to describe the hands. Suffice it to say that the hands listed along the abscissa from left to right have decreasing probability of beating a pair of sevens as do the row hands listed from bottom to top.

The experiment for which data are shown contrasted two different response modes, a rating response in the top panel and a betting response in the bottom panel. As can be seen, the data from both conditions show the characteristic "fan shape" of a multiplicative process (cf. Anderson, 1981, Section 1.4). This result suggests that poker players consider the likelihood of beating a pair of opposing hands to be proportional to the product of the likelihoods of beating the single hands.

The question of interest for present purposes is how these data were generated. Although poker players are generally more familiar with probabilistic notions than nongamblers, there did not appear to be any indication that they were assigning probabilities to the hands and multiplying them. Instead, they appeared to be using an analog serial fractionation strategy that is mathematically equivalent to multiplying (Lopes, 1976; Lopes & Ekberg, 1980).

In serial fractionation, the judge is assumed to anchor the judgment at the

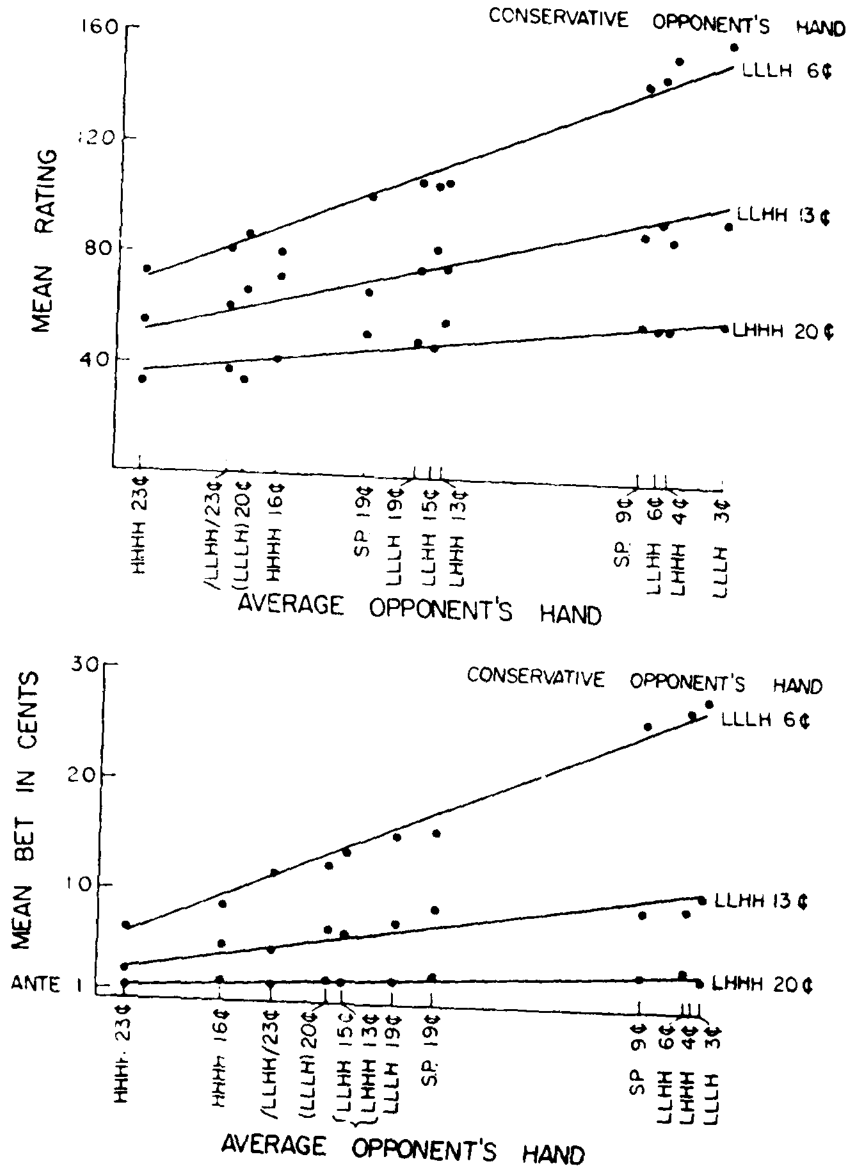


Figure 5. Example of multiplicative data from a pair of joint probability tasks. From Lopes, 1976.

value of one stimulus and then adjust this value downward (i.e., toward zero on the scale) directly in proportion to the value of another stimulus. If there are more than two stimuli to be integrated, this proportional fractionation operation is repeated until all stimuli have been processed. Thus, serial fractionation is simply a special case of anchoring and adjustment in which the adjustments are always downward and always in proportion to the value of the stimulus being integrated.

As it happened, subjects' use of the serial fractionation strategy was directly observable in the bet condition of the poker experiment described above. The experiment was run manually without computer assistance. In the task the subjects were required to bet between 1 and 30 cents that they could beat each pair of opposing hands. The hands were printed separately on 4 x 6 cards and were displayed simultaneously on wooden stands that were located side by side in front of the subject. Between the subject and the stands was a long strip of paper divided into 30 segments and numbered from 1 on the left to 30 on the right. The subject was to indicate his bet by sliding a response marker (i.e., a poker chip) into the segment labeled with his intended bet amount.

A common response pattern for subjects was the following. First, they looked at one of the two stimulus hands. (Since these were printed on fairly large cards, it was possible to know simply from head movements where subjects were looking.) Then they located a position on the betting scale for this first hand using the response marker. Next they looked at the second stimulus hand. And last they moved the response marker downward from its initial position to a final position.

Note that this response pattern appears to be an overt instantiation of serial fractionation. If it were being done on a probability rating scale, the result would be a normatively appropriate response computed without either numerical processes or the necessity of prior knowledge of the normative rule for joint probability. On a bet scale, the response is nonoptimal in the sense that it constitutes a form of matching rather than maximizing. However, the response may be sensible from other perspectives as, e.g., in reducing the variance of outcomes.

Given subjects' clear use of the response scale in generating their bets, a natural question arises concerning whether the process can occur in tasks for which there is no scale to be fractionated physically. There is evidence that the answer is yes. Although most of the studies that have obtained multiplicative data have used physical rating scales, there are some that have not. For example, in another version of the poker task under discussion, subjects typed their bet amounts into a computer console numerically and still produced multiplicative data (Lopes, 1976, Experiment 3). Similarly, Slovic and Lichtenstein (1968, reported in Shanteau, 1975) got multiplicative data using a verbal bidding response.

Joint probability tasks such as the poker task are a special case for serial fractionation in the sense that either stimulus can be used as the anchor. The risky decision task differs from this in that the two parts of a gamble (e.g., 1/6 to win $13) lie on psychologically different scales: the prize or amount to be won or lost lies on the value scale whereas the probability information lies on a probability scale. This differentiation in the informational content of the stimulus items has important consequences for the order in which the items are likely to be processed.

As Lichtenstein and Slovic (1971) noted some time ago, subjects in risky decision tasks tend to use the amount (or prize) to be won as an anchor and then adjust this to account for the probability of winning. This value/probability ordering makes good intuitive sense for serial fractionation since the task is to integrate value and probability information in such a way that the final response lies on the same value dimension as the amount to be won.

For illustration, consider the data in the upper panel of Figure 6. In this experiment Shanteau (1974) asked subjects to rate the worth of verbal bets described by phrases like "somewhat unlikely to win sandals." He used a continuous response scale that ran from "worthless" at the left to "sure thing to win television" at the right. Now consider how a subject might operate on this scale to judge the worth of, e.g., "fairly likely to win watch." The most obvious procedure would seem to be to use "watch" as the anchor and locate its value relative to the end-marker, i.e., the television. This operation would yield a quantity on the value scale that could then be fractionated proportionately in accord with "fairly likely to win." The result would be another quantity that is still on the value scale. More concretely, if we assume that, e.g., for a given subject a watch is about 40% as valuable as a television and "fairly likely to win" is about 80% of certainty, then the serial fractionation process would take the form schematized in the lower panel of Figure 6.

It is important to note that this order of processing is not a mathematical necessity. One could fractionate the response scale first proportionally to the probability information and then proportionally to the relative worth of the amount to be won. However, such a probability/value ordering seems unnatural at best. Not only would such an ordering contradict Lichtenstein and Slovic's (1971)

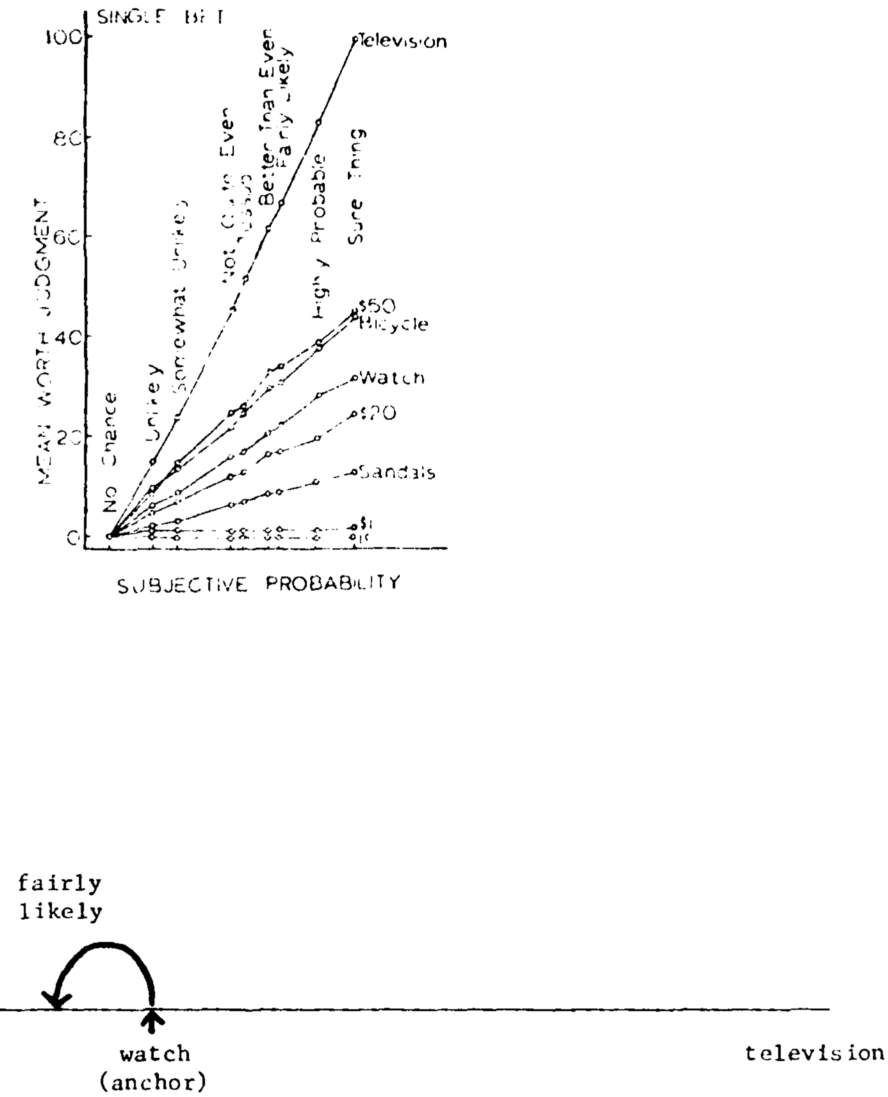


Figure 6. Example of multiplicative data from a risky decision task with illustration of the anchoring and adjustment mechanism. From Shanteau, 1974.

observation that subjects anchor on the amount to be won in judgment tasks, but it would contradict, as well, the finding of Lopes and Ekberg (1980) that value information appears to be processed before probability information when subjects have to choose between gambles and sure things.

Serial fractionation tasks are special in the sense that all adjustments in the process should be downward, toward the zero point on the scale. There are, however, other potentially multiplicative tasks that involve more general scalar-like processes in which adjustments are not constrained to be downward. For example, Anderson and Butzin (1974) gave subjects information about the motivation and ability of athletes and asked them to predict their performance in a college track tryout. They found a multiplicative relationship between motivation and ability. Graesser and Anderson (1974) found a similar relationship between income and generosity when subjects were asked to predict how large a gift a person would give to a worthy cause.

In scalar tasks such as these it is possible that subjects anchor their judgment on an estimate given by one of the quantities and then adjust this according to the other quantity. For example, it would seem reasonable to anchor at the value given by ability in Anderson and Butzin's task or at the value given by income in Graesser and Anderson's task since these quantities lie on the scale of judgment. Adjustment could then be made for motivation and generosity, respectively. There are, however, some difficulties in imagining how scalar multiplication would work for upward adjustments. One possibility would be to fractionate out a portion of the anchor quantity and then add this on so that the final quantity would be, e.g., one-and-a-half times the original. In cases, however, where the incrementing is by whole integers, e.g., increasing a quantity by three, it is not particularly intuitive to me that this is done by successive additions of the anchor quantity.

A question that has arisen especially with respect to multiplicative models is whether human judgment processes display the sorts of consistency and reversibility that characterize formal algebra. In particular, there has been considerable interest in determining whether subjects have dividing-like processes that they can apply when asked for inverse judgments. For example, if

$$\text{Performance} = \text{Motivation} \times \text{Ability}, \tag{12}$$

mathematical algebra would require inferences about ability given motivation and performance to follow the rule

$$\text{Ability} = \text{Performance/Motivation}. \quad (13)$$

Do people's intuitions about the algebra of real world situations agree with this rule? Put in terms of processes the question becomes not only whether subjects have a procedure that is equivalent to analog division, but also whether they have a complete, appropriate, and reversible mapping between the elements and the operators in their judgment system (i.e., values, adjustments, etc.) and the objects and operators of real world task domains (i.e., people giving money to charity or trying out for track meets).

Anderson and Butzin (1974), Graesser and Anderson (1974) and Surber (1980) have all investigated this question. The general conclusion seems to be that natural judgment processes do not display the sorts of consistency and reversibility that characterize formal mathematics. For example, in the Graesser and Anderson experiment, subjects appeared to combine income and generosity multiplicatively to produce judgments of expected gift size. When judging income from gift size and generosity, however, or generosity from gift size and income, they appeared to follow some sort of subtracting rule. Thus, the human judgment system appears to comprise a set of judgmental heuristics that often correspond to simple arithmetical operations but that tend to be both nonreversible and logically unrelated.

Adjustment Procedures and Judgmental Phenomena

Research on algebraic judgment models has two different aspects. One is the investigation and description of classes of integration rules, such as the averaging, relative ratio, and multiplying rules just discussed. The other is the specification of the conditions under which certain more general judgmental phenomena occur. Some of these phenomena represent classic topics in judgment, such as primacy/recency, contrast, assimilation, and halo effects. Other phenomena, however, have come to light only as a result of the power afforded by the algebraic analysis. These more recently described phenomena include initial impression effects, differential weighting, subadditivity, and nonadditivity.

In this section I will show how four of these phenomena (primacy/recency, initial impression effects, differential weighting, and nonadditivity) can be understood in terms of the procedural model. Since most of the work on these phenomena has been in the context of averaging, I will not consider the other models further.

Primacy/Recency

One of the earliest findings in impression formation research was a tendency toward primacy: The earlier the position of an adjective in a trait list, the greater its effect on the judgment. Algebraically, primacy effects consist of a tendency for the weight of the adjectives to be a monotonically decreasing function of serial position.

Several different theoretical interpretations have been offered for primacy including inconsistency discounting, attention decrement, and a generalized form of assimilation-to-expectancy termed directed impression (see Anderson, 1981, section 3.3 for a complete historical summary of this literature). These interpretations were tested in a variety of ways, with the end result being that most studies favored the attention decrement hypothesis.

One of the experimental manipulations that proved to be important conceptually involved modifying the impression formation task in a way that forced subjects to attend to information later in the list. Two effective procedures were to present stimuli serially and have subjects either read all the stimulus items aloud (Anderson, 1968) or make a cumulative response after each stimulus item (Stewart, 1965). Under these conditions, primacy was replaced by recency, thus confirming the attention decrement hypothesis.

Two procedural questions of considerable interest are (a) what processes mediate attention decrement under conditions of ordinary simultaneous presentation, and (b) what operations produce recency in the case of cumulative responding and other attention-demanding conditions? For the primacy results, the best answer seems to be one advanced by Anderson (1981, p. 191), namely, that there is an increasing tendency during judgment for the impression being constructed to crystallize and become resistant to change.

In terms of adjustment processes, the crystallization hypothesis suggests that as a direct function of the cumulative amount of information already processed, the serial adjustment constants become systematically smaller than the 1/n required for arithmetic averaging. This seems perfectly reasonable. In addition, the procedural theory schematized in Figure 1 suggests another source of primacy, namely, that subjects might partially or totally ignore later items in a list if they seem relatively unimportant. One way this could work would be for subjects having a moderately well-crystallized impression merely to scan later items. If the scan does not reveal any item or items that would be likely to produce a large change in the impression, the subject might then choose not to adjust the impression further. Such a process would not only be computationally efficient, it would also capture procedurally the important insight that averages generally become more and more resistant to change as more and more information is included in them.

For recency effects, the situation is somewhat different. To see how, consider once again the number averaging task. Suppose that a subject is given a list of 20 two-digit integers one at a time and asked to give a running average. As noted previously, the adjustment constants in such a task would, ideally, begin with 1/2 and then decrease to 1/3, 1/4, etc. out to the final value of 1/20. A few minutes play with such a running average should convince the reader that the adjustments in such a task typically become very small very quickly.

Now consider the position of a subject in such a task who is required to respond after each stimulus. Sooner or later the subject will reach a psychological "grain size" for adjustments beyond which he will not go; in other words, at some point the subject will begin producing adjustments of relatively constant size that track the direction in which the current adjustment should be made but that greatly over-adjust in terms of magnitude. This, by itself, would produce recency. Indeed, it is difficult to see how a cumulative response requirement could fail to produce recency in tasks involving more than a few stimuli. For the condition in which stimuli are merely pronounced by the subject but not tracked overtly, the "grain size" problem might not hold, in which case recency might reflect some other sort of attentional effect. However, it is at least plausible that the pronouncing condition does cause subjects to produce a cognitively discriminable internal response for each stimulus item, in which case the final result would be essentially identical to that for overt cumulative responding.

Initial Impression Effects

One of the early controversies in the history of impression formation research concerned whether personality impressions follow an adding or an averaging rule. Although there were several impressive results arguing for averaging, some researchers pointed to a common set-size effect as proof of adding (see Anderson, 1981, Section 2, for a summary).

Basically, the set-size effect was that when subjects were given $\underline{n}$ adjectives of identical value and were asked to rate their impressions, the impressions grew more extreme as a function of $\underline{n}$. For example, if all the adjectives were positive, the response would be higher for three adjectives than for two. Obviously if subjects were simply averaging the scale values, the result would not differ for different numbers of adjectives.

As it turned out, the averaging model was able to handle such set-size effects with the assumption that the subject begins each judgment with a neutral initial impression that receives some small, but not zero, weight in the averaging process. As successive adjectives are then integrated, the cumulative weight of the adjectives increases relative to the weight of the initial impression with the result that the response becomes more extreme.

A considerable literature now exists supporting the initial impression construct. However, there are some tasks in which there are no initial impression effects (e.g., number averaging, psychophysical averaging, most tasks in Bayesian inference), so the question may be raised as to when initial impression effects are to be expected and when they are not.

The answer may reflect the subject's understanding of the judgment dimension. In the case of impression formation, for example, the subject is asked to rate the overall likeableness of a person given specific information about traits. It seems self-evident that subjects would recognize that no single trait conveys all the relevant information concerning a person's likeableness. Thus, the tendency for people to anchor at a value somewhat less extreme than the scale value of the anchor stimulus may represent a procedural means for expressing the uncertainty surrounding all the important dimensions not yet described. For tasks such as number averaging, however, the judgment dimension is unidimensional so that there is no apparent need to "hedge" in the anchoring process: the anchor is simply set at the value of the first stimulus processed.

A question of some interest for the procedural model is whether initial impression effects are mediated by pre-judgmental implicit anchoring at the initial impression with subsequent adjustment toward the first-processed stimulus item or whether they are mediated by explicit anchoring at the scale value of the first-processed stimulus item with subsequent adjustment toward neutral. (Note that the latter process would be conceptually similar to "discounting" the anchor stimulus.)

One piece of evidence that bears on this question comes from a controversial experiment (Shanteau, 1975) that tested the averaging hypothesis for inference judgment. In Bayesian inference, nondiagnostic (i.e., neutral) information should have no weight whatsoever in the judgment process. Shanteau used this fact in a clever experiment to get a qualitative test of his averaging model for inference. He presented subjects with various sequences of sample information and asked them for cumulative inferences. He found that subjects tended to adjust their inferences toward neutral after the presentation of nondiagnostic samples. This would be disallowed, of course, by the Bayesian rule, but it is exactly what one would expect of an averaging rule.

Wallsten (1976a) objected to Shanteau's analysis by pointing out that the averaging model only accounted for the data when the nondiagnostic information followed one or more pieces of diagnostic information. When the nondiagnostic sample was given first, it appeared to have no weight at all in the judgment. Specifically, the cumulative response to a two-item series in which a nondiagnostic sample preceded a sample of given diagnosticity was essentially identical to the response to a one-item series consisting only of the diagnostic sample. This result is in accord with Bayes' theorem and not with an averaging rule.

These data suggest a tendency for people to resist anchoring at a neutral value even when the potential anchor stimulus is an explicitly presented stimulus item. If this is so, it would suggest that the explicit procedure for initial impression effects would be more plausible than the implicit procedure since the latter would require subjects to anchor not only at a neutral value, but at an implicit neutral value to boot. Obviously, the case for explicit anchoring is not strong, especially as it rests on generalizing a result from a task in which initial impression effects do not ordinarily occur[1] to a task in which they are quite common. It seems plausible, in fact, that nondiagnosticity (as exists in the inference domain) might be treated quite differently than neutrality (as exists in the trait adjective domain). Since nondiagnosticity is literally content-less relative to the issue at hand (i.e., which hypothesis is supported), it may be relatively easily ignored when it occurs in the first position. In contrast, neutrality (i.e., that people tend to be averagely likeable) has content relative to likeableness. Hence, we may be willing to anchor on it both when it is presented explicitly and when it merely exists as an initial impression.

Differential weighting

As stated above, the most common form of averaging model is constant weighted averaging. Ordinarily, averaging models are expressed with reference to designs in which the various factors represent the serial positions in which stimuli appear in a temporal series or in a printed list. The equation describing the constant weighted averaging model for a factorial design would be written,

$$R_N = \frac{\sum_{i=1}^{N} w_i s_{ij}}{\sum_{i=1}^{N} w_i} \qquad (14)$$

This shows that the factor weights, w_i, vary only with the factor, i, whereas the scale values, s_{ij}, vary with both the factor and the level, j, within factor.

There are several theoretical and practical factors that make it desirable, when possible, to set up experiments so that constant weighting will hold. However, the averaging rule can be written more generally,

$$R_N = \frac{\sum_{i=1}^{N} w_{ij} s_{ij}}{\sum_{i=1}^{N} w_{ij}} \qquad (15)$$

to allow weights to vary not only between factors, but within factors as well. In this form the averaging rule would be able to capture virtually any psychological factor that might affect the weight of a stimulus item, e.g., reliability, source credibility, novelty, extremity, and so forth.

A form of differential weighting that has appeared fairly frequently in the literature is what has been termed a "negativity effect" (Anderson, 1981, Section 4.4.2). The observation is that stimulus weights seem to increase as a function of their negativity. In a factorial plot this distorts parallelism by causing a convergence of the data points for stimuli containing negative items.

For illustration, consider a stimulus set generated by factorial combination of three sets of adjectives:

Factor 1	Factor 2	Factor 3
careful	intelligent	friendly
careless	gullible	cruel

The stimulus combinations that would result would be,

intelligent careful friendly	gullible careful friendly	intelligent careless friendly	gullible careless friendly
intelligent careful cruel	gullible careful cruel	intelligent careless cruel	gullible careless cruel

For the stimuli in the upper row, there are clear differences in the scale values of the adjectives. There may also be differences in importance or salience.

For example, intelligence and friendliness might be considered more important than carefulness in judging the likeableness of a person. However, these particular differences in weight would probably not be great enough to prevent one from simply integrating the adjectives in order as they appear in the stimulus, i.e., using the default order. For the stimuli in the bottom row, however, the situation is different. The adjective "cruel" is both extremely negative and highly salient. Thus, a common pattern would probably be for subjects to judge these stimuli by anchoring on "cruel" and adjusting for the other two adjectives.

Now consider the effect that such an ordering difference would have in the analysis of the factorial design. Ordinarily, the weights of the various adjectives would reflect their serial position in the stimulus so that $w_1 > w_2 > w_3$. In the present case, however, the adjective "cruel," which by virtue of its serial position has weight w_3, is pushed to the top of the processing order where it receives weight w_1. In contrast, its factor-mate, "friendly," is processed in the default order so that it always receives weight w_3. Thus, a negativity effect would occur in which, for example, there would be less difference between the responses for careful/intelligent/cruel and careless/intelligent/cruel than for careful/intelligent/friendly and careless/intelligent/friendly. (Note that ordering effects need not be the only source of increased weighting for salient adjectives. The possibility of direct weighting effects will be discussed below.)

The algebraic differences between Equation 14 and Equation 15 are very small, amounting to no more than a restriction on whether the stimuli within an experimental design factor can vary on weight or not. Are the conceptual differences equally small? In terms of the serial adjustment model, the answer is yes. Subjects do not "have" a constant weighted averaging process and a differential weighted averaging process that they apply selectively to different tasks. Instead, they have a general judgment procedure that functionally produces averages. Whether those averages show constant weighting or differential weighting is not primarily the result of what the subject does, but rather reflects how the experimenter constructs and analyzes the stimulus design.

This point can be illustrated by consideration of an experiment by Anderson and Lopes (1974) in which subjects made judgments of the occupational proficiency of various people (e.g., lawyer and plumber) based on information about their degree of reliability (average vs. extreme) and their degrees (very, moderately, or not-very) of competence on skill factors (e.g., persuasiveness and mechanical ability). In this experiment it was anticipated that the relative weight of the

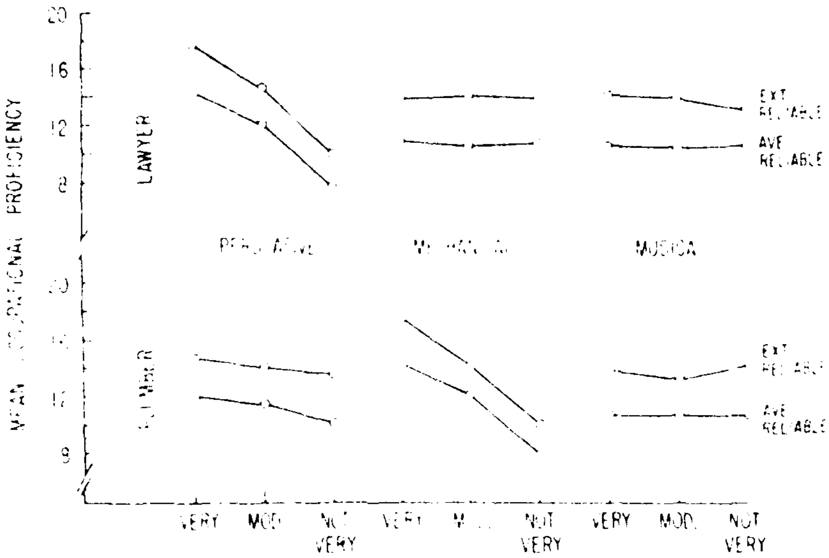


Figure 7. Example of differential weighting in averaging. From Anderson and Lopes, 1974.

competence factors would reflect the importance of those factors to the profession being judged.

As can be seen in Figure 7, this is basically what happened. Taking the case of persuasiveness, for example, we can see that the slopes of the two curves for reliability are very steep and very close together when reliability and persuasiveness are combined for lawyers. When they are combined for plumbers, however, the curves are further apart and virtually flat. Both these results reflect the increased weight assigned to persuasiveness for lawyers. Likewise, the importance of mechanical ability for plumbers is indicated by the steepness and closeness of the curves for reliability when reliability and mechanical ability are combined.

How are these effects produced? According to the serial adjustment model the effects would, at least in part, reflect processing order. (A further source of weighting effects will be discussed below.) For example, if information is given about a lawyer's persuasiveness and reliability, the judgment would probably be anchored on the persuasiveness value and adjusted for reliability. Given primacy, we would, therefore, expect the weight for persuasiveness to be greater than the weight for reliability. For plumbers, however, the anchoring would be on reliability with adjustment (if any occurred at all) for persuasiveness. The result would be greater weight for reliability.

The judgments in this experiment were expected to conform to a constant weighted averaging rule and, with the exception of a small negativity effect that will be discussed below, they did so. But consider how the situation would have differed if the lawyer/plumber designs had been set up as follows:

Reliability Factor	Skill Factor
extremely reliable	very persuasive
averagely reliable	very mechanical

From the subjects' point of view, the stimuli would be identical to some of those judged in the previous design. Presumably they would process them in the same way. But from the design point of view, we would say that the subjects were performing differentially weighted averaging since, for a given profession, the weights on the two levels of the skill factor would be greatly different. Thus, psychologically speaking, there is only one averaging process. It reveals itself in different guises, however, depending on the stimulus design.

A final point should be made about weighting effects that are mediated directly rather than by order of processing. Consider, again, the data in

Figure 7. Note that for the four panels in which there is a job-relevant adjective (e.g., persuasiveness for lawyers, mechanical ability for plumbers, etc.) there is a tendency for the curves to converge at the low value of the adjective. This is, of course, a negativity effect as was described above. But we might wonder why it occurs in these particular designs since, presumably, the skill information is being processed first for all three levels due to its relevance. One answer might be that subjects have, in addition to ordering procedures, other procedures that modify the magnitude of the adjustment constant depending on the relative importance of the stimulus being integrated. If this is so, then the adjustments from "not very persuasive" might involve a smaller adjustment constant than, say, adjustments from "very persuasive." Such direct weighting procedures seem feasible, in general, and they would be necessary practically for proper weighting of important information that was received late in a temporal series. They would also be necessary for proper weighting of unimportant information occurring in short lists where late ordering alone would not be able to give them small enough weight.

Nonadditivity

Constant weighted averaging rules and relative ratio rules both display the mathematical property of ordinal independence. This refers to the fact that when these rules are used to generate data for factorial combinations of stimulus values, the rank order of data within rows (and columns) is the same for all rows (and columns). Multiplying rules can also display ordinal independence when they are restricted to stimulus values which are all positive, as would be the case for the kinds of multiplying models discussed in this paper if we disallowed stimuli of zero value. Ordinal independence does not, however, hold in general for differentially weighted averaging. Although particular data sets may display the property, other data sets may show one or more disordinal interactions (i.e., crossover interactions).

Ordinal independence is important for present purposes because it provides a way to test for the presence of averaging processes without relying on the quantitative properties of the responses. In other words, if violations of ordinal independence can be found within a data set, then it is possible to conclude positively (within the limits of uncertainty imposed by unreliability in the data) that the rule that produced the data is an averaging process and not some other additive rule operating with a response transformation that mimics averaging results.

Tests of additivity have an important place in testing whether people in Bayesian inference tasks are relying on processes that are more averaging-like or more ratio-like. Two experiments were run in the present project (Lopes, 1982b) that attempted to manipulate weighting factors in such a way that differential weighting would be obtained in the machine maintenance task that has been discussed before. The major change in the new task was to designate samples not only in terms of the number of rejected parts that they estimated, but also in terms of the reliability of the estimate.

The results of the experiments were somewhat equivocal. At the group level, there were no violations of additivity. At the single subject level, however, there were numerous violations, suggesting that the judgment process was ordinally distinct from a relative ratio rule. Unfortunately, the forms of nonadditivity that were obtained included not only the particular type of crossover interaction that would be expected in differentially weighted averaging, but two other types of interaction as well. Of these unexpected crossover types, one was fairly easily accounted for in terms of the direction in which adjustments were made. The other one, however, did not lend itself to ready explanation, thus leaving open the possibility that the process might be more complex than heretofore suspected.

Since nonadditivity is an important topic, it is worthwhile showing how it occurs in a serial adjustment process. For this purpose, I will use data from an experiment testing an averaging model for similarity judgment (Lopes & Oden, 1980). In the experiment, subjects were given pairs of kinship terms such as mother/aunt, grandfather/girl-cousin, brother/sister, and so forth, and were asked to rate the similarity between the two terms. The model being tested was that similarity would vary inversely with a weighted average of the stimulus differences on each of a set of *i* relevant dimensions or attributes:

$$\text{Similarity}(a,b) = \alpha - \frac{\sum_{i=1}^{N} w_{iab} d_{iab}}{\sum_{i=1}^{N} w_{iab}} \qquad (16)$$

In this equation, w_{iab} are the weights, d_{iab} are the differences, and α is a scaling constant. The dimensions that were actually used in the model analysis were sex, age, and a lineality dimension termed "immediacy." How immediacy was decided upon need not concern us here. Suffice it to say that immediacy is an index of closeness of kin as measured from the viewpoint of ego, i.e., the judge in the task. It runs, from closest kin to most distant kin, as follows: parent, sibling, grandparent, auntuncle, great auntuncle, and cousin.

In the experiment, there were several dramatic crossover interactions that appeared in both the single subject data and the group data. Two of these are shown in Figure 8. Taking the crossover on the left as an example, notice that uncles and girl-cousins (U/GC) are considered to be more similar than fathers and girl cousins (F/GC). This makes intuitive sense: Although both pairs have the same difference on the sex dimension and on the age dimension, uncles and girl-cousins are closer on the immediacy dimension than fathers and girl-cousins. A reversal, however, occurs for greatuncles and girl-cousins versus grandfathers and girl-cousins. This is apparent when one considers that for these pairs also the sex differences and age differences are the same, and only the immediacy difference varies. In this case, however, the pair with the larger immediacy difference, grandfather and girl-cousin (GF/GC), is seen as more similar than the pair with the smaller immediacy difference, greatuncle and girl-cousin (GU/GC).

How do unintuitive patterns such as this occur? It is useful to think about the process in terms of the serial adjustment model. Table 1 gives some reasonable weights and scale values that might be applicable in the present case. Note that the weights tend to decrease as a function of the magnitude (or salience) of the difference. (In the full analysis of the experiment, the weighting function was actually curvilinear with large weights being assigned both to large differences and to zero differences as when both terms had the same sex or the same age. Since in the present example there are no identities, this complication can be safely ignored.)

Table 2 shows how these weights and scale values would operate in a serial adjustment process of the sort described algebraically by Equation 4. If one reads across the columns and then down the rows, the sequential anchoring and adjustment operations are revealed. For example, for the stimulus pair GU/GC, the largest dimensional difference is for sex, with a scale value of 95. Thus, the anchor is set to 95. This is adjusted next by the value for age, which is the next largest difference (i.e., 90). The adjustment process proceeds by noting the discrepancy between the old value (95) and the new value (90) and adjusting downward about half-way due to the relatively equal saliences of the two items. The final adjustment is for the very small immediacy difference (10). Although the discrepancy between the old value (92.52) and the new value (10) is quite large

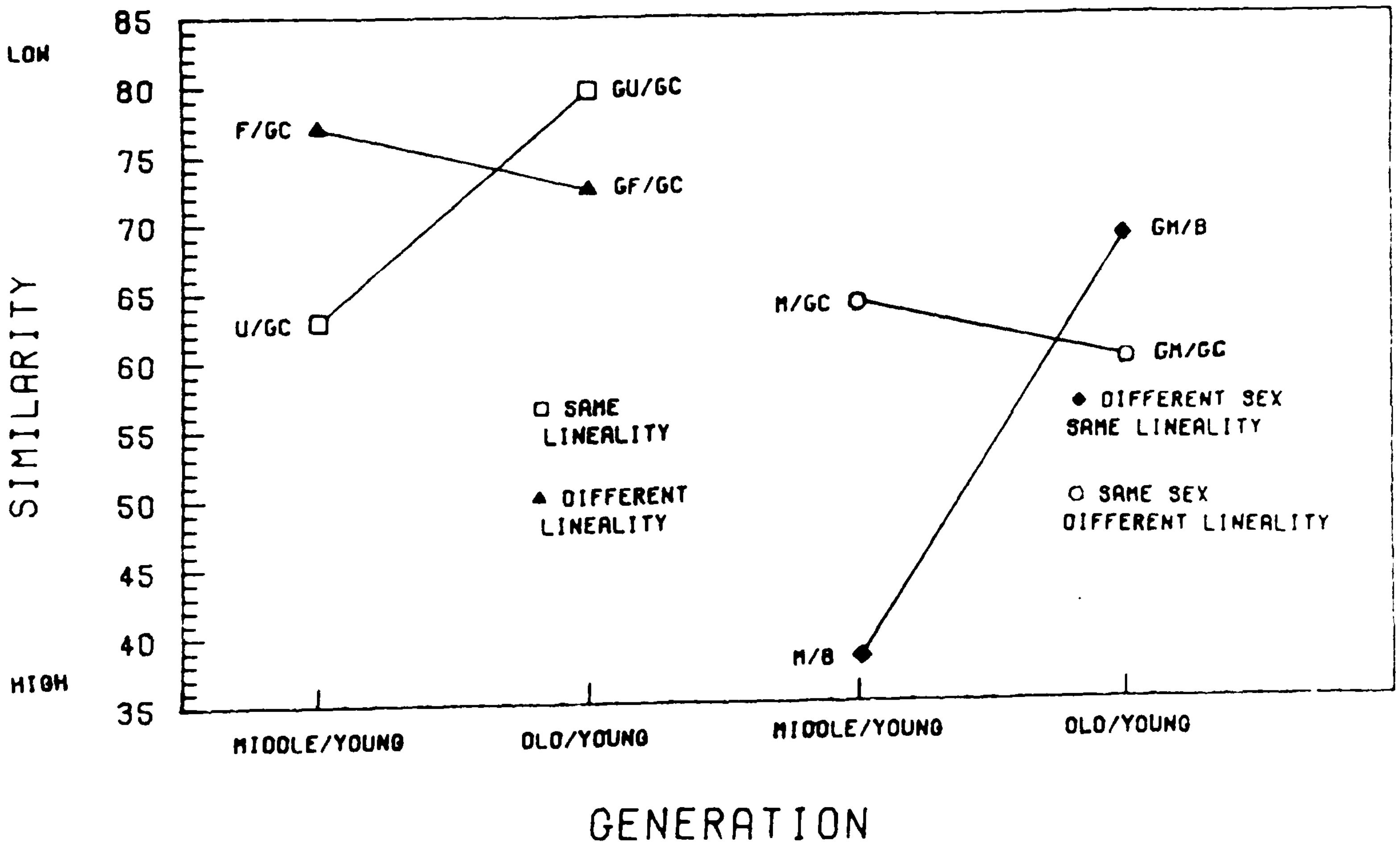


Figure 8. Examples of disordinal interactions (nonadditivities) in a similarity judgment task. From Lopes and Oden, 1980.

Table 1

Hypothetical Weights and Scale Values

for Crossover Interaction

DIMENSION AND DIFFERENCE	SCALE VALUE OF DIFFERENCE	WEIGHT OF DIFFERENCE
SEX		
male vs. female	95	51.0
AGE		
young vs. old	90	50.0
young vs. middle	40	7.5
IMMEDIACY		
cousin vs. parent	90	50.0
cousin vs. grandparent	70	40.0
cousin vs. auntuncle	30	5.0
cousin vs. great auntuncle	10	1.0

Table 2

Steps in Serial Adjustment Model Producing Crossover Interaction

STIMULUS PAIR	JUDGMENT OPERATION	DIMENSION INFORMATION	VALUE OF DIFFERENCE	OLD INTERIM VALUE	DISCREPANCY MAGNITUDE	ADJUSTMENT MAGNITUDE	NEW INTERIM VALUE
GU/GC	anchor	sex	95	-----	-----	-----	95.00
	adjust	age	90	95.00	-5.00	-2.48	92.52
	adjust	immediacy	10	92.52	-82.52	-.81	91.71
							** Final **
GF/GC	anchor	sex	95	-----	-----	-----	95.00
	adjust	age	90	95.00	-5.00	-2.48	92.52
	adjust	immediacy	70	92.52	-22.52	-6.39	86.13
							** Final **
U/GC	anchor	sex	95	-----	-----	-----	95.00
	adjust	age	40	95.00	-55.00	-7.05	87.95
	adjust	immediacy	30	87.95	-57.95	-4.56	83.39
							** Final **
F/GC	anchor	sex	95	-----	-----	-----	95.00
	adjust	immediacy	90	95.00	-5.00	-2.48	92.52
	adjust	age	40	92.52	-52.52	-3.63	88.89
							** Final **

(-82.52), the adjustment is very small since this small difference in immediacy is not very salient compared to the other two differences.

Similar processes occur for the other three values. In every case the information is processed in order of the magnitude of the dimensional difference while the magnitude of the adjustment constant (i.e., the weights in Table 1) reflect the salience of that difference. The result is a set of similarity judgments that violate ordinal independence, thus testifying to the differential weighting in the process.

Note that in this example, there would have been no mathematical reason to require the order of processing to proceed according to salience. The non-additivity comes directly from the differential weighting and, hence, would appear for any serial instantiation of Equation 4. However, to the degree that people tend to process information in an order that runs from most important to least important when they can (i.e., taking first things first), primacy effects would tend to enhance whatever differential weighting was already present and, hence, would be expected to encourage nonadditivity.

Epilogue

In this paper, I have attempted to integrate some hypotheses about the procedures that people use during judgment with the vast literature already extant on the algebraic properties of judgment data. I have not supposed that these procedural hypotheses in any way replace or deny the earlier algebraic work since that work provided the empirical foundation that made it reasonable to look for the specific psychological operations for computing various sorts of algebraic results. Instead, I have seen this work as complementing the algebraic work by suggesting the fine structure of the processes described by the algebra.

All of the research performed on the project for which this is the final report concerned the Bayesian inference task. Although no one would doubt that inference is an important psychological skill, many might judge that the Bayesian task was played out psychologically long years ago when so much research was generated on the topic of conservatism. I would hope, however, that the present research has contributed some new information about the procedures that mediate the appearance of normatively inappropriate averaging-like processes in Bayesian tasks.

The three separate experimental approaches taken to the topic in this research have each contributed differently to the project. The first two experiments were primarily aimed at (and succeeded in) confirming the existence of the kinds of directional errors of adjustment that would be expected from an averaging process (Lopes, 1981). The second two experiments provided a practical means to test the procedural analysis by seeing whether it could be used to engineer "better Bayesians" out of naive subjects (Lopes, 1982a). The result was a clear yes. The last two experiments attempted, but did not entirely succeed, in demonstrating the qualitative necessity of an averaging formulation for describing the judgment process (Lopes, 1982b). Although there were data from some subjects to support the hypothesized nonadditivity, data from other subjects suggested that the process may be more complicated than supposed, particularly when stimulus samples differ in reliability.

Of these three subprojects, the one that pleased me the most was the one on building better Bayesians -- the debiasing project. Perhaps this was merely because the project succeeded in doing what most other approaches to debiasing inference have failed at. Everyone loves success. But I think the deeper reason is that the debiasing effort was accomplished by decomposing the judgment situation in psychological terms and then communicating to subjects in those same psychological terms what errors they might produce in their judgments and what procedures they might use to avoid error. It was reinforcing, indeed, to see a cognitive analysis do what other, more nearly "black box" analyses had failed to accomplish.

Quantitative judgment is both commonplace and fundamental; it underlies our most basic thoughts concerning the quantitative relationships among the complex objects and situations in our social and physical environments. At the present time a great deal is known about the content and algebraic structure of judgment data. Little is known, however, about the cognitive mechanisms that generate such data. I hope the present research will move us a little closer to a procedural understanding of the human judgment system.

References

Anderson, N. H. Application of a linear-serial model to a personality-impression task using serial presentation. Journal of Personality and Social Psychology, 1968, 10, 354-362.

Anderson, N. H. Averaging model analysis of set size effect in impression formation. Journal of Experimental Psychology, 1967, 75, 158-165.

Anderson, N. H. Foundations of information integration theory. New York: Academic Press, 1981.

Anderson, N. H., & Butzin, C. A. Performance = Motivation x Ability: An integration-theoretical analysis. Journal of Personality and Social Psychology, 1974, 30, 598-604.

Anderson, N. H., & Lopes, L. L. Some psycholinguistic aspects of person perception. Memory and Cognition, 1974, 2, 67-74.

Anderson, N. H., & Shanteau, J. C. Information integration in risky decision making. Journal of Experimental Psychology, 1970, 84, 441-451.

Beach, L. R., & Peterson, C. R. Subjective probabilities for unions of events. Psychonomic Science, 1966, 5, 307-308.

Beach, L. R., Wise, J. A., & Barclay, S. Sample proportions and subjective probability revisions. Organizational Behavior and Human Performance, 1970, 5, 183-190.

Birnbaum, M. H. Controversies in psychological measurement. In B. Wegener (Ed.) Social attitudes and psychophysical measurement. Hillsdale, N.J.: Lawrence Erlbaum, 1982, in press.

Dansereau, D. F., & Gregg, L. W. An information processing analysis of mental multiplication. Psychonomic Science, 1966, 6, 71-72.

Einhorn, H. J., & Hogarth, R. M. A theory of diagnostic inference: I. Imagination and the psychophysics of evidence. Technical report, University of Chicago Graduate School of Business, June 1982.

Graesser, C. A., & Anderson, N. H. Cognitive algebra of the equation: Gift size = Generosity x Income. Journal of Experimental Psychology, 1974, 103, 692-699.

Hitch, G. J. The role of short-term working memory in mental arithmetic. Cognitive Psychology, 1978, 10, 302-323.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. F. Prior expectancy and behavioral extremity as determinants of attitude attribution. Journal of Experimental Social Psychology, 1971, 7, 59-80.

Lichtenstein, S. C., & Slovic, P. Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology, 1971, 89, 46-55.

Lopes, L. L. A unified integration model for "Prior expectancy and behavioral extremity as determinants of attitude attribution." Journal of Experimental Social Psychology, 1972, 8, 156-160.

Lopes, L. L. Averaging rules and adjustment processes: The role of averaging in inference. Technical report, Wisconsin Human Information Processing Program (WHIPP 13), Madison, WI, December 1981.

Lopes, L. L. Individual strategies in goal-setting. Organizational Behavior and Human Performance, 1976, 15, 268-277.

Lopes, L. L. Model-based decision and inference in stud poker. Journal of Experimental Psychology: General, 1976, 105, 217-239.

Lopes, L. L. Nonadditivity in inference judgments. Technical report, Wisconsin Human Information Processing Program (WHIPP 16), Madison, WI, November, 1982. (b)

Lopes, L. L. Procedural debiasing. Technical report, Wisconsin Human Information Processing Program (WHIPP 15), Madison, WI, October 1982. (a)

Lopes, L. L., & Ekberg, P. H. S. Test of an ordering hypothesis in risky decision making. Acta Psychologica, 1980, 45, 161-168.

Lopes, L. L., & Johnson, M. D. Judging similarity among strings described by hierarchical trees. Acta Psychologica, 1982, 51, 13-26.

Lopes, L. L., & Oden, G. C. Comparison of two models of similarity judgment. Acta Psychologica, 1980, 46, 205-234.

Marks, D. F., & Clarkson, J. K. An explanation of conservatism in the bookbag-and-pokerchips situation. Acta Psychologica, 1972, 36, 145-160.

Marks, D. F., & Clarkson, J. K. Conservatism as non-Bayesian performance: A reply to De Swart. Acta Psychologica, 1973, 37, 55-63.

Oden, G. C. A fuzzy logical model of letter identification. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 336-352.

Shanteau, J. C. An additive model for sequential decision making. Journal of Experimental Psychology, 1970, 85, 181-191.

Shanteau, J. C. Averaging versus multiplying combination rules of inference judgment. Acta Psychologica, 1975, 39, 83-89.

Shanteau, J. C. Component processes in risky decision judgments. Journal of Experimental Psychology, 1974, 103, 680-691.

Shanteau, J. C. Descriptive versus normative models of sequential inference judgment. Journal of Experimental Psychology, 1972, 93, 63-68.

Shuford, E. H. A comparison of subjective probabilities for elementary and compound events. University of North Carolina Psychometric Laboratory, Report #20, January, 1959.

Slovic, P. Structural determinants of cue utilization in judgment: A polarization and adjustment model for the computation of weighted averages. Unpublished manuscript, 1967.

Slovic, P., & Lichtenstein, S. Relative importance of probabilities and payoffs in risk taking. Journal of Experimental Psychology Monograph, 1968, 78, 1-18.

Surber, C. F. The development of reversible operations in judgments of ability, effort, and performance. Child Development, 1980, 51, 1018-1029.

Tversky, A. Utility theory and additivity analysis of risky choices. Journal of Experimental Psychology, 1967, 75, 27-36.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Wallsten, T. S. A note on Shanteau's averaging versus multiplying combination rules of inference judgment. Acta Psychologica, 1976, 40, 325-330. (a)

Wallsten, T. S. Using conjoint-measurement models to investigate a theory about probabilistic information processing. Journal of Mathematical Psychology, 1976, 14, 144-185. (b)

Footnote

[1]There were some nonadditivities in the Bayesian tasks used in Lopes (1982b) that seemed to require something like an initial impression for proper interpretation. These, generally, were hypothesized to involve a weaker anchor value (i.e., nearer neutral) for unreliable samples than for reliable samples. This would be consistent with an initial impression hypothesis in which the unreliable samples receive less weight than reliable samples.

DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary of Defense (C^3I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217 (2 cys.)

Aviation & Aerospace Technology Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Communication & Computer Technology Programs
Code 240
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology & NeuroBiology Programs
Code 441NB
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Tactical Development & Evaluation Support Programs
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Department of the Navy

Manpower, Personnel & Training Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR K. Hull
Code 230B
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. J. Lester
ONR Detachment
495 Summer Street
Boston, MA 02210

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Department of the Navy

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D. C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Robert E. Conley
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

Dr. Robert G. Smith
Office of the Chief of Naval
  Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Dr. Andrew Rechnitzer
Office of the Chief of Naval
  Operations, OP952F
Naval Oceanography Division
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Department of the Navy

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation
  Group
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Ocean Engineering Department
Naval Ocean Systems Center
San Diego, CA 92152

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

Department of the Navy

HQS, U. S. Marine Corps
ATTN: CCA40 (Major Pennell)
Washington, D. C. 20380

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 92055

Chief, C^3 Division
Development Center
MCDEC
Quantico, VA 22134

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D. C. 20361

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 03416
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Department of the Navy

Mr. Milon Essoglou
Naval Facilities Engineering Command
R & D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

CDR Robert Biersner
Naval Medical R & D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Commander, Naval Air Force,
U. S. Pacific Fleet
ATTN: Dr. James McGrath
Naval Air Station, North Island
San Diego, CA 92135

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

<u>Department of the Navy</u>

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. S. Schiflett
Human Factors Section
Systems Engineering Test
Directorate
U. S. Naval Air Test Center
Patuxent River, MD 02670

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Anapolis Division
Annapolis, MD 21402

Mr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

<u>Department of the Navy</u>

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Office of the Chief of Naval
Operations (OP-115)
Washington, D. C. 20350

<u>Department of the Army</u>

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Dr. Edgar M. Johnson
Technical Director
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U. S. Army Human Engineeting Labs
Aberdeen Porving Ground, MD 21005

<u>Department of the Air Force</u>

U. S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D. C. 20332

AFHRL/LRS TDC
Attn: Susan Ewing
Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Department of the Air Force

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Foreign Addresses

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
CANADA

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology Establishment
Teddington, Middlesex TW11 OLN
ENGLAND

Director, Human Factors Wing
Defense & Civil Institute of Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
CANADA

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
ENGLAND

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 copies)

Dr. Craig Fields
Director, System Sciences Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. M. Montemerlo
Human Factors & Simulation Technology, RTE-6
NASA HQS
Washington, D. C. 20546

Other Government Agencies

Dr. J. Miller
Florida Institute of Oceanography
University of South Florida
St. Petersburg, FL 33701

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Human Resources Research Office
300 N. Washington Street
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Professor Howard Raiffa
Graduate School of Business Administration
Harvard University
Boston, MA 02163

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

<u>Other Organizations</u>

Dr. Ralph Dusek
Administrative Officer
Scientific Affairs Office
American Psychological Association
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Avenue, N. W.
Washington, D. C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert C. Williges
Department of Industrial Engineering
and OR
Virginia Polytechnic Institute and
State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Deborah Boehm-Davis
General Electric Company
Information Systems Programs
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

<u>Other Organizations</u>

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
410 Pine Street, S. E.
Suite 300
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, N. C. 27411

Dr. A. K. Bejczy
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91125

Other Organizations

Dr. Stanley N. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Laboratory
3716 S. Hope Street
Los Angeles, CA 90007

Other Organizations

Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Psychological Documents (3 copies)
ATTN: Dr. J. G. Darley
N 565 Elliott Hall
University of Minnescta
Minneapolis, MN 55455

DAT

FILM